NEURAL NET APPROACHES FOR EVENT LOCATION IN THE DETECTOR MODULES

A. Connors[1], H. Aarts[2], K. Bennett[3], A. Deerenberg[2] *, R. Diehl[4], J.W. den Herder[2], W. Hermsen[2], G. Lichti[4], J. Lockwood[1], M. Loomis[1], J. Macri[1], M. McConnell[1], D. Morris[1], K. Reinhard[1], J. Ryan[1], V. Schönfelder[4], G. Simpson[1], B.N. Swanenberg[2], H. Steinle[4], A. Strong[4], M. Varendorff[4], C. de Vries[2], W.E. Webber[1], C. Winkler[3], V. Zeitlmeyer[4]

1. University of New Hampshire, Durham, NH, USA

2. ROL – Laboratory for Space Research, Leiden, The Netherlands

3. Space Science Dept. of ESA, ESTEC, Noordwijk, The Netherlands

4. Max-Planck Institut für Extraterrestrische Physik, Garching, Germany

1 INTRODUCTION

From the description of the Compton telescope given previously (Schönfelder *et al.*, this volume), one can see that the accuracy with which one determines the position of a cosmic gamma-ray source depends not only on the measurements of the energy deposited in the upper (D1) and lower (D2) detectors, but also on how accurately one estimates the (X, Y, Z) positions of each gamma-ray or neutron interaction (an *event*). If nothing were known about the position of each event except in which module it occured, it would increase the uncertainty in the position of a source by on the order of 10°. Within each COMPTEL module, one extracts position information from comparisons of relative intensities of signals in the photomultipier tubes. This technique was introduced in the 1950's for medical imaging by Anger (1958),

* deceased

Data Analysis in Astronomy IV, Edited by V. Di Gesù *et al.*
Plenum Press, New York, 1992

and later was adapted to astrophysical applications (Zych *et al.* 1983; Schönfelder *et al.* 1984; Stacy 1985).

In practice, one cannot obtain exact (X, Y, Z) position information from the measured PMT signals. First, there are processes which "smear" the signal, such as statistical fluctuations within the PMTs, or multiple scatterings and partially absorbed events within the detector modules. Second, as is illustrated in the next section, the function describing event interaction position, (X, Y, Z), as a function of the relative intensities of PMT signals, is sometimes multi–valued. Also, this function is not easily expressed analytically. The problem then becomes to reliably and rapidly calculate a best estimate of a non–linear, non–analytic, vector–valued function. This is an ideal application for a neural net. For COMPTEL, we chose to implement CMAC, a readily available software neural net. This technique gave us roughly a factor of 20 to 50 improvement in the speed of processing events, compared to methods that had been tried earlier (McConnell 1990; R. Freuder 1991).

In the next section, we present an overview of the responses of the D1 and D2 detectors. In the third section, we review function approximation and interpolation as it applies to neural nets. In the final section, we display some results of COMPTEL D1 and D2 event location.

2 CELL RESPONSE OVERVIEW

2.A (X, Y) *RESPONSE*

As part of the overall COMPTEL calibration procedure, a source–plus–collimator was moved across the top of each COMPTEL module, and the signals from each PMT were recorded, for about 10^4 events per position from $\sim 10^3$ (X, Y) positions per module. In Figures 1 and 2, we display contour plots of the D2 and D1 relative PMT responses as a function of (X, Y) position within each module.

If PMT_j is the overall signal from the j^{th} photomultiplier tube, let the PMT ratio R_j be defined by

$$R_j \equiv N \times \frac{PMT_j}{\sum_{k=1}^{N} PMT_k}, \tag{1}$$

where N is the total number of PMTs in the cell (8 for D1 and 7 for D2).

Figure 1.A and 1.B show contour plots of R_1 and R_2, the mean PMT ratios in PMT 1 and PMT 2, calculated from the D2 mapping data. In this figure, PMT 1 is located in the center of the module, and PMT 2 at an angle of $\frac{\pi}{6}$ radians from the +X axis. The D2 R_1 response resembles a constant plus a Gaussian in r, the distance from its center, with width on the order of 5 cm; while the D2 R_2 response is more asymmetric.

Figure 2.A shows a contour plot of R_1, the mean PMT ratio for PMT 1, taken from D1 mapping data. In this mapping data coordinate system, the center of the face of PMT 1 is located at a radius of 13.5 cm from the center of the module, and at an angle of $\frac{5}{8}\pi$ radians measured counterclockwise from the +X axis ($\frac{\pi}{8}$ from the +Y axis). The D1 PMT response can be described roughly as the sum of a Gaussian

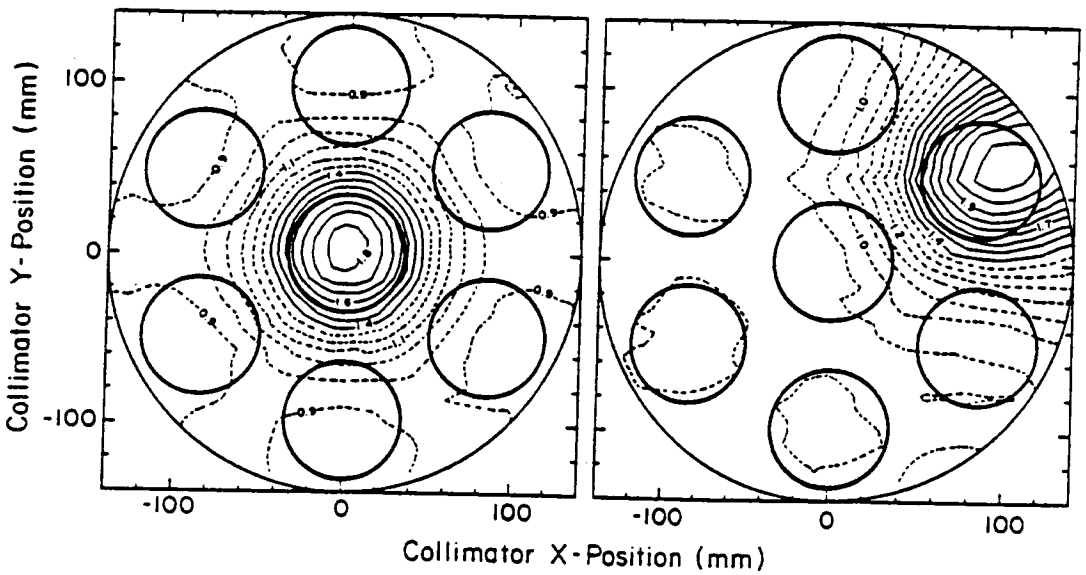

Figure 1.A (left). The relative intensity of the signal received in PMT1 (mean PMT 1 ratio, or R_1); and Figure 1.B (right): the relative intensity of the signal received in PMT2 (mean PMT 2 ratio, or R_2), as a calibration source is moved across a D2 module. Contours range from 0.80 to 1.91 with a contour interval of 0.08. Dashed lines indicate the intensity is less than the median value. Thicker lines indicate the module edge and the photomultiplier tube faces.

of width ~ 3 cm in r, the distance from the center of the PMT, plus an exponential in r with decay rate of ~ 30 cm. In Figure 2.B we display a cross-section of the D1 PMT1 response around $X = 0$. (For this figure the mean ratio for PMT 1, R_1, was calculated from mapping data points within $\pm 7°$ of the Y-axis.)

2.B Z RESPONSE

Although there was good (X, Y) data, no measurements were available on Z, the height of each interaction within a module. To indirectly infer Z, COMPTEL used a principle components analysis (Finger 1987; Varendorff 1987). Several thousand events from each (X, Y) collimator position were used to calculate the covariance matrix of the PMT signals. Then an eigenvector decomposition was performed. For that (X, Y), the direction of the eigenvector with the greatest eigenvalue was designated the λ direction. In many cases this direction of greatest variation was associated with Z, the interaction depth (Varendorff 1987; Loomis 1991). In practice, therefore, COMPTEL event location estimates (X, Y, λ) rather than (X, Y, Z) from the PMT signature.

Since the D1 PMTs are positioned symmetrically on the walls of each D1 module, it is clear there is limited depth information. One expects the PMT response to be symmetric in Z about the middle of the cell. For D2 cells, using a light propogation model originally developed by Zeitlmeir (1988), Loomis (1991) has demonstrated that interaction height is also not a monotonic function of λ and cannot be unambiguously inferred from the PMT signature. This is displayed in Figure 3.

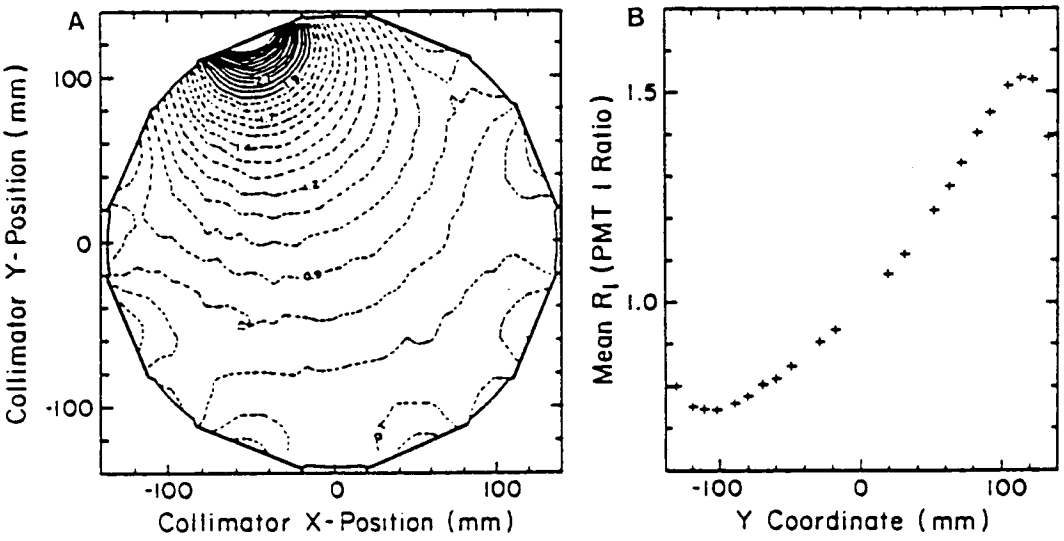

Figure 2.A and 2.B. The relative intensity of the signal received in PMT1 (mean PMT 1 ratio, or R_1) as a calibration source is moved across a D1 module. Thicker lines indicate the module edge and the photomultiplier tube faces. Contours range from 0.56 to 2.86 with a contour interval of 0.08; dashed lines indicate the intensity is less than the median value. The plot on the right, 2.B, is a cross section of the contour plot to the left, taken at $X = 0$.

2.C INCOMPLETE POSITION INFORMATION

In Figure 2.A, notice that the inverse function — the (X, Y) position as a function of the D1 PMT ratios — will be double–valued near the edges of the cell, betwen the phototubes. This is illustrated further in Figure 2.B, the cross–section of Figure 2.A, taken at $X = 0$ mm. In other words, even for the case of infinite signal–to–noise, the inverse functions would be ill–defined, so one has incomplete position information for (X, Y) as well as for Z.

Further, in practice one does not have infinite signal–to–noise. Even when physical effects such as multiple scatterings and escaping photons and electrons are ignored, one expects statistical fluctuations in the PMT signals (Engstrom 1980). At best, one infers a certain probability that an event occurred at (X, Y, λ), given the measured PMT signals. This full probability distribution was deemed too cumbersome for COMPTEL event processing. One therefore summarizes this information with a single position estimator, (X^*, Y^*, λ^*). Typical estimators can be: the most probable position, given the PMT signature; the expectation value, which gives the minimum rms error; the value which maximizes the information entropy, which makes the fewest assumptions about missing information; or some completely different kind of estimator. Recall that for non–Gaussian probability distributions, these can give very different numerical results. A minimum χ^2 search method, such as the one used previously (McConnell 1991; Varendorff 1987), looks for the most probable position; while as the neural net tries to minimize the RMS error, it comes closer to approximating the expectation value of (X, Y, λ). Note that for COMPTEL, since the PMT signature is not a unique indicator of position, one does not expect the distribution of (X^*, Y^*, λ^*), calculated for a large number of events, to trace out the true probability of finding an event at (X, Y, λ), no matter which estimator is chosen.

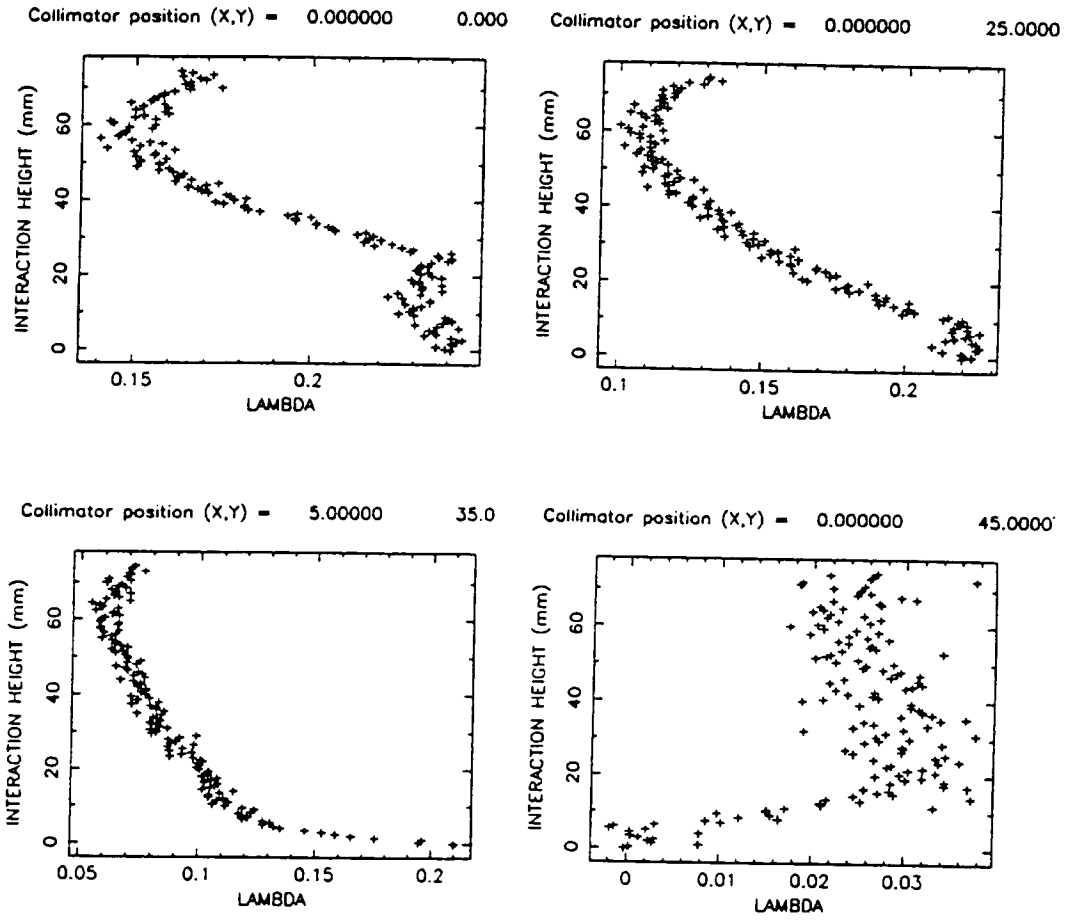

Figure 3. For D2, interaction height as a function of λ is displayed from Monte Carlo simulations for four different collimator positions. From Loomis (1991).

3.A OVERVIEW

Once a position estimator is chosen, to be practical, one needs to find a fast way of approximating (X^*, Y^*, λ^*) as a function of the PMT signals, and of interpolating between the $\sim 10^3$ discrete mapping data points. In general, function approximation and interpolation methods can be classified as either local or non–local. Non–local methods include expansion in eigenfunctions, such as Fourier series, or Legendre polynomials. Sometimes one can cleverly choose a basis function so that the series converges very fast and the approximation is accurate using only a few terms. Local methods can be more flexible, especially when there is no analytic expression for the function of interest (or when an analytic expression would take too long to compute). Local methods include piecewise continuous polynomials, such as cubic splines. For example, most contour plots of unevenly spaced data (such as the contour plots in Figures 1, 2, 4, and 5) are produced using a standard two–dimensional Akima spline algorithm (Clare, Kennison, and Lackman 1987; IMSL 1987; Tennant 1989; and references therein). It requires a look–up table roughly of dimension twice the number of the data points. Another popular local method is to sum Gaussians of varying widths. (See, for example, Bendillini 1991.) This approximation is infinitely continous and differentiable, which is sometimes an advantage. However in some cases it can require storage space for a great many components. An even simpler algorithm, but one that requires even more storage space, is the sum of many constant components. For example, if one were to approximate the one–dimensional function in Figure 2.B by this method, one would divide the Y–axis into discrete bins and associate a weight A_j with the j^{th} bin, so that

$$R_1(Y) \approx \sum_{j=0}^{M-1} A_{S(Y)+j}, \tag{2}$$

with S(Y) the first bin associated with an input value Y; and the number of bins M is chosen by the user to give a convenient overlap between adjacent points. Note that if two inputs, Y_1 and Y_2, are assigned indexes S_1 and S_2 that are H bins apart, the number of overlapping bins is:

$$OVERLAP = \max(0, M - H); \quad H = ||S_1 - S_2|| \tag{3}$$

Once the weights A_j are assigned, this method can be extremely fast.

3.B NEURAL NETS

Of course, all the local methods sketched above can be formulated as Neural Nets. (See, for example, Poggio and Girosi 1990.) Rather than relying on algorithmic complexity, the neural network (or connectionist) philosophy emphasizes using many extremely simple elements, connected together, to represent complicated, non–linear,

functions. They can be particularly useful when one has many examples but no analytic form for a function. The local methods described in the previous section are all classified as single layer localized receptive field networks, with no hidden layer. Equations 2 and 3 describes CMAC, or Cerebellar Model Arithmetic Computer, which is the neural net currently used in COMPTEL event location. It gives a real vector output in response to discrete state (i.e. integer) vector input, from a look–up table (Miller, Glanz, and Kraft 1990; 1987).

CMAC is categorized as having *fixed receptive field centers*. In practice, this means that the mapping from the real–valued PMT ratios to integer look–up table indexes is set ahead of time by the user, independent of the data. For historical reasons these look–up table indexes are called *state vectors* $\vec{S}$ (see Albus 1972, 1975). COMPTEL currently uses a roughly logarithmic function with the spacing between bins set by a *Scale Parameter*. A smaller scale parameter means a smaller bin size, or finer mesh, is used to map the real–valued input into indexes for the look–up table of weights A_j; a larger scale parameter implies a larger bin size, or coarser mesh. For example see Reinhard (1989), Figure 7.1, which uses a scale parameter of .05.

CMAC also uses *fixed receptive field widths*. This means the number of cells summed to form the approximation, M, is also set by the user when the net is made, independent of the data. Since one uses a very large number of cells (typically 50,000), as one is *training* the neural net, one is solving for a large number of weights, A_j. By fixing both the widths and the centers of the receptive fields, one greatly simplifies the optimization procedure. If all three were free to vary, there is no guarantee that any training method would converge.

To reduce the large storage space required one maps the table of weights A_j to a compressed version A'_j. However CMAC retains the convenient property that, even with multi–dimensional input, two inputs $\vec{Y}_1$ and $\vec{Y}_2$ that are close together, having look–up table indexes $\vec{S}_1$ and $\vec{S}_2$, will have an overlap of:

$$OVERLAP = \max(0, M - H); \quad H = ||\vec{S}_1 - \vec{S}_2||. \tag{4}$$

This allows some interpolation between adjacent data points. Note that the amount of overlap depends on both the number of cells to be summed, M, and the mapping from the real inputs $\vec{Y}$ to the integer indexes $\vec{S}$.

With CMAC the weights A'_j are determined by supervised incremental learning. All weights in the net are initialized to zero; then pairs of input and output vectors (termed *exemplars*) are presented to the net. For COMPTEL, the inputs are PMT ratios from the mapping data, and the outputs are the corresponding collimator locations. The weights are adjusted to minimize the square of the distance between the example output vector and its neural net approximation, using a standard Widrow-Hoff (gradient descent) technique.

Let $\vec{S}_i$ represent the ith vector–valued input; $\mathbf{W}$ the array of weights; $\vec{X}_i$ the ith vector–valued output; $\vec{f}(\vec{S}_i, \mathbf{W})$ the estimator of the output; and E_T the total

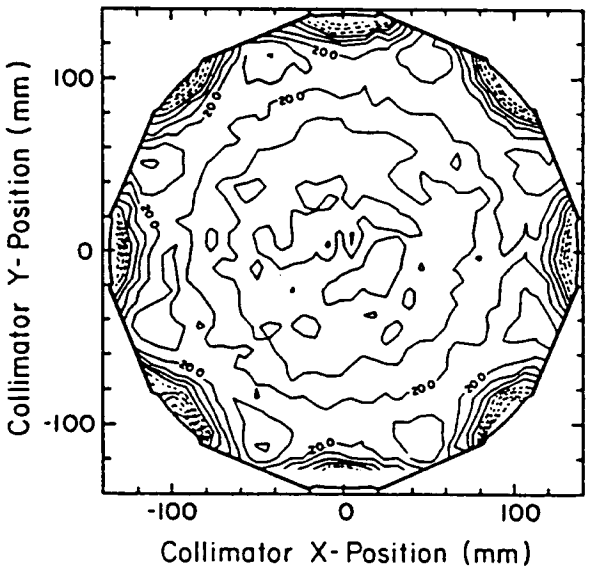

Figure 4. Contour plot of (X, Y) position resolution, in mm, across a D1 module. Contours range from 5 mm to 65 mm with a contour interval of 5 mm. Dashed lines indicate regions with resolution worse than the median. Thicker lines indicate the module edge and the photomultiplier tube faces.

square error. Then

$$E_T = \sum_{i=1}^{M} (\vec{X}_i - \vec{f}(\vec{S}_i, \mathbf{W}))^2. \tag{5}$$

If one were to minimize the total error, one would find an extremum by setting all components of the gradient of the error E_T with respect to the weights $\mathbf{W}$ to zero, and solving the simultaneous equations. In a neural nets context, this is called *batch learning*. This involves inverting a very large, sparse, matrix — a numerically awkward procedure. (COMPTEL currently uses an array of 50000 weights.)

Instead, CMAC uses *incremental learning*. In contrast with batch learning, in which the gradient of the total error is set to zero to find an extremum, in standard Widrow-Hoff incremental learning, one adjusts the weights $\mathbf{W}$ based on the incremental error E_i due to the ith exemplar $(\vec{S}_i, \vec{X}_i)$:

$$E_i = (\vec{X}_i - \vec{f}(\vec{S}_i, \mathbf{W}))^2,$$

and

$$\mathbf{W_{jk}(i)} \equiv \mathbf{W_{jk}(i-1)} - \frac{1}{2}\beta \frac{\partial E_i}{\partial \mathbf{W_{jk}}}. \tag{6}$$

The parameter β is called the learning rate, and $0 < \beta < 1$. For the neural nets used in COMPTEL, β is typically 0.05.

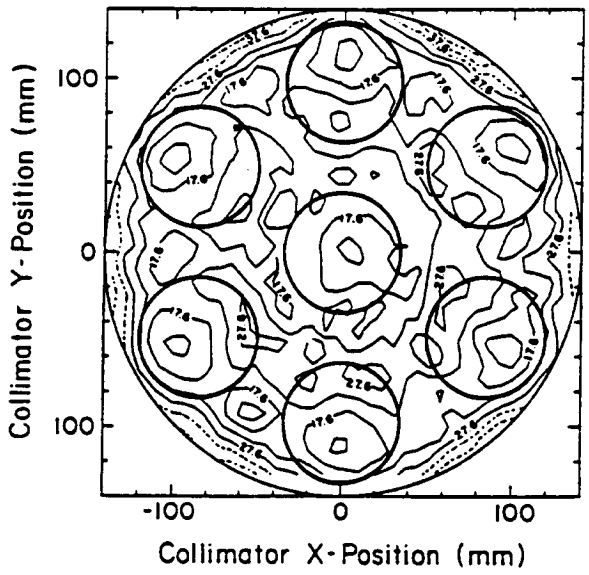

Figure 5. Contour plot of (X, Y) position resolution, in mm, across a D2 module. Contours range from 10.0 to 47.6 with a contour interval of 5. Dashed lines indicate regions with resolution worse than the median. Thicker lines indicate the module edge and the photomultiplier tube faces.

4 RESULTS

Using the neural net method, the root mean square error over the whole module in (X, Y) location was 25 mm for D1 and 23 mm for D2, while the root mean square error in λ estimation was 0.190 for D1 and less than half that, or 0.085, for D2, where λ ranges from -1 to $+1$. In Figures 4 and 5 we display contour plots of the (X, Y) resolution (in mm) across a D1 and a D2 module. Notice that the resolution tends to be best near a photomultiplier tube, and worst around the edges, where the inverse function (i.e. event position as a function of relative PMT signals) is ill–defined. (These two figures were made from mapping data, using neural nets trained on $\sim 10^6$ exemplars, with $M = 50$ cells summmed to form the approximations, and with Scale parameters of 0.05 and 0.02 for D2 and D1, respectively.)

This is also visible in the scatter plots of event locations displayed in Figures 6 and 7. (These two figures were made using data from a 6 MeV CaF source 10 meters from the telescope, at an angle of 10° from the vertical.) Also, there is some slight clustering around the centers of each of the PMTs in both the D1 and D2 modules. This is thought to be due to a tendency of CMAC to broaden sharp peaks (i.e. *overlearning* on regions with steep gradients).

Our studies so far indicate that non-uniformities at this level do not seem to affect the overall COMPTEL angular resolution. In Figure 8, we display a plot of the Angular Resolution Measure (A.R.M.) distribution, for the 6 MeV data described above. (The A.R.M is defined as the angular distance between the inferred photon arrival direction and the true source position.) We show the A.R.M. distribution for

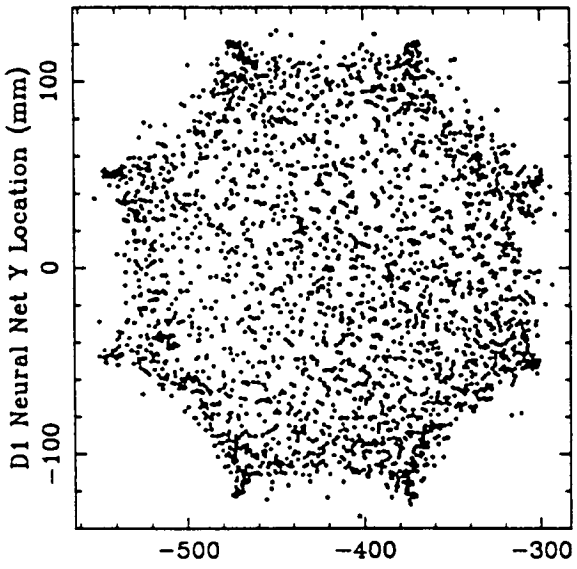

Figure 6. Scatter plot of (X, Y) event locations in a D1 module, from a 6 MeV full telescope run.

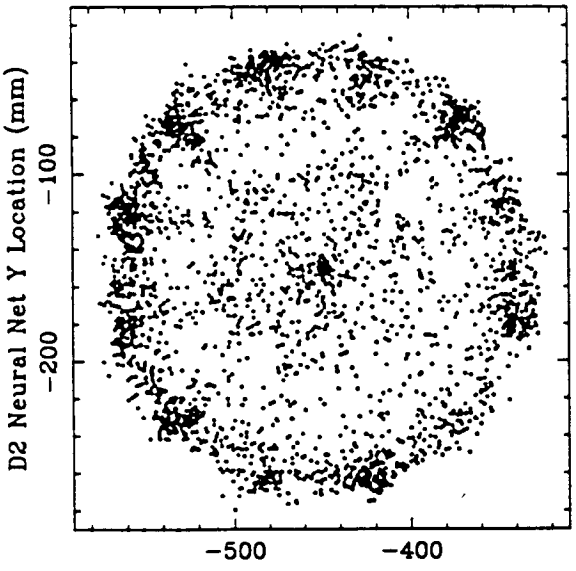

Figure 7. Scatter plot of (X, Y) event locations in a D2 module, from the same 6 MeV full telescope run.

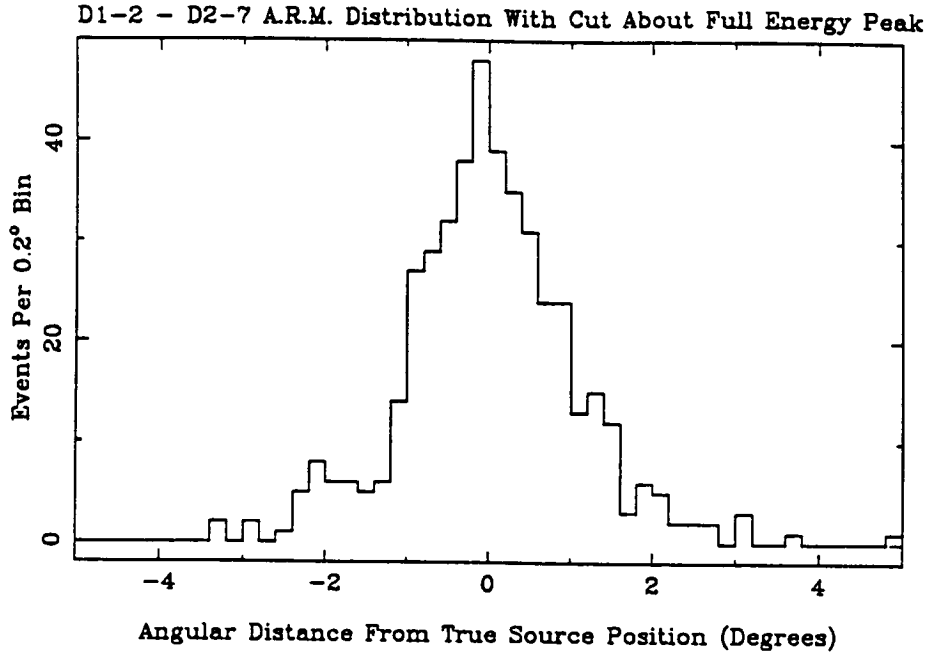

Figure 8. 6 MeV photopeak A.R.M. distribution, in events per 0.2° bin.

events with total energy of 6 ± 0.1MeV. This effectively includes only gamma–rays that have deposited their full energy in the D1 and D2 detectors, and is a reasonable way to observe the effect of just the event location on COMPTEL's angular resolution. The one σ width is about 0.9°.

Advances in electronic computation can change not only the speed with which we calculate numeric answers, but also the ways in which we formulate the questions, and even what kinds of problems are considered interesting to ask. In the use of "connectionist" techniques such as neural nets, or in Bayesian and Entropic analyses of the information content of data, are we beginning to see qualitative changes in how we approach astrophysical data analysis? The coming years may be very interesting.

5 ACKNOWLEDGEMENTS

Much of the summary of neural nets relies on notes from *Advanced Topics in Systems Engineering: Artificial Neural Networks*, as taught by M. J. Carter at the University of New Hampshire.

6 REFERENCES

Albus, J. S., 1972, Ph.D. Thesis, University of Maryland.

Albus, J. S., 1975, *J. Dyn. Sys., Meas., Contr.*, **97**, 220.

Anger, H. O., 1958, *Rev. Sci. Instr.*,, **29**, 27.

Bendellini, O., 1991 *Ap. J.*, **366**, 599.

Clare, F., Kennison, D., and Lackman, R., 1987, *NCAR Graphics User's Guide*, (National Center for Atmospheric Research: Boulder, Co).

Engstrom, R. W., 1980, *RCA Photomultiplier Handbook*, (RCA Corporation Electro Optics and Devices: Lancaster, PA).

Finger, M., 1987, Ph.D. Thesis, California Institute of Technology.

IMSL 1987, *User's Manual: Math/Library Fortran Subroutines for Mathematical Applications.*

Freuder, R., 1991, private communication.

Lampton, M., Margon, B.,and Bowyer, S., 1976, *Astrophysical Journal,* **208,** 177.

Loomis, M., 1991, U.R.O.P. Report, University of New Hampshire.

McConnell, M., 1990, *Comparisons of Neural Net vs Conventional Location Techniques,* COM-TN-UNH-MMG-074.

Miller, W. Thomas, Glanz, Filson H., and Kraft, L. Gordon, 1987, *International Journal of Robotics Research,* **Vol. 6, No. 2,** 84.

Miller, W. Thomas, Glanz, Filson H., and Kraft, L. Gordon, 1990, *Proc. of the IEEE,* **Vol. 78, No. 10,** 1561.

Poggio, T., and Girosi, F., 1990, *Proc. of the IEEE,* **Vol. 78, No. 9,** 1481.

Schönfelder, V., it et al., 1984, *ISEE Trans. Nucl. Sci.,* **NS-31,** 766.

Shafer, R. A., Haberl, F., Arnaud, K. A., 1990, *XSPEC: An X-Ray Spectral Fitting Package User's Guide,* (EXOSAT Observatory: European Space Agency, Noordwijk).

Stacy, J. G., 1985, Ph.D. Thesis, University of Maryland.

Tennant, A. F., 1989 *The QDP/PLT User's Guide,* (Space Science Laboratory: Marshall Space Flight Center).

Reinhard, Kent, 1989, M.S. Thesis, University of New Hampshire.

Varendorff, Martin, 1987, Diplomarbeit, MPE.

Zeitlmeir, Valentine, 1988, Diplomarbeit, Augsburg.

Zych, A. D., Tumer, O. T., and Dayton, B., 1983, *IEEE Trans. Nucl. Sci.,* **NS-30,** 383.