Proceedings of ICRC 2001: 2935 © Copernicus Gesellschaft 2001

ICRC 2001

Fast pattern recognition trigger for atmospheric Cherenkov telescopes

A. A. Vardanyan[1], S. A. Chilingarian[1], W. Eppler[2], and H. Gemmeke[2]

[1]Cosmic Ray Division, Yerevan Physics Institute, Alikhanyan Brothers st. 2,Yerevan 375036, Armenia
[2]Hauptabteilung Processdatenverarbeitung und Electronik, Forschungszentrum Karlsruhe, 76021 Karlsruhe, Germany

Abstract. The ambitions to bridge the energy gap between ground based and satellite borne detectors requires to decrease the threshold of Cherenkov telescopes down to several tens of GeV. The images corresponding to such low energies and registered with high angular resolution will lead to rather complicated disconnected images. The standard second-momentum analysis will not be so effective as for images detected with less angular resolution and/or more compact mirrors and high incident energies above 300 GeV. Since the trigger rate at low thresholds can reach 1 MHz, the main tasks for an "intelligent" trigger are signal pattern recognition and background rejection. We propose to use the hardware neurochip SAND/1 (Simple Applicable Neural Device) as fast "intelligent" Pattern Recognition Trigger (PRT). In addition to decrease the registered event rate down to several kHz, the PRT will reject muon and hadron backgrounds online at present only possible off-line. Using a special board of hardware neural accelerators and evolutionary network training strategies we construct a PRT which meets both timing and pattern recognition requirements.

1 Introduction

The Atmospheric Cherenkov Telescopes (ACT) register the light images of Extensive Air Showers (EAS) during moonless dark nights. The goal is to detect the γ-ray "point sources", usually very distant Active Galactic Nucleus emitted intensive fluxes of TeV gammas.

The new generation of ACTs have huge light receivers, formed of up to 500 sensitive elements (pixels). The useful "signal" events (γ initiated EAS), are contaminated by different background images. The sources of fake signals are: the night sky light, local muons, EAS initiated by hadrons (the latter became essential because of very low threshold of new telescopes, which equals to few tenths of GeV).

Therefore the background rejection problem will require "intelligent" on-line trigger and sophisticated off-line image builder and recognizer.

The first level trigger rate for the second generation ACT's like MAGIC Telescope (Martinez M. et al., 1999) at low thresholds ($\sim$ 4 photoelectrons) can reach 1 MHz for each channel. This huge amount of data has to be reduced down to a few kHz by the second and third level triggers. For more than 100-fold reduction we investigate the possibility to use the MiND PCI board (with 4 SAND/1 chips installed and processing in parallel)(Fischer T., et al., 1996) as a fast "intelligent" trigger.

In order to be able to implement such a specialized device for γ-astronomy it is necessary to design combined software-hardware system for net training, to find the optimal method of generalization from examples. This involves specialization of network architecture, objective function and learning strategy.

We can't expect that the common techniques of pedestal extraction and image reduction to second order moments will provide the necessary level of the background rejection. The use of all distinctive information contained in the pixels will provide the possibility to enlarge signal-to-noise ratio. Working with such huge inputs (as compared with previous analysis of Whipple telescope data, when only 4-5 Hillas parameters were used (Chilingaryan A. A., 1994, 1995)), requires adequate network training algorithms and powerful training accelerators.

In this paper we'll demonstrate that the MiND board meets well tight timing requirements.

2 The SAND/1 chip and MiND PCI board

The SAND/1 chip designed for accelerating various neural net applications is a digital hardware realization (built with $0.8\mu m$ CMOS technology) of NN models based upon the principle of systolic array.

For the stand alone operation it requires only a few external components, such as:

Correspondence to: A. A. Vardanyan (aro@crdlx5.yerphi.am)

- Look Up Table LUT - for non-linear transfer function calculation
- Memory (WRAM) - for storing NN weights and intermediate data,
- Sequencer - for the overall memory management as well as the control of SAND/1 itself.

Table 1. *The principal features of SAND/1*

Maximum Operation Clock - 40MHz
Calculation of scalar product and vector distance
Extreme value search (minimum and maximum)
CUT-function with over/underflow recognition
On-line adaptation of arithmetic precision
Activation function as look-up table external to chip
Parallel processing support
The following neural networks are supported:
– Multilayered Perceptron (MLP)
– Radial Basis Function (RBF)
– Self Organizing Maps (SOM)
Cascadable architecture:
– 16 bit weights and input activities
– 40 bit internal precision
– Processing of packets consisting of 4 data words
– Max. 65K weights for any configuration of NN
Data I/O:
– Input activities normalized to the range -1.0 ... +1.0
– 8 fixed-point formats available for the weights (0.25 ... 128)
– 2 scalable output formats: linear output or any transfer function
– Continuous data flow on the weight and activity busses (max. 100Mb/s)

At a maximum clock frequency the single SAND/1 chip achieves a performance of 200 MCPS (Mega Connections Per Second).

To take the advantage of such a specialized neuro-chip and to use it in real-time applications such as Pattern Recognition, Image Processing, Control Engineering, Fast Intelligent Triggers for High Energy Physics Experiments, the MiND (Multipurpose integrated Neural Device) was designed to integrate the SAND/1 on a PCI bus based device to be compatible with todays advanced architectures and to provide an easy to use design via SAND/1 to PC connection and hardware control via soft routines. Another great advantage of MiND board is that it integrates four SAND/1 chips connected in parallel and attains a further increase of performance up to 800 MCPS.

3 Neural Net Technique for Background Rejection

The inverse problem to be solved can be represented as follows:

$$\begin{array}{ccc} \textit{Simulated data} & & \textit{Experimental data} \\ A_{s/b}(N_j) & \Longrightarrow & ?(N_j) \end{array}$$

where the $A_{s/b}$ is the particle type (A_s - γ(”signal”), A_b - $hadron$(”background”)) inducing the extensive air shower in the atmosphere, N_j is the image registered by Photomultiplayer (PM) matrix. This is a data classification problem and for its solution we propose to use a Neural Network classifier.

Neural Networks (NN) represent very simple structures composed of processing elements (nodes) and connections (weights). NN belongs to the general class of non-parametric methods that do not require any assumption about the parametric form of the statistical model they use. The use of NN classifier is justified by the following reasons:

- NN is non-parametric technique appropriate for classification and background rejection problems
- NN is able to treat multidimensional input data (e.g.400 pixel information).
- all distinctive information contained in the pixels will provide a possibility to achieve higher level of background rejection as compared to the common technique of image reduction to second order moments
- NN is very fast in application phase (possibility of online data analysis)

The central issue of Neural Networks is the bounded mapping of n-dimensional input to the m-dimensional(for classification or background rejection tasks one dimensional class assignments) output:

$$f_{\boldsymbol{W}}(N_j) \rightarrow \ A^*_{s/b}$$

The functional form of f is tuned by iteratively changing $\boldsymbol{W}$ - NN parameters(weights) during NN training process, $A^*_{s/b}$ is the decision on primary particle type made by trained NN.

- The NN training consists in iterative processing of simulated examples
- The goal of the NN training is to find $\boldsymbol{W}$ that provides the minimum of the Error(Quality) Function:

$$Q(\boldsymbol{W}, \boldsymbol{N_j}) = \frac{1}{M_{ev.}} \sum_{m=1}^{M_{ev.}} (OUT_m(N_j) - TRUE_m)^2 * V_m;$$

where $M_{ev.}$ is number of simulated events; V_m- event weight, OUT_m is the actual output of NN; $TRUE_m$ is the corresponding goal value of m-th input vector from training set and V_m is the event weight.

For classification purposes this mapping takes a special form with aim to ”shift” different classes of training samples from each other as much as possible. For different classes we use one output variable. Therefore the ”goal” output

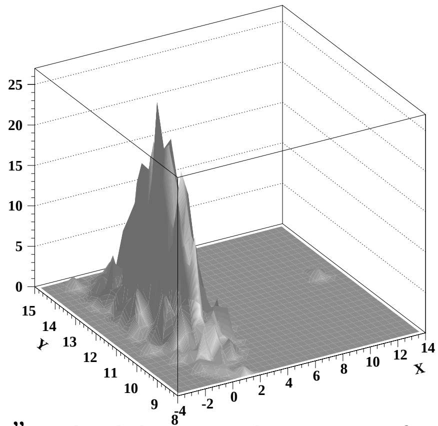

Fig. 1. *"γ Induced Shower" Incident on Matrix of 20x20*

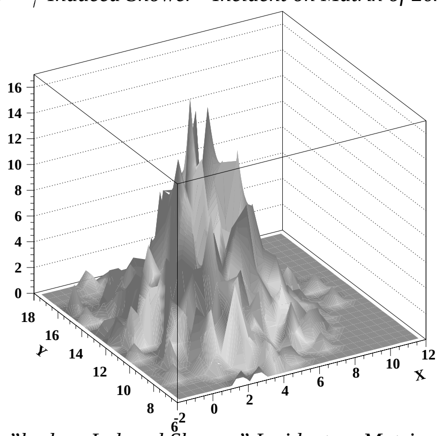

Fig. 2. *"hadron Induced Shower" Incident on Matrix of 20x20*

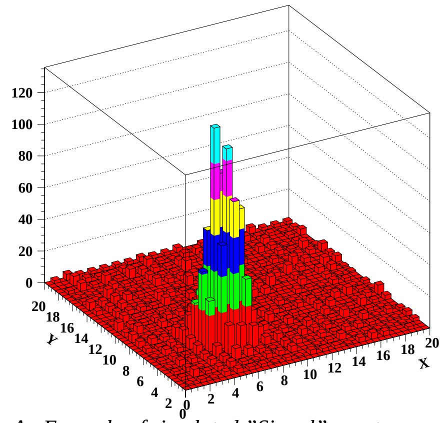

Fig. 3. *An Example of simulated "Signal" event*

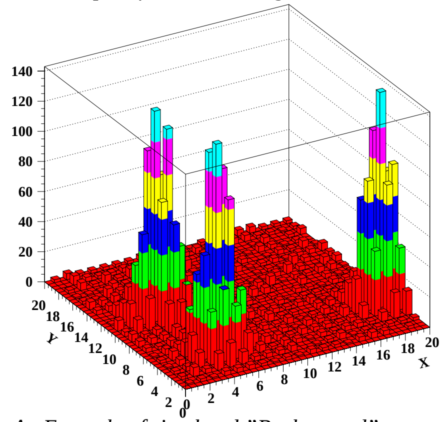

Fig. 4. *An Example of simulated "Background" event*

$TRUE(k)$ for events of the k-th category could be chosen as: TRUE(k) $= \frac{k-1}{K-1}, \quad k = 1, K$, where K is total number of classes.

In the case of two classes, i.e. signal and background events, the "goal" outputs, as one can easily see, are equal to 0 and 1. The actual events classification is performed by comparing the obtained output value with the "goal" one. We expect, that the data flow passing through the trained net will be divided in two clusters concentrated in the opposite regions of the $(0, 1)$ interval. Choosing an appropriate point in this interval (*the so-called decision point c*), controls the relation between two kinds of misclassification committed by decision rule), the classification procedure can be defined as follows:

$$\text{OUT}(N_j) \left\{ \begin{array}{ll} < c & , N_j \text{ is classified as signal,} \\ \geq c & , N_j \text{ is classified as background,} \end{array} \right\}$$

where OUT(N_j) is the output node response for a particular experimental measurement N_j.

4 Simulation Model

For the test of our method we construct very simple signal (showers initiated by the γ-quanta) and background (showers initiated by $hadrons$ and "Night Sky Background") model.

The camera consist of 400 pixels (20x20) arranged in square. The images are generated by means of Gaussian distributions.

Briefly we use the following specifications for our test computation experiment:

– *"γ induced showers" (signal)*: random numbers from 2-dimensional Gaussian population with σ=1 (Fig. 1).
– *"hadron induced showers" (background)*: random numbers from 3 different 2-dimensional Gaussian populations with σ=2 (Fig. 2).
– *"Night Sky Background"*: random numbers uniformly distributed on 20x20 matrix.
– 1000 random points for each Gaussian were generated, 1000 events per class were simulated.
– number of overall points (hits) from all distributions in each pixel was calculated (Figures 3 and 4).
– information from all 400 pixels was used for the NN input.

For the "background" images the variance of the Gaussian was taken larger than for the "signal" images, mimicking longer and broader hadron shower Cherenkov images (see figures 1, 2, 3, 4,). So, only shape information of images was used for the discrimination. 1000 simulated events for "signal" and "background" were generated and used for net training.

5 Results and Conclusion

Using such simplified simulation and each pixel information as NN input we demonstrate that SAND/1 gives reasonable background rejection. Data processing is fast enough for on-line implementation.

The table 2 demonstrates that the MiND board is capable to process the single event with the NN configuration of 400x64x1 approximately in 1 milisecond, which means that the MiND board is able to treat with trigger rates up to 1kHz.

Table 2. *Processing time for single event, with net configuration: 400x64x1, for SAND chip and Pentium II 450MHz CPU*

SAND/1	Pentim II 450MHz
0.97 milliseconds	8.8 milliseconds

Even using fastest modern serial CPU it is very hard to train very big networks for recognition of the pattern of extensive air showers initiated by $\gamma-$quanta or $hadrons$ of primary cosmic ray flux. To train such a huge networks with big amount of MC data to have a sufficient generalization capabilities, days and ever weeks of calculations will be required. So, the performance is neccessary not only for on-line implementation, but for acceleration of the training as well.

The special mode of ANI (Chilingarian A. A., 1998) package was developed for substitution of the Neural Network artificial simulator by the hardware realization. Incorporation of such a specialized device in the learning algorithm, gives a significant benefit in time of large amount of events processing. Taking into account that the learning is an iterative process (sometimes hundred thousands of iterations are required in order to achieve the required level of generalization), the advantage of combined software-hardware implementation of learning is obvious.

The figure 6 demonstrates the efficiency of using such a specialized device. In this figure the single event processing time ratio is plotted for Pentium II CPU at 450 MHz and SAND/1 chip. As one can see the SAND/1 chip processes 5 to 30 times faster depending on the total number of NN weights (reaching the peak performance at most important case of very large NN).

To achieve the "signal" and "background discrimination level as shown on figure 5 ($\sim$ 95% of correct classifications demonstrates the adequateness of used net training methods for such huge amount of network weights.) 108000 iterations were required to train the network. Using MiND board 8 hours were spent to perform the training, the Pentium II 450MHz CPU spent more then 5 days to do the job. Of course, more realistic simulations are necessary for definite conclusions and performance estimates.

Acknowledgements. We would like to thank A. Chilingarian for providing ANI program package and NN training algorithms optimized for background rejection tasks in γ-ray astrophysics experiments. We thank E. Lorenz, R. Mirzoyan and all participants of the workshop "MAGIC Triggers and Associated Problems" for their interest in reported results and useful discussions. This work has been partly supported by NATO grant NIG-975436 and by research grant N 00-15 of the Armenian government.

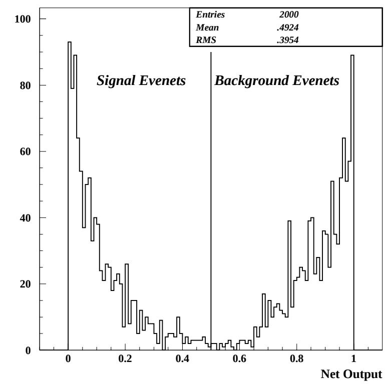

Fig. 5. *NN output distribution(left - "signal" events, right - "background"*

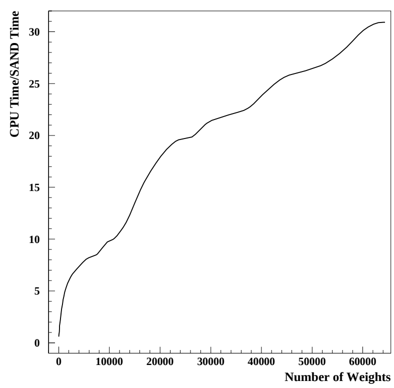

Fig. 6. *CPU Time/SAND Time ratio for single event processing for different number of NN weights*

References

Chilingaryan A. A. Neural classification technique for background rejection in high energy physics experiments. *Neorocomputing*, **6**:497, 1994.

Chilingaryan A. A. Detection of weak signals against background using neural network classifiers. *Pattern Recognition Letters*, **16**:333, 1995.

Chilingarian A. A. Analysis and Nonparametric Inference in High Energy Physics and Astroparticle Physics, 1998. Program Package ANI, (Users Manual, unpublished) http://crdlx5.yerphi.am/proj/ani.

Fischer T., Eppler W., Gemmeke H., Kock G., Becher T. The SAND Neurochip and its Embedding in the MiND System. (Techreport).

Fischer T., Eppler W., Gemmeke H., Menchikov A. Novel Digital Neural Hardware For Trigger Applications IN Particle Physics. In *Proc. of* 2^{nd} *Workshop on Electronics for LHC Experiments*. Balatonfured, Hungary, 1996.

Martinez M. for the MAGIC Collaboration. The MAGIC Telescope Project. In *Proc. 26th ICRC*, vol. 5, p. 219. Salt Lake City, Utah, 1999.