Neural Online Filtering Based on Preprocessed Calorimeter Data

Rodrigo Coura Torres, Danilo Enoque Ferreira de Lima, Eduardo Furtado de Simas Filho, José Manoel de Seixas

Abstract—Aiming at coping with LHC high event rate, the ATLAS collaboration has been designing a sophisticated three-level online triggering system. A significant number of interesting events decays into electrons, which have to be identified from a huge background noise. This work proposes a highly-efficient L2 electron / jet discrimination algorithm based on artificial neural processing fed from preprocessed calorimeter information. The feature extraction part of the proposed system provides a ring structure for data description. Energy normalization is later applied to the rings, making the proposed system usable for a broad energy spectrum. Envisaging data compaction, Principal Component Analysis and Principal Component of Discrimination are compared in terms of both compaction rates and classification efficiency. For the pattern recognition section, a fully-connected feedforward artificial neural network was employed. The proposed algorithm was able to achieve an electron detection efficiency of 96% for a false alarm of 7%.

Index Terms—Online Filtering, Calorimetry, Feature Extraction, Signal Compaction, Particle Identification, Neural Networks, Principal Components Analysis, Principal Components of Discrimination.

Fig. 1: The ATLAS detector.

I. INTRODUCTION

PARTICLE colliders are often used in high-energy physics experiments. By accelerating and colliding particles, studies about the structure of matter can be performed. The *Large Hadron Collider* (LHC) [1] is the largest particle accelerator in the world. Having a circumference of $\sim$ 27 km, LHC will be colliding proton beams with a centre-of-mass energy of 14 TeV and an unprecedented luminosity of 10^{34} cm^{-2} s^{-1}. For increasing the probability of detecting interesting physics, LHC will provide a bunch crossing rate of 40 MHz.

The LHC ring has multiple collision points. At these sites, detection laboratories were built in order to analyze the subproducts originated from collisions. The largest LHC detector, ATLAS [2], comprises tracking, calorimetry and muon detection systems (see Fig. 1). Due to its segmentation and granularity, each bunch crossing will produce $\sim$ 1.5 MBytes of information, resulting in a data stream of $\sim$ 60 TBytes/s. Since it is impossible to store such a quantity of data, and due to the fact that most data is from the huge background noise, this rate must be reduced to $\sim$ 200 Hz in order to be
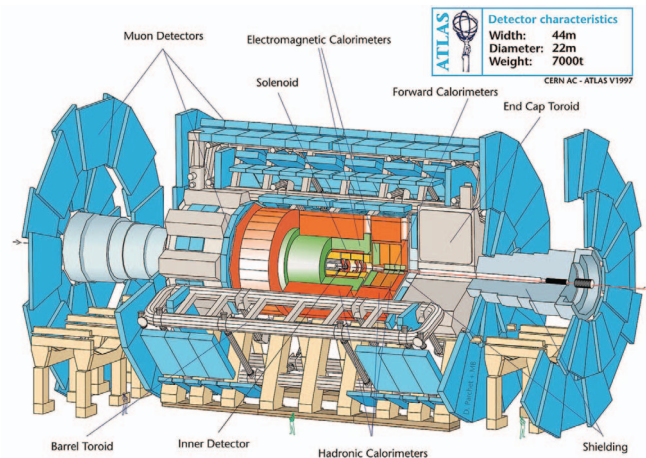
stored in mass storage devices for further offline analysis. For such rate reduction, an online, three-level triggering system has been designed. The first-level trigger will reduce the input rate to $\sim$ 75 kHz analyzing coarse information from both calorimeter and muon systems. The second-level trigger will validate the events approved by the first level, using full detector granularity, reducing the event rate to $\sim$ 2 kHz. The third level will analyze all available information of each event, providing the final online trigger decision.

High-energy electrons are among the most powerful indicators of interesting physics processes [3]. However, in hadron colliders, the interesting electrons are outnumbered by jets in the same energy range by several orders of magnitude [4]. The success of a hadron collider experiment often depends on the electron detection efficiency in such difficult conditions.

The ATLAS triggering system strongly relies in calorimetry. The ATLAS calorimeter is composed by electromagnetic and hadronic sections. Both sections are divided into three segments with distinct granularity and depth. In front of the electromagnetic calorimeter there is also a pre-shower detector, resulting in a total of seven calorimeter layers.

This paper presents a calorimeter based electron / jet separation algorithm that is a candidate for operating at the second level of the ATLAS triggering system. For reducing input data dimension, and revealing relevant variables, a feature extraction step is first performed. For this, calorimeter data are described through a set of energy-normalized concentric ring sums. For data compaction, Principal Component Analysis (PCA) [5] and Principal Component of Discrimination (PCD) [6] are evaluated in terms of both compaction rates
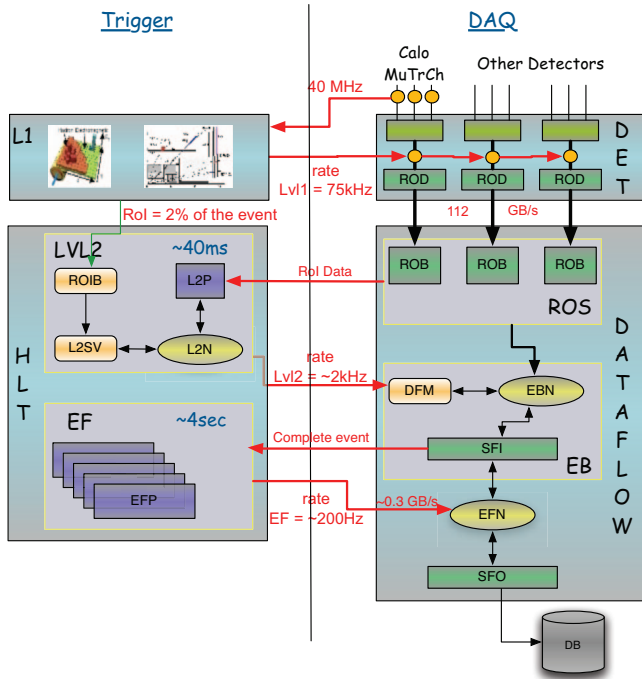
Fig. 2: Schematic view of the ATLAS triggering system.

and classification efficiency. Moreover, as signal decorrelation is performed either linearly (PCA) or nonlinearly (PCD) in the preprocessing phase, the classifier complexity can be reduced, which also favors signal preprocessing adoption in level two trigger strategy. For comparison, components were extracted in a segmented way, i.e., by analyzing each calorimeter layer individually, and non-segmented (concatenating layers).

For the pattern recognition section, an artificial neural network [7] was employed. Neural networks are fast to execute, and can provide nonlinear cuts in a high-dimensional space, providing better pattern separation for complex problems when compared to linear decision systems. Also, neural networks are resilient to noise and missing data, which is attractive for electron / jet discrimination.

This paper is organized as it follows: Sec. II will briefly present the ATLAS triggering system. Next, in Sec. III, the basic concepts behind PCA and PCD are introduced. In Sec. IV, the structure of the proposed electron / jet separation system is detailed. Results obtained from such system are presented in Sec. V. Finally, conclusions are derived in Sec. VI.

II. THE ATLAS TRIGGERING SYSTEM

The triggering system of the ATLAS detector was designed to cope with the huge data stream produced by each LHC collision, reducing it to feasible levels. The ATLAS Trigger [8] comprises three sequential levels of analysis (see Fig. 2).

The first level (L1) will be responsible for the preliminary rejection by analyzing both calorimeter and fast muon systems. To cope with a maximum latency of $2.5 \mu s$, it is implemented in low-programability devices (FPGA), and operates with detector coarse granularity. During its analysis, L1 marks the so-called *Regions of Interest* (RoI) of accepted events, which

contain the η and ϕ directions of the identified L1 objects, as well as the transverse momentum thresholds that have been passed. At the end of this level, the event rate will be reduced to ~ 75 kHz.

The second level (L2) is fed from the RoI approved by L1, and is responsible for validating L1 decision. L2 includes the tracking system information and analyzes each RoI in its full granularity, reducing the event rate to ~ 2 kHz. Since L2 takes, in average, 40 ms to analyze each RoI, a set of ~ 500 general-purpose server computers (PC) running Linux, connected by gigabit ethernet switches are employed. The algorithms developed for this level were implemented in high-level computing language (C++), allowing the development of more complex filtering algorithms.

The last level (also known as the *Event Filter*) provides the final trigger decision, requiring $\sim 1,600$ server computers to cope with an average processing time of ~ 4 sec per event. The events approved by this level (~ 200 Hz) are forwarded to mass storage devices for further offline analysis.

III. SIGNAL COMPACTION

The method of principal components is primarily a feature extraction technique that obtains a linear or nonlinear transformation of a group of correlated variables such that certain optimal conditions are achieved [9]. The transformation is designed in such a way that the dataset may be represented by a reduced number of components, which still retain most of its relevant information.

A. Principal Components Analysis

Principal Components Analysis (PCA) is a widely used technique for feature selection in statistical pattern recognition problems. Considering a random vector $\mathbf{x} \in \mathrm{R}^N$, principal components may be extracted by *Single Value Decomposition* (SVD) [10] of the correlation matrix $\mathbf{R}$ of $\mathbf{x}$. This implies solving the following equation system

$$\mathbf{U}^T \mathbf{R} \mathbf{U} = \mathbf{\Lambda} \qquad (1)$$

$$|\mathbf{U} - \lambda \mathbf{I}| = 0 \qquad (2)$$

where $\mathbf{U}$ is an orthonormal matrix containing the eigenvectors of $\mathbf{R}$, $\mathbf{\Lambda}$ is a diagonal matrix, whose elements $\lambda_1, \lambda_2, ..., \lambda_N$ are the eigenvalues of $\mathbf{R}$, and $\mathbf{I}$ is the identity matrix. Once the eigenvectors are extracted, they can be ranked by their associated eigenvalues, forming the principal component projection base. Usually, prior to the PCA extraction, the random vector $\mathbf{x}$ is centralized by having its mean removed ($\mathbf{x} - \bar{\mathbf{x}}$), so that the correlation matrix becomes the covariance matrix. In such condition, the eigenvalues give the amount of energy (data variance) retained by the corresponding component.

Therefore, PCA provides a projection matrix $\mathbf{U} \in \mathrm{R}^{N \times N}$, where the components are ranked by the amount of statistical variance they retain from $\mathbf{x}$. By discarding the least energetic components, one can achieve dimension reduction in an optimal way, in the mean squared error (MSE) sense.
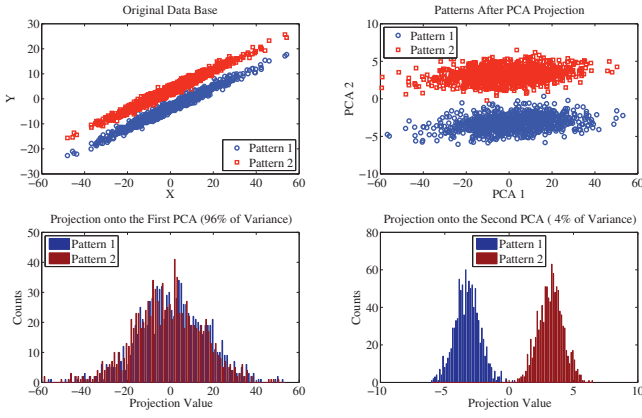
Fig. 3: An example for which the most energetic PCA does not provide useful information for pattern recognition.
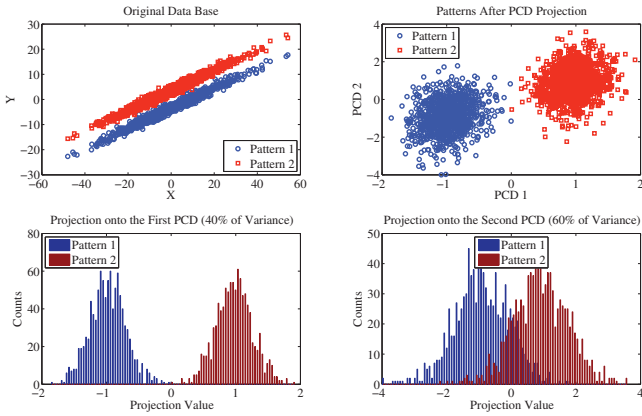

Fig. 4: Example on how PCD exploits directions useful for pattern recognition.

However, for pattern recognition problems, not always the best reconstruction corresponds to optimal discrimination performance. Therefore, one can naively discard minor components although they may point to directions where the boundaries of each pattern are well distinguishable. This is illustrated in Fig. 3. In this example, the first PCA component, although retaining 96% of the process variance, does not provide any useful information for discrimination. On the other hand, the boundaries of each pattern are very well established when we look at the projection onto the second PCA component, which carries only 4% of the original information. Therefore, in classification problems, it may be more efficient to look to the most discriminating components.

B. Principal Components of Discrimination

The Principal Components of Discrimination (PCD), instead of finding directions that minimize the data representation error, searches for directions that maximize discrimination efficiency. Thus, for the example shown in Fig. 3, Fig. 4 shows the projection onto PCD components. As it can be seen, PCD correctly placed its first component in the direction that better reveals the boundaries of each pattern.
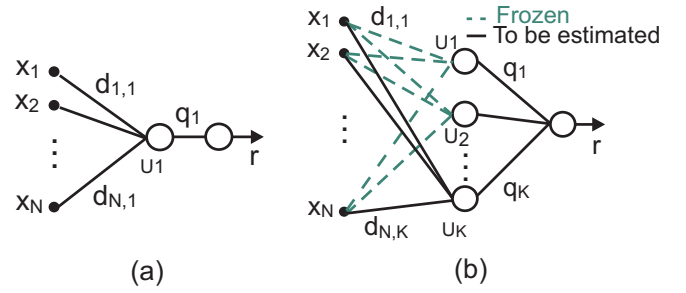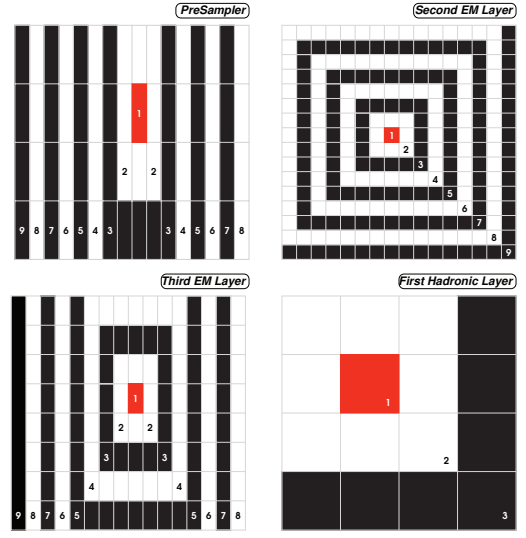

Fig. 5: PCD extraction procedure.

Fig. 6: Rings structure.

PCD components may be extracted using artificial neural networks. Considering that the incoming events are in R^N, a multi-layer perceptron network with a single neuron in the hidden layer and a nonlinear activation function (in our case, the hyperbolic tangent) can extract the first component. When the training phase is completed, the weight values of the hidden node contain the direction of the first PCD. Next, a new node is inserted in the hidden layer, and the weight vector connecting to the first node is frozen. The new neural network is then trained, furnishing the second component. This procedure is, then, carried out until the addition of new components does not improve much the discrimination performance.

IV. THE ELECTRON / JET SEPARATION ALGORITHM

The electron / jet separation algorithm comprises feature extraction and pattern recognition sections. Moreover, it is fed from a L1 approved RoI. Each RoI has, in average, 1,000 cells, providing a highly sparse data space. In order to retrieve intelligent information from the RoI, while maintaining its physical interpretation, a topological pre-processing is performed. By knowing that electromagnetic showers have smaller lateral dispersion than hadronic showers [4], for each calorimeter layer, a set of concentric rings [11] are produced, centered at the hottest cell of each layer (see Fig. 6). Next,

TABLE I: Number of rings in each calorimeter layer.
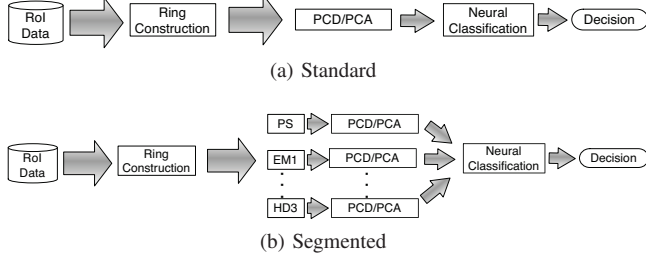
Layer	PS	EM1	EM2	EM3	HD1	HD2	HD3
# Rings	8	64	8	8	4	4	4

(a) Standard

(b) Segmented

Fig. 7: Processing flow for both approaches.

Fig. 8: Datasets transverse energy distribution for the events reaching L2.

the cells belonging to a given ring are summed up, producing a single value. A total of 100 rings are produced, organized as presented in Tab. I.

In order to reduce the energy influence in our analysis, an energy-based normalization by the total energy is performed:

$$r'_i = \frac{r_i}{\sum_{j=1}^{100} r_j} \tag{3}$$

where r_i is the original i-th ring sum.

Once the ring sums are computed, the feature extraction progresses by two different ways, as illustrated in Fig. 7: in the standard case (Fig. 7a), the 100 ring sums are considered as a single input vector, which means that both PCA and PCD extraction and projection phases are performed by analyzing all layers as a whole. In the segmented case (Fig. 7b), the PCA and PCD are extracted for each layer individually, meaning that each layer will have their own projection matrix. Once the projection is performed in such per-layer basis, the resulting components are concatenated, producing the final vector containing the desired components to be used by the classifier. The greatest advantage of such segmented approach, when compared to the standard case, is that it is possible to know how each calorimeter segment constitutes to the classifier performance.

Once the feature extraction section is performed, the resulting vector is fed into a feedforward neural network which implements the pattern recognition section. Previous analysis pointed out that 10 nodes in the hidden layer are enough. Furthermore, both hidden and output layers use hyperbolic tangent as the activation function. The training algorithm is the *Resilient Backpropagation* [12].

V. RESULTS

Monte Carlo simulation data were used for evaluating the performance of the proposed L2 algorithm. Simulation data comprised $\sim$ 160k single electrons with transverse energy (E_T) between 7 and 80 GeV and $\sim$ 100k QCD dijets events, which contain at least one e/γ candidate with $E_T > 17$ GeV. Both datasets were initially pre-filtered by the standard L1 algorithm, considering energy, EM and HAD isolation [13]. A total of $\sim$ 140k electron and $\sim$ 13k fake electron (jets) RoI
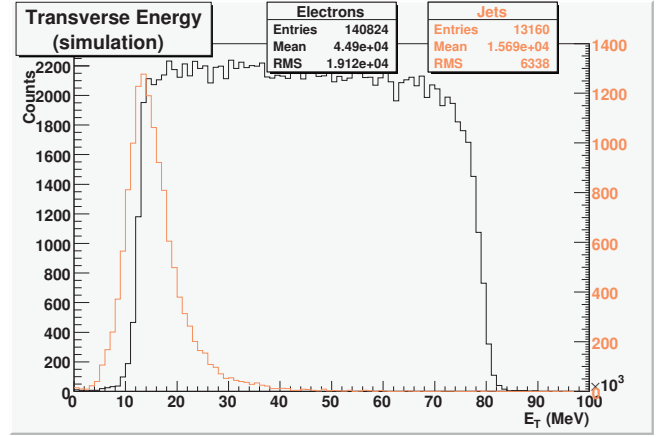
passed L1 cuts and reached the L2 system. Fig. 8 shows the transverse energy distribution for both datasets. For developing and validating the proposed system, the datasets were halved. The first half (training set) was used for system development (PCA/PCD extraction and neural network training). The second half was used to avoid overtraining [7] of the neural classifier design, as well as for validating the developed algorithm regarding its detection efficiency.

For comparing the results obtained from each approach, the Receiver Operating Characteristics (ROC) curve [14] will be used. The ROC curve is generated by analyzing the electron and jet probability density functions (PDF) estimation obtained at the classifier's output, when fed from these two datasets. By varying the decision threshold over the dynamic range of the classifier's output, it is possible to compute, for a given threshold value, the electron detection probability (referred to as *Detection Efficiency*), and also the probability of background noise misclassification (referred to as *False Alarm*).

A. Component Selection by the Number of PCA Components

The number of PCA components was chosen using the *Average Root* criterium [9]. This criterium establishes that the i-th PCA component should be retained if

$$\lambda_i > 0.7 \times \bar{\lambda} \tag{4}$$

where λ_i is the eigenvalue associated to the i-th PCA component, and $\bar{\lambda}$ is the average value of the eigenvalues. Tab. II shows the number of PCA components retained for both each layer (segmented case) and for the standard (non-segmented) case. The table also shows that the amount of variance retained in each case is quite high, showing that the information is not spread over the whole space. The exceptions are EM3, HD1, HD2 and HD3. Since these layers sample little energy for L1 qualified events, most of the information read from these layers is composed by uncorrelated noise. This results in more spread information through the data space, reducing the compaction capabilities of PCA.

After selecting the number of PCA components for both segmented and non-segmented approaches, we forced the

533

TABLE II: Number of components selected by the Average Root criterium for both standard and segmented approaches.

Approach	Segmented								Standard
Layer	PS	EM1	EM2	EM3	HD1	HD2	HD3	Total	Total
# PC	2	5	2	2	2	2	1	16	10
% Variance	95.1	96.6	93.2	68.9	79.2	79.2	72.1	N/A	95.3
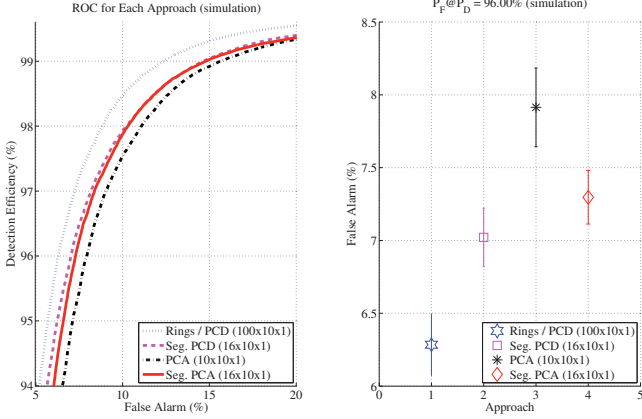
Fig. 9: Results employing PCA-based component selection.

Fig. 10: Relevance mapping for the segmented (left) and standard (right) approaches for both PCD and PCA.

TABLE III: Number of components selected by the Relevance Mapping criterium for both standard and segmented approaches.

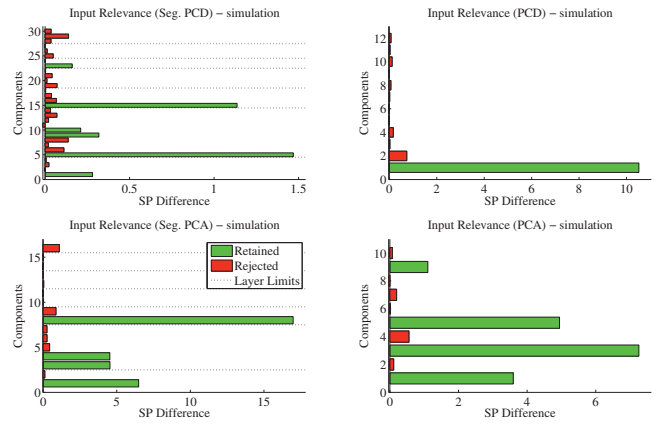| Approach	Segmented								Standard
Layer	PS	EM1	EM2	EM3	HD1	HD2	HD3	Total	Total
# PCA	1	3	0	0	0	0	0	4	4
# PCD	1	3	1	0	1	0	0	6	1

number of PCD components to match exactly the number of PCA components retained for each case. In Fig. 9, the ROC curves obtained for each approach are displayed, considering both PCA and PCD analysis. As is quite noticeable, the PCA performance is poorer with respect to PCD. For instance, for the standard case, considering a fixed electron detection probability of 96%, PCD achieved a false alarm ~ 1.6 percentage points (pp) lower than its PCA counterpart, even though PCA components recover $\sim 95\%$ of data variance.

B. Component Selection by Relevance Mapping

In order to optimize the minimum number of components needed for efficient classification, the relevance mapping was employed [15]. This technique evaluates, for a given developed classification system, the impact of each component to the classifier detection efficiency. Here, the SP product was employed for measuring component relevance. The SP value is given by

$$SP = 100 \times \sqrt{\sqrt{P_e \times P_j} \times \frac{(P_e + P_j)}{2}} \tag{5}$$

where P_e and P_j are the electron and jet detection probabilities. The maximum SP value indicates a good threshold value for the ROC curve, in which high discrimination efficiency is obtained for electrons and also high rejection rates are obtained for background noise. The relevance for a given component x_i may be computed through

$$R_i = SP(\mathbf{x}) - SP(\mathbf{x}|_{x_i = \bar{x}_i}) \tag{6}$$

where $\bar{x}_i$ is the average value of the i-th component. The higher the relevance of a given component, the poorer the classifier will perform without it. In this work, a component is considered relevant if its relevance is at least 10% of the

relevance of the most relevant component for a given design approach.

The results for the relevance mapping can be visualized in Fig. 10. As it can be seen, relevance pointed out that, for the PCA case, the first components are not necessarily the ones which matter the most for classification. It is interesting to notice, that, for the PCA case, some components provided a negative relevance, which shows that they are actually degrading the classifier performance. On the other hand, since PCD are ranked by their discriminating power, the first PCD components are always the most relevant components for the classifier. It is presented in Tab. III the number of retained components based on relevance mapping.

In order to analyze the most relevant component projection for PCA and PCD, considering the standard approach, electrons and jets were projected onto such privileged directions (see Fig. 11a and Fig. 11b). It becomes clear the advantage of PCD in discrimination tasks. Fig. 12 shows the ROC curve obtained when varying a decision threshold over each projection. As expected, the ROC points a clear advantage of PCD over PCA.

The neural classifiers were then retrained, fed only from the relevant components (Tab. III). Despite the input layer, all other parameters regarding the classifier topology (number of hidden nodes, activation function, etc.) were kept unchanged. The results obtained for each case can be observed in Fig. 13. As a reference, it is also shown the result obtained when training a classifier fed only from the ring sums (without any principal component projection or relevance selection).

PCA (both segmented and standard cases) suffered from further component reduction, when compared to their PCD counterparts. This happens because PCD can compact discrim-
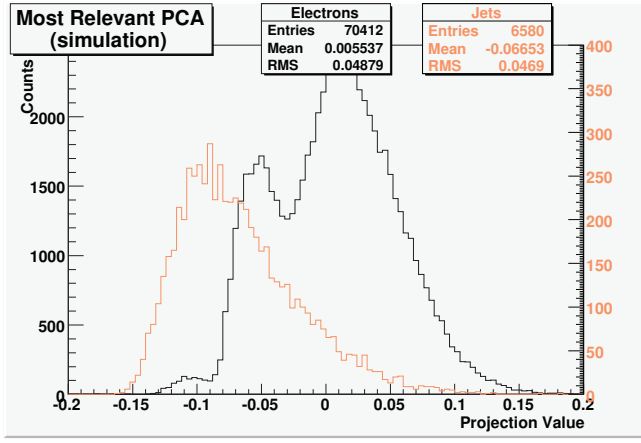
(a) PCA
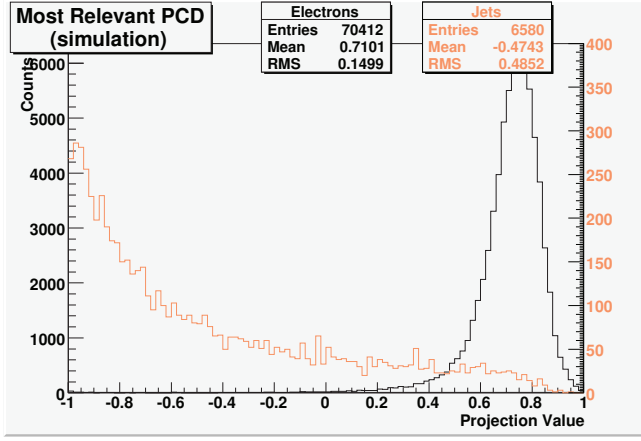
(b) PCD

Fig. 11: Electron and jet projections onto the most relevant PCA and PCD components.

Fig. 12: ROC curves for the most relevant PCA and PCD components.
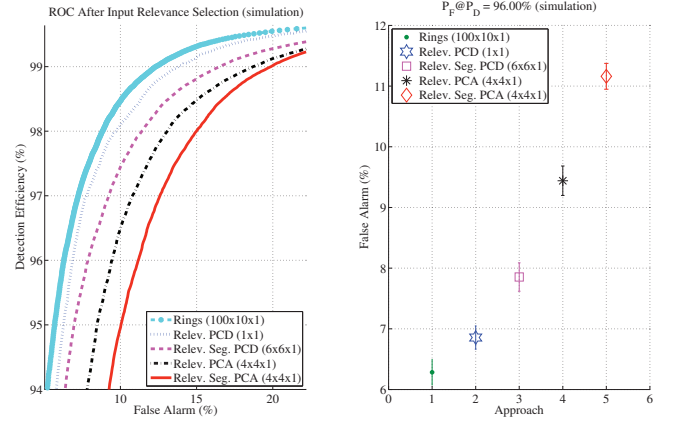
Fig. 13: Results for each approach after relevance selection.

inating information in fewer components, meaning that it will suffer smaller efficiency reduction when multiple components are removed at the same time. Both PCD based approaches manage to provide ROC curves extremely close to the reference curve, showing that PCD is more resilient when multiple components are removed.

The standard PCD case provided the best classification efficiency and also the highest compaction rate, being able to have a false alarm difference of only 0.57 pp from the reference curve, for an electron efficiency of 96%. This interesting result also points out that since a single PCD provided a good result, linear discriminants [16] could be employed, resulting in further design simplification to the classification section.

Although the standard PCD case provided the best detection efficiency, this approach has the drawback that it is impossible to evaluate how much each segment is contributing to the classifier output. On the other hand, the segmented PCD approach provides such information, while achieving good discrimination capabilities, being more useful during nominal trigger operation.

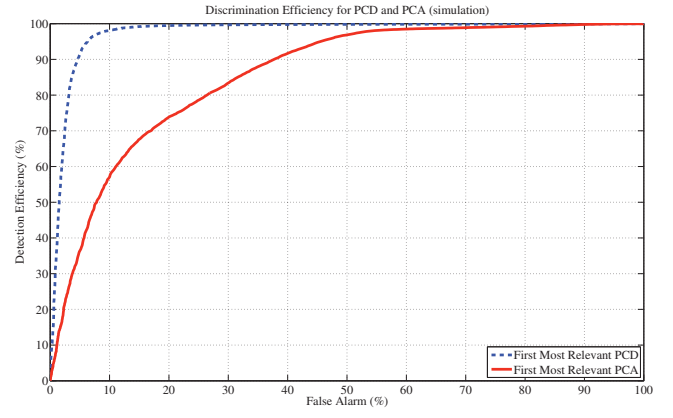Fig. 14 presents operational analysis considering the segmented PCD case after the relevance mapping selection (six components). The upper-left plot shows the neural network output for electrons and jets. It can be observed that both patterns are very well separated. The upper-right plot shows the detection efficiency and false alarm obtained as a function of the transverse energy. The plot shows higher false alarm values after 20 GeV, which happens since the amount of jet events with $E_T > 20$ GeV is quite small. Next, the lower-left plot shows the performance as a function of η, where it can be observed a very good efficiency over that coordinate, with some reduction in $|\eta| \approx 1.5$, which corresponds to the calorimeter crack region [17]. Finally, the lower-right plot shows how performance varies with respect to ϕ, which, as expected, is constant.

VI. CONCLUSIONS

This paper presented an electron / jet candidate algorithm for operating in the second level of the ATLAS online triggering system based on calorimetry. In order to select intelligent information, while maintaining physics interpretation, a topological pre-processing based on ring sums was employed. Further, efficient signal compaction was obtained through both PCA and PCD. PCA and PCD components were extracted both by considering the 100 rings as a single input vector, as well as treating each calorimeter layer individually. The main advantage from a segmented analysis is that it is possible to
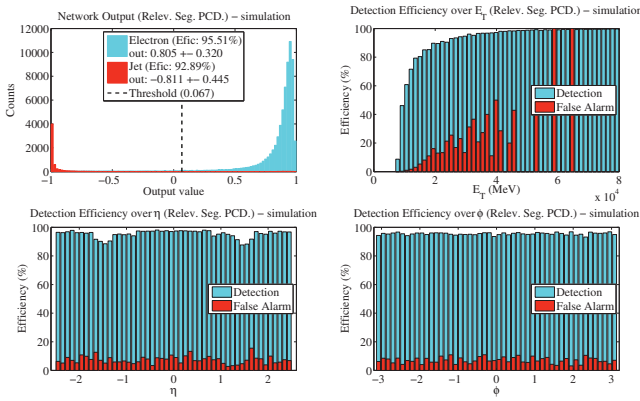
535

Fig. 14: Neural classifier Output, E_T, η and ϕ analysis for the segmented PCD case with relevance selection.

identify how much each calorimeter layer is contributing to the discrimination system performance.

When PCA and PCD were forced to have the same number of components, PCD managed to outperform PCA both in the segmented and non-segmented approaches, showing that this technique was able to retrieve the most relevant information from the discrimination point of view.

PCD also performed better than PCA when components were selected based on the relevance mapping. The relevance analysis showed that only one PCD for the standard case suffices for good classification performance, pointing out that linear discriminants might be a good solution for the classification section, bringing further simplification to the proposed algorithm, with direct impact on speed performance.

Finally, it was shown that relevance mapping can act as a knob, allowing system operators to optimally decide which components to discard, when faster execution is required, minimizing classification efficiency losses.

REFERENCES

[1] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 1, no. 2008 JINST 3 S08001, August 2008.

[2] G. Aad, E. Abat, J. Abdallah, A. A. Abdelalim, A. Abdesselam, and O. Abdinov, "The ATLAS experiment at the CERN large hadron collider," *Journal of Instrumentation*, vol. 1, no. 2008 JINST 3 S08003, August 2008.

[3] G. Altareli and L. Di Lella, *Proton-Antiproton Collider Physics*. World Scientific Publishing, 1989.

[4] R. Wigmans, *Calorimetry: Energy Measurement In Particle Physics*. Oxford, 2000.

[5] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.

[6] L. P. Calôba, J. M. Seixas, and F. S. Pereira, "Neural discriminating analysis for a second level trigger system," in *Computing in High Energy Physics*, Rio de Janeiro, Brazil, 1995, pp. 870–874.

[7] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Prentice Hall, 2008.

[8] K. Kordas *et al.*, "The atlas data acquisition and trigger: concept, design and status," *Nuclear Physics B - Proceedings Supplements*, vol. 172, pp. 178 – 182, 2007.

[9] J. E. Jackson, *A User's Guide to Principal Components*. John Wiley & Sons, 1991.

[10] S. Haykin, *Adaptive Filter Theory*, 3rd ed., ser. Information and System Sciences, T. Kailath, Ed. Prentice Hall, 1996.

[11] A. dos Anjos, R. C. Torres, J. M. Seixas, B. C. Ferreira, and T. C. Xavier, "Neural triggering system operating on high resolution calorimetry information," *Nuclear Instruments and Methods in Physics Research*, vol. 559, pp. 134–138, May 2006.

[12] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in *Proceedings of the IEEE International Conference on Neural Networks*, San Francisco, USA, 1993, pp. 586–591.

[13] R. Achenbach *et al.*, "First data with the atlas level-1 calorimeter trigger," in *IEEE Nuclear Science Symposium Conference Record*, Dresden, Germany, October 2008, pp. 1851–1858.

[14] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2001, vol. 1.

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

[16] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Wiley-Interscience, 2004.

[17] ATLAS Collaboration, "Calorimeter performance technical design report," CERN, Report CERN/LHCC 96-40, january 1997.