Self-organized mapping of calorimetry information for high efficient online electron/jet identification in ATLAS

Eduardo Simas Filho, Jose M. Seixas[*]**, Luiz Caloba**

Signal Processing Laboratory, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
E-mail: `esimas@lps.ufrj.br, seixas@lps.ufrj.br, caloba@lps.ufrj.br`

In this work a self-organizing map (SOM) is applied for electron/jet identification in the high-level trigger of ATLAS detector, using calorimeter data mapped over Regions of Interest (RoI). For feature extraction the SOM is unsupervised trained and the resulting map is slightly modified through Learning Vector Quantization (LVQ) algorithm for optimal signal identification. In practice, calorimeter designs exhibit nonlinearities so a nonlinear feature extraction technique, like SOM, is attractive. The hypothesis testing is performed by a neural classifier implemented through a Multi-layer Perceptron (MLP) architecture. Using the proposed technique for an electron/jet discrimination in the second level trigger of ATLAS, 98% of electrons are correctly identified for a misclassification of jets below 3%, which outperforms the baseline algorithm presently used in ATLAS.

---

[*]Speaker.

1. Introduction

The Large Hadron Collider (LHC) is under construction at CERN [1], and will be colliding bunches of protons at every 25*ns* by the year of 2008. In order to search for the Higgs particle and other new phenomena a new energy range will be explored by LHC. When operating at full capacity (with a luminosity of $10^{34}cm^{-2}s^{-1}$) LHC will produce 40×10^6 events/s. The total detector information per second will be near 60 *Tbytes* (1,5 *MB*/event). Considering this amount of data, the filtering procedure must be performed online, under short latency times. Despite the very high event rate, the interesting channels will rarely occur, so that the design of an efficient triggering system is very important to guarantee that valuable information will not be lost.
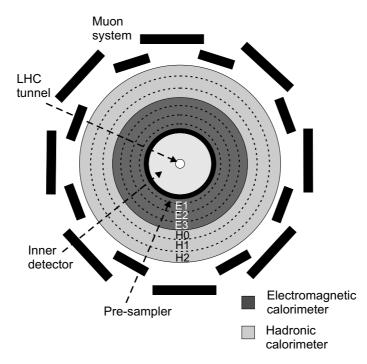

Figure 1: The ATLAS detector diagram.

Placed at one of the LHC collision points, the ATLAS detector has a cylindric format with the LHC tunnel as central axis [2]. One important information that guides the particle identification process is the energy deposition profile measured at the calorimeters. As illustrated in Figure 1, the ATLAS calorimeter system is split into electromagnetic (e.m.) and hadronic sections and has seven layers, four e.m. and three hadronic. The granularity, varies for the different calorimeter layers, see Table 1. The pseudo-rapidity ($\eta = -\log\tan(\theta/2)$, where $\theta = \arctan(x/z)$) and azimuthal angle ($\phi = \arctan(x/y)$) are components of the ATLAS coordinates system, as illustrated in Figure 2.

The ATLAS online trigger system has three sequential filtering levels as shown in Figure 3, the complete event data is available until an accept/reject decision is made. The detector information is obtained from the calorimeters, muon chambers and tracking (inner detector) systems (see Figure 1). The first-level (LVL1) receives the full LHC event rate and shall reduce it to 75*kHz*. The decision must be taken in less than 2.5*μs*. Calorimeter data with reduced granularity and hardware implementation are used in LVL1 fast decision taking process. The High-Level Trigger (HLT),

Table 1: Different granularity of the calorimeter layers [2].

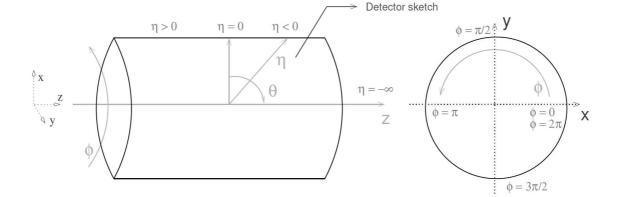| PRESAMPLER	Barrel	Endcap
Granularity ($\Delta\eta \times \Delta\phi$)	$0,025 \times 0,1$	$0,025 \times 0,1$
ELECTROMAGNETIC	**Barrel**	**Endcap**
Granularity ($\Delta\eta \times \Delta\phi$)		
Layer 1	$0,003 \times 0,1$	$0,003 \times 0,1$ a $0,1 \times 0,1$
Layer 2	$0,025 \times 0,025$	$0,025 \times 0,025$ a $0,1 \times 0,1$
Layer 3	$0,05 \times 0,025$	$0,05 \times 0,025$
HADRONIC	**Barrel**	**Extended barrel**
Granularity ($\Delta\eta \times \Delta\phi$)		
Layers 1 e 2	$0,1 \times 0,1$	$0,1 \times 0,1$
Layer 3	$0,2 \times 0,1$	$0,2 \times 0,1$

Figure 2: ATLAS coordinates system [3].

which comprises both the second-level (LVL2) and the event filter (third-level), receives from the first-level the primary location of regions in the detector where interesting information was identified. These detector regions are known as Regions of Interest (RoI). The HLT is implemented in software, processed in parallel by thousands of PCs, and must reduce the event rate to 100 *Hz*. The overall processing time is about one second and LVL2 latency shall not exceed 10ms.

Among interesting objects for LHC, the most frequent for searching the Higgs particle are electrons, photons and muons [2]. Electrons are immerse in an intense background noise of jets. In terms of calorimetry, jets that arrive to the LVL2 trigger tend to fake electron signatures (highly concentrated energy deposition profiles in the e.m. section and almost no energy left in the hadronic calorimeter). Therefore, the electron/jet channel is very important for the detector overall performance.

As proposed on [5], here the calorimeter RoI data are pre-processed and formatted into concentric rings before being used for particle identification. As illustrated on Figure 4, at each calorimeter layer, the most energetic cell is defined as the first ring, and the following rings are formed around this cell, making the classification procedure independent of the impact point. The energy measure-
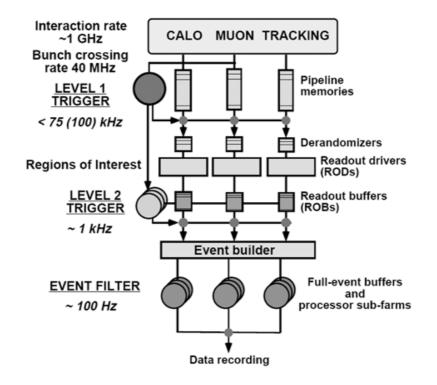
Figure 3: ATLAS trigger system architecture, extracted from [4].

ments that fells belonging to each ring are summed up and normalized. Examples of ring formatted signals for typical electron and jet are illustrated on Figures 5 and 6, the calorimeter layers are limited by vertical dotted lines. For each event, the RoI size may vary, but for particle discrimination purposes a fixed size 0.4×0.4 area in the $\eta \times \phi$ plane carries enough physics. Considering this fixed RoI, the ring formatted signals comprise 100 components.
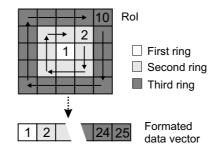

Figure 4: Ring formatting of RoI cells.

This paper proposes a new particle discrimination strategy for ATLAS calorimeter based trigger. For feature extraction, a Self-organizing Map (SOM) [6] is applied. The topologically organized mapping is further adjusted through a Learning Vector Quantization (LVQ) [7]: a vector basis is formed by the neuron weights, and the LVQ algorithm uses the target information to move these vectors to new positions, maximizing the classification performance. After the feature extraction procedure, a neural classifier, through a Multi-layer Perceptron (MLP) architecture [8], performs
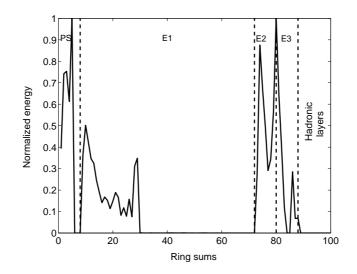
Figure 5: Ring formatted signals for typical electron, E_i represents an e.m. layer.
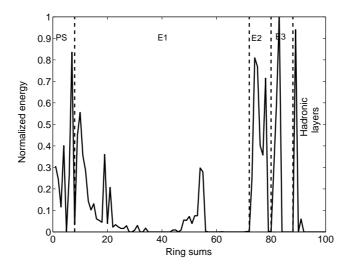

Figure 6: Ring formatted signals for typical jet, E_i represents an e.m. layer.

the hypothesis testing using as input information the mapping obtained by the SOM network after LVQ adjustment. The algorithm overall processing time cannot exceed the limitations imposed for the HLT.

On a previous work [5], a neural classifier was applied directly to the ring formatted signals and good results were obtained on electron/jet classification. On the present work, using the SOM feature extraction a considerable improvement on the discrimination performance was achieved, with a small increase on the computational cost.

2. Feature extraction

A method that combines unsupervised and supervised training routines was applied for feature extraction on the ring formatted calorimeter signals. Self Organized Maps were trained without

knowledge of the particle type and further adjusted, on a supervised way, through the Learning Vector Quantization algorithm. A MLP classifier uses the LVQ adjusted mapping to perform particle discrimination. Figure 7 describes the proposed technique.
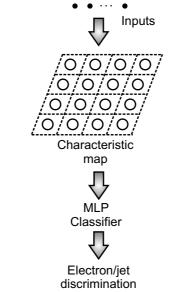

Figure 7: Proposed procedure for particle discrimination.

2.1 Self organizing maps

The Self Organizing Map is an unsupervised trained neural network that realizes a topological organization of the input data set. The SOM were first proposed by Teuvo Kohonen in 1982 [6], and transforms a k-dimensional continuous input space into a discrete characteristics map (generally bidimensional). Each neuron of the map is fully connected to all inputs. SOM compacts information while preserving topological relations of the input data set. SOM networks are widely applied on different signal processing tasks like fault diagnosis, computer vision, statistical pattern recognition, speech processing, noise cancelation [9] and nonlinear independent component analysis [10, 11].

Basically three processes are involved in the SOM training phase: competition (for each input vector there is only one activated neuron), cooperation (the winner neuron determines a neighborhood of excited neurons) and adaption (the map weights are adjusted to reenforce the winner neuron response to a given input pattern) [8].

The inputs are fully connected to each neuron, and the map weights are calculated iteratively through:

$$w_j(n+1) = w_j(n) + \eta(n)h_{ij}(n)(x(n) - w_j(n)) \tag{2.1}$$

where $\eta(n)$ is the learning rate and $h_{ij}(n)$ the neighborhood function, which is defined as:

$$h_{ij}(n) = \exp(-d_{ij}^2/2\sigma^2(n)) \tag{2.2}$$

After the training procedure has converged, the network outputs are obtained for each neuron i through equation 2.3:

$$u_i = \mathbf{x}^T \mathbf{w}_i \tag{2.3}$$

where $\mathbf{w}_i$ is a vector formed by the weights w_{ij} that connect the input x_j to the neuron i. Self organizing maps belong to a class of algorithms of vector quantization, searching for a fixed number of vectors (or code-words) that better describe the input data set. The vectors $\mathbf{w}_i$ form the SOM code-book for the input data set.

For a specific input vector $\mathbf{x_a}$, the highest value of u is obtained for the neuron i that minimizes the distance $|\mathbf{x_a} - \mathbf{w}_i|$ [12], in other words, $\mathbf{w}_i$ is the code-word closest to $\mathbf{x_a}$.

If a winner-takes-all operation of the SOM is preferred, the outputs may now be defined as:

$$\begin{aligned} y_i &= 1, \ u_i > u_j \forall j \neq i \\ y_i &= 0, \quad \text{other case} \end{aligned} \tag{2.4}$$

Considering equation 2.4, the input vector $\mathbf{x_a}$ is mapped into the SOM neuron that better describes it. As the map is topologically ordered, similar input patterns are mapped into neighbor regions.

When applied to classification problems, it is expected that the SOM concentrates the different inputs in opposite sides of the map, facilitating the hypothesis testing procedure.

2.2 Learning vector quantization

Vector quantization (VQ) is a coding technique in which a input data set is mapped into a finite group of representative vectors. [13]. The k-dimensional input space is divided into a finite number of regions and the quantizer maps $\mathbb{R}^k$ into a finite subset Y of $\mathbb{R}^k$:

$$Q : \mathbb{R}^k \to \mathbf{Y} \tag{2.5}$$

where $\mathbf{Y} = \{y_1, y_2, ..., y_k\}$ is the code-book. For each code-word y_i there is a partition R_i of the input space that satisfies:

$$R_i = Q^{-1}(\mathbf{y_i}) = \{\mathbf{x} \in \mathbb{R}^k : Q(\mathbf{x}) = \mathbf{y}_i\} \tag{2.6}$$

$$\bigcup_{i=1}^{N} R_i = \mathbb{R}^k, \ R_i \bigcap R_j = 0, \ i \neq j \tag{2.7}$$

When a quantizer presents minimal distortion it is called **Voronoi quantizer**. In this case the input space is partitioned according to the nearest neighbor (NN) rule, and its partitions are denominated Voronoi cells [14].

The code-words (or Voronoi vectors) can be approximately estimated by the SOM through unsupervised learning procedure and the code-book is formed by the synaptic weights. For classification purposes, the Learning Vector Quantization (LVQ) algorithm [7] can be used to improve the discrimination performance. Through a supervised procedure (using target information), LVQ slightly adjusts the code-words location (obtained by the SOM) to improve the map clustering.

The LVQ algorithm, as formulated on [7], selects randomly a input vector $\mathbf{x}$ and verifies whether or not its class type $\mathscr{C}_{\mathbf{x_i}}$ is the same as the one the Voronoi vector $\mathbf{w_c}$ belongs to. In case both classes are the same, $\mathbf{w_c}$ is moved towards $\mathbf{x}$:

$$\mathscr{C}_{\mathbf{w_c}} = \mathscr{C}_{\mathbf{x_i}} \rightarrow \mathbf{w_c}(n+1) = \mathbf{w_c}(n) + \alpha[\mathbf{x} - \mathbf{w_c}(n)] \tag{2.8}$$

where α is the learning rate ($0 < \alpha < 1$). Otherwise, $\mathbf{w_c}$ is moved away from $\mathbf{x}$:

$$\mathscr{C}_{\mathbf{w_c}} \neq \mathscr{C}_{\mathbf{x_i}} \rightarrow \mathbf{w_c}(n+1) = \mathbf{w_c}(n) - \alpha[\mathbf{x} - \mathbf{w_c}(n)] \tag{2.9}$$

Some examples of LVQ application for signal compaction and classification can be found in [7, 15].

3. Hypothesis testing

The map outputs, after LVQ adjustment, were used to feed a Multi-layer Perceptron (MLP) neural classifier [8], trained through the resilient back-propagation (RPROP) algorithm [16]. The network used a single hidden layer and one neuron in the output. The number of hidden neurons was chosen after testing exhaustively the discrimination performance of each network. A network with four hidden neurons presented better results and was used for the obtained maps.

The particle discrimination process should be energy independent, so an energy normalization was performed on input vectors. Here, data were normalized in a segmented way during the ring formatting procedure.
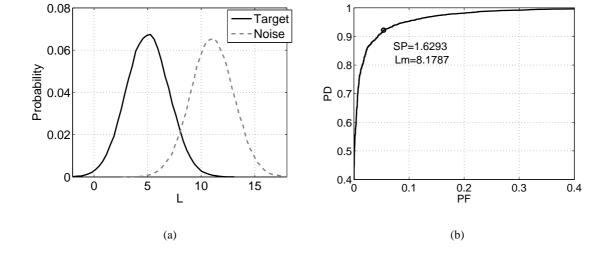
4. Results

(a)   (b)

Figure 8: Examples of (a) classes probability distribution and respective (b) ROC curve.

The Receiver Operating Characteristic (ROC) and the SP product were both used as figures of merit of the particle discrimination performance. Considering a binary decision problem, where the classes distribution are shown in Figure 8(a), the respective ROC curve [17] (see Figure 8(b)) illustrates how both the detection (P_D) and false alarm (P_F) probabilities vary as the decision threshold

(L) changes. For the problem illustrated in Figure 8(a), if the probability distribution function of the target class is defined as: $P_T(L)$, then $PD = \int_{-\infty}^{L_m} P_T(L)dL$, and $PF = \int_{L_m}^{\infty} P_T(L)dL$, where L_m is the chosen threshold.
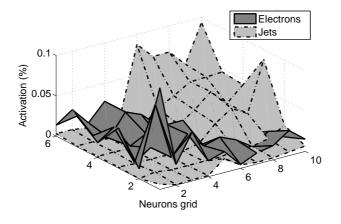
In our particular problem, electron signature represents the target signal to be detected and jet corresponds to false alarm. As the interesting events are very rare at LHC, high P_D is desired for the online triggering operation. Low P_F is also essential for the classifier design, as the huge background noise has to be rejected, as much as possible, to allow offline data analysis on clean data.

The SP product is defined as [5]:

$$(Ef_e + Ef_j) \times (Ef_e \times Ef_j) \tag{4.1}$$

where Ef_e is the detection efficiency for electrons and Ef_j is the corresponding efficiency for jets. The maximum SP is 2, when $Ef_e = Ef_j = 1$, and to obtain high SP values the efficiencies of both classes must be close to 1. As the SP collapses for either low P_D or high P_F (low Ef_j), this figure of merit is valuable for our application. The maximum SP ($SP \approx 1.63$) an the respective threshold value ($L_m \approx 8.18$) for the classes of Figure 8(a) are shown in Figure 8(b).

The database used in this work, which comprises 22581 electron and 7509 jet signatures, was obtained trough Monte Carlo simulation for proton-proton collisions. The detector characteristics and the first level trigger effects were considered during the simulations [2].

Figure 9: Activation probability for electrons and jets on a 6x10 neuron map after unsupervised training.

The database was divided, in approximately equal parts, into training and testing sets. The training set was used to develop and train the algorithms, and the testing events were applied to verify the proposed method performance. Two maps (6x10 and 8x10 neurons) were tested. Considering a winner-takes-all operation of the SOM (see equation 2.4), as illustrated on Figure 9 (for the 6x10 map), the SOM was able to concentrate the particle types on opposite sides of the map, and moreover, the electron signatures projections are confined in a thinner region. The 8×10 map showed similar behavior.

Using the LVQ algorithm to further adjust the map weights, by means of supervised training, the SOM outputs were shifted towards the map edges (see Figure 10). Therefore the border between the mapped regions assigned to electrons and jets was reduced as a result of LVQ.
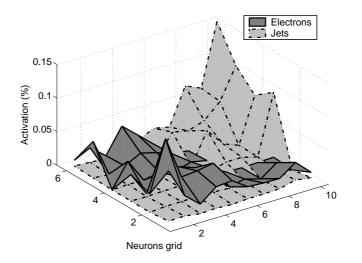
Figure 10: Activation probability for electrons and jets on a 6x10 neuron map after LVQ adjustment.

Better particle discrimination performance was achieved using equation 2.3 to compute the Map outputs. Through this approach, for each ring formatted event, a $N \times M$ mapping is obtained by the SOM and used to feed a MLP neural classifier. Through the proposed method 98% of electrons are correctly identified for a misclassification of jets below 3%. The better calculated SP product was 1.86 (for a 6×10 map).
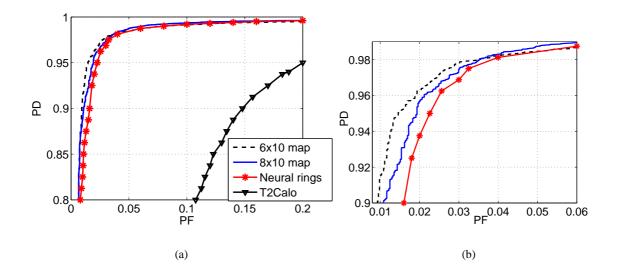

Figure 11: (a) ROC curves for the proposed technique (6x10 and 8x10 maps), Neural_Ringer and T2Calo, (b) detail for better visualization.

Different e^-/j discriminators are compared using this simulated second-level trigger database (testing set). The baseline algorithm for e^-/j discrimination used at ATLAS (T2Calo) [4] extracts, directly from calorimeter measurements, parameters that estimate the shape of the energy deposition profile. Thresholds on these parameters perform the particle discrimination. The

Discrimination Technique	SP	PD (%)	PF (%)	PD (%)	PF (%)	False electrons (jets / second)
T2Calo	1.4	81.1	3	95	20	5000
Neural_Ringer	1.82	96.8	3	95	2.2	550
8x10 map	1.84	97.5	3	95	1.8	450
6x10 map	1.86	98.2	3	95	1.4	350

Table 2: Comparison of detection (PD) and false alarm (PF) probabilities and SP products obtained through different discrimination techniques.

Neural_Ringer [5] is another particle discrimination procedure that is implemented in the ATLAS software platform (ATHENA). Using a MLP neural classifier operating over the ring formatted signals, Neural_Ringer algorithm achieved better discrimination performance and similar computational cost when compared with T2Calo. In Figure 11(a), the performance of the proposed method is compared to T2Calo and Neural_Ringer discriminators.

Analyzing the ROC curves (Figures 11(a) and 11(b)), T2Calo presents the worst performance. The feature extraction through SOM maps improves slightly the discrimination performance of the neural classifier fed from ring sum signals.

Table 2 compares the detection probability (PD) for a fixed false alarm probability (3%) and the maximum SP products obtained from the different discrimination techniques, it is also shown PF values for a fixed PD=95%. The proposed approach is able to increase the detection efficiency on 1.4 and 17.1 percent points, when compared, respectively, to Neural_Ringer and T2Calo. Considering LHC operating conditions at high luminosity, 25000 jets/second are expected at ATLAS high-level trigger [4]. It means that, for a fixed PD, 1% increase of PF implies on recording more 250 false electrons per second. As shown in Table 2, for PD=95%, the proposed technique reduces the false alarm (less 200 and 4650 jets/sec when compared, respectively, to Neural_Ringer and T2Calo), providing cleaner data for offline analysis.

Electron efficiency (E_{ef}) and jet false alarm (J_{FA}) of the proposed method are compared for different values of η and ϕ in Figures 12(a) and 12(b) for a 6×10 map (similar behavior was obtained using the 8×10 map). Considering different values of η (Figure 12(a)), E_{ef} falls for $\eta > 1.5$ due to a gap on calorimeter sensing elements that exists in this region. Through this gap pass maintenance and communication cables to the inner detector. J_{FA} does not appear to be significantly affected by the calorimeter gap. Both E_{ef} and J_{FA} are uniformly distributed in ϕ (Figure 12(b)), this result was expected as the detector is symmetric in this axis.

5. Conclusions

A novel signal processing procedure was proposed for electron/jet discrimination at ATLAS second-level trigger. Self-organizing maps were used for feature extraction, and MLP classifiers performed the hypothesis testing. Using this technique 98% of electrons were correctly identified for a misclassification of jets below 3%. This outperforms the current baseline design (T2Calo Algorithm). The good performance points out that the combination of ring structured data formatting and nonlinear signal processing technique is very efficient for electron/jet discrimination.
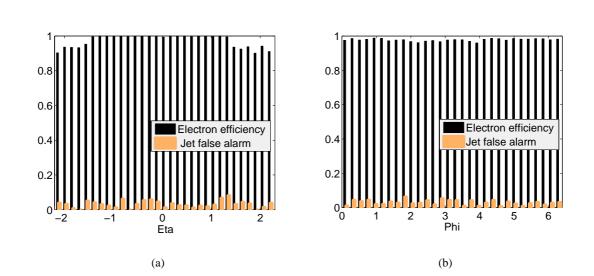
(a)   (b)

Figure 12: Electron efficiency and jet false alarm for different values of (a) η and (b) ϕ (6×10 map).

A proper study of the processing time is needed for the proposed algorithm, but a simple comparison with the Neural_Ringer discriminator indicates that the HLT requirements may be satisfied. On the work [5], it was demonstrated that the ring formatting and neural classification algorithms, implemented at the HLT software platform (ATHENA) are faster than T2Calo. The proposed technique adds only a matrix multiplication to the neural rings, and so, no significant increase on computational cost is expected.

On this work, energy measurements from different calorimeter layers were combined to compose a single input signal used for particle identification. A segmented feature extraction, as proposed on [18], where the energy measurements from each calorimeter layer are processed independently, preserves the different characteristics of each calorimeter layer and usually presents better discrimination performance. A study is been conduced to design and train a different SOM for feature extraction at each layer.

Acknowledgements

References

[1] M.J. Price, *The LHC Project*, Nuclear Instruments and Methods in Physics Research, 2002.

[2] ATLAS-Collaboration, *ATLAS Detector and Physics Performance Technical Design Report, Volume 1*, CERN, May 1999.

[3] A. Anjos, *Online Filtering System Operating on a High Event Rate Environment*, PhD Thesis, Federal University of Rio de Janeiro, Rio de Janeiro-Brazil, 2006.

[4] ATLAS-Collaboration, *ATLAS High-Level Trigger, Data Acquisition and Controls Technical Design Report*, CERN May, 2003.

[5] A. Anjos, R.C. Torres, J.M. Seixas, B.C. Ferreira, T.C. Xavier, *Neural Triggering System Operating on High Resolution Calorimetry Information*, Nuclear Instruments and Methods in Physics Research, No. 559, pp. 134-138, 2005.

[6] T. Kohonen, *The Self-Organizing Map*, Proceedings of the IEEE, Vol. 78, No. 9, pp. 1464-1480, September 1990.

[7] T. Kohonen, *Improved Versions of Learning Vector Quantization*, International Joint Conference on Neural Networks, Vol. 1, pp. 545-550, 1990.

[8] S. Haykin, *Neural Networks, Principles and Practice*, Bookman, 2001.

[9] T. Kohonen, E. Oja, O. Simula, A. Visa, J. Kangas, *Engineering Applications of the Self-Organizing Map*, Proceedings of the IEEE, Vol. 84, No. 10, pp. 1358-1384, October 1996.

[10] P. Pajunen, A. Hyvarinen, J. Karhunen, *Nonlinear Blind Source Separation by Self-organizing Maps*, Proceedings of the International Conference on Neural Information Processing, pp. 1207-1210, 1996.

[11] M. Haritopoulos, H. Yin, N. M. Allinson, *Image Denoising using Self-organizing Map-based Nonlinear Independent Component Analysis*, Neural Networks Journal, Vol.15, pp. 1085-1098, 2002.

[12] T. Kohonen, *Self Organizing Maps*, Springer-Verlag, 3rd Ed., 2001.

[13] A. Gersho, *On the Structure of Vector Quantizers*, IEEE Transactions on Information Theory, Vol. IT-28, No.2, pp. 157-166, March 1982.

[14] R. M. Gray, *Vector Quantization*, IEEE ASSP Magazine, pp. 4-29, 1984.

[15] J. S. Baras, S. Dey, *Combined Compression and Classification with Learning Vector Quantization*, IEEE Transactions on Information Theory, Vol.45, No. 6, pp. 1911-1920, 1999.

[16] M. Riedmiller, H. Braun, *A Direct Adaptive Method for Faster Backpropagation Learning, the RPROP algorithm*, Proceedings of the IEEE International Conference on Neural Networks, pp. 586-591, 1993.

[17] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, John Wiley, 2001.

[18] E. Simas, L. Caloba, J. M. Seixas, *Segmented Independent Component Analysis for Online Filtering Using Highly Segmented Detectors*, Proceedings of the 7th International Conference on Intelligent Systems Design and Applications - ISDA07, October 2007 (to be presented).

PoS(ACAT)055