Monte Carlo and Analysis Techniques for the Sudbury Neutrino Observatory

Stephen John Brice
Balliol College, Oxford University

Thesis submitted for the degree of Doctor of Philosophy
Trinity Term 1996

Abstract

The Solar Neutrino Problem has been troubling physicists for almost 30 years. Its original form, throughout the '70s and early '80s, was a deficit in the rate of solar neutrino capture on ^{37}Cl below that expected from standard particle and solar physics. Subsequent experiments using capture on ^{71}Ga and electron scattering have served to strengthen the problem to the extent that a solar physics solution is all but ruled out and one must look to new particle physics for the answer.

The option of claiming that the experiments are simply wrong is now also untenable, with at least two of the four having to be rejected to remove the problem. Particular confidence can be placed in the results of the gallium experiments as they have both used a ^{51}Cr electron capture ν_e source as a very powerful check on their efficiencies for extracting and detecting the ^{71}Ge atoms that result from neutrino capture. The high γ activity of this source creates a potential radiation hazard and makes it necessary to simulate γ transport through the geometry and media of the shielding with the aid of the EGS4 (Electron Gamma Shower) code system.

Possibly the best hope for a resolution to the Solar Neutrino Problem rests with the Sudbury Neutrino Observatory (SNO), a heavy water Čerenkov detector currently being constructed in Canada. It has the advantage over all previous solar neutrino detectors of being able to detect both neutral and charged current events as well as measuring the spectrum at the high energy end of the neutrino flux.

Like most modern day high energy physics experiments the analysis of data from SNO will rely on an accurate and comprehensive Monte Carlo simulation of the detector. Since the experiment relies on the detection of a rather small number of Čerenkov photons it is necessarily a rather delicate optical device which makes the proper simulation of the detector geometry a crucial aspect of the Monte Carlo. The results from SNO will also depend critically on fitting a position and/or a direction to the events which are detected. Such fits are needed principally for the imposition of fiducial cuts and the assessment of event direction distributions. Although simple in principle, fitting algorithms are complicated by the need to identify and account for PMT hits which have resulted from noise or scattered light.

Perhaps the most important result from SNO will be the number of neutral and charged current events it detects. A number of ways of extracting these two numbers are possible, but one particularly appealing method involves the use of a pattern recognition technique to distinguish between the characteristic PMT hit patterns that result from the two event classes. Such an analysis is able to extract charged and neutral current rates to an accuracy of $\sim$5% with only one year of data.

Acknowledgements

Thanks are due to the entire SNO group at Oxford for their interest in this work and for the creation of such a productive working environment. I am particularly indebted to my supervisor, Dave Wark, for considerable advice and guidance. His tireless, infectious spirit and encyclopedic knowledge of subjects ranging from particle physics, through genetics and cell biology, to history and politics have made him a delight and a privilege to work with. Nick West deserves special thanks for his guidance in matters of computer programming and his unwavering patience and good humour. I have also benefited greatly from discussions with Nick Jelley and, latterly, Mike Bowler. They have provided some real object lessons in scientific rigour and the conduct of research.

Martin Moorhead has contributed a number of useful ideas, as has Matthew Thorman, whose recent work has filled in many of the gaps in this thesis. Thanks are due also to Mike Lay for advice on EGS simulations and geometry coding and to Richard Taplin for his interest in my work and corresponding willingness to explain his own.

Several members of the physics support staff at Oxford deserve a special mention: Pete Gronbech and John Macallister for maintaining an excellent computer network, Beverly Rogers and Del Batts for their tolerance of my bizarre travel arrangements and financial transactions, and all the canteen staff whose admirable credit facilities have frequently saved me from starvation.

On a more personal note I am fortunate to have a number of longstanding friends whose interests lie close to my own. Richard Maunder, in particular, has contributed in ideas and questions. I am also especially grateful to my long-suffering flat-mate, Andy Forbes. Few others would have been so tolerant and good natured in the face of my intolerance and mood swings. Finally, and perhaps most importantly, I should like to thank my family for their unstinting support and encouragement.

Contents

Chapter 1

The Solar Neutrino Problem

The adventure of the sun is the great natural drama by which we live, and not to have joy in it and awe of it, not to share in it, is to close a dull door on nature's sustaining and poetic spirit.

HENRY BESTON
Midwinter, The Outermost House (1928)

Willingly would I burn to death like Phaeton, were this the price for reaching the sun and learning its shape, its size, and its substance

EUDOXUS C. 408-355 B.C.

1.1 History

SPECULATION AS to the nature and mechanisms of the sun date back as far as human history itself. By the late 15^{th} century it had been claimed that the sun was just another star and the next four centuries saw a steady accumulation of knowledge and inference as to its constitution and evolution, culminating in the classic works of Eddington [Eddington 20, Eddington 26]. Although these papers had a great deal to say about the hydrodynamics and makeup of stars they could only speculate as to the kind of power generating mechanism that could produce so much energy over such a long time. In another classic paper at the end of the next decade, Bethe proposed nuclear fusion as the mechanism and described in some detail the necessary chain of interactions [Bethe 39]. This paper made no explicit mention of neutrinos (although the Fermi theory of β decay was five years old by this time), but by 1948 it had been realised that the sun would be a prolific source of these, as yet unobserved, particles [Crane 48]. Since it was believed at this time that the only significant solar neutrino production came from the proton-proton fusion that initiates the p-p chain (see later) the detection of such low energy solar neutrinos was not considered feasible and attention focused, with considerable success, on the detection of neutrinos from reactor piles.

The situation changed in early 1958 when it was reported that the cross-section for the reaction $^3He(\alpha,\gamma)^7Be$ (until then a very minor player in the p-p chain) was a factor of 1000 larger than had been expected [Holmgren 58]. The ^7Be so produced in the sun could then capture a proton to form ^8B which could β decay and produce high energy neutrinos (up to ~15 MeV). This high energy flux offered the possibility of detecting solar neutrinos and in 1962 a collaboration, headed by Ray Davis, formed to do just that, using the capture on ^{37}Cl in the form of 100,000 gallons of C_2Cl_4 as a detecting reaction. At the same time the first calculation of neutrino fluxes from a detailed model of the sun was published [Bahcall 63]. When the first results from the chlorine experiment were published in 1968 there was significant disagreement with the predicted rate, the measured flux being considerably lower [Davis 68, Bahcall 68]. Over the subsequent 30 years of ever more precise solar modelling and neutrino detection this disagreement between prediction and measurement has come to be known as the Solar Neutrino Problem and is the subject of this chapter. The next section describes the Standard Solar Model used to predict the neutrino fluxes and is followed by a section detailing the four experiments to date which have measured these fluxes, a detailed discussion of the Solar Neutrino Problem, and a section offering some possible solutions to it. The chapter closes with a brief look at the future prospects for solar neutrino physics.

1.2 The Standard Solar Model

Solar models attempt to simulate the reactions and transport phenomena of the sun and so produce a coherent picture of its internal workings. Based on a realistic set of initial conditions and simple evolution equations such models must reproduce the observable features of the sun and make predictions of other observables which might subsequently be subject to experimental measurement. A Standard Solar Model (SSM) is one which uses widely accepted physics and initial conditions, as distinct from a Nonstandard Model where the inputs or evolution equations are adjusted to varying degrees of plausibility (usually to achieve the goal of a diminished neutrino flux).

Simply stated the sun, like all other main sequence stars, fuels itself by burning hydrogen nuclei to form helium nuclei according to the reaction

$$4p + 2e^- \rightarrow {}^4He + 2\nu_e \quad + 26.731 \text{ MeV}$$

In this way ~600 million tons of hydrogen is burned every second in the solar core. In the sun the conversion is mainly achieved through a sequence of reactions known as the p-p chain which is illustrated in Fig. 1.1. Another sequence, the CNO cycle, where the conversion is catalysed by carbon, oxygen, and nitrogen, also plays a small role (~1.5% of the energy generation).

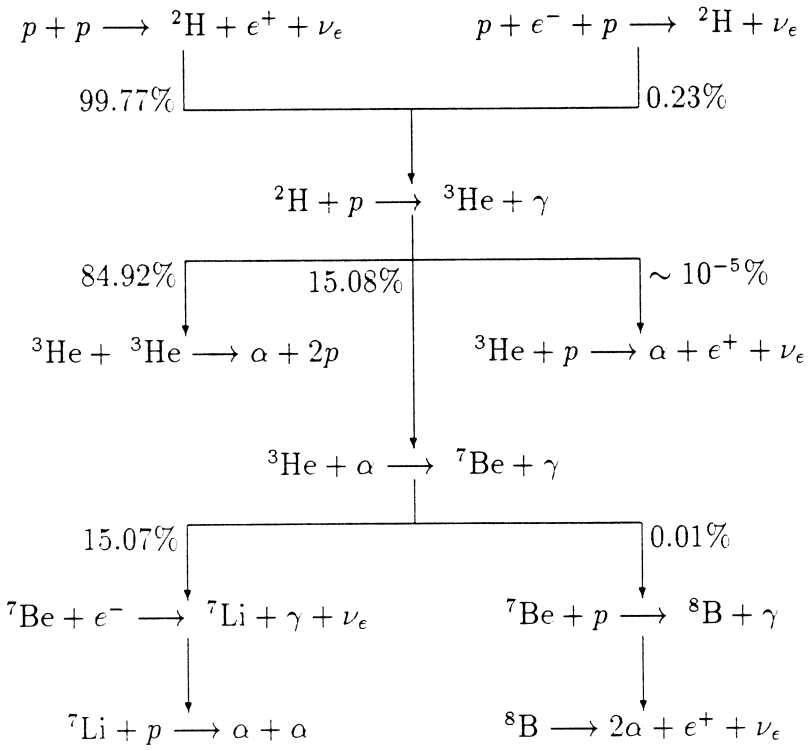


Figure 1.1: The p-p chain (from [Bahcall 88b]).

1.2.1 The Equations of Stellar Evolution

The basic equations of stellar evolution involve relatively simple physics. Since the measured oblateness of the sun ($R_{\mathrm{equatorial}}/R_{\mathrm{polar}}$ - 1) is $\leq 2 \times 10^{-5}$ and solar rotation is neglected then spherical symmetry is assumed. The effects of mass loss and magnetic fields are also neglected and until recently material diffusion was also left out (its inclusion in the work of [Bahcall 95] produced only a minor change in the results).

The principal evolution equation expresses the condition of hydrostatic equilibrium within the sun, where the expansive forces of radiation and thermal pressure are exactly balanced by the contractive force of gravity. (If this condition were not true to high accuracy the sun would collapse in less than an hour!) If $P(r)$ and $\rho(r)$ are the pressure and density at a radius r and $M(r)$ is the mass enclosed by this radius, the condition of hydrostatic equilibrium can be expressed in the usual way as

$$\frac{\mathrm{d}P}{\mathrm{d}r} = -\frac{GM(r)\rho(r)}{r^2} \tag{1.1}$$

As mentioned, the pressure is the sum of radiative and thermal contributions. Since the solar core is almost completely ionised its material behaves like an ideal gas, but with corrections having to be made for electron degeneracy and screening. This means that the thermal pressure as a function of temperature and density is given by the usual ideal gas equation multiplied by a factor of $(1+D)$, where D incorporates the corrections. Adding to this the usual expression for radiative pressure gives the equation of state

$$P(r) = \frac{aT^4}{3} + \frac{\rho kT(1+D)}{\mu m_{\mathrm{H}}} \tag{1.2}$$

where k is the Boltzmann constant, $a \equiv \frac{8\pi^5 k^4}{15c^3h^3}$ is the Stefan-Boltzmann constant, and μ is the mean molecular mass (see [Schwarzschild 58] for a discussion of the equation of state of a stellar plasma).

Luminosity is produced in the sun almost entirely from nuclear, but also from mechanical sources. If E_{nuclear} is the specific nuclear energy production rate at a radius r and S is the specific entropy at that radius then, by considering a spherical shell of thickness $\mathrm{d}r$, the luminosity $\mathrm{d}L_r$ that it produces can be written as

$$\frac{\mathrm{d}L_r}{\mathrm{d}r} = \rho(4\pi r^2)\left[E_{\mathrm{nuclear}} - T\frac{\mathrm{d}S}{\mathrm{d}t}\right] \tag{1.3}$$

The core of the sun is believed to be stable against convection and thermal perturbations ([Bahcall 89] and references therein). Consequently the net flux of energy from the core to the surface of the sun that is induced by the temperature gradient is predominantly in the form of photon diffusion. This simplifies the equation governing energy transport so that it has the form

$$L_r = -(4\pi r^2)\frac{ac}{3}\frac{1}{\kappa\rho}\frac{\mathrm{d}T^4}{\mathrm{d}r} \tag{1.4}$$

where κ is the total opacity which breaks into radiative and conductive components according to $\kappa^{-1} = \kappa_{\mathrm{rad}}^{-1} + \kappa_{\mathrm{cond}}^{-1}$. In the central regions of the sun the radiative contribution dominates.

Completing this set of coupled, non-linear, differential equations governing solar evolution is the boundary condition applied to the outer layer. Here convective equilibrium is assumed resulting in equations connecting the temperature and pressure of the outer region. These outer layers of the sun are more complex than this assumption allows for, but this matters little as the conditions of the solar core are very insensitive to the particular outer boundary condition that is chosen. The difference between the most careful and crudest treatment of the solar convective zone corresponds to at most a 2% change in the calculated solar neutrino fluxes [Bahcall 88b].

1.2.2 Input Parameters and Initial Conditions

Whilst Eqns. 1.1 to 1.4 provide the framework for the time evolution of the Standard Solar Model, several parameters and initial conditions must also be specified.

Chemical Abundances

The chemical abundances of the elements affect the radiative opacity and hence the temperature-density profile of the sun. These abundances are usually expressed as three numbers; X, the mass fraction of hydrogen, Y, the mass fraction of helium, and Z, the mass fraction of elements heavier than helium. To calculate the opacities Z is broken down into individual elemental abundances and modern Solar Models use computer codes to calculate opacities as a function of elemental abundance, temperature, and density. To estimate these abundances two assumptions are made; first the sun is assumed to have been chemically homogenous when it arrived on the main sequence, and second the present solar surface abundances are assumed to reflect the initial numbers for elements at least as heavy as carbon (the relatively low temperatures ensure that nuclear reaction rates are negligible at the solar surface). These two assumptions enable solar spectroscopy to be used to pin down the heavy element mass fractions (for details see [Grevesse 84] and [Aller 86]). The elemental abundances of meteorites can also be used to estimate primeval solar abundances. Until recently there was significant disagreement between the spectroscopic and meteoritic numbers, but within the last three years consensus has been reached [Grevesse 93b, Grevesse 93a].

The mass fraction ratio Z/X is also a crucial input parameter. Most solar models of the '80s used the value from either [Grevesse 84] or [Aller 86] (0.02765 and 0.02739 respectively). The latest best estimate comes from [Grevesse 93b] and is $0.0245(1.000 \pm 0.061)$ (see [Bahcall 95] for the method of estimating the uncertainty).

Radiative Opacity

Since the temperature profile and hence nuclear reaction and neutrino production rates depend very strongly on the radiative opacity it is crucial that it be calculated as accurately as possible. It is fortunate for solar modelling that radiative opacities play an important role in the development of thermonuclear devices and as a result much effort has been expended in developing codes to calculate them. Two major and independent programs have been written by the Los Alamos [Huebner 86] and Livermore National Laboratories [Rogers 92]. The OPAL code of Livermore is the more recent and accurate, but is in good agreement with the earlier Los Alamos results. These codes rely on the chemical composition of the solar interior and include all of the relevant statistical mechanics and atomic physics [Huebner 86].

Reaction	ν Label	Q (MeV)	$S(0)$ (keV b)	dS/dE (b)
$p(p,e^+\nu_e)^2H$	pp	1.442	4.07×10^{-22}	4.52×10^{-24}
$p(pe^-,\nu_e)^2H$	pep	1.442	-	-
$^2H(p,\gamma)^3He$	-	5.494	2.5×10^{-4}	7.9×10^{-6}
$^3He(p,e^+\nu_e)^4He$	hep	19.795	8×10^{-20}	-
$^3He(^3He,2p)^4He$	-	12.860	$5.15 \times 10^{+3}$	-9×10^{-1}
$^3He(^4He,\gamma)^7Be$	-	1.586	5.4×10^{-1}	-3.1×10^{-4}
$^7Be(p,\gamma)^8B$	-	0.137	2.43×10^{-2}	-3×10^{-5}
$^7Be(e^-,\nu_e)^7Li$	^{7}Be	0.862	-	-
	^{7}Be	0.384	-	-
$^8B(e^+\nu_e)^8Be$	^{8}B	17.980	-	-

Table 1.1: Values of the cross-section parameters for some reactions of the p-p chain (taken from [Bahcall 89]).

In the core of the sun one can be confident that the opacity is accurately known. In this region it is dominated by photon scattering off free electrons and inverse bremsstrahlung from protons and alpha particles; all processes that can be simply calculated using quantum and statistical mechanics. The opacity contribution from heavier elements is somewhat harder to calculate as the abundances are subject to significant uncertainties, there may not be complete ionisation and the atomic physics is rather more complex.

Nuclear Cross-Sections

The rates for nuclear fusion in the solar interior are dominated by Coulomb barriers. The typical thermal energies of the particles are only a few keV, whereas barrier heights are in the MeV range. These conditions produce rates too low to be measured in laboratory experiments and so some form of extrapolation is necessary to compute the low energy cross-sections. One common parameterisation for strongly exothermic reactions is to set

$$\sigma(E) = \frac{S(E)}{E}\exp(-2\pi\eta)$$

where

$$\eta(E) = \frac{\alpha z Z}{v/c}$$

for two ions of charges ze and Ze which collide with relative velocity v giving a kinetic energy E in the centre of mass frame. The factor $\exp(-2\pi\eta)$ comes from the Gamow penetration of the Coulomb barrier whilst the $1/E$ dependence comes from the (λ^2) overlap of the particle wavefunctions. $S(E)$ is a slowly varying function expressing the small energy dependence of the matrix element. Values of $S(E)$ can be measured at MeV energies in the laboratory and are then extrapolated to the zero energy values used in solar modelling. Values of the cross-section parameters for some of the p-p chain reactions are given in Table 1.1.

1.2.3 Calculational Procedure

The calculational procedure of a solar model starts with a main sequence star that has a homogenous composition. An initial guess is made as to the hydrogen mass fraction X and an entropy-like variable S. (These are the quantities used by Bahcall in, for example, [Bahcall 88b]. Other authors use different variables). Formulating the solar evolution expressions as difference equations the model is iterated forward in discrete time steps until the present age of the sun (4.6×10^9 years) is reached. Usually between 5 and 10 steps are used. As hydrogen burns in the deep interior, providing both luminosity and support against gravitational collapse, the composition becomes inhomogeneous and other parameters such as opacity alter accordingly. When the model reaches the present age of the sun the mass, luminosity, and radius that it predicts are calculated and compared with the measured values. The whole process is then repeated with altered values of X and S until the predictions for $M_\odot$, $L_\odot$, and $R_\odot$ are in sufficient agreement with observation (better than 1 part in 10^5). This final solar model predicts values for the hydrogen, helium, and heavy element mass fractions, the present complete set of physical variables within the sun, the spectrum of acoustic oscillation frequencies observed on the solar surface, and the neutrino fluxes.

1.2.4 The Neutrino Fluxes

Shown in Fig. 1.2 are the Standard Solar Model predictions for the neutrino fluxes as a function of energy. For continuous spectra the units are $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}\ \mathrm{MeV}^{-1}$ and for line spectra $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$. The solid lines indicate p-p chain fluxes, with the CNO cycle fluxes as dashed lines. Table 1.2 shows predictions of the total neutrino flux from each of the relevant reactions of the p-p chain and CNO cycle. The column labelled [Bahcall 95] is one of the latest models which includes the effects of helium and heavy element diffusion (unlike all previous Standard Solar Models) as well as refined opacity and cross-section data. The columns labelled [Bahcall 88b] and [Turck-Chièze 88] were viewed by many physicists at the time (1988) as representing the best guess and lower bound of the SSM fluxes respectively. Where shown the errors represent effective 3σ limits. It should be noted that the distribution of errors is not well understood and cannot be assumed

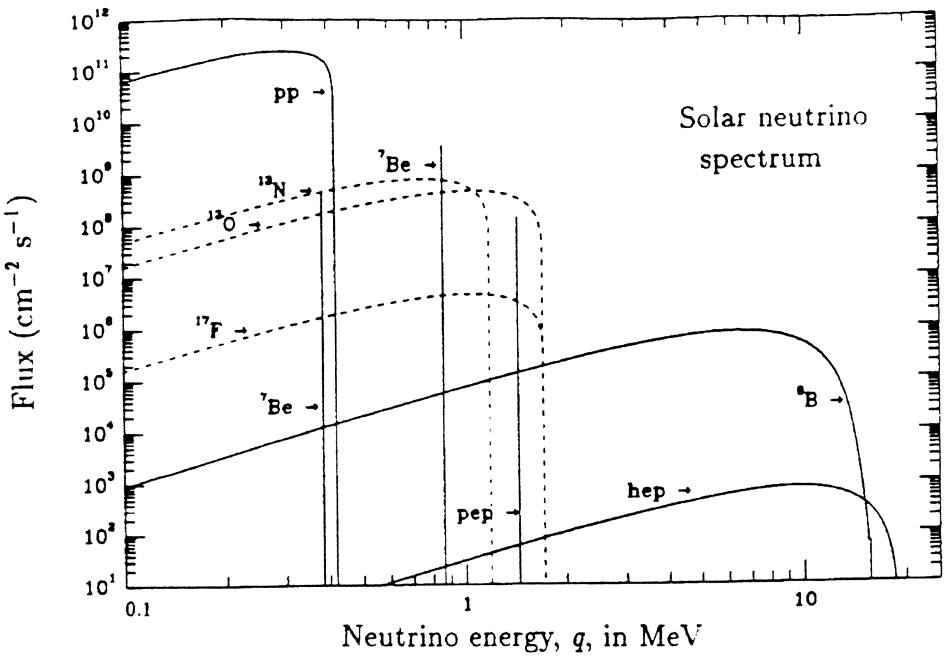


Figure 1.2: The Standard Solar Model prediction for the energy spectra of solar neutrinos (taken from [Bahcall 88b])

Gaussian.

The uncertainties in the neutrino fluxes stem principally from two sources

Opacity:- The calculation of the heavy element contribution to the opacity is particularly complex, as noted earlier. In addition it is sensitive to the ratio of heavy element and hydrogen mass fractions (Z/X), a number which is deduced from meteoritic values and spectroscopic measurements of the solar surface.

The $S(0)$ values:- For the ^{3}He(^{3}He,2p)^{4}He, ^{3}He(^{4}He,γ)^{7}Be, and ^{7}Be(p,γ)^{8}B reactions the $S(0)$ factor in the cross-section has to be extrapolated from laboratory measurements at higher (MeV) energies, where the rate is high enough to permit measurement. The ^{7}Be(p,γ)^{8}B cross-section measurement is further complicated by the radioactivity of the target.

These uncertainties feed through into uncertainties in the 'branching ratios' of the p-p chain (as shown in Fig. 1.1). Since the fusion reactions have to penetrate a sizeable Coulomb barrier one powerful way in which the opacity and cross-section uncertainties propagate is through their effect on the core temperature of the sun. By analysing the output of 1000 solar models, where small changes were made to the input parameters of each, [Bahcall 88b] has derived the empirical

ν flux	[Bahcall 95]	[Bahcall 88b]	[Turck-Chièze 88]
$p(p,e^+\nu_e)^2H$ ($\times 10^{10}$)	$5.91(1.00^{+0.01}_{-0.01})$	$6.0(1.0^{+0.02}_{-0.02})$	5.98
$p(pe^-,\nu_e)^2H$ ($\times 10^8$)	$1.40(1.00^{+0.01}_{-0.02})$	$1.4(1.0^{+0.05}_{-0.05})$	1.3
$^3He(p,e^+\nu_e)^4He$ ($\times 10^3$)	1.21	7.6	-
$^7Be(e^-,\nu_e)^7Li$ ($\times 10^9$)	$5.15(1.00^{+0.06}_{-0.07})$	$4.7(1.0^{+0.15}_{-0.15})$	4.18
$^8B(e^+\nu_e)^8Be$ ($\times 10^6$)	$6.62(1.00^{+0.14}_{-0.17})$	$5.8(1.0^{+0.36}_{-0.36})$	3.8
$^{13}N(e^+\nu_e)^{13}C$ ($\times 10^8$)	$6.18(1.00^{+0.17}_{-0.20})$	$6.1(1.0^{+0.49}_{-0.49})$	-
$^{15}O(e^+\nu_e)^{15}N$ ($\times 10^8$)	$5.45(1.00^{+0.19}_{-0.22})$	$5.2(1.0^{+0.58}_{-0.58})$	-
$^{17}F(e^+\nu_e)^{17}O$ ($\times 10^6$)	$6.48(1.00^{+0.15}_{-0.19})$	$5.2(1.0^{+0.46}_{-0.46})$	-

Table 1.2: Different model predictions for the solar neutrino fluxes. The units are $cm^{-2}\ s^{-1}$.

power law dependence of each neutrino flux on the core temperature, expressed in the form $\phi_\nu \propto \sim T^\beta$, where β is -1.2 for pp neutrinos, 8 for ^{7}Be neutrinos, and 18 for ^{8}B neutrinos (see Table 1.1 for the reactions corresponding to the mnemonics). The strong dependence on temperature of the ^{8}B flux stems from the Gamow penetration factor of the $^7Be(p,\gamma)^8B$ reaction, which produces the ^{8}B, being a strong function of temperature. The pp neutrino flux is far better constrained (see the uncertainties in Table 1.2) because of the measured luminosity of the sun. If one accepts that the p-p chain produces the sun's luminosity then every termination of the chain contributes ~26 MeV to the luminosity. Each termination also produces at most 2 and at least 1 pp neutrino, if one assumes complete ignorance of all the cross-sections of the chain. Incorporating even a vague knowledge of the p-p chain cross-sections then constrains the pp flux quite tightly.

1.3 Solar Neutrino Detection

To date four experiments have reported results of solar neutrino detection; three radiochemical experiments, using ^{37}Cl or ^{71}Ga as a target, and one water Čerenkov detector. This section reviews these experiments and their results. At this stage it is also convenient to introduce the 'SNU'. Originally coined as a pun by [Bahcall 69], it is the unit conventionally used to express solar neutrino capture rates and is defined to be 10^{-36} captures per target atom per second.

Flux	[Bahcall 95] (SNU)	[Bahcall 88b] (SNU)
pp	0.00	0.0
pep	0.22	0.2
hep	-	0.03
^{7}Be	1.24	1.1
^{8}B	7.36	6.1
^{13}N	0.11	0.1
^{15}O	0.37	0.3
^{17}F	-	0.003
Total ($\pm 3\sigma$)	$9.3^{+1.2}_{-1.4}$	$7.9^{+2.6}_{-2.6}$

Table 1.3: Capture rates predicted by two Standard Solar Models for the ^{37}Cl detector.

1.3.1 The Chlorine Experiment

The Chlorine experiment [Cleveland 95] has been taking data for almost 30 years and is located ~4850 feet below ground in the Homestake Gold Mine in South Dakota. The detector itself is a tank containing 615 tons of liquid perchlorethylene (C_2Cl_4) which amounts to about 2.2×10^{30} target atoms of ^{37}Cl. The detecting reaction is

$$\nu_e + {}^{37}\mathrm{Cl} \rightarrow \epsilon^- + {}^{37}\mathrm{Ar}$$

which has a threshold of 0.814 MeV making the experiment sensitive to all the neutrino fluxes bar that from the pp reaction (see Fig. 1.2). As evidenced by Table 1.3, however, the presence of a 4.98 MeV excited state in ^{37}Ar that is the isobaric analogue of the ^{37}Cl ground state means that the high energy ^{8}B neutrinos dominate the capture rate [Bahcall 64b].

At the beginning of each exposure a small quantity of ^{36}Ar or ^{38}Ar is added to the tank to act as a carrier. A couple of months later this argon, together with the few atoms of ^{37}Ar produced by solar neutrinos, is removed by circulating through the tank a large volume of helium. The argon is then extracted from this gas flow by a charcoal trap cooled to liquid nitrogen temperatures. The volume of argon removed serves as a measure of the extraction efficiency of the process (usually 90 - 95%). The charcoal trap is heated to desorb the argon which is then passed over heated titanium to remove reactive elements and undergoes gas chromatography to remove any other unreactive gases (principally krypton and xenon). The remaining Ar is mixed with a tritium free methane counter gas and loaded into a small proportional counter which detects the Auger electrons from the electron capture decay of ^{37}Ar.

The counting of the ^{37}Ar, which has a half life of 35 days, proceeds for about 8 months in proportional counters made from zone refined iron and calibrated with X-rays from a ^{55}Fe source. Background is reduced by a passive shield and an external sodium iodide detector for rejecting cosmic rays. Events are also rejected which do not have a pulse height and rise time consistent with an Auger electron. The accepted counts are fitted to a flat background and an exponential decay with 35 day half life using a maximum likelihood technique.

A number of tests have been performed to verify the assumption that the efficiency for recovering the ^{36}Ar or ^{38}Ar carrier is the same as that for extracting the neutrino induced ^{37}Ar. Doubts arise because the neutrino capture produces an energetic $^{37}Ar^{+}$ ion ejected from the C_2Cl_4 parent, and the subsequent behaviour of this ion could conceivably differ from that of the atomic carrier gas. To check that this is not the case C_2Cl_4 labelled with ^{36}Cl is introduced into the tank. The β decay of this isotope produces an $^{36}Ar^{+}$ ion with approximately the same recoil momentum as the $^{37}Ar^{+}$ from neutrino events. Using the normal extraction procedure the efficiency for removing this ^{36}Ar is measured by activation analysis to be $100 \pm 3\%$, a number consistent with the $\sim$95% inferred from other tests.

Background levels of ^{37}Ar come primarily from the ^{37}Cl(p,n)^{37}Ar reaction where the protons come from two sources; cosmic ray events and fast neutrons from the rock wall which undergo the ^{35}Cl(n,p)^{35}S reaction. The rate of cosmic induced ^{37}Ar is estimated by exposing 600 gallon tanks of C_2Cl_4 at higher levels in the mine and extrapolating down to the 4850 ft level. This yields a cosmic induced background rate of 0.4 ± 0.16 SNU [Fireman 84], a number that is consistent with theoretical estimates. The second source of protons, production from rock neutrons, has a reaction threshold of $\sim$1 MeV and the resulting ^{37}Ar production rate can therefore be made negligible by flooding the cavity surrounding the tank with water which acts as a neutron moderator.

Shown in Fig. 1.3 are the Standard Solar Model predictions for the event rate in the chlorine experiment over the last 30 years and the measured event rate also as a function time. The data are taken from [Bahcall 89] and [Cleveland 95]. It can be seen that the measured rate has always been consistently well below the theoretical rate. The latest Standard Solar Model prediction [Bahcall 95] and experimental value [Cleveland 95] are

$$\text{Experiment} = 2.55 \pm 0.17(\text{stat}) \pm 0.18(\text{syst}) \text{ SNU } (1\sigma)$$

$$\text{Theory} = 9.3^{+1.2}_{-1.4} \text{ SNU } (3\sigma)$$

The discrepancy is clear. One final comment should also be made on the celebrated weak time correlation between the chlorine experiment capture rate and the inverse of the sunspot number that seemed to be emerging in the late '80s. If confirmed such a correlation would imply a neutrino magnetic moment as so-

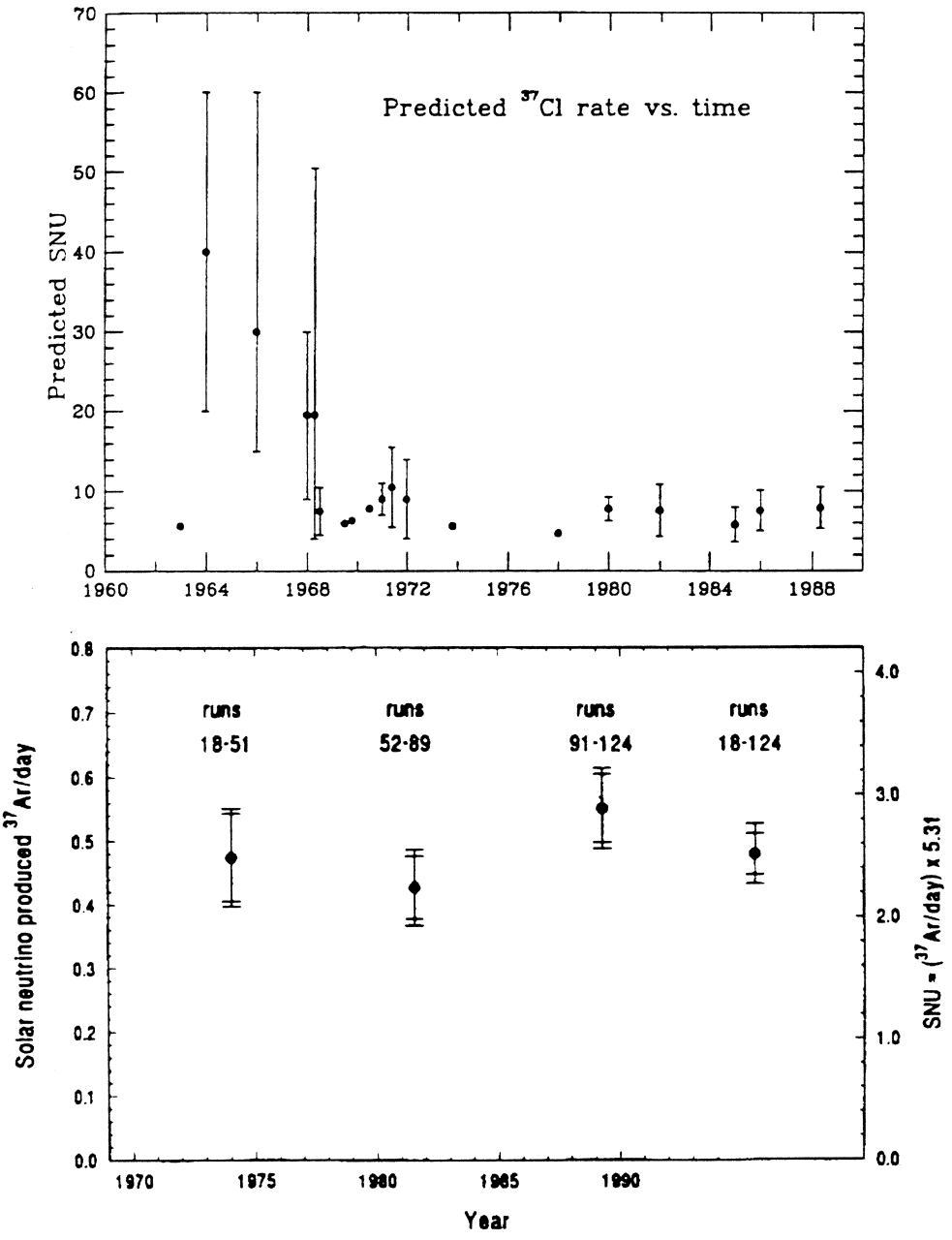


Figure 1.3: (Top) The Standard Solar Model predictions for the event rate in the chlorine experiment over the last 30 years (taken from [Bahcall 89]). (Bottom) The measured event rates over the last 30 years (taken from [Cleveland 95]).

lar magnetic fields are known to vary with sunspot number. With the taking of subsequent data, however, this correlation has now weakened.

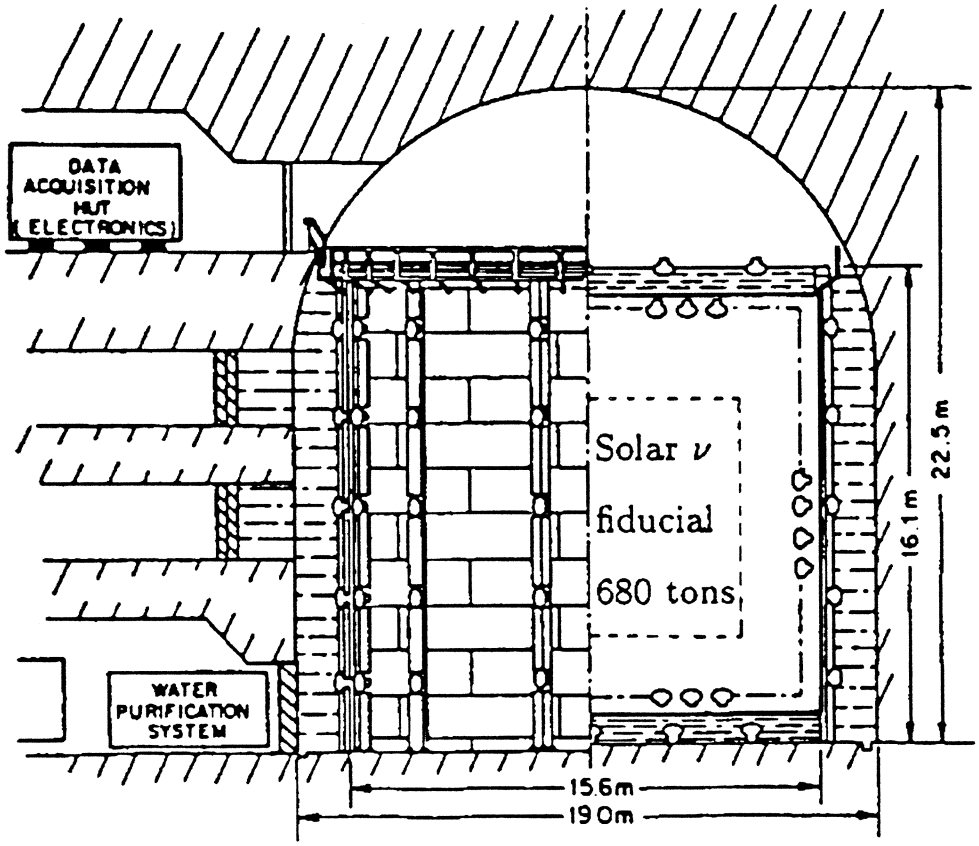


Figure 1.4: A schematic diagram of the Kamiokande II detector (taken from [Nakamura 93]).

1.3.2 Kamiokande II

The Kamiokande I detector [Hirata 91] was built to search for proton decay. In 1985 it was upgraded with better background vetoing, a more sophisticated water purification system and improved electronics to become Kamiokande II, a detector able to measure the high energy solar neutrino flux.

Kamiokande II detects solar neutrinos through the electron scattering reaction

$$\nu_x + e^- \rightarrow \nu_x + e^-$$

in a cylindrical inner volume containing 2140 tonnes of water. The scattered electrons produce Čerenkov light and this is detected by 948 20-inch photomultiplier tubes which surround the inner water volume and give a geometric coverage of ~20%. The number of hit PMTs in an event gives a measure of the electron's energy and the hit pattern reflects its direction. Outside these PMTs is a second volume of water (total mass 2800 tonnes) viewed by a further 123 PMTs. This acts

Fiducial volume	680 tonnes
Trigger efficiency	50% at 5.2 MeV
	90% at 7.5 MeV
Energy resolution	$19/(T_e/10\text{MeV})$ %
Vertex resolution	1.5m at 10 MeV
Direction resolution	28° at 10 MeV

Table 1.4: Some important detector parameters of Kamiokande II.

as an anti-coincidence veto for cosmic muons as well as a passive shield to reduce the background from the cavity walls. The entire detector, shown schematically in Fig. 1.4, is located 1 km (equivalent to 2700 metres of water) below ground in the Kamioka mine in Japan.

The detector background comes from the natural radioactivity of the water, γs from the rock of the cavity wall, and the β decays of cosmic muon induced spallation products. These backgrounds enforce an experimental energy threshold corresponding to a 7.5 MeV electron. This means that Kamiokande II is only sensitive to the high energy ^{8}B neutrino flux. The activity of the water is dominated by the decays of ^{214}Bi, which has an endpoint of 3.26 MeV and is part of the ^{238}U decay chain (see Appendix F). This ^{214}Bi comes from the radium dissolved in the water, the levels of which are controlled by extraction in the water purification system. A second, and rather more troublesome source, is leakage of radon gas into the water which can diffuse in from the air surrounding the detector as well as from the components of the water system. It is combatted by sealing the entire purification and detector system and filling with a radon free cover gas as well as adding a degasser stage to the water system. The flux of γs from the rock is reduced by $\sim$3 orders of magnitude by the outer water volume and by another factor of 400 by limiting the analysis to a central fiducial volume of 680 tonnes. Those γs that do make it into the fiducial volume tend to be heading radially inwards, a signature which can be used in their rejection. Cosmic ray muons are not hard to identify as they deposit a large amount of energy, however, they also have the effect of producing short lived isotopes either by fragmentation of an oxygen nucleus, capture of a π^- produced by the inelastic scattering of a muon, or by capture of a stopped muon. These spallation products require that the detector be vetoed for $\sim$50 ms after each cosmic muon event.

The energy response of the detector is calibrated in three ways:

- The reaction Ni(n,γ)Ni′ produces γs up to $\sim$9 MeV whose subsequent Compton scattering gives high energy electrons. By collimating the γs these electrons can also be fairly well collimated and the resolution of the direction fit checked.

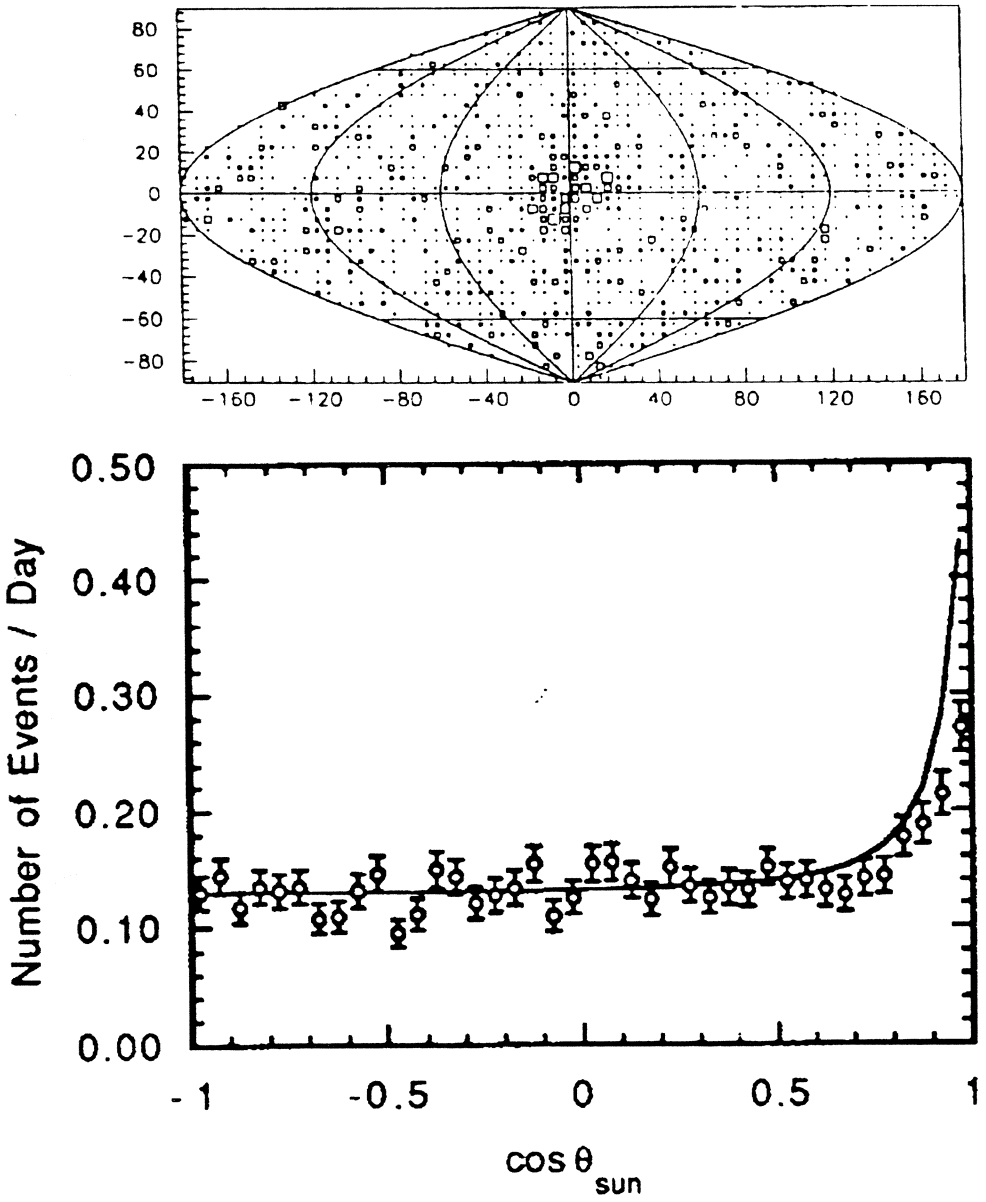


Figure 1.5: (Top) The image of the sun taken by means of neutrinos in a celestial coordinate system where the sun is at the centre. The data are divided into 4° by 4° bins and a box is plotted in each bin with a size weighted by the number of events in the bin. (Bottom) The angular distribution of events relative to the solar direction. The solid curve is the SSM prediction. (Both plots taken from [Suzuki 95]).

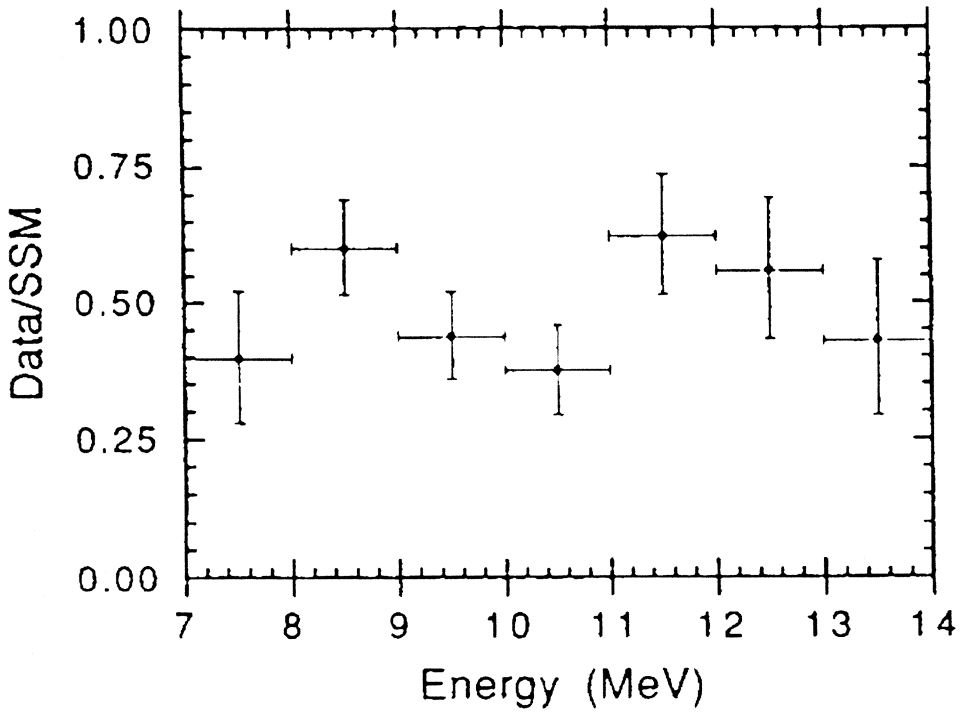


Figure 1.6: The energy distribution of events in Kamiokande II normalised to the SSM prediction (taken from [Suzuki 95]).

- The decay of stopped muons produces high energy Michel electrons whose spectrum is well known.
- The decay of muon induced spallation products has a high energy tail which can be used so long as the relative abundances of the products can be estimated.

The results of these energy calibrations as well as some other detector parameters are listed in Table 1.4.

The raw detector data are passed through a series of cuts. Events are only accepted which lie in the energy range ~7.5 to 20 MeV (these numbers vary from one analysis to another) and reconstruct within the fiducial volume. Any event occurring within ~50 ms of a muon and sufficiently close to it is rejected as are events in the outer metre of the fiducial volume which are travelling inwards, under the suspicion that they are due to external γs. The remaining data is tested for its direction correlation to the sun (the electron scattering angular distribution is strongly forward peaked) by fitting the event direction distribution to a flat background plus a peak pointing away from the sun that is a convolution of the

electron scattering angular distribution and the detector's angular response.

Fig. 1.5 shows the direction distribution of the Kamiokande II events which pass the above cuts. The upper plot shows the reconstructed directions in a celestial coordinate system where the sun is at the centre. This plot is quite simply a neutrino image of the sun. The excess of events in the solar direction is clear. The lower plot gives the angular distribution of the events relative to the solar direction. The peak pointing away from the sun is again clear, but is significantly smaller than the Standard Solar Model prediction which is shown as a solid line on the plot. Fig. 1.6 shows the energy distribution of the accepted Kamiokande II events normalised to the prediction of the Standard Solar Model. The data points are consistent with a flat line i.e. consistent with the predicted energy distribution, but the number of events is approximately half the expected value.

Assuming a ^{8}B β decay spectrum the total ^{8}B neutrino flux can be inferred from the Kamiokande II measurements. Its value [Suzuki 95] and the SSM prediction [Bahcall 95] are

$$\text{Experiment} = [2.89\ ^{+0.22}_{-0.21}(\text{stat}) \pm 0.35(\text{syst})] \times 10^6\ \text{cm}^{-2}\text{s}^{-1}\ (1\sigma)$$

$$\text{Theory} = [6.62\ ^{+0.93}_{-1.13}] \times 10^6\ \text{cm}^{-2}\text{s}^{-1}\ (3\sigma)$$

Again the deficit in the measure solar neutrino flux is obvious.

1.3.3 The Gallium Experiments

There are currently two experimental groups, the SAGE and GALLEX collaborations, which are detecting solar neutrinos through the reaction

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

The great appeal of this reaction is its low threshold energy of 233 keV which enables it to detect the dominant solar pp neutrinos, whose energy has an endpoint of 420 keV. Since the pp neutrino flux is quite tightly constrained by the solar luminosity a measured flux that differed significantly from the prediction would be clear evidence for new physics. The contribution to the predicted capture rate of the different neutrino fluxes from two Standard Solar Models is given in Table 1.5.

Although differing in several important ways the SAGE and GALLEX experiments have a common basic procedure, somewhat similar to that of the chlorine experiment.

1. A small amount of stable germanium is added to the gallium to act as a carrier.

Flux	[Bahcall 95] (SNU)	[Bahcall 88b] (SNU)
pp	69.7	70.8
pep	3.0	3.0
hep	-	0.06
^{7}Be	37.7	34.3
^{8}B	16.1	14.0
^{13}N	3.8	3.8
^{15}O	6.3	6.1
^{17}F	-	0.06
Total ($\pm 3\sigma$)	137^{+8}_{-7}	132 ± 7

Table 1.5: ^{71}Ga capture rates predicted by two Standard Solar Models.

2. The gallium is left for 2-4 weeks.

3. The germanium (neutrino produced plus carrier) is extracted from the gallium and converted into the gas germane (GeH_4).

4. ^{71}Ge decays by electron capture with a half life of 11.43 days. The germane is mixed with xenon and placed in a small proportional counter. The Auger electrons from ^{71}Ge decay (K shell - 10.37keV, L shell - 1.17keV) are detected and the amount of ^{71}Ge produced is inferred, using the amount of stable germanium removed to measure the extraction efficiency.

The SAGE experiment [Abdurashitov 95] is situated at the Baksan Neutrino Observatory beneath Mount Andyrchi in the Caucasus Mountains, which provides shielding equivalent to 4700 metres of water. Their gallium is in liquid metallic form (the melting point of gallium is 29.8°) and the total mass of 57 tonnes (initially 30 tonnes) is held in a series of Teflon lined reactor vessels. The extraction of the germanium begins by mixing a solution of hydrochloric acid and hydrogen peroxide into the gallium [Abazov 91]. This produces an aqueous phase containing the germanium which is syphoned off and reduced in volume by vacuum evaporation. The addition of more hydrochloric acid and a purge with argon then removes the germanium as $GeCl_4$. After a further reduction in volume, this chloride is used to synthesise germane, which is mixed with xenon to form the counter gas. As measured by the amount of germanium carrier removed the overall extraction efficiency is 80-90%. Counting proceeds for several months in miniature proportional counters surrounded by a sodium iodide veto shield which is itself inside a large passive shield. The counters are calibrated with a number

of sources, most notably ^{55}Fe which produces 5.9 keV x-rays and a cerium source which excites the production of 1.03, 5.09, and 9.75 keV x-rays in the counter gas.

The GALLEX experiment [Paoluzi 89] is located in Hall A of the Gran Sasso National Laboratory in Italy where the rock overburden provides a shielding equivalent to 3300 metres of water. Their gallium is in the form of 101 tonnes of acidified gallium chloride (equivalent to 30.3 tonnes of gallium) which makes the chemistry of the germanium extraction somewhat easier. The acidity produces a high concentration of chloride ions which ensures that the germanium is in the form of $GeCl_4$. This is extracted with ~99% efficiency by purging with a large volume of nitrogen. The germanium chloride is then converted into germane and mixed with xenon to form the counter gas. The counting and calibration are conducted in much the same way as the SAGE experiment but with the option of using a passive copper shield in place of the sodium iodide counter.

The principal sources of background in the proportional counters are cosmogenically produced ^{68}Ge which decays by electron capture with a half life of 288 days, radon gas diffusing into detector components, tritium in the germane, and cosmic ray muons. The ^{68}Ge can be removed by repeated extractions from the gallium on its arrival underground, a process that provides a useful cross check on the germanium extraction efficiency. By sealing the experimental areas the levels of radon can be reduced by several orders of magnitude and the use of low tritium water in the synthesis of germane reduces the tritium count rate to < 0.01 counts per day. The rock overburden reduces the contribution of cosmic muons to 3.7 ± 1.1 SNU in GALLEX and to a negligible level in SAGE. In addition to these methods of background reduction a great deal of further elimination is achieved by applying energy and rise time cuts to the pulses from the proportional counters. The other major background concern is the production of ^{71}Ge from non-neutrino sources. The only significant mechanism for this is ^{71}Ga(p,n)^{71}Ge with protons resulting from cosmic muon interactions and (α,p) or (n,p) reactions. The rate of ^{71}Ge production from these mechanisms can be estimated from the rate of production of ^{69}Ge, which is not produced by neutrinos in any significant quantity.

Recently both experiments have carried out a very powerful cross check on their interaction cross-sections and extraction efficiencies by constructing very powerful (~1.5 and ~0.5 MCi) ^{51}Cr sources. This isotope decays by electron capture producing neutrinos of 751 keV (90%) and 426 keV (10%). By exposing the gallium to this source and knowing its activity the amount of ^{71}Ge that is counted can be compared with that which would be expected given the various extraction and detection efficiencies and the interaction cross-sections. Both experiments express the result of this cross check as the ratio R of measured to expected counts. Their results are [Vignaud 95, Elliott 96]

GALLEX	$R = 0.97 \pm 0.11$	(1σ stat. & syst.)
SAGE	$R = 0.93^{+0.15}_{-0.17}$	(1σ stat. & syst.)

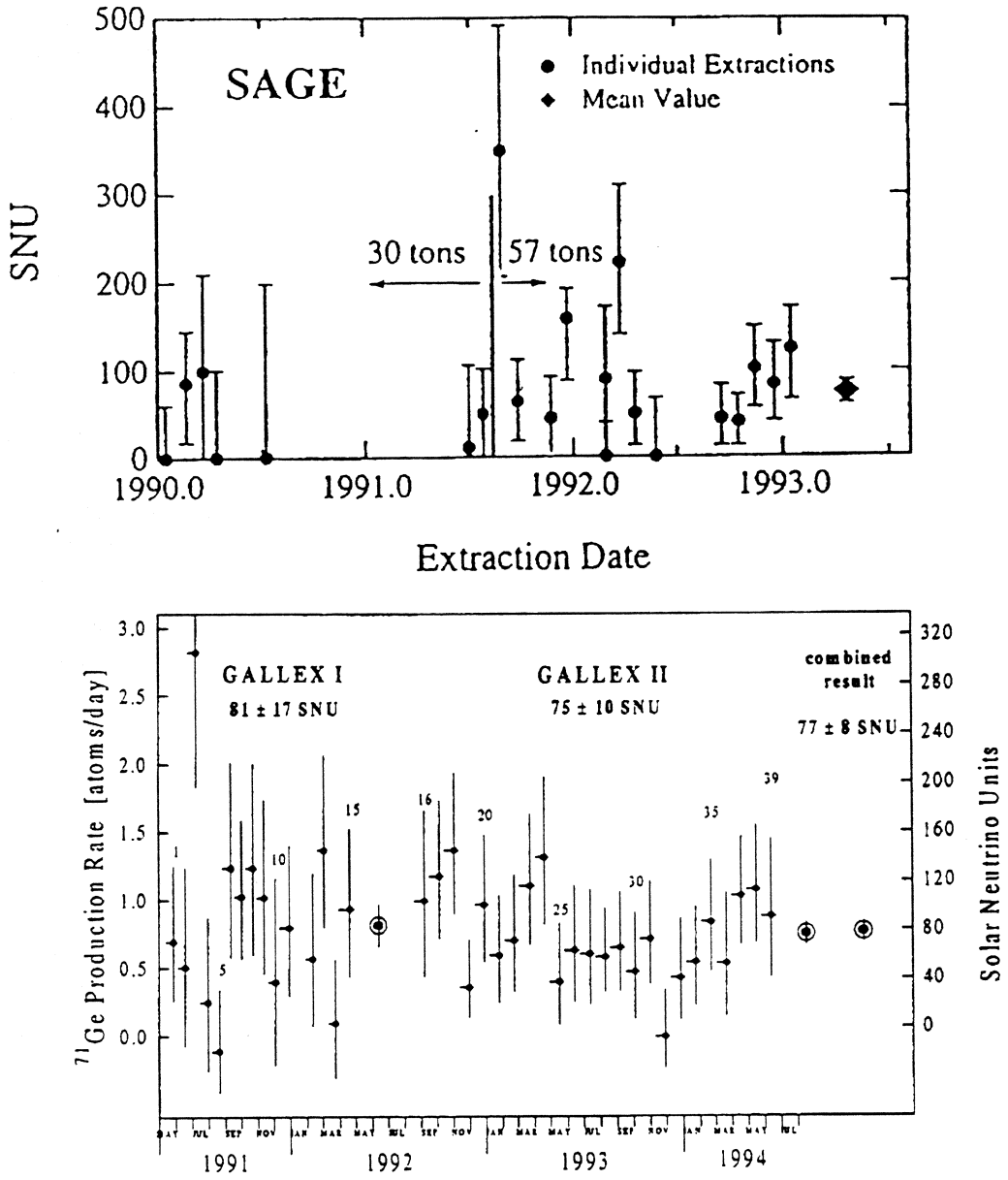


Figure 1.7: (Top) The solar neutrino capture rate inferred from each extraction of the SAGE experiment and the combined result (taken from [Abdurashitov 95]). (Bottom) The same for the GALLEX experiment (taken from [Anselmann 95]).

Both groups point out that the use of a ^{51}Cr source is not strictly a calibration experiment as the various efficiencies are thought to be known to much greater accuracy than could be inferred from the uncertainties above. The agreement between experiment and prediction does, however, give great confidence that the

	Measured (1σ)	Theoretical (3σ)
Chlorine (SNU)	$2.55 \pm 0.17 \pm 0.18$	$9.3\ ^{+1.2}_{-1.4}$
Kamiokande ($\times 10^6 cm^{-2} s^{-1}$)	$2.89\ ^{+0.22}_{-0.21} \pm 0.35$	$6.62\ ^{+0.93}_{-1.13}$
SAGE (SNU)	$74\ ^{+13}_{-12}\ ^{+5}_{-7}$	$137\ ^{+8}_{-7}$
GALLEX (SNU)	$77.1 \pm 8.5\ ^{+4.4}_{-5.4}$	$137\ ^{+8}_{-7}$

Table 1.6: The results and SSM predictions of the four solar neutrino experiments.

solar neutrino results from SAGE and GALLEX are accurate. The ^{51}Cr source is discussed further in Chapter 3.

The neutrino capture rate is extracted from the proportional counter pulse rate by fitting it to a flat background and a component decaying away with the 11.4 day half life of ^{71}Ge. The results for each extraction and the combined result for SAGE and GALLEX are shown in Fig. 1.7. These combined results [Abdurashitov 95, Anselmann 95] and the SSM prediction [Bahcall 95] are

$$\text{SAGE} = 74\ ^{+13}_{-12}(\text{stat})\ ^{+5}_{-7}(\text{syst})\ \text{SNU}\ (1\sigma)$$

$$\text{GALLEX} = 77.1 \pm 8.5(\text{stat})\ ^{+4.4}_{-5.4}(\text{syst})\ \text{SNU}\ (1\sigma)$$

$$\text{Theory} = 137\ ^{+8}_{-7}\ \text{SNU}\ (3\sigma)$$

Once again there is a clear deficit in the measured solar neutrino flux.

1.4 The Problem

In Table 1.6 the experimental results and Standard Solar Model predictions for the four experiments are brought together. The discrepancy is clear and constitutes the Solar Neutrino Problem. In recent years, with the improved accuracy of the gallium experiments, a much more powerful statement of the problem has been made possible. This analysis has been carried out by a number of authors [Castellani 94a, Castellani 94b, Bahcall 94], but the argument of [Parke 95] is repeated below as it is particularly clear and simple. The following assumptions are made:

1. The p-p chain is the dominant energy source of the sun.

2. The solar luminosity will be approximately constant over the next few million years (i.e. the sun is in quasi-equilibrium).

3. The neutrino cross-sections for the three types of experiments are accurate.

4. The neutrinos are unaffected in the passage from sun to earth.

Note that beyond the p-p chain and luminosity assumptions there is no reliance on a detailed solar model.

The solar neutrino flux and the experimental results are dominated by the pp, ^{7}Be, and ^{8}B neutrinos. Considering only these contributions three different terminations of the pp chain are possible

$$\begin{aligned} 4p + 2e^- &\rightarrow {}^4\mathrm{He} + 2\nu_e^{pp} \\ &\rightarrow {}^4\mathrm{He} + \nu_e^{pp} + \nu_e^{^7\mathrm{Be}} \\ &\rightarrow {}^4\mathrm{He} + \nu_e^{pp} + \nu_e^{^8\mathrm{B}} \end{aligned}$$

(Including the pep, hep, and CNO neutrinos only serves to strengthen the argument at the expense of making it more complex.) The total energy release in these reactions is 26.731 MeV, but the ν_e^{pp}, $\nu_e^{^7\mathrm{Be}}$, and $\nu_e^{^8\mathrm{B}}$ carry off on average 0.265, 0.861, and 7 MeV respectively. The contribution to the photon luminosity of the three reactions is therefore 26.2, 25.6, and 19.5 MeV. The solar luminosity $L_\odot$ can therefore be written in terms of the three neutrino fluxes Φ^i

$$L_\odot = 13.1\left(\Phi^{pp} - \Phi^{^7\mathrm{Be}} - \Phi^{^8\mathrm{B}}\right) + 25.6\Phi^{^7\mathrm{Be}} + 19.5\Phi^{^8\mathrm{B}} \tag{1.5}$$

It is convenient to normalise the neutrino fluxes to those of a Standard Solar Model and Parke uses the results of [Bahcall 92].

$$\phi^i = \frac{\Phi^i}{\Phi^i_{\mathrm{SSM}}}$$

In these normalised units the luminosity constraint of Eqn. 1.5 becomes

$$1 = 0.913\phi^{pp} + 0.071\phi^{^7\mathrm{Be}} + 4 \times 10^{-5}\phi^{^8\mathrm{B}} \tag{1.6}$$

Using the neutrino interaction cross-sections and the assumption of unaffected neutrino passage, the contributions to the expected event rates in the chlorine, water, and gallium experiments from the pp, ^{7}Be, and ^{8}B neutrinos can be written in the normalised units as

$$\begin{aligned} S^{th}_{\mathrm{Cl}} &= 1.2\phi^{^7\mathrm{Be}} + 6.2\phi^{^8\mathrm{B}} \ \ (\mathrm{SNU}) \\ S^{th}_{\mathrm{H_2O}} &= \Phi^{^8\mathrm{B}}_{\mathrm{SSM}}\phi^{^8\mathrm{B}} \\ S^{th}_{\mathrm{Ga}} &= 71\phi^{pp} + 36\phi^{^7\mathrm{Be}} + 14\phi^{^8\mathrm{B}} \ \ (\mathrm{SNU}) \end{aligned}$$

The uncertainty in the interaction cross-sections and hence in the above equations is estimated to be a few percent. The luminosity constraint of Eqn. 1.6 can be

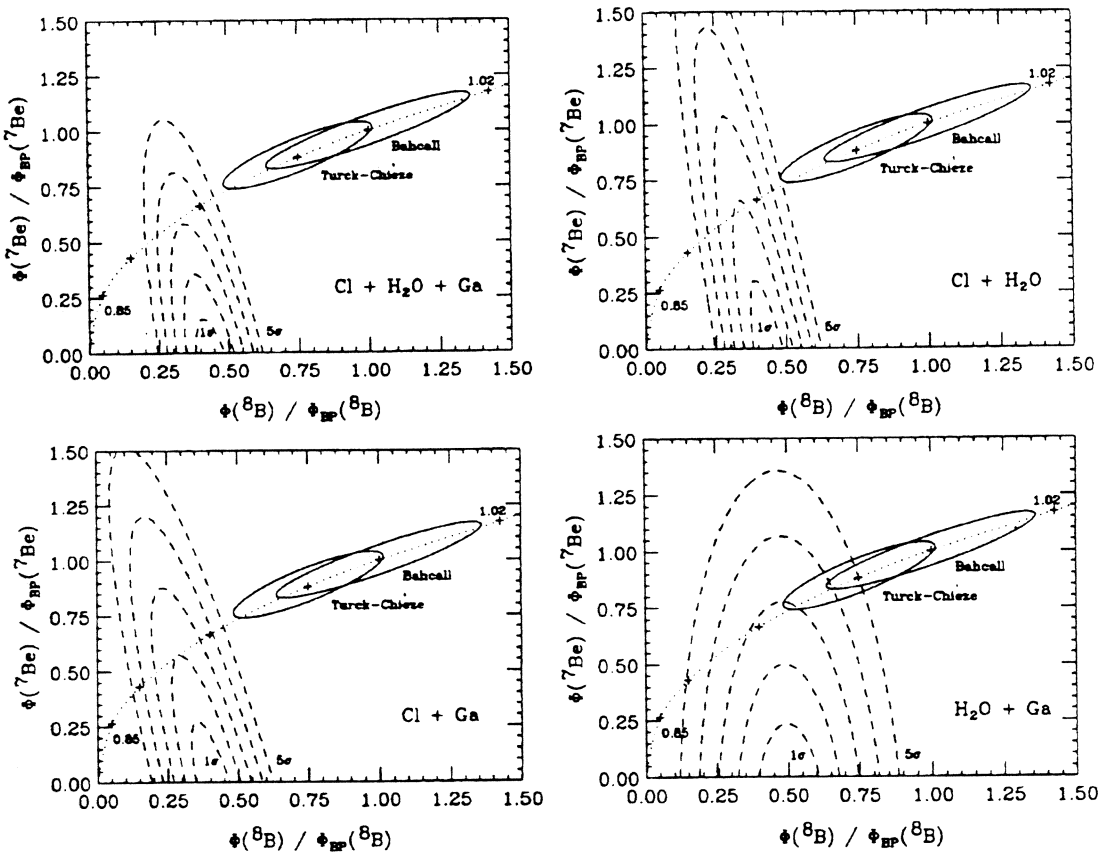


Figure 1.8: The results of fitting the ^{7}Be and ^{8}B neutrino fluxes using (Top Left) all three experimental results, (Top Right) only the chlorine and water results, (Bottom Left) only the chlorine and gallium results, and (Bottom Right) only the water and gallium results. See the text for details (taken from [Parke 95]).

used to eliminate ϕ^{pp} from the gallium result and so all three experimental rate predictions can be written purely in terms of the ^{7}Be and ^{8}B neutrino fluxes

$$S^{th}_{\rm Cl} = 1.2\phi^{^7{\rm Be}} + 6.2\phi^{^8{\rm B}} \ \ ({\rm SNU}) \tag{1.7}$$

$$S^{th}_{\rm H_2O} = \Phi^{^8{\rm B}}_{\rm SSM}\phi^{^8{\rm B}} \tag{1.8}$$

$$S^{th}_{\rm Ga} = 78 + 30\phi^{^7{\rm Be}} + 14\phi^{^8{\rm B}} \ \ ({\rm SNU}) \tag{1.9}$$

Parke uses the results available in late 1994 for the four experiments [Suzuki 95, Cleveland 95, Kirsten 95, Nico 94]. Condensing the SAGE and GALLEX results into one gallium result and combining statistical and systematic errors in

quadrature gives the experimental rates

$$S_{\mathrm{Cl}}^{ex} = 2.55 \pm 0.25 \ \text{(SNU)} \tag{1.10}$$

$$S_{\mathrm{H_2O}}^{ex} = (0.51 \pm 0.072)\Phi_{\mathrm{SSM}}^{^8\mathrm{B}} \tag{1.11}$$

$$S_{\mathrm{Ga}}^{ex} = 74 \pm 9.5 \ \text{(SNU)} \tag{1.12}$$

By equating these results with the theoretical expressions of Eqns. 1.7 to 1.9 the two fluxes $\phi^{^7\mathrm{Be}}$ and $\phi^{^8\mathrm{B}}$ may be fitted. The fit is carried out in four different ways: using all three results together, and using the three different pairs of results. Since the best fit value for $\phi^{^7\mathrm{Be}}$ turns out to be negative in all four fits it is constrained to be positive (i.e. physically sensible).

The results of the four fits are shown in Fig. 1.8 and are presented as contour plots of χ^2 as a function of $\phi^{^7\mathrm{Be}}$ and $\phi^{^8\mathrm{B}}$. The 1σ through to 5σ contours are shown as are the predictions of two Standard Solar Models [Bahcall 92, Turck-Chièze 93] with their effective 3σ bounds. Also shown as a dotted line is the prediction of an *ad hoc* solar 'model' where the core temperature is a free parameter. The crosses on the line indicate values of the temperature which are 0.85, 0.90, 0.95, 0.984, 1.00, and 1.02 times the prediction of [Bahcall 92].

The two Standard Solar Models are rejected at the many σ level and even the model where the solar temperature is free gets excluded to at least the 3σ level. The results are, of course, more dramatic when all three experiments are included but the conclusions change little when the restriction to any pair is enforced. The simple argument of this section produces a much more powerful statement of the Solar Neutrino Problem than is embodied in the separate experimental results and predictions of Table 1.6.

1.5 The Solution ?

Fig. 1.8 of the previous section leads to the inevitable conclusion that at least one of the four assumptions underpinning the derivation must be wrong.

The first assumption states that the pp chain is the dominant solar energy generating mechanism. Rejecting this most fundamental of assumptions would be almost entirely unjustifiable. The Standard Solar Model reproduces every measurable solar quantity, beside the neutrino fluxes, to a considerable accuracy. Most notably the sun's oscillation mode frequencies for which the latest model [Bahcall 95] predicts values that are in exceptional agreement with the helioseismological data [Dziembowski 94, Guenther 93]. The helium surface mass fraction ratio is also predicted by solar models and the latest value of 0.247 [Bahcall 95] agrees with the value of 0.242 $\pm$ 0.003 inferred from helioseismology [Hernandez 94] (excluding element diffusion from the model gives a predicted helium surface mass fraction of 0.268, in significant disagreement with experiment

[Bahcall 95]). No plausible solar model has yet been constructed which correctly predicts solar observables and does not rely on the pp chain as the major energy generating mechanism.

The second assumption relies on the sun being in quasi-equilibrium so that its luminosity is constant over the time-scale of a few million years. The assumption is required because the argument of the previous section ties the current solar luminosity to the current neutrino fluxes and the time-scale for thermal energy diffusion from the core to the surface of the sun is $\sim 10^7$ years whereas the neutrinos traverse the sun in a few seconds. Once again constructing a solar model which is not in main sequence equilibrium involves stretching the limits of plausibility and furthermore changes in solar luminosity would have had an impact on terrestrial life, not something that is observed in the geological or fossil record.

Until the recent improvement in the statistical uncertainties of the gallium and water experiments one method of solving the Solar Neutrino Problem was to construct a solar model with a lower than standard core temperature. This makes the Coulomb barrier penetration of the p-p chain fusion reactions more difficult and so lowers the reaction rates and neutrino fluxes. Such models had to rely on non-standard physics and for this reason alone were disfavoured. Methods of lowering the temperature included reducing the opacity by altering the elemental abundances, using the effects of strong magnetic fields or rapid rotation to alter the equation for hydrostatic equilibrium, introducing turbulent diffusion and thermal instabilities into the solar core, or using WIMPs (Weakly Interacting Massive Particles) as an additional mechanism for removing energy from the centre of the sun (see [Bahcall 89] for a review of such Non-standard Solar Models). However, as Fig. 1.8 shows, even if one is allowed complete freedom to set the core temperature of the sun it is still not possible to solve the Solar Neutrino Problem.

The third assumption states that the cross-sections of the detection reactions of the four experiments are accurately known. This is a diplomatic way of saying that their final results and uncertainties are correct and not subject to some unknown systematic error. Fig. 1.8 makes it very clear that to solve the problem it is necessary to throw out not one, but two experimental results. All four experiments are ground-breaking and exceptionally difficult, but have been subject to exhaustive checks and cross checks and it is unlikely that two of the four results are wrong.

If one can trust the first three assumptions then it is necessary to look to the fourth for a solution to the Solar Neutrino Problem. This fourth assumption requires that nothing happen to the neutrinos in their passage from sun to earth and a number of mechanisms which violate this requirement and solve the problem have been proposed. If the neutrino were to decay then this could produce a deficit [Acker 94], but it requires new particles and/or interactions and is made even more unlikely by the observation of neutrinos emitted from the supernova 1987A [Hirata 88]. If the neutrino were to have a large magnetic moment then left handed neutrinos could be flipped into right handed particles by the solar magnetic

field and so become insensitive to the weak interaction [Cisnero 71]. This requires a neutrino magnetic moment which is right at the experimental upper bound and many orders of magnitude higher than the predictions of conventional models [Bahcall 89]. The most plausible way of solving the problem and the mechanism favoured by the majority of physicists involves the matter enhanced oscillation of electron neutrinos into neutrinos of other flavours (The MSW effect). These other flavours are not detectable by the current experiments and so could account for the measured deficit (Kamiokande can see the electron scattering of muon and tau neutrinos, but cannot distinguish these events from ν_e scatterings). The next section describes the simplest form of neutrino oscillation, that which may occur in vacuum. It is not itself a very likely solution to the Solar Neutrino Problem, but its description is needed for the discussion of matter enhanced oscillations which follows.

1.5.1 Vacuum Oscillations

The first suggestion that electron neutrinos may oscillate to other flavours on their way from sun to earth came from Pontecorvo [Pontecorvo 68, Gribov 69] as a development of earlier work on oscillations from neutrino to antineutrino [Pontecorvo 58b, Pontecorvo 58a].

To simplify the discussion, without altering the basic conclusions, consider two Dirac flavour states ν_e and ν_μ (a tau or even a sterile neutrino would do just as well as a muon neutrino). These two flavours are eigenstates of the weak interaction Hamiltonian, but they do not have to be eigenstates of the Hamiltonian for vacuum propagation. In which case they will be related to the vacuum mass eigenstates, ν_{1v} and ν_{2v}, by a simple unitary transformation

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = U_v \begin{pmatrix} \nu_{1v} \\ \nu_{2v} \end{pmatrix} \quad \text{where} \quad U_v \equiv \begin{pmatrix} \cos\theta_v & \sin\theta_v \\ -\sin\theta_v & \cos\theta_v \end{pmatrix} \tag{1.13}$$

where θ_v and U_v are the lepton sector analogues of the Cabibbo angle and CKM matrix of the quark sector (for two rather than three flavours). The vacuum mass eigenstates, being by definition eigenstates of the Hamiltonian for vacuum propagation, must satisfy

$$\begin{aligned} i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} \nu_{1v} \\ \nu_{2v} \end{pmatrix} &= \begin{pmatrix} E_1 & 0 \\ 0 & E_2 \end{pmatrix}\begin{pmatrix} \nu_{1v} \\ \nu_{2v} \end{pmatrix} \\ &= \frac{1}{2p}\begin{pmatrix} m_1^2 & 0 \\ 0 & m_2^2 \end{pmatrix}\begin{pmatrix} \nu_{1v} \\ \nu_{2v} \end{pmatrix} \end{aligned} \tag{1.14}$$

which is simply a Schrödinger equation where the ultra-relativistic result $E \approx m^2/2p$ has been used. p is the common momentum of the vacuum states, m_1 and m_2 are their masses. (Giving the two states different energies and a common

momentum rather than different momenta and a common energy is an arbitrary choice and both are valid approximations to the proper wave packet treatment [Kayser 81]). The solution to the wave equation above is

$$\begin{aligned} \nu_{1_V}(t) &= e^{\frac{-im_1^2 t}{2p}} \nu_{1_V}(0) \\ \nu_{2_V}(t) &= e^{\frac{-im_2^2 t}{2p}} \nu_{2_V}(0) \end{aligned} \tag{1.15}$$

If m_1 and m_2 are different then the two vacuum mass eigenstates will propagate at different speeds. This means that a neutrino created as ν_e by a weak interaction and with an initial $\cos\theta_V$ component of ν_{1_V} will have a different mix of ν_{1_V} and ν_{2_V} when it is detected some distance away and, therefore, a different probability of being detected as a ν_e. This probability is the square of the overlap between the initial and final neutrino wavefunctions, $P(\nu_e \rightarrow \nu_e) = |\langle \nu_e(t) | \nu_e(0) \rangle|^2$, and can be written as

$$\begin{aligned} P(\nu_e \rightarrow \nu_e) &= \left| (1\ 0)\, U_V^\dagger \begin{pmatrix} e^{\frac{-im_1^2 t}{2p}} & 0 \\ 0 & e^{\frac{-im_2^2 t}{2p}} \end{pmatrix} U_V \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right|^2 \\ &= 1 - \frac{1}{2}\sin^2 2\theta_V \left[1 - \cos\left(\frac{\Delta m^2 t}{4p} \right) \right] \\ &= 1 - \sin^2 2\theta_V \sin^2\left(\frac{\pi L}{L_V} \right) \end{aligned} \tag{1.16}$$

where $L_V \equiv 4\pi p/\Delta m^2$ is the vacuum oscillation length, $\Delta m^2 \equiv m_2^2 - m_1^2$ is the difference in the squared masses of the two vacuum mass states, and L is the earth-sun distance.

To solve the Solar Neutrino Problem using the ν_e flux suppression embodied in Eqn. 1.16 requires some fine tuning of the parameters θ_V and Δm^2. Table 1.6 shows that it is not sufficient to have a uniform suppression independent of energy, and to get an energy dependent suppression Δm^2 must be carefully chosen so that L_V is close to L (2×10^{-12} eV$^2 < \Delta m^2 < 4 \times 10^{-10}$ eV^2). In addition θ_V must be very large so that the overall level of suppression is sufficient. The fine tuning of Δm^2 and the value of θ_V considerably larger than the mixing angles seen in the quark sector make this solution to the problem very unappealing.

With a more general choice of $\Delta m^2 > 4 \times 10^{-10}$ eV^2 there will be many oscillation lengths between the sun and earth and all phase information will be lost (an effect also achieved by the integration over the neutrino energy spectrum). In this case $P(\nu_e \rightarrow \nu_e)$ becomes the simple classical probability for an electron neutrino to be created as a particular vacuum mass eigenstate and then for that mass state to be subsequently detected as an electron neutrino. In other words:

$$
\begin{aligned}
P(\nu_e \to \nu_e) &= \sum_i P(\nu_e(0) \to \nu_{i\mathrm{v}}(0))\; P(\nu_{i\mathrm{v}}(t) \to \nu_e(t)) \\
&= (1\;0)\begin{pmatrix} \cos^2\theta_{\mathrm{v}} & \sin^2\theta_{\mathrm{v}} \\ \sin^2\theta_{\mathrm{v}} & \cos^2\theta_{\mathrm{v}} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos^2\theta_{\mathrm{v}} & \sin^2\theta_{\mathrm{v}} \\ \sin^2\theta_{\mathrm{v}} & \cos^2\theta_{\mathrm{v}} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= 1 - \frac{1}{2}\sin^2 2\theta_{\mathrm{v}} \qquad (1.17)
\end{aligned}
$$

(The same result can be obtained by averaging the $\sin \pi L/L_{\mathrm{v}}$ term of Eqn. 1.16 over many oscillation lengths).

No matter what the choice of θ_{v}, Eqn. 1.17 cannot be used to solve the Solar Neutrino Problem, but simple modifications of it to include the effects of matter produce the MSW effect, a mechanism with considerable power and aesthetic appeal.

1.5.2 Matter Enhanced Oscillations

In 1986 Mikheyev and Smirnov, following earlier work by Wolfenstein, proposed a natural and beautiful mechanism that extended the vacuum oscillation framework to include the effect of propagation through matter [Mikheyev 86a, Mikheyev 86b, Wolfenstein 78, Wolfenstein 79]. Requiring no more new physics beyond the mixing and mass difference of the vacuum oscillation scenario, this so called MSW effect can achieve huge ν_e flux suppression over an enormous range of θ_{v} and Δm^2 in a way that is not only natural, but is actually to be expected (see [Kuo 89] for a particularly comprehensive discussion).

Remembering that the weak flavour eigenstates are related to the vacuum mass states through the unitary transformation

$$
\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = U_{\mathrm{v}} \begin{pmatrix} \nu_{1\mathrm{v}} \\ \nu_{2\mathrm{v}} \end{pmatrix} \quad \text{where} \quad U_{\mathrm{v}} = \begin{pmatrix} \cos\theta_{\mathrm{v}} & \sin\theta_{\mathrm{v}} \\ -\sin\theta_{\mathrm{v}} & \cos\theta_{\mathrm{v}} \end{pmatrix} \qquad (1.18)
$$

then the vacuum wave equation (Eqn. 1.14) can be written in the flavour basis as

$$
i\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \frac{1}{2p}\; U_{\mathrm{v}} \begin{pmatrix} m_1^2 & 0 \\ 0 & m_2^2 \end{pmatrix} U_{\mathrm{v}}^{\dagger} \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} \qquad (1.19)
$$

where a distance differential has replaced the time differential under the assumption that the neutrinos travel at close to the speed of light. Eqn. 1.19 is valid for neutrino propagation through a vacuum, but when passing through matter interaction terms must be added to the Hamiltonian on the right hand side. The neutral current affects all weak flavours equally and so results in an additional term which is proportional to the identity matrix and so has no discernible effect

on propagation. The effect of the charged current interaction on neutrino scattering, however, does need to be considered. If the sun had equal electron and muon densities then the rate of neutrino charged current scattering from these particles would be the same irrespective of whether the neutrino had an electron or muon flavour, and, once again, an ineffectual term proportional to the identity matrix would be added to the Hamiltonian. However, the sun has no significant muon density and an asymmetry in the Hamiltonian is introduced since ν_e can scatter from electrons via the charged current, but ν_μ cannot. This symmetry breaking is at the root of the MSW effect and it introduces a term $A \equiv 2\sqrt{2}G_F N_e p$ into the Hamiltonian, where N_e is the electron number density and G_F is the Fermi coupling constant. This term was first derived by [Wolfenstein 78] and alters the flavour basis wave equation to

$$i\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} = \frac{1}{2p}\left[U_{\mathrm{v}}\begin{pmatrix}m_1^2 & 0\\ 0 & m_2^2\end{pmatrix}U_{\mathrm{v}}^{\dagger} + \begin{pmatrix}A & 0\\ 0 & 0\end{pmatrix}\right]\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} \qquad (1.20)$$

Having converted to the flavour basis in order to add the matter term, A, the next step is to try to convert back to a basis in which the matrix in the square bracket on the right hand side of the wave equation 1.20 is diagonal. The resulting equation would be analogous to Eqn. 1.14 and would describe the way the matter mass eigenstates (as opposed to the vacuum mass eigenstates) propagated independently through the sun. If this conversion could be achieved then the subsequent mathematics would be directly analogous to that of vacuum oscillations. There is, however, a complexity which affects this attempted change of basis states in a very important way which will be returned to later. To make the change of basis states the matrix within square brackets on the right hand side of Eqn. 1.20 is diagonalised resulting in

$$i\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} = \frac{1}{2p}U_{\mathrm{m}}\begin{pmatrix}M_1^2 & 0\\ 0 & M_2^2\end{pmatrix}U_{\mathrm{m}}^{\dagger}\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} \qquad (1.21)$$

This defines the unitary matrix U_{m} transforming the instantaneous matter mass eigenstates into the weak flavour eigenstates, in a way analogous to Eqn. 1.13 for the vacuum mass eigenstates

$$\begin{pmatrix}\nu_e\\ \nu_\mu\end{pmatrix} = U_{\mathrm{m}}\begin{pmatrix}\nu_{1\mathrm{m}}\\ \nu_{2\mathrm{m}}\end{pmatrix} \quad \text{where} \quad U_{\mathrm{m}} \equiv \begin{pmatrix}\cos\theta_{\mathrm{m}} & \sin\theta_{\mathrm{m}}\\ -\sin\theta_{\mathrm{m}} & \cos\theta_{\mathrm{m}}\end{pmatrix} \qquad (1.22)$$

where the diagonalisation defines θ_{m} to be

$$\tan 2\theta_{\mathrm{m}} = \frac{\Delta m^2 \sin 2\theta_{\mathrm{v}}}{\Delta m^2 \cos 2\theta_{\mathrm{v}} - A} \qquad (1.23)$$

and the instantaneous matter mass eigenvalues as

$$M_{2,1}^2 = \frac{1}{2}\left[(\Sigma m^2 + A) \pm \left((A - \Delta m^2 \cos 2\theta_{\mathrm{v}})^2 + (\Delta m^2 \sin 2\theta_{\mathrm{v}})^2\right)^{\frac{1}{2}}\right] \qquad (1.24)$$

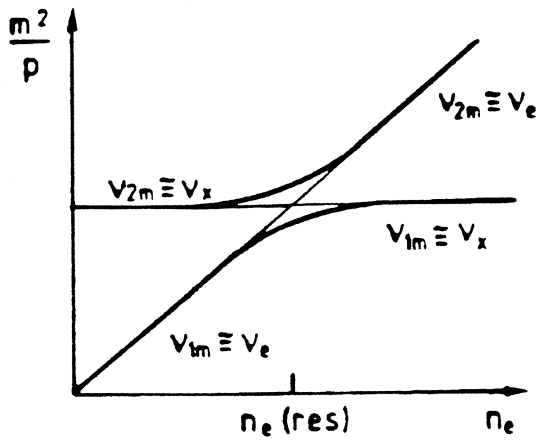


Figure 1.9: The two matter masses M_1 and M_2 as a function of local electron density are shown as the lower and upper solid curves respectively.

where $\Sigma m^2 \equiv m_1^2 + m_2^2$.

In rearranging Eqn. 1.21 to form a wave equation for the instantaneous matter mass eigenstates $\nu_{1\mathrm{m}}$ and $\nu_{2\mathrm{m}}$, analogous to Eqn. 1.14 for the vacuum mass eigenstates, the complexity referred to earlier reveals itself. It is not possible to bring U_m over to the left hand side of the equation and inside the differential as usual, because the matter mixing angle θ_m depends upon the electron density through A, and this density varies with position in the sun. In other words θ_m is not independent of x. Rearranging Eqn. 1.21 therefore produces

$$
\begin{aligned}
i\frac{\mathrm{d}}{\mathrm{d}x}\begin{pmatrix}\nu_{1\mathrm{m}}\\ \nu_{2\mathrm{m}}\end{pmatrix} &= \left[\frac{1}{2p}\begin{pmatrix}M_1^2 & 0\\ 0 & M_2^2\end{pmatrix} - U_\mathrm{m}^\dagger i\frac{\mathrm{d}}{\mathrm{dx}}U_\mathrm{m}\right]\begin{pmatrix}\nu_{1\mathrm{m}}\\ \nu_{2\mathrm{m}}\end{pmatrix} \\
&= \frac{1}{2p}\begin{pmatrix}M_1^2 & -2ip\dfrac{\mathrm{d}\theta_\mathrm{m}}{\mathrm{d}x}\\ 2ip\dfrac{\mathrm{d}\theta_\mathrm{m}}{\mathrm{d}x} & M_2^2\end{pmatrix}\begin{pmatrix}\nu_{1\mathrm{m}}\\ \nu_{2\mathrm{m}}\end{pmatrix} \qquad (1.25)
\end{aligned}
$$

where the last step follows directly from the substitution of the explicit matrix U_m as defined by Eqn. 1.22.

The position dependence of θ_m ensures that the wave equation for the matter mass states is not diagonal and for this reason $\nu_{1\mathrm{m}}$ and $\nu_{2\mathrm{m}}$ are not true eigenstates of the matter Hamiltonian, but are referred to as instantaneous eigenstates. Since the two matter masses M_1 and M_2 depend on A they vary with the local electron density. The two solid curves of Fig. 1.9 show M_1 and M_2 as a function of the local electron density N_e.

For the high electron densities at the core of the sun where the neutrinos are

produced the A term will be very large. For a wide range of θ_v and Δm^2 choices it will dominate Eqn. 1.23 and make $\theta_m \approx \pi/2$. Eqn. 1.22 shows that in that case the electron neutrino is almost entirely in the ν_{2m} state on creation. If the electron density changes sufficiently slowly that the off-diagonal elements of Eqn. 1.25 can be neglected then ν_{1m} and ν_{2m} are genuine eigenstates of the matter Hamiltonian and the neutrino initially created in the ν_{2m} state will stay in that state as it propagates through the sun. This propagation is through ever diminishing electron density and so the neutrino, starting on the upper curve of Fig. 1.9 and on the right hand side, moves to the left down the curve as N_e decreases. As it moves the rules are being changed: the decreasing N_e progressively lowers θ_m through Eqn. 1.23 and the ν_{2m} state, initially pure ν_e, picks up an ever increasing component of ν_μ until, on exiting the sun, $N_e = 0$ and $\theta_m = \theta_v$, giving the neutrino an almost complete muon flavour (for a small vacuum mixing angle). In this way conversion of electron into muon neutrinos can occur.

The constraint that the off-diagonal elements of Eqn. 1.25 are negligible is known as the adiabatic condition. As the neutrino moves through the electron density of the sun $d\theta_m/dx$ will be greatest, and the adiabatic condition most likely to breakdown, when $\theta_m = \pi/4$. This is known as the resonance point, for here the electron and muon neutrino states are maximally mixed combinations of ν_{1m} and ν_{2m}. Looking at Eqn. 1.23 it can be seen that the resonance occurs when $A = \Delta m^2 \cos 2\theta_v$. By explicitly doing the $d\theta_m/dx$ differentiation at the resonance the adiabatic condition can be expressed, using the adiabaticity parameter γ, as

$$\gamma \equiv \frac{\Delta m^2 \sin^2 2\theta_v}{2p \cos 2\theta_v \left| \frac{1}{N_e} \frac{dN_e}{dx} \right|_{res}} \gg 1 \qquad (1.26)$$

If the adiabaticity constraint of Eqn. 1.26 is not satisfied then the off-diagonal elements of the wave equation (Eqn. 1.25) bring about mixing between ν_{1m} and ν_{2m} at the resonance point and there will be a probability P_c for the ν_{2m} state to jump or cross to being a ν_{1m}. Another way to view this level crossing is to note that at the resonance the square root in Eqn. 1.24 that defines the matter masses of the two states ν_{1m} and ν_{2m} is at its minimum and the two masses are separated by a very small amount. The breakdown of the adiabatic constraint corresponds to an inability of the neutrino wavefunction to 'keep up' with its changing surroundings and this can induce something akin to a quantum mechanical tunnelling between the two matter states when their masses are very close. The crossing probability P_c is not a simple quantity to calculate, but approximations can be made. One particularly simple expression was introduced by [Parke 86], drawing on earlier work by [Zener 32] and others on level crossing in atomic collisions.

$$P_c = e^{-\frac{\pi}{2}\gamma} \qquad (1.27)$$

The effect of the level crossing is to counteract that of the adiabatic transition since the neutrinos which cross to the ν_{1m} state emerge from the sun with an almost entirely electron flavour.

The suppression probability $P(\nu_e \rightarrow \nu_e)$ for the MSW effect can be calculated in much the same way as the vacuum expression of Eqn. 1.17, but with the addition of the level crossing probability. Assuming again that the oscillations are such as to lose any phase information then $P(\nu_e \rightarrow \nu_e)$ is the probability that the initial electron neutrino is created in a particular matter mass state, combined with the probability P_c that this mass state jumps to the other mass state, and finally the probability that the emerging, zero electron density, mass state is detected as an electron neutrino.

$$\begin{aligned} P(\nu_e \rightarrow \nu_e) &= \sum_i P(\nu_e(0) \rightarrow \nu_{im}(0))\, P(\text{crossing})\, P(\nu_{im}(t) \rightarrow \nu_e(t)) \\ &= (1\ 0) \begin{pmatrix} \cos^2\theta_v & \sin^2\theta_v \\ \sin^2\theta_v & \cos^2\theta_v \end{pmatrix} \begin{pmatrix} (1-P_c) & P_c \\ P_c & (1-P_c) \end{pmatrix} \begin{pmatrix} \cos^2\theta_m & \sin^2\theta_m \\ \sin^2\theta_m & \cos^2\theta_m \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{1}{2} + \left(\frac{1}{2} - P_c\right) \cos 2\theta_v \cos 2\theta_m \end{aligned} \tag{1.28}$$

where the value of θ_m should be taken at the neutrino creation point.

Each of the three terms P_c, $\cos 2\theta_v$, and $\cos 2\theta_m$ can be expressed as a function of $p/\Delta m^2$ and θ_v. Fig. 1.10 shows $P(\nu_e \rightarrow \nu_e)$ as a function of $p/\Delta m^2$ for two different values of $\sin^2 2\theta_v$. Its form can be understood fairly simply. If $p/\Delta m^2$ for a neutrino is smaller than the critical value (5 MeV / 10^{-4} eV^2) then the resonant electron density $N_{e\text{res}}$ is greater than the electron density at the centre of the sun. This means that θ_m is less than $\pi/4$ at the creation point of the electron neutrino and it will be predominantly in the ν_{1m} matter state, staying in this state on exiting the sun, and being detected in an almost entirely electron neutrino state. The survival probability in this case will be very close to one. For larger values of $p/\Delta m^2$ the survival probability plummets as the electron neutrinos are created predominantly as ν_{2m} and will almost entirely convert adiabatically to muon neutrinos. Moving higher still in $p/\Delta m^2$ the adiabatic condition of Eqn. 1.26 starts to break down and level crossing from ν_{2m} to ν_{1m} becomes important, counteracting the adiabatic transition and steadily increasing the ν_e survival probability. At very large values of $p/\Delta m^2$ the non-adiabatic level crossing completely dominates and the survival probability returns to unity. The effect of increasing $\sin^2 2\theta_v$ is to delay the breakdown of the adiabatic condition (see Eqn. 1.26) and so widen the suppression trough, as can be seen from the two curves in Fig. 1.10. The depth of the trough is determined by the amount of vacuum mixing and so increasing $\sin^2 2\theta_v$ also has the effect of making the trough some shallower.

Using the MSW survival probability there are three distinct ways the Solar Neutrino Problem might be solved. These can be related to positions on Fig. 1.10 as well as to which of the three multiplied terms in Eqn. 1.28 dominates. For a particular choice of Δm^2 the solar neutrino energy spectrum occupies a band in $p/\Delta m^2$. By changing Δm^2 this band moves horizontally across Fig.

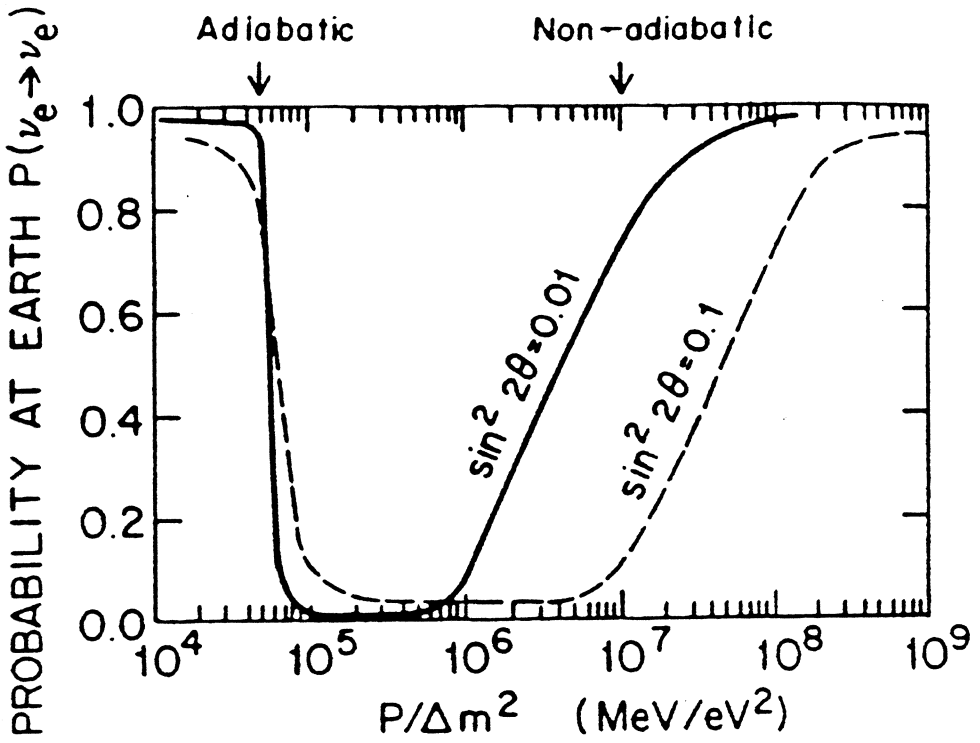


Figure 1.10: The probability $P(\nu_e \rightarrow \nu_e)$ that an electron neutrino created at the centre of the sun is detected as an electron neutrino as a function of $p/\Delta m^2$ for two different values of $\sin^2 2\theta_v$.

1.10. When the band straddles the adiabatic threshold, the term $\cos 2\theta_m$ of Eqn. 1.28 dominates, and a strongly energy dependent suppression can be achieved. Similarly when the band straddles the non-adiabatic rise of Fig. 1.10, and the P_c term of Eqn. 1.28 dominates, a strong energy dependent suppression can again be brought about. The third way of solving the problem is to have a large vacuum mixing angle and the band sitting entirely within the trough of Fig. 1.10. In this case the $\cos 2\theta_v$ term of Eqn. 1.28 dominates and a large, but almost energy independent suppression arises. These three situations are known as the adiabatic, the non-adiabatic, and the large angle solutions respectively.

The usual way to present the suppressions of different MSW scenarios is to plot experimental iso-rate contours on the $\Delta m^2 - \sin^2 2\theta_v$ plane. Each experiment will have its own set of contours as they each have different detection cross-sections and thresholds. Fig. 1.11 shows a pair of such contours for each different type of experiment. Given SSM neutrino fluxes, values of Δm^2 and $\sin^2 2\theta_v$ that fall between the two contours produce a suppression that accounts for the experimental result in each of the three cases. The triangular shape of the contours arises

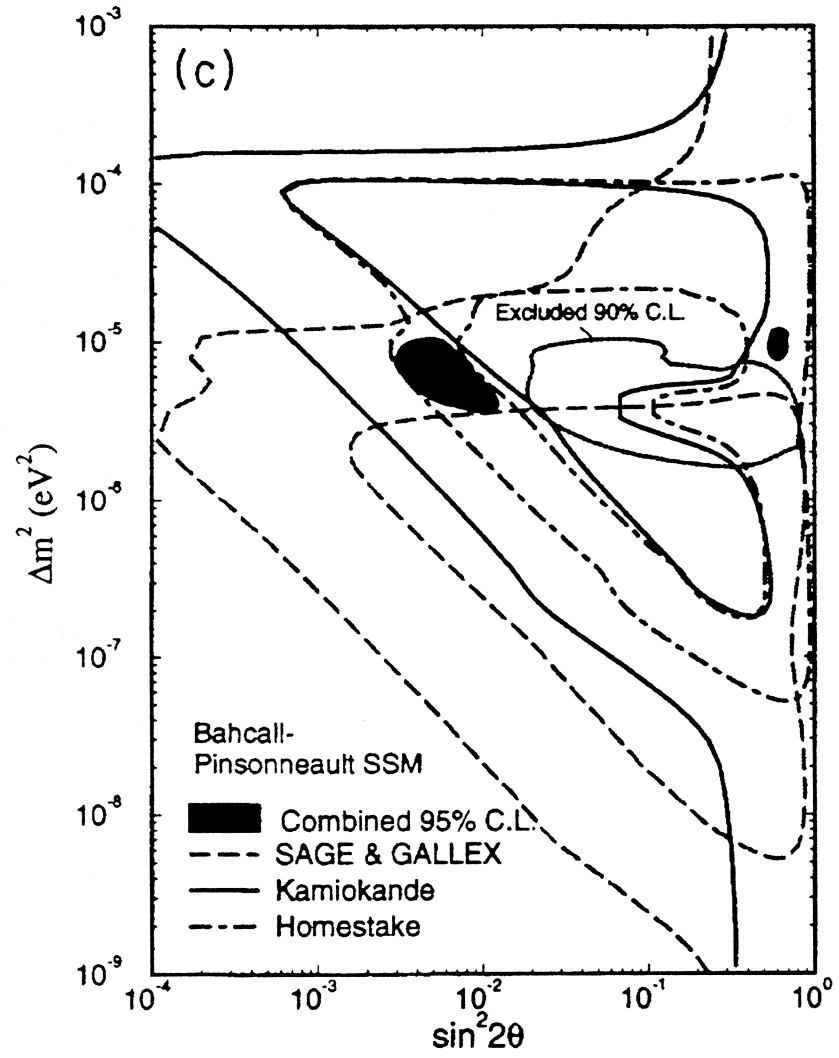


Figure 1.11: The MSW exclusion plot, showing the areas of $\Delta m^2 - \sin^2 2\theta_\nu$ space allowed by each of the three different types of experimental results and the area allow by the combined result. Also shown is the area excluded by the non-observation of a day-night effect in Kamiokande II (taken from [Hata 94]).

from the three different ways of producing a suppression that were listed earlier. An adiabatic solution is essentially independent of the vacuum mixing angle since Δm^2 determines how much of the neutrino energy spectrum is on either side of the adiabatic transition of Fig. 1.10. This produces the horizontal upper contour of the triangle. The large angle solution is basically dependent on the value of θ_ν alone and so produces the vertical contour of the triangle. The non-adiabatic solution depends on Δm^2 to position the neutrino energy band across the non-

adiabatic rise of Fig. 1.10, but it also depends on θ_v to determine the location and shape of the rise. This dependence is made explicit by looking at Eqns. 1.26 and 1.27. A fixed crossing probability relates Δm^2 and θ_v by fixing γ. Integrating out the energy dependence, and assuming that the vacuum mixing angle is small enough to make $\cos 2\theta_v \approx 1$, the fixed γ makes $\Delta m^2 \sin^2 2\theta_v \approx$ const i.e. a straight diagonal line on a log-log plot of these two quantities. In this way the approximately diagonal side of the triangle is formed.

Also shown in solid black in Fig. 1.11 are the two regions of the plane that can account for all the experimental data. The larger, and more likely solution, is non-adiabatic for the high energy neutrinos (adiabatic for the low energy) with the small, all but excluded, area being a large angle solution. Also shown is a region of the parameter space excluded by the fact that Kamiokande do not see a day-night effect. This effect can occur when the MSW process converts muon back into electron neutrinos as they pass through the material of the earth. This produces a difference between the day and night suppressions, giving the effect its name. One final point should be made that addresses a common misconception. The small size of the allowed parameter space of Fig. 1.11 is not a result of any need to fine tune parameters *a priori*, but simply expresses the fact that the parameters are well constrained by experiment.

1.6 The Future

Although the existing four solar neutrino experiments have contributed greatly to our understanding of the sun, they are not sufficient to answer the Solar Neutrino Problem, or to probe much more deeply into solar mechanisms. The next generation of experiments needs to have higher count rates, better energy resolution, and a sensitivity to neutral current processes to probe the total rather than just the ν_e flux. Over the next few years three experiments that together have these properties will be important; Superkamiokande, Borexino, and the Sudbury Neutrino Observatory.

The Superkamiokande experiment [Suzuki 94], which has just begun taking data (April 1996), is essentially a large version of Kamiokande with a total of 50,000 tonnes of water, reduced to a fiducial mass of 22,000 tonnes. Like its predecessor it is situated in the Kamioka mine in Japan, but is housed in a much larger cavity; a cylinder 41 metres in height and 39 metres in diameter. The water detecting volume is viewed by 11,200 50cm diameter PMTs which have much better timing characteristics and radioactive purity than those used in the earlier experiment and the detector as a whole will have much lower radioactive contamination. This is planned to enable the threshold to be lowered from ~7 to 5 MeV, with the improved PMT timing enabling a better rejection of cosmic spallation followers. The larger volume and lower threshold increases the SSM event rate to ~8,400 yr^{-1} and will allow Superkamiokande to make a much better measure-

ment of the electron energy spectrum that results from the electron scattering of ^{8}B neutrinos.

The Borexino detector [Alimonti 93], which will be built at the Gran Sasso Laboratory in Italy, consists of 300 tonnes of liquid scintillator surrounded by ~4.5m of water shielding and viewed by 1700 20cm PMTs. The aim is to observe the electron scattering of ^{7}Be neutrinos and the low energy of this flux (0.862 and 0.384 MeV) makes it necessary to use a scintillator rather than the relatively low efficiency Čerenkov process. The rate is high (~50 day^{-1} with a threshold at 0.250 MeV), but the lower energy makes radioactive background even more of a problem than usual in solar neutrino experiments. ^{238}U and ^{232}Th at the 10^{-16} g/g concentration level are required, and ^{40}K at the 10^{-14} level. Early results from the proposed Borexino purification and assay system show that this is possible with the measured concentrations of ^{238}U and ^{232}Th being $(5.2 \pm 1.2) \times 10^{-16}$ g/g and $(2.8^{+1.7}_{-1.2}) \times 10^{-16}$ g/g respectively [Ranucci 95]. A pure measurement of the ^{7}Be flux would fill a gap left by the other solar neutrino experiments and provide a critical test of many non-standard neutrino models.

The Sudbury Neutrino Observatory (SNO) [Ewan 92] is currently under construction near Sudbury in Ontario, Canada. It is a Čerenkov detector that uses heavy water as a target thus enabling it to detect neutral as well as charged current events and measure the total and electron neutrino fluxes separately. The observation of neutrinos of non-electron flavour would be a clear signature for flavour changing neutrino oscillations. In addition the charged current neutrino capture on the deuteron makes possible an accurate measurement of the ^{8}B neutrino flux as a function of energy. The experiment is described fully in Chapter 2.

Chapter 2

The Sudbury Neutrino Observatory

The sun, though it passes through dirty places, yet remains as pure as before

FRANCIS BACON
The Advancement of Learning (1605)

Observatory. n. A place where astronomers conjecture away the guesses of their predecessors

AMBROSE BIERCE
The Devils Dictionary (1881-1911)

2.1 Overview

ARGUABLY THE best hope for a resolution to the Solar Neutrino Problem rests with the Sudbury Neutrino Observatory (SNO) [Aardsma 87, Ewan 92, SNO 87b]. This experiment will detect the interactions of neutrinos with 1000 tonnes of high purity D_2O. This heavy water is housed in an acrylic vessel of 6 m radius and surrounded by 7000 tonnes of H_2O shielding. Čerenkov light from the various interactions is detected by a sphere of photomultiplier tubes housed on a geodesic structure and located in the light water. In order to shield from cosmic rays the detector is being constructed at a depth of 2070 m in a nickel mine near Sudbury in Ontario, Canada.

In contrast to all previous solar neutrino experiments SNO has more than one detection channel. With the sensitivity to charged current events the flux of electron neutrinos can be determined as can their energy spectrum. In addition SNO's ability to detect neutral current interactions also means that the total solar neutrino flux from all the three weak flavours can be pinned down.

Channel Name	Reaction	Q-value (MeV)	σ at 10 MeV (10^{-42} cm^2)	Angular Distribution
Charged Current	$\nu_e + d \rightarrow p + p + \epsilon^-$	-1.44	2.7	$1 - \frac{1}{3}\cos\theta_{\mathrm{sun}}$
Electron Scattering	$\nu_x + \epsilon^- \rightarrow \nu_x + \epsilon^-$	zero	0.092	forward peaked
Neutral Current	$\nu_x + d \rightarrow \nu_x + p + n$	-2.22	1.1	flat - neutron is detected

Table 2.1: Neutrino interactions in the SNO detector

2.2 The Čerenkov Effect for Electrons in Water

Čerenkov radiation is emitted when a charged particle traverses a medium at a speed which is greater than the speed of light in that medium. The number of Čerenkov photons emitted per unit track length per unit wavelength is given by [Jelley 58]

$$\frac{\mathrm{d}^2N}{\mathrm{d}x\mathrm{d}\lambda} = \frac{2\pi\alpha Z^2}{\lambda^2}\left(1 - \frac{1}{n^2\beta^2}\right)$$

where α is the fine structure constant, $\beta = v/c$, n is the refractive index of the medium, and Z is the charge of the relativistic particle. The light is emitted at the Čerenkov angle, $\theta_C = \cos^{-1}(1/n\beta)$, relative to the direction of the particle. For an electron or positron in water with 5-15 MeV of kinetic energy about 220 Čerenkov photons are emitted per MeV between 250 and 600 nm wavelength and at a Čerenkov cone angle of about 42°. When the speed of the charged particle drops below that of light in the medium no Čerenkov light is emitted. For electrons in water this cutoff occurs at a kinetic energy of ~262 keV.

2.3 Neutrino Interactions

The SNO detector is sensitive to neutrinos (as opposed to anti-neutrinos) via three different interactions. These are summarised in Table 2.1 and their cross-sections as a function of neutrino energy are shown in Fig. 2.1.

2.3.1 The Charged Current Interaction

Although strictly referring to the class of weak interactions involving the $W^{\pm}$ the term *charged current interaction* is used in a SNO context to refer to the reaction

$$\nu_e + d \rightarrow p + p + e^-$$

Since the protons emerge essentially stationary from the interaction the electron is emitted with the energy of the neutrino minus the Q value of 1.44 MeV. The yield of Čerenkov photons and the consequent number of PMT hits is closely linear with the electron's initial energy so long as it is well above the Čerenkov threshold. This means that the charged current interaction can be used to measure the ν_e energy spectrum as well as the integral flux.

The charged current reaction has a pure Gamow-Teller matrix element resulting in an electron angular distribution relative to the neutrino (solar) direction given by [Bahcall 64a]

$$I(\theta_{\text{sun}}) = 1 - \frac{1}{3}\left(\frac{v}{c}\right)\cos\theta_{\text{sun}} \approx 1 - \frac{1}{3}\cos\theta_{\text{sun}}$$

where v is the initial speed of the electron. This distribution could be used to demonstrate that the interactions observed are actually due to neutrinos emitted from the sun, but it is possible that the rather weak asymmetry will be washed out by the angular resolution of the direction fit (about which more is said in Chapter 5). The charged current cross-section as a function of energy has been calculated by [Ying 92, Tatara 90, Doi 92].

2.3.2 The Electron Scattering Interaction

The electron scattering reaction is the same as that used by Kamiokande II

$$\nu_x + e^- \rightarrow \nu_x + e^-$$

It provides an unambiguous signature of solar neutrino detection as the initial direction of the electron is sharply peaked away from the sun (the half-angle of the cone containing 90% of the electrons is 14.6° [Bahcall 87]) and the effect should be obvious. Being an elastic collision the reaction has no threshold and its differential cross-section is one of the staples of Electroweak theory [t'Hooft 71].

$$\frac{d\sigma}{dE_e} = \frac{2G^2 m_e^2}{\pi}\left[g_L^2 + g_R^2\left(1 - \frac{E_e}{E_\nu}\right)^2 - g_L g_R\left(\frac{E_e}{E_\nu^2}\right)\right]$$

where $g_L = (\pm\frac{1}{2} + \sin^2\theta_W)$ and $g_R = \sin^2\theta_W$. The upper sign applies to ν_e scattering and the lower to ν_μ scattering [Weinberg 67].

The cross-section for electron neutrinos is approximately a factor of 7 larger than that for the other weak flavours as the electron flavour has a $W^{\pm}$ channel open

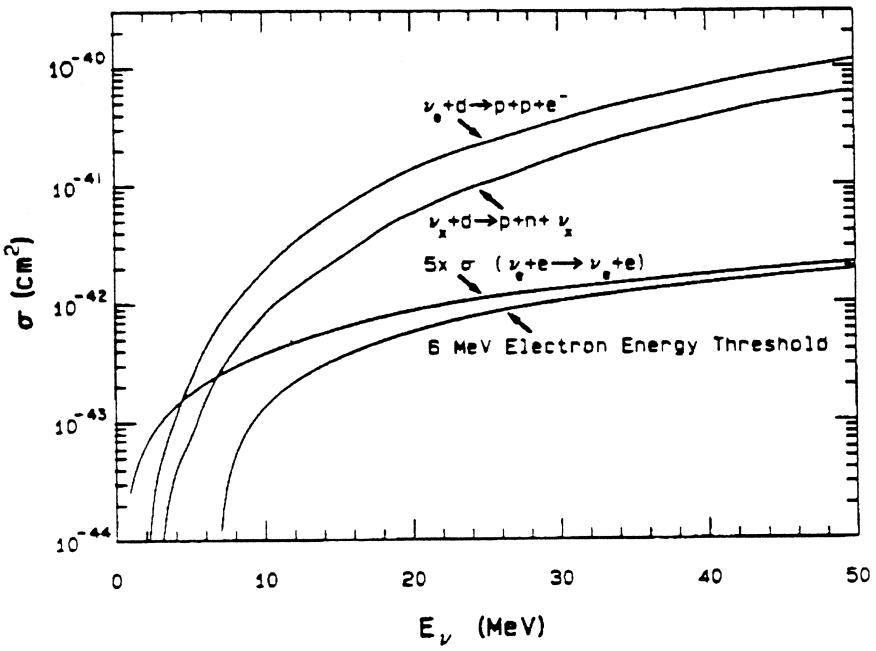


Figure 2.1: The cross-sections as function of energy of the three SNO neutrino interactions

to it as well as the Z^0 channel available to all three neutrino types. As indicated in Table 2.1 the cross-section for electron scattering is somewhat smaller than those of the other two reactions but this is somewhat compensated for by the fact that there are 5 times as many electrons as deuterons in the D_2O molecule and that electrons are also available in the light water.

2.3.3 The Neutral Current Interaction

As with the charged current reaction the phrase *neutral current interaction* is used in a narrow sense by the SNO collaboration to mean the neutrino induced breakup of the deuteron

$$\nu_x + d \rightarrow \nu_x + p + n$$

This reaction is the primary reason for the choice of D_2O as a target material. It is equally sensitive to all of the three weak neutrino flavours and so the measurement of its rate determines the total weak solar neutrino flux irrespective of vacuum or MSW oscillations (excepting those involving sterile neutrinos). The detection of the neutral current interaction depends upon the capture of the final state neutron. Two methods will be used to achieve this: capture on ^{35}Cl with the subsequent gamma cascade producing Čerenkov light via Compton scattered electrons, and

Reaction	SSM		Vacuum	MSW	
	full	$\times\frac{1}{3}$	mixing	Adiabatic	Non-adiabatic
$\nu_e + d \rightarrow p + p + e^-$	9750	3255	3260	3750	3750
$\nu_x + e^- \rightarrow \nu_x + e^-$	1100	360	480	645	488
$\nu_x + d \rightarrow \nu_x + p + n$	2800	933	2800	2800	2800

Table 2.2: Neutrino interaction rates in SNO (per kilotonne-year). A threshold equivalent to an electron energy of 5 MeV is assumed.

capture on ^{3}He within proportional counters which detect the resulting breakup into deuteron and triton. Any method of picking up the neutron almost inevitably involves its thermalisation with the loss of any direction or energy information. Cross-sections for the neutral current interaction have been recently calculated by [Ying 92, Tatara 90, Doi 92, Bahcall 88a].

2.3.4 Anti-Neutrino Interactions

In addition to the three neutrino reactions above, SNO is also sensitive to anti-neutrinos via two further interactions. It is unlikely that these will be of use when studying the solar neutrino flux, but should a galactic supernova occur vast numbers of anti-neutrinos will be emitted during its burst and cooling phases [Burrows 90]. The two anti-neutrino interactions are the inverse β decay of the proton

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

and the analogue of the charged current interaction

$$\bar{\nu}_e + d \rightarrow n + n + e^+$$

The first reaction occurs in the light water and is identified by the Čerenkov light from the positron (the neutron capture on H will probably be undetectable) and the same applies to the second reaction but with the additional detection of two neutrons in the heavy water.

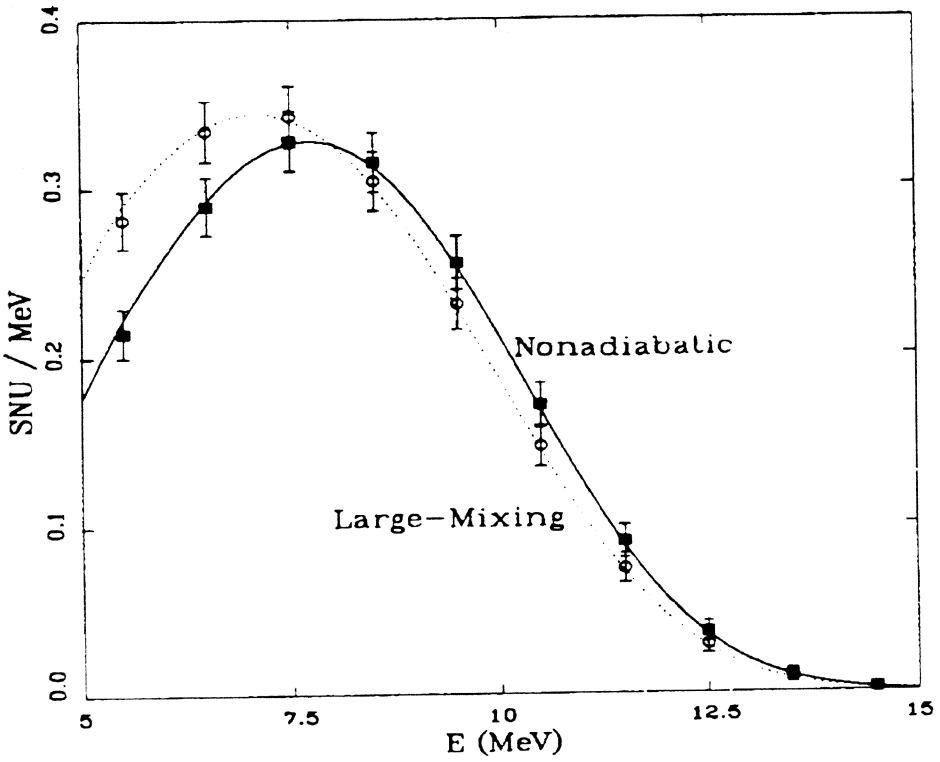


Figure 2.2: The predicted ν_e energy spectra for two favoured MSW scenarios (taken from [Bludman 92]).

2.4 Overview of Data Analysis

2.4.1 Solar Neutrinos

The annual event rate for the three solar neutrino interactions are given in Table 2.2. These rates are taken from the original SNO proposal [SNO 87b] where a threshold equivalent to an electron of 5 MeV is used. Since that time there have been improvements in the solar models used to predict the neutrino flux and more accurate calculations of the interaction cross-sections. Appendix E provides updated event rates where these improvements have been included. (It should be noted, however, that this appendix applies a different threshold, one of 60 PMT hits, suitable for the analysis of Chapter 7).

In its solar neutrino analysis SNO will principally look to answer two questions. Since the MSW and other hypothesised effects distort the neutrino spectra the first question is how does the ν_e spectrum inferred from the energy of the electrons

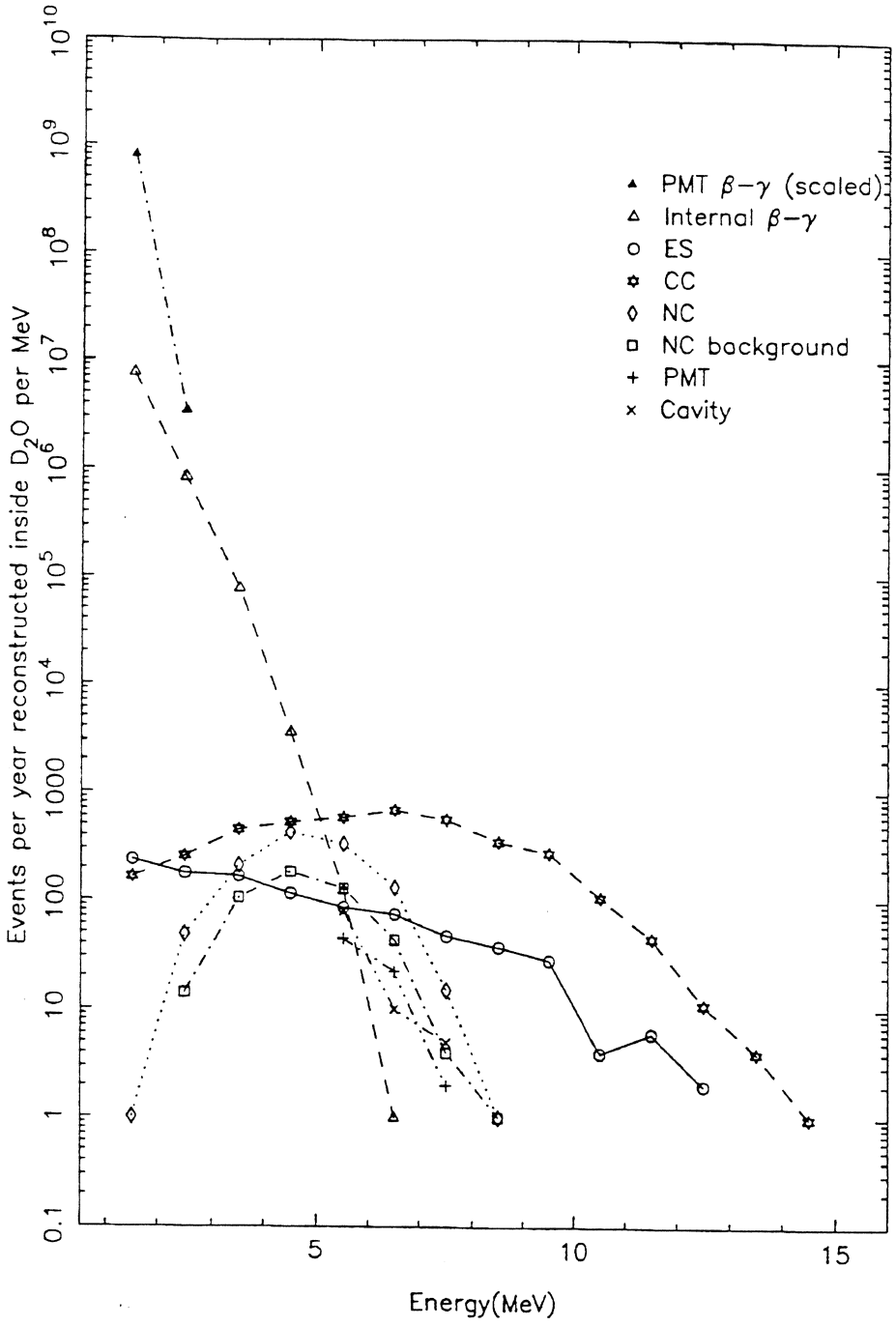


Figure 2.3: Event energy spectrum from 1 year of pure D_2O fill (taken from [Skensved 91]).

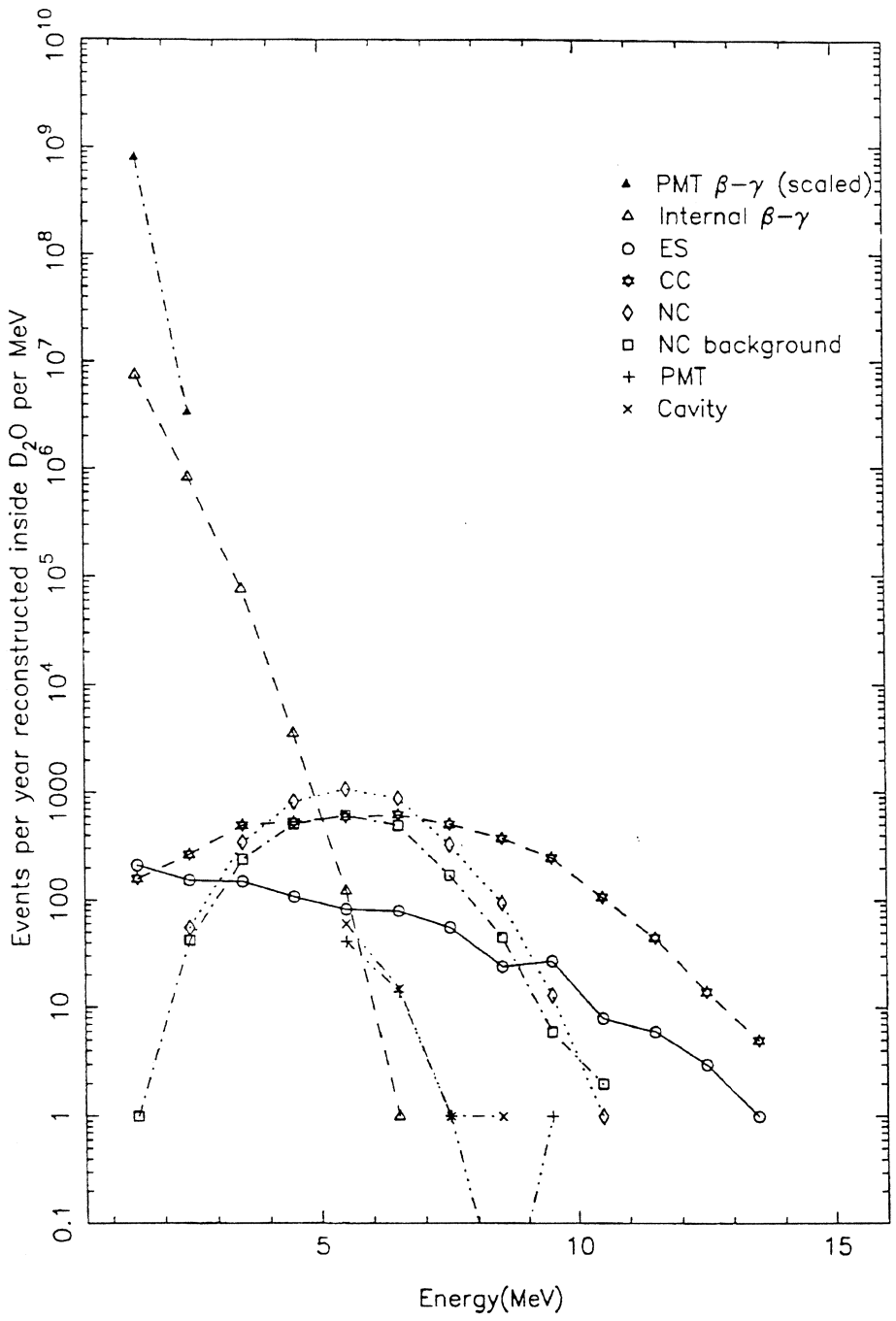


Figure 2.4: Event energy spectrum from 1 year of pure D_2O plus salt fill (taken from [Skensved 91]).

from charged current events compare with that expected from ^{8}B decay? The predicted SNO energy spectrum broken into its various components for an acrylic vessel filled with pure D_2O is shown in Fig. 2.3 and the predicted ν_e energy spectra for two MSW scenarios is illustrated in Fig. 2.2 [Bludman 92]. Below ~5 MeV the contribution to the SNO spectrum from radioactive contaminants dominates completely and above ~7.5 MeV the spectral distortion of any likely MSW effect becomes negligible. This leaves a window between ~5 and ~7.5 MeV where any distortion might be observed. Unfortunately Fig. 2.3 shows that this region is just where the neutral current event rate is high. In pure D_2O these neutral current events stem from the neutron being captured on a deuteron with the subsequent emission of a 6.25 MeV γ. It will be possible to subtract this component using a separate measurement of the neutral current event rate (see later) combined with knowledge of the neutron production rate from other sources, but this may result in a statistically weak statement of any distortion in the remaining charged current spectrum. A better procedure would be to remove the neutrons altogether by doping the D_2O with an isotope of high neutron capture cross-section and no subsequent γ or β decay. An ideal candidate for this poisoning is ^{6}Li and calculations indicate that only ~10 kg in the form of $LiNO_4$ need be added [Moorhead 95b]. The removal of all neutrons in this way has one very sizeable drawback: if a supernova were to occur during a poison run the detector would have no sensitivity to neutral current events, an eventuality that no-one wishes to contemplate.

The second question to be addressed by SNO concerns the number of neutral and charged current events. From the charged current rate the electron neutrino flux can be inferred and the total neutrino flux can be obtained from the neutral current rate. Since the sun only produces electron neutrinos in any significant quantity these two event rates provide a solar model independent way of determining whether the MSW (or vacuum) oscillation scenario is the solution to the solar neutrino problem (they also address other hypotheses, but in a somewhat more model dependent way). With the total and electron neutrino fluxes pinned down by the charged and neutral current interaction rates it is then possible to make a prediction for how many electron scattering events SNO should see and to compare this with how many are actually seen. This will probably constitute no more than a consistency check as the number of electron scattering events will be low and only a small fraction will be due to muon or tau neutrinos. The charged and neutral current cross-sections are known to 10%, but their ratio is known to 0.5%. As a consequence SNO will primarily look at the ratio of the neutral and charged current rates. This involves some loss of information, but a sizable gain in accuracy. Three distinct methods are available to determine this ratio, all of which will be used.

Salt In - Salt Out The default method of detecting the neutral current interaction is to add ^{35}Cl to the D_2O in the form of $MgCl_2$. Monte Carlo calcula-

tions indicate that with ~2 tonnes of salt added around 80% of the neutrons produced in the heavy water will capture on ^{35}Cl [Lyon 95, SNO 87a] (the cross-section for which is 43 b). The capture results in a γ cascade totalling 8.6 MeV (see Appendix D). The γs Compton scatter and the Čerenkov light from the resulting electrons produces a signal at the PMTs. By naively taking the event rate from a salt fill run and subtracting the rate from a pure D_2O run one could obtain the neutral current interaction rate, but this assumes that the backgrounds have not changed between the two runs, an assumption that is not likely to be true. A better approach is to compare the energy spectra, and position and direction distributions of events occurring during a salt fill and during a pure D_2O fill by carrying out a simultaneous fit to charged current, neutral current, electron scattering, and background curves. Fig. 2.4 shows a predicted SNO energy spectrum with salt added to the D_2O and should be compared with the pure heavy water spectrum of Fig. 2.3.

Discrete Neutral Current Detectors With the Salt In - Salt Out option requiring a fit to extract the neutral current rate it would also be desirable to have a technique that could identify neutral current reactions on an event by event basis. It is therefore proposed to have a detector run where an array of proportional counters is placed in the D_2O [Robertson 92]. Each counter is filled with a mixture of ^{3}He, Xenon, and Methane and detects neutrons via the reaction

$$n + {}^3\mathrm{He} \rightarrow t + p + 0.764 \text{ MeV}$$

with a thermal neutron cross-section of 5327 b. Each cylindrical counter has a diameter of 5cm and to minimise background the body is formed from ultra pure nickel with a thickness of 0.4mm. The array of counters has a total length of ~900m and is arranged in a square grid of 1 metre spacing with each counter string being anchored to the bottom of the acrylic vessel. The expected efficiency of these detectors is ~42%. As well as the advantage of event by event neutral current detection these counters do not involve the Čerenkov process and so give a measurement of the SNO neutron flux largely decoupled from that of the salt option. Pulse shape discrimination has been shown to eliminate the background from βs and αs with an acceptable loss in efficiency [Hime 95], but great care must be taken with the radioactive contamination of the counters to ensure that the neutron producing background they add is kept to an absolute minimum.

Pattern Recognition During a salt fill run the signal from charged current events results from the Čerenkov light of a single electron whereas neutral current signals from the γ cascade of ^{36}Cl come from the Čerenkov light of usually more than one electron. This means that the pattern of hits

Reaction	Target medium	Events in ν_e burst (kt^{-1})	Events in cooling phase (kt^{-1})
$\nu_e + d \rightarrow p + p + e^-$	D_2O	10	33
$\nu_x + e^- \rightarrow \nu_x + e^-$	D_2O/H_2O	1	16
$\nu_x + d \rightarrow \nu_x + p + n$	D_2O	6	760
$\bar{\nu}_e + d \rightarrow n + n + e^+$	D_2O	0	20
$\bar{\nu}_e + p \rightarrow n + e^+$	H_2O	0	120

Table 2.3: Predicted number of events in SNO from a stellar collapse at 10 kpc. at 100% detection efficiency. Data is taken from [Burrows 92].

on the PMT sphere from the two event classes should show characteristic differences. Using a pattern recognition technique such differences can be exploited to provide either an event by event or a statistical discrimination of neutral from charged current events. This is the subject of Chapter 7 where it is shown to be successful for a statistical fit, but inadequate for event by event discrimination. The technique has the advantage of only needing a year of salt fill data (although data from a poison run is shown to be highly desirable in addition) as well as not requiring any additional hardware or complex installation procedures.

2.4.2 Supernova Neutrinos

Although primarily designed as a solar neutrino experiment SNO is also a good detector of neutrinos from galactic supernovae [Burrows 90, Burrows 93](but with predictions of the galactic supernova rate that vary between 1 in 10 and 1 in 100 years it is hard to estimate the chances of SNO seeing one). Table 2.3 gives the expected number of events in each class that would result from a stellar collapse at a distance of 10 kpc (the data is taken from [Burrows 92]). In the analysis of such an event the first task would be a search for a ν_τ mass. A tau neutrino with a rest mass of 30 eV emitted at 15 MeV from a supernova at 10 kpc would have a time of flight to earth 2 seconds longer than that of an approximately massless electron neutrino. This would result in a cooling phase resolved into two distinct pulses provided neutral and charged current interactions could be separately identified. This would imply that a D_2O volume strung with neutral current counters would be the detector run of choice when conducting a supernova search, but the separation could also be performed by a pattern recognition technique applied to a salt fill. Discussion of this latter option is continued in Chapter 7.

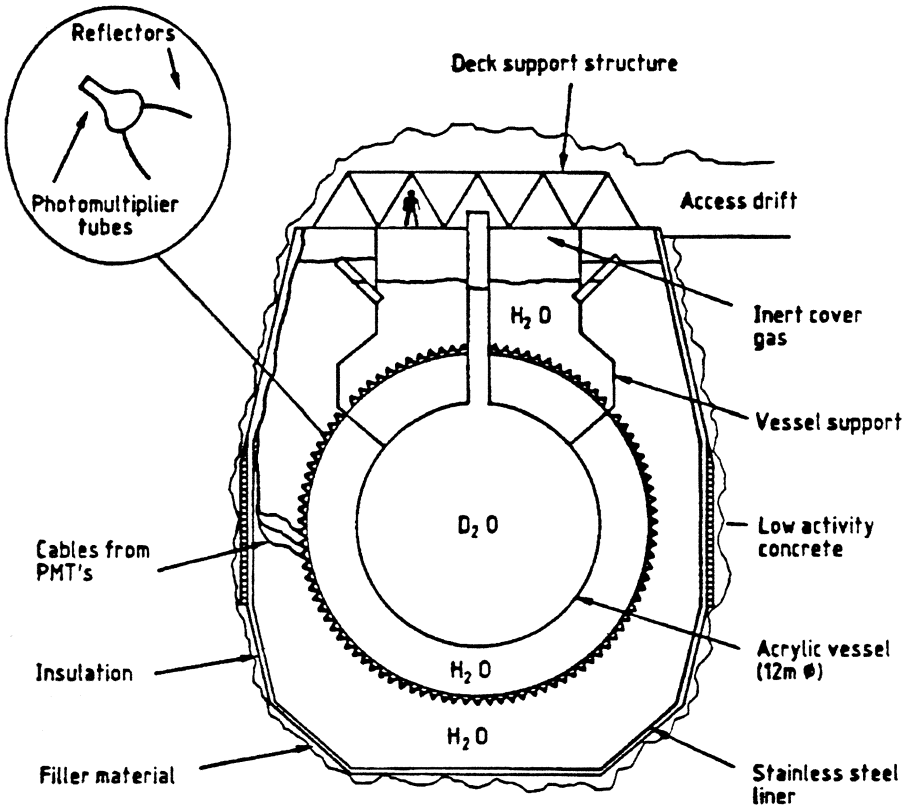


Figure 2.5: Outline of the SNO detector.

2.5 Detector Description

The SNO detector is sited in the Creighton nickel mine of INCO Ltd. near Sudbury, Ontario. The cavity housing the detector is barrel shaped, being 21.4m in diameter at its widest point, 30m high and excavated from norite rock. This rock, similar to granite, is particularly suitable. It has a lower uranium and thorium content than other igneous rocks and its stability enables the excavation of large cavities. An outline of the SNO detector and cavity is shown in Fig. 2.5. It is located at a depth of 2070 m, an overburden that has the stopping power equivalent to about 6000 m of water. This depth ensures that the only significant component of the cosmic ray flux (besides neutrinos) is provided by high energy muons. This flux is small with only about one muon per hour passing through the active volume and only one per day stopping in the detector. Although the muon induced spallation products from ^{16}O will have to be accounted for the low flux means that SNO will have none of the dead time or cosmic rejection problems of the Kamiokande II experiment, where the muon rate was a factor of $\sim$200 greater.

The Čerenkov light from relativistic electrons and muons is detected by an array of ~9600 photomultiplier tubes mounted on 750 panels which are themselves arranged on a geodesic support structure (three-frequency icosahedron). This structure is constructed from struts of 4 inch stainless steel tubing and has an approximate radius of 8.5 m. The ~1.5 m of light water outside the PMT sphere shields the tubes from γs originating at the rock wall and the 2.5 m of water between the PMTs and the acrylic vessel shields the heavy water from the activity of the tubes. For reasons discussed later every detector component must be constructed from low activity materials and the PMTs present a particular problem. As a result of its raw materials glass is usually quite active. The glass bulbs of the PMTs are made from Schott 8246, a special low activity melt, prepared especially for the SNO experiment (assays of the glass are presented later). The tubes themselves are Hamamatsu R1408 photomultipliers with a photocathode diameter of 20cm. The operating voltage produces a charge gain of 10^7 for single photoelectrons and the high voltage between cathode and first dynode ensures the tubes have a low transit time spread of ~1.6 ns [Boardman 92]. This quality of timing assists in the reconstruction of interaction vertices, a subject returned to later in this chapter and discussed fully in Chapter 5. The effect of the earth's magnetic field on the electron optics of the PMTs is all but removed by cancelling field coils surrounding the cavity.

To increase the optical coverage area, reflecting non-imaging concentrators [Welford 89] are mounted on each PMT which increase the solid angle subtended by the tubes around the D_2O from 40% to 70%. By using concentrators the number of detected Čerenkov photons is increased by 65%. They are designed to accept photons within a cone of half angle 56.4°, so that each photomultiplier cannot see its near neighbours [Moorhead 92]. This blinkering effect greatly reduces the background signal from γs generated near the PMT support structure. The cone angle is chosen so that the fiducial volume within a radius of 7 m is visible to all tubes. The reflectors are manufactured from strips of highly polished aluminium used as a substrate for several layers of dielectrics to achieve high durability and reflectivity. The strips are mounted in a plastic holder to ensure that they conform to the computed shape.

The electronic systems to record the analogue signals from the photomultiplier tubes record their size and absolute time. The charge deposited in a PMT signal is recorded by a dual range ADC which operates between 0.05 and 1000 photoelectrons. The trigger operation is essentially free from deadtime and the discriminator threshold can be set in the range 0.1 to 0.6 photoelectrons. Whenever the number of channel triggers in a fixed time exceeds a fixed level a global trigger initiates the acquisition and recording of data. The time resolution of a single channel discriminator is less than 1 ns and signals are recorded to the nearest 0.2 ns. The absolute (universal) time is recorded to an accuracy of 1 μs to aid in the identification of a supernova in conjunction with other detectors. If a supernova were to occur the electronic system would be able to record a burst of

Isotope	Series	Total $\beta\gamma$ energy	Largest γ energy
^{208}Tl	^{232}Th chain	5.00 MeV	2.614 MeV (36%)
^{214}Bi	^{238}U chain	3.26 MeV	2.445 MeV (1.5%)

Table 2.4: β decay isotopes with γs above the deuteron breakup threshold.

at least 10^5 events in one hour.

The acrylic vessel containing the D_2O is a spherical shell of 6m radius and 5.5cm thickness, topped by a cylindrical acrylic chimney through which the heavy water is entered and removed. The spherical part is fabricated from a brick-work of 122 tiles 10 of which are thickened for attachment to the ropes from which the vessel is suspended. The geometry is discussed in more detail in Chapter 4. The acrylic used to manufacture the vessel has two important properties; very low levels of thorium and uranium and very good ultra-violet transmission. Both these properties stem from the fact that the polymethylmethacrylate monomer is kept free from contamination until polymerisation. No stabiliser chemicals or other additives which attenuate UV light are added. What results is a glassy acrylic with good optical transmission above about 325 nm for a vessel thickness of 5.5cm [Zwinkels 90].

The 1000 tonnes of D_2O within the acrylic vessel is on loan from AECL and has a purity of 99.92%. This very low level of light water contamination is important for the detection of the neutron from the neutral current interaction. In less pure heavy water this neutron would quickly be captured by a hydrogen nucleus and produce a 2.2 MeV γ that would be almost impossible to distinguish from background. The levels of radioactive contamination tolerable in the heavy water are the lowest of any part of the detector as it comprises the active volume. These levels are discussed later.

2.6 Backgrounds and Radioactive Purity

With the exception of cosmic ray induced events the backgrounds in the SNO detector can be divided into three classes; $\beta\gamma$ decays, high energy γs, and neutrons.

2.6.1 $\beta\gamma$ Background

The $\beta\gamma$ decays of various isotopes within the detector present the major Čerenkov background [Skensved 94], the β producing the Čerenkov light directly and the γs via Compton electrons. Figs. 2.3 and 2.4 show that at ~5 MeV the contribution

to the SNO energy distribution from such decays rises almost vertically and completely dominates below that energy. This 'wall' arises from two isotopes; ^{208}Tl in the thorium decay chain and ^{214}Bi in the uranium chain (see Table 2.4). At even lower energies the decays of ^{40}K dominate. These three decays and the chains of which two of them are part are detailed in Appendix F.

2.6.2 γ Background

The labels 'PMT' and 'Cavity' in Figs. 2.3 and 2.4 refer to the signals produced by the Čerenkov light originating from single high energy γs in the detector. These γs have their origin in $(\alpha,p\gamma)$ or $(\alpha,n\gamma)$ reactions in high density materials like the rock wall of the cavity or the PMT support structure [Heaton 92]. Figs. 2.3 and 2.4 show that their contribution to the energy spectrum is expected to be relatively minor.

2.6.3 n Background

Since the signal for a neutral current event is a free, thermal neutron in the detector any event which produces a neutron is absolutely indistinguishable from the neutral current reaction. The only such event of any importance is the photodisintegration of a deuteron by a γ above the threshold energy of 2.2 MeV. There are two important sources of such γs; the ^{208}Tl and ^{214}Bi decays mentioned earlier in connection with the $\beta\gamma$ background. These high γ energies and their branching ratios are given in Table 2.4.

2.6.4 Radioactive Purity

The three backgrounds described above place very strict limits on the radioactive purity of all of the detector components. These limits are most stringent in the active volume of the D_2O and become more relaxed with distance from this region.

High energy γs from stripping reactions with incident αs are only important in high density media (in media of low density, such as water, the α multiply scatters down in energy before it has a chance to interact with a nucleus, and what nuclei are present are of low mass). Therefore, it is only the PMT support structure and rock wall that produce high energy γs in any quantity with the αs being provided by decays in the ^{238}U and ^{232}Th chains. A typical example is the 6.915 MeV γ that comes from the ^{27}Al$(\alpha,p\gamma)^{30}$Si reaction in the aluminium reflectors. The light water between the PMT shell and the cavity wall ensures that γs from the rock have little chance of creating a trigger, and the water between the PMTs and the acrylic vessel means that very few PMT support structure γs can produce neutrons in the D_2O, although they can produce Čerenkov signals through their Compton electrons. Monte Carlo and measurements of the γ flux in the cavity suggest that the flux of γs from the rock is of the order 1 per week over

	^{40}K	^{238}U	^{232}Th
University of Guelph	25.9 ± 5.6	47.6 ± 4.8	19.9 ± 2.5
Birkbeck College	20.3 ± 2.9	38.8 ± 4.6	10.1 ± 1.5

Table 2.5: Concentration of contaminants in Schott 8246 glass in ppb by mass.

5 MeV in the D_2O, with a similar rate from γs originating in the PMT support structure (which is constructed from stainless steel particularly low in uranium and thorium). An event fitter and fiducial cut eliminates the vast majority of these events and further rejection is made possible with an additional direction fit as any γ which has traversed tens of radiation lengths to make it into or near the D_2O region must be travelling radially inwards. Detector runs with a light water or air filled acrylic vessel should also provide information on the external high energy γ rate.

With single high energy γs kept in check principally with sufficient light water shielding the main focus of effort in background reduction is the purging and subsequent assay of ^{232}Th and ^{238}U decay chain isotopes and, to a lesser extent, ^{40}K from detector components. The glass used to manufacture the PMT bulbs is a particular concern and is required to have less than 100 ppb of thorium and uranium. Table 2.5 shows that this requirement is met by the Schott 8246 melt used in the SNO tubes where the assays were made at the University of Guelph and Birkbeck College using γ spectroscopy.

The acrylic vessel with its close proximity to the heavy water has a tighter requirement on thorium and uranium contamination. Measurements of the raw acrylic indicate levels between 0.1 and 0.5 ppt but with significant variation from sample to sample [Earle 94]. These numbers are within the $\sim$1 ppt limit set by background neutron production in the D_2O, but there is also the danger of introducing contamination during the vessel fabrication process which necessitates the imposition of strict controls on the handling procedures and materials used to clean and bond the acrylic panels.

The standard solar model predicts approximately 18 neutrons per day produced in the heavy water and so the rate of neutron production by photodisintegration should ideally be 1 per day or less. This places the limit of thorium and uranium contamination in the D_2O at the unprecedented levels of 3.7×10^{-15} and 45×10^{-15} respectively with the constraints relaxed by an order of magnitude for the H_2O. The purification plants for light and heavy water are complex and include ion exchange and reverse osmosis stages, a vacuum degasser to purge radon, the use of MnO_2 beads for the removal of radium, and an ultrafiltration stage seeded with hydrous titanium oxide to remove radium, ^{228}Th and ^{212}Pb. A number of assay techniques have also been developed

- Extraction of ^{222}Rn by vacuum degassing, followed by α counting in a Lucas cell (an acrylic cell lined with zinc sulphide scintillator).
- Extraction of ^{224}Ra and ^{226}Ra on MnO_2 beads, followed by the emanation and α counting of the radon daughters.
- Concentration of ^{212}Pb-EDTA complex by reverse osmosis, followed by the precipitation of lead chromate to make a thin source for β-α coincidence counting.
- Extraction of ^{224}Ra, ^{226}Ra, ^{228}Th, and ^{212}Pb by seeded ultrafiltration, followed by the use of a liquid scintillator for β-α coincidence counting.

It is likely that purification and assay of the light and heavy water will be a continuous process throughout the lifetime of the experiment as ultra pure water is a very good solvent and there will be constant leaching of contaminants from various detector components into the water.

The problem with many of these assay techniques is that they do not assay the thorium or uranium chains at the points where they cause trouble in the detector (see Appendix F). In such cases the assumption of secular equilibrium has to be made. At the end of Chapter 7 is a brief discussion of a pattern recognition technique which has shown considerable success in assaying the thorium and uranium chains right at the ^{208}Tl and ^{214}Bi decays which cause all the trouble and does not require the extraction and concentration of water samples [Chen 95, Chen 96]. A simpler analysis consists of looking at the position of the background wall (see Fig.s 2.3 and 2.4) and using that to determine the total amount of ^{208}Tl and ^{214}Bi [Frati 94c].

2.7 Detector Response

2.7.1 Event Fitting

One of the most important background rejection stages is the fitting of event locations using the positions and timings of the hit PMTs within a trigger followed by the application of a fiducial cut. For instance, the event rates due to PMT $\beta\gamma$ events can be dramatically reduced in this way. In addition to facilitating simple cuts different event classes have different event position distributions and fits to these distributions can be used to glean information on the rate of each class (a typical example are the neutron capture events where the neutron came from a deuteron photodisintegration whose incident γ came from the acrylic. Such events peak noticeably around the acrylic and fall off with radius with a width of $\sim$50 cm). The event position fitters currently in operation have a resolution of $\sim$30 cm when applied to Monte Carlo data.

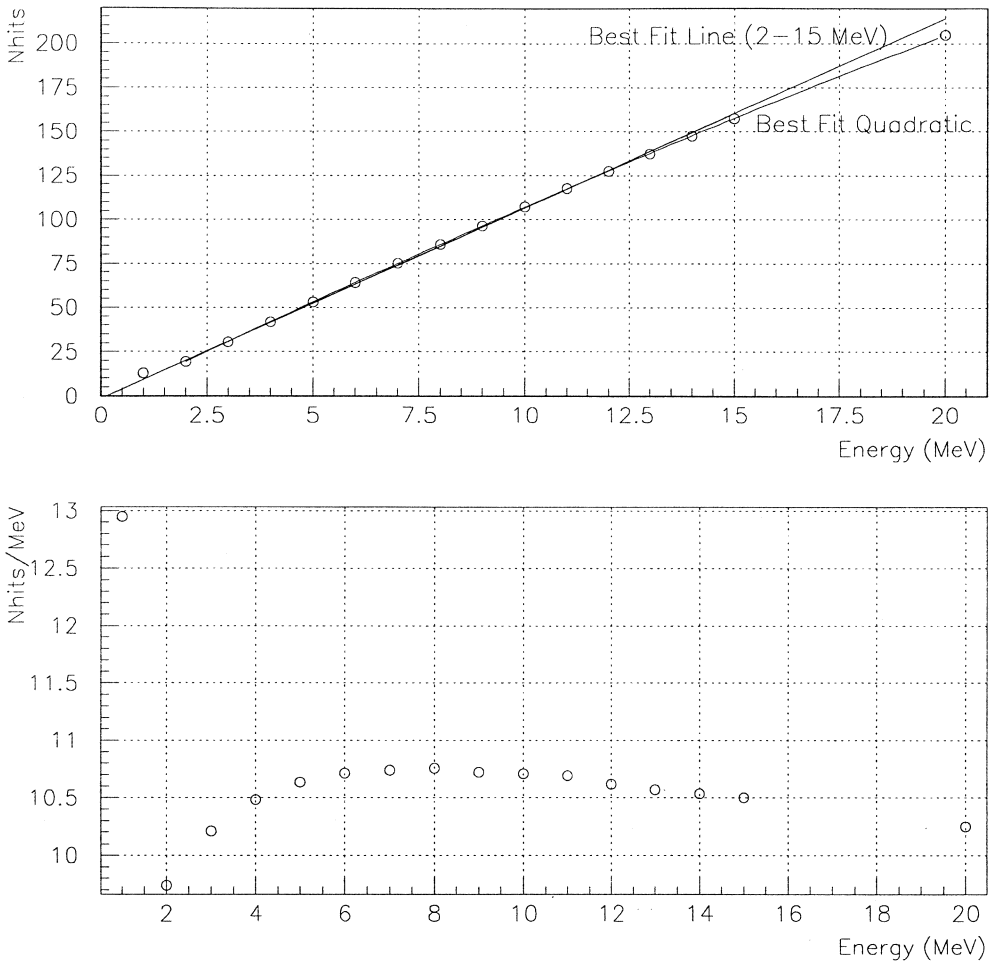


Figure 2.6: The mean number of hit PMTs in a single electron event as a function of electron energy. The events are isotropic in direction and set off throughout the D_2O (plots taken from [Lay 96]).

Direction as well as position fits are useful as some event classes, most notably charged current and electron scattering events, have characteristic angular distributions. Fits to a direction only make sense for single electron events and rely on the shape of the Čerenkov cone, which has a half angle of 42° for a relativistic particle in water. Such direction fits are badly affected by the multiple scattering of the electron which is the cause of the rather poor (~20°) resolution. Further discussion of fitters is deferred until Chapter 5 where they are dealt with in full.

2.7.2 Calibration

The process of understanding the data from the SNO detector relies heavily on Monte Carlo which, in turn, is dependent on calibration of the detector and

Source reaction	Energy	Comment
$^{7}Li(\alpha,\gamma)^{11}B$	E_γ = 10.3 MeV	
$^{3}H(p,\gamma)^{4}He$	E_γ = 20 MeV	
$^{11}B(p,\gamma)^{12}C$	E_γ = 11.7 + 4.4 MeV	Requires 163keV protons
$^{16}N \rightarrow {}^{16}O + e^- + \bar{\nu}_e$	Q = 10.4 MeV	26% B.R., τ = 7.13 s
	Q= 4.3MeV, E_γ= 6.1MeV	68% B.R.
$^{8}Li \rightarrow {}^{8}Be + e^- + \bar{\nu}_e$	Q = 13.1 MeV	τ = 0.84 s
^{252}Cf fission	$E_n \approx 1$ MeV	3 - 4 neutrons per fission

Table 2.6: Some of the proposed SNO calibration sources.

more general consistency checks between simulated and real data. The process of detector calibration breaks into two stages [Hallin 93]. First the determination of various parameters of the data acquisition system such as quantum efficiencies, timing offsets, and pedestals and the detector optics such as water and acrylic transmissions, and reflection probabilities. Secondly, the calibration of the detector's energy response, position and direction fitting resolution, and more general checks on the PMT hit patterns from events.

The first stage uses an optical source, most probably a laser connected to a diffuser ball. The detector optics can be checked by comparing the PMT responses with the source in different positions [Frati 93, Radcliffe 93] and, perhaps, with different fills of the acrylic vessel (the optical properties of the acrylic can most easily be extracted using the source in an air fill). Such source repositioning can also be used to fix the PMT quantum efficiencies which need to be measured in situ. Experience with other water Čerenkov detectors suggests that an overall normalising factor is required to reconcile the Monte Carlo overall efficiency with that of the detector. The IMB collaboration [Becker-Szendy 93] used through-going muons to do this and found that a factor of 0.52 was needed [Svoboda 85], Kamiokande similarly needed a factor of 0.85.

The second stage uses a variety of sources, mainly because the ideal source (a directional, monoenergetic source of electrons in the energy range 5-20 MeV) cannot be constructed. The first task of this stage is an energy calibration. Because of the intrinsic linearity of the Čerenkov process, water Čerenkov detectors have a remarkably linear energy response. Fig. 2.6 shows the Monte Carlo prediction for this response in SNO. The two plots show the mean number of hits and mean number of hits per MeV of single electron events as a function of the electron's energy. The linearity in the region of interest to SNO is very good with about 10.5 hits per MeV. At low energies the curve dips below the straight line due to the effect of the Čerenkov cutoff at 262 keV, and at high energies the curve dips again this time because the plot shows number of hit tubes not number of photo-

electrons and at high energies the probability of multiply hitting a PMT becomes significant [Lay 96]. The resolution at any particular energy is dominated by the statistical fluctuations in the number of tubes which are hit. This means that for a mean of $N(E)$ hits from an electron of energy E the Poisson distribution predicts a standard deviation of $\sqrt{N}$ hits. For a 5 MeV electron this corresponds to a 14% resolution. The LSND experiment were able to use the Michel spectrum from stopped cosmic ray muons to calibrate their energy response [Louis 95], unfortunately the depth of SNO makes this impossible and so β and γ decay sources placed in the heavy water have to be used. There are two problems using a β decay source to calibrate the energy response; firstly the high energies required mean that the source isotope will be short lived and so must be produced as well as decay in situ, secondly the electrons will scatter whilst leaving the source and so distort their energy spectrum with Monte Carlo having to be trusted to predict this distortion. SNO principally uses one such β decay source, ^{8}Li [Earle 93]. It is produced by neutron capture in a pit above the detector and then ferried down tubing in an aerosol into the D_2O volume. γ decay sources are also used in the energy calibration, most notably the 20 MeV γ from the reaction ^{3}H(p,γ)^{4}He where the protons are accelerated onto a tritiated target [Poon 93]. These and other possible sources are summarised in Table 2.6.

Calibration sources are also required for more general checks on the Monte Carlo. In particular the pattern recognition techniques discussed in Chapters 6 and 7 may need to use Monte Carlo data making it imperative that the hit patterns from different event classes are accurately simulated. In addition to the β and γ sources already mentioned SNO uses a ^{252}Cf fission source to produce neutrons in the detector. Depending on the detector run these neutrons capture on deuterium, ^{35}Cl, and/or the ^{3}He of the neutral current detectors producing as many neutron capture events as are required. Some further aspects of calibration and parameter setting are discussed in Chapter 7 along with a proposal for how calibration data might be used to optimise the Monte Carlo. The calibration checks on event position fitters can be carried out with either the laser ball or the β sources, but checks on event direction fitters present more of a problem. Here reliance will probably have to be made on collimated γ sources, but the assumption of single electron events breaks down in this case as a γ can easily produce more than one significant Compton electron.

2.8 SNOMAN

The SNOMAN (SNO Monte Carlo and ANalysis) code is a comprehensive simulation and analysis package for the SNO detector [SNOMAN 95b, SNOMAN 95a, Lay 91]. Written by the collaboration it models the production and transport of all relevant particles through the geometry and media of the detector, simulating the PMT responses and providing a bank of subsequent analysis tools.

At the core of the program is the data structure set up by the ZEBRA memory manager [Brun 93]. Separate software units, carrying out semi-independent tasks, communicate through this structure which houses all relevant information. This modularity makes for more easily altered code and is vital for distributed development across many institutions.

The work of this thesis is rooted in SNOMAN. The source simulation of Chapter 3 uses the EGS4 code system [Nelson 85] to model the transport of electrons, positrons, and gammas. This code also forms the software unit that transports these particles in SNOMAN. The design and implementation of the geometry software unit of SNOMAN is covered in Chapter 4 and Chapter 5 discusses the fitter modules of SNOMAN with the introduction of a new fitting algorithm. Finally Chapters 6 and 7 rely entirely on the simulations of various event classes provided by SNOMAN to carry out their analyses.

Chapter 3

Modelling the SAGE ^{51}Cr Source

Better one safe way than a hundred unreckonable

AESOP
Fables, The Fox and the Cat

No man was ever great by imitation

SAMUEL JOHNSON
Rasselas (1759)

3.1 Calibrating the Gallium Experiments

THE TWO gallium experiments, SAGE and GALLEX, detect solar neutrinos through the reaction:-

$$\nu_e + {}^{71}\mathrm{Ga} \rightarrow {}^{71}\mathrm{Ge} + e^-$$

This reaction has a threshold of 233 keV and is principally sensitive to the pp neutrinos as described in Chapter 1. The germanium production rate from solar sources is of the order of 1 atom per day and so the extraction procedure, which occurs about once a month, has to isolate and detect a mere handful of atoms from many tons of target material. This Herculean task is achieved through a complex sequence of chemical procedures with final counting taking place in a small proportional chamber. With a required isolation factor of $\sim 10^{28}$ it is little wonder that doubts have been raised as to how well the extraction and detection efficiencies are known. Extensive studies of these efficiencies have been made by both collaborations, with experiments testing each stage of the extraction procedure from doping the gallium with a known amount of germanium right through to calibrating the proportional counters with X-rays from xenon. In addition, the low threshold of the detecting reaction above means that is it possible to carry out a global check on the efficiencies and the cross-section by inserting

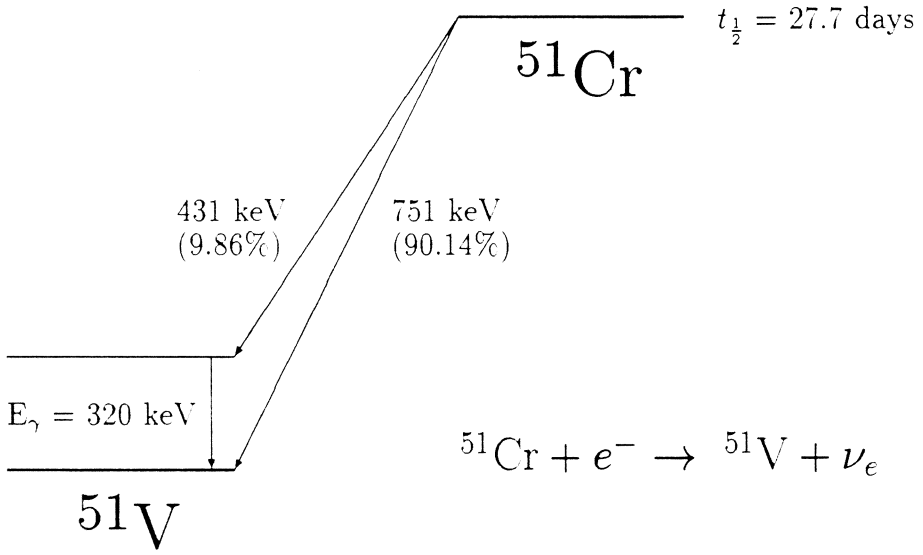


Figure 3.1: The electron capture decay of ^{51}Cr

into the Gallium a neutrino source of known strength and comparing the number of ^{71}Ge atoms subsequently counted with that predicted.

The electron capture decay of ^{51}Cr produces neutrinos of 751 keV (90%) and 426 keV (10%) as illustrated in Fig. 3.1. These energies are very similar to those of the pp and ^{7}Be neutrinos. The ^{51}Cr half life of 27.7 days also means that a source constructed from this isotope will maintain its activity for long enough that a measurement can be made, but will decay away sufficiently quickly that the corresponding drop off in ^{71}Ge production can be detected. Principally for these two reasons both SAGE and GALLEX chose to construct ^{51}Cr calibration sources.

The GALLEX source, with an initial activity of ~1.52 MCi, was inserted into their Gallium tank on June 23, 1994. The exposure continued for ~4 months and the number of ^{71}Ge atoms counted, expressed as the ratio R with the expected number, is [Vignaud 95]

$$\text{GALLEX} \quad R = 0.97 \pm 0.11 \quad (1\sigma \text{ stat. \& syst.})$$

(This is an update from the result of 1.04 ± 0.12 [GALLEX 95].)

Due to size constraints, the SAGE source was somewhat less powerful. Its initial activity was ~0.509 MCi. Their exposure started on December 26, 1994 and after 5 extractions the value of their ratio R is [Elliott 96]

$$\text{SAGE} \quad R = 0.93^{+0.15}_{-0.17} \quad (1\sigma \text{ stat. \& syst.})$$

These two numbers are both clearly consistent with 1 and so provide a global check on the extraction and detection procedures used by each experiment. They

lead to the conclusion that the deficit of solar neutrinos detected by SAGE and GALLEX are not experimental artifacts.

This chapter concerns itself, not with the actual source experiments themselves, but with the earlier stage in 1993 when the SAGE ^{51}Cr source and handling procedures were being designed. In order to obtain a sufficient Ga reaction rate the source needed to be exceptionally active (the original design was for 1 MCi) and the 320 keV γ that accompanies the 10% decay branch, together with γs from impurities, meant that it would also be exceptionally γ active and so pose a significant radiation hazard. This hazard could have been eliminated with sufficient shielding, but the size of the entry port to the SAGE Gallium tanks, together with the dimensions of the calorimeter used to measure the source activity, meant that strict constraints were imposed on the size of the source and hence on the amount of shielding that could be used. It was therefore necessary to model the γ flux from the proposed source design and to quantify the radiation hazard. This would enable suitable handing procedures to be implemented, or, possibly, the redesign of the source (although this latter option did not prove possible to any significant degree as the size and activity constraints were so stringent). This chapter details the method and results of modelling the SAGE source.

3.2 The SAGE Source Design

3.2.1 The Original Geometry

In the original design the SAGE source comes in two parts. The source itself with its shielding is that which is inserted into the Gallium and is shown in Fig. 3.2. For the purposes of transportation the source is housed inside a shipping container, shown in Fig. 3.3, which provides considerably more shielding. These two parts are referred to in this chapter as the inner and outer source respectively.

The ^{51}Cr is in the form of 44 rods of length 1.77" and diameter 0.286". These rods are stacked in pairs and the 22 pairs held in a matrix of 'Heavymetal', a machinable alloy of Tungsten (as shown on the right of Fig. 3.2). The cylindrical Heavymetal matrix is housed in a stainless steel can of height 3.7" and diameter 1.778" which is itself surrounded by more Heavymetal and a further steel can (the left of Fig. 3.2). This gives the inner source a height of 5.52" and a diameter of 3.256". Being predominantly Tungsten the Heavymetal provides considerable shielding, but the hole in the top of the source should be noted. Its purpose is to facilitate remote handling, but its effect on the shielding, as is shown later, is to produce a 'torch beam' of γs. It should also be noted that the chromium and Heavymetal matrix has a 14-fold reflectional symmetry while the rest of the inner source is cylindrically symmetric. This fact can be used in the Monte Carlo to speed up operation.

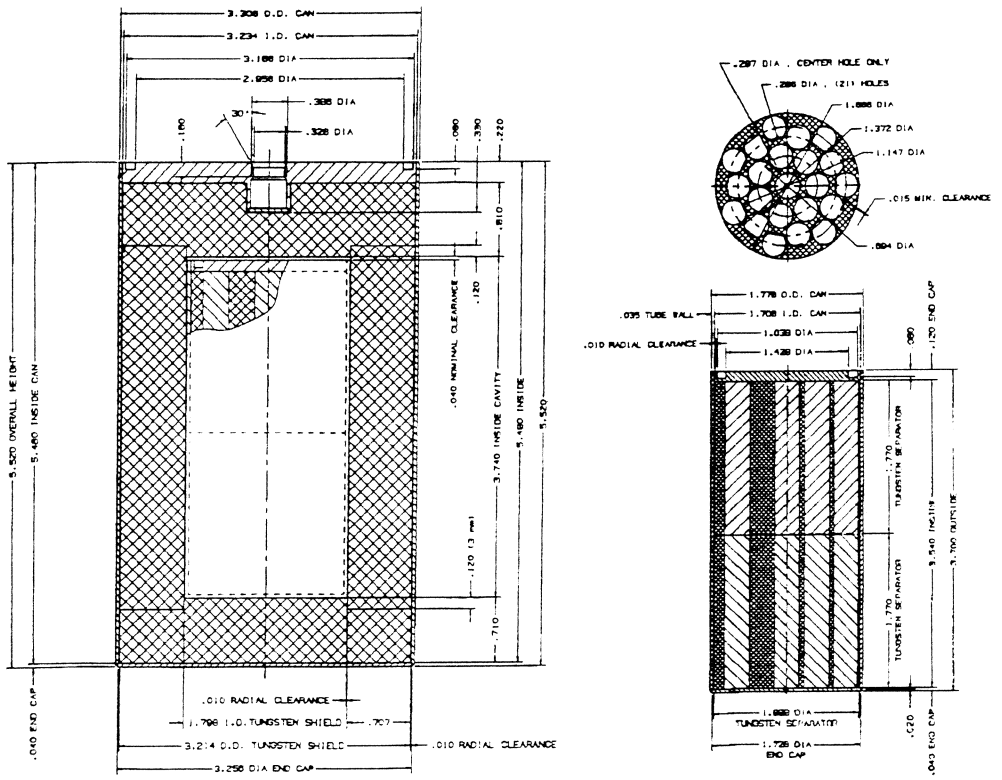


Figure 3.2: The SAGE source and shielding (the inner source) [Teasdale 92].

The shipping container or outer source is shown in Fig. 3.3. Although the detailed design is rather complex, it can be seen that the outer source surrounds the inner source with what is basically a cylinder of Heavymetal of height 9.50" and diameter 7.62" truncated by a vertical plane 3.25" from the axis of the outer cylinder. This outer cylinder is itself offset by 0.56" from the axis of the inner source. Besides a single reflectional symmetry the outer source has no helpful geometric properties that might form the basis of a Monte Carlo speed-up trick.

3.2.2 The Reconfigured Geometry

Simulations with the geometry of the previous section yielded extremely high radiation dose rates from the inner source. This fact, coupled with the desire not to use all the available chromium in one source, led to a decrease in the number of chromium rods from 44 to 24 and a consequent decrease in the activity of the source from ~1 MCi to ~0.5 MCi. This meant a redesign of the rod housing (on the right of Fig. 3.2) in order to optimise the amount of shielding. The dimensions of the housing were not alterable, but the rod configuration within it was, and, after a number of simulations it was determined that the optimal arrangement of

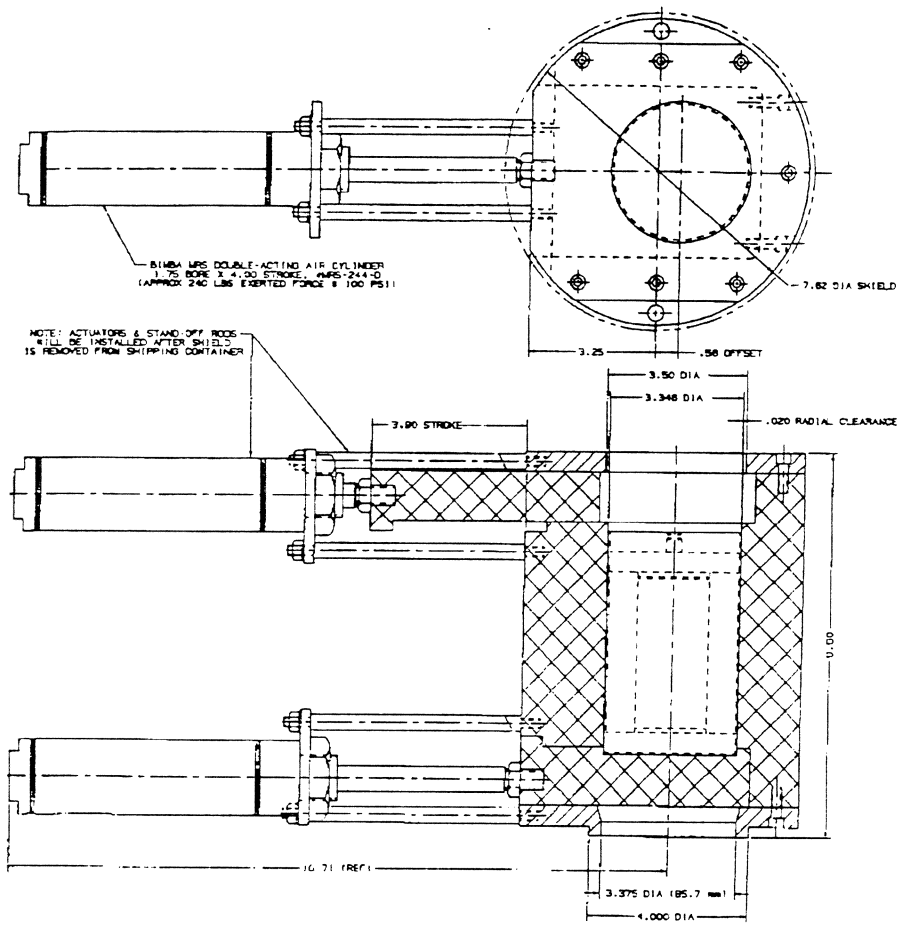


Figure 3.3: The shipping container for the SAGE source (the outer source) [Teasdale 92].

the 24 rods was in two close packed rings, each of 12 rods, where the central radius of the inner ring was 0.36". This ring arrangement gave the inner source a 12 fold symmetry as opposed to the 14-fold symmetry of the original configuration. The rest of the inner and outer source designs were unaltered. The rest of this chapter deals exclusively with the reconfigured geometry.

3.2.3 The Activity

The ^{51}Cr rods are produced by the neutron activation of chromium highly enriched in ^{50}Cr. The enrichment uses a centrifugal isotope separation process [Popov 95] applied to natural chromium and results in the ^{50}Cr fraction rising from 4.3% to above 90%. The chromium rods are then irradiated and the source fabricated at

Imp.	Concentration (% by mass)			No. γs modelled	Energy range of γs (MeV)
	n activation	ICP-MS	JMS		
^{51}Cr	90	90	90	1	0.320
^{76}As	4.5×10^{-5}	$< 2.0 \times 10^{-4}$	$< 1.0 \times 10^{-4}$	22	0.559 - 2.655
^{60}Co	2.3×10^{-4}	1.6×10^{-4}	9.0×10^{-5}	2	1.173 - 1.332
^{64}Cu	1.2×10^{-3}	2.9×10^{-3}	2.9×10^{-3}	1	1.346
^{59}Fe	2.7×10^{-2}	3.4×10^{-2}	1.2×10^{-2}	4	0.143 - 1.292
^{72}Ga	2.9×10^{-3}	$< 2.5 \times 10^{-3}$	7.1×10^{-4}	34	0.601 - 2.844
^{140}La	2.0×10^{-3}	$< 3.0 \times 10^{-5}$	$< 1.0 \times 10^{-5}$	17	0.323 - 2.547
^{24}Na	1.0×10^{-3}	—	2.4×10^{-4}	2	1.361 - 2.754
^{124}Sb	5.5×10^{-5}	$< 3.0 \times 10^{-5}$	$< 1.0 \times 10^{-5}$	22	0.400 - 2.092
^{46}Sc	4.0×10^{-4}	8.0×10^{-5}	6.0×10^{-5}	2	0.889 - 1.121
^{182}Ta	3.0×10^{-3}	5.0×10^{-4}	7.3×10^{-4}	30	0.100 - 1.374
^{187}W	2.0×10^{-3}	6.1×10^{-3}	5.4×10^{-3}	21	0.134 - 0.879
^{65}Zn	2.0×10^{-3}	$< 3.0 \times 10^{-3}$	$< 5.0 \times 10^{-4}$	1	1.116

Table 3.1: The important impurities in the ^{51}Cr rods. The concentrations are those three days after neutron irradiation is halted.

the BN-350 fast breeder reactor at Actau, Kazakhstan.

As well as producing ^{51}Cr the neutron irradiation also activates many of the impurities in the rods. Although the impurity concentration is far below that of the chromium, the lifetimes and higher energy γ decays of some of the isotopes produced make a considerable contribution to the γ flux from the source. This was not at first realised, but emerged conclusively from the simulations. The impurity concentrations were first determined via neutron activation [Bowles 93] of the rods and the results for the important impurities are given in the second column of Table 3.1. Simulations with these concentrations yielded a very high dose rate almost entirely due to the impurities. More careful measurements of the rods were then made using ICP-MS [Timofeyev 93] and an alternative mass spectroscopy technique [Gavrin 93]. The results are given in columns three and four of Table 3.1 and indicate generally lower concentrations. The remaining columns give the number of separate γs that are modelled in each decay and their energy range [Lederer 78].

3.3 The Monte Carlo

3.3.1 The Physics

Using the lifetimes and γ branching ratios of the relevant impurities and taking the mass concentrations from Table 3.1 enables the production rate of each γ of a particular energy to be calculated. Adding up the numbers in column 5 of Table 3.1 shows that there are 159 such γs each with its own production rate. The Monte Carlo operates by selecting a γ energy with a probability proportional to the production rate and then propagating it until it leaves the source or is absorbed. Since the task of the simulation is the assessment of radiation safety, where a mass fraction is only known to an upper limit the Monte Carlo uses that limit as the isotope concentration.

The propagation of the γ is carried out by the EGS4 (Electron and Gamma Shower) code system [Nelson 85] which deals with the coupled transport of electrons and photons in media and a geometry specified by the user. It forms the core of the Monte Carlo around which are built the dimensions, geometry, and composition of the source, in addition to routines extracting the required information.

The EGS4 code is informed of the geometry of the source by the routine HOWFAR and its attendant subroutines. The electrons and photons are propagated in energy loss/distance steps with the various interaction probabilities and results being calculated and implemented at each step. Once the next step length has been calculated, but before it is carried out, EGS4 checks with HOWFAR to see whether the proposed step will take the particle across a geometry boundary. If so, the particle is propagated to that boundary, the media and resultant cross-sections are changed, and the propagation continues with the next step. Information is extracted from the electron and gamma showers by the AUSGAB routine which is called after each step and the required data, if needed, can be logged.

EGS4 relies on a number of non-physical parameters which need to be set externally by the user:-

AE:- This global cutoff determines the energy below which an electron is no longer tracked or a secondary electron created. In each case the remaining energy is factored into continuous energy loss. Any electron created by the γs in the media of the source will lose energy very rapidly and so it would seem prudent to set AE very high forcing the electron to deposit its energy immediately and not waste CPU time tracking it. However, the Monte Carlo uses the γ energy deposition in an inch of water to estimate a dose rate (this is described in a later section). The distance travelled by a 1 MeV electron in water is $\sim$2 cm and so the assumption of immediate electron energy deposition would lead to an overestimated dose rate. For this reason AE is set to a value of 150 keV (kinetic), low enough to properly track the electrons in water, but high enough not to use too much CPU time doing it.

ECUT:- This electron energy cutoff is similar to AE but only applies to the discarding of electrons already being tracked. It has no impact on whether a secondary electron is created. A separate ECUT can be set for each media in the simulation and so it is set high (1 MeV kinetic) for chromium, steel, and tungsten causing any electrons to deposit their energy immediately and not be tracked, and low (150 keV kinetic) for water where any electron should be properly tracked.

AP:- This is the same as AE but applies to γs. Since the γs need to be simulated as accurately as possible and since tracking a γ is not nearly as CPU time consuming as tracking an electron of the same energy, AP is set very low (10 keV).

PCUT:- This is the same as ECUT but applies to γs. For the same reasons as AP it is set to 10 keV.

ESTEPE:- This is the maximum continuous energy loss (expressed as a percentage) allowed in any one step when tracking a charged particle. It has no effect on γs, but is very important if an accurate simulation of electron transport is required [Rogers 84]. Consequently its value is only important for the water region and extensive studies [Lay 94] on the transport of electrons in the water regions of the SNO detector indicate that an appropriate value is 2%.

3.3.2 The Geometry

The length of tungsten required to attenuate a beam of ~0.3 MeV γs by a factor of a half is ~0.25cm and the average path length from the chromium to the outside of the inner source is ~5cm. This means that the probability of a γ emitted from the chromium making it out of the inner source is roughly 1 in 10^6. This probability is the whole point of the shielding, but it means that a Monte Carlo geometry implemented as a simple and faithful reproduction of the true geometry, setting off γs in the rods and recording those which made it out of the inner and/or outer sources, would require months of CPU time to obtain a statistically significant number of exiting γs. For this reason two tricks are applied which speed up the execution of the Monte Carlo: one applied to the inner source and one to the outer source.

The Inner Source Trick

As was noted in Sec. 3.2.2, the inner source possesses a 12-fold reflectional symmetry. A factor of 12 in speed and minimal loss of accuracy can be gained by just modelling a $\frac{2\pi}{12}$ slice with the wedge sides perfectly reflecting all electrons and photons. The position, direction, and energy of all particles leaving the top,

bottom, and side are logged and then this output can be reflected round 12 times to generate a simulated flux from the whole inner source. This procedure does generate a slight inaccuracy in the case when two particles resulting from one initial photon would have exited the inner source from two different wedges, but this event is incredibly unlikely and is washed out by the enormous number of γ decays which are simulated and the resulting high statistics.

In addition to modelling the inner source, the Monte Carlo embeds it in a cylinder of water which is an inch larger in all directions. Once a particle has exited the inner source, and its properties have been logged, it then passes through this water region and the energy it deposits there is also logged. The inch of water is supposed to mimic a human hand clasping the shielding and the energy deposition is used to calculate the dose rates from various parts of the inner source.

The Outer Source Trick

The outer source does not possess the appealing symmetries of the inner source and so must be modelled in its entirety. Yet the degree of shielding it achieves and the consequent photon exit probability means that it is not possible to obtain a large enough sample of exiting photons in a reasonable time by starting them off from the chromium rods. To get around this the outer source is handled by a separate Monte Carlo. For every photon leaving the inner source in its Monte Carlo, N photons with the same position, direction, and energy are started into the outer source Monte Carlo. This affords a speed up by a factor of N, but may result in a Monte Carlo that is susceptible to statistical glitches in the inner source output distribution away from its true flux distribution. Since the chance of a photon making it through the outer source without an interaction of some kind is extremely low any such glitches will be smoothed out by the subsequent passage of the photon.

As with the inner source Monte Carlo, particles exiting the outer source have their position, direction, and energy logged before being passed through an inch of water where the energy deposition is recorded to facilitate a dose rate calculation.

3.4 Simulation Results

3.4.1 The γ Spectra

With the constraint imposed by CPU time, 5.6 μs of real time output from the inner source can be simulated and an outer source multiplication factor of $N = 100$ boosts this up to 560 μs of gammas from the shipping container. Shown in Fig. 3.4 are the γ spectra emitted from the Cr rods, inner source, and outer source in 1 μs (each spectrum being scaled to this time period). As was mentioned earlier the concentrations in column 2 of Table 3.1 are superseded by those of columns 3

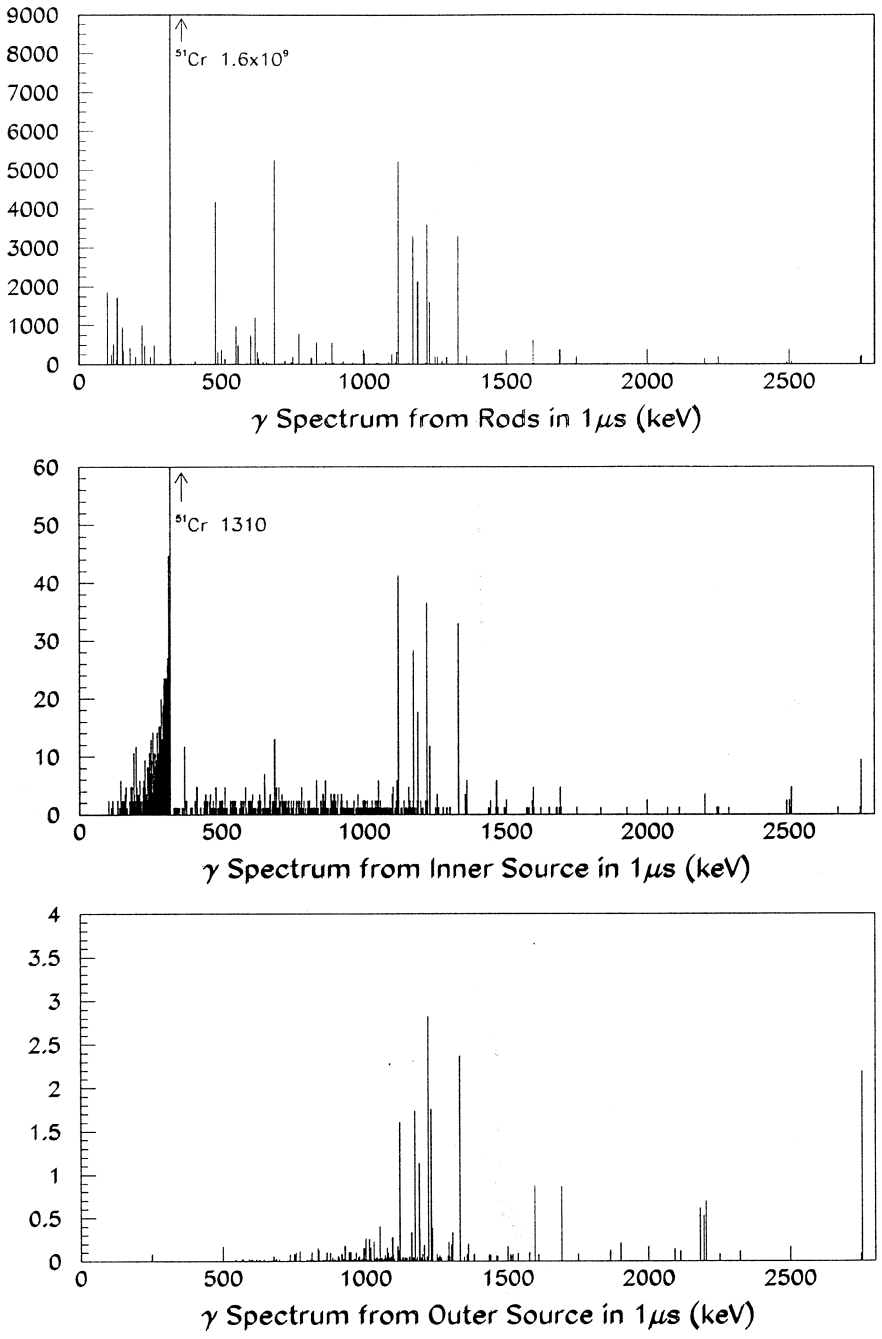


Figure 3.4: The γ spectra from the Cr rods, the inner source, and the outer source. Each spectrum is scaled to correspond to 1 μs of output. Details of the lines are given in the text.

and 4 and it is these latter mass fractions which are used to produce Fig. 3.4. Where the numbers in the two columns conflict the larger of the two is used.

The most notable aspect of the spectrum from the rods is, of course, the huge line from ^{51}Cr at 320 keV, a factor of 10^6 larger than any of the impurities. The two strong lines at 480 and 686 keV come from ^{187}W and the cluster of six lines between 1100 and 1600 keV from ^{182}Ta and ^{60}Co.

Moving to the inner source spectrum the ^{51}Cr line is still strong but has been dramatically attenuated as evidenced by the height of the line and the slew of scattered down γs at lower energies. The two ^{187}W lines have almost disappeared whereas ^{182}Ta and ^{60}Co are still strong. This is to be expected as the total γ interaction cross-section is dominated by the photoelectric effect at these energies in tungsten and assuming roughly an E^{-3} dependence [Krane 88] the cross-section drops by about an order of magnitude over the energy range of the three spectra.

The outer source spectrum dramatically displays this variation in cross-section with the complete disappearance of the ^{51}Cr line and with it almost all γs below 500 keV. The cluster of six ^{182}Ta and ^{60}Co lines are dominant and they have now been joined by lines from ^{140}La, ^{124}Sb, and ^{72}Ga between 1600 and 2200 keV and a strong high energy γ at 2.754 keV from ^{24}Na.

These three plots demonstrate a salutary and perhaps not obvious point: impurities that contribute one part per million of the intrinsic activity of a source can come to dominate the external source activity given only a small difference in cross-section - 'never underestimate the power of an exponential law' (irresistible pun!).

3.4.2 The Radiation Hazard

Once they have left the inner source and outer source geometries, any emitted γs are passed through an inch of water. The energy they deposit is recorded and, with the assumption that water is an adequate substitute for flesh, a radiation dose rate from the inner and outer source surfaces can be calculated. The standard unit for such dose rates from γs is the Rad hour^{-1} (where the Rad is a measure of energy deposition per unit mass and is defined to be 10^2 erg g^{-1} [Shapiro 90]). When dealing more generally with dose rates from ionising radiations the Rad is usually weighted by a Quality Factor. This weighting tries to account for the differing degrees of biological damage caused by various radiations of the same energy. It is defined to be 1 for γs and so is not a concern in this work.

In calculating dose rates from the ^{51}Cr source no attempt is made to quantify uncertainties. The variation between columns 2, 3, and 4 of Table 3.1 indicates an uncertainty in the impurity concentrations that is far greater than any uncertainties arising from the Monte Carlo and upper and lower bounds on each mass fraction measurement are not provided. In addition, all concentrations are modelled at their upper limits in an attempt to quantify an upper limit to the radiation hazard. These two facts preclude the assignment of any meaningful dose

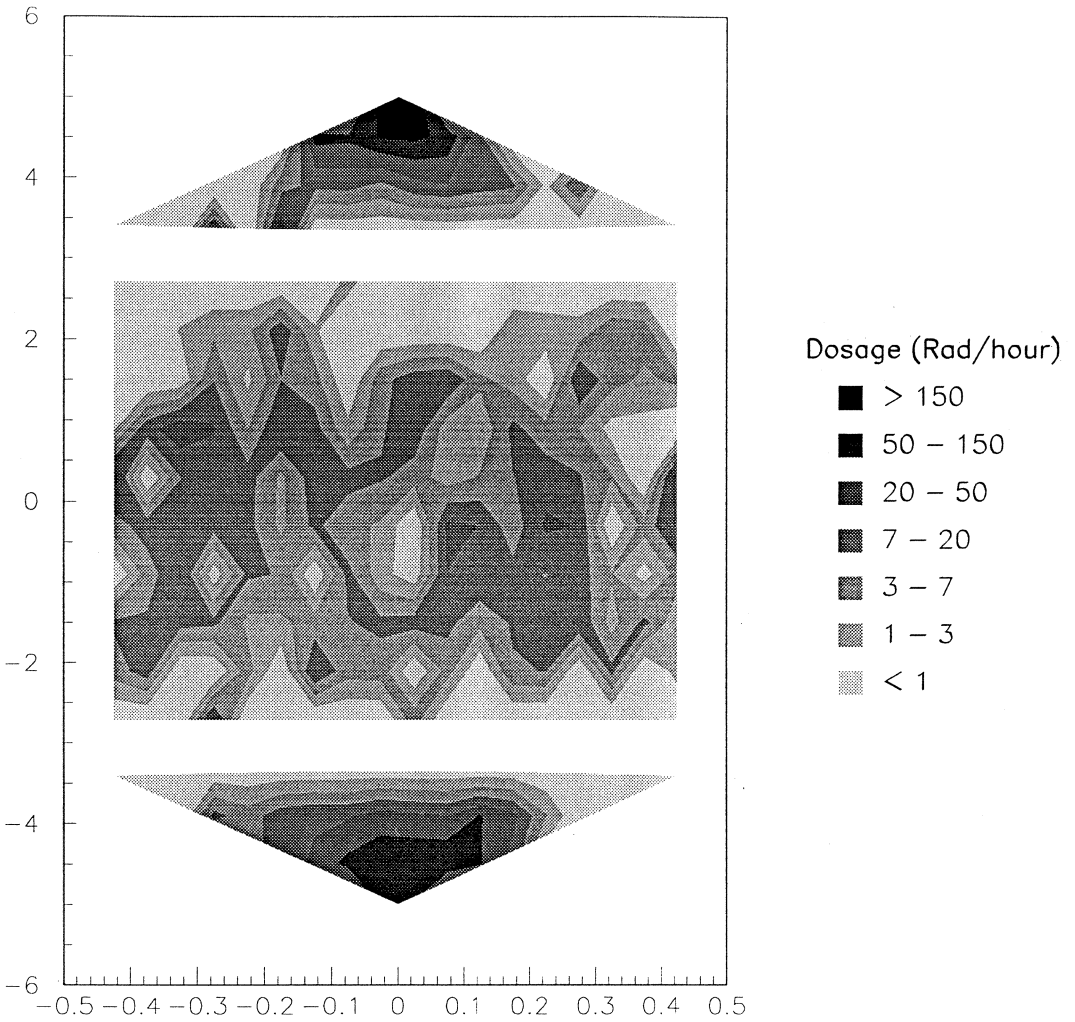


Figure 3.5: The radiation dose rates from the $\frac{2\pi}{12}$ slice of the inner source. The surface is folded out flat with the top segment, curved side, and bottom segment stacked vertically. The axes give the dimensions of the inner source in inches (it has been stretched horizontally).

rate uncertainties. As has already been stated, the results presented use whichever concentration is higher in columns 3 and 4 of Table 3.1 and the reconfigured arrangement of 24 chromium rods.

Shown in Fig. 3.5 is a contour map of the dose rates from the inner source segment. The surface of the $\frac{2\pi}{12}$ slice is unfolded flat with the top, curved side, and bottom stacked vertically. The most noticeable feature is the high dose rate from the top of the inner source. This results from the hole in the Heavymetal that was noted in Sec. 3.2.1 and is shown in Fig. 3.2. The average dose rate from

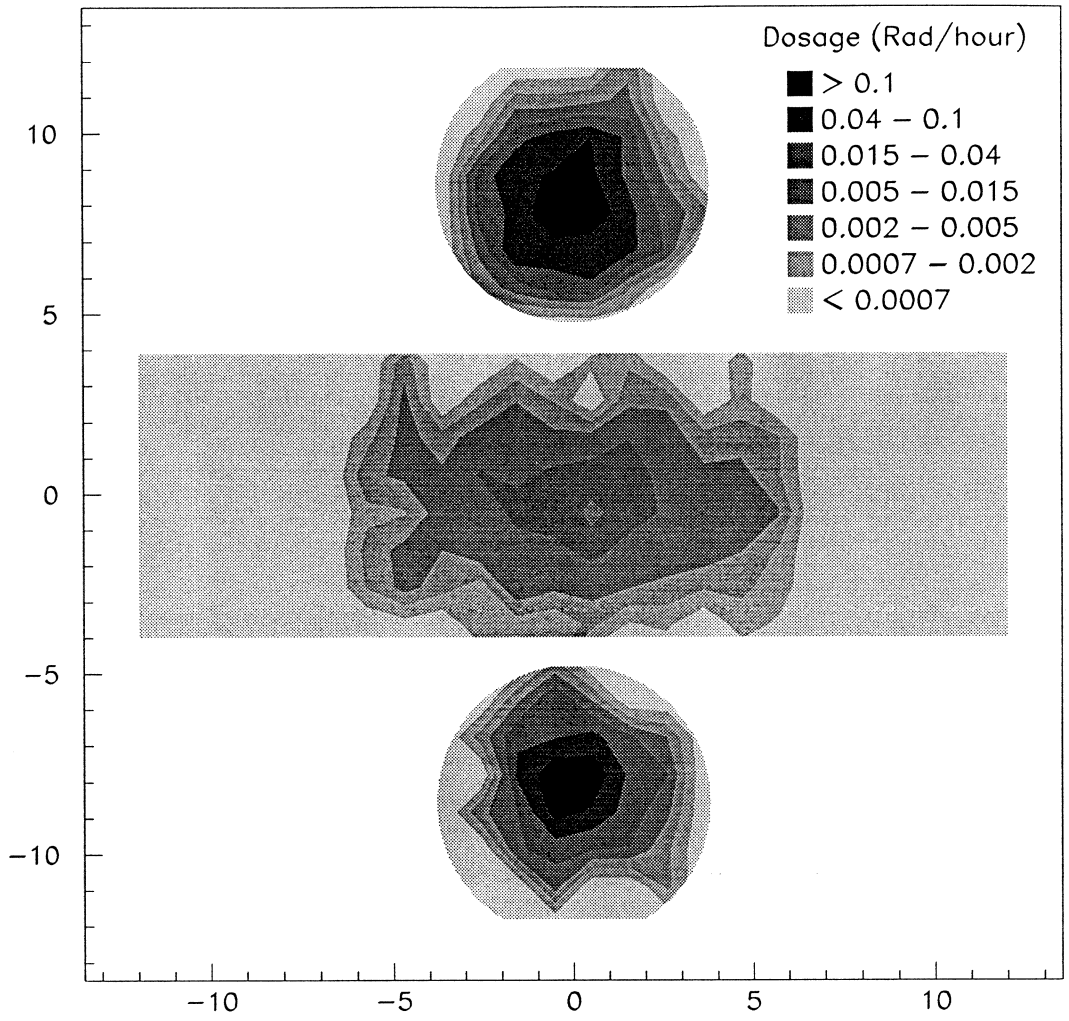


Figure 3.6: The radiation dose rates from the whole outer source. The surface is folded out flat with the flat top, curved side, and flat bottom stacked vertically. The flat truncated part of the curved side (see Fig. 3.3) forms either end of the central strip. The axes give the dimensions of the source in inches.

a circular area of radius 0.5" at the centre of the top of the inner source is 258 Rad hour^{-1}. This and other average dose rates are given in Table 3.2. The curved side of the inner source is relatively benign and has an average dose rate from its entirety of 10 Rad hour^{-1}. This rate rises again on the bottom of the source where the shielding is somewhat thinner. The mean dose rate from a circle of radius 2" at the centre of the bottom is 37 Rad hour^{-1}. Some structure is evident in the dose rates from the side. Whilst much of the detail is due to statistical fluctuations, there is some evidence for the vertical bars of high and low dose rate

Inner Source Dose Rates (Rad/hour)

	^{51}Cr	Imp.	Total
Hole in Top (to radius 0.5")	241	17	258
Whole of Side	4	6	10
Bottom (to radius 1.2")	33	4	37

Outer Source Dose Rates (Rad/hour)

	^{51}Cr	Imp.	Total
Top (to radius 1.7")	-	0.049	0.049
Side (circle radius 5.0")	-	0.022	0.022
Bottom (to radius 2.2")	-	0.032	0.032

Table 3.2: The dose rates (in Rad/hour) from the different parts of the inner and outer sources broken up into contributions from ^{51}Cr and impurities.

which would result from the chromium rod structure within the inner source.

The dose rate contour map for the outer source is shown in Fig. 3.6. The map displays the whole outer source surface with the curved side unfolded in such a way that the flat truncated surface (see Fig. 3.3) is at either end. The dose rates are, of course, much lower than those of the inner source but also show far less variation. The hottest part is the top which still maintains the effect of the hole in the inner source. The average dose rate from a 1.7" circle in the middle of the top is 0.049 Rad hour^{-1}. The curved side of the outer source does not have the uniformity in dose rate displayed by the side of the inner source. This results from the offsetting of the inner source from the cylinder axis of the outer source, as shown in Fig. 3.3. The offset is necessary because of the truncated side of the outer source, but the contour map shows that it is too great and results in a hot spot on the side opposite the truncation. Enclosing this hot spot with a circle of radius 5", the average dose from the region is 0.022 Rad hour^{-1}. Once again the dose rate increases on the bottom of the source as a result of less shielding. The mean rate from a circle of radius 2.2" in the middle of the base of the outer source is 0.032 Rad hour^{-1}.

The average dose rates mentioned in the previous two paragraphs are brought together in Table 3.2 where they are broken down into ^{51}Cr and impurity contributions. For the inner source the ^{51}Cr dose rate dominates (except for the curved side where there is enough shielding to ensure that the impurity rate is higher). This is only to be expected given the second spectrum of Fig. 3.4. The outer source γ spectrum shows a complete absence of the ^{51}Cr line and this is reflected

in the break down of Table 3.2 where the outer source dose rate comes entirely from impurities.

To put the numbers in Table 3.2 into some perspective some comparisons are in order. The recommended γ radiation exposure limit is 5 Rad yr^{-1} [ICRP 90], but this is not a particularly useful number for comparison, as a dose absorbed in a short period of time will do far more damage than the same dose spread over a year. It is estimated [UNSCEAR 77] that a single, whole body, dose of 350 Rad produces 50% mortality and a single dose of 100 Rad delivered to the reproductive system leads to a doubling in the probability of genetic mutations. Comparison with these latter numbers are more relevant and leads to the conclusion that whilst the outer source is relatively safe, the inner source, and more particularly the hole in the top, poses a considerable radiation hazard.

3.5 Discussion

Even with the optimised rod reconfiguration simple calculations had indicated an inner source radiation hazard before the simulation was carried out and, as a result, the ability to remotely handle the source had been included in the design. The simulation results meant that this ability was now an absolute requirement, but reducing the contact dose rates was also desirable. This might have been possible in several ways [Bowles 93]

- The most obvious way would have been to add more shielding, but this was not possible due to the size constraints imposed by the Gallium tank entry port and the calorimeter used to assess the activity of the source. Of these two the calorimeter size was the more important; increasing the source dimensions by even a millimetre would have meant building a completely new calorimeter, a task that was considered too expensive and time consuming.

- Uranium has almost the same density as tungsten but a higher Z. With the photoelectric effect having an approximate Z^4 dependence it would, therefore, afford more shielding for the same volume. The proposal to replace the heavy metal with uranium was attractive but involved the additional production of ^{71}Ge (over and above that initiated by neutrinos) from fission neutrons via a ^{69}Ga(n,p) followed by ^{71}Ga(p,n). Although this rate of ^{71}Ge production was calculated to be very low the existence of the process detracted from the credibility of the calibration and so uranium shielding was dropped.

- Having the ^{51}Cr in the form of rods was not particularly optimal. By packing the chromium closer to the centre of the source greater shielding could be achieved. This meant using either a single solid cylinder of chromium or flat disks which could be stacked into a cylinder. The single cylinder was not

feasible due to self shielding during the neutron irradiation, and the disks had to be abandoned because of difficulties in fabrication and constraints imposed by the neutron irradiation port of the reactor.

- By taking the chromium and splitting it between two sources, each with the design of Figs. 3.2, and 3.3, all of the physical constraints could be satisfied and a considerable drop in dose rate achieved. However, this proposal involved the effective doubling of what were already complex logistics and so was deemed infeasible.

This inability to seriously redesign the source gave vital importance to the careful planning and execution of remote handing techniques.

One further comment on the γ activity of the inner source is necessary. This chapter deals exclusively with the negative aspects of this activity, but the spirit of fairness demands mention of one positive aspect. The large ^{51}Cr line visible in the second spectrum of Fig. 3.4 meant that the source activity (a vital number in the calibration) could be measured both by calorimetry and spectroscopy, giving a very desirable redundancy check.

When the work described in this chapter was being carried out, it was originally intended that some of the Monte Carlo speed-up techniques discussed in Sec. 3.3.2 would be applied to the SNO simulation. The PMT $\beta\gamma$ problem in SNO would have been the major application. This problem is one of CPU time and results from the activity of the PMT glass and dynode stack producing hundreds of millions of $\beta\gamma$ events per year only a tiny fraction of which produce enough PMT hits to be a background concern. Without any *a priori* criteria for the way in which such high NHIT events are produced, many millions must be simulated to extract the thousand or so which are of interest. Using the kind of symmetry based techniques described in Sec. 3.3.2 would afford considerable speed up, but could be dangerous when applied to SNO as they would mean assuming a cylindrical symmetry for the detector and one could not be sure that this assumption was not affecting the PMT $\beta\gamma$ events in some way. This concern meant that the speed-up techniques were not carried over to SNO and, besides, the massive increase in workstation CPU power over the last few years has meant that the PMT $\beta\gamma$ problem can now be handled by the brute force application of computer time.

Chapter 4

The SNOMAN Geometry

Ptolemy once asked Euclid whether there was any shorter way to a knowledge of geometry than by a study of The Elements, whereupon Euclid answered that there was no royal road to geometry

PROCLUS DIADOCHUS 5th C. A.D.
Commentaria in Euclidem

God eternally geometrizes

PLATO
Timaeus

4.1 The Overall Structure

THE SNOMAN (SNO Monte Carlo and ANalysis) code is designed to handle the data analysis tasks of the SNO experiment [Lay 91, SNOMAN 95b, SNOMAN 95a]. As the name implies it tackles the simulation and interpretation of events in a single software package. Fundamental to the design of the code is modularity. This is vital in any large piece of software written by many authors and facilitates easy alteration and addition to the code. Primary amongst these modules is the geometry unit [Wark 93] whose task is to describe the shape, size, and position of the numerous detector components.

The design of the geometry routines within SNOMAN is driven by two fundamental requirements: those of efficiency and adaptability. These routines are central to the operation of the code and their speed must therefore be optimised. Countering this requirement, to some extent, is the desire for SNOMAN to be flexible to addition and alteration, and consequently for the geometry code itself to be modular and the calling sequences to be standardised.

4.1.1 What Is Required Of The Geometry?

When called, the task of the geometry module is to answer one of the following requests:-

NEXT:- Given a position what is the distance to the next boundary in a given direction ?

NEAR:- Given a position what is the nearest distance to any boundary ?

PICK:- Given a detector region pick a point randomly in that region.

(The concept of detector region is defined in Sec. 4.1.2).

To answer these requests the geometry module is fed some or all of the following information on each call that is made to it

- The request type (one of the list above)
- A position (X,Y,Z)
- A direction (u,v,w)
- A region code

4.1.2 Regions and Boundaries

Region Codes

A *detector region* is defined to be an homogenous (i.e. single medium), but not necessarily continuous, volume within the detector. Examples would be the D_2O region or the nickel outer volumes of the discrete neutral current detectors (NCDs). To each different detector region is attached a unique three digit *region code* by which the geometry identifies it. A detector region may comprise an array of identical volumes (the nickel NCD volumes above provide an example) in which case a particular volume is referred to as a *detector element.* Each detector element is identified by a four digit *serial number* in addition to the region code. For detector regions which are not in the form of arrays the terms detector region and detector element are synonymous and the serial number defined to be zero.

Boundaries and Region Type Priorities

By their nature region boundaries are shared between two regions. In order to ensure consistency and minimise both the amount of code and its interdependence it is best if each boundary is dealt with by one piece of code. However the geometry routines deal with regions so the question becomes one of how to assign boundaries to regions. The solution is to give each region a priority and to assign the boundary to the higher priority region. The rules for assigning priorities to regions are as follows:-

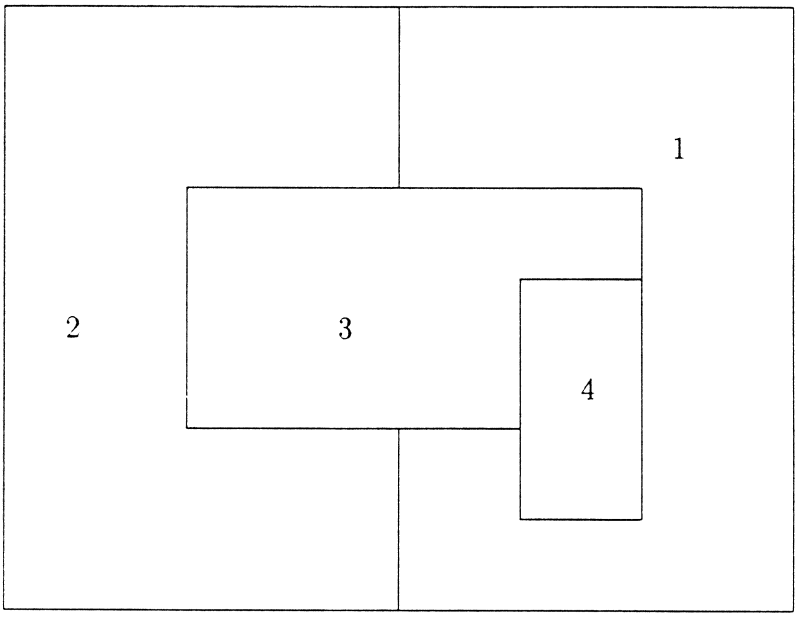


Figure 4.1: An example of a four region embedding geometry. The numbers label the regions as well as giving their priorities (taken from [SNOMAN 95b]).

- If 2 regions share a boundary they have different priorities.
- Where its is simplest to describe the geometry as one element partially embedded in another (e.g. the chimney embedded in the acrylic vessel) the embedded element has higher priority.

In this scheme each region is defined by 2 types of boundary:-

EXTERNAL BOUNDARY to region of lower priority.

INTERNAL BOUNDARY to region of higher priority.

As an illustration consider Fig. 4.1 where the numbers represent both the regions and their priorities. In this diagram:-

Region	Internal Boundary
1	2 3 4
2	3
3	4
4	None

4.1.3 The Building Blocks

The structure of the geometry within the SNO detector can be broken down into three levels.

Primitive Level:- This is the most basic level. A detector element is constructed by the logical operations of OR and NOT applied to a number of primitive shapes such as spheres, cylinders, and blocks. These primitives are rotated and translated into place to form a detector element.

Detector Element Level:- Once constructed by primitive operations the detector element is translated and rotated into its proper place in the detector. If the element is part of an array this process is repeated the appropriate number of times.

Detector Level:- This level consists of the detector in its entirety.

With each of these levels is associated a coordinate system. A primitive is described in a Cartesian system whose origin is at the primitive centre. This system is rotated and translated to the detector element coordinate system centred on the element. Finally a transformation to the detector system is carried out where the origin is now at the centre of the D_2O sphere.

4.1.4 The Software Structure

The physical three layer structure of the previous section is implemented as a six layer software structure, the routines in each layer being identified by a particular prefix.

GE_*:- These are the top level geometry routines and there is one for each of the three request types (i.e. GE_NEXT, GE_NEAR, and GE_PICK). Their task is simply to ascertain which detector region the current point is in (or within which element a point needs to be picked) and to call the corresponding GED routine.

GED_*:- There is one GED routine for each detector region (e.g. GED_D2O, GED_H2O) and it controls the subsequent processing for a point that is in, or needs to be picked in, that region. The GED passes on the geometry request to its own GEDX to answer the question for its external boundary. It then calls the GEDX routines of detector regions that contribute to its internal boundaries to get their answers to the question. An analysis is made of these various answers and from this emerges The Answer to the request made of the geometry module.

GEDA_*:- These routines only exist for detector regions that comprise an array of detector elements. If such a region consisted of 1000 elements then a GED routine would have to call the appropriate GEDX 1000 times, once for each element. To prevent this CPU intensive process the GED routine instead calls the appropriate GEDA routine which isolates a subset of the array, in the form of a list of serial numbers, and then repeatedly calls the appropriate GEDX, but only for that subset. The GEDA routines are not integral to the software structure, but are inserted between the GED and GEDX levels merely optimise the speed of the code.

GEDX_*:- There is one GEDX for each detector element and its task is to answer the geometry request for the external boundaries of that element. It does this by passing on the request to the GEP routines of the primitives that comprise the element. The answers to the request that are provided by each primitive are analysed to provide an answer for the external boundaries of the detector element as a whole. This answer is passed up to the GED level for further analysis as described above.

GEP_*:- There is one GEP routine for each type of primitive (for example GEP_SPHERE, GEP_CYLINDER). The geometry request is passed right down to the GEP routines and answered at this primitive level. The answer is passed up to the GEDX level for further analysis as described above. See Appendix A for details of the primitives used by the geometry code.

GEB_*:- These are the routines that do the analysis of the answers at the primitive level to form the GEDX answer and at the GEDX level to form the GED answer.

The above list is a dense and rather general description of the processing sequence carried out by the geometry module. A specific example should help to clarify the logic.

Consider the two dimensional geometry of Fig. 4.2. Each detector element in this example is labelled by a two digit number, the first digit is the region code and the second the serial number. The current particle position is indicated by the black dot in region 30 and the geometry module is being called to ascertain the nearest distance from the current point to any boundary, i.e. a NEAR request.

GE_NEAR:- Due to the nature of the request the geometry module is entered with a call to GE_NEAR which is passed the current position and detector element. GE_NEAR then calls GED_REG3, the GED routine corresponding to the particle's current region, which coordinates the subsequent geometry processing.

GED_REG3:- GED_REG3 knows that it has an external boundary and as well as internal boundaries formed by detector element 40 and detector element

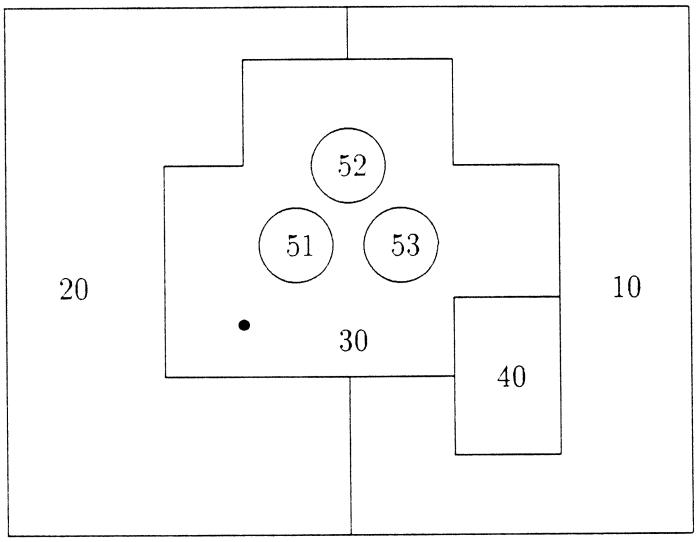


Figure 4.2: The 2D geometry for the NEAR example. Each detector element is labelled by a two digit number. The first digit gives the region code, the second the serial number. The point being considered is the black dot in region 30.

array 5*. First it calls GEDX_REG3 to answer the NEAR request for its external boundaries.

GEDX_REG3:- GEDX_REG3 knows that the external boundaries of detector element 30 are formed from two rectangular primitives combined with a logical OR. It calls GEP_RECTANGLE once for each primitive and is returned a NEAR distance for each. Based on the logic with which the primitives are combined the relevant GEB routine selects one of the two primitive answers as the NEAR distance to detector element 30.

GED_REG3:- Having obtained the NEAR distance to its external boundary, GED_REG3 proceeds to request the NEAR distances to its internal boundaries. First it calls GEDX_REG4.

GEDX_REG4:- GEDX_REG4 knows that its external boundary is formed by a single rectangle. It calls GEP_RECTANGLE obtains a NEAR distance and then returns with this answer back to GED_REG3.

GED_REG3:- Taking the NEAR answers from GEDX_REG3 and GEDX_REG4 a GEB routine selects whichever is smallest and GED_REG3 keeps this as its best answer to the NEAR request so far. It has one further internal

boundary to check and this is with the detector element array of region 5. It consequently calls the array routine GEDA_REG5.

GEDA_REG5:- GEDA_REG5 knows about the locations and sizes of the three detector elements making up the array as well as knowing the current particle position. Based on this information it calculates that the NEAR distance to the array must be to the element with serial number 1 (i.e. detector element 51). It therefore calls GEX_REG5 only once telling it to return an answer for detector element 51.

GEDX_REG5:- GEDX_REG5 knows that each element of the array is formed by a circle and so calls GEP_CIRCLE to return a NEAR distance for the array element asked for by GEDA_REG5. This answer is passed back up to GED_REG3.

GED_REG3:- Once again the smallest NEAR distance so far and the answer provided by GEDX_REG5 are compared by a GEB routine and the smallest is kept. Since there are no further internal boundaries to check this comparison yields the final answer to the geometry request and the result is passed back up to GE_NEAR.

GE_NEAR:- The top level geometry routine simply takes the result provided by GED_REG3 and offers it back to the routine which called the geometry module as the answer to the request.

4.2 Two Complex Examples

This section describes the implementation within the SNOMAN geometry of two of the most complex parts of the detector: the acrylic tiles, and the belly plates and ropes. The tiles are dealt with first and then the plates, in each case the shapes, sizes, and positions of the elements are covered followed by their break down into primitives and then their replication to form arrays.

4.2.1 The Acrylic Tiles

The acrylic vessel separating the D_2O and H_2O has the form of a spherical shell topped by a cylindrical chimney. Ignoring the chimney, it is formed by 'gluing' together a brick-work of spherical tiles as shown in Fig. 4.3. These tiles sit at a radius of 600cm and have a radial thickness of 5.5cm. Viewed in spherical polar coordinates the boundary of each tile is determined by two surfaces of constant radius, two cones of constant polar angle and two planes of constant azimuthal angle. The tiles are arranged in 9 horizontal bands with 11, 13, or 20 tiles in each band, an endcap at the bottom, and a hole at the top where the chimney connects. The 'glue' between the tiles is ~1cm thick along the surface of constant

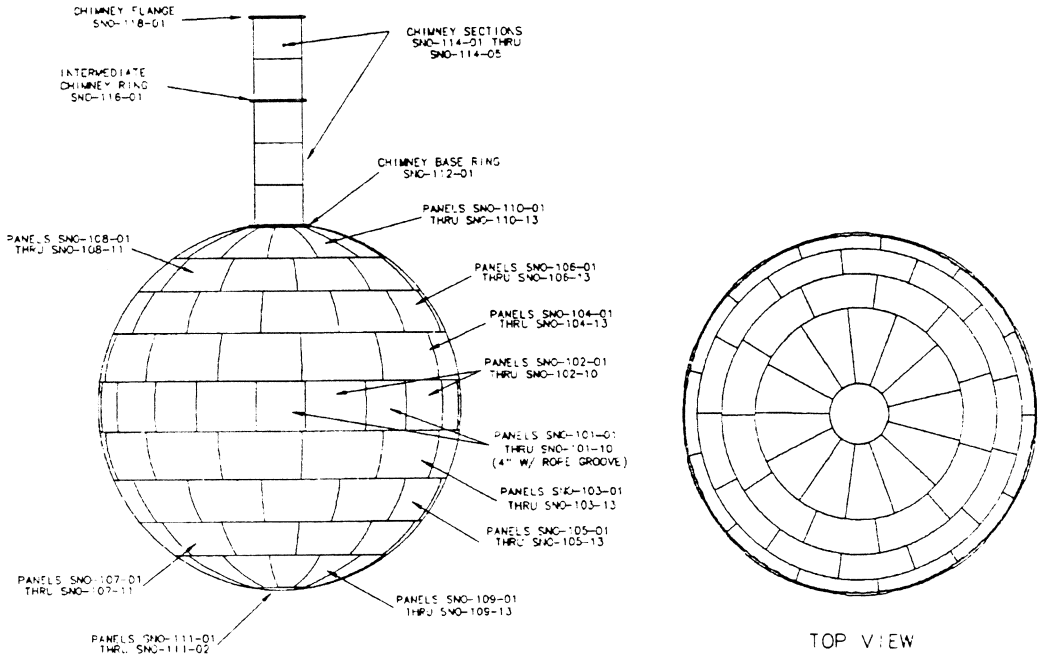


Figure 4.3: The arrangement of tiles making up the acrylic vessel. Different numerical codes refer to different tiles shapes.

radius. 10 of the 20 tiles in the equatorial band are in fact belly plates. These are not shown in Fig. 4.3 for clarity, but are described later.

Primitive Composition - GEDX

An acrylic tile is composed of a single spherical tile primitive (see Appendix A). There are 10 different tile shapes (one in each of the 9 bands and the endcap pieces), but these are all handled by the same spherical tile primitive the only difference being the bounding polar and azimuthal angles which the primitive is fed.

Array Description - GEDA

The description of a set of detector elements as an array is dealt with by the appropriate GEDA routine (as detailed in Sec. 4.1.4). For each of the three request types (NEXT, NEAR, and PICK) this routine must provide a list of serial numbers of the detector elements that the subsequent GEDX routine must check.

The acrylic tile GEDA routine forms its serial number list for the NEXT request by finding where the given point and direction will intersect the spherical

shell of radius 600cm and thickness 5.5cm (there will be 0, 2, or 4 such intersections). By comparing the polar and azimuthal coordinates of these points with the bounding angles of the tiles the list of candidate serial numbers is formed.

For the NEAR request the acrylic tile GEDA routine can narrow the serial number list down to a single tile. This is found by calculating the polar and azimuthal coordinates of the given point and seeing which tile has angle bounds that enclose or are nearest these coordinates.

The GEDA routine handles the PICK request by randomly selecting the particular tile within which the point should be picked. The polar and azimuthal angle bounds of a tile determine its volume (all tiles have the same radius and thickness) and this volume is used as a weighting factor in the random selection. This ensures that any point within the acrylic tile array has equal probability of being picked.

4.2.2 The Belly Plates and Ropes

The acrylic vessel is suspended by 10 pairs of ropes. Each pair is, in fact, a single rope looped through one of 10 belly plates on the equatorial band of acrylic tiles. This arrangement is shown in the upper part of Fig. 4.4 which illustrates the belly plates and ignores the tiles for reasons of clarity. The detailed design of a single belly plate is shown in the lower part of Fig. 4.4. It is essentially a thickened acrylic tile with a radial extent of 11cm. Near the edges of the plate this radial thickness is dropped to the 5cm acrylic tile thickness by four bevelled edges on the both the inner and outer faces. In order to house the rope a groove is cut in the front face of the belly plate. In cross section, perpendicular to its direction, this rope groove is approximately an ellipse wholly enclosed by acrylic with a slot cut through to the outer face of the plate. (The units of Fig. 4.4 are inches. In an experiment jointly designed by American, Canadian, and British institutions the number 2.54 takes on a very special importance!)

Primitive Composition - GEDX

The base primitive for constructing the belly plate is a spherical tile of thickness 11cm. A good approximation to the bevelled edges is made by combining this tile with an appropriately shaped and positioned octahedron using the logical AND operation. The thinner outer part of the plate is formed by adding, with a logical OR, another spherical tile this time of thickness 5cm.

Having constructed the basic belly plate it remains to form the rope groove. As can be seen in Fig. 4.4 the groove comprises two straight parts joined by a half toroidal curved section. For the two straight parts the groove itself is an elliptical cylinder and the slot is formed by a block primitive. There is an analogous combination to form the curved part, with the groove being a half elliptical torus and the slot a half rectangular torus. The rope itself is formed in

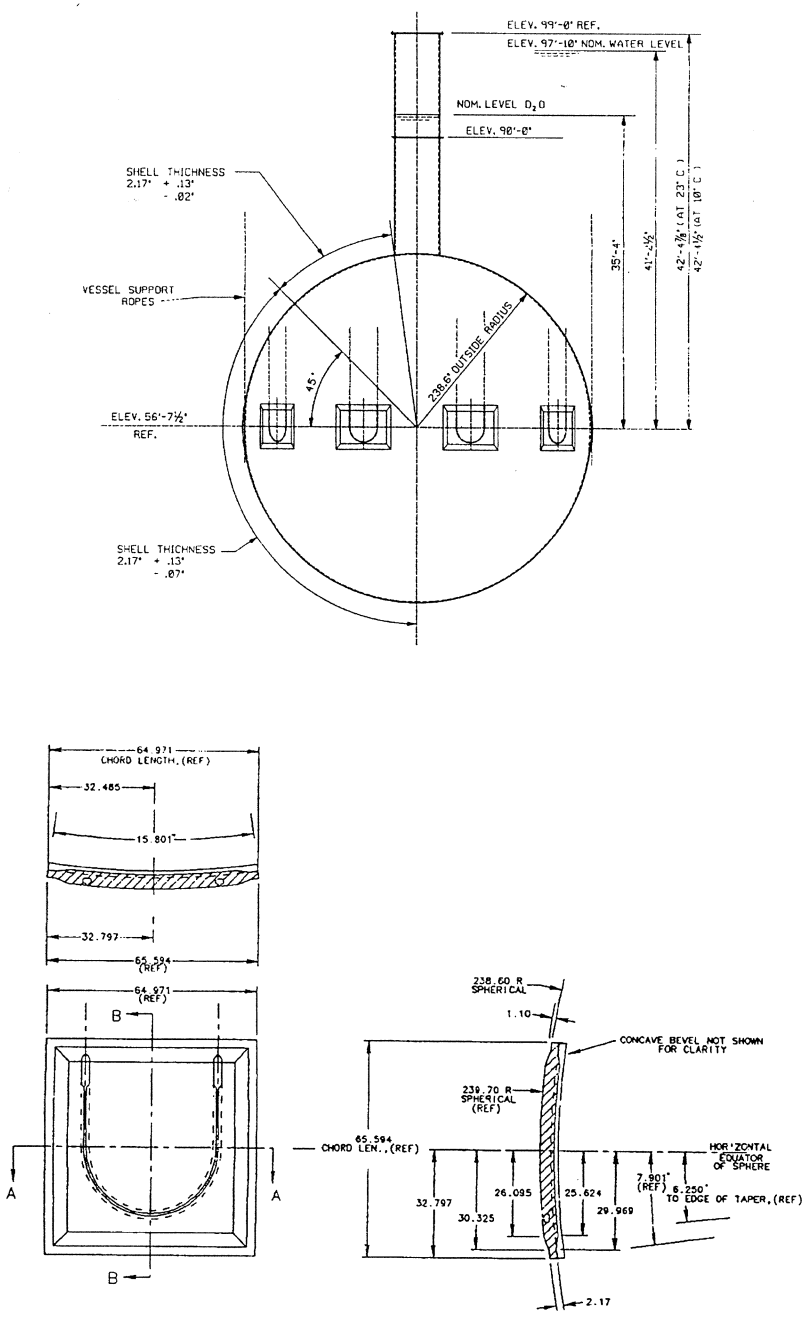


Figure 4.4: The belly plates and ropes. The top diagram shows the positioning of the ten plates and ropes around the equator of the acrylic vessel. The lower diagram gives a more detailed picture of a belly plate. All dimensions are in inches.

much the same way with two very long cylinders linked by a half torus. These various primitives are detailed in Appendix A.

A belly plate and rope is clearly a rather complex construction, involving 12 primitives of 8 different types. It can confidently be claimed that no other object modelled by the SNOMAN geometry is as complicated and so the belly plates provide a very adequate proving ground for the geometry code.

Array Description - GEDA

Since the belly plates are so complex dealing directly with their shape in the GEDA array routine would be hopelessly time consuming and error prone. For this reason the GEDA routine constructs an imaginary sphere around each belly plate and encloses each pair of ropes in an imaginary cylinder. The serial number list for the NEXT request is constructed by looking for intersections (if any) with these imaginary spheres and cylinders. Similarly the NEAR request looks for the closest sphere or cylinder and forms a list comprising but a single serial number. The PICK request is even simpler and need only select a random integer between 1 and 10 since each belly plate is identical.

4.2.3 Results from SNOMAN

A very powerful way of checking that the geometry is properly coded in SNOMAN is to run many particles through the Monte Carlo and record the position and nature of every boundary crossing. By plotting these crossing positions pictures of the SNOMAN geometry can be built up. Fig. 4.5 shows examples of such pictures illustrating the acrylic tiles, and belly plates and ropes.

In the top plot of Fig. 4.5 intersections with the 'glue' between the tiles and with the acrylic vessel chimney are shown as black dots, those with the belly plates as blue dots, and those with the ropes as red dots. (Fig. 4.5 loses significant impact in black and white.) Plotting intersections with the 'glue' allows the three dimensional tiles structure to appear which should be (and is) a near perfect replica of Fig. 4.3. The SNOMAN belly plate geometry is shown in more detail in the lower plot of Fig. 4.5. On the left is a front view of the plate with the rest of the acrylic vessel in green, intersections with the plate in blue, and with the ropes in red. The plunging of the ropes into the belly plate can be seen and their whole course can be followed because of the slot cut in the plate. The curved boundaries of the plate on the left of Fig. 4.5 result from the planar projection of a curved surface. The cross section of the rope groove is shown in detail on the right of Fig. 4.5 with plate and groove intersections in blue and rope intersections in red. The fuzzy boundaries are a natural consequence of the curvature of the plate. Fig. 4.5 provides evidence that the SNOMAN geometry is accurate from scales of 1000cm to 0.1cm. Such accuracy is necessary when it is considered how sensitive an optical device the SNO detector really is.

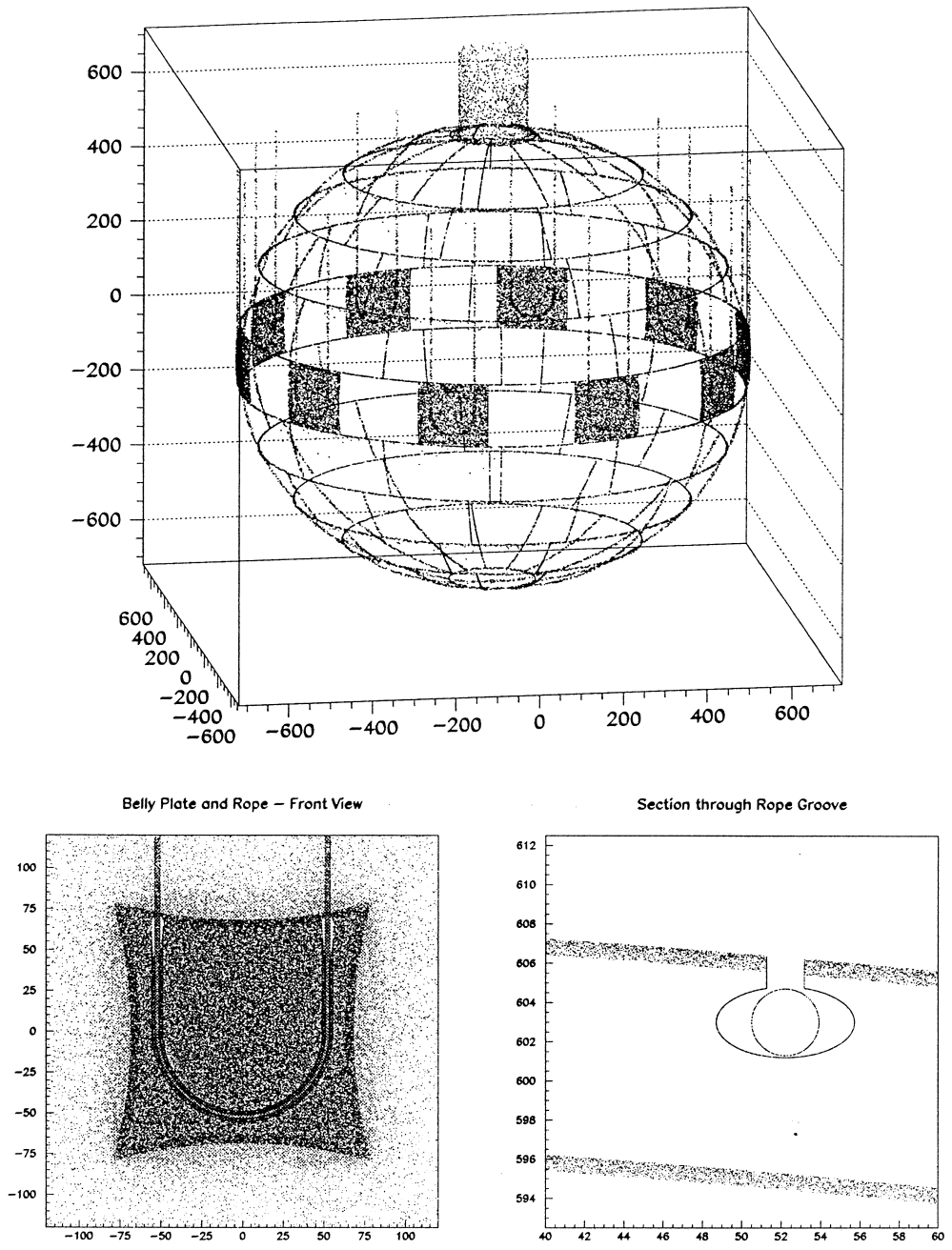


Figure 4.5: (Top) An intersection plot of the acrylic tiles, belly plates, and ropes. (Bottom) Intersection plots of a belly plate and ropes. On the left is a front view of an entire plate, on the right a detailed view of the rope groove in cross section. To see the detail this figure should be in colour.

Chapter 5

The Elastic Fitter

The effort really to see and really to represent is no idle business in the face of the constant force that makes for muddlement

HENRY JAMES
Prefaces [1907-1909]. What Maisie Knew

Everything is vague to a degree you do not realise 'til you have tried to make it precise

BERTRAND RUSSELL
The Philosophy of Logical Atomism

5.1 Introduction

WHEN CONSIDERING the solar neutrino analysis of raw events coming from the SNO detector the signal to noise ratio, with no processing or filtering, will be extremely low (here signal is considered to mean solar events, with noise coming principally from local radioactivity). The most simple processing scenario first involves the raw data being passed through an NHIT filter where only those events comprising more than some threshold number of PMT hits are accepted (a threshold set at 50 hits should produce a signal to noise ratio of order unity in this first NHIT bin). The next filtration step would reject all events that did not originate from some region of the detector (most likely the D_2O). This step requires a fitter algorithm that takes the hit timing and position information and produces a fitted event location.

The scenario of the previous paragraph is only one way in which the analysis of SNO data might begin. Other approaches will require the isolation of certain background classes, the imposition of an NHIT window rather than a threshold, the need for a fitted position as a weighting rather than rejection factor, the use of a fitted event direction, and so on. Common to almost all of these alternative approaches, however, will be some kind of event rejection based on NHIT and the need for a fitted event position and/or direction.

A number of event fitters have been developed for SNO purposes [Frati 94b, Anderson 94, Klein 94, Moorhead 95a]. This chapter begins by describing some general fitter considerations and then the Time and Quad fitters, which have become benchmarks for new algorithms. It goes on to introduce the Elastic fitter, describing the technique in general and then its application to an event position fit and an event direction fit.

5.2 General Fitter Considerations

5.2.1 Position Fits - The Residual

All SNO position fitters to date are based on the same principle. If $\underline{\mathbf{r}}_i$, t_i are the position and time of a PMT hit, $\underline{\mathbf{r}}_{\rm ev}$, $t_{\rm ev}$ are the creation position and time of the photon that caused the hit, and v is the speed of light in water, then

$$R_i \equiv |\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\rm ev}| - v(t_i - t_{\rm ev}) = 0 \tag{5.1}$$

where R_i is known as the residual for the PMT hit labelled by i. This equation assumes a number of things:-

1. The PMTs have perfect position and timing resolution.
2. The photon travels in a straight line with a speed independent of wavelength.
3. The photon travels from its point of creation to the detecting PMT without scattering or reflecting.
4. The PMTs are noiseless.

All four of these assumptions are violated to varying degrees in the SNO detector. In the case of item 1 the PMTs have a photocathode of diameter ~20cm and an RMS uncertainty in the photon arrival time of ~1.6ns. The variation of 22.1 to 22.5 cm ns^{-1} in the water speed of detectable Čerenkov photons in addition to their passage through the different medium of the acrylic vessel will ensure that item 2 is not true (but it is demonstrated in Appendix B that the latter effect will almost always be negligible). Scattering and reflection from the various detector components means that item 3 will be violated and item 4 does not hold up as the PMTs have a dark current that amounts to a firing rate of ~1500Hz.

With $\underline{\mathbf{r}}_{\rm fit}$ and $t_{\rm fit}$ substituted for $\underline{\mathbf{r}}_{\rm ev}$ and $t_{\rm ev}$ in Eqn. 5.1 residual based fitters attempt to minimise in some way the residual of each PMT hit in an event and hence obtain values for $\underline{\mathbf{r}}_{\rm fit}$ and $t_{\rm fit}$. The four items above ensure that an $\underline{\mathbf{r}}_{\rm fit}$ and $t_{\rm fit}$ cannot be found that will set every residual to zero and this produces the finite position resolution of any fitter. Whilst the violations of items 1 and 2 produce relatively minor deviations from Eqn. 5.1 detected photons which have violated items 3 or 4 can show gross deviations and are the major headache for any fitter

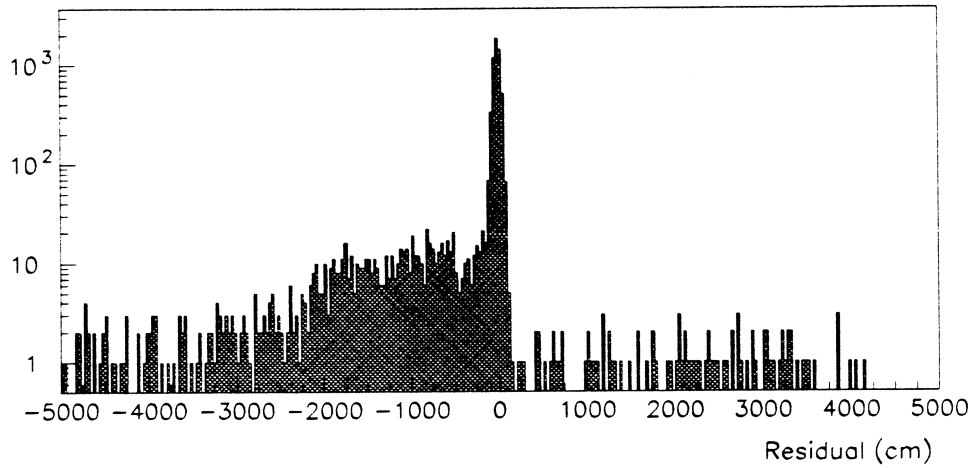


Figure 5.1: The values of the residual for 7000 photons (plus noise hits) generated in the D_2O by SNOMAN.

based on the principle of Eqn. 5.1. For this reason the quality of a residual based fitter is almost entirely determined by its ability to identify and eliminate random noise hits and late hits due to photons that have scattered or reflected. Fig. 5.1 plots the value of the residual for 7000 photons generated throughout the D_2O (simulated by SNOMAN). The flat background of 1 or two hits per bin is due to the noise of item 4. The large number of very negative residuals (i.e. late PMT arrival times) is due to scattered or reflected light, and the narrow peak at zero comes from 'good' hits that have travelled straight from their creation to their detection points, the width of the peak being almost entirely due to the PMT timing resolution.

In fitting the same $\underline{\mathbf{r}}_{\text{fit}}$ and t_{fit} to all the residuals of an event the implicit assumption is being made that all the light originates from a single point in space and at a single time. With the exception of cosmic ray events, where different techniques will be needed, this is reasonable. The time over which Čerenkov light is emitted by a 10 MeV electron is smaller than the timing resolution of the PMTs and the region over which it is emitted is smaller than v times the timing resolution, even when considering Compton scatters from most γ cascades.

Violations of item 3 together with the fact that late hits are more likely to occur outside the Čerenkov cone also combine to produce a systematic shift in the fitted position away from the true position and along the electron's direction vector. This shift is known as fitter pull and is discussed in Appendix C.

5.2.2 Direction Fits - The Cone

Fitting a direction to an event only makes sense if the hits originate from a single electron, in which case the aim is to infer either the initial or some mean electron direction. It is not known *a priori* whether an event stems from one electron or many but a direction fit algorithm must assume the former. All direction fits are also based on one principle: the axis of the Čerenkov cone as the electron propagates is given by the particle's instantaneous direction and so the directions of the PMT hits from a fitted event position can be used to infer a single electron direction. The problem is that electrons of between 5 and 10 MeV will undergo considerable direction changes as they propagate and multiply scatter in the electric field of the surrounding atoms. Because the time between the 'birth' and 'death' of an electron is smaller than the intrinsic time resolution of the PMTs the creation order of the photons which cause PMT hits cannot be inferred and so the direction fit reflects some average electron direction which may be very different from the initial direction one is usually interested in.

5.2.3 Position Fits - The Cone

Besides its use in direction fitting the Čerenkov cone could be used to fit an event position by simply placing a cone of the correct opening angle in the position and direction that best accounts for the observed hit positions and times. All attempts to date at applying this method to SNO have resulted in algorithms greatly inferior to those based on Eqn. 5.1 for the simple reason that the Čerenkov cone is badly smeared by electron multiple scattering. (Such a procedure is used in RICH detectors for particle ID, trajectory and energy fitting [Adam 95, Bachler 94], but multiple scattering is negligible there as the energies involved are far greater.) Another possibility would be to combine a cone fit with a method based on PMT residuals and this idea is returned to in a later section of this chapter as a possible way of combating the fitter pull of residual based algorithms.

5.3 Benchmark Fitters

5.3.1 The Time Fitter

The Time Fitter is the simplest way of minimising the PMT residuals of an event and so obtaining a fit position and time. The method is just a χ^2 minimisation

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{|\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\text{fit}}| - v(t_i - t_{\text{fit}})}{\sigma_i} \right]^2$$

The sum is over the N PMT hits of an event with σ_i being v times the timing resolution of the tubes (this being the width of the peak in Fig. 5.1). The χ^2 is

minimised with respect to $\underline{\mathbf{r}}_{\text{fit}}$ and t_{fit} using the Levenberg-Marquardt algorithm [Press 92]. Hits with large residuals are then removed under suspicion of being due to noise or scattered light and the minimisation is conducted again using the resulting subset of hits. This process of removal and re-fitting is repeated until all remaining hits have satisfactory residuals.

As a method for eliminating 'bad' hits the removal and re-fitting procedure is quite poor. If the first minimisation produces an inaccurate fit because of 'bad' hits then many 'good' hits will have large residuals and be removed from subsequent minimisations thus exacerbating the problem. Nevertheless, the fit position that results after a few re-fits is usually more accurate than that obtained by the first minimisation. The Time Fitter position accuracy is compared with that of other algorithms later in the chapter.

The Time Fitter has appended to it the simplest of direction fitters. Once a fit position has been obtained the unit vector from the fit position to each PMT hit is found and the mean of these is taken as the fit direction. Subsequent direction fit algorithms have improved on this technique substantially.

5.3.2 The Quad Fitter

The Quad Fitter [Frati 94b, Frati 94a, Thorman 94] operates under a somewhat different paradigm to the Time Fitter: that of the Hough Transform [Duda 73]. Taking four hits from an event and applying Eqn. 5.1 to each of them gives 4 equations in the 4 unknowns $\underline{\mathbf{r}}_{\text{fit}}$ and t_{fit}. These equations are solved to produce an $\underline{\mathbf{r}}_{\text{fit}}$ and t_{fit} for those four hits, the solution being termed a quad point. By taking every different combination of four hits a large number of such quad points are found.

$$\text{Number of quad points} = \frac{N!}{4!(N-4)!} \approx N^4$$

where N is the number of PMT hits in the event. These quad points form a cloud in a 4D space (x,y,z,t) which is then searched for the location of maximum point density using the AMOEBA algorithm [Press 92]. This maximum density location is the fitted position and time.

'Bad' hits tend to produce quad points that are widely dispersed throughout the quad cloud and so have little or no effect on the maximum density location. Many four hit combinations that contain 'bad' hits don't even contribute to the quad cloud as their $\underline{\mathbf{r}}_{\text{fit}}$ and t_{fit} solutions turn out to be imaginary (solving the four equations involves a quadratic). In this way the Quad Fitter has an inbuilt method of neglecting or at least penalising 'bad' hits. This method proves to be more effective than that of the Time Fitter and the improved position resolution is discussed later in the chapter.

5.4 The Elastic Fitter Algorithm

The Elastic Fitter algorithm is a development of the Elastic Arms algorithm of Ohlsson, Peterson, and Yuille [Ohlsson 92, Ohlsson 93] and is also related to an earlier neural network technique [Gyulassy 91]. The Elastic Arms algorithm was designed to fit TPC data to straight or helical tracks. It has been changed here to perform a combined position and direction fit with the respective minimisations being largely decoupled. The next section will describe the algorithm in general terms with the following two sections dealing with its application to a position and to a direction fit respectively.

5.4.1 A General Description

Given a set of hits (TPC, PMT, etc.) labelled by i and forming an event, the task is to fit them to some track/vertex defined by a set of parameters π_k (in the case of a SNO position fitter these parameters are $\underline{\mathbf{r}}_{\text{fit}}$ and t_{fit}). To do this one first constructs the square of the minimum distance from each hit to the track/vertex. This minimum distance is assumed to be Gaussian distributed about zero and its square is denoted by M_i. The desired fit could then most simply be carried out by minimising

$$\chi^2 = \sum_i \frac{M_i}{\sigma_i^2} \tag{5.2}$$

with respect to the parameters π_k, where σ_i is a measure of the Gaussian error on each hit. Such minimisations are plagued by problems with local minima and, more seriously, by the effects of noise and 'bad' hits (as has already been discussed in connection with the Time Fitter). But suppose, through divine insight, it were possible to identify these bad hits. One could then define a weight S_i for each hit where

$$\begin{aligned} S_i &= 1 \quad \text{for a true hit} \\ S_i &= 0 \quad \text{for a bad hit} \end{aligned}$$

and then minimise the quantity

$$E' = \sum_i S_i M_i \tag{5.3}$$

where the summation still extends over all the hits in an event and the σ_i are all assumed to be the same.

In the absence of divine insight, suppose some method can be constructed to enable the algorithm to decide for itself which hits are good and which are bad and set S_i accordingly. In this case the function of Equation 5.3 is the wrong function to minimise, as its global minimum can most easily be achieved by saying that EVERY hit is bad and setting $S_i = 0$ for then all. What is needed is a penalty

term in the function which ensures that there is some cost to assigning a bad hit. Calling this cost λ the correct minimisation function is

$$E = \sum_i [\, S_i M_i + (1 - S_i)\lambda \,] \tag{5.4}$$

The optimal fit parameters π_k for a given event are those which minimise E and the simplest way to do that is iteratively using gradient descent. Starting with some initial values of π_k the fit parameters are successively altered by $\Delta\pi_k$ given by

$$\Delta\pi_k \equiv -\eta \frac{\partial E}{\partial \pi_k} = -\eta \sum_i S_i \frac{\partial M_i}{\partial \pi_k} \tag{5.5}$$

where η is an update parameter controlling the speed of descent.

It is now necessary to construct a method by which the algorithm can set the weights S_i for itself. The first step is to notice the close analogy between the minimisation function of Equation 5.4 and the total energy of a system consisting of N particles each of which has two energy states available to it (a physical picture would be that of a lattice of spin $\frac{1}{2}$ particles). In this case $S_i = 1$ would indicate the occupation by the i^{th} particle of the level with energy M_i and $S_i = 0$ the occupation of the level with energy λ. The λ level has fixed energy, but the M_i level has an energy that depends upon the occupations and energies of all the other particles and is also different for each particle. Pushing the analogy further one can consider the system of particles existing in an environment of finite temperature. To calculate various properties of the system of particles it is usually only necessary to consider the average energy level occupation $< S_i >_T$ at a particular temperature and this is done by using the Boltzmann distribution. With $\beta \equiv \frac{1}{T}$, one can therefore define

$$\begin{aligned} V_i \equiv < S_i >_T &= \frac{1 \times e^{-\beta M_i} + 0 \times e^{-\beta\lambda}}{e^{-\beta M_i} + e^{-\beta\lambda}} \\ &= \left[1 + e^{-\beta(\lambda - M_i)}\right]^{-1} \end{aligned} \tag{5.6}$$

One can now take the results of pushing the energy level analogy back into the the fitting algorithm and instead of using S_i in the gradient descent use $< S_i >_T$. So that

$$\Delta\pi_k = -\eta \sum_i V_i \frac{\partial M_i}{\partial \pi_k} \tag{5.7}$$

where

$$V_i = \left[1 + e^{-\beta(\lambda - M_i)}\right]^{-1} \tag{5.8}$$

This weighting factor V_i can be derived in many ways (the working above is but one of them, another is given in [Ohlsson 92]) and is the key to the algorithm. It has the same sigmoid form that reappears in Chapters 6 and 7 in a different

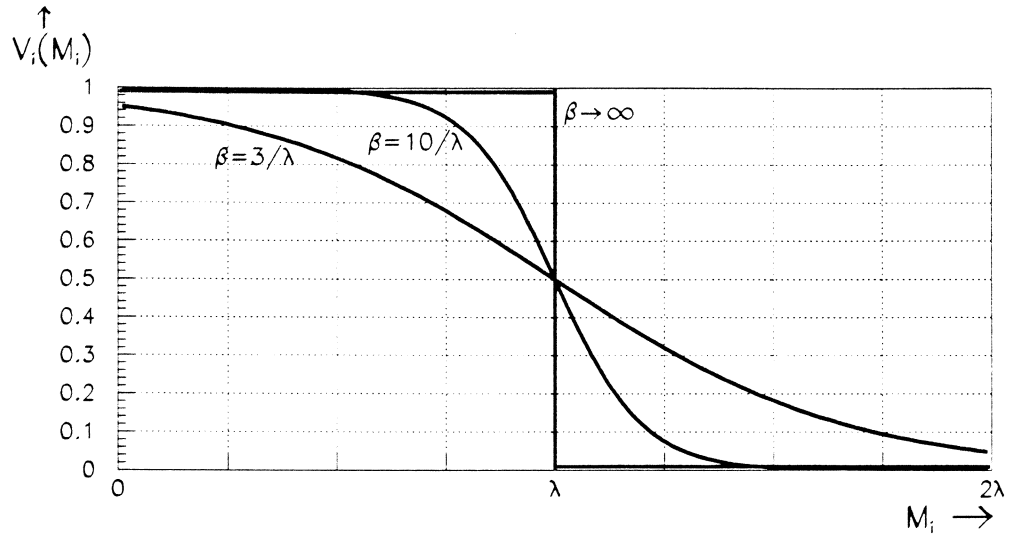


Figure 5.2: The weighting factor V_i as a function of the squared distance measure M_i for various values of the temperature parameter β.

context and is shown in Figure 5.2 as a function of M_i for various values of the temperature parameter β. Values of M_i less than λ receive a weight close to 1, whereas values above λ are effectively discounted from the fit with a weight close to 0. It can be seen that λ is behaving as a threshold, with hits whose M_i is greater than this threshold being discounted as bad hits. Notice also that the sharpness of the threshold increases as the temperature T is lowered (i.e. as β is raised). This suggests an amendment to the minimisation strategy. At the start of the gradient descent iteration the initial values of the fit parameters π_k may not be very accurate and so a slow threshold between good and bad hits (i.e. a high temperature T_{initial}) is necessary so that true hits are not completely rejected merely because of this inaccuracy. As iteration proceeds and the fit improves then this *elasticity* in the threshold (the property which gives the fitter its name) should be reduced (i.e. T should be lowered) so that hits are neglected whose M_i values one can now be confident are bad. This lowering of temperature as the fit proceeds should not continue to $T = 0$ but stop at some final temperature T_{final} that reflects the uncertainty in the value of λ that should be used. This procedure is very like the process of simulated annealing (whose name comes from the analogy with the slow cooling of a metal to give it greater strength). Such annealing methods are becoming widely used principally because they greatly increase the chance of finding the global rather than a local minimum of a function.

There is one further detail needed to complete the algorithm. Close to a minimum the minimisation function E has an approximately parabolic shape, In order

to avoid oscillation from one side of the minimum to the other, as the iteration nears its end, it is necessary to progressively lower the step length parameter η. This should only start once T_{final} has been reached and will enable faster convergence.

5.4.2 The Position Fit

To apply the Elastic algorithm to a specific situation one needs to specify the squared distance measure M_i and the fit parameters to be obtained. For the position fit M_i^{pos} is simply the squared residual for each hit, with a constant δ^{pos} subtracted before squaring.

$$M_i^{\text{pos}} = [\ |\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\text{fit}}| - v(t_i - t_{\text{fit}}) - \delta^{\text{pos}}\ ]^2 \tag{5.9}$$

The subtraction of δ^{pos} is included in order to have the bracketed quantity above evenly distributed about zero. Photons which have suffered grazing incidence reflections, small angle scatters, or the lensing effect of the acrylic vessel (see Appendix B) will tend to pull the peak of Fig. 5.1 to slightly negative values and δ^{pos} is an attempt to bring it back to zero.

Plugging the definition of M_i^{pos} into equation 5.7 the position fit proceeds by first picking some initial values for $\underline{\mathbf{r}}_{\text{fit}}, t_{\text{fit}}$ and then updating them with each iteration according to

$$\begin{aligned} \Delta\underline{\mathbf{r}}_{\text{fit}} &= +2\eta^{\text{pos}} \textstyle\sum_i V_i^{\text{pos}}\ [\ |\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\text{fit}}| - v(t_i - t_{\text{fit}}) - \delta^{\text{pos}}\ ]\ \frac{\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\text{fit}}}{|\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\text{fit}}|} \\ v\Delta t_{\text{fit}} &= -2\eta^{\text{pos}} \textstyle\sum_i V_i^{\text{pos}}\ [\ |\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\text{fit}}| - v(t_i - t_{\text{fit}}) - \delta^{\text{pos}}\ ] \end{aligned} \tag{5.10}$$

with V_i^{pos} defined as before to be

$$V_i^{\text{pos}} = \left[\ 1 + e^{-\beta^{\text{pos}}(\lambda^{\text{pos}} - M_i^{\text{pos}})}\ \right]$$

Iteration proceeds with T^{pos} (i.e. $1/\beta^{\text{pos}}$) decreasing from $T^{\text{pos}}_{\text{initial}}$ to $T^{\text{pos}}_{\text{final}}$ and then with η^{pos} being lowered.

5.4.3 The Direction Fit

To define M_i^{dir} for the direction fit one starts by specifying $R(\theta_{\text{fit}}, \phi_{\text{fit}})$ as the matrix that rotates a coordinate system so that its polar axis is along the Čerenkov cone axis (which is given by θ_{fit} and ϕ_{fit} in detector coordinates). One can then define a transformed PMT hit position $\underline{\mathbf{r}}'_i$

$$\underline{\mathbf{r}}'_i = R(\theta_{\text{fit}}, \phi_{\text{fit}})(\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\text{fit}}) \tag{5.11}$$

In this transformed coordinate system the fit position is at the coordinate origin and the fitted Čerenkov cone axis lies along the polar axis. This is shown in

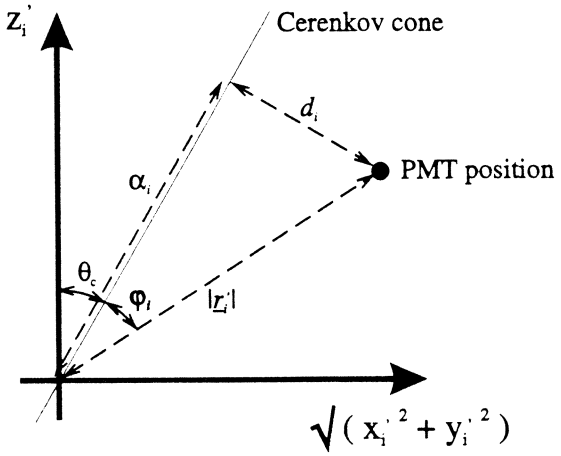


Figure 5.3: The transformed space where the PMT hit is at $\underline{\mathbf{r}}_i'$, the diagonal line is the Čerenkov cone, d_i is the nearest distance from the PMT hit to the cone, and ϕ_i is the smallest angle between the Čerenkov cone and the PMT hit.

Fig. 5.3 which also indicates the minimum PMT - cone distance d_i. Some simple geometry leads to

$$d_i = \cos\theta_c\sqrt{(x_i'^2 + y_i'^2)} - \sin\theta_c z_i' \tag{5.12}$$

This means the sin of the angle ϕ_i between the Čerenkov cone and the PMT hit is

$$\sin\phi_i = \frac{d_i}{|\underline{\mathbf{r}}_i'|} = \frac{\cos\theta_c\sqrt{(x_i'^2 + y_i'^2)} - \sin\theta_c z_i'}{\sqrt{(x_i'^2 + y_i'^2 + z_i'^2)}} \tag{5.13}$$

Also show in Fig. 5.3 is the distance α_i along the cone of the cone point nearest to the PMT hit position. It is given by

$$\alpha_i = \sin\theta_c\sqrt{(x_i'^2 + y_i'^2)} + \cos\theta_c z_i' \tag{5.14}$$

PMT hits which have a negative alpha, i.e. hits which are behind the cone should be left out of the fit as they can badly skew the result. One can now define the minimum squared distance M_i^{dir} as $\sin^2\phi_i$ i.e.

$$\begin{aligned} M_i^{\text{dir}} &= \frac{[\,\cos\theta_c\sqrt{(x_i'^2 + y_i'^2)} - \sin\theta_c z_i'\,]^2}{[\,x_i'^2 + y_i'^2 + z_i'^2\,]} && \alpha > 0 \\ &= \infty && \alpha < 0 \end{aligned} \tag{5.15}$$

With M_i^{dir} defined one can once again plug it into Equation 5.7 and produce the iterative update rules for the fit parameters θ_{fit} and ϕ_{fit}. With $\alpha_i < 0$ then

$\Delta\theta_{\rm fit} = \Delta\phi_{\rm fit} = 0$, and with $\alpha_i > 0$ the updates are

$$\Delta\theta_{\rm fit} = -\eta^{\rm dir} \sum_i V_i^{\rm dir} V_i^{\rm pos} \frac{2d_i}{\sqrt{(x_i'^2 + y_i'^2)}} \frac{\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\rm fit}}{|\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\rm fit}|^2} \bullet \frac{\partial R^{\rm T}}{\partial \theta_{\rm fit}} \begin{pmatrix} x_i' \cos\theta_{\rm c} \\ y_i' \cos\theta_{\rm c} \\ -\sqrt{(x_i'^2 + y_i'^2)} \sin\theta_{\rm c} \end{pmatrix}$$

$$\Delta\phi_{\rm fit} = -\eta^{\rm dir} \sum_i V_i^{\rm dir} V_i^{\rm pos} \frac{2d_i}{\sqrt{(x_i'^2 + y_i'^2)}} \frac{\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\rm fit}}{|\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{\rm fit}|^2} \bullet \frac{\partial R^{\rm T}}{\partial \phi_{\rm fit}} \begin{pmatrix} x_i' \cos\theta_{\rm c} \\ y_i' \cos\theta_{\rm c} \\ -\sqrt{(x_i'^2 + y_i'^2)} \sin\theta_{\rm c} \end{pmatrix} \quad (5.16)$$

where $R^{\rm T}$ is the transpose of $R(\theta_{\rm fit}, \phi_{\rm fit})$.

One deviation from the general algorithm occurs in the equations above. The position weighting factor $V_i^{\rm pos}$ has been included in addition to the direction weighting factor $V_i^{\rm dir}$. This makes sense as hits which the position fit declares to be bad should also be excluded from the direction fit. The direction fit has not been allowed to impact the position fit in the same way because it is significantly less precise.

5.5 Setting the Free Parameters

The largest single disadvantage of the Elastic fitter is the number of free parameters controlling the algorithm which need to be externally set. Some have a strict dependence on the physics and properties of the detector, others reflect the average properties of the minimisation function. These parameters are listed below with indications of how they should be set.

Position Fit

$T_{\rm initial}^{\rm pos}$:- This is the initial temperature of the simulated annealing. There are no strict guidelines to determine how it should be set, but a value of 10-50 times the final temperature $T_{\rm final}^{\rm pos}$ is found to work well.

$T_{\rm final}^{\rm pos}$:- The final temperature of the simulated annealing can be set with reference to Eqn. 5.6. The exponential in $V_i^{\rm pos}$ is effectively a Gaussian in $\pm\sqrt{M_i^{\rm pos}}$ and so $\sqrt{(2T_{\rm final}^{\rm pos})}$ should be the width of this Gaussian. Fig. 5.4 shows the distribution of the residual (i.e. $\pm\sqrt{M_i^{\rm pos}}$) for the true vertex position of 100 electrons with a charged current energy spectrum. It is a blow up of the peak of Fig. 5.1. The width indicated is 36.3cm (which is just v times the timing jitter of the PMTs), meaning that $T_{\rm final}^{\rm pos}$ should be set to ~650cm^2.

$\kappa_T^{\rm pos}$:- This is the rate at which the temperature $T^{\rm pos}$ decreases. From one iteration to the next the annealing temperature decreases by a factor of κ_T.

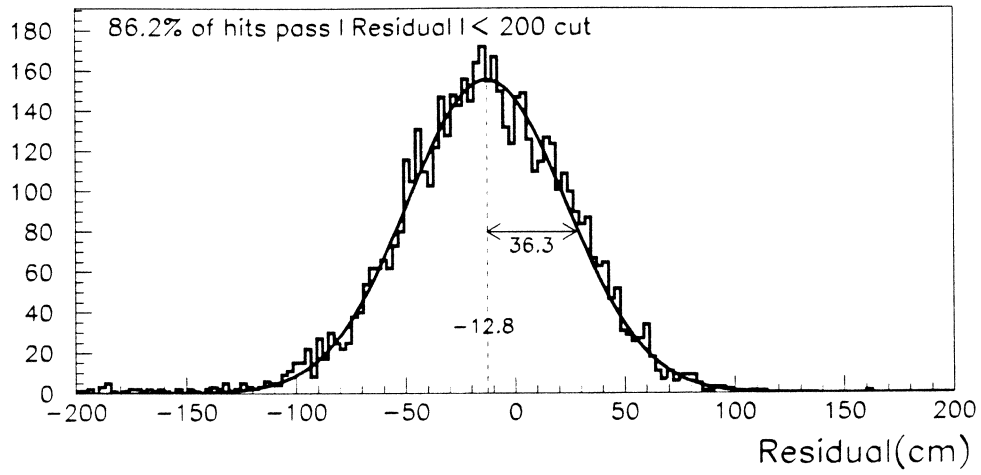


Figure 5.4: The distribution of the residual $|\underline{\mathbf{r}}_i - \underline{\mathbf{r}}_{ev}| - v(t_i - t_{ev})$ for the true vertices of 100 charged current events (~7000 photons) isotropic in the D_2O.

Once again there are no strict guidelines, as its value depends on the shape of the minimisation surface, but a figure of 0.985 works well.

λ^{pos} :- Events with squared residuals (i.e. M_i^{pos}) greater than λ^{pos} should be considered bad hits. From Fig. 5.4 it can bee seen that a good value for this residual is 200cm and Fig. 5.1 shows that this is approximately where the late hits start to have a serious effect. λ^{pos} should therefore be set at 40000cm^2.

δ^{pos} :- As stated before the algorithm works best if the distance measure being minimised is Gaussian distributed about zero. As can be seen in Fig. 5.4 the residual for a true vertex is nicely Gaussian distributed but has an offset of -12.8cm. To try to counter it δ^{pos} should be set to -12.8cm.

η^{pos} :- The size of the iteration step length parameter η^{pos} is dependent upon the shape of the minimisation function. Experimentation shows that a value of 0.01 works well.

κ_η^{pos} :- This is the rate at which the step length η^{pos} decreases once T_{final}^{pos} has been reached. From one iteration to the next the η^{pos} decreases by a factor of κ_η^{pos}. A value of 0.985 is suitable.

initial $\underline{\mathbf{r}}_{fit}$, t_{fit} :- One has to provide the algorithm with starting values of the fit variables. These should be produced by another fitter. With the Elastic fitter being simply a fancy Time fitter, using an initial Time fitted vertex is unsuitable as a bad hit which pulls a Time vertex is also likely to pull an

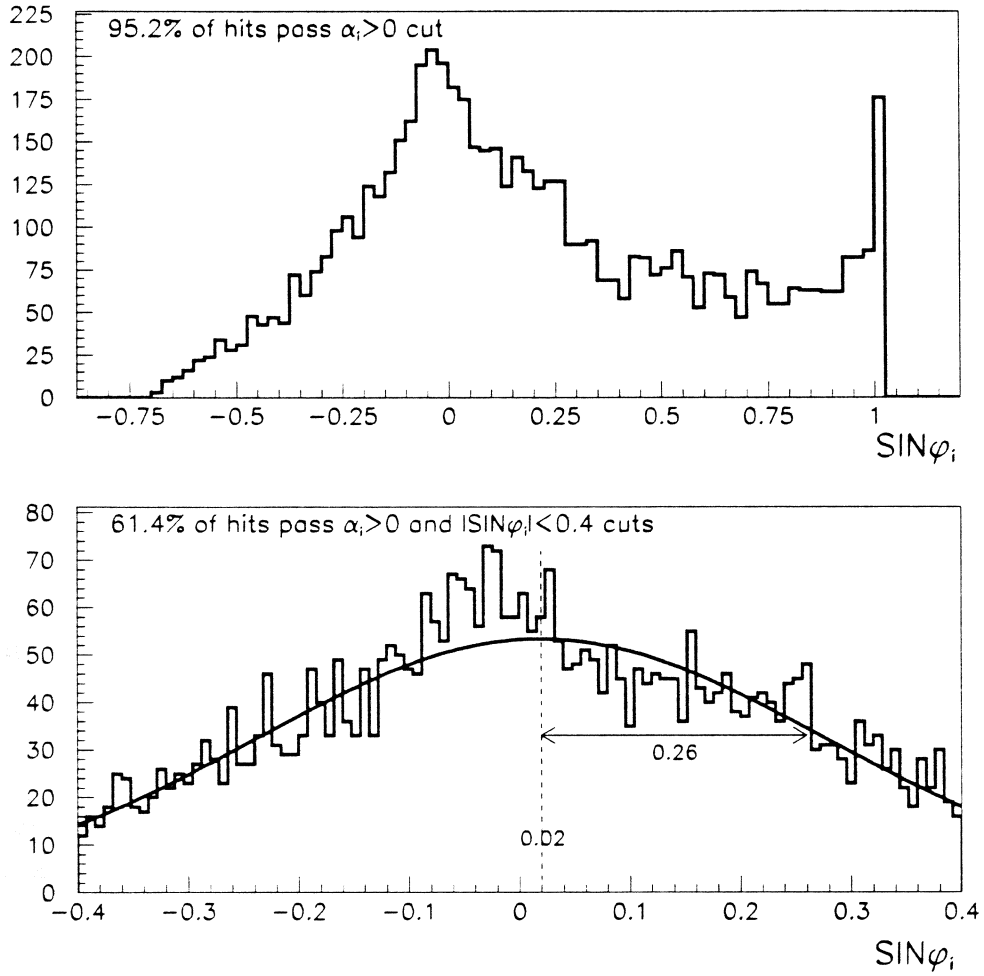


Figure 5.5: The distribution of the SIN of the smallest angle between the PMT hit position and the true, initial Čerenkov cone for the true vertices of 100 charged current events (~7000 photons) isotropic in the D_2O.

Elastic vertex if it is allowed to initially dominate. A Quad fitted vertex is suitable, however.

ftol^{pos} :- The iteration should be stopped once the minimisation function E differs fractionally by less than ftol^{pos} from one iteration to the next. A suitable value for ftol^{pos} is 10^{-6}.

Direction Fit

$T^{dir}_{initial}$:- As with the position fit a value of 10-50 times the final temperature T^{dir}_{final} is found to work.

$T^{\text{dir}}_{\text{final}}$:- Fig. 5.5 shows the distribution of $\sin\phi_i$ (i.e. $\pm\sqrt{M_i^{\text{dir}}}$) for the true vertex direction of 100 electrons with a charged current energy spectrum. The width indicated is 0.26 and the same logic that went into determining $T^{\text{pos}}_{\text{final}}$ means that $T^{\text{dir}}_{\text{final}}$ should be set to 0.035.

κ_T^{dir} :- For the direction temperature decrease rate there are, again, no strict guidelines, but a value of 0.985 works well.

λ^{dir} :- Events with $\sin^2\phi_i$ (i.e. M_i^{dir}) greater than λ^{dir} should be considered bad hits for the direction fit. From Fig. 5.5 it can be seen that a good value for the $\sin\phi_i$ cutoff is 0.4, meaning that λ^{dir} should be set at 0.16.

η^{dir} :- Experimentation shows that a value of 0.1 works well.

κ_η^{dir} :- From one iteration to the next the η^{dir} decreases by a factor of κ_η^{dir}. A value of 0.985 is again suitable.

initial θ_{fit}, ϕ_{fit} :- Multiple scattering ensures that the direction fit is unavoidably less accurate than the position fit. The accuracy of the initial values of θ_{fit}, and ϕ_{fit} is therefore less important than that of their position fit counterparts. A simple mean of the hit directions from some fitted vertex is sufficient.

ftol^{dir} :- This can be set to be the same as ftol^{pos}, i.e. to 10^{-6}.

5.6 Results

All results presented in this section are generated using SNOMAN version 2.08. Shown in Figs. 5.6 and 5.7 are the results of fitting 1500 5MeV electron events isotropically distributed in the D2O (the number of events which fail to fit is negligible). The histograms show the distance from true to fitted vertex for the Quad Fitter, the Time Fitter with an initial Quad vertex, and the Elastic Fitter with the same initial Quad vertex. Fig. 5.6 has a cut at 100cm and Fig. 5.7 a cut at 300cm.

When comparing fitters via these distance error histograms two criteria are relevant:-

1) The extent of the tail out to large distances

2) The proximity of the main peak to zero.

A long tail is a result of bad hits being incorporated in a fit and skewing the result whereas the location and sharpness of the main peak depends loosely upon how much of the available hit information is used by the fitter (here there is a vague analogy with a $\sqrt{N}$ statistical error).

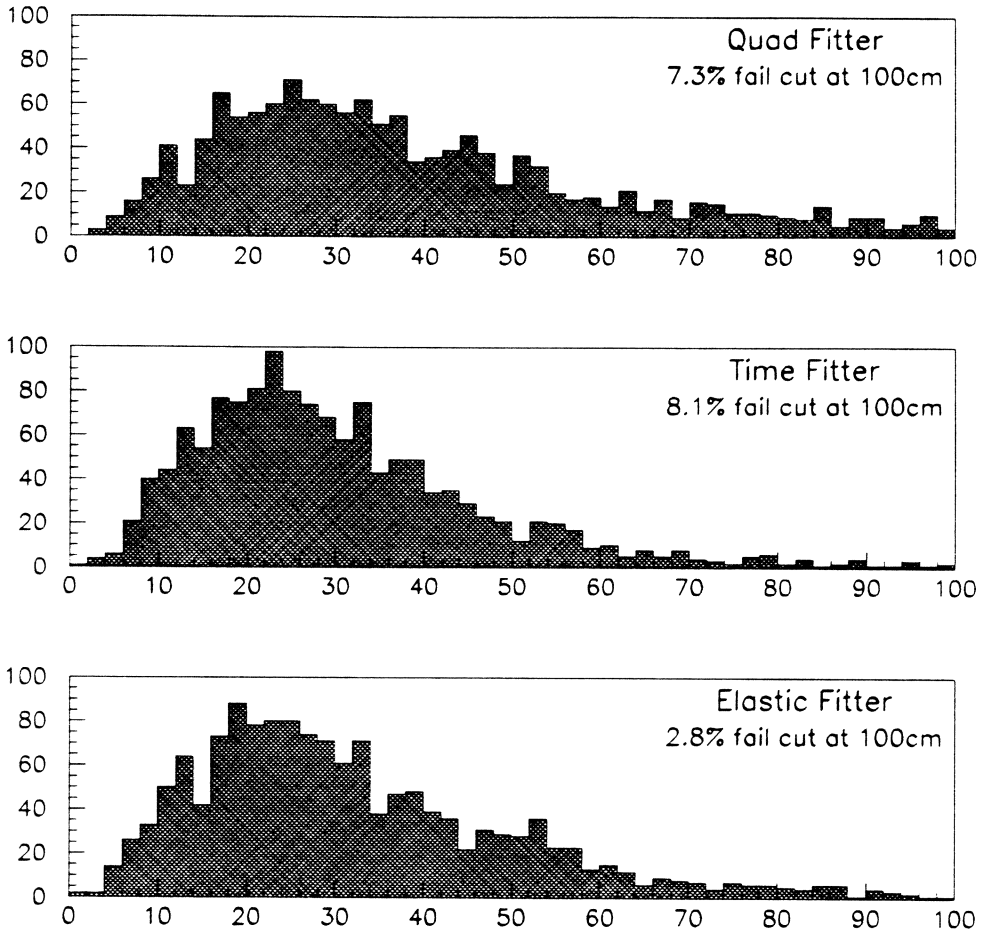


Figure 5.6: The distance (in cm) between the fitted and true vertices for 1500 5MeV electron events isotropic in the D2O and using the Quad Fitter, the Time Fitter with an initial Quad vertex, and the Elastic Fitter with the same initial Quad vertex. A cut is made at 100cm

Typically the Quad Fitter performs well when the first criterion is considered, with all of the 1500 fits getting within 300cm of the true vertex. This is because bad hits tend to either give imaginary results from the quadratic whose solution determines a quad point or produce quad points far from the main cluster. This gives the algorithm its inbuilt method for reducing the effect of bad hits, as was mentioned earlier. Its main peak, however, is comparatively spread out, with 7.3% of the fits failing to get within 100cm of the true vertex. The same technique which damps out bad hits also ensures that not all the good hit information is used and so the fitter does not to perform so well under the second criterion.

The Time Fitter, even when given a Quad fitted initial vertex, has a very long tail, with 4.7% of the fits further than 300cm from the true vertex. Its technique

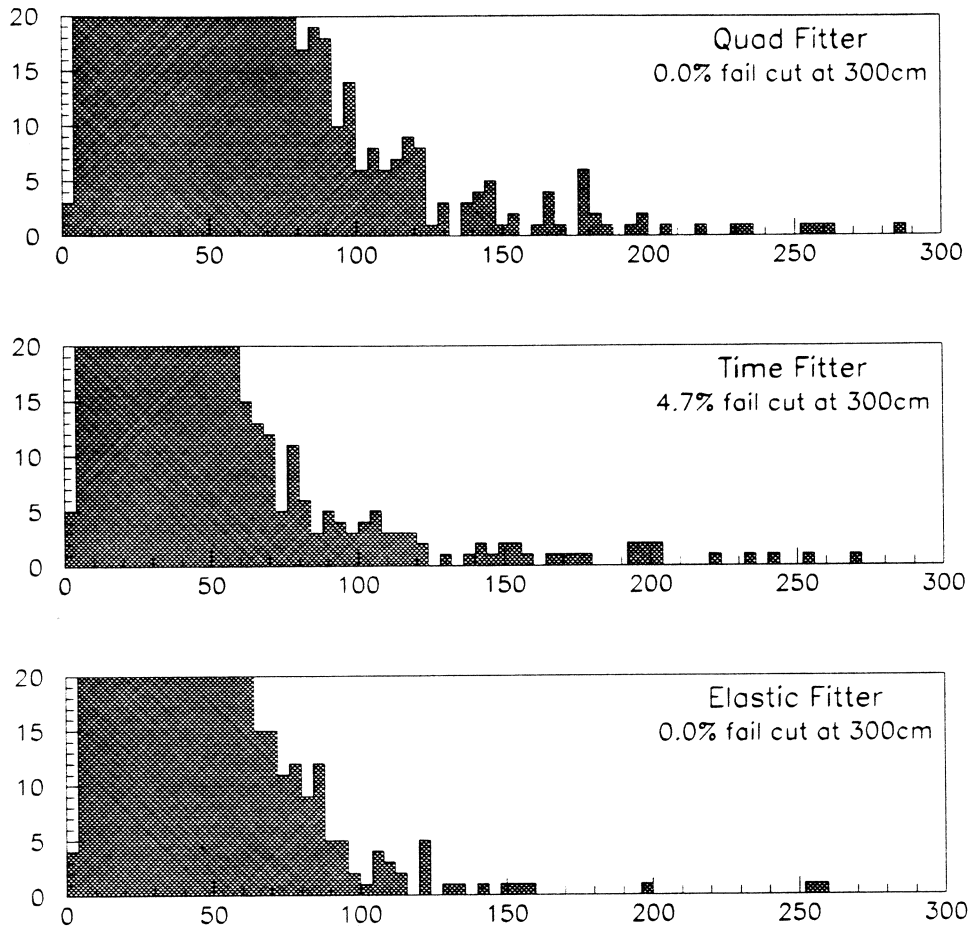


Figure 5.7: The distance (in cm) between the fitted and true vertices for 1500 5MeV electron events isotropic in the D2O and using the Quad Fitter, the Time Fitter with an initial Quad vertex, and the Elastic Fitter with the same initial Quad vertex. A cut is made at 300cm

of fitting, removing hits with a bad χ^2, re-fitting etc. is not satisfactory either aesthetically or practically. However, when the fit is good the Time Fitter makes better use of the available statistics (i.e. it uses more of the hits) giving it a sharper peak closer to zero than the Quad Fitter.

The Elastic Fitter performs well under both criteria. With its V_i weighting it has good noise rejection and a minimal tail, and with the similarity of the algorithm to a χ^2 fit it has the Time Fitter's sharp peak close to zero.

Fig. 5.8 compares the direction fit results for the Time Fitter (a simple vector sum of the PMT hit directions) and the Elastic Fitter. Here the improvement is quite dramatic, but this should be no surprise given the feeble competition provided by the Time Fitter's direction fitting algorithm.

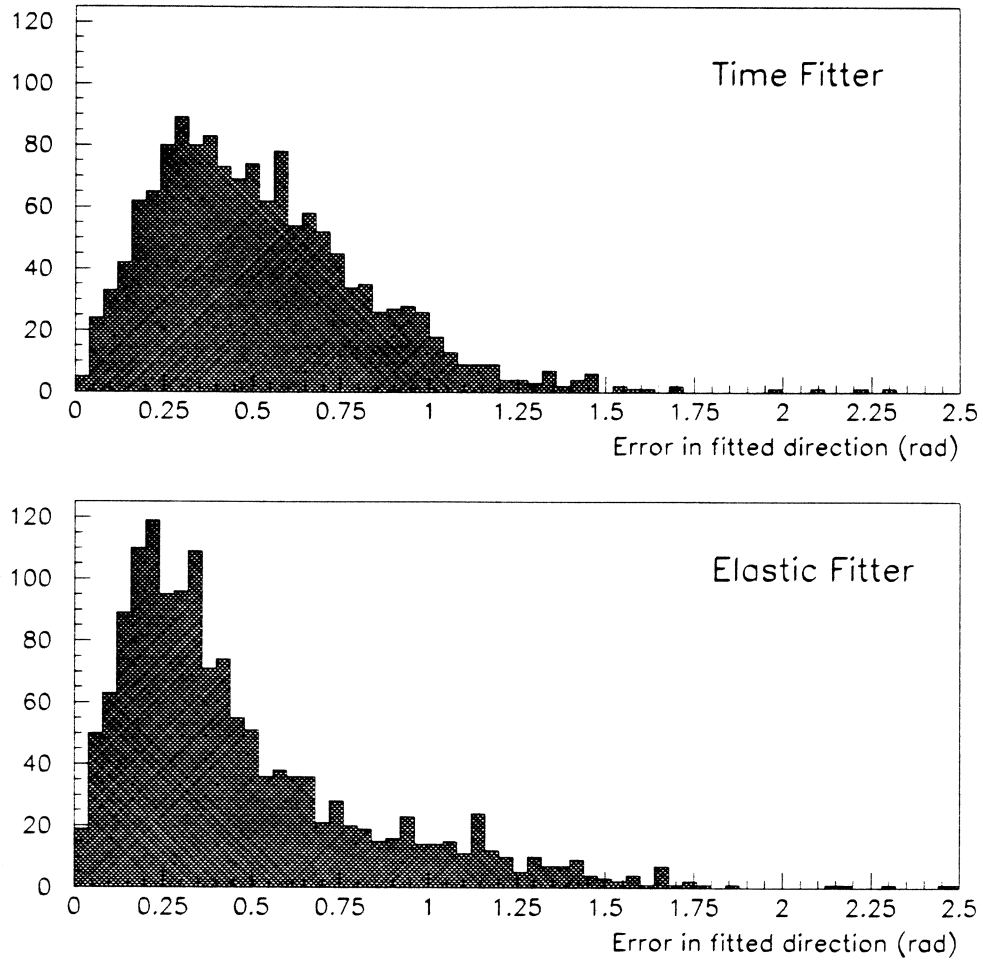


Figure 5.8: The angle (in radians) between the true and fitted direction for the Time Fitter and the Elastic Fitter.

Whilst the improvement in performance of the Elastic Fitter over both the Quad and Time Fitters is clear it will probably be concluded that the complexity of the algorithm is not outweighed by the quality of the fit. However, the techniques used by the Elastic Fitter may provide inspiration for further work.

Shown in Table 5.1 are the results of timing studies on the three fitters under consideration. It can be seen that the Elastic Fitter has a speed comparable to that of the Time Fitter and considerably faster than that of the Quad Fitter. However, as long as the Elastic algorithm needs an initial Quad fit the speed difference between the two is irrelevant. It is clear that for the Elastic Fitter to be viable an alternative initial fit needs to be found. Care must be taken when making these time comparisons, however, as they are specific to 5MeV electron events. The Time and Elastic fitter CPU times are independent or, at worst,

Fitter Method	CPU time (in units of the Time time !!)	% of fits outside 100cm	% of fits outside 300cm
Quad	122	7.3%	0.0%
Time	1	8.1%	4.7%
Elastic	4	2.8%	0.0%

Table 5.1: Timing and fit quality results of the three fitters

linearly dependent on the number of hits, but the size of the Quad cloud and hence the Quad CPU time depends on the $\sim 4^{\text{th}}$ power of NHIT. Therefore, for event data sets where the mean NHIT is greater than 50 the disparity between the Quad CPU time and the others will be much greater.

5.7 Comments and Further Developments

In Appendix B it is argued that late hits within the Čerenkov cone tend to push the fit position back along the track direction and such hits outside the cone pull it along the track direction. With more PMTs outside the Čerenkov cone than inside the usual effect of such fitter pull is to drag the fit position along the direction of the electron. It can also be shown [Klein 94] that a vertex determination made by fitting the hits to a Čerenkov cone tends to push the vertex position back along the electron's direction i.e. in the opposite direction to the pull seen in residual based fitters. [Klein 94] explains this as being due the higher PMT density per unit solid angle as the fit position is moved back against the electron's direction. The original thinking behind the Elastic Fitter was to try to use these two opposing pulls to cancel each other by carrying out the position fit with the distance measure M_i^{tot}

$$M_i^{\text{tot}} = M_i^{\text{pos}} + \gamma M_i^{\text{dir}} \tag{5.17}$$

where γ would be a parameter determining the relative weighting that should be applied to the position and direction components of M_i^{tot}. M_i^{pos} and M_i^{dir} would have their previous definitions and so the event position determination would be a combined residual and cone fit. By carefully choosing the value of γ it was hoped that the two opposing pulls would tend to cancel each other. Unfortunately, and perhaps not surprisingly, this proved not to work too well, the reason being that compared to the sharp precision of a vertex determination produced using M_i^{pos}, a vertex fit using M_i^{dir} is a very blunt instrument and setting γ large enough to effect a partial cancellation of the pulls meant that the accuracy of the position fit was severely compromised. Although this implementation of pull cancellation proved ineffective the basic idea may warrant further study.

In its original incarnation as the Elastic Arms Algorithm the fitter was constructed to simultaneously fit multiple tracks in a TPC event. Although modified

in this chapter, the Elastic Fitter maintains the capability to be simply extended to fit multiple vertices and directions. The extension principally involves defining an M_i for each track/vertex with the λ parameter and the weighting factor making decisions as to which track/vertex a particular hit should be attributed to in the same way that they make the good hit/bad hit decision in the work of this chapter. Such an extension may possibly prove useful when fitting cosmic ray events where the Čerenkov light originates from a large region.

5.8 Conclusion

As has already been stated, the quality of fit produced by the Elastic Fitter may not be enough of an improvement to outweigh the complexity of using the algorithm, but it is certainly true that the technique of simulated annealing, the weighting method used to eliminate bad hits, and the possibility of extending the algorithm to handle multiple vertices demand further attention.

Chapter 6

The Feedforward Neural Network

There is nothing to do with computers that merits a PhD

MAX NEWMAN
Head of the Manchester group which built an early computer

If the human brain were simple enough for us to fully understand it, we would be too simple to do the job

E. PUGH

6.1 History

THE HISTORY of the modelling of brain cell operation dates back to 1943 [McCulloch 43] where the cells were modelled as interconnected two-state units with a number of inputs and one output. Each unit calculated a weighted sum of inputs and switched on if this sum exceeded a threshold, switched off otherwise. The outputs of most units were fed into the inputs of others. It was realised that with an appropriate choice of weights such networks could be designed to perform specific calculations [Minsky 67]. Weight setting algorithms were designed based on a network of two layers of threshold units (known as a perceptron), the output of the first layer feeding into the second. It was soon revealed, however, that such networks were severely limited in their capabilities, being only able to solve linear problems [Minsky 69] and it was at this point that development of this algorithm ceased for almost 20 years.

In the 1970's attention focused on networks of fully connected units, where the output of every unit fed into the input of another. Such networks operated as content addressable memories, where the weights were able to store and recall patterns of activation of the units and were strongly allied to theories of learning

in the brain [Hebb 49]. Another line of research led to the modelling of such networks as a system of magnetically interacting spins and the power of statistical mechanics was brought to bear on the analysis [Caianiello 61].

It was 1985 that saw the great leap forward in the field of neural networks. In this year a number of authors [Rumelhart 86a, Rumelhart 86b, Parker 85, Le Cun 85] picked up the perceptron algorithm where it had been left in 1969 and extended it by using more than two successively connected layers, changing the threshold operation of a unit into a sigmoid function of its weighted inputs, and developing the backpropagation weight setting algorithm. The sigmoid function and increased number of layers meant that the network could handle non-linear problems with backpropagation providing simple and powerful optimisation of the network weights. In the 10 years since 1985 there has been an explosion in the number of neural network algorithms and applications and the field continues to expand rapidly.

6.2 Introduction

In the last five years neural network techniques have begun to be applied to high energy physics problems [Denby 92, Peterson 92]. The parallel nature of the algorithms has produced fast hardware triggers [Denby 95, Moeck 95] and their pattern recognition ability has been used in offline event analysis [Babbage 92, DELPHI 92]. In the context of SNO such techniques are crying out to be used for event classification via hit pattern recognition. In this chapter one such neural net paradigm is introduced: the feedforward network, which is by far the most popular method and is ideally suited to the questions SNO needs answered.

Two kinds of pattern recognition are relevant to the experiment. Event by event classification is needed to identify the nature of individual hit patterns and may be useful when analysing supernova events. For dealing with the solar neutrino flux statistical classification is more relevant and involves taking a sample of hit patterns and finding the fraction that belongs to each event class.

The probability for the occurrence of a galactic supernova during the lifetime of the SNO experiment is fairly low (estimates for the rate vary between 1 in 10 and 1 in 100 years). Should one occur it is predicted that SNO will see $\sim$1000 events and so an event class analysis might proceed by binning the events in time slices and conducting a statistical analysis on each bin. In this case an event by event analysis would not strictly be necessary. An event by event technique might be needed for small time scale analyses of background variations or monthly checks on the solar neutrino flux(es), but it seems likely that all the 'big' questions could be attacked by statistical analyses. For this reason this chapter and particularly Chapter 7 focus on the development and application of statistical classification. A method for event by event pattern recognition is outlined, but not comprehensively developed or applied.

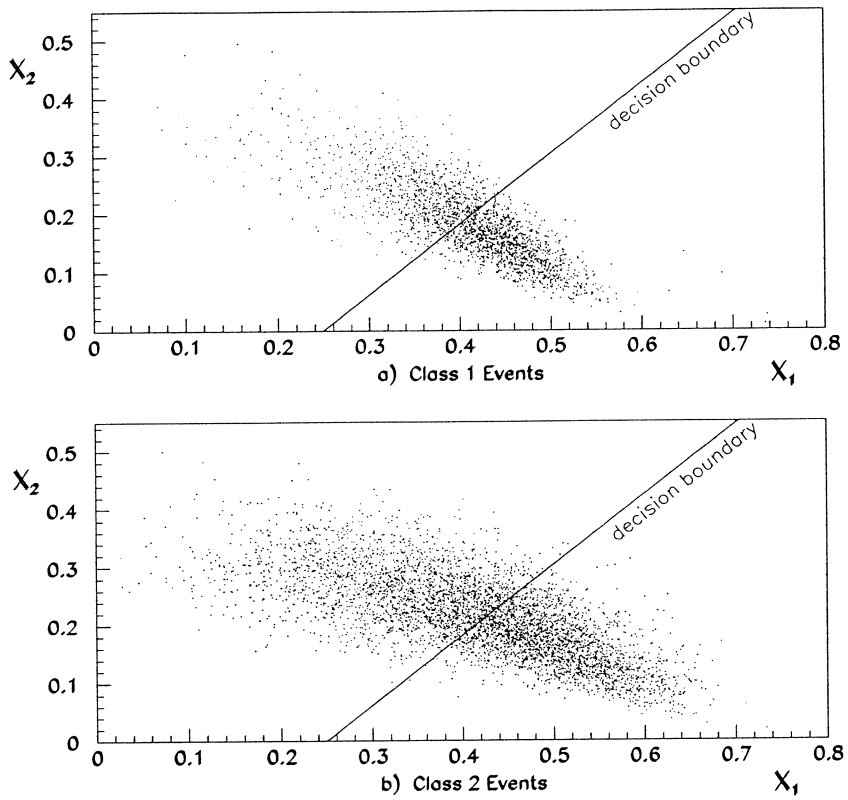


Figure 6.1: Locations in 2D parameter space of a) 5000 Class 1 events b) 5000 Class 2 events

This chapter lays the basis for the results presented in Chapter 7 and its structure separates into three parts. First the feedforward neural net algorithm is described. The clearest way of doing this is via its application to the event by event problem. In the second section the procedure for event by event classification is fleshed out with some ideas on error analysis. The third section moves on to detail the method used to handle the statistical classification task.

6.3 The Algorithm

Given a single PMT hit pattern the task of event classification breaks into two parts: The extraction of a set of parameters from the hit pattern and the classification of the hit pattern based on where this parameter vector lies in parameter space. This classification is best conducted by assigning regions of parameter space to particular event classes. The following sections assume that a set of

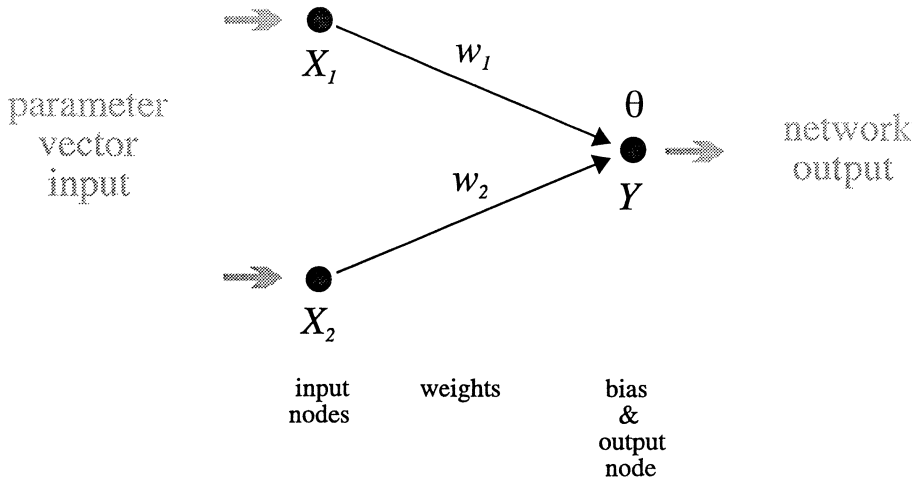


Figure 6.2: A neural network for the simple 2 parameter/2 class example

parameters have been chosen and go on to describe the way in which a neural network carries out the classification.

6.3.1 The Simplest Case

It is convenient to start with the simplest possible example and then to generalise it. Consider the extraction of two parameters from each hit pattern and the classification of hit patterns into Classes 1 and 2. As shown in Fig. 6.1 the two parameters extracted from a hit pattern then describe a point in the 2D parameter space. It can be seen, for the parameters chosen, that the two event classes fall in different but overlapping regions of the space. The best extraction parameters are those for which this overlap is minimised.

The simplest way of dividing the parameter space of Fig. 6.1 into two regions (one for each of the two event classes) is with a linear cut or decision boundary as shown. Events lying above and to the left of the cut are then classified as Class 1 and those below and to the right as Class 2. It can also be seen that with the rather poor parameters chosen a lot of misclassification errors are made in this example. The next two subsections describe how a simple neural network can assign a region/class to the parameters of a hit pattern, but also how the network can optimally place the decision boundary to minimise the misclassification error.

The Operation of a Feedforward Neural Network

A neural network to handle the 2 parameter/2 class example is shown schematically in Fig. 6.2. The parameters are fed into and held in the input nodes X_i. The weights w_i enable a weighted sum of the inputs to be fed into the output node Y. A bias θ is subtracted from the sum and a non-linear sigmoid function

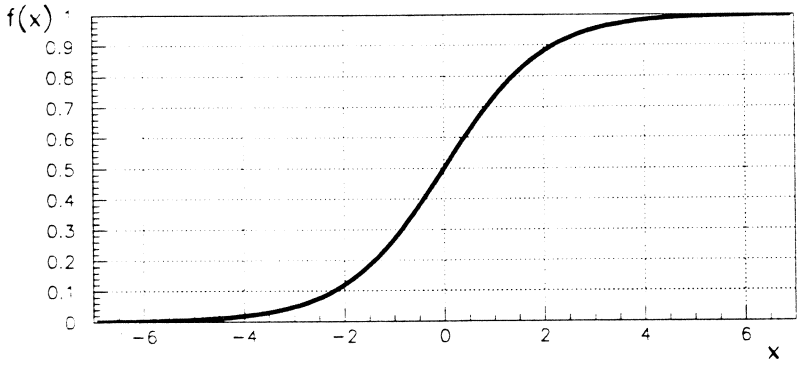


Figure 6.3: The sigmoid function: $f(x) = [1 + \exp(-x)]^{-1}$

(see Fig. 6.3) of this result is calculated by the output node to produce the output of the network. The network is then carrying out the function:

$$Y(\underline{\mathbf{X}}) = \left[1 + \exp - \left(\sum_i w_i X_i - \theta\right)\right]^{-1} \tag{6.1}$$

where $\underline{\mathbf{X}}$ is the 2D vector of extracted parameters. An output of $Y = 0.5$ then corresponds to

$$\sum_i w_i X_i = \theta$$

i.e. the input parameters specify a point that lies on a line in parameter space defined by w_i and θ. An output of $Y > 0.5$ means that the input parameters lie on one side of this line and $Y < 0.5$ shows that they lie on the other side. Therefore, given a linear decision boundary defined by w_i and θ, the network provides a way of determining the region within which the extracted parameter point lies.

Training a Feedforward Network

Given a decision boundary in the parameter space the simple network described in the previous subsection is able to assign the hit pattern to one of the two classes based on whether the output of the net is greater or less than 0.5. On its own this is nothing special, but there also exists a network training technique, known as backpropagation, that enables the optimal placing of the decision boundary i.e the optimal setting of w_i and θ.

Training the network to set its internal parameters requires the use of a training set of hit patterns. This set should contain equal numbers of Class 1 and Class 2

events to avoid *a priori* biasing, and the true class of each event must be known. If a network output of $Y > 0.5$ is to signify a Class 1 event and $Y < 0.5$ a Class 2 then this enables a target network output to be assigned to each hit pattern of the training set:

For true Class 1 events Target T = 1.0
For true Class 2 events Target T = 0.0

With p labelling the hit patterns in the training set then X_i^p are the extracted parameters from hit pattern p and T^p is the target output that the network should try to achieve for that pattern. With these designations it is then possible to define a training error E for a particular set of w_i and θ values.

$$E(w_i, \theta) = \frac{1}{2} \sum_p [Y(\underline{\mathbf{X}}^p) - T^p]^2 \tag{6.2}$$

With a network achieving perfect classification then the network output $Y(\underline{\mathbf{X}}^{\mathrm{P}})$ for each hit pattern p will be equal to the target output T^p for that pattern and the training error E will be zero. The more the network outputs differ from their target values then the higher the value of the training error. This provides a criterion for the setting of the internal parameters of the network. The optimal values of w_i and θ (i.e. the optimal positioning of the decision boundary) are those which minimise the training error E. Starting with w_i and θ set randomly between -1 and 1 they are then updated by gradient descent of the error function.

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

$$\Delta \theta = -\eta \frac{\partial E}{\partial \theta} \tag{6.3}$$

The η parameter above is known as the learning rate and controls the size of the update steps. Iteration continues until the error function has converged to its minimum value. Although gradient descent is the simplest weight update method it works well in most cases. If convergence is slow or there are problems with local minima of the error function then extensions to gradient descent are available (e.g. the addition of a momentum term [Hertz 91]). Calculating the update partial derivatives explicitly yields

$$\Delta w_i = \eta \sum_p [Y(\underline{\mathbf{X}}^p) - T^p][Y(\underline{\mathbf{X}}^p) - 1]Y(\underline{\mathbf{X}}^p)X_i^p$$

$$\Delta \theta = -\eta \sum_p [Y(\underline{\mathbf{X}}^p) - T^p][Y(\underline{\mathbf{X}}^p) - 1]Y(\underline{\mathbf{X}}^p) \tag{6.4}$$

For the simple case under consideration these two subsections have shown how a neural network can optimally set a linear decision boundary in parameter space by altering its internal weights and biases and then with these internal parameters frozen can assign a class to any new parameter vector presented to it.

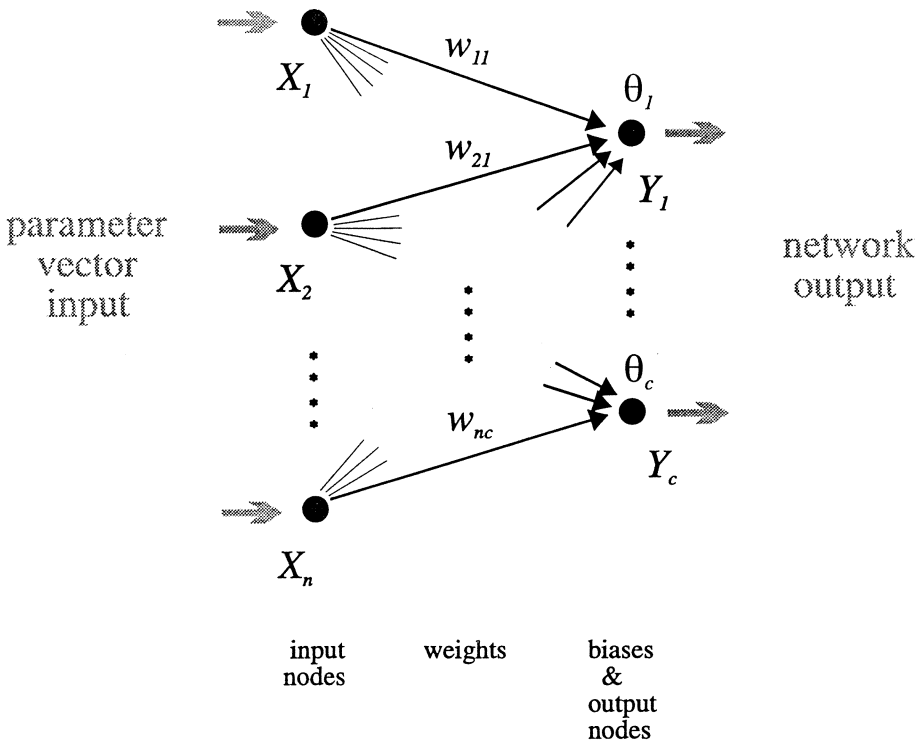


Figure 6.4: The extension of the simple example to n inputs and c outputs

6.3.2 Extending the Simplest Case into Something Useful

What has been described so far is no more than a linear discriminant and there exist a number of equally effective and much simpler statistical techniques for performing the same task (e.g. the Fisher discriminant [Schalkoff 92]). The simplest case can easily be extended to handle more than two extracted parameters and an increased number of event classes, but the real advantage of using a feedforward net, and the point at which it leaves the competition behind, comes when it is generalised to form non-linear or curved decision boundaries (see [Becks 92, Mazzanti 93] for comparisons of neural and classical algorithms applied the problems of high energy physics analysis).

More Inputs and More Outputs

The first generalisation required of the simplest case is to enable the network to deal with more than two extracted parameters. With n parameters then the number of input nodes X_i and the number of weights w_i increases from 2 to n and all the mathematics is identical. This is a trivial extension and simply

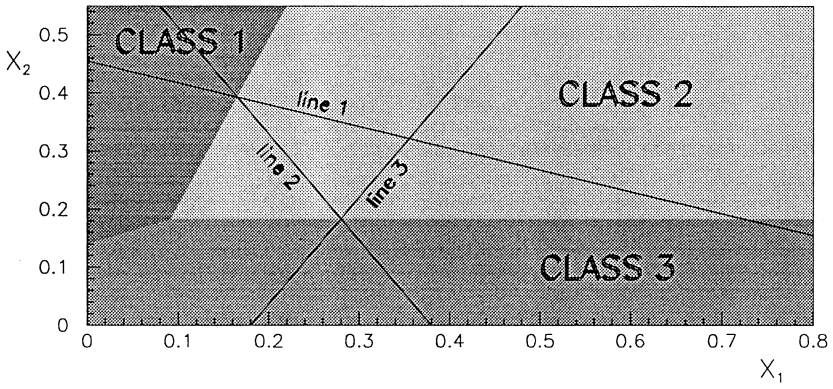


Figure 6.5: The decision regions for the 2 parameter/3 class extension to the simple example. Each output k has an associated line k given by $\sum_i w_{ik}X_i = \theta_k$. Class k then corresponds to the region which is most distant from line k i.e. its distance is most positive.

involves moving from a 2 dimensional parameter space with a single line cut to an n dimensional space with a single hyperplanar cut.

With more than two event classes the required extensions are a little more involved but still simple. To distinguish c classes then c output nodes are required (as opposed to 1 node for the 2 class example which is a special case) as is illustrated in Fig 6.4. There are now weights w_{ik} connecting every input to every output where i labels the input node and k the output node and each output has its own bias θ_k. The vector of output values Y_k is calculated from the inputs X_i by the logical extension of Eqn. 6.1

$$Y_k(\underline{\mathbf{X}}) = \left[1 + \exp - \left(\sum_i w_{ik}X_i - \theta_k\right)\right]^{-1} \tag{6.5}$$

Each output Y_k corresponds to a particular class and an input parameter vector, once fed through the net, is assigned to the class corresponding to the highest output. To see more graphically how this works consider the two parameter simple example, but now with classification into three event classes i.e. a network with two input and three output nodes. For each output node the equation $Y_k = 0.5$ once again corresponds to a line in parameter space as is shown in Fig. 6.5.

For each output the argument of its sigmoid function

$$\sum_i w_{ik}X_i - \theta_k$$

is simply the signed distance from this line. Since the sigmoid is a monotonically increasing function then each output value Y_k is a nonlinear measure of the signed distance of the input parameter vector from the line labelled by k. Therefore, by assigning the class of an input vector to that of its highest output the parameter space is broken up into regions. As shown in Fig. 6.5 for the 3 output example each region is labelled by the line from which it is furthest i.e. region k corresponds to output Y_k being highest. It is in this way that the weights w_{ik} and biases θ_k break up the n dimensional parameter space into c regions.

To train such a multi-output network now requires a c dimensional target vector for each hit pattern. This vector has 1 in the slot of the target class and all other entries are zero. An error function is defined as a direct extension of Eqn. 6.2 with the output $\underline{\mathbf{Y}}$ and target $\underline{\mathbf{T}}^p$ now being vectors.

$$E(w_{ik}, \theta_k) = \frac{1}{2} \sum_p [\underline{\mathbf{Y}}(\underline{\mathbf{X}}^p) - \underline{\mathbf{T}}^p]^2 \tag{6.6}$$

The optimal setting of w_{ik} and θ_k then proceeds by gradient descent of this error function as before.

Non-linear Boundaries and Hidden Nodes

Even with the extension to n extracted parameters and c event classes the network described so far is still just a linear discriminant with hyperplanar region boundaries. With a further extension the feedforward network becomes a non-linear classifier with curved region boundaries. This is achieved by inserting a layer of hidden nodes between the input and output nodes as shown in Fig. 6.6.

The network has n inputs X_i, h hidden nodes H_j and c outputs Y_k. The input parameters are fed through to the output nodes by the logical extension of the previous examples. Each hidden node calculates a weighted and biased sigmoid of the inputs with weights w_{ij} and biases θ_j and then each output node calculates a weighted and biased sigmoid of the values from the hidden nodes using weights W_{jk} and biases Θ_k. Overall then the network executes the function

$$Y_k(\underline{\mathbf{X}}) = \left[1 + \exp - \left(\sum_j W_{jk} \left[1 + \exp - \left(\sum_i w_{ij} X_i - \theta_j\right)\right]^{-1} - \Theta_k\right)\right]^{-1} \tag{6.7}$$

Alternatively this can be expressed as

$$H_j(\underline{\mathbf{X}}) = \left[1 + \exp - \left(\sum_i w_{ij} X_i - \theta_j\right)\right]^{-1}$$

$$Y_k(\underline{\mathbf{X}}) = \left[1 + \exp - \left(\sum_j W_{jk} H_j(\underline{\mathbf{X}}) - \Theta_k\right)\right]^{-1} \tag{6.8}$$

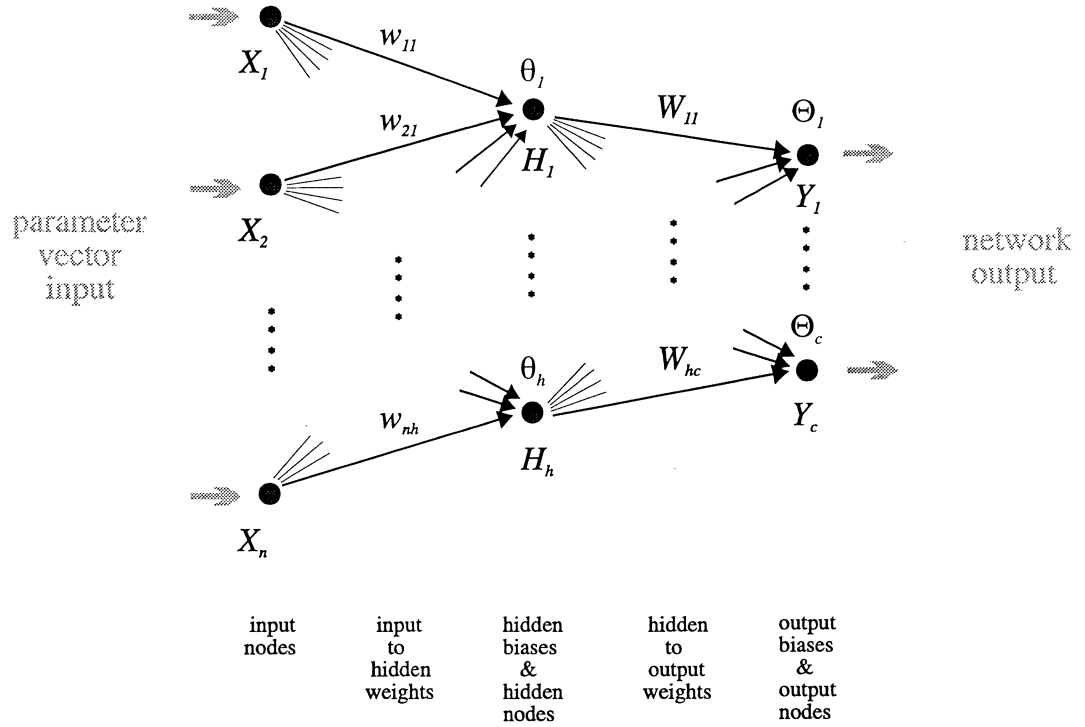


Figure 6.6: A fully fledged feedforward network with n input, h hidden, and c output nodes

With two input parameters the straight lines of Fig. 6.5 now become more general curves with corresponding curved region boundaries. This is entirely due to the use of the sigmoid function which, until now, has been an unnecessary encumbrance. To demonstrate this mathematically is rather messy and not particularly illuminating.

The training of such a multilayer network uses the error function of Eqn. 6.6

$$E(w_{ij}, \theta_j, W_{jk}, \Theta_k) = \frac{1}{2} \sum_p [\underline{\mathbf{Y}}(\underline{\mathbf{X}}^p) - \underline{\mathbf{T}}^p]^2 \tag{6.9}$$

and the weights and biases are updated by the usual gradient descent equations

$$\Delta W_{jk} = -\eta \frac{\partial E}{\partial W_{jk}}$$

$$\Delta \Theta_k = -\eta \frac{\partial E}{\partial \Theta_k}$$

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

$$\Delta\theta_j = -\eta \frac{\partial E}{\partial \theta_j} \tag{6.10}$$

It is instructive to calculate these partial derivatives explicitly. First make the definitions

$$\delta_k^p \equiv Y_k^p[Y_k^p - 1][Y_k^p - T_k^p] \tag{6.11}$$

$$\delta_j^p \equiv H_j^p[H_j^p - 1]\sum_k W_{jk}\delta_k^p \tag{6.12}$$

where δ_k^p is known as the error of output node Y_k for pattern p and δ_j^p is the error of hidden node H_j for pattern p. With these assignments then the updates for the weights and biases become

$$\begin{aligned}
\Delta W_{jk} &= \eta \sum_p \delta_k^p H_j^p \\
\Delta\Theta_k &= -\eta \sum_p \delta_k^p \\
\Delta w_{ij} &= \eta \sum_p \delta_j^p X_i^p \\
\Delta\theta_j &= -\eta \sum_p \delta_j^p
\end{aligned} \tag{6.13}$$

These equations have a rather satisfying interpretation. The updates for the hidden to output weights W_{jk} are formed by combining a hidden node value H_j^p with an output node error δ_k^p. These output node errors are then weighted and summed to form hidden node errors. The input to hidden weights w_{ij} are then updated by combining an input node value X_i^p with a hidden node error δ_j^p. Overall then network training proceeds by propagating extracted parameter vectors from input to output and then backpropagating errors from output to input in order to update the weights and biases. Hence the algorithm is known as a feedforward neural network with backpropagation.

The extension of the network to include more than one hidden layer is perfectly feasible, but almost certainly unnecessary for SNO purposes. This follows principally from a theorem which states that any functional mapping can be approximated to arbitrary accuracy by a feedforward neural network with just one hidden layer [Cybenko 89]. The proof of this uses the fact that the sigmoids of an infinite number of hidden nodes form a complete set of functions. Second and third hidden layers are used, however, in many applications as a task can often be accomplished with less hidden nodes overall by using more hidden layers. Since the extraction of parameters from hit patterns constitutes a significant amount of preprocessing the decision regions required of a network for SNO event classification are likely to be singly connected, simple spaces requiring only one hidden layer to form them.

The inclusion of a hidden layer has introduced a second externally set parameter into the network algorithm (the learning rate η is the first) - the number of

hidden nodes has to be chosen. With too few hidden nodes there is insufficient flexibility to form the necessary decision boundaries and with too many a process known as overtraining can occur where the network starts to learn the features of individual input patterns rather than the characteristics of the probability distributions from which they are drawn. To understand this overtraining an analogy with curve fitting is useful. Any physicist knows the danger of fitting a set of data points with a curve that has too many adjustable parameters. What results is a curve that fits the points perfectly but says very little about the underlying distribution from which the data was taken. Just the same problem occurs with the adjustable weights and biases of the network and the number of training patterns. To avoid overtraining the size of the training set of input vectors must be significantly larger than the number of weights and biases of the network. If W is the number of weights and biases and ϵ is average network error per input pattern (the network error being given by Eqn. 6.9 for a set of patterns independent of those used to train the network) then it can be shown [Hertz 91] that the number of training patterns P should satisfy

$$P > \sim \frac{W}{\epsilon} \tag{6.14}$$

in order to avoid overtraining.

Whilst there exist some very elegant pruning and growing techniques to kill off or add hidden nodes during the training process, a network for SNO event classification probably has no need of them. This is once again due to the simplicity of the required decision boundaries and the limited number of hidden nodes required to form them. Typically the task of SNO event classification requires a number of hidden nodes that is of the same order as the number of input nodes and often considerably less. So long as the inequality of Eqn. 6.14 is satisfied then the classification ability of the network will not be affected by having more hidden nodes than is strictly necessary.

6.4 Event by Event Hit Pattern Recognition

6.4.1 The Task

Given a single PMT hit pattern the task of event by event classification breaks into two parts:

1. Extract a number of parameters from the hit pattern. These parameters should describe the pattern and capitalise on the intrinsic differences between patterns from the various event classes.

2. Break up the parameter space into fixed regions with each region being assigned to an event class. The parameter values extracted from a hit pattern

then specify a point in parameter space and the region within which this point falls then gives the event class.

The procedure outlined above hinges on the appropriate choice of extracted parameters. This choice is far from obvious, and is dealt with in the next chapter.

The preceding section has shown how a neural network can optimally break up the extracted parameter space and produce a classification of a hit pattern. Some additional steps now need to be added to turn this into a useful event by event system.

6.4.2 A Useful Result

For a network trained on an infinitely large pattern set then the output $Y_k(\underline{\mathbf{X}}^p)$ for a particular pattern is the probability of that pattern belonging to class k [Gish 90]. To demonstrate this consider the network error of the output corresponding to Class k

$$E_k = \frac{1}{2}\sum_p [Y_k(\underline{\mathbf{X}}^p) - T_k^p]^2$$

For an infinitely large training set this error becomes

$$E_k = \frac{1}{2}\int \left(P(\text{Class } k|\underline{\mathbf{X}})\,[Y_k(\underline{\mathbf{X}}) - 1]^2 + P(\overline{\text{Class } k}|\underline{\mathbf{X}})\,[Y_k(\underline{\mathbf{X}})]^2 \right) P(\underline{\mathbf{X}})d\underline{\mathbf{X}}$$

where $P(\underline{\mathbf{X}})$ is the probability distribution function of the input $\underline{\mathbf{X}}$, $P(\text{Class } k|\underline{\mathbf{X}})$ is the probability of the input $\underline{\mathbf{X}}$ belonging to Class k, and $P(\overline{\text{Class } k}|\underline{\mathbf{X}})$ is the probability that it doesn't. A necessary and sufficient condition for a trained network is that the derivative of this error with respect to the output Y_k be zero. Therefore

$$\int \left(P(\text{Class } k|\underline{\mathbf{X}})\,[Y_k(\underline{\mathbf{X}}) - 1] + P(\overline{\text{Class } k}|\underline{\mathbf{X}})\,Y_k(\underline{\mathbf{X}}) \right) P(\underline{\mathbf{X}})d\underline{\mathbf{X}} = 0$$

Since $P(\underline{\mathbf{X}})$ is arbitrary this equation requires that the term in brackets be zero and so

$$Y_k(\underline{\mathbf{X}}) = \frac{P(\text{Class } k|\underline{\mathbf{X}})}{P(\text{Class } k|\underline{\mathbf{X}}) + P(\overline{\text{Class } k}|\underline{\mathbf{X}})}$$

Since the sum of the two conditional probabilities must be unity then

$$Y_k(\underline{\mathbf{X}}) = P(\text{Class } k|\underline{\mathbf{X}}) \tag{6.15}$$

This shows that the actual numerical value of the output corresponding to Class k is the probability of the input vector belonging to Class k. In the parlance of Bayesian statistics the network is calculating an *a posteriori* probability for each class and as such is behaving as an optimal Bayesian discriminant [Duda 73]. Although this has been demonstrated for an infinitely large training set it can be

shown that as the size of a training set increases then convergence to Eqn. 6.15 is quite rapid and the result is essentially true for even a moderately large number of training patterns. This means that on an event by event basis the network will assign events to particular classes and give a figure for its confidence in that assignment.

When training a network with a view to event by event classification one is not so much interested in getting the network outputs as close to their targets as possible, but rather getting the class assignment of the network right. In this case one would like the minimisation of the network error to concentrate on the few bad outputs which are far from their targets, rather than pushing the many good outputs even closer to their targets. To bias the training in this way one can simply remove from the backpropagation procedure any outputs which are within some distance of their target (0.1 would be a typical distance to use). Applying such a bias invalidates the Bayesian derivation above and networks trained in such a way have outputs that do not give the *a posteriori* probability. It is therefore necessary to make a choice between slightly worse classification performance, but with the assignment of a confidence probability, and improved performance, but with no probability attached.

6.4.3 Quantifying Overall Network Performance

Once the network has been trained its overall performance can be tested using a labelled data set independent of that used for training. The results of this testing are best displayed using the charmingly named *confusion matrix*. An example of this for a testing set comprising 5000 Class 1 and 5000 Class 2 events is shown in Fig. 6.7 (this example can be extended from 2 to any number of classes). The rows of the matrix are labelled by the true class of the events and the columns by the network assignment. The matrix entry (i, j) is the number of events that belong to class i and which the network assigned to class j. The goal of perfect event by event hit pattern classification is then a diagonal confusion matrix.

		Network Assignment	
		1	2
True Class	1	3684	1316
	2	1274	3726

Figure 6.7: Example of a confusion matrix for a testing set consisting of 5000 Class 1 and 5000 Class 2 events

Different authors extract different numbers from the confusion matrix as indicators of overall network performance. In this work the following definitions are made:-

$$\text{Class 1 purity} = \text{P(correct|Class 1)} \approx \frac{\text{number of class 1 events correct}}{\text{number of class 1 events}}$$

$$\text{Class 2 purity} = \text{P(correct|Class 2)} \approx \frac{\text{number of class 2 events correct}}{\text{number of class 2 events}}$$

$$\text{Overall purity} = \text{P(correct)} \approx \frac{\text{number of events correct}}{\text{number of events}}$$

where P(correct|Class k) is the probability that the network gets the assignment right given that the event was truly Class k and P(correct) is just the probability that the network gets its right (with the constraint that there are an equal number of events in each class).

Focusing on the overall purity, if p is the true probability that the network assigns a randomly selected pattern correctly then the number of correct assignments C from a testing set of N patterns has a binomial probability distribution

$$P(C) = \frac{N!}{C!(N-C)!} p^C (1-p)^{N-C} \tag{6.16}$$

The expectation value of C is then Np and so $\frac{C}{N}$ is an estimate of the overall purity where C is given by the trace of the confusion matrix. Since the variance of the binomial is

$$\sigma^2 = Np(1-p)$$

then it is possible to put an error on the overall purity estimate by using the unbiased estimate of the width of the distribution. Bringing all this together then the overall purity can be estimated as

$$\text{Overall purity} = \frac{C}{N} \pm \left[\frac{1}{N-1}\frac{C}{N}\left(1-\frac{C}{N}\right)\right]^{\frac{1}{2}} \tag{6.17}$$

where the 1σ error is understood to be the width of a binomial rather than the usual Gaussian width. Analogous procedures can be followed to obtain the Class 1 and Class 2 purities and the formalism can easily be extended to networks distinguishing more than two classes. For the testing results of Fig. 6.7 the overall purity is $(74.1 \pm 0.4)\%$.

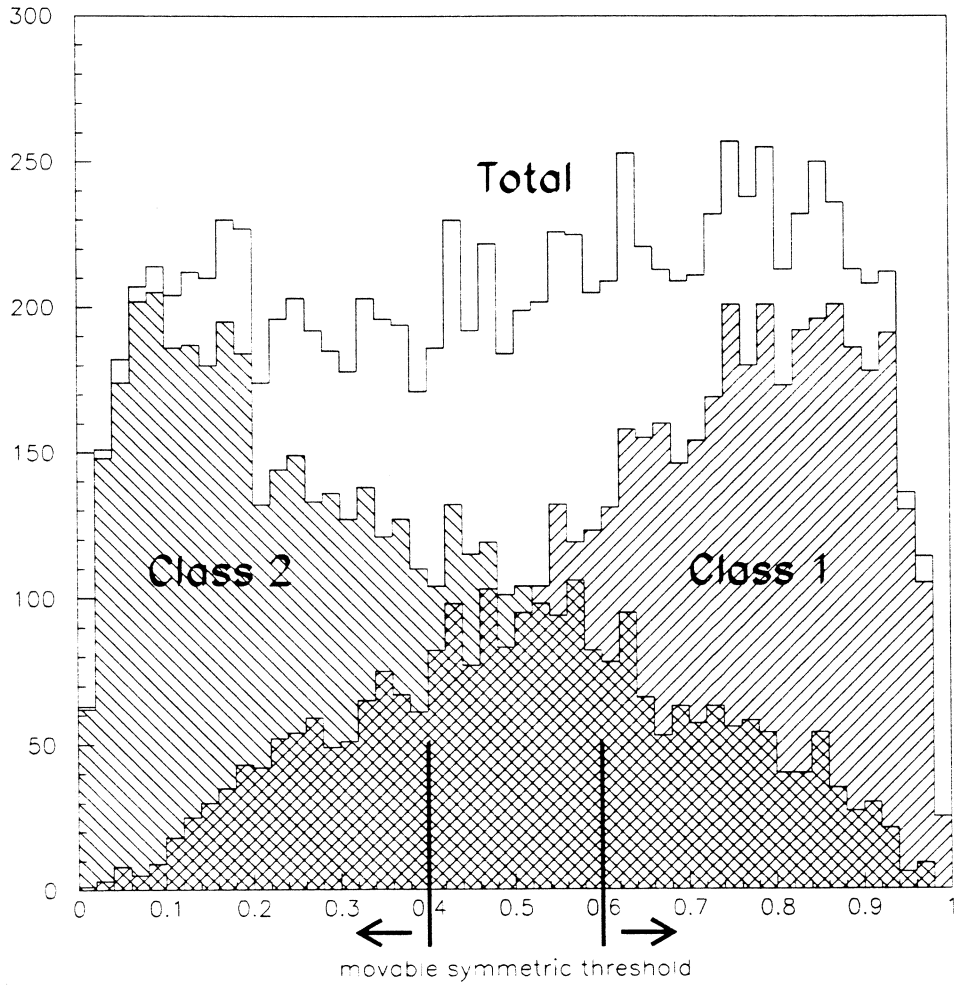


Figure 6.8: A histogram of the testing set output values of a network with one output node trained to distinguish between Class 1 and Class 2 events. The two hatched histograms show the output values for the 5000 true Class 1 and 5000 true Class 2 events. The sum of these two histograms is also shown.

Fig. 6.8 shows a histogram of the network output values for the same testing set used to produce the confusion matrix of Fig. 6.7. The network has a single output and has been trained to distinguish between Class 1 and Class 2 events. The testing set consists of 10000 events, 5000 from each class. Events with an output value greater than a half are assigned to Class 1 and those with an output less than a half to Class 2 and in this way the confusion matrix of Fig. 6.7 is produced.

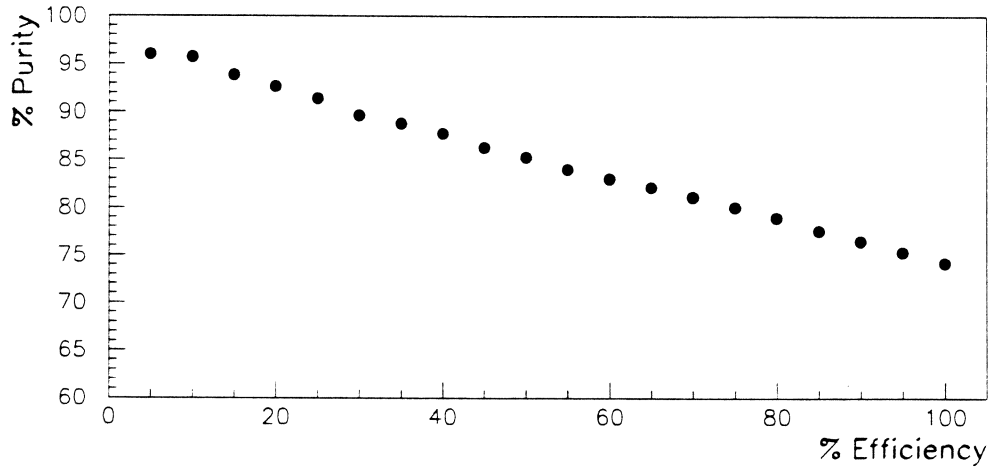


Figure 6.9: The variation of network purity with overall efficiency for the testing set of outputs from Fig. 6.8 as the symmetric threshold is moved. The error bars vary from 0.4% to 1% and are not shown.

As was shown in the last section the numerical value of an output for a particular input pattern is a measure of the network's confidence in the class assignment it has made. For the single output network of Fig. 6.8 output values around 0.5 indicate that the network is not confident about the class assignment and confidence increases as the output value tends to zero or one. For a multi-output network the closer a particular output value is to unity the more confident is the class assignment. This provides a method of increasing overall purity at the expense of overall efficiency. A multi-output network assigns an input pattern to the class corresponding to the highest output value, but if this value fails to exceed some threshold then the event can be thrown out as being unclassifiable. Similarly, a pair of thresholds for a single output network can be moved out symmetrically from 0.5 as shown in Fig. 6.8 so that events falling between the two thresholds are discarded. The efficiency of a network at a particular threshold setting is simply the fraction of events which are not discarded. As the threshold becomes more stringent and efficiency drops then the overall purity will increase. This purity can still be calculated from Eqn. 6.17 with the appropriate values for C and N. Fig. 6.9 shows how network purity varies with efficiency as the symmetric threshold is moved for the example outputs of Fig. 6.8. In should be born in mind that, unless the output histograms for true Class 1 and true Class 2 events (see Fig. 6.8) are symmetric about 0.5, separate Class 1 and Class 2 efficiencies should be used.

6.5 Statistical Hit Pattern Recognition

The preceding sections have shown how a feedforward neural network can achieve event by event hit pattern classification. The next sections will show how these techniques can be adapted to achieve statistical hit pattern recognition.

6.5.1 The Task

Given a set of hit patterns the task is to ascertain the fraction of the set which belongs to each event class. This is achieved in three parts:

1. Decide on a set of extraction parameters and train a network in just the same way as for event by event classification.

2. Take a labelled set of data, independent of that used for training, and feed it through a trained network. With the network outputs from this data form a histogram for each event class (hereafter known as a PDF histogram).

3. Take the real data set and feed it through the trained network. Fit the resulting distribution of output values to a linear combination of PDF histograms. Hence find the fraction of the data set belonging to each event class.

Once again it will be assumed that Part 1 has been achieved and the next section will go on to describe Parts 2 and 3. First, however, a diversion into the handling of networks with more than one output is required.

6.5.2 PDFs from Networks with More than One Output

When two classes of event are separated by a network only one output is needed and events from each class tend to cluster near 0 or 1 in the output space as was indicated in Fig. 6.8. To separate c classes it is conventional (though not strictly necessary) to use a network with c outputs (as was described in Sec. 6.3.2). The trained network then tends to cluster events from a particular class in such a way that a particular output tends to be high, i.e. each individual output deals with a particular class. If an event is deemed by the network to belong to a particular class then a certain output will be high, otherwise it will be low.

It was shown in Sec. 6.4.2 that, for a large training set, the actual numerical value of an output is the posterior probability of the event belonging to that class. Normalisation of the probability then means that for any event fed into the network the sum of the output values must equal unity (or be very close to unity given that the training set is not infinite). Individual output values are also constrained to lie between 0 and 1.

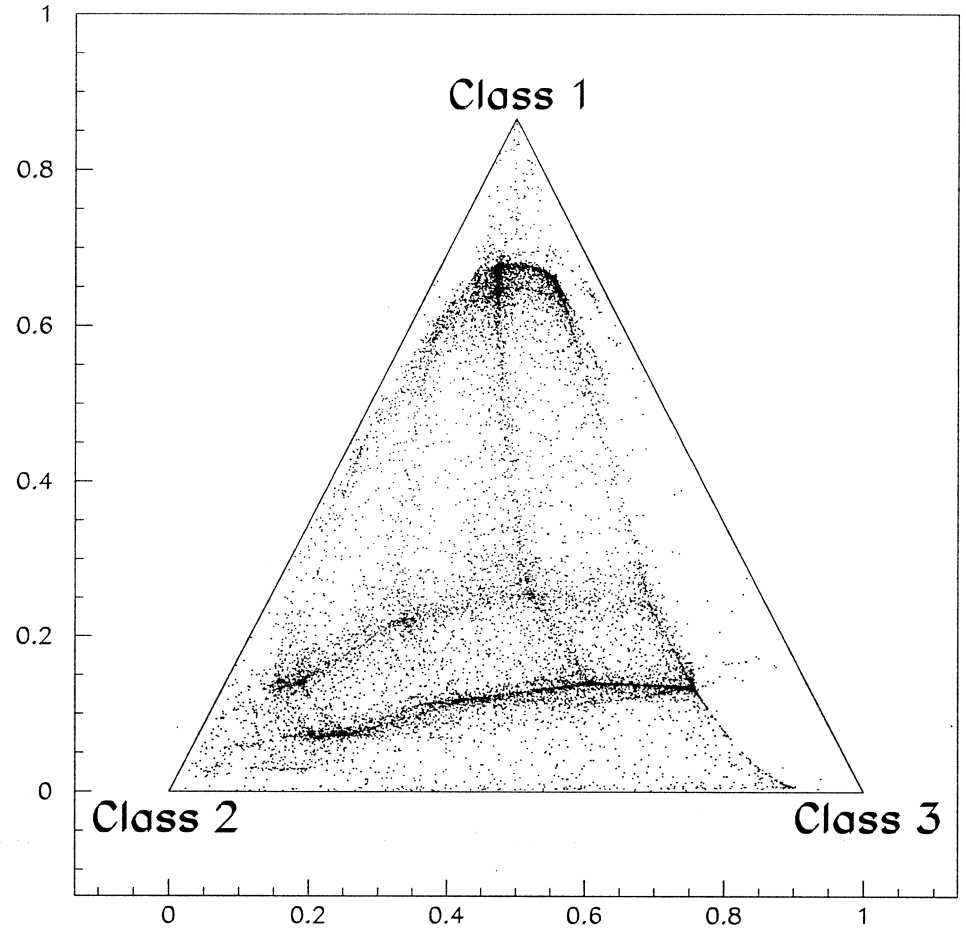


Figure 6.10: A scatter plot of the output of a network designed to distinguish between three classes. Plotted are the two independent numbers produced by transforming the 3D output vector according to Eqn.6.18.

For a network with c outputs labelled Y_k these constraints can be written

$$\sum_{k=1}^{c} Y_k = 1$$

$$0 < Y_k < 1$$

This means that events are constrained to lie on an $c-1$ dimensional subspace (a hyperplane) of the c dimensional output space. To take a three output network separating three classes as an example, events are constrained to lie on the plane $Y_1 + Y_2 + Y_3 = 1$ in the positive octant i.e. within a two dimensional equilateral triangle. Each of the three classes will tend to cluster near its own corner of this

triangle. The output of the network can therefore be described by two Cartesian coordinates in the plane of the triangle.

By taking the 3D output vectors (Y_1, Y_2, Y_3) and applying the following transformation

$$\frac{1}{Y_1 + Y_2 + Y_3} \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{\sqrt{3}}{2} & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \tag{6.18}$$

then such a 2D vector is produced as shown in Fig. 6.10. This scatter plot also shows the clustering of each class near its own corner and is the three class analogue of the two class PDFs of Fig. 6.8.

6.5.3 Getting PDF Histograms

The network topology and training for statistical pattern recognition is identical to that already described for event by event recognition. It is convenient, however, to interpret the resulting algorithm in a somewhat different way. Rather than viewing the network as making a decision on each input pattern based on the network output it is better to think of the network as carrying out a transformation of the extracted parameters from the n dimensional input space to an $c-1$ dimensional output space. The training process then maximises the separation in output space of the probability distribution functions (PDFs) for each event class. From this viewpoint the network is merely a functional mapping.

Once the network has been trained, the PDF histograms for each event class are formed. This is done by simply feeding events through the network whose class is known and histograming the resulting outputs, one histogram for each event class. The two histograms labelled Class 1 and Class 2 in Fig. 6.8 are examples of such PDFs, in this case for a single output network.

6.5.4 Fitting

The procedure for extracting the fraction of events belonging to each class in a set of real data simply consists of fitting the real data network output histogram to a linear combination of the PDF histograms. This can be done most easily with a χ^2 fit. With the following assignments:-

N = Number of real data events
G_i = Network output histogram of real data (unnormalised)
g_i = Network output histogram of real data (normalised to unity)

N^k = Number of training events belonging to class k
F_i^k = PDF histogram for class k (unnormalised)
f_i^k = PDF histogram for class k (normalised to unity)

A^k = Number of events in the real data set belonging to class k
α^k = Fraction of events in the real data set belonging to class k

then the χ^2 function can be written as

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

where

$$\begin{aligned} \sigma_i^2 &= \sigma_{G_i}^2 + \textstyle\sum_k (A^k \sigma_{f_i^k})^2 \\ &= G_i + \textstyle\sum_k \frac{A^k}{N^k} A^k f_i^k \\ &= N \left[g_i + \textstyle\sum_k (\frac{N}{N^k} \alpha^k) \alpha^k f_i^k \right] \end{aligned}$$

Entirely using normalised histograms the χ^2 can therefore be written as

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

This function is then minimised with the constraint that the α^k sum to unity to find the fraction of events in each class. Statistical uncertainties can be obtained from the error matrix in the usual way [Eadie 71].

In addition to using χ^2 each fit can also be carried out using a maximum likelihood technique specifically designed for fits of one histogram to a linear combination of others where there are statistical uncertainties on all histograms [Eberhard 93]. As well as allowing the α^k to vary this technique regards the f_i^k as estimators for the true PDF histograms and allows their values to be tweaked according to Poisson statistics.

6.6 Summary

Event by Event Analysis

1. Chose a set of parameters to extract from each hit pattern.
2. Train the network.
3. Feed data events through the network one by one. The network output assigns each event to a class and gives a measure of its confidence in the assignment.

Statistical Analysis

1. Chose a set of parameters to extract from each hit pattern.
2. Train the network.
3. Take a fresh set of labelled events, feed them through the network, and form an output PDF histogram for each event class.
4. Feed the data events through the network, form the data output histogram, and fit it to a linear combination of the PDF histograms thus obtaining the fraction of data events in each class.

Chapter 7

Neural Network Analysis Results

Art is the imposing of a pattern on experience, and our aesthetic enjoyment in recognition of that pattern.

ALFRED NORTH WHITEHEAD
Dialogues of Alfred North Whitehead, 1953, Chapter 29

Appearances to the mind are of four kinds. Things either are what they appear to be; or they neither are, nor appear to be; or they are and do not appear to be; or they are not and yet appear to be. Rightly to aim in all these cases is the wise man's task.

EPICTETUS c. 50-120
Discourses, Book 1, Chapter 27

7.1 Introduction

THE PREVIOUS chapter has prepared the theoretical foundations of the feed-forward neural network, and defined and detailed the concepts of event by event and statistical hit pattern analysis. In this chapter this framework is applied to the task of separating signal classes in the SNO detector. Neural techniques have been applied to the analysis of Čerenkov images before (e.g. [Akkila 93, Altherr 92, Vazquez 92]), but almost entirely to data from air shower or RICH detectors where the energy regime, detector geometry, and particle tracks and multiplicities are radically different to those of SNO and so require very different analysis approaches. The work of this chapter almost exclusively centres on the statistical analysis of solar neutrino signals, but with a brief mention of event by event separation.

The chapter starts by isolating those signal classes which are of interest and giving their interaction rates. Then the parameters to be extracted from each hit pattern are discussed and defined. The final preparatory step is to outline a few likely SNO detector run scenarios and their effect on extracted parameters and class separations. After this results are presented, first for the event by event

case and then for a statistical separation. An analysis of the network mapping and a preliminary discussion of systematics are presented before the chapter concludes with sections on calibrating the Monte Carlo and further applications of the network algorithm.

7.2 Relevant Event Classes

7.2.1 Isolating the Classes

The first task is to decide which event classes should be considered in the analysis. Monte Carlo studies [Skensved 94] indicate that with an NHIT threshold at 60 all backgrounds (except those producing neutrons) are negligible. NHIT spectra for the major signal and background classes are shown in Fig. 7.1. The analysis is therefore conducted with this 60 hit threshold and means that only 4 classes of event need to be considered. These are precisely defined to be:-

(CC)	Charge Current	$\nu_e + {}^2\mathrm{H} \rightarrow p + p + e^-$
(ES)	Electron Scattering	$\nu_x + e^- \rightarrow \nu_x' + e^{-\prime}$
(NC)	Chlorine Neutron Capture	$n + {}^{35}\mathrm{Cl} \rightarrow {}^{36}\mathrm{Cl} + \gamma\mathrm{s}$
(ND)	Deuteron Neutron Capture	$n + {}^2\mathrm{H} \rightarrow {}^3\mathrm{H} + \gamma$

Since these studies are entirely conducted with Monte Carlo data, and the network algorithm is sensitive to quite fine details of the hit patterns, it is important to satisfy oneself that the particle trajectories and subsequent angular distributions of Čerenkov light are accurately simulated for the above event classes. This is a lengthy and ongoing process, one particular aspect (the modelling of the Cl gamma cascade) is outlined in Appendix D.

7.2.2 Signal Rates

The approximate number of events per year which satisfy the NHIT cut and reconstruct within a radius of 7m can be obtained using SNOMAN and is shown in Table 7.1 (see Appendix E for the justification of these numbers). Standard Solar Model and best fit MSW rates are given for both a pure D_2O and a chlorine fill. These numbers do not include background neutrons.

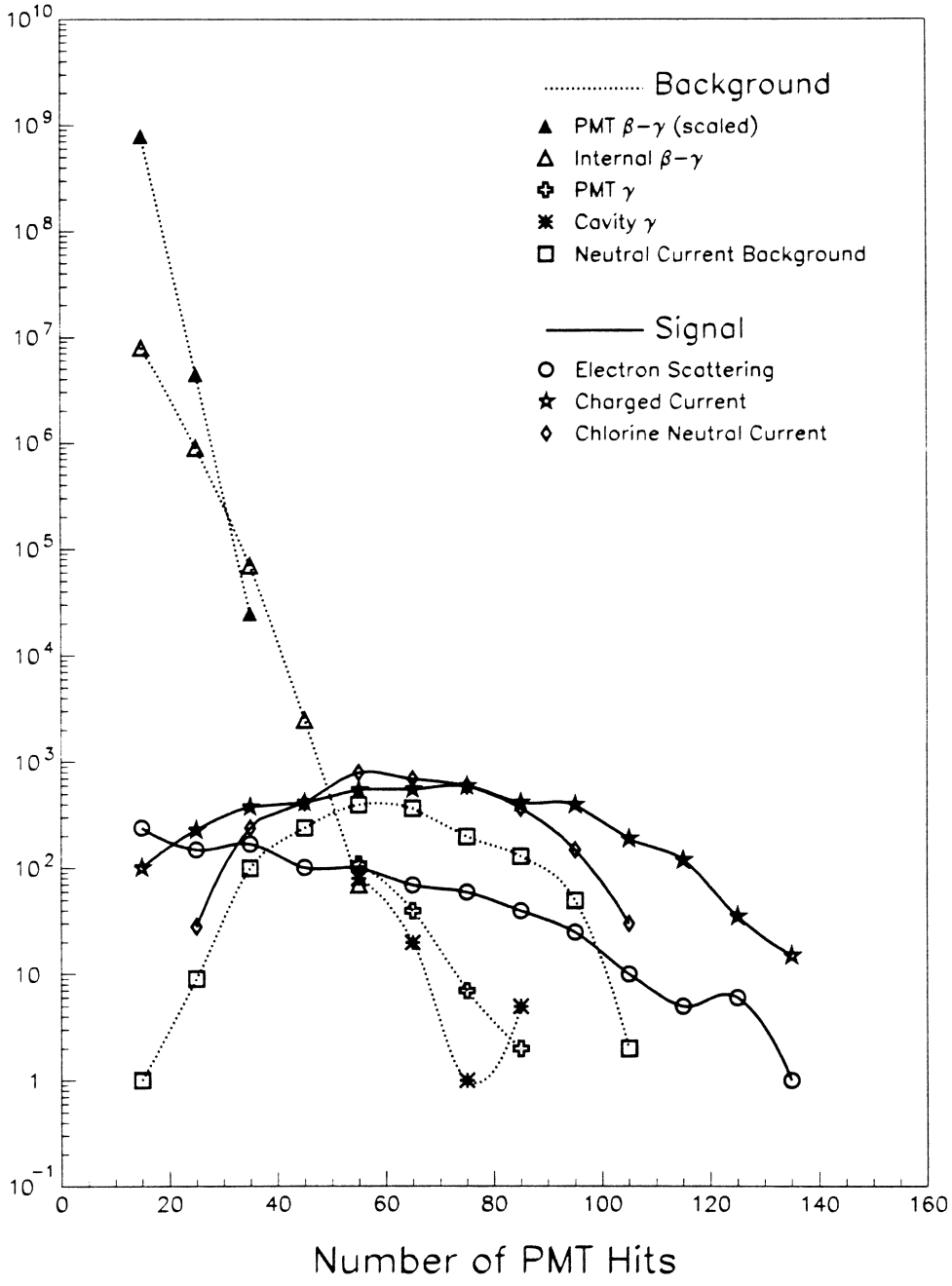


Figure 7.1: The NHIT spectra of the major signal and background classes for 1 year of running with salt added to the D_2O.

		CC	ES	NC	ND
1 Year Cl	SSM	8832	1208	2585	62
1 Year Cl	MSW	3397	536	2585	62
1 Year D_2O	SSM	8832	1208	0	634
1 Year D_2O	MSW	3397	536	0	634

Table 7.1: The number of events with NHIT > 60 which reconstruct within 7m

The numbers are those used in the analysis of this chapter and are not intended to be completely accurate, but rather to represent the spread in rates that SNO can probably expect.

7.3 Hit Pattern Features

Having decided upon the event classes, the next step is to define a set of parameters which will describe each hit pattern. In order to do this it is necessary to delineate the hit pattern features which are relevant. One can then go on to find parameters which cover each feature. Of the four event classes that are to be separated the most important are the CC and NC events. The hit patterns from CC result from a single electron and those from NC from usually more than one electron. It is this difference, combined with the contrasting NHIT and direction distributions, that determines which hit pattern features are relevant. The complete list comprises:-

- **Spatial Hit Distribution**:- This feature covers the positions of the hits that comprise an event. If the Čerenkov light is considered to arise from a point (an approximation that is reasonably valid even for gamma cascades and $\beta\gamma$ events, given the resolution of SNO) then the entire spatial hit distribution is held in the set of θ and ϕ values, giving the direction of each hit from the event position.

- **Temporal Hit Distribution**:- If all the Čerenkov light genuinely arises from a point then the information held in the PMT hit times is redundant given that held in the distances of each hit from the event position (this excludes scattered and reflected light). However, this pointlike approximation is known to break down for some event classes and in such cases there will be additional information in the temporal hit distribution.

- **Spatio-Temporal Hit Distribution**:- There will also be useful information in the correlations between hit positions and hit times and there should be parameters which mix spatial and temporal information.

- **Solar Angle**:- The charged current and electron scattering cross-sections have a distinctive angular dependence. This means that there is useful

information in the angle between the fitted direction of an event and the direction of the sun.

- **Number of Hits**:- Different event classes have very different NHIT spectra, but great care must be taken with this parameter, as described below.
- **Radial Position of Event**:- The distribution of event radius for neutral current events shows a characteristic drop off as the acrylic vessel is approached. This occurs because a neutron travels between about 30 and 50 cm before capturing in salt filled D_2O, but a much shorter distance when the transport is through light water. There is therefore a smaller volume from which neutrons are emitted such that they capture on the outskirts of the D_2O than there is for neutrons which capture well within the heavy water. This is behaviour not seen in the other signal classes

In the analysis of this chapter all of the above event features have been used in characterising hit patterns with the exception of the radial position of the event. This potentially valuable feature has been omitted because it is distorted, to an *a priori* unknown degree, by the neutral current background originating in the acrylic vessel.

The NHIT spectra can be predicted for all event classes with the exception of charged current and electron scattering events. For these two an MSW effect may distort the ν_e spectrum to an *a priori* unknown degree. This means that unless the network can be trained on real data sets the NHIT parameter should probably not be used. It also means that great care must be taken to ensure that none of the other parameters have a strong NHIT dependence. That being said a certain level of NHIT dependence is unavoidable. For a fixed underlying angular distribution of Čerenkov light, the more hits an event has the closer the hit pattern will approach the underlying distribution. This means that the spread in the values of a spatial parameter for a particular event class will become smaller as the energy/NHIT of the event is increased. This introduces an irreducible, though minor, energy dependence into the spatial and spatio-temporal parameters. Sec. 7.8 discusses the small systematic error that this NHIT dependence induces when the true ν_e spectrum is distorted.

To a good approximation the detector can be taken to be spherically symmetric and, with the exception of solar correlations, the distribution of event directions is isotropic. With the solar directional correlations being taken care of by the solar angle parameter, these facts strongly suggest that the parameters describing the spatial hit distribution should be rotational invariants. This means that a spatial hit pattern occurring on one side of the detector will have exactly the same extracted parameters as the same hit pattern transformed by a rotation about the event position. This rotational invariance gives a fixed number of spatial parameters far greater descriptive ability, as well as removing from them the errors inherent in a fitted direction.

7.4 The Extracted Parameters

This section takes each feature outlined in Sec. 7.3 and describes the parameter or parameter set which encapsulates the information. The first step in extracting parameters from a hit pattern is to fit a position for the event (a direction is also needed for the solar angle parameter). This position is then used as the coordinate origin.

7.4.1 Spatial — The Harmonic Parameters

First define the spatial hit pattern function:-

$$f(\theta,\phi) = \frac{1}{N}\sum_{i=1}^{N} \delta(\cos\theta - \cos\theta_i)\,\delta(\phi - \phi_i) \tag{7.1}$$

which is a function of the polar and azimuthal angles of each hit with respect to the event position with a delta spike centred on each hit. With N as the number of hits this function is the spatial hit pattern. Next decompose this function into its spherical harmonic components. (This procedure is closely linked to that of [Proriol 92]).

$$f(\theta,\phi) = \sum_{l,m} \alpha_{lm} Y^*_{lm}(\theta,\phi) \tag{7.2}$$

The α_{lm} can be found in the usual way

$$\begin{aligned} \alpha_{lm} &= \int f(\theta,\phi) Y_{lm}(\theta,\phi)\, d\Omega \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N} Y_{lm}(\theta_i,\phi_i) \end{aligned}$$

The next step is to define a set of rotationally invariant quantities β_l

$$\begin{aligned} \beta_l &= \textstyle\sum_m |\alpha_{lm}|^2 \\ &= \tfrac{1}{N^2}\textstyle\sum_{i,j}\sum_m Y_{lm}(\theta_i,\phi_i) Y^*_{lm}(\theta_j,\phi_j) \\ &= \tfrac{2l+1}{4\pi N^2}\textstyle\sum_{i,j} P_l(\cos\omega_{ij}) \end{aligned}$$

where ω_{ij} is the angle between the hits labelled i and j. This last line makes explicit the fact that the β_l quantities are rotationally invariant. However, the double summation includes the combination of each hit with itself, and writing this part explicitly yields

$$\beta_l = \frac{2l+1}{4\pi N^2}\left[N + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} P_l(\cos\omega_{ij})\right] \tag{7.3}$$

It is clear from this last equation that the β_l quantities are explicitly dependent upon the number of hits. This dependence can simply be subtracted away to give

$$\begin{aligned}\beta_l' &= \frac{N}{N-1}\left(\beta_l - \frac{2l+1}{4\pi N}\right) \\ &= \frac{2l+1}{2\pi N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} P_l(\cos\omega_{ij})\end{aligned}$$

These β_l' quantities are the harmonic parameters which encode a spatial hit pattern in a rotationally invariant way.

7.4.2 Temporal — The Residual Parameters

First define the residual R_i for the i^{th} hit

$$R_i = |\mathbf{r}_{\text{fit}} - \mathbf{r}_i| - v(t_{\text{fit}} - t_i) \tag{7.4}$$

where $\mathbf{r}_{\text{fit}}$ and t_{fit} are the fitted event position and time, $\mathbf{r}_i$ and t_i are position and time of the i^{th} hit, and v is the speed of light in water. This residual quantity is the difference between the vertex to hit distance as measured by the vertex and hit positions and the vertex to hit distance as measured by the vertex and hit times. It is the same quantity that is minimised by the residual based fitters of Chapter 5.

The mean of the residual distribution of an event can be defined in the usual way.

$$\overline{R} = \frac{1}{N}\sum_{i=1}^{N} R_i \tag{7.5}$$

as can the central moments of the distribution

$$\mu_l = \frac{1}{N-1}\sum_{i=1}^{N}\left[R_i - \overline{R}\right]^l \tag{7.6}$$

where l is integer.

With μ_0 defined as $\overline{R}$, the μ_l are the residual parameters encoding the timing information of the hit pattern.

7.4.3 Spatio-Temporal — The Residual Harmonic Parameters

By incorporating the residual for each hit within the spatial hit pattern function defined earlier the **full** hit pattern function can be defined:-

$$F(\theta,\phi) = \frac{1}{N}\sum_{i=1}^{N} R_i\,\delta(\cos\theta - \cos\theta_i)\,\delta(\phi - \phi_i) \tag{7.7}$$

This particular form for incorporating timing information is somewhat arbitrary and the job could be done in a number of other ways. The form chosen here is merely the simplest. Just as with the harmonic parameters decompose this function into spherical harmonic components.

$$F(\theta,\phi) = \sum_{l,m} A_{lm} Y^*_{lm}(\theta,\phi) \tag{7.8}$$

and define the rotationally invariant quantity γ_l

$$\gamma_l = \sum_m |A_{lm}|^2 \tag{7.9}$$

Manipulate and remove the NHIT dependence as before to give

$$\gamma'_l = \frac{2l+1}{2\pi N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} R_i R_j P_l(\cos\omega_{ij}) \tag{7.10}$$

These γ'_l quantities are the residual harmonic parameters which encode the hit pattern along with its timing information in a rotationally invariant way.

7.5 Analysis Scenarios

Having decided on the classes to separate and the parameters to extract from hit patterns, it is necessary to obtain labelled training sets of data in order to optimise the network mapping. It is clearly preferable that these should come from the detector as opposed to Monte Carlo, and for the ND and NC event classes this can be achieved with a neutron source during pure D_2O and chlorine runs respectively. For the CC and ES classes the situation is a little more tricky. It is probably the case that a *8* decay source will not be an adequate mimic of these classes because of the light scattering effects of the source itself and the sensitivity of the network algorithm to such differences (this sensitivity is, after all, the origin of the algorithm's power). It is, however, possible to obtain a pure data set of mixed CC and ES events by adding a neutron poison to the D_2O [Moorhead 95b]. This would enable training of the network on data entirely derived from the detector itself, with all the credibility and avoidance of systematics that that entails. However, powerful reasons exist for not adding a poison (the occurrence of a supernova during a neutral current insensitive poison run dogs the nightmares of all SNO physicists!) and for this reason CC and ES training sets may have to come from Monte Carlo. It should also be noted that whilst the NC and ND events may show sufficient differences to be separated by a neural algorithm there appears to be no point in attempting this to obtain neutrino flux results as both classes indicate the same thing - the presence of a free neutron in the detector. Attempting to separate NC from ND would, however, be a useful systematic check as the ratio of their rates should simply be the ratio of their known capture cross-sections.

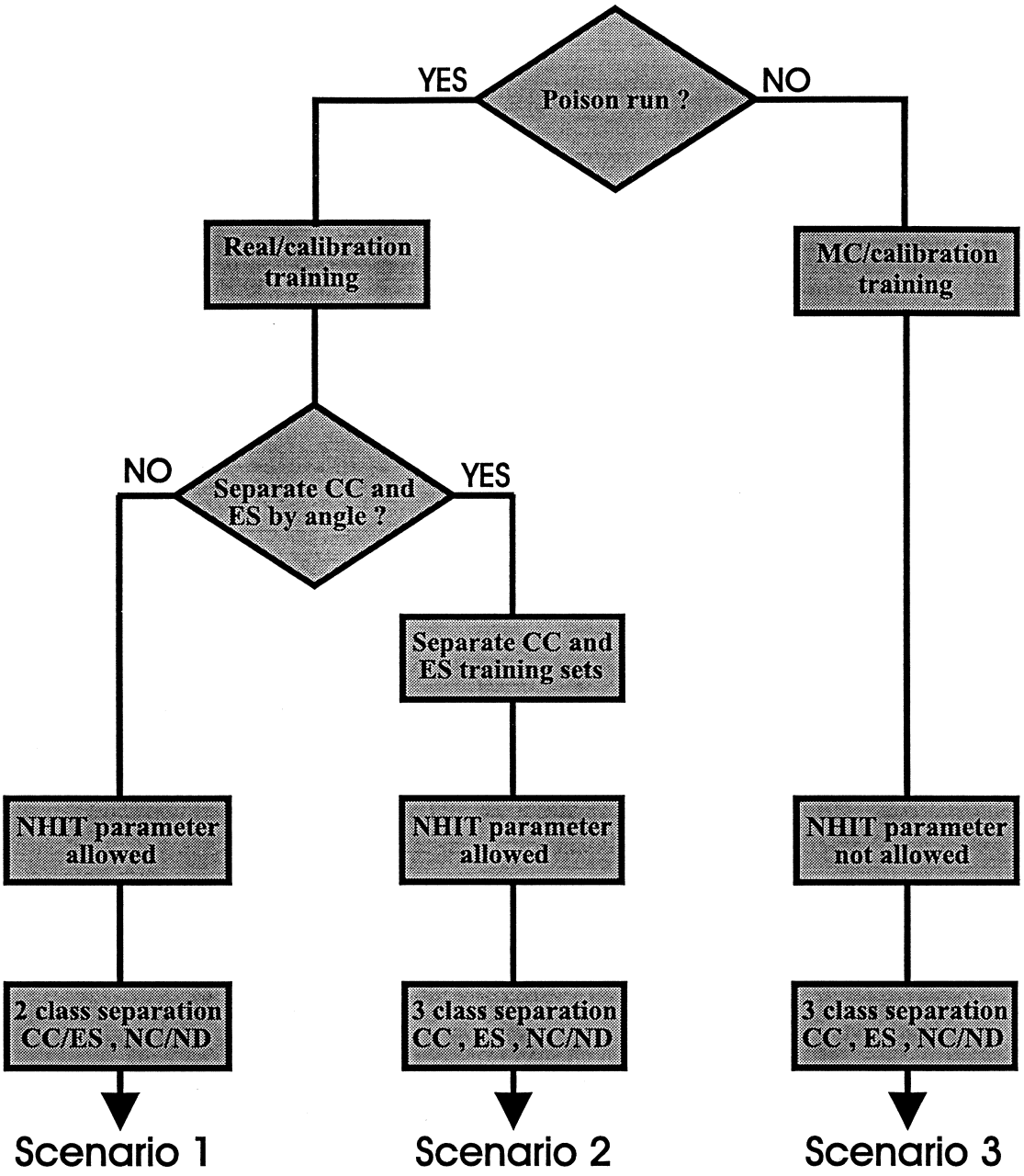


Figure 7.2: A flow chart defining the three operating scenarios assumed in the analysis of this paper

With the preceding paragraph in mind the analysis is conducted under the three different scenarios defined in the flow chart of Fig. 7.2.

- **Scenario 1** With a poison run and a neutron source calibration then a training set of mixed CC/ES and a set of mixed NC/ND can be obtained. A network can then be trained to separate these two classes and NHIT information can be used since real CC/ES data is available.

- **Scenario 2** Since the angular cross-section for ES events is so sharply forward peaked it should be possible to take the mixed CC/ES data set from a poison run and separate it on an event by event basis into a CC and an ES data set by simply using the angle between fitted direction and the sun. A network can now be trained to distinguish between the three classes CC, ES, NC/ND with NHIT information also allowed.

- **Scenario 3** If a poison run is not carried out then CC and ES training sets must be obtained from Monte Carlo meaning that NHIT information should not be used as the true spectrum is uncertain. A network is trained in this case to distinguish the three classes CC, ES, NC/ND.

Scenario 1 is clearly the most preferable as it uses data derived entirely from the real detector and for the same reason Scenario 3 is to be least favoured.

Scenario 2 has been included for completeness only and is seriously flawed on two counts: firstly it is not clear that a good enough direction fit can be made to reliably separate CC and ES on a event by event basis, and secondly there will probably be insufficient ES events to properly train a network to pick out this class alone.

7.6 Results

When optimising the network function, training data sets are used in the analysis which contain 5000 events for each class. So a network for two class separation is trained with 10000 events and one for three classes uses 15000 events. Such numbers of events are easily obtained from Monte Carlo or calibration, but may not be possible with a year of poison data. The figure of 5000 is chosen so as to completely avoid overtraining (see Chapter 6), but this number could be taken as low as 2000 and still produce the same network mapping function. Consequently the use of more CC/ES training events in this paper than will probably be available from a poison run is not deemed to be a problem.

Throughout the analysis the data has been produced using SNOMAN 2_08 with the vertex fit being made with the Grid Fitter [Moorhead 95a] and the direction fit using the Elastic Fitter of Chapter 5.

The extracted parameters used are:-

	9	Harmonic parameters
	9	Residual parameters
	9	Residual harmonic parameters
	1	Solar angle parameter
and, where allowed,	1	NHIT parameter

This amounts to 28 or 29 parameters.

The network architectures have 20 hidden nodes and either 1 or 3 output nodes depending upon the number of classes that are separated. The three different analysis scenarios therefore use network architectures defined by Table 7.2.

	No. input nodes	No. hidden nodes	No. output nodes
Scenario 1	29	20	1
Scenario 2	29	20	3
Scenario 3	28	20	3

Table 7.2: Network architectures for the three analysis scenarios

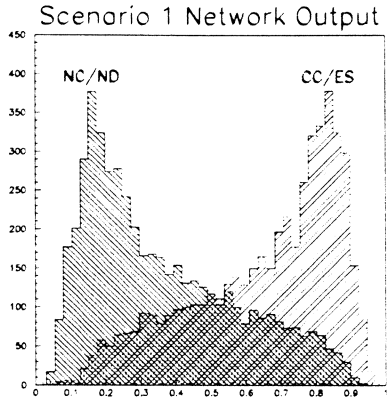
For event by event analysis the results are not encouraging and the technique is not pursued further. For statistical analysis excellent results are obtained. This analysis is conducted for two detector configurations (Cl and pure D_2O fill) in each case all three analysis scenarios are investigated (Scenarios 1, 2, and 3 as defined by Fig. 7.2) for both types of neutrino flux assumptions (SSM and MSW as defined by Table 7.1). This gives a total of 12 fits with each histogram being divided into approximately 100 bins. Both the χ^2 and Maximum Likelihood techniques of Chapter 6 are used on each fit. In all cases the two methods give the same results to within statistical uncertainties.

All training and testing of the network is carried out with an excellent neural network simulator known as SNNS which is produced by the University of Stuttgart [Zell 95]. It is available free from `ftp.informatik.uni-stuttgart.de` via anonymous FTP, and handles feedforward networks as well as many other neural algorithms.

7.6.1 Event by Event Results

Shown in Fig. 7.3 are the network output histograms and confusion matrix for a network trained to distinguish CC/ES from NC/ND i.e Scenario 1. The testing set of data consists of 5000 events from each of the two classes.

The overall purity of 75% means that approximately a quarter of event by event class assignments will be wrong and this isn't good enough for practical



		Network Assignment	
		CC/ES	NC/ND
True Class	CC/ES	3809	1191
	NC/ND	1295	3705

Overall Purity = $(75.1 \pm 0.4)\%$

Figure 7.3: Network output histograms and confusion matrix for a network trained according to Scenario 1

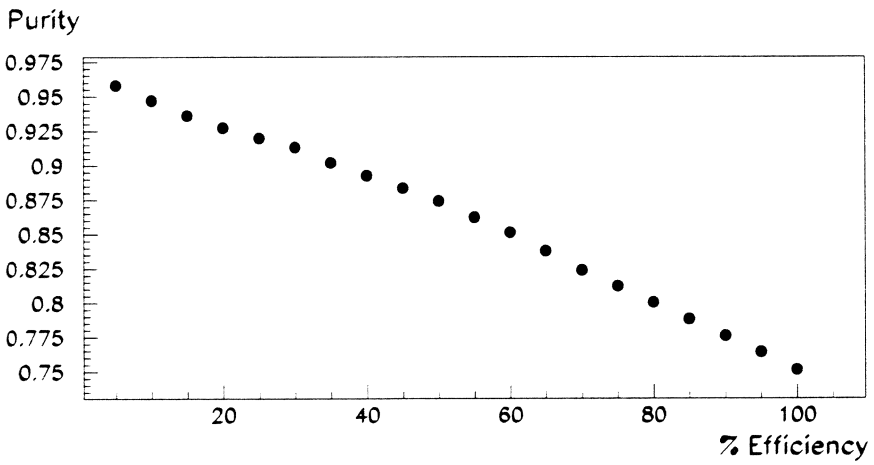


Figure 7.4: The overall purity versus efficiency plot for a Scenario 1 network. (See Sec. 6.4.3 for the definitions of purity and efficiency.)

application. This conclusion holds for Scenarios 2 and 3 as well, where the overall purities are 74% and 72% respectively. Even with the increase in overall purity achieved by lowering efficiency (as shown in Fig. 7.4) respectable purities of ~95% are only reached once the efficiency has fallen to ~10%. Event by event analysis is therefore not deemed viable with this choice of event classes and extraction parameters and is not pursued further in this work. As was stated at the start of Chapter 6 all of the 'big' questions can be handled with statistical classification and the rest of this chapter deals exclusively with that task.

7.6.2 Statistical Results:- 1 Year Cl Fill

The results of the fits to 1 year of Cl fill data and the corresponding PDF histograms and scatter plots are shown in Fig. 7.5.

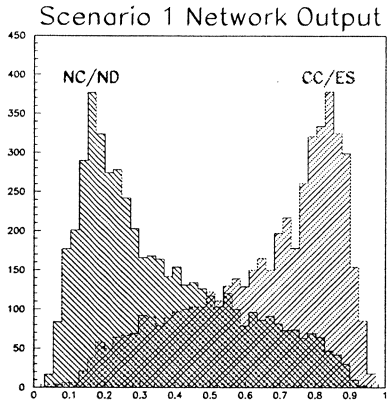
It should be noted that with the fit class fractions constrained to sum to unity the uncertainties for each class are NOT independent. In all the results presented the ES and NC/ND uncertainties are independent (subject to off diagonal error matrix elements) whereas the CC or CC/ES uncertainty quoted is the RMS of the ES and NC/ND errors. The results show that, given a year of Cl fill data, the CC rate can be determined to ~2% statistical accuracy, the ES rate to ~7%, and the NC/ND rate to ~3%. It is interesting to note the increased uncertainties when going from Scenario 2 to Scenario 3, which differ only in that NHIT is allowed in 2, but not in 3. This indication that NHIT is an important separating parameter is born out by a comparison of the PDF scatter plots for Scenarios 2 and 3. The separation in the plot for Scenario 2 is clearly superior to the rather compressed distribution of points in the Scenario 3 plot. That being said, the statistical uncertainties of Scenarios 2 and 3 are pretty close and show that NHIT, although important, is not crucial.

7.6.3 Statistical Results:- 1 Year Pure D_2O Fill

The results of the fits to 1 year of pure D_2O data and the corresponding PDF histograms and scatter plots are shown in Fig. 7.6.

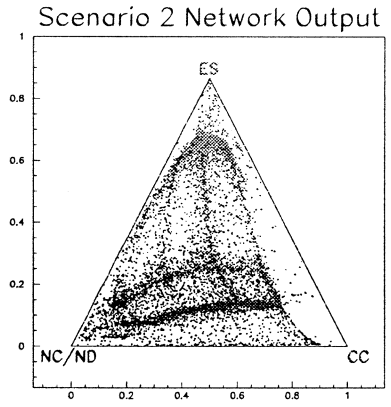
As with the results from the Cl fill, with a year of pure D_2O data, the CC and ES rates can be determined to ~2% and ~7% respectively. The smaller number of ND events means, however, that the uncertainty in their rate increases to ~15%.

On the face of it the results from the D_2O run are surprisingly good, comparing very favourably with the Cl results. This surprise is based on the assumption that what enables CC and NC events to be separated is the difference between an event that produces a single particle and one that produces more than one, be they electrons or gammas. Indeed, an analysis of the Klein-Nishina formula that governs Compton scattering indicates that one can usually expect more than one significant electron from the γ cascade of an NC event, but less often from the single 6.25 MeV γ of an ND event [Bowler 96]. So why is the ND separation so good? It must admitted that the answer is not yet known for sure, but probably lies in the NHIT spectra. A comparison of Figs. 2.3 and 2.4 shows that the ND NHIT spectrum is more unlike the CC spectrum than is the NC NHIT distribution, being narrower and peaking at a lower value. It is possible that the unavoidable, slight NHIT dependence of the spatial parameters enables the network to distinguish ND from CC where they differ most significantly at high NHIT, whilst not introducing much of a systematic error when the ν_e spectrum is distorted at the low NHIT end (see Sec. 7.8). This hypothesis is speculative and is currently being investigated [Bowler 96, Thorman 94].



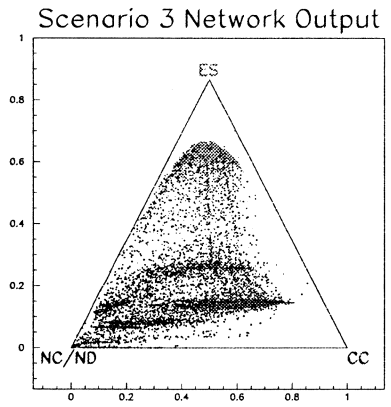
Scenario 1 Fit

	CC/ES	NC/ND
SSM fit	10119(116)	2568(116)
SSM true	10040	2647
MSW fit	3963(90)	2617(90)
MSW true	3933	2647



Scenario 2 Fit

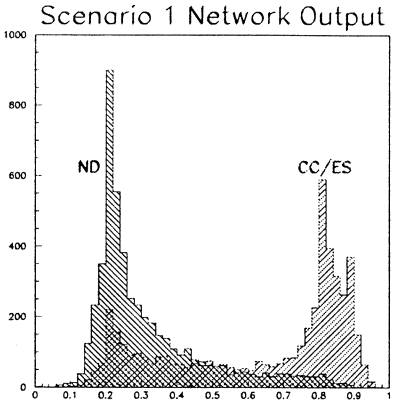
	CC	ES	NC/ND
SSM fit	8871(132)	1220(69)	2596(113)
SSM true	8832	1208	2647
MSW fit	3415(99)	525(47)	2640(87)
MSW true	3397	536	2647



Scenario 3 Fit

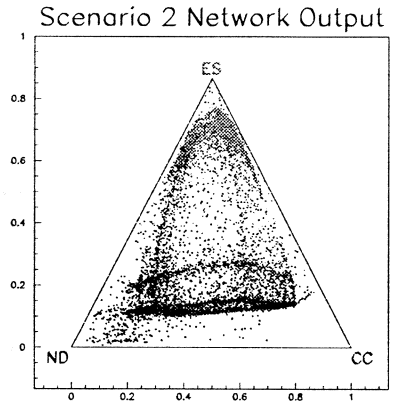
	CC	ES	NC/ND
SSM fit	8837(143)	1236(70)	2614(125)
SSM true	8832	1208	2647
MSW fit	3437(108)	549(48)	2594(97)
MSW true	3397	536	2647

Figure 7.5: The fit results and PDF histograms and scatter plots for each of the three analysis scenarios applied to Cl fill data



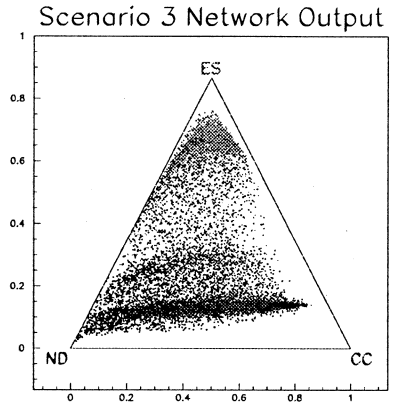
Scenario 1 Fit

	CC/ES	ND
SSM fit	10157(105)	517(105)
SSM true	10040	634
MSW fit	3838(73)	729(73)
MSW true	3933	634



Scenario 2 Fit

	CC	ES	ND
SSM fit	8893(124)	1220(67)	561(104)
SSM true	8832	1208	634
MSW fit	3334(85)	541(46)	692(72)
MSW true	3397	536	634



Scenario 3 Fit

	CC	ES	ND
SSM fit	8818(141)	1211(67)	645(125)
SSM true	8832	1208	634
MSW fit	3306(96)	536(45)	725(85)
MSW true	3397	536	634

Figure 7.6: The fit results and PDF histograms and scatter plots for each of the three analysis scenarios applied to pure D_2O data

7.7 Discussion

7.7.1 Picking Apart the Network Mapping

The most distinct feature of the scatter plots of Figs 7.5 and 7.6 is the amount of detailed structure within them. Shown in Fig. 7.7 are a smoothed histogram version of the Scenario 2 scatter plot of Fig. 7.5, the scatter plot itself, and a simple diagram picking out the line structures within the plot. The histogram is composed of seven ridges with peaks occurring where the ridges intersect. By cutting on the events that form each ridge and then looking at their extracted parameters (i.e. their network inputs) it is possible to roughly identify the ridges with particular properties of hit patterns. These are outlined below:-

- **Ridge 1:** Events pointing directly away from the sun.
- **Ridge 2:** Events pointing in the hemisphere away from the sun.
- **Ridge 3:** Events pointing in the hemisphere towards the sun.
- **Ridge 4:** Events which are abnormally isotropic.
- **Ridge 5:** No clear characteristics besides the events being somewhat more isotropic.
- **Ridge 6:** No clear characteristics besides NHIT being quite high.
- **Ridge 7:** Events with a large number of hits.

These gross properties of the ridges are exactly what would be expected. The ES events are picked out entirely on the basis of their sharp forward peaking in a direction away from the sun, with a weaker distinction between forward and backward hemispheres for the CC and NC/ND classes. Differences between isotropy and NHIT then serve to distinguish between CC and NC/ND events.

The preceding paragraph begs the question:- If such simple characteristics are all the network is using then why not avoid the neural network technique entirely and construct some cuts based on solar angle, NHIT and isotropy? This could certainly be done and may well prove successful, but there seems to be little point. The network technique is relatively simple and straightforward and, as Ridges 5 and 6 above demonstrate, does use more than the three simplest parameters. Such cuts, once optimised, may be just as mathematically complex as the function carried out by the network and there would still be the necessity of obtaining pure data sets from the detector or Monte Carlo.

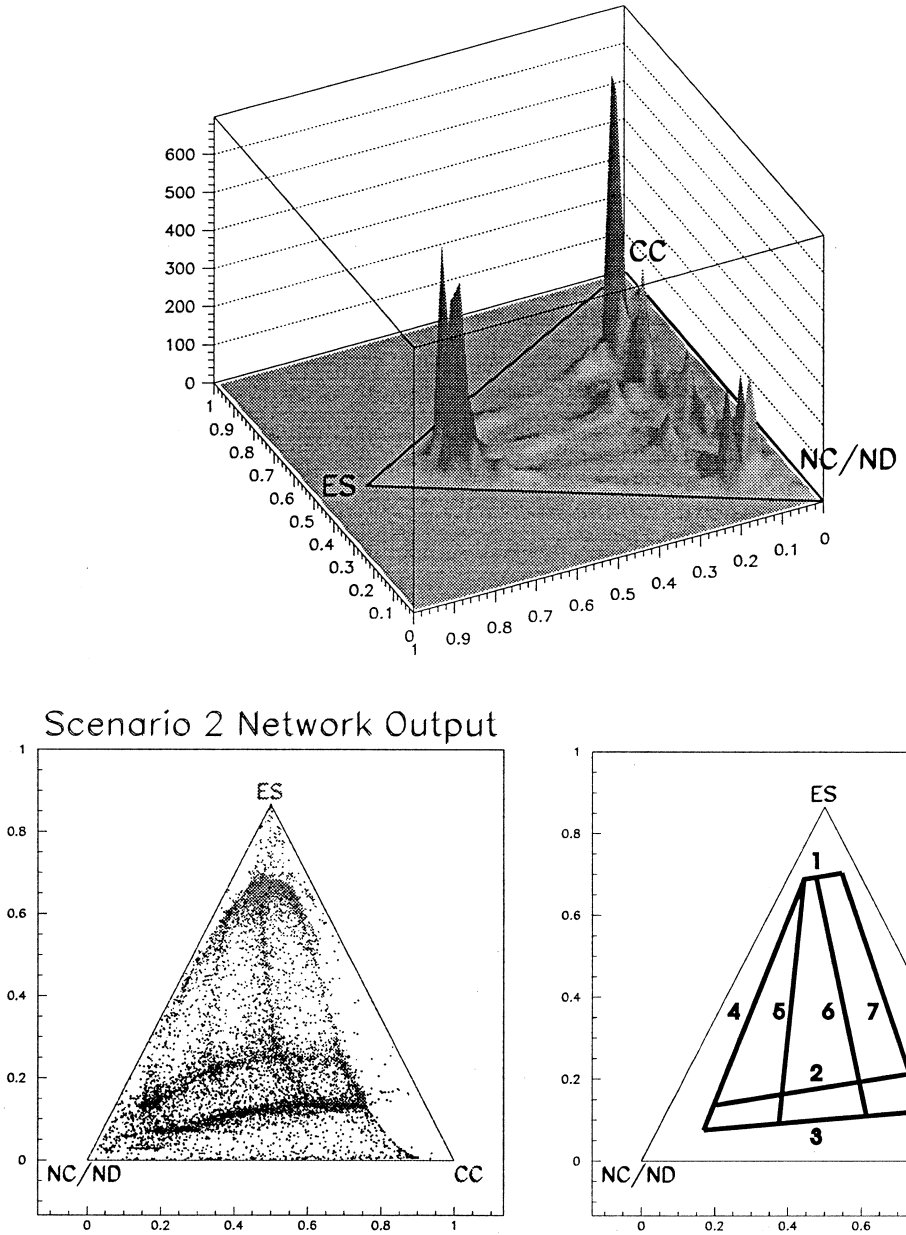


Figure 7.7: On top is a combined and smoothed histogram of the 3 PDFs of Scenario 2 applied to Cl fill data. Below and on the left is a reproduction of the corresponding scatter plot with the diagram on the right picking out the key features.

7.7.2 High but Known Scattering

The analyses of Figs. 7.5 and 7.6 are carried out with the known or best guess values for the light scattering and absorption of the various detector components included in the Monte Carlo. These may or may not prove to be accurate and so a check is carried out where the analysis of Fig. 7.5 is repeated but with the amount of Rayleigh scattering in the Monte Carlo increased by an order of magnitude for all the training, fitting and 'real' data sets. Rayleigh scattering is chosen not because it is uncertain to an order of magnitude, but because it is a simple mimic of other forms of scattering. Fig. 7.8 shows the results of this high scattering analysis.

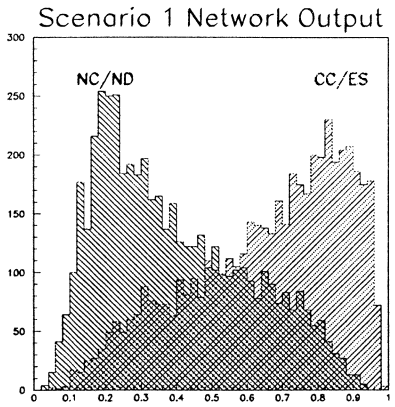
It is clear from each of the three plots in Fig. 7.8 that the degree of separation is poorer than that of their counterparts in Fig. 7.5. This is only to be expected as the more scattering there is in the detector the more isotropic the hit patterns of every event class become. That being said the separation is still pretty good and this is born out by the fit results which are accurate and have uncertainties only slightly higher than those of Fig. 7.5. This is therefore an impressive demonstration of the fact that the network technique should work in any reasonable detector conditions so long as those conditions are known.

7.8 Systematics

The neural technique has shown good statistical uncertainties and a resilience to detector conditions, but nothing has yet been said about its susceptibility to systematic errors. Such errors can be defined as systematic differences between the data sets used to form the PDF histograms and the data set to which these PDFs are fitted. It is here that the true benefit of using CC and ES data from a poison run rather than Monte Carlo comes in. With data from a poison run and NC/ND data from calibration the only systematics that can occur will be due to the detector conditions changing between the PDF data being taken and the real data being taken. Although this can happen the leeway for temporal changes should be smaller than that for differences between Monte Carlo and reality.

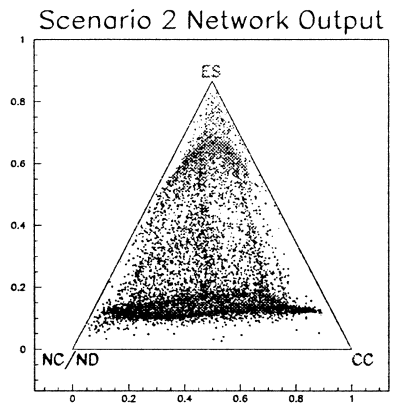
There are many effects which could contribute to a systematic error and they will all have to be studied (or at least as many as can be identified). In this work scattering in D_2O and H_2O is looked at as a first pass at assessing the level of systematic error and the results of systematically distorting the ν_e energy spectrum are reported.

A network is trained and PDF histograms formed using data produced with the default level of scattering. These PDFs are then fitted to 'real' data sets produced with increased Rayleigh scattering. The discrepancy between true and fitted numbers then gives the systematic error. Increases in Rayleigh scattering are chosen as a mimic for more general scattering increases. Fig. 7.9 shows four



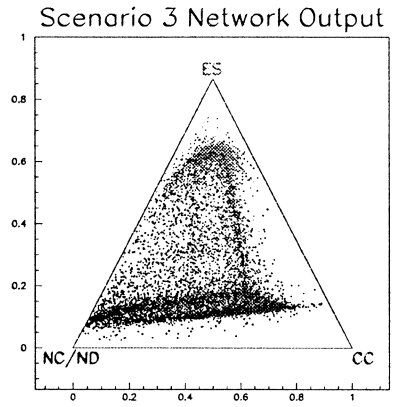
Scenario 1 Fit

	CC/ES	NC/ND
SSM fit	10062(134)	2625(134)
SSM true	10040	2647
MSW fit	3791(98)	2789(98)
MSW true	3933	2647



Scenario 2 Fit

	CC	ES	NC/ND
SSM fit	8782(153)	1223(71)	2682(136)
SSM true	8832	1208	2647
MSW fit	3313(112)	530(49)	2737(100)
MSW true	3397	536	2647



Scenario 3 Fit

	CC	ES	NC/ND
SSM fit	8865(166)	1213(69)	2609(151)
SSM true	8832	1208	2647
MSW fit	3330(125)	556(49)	2694(115)
MSW true	3397	536	2647

Figure 7.8: A repeat of the analysis of Fig. 7.5, but with Rayleigh scattering increased by an order of magnitude for all training, fitting, and 'real' data sets.

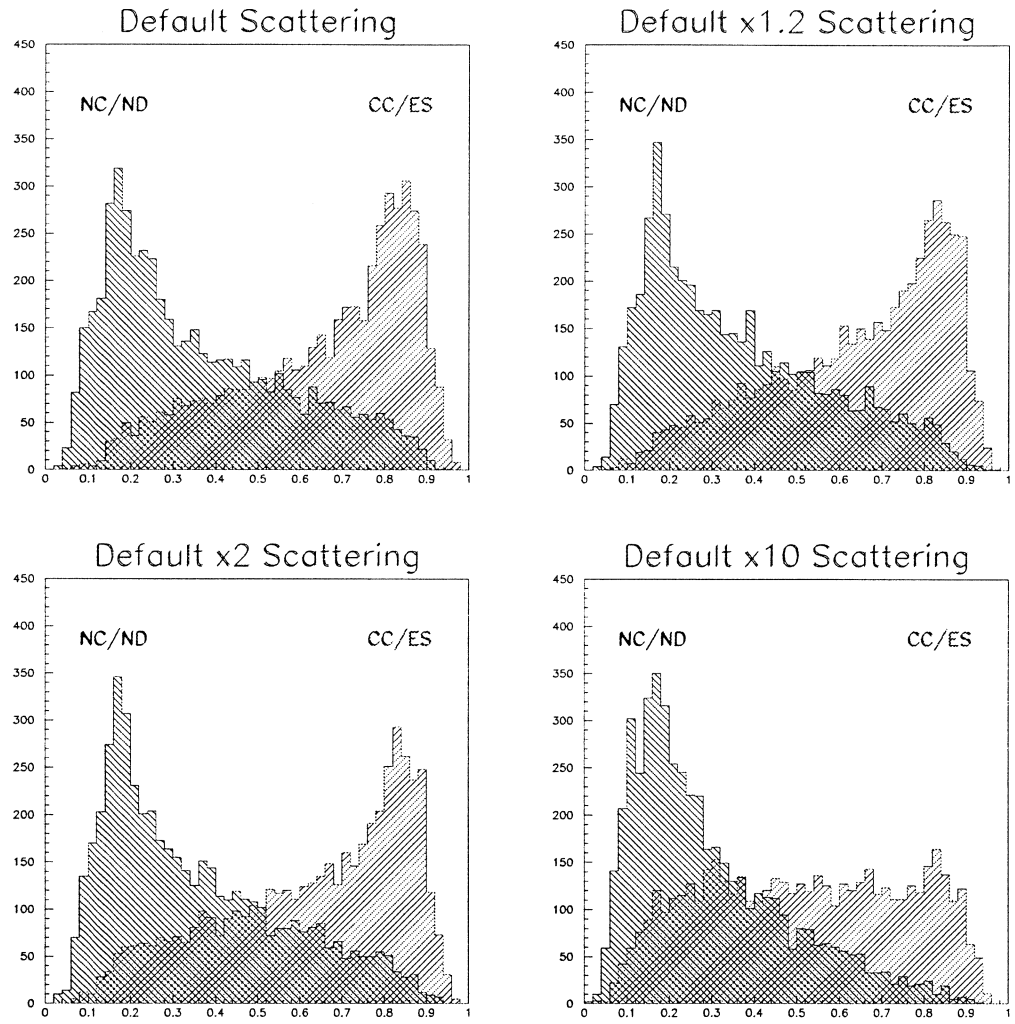


Figure 7.9: Four plots showing the network output of data sets of increasing levels of scattering for a network trained on the default scattering level.

histograms of the 'real' data using Scenario 1 at four different levels of scattering. Each data set has been fed through the same network trained on the default scattering level. The observed degradation with increased scattering is entirely predictable as more scattering produces more isotropic hit patterns. Thus both event classes NC/ND and CC/ES get pulled to the left, with NC/ND, the class that shows more isotropy being little affected, but CC/ES being gradually smeared flat. The results of the fits are shown in Table 7.3. Default scattering PDFs are fitted to 'real' data with 1.2 and 2 times the default scattering level. The systematic errors made are ~2% and ~5% respectively for CC events, ~6% and ~7% for ES events, and ~4% and ~15% for NC/ND events. Two features should be noted: The NC/ND numbers are consistently high and the CC numbers are

Scenario 1 Fit

	CC/ES	NC/ND
×1.2 scattering SSM fit	10042(119)	2645(119)
×2.0 scattering SSM fit	9583(124)	3104(124)
SSM true	10040	2647
×1.2 scattering MSW fit	3840(91)	2740(91)
×2.0 scattering MSW fit	3544(92)	3036(92)
MSW true	3933	2647

Scenario 2 Fit

	CC	ES	NC/ND
×1.2 scattering SSM fit	8754(134)	1204(69)	2729(115)
×2.0 scattering SSM fit	8401(138)	1151(68)	3135(120)
SSM true	8832	1208	2647
×1.2 scattering MSW fit	3295(98)	491(46)	2794(87)
×2.0 scattering MSW fit	3031(100)	482(45)	3067(189)
MSW true	3397	536	2647

Scenario 3 Fit

	CC	ES	NC/ND
×1.2 scattering SSM fit	8834(144)	1209(70)	2644(126)
×2.0 scattering SSM fit	8105(153)	1125(68)	3457(137)
SSM true	8832	1208	2647
×1.2 scattering MSW fit	3380(108)	482(46)	2718(98)
×2.0 scattering MSW fit	2917(111)	466(45)	3197(101)
MSW true	3397	536	2647

Table 7.3: Fits of PDF histograms formed with the default level of scattering to 'real' data with increased scattering.

consistently low. This follows naturally from the slewing observed in Fig. 7.9. Secondly the ES numbers are only slightly affected by the systematic (the high percentage error arises from the low number of ES events). This is also no surprise as the ES separation is almost entirely based upon the fitted solar angle and the levels of scattering are not yet high enough to have a serious effect on the quality

of direction fit. It is claimed that the overall level of scattering in the detector will be known to within 10% with a lightball calibration. A conservative estimate of the systematic error due to detector scattering would therefore be about 2% and it is thought likely that this will form the largest contribution to the systematic error budget.

It was noted in Sec. 7.3 that the harmonic and residual-harmonic parameters (extracting spatial and spatio-temporal information respectively) have an irreducible, but slight NHIT dependence as a result of the hit pattern tending towards the underlying distribution of Čerenkov light as the energy of a single electron is increased. By producing CC and ES events from distorted ν_e spectra [Thorman 96] has carried out checks to see what level of systematic error is induced by training a network on an undistorted data set and then fitting undistorted PDF histograms to 'real' data histograms which have the distortion that results from the best fit MSW parameters. The systematic error induced is small for all three classes (CC, ES, NC/ND), almost always well below the 1σ statistical uncertainty. One can therefore tentatively conclude that if the MSW effect operates and if the existing solar neutrino experiments are correct then the spectral distortion will induce a minimal systematic.

7.9 Calibrating the Monte Carlo

If data from a poison run is not available the network training will have to rely on Monte Carlo data. In this case the accuracy of the simulation is of paramount importance, and is crucial anyway for the many other ways in which SNOMAN will be used. Apart from the obvious possibility of neglected or improperly coded physics, the simulation accuracy will depend upon the correct values of various physical and code parameters such as the level of light scattering and absorption in the D_2O, H_2O, and acrylic, the reflectivity of the concentrators, unphysical parameters in the EGS4 code, and so on. Whilst many of these parameters can be obtained from theory or lab bench measurements of detector components, some values will only be obtainable from in situ calibration with a variety of sources. It could be that the parameters to be optimised will not interfere greatly with each other and can be simply extracted from the calibration data. What is, perhaps, more likely is that a number of parameters whose effects can mask or interfere with each other will need to be found and cannot be simply extracted. It is here that there may be an application for a neural network.

The principle is most clearly illustrated with a concrete example: the intended use of a ^{8}Li and a ^{252}Cf source to produce electrons and neutrons respectively in the detector. ^{8}Li is an isobaric analogue of ^{8}B (the isotope whose decay produces the neutrinos that SNO will detect) and its β decay with an endpoint of 13 MeV has a spectrum similar to that expected from the CC reaction. ^{252}Cf is a fission neutron source and can be used to produce a large number of events which are known to be

(n,γ) reactions with chlorine or deuterium i.e. NC or ND reactions. Although ^{8}Li, with its different energy spectrum and source scattering effects, will probably not be a good enough mimic of the CC electrons, hit patterns from this source can be compared with Monte Carlo ^{8}Li hit patterns and any differences used to optimise simulation parameters. How should this hit pattern comparison be carried out? A very powerful way would be to train a network to distinguish between ^{8}Li and ^{252}Cf calibration events and so produce two PDF histograms, one for each source. Monte Carlo ^{8}Li and ^{252}Cf events can then be fed through the trained network and the resulting histograms compared with the calibration histograms. This comparison could take the form of a χ^2 test and many Monte Carlo data sets with different parameter settings could be generated and fed through the network until the χ^2 was sufficiently good. Perhaps this whole procedure could be wrapped up as a genetic algorithm [Goldberg 89, Holland 75] to provide a more directed choice of Monte Carlo parameters with each successive set of simulations (a method for doing this outlined in [Hahn 92, Hahn 93]).

Alternatively a network could be trained to distinguish between real and Monte Carlo data of the same source (this is briefly discussed in [Garrido 92]). In this case the Monte Carlo would be run many times with different parameters and with each data set a fresh training of the network is carried out. An optimisation algorithm (perhaps genetic) would be used to achieve simulation parameters that minimised the difference between the Monte Carlo and real data output histograms. This algorithm would be acting in direct opposition to the network training: whilst one is trying to bring the histograms together the other is attempting to separate them, the result of the struggle being an optimal set of simulation parameters.

Such procedures will involve some effort and considerable CPU time and may constitute an over-reaction to what proves to be a simple problem. Only further work will tell.

7.10 Further Applications

The success of the neural technique in separating CC, ES, and NC/ND events has led to thoughts of applying it to other hit pattern identification problems in SNO. Work is currently proceeding on the separation of hit patterns from ^{208}Tl and ^{214}Bi $\beta\gamma$ events that fall within a window from 30 to 40 hits [Chen 95]. The idea, suggested by Martin Moorhead [Moorhead 95a], is that the Tl events can easily get above 30 hits with a β that doesn't necessarily have to be close to its 4.99 MeV endpoint, a few low energy γs and the 2.6 MeV γ (see Appendix F for the Thorium and Uranium decay chains). In order for the ^{214}Bi event to get above 30 hits the β has to be pretty close to its 3.27 MeV endpoint and there can't be any accompanying γs. Consequently these two types of events produce Čerenkov light from more than one electron and from a single electron respectively and so ought to show characteristic differences in their hit patterns. This has proven to be the

case and reasonable separation has been achieved using the same neural technique as outlined in this paper. This work has even been successfully extended to a six class separation [Chen 96] where ^{208}Tl and ^{214}Bi $\beta\gamma$ events are identified in three detector regions: D_2O, acrylic vessel, and H_2O. As well as using the intrinsic differences between the two $\beta\gamma$ decays this technique exploits the differing impact of the acrylic vessel on Čerenkov light originating from the three regions.

The more refined the SNO position/direction fitting algorithms and hit pattern recognition techniques get, the more entwined the two tasks become. To improve position fits beyond those of the Time and Quad fitters it is necessary to bring in information about the type of event being fitted. This is the case with fitters that use some kind of PDF as the Grid Fitter and, to a lesser extent, the Elastic Fitter do. On the other hand any hit pattern recognition technique will greatly benefit from position and direction fit information. In the analysis of this work such information enables individual PMT hits to be specified by three numbers (θ, ϕ, and t) rather than four (x, y, z, t). This reduction in the number of degrees of freedom produces a better separation of event classes. It seems that one is being powerfully drawn towards a technique that will both fit an event and perform hit pattern recognition as one 'organic' process. Such a technique may not be realisable, but certainly warrants investigation. The best place to start is with the receptive fields neural network technique [Hertz 91] which has been used by the U.S. Postal Service to develop a neural network for zip code recognition on envelopes [Le Cun 89b, Le Cun 89a, Keeler 90] a problem whose translational invariance is analogous to the rotational invariance of the SNO detector. A receptive fields technique has also been used on SNO data [Dong 93a, Dong 93b], but applied to the somewhat less ambitious task of distinguishing Čerenkov hit patterns from random coincidences.

It was concluded in Sec. 7.8 that the distortion in the ν_e energy spectrum that results from the best fit MSW parameters induces a minimal systematic error in the fitted event class fractions. What if, for some reason the distortion were much larger? In that case the systematic error would almost certainly become significant and it would be necessary to adapt the analysis accordingly. The best way to avoid this systematic is to conduct separate network analyses in NHIT slices, say from 60-65 hits, 66-70 hits, etc. This also has the advantage of producing results for the CC NHIT spectrum in addition to providing the overall numbers of CC and NC events. Investigations of this alternative approach are currently underway.

Chapter 8

Final Conclusions

This thesis covers a number of issues within the broad heading of Monte Carlo and Analysis Techniques for the Sudbury Neutrino Observatory. Chapters 3 and 4 handle two Monte Carlo subjects. In the SAGE ^{51}Cr source analysis of Chapter 3 it is shown that the calibration device poses a substantial radiation hazard and makes clear the necessity of remote handling procedures. Simulation speed-up techniques based on geometric symmetries are also developed and, whilst these have not found an application in the SNO Monte Carlo, they may yet prove useful. Chapter 4 details the coding of the detector geometry within SNOMAN, the SNO Monte Carlo. It is shown how a simple modular structure can produce a simulation of considerable detail, accurately modelling the SNO detector froms scales of 1000cm to 0.1cm.

Chapters 5, 6, and 7 cover aspects of data analysis within the SNO experiment. In Chapter 5 a new algorithm for fitting event positions and directions is introduced and it is shown to be more accurate than the benchmark fitters, though considerably more complex. Such seemingly incremental advances in fit quality are important for the application of fiducial cuts, the better extraction of hit pattern features, and the introduction of new ideas ripe for further development. Chapters 6 and 7 are certainly the most important of this thesis. They describe how a neural network technique can be used to extract the characteristic features of hit patterns from different classes of event. This enables the two most important event classes, the charged and neutral current interactions, to be statistically separated and hence the flavour makeup of the solar neutrino flux to be analysed. If the MSW effect operates its influence should be clearly visible in such a charged/neutral current hit pattern analysis.

Appendix A

Geometry Primitives

Each request (NEXT, NEAR, or PICK) made of the geometry module in SNOMAN is passed down to be answered at the level of the geometry primitive, as described in Chapter 4. Working in a coordinate system where the origin is at the centre of the primitive (or some similar position) the GEP routine is given the position and direction of the particle and must return an answer to one of the three requests

NEXT:- What is the distance to the primitive boundary in the given direction ?

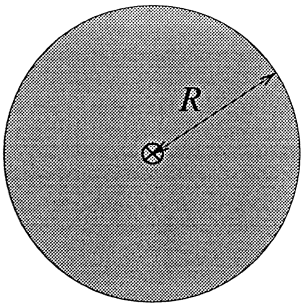
NEAR:- What is the nearest distance to the primitive in any direction ?

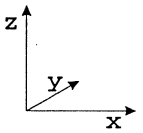
PICK:- Pick a position randomly within the primitive.

This appendix briefly describes how each of these requests is answered for each of the nine geometry primitives in SNOMAN. The diagrams show the parameters defining the dimensions of the primitives with the crossed circle indicating the location of the origin of the coordinate system. The particle position is referred to as $\underline{\mathbf{r}}$ or (x,y,z), the direction as $\underline{\mathbf{u}}$ or (u,v,w), a general point in space as (X,Y,Z), and the NEXT and NEAR distances as D and N respectively.

With some of the more complex primitives it is necessary to solve a quartic equation to answer the request. The general solution of such equations has a long and distinguished history, but the method adopted by the primitive routines is a more recent technique specifically designed for implementation in a computer program [Cashwell 69]. Many computational tricks for speeding up such geometry calculations can also be found in [Glassner 89].

SPHERE





NEXT

The two candidate NEXT distances come directly from the two intersections of a line with a sphere. They are the solutions of the quadratic

$$|\underline{\mathbf{u}}|^2 D^2 + 2\underline{\mathbf{r}}.\underline{\mathbf{u}} D + |\underline{\mathbf{r}}|^2 - R^2 = 0$$

NEAR

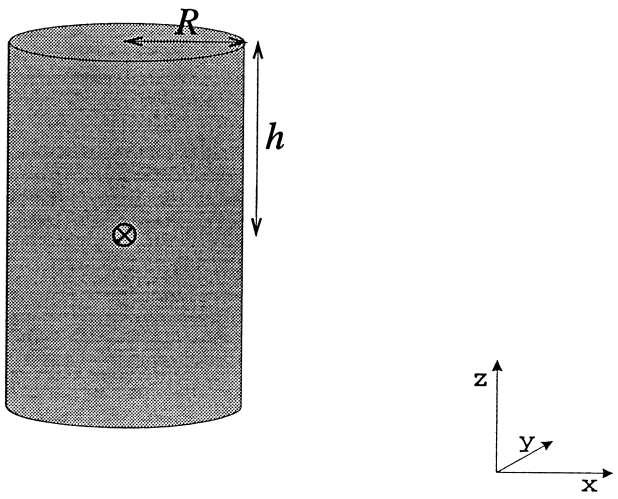
The near distance is simply the radius of the particle minus the radius of the sphere

$$N = |\underline{\mathbf{r}}| - R$$

PICK

When picking a point within a sphere it is actually faster to pick within the enclosing cube and reject the position and pick again if it is outside the sphere. This avoids the CPU time consuming calculation of trigonometric functions that would result from the picking of a point in spherical polar coordinates.

CYLINDER



NEXT

The first step is to check for the NEXT distances to an infinite cylinder by solving the quadratic

$$(u^2 + v^2)D^2 + 2(xu + yv)D + x^2 + y^2 - R^2 = 0$$

then checking to see whether the intersection points have a Z coordinate between $-h$ and $+h$. Then a similar procedure is followed for the top and bottom of the cylinder, checking the NEXT distances to the planes at $Z = -h$ and $Z = +h$ and seeing whether these intersections fall within the circle $X^2 + Y^2 = R^2$.

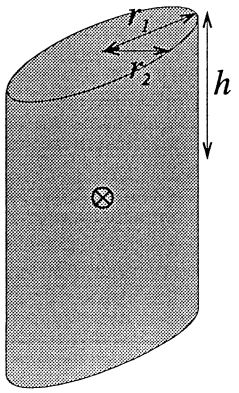
NEAR

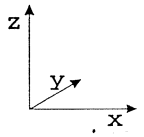
The whole space is subdivided into regions by cones and planes in such a way that the region within which a point falls determines whether the nearest distance to the cylinder is to the flat top or bottom, the curved side, or to the top or bottom circular edges. Once the region has been ascertained the distance is simple to calculate.

PICK

A point is picked using cylindrical polar coordinates i.e. uniformly in r^2, ϕ, and z.

ELLIPTICAL CYLINDER





NEXT

The first step is to check for the NEXT distances to an infinite elliptical cylinder by solving the quadratic

$$\left[\left(\frac{u}{r_2}\right)^2+\left(\frac{v}{r_1}\right)^2\right]D^2+2\left[\frac{xu}{r_2^2}+\frac{yv}{r_1^2}\right]D+\left(\frac{x}{r_2}\right)^2+\left(\frac{y}{r_1}\right)^2-1=0$$

then checking to see whether the intersection points have a Z coordinate between $-h$ and $+h$. Then a similar procedure is followed for the top and bottom of the elliptical cylinder, checking the NEXT distances to the planes at $Z=-h$ and $Z=+h$ and seeing whether these intersections fall within the ellipse $\left(\frac{X}{r_2}\right)^2+\left(\frac{Y}{r_1}\right)^2=1$.

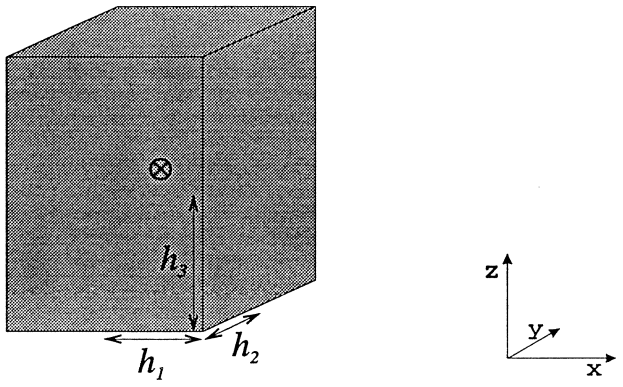
NEAR

The whole space is subdivided into regions by elliptical cones and planes in such a way that the region within which a point falls determines whether the nearest distance to the elliptical cylinder is to the flat top or bottom, the curved side, or to the top or bottom elliptical edges. Once the region has been ascertained the distance is simple to calculate.

PICK

A point is picked in the box enclosing the elliptical cylinder and rejected and re-picked if it falls outside the cylinder.

BLOCK



NEXT

The NEXT distances are found by checking the line intersections with the planes at $X = \pm h_1$ and then seeing whether these intersections fall within the rectangle in the YZ plane formed by the adjacent sides. This process is repeated for the planes at $Y = \pm h_2$ and $Z = \pm h_3$.

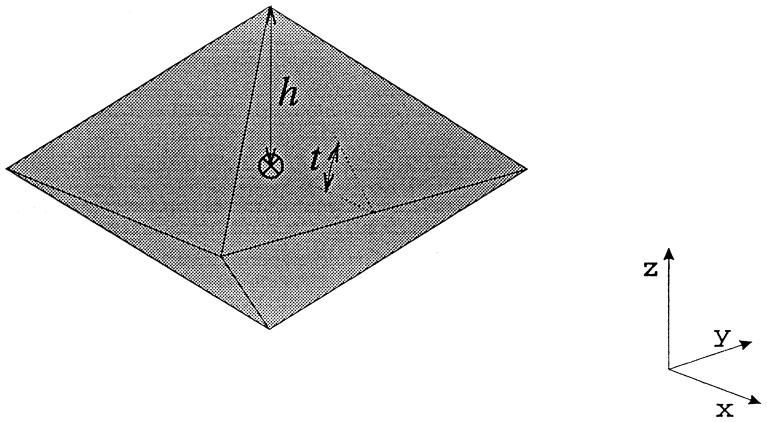
NEAR

The calculation of the NEAR distance utilises the symmetry of the block. The particle position is reflected in the planes $X = 0$, $Y = 0$, and $Z = 0$ until it is in the positive octant. The whole of the positive octant is divided into regions with planar boundaries. The region within which the reflected point falls determines whether the NEAR distance is to a particular flat side or straight line edge. Once this has been determined the NEAR distance calculation is simple.

PICK

The x, y, z coordinates of the point are picked uniformly in the range determined by h_1, h_2, and, h_3.

OCTAHEDRON



NEXT

The octahedron is restricted to one defined by the height of the apex in the Z direction and the angle made with the XY plane by any of the eight planes. The NEXT distances are obtained by calculating the intersection of the given point and direction with each of the eight planes making up the octahedron and for each plane checking that the intersection is between the three planes defining that face of the octahedron.

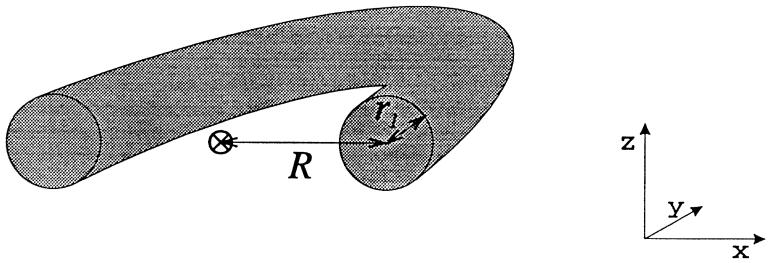
NEAR

The calculation of the NEAR distance utilises the symmetry of the octahedron. The particle position is reflected in the planes $X = 0$, $Y = 0$, $Z = 0$, and $X = Y$ until it is in the $X > Y$ region of the positive octant. The $X > Y$ part of the positive octant is divided into regions with planar boundaries. The region within which the reflected point falls determines whether the NEAR distance is to a particular flat side or straight line edge. Once this has been determined the NEAR distance calculation is simple.

PICK

The x, y, z coordinates of the point are picked uniformly in the block enclosing the octahedron and then rejected and re-picked if they do not fall within the octahedron.

HALF TORUS



NEXT

The NEXT distances are found by first finding the intersection distances of a line with the complete torus which has the Cartesian equation

$$\left[\sqrt{(X^2+Y^2)}-R\right]^2+Z^2=r_1^2$$

These four distances emerge as the solution to a rather ugly quartic equation. Intersection points in the negative Y region are thrown away and then intersections with the $Y=0$ plane are checked, these points also being thrown away if they fall outside the two circular disks in the that plane.

NEAR

To calculate the NEAR distance for a particle with $y>0$ it is only necessary to work in the plane containing the origin, the particle position, and the Z axis. The problem then reduces to finding the NEAR distance to a circle in this plane, yielding the result

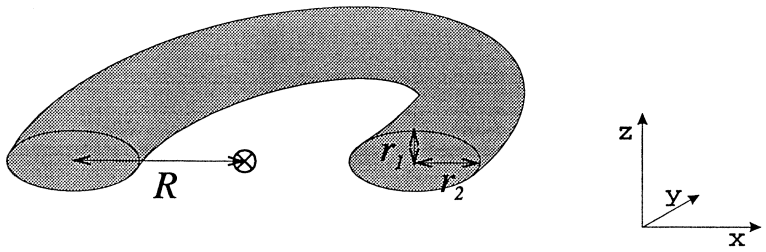
$$N=\sqrt{\left[\left(\sqrt{(x^2+y^2)}-R\right)^2+z^2\right]}-r_1$$

For a particle in the $y<0$ region the whole space is broken into regions determining whether the NEAR distance is to one of the flat ends or to one of the circular edges. Once the region is known the NEAR distance can be easily calculated.

PICK

The z coordinate is first picked uniformly in the range $-r_1$ to $+r_1$. Then a radius ρ is picked uniformly in ρ^2 in the range $R-r_1$ to $R+r_1$. These point are rejected and re-picked if they do not fall within the circle of radius r_1, and centre $\rho=R$. An angle ϕ is picked uniformly in the range 0 to π. The x and y coordinates of the point are then given by $\rho\cos\phi$ and $\rho\sin\phi$ respectively.

ELLIPTICAL HALF TORUS



NEXT

The NEXT distances are found by first finding the intersection distances of a line with the complete elliptical torus which has the Cartesian equation

$$\left[\frac{\sqrt{(X^2+Y^2)}-R}{r_2}\right]^2+\left[\frac{Z}{r_1}\right]^2=1$$

These four distances emerge as the solution to a rather ugly quartic equation. Intersection points in the negative Y region are thrown away and then intersections with the $Y=0$ plane are checked these points also being thrown away if they fall outside the two elliptical disks in that plane.

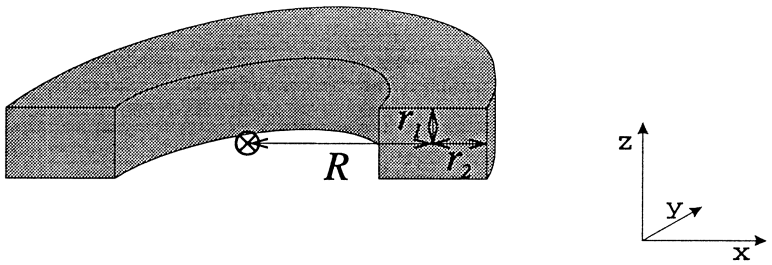
NEAR

To calculate the NEAR distance for a particle with $y>0$ it is only necessary to work in the plane containing the origin, the particle position, and the Z axis. The problem then reduces to finding the NEAR distance to a ellipse in this plane and once again results in a quartic equation. For a particle in the $y<0$ region the whole space is broken into regions determining whether the NEAR distance is to one of the flat ends or to one of the elliptical edges. Once the region is known the NEAR distance can be easily calculated.

PICK

The z coordinate is first picked uniformly in the range $-r_1$ to $+r_1$. Then a radius ρ is picked uniformly in ρ^2 in the range $R-r_2$ to $R+r_2$. These point are rejected and re-picked if they do not fall within the ellipse of semi-axes r_1 and r_2, and centre $\rho=R$. An angle ϕ is picked uniformly in the range 0 to π. The x and y coordinates of the point are then given by $\rho\cos\phi$ and $\rho\sin\phi$ respectively.

RECTANGULAR HALF TORUS



NEXT

The rectangular half torus is the region carved out by the two planes $Z = \pm r_1$, the two infinite cylinders of radii $R + r_2$ and $R - r_2$, and the plane $Y = 0$. The NEXT distances are found by calculating the intersections with the two cylinders and checking that they fall within the three planes, calculating the intersections with the two Z planes and checking that they fall within the cylinders and the Y plane, and finally repeating the process for intersections with the $Y = 0$ plane.

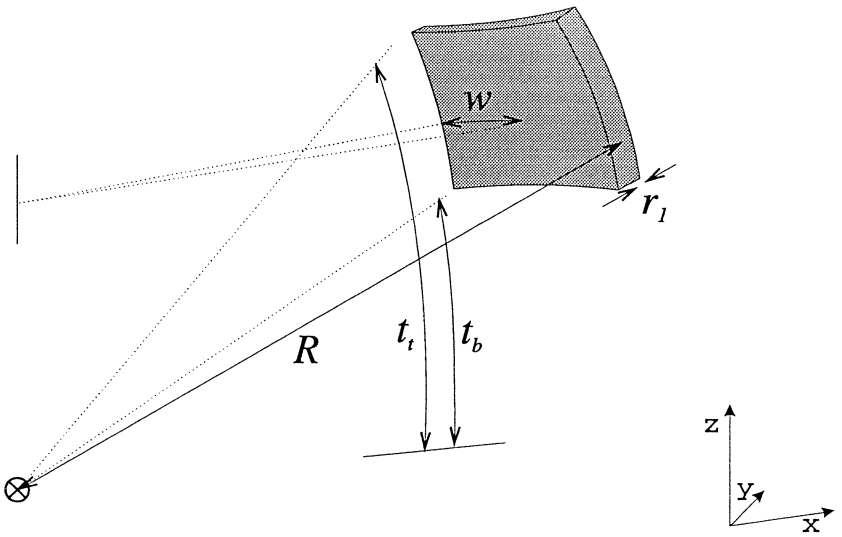
NEAR

To calculate the NEAR distance for a particle with $y > 0$ it is only necessary to work in the plane containing the origin, the particle position, and the Z axis. The problem then reduces to finding the NEAR distance to a rectangle in this plane. For a particle in the $y < 0$ region the whole space is broken into regions determining whether the NEAR distance is to one of the flat ends or to one of the straight edges. Once the region is known the NEAR distance can be easily calculated.

PICK

The z coordinate is first picked uniformly in the range $-r_1$ to $+r_1$. Then a radius ρ is picked uniformly in ρ^2 in the range $R - r_2$ to $R + r_2$ and an angle ϕ is picked uniformly in the range 0 to π. The x and y coordinates of the point are then given by $\rho \cos\phi$ and $\rho \sin\phi$ respectively.

SPHERICAL TILE



NEXT

The spherical tile is most easily described in spherical polar coordinates, where it is the region carved out by two spheres of constant radius $R + r_1$ and $R - r_1$, two cones of constant polar angle t_t and t_b, and two planes of constant azimuthal angle $+w$ and $-w$. The NEXT distances are found by calculating the intersections with the two spheres and checking that they fall within the polar and azimuthal angle bounds, then repeating the analogous process for the two cones and the two planes.

NEAR

The NEAR distance is found by breaking up the whole space into regions bounded by cone and planes. The region within which the particle position falls determines whether the NEAR distance is to a curved side, a flat side, a curved edge etc. Once the region has been established the NEAR distance is easy to calculate.

PICK

The position is picked in spherical polar coordinates i.e. uniform in r^3, $\cos\theta$, and ϕ between the appropriate upper and lower bounds determining the extent of the tile.

Appendix B

The Acrylic Vessel as a Lens

Every SNO position and direction fitter to date makes the assumption that the Čerenkov photons travel in a straight line with the speed of light in water from their point of origin to their point of detection. This assumption is clearly not true for photons that have scattered or reflected and this is why most fitters attempt to eliminate such light. It is also untrue for any photon which has passed through the acrylic vessel as the differing refractive indices of water and acrylic will produce refraction at each interface. If the vessel were planar this refraction would simply produce a lateral offset in the trajectory of a photon crossing it, but with an approximately spherical vessel the surface normal at the point where the photon exits the acrylic will be slightly different to the entry point normal thus producing a change in direction as well. This is illustrated by Fig. B.1 where a photon starts with the position and direction of the double arrow, hits the acrylic vessel with angle of incidence θ_i, travels a distance x within the vessel, and exits to travel a distance d_o to the PMT sphere. Without the refraction the photon would have taken the dashed route in the figure and travelled a distance d_i from its acrylic entry point to the PMT sphere. To assess the impact of this lensing effect of the acrylic vessel on the assumption of straight line travel it is convenient to calculate two quantities which are functions of the incident angle θ_i. They are the distance D between where the photon hits the PMT sphere and where it would have done if it had travelled in a straight line, and the difference T in PMT arrival times of these two trajectories.

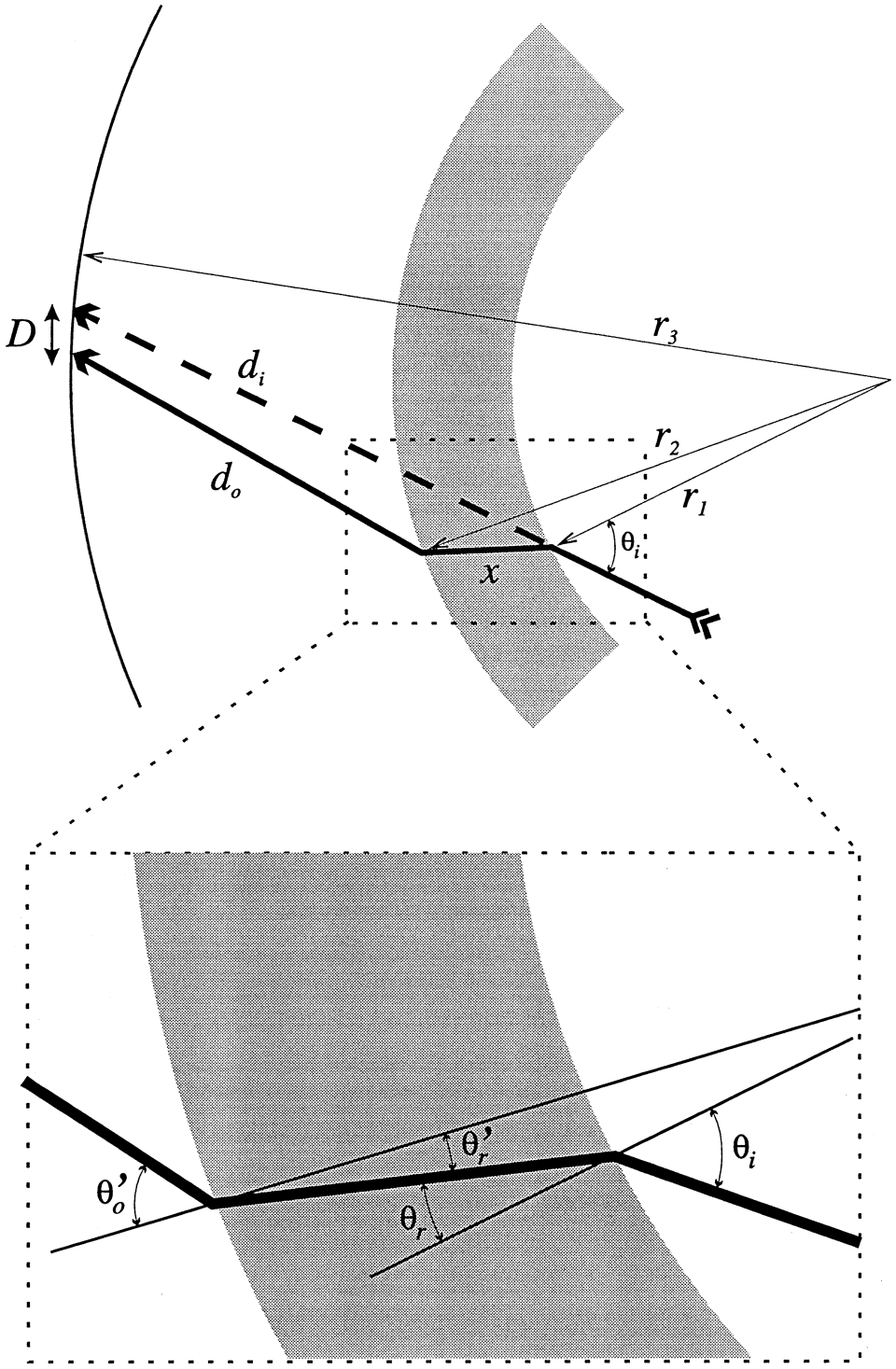


Figure B.1: The lensing effect of the acrylic vessel.

With reference to Fig B.1 the following quantities are defined:-

α	Ratio of the refractive indices of water and acrylic (i.e. $\frac{n_{\text{water}}}{n_{\text{acrylic}}}$).
v_{water}	Speed of light in water.
r_1	Inner radius of acrylic vessel.
r_2	Outer radius of acrylic vessel.
r_3	Radius of PMT sphere.
θ_i	Incident angle of photon with respect to acrylic entry normal.
θ_r	Acrylic angle of photon with respect to acrylic entry normal.
θ_r'	Acrylic angle of photon with respect to acrylic exit normal.
θ_o'	Exit angle of photon with respect to acrylic exit normal.
x	Distance travelled by photon in the acrylic.
d_o	Distance travelled by photon from leaving acrylic to hitting PMT sphere.
d_i	Straight line distance the photon would have travelled from entering acrylic to hitting PMT sphere.

It is assumed that the refractive indices of light and heavy water are the same. Given α, r_1, r_2, r_3, Snell's law, and some trigonometry D and T can be found as functions of θ_i alone. These functions are particularly messy and so the following derivation will express various lengths and angles in terms of θ_i and give D and T in terms of these quantities.

Considering vector components parallel and perpendicular to the length x then

$$\begin{aligned} D^2 &= [x + d_o\cos(\theta_o' - \theta_r') - d_i\cos(\theta_i - \theta_r)]^2 + \\ &\quad [d_o\sin(\theta_o' - \theta_r') - d_i\sin(\theta_i - \theta_r)]^2 \end{aligned} \tag{B.1}$$

and

$$T = \frac{1}{v_{\text{water}}}\left[d_o + \frac{x}{\alpha} - d_i\right] \tag{B.2}$$

Using Snell's law

$$\sin\theta_r = \alpha\sin\theta_i \tag{B.3}$$

Using the sine rule applied to the triangle formed by x, r_1, and r_2

$$\frac{\sin\theta_r'}{r_1} = \frac{\sin(\pi - \theta_r)}{r_2}$$

therefore

$$\sin\theta_r' = \alpha\frac{r_1}{r_2}\sin\theta_i \tag{B.4}$$

Using Snell's law again

$$\sin\theta_o' = \frac{1}{\alpha}\sin\theta_r'$$

therefore

$$\sin\theta_o' = \frac{r_1}{r_2}\sin\theta_i \tag{B.5}$$

Applying the cosine rule to the triangle formed by x, r_1, and r_2

$$r_2^2 = r_1^2 + x^2 - 2r_1 x\cos(\pi - \theta_r)$$

therefore

$$x = \sqrt{\left[r_2^2 - r_1^2\sin^2\theta_r\right]} - r_1\cos\theta_r$$

and so

$$x = \sqrt{\left[r_2^2 - \alpha^2 r_1^2\sin^2\theta_i\right]} - \sqrt{\left[r_1^2 - \alpha^2 r_1^2\sin^2\theta_i\right]} \tag{B.6}$$

Another application of the cosine rule to the triangle formed by d_o, r_2, and r_3 gives

$$r_3^2 = d_o^2 + r_2^2 - 2d_o r_2\cos(\pi - \theta_o')$$

therefore

$$d_o = \sqrt{\left[r_3^2 - r_2^2\sin^2\theta_o'\right]} - r_2\cos\theta_o'$$

and so

$$d_o = \sqrt{\left[r_3^2 - r_1^2\sin^2\theta_i\right]} - \sqrt{\left[r_2^2 - r_1^2\sin^2\theta_i\right]} \tag{B.7}$$

A final application of the rule to the d_i, r_1, r_3 triangle gives

$$r_3^2 = d_i^2 + r_1^2 - 2d_i r_1\cos(\pi - \theta_i)$$

therefore

$$d_i = \sqrt{\left[r_3^2 - r_1^2\sin^2\theta_i\right]} - \sqrt{\left[r_1^2 - r_1^2\sin^2\theta_i\right]} \tag{B.8}$$

Equations B.3,B.4,B.5 give the three angles and equations B.6,B.7,B.8 the three lengths purely in terms of θ_i. These are plugged into the expressions for D and T and the resulting curves are shown in Fig. B.2 for both an air fill and a water fill. These curves indicate that at large angles of incidence there are significant deviations from straight line travel in both the arrival position and arrival time of the photon, the situation being slightly worse for air, with its lower refractive index, than it is for a water fill.

It must be born in mind that for photons generated isotropically in the D_2O region all angles of incidence are not equally likely. In fact small angles will be far more probable than large ones. The probability distribution for θ_i can be easily obtained. For a photon generated at a radius r and in a direction that makes an angle β to the radius vector some simple geometry shows that

$$\sin\theta_i = \frac{r}{r_1}\sin\beta \tag{B.9}$$

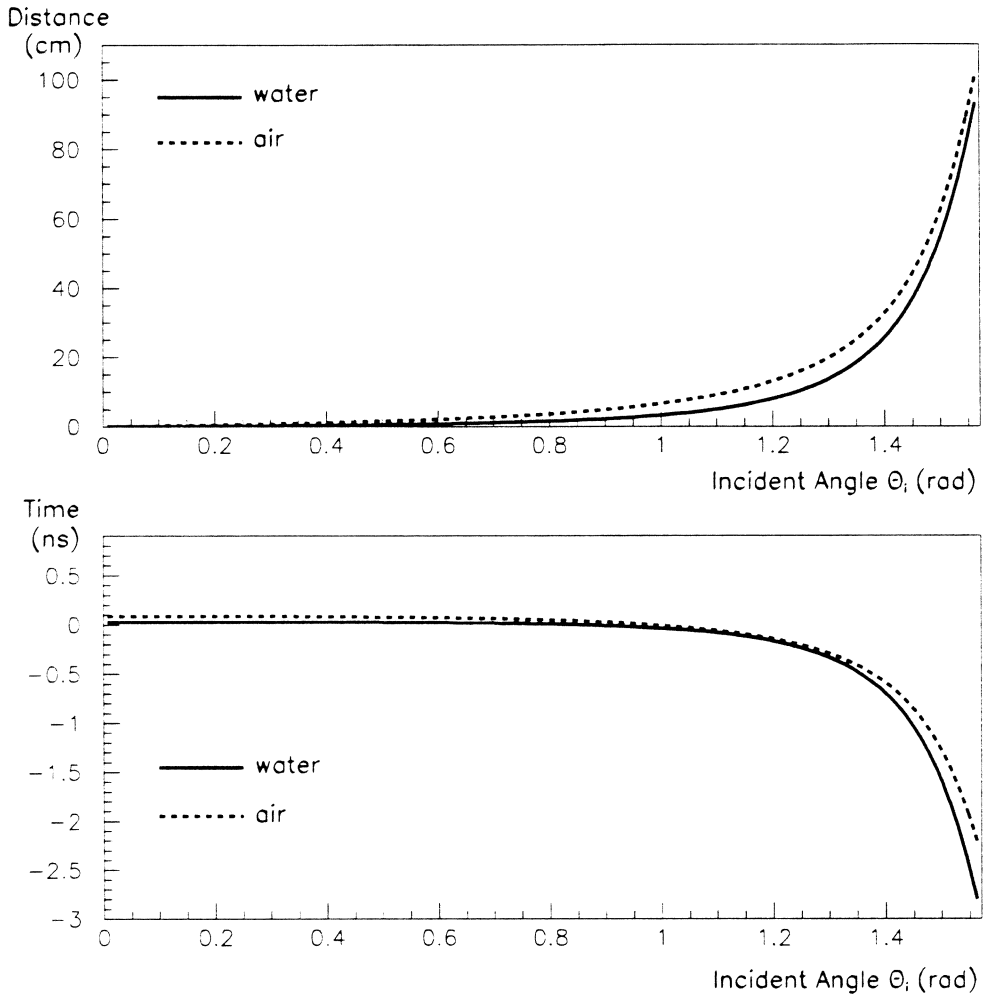


Figure B.2: The curves expressing D and T as functions of the incident angle θ_i for a detector filled with water and one filled with air.

The probability distributions for r and β are the familiar expressions from spherical geometry

$$\frac{dP}{dr} \propto r^2$$

$$\frac{dP}{d\beta} \propto \sin\beta$$

The joint probability distribution for a photon having an angle of incidence θ_i and a direction β is

$$\frac{d^2P}{d\theta_i\, d\beta} = \frac{d^2P}{dr\, d\beta}\frac{dr}{d\theta_i}$$

$$\propto r^2 \sin\beta \frac{dr}{d\theta_i}$$

From Eqn. B.9

$$\frac{dr}{d\theta_i} = \frac{r_1 \cos\theta_i}{\sin\beta}$$

giving

$$\frac{d^2P}{d\theta_i\, d\beta} \propto r^2 r_1 \cos\theta_i$$

For a particular choice of θ_i, β can vary between the limits θ_i and $\pi - \theta_i$. Expressing r in terms of β and integrating out over β gives

$$\frac{dP}{d\theta_i} \propto r_1 \cos\theta_i \int_{\beta_{\min}}^{\beta_{\max}} r^2\, d\beta$$

$$\propto r_1^3 \sin^2\theta_i \cos\theta_i \int_{\theta_i}^{\pi-\theta_i} \frac{1}{\sin^2\beta}\, d\beta$$

$$\propto 2r_1^3 \sin\theta_i \cos^2\theta_i$$

The probability distribution for θ_i given photons generated isotropically in position and direction is therefore

$$\frac{dP}{d\theta_i} d\theta_i \propto d(\cos^3\theta_i) \tag{B.10}$$

By sampling θ_i according to this distribution and calculating the corresponding Ds and Ts the histograms of Fig. B.3 are produced. Given the ~1.6ns timing jitter of the PMTs and the photocathode diameter of ~30cm, they show that for photons generated isotropically in position and direction in the D_2O region the deviations from straight line travel are negligible. For a water fill only 1.6% of the PMT hits (3.5% for an air fill) are offset by more than 15cm.

For photons generated at large radii, however, the incident angle θ_i will tend to be larger than usual and with it the typical distance and time deviations. For $r > 0.95r_1$ 11% of hits are deviated by more than 15cm for a water fill and 24% for an air fill. For fitters assuming straight line photon propagation this effect could lead to systematic direction dependent shifts in the fitted position for events at large radii. It should also be noted that the assumption of a spherical PMT shell at a radius of 850m is not quite true. In fact the PMTs are held in a geodesic

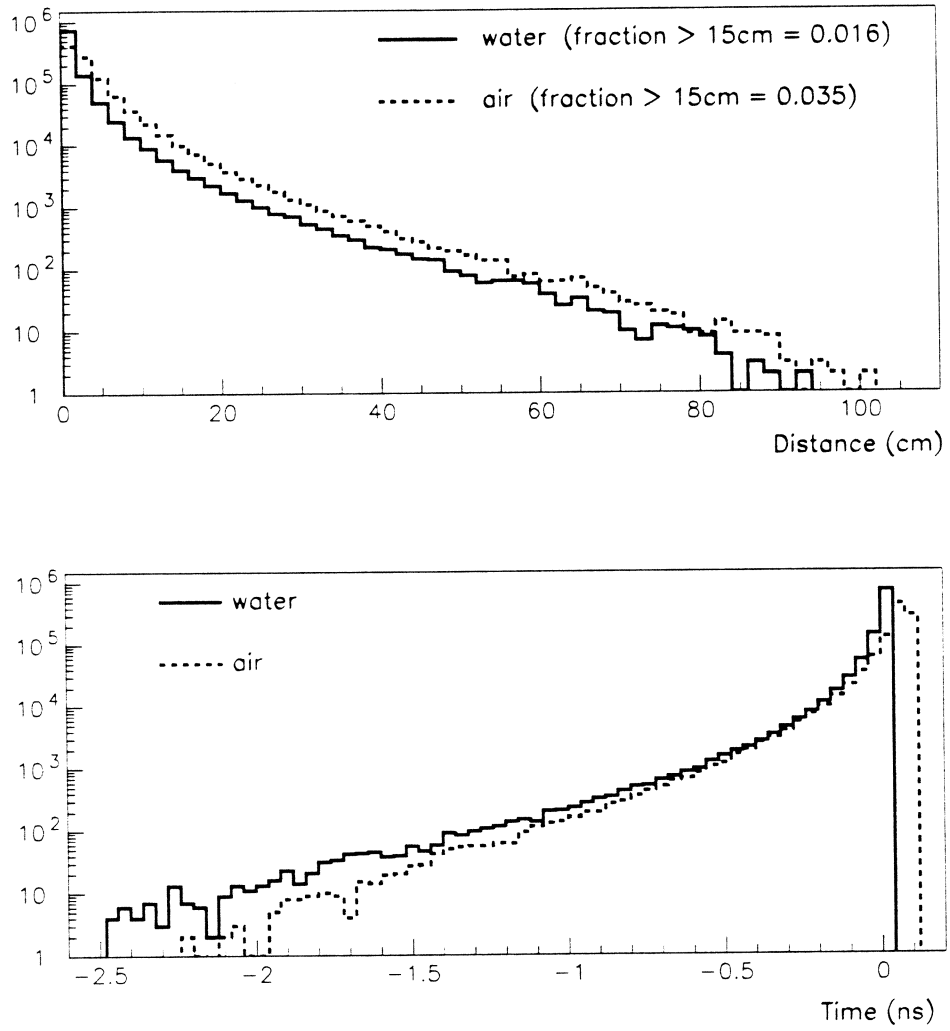


Figure B.3: Histograms of the distance and time deviations from straight line travel for photons generated isotropically in position and direction in the D_2O region.

structure and are at radii of between ~840 and ~860m. This will have a slight smearing effect on the D and T curves.

It can therefore be concluded that the assumption of photon straight line travel that is made by all existing fitters is reasonable, but its effect on events occurring at large radii should be further investigated.

Appendix C

Fitter Pull

All SNO position fitters to date are based on minimising the PMT residuals in some way (where the residual is defined by Eqn. 5.1). It is noted in Chapter 5 that all such fitters are susceptible to a systematic shift of the fitted position away from the true vertex and along the electron direction. This shift is known as fitter pull and is illustrated in Fig. C.1 which shows a histogram of the fraction of the fit error vector (i.e. $\underline{\mathbf{r}}_{\mathrm{fit}} - \underline{\mathbf{r}}_{\mathrm{ev}}$) which lies along the electron direction for 1500 5MeV electrons generated by SNOMAN and fitted with the Time Fitter. The quantity actually plotted is

$$\frac{(\underline{\mathbf{r}}_{\mathrm{fit}} - \underline{\mathbf{r}}_{\mathrm{ev}}).\underline{\mathbf{u}}}{|\underline{\mathbf{r}}_{\mathrm{fit}} - \underline{\mathbf{r}}_{\mathrm{ev}}|}$$

The large peak at +1 indicates a strong tendency for the fitted position $\underline{\mathbf{r}}_{\mathrm{fit}}$ to be pulled away from the true position $\underline{\mathbf{r}}_{\mathrm{ev}}$ and along the electron direction $\underline{\mathbf{u}}$. The smaller peak at -1 shows a weaker tendency for the pull to be against the electron direction.

Using a special case (electrons travelling radially) this appendix provides strong evidence that fitter pull is a simple and inevitable consequence of two facts: more light arrives late than arrives early and a hit placed randomly on the PMT sphere is more likely to be outside a particular Čerenkov cone than inside. An alternative treatment of fitter pull is given in [Skensved 90].

An electron travelling radially with direction vector $\underline{\mathbf{u}}$ will produce a Čerenkov cone that intersects the PMT sphere in a perfect circle. The situation is illustrated in two dimensions by Fig. C.2 where the circle represents the PMT sphere and the two solid straight lines indicate the Čerenkov cone. Under perfect conditions the photons will form a circular hit pattern each of them having the same arrival time t_{a}. A residual based fitter applied to this hit pattern will be able to determine that the event lies along the dashed line (the perpendicular bisector of the hit pattern circle), but its position along the line will be completely undetermined.

Consider a single additional hit at $\underline{\mathbf{r}}_{\mathrm{b}}$, with time t_{b} that does not lie on the hit pattern circle. This additional hit will enable the fitter to pin down an event position on the dashed line. If the light travels straight from its point of creation to

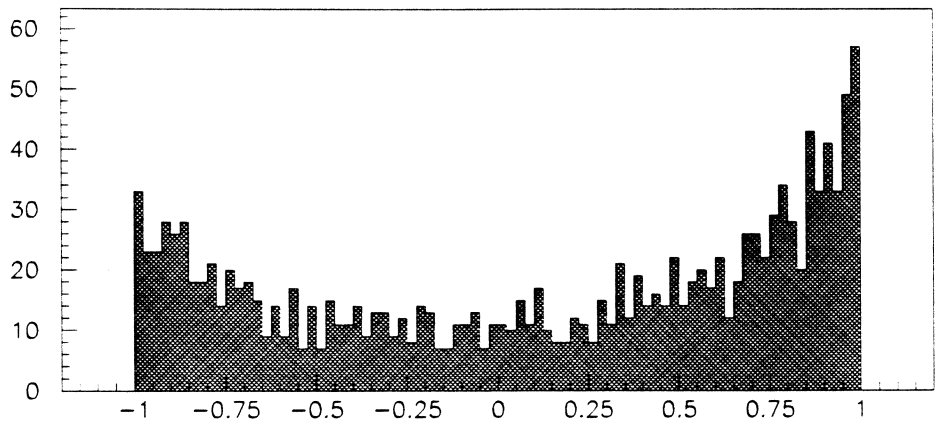


Figure C.1: . The fraction of the fit error vector which lies along the true electron direction. With $\underline{\mathbf{r}}_{\mathrm{fit}}$ as the fit position, $\underline{\mathbf{r}}_{\mathrm{ev}}$ as the true event position, and $\underline{\mathbf{u}}$ as the true electron direction, the quantity plotted is $(\underline{\mathbf{r}}_{\mathrm{fit}} - \underline{\mathbf{r}}_{\mathrm{ev}}).\underline{\mathbf{u}} / |\underline{\mathbf{r}}_{\mathrm{fit}} - \underline{\mathbf{r}}_{\mathrm{ev}}|$. The data comprises 1500 5MeV electron generated by SNOMAN and fitted with the Time Fitter.

$\underline{\mathbf{r}}_{\mathrm{b}}$ the time of flight inferred from t_{b} will correspond to the distance travelled and the fitted position will be exactly right. If t_{b} is earlier or later than this correct time the fitted position will be shifted away from the correct position, but the shift will be along the dashed line where the fit is constrained to lie by the other PMT hits. A shift along the electron direction $\underline{\mathbf{u}}$ means that the quantity $\underline{\mathbf{u}}.\underline{\mathbf{r}}_{\mathrm{fit}}$ increases no matter where the event is and a shift against the electron direction will decrease $\underline{\mathbf{u}}.\underline{\mathbf{r}}_{\mathrm{fit}}$. In this radial electron case the fitter pull induced by the extra hit being out of time by dt_{b} can therefore be defined as $d(\underline{\mathbf{u}}.\underline{\mathbf{r}}_{\mathrm{fit}})$.

If $\underline{\mathbf{r}}_{\mathrm{a}}$ is a particular hit on the circular Čerenkov cone/PMT sphere intersection the following two relations will hold for the fitted position and time on the dashed line if a residual based fitter is used

$$\begin{aligned} v(t_{\mathrm{a}} - t_{\mathrm{fit}}) &= |\underline{\mathbf{r}}_{\mathrm{a}} - \underline{\mathbf{r}}_{\mathrm{fit}}| \\ v(t_{\mathrm{b}} - t_{\mathrm{fit}}) &= |\underline{\mathbf{r}}_{\mathrm{b}} - \underline{\mathbf{r}}_{\mathrm{fit}}| \end{aligned}$$

Subtracting one from the other

$$v(t_{\mathrm{a}} - t_{\mathrm{b}}) = |\underline{\mathbf{r}}_{\mathrm{b}} - \underline{\mathbf{r}}_{\mathrm{fit}}| - |\underline{\mathbf{r}}_{\mathrm{a}} - \underline{\mathbf{r}}_{\mathrm{fit}}|$$

If t_{b} is allowed to vary and with it the fit position and time then

$$v\, dt_{\mathrm{b}} = \left[\frac{\underline{\mathbf{r}}_{\mathrm{a}} - \underline{\mathbf{r}}_{\mathrm{fit}}}{|\underline{\mathbf{r}}_{\mathrm{a}} - \underline{\mathbf{r}}_{\mathrm{fit}}|} - \frac{\underline{\mathbf{r}}_{\mathrm{b}} - \underline{\mathbf{r}}_{\mathrm{fit}}}{|\underline{\mathbf{r}}_{\mathrm{b}} - \underline{\mathbf{r}}_{\mathrm{fit}}|} \right] . \, d\underline{\mathbf{r}}_{\mathrm{fit}}$$

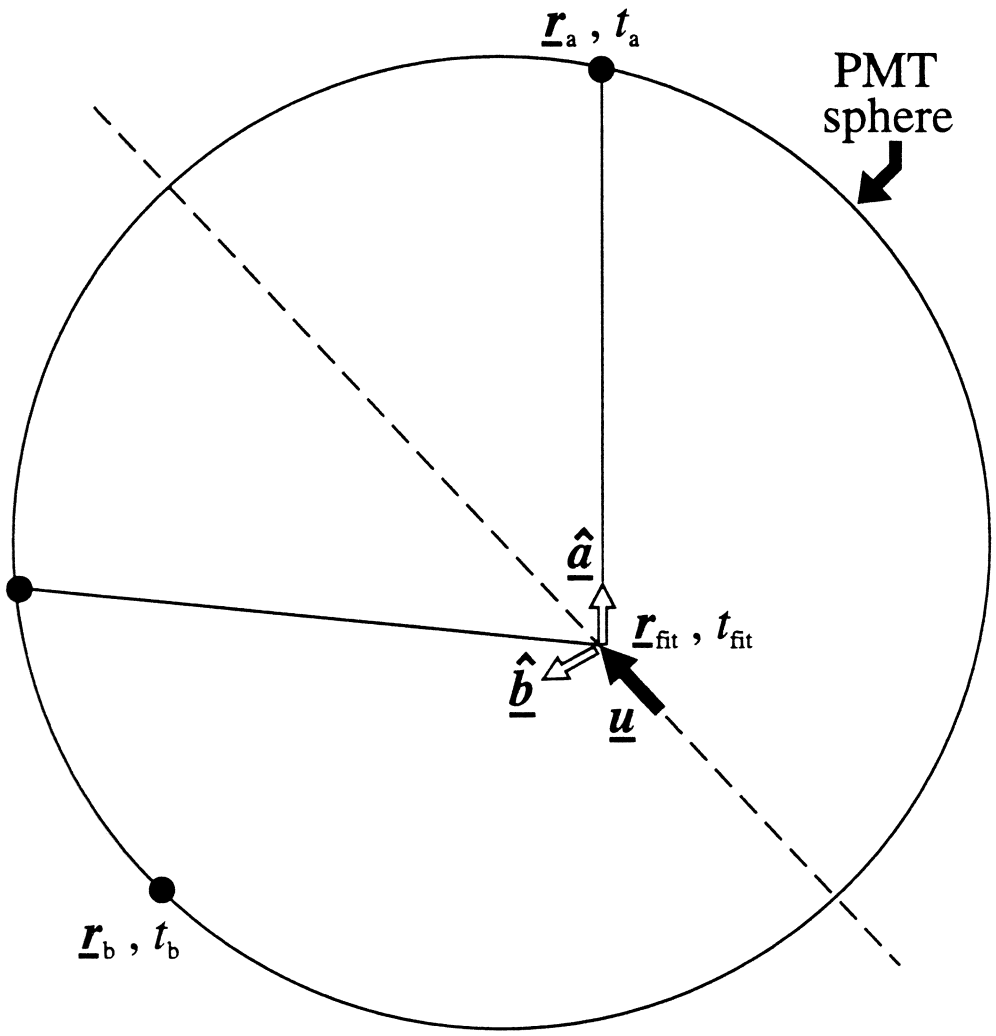


Figure C.2: The set up for a simple demonstration of fitter pull.

with $\hat{\mathbf{a}}$ defined to be the unit vector from $\underline{\mathbf{r}}_{fit}$ to $\underline{\mathbf{r}}_a$ and $\hat{\mathbf{b}}$ the same from $\underline{\mathbf{r}}_{fit}$ to $\underline{\mathbf{r}}_b$ this equation can be written

$$v\, dt_b = \left[\hat{\underline{\mathbf{a}}} - \hat{\underline{\mathbf{b}}}\right] . \, d\underline{\mathbf{r}}_{fit}$$

It has already been claimed that any movement of the fitted position due to changes in t_b must be along $\underline{\mathbf{u}}$, in other words

$$d\underline{\mathbf{r}}_{fit} = \underline{\mathbf{u}}\, d(\underline{\mathbf{r}}_{fit}.\underline{\mathbf{u}})$$

and so

$$d(\underline{\mathbf{r}}_{fit}.\underline{\mathbf{u}}) = \frac{v\, dt_b}{(\hat{\underline{\mathbf{a}}} - \hat{\underline{\mathbf{b}}}).\underline{\mathbf{u}}} \qquad \text{(C.1)}$$

Since $\hat{\underline{\mathbf{a}}}$ is a unit vector along the Čerenkov cone and $\underline{\mathbf{u}}$ is the unit vector along the cone axis, if $\underline{\mathbf{r}}_{\mathrm{b}}$ lies within the cone $(\hat{\underline{\mathbf{a}}} - \hat{\underline{\mathbf{b}}}).d\underline{\mathbf{u}}$ will be negative, otherwise it will be positive. This leads to

$$\underline{\mathbf{r}}_{\mathrm{b}} \text{ within cone} \quad d(\underline{\mathbf{r}}_{\mathrm{fit}}.\underline{\mathbf{u}}) = -\left|\frac{v}{(\hat{\underline{\mathbf{a}}} - \hat{\underline{\mathbf{b}}}).\underline{\mathbf{u}}}\right| dt_{\mathrm{b}}$$

$$\underline{\mathbf{r}}_{\mathrm{b}} \text{ outside cone} \quad d(\underline{\mathbf{r}}_{\mathrm{fit}}.\underline{\mathbf{u}}) = +\left|\frac{v}{(\hat{\underline{\mathbf{a}}} - \hat{\underline{\mathbf{b}}}).\underline{\mathbf{u}}}\right| dt_{\mathrm{b}}$$

Light that is out of time can be divided into noise hits which are equally distributed through time and scattered/reflected light which is always late. This tendency towards late rather than early light means that dt_{b} will tend to be positive. With the assumption that hits from scattered/reflected light have equal probability to show up anywhere on the PMT sphere the Čerenkov angle ensures that they are more likely to occur outside the Čerenkov cone than inside. This means that the latter of the two cases above will be more frequent than the former with dt_{b} tending to be positive. Therefore $d(\underline{\mathbf{r}}_{\mathrm{fit}}.\underline{\mathbf{u}})$ will tend to be positive and the fitter pull will be along the electron's direction of motion producing the peak at +1 in Fig. C.1. The peak at -1 in the plot is probably due to events where the late hit(s) happen to fall within the Čerenkov cone.

The preceding paragraphs have used a somewhat idealised case to demonstrate why fitter pull occurs. The extension to the general case (unconstrained $\underline{\mathbf{u}}$) is rather more difficult as the fit position degeneracy along the dashed line of Fig. C.2 no longer occurs and it is also possible that for events in certain positions the lensing effect of the acrylic vessel (described in Appendix B) may play a part.

Appendix D

Simulating The Cl Gamma Cascade

If Monte Carlo data are to be used in a neural network analysis it is crucial that the simulation be as accurate as possible. It is particularly important to ensure the accuracy of simulation of the four reactions CC, ES, NC, and ND (see Sec. 7.2 for the definitions of these acronyms). The transport simulation of the single electron from CC and ES events has been looked at by [Bowler 95b, Bowler 95a] and verification of the gamma interactions that follow NC and ND events is underway. This appendix deals with the production simulation of the γ cascade from the NC reaction.

When neutrons capture on ^{35}Cl they form an excited state of ^{36}Cl at 8.6 MeV. The nucleus subsequently γ cascades down to the ground state. Between the capture and ground states it is estimated that there are $\sim$100 levels [Lederer 78] and the energies of only the first 27 are given in the literature. There is much data on the gamma cascade [Endt 57, Draper 61, Bartolomew 67, Lone 81], but it is always presented in the form: number of γs of a particular energy emitted per 100 neutron captures. Given this data the simplest way to simulate the cascade would be to form a histogram of gamma energy frequencies and to sample from it. However, this is unsatisfactory for two reasons. Firstly it is necessary for the network analysis that each event conserve energy i.e. the total energy of the cascade must be 8.6 MeV. Secondly the individual energies within the simulated cascade must match as closely as possible those of the real event (four 2 MeV γs produce a very different hit pattern to two 4 MeV γs). The first objection is easily overcome by sampling in such a way that the total energy of the γs sums to 8.6 MeV. The second objection cannot be so easily avoided and results in the rejection of the histogram sampling method.

The method actually adopted to simulate the cascade involves considerable effort, but does produce an accurate mimic on an event by event basis. The most recent data on the cascade energies and frequencies [Lone 81] lists information on 449 γs. These energies are matched up with transitions between particular levels

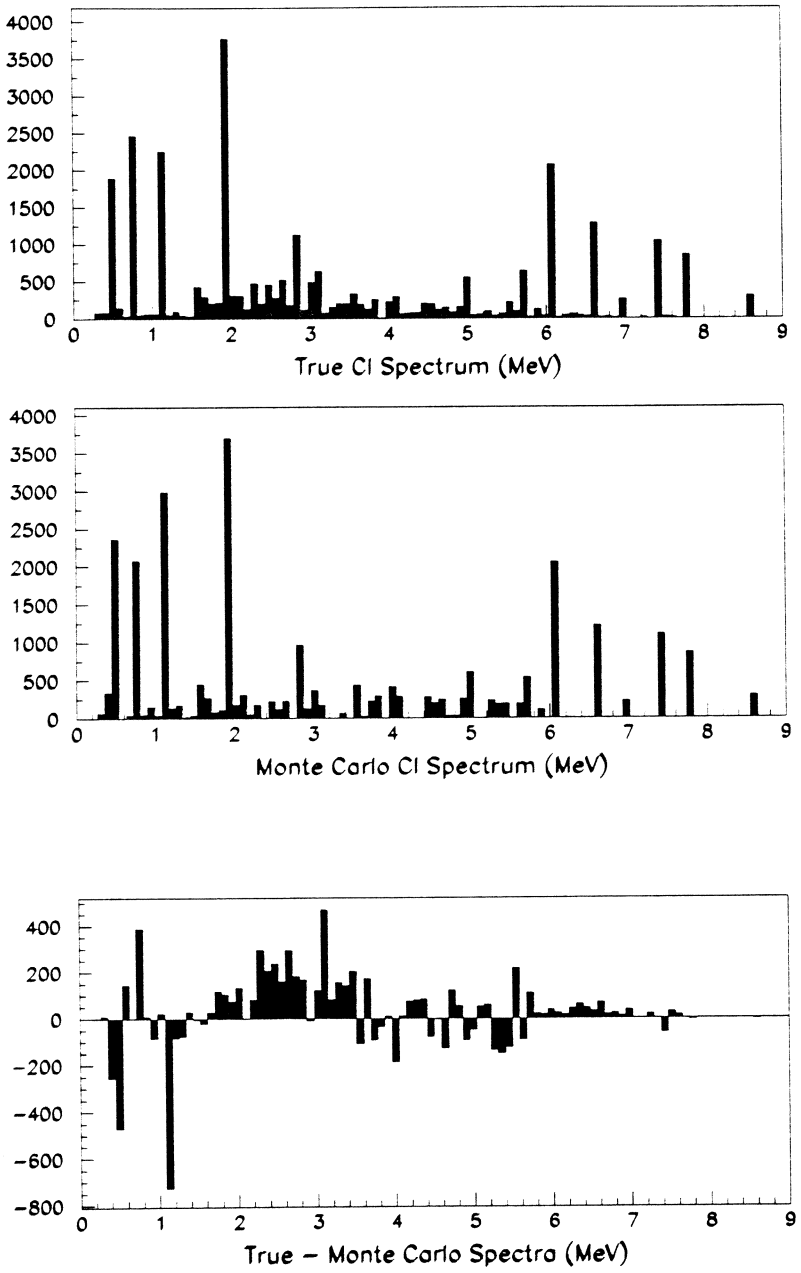


Figure D.1: The top two plots show the γ energy spectra for the true chlorine cascade and the simulated cascade each using 10000 events. The third plot shows the difference between them.

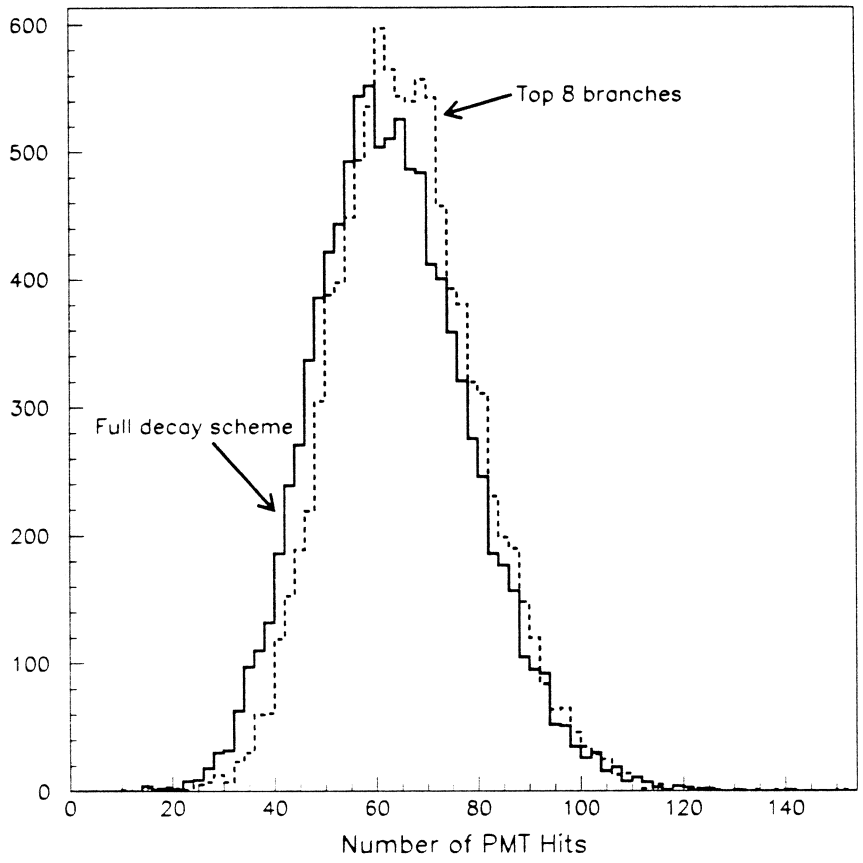


Figure D.2: The NHIT spectra for 10000 Cl cascades using the full simulation and using only the top 8 branches.

(data can only be found on the excitation energies of the first 27 levels). As a result a branching ratio can be calculated (with some effort) for the transition from each level to each of the levels below it. Only about 20% of the listed γ energies can be assigned to particular transitions and only about 50% of the possible transitions between the known excitation levels have corresponding γ energies, but all the major γs are accounted for and where γ energies cannot be accounted for there are always transitions close in energy and rate of occurrence that can. The γs from neutron capture on chlorine can then be modelled as an explicit cascade down through the energy levels of ^{36}Cl with the branching ratios being used to sample the transition from the current energy level to the next.

Data also exists on the angular correlations between subsequent γs in the five most likely cascades [van Middelkoop 66]. Although these correlations are tiny they are included in the simulation on the grounds that approximations to physical processes should only be made after balancing coding effort and CPU time with the likely effect of the approximation.

A proper assessment of the accuracy of the cascade simulation has to wait for the calibration data from a neutron source in SNO, but a number of checks can be carried out before then. Fig. D.1 compares the true energy spectrum of the chlorine cascade with the spectrum resulting from the simulation. The true spectrum is taken from the data in [Lone 81] and is the proposed sampling histogram that was rejected earlier. Both histograms are formed from 10000 neutron captures and the difference between them is shown as a third plot. It can be seen that the agreement is very good and only falls down at the low energy end where the γs are very inefficient at producing Čerenkov light.

The necessity of an exhaustive cascade simulation is illustrated by Fig. D.2. Here the NHIT spectra of 10000 cascades using the full cascade simulation and using only the top 8 branches are compared. A small but significant difference can be observed. The top 8 branches typically contain a single γ of $\sim$6 MeV and one or two low energy γs whereas the full simulation contains a higher proportion of cascades where the individual energies are all around $\sim$3 MeV. The fact that such low energy γs are less efficient at producing Čerenkov light results in the shift of the full simulation histogram to lower NHIT.

Appendix E

SNO Signal Rates

This appendix justifies the SNO signal rates used in the analysis of Chapter 7. Event rates are derived for standard solar model and best fit MSW neutrino fluxes. The calculations are by no means exact and no attempt is made to quantify uncertainties. The numbers derived are merely estimates with the two flux assumptions providing likely upper and lower bounds to the rates that SNO will see.

E.1 Standard Solar Model Rates

The ^{8}B spectrum averaged cross-sections for the charged current interaction (CC) and neutral current disintegration of the deuteron (DD) are given by [Ying 92] and that for electron scattering (ES) is derived in [Bahcall 87]. They are:-

$$\begin{aligned} \langle\sigma_{CC}\rangle_{^8B} &= 1.15 \times 10^{-42}\ \mathrm{cm}^2 \\ \langle\sigma_{DD}\rangle_{^8B} &= 4.44 \times 10^{-43}\ \mathrm{cm}^2 \\ \langle\sigma_{ES}\rangle_{^8B} &= 6.08 \times 10^{-44}\ \mathrm{cm}^2 \end{aligned}$$

The most recent calculation of the total ^{8}B neutrino flux includes the effects of heavy metal and helium diffusion in the sun and is given in [Bahcall 95]. It is:-

$$\phi_{^8B} = 6.5 \times 10^6\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$$

Calculating the number of deuteron nuclei in 1000 tonnes of D_2O and also the number of electrons in this mass plus that in the first meter of H_2O outside the acrylic vessel enables the total SNO event rates per year to be calculated:-

$$\begin{aligned} R_{CC} &= 14200\ \mathrm{yr}^{-1} \\ R_{DD} &= 5480\ \mathrm{yr}^{-1} \\ R_{ES} &= 5970\ \mathrm{yr}^{-1} \end{aligned}$$

These are the raw event rates over the whole neutrino spectrum. To get the observed event rates various thresholds and efficiencies must be factored in and to do this requires the use of the SNOMAN Monte Carlo.

First of all the rate for deuterium disintegration by a neutrino must be turned into rates for chlorine (NC) and deuteron (ND) gamma cascades for both salt fill and pure D_2O detector runs. Using SNOMAN the fraction of neutrons which do not leak out of the D_2O is found to be

$$\begin{aligned} \text{Salt fill } n \text{ non-leakage fraction} &= 0.89 \\ \text{Pure } D_2O \; n \text{ non-leakage fraction} &= 0.59 \end{aligned}$$

The higher fraction for salt fill reflects the higher neutron capture cross-section of chlorine. Comparing the thermal capture cross-sections of ^{35}Cl, ^{2}H, ^{1}H, and the various other isotopes in the D_2O and factoring in the relative amount of each isotope gives the fraction of captures within the D_2O volume which can be attributed to ^{35}Cl and ^{2}H.

Salt Fill

$$\begin{aligned} ^{35}Cl \text{ capture fraction} &= 0.93 \\ ^{2}H \text{ capture fraction} &= 0.038 \end{aligned}$$

Pure D_2O

$$\begin{aligned} ^{35}Cl \text{ capture fraction} &= 0.00 \\ ^{2}H \text{ capture fraction} &= 0.58 \end{aligned}$$

R_{DD} multiplied by these numbers together with the appropriate non-leakage fraction gives the raw NC and ND event rates for salt fill and pure D_2O detector runs. Together with the CC and ES rates from above these are

Salt Fill

$$\begin{aligned} R_{CC} &= 14200 \text{ yr}^{-1} \\ R_{ES} &= 5970 \text{ yr}^{-1} \\ R_{NC} &= 4536 \text{ yr}^{-1} \\ R_{ND} &= 185 \text{ yr}^{-1} \end{aligned}$$

Pure D_2O

$$\begin{aligned} R_{CC} &= 14200 \text{ yr}^{-1} \\ R_{ES} &= 5970 \text{ yr}^{-1} \\ R_{NC} &= 0 \text{ yr}^{-1} \\ R_{ND} &= 1888 \text{ yr}^{-1} \end{aligned}$$

Finally SNOMAN is used to find the fraction of events in each class that produce more than 60 PMT hits and reconstruct within a 7 meter radius. These fractions are

Fraction of events with NHIT > 60 and RFIT < 7m	
CC	0.622
ES	0.202
NC	0.570
ND	0.336

These multipliers are applied to give the observed SSM event rates used in the analysis of this work.

		CC	ES	NC	ND
1 Year Cl	SSM	8832	1208	2585	62
1 Year D_2O	SSM	8832	1208	0	634

E.2 Best Fit MSW Rates

The calculation of the SNO event rates resulting from best fit MSW neutrino fluxes is rather less rigorous and amounts to simply multiplying the SSM rates by a suppression factor. For the NC and ND classes there is no suppression as the neutral current interaction is insensitive to neutrino flavour. For the CC and ES classes the suppression factors are the calculated non-adiabatic MSW suppression factors of [SNO 87b]. The factor for ES is consistent with the drop in rate observed by the Kamiokande II experiment [Kamiokande 90]. They suppression factors used are:-

$$\begin{aligned} \text{CC suppression factor} &= 0.385 \\ \text{ES suppression factor} &= 0.444 \end{aligned}$$

These factors are applied to give the observed MSW event rates used in the analysis of this work.

		CC	ES	NC	ND
1 Year Cl	MSW	3397	536	2585	62
1 Year D_2O	MSW	3397	536	0	634

Appendix F

^{40}K, ^{232}Th and ^{238}U Decay Chains

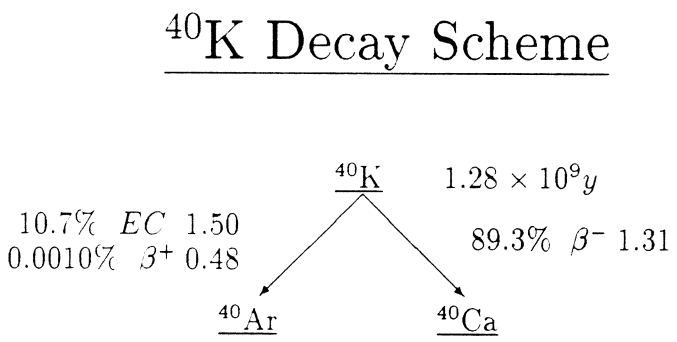


Figure F.1: The potassium decay chain. All Q values are in MeV (data from [Lederer 78]).

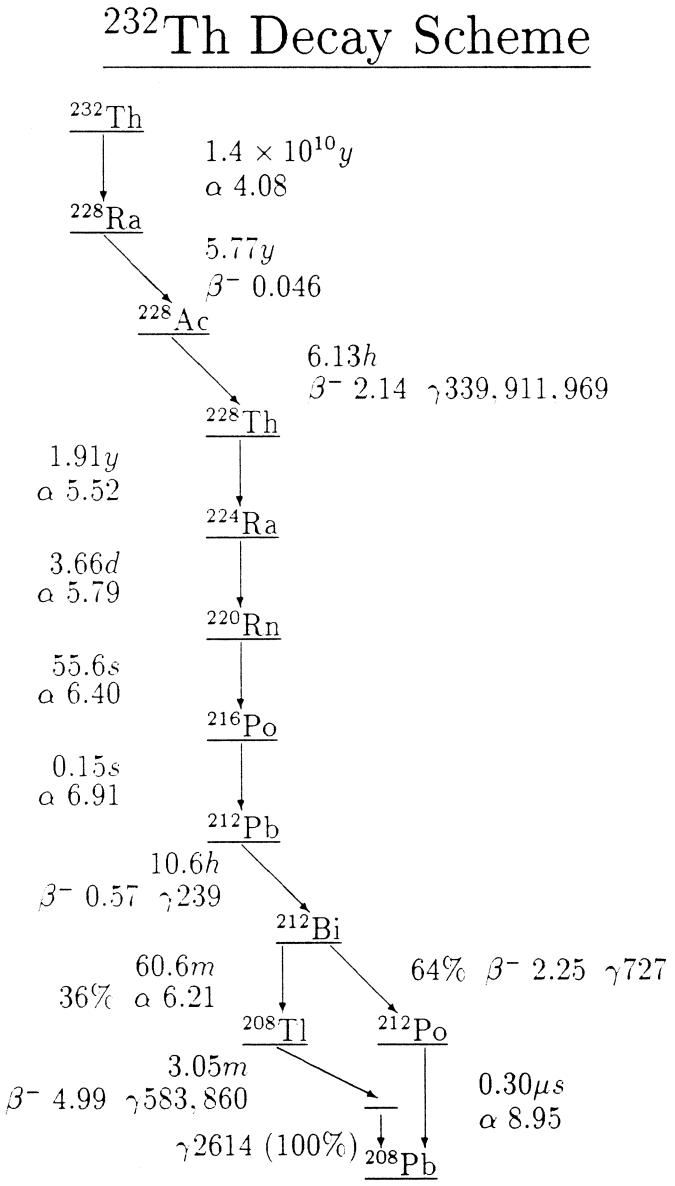


Figure F.2: The thorium decay chain. All half-lives are shown, with the Q values of beta and alpha decays in MeV, and gamma rays in keV. (Taken from [Ferraris 92]).

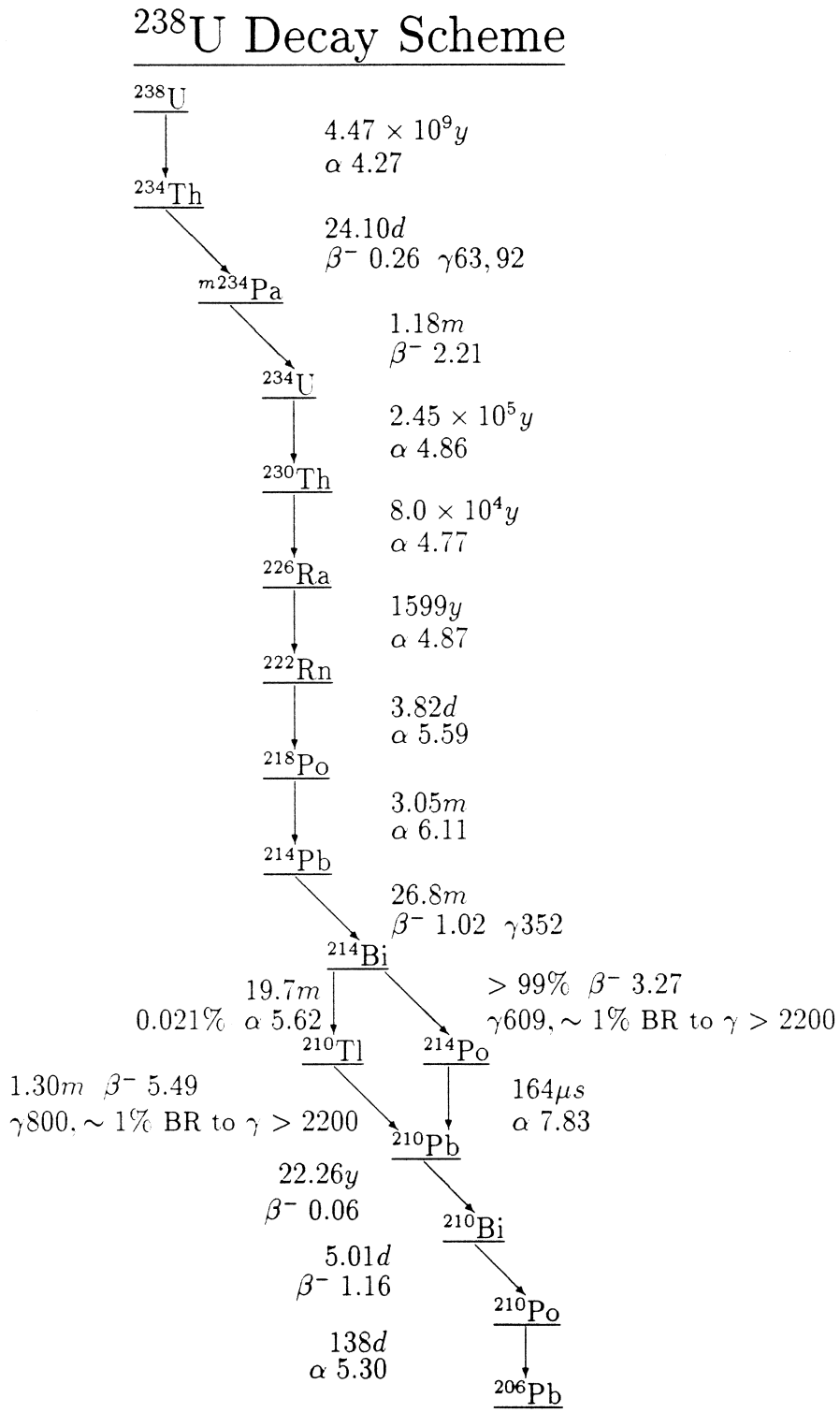


Figure F.3: The uranium decay chain. All half-lives are shown, with the Q values of beta and alpha decays in MeV, and gamma rays in keV. (Taken from [Ferraris 92]).

Bibliography

[Aardsma 87] G. Aardsma *et al. A Heavy Water Detector to Resolve the Solar Neutrino Problem.* Physics Letters B, vol. 194, pages 321–325, 1987.

[Abazov 91] A.I. Abazov *et al. Search for Neutrinos from the Sun Using the Reaction* $^{71}Ge(\nu_e, e^-)^{71}Ge$. Physical Review Letters, vol. 67, pages 3332–3335, 1991.

[Abdurashitov 95] J.N. Abdurashitov *et al. Results from SAGE II.* Nuclear Physics B (Proc. Suppl.), vol. 38, pages 60–67, 1995.

[Acker 94] A. Acker & S. Pakvasa. *Solar Neutrino Decay.* Physics Letters B, vol. 320, pages 320–322, 1994.

[Adam 95] W. Adam *et al. The Ring Imaging Čerenkov Detectors of DELPHI.* IEEE Transactions on Nuclear Science, vol. 42, no. 4, pages 499–504, 1995.

[Akkila 93] T. Akkila *et al. An Analog Neural Network Hardware Solution to a Čerenkov Ring Imaging Particle Identifier.* Nuclear Instruments and Methods in Physics Research A, vol. 327, pages 566–572, 1993.

[Alimonti 93] A. Alimonti *et al. A New Solar Neutrino Detector.* Nuclear Physics B (Proc Suppl.), vol. 32, pages 149–155, 1993.

[Aller 86] L.H. Aller. *Chemical Abundances.* In A. Dalgarno & D. Layzer, editeurs, Spectroscopy of Astrophysical Plasmas, pages 89–124. Cambridge University Press, 1986.

[Altherr 92] T. Altherr & J. Seixas. *Čerenkov Ring Recognition Using a Non-Adaptable Network.* Nuclear Instruments and Methods in Physics Research A, vol. 317, pages 335–338, 1992.

[Anderson 94] Tom Anderson. *Box Fitter*, 1994. SNO Internal Report.

[Anselmann 95] P. Anselmann *et al.* *GALLEX Solar Neutrino Observations: Complete Results for GALLEX II.* Physics Letters B, vol. 357, pages 237–247, 1995.

[Babbage 92] Wayne S. Babbage & Lee F. Thompson. *The Use of Neural Networks in γ-π^0 Discrimination.* Nuclear Instruments and Methods in Physics Research A, vol. 330, pages 482–486, 1992.

[Bachler 94] J. Bachler *et al.* *Reconstruction of Čerenkov Rings in Imaging Detectors.* Nuclear Instruments and Methods in Physics Research A, vol. 343, no. 1, pages 273–275, 1994.

[Bahcall 63] J.N. Bahcall, W.A. Fowler, I. Iben & R.L. Sears. *Solar Neutrino Flux.* Astrophysical Journal, vol. 137, pages 344–346, 1963.

[Bahcall 64a] J.N. Bahcall. *Neutrino Spectroscopy of the Solar Interior.* Physics Letters, vol. 13, pages 332–333, 1964.

[Bahcall 64b] J.N. Bahcall. *Solar Neutrino Cross-Sections and Nuclear Beta Decay.* Physical Review B, vol. 135, pages 137–146, 1964.

[Bahcall 68] J.N. Bahcall, N.A. Bahcall & G. Shaviv. *Present Status of the Theoretical Predictions for the ^{37}Cl Solar-Neutrino Experiment.* Physical Review Letters, vol. 20, pages 1209–1212, 1968.

[Bahcall 69] J.N. Bahcall. *What Next with Solar Neutrinos?* Physical Review Letters, vol. 23, pages 251–254, 1969.

[Bahcall 87] J.N. Bahcall. *Neutrino-Electron Scattering and Solar Neutrino Experiments.* Reviews of Modern Physics, vol. 59, no. 2, pages 505–521, 1987.

[Bahcall 88a] J.N. Bahcall, K. Kubodera & S. Nozawa. *Neutral Current Reactions of Solar and Supernova Neutrinos on Deuterium.* Physical Review D, vol. 38, pages 1030–1039, 1988.

[Bahcall 88b] J.N. Bahcall & R.K. Ulrich. *Solar Models, Neutrino Experiments, and Helioseismology.* Reviews of Modern Physics, vol. 60, no. 2, pages 297–372, 1988.

[Bahcall 89] J.N. Bahcall. Neutrino Astrophysics. Cambridge University Press, 1989.

[Bahcall 92] J.N. Bahcall & M.H. Pinsonneault. *Standard Solar Models, With and Without Helium Diffusion, and the Solar Neutrino Problem.* Reviews of Modern Physics, vol. 64, pages 885–926, 1992.

[Bahcall 94] J.N. Bahcall. *Two Solar Neutrino Problems.* Physics Letters B, vol. 338, pages 276–281, 1994.

[Bahcall 95] J.N. Bahcall & M.H. Pinsonneault. *Solar Models with Helium and Heavy Element Diffusion*, 1995. Preprint.

[Bartolomew 67] G.A. Bartolomew *et al. Compendium of Thermal Neutron Capture γ ray Measurements: Part 1 $Z < 46$.* Nuclear Data, no. 3, pages 434–440, 1967.

[Becker-Szendy 93] R. Becker-Szendy *et al. IMB-3 A Large Čerenkov Detector for Nucleon Decay and Neutrino Interactions.* Nuclear Instruments and Methods in Physics Research A, vol. 324, pages 363–382, 1993.

[Becks 92] K.H. Becks *et al. B-quark Tagging Using Neural Networks and Multivariant Statistical Methods: A Comparison of Both Techniques.* Nuclear Instruments and Methods in Physics Research A, vol. 329, pages 501–517, 1992.

[Bethe 39] H.A. Bethe. *Energy Production in Stars.* Physical Review, vol. 55, pages 434–456, 1939.

[Bludman 92] S.A. Bludman *et al. The Implications of Combined Solar Neutrino Observations and Their Theoretical Uncertainties.* Physical Review D, vol. 27, no. 6, pages 2220–2233, 1992.

[Boardman 92] R.J. Boardman, 1992. D.Phil. Thesis, Oxford University.

[Bowler 95a] M.G. Bowler. *Angular Distribution of Čerenkov Light from Electrons Both Produced and Stopping in Water*, 1995. SNO-STR-96-007.

[Bowler 95b] M.G. Bowler. *Effects of Electron Scattering on Čerenkov Light Output*, 1995. SNO-STR-96-006.

[Bowler 96] M.G. Bowler, 1996. Private Communication.

[Bowles 93] T.J. Bowles, 1993. Private Communication.

[Brun 93] R. Brun & J. Zoll. *ZEBRA User Guide*, 1993. CERN Program Library entry Q100.

[Burrows 90] A. Burrows. *Neutrinos from Supernova Explosions.* Annual Reviews of Nuclear and Particle Science, vol. 40, pages 181–212, 1990.

[Burrows 92] A. Burrows, D. Klein & R. Gandhi. *The Future of Supernova Neutrino Detection.* Physical Review D, vol. 45, pages 3361–3385, 1992.

[Burrows 93] A. Burrows, D. Klein & R. Gandhi. *Supernova Neutrino Bursts, the SNO Detector, and Neutrino Oscillations.* Nuclear Physics B (Proc. Suppl.), vol. 31, pages 408–412, 1993.

[Caianiello 61] E.R. Caianiello. *Outline of a Theory of Thought and Thinking Machines.* Journal of Theoretical Biology, vol. 1, pages 204–235, 1961.

[Cashwell 69] E.D. Cashwell & C.J. Everett. *Intersection of a Ray with a Surface of Third or Fourth Degree*, 1969. Los Alamos Internal Report.

[Castellani 94a] V. Castellani, S. Degl'Innocenti, G. Fiorentini, M. Lissia & B. Ricci. *Future Solar Neutrino Spectroscopy and Neutrino Properties.* Physics Letters B, vol. 324, pages 425–432, 1994.

[Castellani 94b] V. Castellani, S. Degl'Innocenti, G. Fiorentini, M. Lissia & B. Ricci. *Neutrinos from the Sun: Experimental Results Confronted with Solar Models.* Physical Review D, vol. 50, no. 8, pages 4749–4761, 1994.

[Chen 95] X. Chen. *A Monte Carlo Study of $\beta\gamma$ Backgrounds from ^{208}Tl and ^{214}Bi.* 1995. SNO Internal Report.

[Chen 96] X. Chen. *Neural Network and Statistical Study of the ^{208}Tl and ^{214}Bi Decay Events in SNO*, 1996. SNO-STR-96-004.

[Cisnero 71] A. Cisnero. *Effect of Neutrino Magnetic Moment on Solar Neutrino Observations.* Astrophysics and Space Science, vol. 10, pages 87–92, 1971.

[Cleveland 95] B.T. Cleveland *et al. Update on the Measurement of the Solar Neutrino Flux with the Homestake Chlorine Detector.* Nuclear Physics B (Proc. Suppl.), vol. 38, pages 47–53, 1995.

[Crane 48] H.R. Crane. *The Energy and Momentum Relations in Beta-Decay, and the Search for the Neutrino.* Reviews of Modern Physics, vol. 20, pages 278–295, 1948.

[Cybenko 89] G. Cybenko. *Approximation by Superpositions of a Sigmoidal Function.* Mathematics of Control, Signals, and Systems, vol. 2(4), pages 303–314, 1989.

[Davis 68] R. Davis Jr., D.S. Harmer & K.C. Hoffman. *Search for Neutrinos from the Sun.* Physical Review Letters, vol. 20, pages 1205–1209, 1968.

[DELPHI 92] The DELPHI Collaboration. *Classification of the Hadronic Decays of the Z^0 into b and c Quark Pairs using a Neural Network.* Physics Letters B, vol. 295, pages 383–395, 1992.

[Denby 92] Bruce Denby. *Tutorial on Neural Network Applications in High Energy Physics: A 1992 Perspective.* In Proceedings of the Second International Workshop on Software Engineering, Artificial Intelligence, and Expert Systems for High Energy and Nuclear Physics, pages 287–325, La Londe les Maures, France, 1992.

[Denby 95] B. Denby *et al. Performance of the CDF Neural Network Electron Isolation Trigger.* Nuclear Instruments and Methods in Physics Research A, vol. 356, pages 485–492, 1995.

[Doi 92] M. Doi & K. Kubodera. *Astrophysical Neutrino Reactions on the Deuteron.* Physical Review C, vol. 45, pages 1988–1995, 1992.

[Dong 93a] D.W. Dong & Y. Chan. *Neural Network for Recognizing Čerenkov Radiation Patterns*, 1993. SNO Internal Report.

[Dong 93b] D.W. Dong & Y. Chan. *Three Layer Network for Identifying Čerenkov Radiation Patterns.* In World Neural Network Congress, pages 312–315, Oregon, 1993.

[Draper 61] James E. Draper & Allan A. Fleischer. *Investigations of the Reaction $^{35}Cl(n,\gamma,\gamma')^{36}Cl$.* Physical Review, vol. 122, no. 5, pages 1585–1589, 1961.

[Duda 73] Richard O. Duda & Peter E. Hart. Pattern Classification and Scene Analysis. Wiley-Interscience, 1973.

[Dziembowski 94] W.A. Dziembowski *et al. A Seismic Model of the Sun's Interior.* Astrophysical Journal, vol. 432, pages 417–426, 1994.

[Eadie 71] W.T. Eadie *et al.* Statistical Methods in Experimental Physics. North-Holland, 1971.

[Earle 93] E.D. Earle. *Calibration Conceptual Design*, 1993. SNO Internal Report.

[Earle 94] E.D. Earle, R. Deal & E. Gaudette. *Optical and Th Measurements of RPT Acrylic Samples*, 1994. SNO-STR-94-010.

[Eberhard 93] P. Eberhard, G. Lynch & D. Lambert. *Fits of Monte Carlo Distributions to Data.* Nuclear Instruments and Methods in Physics Research A, vol. 326, pages 574–580, 1993.

[Eddington 20] A.S. Eddington. *The Internal Constitution of the Stars.* Observatory, vol. 43, pages 341–358, 1920.

[Eddington 26] A.S. Eddington. The Internal Constitution of the Stars. Cambridge University Press, 1926.

[Elliott 96] S.R. Elliott *et al.* *Preliminary Results from the Russian-American Gallium Experiment Cr-neutrino Source Measurement.* University of Washington Preprint, 1996. to appear in the Proceedings of the IV International Workshop on Theoretical and Phenomenological Aspects of Underground Physics.

[Endt 57] P.M. Endt & C.M. Braams. *Nuclear Energy Levels, $Z = 11$ to 20.* Reviews of Modern Physics, vol. 29, pages 727–728, 1957.

[Ewan 92] G. Ewan *et al.* *The Sudbury Neutrino Observatory.* Nuclear Instruments and Methods in Physics Research A, vol. 314, pages 373–379, 1992.

[Ferraris 92] A.P. Ferraris, 1992. D.Phil. Thesis, Oxford.

[Fireman 84] E.L. Fireman, B.T. Cleveland, R. Davis & J.K. Rowley. *Cosmic Ray Depth Studies at the Homestake Mine with $^{39}K + ^{37}Ar$ Detectors.* In Solar Neutrinos and Neutrino Astronomy: Proceedings of the Homestake Conference, pages 22–31. AIP, 1984.

[Frati 93] W. Frati, A. Hallin & T. Radcliffe. *Offline Analysis of the Laser Calibration Runs*, 1993. SNO Internal Report.

[Frati 94a] Bill Frati. *Quad Fitter Update - I*, 1994. SNO Internal Report.

[Frati 94b] Bill Frati & Richard Van de Water. *Quad Fitter Introduction - Reconstruction and Pattern Recognition*, 1994. SNO-STR-94-030.

[Frati 94c] W. Frati. *Determination of Uranium and Thorium Contamination in the D_2O from the NPMT Spectrum Above 1.4 MeV*, 1994. SNO-STR-94-005.

[GALLEX 95] The GALLEX Collaboration. *First Results from the ^{51}Cr Source Experiment with the GALLEX Detector.* Physics Letters B, vol. 342, pages 440–450, 1995.

[Garrido 92] L. Garrido, V. Gaitan & M. Serra-Racart. *Bayesian Interpretation of the Neural Net Output and Its Application to Test the Agreement Between Two Empirical Distributions.* In Computing in High Energy Physics, pages 661–664, 1992.

[Gavrin 93] V.N. Gavrin, 1993. Private Communication.

[Gish 90] H. Gish. *A Probabilistic Approach to the Understanding and Training of Neural Network Classifiers.* In Proceedings of the 1990 IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, pages 1361–1364, April 1990.

[Glassner 89] A.S. Glassner. An Introduction to Ray Tracing. Academic Press, 1989.

[Goldberg 89] David E. Goldberg. Genetic Algorithms in Search, Optimisation, and Machine Learning. Addison-Wesley, 1989.

[Grevesse 84] N. Grevesse. *Accurate Atomic Data and Solar Photospheric Spectroscopy.* Physica Scripta, vol. T8, pages 49–58, 1984.

[Grevesse 93a] N. Grevesse & A. Noels. *Atomic Data and the Spectrum of the Solar Photosphere.* Physica Scripta, vol. T47, pages 133–138, 1993.

[Grevesse 93b] N. Grevesse & A. Noels. *Cosmic Abundances of the Elements.* In N. Prantzos, E. Vangioni-Flam & M. Cassé, editeurs, Origin and Evolution of the Elements, pages 15–25. Cambridge University Press, 1993.

[Gribov 69] V. Gribov & B. Pontecorvo. *Neutrino Astronomy and Lepton Charge.* Physics Letters B, vol. 28, pages 493–496, 1969.

[Guenther 93] D.B. Guenther, M.H. Pinsonneault & J.N. Bahcall. *The Effects of Helium Diffusion on Solar p-Mode Frequencies.* Astrophysical Journal, vol. 418, pages 469–475, 1993.

[Gyulassy 91] Miklos Gyulassy & Magnus Harlander. *Elastic Tracking and Neural Network Algorithms for Complex Pattern Recognition.* Computer Physics Communications, vol. 66, pages 31–46, 1991.

[Hahn 92] S. Hahn, K.H. Becks & A. Hemker. *Optimising Monte Carlo Generator Parameters Using Genetic Algorithms.* In Proceedings of the Second International Workshop on Software Engineering, Artificial Intelligence, and Expert Systems for High Energy and Nuclear Physics, pages 255–265, La Londe les Maures, France, 1992.

[Hahn 93] S. Hahn, K.H. Becks & A. Hemker. *Solving Optimisation Problems with Evolutionary Algorithms.* In Proceedings of the Third International Workshop on Software Engineering, Artificial Intelligence, and Expert Systems for High Energy and Nuclear Physics, pages 241–252, 1993.

[Hallin 93] A.L. Hallin. *Calibration Issues and Plan*, 1993. SNO Internal Report.

[Hata 94] N. Hata & P. Langacker. *Solar Model Uncertainties, MSW Analysis, and Future Solar Neutrino Experiments.* Physical Review D, vol. 50, no. 2, pages 632–660, 1994.

[Heaton 92] R. Heaton, H. Lee & B.C. Robertson. *α-Particle Induced High Energy γ-Rays from Light Elements*, 1992. SNO Internal Report.

[Hebb 49] D.O. Hebb. The Organisation of Behavior. New York: Wiley, 1949.

[Hernandez 94] F.P. Hernandez & J. Christensen-Dalsgaard. *The Phase Function of Stellar Acoustic Oscillations.* Monthly Notices of the Royal Astronomical Society, vol. 269, pages 475–492, 1994.

[Hertz 91] John Hertz, Anders Krogh & Richard G. Palmer. Introduction to the Theory of Neural Computation. Addison Wesley, 1991.

[Hime 95] A. Hime, 1995. Private communication.

[Hirata 88] K.S. Hirata *et al. Observation in the Kamiokande II Detector of the Neutrino Burst from Supernova SN1987A.* Physical Review D, vol. 38, no. 2, pages 448–458, 1988.

[Hirata 91] K.S. Hirata *et al. Real-Time, Directional Measurement of ^8B Solar Neutrinos in the Kamiokande II Detector.* Physical Review D, vol. 44, pages 2241–2260, 1991.

[Holland 75] J. H. Holland. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. University of Michigan Press, 1975.

[Holmgren 58] H.P. Holmgren & R. Johnston. *$^3He(\alpha,\gamma)^7Be$ Reaction.* Bulletin of the American Physical Society, vol. II, no. 3, page 26, 1958.

[Huebner 86] W.F. Huebner. *Atomic and Radiative Processes in the Solar Interior.* In P.A. Sturrock, T.E. Holzer, D.M. Mihala & R.K. Ulrich, editeurs, Physics of the Sun, volume I, pages 33–76. Reidel, 1986.

[ICRP 90] ICRP. Recommendation of the International Commision on Radiological Protection, Publication 60. Pergamon Press, 1990.

[Jelley 58] J.V. Jelley. Čerenkov Radiation. Pergamon Press, London, 1958.

[Kamiokande 90] The Kamiokande Collaboration. *Results from One Thousand Days of Real-Time, Directional Solar Neutrino Data.* Physical Review Letters, vol. 65, no. 11, pages 1297–1300, 1990.

[Kayser 81] B. Kayser. *On the Quantum Mechanics of Neutrino Oscillation.* Physical Review D, vol. 24, no. 1, pages 110–116, 1981.

[Keeler 90] J.D. Keeler, D.E. Rumelhart & W. Leow. *Integrated Segmentation and Recognition of Hand-Printed Numerals.* In Advances in Neural Information Processing Systems III, pages 557–563, 1990.

[Kirsten 95] T. Kirsten *et al. Update of GALLEX Solar Neutrino Results and Implications.* Nuclear Physics B (Proc. Suppl.), vol. 38, pages 68–76, 1995.

[Klein 94] Joshua Klein. *Maximum Likelihood Fitting to Both Time and Angle*, 1994. SNO-STR-94-055.

[Krane 88] K.S. Krane. Introductory Nuclear Physics. John Wiley and Sons, Inc., 1988.

[Kuo 89] T.K. Kuo & J. Pantaleone. *Neutrino Oscillations in Matter.* Reviews of Modern Physics, vol. 61, no. 4, pages 937–979, 1989.

[Lay 91] M.D. Lay, D.L. Wark & N. West. *It's Alive - The Birth of SNOMAN*, 1991. SNO-STR-91-077.

[Lay 94] M.D. Lay. *Creation and Detection of Čerenkov Light in the Sudbury Neutrino Observatory*, 1994. D.Phil. Thesis, Oxford.

[Lay 96] M.D. Lay. *Calibration and PMT Backgrounds in SNO*, 1996. SNO Internal Report.

[Le Cun 85] Y. Le Cun. *Une Procédure d'Apprentissage pour Réseau à Seuil Assymétrique.* In Cognitiva 85: A la Frontière de l'Intelligence Artificielle des Sciences de la Connaissance des Neurosciences, pages 599–604, Paris, 1985. Paris: CESTA.

[Le Cun 89a] Y. Le Cun *et al. Backpropagation Applied to Hand-Written Zip Code Recognition.* Neural Computation, vol. 1, pages 541–551, 1989.

[Le Cun 89b] Y. Le Cun *et al. Handwritten Digit Recognition with a Back-Propagation Network.* In D.S. Touretzky, editeur, Advances in Neural Information Processing Systems II, pages 396–404, 1989.

[Lederer 78] C.M. Lederer *et al.* Table of Isotopes. Wiley-Interscience, 1978.

[Lone 81] M.A. Lone, R.A. Leavitt & D.A. Harrison. *Prompt γs from Thermal Neutron Capture.* Atomic and Nuclear Data Tables, vol. 26, no. 6, pages 531–532, 1981.

[Louis 95] W.C. Louis. *Neutrino Oscillation Studies at LAMPF.* Nuclear Physics B (Proc. Suppl.), vol. S38, pages 229–234, 1995.

[Lyon 95] J. Lyon, 1995. Private communication.

[Mazzanti 93] P. Mazzanti & R. Odorico. *Bottom Jet Recognition by Neural Networks and Statistical Discriminants.* Zeitschrift für Physik C, vol. 59, pages 273–282, 1993.

[McCulloch 43] W.S McCulloch & W. Pitts. *A Logical Calculus of Ideas Immanent in Nervous Activity.* Bulletin of Mathematical Biophysics, vol. 5, pages 115–133, 1943.

[Mikheyev 86a] S.P. Mikheyev & A.Yu. Smirnov. . Soviet Journal of Nuclear Physics, vol. 42, page 913, 1986.

[Mikheyev 86b] S.P. Mikheyev & A.Yu. Smirnov. *Resonant Amplification of ν Oscillations in Matter and Solar Neutrino Spectroscopy.* Nuovo Cimento C, vol. 9, pages 17–26, 1986.

[Minsky 67] M.L. Minsky. Computation: Finite and Infinite Machines. Englewood Cliffs: Prentice-Hall, 1967.

[Minsky 69] M.L. Minsky & S.A. Papert. Perceptrons. Cambridge: MIT Press, 1969.

[Moeck 95] J. Moeck *et al.* *Artificial Neural Networks as a Second Level Trigger at the H1 Experiment - Performance Analysis and Results.* In Proceedings of the Fourth International Workshop on Software Engineering, Artificial Intelligence, and Expert Systems for High Energy and Nuclear Physics, Pisa, Italy, 1995.

[Moorhead 92] M.E. Moorhead, 1992. D.Phil. Thesis, Oxford University.

[Moorhead 95a] M. Moorhead. *Grid Fitter for Reducing Tails in Spatial Distributions*, 1995. SNO Internal Report.

[Moorhead 95b] M. Moorhead *et al.* ^{6}*Li Neutron Poison for SNO*, 1995. SNO-STR-95-056.

[Nakamura 93] K. Nakamura. *Recent Results from Kamiokande Solar Neutrino Observations.* Nuclear Physics B (Proc. Suppl.), vol. 31, pages 105–110, 1993.

[Nelson 85] W.R. Nelson, H. Hirayama & D.W.O. Rogers. *The EGS4 Code System*, 1985. SLAC Report 265.

[Nico 94] J. Nico *et al.* *Results from SAGE II.* In P.J. Bussey & I.G. Knowles, editeurs, Proceedings of the XXVII Conference on High Energy Physics, Glasgow, pages 965–967. IOP Publishing, 1994.

[Ohlsson 92] Mattias Ohlsson, Carsten Peterson & Alan Yuille. *Track Finding with Deformable Templates – The Elastic Arms Approach.* Computer Physics Communications, vol. 71, No.1–2, pages 77–98, 1992.

[Ohlsson 93] Mattias Ohlsson. *Extensions and Explorations of the Elastic Arms Algorithm.* Computer Physics Communications, vol. 77, No.1, pages 19–32, 1993.

[Paoluzi 89] L. Paoluzi. *The Gran Sasso National Laboratory.* Nuclear Instruments and Methods in Physics Research A, vol. 279, pages 133–136, 1989.

[Parke 86] S.J. Parke. *Non-Adiabatic Level Crossing in Resonant Neutrino Oscillations.* Physical Review Letters, vol. 57, pages 1275–1278, 1986.

[Parke 95] S. Parke. *Status of the Solar Neutrino Puzzle.* Physical Review Letters, vol. 74, no. 6, pages 839–841, 1995.

[Parker 85] D.B. Parker. *Learning Logic*, 1985. Center for Computational Research in Economics and Management Science, MIT. Technical Report TR-47.

[Peterson 92] Carsten Peterson. *Pattern Recognition in High Energy Physics with Neural Networks.* In L. Cifarelli, editeur, QCD at 200TeV, pages 149–163, 1992.

[Pontecorvo 58a] B. Pontecorvo. *Inverse Beta Processes and Nonconservation of Lepton Charge.* Soviet JETP, vol. 7, pages 172–173, 1958.

[Pontecorvo 58b] B. Pontecorvo. *Mesonium and Antimesonium.* Soviet JETP, vol. 6, pages 429–431, 1958.

[Pontecorvo 68] B. Pontecorvo. *Neutrino Experiments and the Problem of Conservation of Lepton Charge.* Soviet JETP, vol. 26, pages 984–988, 1968.

[Poon 93] A. Poon & R.G.H. Robertson. *Schedule for the LANL $^3H(p,\gamma)^4He$ Calibration Source*, 1993. SNO-STR-93-047.

[Popov 95] G.E. Popov *et al.* *Centrifugal Enrichment of Chromium-50 for Experiments on Detecting Solar Neutrinos.* Nuclear Instruments and Methods in Physics Research A, vol. 362, pages 532–537, 1995.

[Press 92] W.H. Press *et al.* Numerical Recipies in Fortran: The Art of Scientific Computing. Cambridge University Press, 2 edition, 1992.

[Proriol 92] J. Proriol. *Pattern Recognition: Invariants in 3D.* In Proceedings of the Neural Networks Conference: From Biology to High Energy Physics, volume 3(Supp. 1992), pages 271–276. World Scientific, 1992.

[Radcliffe 93] T. Radcliffe. *Extraction of Optical Properties of the SNO Detector from in situ Measurements*, 1993. SNO Internal Report.

[Ranucci 95] G. Ranucci, 1995. talk given at The International Europhysics Conference on High Energy Physics, Brussels.

[Robertson 92] R.G.H. Robertson *et al.* *Neutral Current Detection in the Sudbury Neutrino Observatory*, 1992. Los Alamos funding request FIN-94-ER-E324.

[Rogers 84] D.W.O. Rogers. *Low Energy Transport with EGS.* Nuclear Instruments and Methods in Physics Research A, vol. 227, pages 535–548, 1984.

[Rogers 92] F.J. Rogers & C.A. Iglesias. *Radiative Atomic Rosseland Mean Opacity Tables.* Astrophysical Journal Supplement, vol. 79, pages 507–568, 1992.

[Rumelhart 86a] D.E. Rumelhart, G.E. Hinton & R.J. Williams. *Learning Representations by Back-Propagating Errors.* Nature, vol. 323, pages 533–536, 1986.

[Rumelhart 86b] D.E. Rumelhart, J.L. McClelland & the PDP Group. Explorations in the Microstructue of Cognition, Volume 1: Foundations. Cambridge: MIT Press, 1986.

[Schalkoff 92] Robert Schalkoff. Pattern Recognition; Statistical, Structural, and Neural Approaches. Wiley, 1992.

[Schwarzschild 58] M. Schwarzschild. Structure and Evolution of the Stars. Princeton University Press, 1958.

[Shapiro 90] J. Shapiro. Radiation Protection. Harvard University Press, 3rd edition, 1990.

[Skensved 90] P. Skensved & B.C. Robertson. *Miscellaneous Effects*, 1990. SNO-STR-90-028.

[Skensved 91] P. Skensved & B.C. Robertson. *Monte Carlo Status Report*, 1991. SNO-STR-91-073.

[Skensved 94] P. Skensved & B.C. Robertson. *Summary of Backgrounds in SNO*, 1994. SNO-STR-94-013.

[SNO 87a] The SNO Collaboration. *Collection of Annexes in Support of the Main Proposal*, 1987. SNO-STR-87-012.

[SNO 87b] The SNO Collaboration. *The Sudbury Neutrino Observatory Proposal*, 1987. SNO-STR-87-12.

[SNOMAN 95a] The SNOMAN Authors. *The SNOMAN Programmers Manual, Version 2_08*, 1995.

[SNOMAN 95b] The SNOMAN Authors. *The SNOMAN Users Manual, Version 2_08*, 1995.

[Suzuki 94] Y. Suzuki. *The Superkamiokande Project.* Nuclear Physics B (Proc. Suppl.), vol. 35, pages 273–275, 1994.

[Suzuki 95] Y. Suzuki. *Kamiokande Solar Neutrino Results.* Nuclear Physics B (Proc. Suppl.), vol. 38, pages 54–59, 1995.

[Svoboda 85] R.C. Svoboda, 1985. Ph.D. Thesis, University of Hawaii.

[Tatara 90] N. Tatara, Y. Kohyama & K. Kubodera. *Weak Interaction Processes on Deuterium: Muon Capture and Neutrino Reactions.* Physical Review C, vol. 42, pages 1694–1717, 1990.

[Teasdale 92] W.A. Teasdale, 1992. Engineering drawings for the SAGE calibration source and shipping assembly.

[t'Hooft 71] G. t'Hooft. *Prediction for Neutrino-Electron Cross-Sections in Weinberg's Model of Weak Interactions.* Physics Letters B, vol. 37, pages 195–196, 1971.

[Thorman 94] M.D. Thorman. *The Application of a Quad Fitter to SNO Data*, 1994. SNO Internal Report.

[Thorman 96] M.D. Thorman, 1996. Private Communication.

[Timofeyev 93] P.V. Timofeyev. *Report on the Measurements of the Chemical Content of Impurites and Isotopic Content of Metallic Chromium Samples for the Neutrino Source*, 1993. Troitsk Internal Report.

[Turck-Chièze 88] S. Turck-Chièze, S. Cahen, M. Cassé & C. Doom. *Revisiting the Standard Solar Model.* Astrophysical Journal, vol. 335, pages 415–424, 1988.

[Turck-Chièze 93] S. Turck-Chièze & I. Lopes. *Toward a Unified Classical Model of the Sun: On the Sensitivity of Neutrinos and Helioseismology to the Macroscopic Physics.* Astrophysical Journal, vol. 408, pages 347–367, 1993.

[UNSCEAR 77] UNSCEAR. United Nations Scientific Committee on the Effects of Atomic Radiation Report to the General Assembly. New York, United Nations, 1977.

[van Middelkoop 66] G. van Middelkoop & P. Spilling. *Gamma-Gamma Angular Correlation Measurements in the $^{35}Cl(n,\gamma)^{36}Cl$ Reaction.* Nuclear Physics, vol. 77, pages 267–275, 1966.

[Vazquez 92] R.A. Vazquez, F. Halzen & E. Zas. *Improving the Čerenkov Imaging Technique with Neural Networks.* Physical Review D, vol. 345, no. 1, pages 356–361, 1992.

[Vignaud 95] D. Vignaud *et al.*, 1995. talk given at the IV International Workshop on Theoretical and Phenomenological Aspects of Underground Physics.

[Wark 93] D.L. Wark & N. West. *Proposal for New SNOMAN Geometry Software*, 1993. SNO Internal Report.

[Weinberg 67] S. Weinberg. *A Model for Leptons.* Physical Review Letters, vol. 19, pages 1264–1270, 1967.

[Welford 89] W.T. Welford & R. Winston. High Collection Non-Imaging Optics. Academic Press, 1989.

[Wolfenstein 78] L. Wolfenstein. *Neutrino Oscillations in Matter.* Physical Review D, vol. 17, pages 2369–2374, 1978.

[Wolfenstein 79] L. Wolfenstein. *Neutrino Oscillations and Stellar Collapse.* Physical Review D, vol. 20, pages 2634–2635, 1979.

[Ying 92] S. Ying, W.C. Haxton & E.M. Henley. *Charged and Neutral Current Solar Neutrino Cross Sections for Heavy Water Čerenkov Detectors.* Physical Review C, vol. 45, no. 4, pages 1982–1987, 1992.

[Zell 95] Andreas Zell *et al. SNNS User Manual, Version 4.0.* University of Stuttgart, Institute for Parallel and Distributed High Performance Systems, 1995.

[Zener 32] C. Zener. *Non-Adiabatic Crossing of Energy Levels.* Procedings of the Royal Society of London, Series A, vol. 137, pages 696–702, 1932.

[Zwinkels 90] J.C. Zwinkels, W.F. Davidson & C.X. Dodd. *Optical Properties of UV Transmitting Acryics for use in a Heavy Water Čerenkov Detector.* Applied Optics, vol. 29, no. 22, pages 3240–3248, 1990.