Pattern Recognition for Event-Type Identification in the SNO Detector SNO-STR-94-019

T. J. Radcliffe
Dept. of Physics
Queen's University at Kingston
Ontario CANADA K7L 3N6

June 22, 1994

1 Introduction

The Hough Transform (HT) is a pattern recognition algorithm used to extract features from images [1]. This paper describes the application of a Circular Hough Transform (CHT) to extraction of event-type information from the PMT hit pattern of CC and NC (Cl(n,γ)) events. Preliminary results indicate that the CHT is capable of:

1. Accurately estimating the number of NC events in the spectrum on a statistical basis (that is, with no event-by-event identification).
2. Identifying $\sim$ 35% of NC events on an event-by-event basis with a few percent contamination by CC events.
3. A faint possibility of determining the fraction of *single* γ ray events in the spectrum with pure D_2O, particularly internal β/γ events between 30 and 50 hits.

All tests described here were performed using the standard data set from the Queen's Monte Carlo. Results for pure CC events are from the file cc_std.bin.1 and for pure NC events are from nc_std.bin.1. Each of these files contains 1000 events. The final testing was done on the file mixture_std.bin.1, which contains approximately 1000 events of each type. The real data may contain many fewer NC events than CC events, which makes the NC events more difficult to extract. Also, no internal β/γ events are included in the current analysis CC vs. NC events. For sufficiently high thresholds this should not make a difference. For low thresholds there is the possibility that the CHT can distinguish internal β/γ events from CC events. Further quantification of these results requires a more realistic test data set.

2 Circular Hough Transform

The CHT is applied to SNO detector hit patterns by first fitting the event (in this case with a modal/maximum likelihood fitter that will be described elsewhere) and finding the direction of the event based on the average direction of photons that hit within 4.8 nS of their expected arrival time. The hit pattern is then projected onto a plane that is 450 cm away from the reconstructed vertex, and whose normal is the reconstructed direction of the event. It is possible to perform the CHT in the spherical geometry of the detector, but much more complicated, so projection is an important simplification.

A circle in a plane can be parameterized by the position of its centre. To find a circle of fixed radius (about 400 cm for a 41 degree Čerenkov cone in the plane of projection) the Hough transform determines all the circles of that radius that a given PMT hit might lie on, and the circle that contains the most hits is deemed to be the circle defining the Čerenkov cone. This defines a new direction for the event, and by iterating this procedure until the event direction does not change the angular resolution of the detector is slightly improved (see Figure 1.)

Events that contain a single electron have a slightly more well-defined Čerenkov cone than either neutron captures on Cl or single γ rays. This means that the best circle (the one that has the most PMTs lying near it) for these events is slightly better than either of the other cases, and this fact can be used to distinguish the cases.

The CHT has been implemented using an array of pixels as shown in Figure 2. For each PMT that meets the time cut, the centre pixel is found, and a circle drawn in the array. This circle gives the possible locations of the centre of the circle that produced that PMT hit. Each pixel acts as an accumulator, summing up the weights of the circles that contribute to it. The pixel that has the most PMT hits contribute to it is taken as the central pixel. Once this procedure has been iterated to find the best direction of the event a more complete pattern is searched for. This pattern consists of a series of concentric circles of varying radius, weighted to reflect the angular distribution of a single-electron event. Using this more complete pattern was found to significantly improve the performance of the algorithm. In all cases the fraction of PMT hits that contributed to the best pattern was used as the final diagnostic. This has the advantage of being essentially independent of the number of hits, which makes comparison of diverse data sets simpler. This fraction is shown plotted against nhit for CC events in Figure 3 for a 50 cm pixel size. Pixel sizes of 100 cm, 50 cm and 25 cm were tested, with similar results. All results presented in this paper are for a 50 cm pixel size.

The angular distributions in question are shown in Figure 4. The true single-electron spectrum (for 7 MeV electrons at the centre of the vessel) shows a pronounced spike at 41 degrees. The NC events show a uniform distribution relative to the "true" event direction (an ambiguous designation at best) and a more sharply spiked distribution relative to the fitted direction. The distribution of CC events relative to the fitted direction is more sharply spiked still, and the problem of any pattern recognition algorithm is to distinguish between these two distributions.

3 Things That Don't Work So Well

A number of things other than the CHT were investigated, and are reported here for completeness. Looking at Figure 4, several obvious diagnostics for event-type discrimination come to mind. One is the likelihood that the angular distribution of in-time PMT hits in a given event are drawn from the expected distribution for CC events. Another is the ratio of the likelihoods for CC and NC events. A third is simply the fraction of PMT hits that fall between 0.6 and 0.8 in $\cos(\theta)$. Another is to look at the asymmetry in the event: the fraction of PMT hits in the forward direction.

None of these diagnostics work particularly well. The likelihood that a given distribution was drawn from the expected distribution gives some discrimination, but the peaks in the likelihood curves for the two event types, shown in Figure 5, are separated by only 0.8 σ. The ratio of the likelihoods has a larger width for about the same separation, and the ratio of the number of counts in the spike region shows a separation of only 0.7 σ between the peaks for the two types of event. The fraction of hits in the forward direction (relative to the fitted position and direction, which is all we know about) does not distinguish between event types at all.

Several of these diagnostics do contain some information on event type, and a more sophisticated scheme that uses this information will be the subject of future investigation. For the present, they will be ignored.

4 Fitting Procedure

In all cases described below events were fitted using a modal/maximum likelihood fitter to be described in detail elsewhere. Briefly, the fitter finds a point where the number of PMTs hit at close to the expected time is a maximum. Then, starting from this point, the point at which the distribution of the difference between the expected and actual arrival times looks most like a Gaussian with a width of 1.6 nS is found. In both cases the Numerical Recipes amoeba code is used to do the maximization. For 7 MeV electrons uniformly distributed in the D_2O the fitter has a resolution transverse to the reconstructed direction of 17.4 cm and in the longitudinal direction of 21.7 cm. These values are both somewhat worse than the values for the time fitter. The systematic pull in the longitudinal direction is 4.2 cm, significantly better than the time fitter. It is not clear why this fitter performs better than the time fitter: naively, one would expect a similar amount of pull from maximum likelihood as from least-squares.

The event direction was determined by taking first the average PMT direction from the fitted position, and then iterating on the CHT for a single Čerenkov ring as described above. Finally, the CHT value for a weighted ring pattern was calculated, and it is this value that is used to discriminate between CC and NC events. Because of a funny normalization for the weights this value is numerically equal to about half the fraction of the PMT hits that contribute to the best ring.

For the purposes of fitting the event only PMTs that are hit within 4.8 nS of the expected arrival time are used. However, in the following, when I refer to the number of hits I mean the full number of hits in the event, not

Event Type	Mean	σ
CC	0.1008 $\pm$ 0.0006	0.0170 $\pm$ 0.0004
NC	0.0815 $\pm$ 0.0006	0.0173 $\pm$ 0.0004
Internal β/γ	0.09111 $\pm$ 0.0004	0.0201 $\pm$ 0.0004
7 MeV γs	0.0948 $\pm$ 0.0006	0.0172 $\pm$ 0.0004

Table 1: Best Fits to CHT Distributions for Different Event Types

the number that were used in finding its position and direction.

5 CHT for Known CC and NC Events

The CHT results for known CC and NC events are shown in Figure 6. Gaussian fits to the curves are also shown – the reduced χ^2 is less than one in all cases. The best fit parameters are shown in Table 1. The peaks are separated by 1.1 σ. The fits to these curves are used as inputs to the event-type discriminator, and so to apply this technique to real data there must be a way to generate these curves. The NC curve can be generated from a neutron source at the centre of the detector. The CC curve must be generated with an electron source: as discussed below, single γ rays are sufficiently different from single electrons to prevent them from being used as source for this technique. Thus, the use of the CHT is critically dependent on the ^{8}Li source.

6 Results

There are two ways to extract the number of neutral current events from the detector spectrum using the CHT. One is to generate the CHT curve for all events above some threshold, and then fit that curve with the two CHT shapes extracted from the pure CC (i.e. ^{8}Li source) and pure NC (neutron source) data. This was done by maximizing the likelihood that the given distribution of CHT values was drawn from a weighted sum of the two Gaussians. Minimizing the mean-square error on the tails of the distribution also worked well.

This procedure is illustrated in Figure 7. The CHT distribution for the mixed events does look like the sum of the two shapes, and best fit is found for about 1000 events in each shape. A more interesting way of looking at this is to consider the number of NC events above a particular threshold in the number of hits. This is shown in Figure 8, where it is compared to the actual shape of the NC spectrum. As can be seen, the method gives an estimate of the number of NC events in the spectrum that agrees with the expected number within error. The estimated number of NC events in the mixture above 50 hits is 879 $\pm$ 30, and the expected number is 901. There is some loss in statistical precision in going to higher thresholds, but even at a threshold of 70 hits the estimated value is 387 $\pm$ 20 compared to an expected value of 416. The number of CC events found in the mixed data is systematically lower than the expected number by about 1 σ. Because the precise number of NC events in the mixed data is not known, it is not clear how much of this discrepancy is real.

The second method of determining the number of NC events is to attempt an event-by-event identification. This was done by setting a threshold on the CHT value of 0.075. Events below this were considered NC events. Applying this cut to pure NC events indicates that 35 % survive it, and 7 % of CC events survive it, for an enhancement of a factor of 5. Applying this cut to the mixed events yields the spectrum shown in Figure 9. No cut was placed on the number of hits. For comparison, the pure NC spectrum scaled by a factor of 0.35 is also shown.

7 Single γ Rays

In the course of investigating CC and NC events, the CHT curve for 7 MeV γ rays at the centre of the vessel was generated, in the hope that it would look like the CC curve so that the ^{8}Li source would not be needed to analyze the real data. It was found that this yielded a Gaussian fit with a mean of 0.0948 and $\sigma = 0.0172$, significantly different from (and about half-way between) the curves for CC and NC events.

This suggests that the present technique may be suitable to either extracting the number of 6.25 MeV γ rays from neutron capture on D_2O, or in identifying some of the internal β/γ background events at high energy, which would allow us to understand the shape of the tails near the threshold. The Gaussian fit parameters to the internal β/γ spectrum with nhit ≥ 30 is shown in Table 1. This fit is even more different from the CC shape than 7 MeV γ

Threshold	Expected CC	Expected β/γ	Extracted CC	Extracted β/γ
30	75	1166	70	1171
35	73	403	0	476
40	70	117	48	139
45	69	27	57	39
50	68	6	59	15

Table 2: Expected and Extracted CC and Internal β/γ Events for 10 Days at 1/3 SSM

events are. The standard data files contain only 10,000 internal β/γ events, which is equivalent to about 10 days running, during which time there will be 82 CC events at 1/3 the standard solar model rate. A mixed data file with 82 CC events and 10,000 internal β/γ events was made up and analyzed to extract the CHT shapes for the two event types. The results of this process are shown in Table 2.

The low statistics make it difficult to draw any conclusions from this exercise. The number of CC events does not look very well-determined, but the results are sufficiently suggestive to warrant further work.

A second source of single γ rays in pure D_2O is neutron captures on deuterium. This happens to about 1/4 of all neutrons created in the D_2O: the rest leak out or are captured on hydrogen. An attempt was made to distinguish these γ rays from CC events, but as shown in Table 1 the peaks are separated by only 0.35 σ, and so the two shapes could not be extracted from mixed data cleanly.

8 Conclusion

It has been shown that pattern recognition using the CHT is capable of distinguishing NC from CC events on a statistical basis, and in a more limited way on an event-by-event basis. It is possible that the shape of the internal β/γ tail can be extracted using the same method. These results depend on the following assumptions:

1. We can get the CHT curve for CC events from the ^{8}Li source

2. We can get the CHT curve for NC events from a neutron source

3. The Monte Carlo angular distribution of hits is not wildly more well-defined than the angular distributions we will observe

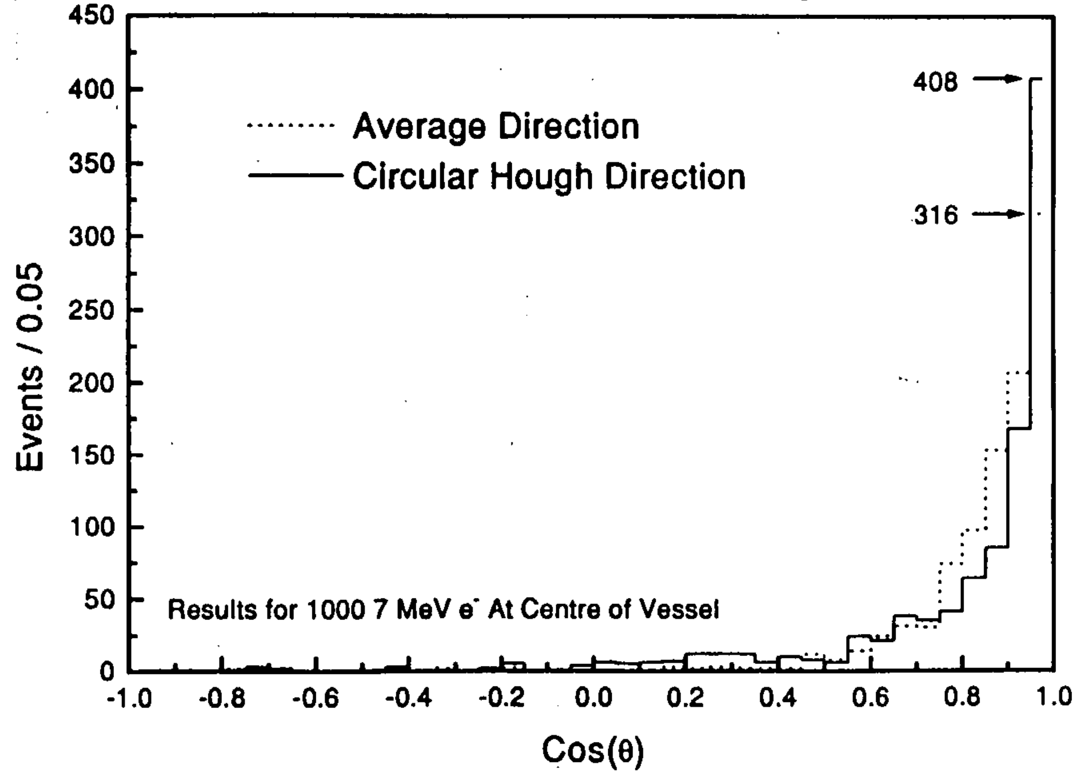
Also, the event-by-event extraction depends on there being comparable numbers of CC and NC events, a scenario that is likely only if neutrino oscillations are real. The statistical comparison, however, is much less sensitive to this effect, and is still capable of yielding a systematic check in the form of the extracted NC spectral shape as the threshold is raised.

If the angular distributions found in the Queen's Monte Carlo reflect those of the real data, the CHT will allow us to extract the NC and CC signals from the spectrum independently of any knowledge of the CC shape.

References

[1] J. Illingworth and J. Kittler, A Survey of the Hough Transform, Computer Vision, Graphics and Image Processing 44 (1988) 87 - 116

Figure 1: Angular Resolution of Circular Hough Transform



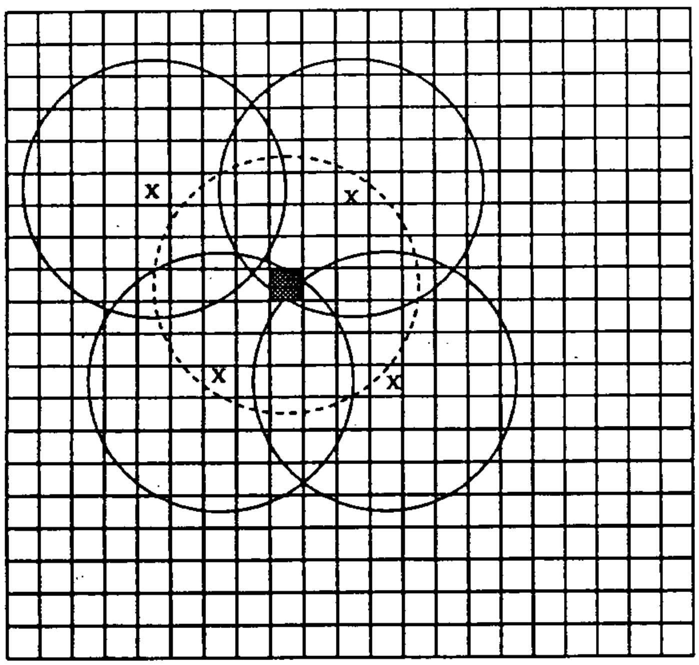


Figure 2: Illustration of Cirular Hough Transform

X Projected PMT hits

--- Best Circle

— Possible Centres for each hit

Figure 3: CHT vs. nhit for CC Events

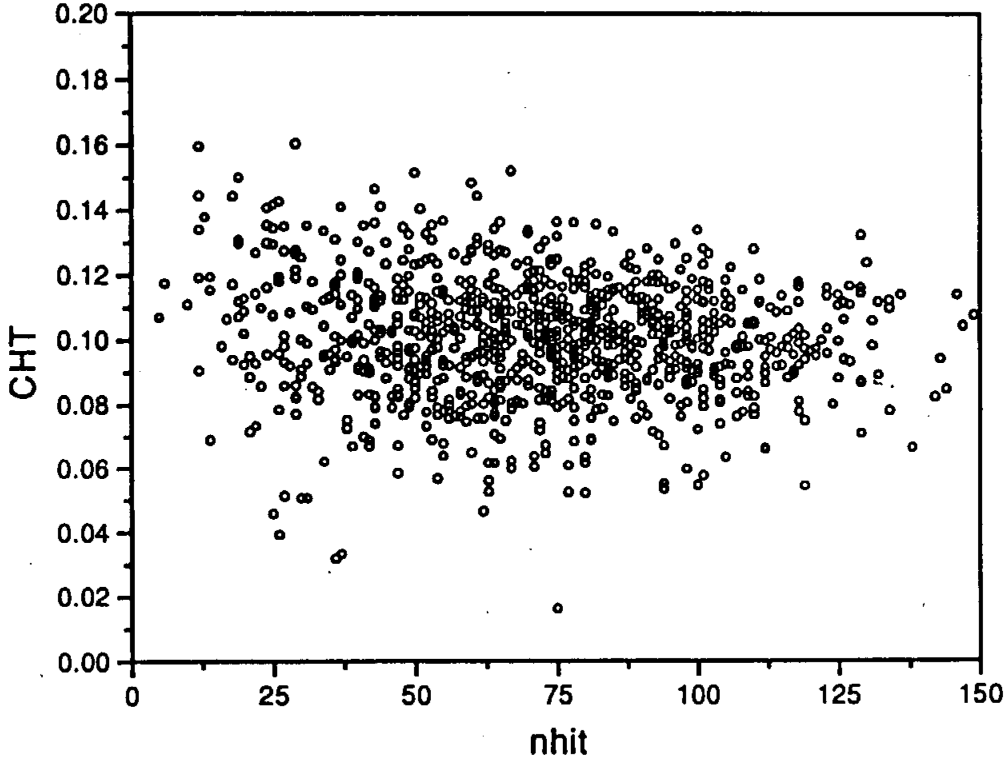


Figure 4: Angular Distributions for CC and NC Events

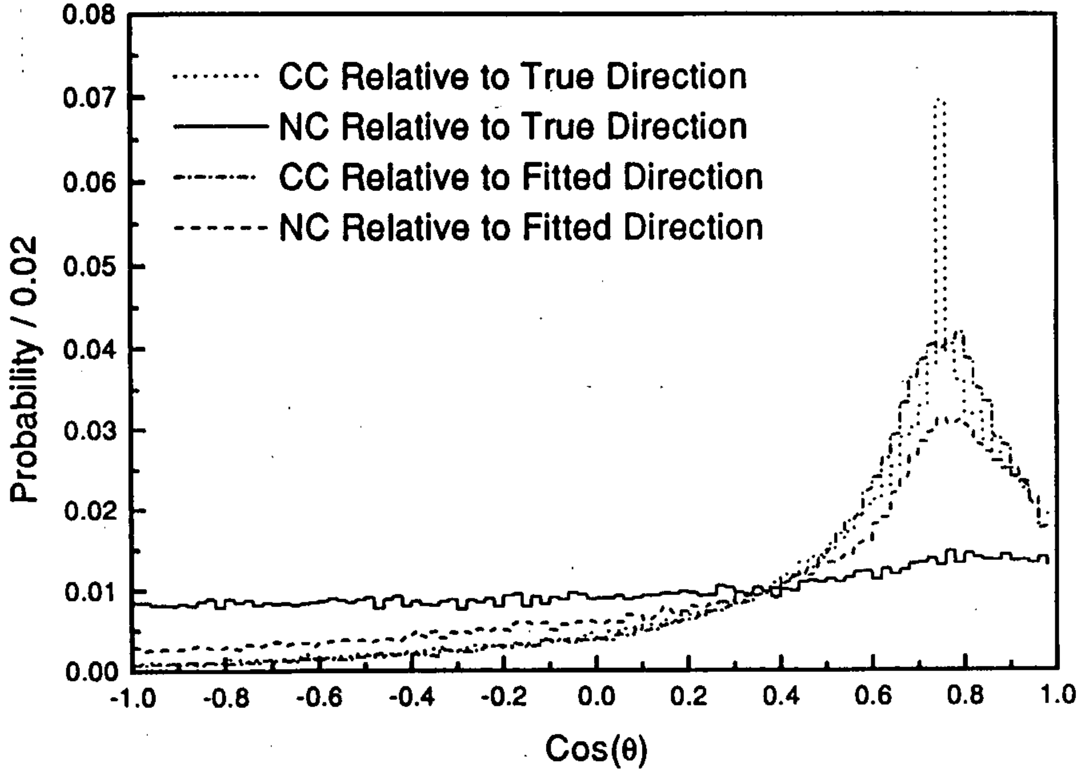


Figure 5: Likelihood for NC and CC Events to be Drawn from CC Distribution

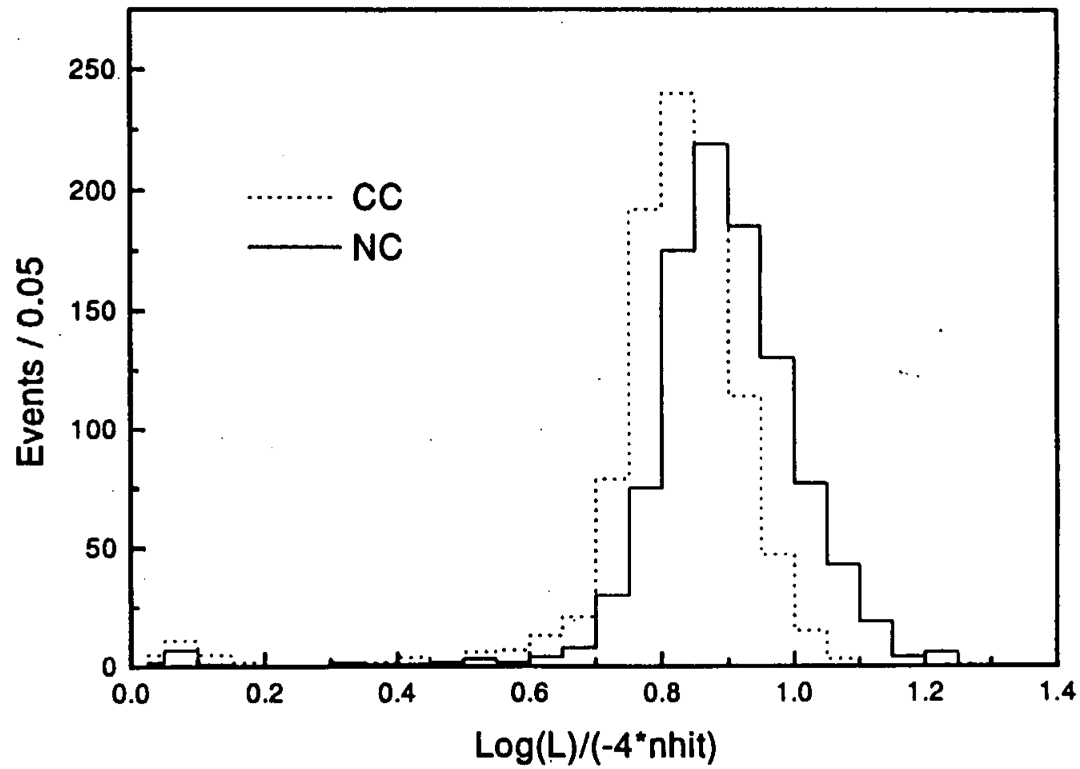


Figure 6: CHT Curves for CC and NC Events

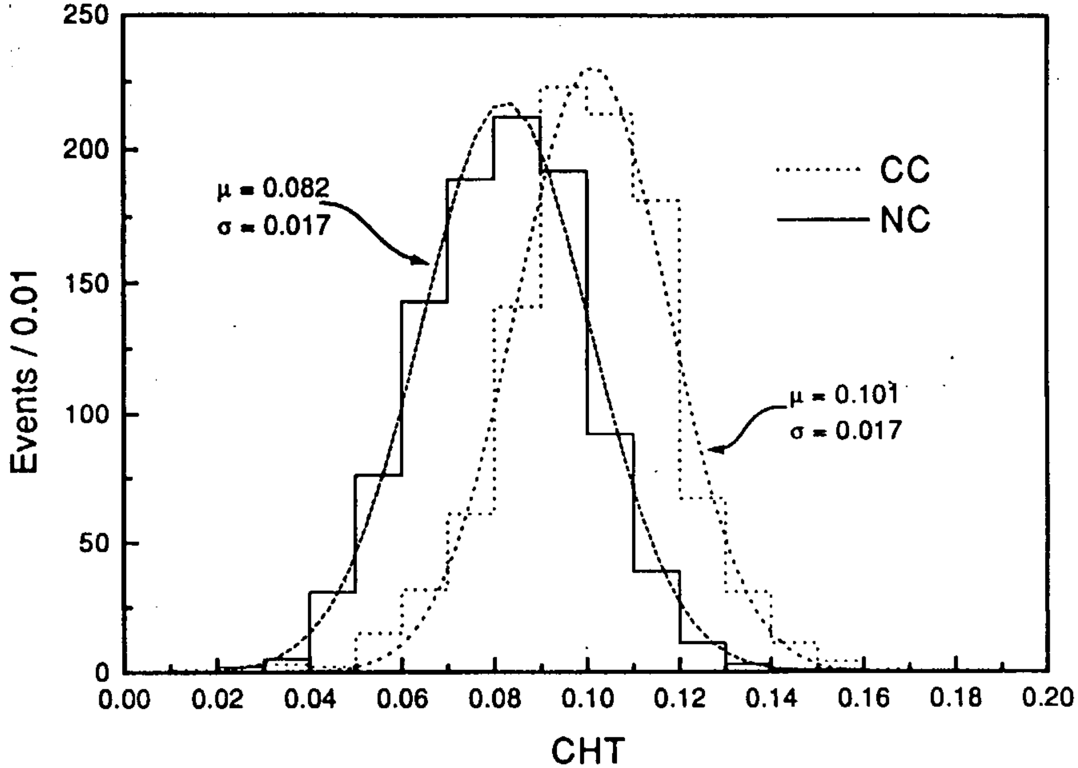


Figure 7: Comparison of CC, NC and Mixture CHT Distributions

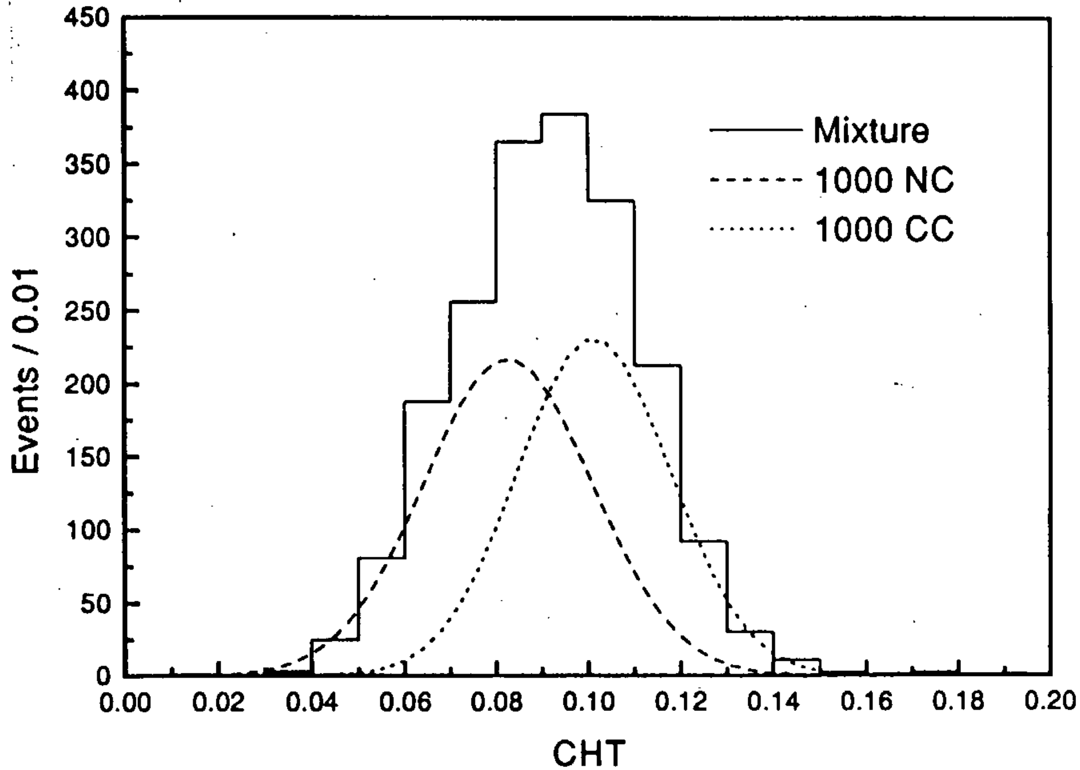


Figure 8: Statistical Extraction of NC Events from Mixture

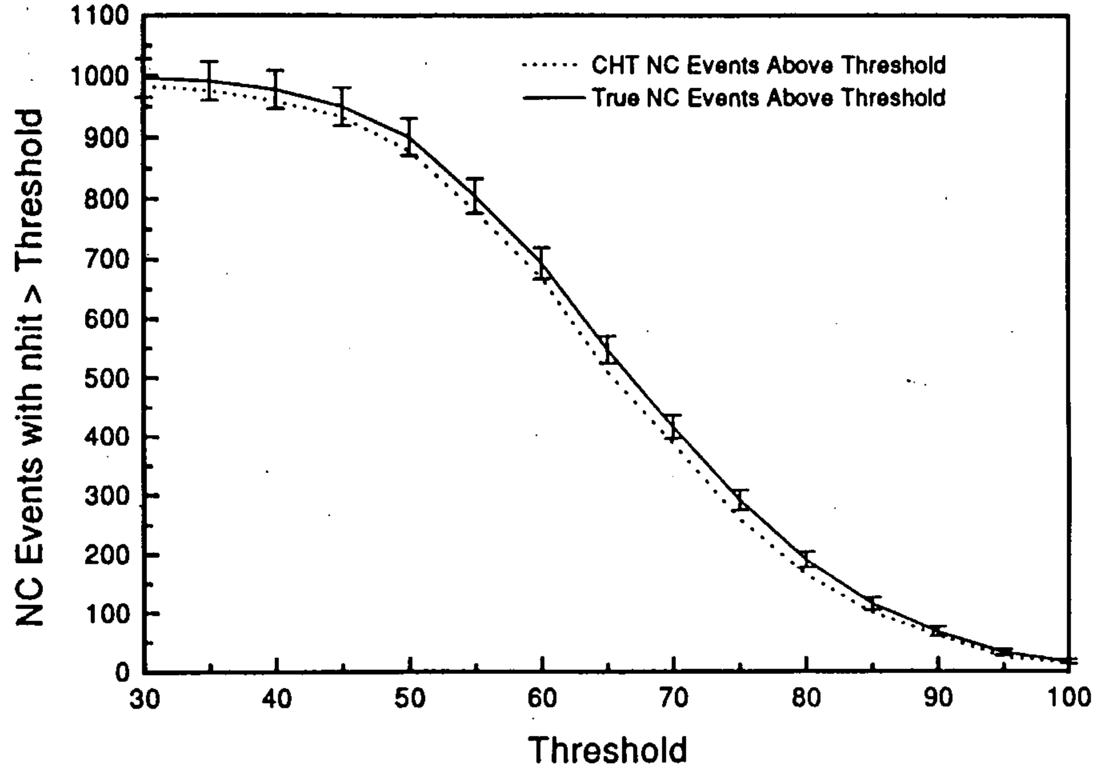


Figure 9: Event-by-event CHT Exctraction of NC Spectrum