

# Optimized Calorimeter Signal Compaction for an Independent Component based ATLAS Electron/Jet Second-Level Trigger

---

**Eduardo F. Simas Filho<sup>1,2</sup>, José M. Seixas<sup>\*1</sup> and Luiz P. Calôba<sup>1</sup>**

<sup>1</sup>*Signal Processing Laboratory, COPPE/Poli, Federal University of Rio de Janeiro, Brazil.*

<sup>2</sup>*Federal Institute of Education, Science and Technology, Simões Filho, Bahia, Brazil.*

*E-mail: esimas@lps.ufrj.br, seixas@lps.ufrj.br, caloba@lps.ufrj.br*

The ATLAS online trigger system has three filtering levels and relies very much on calorimeter information, which is segmented into seven detection layers. Due to differences both in depth and cell granularity of these layers, trigger algorithms may benefit from performing feature extraction at the layer level. This work addresses electron/jet separation at the second level (LVL2) filtering restricted to calorimeter data. Segmented Independent Component Analysis (SICA) is applied over the calorimeter layers in order to extract relevant features for particle identification. The number of independent components to be extracted from a Region of Interest (RoI) is estimated through different signal compaction strategies, such as Principal Component Analysis, Nonlinear Principal Component Analysis and Principal Components for Discrimination. These compaction techniques are evaluated with respect to dimensionality reduction (and processing speed) and classification efficiency. The hypothesis testing is performed by a Multi-Layer Perceptron classifier fed from the segmented independent components. It is shown that the proposed discriminators outperform the baseline design for ATLAS second-level trigger system, achieving a detection efficiency of 99% for a rejection factor smaller than 2%.

*XII Advanced Computing and Analysis Techniques in Physics Research  
November 3-7 2008  
Erice, Italy*

---

<sup>\*</sup>Speaker.

### 1. Introduction

The Large Hadron Collider (LHC) is the most powerful particle accelerator ever built [1]. In the search for new physics channels, the LHC will combine high energy levels (up to 14 TeV) and high interaction rate (1 GHz when operating at high luminosity). Bunches of protons will be accelerated and collisions are expected to occur at every 25 ns, generating a huge amount of information. However, only a small fraction of these data will contain relevant information.

The ATLAS detector [2] is one of the general purpose experiments of the LHC. Considering the LHC interaction rates and detector segmentation, 52 Tbytes/s will be produced (1.3 Mbyte per collision). As it is impossible to record all detector data, an efficient online filtering system is required to reject the background noise as much as possible while retaining most of the relevant signatures. In ATLAS, the filtering (triggering) is performed through three cascaded filtering levels [3].

For ATLAS, electron identification is a very important task, as Higgs signatures can be found efficiently through decays that produce electrons [2]. The calorimeters provide an important information for electron identification, as they measure the energy deposition profiles for incoming particles. The calorimeter system (see Figure 1-a) is segmented into four electromagnetic (PS, E1, E2 and E3) and three hadronic (H0, H1 and H2) layers, producing more than 100,000 readout channels. Typical electrons present energy deposition profiles highly concentrated in the electromagnetic sections and no energy left in the hadronic layers. A huge background noise for electron identification arises due to the existence of a class of hadronic jets which present energy deposition profiles very similar to typical electrons.

In this work, the Independent Component Analysis (ICA) model [4] is applied to extract, from ATLAS calorimeter signals, relevant features for electron identification. As a pre-processing step, raw data from the calorimeter is formatted into concentric rings and, within each ring, the energy is summed and normalized. The proposed method is applied at the second-level trigger (LVL2).

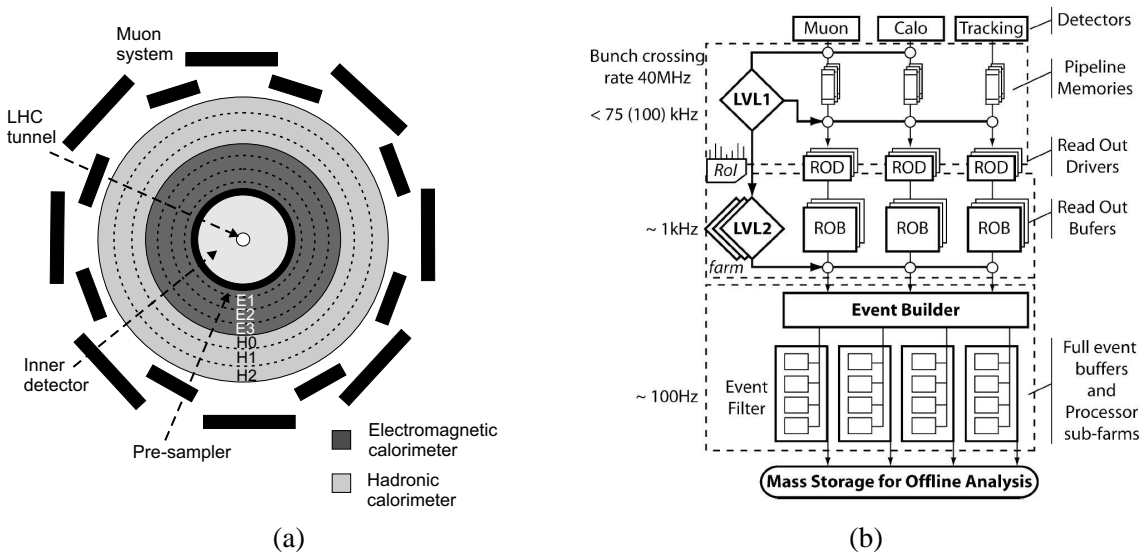


Figure 1: ATLAS (a) sub-detectors and (b) triggering system architecture.

## 2. ATLAS Trigger and Calorimetry Overview

The ATLAS filtering (trigger) system [3] comprises three sequential decision levels (see Figure 1-b), in which the amount of non-relevant signals retained in temporary memories is gradually reduced until the final storage/rejection decision is taken. The first-level is implemented in hardware, as a very fast decision (within 2.5 ms) is required. The second-level (LVL2) and the event filter (EF) are both implemented in software and processed in parallel by thousand of PC-like processors [3]. The latencies for LVL2 and EF are, respectively, 40 ms and 4 s.

The LVL1 uses detector information with coarse-grained granularity to identify the Regions of Interest - RoI (detector regions where relevant interactions may have occurred). The LVL2 receives from the LVL1 RoI location and uses full calorimeter granularity to produce an accept/reject decision for each incoming event. The particle discrimination procedure at LVL2 is split into two phases: feature extraction, where relevant information is extracted from the measured signals, and hypothesis testing, where particle discrimination is performed.

The baseline electron/jet discriminator at LVL2 (T2Calo) is based on classical statistical approach. Using calorimeter information, parameters from the energy shower are computed and the comparison with thresholds (previously determined considering typical characteristics of electrons and jets) allows particle identification.

Using a neural classifier fed from ring structured calorimeter data (ring construction process will be described in the next section), an alternative electron/jet LVL2 trigger algorithm (Neural-Ringer) was proposed in [5]. The Neural-Ringer is already operational at the HLT software platform (ATHENA) and, if compared to T2Calo, presents higher discrimination efficiency.

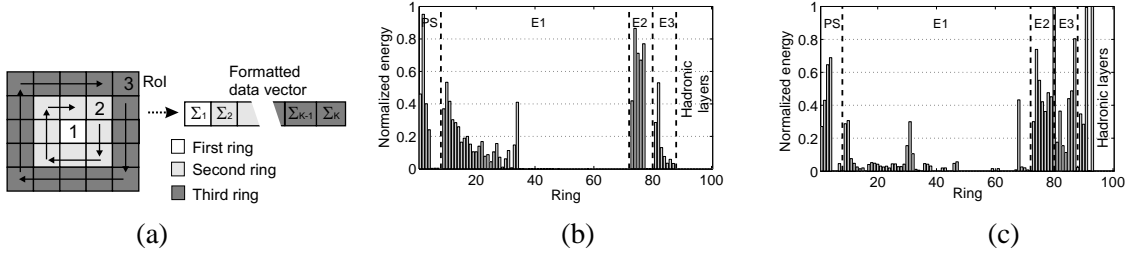
## 3. Segmented ICA-based Calorimeter Trigger

As proposed in [5], calorimeter (Region of Interest) RoI data is pre-processed and formatted into concentric energy rings (see Figure 2-a). For each calorimeter layer, the most energetic cell is considered as the first ring. The next rings are sequentially formed around the first one (this process goes on in a fixed size area of  $0.4 \times 0.4$  in the  $\eta \times \phi$  plane, centered at the most energetic cell, which is assumed to carry enough physics for particle characterization). Finally, ring signals are obtained by summing the energy of the cells belonging to a given ring. The ring energy is normalized within each layer through a sequential normalization procedure. The normalized energy of ring  $i$  at layer  $L$  ( $E_{Li}^{(N)}$ ) is defined as:

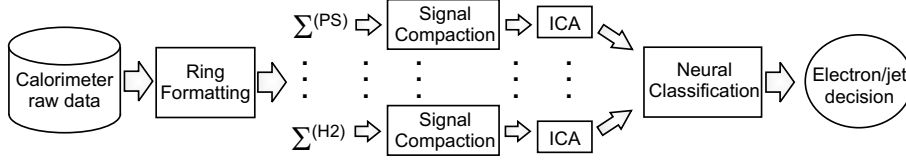
$$E_{Li}^{(N)} = \frac{E_{Li}}{\sum_{j=i}^I E_{Lj}} \quad (3.1)$$

where the sum runs from  $i = 1$  and  $i = I$ , are respectively the central (first) ring and the outermost (last) ring. This procedure aims to emphasize the influence of the outer rings (on a per layer basis), which was proved to improve electron/jet discrimination efficiency [5].

Ring signals from each layer are arranged together into a single vector. Considering this  $0.4 \times 0.4$  region in the  $\eta \times \phi$  plane, the total number of rings (built up from the seven calorimeter layers) amounts to 100.



**Figure 2:** (a) Ring built-up procedure, (b) ring-formatted signature of typical electron and (c) jet. Here, calorimeter layers are limited by vertical dotted lines.



**Figure 3:** Proposed Segmented ICA-based classifiers.

This procedure makes signal description independent from the impact point in the detector, compacts RoI information by a factor of 6 (from  $\approx 600$  cells up to 100 rings) and preserves the energy deposition profile. Ring signals from typical electron and jet are illustrated in Figure 2. One can see that the signal (electron) and background noise (jet) present similar patterns and their discrimination is not a trivial task.

As illustrated in Figure 3, signal compaction algorithms are applied to the ring structure in order to reduce signal dimensionality. This allows more accurate estimation of the independent components and a faster trigger decision (as fewer information is required at the hypothesis testing). In order to cope with different characteristics of each calorimeter layer, such as cell granularity and sensing elements, in this work, the feature extraction procedure (ring construction + signal compaction + ICA) is performed in a segmented way (at the layer level). For the hypothesis testing phase, a supervised MLP neural network [9], fed from the segmented independent components (SIC), is used. The SIC ( $\mathbf{y}_i$ ) from layer  $i$  ( $i=\text{PS}, \text{E1}, \text{E2}, \text{E3}, \text{H0}, \text{H1}$  and  $\text{H2}$ ) were concatenated into a single vector  $\mathbf{Y} = [\mathbf{y}_{\text{PS}}, \dots, \mathbf{y}_{\text{H2}}]^T$ , which was used as the input to the neural discriminator. Details of each processing block are given in the next sub-sections.

### 3.1 Independent Component Analysis

Independent Component Analysis (ICA) is a multidimensional (linear) signal processing technique that searches for a transformation of data, so that their essential structure becomes somehow more accessible [4]. For this, the transformed variables are restricted to be statistically independent.

In the ICA model, the measured signals  $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$  are considered to be generated through a linear combination of the independent source signals  $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$ :

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (3.2)$$

where  $\mathbf{A}$  is the mixing matrix. For the standard case  $N=K$  and thus  $\mathbf{A}$  is square. The ICA model has been widely applied in a variety of signal processing tasks [6].

The purpose of ICA is to estimate the source signal  $\mathbf{s}$  and the mixing matrix  $\mathbf{A}$  using only the observed data  $\mathbf{x}$ . This can be achieved through an inverse model [4]:

$$\mathbf{y} = \mathbf{W}\mathbf{x}, \quad (3.3)$$

where the coefficients of estimated matrix  $\mathbf{W}$  are obtained by considering that the source signals are statistically independent (or at least as independent as possible). If  $\mathbf{y} = \mathbf{s}$  the sources are perfectly recovered and thus  $\mathbf{W} = \mathbf{A}^{-1}$ .

There are many mathematical methods for calculating the coefficients of matrix  $\mathbf{W}$ . Considering two random vectors  $\mathbf{y}_1$  and  $\mathbf{y}_2$ , they are statistically independent if and only if [7]:

$$p_{\mathbf{y}_1, \mathbf{y}_2}(\mathbf{y}_1, \mathbf{y}_2) = p_{\mathbf{y}_1}(\mathbf{y}_1)p_{\mathbf{y}_2}(\mathbf{y}_2) \quad (3.4)$$

where  $p_{\mathbf{y}_1}(\mathbf{y}_1)$ ,  $p_{\mathbf{y}_2}(\mathbf{y}_2)$  are respectively the probability density function (pdf) of  $\mathbf{y}_1$  and  $\mathbf{y}_2$  and  $p_{\mathbf{y}_1, \mathbf{y}_2}(\mathbf{y}_1, \mathbf{y}_2)$  is their joint pdf. In typical blind signal processing problems, there is very little information on the source signals and so the pdf estimation is a very difficult task. Considering this, alternative independence measures (derived from the central-limit theorem, the nonlinear decorrelation property and also from the information theory) are usually applied during the search for independent components [4].

There are some indeterminacies in the ICA model: the order of the extracted components can change and scalar multipliers (positive or negative) may modify the estimated components. Fortunately, these limitations are insignificant in most applications. In this work, the FastICA algorithm [6] was applied for independent component estimation.

### 3.2 Signal Compaction

In our particular problem, it is not expected that the number of sources (i.e. independent components) is as large as the number of observed signals (100 rings). Considering this, signal compaction is performed as a pre-processing step for ICA. The aim is to reduce the number of signals from  $K=100$  (total number of rings) to  $N$  (number of independent components to be estimated), where it is expected that  $N \ll 100$ .

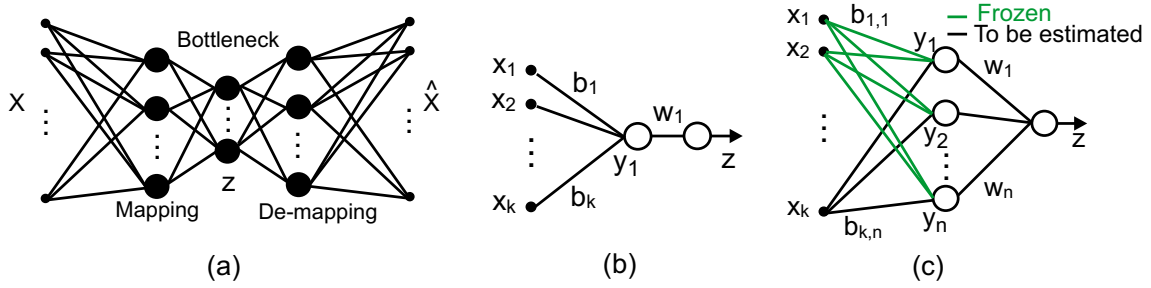
In this work, three different compaction strategies were applied to the ring signals: Principal Component Analysis (PCA), Nonlinear Principal Component Analysis (NLPCA) and Principal Components for Discrimination (PCD).

#### 3.2.1 Principal Component Analysis

Principal Component Analysis (PCA) [8] is a statistical signal processing technique which looks for a linear representation of the input signals in which the energy is often concentrated on a small number of components. This signal representation technique uses second order statistics to search for a subspace where the projections  $\mathbf{z}_i = \mathbf{b}_i^T \mathbf{x}$  ( $i=1, \dots, K$ ) of a zero-mean random vector  $\mathbf{x}$  in direction  $\mathbf{b}_i$  are non-correlated.

The first direction  $\mathbf{b}_1$ , can be computed through the maximization of equation 3.5 [4]:

$$J_1^{PCA}(\mathbf{b}_1) = E[z_1^2] = E[(\mathbf{b}_1^T \mathbf{x})^2] = \mathbf{b}_1^T \mathbf{C}_x \mathbf{b}_1, \quad (3.5)$$



**Figure 4:** Neural models do estimate (a) NLPCA, (b) the first PCD and (c) the n-th PCD.

where  $C_x$  is the covariance matrix of  $\mathbf{x}$ . The PCA for the vector  $\mathbf{x}$  is equivalent to the eigenvalue decomposition of the matrix  $C_x$  [9]. In the PCA transformation, components are sorted according to their energy, with the first component being the most energetic one (i.e. maximum variance). Signal compaction is obtained by retaining only the first  $N$  components (typically,  $N \ll K$ ).

### 3.2.2 Nonlinear Principal Component Analysis

Nonlinear Principal Component Analysis (NLPCA) is a nonlinear extension of PCA. In NLPCA, the mapping from measured data  $\mathbf{x}$  to principal component vector  $\mathbf{z}$  is not restricted to be linear:

$$\mathbf{z} = F(\mathbf{x}), \quad (3.6)$$

where  $F(\cdot)$  is a  $\mathbb{R}^K \rightarrow \mathbb{R}^N$  nonlinear mapping ( $N \ll K$ ). Through NLPCA, one usually obtains higher variance concentration in the initial components with respect to PCA [10].

One way to estimate the nonlinear principal components is through auto-associative neural networks [11]. A multi-layer perceptron (MLP) comprising three hidden layers (see Figure 4-a) is trained using the input vectors as targets. After training convergence, the nonlinear principal components are the bottleneck layer outputs.

When using NLPCA compaction scheme, there are basically two parameters to adjust, the number of bottleneck neurons (i.e. the number of nonlinear principal components -  $N$ ) and the number of mapping neurons ( $MN$ ). For proper data representation, it is often required that  $MN \gg N$ .

### 3.2.3 Principal Components for Discrimination

Considering a classification problem, the purpose of PCD analysis is to obtain a linear projection of the input signals  $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$  that maximizes class separation and data compaction rate simultaneously [12]. The discriminating components  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$  (usually  $N \ll K$ ) can be expressed as:

$$\mathbf{y} = \mathbf{B}\mathbf{x} \quad (3.7)$$

The PCD extraction can be performed through a MLP neural network [9]. For simplicity, considering a binary discrimination process, a network with a single hidden neuron, trained to maximize class discrimination, extracts the first discriminating component (see Figure 4-b). By sequentially adding neurons to the hidden layer and restarting the training procedure, the next components are estimated. The hidden layer weights are trained only for the added neurons (highlighted synaptic lines in Figure 4-c). The estimated weights from the previous steps are kept fixed,

as they represent the directions of the principal components already extracted. The weights of the output layer are adjusted during the training procedure for optimal combination of principal component information at each processing step.

### 3.3 Hypothesis Testing

Here, the hypothesis testing phase at LVL2 is performed by a neural classifier fed from the segmented independent components. The available data set is divided into training, testing and validation sets. The Resilient Back-propagation (R-PROP) [13] algorithm is used in this work to train the MLP classifiers. The network comprises a single hidden layer and one output neuron. The optimum number of hidden neurons (NH) was chosen through the following procedure:

1. Set NH=1;
2. test the discrimination efficiency for five different initializations and chose among them, the network with best discrimination performance ( $NET_{NH}$ );
3. set NH=NH+1 and repeat step 2;
4. repeat step 3 until there is no performance improvement by adding hidden neurons.

When this iteration finishes, chose among  $NET_{NH}$  the value of NH that produces best separation efficiency.

## 4. Results

A simulated database comprising approximately 450,000 electron and 450,000 jet signatures was used to design and test the proposed discriminators for the ATLAS LVL2 trigger. Level-one trigger selection and instrumentation noise effects were considered in these simulations. Here, the performance comparison is performed through the ROC (Receiver Operating Characteristics) curve and SP (Sum-Product) index. The ROC [14] displays how both the detection and false alarm probabilities, respectively  $P_D$  (probability of identifying correctly an incoming electron) and  $P_F$  (probability of assuming that an incoming jet is an electron), vary as the decision threshold changes. The SP index is defined as [5]:

$$SP = \frac{Ef_e + Ef_j}{2} \times \sqrt{(Ef_e \times Ef_j)} \quad (4.1)$$

where  $Ef_e = P_D$  is the detection efficiency for electrons and  $Ef_j = (1 - P_F)$  is the corresponding efficiency for jets. The threshold value that maximizes SP can be considered optimal for trigger operation as it provides both high  $P_D$  and low  $P_F$ .

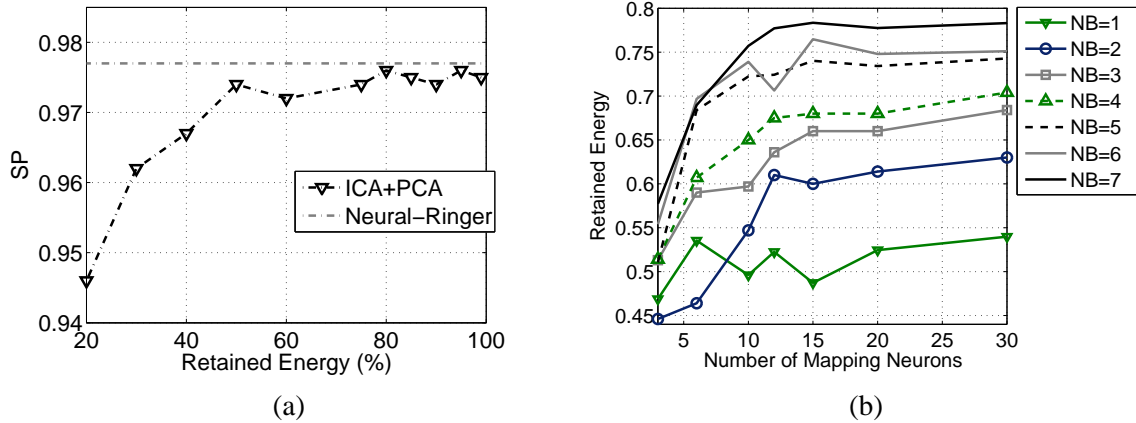
If PCA is applied for signal compaction, it is necessary to determine the optimum number of retained components. In order to do this, the ring structure was compacted using different energy retention levels. For each situation, segmented ICA was computed and used to feed the neural classifier. The number of segmented principal components retained and the SP indexes for different energy preservation levels are depicted, respectively, in Table 1 and Figure 5-a. It can be seen that the SP starts to decrease significantly for energy retention levels lower than 50%. Considering this,

**Table 1:** Preserved components for different SPCA energy retention levels.

Layer	Retained Energy (%)											
	100	99	95	90	85	80	75	60	50	40	30	20
PS	8	7	7	7	6	6	5	4	3	2	1	1
E1	64	62	54	45	36	29	22	15	10	4	3	1
E2	8	7	6	5	4	4	3	2	1	1	1	1
E3	8	7	7	6	6	5	5	4	3	2	1	1
H0	4	3	3	3	3	3	2	2	2	1	1	1
H1	4	3	3	3	3	3	2	2	2	1	1	1
H2	4	3	3	3	3	3	2	2	2	1	1	1
<b>Total</b>	100	92	83	72	61	53	41	31	23	12	9	7

**Table 2:** Number of PCD and respective SP for each calorimeter layer.

	PS	E1	E2	E3	H0	H1	H2	Total
<b>NPCD</b>	6	4	5	4	3	3	3	28
<b>SP</b>	0.685	0.960	0.917	0.674	0.767	0.652	0.457	0.980

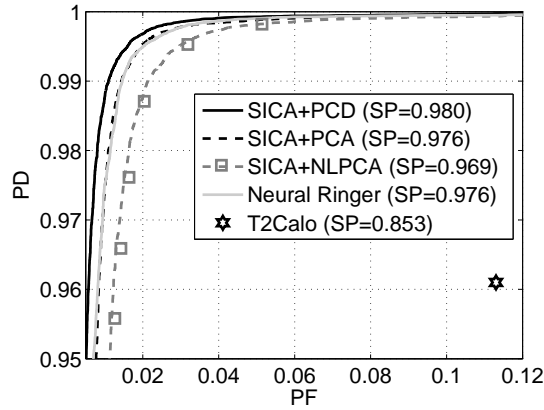

**Figure 5:** (a) SP versus PCA total energy retention levels and (b) energy retention at PS calorimeter layer for different NLPCA network architectures.

it is possible to achieve almost the same efficiency as the Neural-Ringer (which uses 100 rings) using only 23 principal components. The number of hidden neurons used for classification also decreased from nine to five, respectively for the Neural Ringer and the PCA-based discriminator.

Considering NLPCA signal compaction, in this work, the optimum network architecture was chosen considering parameters like energy retention and SP index, through a procedure similar to the one applied for PCA. In Figure 5-b, the energy retention level is computed varying the number of bottleneck (NB) and mapping neurons (for PS layer). The same procedure was repeated for the other six layers. The highest SP obtained with a NLPCA discriminator was 0.969, and the total number of segmented nonlinear principal components was 35.

Here, PCD networks were applied to the ring structure in a segmented way. The number of principal components for discrimination, and the respective SP, estimated for each layer are




**Figure 6:** ROC curves.

**Table 3:** Efficiency for different discriminators, N is the number of components used to feed the neural classifier.

	SICA+PCD	SICA+PCA	Neural Ringer	SICA+NLPCA	T2Calo
<b>SP</b>	0.980	0.976	0.976	0.969	0.853
<b>PD(%)</b>	98,9	97,6	97,5	97,1	96,1
<b>PF(%)</b>	1,6	1,8	1,7	2,3	11,3
<b>N</b>	28	23	100	35	-

illustrated in Table 2. It can be seen that the total number of PCD is 28 (producing SP=0.980).

A performance comparison between the proposed discriminators is shown in Figure 6 and Table 3. It can be seen that the highest efficiency was achieved through the PCD method (using 28 components). PCA pre-processing produced highest signal compaction (23 components, SP=0.976) and NLPCA produced lower levels of compaction and discrimination performance (35 components, SP=0.969).

Considering the processing time requirements for the ATLAS second-level trigger (latency time smaller than 10 ms), the computational performance of T2Calo and the Neural Ringer were compared using ATHENA implementations. The total processing times (considering feature extraction and decision) were  $860 \mu s$  for T2Calo and  $972 \mu s$  for the Neural Ringer (13% higher, but still within LVL2 requirements). A simple comparison with the Neural Ringer indicates that the HLT requirements should be satisfied for the proposed discriminator. Through the proposed approach, a signal compaction block will reduce significantly the number of inputs for the neural classifier, contributing for processing speed improvement during classification. Furthermore, considering PCD compaction, only a matrix multiplication is added in the signal processing routine (the ring signals are multiplied by  $\mathbf{C} = \mathbf{W} \times \mathbf{B}$ , where  $\mathbf{B}$  and  $\mathbf{W}$  are respectively the PCD and ICA matrices), and thus no considerable increase in the computational effort is expected.

## 5. Conclusions

An alternative second-level trigger algorithm based on independent component analysis is pro-

posed in this work for the electron/jet channel of ATLAS detector. The proposed feature extraction procedure operates over calorimeters information and comprises ring structure construction, signal compaction and estimation of the independent components. An artificial neural network discriminator, fed from the independent components, is used for hypothesis testing. Through the proposed method, an electron efficiency of approximately 99% was achieved for less than 2% false alarm, outperforming the baseline discriminator in use.

## Acknowledgements

The authors would like to express their gratitude to CNPq, FINEP, CAPES, FAPERJ (Brazil) and CERN (Switzerland) for the support. We also thank the ATLAS Trigger/DAQ collaboration for providing the calorimeter data and for fruitful discussions concerning this work and M.Sc. Rodrigo Torres for helping in the simulation data production.

## References

- [1] Evans, L., Bryant, P. "LHC Machine". *Journal of Instrumentation*, 3, S08001 (2008).
- [2] ATLAS Collaboration "ATLAS Experiment at CERN Large Hadron Collider". *Journal of Instrumentation*, 3, S08003 (2008).
- [3] Anjos, A. et al. "Configuration of the ATLAS Trigger". *IEEE Transactions On Nuclear Science*, Vol. 53, No. 3, pp. 990-994, June (2006).
- [4] Hyvarinen, A., Karhunen, J., Oja, E. *Independent Component Analysis*. Wiley (2001).
- [5] Anjos, A. et al. "Neural Triggering System Operating on High Resolution Calorimetry Information". *Nuclear Instruments and Methods in Physics Research*, 559, pp.134-138 (2006).
- [6] Hyvarinen, A. and Oja, E. "Independent Component Analysis: Algorithms and Applications". *Neural Networks*, 13(4-5):411-430 (2000).
- [7] Papoulis, A. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill (1991).
- [8] Jolliffe, I. T. *Principal Component Analysis*, Springer (2002).
- [9] S. Haykin: *Neural Networks, Principles and Practice*, Bookman (2001).
- [10] Soares-Filho, W. Seixas, J. M. and Caloba, L. P. "Principal component analysis for classifying passive sonar signals" *IEEE International Symposium on Circuits and Systems*, pp. 592-595, Sydney, Australia (2001).
- [11] Kramer, M. A. "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks". *AIChE Journal*, 37(2): 233-243 (1991).
- [12] Caloba, L.P., Seixas, J.M. and Pereira, F.S. "Neural Discriminating Analysis for a Second-Level Trigger System". *Int. Conf. on Computing in High Energy Physics*, pp. 1-5, Rio de Janeiro, September (1995).
- [13] Riedmiller, M, Braun, H. "A Direct Adaptive Method for Faster Backpropagation Learning, the RPROP algorithm", *Proceedings of the IEEE International Conference on Neural Networks*, pp. 586-591 (1993).
- [14] Van Trees, H. L. *Detection, Estimation, and Modulation Theory*, Part I. Wiley (2001).