Monte Carlo Study of Beta-Gamma Backgrounds from ^{208}Tl and ^{214}Bi

X.Chen,Oxford University
SNO-STR-96-003

January 31, 1996

Abstract

The major background in SNO is from Beta-Gamma decay of ^{208}Tl in the Thorium chain and ^{214}Bi in the Uranium chain.This paper presents some Monte Carlo results using SNOMAN 2_09 development code:

- Neutron capture situation inside detector.
- ^{208}Tl and ^{214}Bi background compared to charged current and neutral current spectrum.
- A neural network is used to distinguished the Beta-gamma decay of ^{208}Tl from the Beta-gamma decay of ^{214}Bi.

1 Introduction

The beta-gamma decay of ^{208}Tl in the Thorium chain and ^{214}Bi in the Uranium chain constitutes the most important background in SNO:

- High energy decay products mimic the electron produced by charged current event so that it determines the value of energy threshold.
- High energy gamma rays from decay produce free neutrons by photodisintegration of deuterons,which is indistinguishable from a neutral current event.

First the study of neutron capture will be presented in section 2.Section 3 is about the ^{208}Tl and ^{214}Bi background wall compared with charged current and neutral current spectrum.In section 4, a feedforward neural network is used to distinguish ^{208}Tl decay from ^{214}Bi decay in D_2O in order to decide their concentration level independently.

2 Neutron Capture Inside Detector

Assuming 2 tons $MgCl_2$ are dissolved in 1000 tons 99.92% pure heavy water(the other 0.08% is light water), a free neutron will be captured by chlorine,deuteron and hydrogen inside detector.

Generate 10000 free NC neutrons isotropically in D_2O, 9268 neutrons produce sink vertices inside the detector.

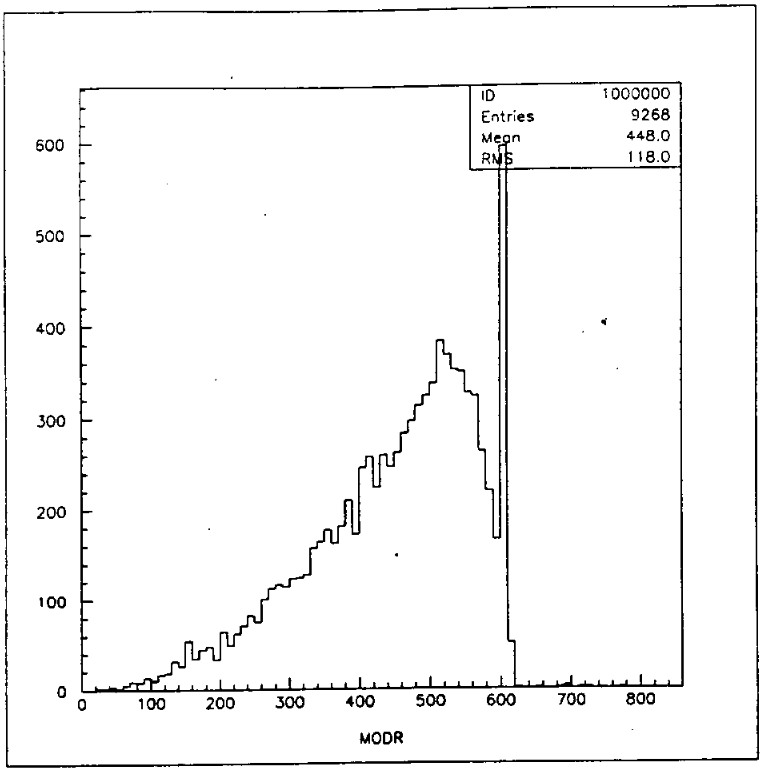

Figure 1: R_distribution of neutron sink vertices,unit=cm

$$capture\ rate = \frac{capture\ event\ number}{total\ event\ number}$$

Cl:

$$capture\ rate = \frac{8271}{10000} = 0.827 \pm 0.0091$$

D:

$$capture\ rate = \frac{258}{10000} = 0.0258 \pm 0.0016$$

H(inside D_2O sphere):

$$capture\ rate = \frac{96}{10000} = 0.0096 \pm 0.00098$$

H(outside D_2O sphere):

$$capture\ rate = \frac{643}{10000} = 0.0643 \pm 0.0025$$

Figure 1 is the R-distribution of neutron sink vertices.Note that there is a valley from 550cm to 600cm.This region contains the largest volume heavy water and so produces many more neutrons than others. No neutrons are from the H_2O and the length of neutron's free length in D_2O(salt included) is roughly half a meter.

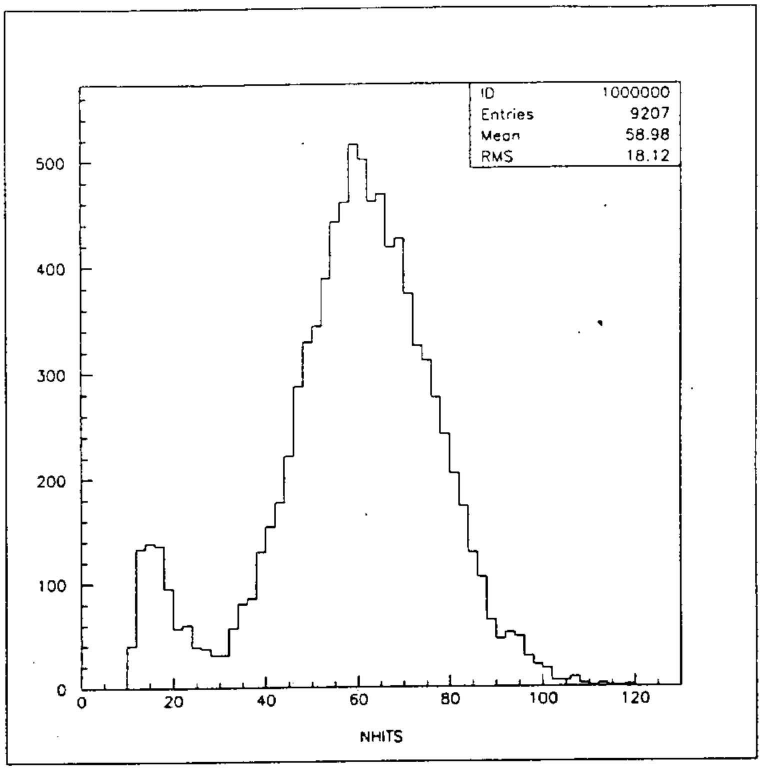

Figure 2: Nhits of Neutron Absorption

The capture of neutrons on H,D and Cl will produce 2,6.25 and 8.6 Mev Gamma rays respectively. As we can see from figure 2, there are two peaks: the large one is caused by high energy gamma rays(6.25,8.6Mev),the small one caused by low energy Gamma rays(2Mev).High energy gamma rays(although sometimes energy is shared by several gamma rays) generated by absorption of neutrons can create new free neutrons by photodisintegration of deuteron. MC result is : 10000 neutrons produce 58 new neutrons.

3 ^{208}Tl and ^{214}Bi background in SNO

The assumed of concentration levels of Thorium and Uranium in different parts of detector are:

region	Thorium Concentration(g/g)	^{208}Tl decay per year
1000 tons D_2O	$11 * 10^{-15}$	505890
30 tons Acrylic Vessel	$4.6 * 10^{-13}$	$6.35 * 10^5$
8000 tons H_2O	$7.52 * 10^{-14}$	$5.2 * 10^6$

region	Uranium Concentration(g/g)	^{214}Bi decay per year
1000 tons D_2O	$11 * 10^{-15}$	$4.38 * 10^6$
30 tons Acrylic Vessel	$6.3 * 10^{-13}$	$7.5 * 10^6$
8000tons H_2O	$8.6 * 10^{-14}$	$5.1 * 10^7$

Assuming ^{232}Th and ^{238}U chain are in secular equilibrium ,then from the White Book:

- ^{232}Th gives 126 2.615Mev γ-rays $ug^{-1}d^{-1}$.

- only 1.54% ^{214}Bi decays to produce 2.445Mev γ-rays.

- ^{238}U produces 16.8 2.445Mev γ-rays $ug^{-1}d^{-1}$.

Charged Current Background.

The background wall from ^{208}Tl and ^{214}Bi decay in D_2O ,Acrylic Vessel and H_2O are drawn with Charged Current and Neutral Current spectra in figure 3(Assuming 4500 CC event per year(1/3 SSM flux) and 5295 NC events per year(Full SSM flux)). The grid fitter was used . CC event number is less than the background event number at 50 Hits.

Neutral Current Background.

Gamma rays greater than 2.223Mev in heavy water can cause deuteron photodisintegration, mimicking the neutrino disintegration of deuteron.Decay of ^{208}Tl produce 2.615 Mev Gamma rays and 1.54% of decay of ^{214}Bi produce 2.445 Mev Gamma rays.The Monte Carlo result is one neutron will be produced by 724.6 2.615Mev or 1153.8 2.445Mev gamma rays.The white book value are 470 and 750 respectively.

Region	Decay type	MC. event	Neutrons produced	Decay/day	Neutrons/day
D_2O	^{208}Tl	70000	105	1386	2.079
D_2O	^{214}Bi	300000	5	12000	0.2
Acrylic Vessel	^{208}Tl	200000	100	1738.8	0.869
Acrylic Vessel	^{214}Bi	1000000	4	20618.6	0.082
H_2O	^{208}Tl	200000	2	$1.4 * 10^4$	0.14
H_2O	^{214}Bi	2000000	0	139636	

4 Neural Network Study of ^{208}Tl and ^{214}Bi Decay in D_2O

4.1 Introduction

Feedforward and error backpropagation neural networks have been used widely in high energy physics(See [1],[2],[3]).In SNO, S.Brice has used it to distinguish CC and NC event in D_2O and got good results[4].

In future SNO experiment, if one subtract the number of event with salt in D_2O by the number of event without salt in D_2O, he will get rid of almost all background and CC event except NC background from ^{208}Tl and ^{214}Bi Decay.Because almost all ^{208}Tl decay will produce 2.615Mev gamma rays enough to make photodisintegration and only 1.54% ^{214}Bi decay can produce 2.445Mev gamma rays enough to make photodisintegration, it is valuable to distinguish ^{208}Tl decay from ^{214}Bi decay so people can calculate NC background. 'Th/U monitoring window' (M.Moorhead[7]) has been proposed to measure the ^{208}Tl and ^{214}Bi concentrations in the D_2O independently by using neural network to recognize different hit patterns of these decays.There are two proposed monitoring windows,only 'high nhits monitoring window' is considered here. It is constituted by two cuts: rfit(distance from the center of detector to the fitted vertex)<400cm and $30 < Nhits < 40$. There are two reasons for choosing this window:1.above 30 nhits ^{214}Bi decay event are mostly single electron event (3.27Mev) whereas ^{208}Tl decay events are always one 2.6145

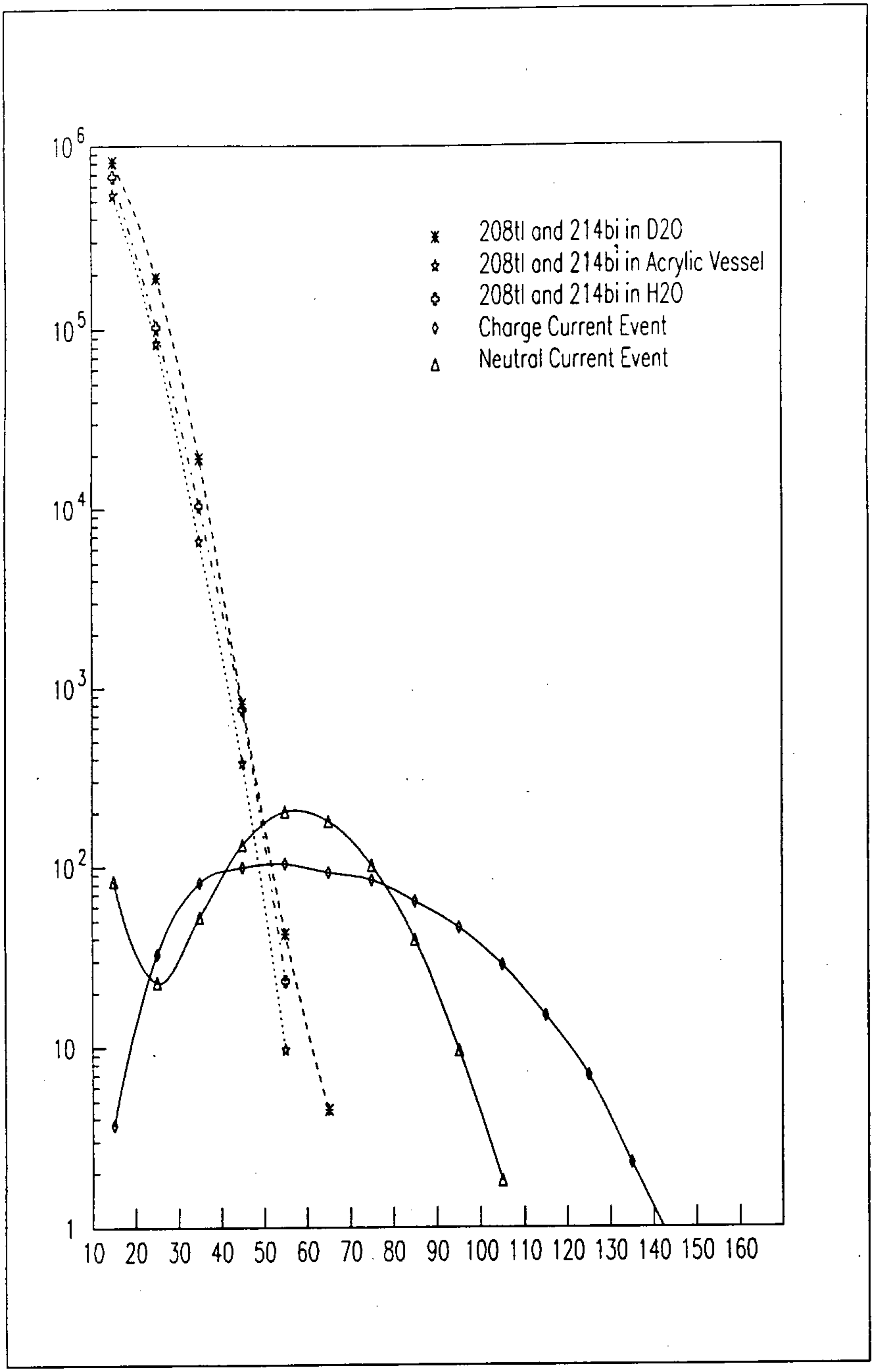

Figure 3: Event energy distribution from a year data,reconstructed inside 600cm per year per 10 hits,unit=nhits

Mev gamma + an electron + one or more lower energy gamma. 2. Hopefully, rfit and nhits cut will let ^{208}Tl decay and ^{214}Bi decay event in D_2O become dominant part in this window.

4.2 Prerequisites

As mentioned above, this "Th/U monitoring window" should be relatively clean of other signals compared to the ^{208}Tl decay and ^{214}Bi decay event in the D_2O. Preliminary study is presented below,using the grid fitter.

type	concentration level(g/g)	No. of events reconstructed in this window
^{208}Tl in D_2O	Th: $11 * 10^{-15}$	8159.3/year
^{214}Bi in D_2O	U: $11 * 10^{-15}$	6789/year
^{208}Tl in Acrylic Vessel	Th: $4.6 * 10^{-13}$	95.2/year
^{214}Bi in Acrylic Vessel	U: $6.3 * 10^{-13}$	150/year
^{208}Tl in H_2O	Th: $7.52 * 10^{-14}$	125/year
^{214}Bi in H_2O	U: $8.6 * 10^{-14}$	637/year
Charged Current Event	CC No:4500	84.9/year

In the future,the fitter can be optimized to improve the purity within this window.For example, the quad fitter + elastic fitter was used instead of grid fitter on ^{208}Tl decay event in the H_2O. It improved the accuracy of the fitted vertices.Then only 80 events per year are reconstructed in this monitoring window.

4.3 Neural Network Study

For elementary knowledge of neural network,see [5].A neural network simulator built by Stuttgart University was used.The simplest feedforward and error-backpropagation neural network was applied.The whole network is made of 3 layers: an input layer(28 input units),a hidden layer (4 hidden units) and an output layer(1 output unit).The hit pattern is represented by 28 input parameters X_{inp},the result pattern is represented by 1 output value Y_{out}. The rule is :

- $0. \leq Y_{out} \leq 1.$.
- If $0. \leq Y_{out} \leq 0.5$,then the neural network decides this event belongs to ^{214}Bi decay .
- If $0.5 \leq Y_{out} \leq 1.$, then the neural network decides this event belongs to ^{208}Tl decay .

4.3.1 Event by Event separation

2000 Tl + 2000 Bi event were used to train the neural network,then the other 2000 Tl + 1000Bi event were used to test the performance of the network. The result is as below:

	$0.5 \leq Y_{out} \leq 1.$	$0. \leq Y_{out} \leq 0.5$
Tl decay event	1118	882
Bi decay event	259	741
Percentage Correct = (62.0 +/- 0.9)%		

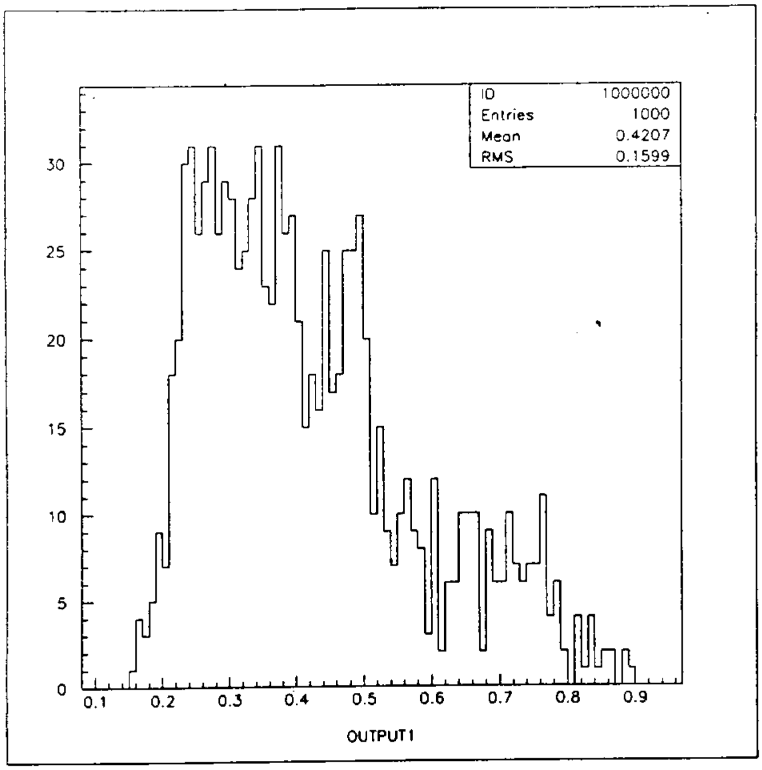

Figure 4: Neural network output of ^{214}Bi event

The output value of network shown in figure 4 and figure 5.We can see neural network can recognize ^{214}Bi events well(the event number of $output < 0.5$ is much larger then the number of $output > 0.5$),but it recognizes ^{208}Tl events poorly (the event number of $output > 0.5$ is only slightly larger then the number of $output < 0.5$).

Defining two variables:

- $$Purity = \frac{number\ of\ correctly\ recognized\ events}{number\ of\ total\ events}$$

- $$Efficiency = \frac{number\ of\ events\ classified}{number\ of\ total\ events}$$

If we decrease the threshold of ^{214}Bi event from 0.5 to 0.4 , increase the threshold of ^{208}Tl event from 0.5 to 0.6 and discard the events when output value is between 0.4 and 0.6 , we will get a good separation(purity is improved) but fewer events(efficiency is decreased).Figure 6 is a plot of purity against efficiency.

4.3.2 Statistical distinguish

From above, the neural network can not distinguish ^{208}Tl and ^{214}Bi event by event very well.But what is important is the total number of ^{208}Tl events and the total number of ^{214}Bi events.So by assuming ^{208}Tl and ^{214}Bi decay are the dominant

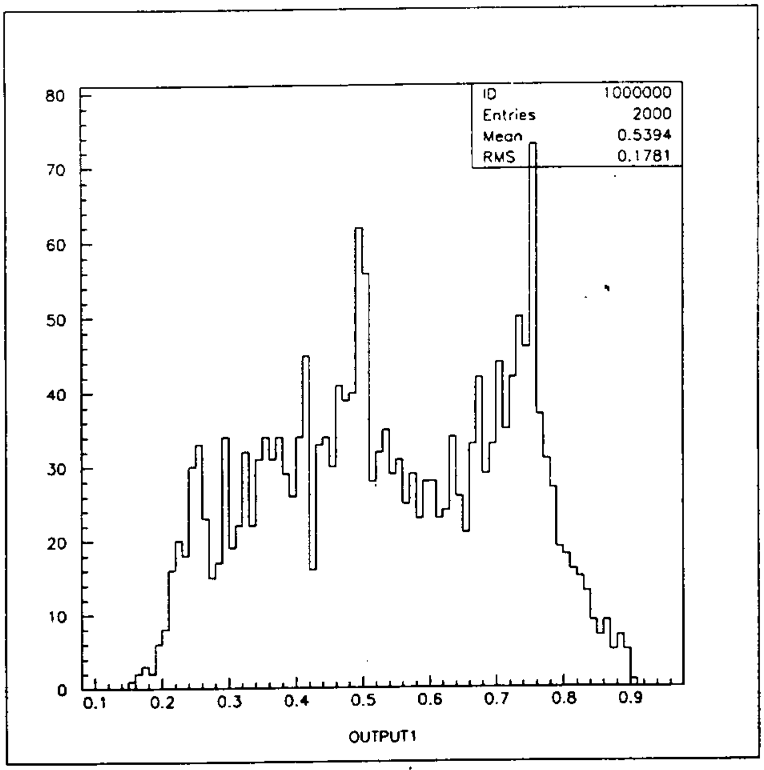

Figure 5: Neural network output of ^{208}Tl event

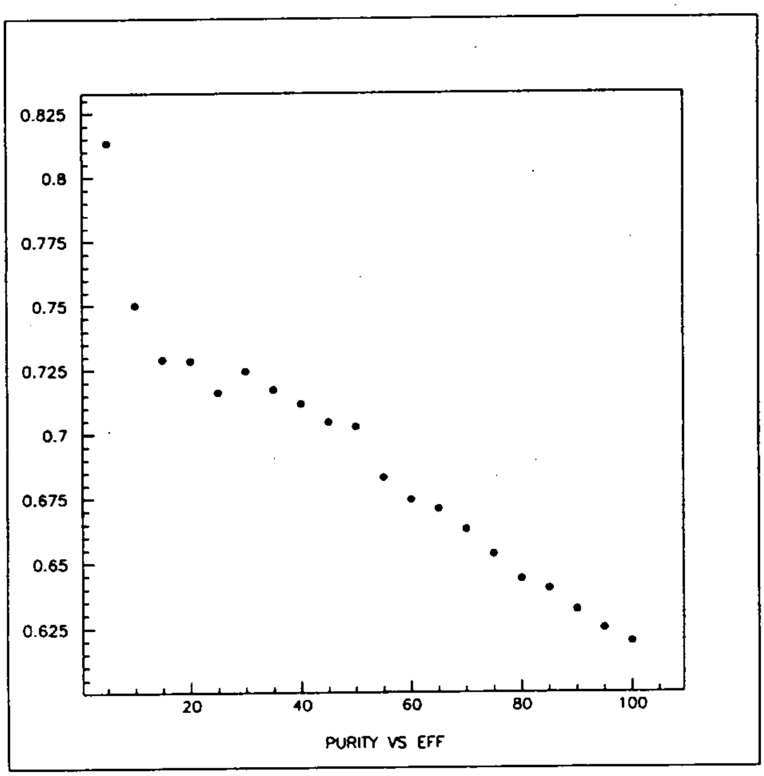

Figure 6: Neural network output of Purity against Efficiency

events inside the window, the task is: Given a set of mixed hit patterns of ^{208}Tl and ^{214}Bi events, find the correct fraction of ^{214}Bi events .

The method is:

1. Train the neural network with separate sets of ^{208}Tl and ^{214}Bi events.
2. Feed the other separate test sets of ^{208}Tl and ^{214}Bi events through the trained network, get the separate output distributions of the two classes events . Bin and normalize these distributions separately to get two calibration distributions.
3. Feed the mixed set of ^{208}Tl and ^{214}Bi events (or real mixed data in the window)through trained network and get the output distribution of mixed set .Bin and normalize this distribution to get the data distribution.
4. Fit the data distribution to some linear combination of two calibration distribution and so get the fraction of ^{214}Bi events in the mixed data.

Define a Chi-square function:

$$\chi^2 = \sum_i \frac{(d_i - (1-a)f_{1i} - af_{2i})^2}{\sigma_{d_i}^2 + (1-a)^2\sigma_{f_{1i}}^2 + a^2\sigma_{f_{2i}}^2}$$

i is bin number.d_i is the normalized data distribution value in the ith bin .f_{1i} is the normalized ^{208}Tl calibration distribution value in the ith bin. f_{2i} is the normalized ^{214}Bi calibration distribution value in the ith bin.

$$\sigma_{d_i}^2 = \frac{d_i}{N}$$

(N is the total number of data events)

$$\sigma_{f_{1i}}^2 = \frac{f_{1i}}{N1}$$

(N1 is the total number of ^{208}Tl calibration events)

$$\sigma_{f_{2i}}^2 = \frac{f_{2i}}{N2}$$

(N2 is the total number of ^{214}Bi calibration events)

If the number of calibration events is much larger then the number of data events, then $\sigma_{d_i}^2$ is much larger than $\sigma_{f_{1i}}^2$ and $\sigma_{f_{2i}}^2$,the error from calibration distribution can be ignored. Then the Chi-square function become simple:

$$\chi^2 = \sum_i \frac{(d_i - (1-a)f_{1i} - af_{2i})^2}{\sigma_{d_i}^2}$$

The normal linear Chi-square fit can be used and the error can be calculated. The MC result:

- train set:2000Tl+2000Bi

- test(calibration) sets:2000Tl ,2000Bi
- data set:400Tl+200Bi
- true Bi fraction: 0.333
- Chi-square fitted Bi fraction:0.32
- statistical error:0.756E-03
- Degree of Freedom:6. Chi-square value:2.53.

There are several thousand Tl and Bi events in the monitoring window per year.It will take enormous CPU time to generate calibration sets much larger than this number.If the calibration set number is as the same order as the data set,the error from calibration set can not be ignored.Two ways can be used to solve this problem.

- Minimize the Chi-square function directly using the Anneal-amoeba method ([6]) which is regarded as a useful way to avoid local minima.
- Change the Chi-square function to :

$$\chi^2 = \sum_i \frac{(1-X_i)^2}{1}$$

$$X_i = 1 - \frac{d_i}{\sqrt{b_i}} + \frac{(1-a)f_{1i} + af_{2i}}{\sqrt{b_i}}$$

$$b_i = \sigma_{d_i}^2 + (1-a)^2\sigma_{f_{1i}}^2 + a^2\sigma_{f_{2i}}^2$$

Then Levenberg-Marquart Method([6]) is used to fit this nonlinear Chi-square function.

Different combination of Tl and Bi events have been tried.The results are as below:

Data set	true Bi-fraction	Anneal method		
		Fitted Bi-fraction	Chi2 value	Degree of freedom
2800Tl+700Bi	0.2	0.199 +/- 3.9E-02	79	65
1000Tl+2000Bi	0.66	0.65 +/- 3.6E-02	65.6	64
1000Tl+1000Bi	0.5	0.52 +/- 4.47E-02	68	59
2000Tl+1000Bi	0.33	0.35 +/- 3.89E-02	78	65
1000Tl+3000Bi	0.75	0.73 +/- 3.34E-02	74	65

Data set	true Bi-fraction	Levenberg-Marquart method		
		Fitted Bi-fraction	Chi2 value	Degree of freedom
2800Tl+700Bi	0.2	0.244 +/- 4.06E-02	63.6	65
1000Tl+2000Bi	0.66	0.63 +/- 3.99E-02	47	64
1000Tl+1000Bi	0.5	0.52 +/- 4.59E-02	51.5	59
2000Tl+1000Bi	0.33	0.37 +/- 4.09E-02	57.5	65
1000Tl+3000Bi	0.75	0.70 +/- 3.73E-02	53.7	65

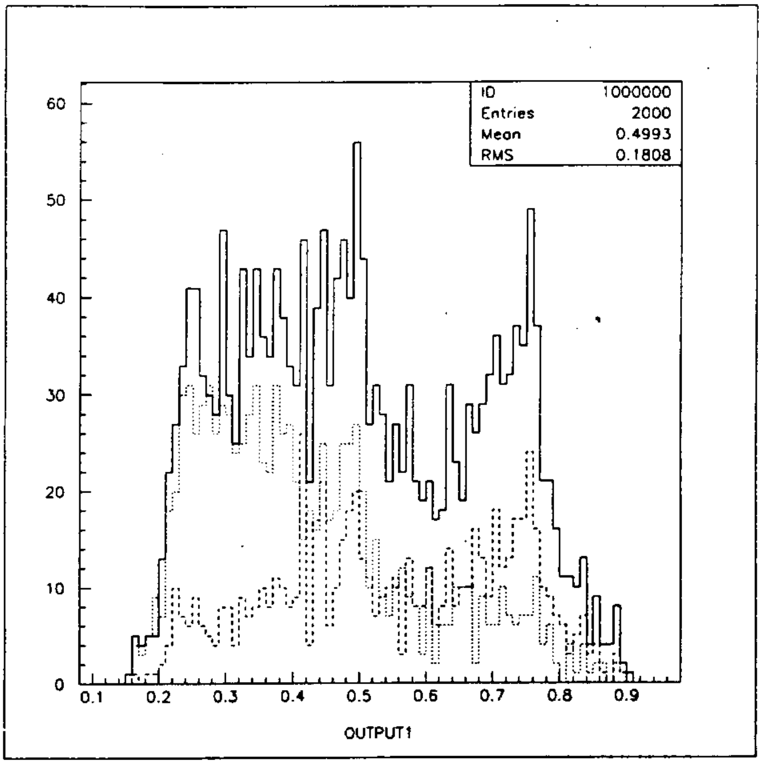

Figure 7: Neural network output of a mixed set of 1000Tl and 1000Bi events,dot line is calibration distribution of Bi and dash line is calibration distribution of Tl

From above ,we can see that both the Anneal and Levenberg-Marquart methods get nearly the same and good results.Figure 7 is neural network output of a mixed set of 1000Tl and 1000Bi events.

4.4 Robustness of Neural Network

Up to now, the calibration set can only come from Monte Carlo.If Monte carlo can not show the truth accurately or something changes from day to day in the detector(eg:the scattering coefficiency of Cerenkov photons in D_2O),it is doubtful whether neural network would still give us correct results.So the robustness of the network should be checked.

If one increases the Isothermal compressibility of both H_2O and D_2O to 10 times of their normal value ,it increases the amount of Rayleigh scattering by a factor of 10 .A test set(1000Tl+500Bi) fed through the original trained network and the result percentage correct was still 62% . But there is a big difference in detail.

Compare Figure 4 and Figure 8 ,we can see now there is a peak just below the 0.5 instead of around 0.25 in Figure 4.Of course this is due to the change of the Isothermal compressibility.But because we define a ^{214}Bi event by output value less then 0.5 ,neural network still get the same result. But the Chi-square test is much more sensitive to these changes than the event by event discrimination. There is no way to fit Figure 8 by Figure 4. Both the Anneal and the non-linear Chi-square fits give a negative result(which is physically impossible).The Chi-square fit can sense any shift of distribution provided bin size is small enough.But if bin size set at is 0.5 and two classes are still well separated ,hopefully the Chi-square fit can still get

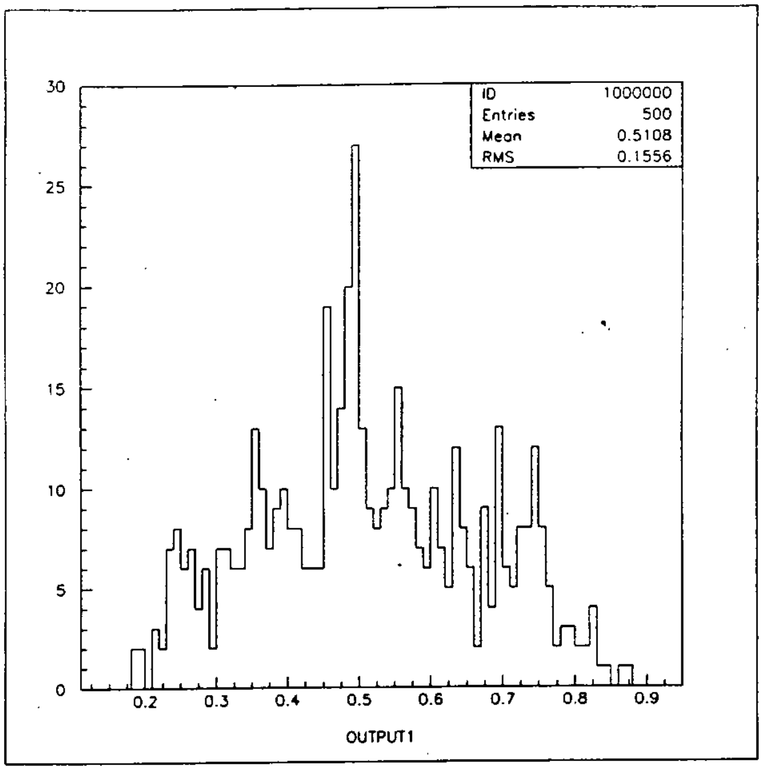

Figure 8: Neural network output of ^{214}Bi events after changing Isothermal compressibility of both H_2O and D_2O to 10 times of their normal value

a good result.

4.5 conclusion

Neural network + Chi-square fit can distinguish ^{208}Tl event and ^{214}Bi event.Initial results are not robust against dramatic changes in the systematics.We are studying this further to try to quantify the effect of believable uncertainties. Obviously true calibration is important.For example, ^{220}Rn in Thorium chain and ^{222}Rn in Uranium chain could be used as a calibration source.Because radon exists as gas which may be easily extracted from Thorium and Uranium and the half life is no more than several days so it would in principle not leave any radioactivity in the detector. In the future, more work will be done to improve the separation of two types of events.

5 Acknowledgement

Thanks to M.Moorhead for introducing the grid fitter , S.Brice for supplying a neural network to SNO and D.Wark for reading this paper.

References

[1] 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,295:383-395,1992)

[2] Bruce Denby, *Tutorial on neural network applications in high energy physics :1992 perspective.* (Proceedings of the Second International Workshop on Software Engineering,Artificial Intelligence,and Expert Systems for High Energy and Nuclear Physics,page 287-325,La Londe Les Maures,France,1992)

[3] Carsten Peterson.*Pattern recognition in high energy physics with neural networks.* (L.Cifarelli,editor,QCD at 200Tev,pages 149-163,1992)

[4] S.Brice *An overview of the Feedforward Neural Network Technique and its application to SNO event Classification*

[5] John Hertz,Anders Krogh and Richard G.Palmer *Introduction To the Theory of Neural Computation.*

[6] William H.Press , Saul A.Teukolsky, William T.Vetterling and Brian P.Flannery *Numerical Recipes in Fortran*

[7] M.Moorhead *Grid Fitter for Reducing Tails in Spatial Distribution*