DD

DFUB 95/12

Study of a MA16-Based Vertex Trigger for the CDF Experiment[#]

R. Odorico^{*}

University of Bologna, Physics Department
INFN, Sezione di Bologna
Via Imerio 46, I-40126 Bologna, Italy

ABSTRACT

sw 95 42

A Neural Network trigger for $b\bar{b}$ events based on the SVT microvertex processor of experiment CDF at Fermilab is presented. It exploits correlations among track impact parameters and azimuths calculated by the SVT from the SVX microvertex detector data. The neural trigger is meant for implementation on the systolic Siemens microprocessor MA16, which has already been used in a neural-network trigger for experiment WA92 at CERN. A suitable set of input variables is found, which allows a viable solution for the preprocessing task using standard electronic components. The response time of the neural-network stage of the trigger, including preprocessing, can be estimated $\sim 10 \mu\text{s}$. Its precise value depends on the quantitative specifications of the output signals of the SVT, which is still in development. The performance of the neural-network trigger is found to be significantly better than that of a conventional trigger exclusively based on impact parameter data.

[#] Postscript file: <http://www.bo.infn.it/preprint/odorico.html>

^{*} E-mail: odorico@bo.infn.it

1. Introduction

The Silicon Vertex Tracker (SVT) [1] is a trigger processor under development for the CDF experiment. It is designed to reconstruct tracks having large enough transverse momenta ($> 1\text{-}2\text{ GeV}/c$) in the Silicon Vertex Detector (SVX) [2] and Central Tracking Chamber (CTC) with enough speed and accuracy to be used at trigger level 2 to tag events containing secondary vertices from b decay. The SVT provides a list of track parameters, which includes the impact parameter of each track. Conventional tagging requires the presence of one or two tracks with impact parameter above some threshold, disregarding other track information.

It is natural to expect that trigger performance should improve when using more of the track information provided by the SVT, in particular track azimuths. That should be especially true for complex events, like those with large values for the total transverse energy, which are e.g. of interest for estimates of the background to the top-quark signal. As to the problem of using such information in an on-line trigger, a convenient solution can be provided by dedicated neural network chips, once a suitable set of input variables for the neural net, to be preprocessed on-line from the SVT output, has been identified.

A preliminary study of this type, based on the analog NeuroClassifier chip, was presented in [3], although the problem of hardware preprocessing of the variables still remains to be settled.

Aim of this note is to present results for an approach to the problem based on the Siemens MA16 neural chip [4-5]. This is a digital chip with 16-bit I/O precision, which with proper handling of the input variable preprocessing can accommodate the entire neural-network stage for the level-2 trigger within $\sim 10\text{ }\mu\text{s}$, if no more than 32 input variables are used. Most of the present study deals with the search of a limited set of easy to compute input variables which allows to exploit correlations among the relevant track parameters within $b\bar{b}$ events.

Section 2 describes the SVT. Section 3 defines the event samples used. Section 4 outlines the MA16 microprocessor and board. Section 5 shows the azimuth-impact parameter correlations for secondary vertex tracks. Section 6 presents the crucial choice of input variables for the MA16 neural network, which are amenable to a viable preprocessing procedure. Section 7 describes the neural network and the Fisher discriminant to be loaded on the MA16 board. Section 8 discusses the results. Section 9 presents the conclusions.

2. SVT Output and SVT Trigger

For each event, that is a $p\bar{p}$ collision, the SVT supplies for all tracks, carrying a transverse momentum above $1\text{-}2\text{ GeV}/c$, the values of their impact parameter D , azimuthal angle ϕ and transverse momentum p_t . The impact parameter D of a track is defined as the minimum distance between the reconstructed trajectory and the beam axis, Fig. 1. All tracks and impact parameters are measured as projections in the transverse plane and no knowledge of the z coordinate, along the beam axis, is assumed.

The conventional SVT jet trigger condition for $b\bar{b}$ events, with b decaying into anything, uses only the impact parameter information provided by the SVT and amounts to the request of one or two tracks having an impact parameter above some minimum value, without further exploitation of the SVT output.

At the moment, the SVT is not yet operating within the CDF experiment. Its final set-up will be determined by the eventual configuration of the SVX upgrade, SVX II [6], which will start operating after the 1998 Main Injector upgrade of the accelerator complex at Fermilab. The simulators of SVX II and of SVT were not available for this study, and thus finite resolution effects and other experimental uncertainties of the vertex detector have been neglected. Once the latter are included, it is natural to expect that they will affect more the performance of the conventional SVT trigger than that of the neural trigger, since the latter relies on a number of individual data points and is thus less sensitive to their individual fluctuations.

3. Samples of Simulated Events

The event samples used for this study have been generated with the standard simulation software package of CDF. That includes GEANT simulation of the apparatus, except for SVX and SVT. The beam spot size has been specified with a σ in both transverse directions of $63\ \mu\text{m}$ (that is relevant for the discrimination of secondary vertices from the primary one). Events were Monte Carlo generated with a parton minimum transverse-momentum cutoff of $50\ \text{GeV}/c$, corresponding to a minimum total transverse energy (at the primary parton level) of $100\ \text{GeV}$. Two separate event samples were used: one for $b\bar{b}$ events and the other for ordinary jetty events. From the latter, $b\bar{b}$ events (amounting to about 2% of the total) were eliminated. The two event samples consisted of about 10,000 events each.

The kinematic conditions chosen for the event generation are strictly related to the aims of the present study. At low transverse energies, the event structure in terms of SVT-reconstructable tracks is rather simple, and thus the margins for improvement by using a trigger sensitive to a larger amount of the available information are rather limited. At large transverse energies, on the other hand, the amount of tracks reconstructed by the SVT becomes substantial and it may pay off to have a more sophisticated exploitation of the global event structure.

4. MA16

The MA16 is a digital, systolic chip developed by Siemens [4-5]. It 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 accommodate 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. 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 [7].

A VME9U board for the MA16 has been realized [8-10]. In it, the MA16 is serviced by nine 16-bit EPROM's, storing the MA16 operating code, the board control file and the neural net parameters. The board is conceived to accommodate 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. The 1st pass input variables are temporarily stored on a RAM (accommodating up to 32 of them), which is normally loaded via a proprietary bus with the output from other boards, but can also be loaded via the VME bus, e.g. from a PC. 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 accommodate two independent nets, one can thus pick up their separate outputs and send them to the associated board output ports. Both output signals are sent to separate threshold comparators, with threshold values preset via the VME bus, and the resulting trigger control bits can be picked up outside the board. The contents of the two output latches and of the trigger control bits can also be read via the VME bus, e.g. from a PC. The board operates at a clock frequency of 50 MHz. The response time for a 1 layer net is under $3\ \mu\text{s}$ and for a 2 layer net under $6\ \mu\text{s}$.

The MA16 board was originally developed as part of a full Neural Trigger module [8-9] operating within the CERN experiment WA92 [11], which looks for the production of beauty particles by a π^- beam at $350\ \text{GeV}/c$ impinging on a fixed target. The Neural Trigger hosted in

the experiment had the task of selecting events, already accepted by the WA92 standard trigger, by exploiting vertex detector information elaborated by the Beauty Contiguity Processor [12] and to accept the events 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 tracks with sum of electric charges equal to +1 or -1. C3 vertices were 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.

5. Azimuth - Impact Parameter Correlations

The values of the impact parameter D and of the azimuthal angle ϕ of a track coming out of a vertex with a distance r_v from the beam axis and azimuth ϕ_v are related by [3] (Fig. 1):

$$D = r_v \sin(\phi_v - \phi) \quad (1)$$

That of course holds true with the same r_v and ϕ_v values for all tracks originating from the same vertex. Since typically $|\phi_v - \phi| \ll 1$, tracks from the same vertex approximately fall on a line in the (D, ϕ) plane with a negative slope:

$$D = -r_v \phi + \text{constant} \quad (2)$$

That can be seen in Fig. 2, where an example of $b\bar{b}$ event is shown and compared with a typical background event.

For the primary vertex, r_v is small: its distribution is determined by the beam spot size. As a consequence the D values of emerging tracks are all small and thus show little dependence on ϕ . For a secondary vertex, r_v is related to the lifetime of the particle whose decay generates the vertex and it is substantial, typically. Thus, emerging tracks approximately distribute themselves along steep lines in the (D, ϕ) plane (Fig. 2). The maximum theoretical value of $|D|$ for such tracks is r_v , but the actual value depends on the vagaries of the track azimuths. That reduces the efficiency of a trigger exclusively based on D . If, on the other hand, the trigger relies on the identification of the (D, ϕ) linear structures associated with secondary vertices, then its performance only depends on having enough spread for the vertex tracks in the (D, ϕ) plane, and becomes less affected by statistical fluctuations.

6. Preprocessing of Input Variables

The choice of variables, encoding the events and to be supplied in input to the trigger discriminator, must satisfy two main criteria: i) they must convey as much as possible of the information relevant for the event discrimination, ii) they must be simple and fast to calculate on-line by means of available electronic components.

In order to get variables sensitive to the presence of linear distributions of tracks in the (D, ϕ) plane, it is natural to think of projecting the tracks on an ordered set of straight lines oriented at increasing angles with respect to the ϕ axis: for a direction orthogonal to that along which the tracks are aligned, the associated track distribution will exhibit a peak, which will broaden and eventually dissolve as one moves away to other directions. Thus picking up the maxima of the track distributions for a set of directions provides a pattern from which a neural net can read the likely presence of secondary vertices.

To have more than one secondary vertex in the event does not affect the above pattern profile appreciably, if the distributions are not too fine grained and the slope increments are not too small. Undesired noise is rather generated by the crowding of small $|D|$ tracks. That can be suppressed either by neglecting tracks with $|D|$ below some minimum value or, more smoothly, by weighting tracks in the distributions according to some function of $|D|$. Another unwanted

effect, affecting the value of the maximum height in the histogram distributions, originates from having tracks from the same secondary vertex sometimes projecting onto separate bins even for the direction which, among those considered, is closest to the theoretical one with maximum sharpness of the peak. To circumvent that, one can simply introduce multiple entries for each track into the histograms: besides the original bin into which the entry is made, the latter is accumulated also into the two adjacent bins. The ensuing loss of resolution is of negligible impact on the result, once contributions from small $|D|$ tracks are suppressed.

In sum, we consider the following n variables. For each $i = 1, n$ we pick up an angle α_i in the (D, φ) plane with respect to the φ axis and distribute tracks according to

$$\omega_i = (D/D_s) \cos(\alpha_i) + \varphi \sin(\alpha_i) \quad (3)$$

where we take $D_s = 250 \mu\text{m}$. The distribution is organized as a 32-bin histogram. For each track, an entry is made in the bin within which ω_i falls and in the two adjacent bins, with a D dependent weight given by the function (Fig. 3):

$$D\text{-weight} = 1 - \frac{x}{e^x - 1}, \quad x = \frac{D}{\Delta} \quad (4)$$

where we take $\Delta = 50 \mu\text{m}$. Finally, the maximum over the histogram is identified with the i th variable, after truncation to an integer. We consider $n = 8$ variables of this type, associated with directions $\alpha_i = 5^\circ (i - 1)$, $i = 1, n$ (i.e. $\alpha_i = 0^\circ, 5^\circ, 10^\circ, \dots, 35^\circ$). Positive values of α_i are enough, since slopes of vertex track distributions are negative, and a maximum angle of 35° is sufficient to see the disappearance of secondary vertex peaks because of the typically high values of the corresponding slopes.

Fig. 4 shows the 32-bin distributions associated with angles $0^\circ, 10^\circ, 20^\circ$ and 30° for the $b\bar{b}$ event of Fig. 2. From it one can see the peaks originated by the tracks of the two secondary vertices which are visible at small angles, and the way they disappear as the angle increases. One can also see how the value of the maximum over a distribution is able to gauge the presence of the peaks.

Fig. 5 shows the mean values of the 8 variables so constructed over the $b\bar{b}$ and the background event samples. The substantial difference of the profiles of the 8-variable patterns for the two event samples provides an indication of their discriminating power.

The 8 variables so defined efficiently convey the presence of D - φ correlations associated with secondary vertices, but they turn out to be less effective in representing the D distribution of tracks. In order to cope with that, we introduce other 5 variables counting the number of tracks with $D > 0, 50, 100, 200, 400 \mu\text{m}$, respectively. We also find it useful to add other 3 variables which are identical to the first 3 of the above 8 D - φ variables, except for the absence of the D -weight (i.e. each track is entered with weight equal to one). In fact, by comparing the values of the same (small angle) variables calculated with and without D -weights, the neural net is apparently able to correct for spurious effects induced by the suppressed contribution of small D tracks. We thus come to a total of 16 input variables, which can be conveniently accommodated by the MA16 board presented above.

As to the hardware implementation of the input variable preprocessing, since it involves at most integer 8-bit multiply+add arithmetics, it can be easily handled by 16-bit addressable PROM's acting as LUT, in particular for the calculation of ω_i . For the histogramming and maximum search, standard DSP's can be used. The D -weight can be tabulated as a 4-bit integer, with a final 4-bit right-shift on the accumulation results. Of course, the 16 variables can be calculated in parallel. The specifics of the hardware implementation depend on the way the SVT output will be made available, but it looks reasonable to estimate that the preprocessing response time will stay within $\approx 4\text{-}5 \mu\text{s}$ and it will be largely dominated by the I/O transfer times, which increase linearly with the number of tracks for the input and with the number of variables for the output.

7. Neural Network

A neural net with one hidden layer of 5 nodes and one output node gives an optimal performance for the 16 input variables specified in the previous section, although the dependence on the number of hidden nodes is not critical. The transfer function has been taken to be the standard sigmoid: $(e^{-x}+1)^{-1}$. We have also considered a Fisher discriminant (i.e. a one-layer net). The two nets have been trained with one half of the simulated event samples described above, i.e. about 5,000 events per class, leaving the residual events for testing, from which the results quoted below are obtained. All tracks carrying a transverse momentum above 1 GeV/c have been used, in order to provide a more complete information to the neural network. For the event samples considered the mean multiplicity of such tracks is of about 20. The training has been done with program NEURAL [13], which includes the effects of the finite precision of the MA16 and of the transfer function LUT. The resulting 2-layer net and the 1-layer Fisher discriminant can be simultaneously loaded on the MA16 board.

8. Results

Fig. 6 shows the results, expressed in the form of enrichment of the signal/background ratio (S/B) versus the corresponding efficiency, for a moving cut on the neural net or Fisher discriminant output. Enrichment is defined as the ratio $(S/B)/(S_0/B_0)$, where S_0 and B_0 are the numbers of signal and background events, respectively, in the original samples, and S and B are the corresponding quantities in the samples surviving the cut. Efficiency is defined as S/S_0 . Also shown are the results for a conventional D trigger, i.e. requesting the existence of a least one, or two, tracks with $p_t > 2$ GeV/c, pseudorapidity $|\eta| < 1$, $|D| < 1000 \mu\text{m}$ and $|D|$ larger than a moving $|D|$ cut.

For the large $\sum |E_T|$ events we are considering, the performance of the neural net trigger looks significantly better than that of a conventional D trigger. For efficiencies above 0.5, a 2-layer net does not improve on a simpler linear Fisher discriminant. But for lower efficiencies, which may become of interest for very high luminosities, the optimal 2-class separation gets far from linear and a 2-layer net largely outperforms the Fisher discriminant.

9. Conclusions

As to be expected, a more complete use of the track information provided by the SVT significantly improves on the performance of a simple-minded impact parameter trigger. The gain is especially appreciable when the complexity of the event information given in output by the SVT is high enough to justify the use of a neural network, as in the case of events with large total transverse energy. It has been shown that the preprocessing of input variables needed for this sake can be designed in a viable way from the point of view of the hardware implementation. Pending a detailed specification of the output signals from the SVT, one may estimate that an implementation of the neural network stage of the trigger on the MA16 microprocessor would require $\sim 10 \mu\text{s}$, including the time needed for the input variable preprocessing.

Acknowledgements

This study has profited from a number of conversations with Bruce Denby, who has originally put forward the idea of employing a neural trigger with the SVT and has coordinated a similar project based on the NeuroClassifier analog chip. Further clarifications on the work already done on the latter project derived from conversations with Dino Salvatore. G. Pauletta and L. Santi have kindly made available the event files used in [3].

References

- [1] G. Punzi and L. Ristori, CDF Note 1872 (1992).
- [2] D. Amidei et al., Nucl. Instr. and Meth. A350 (1994) 73.
- [3] D. Salvatore et al., Proc. 4-th Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, 1995, Pisa, Italy.
- [4] U. Ramacher et al., in Proc. Int. Conf. on Systolic Arrays, Killarney, Ireland, Prentice Hall (1989), p. 277.
- [5] 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).
- [6] J. Antos et al., Nucl. Instr. and Meth. A360 (1995) 118.
- [7] N. Bruels, MA16 -- Development Software, User Manual, Version 1.0, Siemens AG, ZFE T ME2, 81730 Munich, Germany.
- [8] C. Baldanza et al. (ANNETTHER 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, 1993), p. 391.
- [9] C. Baldanza et al. (ANNETTHER Collaboration), University of Bologna report DFUB 94/16 (1994), Nucl. Instr. and Meth. A, in press.
- [10] C. Baldanza et al. (ANNETTHER Collaboration), Proc. 4-th Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, 1995, Pisa, Italy.
- [11] M. Adamovich et al. (WA92 Collaboration), CERN/SPSC 90-10 (1990); Nucl. Phys. (Proc. Suppl.) B 27, 251 (1992).
- [12] G. Darbo and L. Rossi, Nucl. Instr. and Meth. A 329, 117 (1993).
- [13] A. Cherubini and R. Odorico, LVQNET vers. 1.0, Comp. Phys. Commun. 72 (1992) 238; R. Odorico, Program NEURAL, in preparation.

Figure Captions

Fig. 1 - Relation between a track azimuth φ and its impact parameter D , as defined in the SVT. Axes x and y are orthogonal to the beam direction.

Fig. 2 - Distributions of tracks with $p_t > 1$ GeV/c in the (D, φ) plane for a $b\bar{b}$ and a background event with $\sum |E_T| > 100$ GeV (at the parton level). The linear, close to the vertical, distributions present in the $b\bar{b}$ event are originated by two secondary vertices.

Fig. 3 - Assumed dependence of D -weight of track on its impact parameter D . Δ is an adjustable parameter.

Fig. 4 - D -weighted track distributions in the variable $\omega_i = (D/D_s) \cos(\alpha_i) + \varphi \sin(\alpha_i)$ with $\alpha_i = 0^\circ, 10^\circ, 20^\circ, 30^\circ$ ($D_s = 250 \mu\text{m}$) for the $b\bar{b}$ event of Fig. 2. The figure shows how the distribution maximum relates to the presence of peaks associated to secondary vertices in the event.

Fig. 5 - Mean values over the $b\bar{b}$ and the background event samples of the maxima of the track distributions in the variable ω_i for the α_i values shown on the horizontal axis.

Fig. 6 - Results for the enrichment in the signal/background ratio versus the signal efficiency obtained over the $\sim 5,000$ $b\bar{b}$ -signal and $\sim 5,000$ background events with $\sum |E_T| > 100$ GeV used for testing. Results refer to the neural-network and Fisher classifiers and to conventional trigger conditions requesting the presence of one, or two, tracks with impact parameter D above some value. Each point in the graph refers to a specific threshold value for the corresponding trigger control parameter (neural-network output, Fisher variable and D , respectively).

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6