Neural Chip SAND in online data processing of extensive air showers

W. Eppler [a,1], T. Fischer [a], H. Gemmeke [a], A. Chilingarian [b,2], A. Vardanyan [b]

[a] *Forschungszentrum Karlsruhe, Germany*
[b] *Yerevan Physics Institute, Armenia*

Abstract

The neural chip SAND (Simple Applicable Neural Device) was designed to accelerate computations of neural networks at a very low cost basis, due to the fact that only few peripheral chips are necessary to use the neural network chip in applications. Four SAND-chips were implemented on one PCI-board. The board is highly usable for hardware triggers in particle physics. The performance of a SAND-PCI-board is 800 Mega Connections per Second due to four neuro-chips, each with four parallel 16 bit multipliers and 40 bit adders. SAND is able to implement feedforward neural networks with a maximum of 512 input neurons and three hidden layers. Kohonen feature maps and radial basis function networks may be also calculated. The application of the SAND-PCI-board is proposed for cosmic ray physics to allow online analysis of extensive air showers. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Neural network; Trigger; Neural chip; PCI board; Cosmic ray physics

1. Introduction

The Research Center Karlsruhe (FZK), in collaboration with the Institute of Microelectronics, Stuttgart (IMS), developed the neuro-chip SAND (Simple Applicable Neural Device) for on- and off-line data analysis as well as first and second level triggers in astrophysics experiments (KASCADE, ANI, MAGIC, AUGER). Many sophisticated methods were proposed and implemented during the last decade to reveal the characteristic of extensive air showers. But one drawback is still present – the absence of an 'intelligent' adaptive hardware trigger for on-line data analysis. Usually it is a multiplicity or sum-energy trigger with very simple logic, requiring some channels exceed-

ing the chosen threshold value. Now, detailed simulations of an Extensive Air Shower (EAS) developing in the atmosphere and the response of the apparatus will be used to train an Artificial Neural Network. This allows one to implement sophisticated pattern recognition tasks for first level trigger and event builders in modern EAS experiments, like ANI in Armenia or KASCADE in Karlsruhe, Germany, measuring as many parameters of a single event as possible. Information from thousands of electronic channels has to be processed in a very short time. The fast primary energy and primary particle type estimators will be trained by simulations and implemented for on-line analysis.

The expected trigger rate of extensive air shower experiments can be decreased considerably by using neural network hardware triggers. The computing power of SAND is sufficient for the pattern recog-

nition of Cherenkov images as the expected decision time did not exceed few tenths of microseconds. The training of the neural network is carried out with detailed simulations of Extensive Air Shower (EAS) traversing atmosphere and with the response of apparatus (see [1], but also notice [2]).

2. ANI-extensive air shower experiment

The ANI-experiment (Yerevan, Armenia) with its large detector array at mountain altitude measures the longitudinal structure of EAS by observing the arrival time distribution of muons and measures the shower soft component density. The arrangement of detectors is shown in Fig. 1.

The hardware trigger accepts an event if at least in 7 of the 11 trigger detectors (0) the energy deposit is equal to 4 particles. The software trigger condition is given if the 1st ring of software triggers (1) detects more particles than the second one (2). For each event several variables are determined. Three of them are used as input for the analyzing neural net-

work: the number of electrons, the limited number of muons, and the shower age with moliere radius of 30m (temporal distance of detected particles on a circle with 30 m radius). Further variables were rejected by discriminant analysis as they were strongly correlated with the used variables. The training of the network was carried out with detailed Monte Carlo simulations. The multi-layer neural network has two hidden layers with 7 and 5 neurons, respectively. After training, experiments with test data showed the advantages of non-linear neural classifiers (13.3% misclassifications) compared to linear (15.8% missclassifications) or quadratic classifiers (15.2% missclassifications). The neural network approach therefore seems to be good for this experiment. An additional reason for applying this methods is the inherent massive parallelism of neural networks. This means with the use of specialized hardware they may be performed very efficiently. One feedforward pass of the used network with SAND takes 183 ns.

Therefore, for the online event processing in the following the neural hardware accelerator chip SAND is proposed.
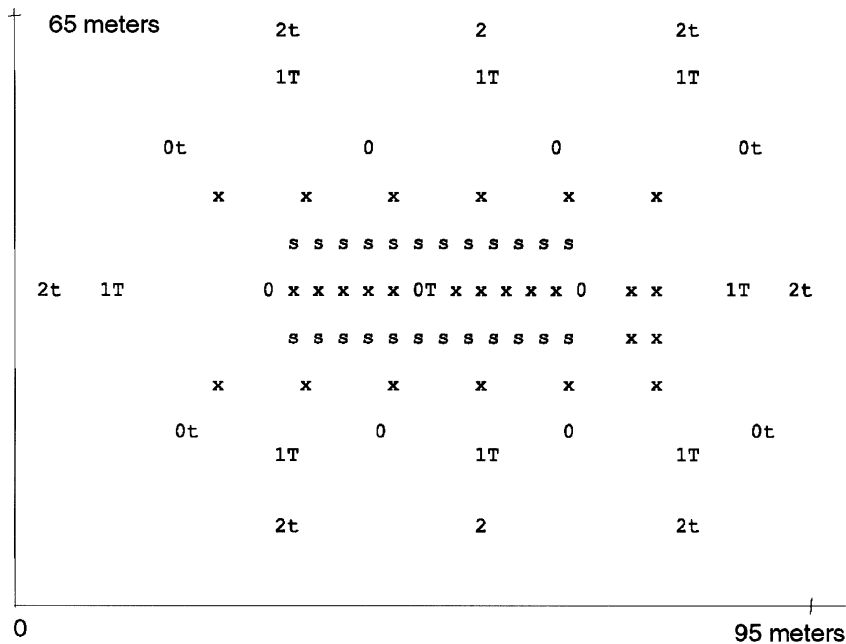

Fig. 1. Arrangement of detectors in the ANI-experiment (not all detectors are shown). Following shortcuts are used: s: detectors with small surface; 0: trigger detectors; t,T: detectors with timing channel; T: timing trigger detectors; 1: 1st ring of software trigger; 2: 2nd ring of software trigger.

3. Performance of SAND

The neural chip SAND (Simple Applicable Neural Device) was designed to accelerate computations of neural networks at a very low cost basis [3]. Only few peripheral chips are necessary to use the neural network chip in applications. On a single neuro-chip four parallel data processing units are implemented. The performance is 200 MCPS with a 50 MHz clock. This leads to a maximum of 600 MOPS when using radial basis function networks. A PCI board works with four SAND chips gaining a peak performance of 2400 MOPS. The chip was designed to be applicable stand-alone, i.e. without additional host. This implied that the neural transfer function was not implemented with an additional microprocessor but with a freely programmable lookup table. Another feature of the chip is the processing of several input patterns in one block. With this the processing performance of the chip could be brought into line with the transfer rate of weights and activities. Each of the parallel working units has a 16 bit adder, a 16 bit multiplier and a 40 bit accumulator. A tricky cut mechanism provides a nearly optimal reduction of the internal data size to the external 16 bit words.

If all neurons of one layer of a feedforward network are regarded, the function of the complete layer can be described as a matrix/vector multiplication. To increase the calculation speed of a neural network, neurons have to work in parallel. On the other hand a high flexibility concerning the structure of neural networks should be ensured. To grant both demands, only neurons within the same layer are processed in parallel, whereas the various layers are processed sequentially. The matrix/vector multiplication is replaced by a matrix/matrix-multiplication by processing four input events in a block. This measure saves three input buses while the processing units work with the same performance.

There are especially three neuro-chips available fulfilling partly similar requirements as SAND: the MA16 of Siemens, CNAPS of Adaptive Solutions and ETANN of Intel. SYNAPSE is a neuro-computer with one MA16. For industrial applications a stand-alone solution without host computer requires many other chips and at least one micro-controller. CNAPS and ETANN have problems with their low precision. CNAPS is working with 8 bit accuracy, or 16 bit with less than half the rate, the analog ETANN computes with approximately 6 bit accuracy. Sometimes poor accuracy may be compensated by non-linear data transformations. But for on-chip training of the neural network a minimal data length of 16 bits seems to be necessary to find the global optimum (but see special example in [4,5] where 4 bits are shown to be accurate using tabu search). There are two well known applications of digital neural network processors in second level triggers: CNAPS [6] in H1 [7,8] and MA16 [9] in WA92 [10]. The neural processor module based on SAND demonstrates a throughput similar to the CNAPS-board (computation time of CNAPS: 8 s, SAND: 5.1 s) and successfully competes with it when the data acquisition system is equipped with an event buffer. Moreover, the module allows processing of higher accuracy input activities (CNAPS: 8 bit, SAND: 16 bit). The SAND processor module shows higher throughput than the trigger module based on MA-16 (computation time of MA16: 5.5 s, SAND: 0.5 s, latency time of MA16: 8 s, SAND: 3.6 s) due to the simultaneous processing of four events and the higher clock frequency of SAND board.

4. Conclusion

SAND performs feedforward networks, Kohonen feature maps and radial basis functions with a very high speed. The central processing unit of the chip was designed in a way that only few additional devices are required compared to previous designs. Therefore it is simply applicable in scientific and industrial applications. Future developments of general purpose micro-processors like Pentium PII from INTEL, K6 from AMD, M2 from Cyrix and others have to be regarded carefully. Their MMX-instructions and the use of parallel integer units on chip enable these devices to very fast matrix multiplications. Up to now the independent parallel transfer of data is a problem so that they cannot compete with the performance of SAND. Because of many restrictions the internal pipeline organization is not appropriate for the fast computation of neural networks. Other processors (e.g., MIPS with MDMX-instructions) or the digital signal processor C6x from TI aim to the same direction but show similar problems. In near future this might change. The SAND chip is produced by the Institute

of Microelectronics, Stuttgart, the PCI-board with four SAND chips by datafactory, Leipzig [11]. Faster versions using full-custom design and supporting a fast hardware learning features are under development at FZK.

References

[1] E. Lorenz, Nucl. Phys. B (Proc. Suppl.) 48 (1996) 494.

[2] S. Omner, S. Westerhoff, H. Meyer, HEGRA collaboration: Search for TeV gamma-rays from extragalactic point sources with neural photon/hadron separation, Nucl. Instr. and Meth. A 389 (1997) 204.

[3] T. Fischer, H. Gemmeke, W. Eppler, A. Menchikov, S. Neusser, Novel digital hardware for trigger applications in particle physics, in: Proc. 2nd Workshop on Electronics for LHC Experiments, Balatonfred, Hungary, 1996.

[4] G. Anzellotti et al., TOTEM: a highly parallel chip for triggering applications with inductive learning based on the reactive tabu search, Int. J. Mod. Phys. C 6 (1995) 555.

[5] S. Dusini et al., The neurochip TOTEM: a case study in HEP, ICHEP97.

[6] Adaptive Solutions, CNAPS product, Information (1995).

[7] D. Goldner et al., Proceedings of 4th International Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, Pisa, Italy, April 3–8, 1995, pp. 333–340.

[8] J.K. Kohne et al., Realization of a second level neural network trigger for the H1 experiment at HERA, Nucl. Instr. and Meth. A 389 (1997) 128.

[9] U. Ramacher et al., Design of a first generation neurocomputer, in: VLSI-Design of Neural Networks, U. Ramacher, U. Rckert (Eds.) (Kluwer Academic Publishers, 1991).

[10] C. Baldanza et al., Nucl. Instr. and Meth. A 376 (1996) 411; Nucl. Instr. and Meth. A 373 (1996) 261.

[11] T. Becher, W. Eppler, T. Fischer, H. Gemmeke, G. Kock, The MiND-project: building, applying and speeding-up neural networks using the SAND-neuroprocessor, in: Proc. of EUFIT '97, H.J. Zimmermann (Ed.) Aachen, Germany.