CHARGED-CURRENT AND NEUTRAL-CURRENT EVENT FRACTION DETERMINATION BASED ON FIT VERTICES, TIME RESIDUALS AND PMT HIT ANGLES FOR THE SUDBURY NEUTRINO OBSERVATORY

By

Christian Winslow Nally

B. Sc. (Engineering Physics) Queen's University. Kingston, Ontario, 1993

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF PHYSICS

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

Apr 1996

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Physics
The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1Z1

Date: 1996-04-24

Abstract

The Sudbury Neutrino Observatory (SNO) aims to improve our understanding of neutrinos and energy-generating processes in the Sun. The central purpose of SNO is to compare the flux of charged-current (CC) and neutral-current (NC) events caused by solar neutrinos. The target in SNO will be 10^6 kg of D_2O .

One plan for enhancing the detection of NC events is by doping the D_2O with ^{35}Cl. This thesis describes maximum likelihood fitters created for the purpose of fitting CC events and for fitting NC events that result from neutron capture by a ^{35}Cl nucleus. For each fit event, likelihood ratios are extracted from these fitters to aid in determining the fraction of each type of event in a data set of unknown mixture.

It is concluded that using both timing and angle information marginally improves the ability to discriminate between CC and NC events, compared with using angle information alone. This is seen in the reduced estimated error for a technique that determines the CC event fraction in a 50/50 mixed set of CC and NC events. For prototype sets with 2000 events each, and a mixed set with 950 of each event type, an event fraction determination based on only angle information produces a fraction estimate of 0.494 ± 0.037. With identically sized data sets, an event fraction determination that includes both timing and hit angle information produces and estimate of 0.493 ± 0.034.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I am especially grateful for the support of my supervisor, Prof. Chris Waltham. I would like to thank Robert Komar for helping with the neural network calculations and very helpful discussions about the rest of the work. Thanks go to Werner Keil for providing a well run LAN, and help with various UNIX 'tidbits'. My apologies go to the users of the UBC nuclear group LAN for my long running, memory hogging process runs. Thanks go to Peter Skensved of Queen's University for the original idea for this project and Steve Brice of Oxford for Figures and helpful discussions about neural networks. Thanks also go to Tom Radcliff of Queen's University, and Joshua Klein of the University of Pennsylvania for valuable conversations.

Chapter 1

Introduction

1.1 Context

This thesis is concerned with an aspect of data analysis for the Sudbury Neutrino Observatory (SNO), currently being built at the 6800 ft level of the Creighton mine in Sudbury, Ontario. The purpose of SNO is to point to the solution of a discrepancy called the 'Solar Neutrino Problem'. This thesis is a contribution to the growing set of studies on analysis techniques that will aid the understanding of the data SNO will produce. The goal of this thesis is to investigate the ability of several parameters to aid in determining the fraction of CC and NC events in a data set with an unknown mixture of the two.

1.2 The Standard Solar Model

The standard solar model (SSM) [Bahcall95] is a model of the sun's physical properties. There are four assumptions that serve as building blocks of the SSM.

- The sun is in hydrostatic equilibrium, with particle and radiation pressures balancing gravitational collapse.
- Energy is transported between layers in the sun by photons and convection of stellar matter.

- Nuclear reactions are driving energy production in the sun.
- The sun started as a homogeneous mixture of gases.

The SSM is built upon these assumptions, and uses as input measurable parameters such as solar luminosity and nuclear cross sections. It includes physical theories for radiation transport and convection of stellar material. Composition changes in layers of the sun stable to convection are primarily due to nuclear reactions, but recent versions of the SSM also account for diffusion processes.

1.3 The Solar Neutrino Problem

Neutrinos are neutral fundamental subatomic particles. They are produced in the sun in connection with the energy generating fusion processes, and in β-decays of fusion products. There are thought to be three kinds of neutrino, one for each of the three generations of subatomic particles, but the current theory of particle physics only allows electron type neutrinos (ν_e) to be created in the sun. Of the four fundamental forces of nature, neutrinos are only known to interact through the weak force, and therefore can pass through large amounts of matter without interacting.

The SSM predicts a value for the flux of neutrinos from the sun. All measurements of the solar neutrino flux have indicated that it is lower than the expected value by a factor of two or three. This discrepancy is called the 'Solar Neutrino Problem' (SNP) [Bahcall89].

There are two general possibilities for a solution to the SNP. On one hand, there might be something fundamentally wrong with the SSM. Otherwise, there might be some new neutrino physics to be discovered. The Sudbury Neutrino Observatory aims to solve the SNP.

Experiment	Flux (SNU)	SSM Prediction (SNU)	Reaction	Threshold (MeV)
Homestake	2.56 ± 0.21	$9.3^{+1.2}_{-1.4}$	$\nu_e + {}^{37}\mathrm{Cl} \rightarrow e^- + {}^{37}\mathrm{Ar}$	0.8
Kamiokande	2.8 ± 0.45 *	6.62 *	$\nu_e + e^- \rightarrow \nu'_e + e^{-\prime}$	7.5
SAGE	69 ± 12	137^{+8}_{-7}	$\nu_e + {}^{71}\mathrm{Ga} \rightarrow e^- + {}^{71}\mathrm{Ge}$	0.23
Gallex	77 ± 10	137^{+8}_{-7}	$\nu_e + {}^{71}\mathrm{Ga} \rightarrow e^- + {}^{71}\mathrm{Ge}$	0.23

Table 1.1: Measurements of the Solar Neutrino Flux. 1 SNU = 10^{-36} interactions per target atom per second. * $\rightarrow$ The Kamiokande flux measurement is given in units of 10^6 $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ rather than in SNU, which would be threshold dependent.

Among the possibilities for new neutrino physics is the possibility of neutrino oscillations. A neutrino would 'oscillate' if it switched from one type into another. A particularly attractive solution to the SNP is the possibility of matter enhanced neutrino oscillations. Dubbed the MSW effect, in honour of its inventors Mikheyev, Smirnov [Mikheyev] and Wolfenstein [Wolfenstein], this effect depends critically on the local electron density. Possible observable effects of the MSW solution include a ν_e energy spectrum that is distorted from the expected spectrum, a significant flux of ν_μ and ν_τ, and day-night variations in the neutrino flux caused by oscillations in the earth.

1.4 Previous Experiments

Four measurements of the solar neutrino flux have been made and all show a flux less than that expected from the SSM. All of these experiments are predominantly sensitive to ν_e and have not made any conclusions about the solar flux of ν_μ and/or ν_τ. Figure 1.1 shows the fluxes of neutrinos from various sources in the sun, according to the SSM. Notice that the flux of low energy neutrinos is very much higher than that of the higher energy ones. This is due to the fact that the higher energy neutrinos are produced in reactions with low branching ratios. The experimental results are tabulated in Table 1.1.

The Homestake experiment [Wildenhain] has been in operation since the 1967, and

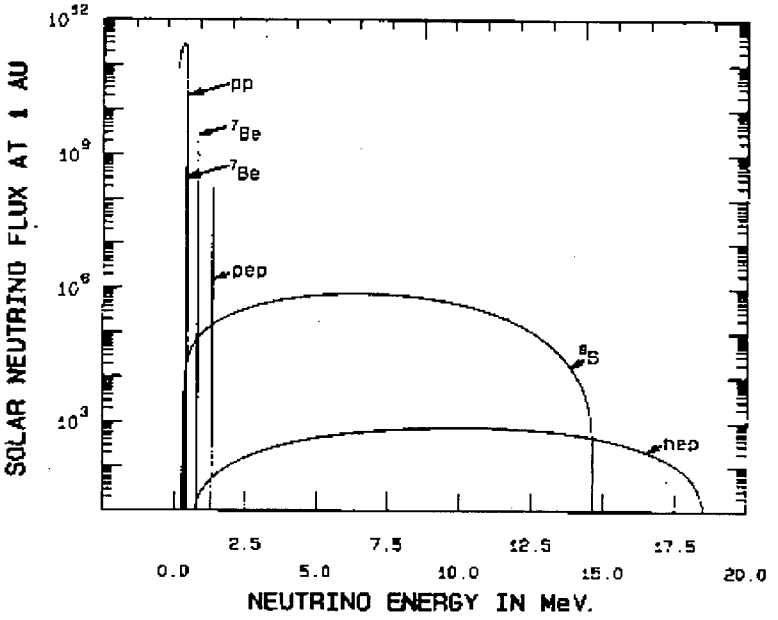

Figure 1.1: The neutrino spectra resulting from the main neutrino generating reactions as predicted by the standard solar model.[Bahcall95] The SNO detector will be sensitive only to neutrinos above some as of yet undetermined threshold near 5 MeV. The flux unit at Earth for the continuous spectra is $cm^{-2}s^{-1}MeV^{-1}$, and for the line spectra $cm^{-2}s^{-1}$.

has undergone almost three decades of careful scrutiny. The Kamiokande experiment [Takeuchi] is a real-time water Cerenkov detector, which allows the time of each neutrino event to be recorded. Information relevant to the direction of each event is gathered as well. The direction information demonstrates that the sun is the actual source of the neutrinos. The SAGE [Gavrin] and GALLEX [Anselmann] experiments confirm that the flux is lower than expected, and are noteworthy because of their low threshold. According to the SSM, the majority of solar neutrinos have an energy lower than 0.42 MeV, and the fact that these two experiments show a low flux rules out the possibility that it is only the higher energy neutrinos that have a flux lower than expected.

1.5 The SNO Detector

The SNO detector will measure the flux of neutrinos from the sun, and produce information designed to point toward a solution to the SNP. The detector, shown in Figure 1.2, will comprise of an acrylic vessel 12.0 m in diameter that will hold ultra pure heavy water (D_2O). There is a geodesic sphere of photomultiplier tubes (PMTs) looking inward on the acrylic vessel. These tubes will detect photons produced by neutrinos reacting in the D_2O.

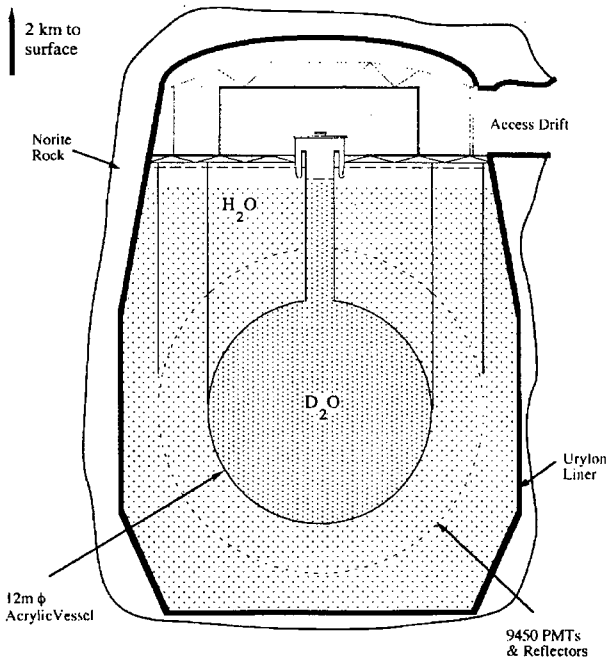

Figure 1.2: Detector Cross Section

Radioactivity in detector materials mimic the neutrino interactions in SNO and pose a problem for the measurement of the solar neutrino flux. Ultra pure light water will fill the cavity outside the acrylic vessel. This acts as a shielding from radioactivity in the

PMT materials. The detector materials themselves have been scrutinized for cleanliness.

1.6 The Physics of the SNO

The SNO detector essentially detects light from fast electrons, where *fast* means faster than the local speed of light in water. This threshold corresponds to the electron having a kinetic energy above 262 keV. The resulting Cerenkov radiation is the lightwave analogy to a sonic boom. Its generation is shown schematically in Figure 1.3. All of the reactions in SNO that are detected by the PMTs end in the creation of electrons that emit Cerenkov radiation.

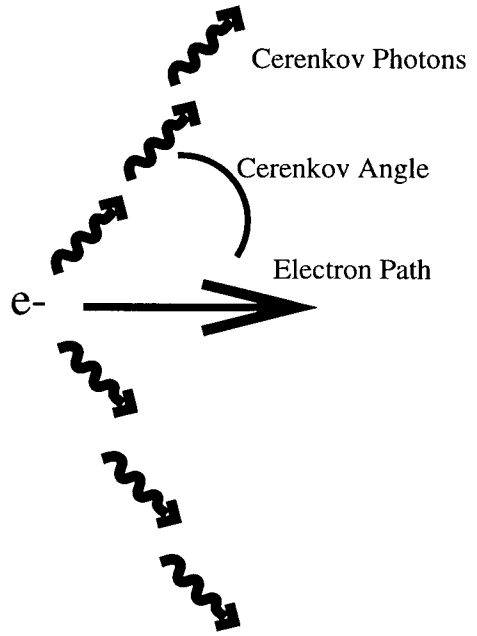

Figure 1.3: Schematic of Cerenkov Production

Using D_2O as a target, instead of relying on the electrons in normal water, provides SNO with some distinct advantages. The cross section for ν interaction is higher due to a

higher centre of mass energy for a nucleon-neutrino pair. D_2O provides higher sensitivity to ν_μ and ν_τ through a reaction that involves the breakup of the deuteron.

There are three reactions of primary interest in SNO.

$$\nu_x + e^- \rightarrow \nu_x + e^- \quad \text{ES}$$
$$\nu_e + d \rightarrow p + p + e^- \quad \text{CC}$$
$$\nu_x + d \rightarrow \nu_x + n + p \quad \text{NC}$$

The first is elastic scattering (ES), where a neutrino collides with an electron in the water. All three types of neutrino scatter from electrons through a neutral current interaction, but ν_e also scatters from electrons through a charged current interaction. Thus the cross section is 6 to 7 times higher for ν_e than for the other two neutrino types. Measurements of the rate of this reaction have not produced direct evidence for neutrino oscillations.

The second reaction, the charged current (CC) reaction, is unique to SNO. It involves the inverse β decay of the deuterium nucleus. A fast electron is ejected, emitting Cerenkov photons. The energy of this electron is correlated with the energy of the incoming neutrino, and this allows the measurement of the CC energy spectrum.

The third reaction, also unique to SNO, is called the neutral current (NC) reaction. It involves the dissociation of the deuteron into a proton and a neutron. This reaction is independent of neutrino type, and SNO will use this reaction as a measurement of the total flux of high energy solar neutrinos. There are two main schemes for the detection of the neutron ejected from this third reaction. In one scheme, ^{3}He neutron counters (NCDs) are placed in the detector. In the other scheme, Cl ions are introduced into the heavy water; ^{35}Cl, which comprises 76 % of natural Cl, is a good neutron detector. When a neutron is captured by a ^{35}Cl nucleus, gamma rays are emitted, as shown in the decay scheme in Appendix B. These gamma rays will subsequently transfer their energy

to electrons, which in turn emit Cerenkov radiation, thus making them detectable by the PMTs in SNO. In the remainder of this thesis, neutral-current events as detected by neutron capture on ^{35}Cl are referred to with the notation NC.

An event in SNO is a set of PMT hits close together in time. A 5 MeV electron will typically result in around 50 PMT hits. The time of each PMT hit is recorded. Often, most of the hits in the event are associated with a single electron. Some of the photons that cause a hit might have scattered on the way to the PMT. Some of the hits might be associated with random PMT firings, called noise hits. A typical event hit pattern is shown in Figure 1.4.

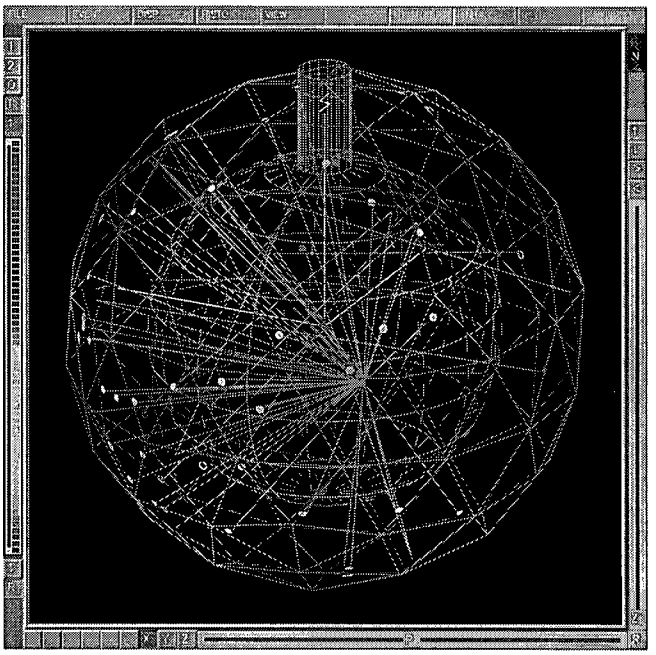

Figure 1.4: Sample PMT Hit Pattern for a single electron created by the SNOMAN Monte Carlo. The electron kinetic energy is 5.0 MeV

This thesis is concerned with discriminating CC events from NC events.

1.7 Comparison of CC and NC Events

For a single e^-, two effects tend to prevent the PMT hits from forming a clean Cerenkov ring. First, the direction of the electron will wander as it slows due to multiple scattering. Second, the angle of Cerenkov emission is a function of e^- energy such that $\cos\theta = \frac{1}{\beta n}$ where θ is the emission angle from the direction of propagation, β is v/c for the e^- and n is the index of refraction. Thus as the e^- slows, the emission angle closes from the $\beta = 1$ limit of 41.4° to 0.0°. These effects obscure the direction resolution to the point where the error in the reconstructed direction as computed by the mean direction to the PMT hits is approximately 25°.

Charged current events are single e^- events with a known direction distribution from the sun.

The essential difference between CC and NC events from the PMTs point of view is the presence of one or more gammas emitted from a neutron capture on the ^{35}Cl nucleus (see the decay scheme in Appendix B). A gamma ray will result in more than one source for Cerenkov radiation if it Compton scatters multiple times. The presence of more than one gamma will compound this effect. This means that for NC events, there will sometimes be multiple sources of Cerenkov radiation. CC events and NC events will also have different energy spectra.

1.8 Searching for a Characterizing Parameter

The SNO detector aims to point to a solution of the SNP by measuring both the flux and spectrum of electron type neutrinos (through the CC reaction) and the total flux of

neutrinos (through the NC reaction). If these fluxes are the same, then the solar neutrinos are electron neutrinos, and the solution of the SNP probably involves modification of the SSM. If the total flux of neutrinos is significantly higher than the electron neutrino flux, then there are a significant number of muon and/or tau type neutrinos and the solution of the SNP would be that neutrinos are mixing among generations after being generated.

In the worst case, we will not know whether a given event is CC or NC at all. In this case, we will be forced to estimate the fluxes of these reactions by performing two separate data runs. One with just pure D_2O, the other with the NC detection enhancing ^{35}Cl. The N_{hits} spectrum for various sources of event is shown in Figures 1.5 and 1.6. Events labeled PMT $\beta - \gamma$ come from radioactivity in the PMTs. Internal $\beta - \gamma$ events are due to radioactivity in the D_2O. Events labeled 'PSUP + Cavity' come from radioactivity in the PMT support structure or from the rest of the SNO cavity. The first is for the pure D_2O case, while the second is for the case where 2500.0 kg of NaCl has been added.

In the ideal (and very unrealistic) case, the exact nature of each event would be known as it occurs. In this case, only a single data run with ^{35}Cl would be required, and independent energy spectra for each event class could be created.

This thesis is concerned with finding event characterizing parameters based on the ratio of a NC likelihood and a CC likelihood. These likelihoods will be calculated using probability densities for PMT timing, and PMT hit angles. The characterization ability of the parameters will be tested by using each parameter to estimate the event fraction in a set with a known mixture of CC and NC events.

Event fraction determination is done by producing distributions of an event characterizing parameter. In the ideal case described above, for CC events this parameter would have values very different from its values for NC events. To determine whether the event was CC or NC, one would only have to evaluate this parameter. In a more realistic case, this parameter will have overlapping distributions for the two event types and

event fraction determination proceeds via statistical analysis with a fit to two prototype distributions.

This thesis ignores the fact that there will be ES events produced by SNO, and also ignores any background events that mimic the CC and NC signals. The analysis presented here would have to be applied to data sets where these event types were already removed, or the analysis would need to be extended to the case where more than two classes of event are present.

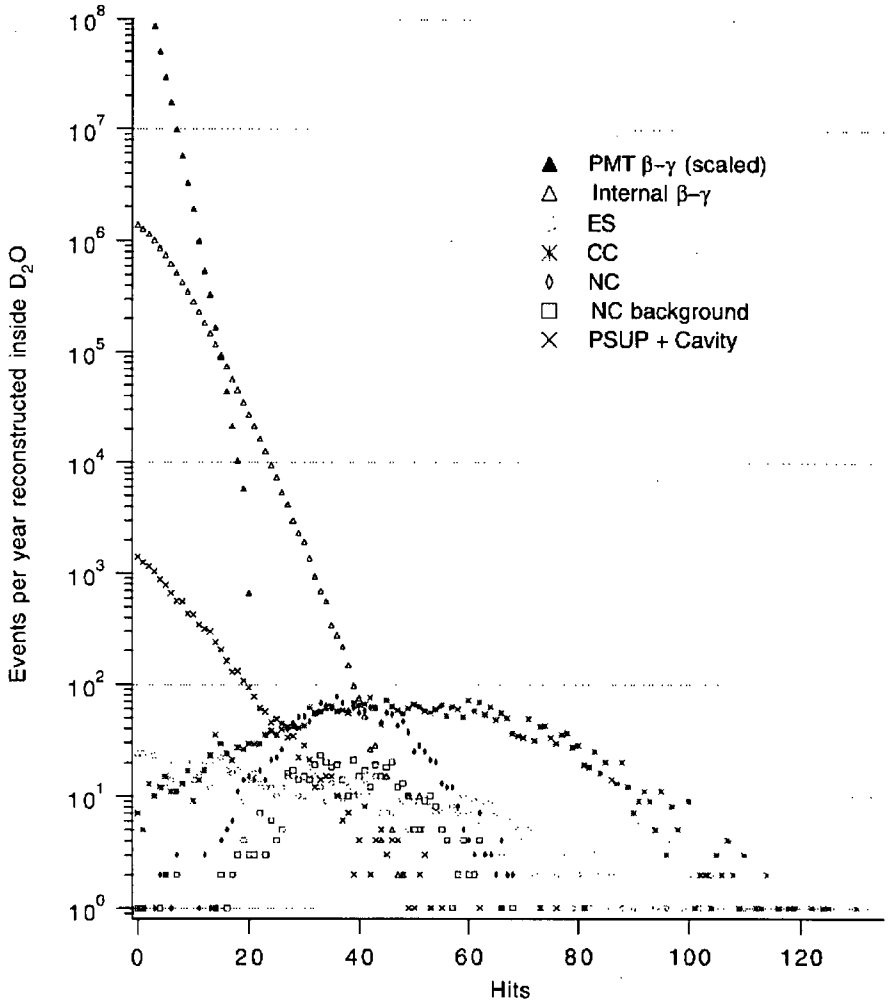

Figure 1.5: N_{hits} spectrum for various event sources for a pure D_2O data run. Fluxes are scaled to 1/3 of the SSM.

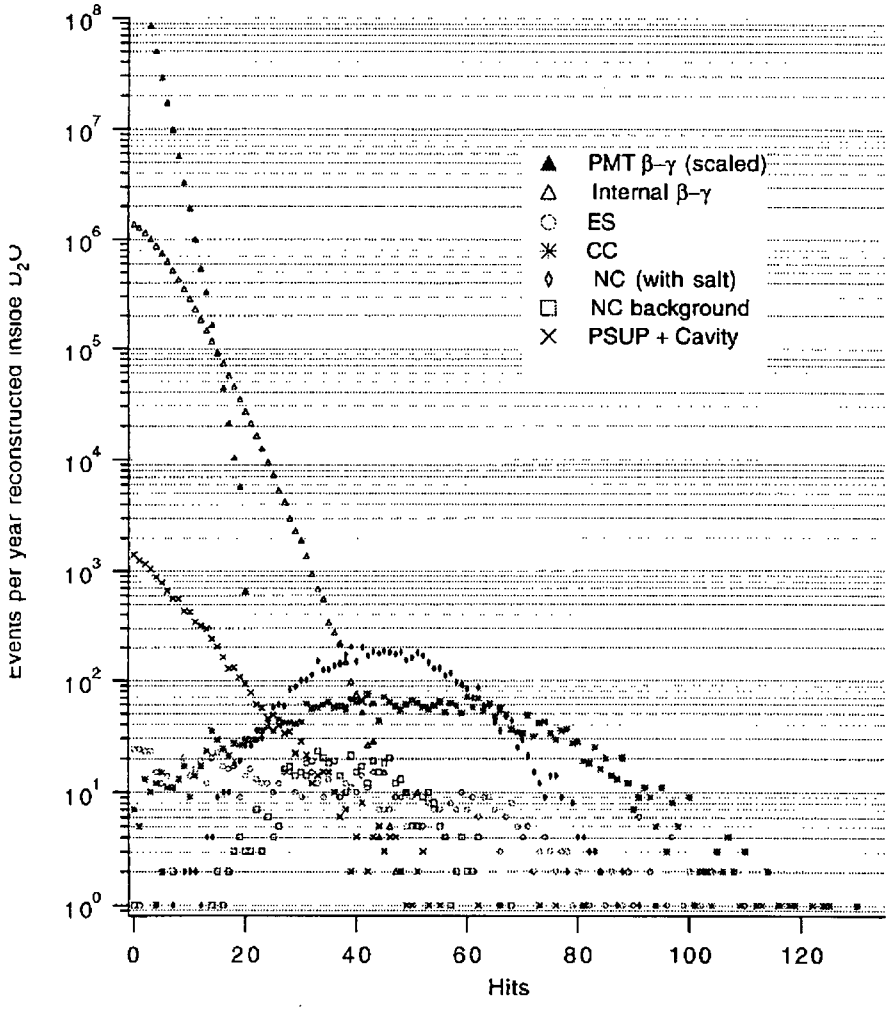

Figure 1.6: N_{hits} spectrum for various event sources for a D_2O data run with NaCl added. Fluxes are scaled to 1/3 of the SSM.

Chapter 2

Mean Opening Angle Classification

This chapter describes a simple parameter that can be used for event type characterization. Results are given for ratio determination using this parameter so that it may be held up as a benchmark for parameters described in later chapters.

2.1 Average Opening Angle

A very simple parameter that distinguishes CC events from NC events is the mean opening angle $\overline{\theta}$ of the PMT hits in an event from the average direction of the hits from a fit vertex position. It is relatively easy to see the physical reason why this parameter distinguishes between CC and NC events. NC events will more often have more than one source of light for PMT hits, and when this is the case, the PMT hits will be distributed in a more isotropic fashion. The $\overline{\theta}$ parameter is simply a measure of the degree to which the distribution of PMT hits is isotropic. Histograms of $\overline{\theta}$ for sets of CC and NC events are shown in Figure 2.1.

2.2 Results of a χ^2 Fit

Given a parameter that provides some ability to distinguish between the two event classes, one is able to estimate the fraction of events of each type in a mixed set by performing a

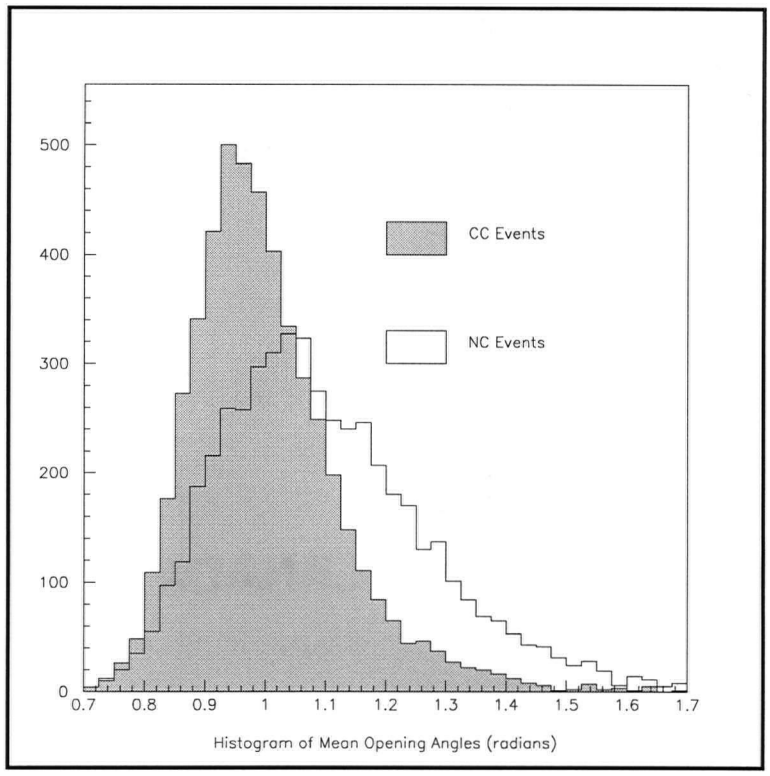

Figure 2.1: Opening Angle ($\overline{\theta}$) Histograms for CC and NC events.

fit of the distribution of the parameter for that mixed set to a sum of two distributions from prototype sets of the two event classes. This method, to be described further in section 3.1, was used in conjunction with $\overline{\theta}$ to determine the most probable number of CC events and NC events in a mixed set with 950 CC events and 950 NC events. Prototype distribution sets were created that included 2000 CC events and 2000 NC events. The $\overline{\theta}$ distribution of the mixed set was fit to a mixture of the two prototype sets with the CC fraction used as the free parameter. This method produced an estimated CC event fraction of 0.521 ± 0.047. The given error corresponds to 1.0 σ statistical error; systematic errors have not been investigated. The remaining chapters of this thesis set out to find a parameter that provides a better basis for determining the fractions of CC events in a

dataset. The remainder of this chapter describes how the $\overline{\theta}$ parameter was discovered as being the important distinguishing parameter among several others.

2.3 Introduction to Neural Networks

The importance of the $\overline{\theta}$ parameter in relation to several others was initially investigated using a Feedforward Neural Net (FNN) classifier. The neural net simulation code that comes with [Masters] was used for the investigations described here. A high quality, free alternative is the Stuttgart Neural Network Simulator (SNNS) [SNNS].

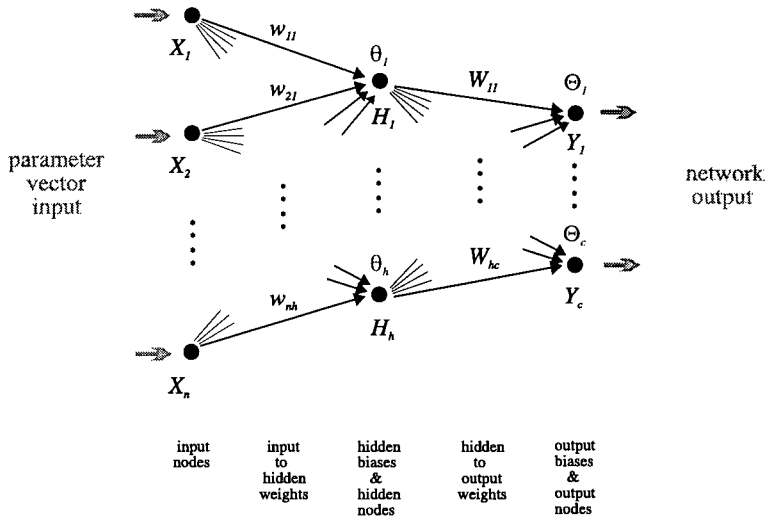

Figure 2.2: Diagram of a Feedforward Neural Net. Reprinted with permission from [Brice95]

The diagram of a FNN is a graphical way of understanding how the FNN algorithm computes its output. An example FNN with one hidden layer is shown in Figure 2.2. This diagram depicts input neurons, hidden neurons and output neurons. Each neuron in the input layer is connected to each hidden neuron, which is, in turn connected to each output neuron. A calculation of the FNN's output is carried out as follows. The

input values are given to the input neurons. The value of the input neuron is multiplied by a value associated with the link between it and a hidden neuron, called the 'weight' of that link. For each hidden neuron, the products of the input neuron values and their associated weights are summed. This is effectively the scalar product of the input vector and the weight vector for a given hidden neuron. The hidden neuron feeds its input through a non-linear function called an activation function. The result of this function is considered to be the hidden neuron's output. A similar process of taking the scalar product of neuron values with weight vectors occurs between the hidden layer and the output layer.

This process is described by the following set of equations. The output of a particular hidden neuron is calculated as

$$H = A(\sum_{i=1}^{n} X_i w_i + \theta)$$

where the sum over i is over each input neuron, n is the number of input neurons, X_i is the value input from a neuron in the input layer, w_i is the weight vector associated with the link to that neuron, θ is the bias value associated with that hidden neuron, and A is the activation function for the neuron. The activation function is a 'sigmoid', that is to say 's-shaped'. The performance of a neural net is often not sensitive to the exact functional form of the activation function. For the trials described here an interpolated logistic function $A(x) = \frac{1}{1+e^{-x}}$ was used. The logistic function is plotted in Figure 2.3.

2.4 Training a FNN to Classify SNO Events

A FNN can be viewed as a function approximator. The function being approximated is from from n to c dimensions, where n is the number of input neurons and c is the number of output neurons. The function is learned through a training procedure based

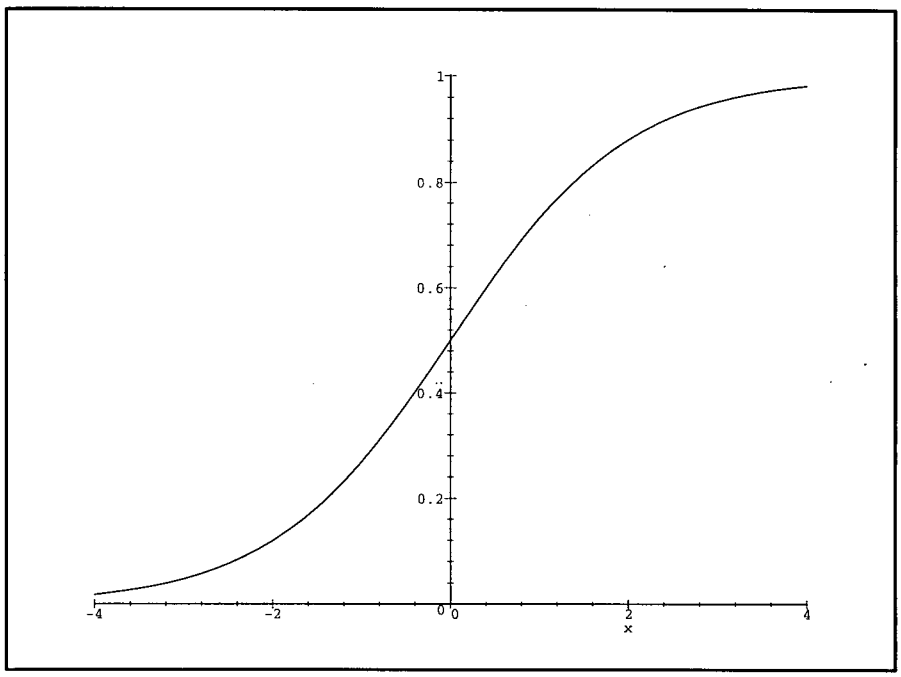

Figure 2.3: The Neural Networks Activation Function

on sets of prototype data. Training occurs by varying the weights and biases until the net approximates the function represented by the training data. A FNN is trained to classify by creating a training set for each class that the network is intended to recognize. Each input vector in the training set is associated with one event such that its components are either parameters that describe an event, such as N_{hits}, or are calculated from the PMT Hit pattern, for example, an error estimate in the fit vertex. Each one of these input vectors is associated with a desired output vector. For the case of distinguishing between two event classes, only one output neuron is required. The FNN will be trained to have a high output value for one event class, and a low output value for the other class.

The goal of training is to minimize the mean square error over a complete data set,

defined as

$$M = 1/n \sum_{e=1}^{n} \sum_{i=1}^{c} (T_{ie} - Y_{ie})^2 \tag{2.1}$$

where e runs over each event in the training set, n is the number of events, i runs over the output neurons, c is the number of output neurons, T_ie is the target value for the ith output neuron for event e, and Y_ie is the observed activation for the ith output neuron for event e. The reason for training the net in this way is that the derivative of M with respect to the weights is easily computed, and this provides a way of updating the weights in such a way as to make them produce output closer to the desired value for that class.

The ability of the network to approximate a given function is constrained by the number of neurons and the number of hidden layers it has. This is similar to fitting many data points to a polynomial. One should be careful to have many more data points in the training set than free parameters in the net. This was not a problem for the investigations presented here because there were several thousand events in the training sets used, and only one or two hidden layers with few a few hidden neurons in each.

2.5 Training Data

Two data sets, produced with SNOMAN, were used for the results reported in this Chapter; different prototype sets are developed for later work. For all sets, event interactions were distributed uniformly throughout the D_2O.

One data set was a combination of various single e^- Monte Carlo runs with various electron energies resulting in a spectrum of N_{hits} similar to that for NC. This was done to minimize the FNNs ability to distinguish between the two event types based on the N_{hits} parameter alone. It was hoped that by doing this, the FNN might develop sensitivity to differences between the two event classes in the form of an N_{hits} dependence of one of the

E_e (MeV)	number out of 10000
4.0	2375
5.0	2250
6.0	2500
7.0	2375
8.0	500

Table 2.1: Electron energies in the single e^- event set. The events from these discrete energy electrons were added together to form an event set with an N_{hits} spectrum spectrum close to that of the NC events.

E_γ (MeV)	number out of 10000
8.578	291
7.790 and 0.788	855
7.414 and 1.165	1042
6.978 and 1.601	226
6.628 and 1.951	302
6.628 and 1.164 and 0.787	152
6.620 and 1.959	801
6.111 and 1.951 and 0.517	1332
6.111 and 1.164 and 0.787 and 0.517	668
5.715 and 2.863	664
1.0 and 2.0 and 2.0 and 3.578	3667

Table 2.2: Various decays included as a simulation of the gammas resulting from neutron capture on ^{35}Cl. The last entry is included to simulate the many decays that were not extracted from the decay scheme.

other input parameters. Each single e^- was started in a random direction. The contents of the data set are shown in Table 2.1.

The other data set was a 'hand made' simulation of NC events, in the sense that it was constructed by combining the results of gamma ray Monte Carlo runs. The contents of this combined set are shown in Table 2.2. This set was created before the physics for neutron capture on ^{35}Cl was built into the SNOMAN Monte Carlo. Angular correlations between the gamma rays in the ^{36}Cl were not taken into account for the data set described here. Each gamma ray was released in a random direction.

N_{hits}	The number of PMT hits for a given event
ΔX	The uncertainty in the position of the fit vertex
ΔT	The uncertainty in the time of the fit vertex
$\overline{\theta}$	The mean of the angle between the direction to each hit and the fit direction
σ_θ	The standard deviation in θ
σ_{time}	The standard deviation in the PMT hit times expected from the hit
χ^2_R	SNOMAN Fitter Reduced χ^2

Table 2.3: Parameters used for event discrimination with the FNN.

2.6 Input Parameters

The input parameters were extracted and/or calculated from the SNOMAN time fitter, a subroutine that returns a fit vertex and direction based on PMT times. The complete list of parameters considered is in Table 2.3. Each parameter was rescaled between 0.0 and 1.0 as a requirement for the use of the neural net.

2.7 Classification Performance

To measure classification performance, a sample from the set of NC events was fed into the classifier. Each event was labeled as being 'NC-like', or 'single e^--like' depending on whether the output neuron value was above or below 0.5. Then a set of the single e^- events were classified in a similar manner. The performance of the classifier was measured by recording the percentage of events that were classified correctly. The results are summarized in Table 2.4. Histograms of the FNN output are shown in Figure 2.4.

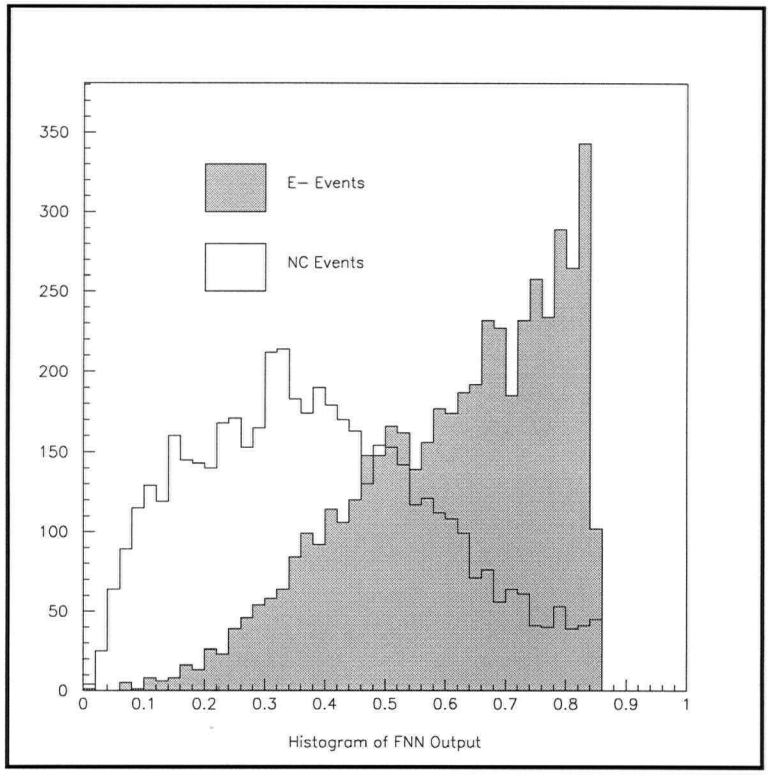

Figure 2.4: Histograms of one of the FNN output neurons. These histograms are for the FNN with all 7 input parameters from . The sum of the output neuron activations was 1.0, so the other output neuron contains no extra information.

2.8 Verification Using PDFs

To verify the results of these investigations, a separate test based on the estimation of the probability density functions (PDFs) of the input parameters was done. Estimation of PDFs is described in Section 4.4. This classifier worked by evaluating its input parameters for each event, and comparing the values of the PDF for each class at that point. The event was deemed to be of the class with the higher PDF value.

When a PDF comparison was done for both $\overline{\theta}$ and $\sigma_{\overline{\theta}}$, the performance matched that of the initial FNN. A PDF comparison using only the $\overline{\theta}$ parameter was performed and

Results	% NC correct	% single e^- correct
FNN all parameters	71%	74%
FNN $\overline{\theta}\sigma_\theta$	67%	76%
PDF $\overline{\theta}\sigma_\theta$	69%	75%
PDF $\overline{\theta}$	64%	75%

Table 2.4: Summary of Results. The similar performance for the two FNN runs shows that the majority of the classification ability is inherent in the $\overline{\theta}$ and σ_θ parameters. Error estimates on these percentages are approximately 0.6%

produced results that were only slightly worse than those from the initial FNN. This suggested that $\overline{\theta}$ is a parameter that provides much of the ability to distinguish between these event types. It was concluded that the FNN was not finding correlations in the other parameters. Histograms for the other parameters appeared, for the most part, to overlap for both classes, so none of those parameters would provide much in the way of classification ability on their own.

Chapter 3

Event Fraction Analysis

This chapter outlines the method used throughout this thesis to perform event ratio determination. It requires a single parameter that has lower values for one event type, and higher values for the other event type. The method is essentially a χ^2 fit of the distribution of the parameter for a mixed set to a sum of the distributions of the parameter for CC and NC events. The one free parameter in the fit is taken to be the CC fraction. The method was adopted from standard statistics by Steve Brice [Brice96].

3.1 Event Fraction Determination

The method used for event fraction determination requires a parameter that can be evaluated for each event. One example of this sort of parameter is the mean opening angle $\overline{\theta}$ discussed previously in chapter 2. Further parameters will be developed in chapters 4 and 6.

The technique implements a fit of the distribution of the parameter for the mixed set to a sum of the two distributions of the same parameter for CC and NC prototype sets.

The quantities of concern in this method are listed in table 3.1. The index $k = 1$ refers to CC events and $k = 2$ refers to NC events. The method is implemented under the constraint that $\alpha^1 + \alpha^2 = 1$, so that there is only one free parameter in the fit. There

Quantity	Description
N	The Number of Events in the Mixed Set
U_i	Bin Contents for the Mixed Set (Unnormalized)
u_i	Bin Contents for the Mixed Set (Normalized)
N^k	Number of Prototype Events in Class k
F_i^k	Prototype Histogram for Class k (Unnormalized)
f_i^k	Prototype Histogram for Class k (Normalized)
A^k	Number of Events in Prototype Set k
α^k	Fraction of Events in Class k
χ^2	The Goodness of Fit Criterion

Table 3.1: Quantities Used in the χ^2 Fit. $k = 1$ refers to CC events, $k = 2$ refers to NC events.

is a contribution to the error estimate in each bin from both the prototype sets and the mixed set.

If the fit were perfect, the bin contents of the mixed set (U_i) would be exactly equal to the contents of the sum of the two prototype distributions ($\sum_k A^k f_i^k$). There will be some expected error in this fit for each bin, related to how many events were accumulated in the mixed set and in each of the prototype sets. These independent errors are added in quadrature, where $\sigma_{U_i}^2$ is the error estimate for the bin contents of the mixed set, and $(A^k \sigma_{f_i^k})^2$ is the error estimate in the scaled bin contents for the kth prototype set.

For the purposes of this thesis, prototype distributions always contain 2000 events each, and the mixed set is always made up of 950 CC events and 950 NC events. In reality, the prototype sets might be produced by a tuned Monte Carlo code and so may have many more events than this. The main thrust of this thesis is to compare the ability of each parameter in distinguishing CC from NC. This is done by comparing the estimated statistical error for an event fraction determination using each parameter. The parameter that produces the lowest statistical error is taken to be the best parameter.

The χ^2 function can then be written as

$$\chi^2 = \sum_i \left[\frac{U_i - \sum_k A^k f_i^k}{\sigma_i} \right]^2 \tag{3.1}$$

where

$$\begin{aligned} \sigma_i^2 &= \sigma_{U_i}^2 + \sum_k (A^k \sigma_{f_i^k})^2 \\ &= U_i + \sum_k \tfrac{A^k}{N^k} A^k f_i^k \\ &= N \left[u_i + \sum_k (\tfrac{N}{N^k} \alpha^k) \alpha^k f_i^k \right] \end{aligned}$$

Or, written entirely in terms of normalized histograms

$$\chi^2 = N \sum_i \frac{\left[u_i - \sum_k \alpha^k f_i^k \right]^2}{\left[u_i + \sum_k (\tfrac{N}{N^k} \alpha^k) \alpha^k f_i^k \right]}$$

3.2 Event Fraction Error Estimation

For the purposes of error estimation the χ^2 was simply tabulated with a 0.001 interval from 0.0 to 1.0. The best fit was selected by looking in the resulting list for the smallest χ^2, and error estimates were produced by finding the values of the parameter that corresponded to a 1.0 increase in the χ^2. This technique was used to arrive at the 0.521 ± 0.047 CC event fraction estimate presented in the previous chapter for $\overline{\theta}$ event fraction determination on a 50/50 mixture.

Chapter 4

Likelihood Ratio Characterization

This chapter describes a characterization parameter formed by taking the ratio of two likelihoods. Likelihoods are calculated using various PDFs, and results are given for the event fraction determination ability of three parameters based on vertices returned by the SNOMAN Time Fitter.

4.1 A Type of Classification Parameter

A likelihood is a measure of how probable obtaining a given set of data is, assuming that it was caused by physics well represented by a certain model. A likelihood is calculated using a PDF in the following way:

$$L_{event}(x, y, z, t) = \prod_{i=1}^{N_{hits}} P_i(x, y, z, t) \tag{4.1}$$

The function $P(x, y, z, t)$ is a PDF that represents the assumption about the origin of the data. A PDF is a representation of our knowledge of the character of a certain event class. One of the PDFs used in this thesis is the time residuals PDF for CC events. Another is the hit angles PDF for NC events. In this thesis, all PDFs are formed from data sets using the Parsen windowing technique, as described in Section 4.4. One can calculate the likelihood of obtaining a given set of event data, assuming that the

event is an NC event. In this case, one would take the event information and calculate the likelihood based on a PDF formed with NC events. Thinking of likelihoods in this way suggests a possible classification parameter. For a given event and an associated fit vertex, one could calculate the likelihood of obtaining the data if the event were NC, and the likelihood of obtaining the data if the event was CC. The characterizing parameter is formed by taking the ratio of the two likelihoods produced in this process. This kind of parameter will be referred to as a Likelihood Ratio (LR), and is defined as $\frac{L_{NC}}{L_{CC}}$.

4.2 Prototype Generation

In order to produce the required PDFs, two event sets were produced with SNOMAN and used as class prototypes. These event files were written to disk in the format that the actual data from SNO will be written to tape. One set was used to create PDFs for the CC class, and the other was used to create the NC class PDFs. The events in these sets were distributed uniformly throughout the D_2O. The NC events were built using SNOMAN version 2.08 which takes into account angular correlations between the emitted gamma rays. This thesis does not investigate whether it is possible to create similar PDFs using calibration data. If this is not possible, then to achieve the results presented in this thesis, one would have to use the SNOMAN Monte Carlo to produce such sets, after having tuned the SNOMAN Monte Carlo to reproduce all calibration data with sufficient precision.

4.3 Time Residual and Angle Extraction

Three different PDF definitions are used in this thesis. Starting from a vertex, one can calculate the $T_{residual}$ for any PMT hit. The time residual for a given PMT hit is defined

as:

$$T_{residual}(x, y, z, t) = T_{hit} - T_{expected} \tag{4.2}$$

where $T_{expected}$ is the time that the PMT would have been hit if light had traveled from an assumed vertex (x, y, z, t) to that PMT.

The PMT hit angle can be measured from the average direction to the PMTs from the vertex in question. SNOMAN, with additional user code, was used to extract PMT hit angles and time residuals from the Monte Carlo seed vertices. Density estimates for the $T_{residual}$ PDF, the hit angle PDF and for the joint PDF in $T_{residual}$ and hit angle were created for both CC and NC events.

4.4 Probability Density Estimation

A histogram is a representation of the distribution of a given parameter. One can observe the salient features rather easily, but one must choose both a bin width, and where to start the binning. When all is said and done, the distribution is 'blocky', and one must be satisfied that choosing a different bin width and starting point will not affect the results of the analysis. Only in the limit of high statistics and small bin widths does the histogram become smooth.

An alternative is to fit a curve to a histogram. This is reasonable when one has some idea what the distribution should look like. If there is reason to believe that the curve is exponential, then fitting an exponential to the histogram will provide a decent estimate. Very often, however, one does not know the form of the distribution, and using this method will impose restrictions on the shape of the data. This method may end up hiding unexpected but relevant features of the shape.

Another method for estimating PDFs is 'Parsen windowing' [Silverman]. It has the

advantage that it lets the data speak more for itself. The estimate is made by summing 'kernel functions', each one centred on a datapoint in the set. A typical kernel function is the normal distribution. The user of this method is left, however, with the problem of choosing the width of the kernel function, h. This parameter is often called the window width. The PDF estimate then takes the following form:

$$P = \frac{1}{nh} \sum_{i}^{n} K(\frac{x - X_i}{h})$$

where n is the number of data points used to estimate the density, x is the point at which the density is being estimated, X_i is the ith data point, and K is the kernel function.

4.5 Choosing a Window Width

There is no prescribed method for choosing a window width in the Parsen windowing scheme. It is often best to just produce various density estimates using different window widths, and choose the best one based on previous knowledge of the density. One knows absolutely nothing about the parameter for which one is producing a density estimate then one has little to go on with regard to choosing a window width. This is reasonable because if one truly knows nothing about the distribution then the best estimate might be delta functions at each of the observed points, but this kind of estimate is rarely reasonable when the parameter is in a continuous space.

The choice of the window width is mixed up with how many observations one has and how fast the density can fluctuate. For a given set of observations, a larger window width results in a smoother density estimate.

One way of getting around this deadlock is to take the window width that would be optimal if the density was a Gaussian distribution with a sample standard deviation equal to that of the observed set and then adjust this value until the estimate fits with some

previous knowledge of how the density should behave. This is different from assuming that the distribution is Gaussian as one would be doing if one fit a Gaussian distribution to a histogram. Instead, this is choosing the window width for a Parsen estimate. If the density is in fact Gaussian with the measured width, this method will produce a very good estimate of it. If there are data points that suggest some other features to the shape, they will have an effect of the estimate. The following choice of h provides us with a method that would be ideal if the density being estimated was actually Gaussian.

$$h_{opt} = (4/3)^{1/5} \sigma n^{-1/5} \tag{4.3}$$

where h_{opt} is the window width that is optimal in the Gaussian limit, n is the number of points in the sample set, and σ is the sample standard deviation.

One problem with the Parsen windowing technique is that it can be computationally intensive. For the case of a Gaussian kernel, the estimate requires the calculation of an exponential once for every point in the prototype set each time the density estimate is computed. Overall speed can be increased by computing a 'lookup' table for the PDF, and then interpolating the PDF estimate from this table.

4.6 One Dimensional PDFs

A time residual PDF was formed for both CC events and NC events using a normal kernel. These PDFs were sampled on a grid from -150.0 ns (i.e. a PMT that was 150 ns early) to 180.0 ns with 3301 samples. These PDFs are shown in Figure 4.1. One can see that the distribution is centred close to 0.0 as expected. The PDF is dominated by the 'in time' PMT hits. Late light, due to scattered and reflected photons, is shown as having a positive time residual. Early PMT hits are due to noise hits. PDFs for the angle of a PMT hit from the mean direction to the PMT hits are shown in Figure 4.2.

A total of 3141 samples were taken from 0. to π radians.

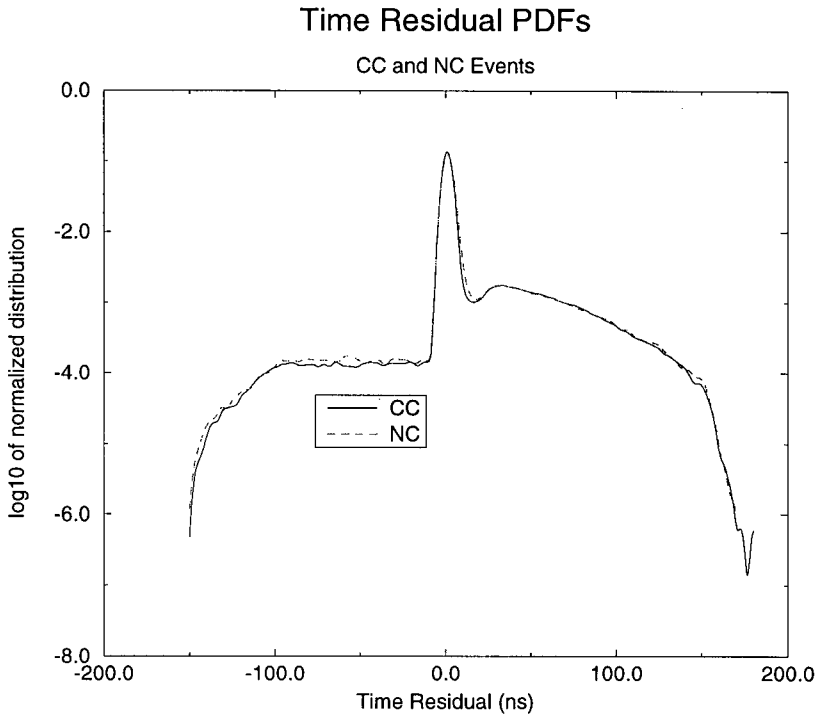

Figure 4.1: Normalized Parsen window estimate of the CC and NC time residual PDFs.

4.7 Two Dimensional PDFs

Estimating a two dimensional PDF using Parsen windowing proceeds in the same fashion as outlined above, expect that the kernel function is now two dimensional. One obvious choice for a kernel function would be a bivariate normal distribution. It is preferable to have a kernel that has the same shape as the data in an overall sense. For instance, consider the effect of a data point on the various PDF sample points. If the data is highly correlated in the two dimensions being investigated, then the effect of a data point on the PDF sample points should not be radially symmetric around the data point. This

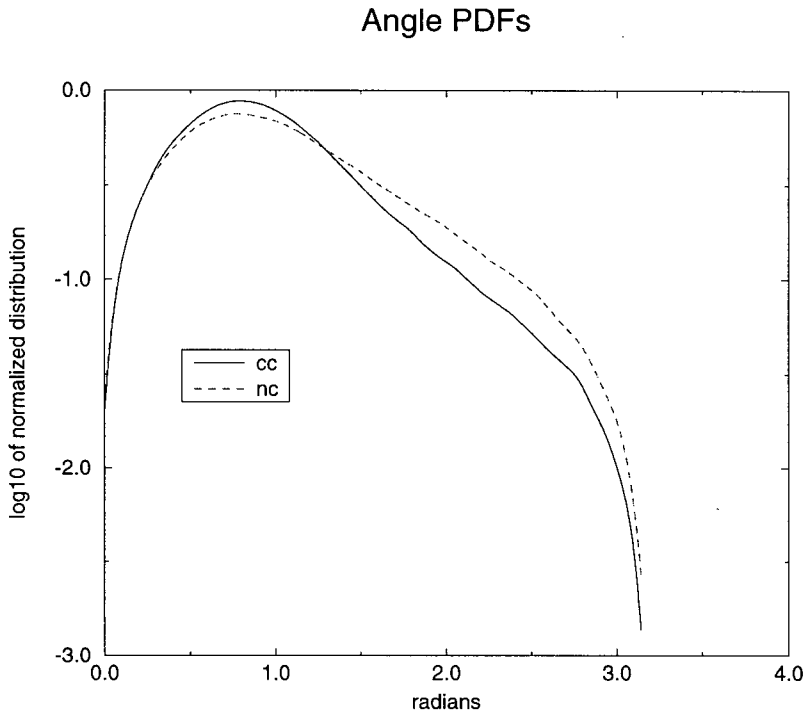

Figure 4.2: Normalized Parsen window estimates of the CC and NC Hit Angle PDFs

can be taken into account by computing the covariance matrix for the data and using it to first transform points into a coordinate set where the data has unit variance and the basis vectors for the new coordinate set lie along the principle axes of the data.

This is achieved by using a PDF estimate of the following form:

$$P = \frac{1}{nh^d\sqrt{(detS)}} \sum_{i=1}^{n} K(h^{-2}(x - X_i)^T S^{-1}(x - X_i))$$

where P is the PDF estimate, x is the PDF sample point vector, X_i is a data point, n is the number of points in the data set, h is the window width, d is the dimension of the space the PDF is in (2 in this case), and K is the kernel function. The sum is over all the points in the prototype set.

Two dimensional PDFs, in $T_{residual}$ and hit angle, were constructed for CC and NC

events using prototype sets sorted in angle and a truncated kernel of the following form:

$$K(x \cdot x) = \begin{cases} 1 - (x \cdot x)^3 & \text{for } x \cdot x < 1 \\ 0 & \text{for } x \cdot x \geq 1 \end{cases} \tag{4.4}$$

where x is the vector from the data point to the PDF sample point.

This allowed a faster computation of the PDF estimate by determining which PMT hits could possibly have any effect at all on PDF sample points with each angle sample coordinate. Also, the second derivative properties of this kernel allow the sampling interval to be larger without a significant increase in PDF estimate error. See [Silverman] for details. The sampled PDF estimates are shown in Figures 4.3 and 4.4. Initially, the window widths calculated with equation 4.3 were found to be too high for the angle components, as seen by over smoothed PDFs. Lower ones were put in by hand until the angle distribution along the time peak tended to zero at the limits of 0 and π, as it should.

4.8 Likelihood Ratio Results

Two datasets with close to 1000 events each were analysed with the SNOMAN time fitter, resulting in a single vertex for each event. For each set of PMT hits and fit vertex, a CC likelihood and an NC likelihood was calculated with each PDF. These likelihoods were used to calculate the three likelihood ratios (LRs) for $T_{residual}$ only, hit angle, or both $T_{residual}$ and hit angle.

Histograms for the likelihood ratios LR_{time}, LR_{angle}, and $LR_{time+angle}$ are shown in Figures 4.5, 4.6, and 4.7.

Event fraction analysis was done using each of LR_{time}, LR_{angle}, and $LR_{timeangle}$. Prototype histograms for each of these parameters were created with 2000 events each. In

Parameter	Extracted CC Fraction
LR_{time}	$0.189 \pm^{0.155}_{0.172}$
LR_{angle}	0.489 ± 0.042
$LR_{timeangle}$	0.497 ± 0.039
$\overline{\theta}$	0.521 ± 0.047

Table 4.1: Results for LR Event Characterization. The LR_{time} parameter clearly cannot be used to form an event characterizing parameter. This was because the histograms for CC and NC events overlap almost exactly for this parameter. These three results are compared with the performance of the $\overline{\theta}$ parameter.

each case, an attempt was made to extract the CC event fraction from a set of data that was known to have 950 CC and 950 NC. Results are summarized in Table 4.1

A method was devised of finding the theoretical limit of how well any given parameter could determine event fractions. When analysing the prototype events, the Monte Carlo vertex was used to calculate the likelihood in each PDF. If any fitter provided vertices that resulted in a better performance than using the Monte Carlo vertices of the events, then it would be because the fitter was systematically doing poorly on one of the event types. This would be a dubious method of creating a characterization parameter and it would be better to understand the systematics of the fitter, which should suggest a better kind of characterization parameter. In this light, the histogram for the $LR_{time+angle}$ parameter calculated using the Monte Carlo vertices is shown in Figure 4.8. Here it can be seen that indeed, we have come close to the limit with the current PDF definitions.

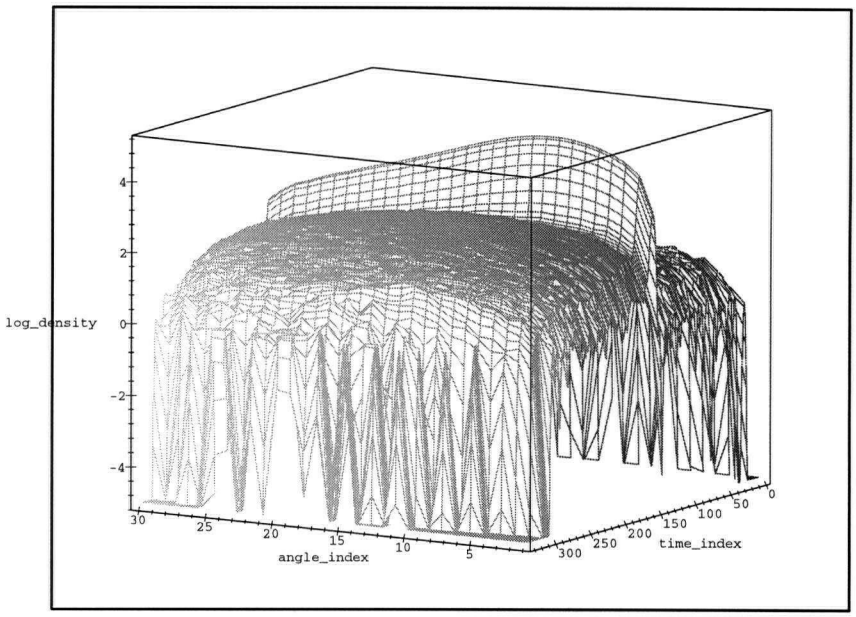

Figure 4.3: CC PDF Estimate in Both Time Residual and Hit Angle. The PDF was generated with a data set with close to 50000 events. The angle index that runs from 1 to 30 corresponds to the angles from 0.0 to π radians. The time residual index from 1 to 330 corresponds to times from -150.0 to 180.0 ns.

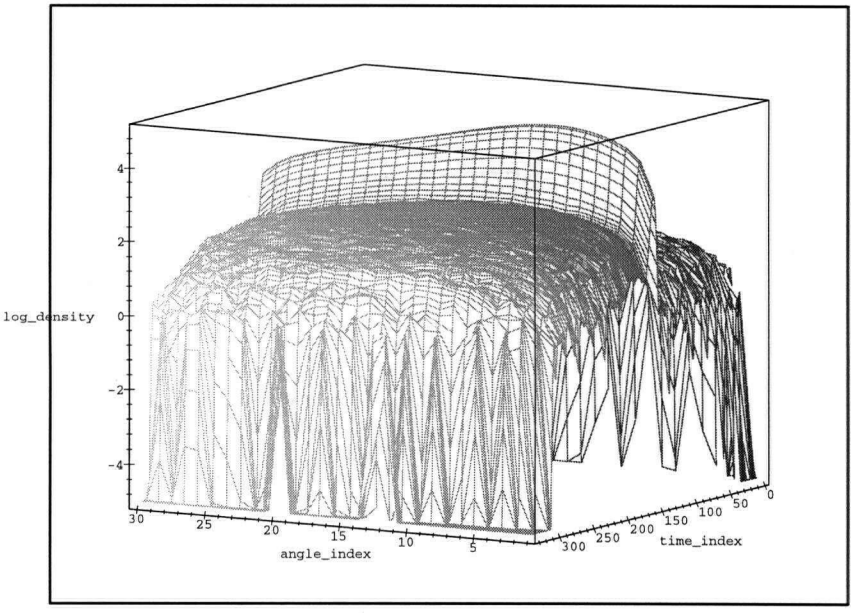

Figure 4.4: NC PDF Estimate in Both Time Residual and Hit Angle. The PDF was generated with a data set with close to 50000 events. The axis values are identical to those in Figure 4.3

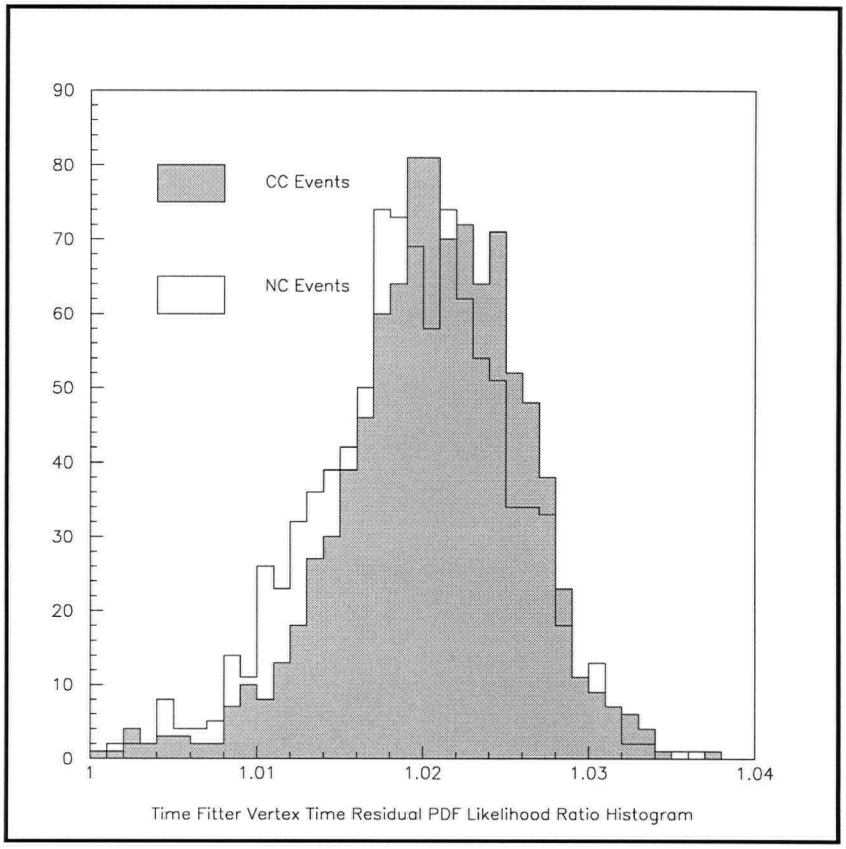

Figure 4.5: Histograms for the likelihood ratio formed using the $T_{residual}$ PDFs (LR_{time})

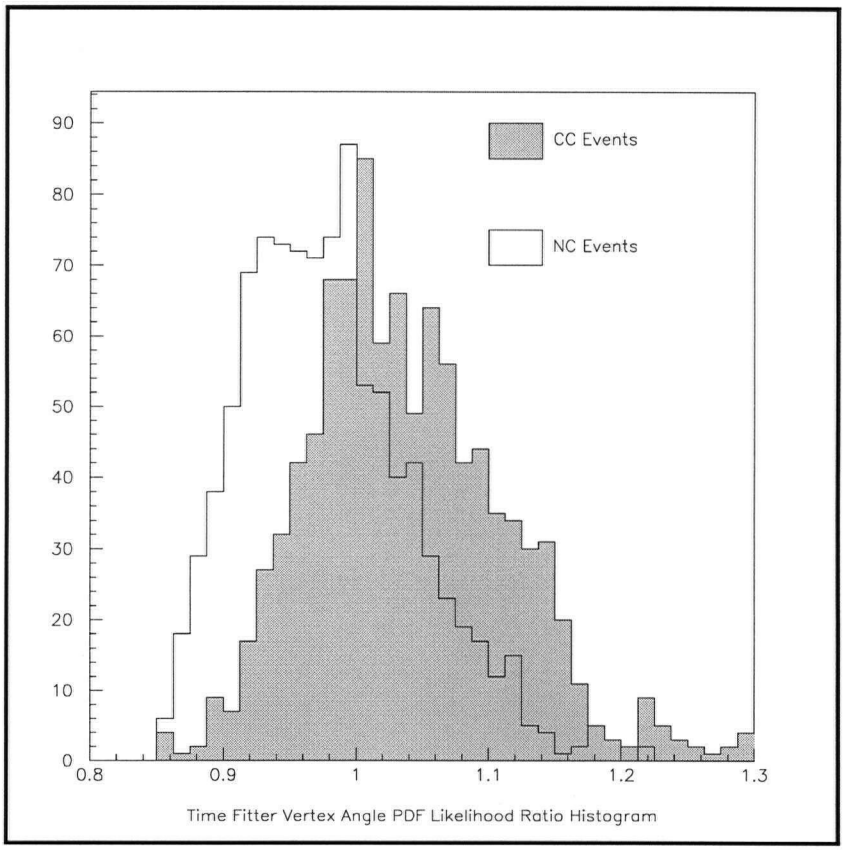

Figure 4.6: Histograms for the likelihood ratio using the hit angle PDFs (LR_{angle})

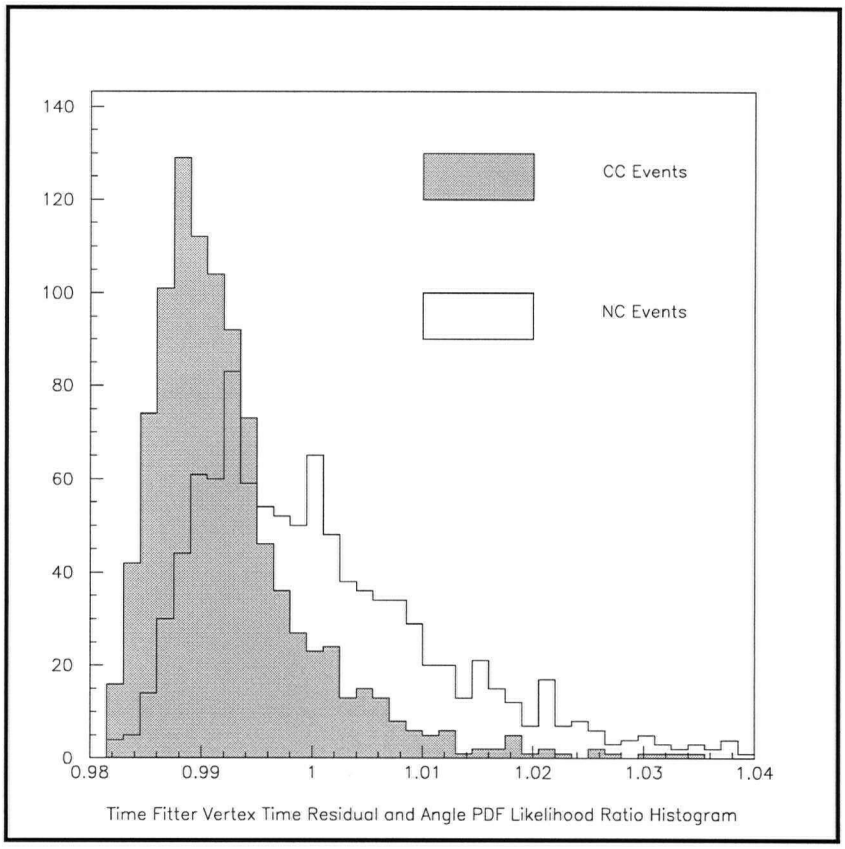

Figure 4.7: Histograms for the likelihood ratio using the joint $T_{residual}$ and hit angle PDFs ($LR_{time+angle}$)

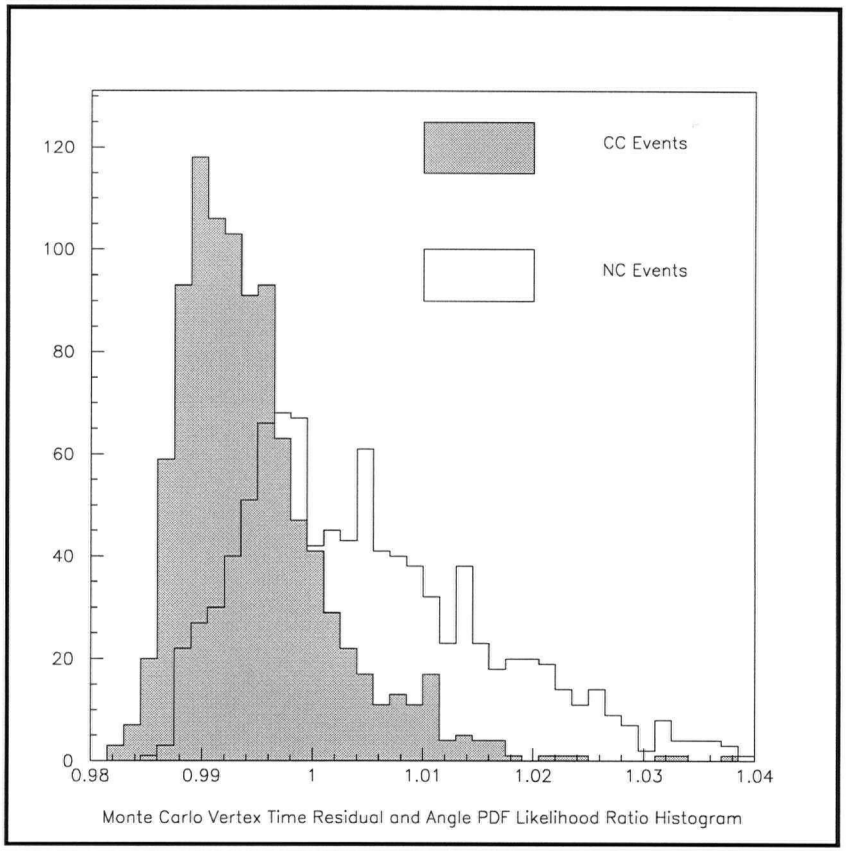

Figure 4.8: Likelihood ratio produced using the Monte Carlo vertices and the $T_{residual}$, hit angle PDFs. Of course, when fitting events, the actual event vertex is not available. This histogram shows the distributions that would result if one could use perfect fitters.

Chapter 5

Maximum Likelihood Fitters

One possible problem with the results presented in Section 4.8 is that the assumptions built into the SNOMAN time fitter might be minimizing the differences between CC and NC events. The SNOMAN time fitter assumes at the outset that there is a single source of Cerenkov light; it is possible that the returned fit vertices for NC events are biased in such a way as to mask their inherent nature as NC events. This possibility motivates the idea of creating a fitter that is designed to fit NC events. It is possible to create such a fitter, and a fitter tailored to CC events using the PDFs already developed in the previous chapter. Maximum likelihood fitting involves searching for a vertex with the highest likelihood as calculated with a given PDF. These fitters were created, and are described in this chapter.

5.1 Building a Maximum Likelihood Fitter

The purpose of an event fitter is to extract information from an event about the physical interaction that caused that set of PMT hits. Any given fitter makes assumptions about the nature of the event. Most SNO fitters constructed to date assume that there is one vertex for the creation of Cerenkov light. This restriction will be dropped here. In this thesis, all information extracted from an event is based on its vertex. The hit angles

are computed by first finding the average direction to the PMT hits from the returned vertex. The hit angles are subsequently measured from this average direction. An event is assumed to have a clearly defined vertex. For CC events, it is the position and time of interaction of a neutrino with a deuteron. For NC events, it is the position and time of a neutron capture on a ^{35}Cl nucleus. This NC vertex will be separated from the Cerenkov generation points by the distance that the emitted gamma rays travel before transferring its energy into electrons through Compton scattering.

Consider the case where a fitter is trying to extract the most likely vertex for a given event. Given an event from SNO, different vertices will have a different likelihood of having occurred for each of the various hypotheses about the nature of the event. If one assumes that the event is CC then one might expect the hit pattern to have certain features. One expects light from CC events to be less isotropic than light from NC events. A method characterizing these features numerically is with a PDF of parameters that are calculated from the set of PMT hits.

One of the important features of the Maximum Likelihood method is that it does not throw away any of the data. The time fitter built into SNOMAN tries to limit itself to PMT hits that are associated with direct hits, as opposed to scatter or noise. This results in 'pull', an effect which results in the fit time being dragged late in order to accommodate scattered light that cannot be tagged as such. When producing PDFs for any parameters, all of the physics of the situation is included. One is not required to throw away any information contained in noise tubes or hits due to scattered photons.

5.2 Likelihood Maximization

A likelihood, even when maximized, can be a very small number because it is a product of many numbers less than one. Instead of trying to find a vertex that maximizes a

likelihood, it is computationally better to find a vertex that minimizes the negative sum of density logarithms. This is what has been done for the fitters described here. Minimization of a function can be difficult, depending on the 'landscape' of the function to be minimized. It can be difficult to find the deepest minimum if there are many different minima. The more valleys there are the more searching and computing one must do to find them.

Two kinds of minimization were used in this thesis. One method of finding the minimum of a function is to use gradient descent. This is a very common technique, and a description and implementation is described in Section 10.5 of [Press]. This technique was used whenever it was believed that a vertex was known that was reasonably close to the global minimum. The other technique is called simulated annealing, and is used whenever there is reason to believe that there are local minima between the initial vertex and the global minimum.

5.3 Simulated Annealing

The name for simulated annealing derives from an analogy with solidification. If a molten metal is cooled rapidly it will often solidify into a phase that is not the lowest energy state for the final temperature. If it is cooled slowly, however, it stands a far better chance of finding the lowest energy phase, because thermal motion over longer periods of time allows the system to explore many configurations.

The analog to temperature while searching for a global minimum of a function is the magnitude of the typical steps that a probing point takes, and the degree to which the algorithm lets the current accepted point jump out of local minima.

Annealing occurs in the following way. First, a starting point for the minimization is chosen, and labeled as the current point. Next, a new point is picked by adding a random

vector to the point, with the components of this vector picked from a distribution. Next, the function is evaluated at the new point, and compared with the function value at the old point. If the new point has a lower function value, it is automatically adopted as the 'current' point for the next step. If the new point has a higher function value, then a random number is drawn. That random number determines whether the new point will be adopted as the current point. This gives the annealer the ability to jump out of a local minimum.

The theory and implementation of various kinds of simulated annealing is described in [Ingber]. Any simulated annealing method involves prescribing three functions: the distribution from which new points are selected, the probability that a new point will be accepted (based on the value of the function being optimized at both the new and old location), and a function that describes the lowering of the temperature with each 'time' step.

5.4 Very Fast Simulated Annealing

A variant of simulated annealing called Very Fast Simulated Annealing (VFSA) is used in this thesis. This algorithm was developed to address some problems with simpler versions of simulated annealing. VFSA allows each model parameter (x, y, z, t) to behave independently, in that there is a separate temperature and a separate temperature schedule for each variable.

There are a number of free parameters to be set when using VFSA. It is through these free parameters that the algorithm can be tuned to fit a specific problem. For the fitters discussed in this thesis, the models are event vertices requiring four parameters (x, y, z, t). For each model description parameter, there is an initial temperature, and a cooling schedule. There is also an initial acceptance temperature, and corresponding

cooling schedule. One must also choose the number of temperatures, and the number of probes to attempt per temperature.

For VFSA in four dimensions, there are 12 free parameters; there are four initial temperatures and four temperature schedule parameters (a pair for each of x,y,z and t), an initial acceptance temperature and corresponding temperature schedule, and finally a number of new points per temperature, and a number of temperatures. For each free parameter that was specific to a model parameter, x, y and z were always given the same value, and t was given a different one, reducing the overall number of free parameters to 8. This is motivated by previous knowledge that since the events were distributed uniformly throughout the D_2O, and directions were chosen isotropically, there is no distinction between the spatial directions. Temperatures are calculated, one after the other, using $T = T_0 e^{-ak^{\frac{1}{4}}}$ where a is the cooling schedule parameter and T_0 is the specific initial temperature and k is the 'time step'.

A pseudo code implementation of simulated annealing is shown in Appendix C. The new model is selected using the following set of equations.

$$u_i \sim [0,1] \tag{5.1}$$

$$y_i = sgn(u_i - \frac{1}{2})T_i[(1+\frac{1}{T_i})^{abs(2u_i-1)} - 1] \tag{5.2}$$

$$v_i^{new} = v_i^{old} + y_i(v_i^{max} - v_i^{min}) \tag{5.3}$$

$$v_i^{min} \leq v_i^{new} \leq v_i^{max} \tag{5.4}$$

where u_i is drawn at random, T_i is the current temperature for function variable i (i=1 corresponding to x etc.), v_i^{old} is the current vertex, v_i^{new} is the vertex about to be probed vertex, v_i^{min} and v_i^{max} are limits set by the user of the code. The perturbing distribution is a function of the annealing temperature, such that as the temperature lowers, the new point will be closer and closer to the current point.

Parameter	Initial Tempurature	Cooling Schedule
(x, y, z)	0.01	2.0
(t)	0.01	1.6378
Probe Acceptance	100.0	1.0

Table 5.1: VFSA parameter values. 'Probe acceptance' refers to the temperature associated with whether or not a newly probed vertex will be accepted if the new function value is worse than the one for the current vertex.

5.5 Selection of VFSA Parameters

The following strategy was used to set the annealing parameters. At first, the probing vertex should sample places all over the detector. Gradually, the probing should be restricted to a finer and finer area, stopping when the resolution limit of the detector has been reached.

In general a large number of probes are required to sufficiently sample the likelihood space. For the results presented here, the temperature was lowered 1000 times, and for each temperature, the likelihood space was sampled 30 times.

The annealing parameters were chosen using some knowledge of the detector, some knowledge from experience with the likelihoods, and some experience with the performance of the annealing as a function of the number of anneal probes per temperature. Values are shown in Table 5.1.

During annealing different initial vertices did produce different results. This is seen as a failure in the choice of annealing parameters. Consider starting from the centre of the detector. When the initial vertex time was chosen to be 25.0 ns before the first PMT hit, as is done in the SNOMAN Time Fitter, then when fitting to a $T_{residual}$ PDF the final vertex time was more than 20 ns too early about half the time. A more intelligent choice of initial vertex time is choosing the time that makes the average time residual to each of the PMT hits zero. This time will often be within 20.0 ns of the actual event

time. When the vertex was annealed with this choice of initial vertex, annealing followed by gradient descent proved a good choice for minimization, and improved fit residuals in the tails. Annealing did improve the number of fits that went very badly. Following SNO lore, a fitter is said to 'Fit the Event in Cleveland' when the fit residual is greater than 10^4 cm. For the smart initial vertex, annealing, and gradient descent combination there were about half as many events in Cleveland, as is seen on the log-log plot in Figure 5.1.

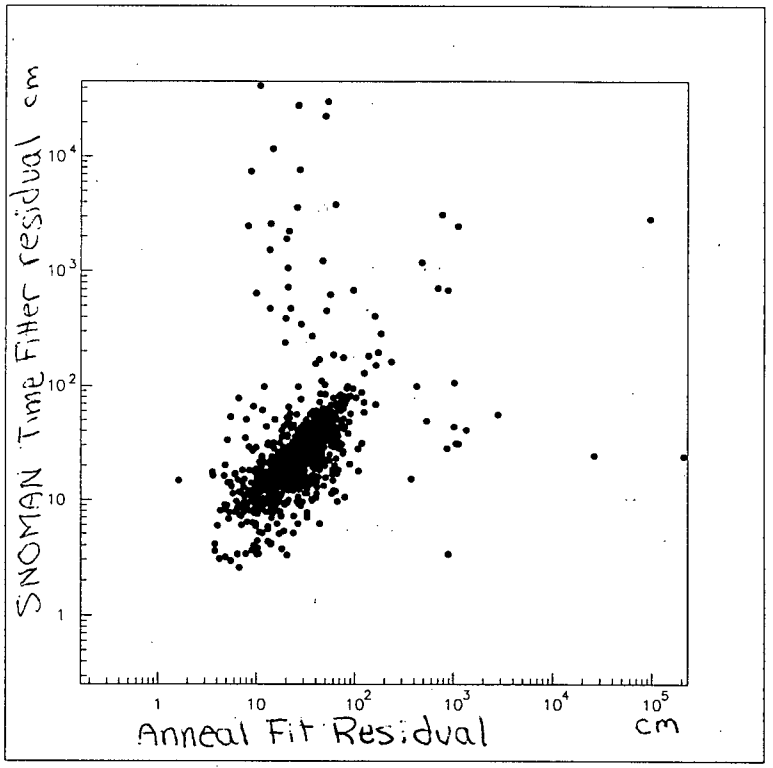

Figure 5.1: Maximum Likelihood fits fewer events 'In Cleveland'.

5.6 Using a Previous Fit as an Initial Vertex

The behaviour of the likelihoods near the global minimum is of great interest. A way of fast tracking to the interesting part of the function is to use a vertex from a previous fit as the seed for a new minimization. Very often, this tends to put the vertex in the global minimum for a new fitter. This method suffers from the problem that the next minimization is biased by the previous fit, but in the face of a very rough likelihood, this was seen as a necessary evil in order to be able to analyse the behaviour of the new likelihoods. This tended to leave in the poor performance of the SNOMAN Time Fitter with regard to fitting in Cleveland, but gains were made is the speed of the algorithm, because annealing was no longer required.

The two dimensional PDFs produced likelihoods that had very many local minima. This is deduced because whenever annealing was used with any initial vertex other than one provided by a previous fit, the subsequent gradient descent minimizer did not move the vertex by more than 1.0 cm or 0.1 ns. This difficulty was bypassed by feeding the result of the $T_{residual}$ fit to the $T_{residual}$ $\overline{\theta}$ fit.

5.7 Fitter Performance

It is impossible to tell whether any given minimization search is finding the global minimum, for indeed if one knew where the global minimum was one wouldn't need to do the search in the first place. In this case however, one does know where the Monte Carlo vertex is, and it is possible to check whether the fitter is finding a likelihood that is better than the likelihood for the Monte Carlo vertex. If this showed that the likelihood for the Monte Carlo vertex was better than the one found, then one would certainly know that the minimization process was not working. Histograms of the difference between

the negative log likelihoods for the Monte Carlo vertex and the fit vertex for the $T_{residual}$ PDF and the $T_{residual}$ and hit angle PDF are shown in Figure 5.2. This shows that the $T_{residual}$ only fit is finding a vertex with a better fit almost all of the time, whereas the $T_{residual}$ and hit angle PDF likelihood search is finding a higher likelihood than the Monte Carlo vertex around 85 % of the time.

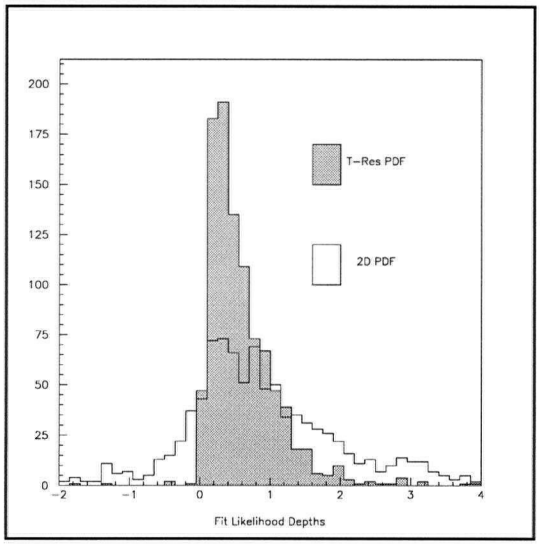

Figure 5.2: These histograms are a measure of the ability of the fitters to find the global minimum. These histograms show the difference between the negative log likelihood at the Monte Carlo vertex and at the returned fit vertex, for each of the Time PDF fitter and the 2D PDF. These histograms were formed using 1000 CC events.

After analysing a set of data, one is able to compare the vertices returned by the fitter with the Monte Carlo vertices that produces the hit patterns. This provides a measure of the ability of the fitter to reproduce the event vertex.

Two event sets, one of 1000 CC events and another with 1000 NC events were run through each fitter.

Figures 5.3 and 5.5 show the distance residuals for CC and NC events respectively. These Figures show that there is little difference in the various fitters ability to find the

event vertex location. A cut is made so that the distance residual is less than 100.0 cm. Figures 5.4 and 5.6 show the time residuals for CC events and NC events respectively. Note that the fit time is centred around the Monte Carlo time for the $T_{residual}$ and angle fits, but is late for the SNOMAN time fit.

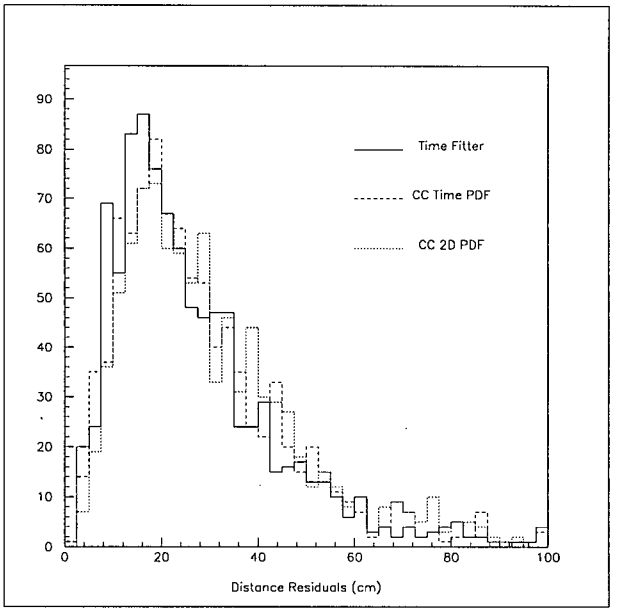

Figure 5.3: CC Distance Residuals

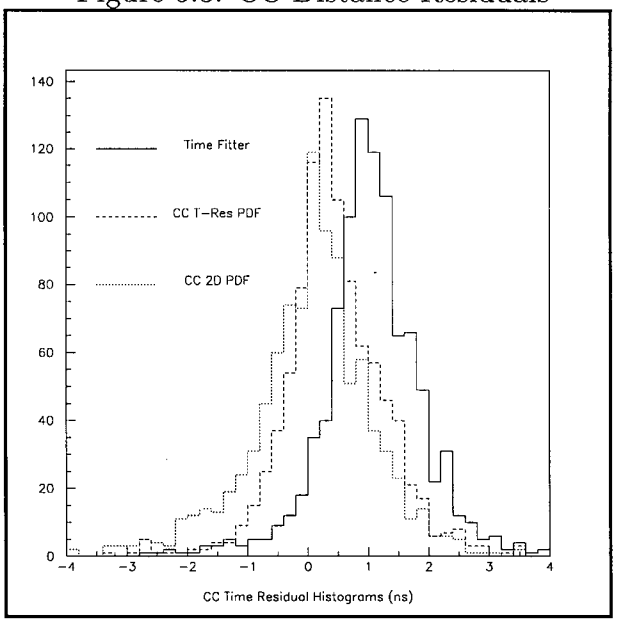

Figure 5.4: CC Time Residuals

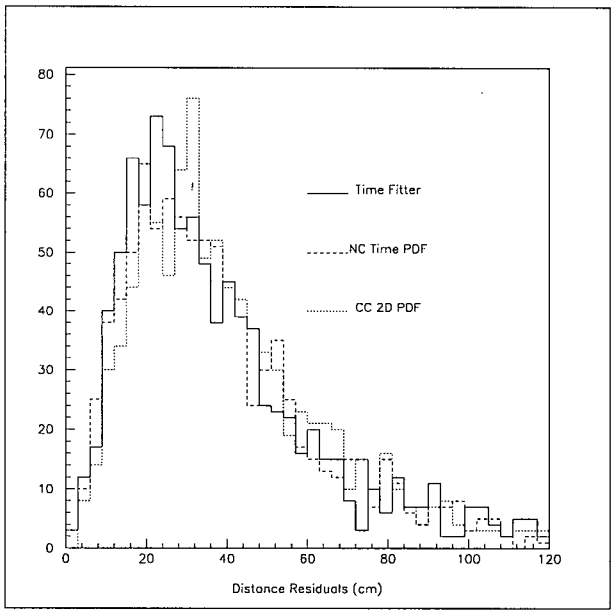

Figure 5.5: NC Distance Residuals

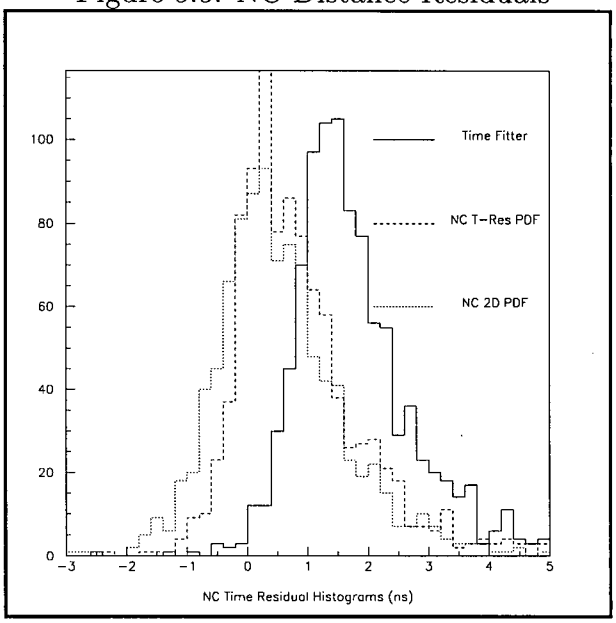

Figure 5.6: NC Time Residuals

Chapter 6

Maximum Likelihood Ratio Characterization

This chapter deals with event fraction determination, and the creation of event characterizing parameters. Given the fitters described in the previous chapter, the MLR characterization parameter is investigated. It is calculated by taking the ratio of the likelihood returned by the NC fitter and the likelihood returned by the CC fitter.

6.1 MLR Results

Event fits returned by the $T_{residual}$ fit and the $T_{residual}$ and angle fit were used to calculate likelihoods and form ratios of the maximum likelihoods (MLRs) . The parameters produced were used as fraction determination parameters. In each case, prototype sets with 2000 events each were used. The distribution of a mixture set, with 950 CC and 950 NC events was then fit against a normalized sum of the two prototype distributions. Results are shown in Table 6.1.

For the case of fitting to the $T_{residual}$ PDFs, the likelihoods used were from the angle only PDFs. A histogram of the characterization parameter is shown in Figure 6.1. For the case of using the two dimensional PDFs for the fit, the likelihoods were calculated

Fit Type	Likelihood PDF	Extracted CC Fraction
2D PDF	2D	0.493 ± 0.034
$T_{residual}$ PDF	hit angle	0.494 ± 0.037
Fit Type	**Parameter**	**Extracted CC Fraction**
SNOMAN Time	LR_{time}	$0.189 \pm^{0.155}_{0.172}$
SNOMAN Time	LR_{angle}	0.489 ± 0.042
SNOMAN Time	$LR_{timeangle}$	0.497 ± 0.039
$T_{residual}$ PDF	$\overline{\theta}$	0.521 ± 0.047

Table 6.1: MLR Event Fraction Determination Results. For fits performed using PDFs, each event was fit as a CC event and then as a NC event. The MLR was produced by taking the ratio of the returned negative log likelihoods for each event type. All previous results are included for comparison.

from the two dimensional PDF. A histogram of this characterization parameter is shown in Figure 6.2.

6.2 Comparison and Discussion

In view of the fact that the 2D PDFs have provided only a factor of two reduction in event fraction error estimate compared with the $\overline{\theta}$ event fraction determination, it is concluded that much of the ability to characterize NC and CC events in the manner described here is contained in PMT hit angle information.

6.3 Possible PDF Improvements

It is clear from the results above that much of the ability of these classification parameters lies in the hit angle information. This highlights the importance of the angle PDFs, and any improvements in the definitions of the angle PDFs may provide direct improvement in the ability of these parameters. More sophisticated definitions of the angle information may be fruitful. This might involve applying Hough transforms [Radcliff] to look for

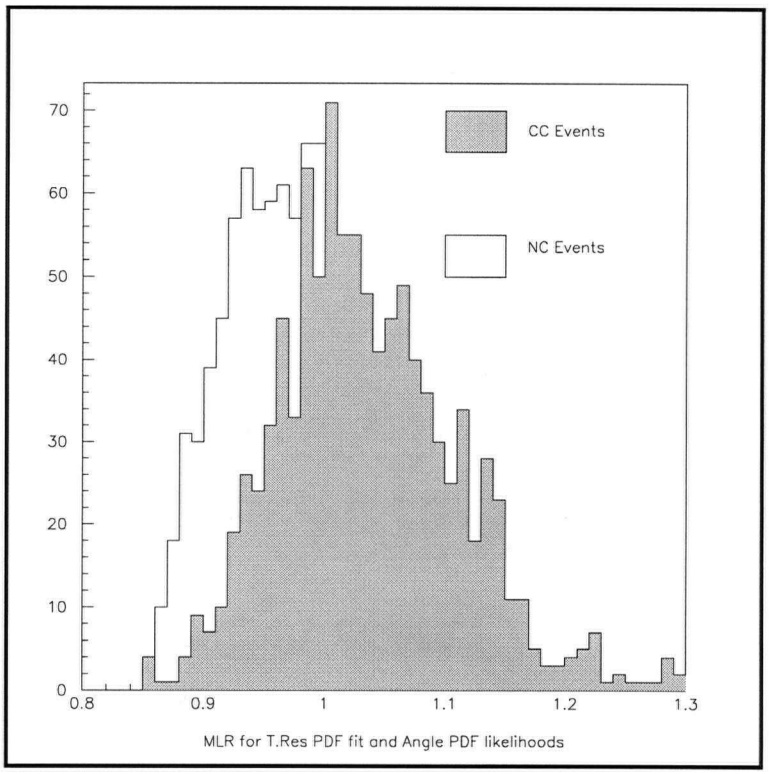

Figure 6.1: Histograms of Ratios of the Maximum Likelihoods for CC and NC events. These ratios were calculated using the vertex returned by a maximum likelihood fit using the $T_{residual}$ PDF. The likelihoods were calculated using the hit angle PDF.

more robust estimates of the direction of the event. It might also involve having separate PDF definitions for each event class. For instance, one could use the initial Monte Carlo direction as a direction definition for single e^- events, but use some other definition for direction for the NC events. This would mean fitting for both the event vertex and direction, as in [Klein] because alternate definitions for direction might not be directly calculable as they are for the mean PMT hit direction definition.

Moving to different PDF definitions for each class opens up a variety of new fitters for development. Also, restricting the analysis to CC and NC event types, each pair of definitions would have to be checked against each other for event fraction determination

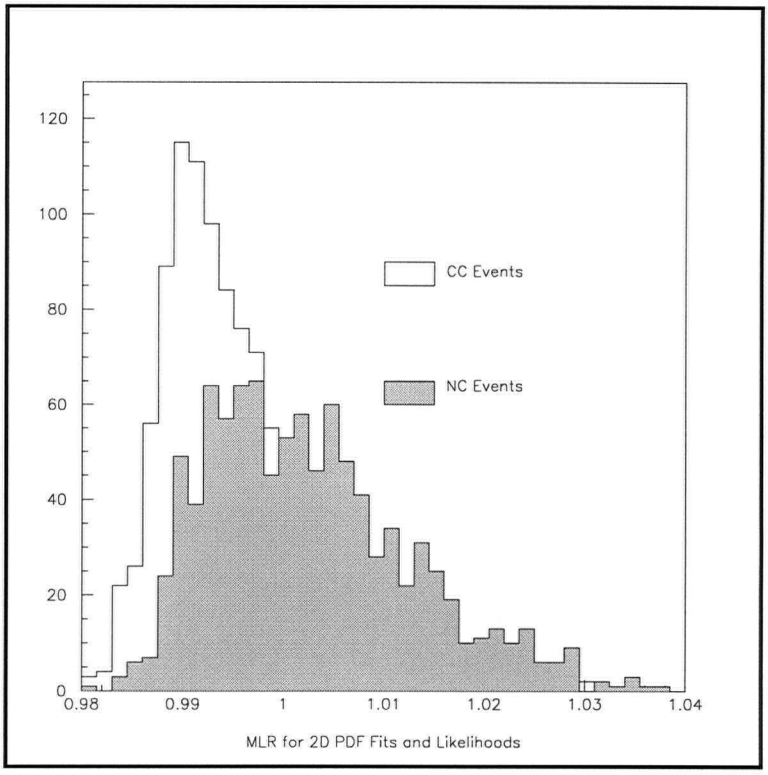

Figure 6.2: Histograms of Ratios of the Maximum Likelihoods for CC and NC events. These ratios were calculated using likelihoods returned by the maximum likelihood analysis fitters. Both the fits and the likelihoods used the two dimensional $T_{residual}$, hit angle PDFs.

ability.

Chapter 7

Conclusion

7.1 Conclusion

For event sets containing only CC and NC events, it has been shown that, compared with using only angle information, there is some additional ability to distinguish between CC and NC events in a joint PDF for $T_{residual}$ and PMT hit angle. The sets used were assumed to contain only CC and NC events. This means that if the analysis presented here were to be used, data would first have to have ES and other background events removed.

Out of all of the parameters investigated in this thesis, the best performance was found for the case where events were fit first with the NC and then with the CC fitter, and the resulting likelihoods, calculated with the two dimensional $T_{residual}$, hit angle PDFs, were used to form a ratio (MLR). This parameter distinguished between CC and NC events such that for prototype sets with 2000 events each, a mixed set with 950 NC and 950 CC events was estimated to have a CC fraction of 0.493 ± 0.034.

7.2 Future Work

A next step in the work presented in this thesis, would be to include more event classes in the analysis. This would allow the analysis of datasets comprised of more than two event types, including such things as ES events, and backgrounds of various kinds.

Extensions of this work should include some estimates of the systematic errors in the event fraction estimates. This would be done by analysing several separate Monte Carlo data sets, and looking at the distribution of results with respect to the exact result.

This work stands to be improved most by investigating improvements to the PDF definitions. This could be done by first looking for a PDF definition that provided as much differentiation in hit angle alone, and then forming the two dimensional PDFs after settling on a definition for the angle PDFs. This is recommended on the basis the vast majority of the discrimination information is in the angle information. One possibility is to use the Hough transform to look for a direction that maximizes the number of PMT hits in a Cerenkov cone, as in [Radcliff].

Given that fitters are naturally produced as a side effect of the creation of these PDFs, it is worthwhile spending some more effort on finding the global minimum of the likelihoods. In particular, it is preferable that these fitters not rely on being fed an initial vertex from another fitter. One possibility is to use a smart initial vertex, and then use simulated annealing over a smaller spatial and temporal neighbourhood, or possibly a grid search as in [Moorehead], followed by gradient descent.

Bibliography

[Mikheyev] S.P. Mikeyev and A. Yu. Smirnov, Nuovo Cimento C9, 17 (1986); Yad. Fiz. 42, 1441 (1985) [Sov. J. Nucl. Phys. 42, 913 (1985)].

[Wolfenstein] L. Wolfenstein, Phys. Rev. D17, 2369 (1978) and Phys. Rev. D20, 2634 (1979).

[Wildenhain] P. Wildenhain, Ph.D. thesis, University of Pennsylvania, 1995.

[Takeuchi] Y. Takeuchi, Ph.D. thesis, Tokyo Institute of Technology, 1995.

[Gavrin] V.N. Gavrin, talk given at TAUP 95 (Conference on Theoretical and phenomenological Aspects of Underground Physics), Sep. 19-23, 1995, Toledo, Spain.

[Anselmann] GALLEX Collaboration, P. Anselmann et al., Phys. Lett. B 357, 237 (1995).

[Silverman] B.W. Silverman *Density Estimation for Statistics and Data Analysis.* Chapman and Hall. 1986.

[Klein] J.R. Klein *Maximum Likelihood Fitting to Both Time and Angle.* SNO-STR-94-055 (Restricted SNO Internal Document)

[Bahcall95] J.N. Bahcall, M.H. Pinsonneault. Rev. Mod. Phys. Oct 95. or J. N. Bahcall and M. H. Pinsonneault, Rev. Mod. Phys. 64, 885 (1992).

[Bahcall89] J.N. Bahcall *Neutrino Astrophysics.* Cambridge University Press. 1989.

[Press] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. *Numerical Recipes in Fortran: The Art of Scientific Computing* Second Edition. Cambridge University Press. 1992.

[Brice95] S. Brice. *An Overview of the Feedforward Neural Network Technique and its Application to SNO Event Classification* SNO-STR-95-037 (SNO internal document)

[Brice96] S. Brice. *The Results of a Neural Network Statistical Event Class Analysis.* SNO-STR-96-001 (SNO internal document)

[SNNS] Stuttgart Neural Network Simulator. See http://vasarely.informatik.uni-stuttgart.de/snns/snns.html or search the WWW for SNNS.

[Masters] Masters, T. *Practical Neural Network Recipes in C++* Academic Press, 1993.

[SNO] SNO Collaboration, G.T. Ewan *et. al.*, *Sudbury Neutrino Observatory Proposal.* Report No. SNO-87-12, 1987 (unpublished)

[Ingber] Ingber, L. *Simulated Annealing: Practice versus theory* Mathl. Comput. Modelling, V 18, N 11, 1993, p 29 - 57.

[Radcliff] Radcliff, T.J. *Three methods for event-type identification in the SNO Detector* SNO-STR-95-002 (SNO Internal Document)

[Moorehead] Moorehead, M.E. *Grid Fitter for Reducing Tails in Spatial Distributions* SNO-STR-95-042 (SNO Internal Document)

[Bartholomew] Bartholomew, Groshev, *et al.* Nuclear Data A3, 1967.

Appendix A

Glossary

CC Charged Current event (CC). In this thesis, this refers to events as simulated by SNOMAN. These are single electron with an energy that has been selected from the expected charged current spectrum.

FNN A Feedforward Neural Network (FNN) is an algorithm loosely based on ideas about how neurons in the human brain compute.

MLA Maximum Likelihood Analysis (MLA) is a technique for finding the most probable value of a parameter associated with a given model for the source of a certain set of data.

NC Neutral Current event (NC). In SNO terms, this is a neutrino event that dissociates the deutron in heavy water, into a proton and a neutron. This thesis concentrates on the scenario where the neutron ejected in this reaction is detected by neu tron capture on ^{35}Cl.

PDF Probability Density Function.

PMT Photomultiplier Tube. An evacuated tube about the size of a human head, that is designed to detect single photons.

SNOMAN The **S**udbury **N**eutrino **O**bservatory's **M**onte carlo and data **AN**alysis package.

SNP Solar Neutrino Problem

SSM Standard Solar Model

VFSA Very Fast Simulated Annealing. An algorithm for finding the global minimum of a function.

Appendix B

NC Gamma Rays

This diagram [Bartholomew] shows the gamma rays that result from neutron capture on ^{35} Cl. This reaction results in a ^{36}Cl nucleus in an excited state. This nucleus reaches the ground state via the emission of gamma rays. There are many energy levels that the nucleus might go through in this transition, however, and the number of gamma rays that result from any given neutron capture depend on the route that the ^{36}Cl nucleus takes to the ground state.

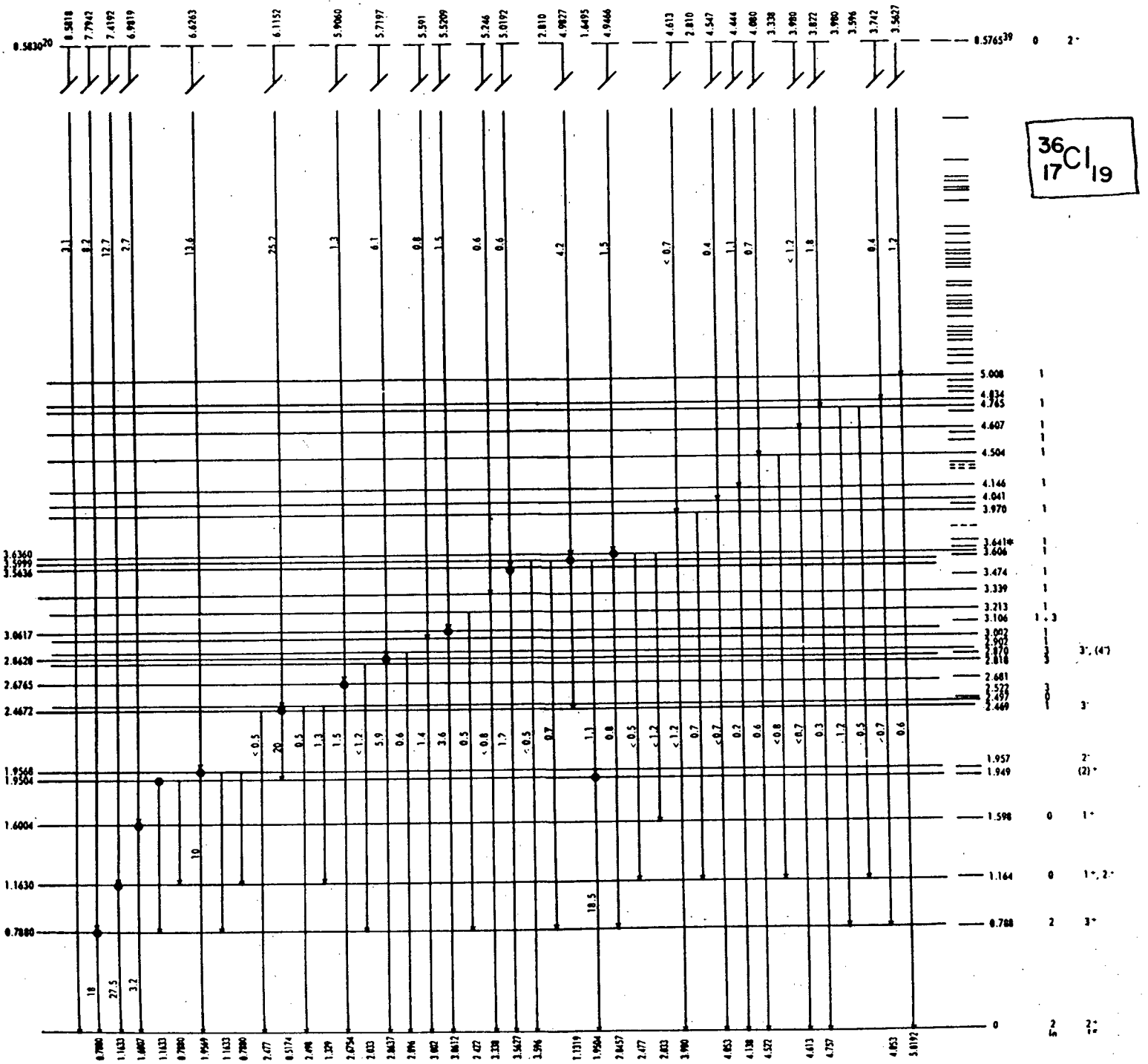

Figure B.1: Gammas from neutron capture on ^{35}Cl

Appendix C

Simulated Annealing Algorithm Outline

The following is an algorithm for the Simulated Annealing. This becomes Very Fast Simulated Annealing (VFSA) when the distributions used to select the new model, and make acceptance decisions are those shown in Section 5.4.

```
loop for each temperature(T)
    loop over number of random moves/temperature
        select new model
        move to new model, or to model limits
        if new model has lower function value, adopt it
        else draw random number and decide whether to adopt it
    end random moves loop
end temperature loop
```