17

Segmented Online Neural Filtering System Based On Independent Components Of Pre-Processed Information

Rodrigo Torres[1], Eduardo Simas Filho[1,2],
Danilo de Lima[1] and José de Seixas[1]
[1] *COPPE / Poli - Federal University of Rio de Janeiro, Brazil*
[2] *Federal Institute of Education, Science and Technology of Bahia, Brazil*

1. Introduction

Data filtering systems are used in different fields of research, aiming at isolating signals of interest from patterns related to a given background noise. Nowadays, in many complex applications, the input data space dimensionality is very high, as well as the incoming data rate. In this case, the difficulty of the input data stream analysis increases significantly. Also, the processing speed plays a critical role when the filtering system is envisaged for online operation. Finally, the signals of interest may rarely occur, forcing the experiment to keep running for a long period of time in order to acquire a reasonable amount of events for better measurement estimation.

In general, online filtering systems should have the following features:

- High detection efficiency for a low false alarm probability.
- Simplified software / hardware implementation.
- Flexibility in order to accomplish possible future requirements.
- Execution speed capable of meeting the desired time requirements.
- Robustness, in order to keep its filtering features through its lifetime of operation.

To cope with such high-input data dimension, feature extraction techniques may be applied in order to isolate the relevant information from the event data description, eventually reducing its dimension. For this, different data compaction techniques have been developed using expert information or / and stochastic processing. Pre-processing schemes based on signal decorrelation (linear or nonlinear) may even reduce the complexity of the classifier (signal against background) design. Finally, in the case where the available information is from a set of sensors, signal pre-processing might also be segmented, better exploiting the available local information.

Statistical processing can play a valuable role in the pattern recognition task, since it can provide better separation cuts than deterministic methods, specially for the case where the problem to be solved presents nonlinear characteristics. By using algorithms based on high-order statistics, it is possible to better estimate the bounds of each pattern, achieving higher detection efficiencies. In many applications, neural networks (Haykin, 2008) may play a role in

signal classification. On the other hand, by reducing the classifier design complexity by means of signal pre-processing, it might be possible to restrict the nonlinear processing implemented by the neural network to perform slight adjustments to the linear signal classification. It may also be the case where signal classification can go linear (through a Fisher Discriminant) (Duda et al., 2004), as a result of a highly-efficient pre-processing scheme.

In the field of experimental high-energy physics, stringent conditions make signal processing a challenge, as there is often a large gap between the experiment requirements and the technology currently available, which forces the development of new technologies. This is particularly the case for modern particle collider experiments, in which particles are accelerated at high speed and put in collision route. Analyzing the resulting collisions products, one can probe deeper into the structure of matter (Perkins, 2000). One important aspect in particle collider experiments is that events of interest are typically very rare, since most of the produced events are from background noise. In addition, the fine-grained segmentation of the particle detectors placed around the collision points for the resulting interaction readout may produce up to terabytes per second of information. Therefore, an online filtering system must be applied for selecting only the interesting physics channels, while rejecting, as much as possible, the huge amount of background noise.

Presently, the Large Hadron Collider (LHC) at CERN (CERN, 2007) is the largest particle accelerator in the world. LHC has a total length of 27 km and will be colliding protons with 14 TeV at their center of mass, at a rate of 40 MHz and at a luminosity of $10^{34}cm^{-2}s^{-1}$ (Evans and Bryant, 2008). Multiple collision points occur around the LHC ring. Around each collision point, a detection laboratory is placed to analyze the sub-products of the collisions. Among such detectors ATLAS (The ATLAS Collaboration, 2008) is the largest one. It comprises multiple sub-detectors, such as tracking, calorimeter and muon detection systems. Due to the detector granularity, each collision produces ~1.5 MBytes of information, resulting in a total rate of ~60 TB/s of information. Therefore, an online filtering system is mandatory for proper ATLAS operation.

This chapter focuses on proposing an efficient data filtering strategy for operating at stringent conditions. It is based on a signal processing scheme that combines expert knowledge with stochastic signal processing techniques for data dimension reduction and relevant feature extraction. The classifier design that implements the final filtering operation (rejection / acceptance of incoming data) is evaluated in terms of complexity and efficiency. For this, the input nodes of the classifier are fed from pre-processed information. The proposed signal processing strategy will be applied in high-energy physics, using the ATLAS detector as a case study.

This chapter is organized as follows: Section 2 briefly describes the pre-processing methods used in the application. Then, in Section 3, the ATLAS filtering system will be explained, and the envisaged application is presented. In Section 4, the obtained results for such application are discussed. Finally, conclusions are derived in Section 5.

2. Signal Pre-Processing

The pre-processing techniques presented in this section focus on performing linear and nonlinear input variable decorrelation. This could make the relevant discrimination features more evident to the classifier, simplifying its design. Furthermore, depending on the power

of the nonlinear decorrelation applied, the classifier could be simplified to the point of a simple linear discriminant.

2.1 Independent Component Analysis

Independent Component Analysis (ICA) is a multidimensional signal processing technique that searches for a linear transformation of the data, so that its essential structure becomes somehow more accessible (Hyvärinen et al., 2001). In ICA, the transformed variables are restricted to be statistically independent.

In the standard ICA model, the measured (observed) signals $\boldsymbol{x}=[x_1,x_2,...,x_N]^T$ are considered to be generated through a linear combination of the independent (unobserved) signals $\boldsymbol{s}=[s_1,s_2,...,s_N]^T$:

$$x_i = \sum_{j=1}^{N} a_{ij} s_j \rightarrow \mathbf{x}=\mathbf{A}\mathbf{s} \tag{1}$$

where $i=1,...,N$ and $\boldsymbol{A}$ is the mixing matrix. The ICA model has been widely applied in a variety of signal processing tasks, see as reference (Choi et al., 2005) and (Moura et al., 2009).

The purpose of ICA is to estimate the independent signals $\boldsymbol{s}$ and the mixing matrix $\boldsymbol{A}$ using only the observed data $\boldsymbol{x}$. This can be achieved through an inverse model:

$$\mathbf{y}=\mathbf{W}\mathbf{x}, \tag{2}$$

where the coefficients of the estimated matrix $\boldsymbol{W}$ are obtained by considering that the components of $\boldsymbol{y}$ are statistically independent (or at least as much independent as possible).

There are some indeterminacies in the ICA model: the order of extraction of the independent components can change and scalar multipliers (positive or negative) may modify the estimated components. Fortunately these limitations are insignificant in most applications.

In some practical signal processing problems, the standard linear ICA model may not be able to properly describe the data. Considering a practical ICA application, both the mixing environment and the sensors may present some nonlinear behavior. Providing a more general formulation, the nonlinear independent component analysis (NLICA) model considers that the measured signals $\boldsymbol{x}$ are formed by a nonlinear instantaneous mixing model (Almeida, 2006):

$$\mathbf{x}=F(\mathbf{s}) \tag{3}$$

where $F(.)$ is a $\boldsymbol{R}^N \rightarrow \boldsymbol{R}^N$ nonlinear mapping (the number of sources is usually assumed to be equal to the number of observed signals). The purpose of NLICA is to estimate an inverse transformation $G(.)$ $\boldsymbol{R}^N \rightarrow \boldsymbol{R}^N$:

$$\mathbf{y}=\mathrm{G}(\mathbf{x}) \tag{4}$$

so that the components of $\boldsymbol{y}$ are statistically independent. If $G(.)=F^{-1}(.)$, the sources are perfectly recovered (Jutten and Karhunen, 2003).

A characteristic of the NLICA problem is that the solutions are non-unique (Jutten and Karhunen, 2003). If $\boldsymbol{u}$ and $\boldsymbol{v}$ are independent random variables, it is easy to prove that $f(\boldsymbol{u})$

and *g(v)*, where *f(.)* and *g(.)* are differentiable functions, are also independent. So, it is clear that, without imposing some restrictions, there is an infinite number of solutions for the inverse mapping *G* in a given application (the problem is ill-posed). Considering this, an unique solution for the nonlinear independent component analysis (NLICA) can not be achieved without some prior information on the mixing model or the sources. A complete investigation on the uniqueness of nonlinear ICA solutions can be found in (Hyvärinen and Pajunen, 1999). NLICA algorithms have recently been applied in different problems such as speech processing (Rojas et al., 2003), (Wei et al., 2006), image denoising (Haritopoulos et al., 2002), chemistry sensor array processing (Duarte et al., 2009).
The minimization of statistical dependence is a main concern for any ICA/NLICA algorithm, as it leads to the estimation of the mixing system (and consequently the independent components). In addition, ICA often requires some pre-processing for data compaction, especially for high-dimension input data space applications. These topics are briefly described in the next subsections. It is also summarized the JADE algorithm, which has been widely used for independent component estimation. Different NLICA approaches are also briefly reviewed.

2.1.1 Statistical Independence

Considering two random vectors v_1 and v_2, they are statistically independent if and only if (Papoulis and Pillai, 2002):

$$p_{\mathbf{v}_1,\mathbf{v}_2}(\mathbf{v}_1,\mathbf{v}_2) = p_{\mathbf{v}_1}(\mathbf{v}_1)p_{\mathbf{v}_2}(\mathbf{v}_2) \tag{5}$$

where $pv_1(v_1)$, $pv_2(v_2)$ are, respectively, the probability density function (pdf) of v_1 and v_2 and pv_1,v_2 (v_1, v_2) is their joint pdf. In typical ICA 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 are usually applied during the search for independent components (Hyvärinen et al., 2001) (Cichocki and Amari, 2002). They are defined next for reference.

2.1.1.1 Nonlinear Decorrelation

Two zero-mean random variables (u_1 and u_2) are said to be (linearly) uncorrelated if their cross-correlation R_{u1u2} is zero (here, $E\{.\}$ is the expectation operator):

$$R_{u_1u_2} = E\left\{u_1 u_2^T\right\} \tag{6}$$

Independent variables are uncorrelated, although, the reciprocal is not always true. Linear correlation is verified by second order statistics, while independence needs higher-order information too (requiring direct or indirect computation of higher-order moments).
Considering $g(u_1)$ and $f(u_2)$ absolutely integrable functions of u_1 and u_2, respectively, it can be proved that if Equation 7 holds for all possible *g(.)* and *f(.)*, than u_1 and u_2 are independent

$$E\left\{g(u_1)f(u_2)\right\} = E\left\{g(u_1)\right\}E\left\{f(u_2)\right\} \tag{7}$$

By choosing *g(.)* and *f(.)* as nonlinear functions, high-order statistical information is (indirectly) accessed. The statistical independence measure provided by Equation 7 is usually called the nonlinear decorrelation between u_1 and u_2 (Cichocki and Unbehauen, 1996).

A practical limitation appears when trying to apply this measure in an ICA algorithm as it is not possible to check all integrable functions *g(.)* and *f(.)*. Thus estimates of the independent components are usually obtained while guaranteeing nonlinear decorrelation between a finite set of nonlinear functions (Cichocki and Unbehauen, 1996).

2.1.1.2 Higher-Order Statistics

Another principle that can be used to estimate the dependence of variables comes from the central limit theorem (McClave et al., 2008): "The sum of two random variables is always closer to a Gaussian distribution than the original variable distributions". As the measured signals (***x***) are considered to be a linear combination of independent sources (***s***), then the measured signals are closer to a Gaussian distribution than the original sources. Thus, the independent components can be obtained through maximization of non-gaussianity (Hyvärinen et al., 2001).

It is known that, for Gaussian random variables, the cumulants of orders higher than two are all zero. Considering this, non-gaussianity (and consequently independence) measures can be obtained by using high-order cumulants, such as the kurtosis K_4, which, for a zero-mean, unit-variance random variable u is defined through (Papoulis and Pillai, 2002):

$$K_4 = E\{u^4\} - 3\left[E\{u^2\}\right]^2 \tag{8}$$

2.1.1.3 Information Theoretic Measures

Alternative statistical independence measures can be obtained from information theory (Mackay, 2002). A basic definition in information theory is the entropy (*H(.)*), which, for a discrete random variable u, is defined as (Shannon, 1948):

$$H(u) = -\sum_i P(u = \kappa_i)\log P(u = \kappa_i) \tag{9}$$

Where κ1, κ2, ..., κ m are all the possible discrete values of u. is that the Gaussian variable has maximum entropy between the random variables of same variance (Hyvärinen et al., 2001). Considering this, entropy can be used as gaussianity measure.

The Negentropy *J(u)* of the random variable u is also applied in the ICA context:.

$$J(u) = H(u_{gauss}) - H(u) \tag{10}$$

where u_{gauss} is a Gaussian random variable with the same mean and variance of u. The advantage of using *J(u)*, instead of *H(u)*, is that it is always non-negative and zero when u is Gaussian. A problem with the computation of both *J(.)* and *H(.)* is the pdf estimation. To avoid this, approximations using high-order cumulants or non-polynomial functions are often applied (Murillo-Fuentes et al., 2004).

The Mutual Information $I(u_1, u_2, ..., u_m)$ between m random variables $u_1, u_2, ..., u_m$ is obtained through Equation 11.

$$I(u_1, u_2, \ldots, u_m) = \sum_{i=1}^{m} H(u_i) - H(\mathbf{v}) \tag{11}$$

It is known that the entropy of the vector v = [u_1, u_2, ..., u_m] is always smaller than the sum of $H(u_i)$, unless the variables are independent. So, minimization of mutual information leads to independence (Hyvärinen et al., 2001).

2.1.2 Signal Decorrelation

The standard ICA model assumes a mixing system where the number of sources and observed signals is the same. In a practical problem, this assumption may not be always true. When there exist more sources than sensors (observed signals), the problem is under-determined and the sources are only recovered approximately through algorithms derived for such situation (Syskind et al., 2006), (Natora et al., 2009). In the case where the number of sources (K) is smaller than the number of observed signals (N), the problem is over-determined and thus some signal compaction algorithm is needed to reduce signal dimensionality. With this purpose, Principal Component Analysis (PCA) is usually applied as a pre-processing for ICA algorithms. Principal Components for Discrimination (PCD) analysis (Caloba et al., 1995) has been introduced as an alternative to PCA, when ICA is applied to classification problems (Simas Filho et al., 2009b).

2.1.2.1 Principal Component Analysis

Principal Component Analysis (PCA) (Jolliffe, 2002) is a statistical signal processing technique that searches for a new representation of the input signals where the energy is concentrated on a small number of components. Using second-order statistics, PCA transformation searches for a vector basis for which the projections $y_i = w_i\, x_i$ of a zero-mean random vector $\boldsymbol{x}$ ($E\{x\}=0$) are uncorrelated and have maximum variance (i.e. composing an orthonormal basis).

The first principal direction $\boldsymbol{w_1}$ can be computed through the maximization of

$$J_1^{PCA}(w_1) = E\{v^2\} = E\{(w_1 x)^2\} = w_1^T C_x w_1 \tag{12}$$

where Cx is the covariance matrix of x and $||w_1||=1$.

PCA transformation is very useful as a pre-processing for ICA as it eliminates second-order dependencies (correlation) between the signals, facilitating the search for independence.

2.1.2.2 Whitening

A zero-mean random vector $\boldsymbol{z}$ is said to be white if their components are uncorrelated and have unit variance (Hyvärinen et al., 2001). This implies that the covariance matrix (and also the correlation matrix) of $\boldsymbol{z}$ equals the identity matrix. Whitening is sometimes called sphering and is a slightly stronger operation than decorrelation. One popular method to perform whitening is the eigenvalue decomposition (EVD) of the covariance matrix (Strang,

2009). In this approach, considering $\boldsymbol{Z}$ the matrix whose columns are the unit-norm eigenvectors of the covariance matrix $\boldsymbol{C}_x$ of a random vector $\boldsymbol{x}$ and $\boldsymbol{D}$ the diagonal matrix of the eigenvalues of $\boldsymbol{C}_x$, the linear whitening transform $\boldsymbol{V}$ is given by:

$$\mathbf{V} = \mathbf{D}^{-1/2}\mathbf{Z}^{T} \tag{13}$$

2.1.2.3 Principal Components of Discrimination

Considering a classification problem, the purpose of PCD analysis is to determine the directions that maximize class separation (Caloba et al., 1995). Typically, PCD provides a higher compaction rate for classification problems with respect to PCA (Simas Filho et al., 2009b).

The PCD analysis can be performed through a Multilayer Perceptron (MLP) neural network (Haykin, 2008). For simplicity, considering binary discrimination, a network with a single hidden neuron, trained to maximize class discrimination, extracts the first discriminating component (see Fig. 1-a). By sequentially adding neurons to the hidden layer and restarting the training procedure, the next components are estimated. The hidden weights are trained only for the added neurons (highlighted synaptic lines in Fig. 1-b). 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 whole training procedure for optimal combination of principal components at each processing step. The PCD extraction continues up to the point where the classification efficiency does not improve significantly by adding more components.

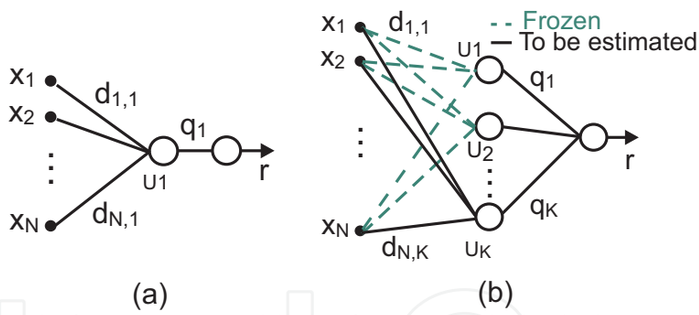


Fig. 1. Neural models for estimating (a) the first and (b) the k-th principal discriminating component.

2.1.3 ICA / NLICA algorithms

Nowadays, there is a number of efficient ICA algorithms, which use, in general, the independence measures mentioned in Section 2.1.1. Among them, JADE (Cardoso and Souloumiac, 1993) is a very popular method. For NLICA, one way to address the ill-posedness of the problem is to restrict the range of allowed nonlinearities, generating structural constrained models for the mixing system and thus unique solutions for the problem (Jutten and Karhunen, 2003). Among these models, we can mention the post-nonlinear (PNL) mixture, which has met a significant practical applicability (Almeida, 2006). There is also a method closely related to the NLICA problem, known as Local ICA, which introduces nonlinear transformations by clustering the dataset into groups of similar

characteristics. After that, linear ICA is applied to data belonging to each cluster producing local independent components.

2.1.3.1 JADE

In JADE (Joint Approximate Diagonalization of Eigenmatrices) algorithm, second and fourth-order statistics are applied for independent component estimation through a tensorial approach. The second-order cumulant (i.e. the covariance matrix) is used to ensure that data are white (uncorrelated). Fourth-order information (through the fourth-order cumulant tensor matrix) produces an independence criterion.

Tensors are considered as a higher-dimensional generalization of matrices or linear operators (Michal, 2008). Cumulant tensors are matrices containing the cross-cumulants. Considering this, the second-order cumulant tensor is the covariance matrix and the fourth-order tensor ($\boldsymbol{T_4}$) is formed by the fourth-order cross-cumulants $cum(u_i,u_j,u_k,u_l)$, which, for zero-mean random variables, is defined as:

$$cum\left(u_i,u_j,u_k,u_l\right)=E\left\{u_i,u_j,u_k,u_l\right\}-E\left\{u_i,u_j\right\}E\left\{u_k,u_l\right\}-E\left\{u_i,u_k\right\}E\left\{u_j,u_l\right\}-E\left\{u_k,u_j\right\}E\left\{u_i,u_l\right\} \quad (14)$$

The fourth-order cumulant tensor $\boldsymbol{T_4}$ is a four-dimensional array, where, for each element $q_{ijkl}= cum(u_i,u_j,u_k,u_l)$, the indexes i ,j ,k ,l vary from 1 to N (where N is the number of signals). The fourth-order cumulant tensor contains all fourth-order information of the data.

JADE estimation criterion is derived through a procedure analogous to diagonalization of the covariance matrix, which produces signal decorrelation. As $\boldsymbol{T_4}$ is a fourth-order counterpart of the covariance matrix, independence can be achieved by diagonalizing $\boldsymbol{T_4}$, as for independent signals the unique non-zero fourth-order cross-cumulant appears when $i=j=k=l$. Analogous to the second-order case, diagonalization of the fourth-order tensor can be achieved through eigenvalue decomposition (EVD) (Strang, 2009).

Using tensorial methods for ICA is theoretically simple, but computing EVD of four-dimensional matrices by ordinary algorithms requires a very large amount of memory and may be computationally prohibitive in some cases. In order to avoid this limitation, methods like JADE were proposed in the literature. JADE algorithm searches for the matrix $\boldsymbol{W}$ that minimizes the sum of the squares of the non-diagonal elements of the output data of $\boldsymbol{T_4^{(y)}}$ (where $\boldsymbol{T_4^{(y)}}$ is the fourth-order cumulant tensor of the output data $\boldsymbol{y}$).

2.1.3.2 Post-Nonlinear ICA

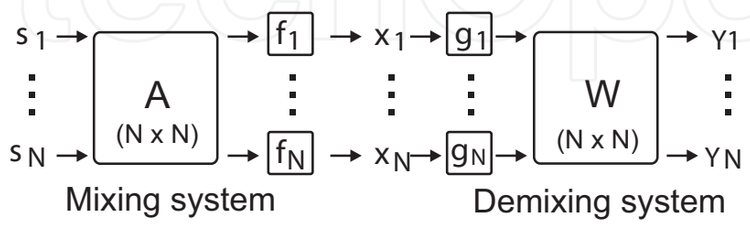


Fig. 2. Post-Nonlinear mixing/de-mixing model.

Post-Nonlinear (PNL) mixtures arise whenever, after a linear mixing process, the sensors present nonlinear behavior. The observed signals can be expressed as (Almeida, 2006):

$$x_i = f_i(\alpha_i) \tag{15}$$

where $\boldsymbol{\alpha}=\boldsymbol{As}$. As stated in Eq. 15, each observed signal x_i is obtained through component-wise nonlinear functions f_i applied to the linearly mixed signals $\boldsymbol{\alpha}_i$. The independent components are obtained by a mirror model:

$$y_i = \mathbf{W}_i g_i(x_i) \tag{16}$$

where $\boldsymbol{W}$ is the de-mixing matrix and g_i the inverse nonlinearity (see Fig. 2). The nonlinear functions are usually estimated through neural networks (MLP) and the de-mixing matrix by a linear ICA algorithm (Taleb and Jutten, 1999).

A limitation of the PNL algorithm is that the number of observed signals is assumed to be equal to the number of sources (square model). This prevents its application to high-dimensional data problems as both the number of parameters and the computational complexity increase exponentially with problem dimensionality.

In order to deal with high-dimensional data, a modified PNL model for the overdetermined case (when there exist more sensors N than sources K) was proposed in (Simas Filho et al., 2009a). As illustrated in Fig. 3, a linear block $\boldsymbol{B}$ is added to the standard PLN mixing model, allowing K<N. Coefficients of matrix $\boldsymbol{B}$ are estimated through signal compaction methods such as PCA and PCD, described in Section 2.1.2. The inverse (demixing) algorithm is thus described using a mirror model: $\boldsymbol{y}=\boldsymbol{W}\,G(\boldsymbol{Dx})$, where $\boldsymbol{y}$ are the estimated sources.

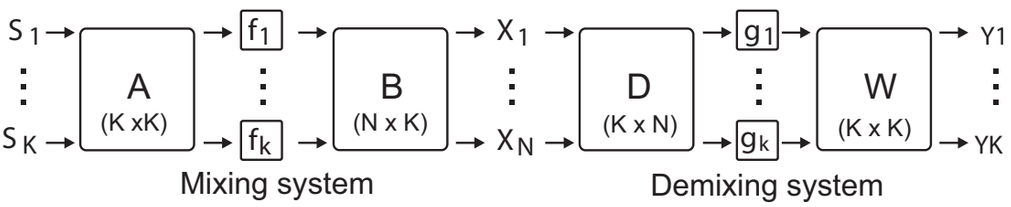


Fig. 3. Modified Post-Nonlinear mixing/de-mixing model.

2.1.3.3 Local ICA

Local ICA (Karhunen et al., 2000), (Jutten and Karhunen, 2003) can be viewed as a compromise between linear and nonlinear ICA. If the ICA model is used for feature extraction, better description of the data set can be obtained while exploring local characteristics. The purpose is to obtain better data representation when compared to linear ICA, while avoiding the high computational cost of the nonlinear models.

In Local ICA model (see Equation 17), a N-dimensional input space $Q \subset R^N$ is divided into a finite number of subsets Q_l, $l=1,...,L$, which satisfy:

$$Q_1 \cup Q_2 \cup \ldots \cup Q_L = Q \tag{17}$$

Clustering is responsible for the overall nonlinear representation. Linear ICA models are applied to data belonging to each cluster ($\boldsymbol{x}^{(l)}$) in order to estimate the local independent components $\boldsymbol{s}^{(l)}=\boldsymbol{B}^{(l)}\boldsymbol{x}^{(l)}$, where $\boldsymbol{B}^{(l)}$ is a local de-mixing matrix.

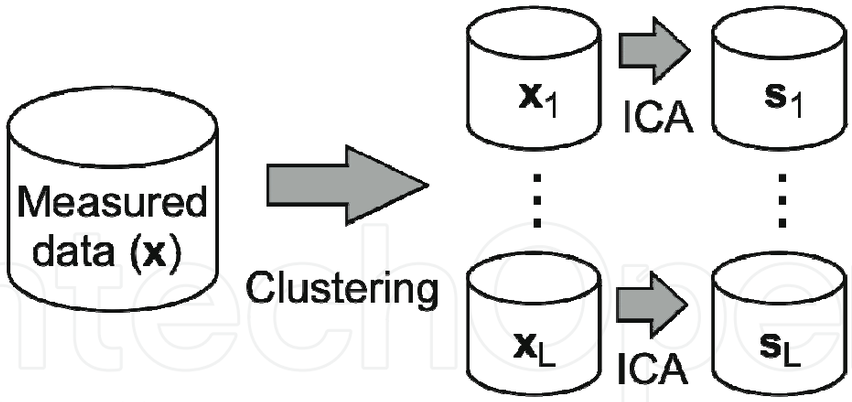


Fig. 4. Local ICA model.

3. The Application

As mentioned in Section 1, the LHC collision rate, together with ATLAS granularity, will result in a data stream of ~60 TB/s, requiring an efficient online filtering system for retrieving the interesting physics channels from dense background noise. This filtering system comprises three cascaded operation levels, applying successive cuts to the incoming data (Riu et al., 2008).

The first level (L1) will receive full data and will reduce the input event rate to ~75 kHz. The first level is responsible for marking the regions in the detector that have effectively been excited. These regions are known as *Region of Interest* - RoI, and will be the only information passed over to the second level analysis.

The second level (L2) will receive the regions of interest marked by the first level and will apply more specific analysis on them. For coping with an average processing time of 40 ms per event, a set of 500 off-the-shelf server processors will be employed, providing a multi-processed environment. The third and last filtering level, also known as Event Filter - EF, will take the final decision on events approved by the previous levels. A highly parallel processing environment composed by ~1600 off-the-shelf server processors will be employed for coping with an average processing time of 4 sec. At the end, a rate of ~200 Hz events will be recorded in mass storage devices for further offline analysis by the physicists.

One of the ATLAS main research goals is to experimentally prove the Higgs boson (Perkins, 2000). Being the Higgs boson highly unstable, it soon decays into more stable particles. Therefore, the physicists will prove its existence not by detecting the Higgs boson directly, but by analyzing its decaying signatures. It is known (Perkins, 2000) that some of such signatures produce electrons at their final state. Therefore, the identification of electrons is of great importance. On the other hand, during proton-proton collisions, a cascade of quark and anti-quark pairs can be produced, which quickly merge into more stable particles, producing a pattern known as jet. These jets may interact with the detector in a manner very similar to electrons, making the correct identification of electrons a tricky process.

Our analysis will focus on the electron / jet separation problem at the second level of the ATLAS filtering system, using calorimeter information. Calorimeters are total absorption detectors (Wigmans, 2000). Typically, they use a (passive) material (iron, lead, for instance) for absorbing entirely the energy of the incoming particle and sample the energy being deposited in the detector by using an active material (scintillating fibers, tiles, for instance).

Calorimeters play a major role in collider experiments as they provide fast response, their energy resolution improves with increasing energy, and they interact with charged and non-charged particles. In addition, they are highly-segmented detectors, so that it is possible to identify particle classes by their energy deposition profile.

The ATLAS calorimetry system is composed by two calorimeter sections (The ATLAS Collaboration, 2008). The electromagnetic (EM) calorimeter is responsible for detecting electrons, positrons and photons. The hadronic calorimeter (HD) is responsible for detecting hadrons (kaons, pions, etc) and it is placed on top of the electromagnetic calorimeter. Both detectors comprise 3 sequential layers with distinct granularity and depth, providing detailed information of incoming particles. The electromagnetic calorimeter has, in addition, a very thin layer in front of it, which is called the pre-sampler (PS). Fig. 5 displays the energy deposition profile for an electron interaction. A region of interest selected by the first-level filtering system amounts, in average, to 1,000 calorimeter cells.

Electrons have the property of depositing their energy in a very punctual way, differently from jets, which, for L2 data, tend to slightly spread their energy over multiple cells within a layer. Therefore, the relevant information relies not at the impact point center, but at its surrounding area. Aiming at exploiting this feature, a topological pre-processing based on ring sums has been tried by some L2 algorithms for data formatting (Torres et al., 2008). In this approach, the cell that samples the highest energy value (also known as the *hottest cell*) is considered the center of our region of interest in each calorimeter layer (seven in total). Then, a set of concentric rings are built around this hottest cell in a pattern similar to the one presented in Fig. 6. It can also be observed in Fig. 6 that, depending on the layer granularity, the rings might not close (incomplete) or even be composed only by strips. Finally, the cells belonging to a given ring are summed up, reducing the final event dimension, without jeopardizing their physics interpretation. This ring procedure is performed on a per layer basis, resulting, at the end, in a total of 100 rings, distributed as shown in Tab. 1.

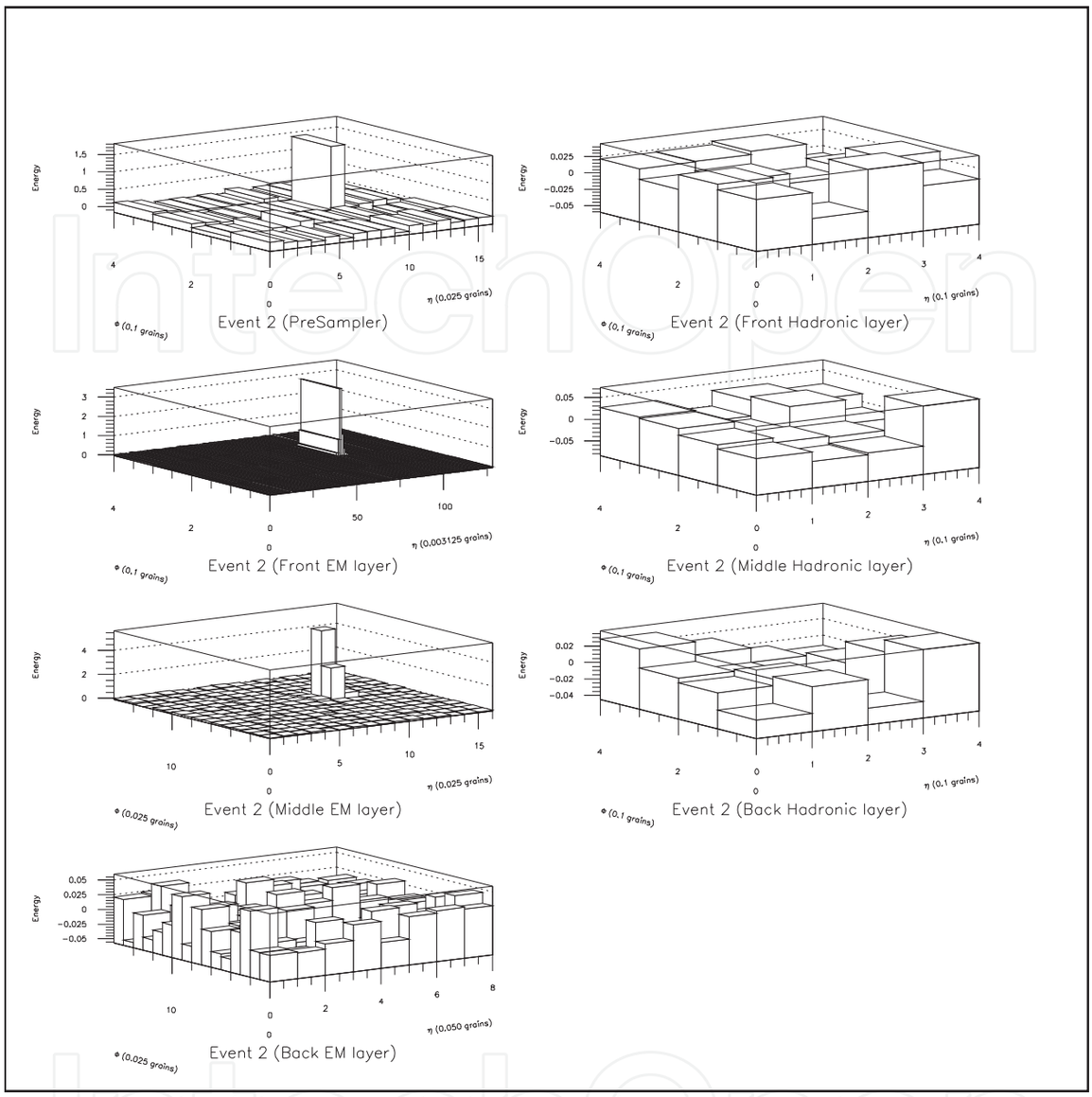


Fig. 5. Example of the segmented calorimeter information obtained from an incoming electron.

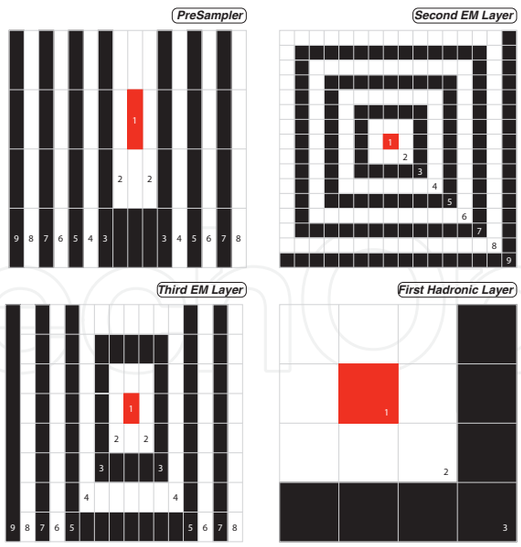


Fig. 6. Ring formatting for calorimetry.

Layer	PS	EM1	EM2	EM3	HD1	HD2	HD3	Total
Rings	8	64	8	8	4	4	4	100

Table 1. Number of rings in each calorimeter layer.

In the electron / jet separation problem, the dynamic range of the sampled energy is very large, therefore, an energy normalization procedure is applied, in order to focus, as much as possible, our analysis at the signal shape, rather than its energy nominal value, resulting in a steady detection efficiency over all the relevant energy spectrum. Also, since the relevant information from the discrimination point of view is known to be off-center, a sequential normalization is employed (dos Anjos et al., 2006). In this procedure, for each calorimeter layer, the normalized energy (E_N) of each ring is given by

$$E_{N_{l,i}} = \frac{E_{l,i}}{E_{tot_l} - \sum_{j=1}^{i-1} E_{l,j}} \tag{18}$$

where $E_{l,i}$ is the original energy of the i-th ring belonging to the l-th layer, and $E_{tot\ l}$ is the total sampled energy by the l-th layer. As a result, successively smaller attenuation factors are applied to the outer rings, but the normalization procedure is resilient enough to keep track of the signal-to-noise ratio, avoiding the amplification of irrelevant information.

4. Results

The available dataset was obtained through Monte Carlo simulation for proton-proton collisions and comprises approximately 470,000 electrons and 310,000 jet signatures. The simulation considers the detector characteristics and the first-level filtering operation. The available data set was approximately equally split into training, validation (stopping

criterion for neural network training based on mean-squared error) and testing (performance evaluation) sets.

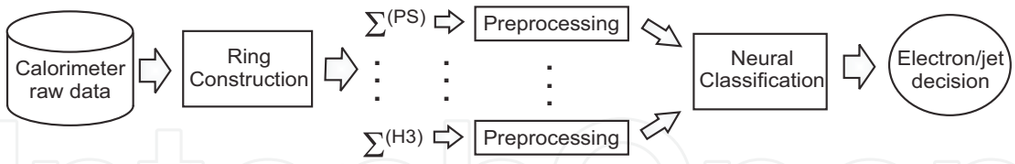


Fig. 7. - Processing chain of the electron/jet separation system.

It is shown in Fig. 7 the block diagram of the electron / jet discriminator. The raw calorimeter data is received and the topological processing based on ring sums is performed. Next, for the segmented case, the rings belonging to a given layer are pre-processed individually and the pre-processed event obtained for each layer is concatenated, generating a single input, which is propagated to the neural network for the pattern recognition. For the non-segmented case, the generated rings are concatenated prior to the pre-processing phase, so that the pre-processing is performed in all 100 rings at once. For the ICA based pre-processing, the JADE algorithm was used, and the clustering algorithm used by local ICA was the k-mean (Duda et al., 2004).
In this work, the Fisher Linear Discriminant (FLD) and supervised Multi-Layer Perceptron (MLP) neural classifiers (single hidden layer) (single hidden layer) were used to perform particle identification over calorimeter information. The neural networks were trained using the Resilient Backpropagation algorithm (Riedmiller and Braun, 1993). In order to compare the discrimination efficiency for the proposed classifiers, both the Receiver Operating Characteristics (ROC) and the SP index were applied. The ROC curve (Van Trees, 2003) shows how the detection probability P_D and false alarm probability P_F vary as the decision threshold changes. The SP index (dos Anjos et al., 2006) is computed through

$$SP = \frac{P_D + P_J}{2} \times \sqrt{\left(P_D \times P_J\right)} \tag{19}$$

where P_J is the efficiency for jets. The threshold value that maximizes the SP provides both high P_D and low P_F.

As mentioned in previous sections, the available calorimeter signatures are topologically pre-processed, generating 100 rings for an incoming event. Considering this, the discrimination system may benefit from signal compaction algorithms as they reduce redundant information and signal dimensionality. Here, compaction was performed through both Principal Component Analysis (PCA) and Principal Components for Discrimination (PCD), using segmented (layer-level) and non-segmented approaches. As the calorimeter system provides highly-segmented information, segmented processing tries to exploit subtle differences in electron and jet energy deposition profiles, which are available at the layer level.

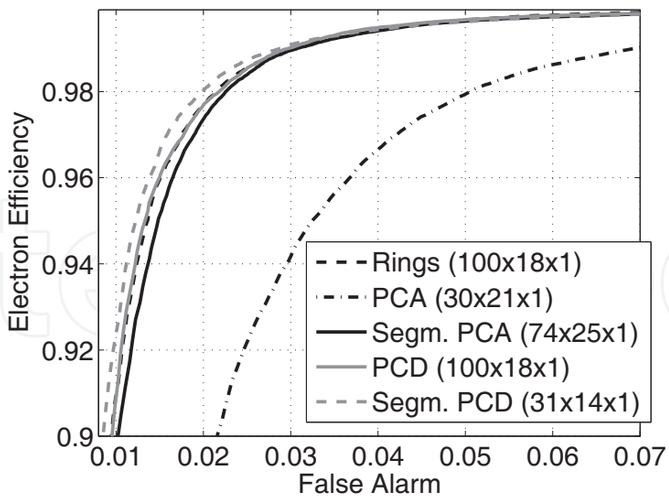


Fig. 8. ROC curves (and respective classifier topology) for segmented and non-segmented feature extraction through PCA and PCD.

Fig. 8 illustrates the discrimination performance for different methods in terms of ROC curves. It can be observed that the segmented approach outperforms the non-segmented one, for both PCA and PCD. It can also be seen that PCD usually presents lower false-alarm when compared to PCA (for the same detection probability) and achieves higher compaction rates (31 components for PCD against 74 components for PCA in segmented processing mode). This is a result of the compaction strategies, as in PCA the purpose is to maximize the energy projection and in PCD the objective is to optimize the discrimination capability of the components. Moreover, PCD uses nonlinear processing to estimate its components, which proves to be efficient in terms of discrimination performance. As it can be seen from Fig. 8, using only 31 components, the PCD performance is even better than processing 100 rings without any further pre-processing.

The (linear) Independent Component Analysis (ICA) model was without any further pre-processingalso applied to ring signals, either without pre-processing or combined with segmented and non-segmented PCA and PCD compaction schemes. Fig. 9 illustrates the ROC curves for different ICA-based discriminators. It can be observed that the segmented feature extraction provides slightly higher discrimination performance when independent components are estimated. Other benefit observed with ICA is that the classifier training procedure usually converges in very few iterations, in contrast to PCA and PCD based discriminators, which, in general, require a larger number of training steps. From Fig. 9, it is also interesting to observe that ICA could be the only pre-processing technique, as the nonlinear decorrelation it provides allows the neural network to perform slightly better in terms of discrimination efficiency.

Considering feature extraction through NLICA (using the modified PNL model) based on PCD projection, the nonlinearities which may arise are expected to be smooth. In a practical design, a calorimeter can exhibit small nonlinearities along the wide dynamic range it has to work on. In view of this, the neural networks used to estimate the inverse nonlinearities are restricted to have small number of hidden neurons and thus can only approximate smooth

nonlinear functions. This also reduces the probability of reaching local minima during the training procedure.

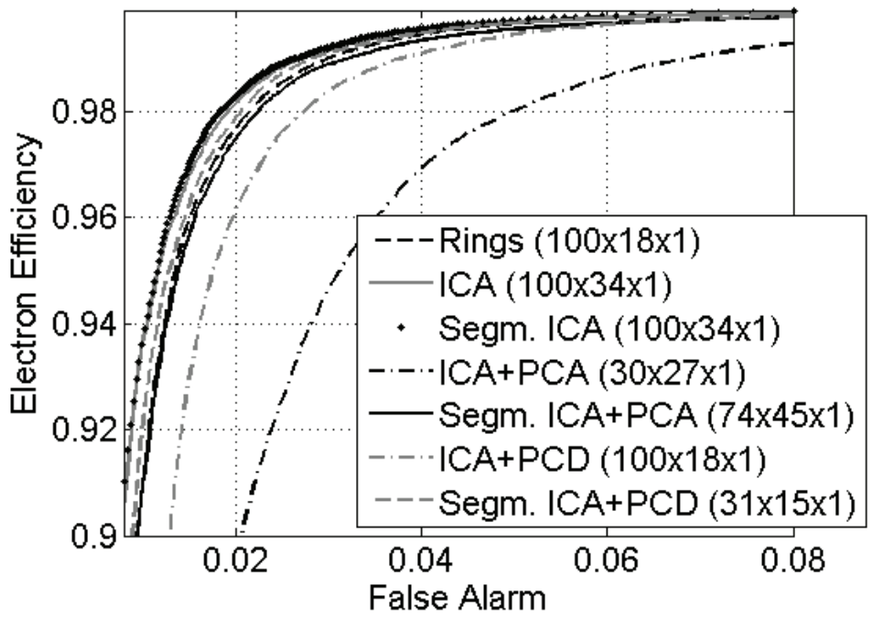


Fig. 9. ROC curves (and respective classifier topologies) for ICA-based discriminators.

In the Local ICA approach, the training data set was initially clustered into two clusters (as there are two possible classes for the incoming particles). As illustrated in Fig. 10, cluster 1 concentrates most of the electron signatures and cluster 2 the jets. After clustering, ICA and ICA with PCD pre-processing were both estimated for data belonging to each cluster. The classifiers were also trained locally, generating two ROCs (one for each cluster). A global optimization algorithm (Genetic Algorithm) (Haupt and Haupt, 2004) was used to search for the optimal combination of the local thresholds, which provides optimum global discrimination.

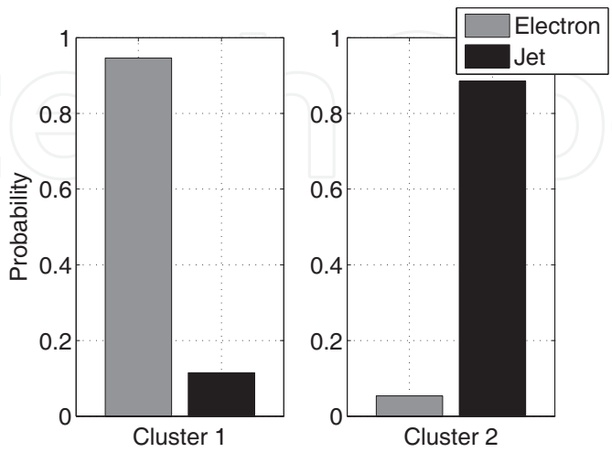


Fig. 10. Concentration of electrons and jets in each cluster for Local ICA pre-processing approach.

Fig. 11 illustrates the discrimination performance obtained through PNL and Local ICA approaches. It can be seen that, compared to the linear ICA model, PNL exhibits slightly poorer performance. On the other hand, Local ICA produces higher discrimination efficiency with respect to the other models when it is performed on PCD directions. For Local ICA, only the optimum point is shown in Fig. 11.

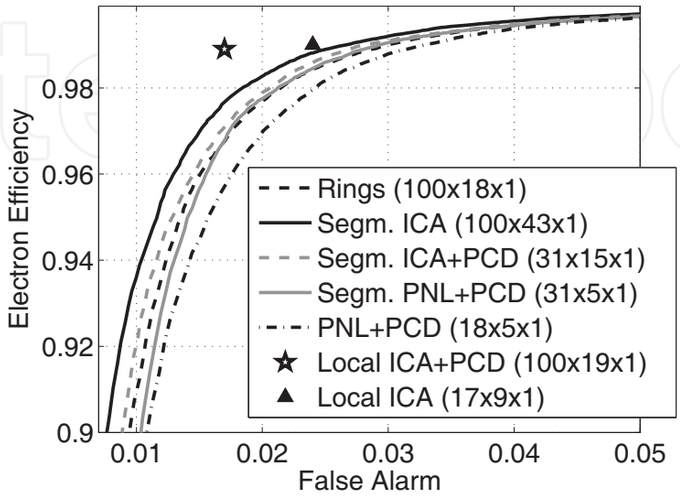


Fig. 11. ROC curves (and respective classifier topology) for NLICA and ICA discriminators.

A summary of the results achieved for each approach can be observed in Tab. 2, where the maximum SP value obtained for each approach is presented. Furthermore, Tab. 3 presents the false alarm probability for a fixed 97% electron detection efficiency.
The classifier complexity is shown in Tab. 4 for each approach. It can be depicted from this table that applying nonlinear decorrelation (PCD and ICA based algorithms) reduces the computational requirements for the classification task. The local ICA based on PCD projections not only achieves better classification efficiency, but it is also very efficient in terms of computational load (~33% reduction with respect to the rings only approach).

Approach	Non-segmented	Segmented
Rings	96.10	---
PCA	93.07	96.04
PCD	96.11	96.28
ICA	96.38	96.45
ICA + PCA	93.21	96.00
ICA + PCD	95.45	96.25
PNL + PCD	95.80	96.20
Local ICA	96.63	---
Local ICA + PCD	97.32	---

Table 2. Maximum SP (× 100) obtained for segmented and non-segmented approaches.

Approach	Non-segmented	Segmented
Rings	1.75	---
PCA	4.20	1.88
PCD	1.77	1.59
ICA	1.55	1.51
ICA + PCA	4.04	1.83
ICA + PCD	2.26	1.68
PNL + PCD	2.02	1.72
Local ICA	1.30	---
Local ICA + PCD	0.82	---

Table 3. False alarm probability (%) for a detection efficiency of 97%.

Approach	Non-segmented	Segmented
Rings	3636	---
PCA	4242	5050
PCD	3636	2828
ICA	6868	8686
ICA + PCA	5454	9090
ICA + PCD	200	3030
PNL + PCD	3934	6768
Local ICA	4438	---
Local ICA + PCD	2418	---

Table 4. Number of total floating point operations per approach.

In order to verify whether a linear classifier suffices, the pre-processed signals were used to feed a linear Fisher Discriminant (FLD), which is proved to be optimal linear discriminators (Duda et al., 2004). Fig. 12 provides a comparison between the discrimination performance obtained through linear (FLD) and nonlinear (MLP) classifiers. It can be seen that the nonlinear decorrelation introduced by ICA was able to improve the discrimination obtained through FLD, providing more separated patterns for different types of particles. Through the proposed pre-processing chain, the results of the linear classifier got closer to the ones obtained by the MLP. However, the nonlinear decorrelation provided by PCD and ICA were still not sufficient to discard a nonlinear classifier in order to achieve optimal detection efficiency. Tab. 5 and Tab. 6 summarize the detection efficiency comparison between linear Fisher and neural discriminants. An important issue is the computational cost, which, for a linear classifier, is much smaller (see Tab. 7) with respect to the nonlinear counterpart. This might be a striking advantage for online filtering.

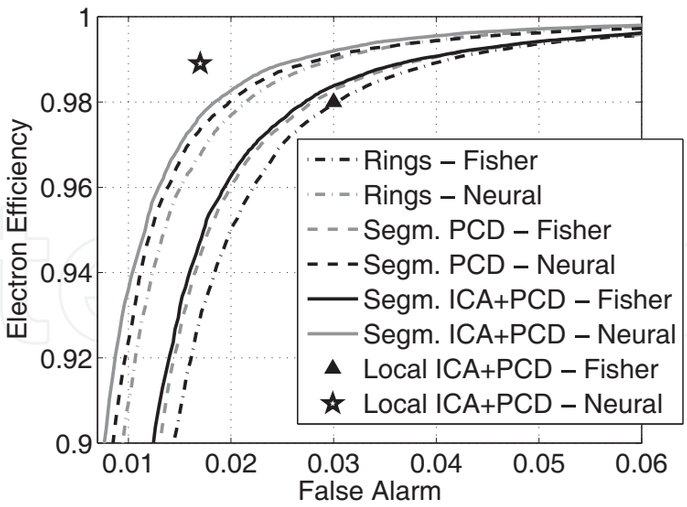


Fig. 12. ROC curves for the linear and neural discriminators.

Approach	Neural	Fisher
Rings	96.10	95.12
Segm. PCD	96.28	95.35
Segm. ICA + PCD	96.25	95.43
Local ICA + PCD	97.32	94.40

Table 5. Maximum SP (× 100) obtained for each approach considered for a linear classifier.

Approach	Neural	Fisher
Rings	1.75	2.50
Segm. PCD	1.59	2.29
Segm. ICA + PCD	1.68	2.22
Local ICA + PCD	0.82	3.00

Table 6. False alarm (%) for a detection efficiency of 97% for each approach considered for a linear classifier.

Approach	Neural	Fisher
Rings	3636	200
Segm. PCD	2828	200
Segm. ICA + PCD	3030	200
Local ICA + PCD	2418	700

Table 7. Number of total floating point operations per approach considered for a linear classifier.

5. Conclusions and Perspectives

Online data filtering in high-dimensional input data space finds application in multiple areas. Depending on the accumulated knowledge about the target problem, combining what is known by experts with high-order stochastic signal processing techniques is being shown to be an efficient design approach. Among the benefits, high signal compaction rates, relevant feature extraction and reduced computational load are often accomplished.

For a very demanding high-energy physics application, it was shown that we could benefit from topological pre-processing, which implements the expert part of the whole pre-processing scheme. Then, adding decorrelation techniques to the signal processing chain provided efficient feature extraction. For this, only 30% of the original data components were required. Further knowledge about the problem pointed out that the segmented signal processing was the right approach. In addition, the overall computational load could significantly be reduced, which was attractive due to the low processing time required by the target application.

Nonlinear independent component analysis achieved the best performance in this case study, which motivates further application investigations. A possibility is to implement it through SOM (Self-Organizing Maps) (Haykin, 2008). In detectors where the arising nonlinearities of practical designs are expected to be small deviations from the linear behavior, it would also be important to restrict the degrees of freedom of the nonlinear component extraction. The independent component analysis is also attractive in facing pile-up effects (Knoll, 1989), which typically decreases discrimination efficiencies in high event rate applications. There is plenty of room for algorithm development in high demanding application scenarios.

6. Acknowledgements

The authors would like to express their gratitude to CNPq, FINEP, CAPES, FAPERJ (Brazil) and CERN (Switzerland) for their financial support. We also thank the ATLAS collaboration at CERN for providing the simulated calorimeter data and for fruitful discussions concerning this work.

7. References

Almeida, L. B. (2006). *Nonlinear Source Separation*, Morgan and Claypool.

Caloba, L., Seixas, J. and Pereira, F. (1995). Neural discriminating analysis for a second-level trigger system, *Proceedings of the International Conference on Computing in High Energy Physics (CHEP95)*, Rio de Janeiro, Brazil.

Cardoso, J. F. and Souloumiac, A. (1993). Blind beamforming for non-gaussian signals, *IEEE Proceedings*- F 140(6): 362–370.

CERN (2007). European organization for nuclear research. URL: http://www.cern.ch

Choi, S., Cichocki, A., Park, H. and Lee, Y. (2005). Blind source separation and independent component analysis - a review, *Neural Information Processing - Letters and Reviews* 6(1).

Cichocki, A. and Amari, S. (2002). *Adaptive Blind Signal and Image Processing*, Willey.

Cichocki, A. and Unbehauen, R. (1996). Robust neural networks with on-line learning for blind identification and blind separation of sources, *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications* (11).

dos Anjos, A., Torres, R. C., Seixas, J. M., Ferreira, B. C. and Xavier, T. C. (2006). Neural triggering system operating on high resolution calorimetry information, *Nuclear Instruments and Methods in Physics Research* 559: 134–138.

Duarte, L. T., Jutten, C. and Moussaoui, S. (2009). Ion selective electrode array based on a bayesian nonlinear source separation method, in T. Adali, C. Jutten, J. Romano and A. Barros (eds), Independent Component Analysis And Signal Separation, 8th International Conference, *Lecture Notes In Computer Science*, Springer, Paraty, Brazil, pp. 662–669.

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

Evans, L. and Bryant, P. (2008). LHC machine, *Journal of Instrumentation* (2008 JINST 3 S08001).

Haritopoulos, M., Yin, H. and Allinson, N. M. (2002). Image denoising using self-organizing map-based nonlinear independent component analysis, *Neural Networks* pp. 1085–1098.

Haupt, R. L. and Haupt, S. E. (2004). *Practical Genetic Algorithms*, 2nd ed,Wiley-Interscience.

Haykin, S. (2008). *Neural Networks and Learning Machines*, 3rd ed, Prentice Hall.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*, John Wiley & Sons.

Hyvärinen, A. and Pajunen, P. (1999). Nonlinear independent component analysis: Existence and uniqueness results, *Neural Networks* 12(3): 429–439.

Jolliffe, I. T. (2002). *Principal Component Analysis*, 2nd ed, Springer.

Jutten, C. and Karhunen, J. (2003). Advances in nonlinear blind source separation, *Proceedings of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation* (ICA2003) pp. 245–256.

Karhunen, J., Malaroiu, S. and Ilmoniemi, M. (2000). Local linear independent component analysis based on clustering, Int. *Journal of Neural Systems* 10: 439–451.

Knoll, G. F. (1989). *Radiation Detection and Measurement*, 2nd ed, John Wiley & Sons.

Mackay, D. J. C. (2002). *Information Theory, Inference and Learning Algorithms*, Cambridge University Press.

McClave, J. T., Sincich, T. and Mendenhall, W. (2008). *Statistics*, 11th ed, Prentice Hall.

Michal, A. D. (2008). *Matrix and Tensor Calculus With Applications to Mechanics, Elasticity and Aeronautics*, 1st ed, Dover.

Moura, N. N., Simas Filho, E. F. and Seixas, J. M. (2009). *Advances in Sonar Signal Processing*, In-Tech, Vienna, Austria, chapter Independent Component Analysis for Passive Sonar Signal Processing, pp. 91–110.

Murillo-Fuentes, J., Boloix-Tortosa, R., Hornillo-Mellado, S. and Zarzoso, V. (2004). Independent component analysis based on marginal entropy approximations, *Proceedings of the World Automation Congress* (16): 433–438.

Natora, M., Franke, F., Munk, M. and Obermayer, K. (2009). Bss of sparse overcomplete mixtures and application to neural recordings, in T. Adali, C. Jutten, J. Romano and A. Barros (eds), Independent Component Analysis And Signal Separation, 8th

International Conference, *Lecture Notes In Computer Science*, Springer, Paraty, Brazil, pp. 459–467.

Papoulis, A. and Pillai, S. U. (2002). *Probability, Random Variables, and Stochastic Processes*, 4th ed, McGraw-Hill.

Perkins, D. H. (2000). *Introduction to High Energy Physics*, 4th ed, Cambridge University Press.

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

Riu, I., Abolins, M., Adragna, et al. (2008). Integration of the trigger and data acquisition systems in ATLAS, *IEEE Transactions on Nuclear Science* 55(1): 106–112.

Rojas, F., Puntonet, C. G. and Rojas, I. (2003). Independent component analysis evolution based method for nonlinear speech processing, *Artificial Neural Nets Problem Solving Methods*, PT II 2687: 679–686.

Shannon, C. E. (1948). A mathematical theory of communication, *The Bell System Technical Journal* pp. 379–423.

Simas Filho, E. F., Seixas, J. M. and Caloba, L. P. (2009a). High-energy particle online discriminators based on nonlinear independent components, in T. Adali, C. Jutten, J. Romano and A. Barros (eds), Independent Component Analysis And Signal Separation, 8th International Conference, *Lecture Notes In Computer Science*, Springer, Paraty, Brazil, pp. 718–725.

Simas Filho, E. F., Seixas, J. M. and Caloba, L. P. (2009b). Optimized calorimeter signal compaction for an independent component based ATLAS electron/jet second-level trigger, *Proceedings of Science* - PoS ACAT08 102.

Strang, G. (2009). *Introduction to Linear Algebra*, 4th ed, Wellesley Cambridge Press.

Syskind, M., Wang, D. L., Larsen, J. and Kjem, U. (2006). Separating underdetermined convolutive speech mixtures, in J. Rosca, D. Erdogmus, J. C. Principe and S. Haykin (eds), Independent Component Analysis And Signal Separation, 8th International Conference, *Lecture Notes In Computer Science*, Springer, Charleston, USA, pp. 674–681.

Taleb, A. and Jutten, C. (1999). Source separation in post-nonlinear mixtures, *IEEE Transactions on Signal Processing* (47): 2807–2820.

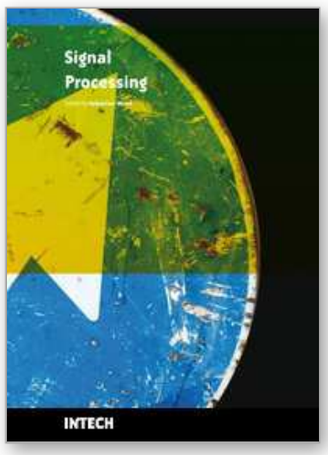
The ATLAS Collaboration (2008). The ATLAS experiment at the CERN large hadron collider, *Journal of Instrumentation* (2008 JINST 3 S08003).

Torres, R. C., Seixas, J. M., dos Anjos, A. and Cunha, D. V. (2008). Online electron/jet neural high-level trigger over independent calorimetry information, *Proceedings of Science* PoS(ACAT)039: 1–15.

Van Trees, H. L. (2003). *Detection, Estimation, and Modulation Theory*, Part I, Wiley.

Wei, C., Khor, L. C., Woo, W. L. and Dlay, S. S. (2006). Post-nonlinear underdetermined ICA by bayesian statistics, in J. Rosca, D. Erdogmus, J. C. Principe and S. Haykin (eds), Independent Component Analysis And Signal Separation, 8th International Conference, *Lecture Notes In Computer Science*, Springer, Charleston, USA, pp. 773–780.

Wigmans, R. (2000). *Calorimetry: Energy Measurement In Particle Physics*, Oxford.



Signal Processing

Edited by Sebastian Miron

ISBN 978-953-7619-91-6
Hard cover, 528 pages
Publisher InTech
Published online 01, March, 2010
Published in print edition March, 2010

This book intends to provide highlights of the current research in signal processing area and to offer a snapshot of the recent advances in this field. This work is mainly destined to researchers in the signal processing related areas but it is also accessible to anyone with a scientific background desiring to have an up-to-date overview of this domain. The twenty-five chapters present methodological advances and recent applications of signal processing algorithms in various domains as telecommunications, array processing, biology, cryptography, image and speech processing. The methodologies illustrated in this book, such as sparse signal recovery, are hot topics in the signal processing community at this moment. The editor would like to thank all the authors for their excellent contributions in different areas of signal processing and hopes that this book will be of valuable help to the readers.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen