DFUB 95/2

RESULTS FROM A NEURAL TRIGGER BASED ON THE MA16 MICROPROCESSOR

C. BALDANZA[b], J. BEICHTER[a], F. BISI[b], N. BRUELS[a], C. BRUSCHINI[c], A. COTTA-RAMUSINO[b], I. D'ANTONE[b], L. MALFERRARI[b], P. MAZZANTI[b], P. MUSICO[c], P. NOVELLI[c], F. ODORICI[b], R. ODORICO[b]*, M. PASSASEO[d], M. ZUFFA[b]

a) SIEMENS AG, ZFE T ME2, 81730 Munich, Germany
b) INFN ANNETTHE, Via Irnerio 46, 40126 Bologna, Italy
c) INFN/Genoa, Via Dodecaneso 33, 16146 Genoa, Italy
d) CERN, 1211 Geneva 23, Switzerland

Received (received date)
Revised (revised date)

Results from a neural-network trigger based on the digital MA16 chip of Siemens are reported. The neural trigger has been applied to data from the WA92 experiment, looking for beauty particles, which have been collected during a run in which a neural trigger module based on Intel's analog neural chip ETANN operated, as already reported. The MA16 has a precision of 16 bits for input variables, 16 bits for weights, 16 bits for scalar multipliers modifying the transfer function shape, 47 bits for thresholds, 53 bits for internal calculations, 38 bits for the output. Of the latter only 16 bits are used in the MA16 board, as input to the transfer function inplemented on 16-bit addressable EPROM's. The MA16 board operated at 50 MHz, yielding a response time for a 16 input variable net of 3 μs for a Fisher discriminant (1-layer net) and of 6 μs for a 2-layer net. Results are compared with those previously obtained with the ETANN trigger.

We have already reported at the previous workshop of this series[1] and in subsequent publications (see e.g.[2]) the results from a neural trigger based on the analog ETANN chip which operated in the experiment WA92 at CERN during the 1993 run. The neural trigger included also two boards based on the digital microprocessor MA16. which were not ready at the time of the run because of delays in the debugging of the board control code. We present here results from the MA16 board of the neural trigger module which has been applied off-line to the same experimental data on which the ETANN board operated on line.

Aims of the Neural Trigger in WA92. WA92 is an experiment at CERN looking for the production of beauty particles by a π^- beam at 350 GeV/c impinging on a Cu target (during the 1993 run)[3]. The Neural Trigger hosted in the experiment had the the task of selecting events, already accepted by the WA92 standard trigger, by exploiting a non-leptonic beauty decay signature and to accept them into a special data stream, meant for early analysis. Specifically, the Neural Trigger was trained to enrich the fraction of events with C3 secondary vertices, i.e. branching into three

*E-mail: odorico@bo.infn.it, 38239::odorico

Paper presented at the 4-th Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High energy and Nuclear Physics, April 4-8, 1995, Pisa, Italy

tracks with sum of electric charges equal to +1 or -1. C3 vertices are sought for further analysis aimed to identify charm and beauty non-leptonic decays. Training was done with previously collected events, certified off-line to contain or not a C3 vertex by the Trident event reconstruction program.

Input from the WA92 Trigger Apparatus. Input to the Neural Trigger was provided by the Beauty Contiguity Processor (BCP)[4], which determined tracks and their impact parameters on-line, using hit locations in the silicon microstrip Vertex Detector. The BCP output, arranged in five 64-bit hit-maps each one corresponding to separate impact parameter windows, was preprocessed within the neural crate to yield 16 input variables for the neural chips.

Neural Trigger Module. The Neural Trigger hardware consists of a crate hosting VME9U boards: i) an Interface Board receiving the five 64-bit words plus a termination word from the BCP and control bits to synchronize with the beam pulse; ii) four Preprocessing Boards hosting two independent Preprocessing Unit each, each one of the latter calculating the input variables referring to a given impact parameter window entering the neural chips (with 5 impact parameter windows only 5 such units are actually used); iii) an ETANN board, hosting two independent ETANN neural chips[5]; iv) two MA16 boards, hosting a MA16 neural chip each; v) a (VME6U) VIC board of CES interfacing the VME bus to a personal computer for control and monitoring operations, and which also allows to simulate on-line running conditions by having recorded experimental inputs passed through the crate with subsequent collection of the corresponding outputs. At the time of running the MA16 boards were not yet operating, waiting for debugging of the microcode controlling the boards. The Neural Trigger output trigger bit was sent to the WA92 trigger apparatus and fuller information about the neural chips responses was sent to the WA92 event recording apparatus where it was written in the event record on tape.

The MA16 Microprocessor. The MA16 is a digital, systolic chip developed by Siemens[6-7]. The MA16 has a precision of 16 bits for input variables, 16 bits for weights, 16 bits for scalar multipliers effectively modifying the transfer function shape, 47 bits for thresholds, 53 bits for internal calculations, 38 significant bits for the output (out of a 48 bit output word). It can accomodate an arbitrarily large number of input variables, the processing time increasing with their number. It operates up to a clock frequency of 50 MHz. It processes four input patterns simultaneously in a pipelined fashion. The latter feature cannot be directly used in our trigger application (where one has to wait for one event at the time), but can be exploited to reduce by a factor of four the clock frequency requirements on parts of the chip periphery on the board, as discussed below. The MA16 is a pure processor, with no memory available to store the neural-network parameters. The latter must be stored on external memories, and must be supplied at the required clock times to the MA16, according to the operating code loaded into the MA16 during processing. The MA16 is supported by development software to generate and to check microcodes for MA16 applications in a UNIX/XView environment,

including a MA16 circuit simulator[8].

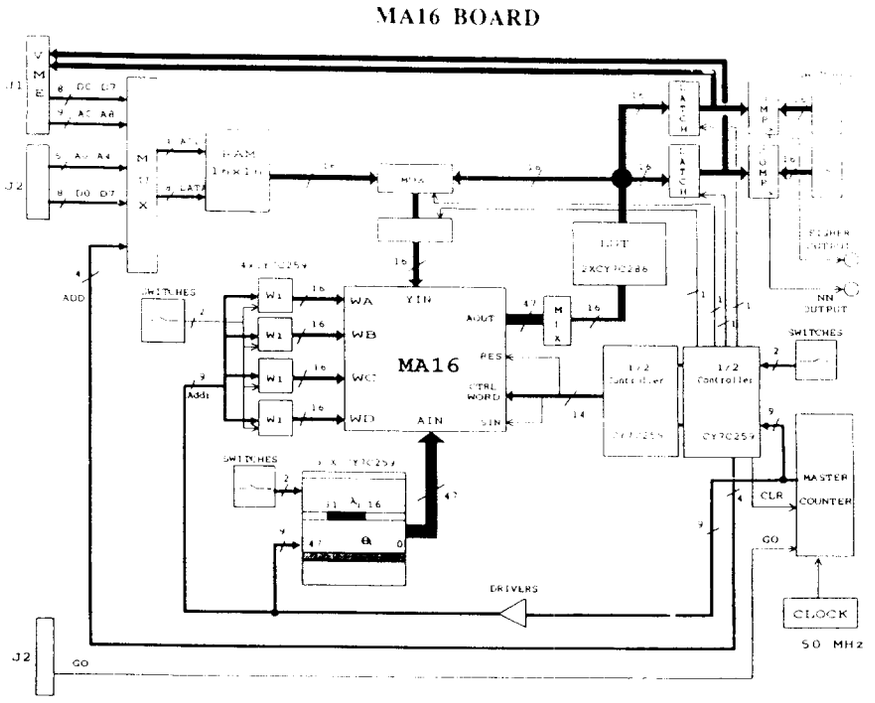

Fig. 1. Block scheme of the MA16 board.

The MA16 Board. The block diagram of the VME9U MA16 board is shown in Fig. 1. The MA16 is serviced by nine 16-bit EPROM's CY7C259 of Cypress Semiconductor, sustaining clock frequencies of up to 83 MHz: i) one storing the MA16 operating code file; ii) one storing the board control file; iii) three storing 16 bits each of the 47-bit words of the file containing the thresholds (Θ_i) and the (16-bit) scalar multipliers (λ_i), which are loaded at different times; iv) four storing the four weight files, each one containing the 16-bit weights of four nodes. The board is conceived to accomodate a 2-layer neural network. The MA16 receives the 16-bit input variables from the board inputs in the 1st pass and in the 2nd pass from its output fed back through a transfer function realized by a Look-Up-Table (LUT) implemented on 16-bit addressable EPROM's CY7C286. The 1st pass input variables are temporarily stored on a RAM (accomodating up to 32 of them), which is normally loaded via a bus with the output from the Preprocessing Boards, but can also be loaded via the VME bus from a PC (for off-line use). From there, they are transferred to input latches, which keep them for 4 clock cycles, during which the MA16 loades each variable 4 times, thus filling the 4 input patterns it can process simultaneously. That leads to a persistence of the MA16 output for

each node for 4 clock cycles. From the 47-bit output (plus a 1-bit overflow flag) a 16-bit slice is cut, truncating the 20 least significant bits and keeping the 15 following bits and the sign bit, which is sent in input to the LUT. The CY7C286 has an address access time of 50 ns, much longer than the clock cycle duration of 20 ns (at 50 MHz), but the persistence of the MA16 output signal for 4 cycles takes care of that. Also, the CY7C286 has a 8-bit output, thus two of them are used to store a 16-bit word. There are two output latches on the signal emerging from the LUT. The board controller can trigger them at the clock times one prefers. By organizing the weights to accomodate two independent nets, one can thus pick up their separate outputs. Each one of the two output latches let the signal persist so that it is made available in output from the board. Both output signals are also sent to separate threshold comparators, with threshold values preset via the VME bus, and the resulting trigger control bits are available in output from the board. The contents of the two output latches can also be read via the VME bus from the PC, for off-line operations. For 16 input variables, the lengths of the files loaded into the CY7C259 EPROM's is less than 512 words, whereas 2048 of them can be stored. Their start address can be preset via the VME bus from the PC, so that four independent sets of net parameters can be used without reprogramming the EPROM's. The board operates at a clock frequency of 50 MHz. The response time for a 1 layer net is under 3 μs and for a 2 layer net under 6 μs.

Results. Two independent neural networks with 16 input variables have been simultaneously loaded on the MA16 board: i) a one layer net, yielding a linear Fisher discriminant, ii) a two layer net, with 5 hidden nodes and one output note, whose response is referred to in the following as the Neural discriminant. C3 acceptance is defined as the fraction of C3 events retained in the neural trigger selected sample out of the full originary sample (of events already accepted by the BCP trigger). C3 enrichment is defined as the ratio of the C3 event fraction in the neural trigger selected sample to the C3 event fraction in the full originary sample. The two top figures in Fig. 2 show the dependences on the Fisher and Neural thresholds of the acceptance and enrichment of C3 events. The results are for three indepedent sets of experimental events, collected a few days apart, each one containing about 300,000 events accepted by the BCP trigger and about 6,000 events with Trident reconstructed C3 vertices[1-2]. To better see the tradeoff between enrichment and acceptance, one can remove the horizontal axis and plot C3 enrichment directly versus the corresponding C3 acceptance, as done in the two bottom figures of Fig. 2. Results from the Fisher and Neural discriminants are qualitatively similar, which indicates that the problem at hand is essentially linearly separable and does not need more than one layer to be handled. The MA16 results are also qualitatively similar to those provided by the ETANN trigger[1-2], whose analog precision is of a few percents. Clearly this neural trigger application is not the best suited to profit of the high precision of the MA16. The fact is that the relevant input variables one extracts from the 64-bit hit-maps in input from the BCP have integer values ranging from 0 to 16 (decimal), with most of them actually varying only between 0 and 1.

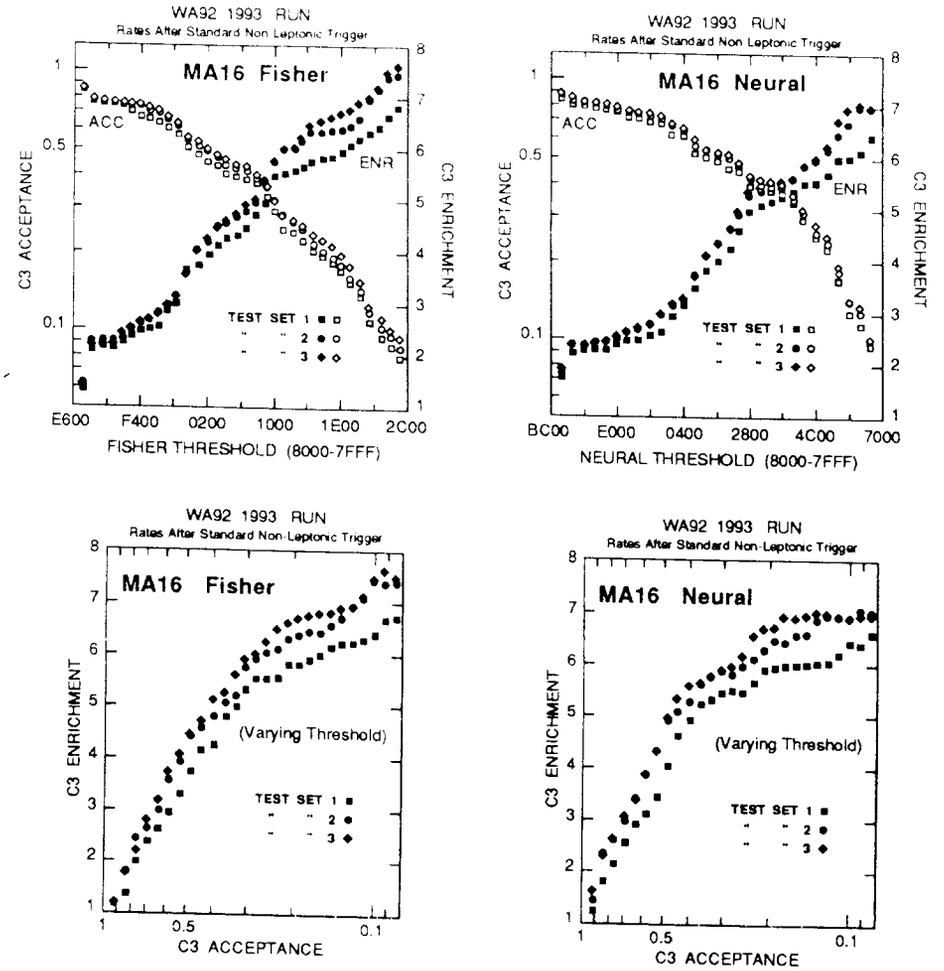

Fig. 2. Results from the MA16 neural trigger for the acceptance and the enrichment of events with C3 vertices, varying the output threshold.

Conclusions. The MA16 part of the Neural Trigger module, which operated on-line during the 1993 run of the WA92 experiment, has been completed and applied off-line to the same data collected and selected by the ETANN board. Differently from ETANN, the MA16 board does not have stability problems to care about and thus its testing does not need real operating conditions. This first neural trigger actually operating in a standard high-energy physics experiment shows the viability and the merits of neural trigger technology. The experimental set-up in which it operated, dealing with a relatively low energy beam, is not the best suited to show its full potential especially in connection with the MA16, since the trigger signature

turns out to be a relatively simple one easily handled by a low level neural network. That is expected to change with the high multiplicity events which will have to be sorted out at the LHC experiments.

Acknowledgements

We would like to thank U. Ramacher (Siemens AG) for supporting the MA16 board project. We would also like to thank the WA92 Collaboration for hosting this demonstration experiment. One of us, R.O., acknowledeges useful conversations with R. Battiti about neural net training procedures and would like to express his appreciation to P. Ribarics for calling his attention to the MA16.

References

1. C. Baldanza *et al.* (ANNETTHE Collaboration), in *Proc. 3rd Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics*, ed. K.-H. Becks and D. Perret-Galix (World Scientific, Singapore, 1990), p. 391.
2. C. Baldanza *et al.* (ANNETTHE Collaboration), University of Bologna report DFUB 94/16 (1994), *Nucl. Instr. and Meth.* **A**, in press.
3. M. Adamovich *et al.* (WA92 Collaboration), CERN/SPSC 90-10 (1990); *Nucl. Phys.* (Proc. Suppl.) **B 27**, 251 (1992).
4. G. Darbo and L. Rossi, *Nucl. Instr. and Meth.* **A 329**, 117 (1993).
5. *80170 Electrically Trainable Analog Neural Network Data Booklet*, Intel Corp., 2250 Mission College Boulevard, Santa Clara, CA 95052-8125, USA.
6. U. Ramacher *et al.*, in *Proc. Int. Conf. on Systolic Arrays*, Killarney, Ireland, Prentice Hall (1989), p. 277;
7. U. Ramacher, J. Beichter and N. Bruels, *MA16 - Programmable VLSI Array Processor for Neuronal Networks and Matrix-based Signal Processing*, User Description, Version 1.4, Siemens AG, ZFE T ME2, 81730 Munich, Germany (available from N. Bruels, E-mail: nico@dali9.zfe.siemens.de).
8. N. Bruels, *MA16 - Development Software*, User Manual, Version 1.0, Siemens AG, ZFE T ME2, 81730 Munich, Germany.