Neural Particle Discrimination for Triggering Interesting Physics Channels with Calorimetry Data

André Rabello dos Anjos and José Manoel de Seixas

Abstract— **High Energy Physics experiments use on-line event validators (triggers) to distinguish known physics from unstudied physics events. This article introduces a triggering scheme for high input rate processors, based on neural networks. The technique is applied to the Electron/Jet discrimination problem, present at the Second Level Trigger of the ATLAS experiment, being constructed at CERN, Switzerland. The proposed solution outperforms the scheme adopted nowadays at CERN, both in discrimination efficiency and performance, becoming a candidate algorithm for implementation at the experiment.**

Keywords— **Neural Networks, Relevance, DSP, Trigger, High Energy Physics**

I. Introduction

In High-Energy Physics experiments (HEP), it's of practice to collide (bunches of) particles in order to produce unstudied physical phenomena. These experiments use the state-of-the-art in signal detection in order to classify events that separate the physics one wants to observe (signal) from the ones that represent already known events (noise).

The event classification is commonly denominated *triggering* in HEP experiments. *Trigger* systems are usually very fast and robust due to the small time window for data processing and to the violent conditions (radiation, years of operation) they are submitted to. In modern experiments, the trigger is composed of a few cascaded sections that can perform slower but better filtering schemes. In this article, we present a neural solution for (part of) the second level trigger of ATLAS (A Toroidal LHC ApparatuS) experiment located at CERN, Switzerland.

A. The ATLAS Experiment

The ATLAS experiment shall be ready to run in 2006. It aims investigate the structure of matter using the state-of-the-art in particle detection technology. For this matter, the ATLAS collaboration is building a very large detector (see Figure 1). The detector will be installed at one of the collision points of the LHC (Large Hadron Collider), which will be colliding bunches of protons at 14 TeV at the center of mass. The LHC can produce 40,000,000 events (collisions) per second and the ATLAS should select, *on-line*, from those, the interesting ones to be recorded for further analysis.

The ATLAS is actually a set of specialized detectors that measure particle (object) features. At the inner part, there are the track detectors, which are responsible for determining the particles trajectory. Surrounding the track detectors are the calorimeters. Calorimeters measure the energy of particles by absorbing their energy while they cross their interior[1]. At last, covering the whole structure, one has muon detectors, signaling the existence of this type of particles (muons) in a event.

Laboratório de Processamento de Sinais - COPPE/EE/UFRJ. **Electronic Mail**: `Andre.Rabello@ufrj.br, seixas@lps.ufrj.br`

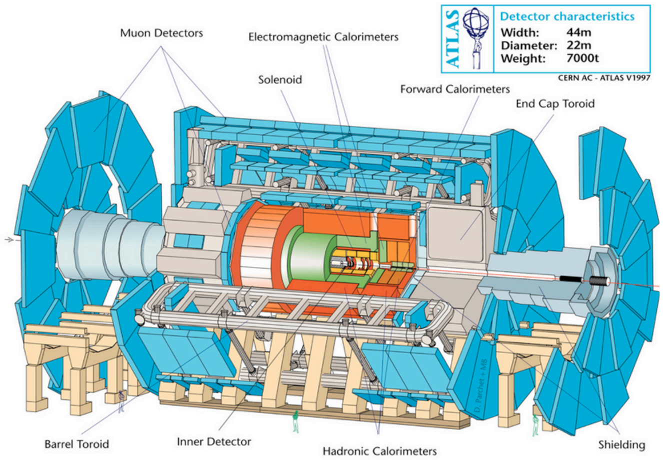

Fig. 1. A sketch of the ATLAS detectors. Notice, in scale, people at the left-bottom corner.

There are two types of calorimeter at ATLAS, forming two sections: electromagnetic and hadronic. The electromagnetic (e.m.) section is responsible for measuring the energy of particles, like electrons and photons, while the other section is responsible for heavier objects, like mesons and jets (showers) of particles. The hadronic section surrounds the e.m. section. All calorimeters of ATLAS are segmented into cells, which helps dimensional positioning at the structure. They are also subdivided into shelves, being the internal ones more segmented (more cells) than the external ones. The arrangement is optimized for the classes of particles the ATLAS experiment intends to detect.

The detector uses a special coordinate system which transforms its toroidal aspect into planes of detectors, stacked one above the other. The coordinates are $\phi = \text{atan}(\frac{x}{y})$, the rotation angle, $\eta = -\log(\tan(\frac{\phi}{2}))$, the *pseudo-rapidity* and the Cartesian coordinate z. This set can pinpoint the exact location of each detector component within the ATLAS structure.

B. ATLAS Trigger

Due to the rare physics to be studied at ATLAS, most collisions will produce already known, studied cases. It's es-

[1]This is possible since particles loose energy while interacting with the calorimeters, decaying into lighter objects.

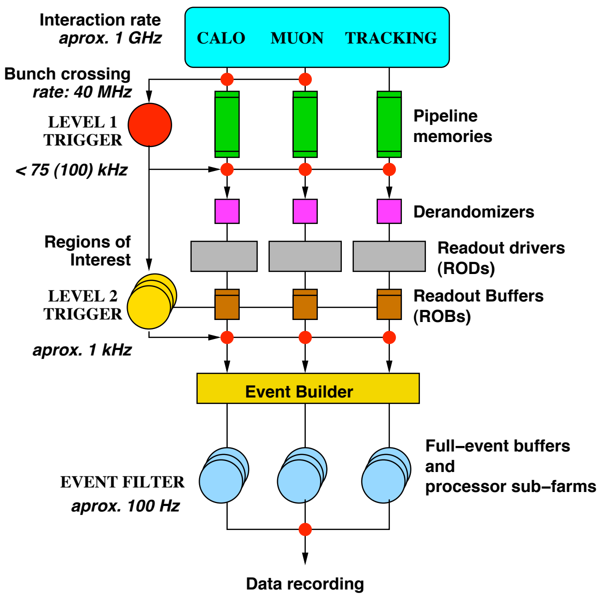

Fig. 2. Sketch of the triggering system of ATLAS.

timated that only a few bunch crossings per day will really produce the physics of interest. Because of these statistics, and to the volume of data produced at each iteration (∼1 megabyte), the ATLAS experiment will rely on a very efficient triggering system that will classify, *on-line*, events that possibly characterize the physics to be studied from those corresponding to known processes [1].

The ATLAS On-line Triggering System (see figure 2 is composed of three classification levels cascade connected. Each of the triggering levels is more complex than the previous one and analyze better the event in a considerably slower manner. The first level is built upon low-programmability hardware and optimized for speed. The second and third levels use fast computing networks from the commodity market. This choice allows the second and third levels to be adapted and maintained along the operation time foreseen for ATLAS.

The whole triggering system works as follows: after an event is produced by LHC and detected by ATLAS, the First Level Trigger (T1), looking only into to the calorimeters and the muon detector, classify the incoming event by searching the sensibilized regions of the detector (also known as Regions of Interest - RoI). The outcome of the search is a number of objects that may characterize an interesting event. This method shall guarantee a reduction on the event rate of 1000:1. The tagged events are sent to Second Level Trigger (T2) for further analysis.

T1 will only analyze each RoI (object) locally, ignoring the occurrence of other objects during the event, and that is why it is fast enough to cope with the high input rate of the experiment. The T2 [2] is the first step of the triggering process where the event will be analyzed globally, looking to each of the objects tagged by T1. The T2 will have at its disposal the data from all detectors (differently from T1) and starts analyzing the event, confirming the analysis performed by T1. Despite the existence of many processing strategies, pre-triggering using calorimetry only data is the most suited, since it allows faster processing of the incoming T2 event rate (100,000 per second). At T3 (the Third Level Trigger), the event is also analyzed globally, with access to all detector data (not only per object analysis as in T2). T2 should allow only 100 to 1,000 events per second to hit T3.

C. *Calorimetry Analysis in T2*

Calorimetry analysis in T2 (pre-trigger) will be able to distinguish many object types produced by the LHC interactions. High energy (momentum) electrons though, are likely to occur much more (∼60-70%). Electrons are commonly confused with jets of particles by T1, because of the way they interact with the calorimeters of ATLAS. For each 25,000 objects defined as electrons by T1, only 1 is likely to be an electron. This sets the background jet rate at T2.

Particles interact with the calorimeters depositing their energy along the penetration direction. As they interact with the calorimeter and loose energy, particles decay into less massive particles, that also interact with the calorimeter, producing a shower [3].

Electrons, differently from jets, loose almost 100% of their energy into the e.m. section. Jets usually go a bit further, being stopped only by the hadronic calorimeters. It happens though, that sometimes jets develop a cascade very much like electrons and seem confusing to T1. T2, despite of that resemblance, can classify electrons a lot better by using the calorimetry full granularity, only available at this triggering level.

An example picture of an electron RoI is shown in Figure 3. This plot shows the amount of energy deposited by the object at each calorimeter layer. The first four layers (left part) correspond to the e.m. calorimeter section, while the remaining three, to the hadronic section.

II. Classical Quantities

Electron/Jet separation is prototyped nowadays using a bi-dimensional plotting-and-cutting strategy. Basically, the RoI data are compacted into four highly discriminating features, also called *Classical Quantities*:

1. **e.m. energy** - The sum of energies deposited in a 0.175 (in ϕ) by 0.075 (in η) window at the e.m. section, centered at the cell with the highest energy deposited at the second e.m. layer;
2. **hadronic energy** - The sum of energies deposited in a 0.2 by 0.2 window centered at the energy with the highest energy deposited at the second e.m. layer section;
3. **e.m. rate** - The rate of the sum of energies deposited in a 0.175 (in ϕ) by 0.075 (in η) window at the second e.m. layer by a sum at a 0.175 by 0.175 window for the same layer. Both windows are taken centered on the cell with the highest energy deposited at the second e.m. layer;
4. **strip rate** - The rate between the difference and sum of the two highest peaks at the first e.m. layer.

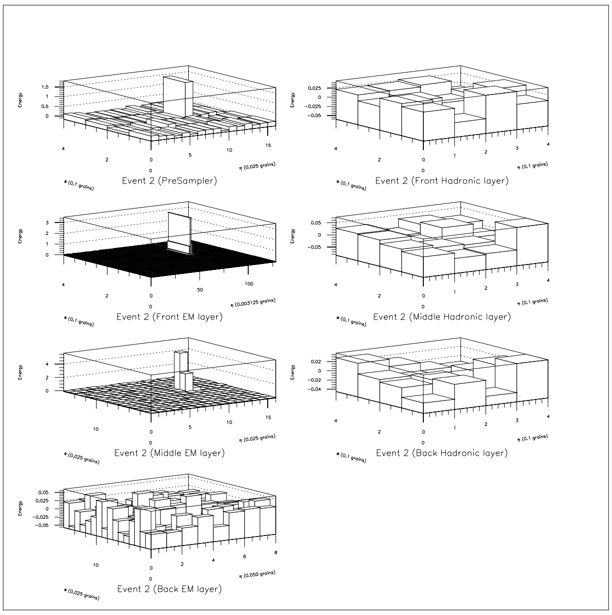

Fig. 3. Electron interaction with the ATLAS calorimeter system, which includes a pre-sampler layer. The layers, are in order, from the center of collision to the outer detector part, from top to bottom, left to right.

Quantities 1 and 2 are highly discriminating because one expects electrons to concentrate all of its energy within the e.m. section, while for jets it's expected that one could observe some energy at the hadronic sections. Also, since electrons are single objects, as opposed to jets, it's expected that their interaction with the e.m. sections would not spread-out a lot its energy. These features can be measured by quantities 3 and 4.

The plotting-and-cutting method is applied by selecting two out of the four variables described and choosing a cut (threshold) based on a test set. This is done for every combination of two variables, generating six cuts. The cuts have to be applied by a specialist, as the cutting order affects the efficiency of the method. It performs well for the data sets, but is not scalable and suffers a lot from input bias.

III. Neural Discrimination

Based on the previously described quantities, it was developed both a linear and a neural discriminator that would perform the Electron/Jet separation task. At this section we also show that another compactation scheme that better preserves the object features can be adopted, outperforming the results achieved with the Classical Quantities.

A. Neural Training and Data Samples

The Electron/Jet classification per RoI basis can be better accomplished using neural networks. This classification method is well suited for the job, since it comprises several characteristics which are interesting to T2:

- Scalability - the neural network can be rescaled to suit changes in calorimetry configuration;
- Easy management and upgrade - environmental changes can be easily incorporated by re-training the network;
- Robustness - in the case of hardware failure, some calorimeter cells may be unavailable for T2. Neural Networks can cope with that, due to its distributed process characteristic;
- Efficiency - Neural Networks can achieve very high discrimination efficiency;
- Fast operation - Neural Networks can be implemented in a variety of platforms (or co-platforms), allowing very fast execution;

The network model of choice for this work was the fully-connected feed-forward three layer perceptron. The training method adopted was back-propagation using variable sized batches with momentum and learning rate decay [4]. The perceptrons use the hyperbolic tangent as activation function. During the neural network training phase, the stop criteria used was proportional to the network discrimination efficiency (maximum product of classification efficiencies, taking in consideration all possible thresholds for the network output), instead of the mean-square-error (MSE). This way, network training would be focused on providing better discrimination capabilities instead of only minimizing the output error.

The data sample consisted of around 600 electrons and 3,600 jets, being equally divided into subsets for training and testing the network performance. The objects in consideration are derived from careful Monte Carlo simulations to reproduce the physics to be studied and have fulfilled T1 selection criteria. Because of that, the number of available electrons is much lower than the one for jets. The RoI size given at the database was the one predefined for electrons, 0.4 in η versus 0.4 in ϕ.

B. Results for the Four Classical Quantities

The *Region of Convergence* [5] for the neural and linear discriminators can be seen at Figure 4. The (jet) false alarm was pre-multiplied by 25 kHz, as this is the estimated background rate.

For the same data set, neural networks have performed a lot better than linear discrimination using the four quantities described. Notice that the plotting-and-cutting method doesn't allow us to plot a precise RoC since the overall discriminator threshold is placed at the sixth-dimensional space, as there are six of them. With optimization though, it can be seen that this method gets close, but is still worse, than using neural networks for this problem.

C. A Different Compactation Approach

As the reader may have noticed in Figure 3, the total number of cells per RoI is rather big (around 1,000). Neural Networks, despite also being suited for such a large input vector, can train and operate faster with smaller ones. Therefore, some sort of data compactation that would not remove important object features is desirable for designing a compact, but efficient classifier. Also, the energy of particles may vary a lot from event to event and a clever

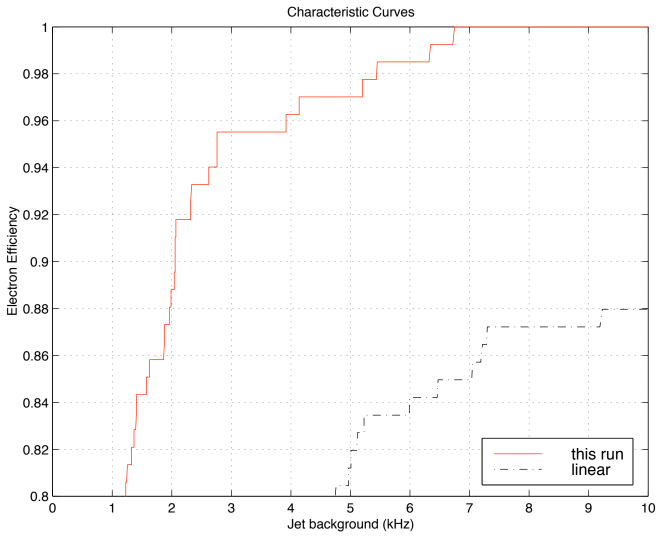

Fig. 4. Results for both linear and neural (represented by the "*This run*" line) discriminators for the four quantities presented.

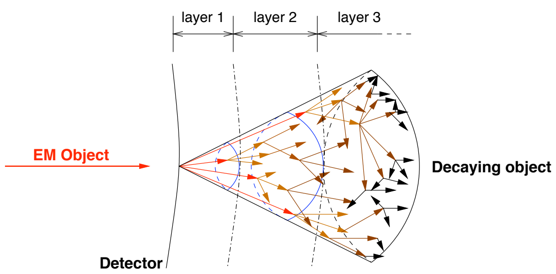

Fig. 5. Object interaction with the calorimeters, as a shower of particles is produced.

normalization scheme has to be implemented, in order for a single network to be able to cope with full energy ranges.

It is reasonable to assume that the objects would decay into less massive ones, generating a shower of particles, isotropically. This shower would resemble a cone, if one takes in consideration all the path the object traverses while interacting with the calorimeters. This *decaying* scheme is simplified in Figure 5.

Based on this observation a new compaction strategy is developed. This scheme is based on adding the energy sampled by cells grouped into concentric rings [6]. It gathers the deposited cell energies at the same layer in a way that "square" rings are formed, starting from the cell with the maximum energy deposition at the current layer (the first ring), as exemplified in Figure 6.

It's clear from the drawings that there will be some opened rings. In fact, the pre-sampler layer does not form proper rings, but columns of cells are added to form the ring sums of this layer. Also, as one goes from the center to the outer part of the layer, the signal/noise ratio is likely to be decreased. The approach adopted at this work was to sum all rings from a determined point on, forming the final ring sum of each layer. All ring sums (58 in total for each RoI) are fed into the neural network input nodes, for

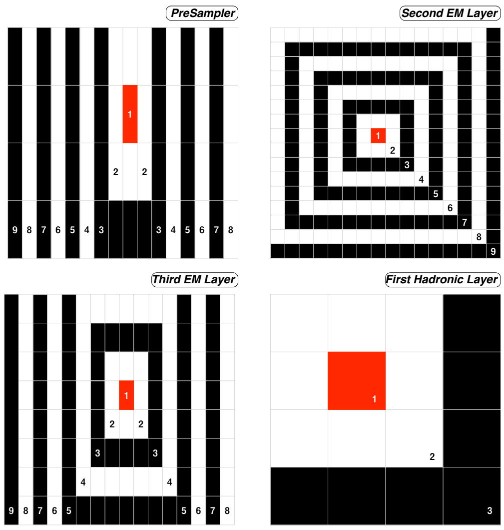

Fig. 6. Building rings with the cells' deposited energies.

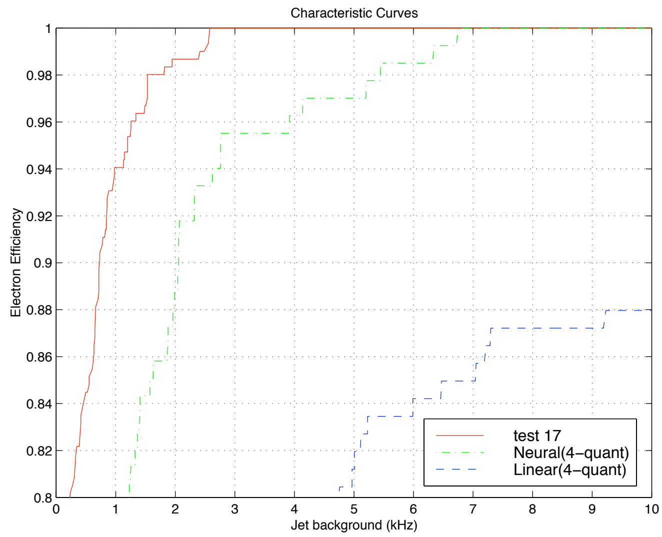

Fig. 7. The RoC for the Electron/Jet discriminator proposed using "square" rings and neural networks (marked with *Test 17*).

performing Electron/Jet discrimination. Normalization of input vectors is performed by using the total RoI energy of all sections included at the RoI for each ring sum [?].

The results applying this compression strategy and normalization scheme can be seen at Figure 7. This result was achieved by using the same strategy for neural training as explained in section III-A. As one can observe, this compression strategy performs a lot better than the one adopted at CERN. The neural network used at this run had only 5 hidden neurons. This result suggests that further compression could be achieved without compromising the discriminator performance.

C.1 Input Relevance

As speed requirements are very tight at T2, a further compactation of the input data space of the neural network was developed. For this, the relevance of each input node

TABLE I
VARIABLE DESCRIPTION FOR THE CHART IN FIGURE 8.

Layer	Chart Position
Pre-sampler	from 1 to 6
First e.m. layer	from 7 to 39
Second e.m. layer	from 40 to 49
First hadronic layer	from 50 to 52
Second hadronic layer	from 53 to 55
Third hadronic layer	from 56 to 58

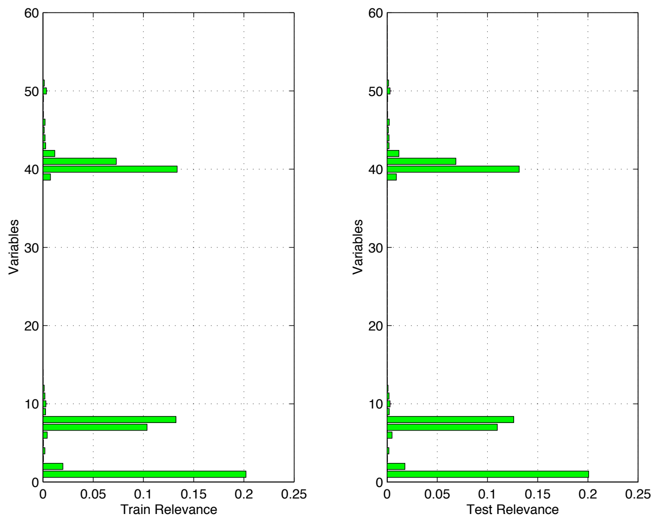

Fig. 8. The relevance values for each input variable for a trained neural network that performs Electron/Jet separation.

is measured [7] at the end of the training phase. This procedure is carried out using the following consideration:

$$R_i = \frac{1}{N}\sum_{j=1}^{N}[output(\overrightarrow{x_j}) - output(\overrightarrow{x_j}|_{x_{j,i}=\overline{x_i}})]^2 \qquad (1)$$

The relevance of each input is the mean value of the squared error computed over all training samples when one substitutes the actual input value by its mean. This relevance mapping allows to measure the impact to the output of the neural network when instantaneous values from a variable are clamped to its mean value. If the neural network output is significantly sensitive for a given variable, its relevance shall be also great, else, this input variable may be considered unimportant to the discrimination task and may be even ignored. Figure 8 shows an example bar chart of the relevance for a 58-input neural discriminator already trained. As one can observe, not all variables seem to be relevant to the discrimination task. The variables shown at the chart are enumerated in Table I.

By applying cuts over the relevance, one may further compact the input space by only retaining input variables that pass the cut. Feeding the input nodes of a new neural classifier with such restricted number of qualified input

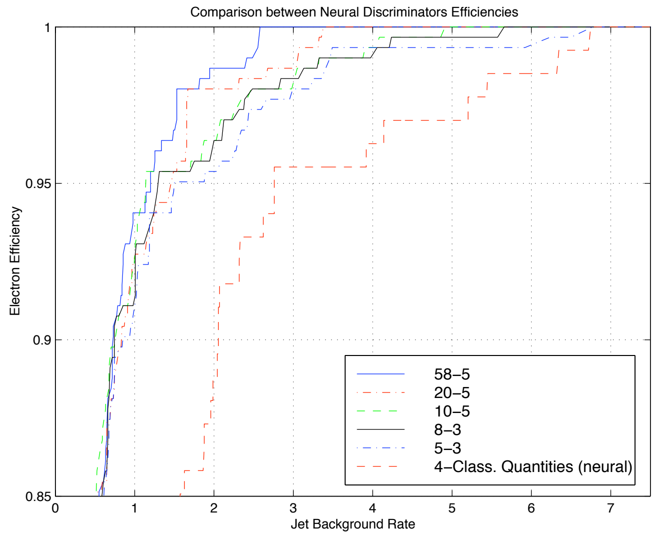

Fig. 9. The RoC for different discriminators (see text). The false alarm for this plot is in kHz.

variables, it may be expected that the discrimination efficiency is not significantly decreased after training. Tests have been conducted cutting from 38 to 53 of the 58 inputs. Figure 9 resumes the results. It shows the RoC of the best discriminators found and the network topology in terms of the number of input nodes (surviving variables) and hidden nodes. All network topologies have a single output neuron.

D. Timing Issues

Due to the timing constraints [2], [8] for T2's computer network operation and T1 acceptance rate (100,000 Hz), each event shall be analyzed in less than 10 ms. This gives no more than 2 ms for RoI processing, as in average, 5 RoI's are produced per event.

After implementing in C language all phases of T2's local processing, ending with electron/jet neural discrimination, it was achieved, with optimized code on a PentiumII@300MHz, 13 ms of total application time, per RoI[2]. This figure does not meet the second level trigger requirements.

Taking a closer look at the algorithm, one can realize that the timing distribution for each separate task of the processing goes as follows: 87% (11.31 ms) for data acquisition, 7% (0.91 ms) for integrity checks, 5.8% (0.75 ms) of the signal processing time is taken for ring extraction and finally, only 0.2% (26 μs) of the time is spent at the neural processing. The first two phases of data processing are common to every algorithm and, therefore, should be the same no matter neural discrimination is performed or not. Around 6% of the initial 13 ms ($\sim 780\mu$s) is spent in particle discrimination.

The implementation of the same algorithm in a ADSP-21062@40MHz of Analog Devices ([9]) shows that the timings could be reduced about 13 times, ending with a 1 ms

[2]This test was performed considering all 58 input rings and with 5 hidden neurons at the classifier

per RoI implementation, still using ADSP's native C compiler, with no optimizations. That figure is far better than the first one and make it feasible an implementation at T2, using such DSP's as co-processing nodes. DSP's are cheap, reliable and, as one can see, very fast, making it a very interesting choice for T2's processing network. Taking the percentage in time for each of the phases for local processing, data acquisition would take 870μs, integrity checks, 70μs, ring extraction 59μs and neural discrimination only 1μs using the DSP implementation.

IV. Conclusions and Outlook

At this work, a neural classification technique for calorimeter data, based on neural networks and topology studies was developed. The technique was applied to the Electron/Jet discrimination problem, present at the Second Level Trigger (T2) of ATLAS.

The neural classifier, besides being faster, is also much more efficient for T2, as could be plotted in Figure 9. That plot also showed that the amount of data needed to perform classification (number of rings) could be reduced (up to 5), still with gain compared to the algorithm proposed at CERN. While tests have not been conducted to prove that there shall be reduction at the overall processing time for electron/jet classification at T2, it's clear from the context that such major change would produce significant impacts to that, also saving ATLAS resources.

Being this the first time the whole local processing algorithm is fully implemented, there was no comparison figures to these results. Despite of that, [10] showed that the CERN equivalent for the last processing phase of electron/jet classification (neural processing) can be performed in around 50μs, also using a PII@300MHz. That number is almost twice the one for neural classification (26μs).

The redundancy (increased number of rings) could be added, by considering more data to be used, making the system more robust, while also increasing the algorithm performance. Finally, the DSP implementation showed that only the thirteenth part of the initial time estimative would be need to implement the whole classification structure. That is far better than using conventional processing (with PC's) and is a realistic choice for T2.

The next step in this work will be the implementation of the whole algorithm at the CERN prototype for T2. It intends to confirm the expectations for neural processing at that subsystem of ATLAS.

V. Acknowledgements

The authors are thankful to CAPES and CNPq, the brazilian funding agencies, for their support to this research. They are also thankful to CERN and, in particular, to Saul Gonzalez, who help with fruitful discussions and provided data sets for this experiment.

References

[1] ATLAS Collaboration, "Trigger performance technical progress report," Tech. Rep., CERN, 1999.

[2] ATLAS Collaboration, "ATLAS daq, ef, lvl2 and dcs technical progress report," Tech. Rep., CERN, 1998.

[3] R. Wigmans, "Advances in hadron calorimetry," *Submitted to the Annual Review of Nuclear and Particle Science*, 1991.

[4] Simon Haykin, *Neural Networks: A comprehensive foundantion*, Prentice Hall, 1999.

[5] H.L. Van Trees, *Detection, Estimation and Modulation*, John Wiley & Sons, 1966.

[6] J.M. Seixas and I.F. Isidro, "A neural electron/jet discriminator based on a highly-segmented calorimeters," *Artificial Intelligence in High-Energy and Nuclear Physics (AIHENP)*, 1999.

[7] A. Gruber, J. Möck, J. Fröchtenicht, D. Goldner, C. Kiesling, T.Klober, H. Kolanoski, T. Krämerkämper, and P. Ribarics, "Performance of backpropagation mentworks in the second level trigger of the h1-experiment," *New Computing Techniques in Physics Research*, 1994.

[8] T. Shears S. Gonzalez, B. Gonzales-Piñero, "First implementation of calorimeter fex algorithms in the level-2 reference software," Lvl2 testbed technical note, CERN, Maio 1999.

[9] Analog Devices, *ADSP-2106x Sharc User's Manual*, secondedition edition.

[10] S. Gonzalez, T. Hansl-Kozanecka, and M. Wielers, "Selection of high-pt electromagnetic clusters by the level-2 trigger of ATLAS," ATLAS-Trigger DAq Note 002, CERN, Fevereiro 2000.