Particle Identification with Neural Networks using a Rotational Invariant Moment Representation

Ralph Sinkus*
Raymond and Beverly Sackler Faculty of Exact Sciences
School of Physics and Astronomy,Tel–Aviv University, Israel

Thomas Voss†
Deutsches Elektronen Synchrotron, Hamburg, Germany

Abstract

A feed-forward neural network is used to identify electromagnetic particles based upon their showering properties within a segmented calorimeter. A preprocessing procedure is applied to the spatial energy distribution of the particle shower in order to account for the varying geometry of the calorimeter. The novel feature is the expansion of the energy distribution in terms of moments of the so-called Zernike functions which are invariant under rotation. The distributions of moments exhibit very different scales, thus the multidimensional input distribution for the neural network is transformed via a principle component analysis and rescaled by its respective variances to ensure input values of the order of one. This increases the sensitivity of the network and thus results in better performance in identifying and separating electromagnetic from hadronic particles, especially at low energies.

*Present address: Philips Research Division Technical Systems Hamburg, Roentgenstrasse 24-26, D 22335 Hamburg, Germany. email: R.Sinkus@pfh.research.philips.com. This research was partly supported by the Israel National Science Foundation and the Israeli Ministry of Science.

†Present address: NORCOM Infosystems, Glockengiesserwall 26, D 20095 Hamburg, Germany

1 Introduction

In high energy physics the identification of particles in the final state of inelastic scattering events is a crucial point in data analysis and reduction. The multi-purpose detectors provide a large range of information, such as momentum measurements of charged particles, energy loss due to ionization and calorimetric measurements of particle energies. In this analysis we consider deep inelastic scattering (DIS) ep events taken with the ZEUS detector [1] at the HERA collider. Here, 27 GeV electrons collide with 820 GeV protons producing a final state with a large particle multiplicity. The kinematic regime of large inelasticities is characterized by a low energetic debris ($\approx$ 5 GeV) of the fragmented proton (the so called current jet) often overlapping with the scattered electron of comparable momentum. It is an unfavorable situation since the electron is not isolated and thus its identification becomes difficult. Neural networks, with their inherent ability to identify complex patterns can provide substantial improvement in identifying non-isolated electrons. This has been shown in our previous publication (see [2]), where the showering properties of the particles in the segmented uranium calorimeter [3] of the ZEUS detector were used to efficiently differentiate between electromagnetic and hadronic particles. The spatial energy distribution of the showers was described by 54 energy readings probing a fixed volume in space. A feed-forward neural network [4] was used to identify, based upon these multi-dimensional information, between the two particle types. It was shown, that when compared to a classical approach, the neural network yielded a superior performance.

The approach used originally had a few weak points: i) not all information provided by the calorimeter was used due to the limitation imposed by a fixed number of energy readings, ii) the number of input variables was considerable and was affecting the performance of the neural network and iii) the efficiency of identifying an electromagnetic object dropped drastically at lower momenta ($\approx$ 5 GeV).

The new approach described in this paper focuses on defining new input variables to the net and on improving the convergence of the error minimization. The energy distribution, corrected for the non-projectivity of the uranium calorimeter, is expanded in terms of moments of two orthonormal function sets. In the plane perpendicular to the initial direction of the particle the so-called Zernike [5] functions are used. In the longitudinal direction the common Legendre functions are used. This way a fixed number of inputs describing the energy distribution of individual showers can be easily defined. The distributions of moments exhibit very different scales. To increase the sensitivity of the neural network towards relatively small moments we transform the distributions by means of a principle component analysis [6] and re-scale each new variable by its corresponding variance.

The new procedure results in an improved identification of scattered electrons in the calorimeter and in particular for the lowest energy region.

2 Preprocessing of Calorimeter Cells

This section describes the preprocessing procedure applied to all energy deposits registered in the ZEUS uranium scintillator calorimeter (CAL) originating from an ep scattering process. An example of a neutral current deep inelastic event, as seen in the ZEUS detector is shown in figure 1. A detailed description of the ZEUS experiment can be found elsewhere [7]. The

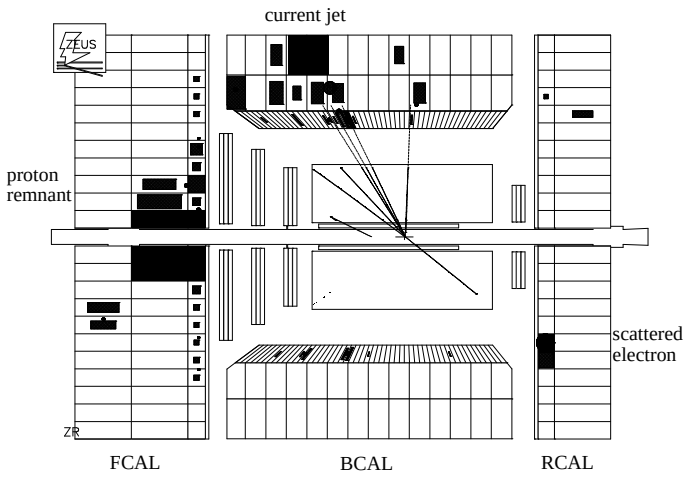

Figure 1: Event picture from the ZEUS detector showing a typical neutral current DIS event for $Q^2 = 310$ GeV^2. The black squares indicate various energy deposits in the CAL.

high resolution CAL is divided into three sections, the forward (FCAL), barrel (BCAL) and rear (RCAL) calorimeters. For normal incidence, the depth of the CAL is 7 interaction lengths in FCAL, 5 in BCAL and 4 in RCAL. The relative thicknesses of uranium and scintillator in the layer structure were chosen to give equal calorimeter response to electrons and hadrons. Under test beam conditions, the energy resolution for electrons was measured [3, 8] to be $\sigma(E)/E = 0.18/\sqrt{E}$ and for hadrons $\sigma(E)/E = 0.35/\sqrt{E}$, where E is in GeV.

Scintillator tiles form towers in depth that are read out on two sides through wavelength shifter bars, light guides and photomultipliers (PMTs). The towers are longitudinally segmented into electromagnetic (EMC) and hadronic (HAC) cells. The towers in FCAL and BCAL each have two HAC cells whereas those in RCAL have one. The depth of the EMC cells is 25 radiation lengths corresponding to one interaction length. Characteristic transverse sizes are 5 cm $\times$ 20 cm for the EMC cells of FCAL and BCAL and 10 cm $\times$ 20 cm for those in the RCAL. The HAC cells are typically 20 cm $\times$ 20 cm in the transverse dimension. The towers are read out by a total of 11836 PMTs. The construction minimizes the possibility for particles from the interaction point (IP) to propagate down the boundaries between modules. Holes of 20 cm $\times$ 20 cm in the center of FCAL and RCAL are required to accommodate the HERA beam pipe. The resulting solid angle coverage is 99.7% of 4π.

Initially the various cells of the CAL are grouped into so-called islands by a clustering algorithm [2]. This leads to clusters of a certain spatial size containing a total energy E_{tot}, which is distributed among the cells associated to the island.

The geometric center of each cell is corrected in one dimension based on the energy imbalance information provided by the two PMTs read-out on either side of each cell. This improves the reconstruction of the spatial shape of the shower in the CAL. In the following the center of a cell (with index i) refers to the point (x_i, y_i, z_i) in space to which the energy contained in this cell E_i has been assigned.

The second step in reconstructing the energy distribution is to transform the shower into a coordinate space corresponding to a projective geometry of the CAL. The transformation cannot be complete due to the finite granularity of the CAL. We assume that the energy of each cell is deposited in one point at the cell center (as defined above). It is also assumed

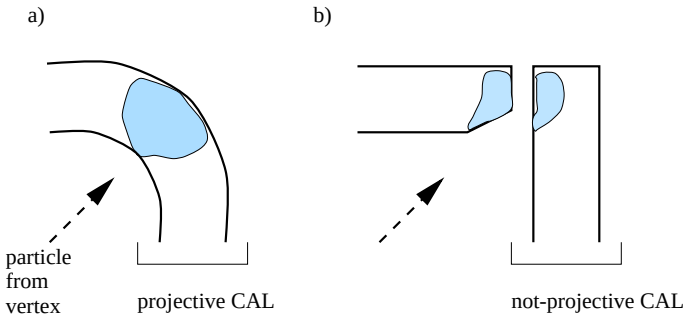

Figure 2: Schematic representation a) of a radially symmetric shower in a projective geometry of a fictitious calorimeter and b) of an extreme example of distortions within the ZEUS detector due to the non projective geometry (in a crack between two parts of the calorimeter).

that the shower develops only radially. The situation is sketched in Fig. 2. If the structure of the CAL were radial with respect to the vertex of the interaction, no distortion of the energy profile would occur (Fig. 2a). Since the uranium CAL of the ZEUS detector exhibits a varying granularity and projectivity we face, in the worst case, the situation depicted in Fig. 2b. The particular realization of the correction method is done as follows (see Fig. 3). Given any cluster somewhere in the CAL, there is no a priori point of reference other than the vertex for this cluster (the vertex is provided by other detectors like the central tracking device). For instance the evaluation of the center of gravity for a given cluster based upon linear energy weighting and centers of cells is wrong in the case of a cluster situated between different parts of the CAL (so-called *super-cracks*). For each cell we evaluate the length Λ (in cm) actually traversed within the uranium CAL by a ray (dotted lines) starting at the vertex (Z_{VTX}) and ending at the cell center. The cell center is now moved on this ray to a point, whose distance to a sphere of radius 1 m around the Z_{VTX} position is equal to Λ. In Fig. 3 the original cell centers are marked by the open symbols while the moved positions are indicated by correspondingly filled symbols. Now the center of gravity (COG) is evaluated based upon a linear energy weighting

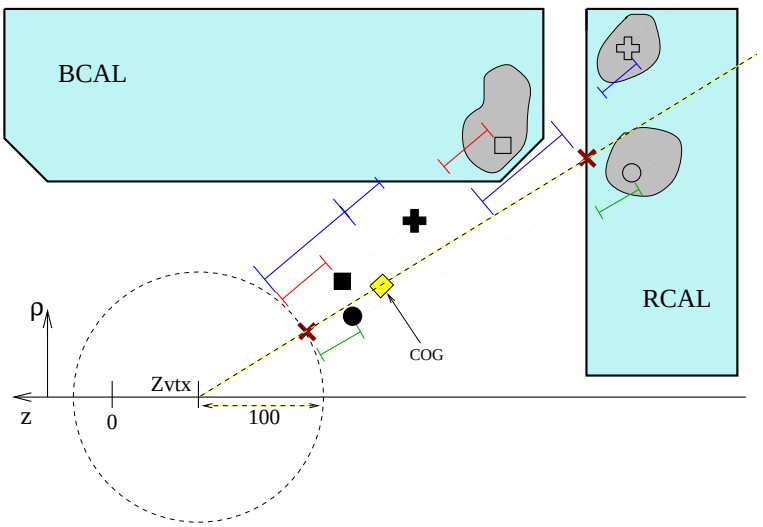

Figure 3: Schematic representation of the position reconstruction for a shower which developed in the boundary region between the barrel (BCAL) and the rear part (RCAL) of the calorimeter. See text for detailed explanation.

(depicted by a diamond). The ray starting at the vertex position and traversing the COG (dashed line) has two unique crossing points : • the surface of the sphere and • the CAL surface.

These two points are marked as small tilted crosses. All points of the cluster are now moved back radially by the distance between the two crossing points to properly recuperate the size of the shower. This is essential as the distance between the surface of the sphere around the vertex is different for clusters in the FCAL, BCAL and RCAL. The new $\acute{z}$-axis is along the ray between the vertex and the COG and has its origin at the first crossing with the CAL surface. The coordinates $\acute{x}$ and $\acute{y}$, or better the cylindrical variables $\acute{\rho}$ and $\acute{\varphi}$, are evaluated in a plane perpendicular to this ray. The orientation of this plane is arbitrary due to the moment expansion chosen here. This is explained in the next section.

3 Moment Representation of the Energy Distribution

The energy distribution of a shower is described in terms of a discrete set of points in phase-space. The number of points vary from shower to shower and therefore this distribution cannot yet be used as an input to a feed-forward neural network. We thus expanded it in terms of moments of an infinite series expansion. In the previous section the new coordinate system $\acute{z}, \acute{\rho}$ and $\acute{\varphi}$ has been introduced and is depicted in figure 4. While $\acute{z}, \acute{\rho}$ are independent of the

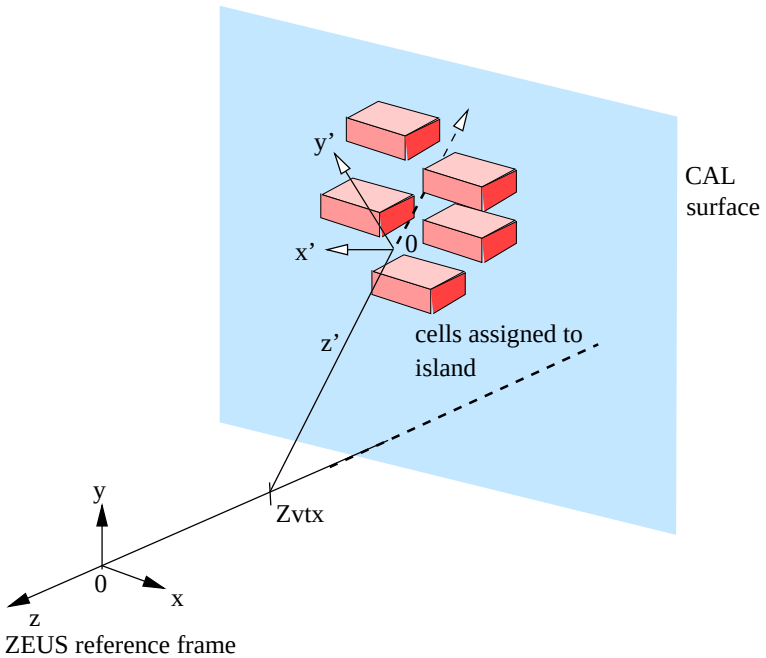

Figure 4: Schematic representation of the new coordinate system assigned to an island $(\acute{x}, \acute{y}, \acute{z})$. The ZEUS reference frame is denoted by (x, y, z).

orientation of the plane perpendicular to the $\acute{z}$-axis, this is not the case for $\acute{\varphi}$. As the mathematical representation in terms of moments should be independent of the actual orientation of the $\acute{x} - \acute{y}$-plane, it is convenient to choose a moment representation which only depends on $\Delta\varphi'_{ik}$, with i and k indices of different cells attached to the cluster. Such properties in two dimensions $(\acute{\rho}, \acute{\varphi})$ are accomplished by the absolute values of the complex *Zernike Moments* [5].

In $\acute{z}$-direction the energy distribution is expanded in moments of the Legendre functions. Both functional sets are orthogonal and normalized to one. The Zernike functions are only defined within the unity circle, while the Legendre functions are defined within the interval $[-1, 1]$. We thus re-scale the cell positions

$$\rho'' \;=\; \acute{\rho}/50 \quad , \quad \varphi'' \;=\; \acute{\varphi} \quad , \quad z'' \;=\; \acute{z}/200 \tag{1}$$

such that they typically lie within the allowed intervals. Cells with values of ρ'' or z'' which are outside the allowed ranges are not used in the evaluation of the moments, while all cells contribute to the total energy of an island (E_{tot}). A moment $\mathcal{A}$ is now determined by 3 integer values n, m and l which correspond to the different orders of the base functions

$$\mathcal{A}_{nml} \;=\; \sqrt{\frac{n+1}{\pi}} \cdot \sqrt{\frac{2l+1}{2}} \;\left|\; \sum_{i=1}^{Ncell} \frac{E_i}{E_{tot}} \cdot V^*_{nm}(\rho''_i, \varphi''_i) \cdot \mathcal{L}_l(z''_i) \;\right| \tag{2}$$

$$V^*_{nm}(\rho''_i, \varphi''_i) \;=\; R_{nm}(\rho''_i) \cdot \exp(-im\varphi''_i) \tag{3}$$

$$R_{nm}(\rho''_i) \;=\; \sum_{s=0}^{(n-m)/2} \frac{(-1)^s \;(n-s)!\; \rho_i^{''n-2s}}{s!\;\left(\frac{n+m}{2}-s\right)!\;\left(\frac{n-m}{2}-s\right)!} \tag{4}$$

$$\text{with } n, m \;\geq 0 \text{ integers }, \; n-m \text{ even and } m \leq n \tag{5}$$

$$\mathcal{L}_l(z''_i) \;=\; \text{standard Legendre function, } l \geq 0 \text{ integer} \tag{6}$$

with E_i the total energy of cell i and where the sum i runs over all cells attached to the cluster ($Ncell$). The value of $\mathcal{A}_{nml}$ represents the moment of order n, m, l. The (0,0,0) moment should in general be equal to

$$\mathcal{A}_{000} = \sqrt{\frac{1}{2\pi}}\,, \tag{7}$$

independent of ρ'', φ'' and z''. In case of large clusters, where some cells are outside the allowed range and not included into the sum of Eq. 2, the (0,0,0) moment is reduced by the fraction of the neglected energy. Therefore it serves to describe the energy leakage outside of the probed region.

4 Transformation and Rescaling of the Moment Distribution

The energy distribution of an island is represented by an infinite series of moments. Typically only terms up to a certain order l_{max}, n_{max} are important. The higher order terms are sensitive to the tails of the energy distributions which may not be properly simulated in the Monte Carlo. Thus it is advisable to truncate the series and only use moments up to a certain order as input for the neural network. We choose to work with $n_{max} = 2$ and $l_{max} = 3$. The choice of $n_{max} = 2$ allows to probe the shower width while the departure from the cylindrical symmetry of the shower, a sign for overlapping showers, is probed by the ϕ dependence of the Zernike moments. The longitudinal shape of the showers is probed up to the third moment of the energy

distribution as this is a strong discriminator between electromagnetic and hadronic showers. In Fig. 5 the distributions of the first four moments are shown, separately for electromagnetic (C_e) and hadronic (C_h) clusters. The widths of the various moment distributions differ significantly, making it very hard for a feed-forward neural network to use all information with equal weight. Thus, there is a need to properly re-scale the input distributions by their respective widths. The moment distribution for electromagnetic clusters is transformed via a *principle component analysis* (PCA) [6] to its main axes. The transformation consists of a translation to the center of gravity of the distribution and a rotation of the initial space such that the product of the new variances is minimal. This allows a correct rescaling and removes possible linear dependencies.

The input distribution consists of all moments up to $n_{max} = 2$ and $l_{max} = 3$ (i.e. 16 moments) and the total energy of the cluster scaled by 20 ($E_{tot}/20$). The new root mean squares (RMS), σ_i, obtained by applying the PCA to a specific clusters belonging to class C_e and ordered by their significance are presented in figure. 6. The linear transformation and the rotation of the PCA obtained with the C_e clusters is then applied to the C_h clusters and the appropriate values of σ_i are also shown in figure. 6. The σ_i corresponding to the C_h class are typically larger when compared to the ones of class C_e, as naturally expected from the different showering properties of hadronic and electromagnetic particles. The overall trend is similar to the σ_i of class C_e, however the ordering by significance does not strictly apply. The inverse of the maximum value of the respective RMS $\max(\sigma_i(C_e), \sigma_i(C_h))$, $i \in [1, 17]$ become the rescaling factors for the new variables. The appropriate values are listed in Table 1.

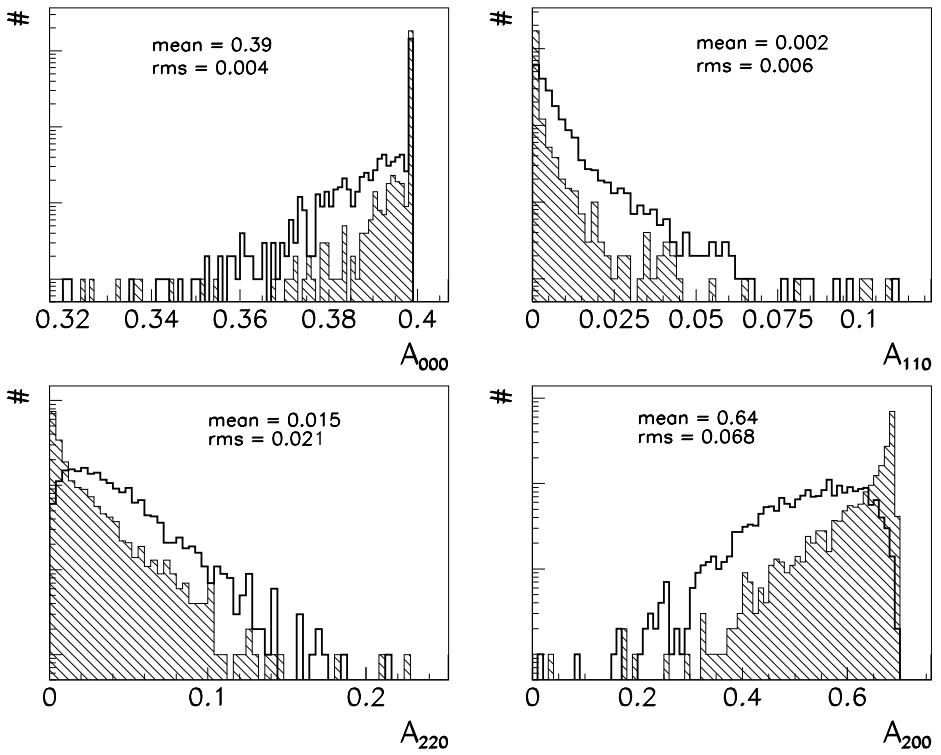

Figure 5: Normalized distributions of the first four moments $\mathcal{A}_{nml}$ for electromagnetic clusters (hatched histogram) and hadronic clusters (thick line). The vertical scale is logarithmic. The quoted means and RMS refer to the distributions of the electromagnetic clusters.

i	factor	i	factor	i	factor
1	0.213E+01	7	0.3527E+02	13	0.21607E+03
2	0.674E+01	8	0.1814E+02	14	0.36567E+03
3	0.528E+01	9	0.4001E+02	15	0.13797E+03
4	0.2747E+02	10	0.9548E+02	16	0.20868E+03
5	0.4350E+02	11	0.13252E+03	17	0.37518E+03
6	0.4837E+02	12	0.8379E+02		

Table 1: Rescaling factors for the transformed input distributions

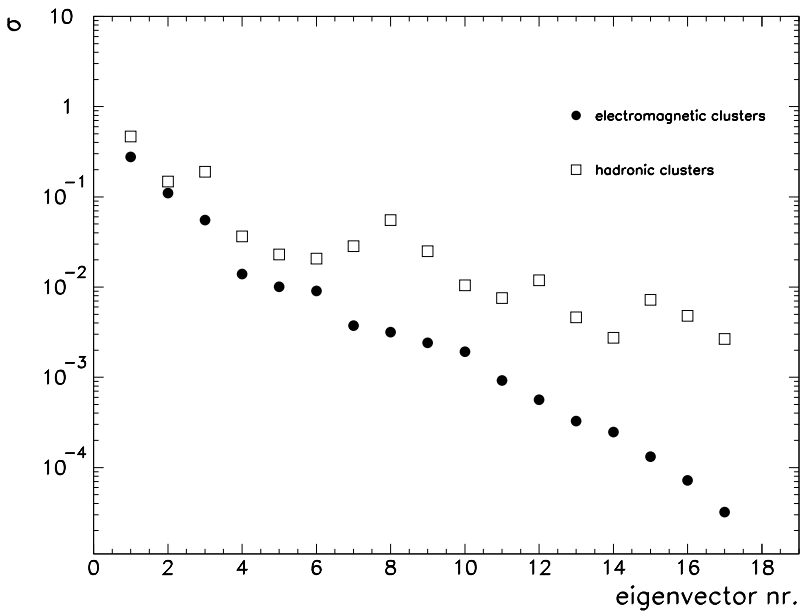

Figure 6: Root mean squares σ_i of the transformed distribution of electromagnetic patterns ordered by significance (full dots). The corresponding σ_i of patterns belonging to the hadronic class are shown as open squares.

5 The Neural Network

We use a feed-forward neural network with error back-propagation (FFBP) to distinguish between electromagnetic and hadronic particles based upon the input variables. The network used in this particular application is described in [4]. In contrast to standard FFBP neural networks, this model introduces an additional parameter denoted as temperature t. Its purpose is to make the convergence of the network towards the global minimum of the error function (E) more likely. The minimization procedure is started at large values of t and the approach to the global minimum of E is achieved by a lowering of t.

It has been shown for a one-dimensional example [4] that the temperature domain corresponding to a quadratic dependence of the error function E on the free parameters of the network is

reached for $t \geq 5$. This result only holds if the values of the input distribution are of the order of one ($\mathcal{O}(1)$). Suppose this requirement is not fulfilled and the values of the input distribution are instead rescaled by one order of magnitude, i.e. $\mathcal{O}(0.1)$. The structure of the error function E becomes scale-invariant only if the temperature t is rescaled by one order of magnitude as well. Otherwise, starting the minimization procedure at $t = 5$ (instead of $t = 0.5$) results in an extremely slow convergence, because E becomes flat for $t \to \infty$. Thus it is very important for an efficient minimization of the error function to ensure that the input distribution is of the order of one. This requirement is achieved by the rescaling procedure described in the previous section.

6 Training of the Neural Network

The energy distribution within each cluster is represented by 16 moments ($l_{max} = 3, n_{max} = 2$). As an additional input parameter we choose the total energy of the cluster scaled by 20 ($E_{tot}/20$). This 17-dimensional vector is transformed by the already established matrices of the PCA and then rescaled by the factors of Table 1. The network is trained with 4000 patterns generated by a Monte Carlo simulation, 2000 belonging to class C_e and 2000 to belonging to class C_h. The number of hidden neurons has been optimized by trial and error. It was found that 5 hidden neurons are sufficient and further increase shows no significant improvement.

After 2000 iterations the minimization procedure is stopped as over-training starts to onset as established by inspecting a statistical independent but similar data sample (cross-validation [9]). The training is performed 10 times with different initial values for the free parameters of the network. In Fig. 7 the means of the temperature and the error function E as a function of the iteration step (epoch) are presented. Also shown are the corresponding values for t and E for the case when the input distributions are NOT transformed and rescaled by the PCA as described in the previous section. Although both methods result in similar values for t after 2000 epochs ($t \approx 0.3 - 0.4$), the values for E differ by about 3%. Thus without the rescaling procedure the convergence to the global minimum of the error function is hindered.

By cutting on the final one-dimensional output value of the neural network we can evaluate the efficiency and purity to identify patterns belonging to class C_e, i.e

$$\text{efficiency} \equiv \frac{N(C_e, Y > Cut)}{N(C_e)}$$

$$\text{purity} \equiv \frac{N(C_e, Y > Cut)}{N(C_e, C_h, Y > Cut)} \quad ,$$

with $N(C_e, Y > Cut)$ representing the number of patterns belonging to class C_e with an output value of the network greater than the required value of the cut. Varying the cut between ± 1 yields a curve in the efficiency-purity plane. These curves are shown in Fig. 8 for both scenarios, i.e. using the bare and the transformed distributions. The curve for the latter case is pushed to larger values of efficiency and purity. Thus the general ability to separate the classes is superior.

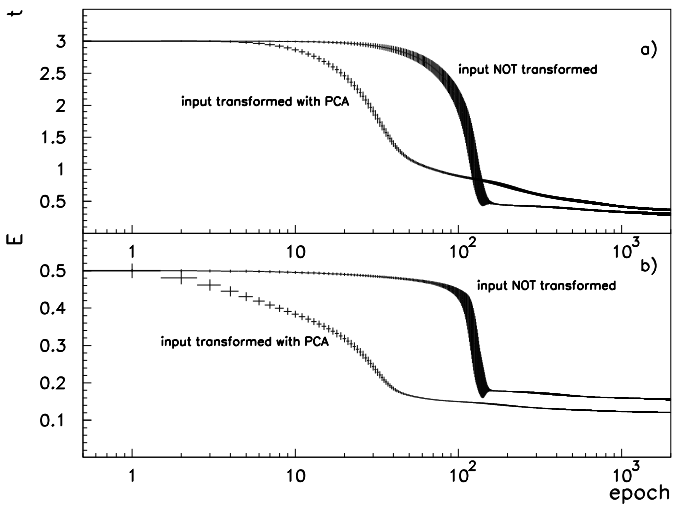

Figure 7: Mean values for the temperature a) and the value of the error function E b) as a function of the iteration (epoch). The results for two different training pattern samples are depicted. In one case the bare 17-dimensional input is used as input for the neural network. In the other case the bare input distributions are transformed and rescaled by means of a PCA.

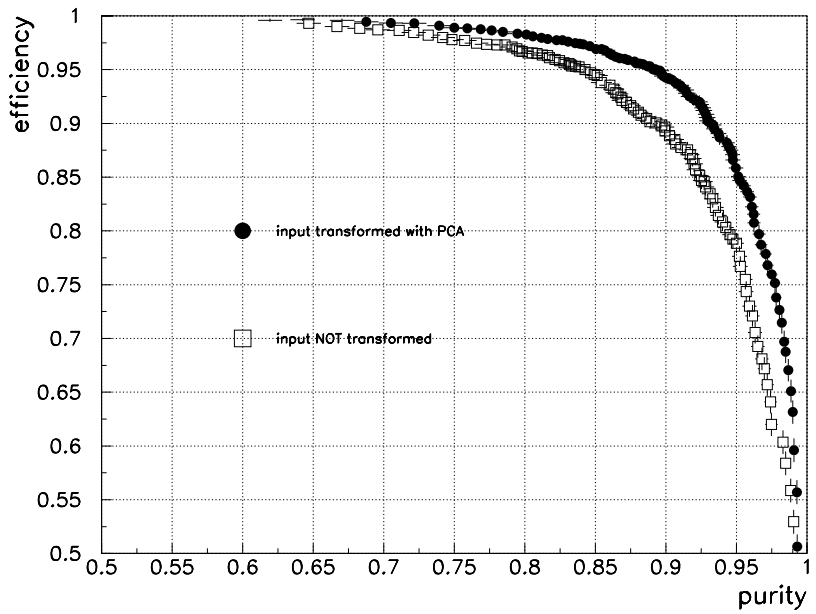

Figure 8: Efficiency versus purity for electron identification as obtained for the neural network trained on the bare moment distributions (open squares) and the neural network trained on the transformed and rescaled distributions.

7 Results

The performance of the trained neural network was check with a Monte Carlo sample of DIS events generated with a virtuality of the scattered lepton greater than $Q^2 > 2.2$ GeV^2 and large inelasticity ($y > 0.4$), a configuration heavily biased towards non-isolated scattered electrons. The efficiency and purity to identify the scattered lepton within an event is depicted in Fig. 9 as a function of the reconstructed energy of the selected cluster. For the selection of the scattered electron within an event we require the cluster to have a probability calculated by the network (Y) greater than $\frac{1}{2}(1+Y) > 0.9$. Among the preselected clusters we select the one with highest probability to be the scattered electron.

We observe very high efficiencies of about $\approx 50\%$ at energies of ~ 5 GeV with purities $\approx 95\%$. This is a big improvement when compared to the previous version of the electron finder which had about $\sim 10\%$ efficiency at 5 GeV with a similar purity. The sensitivity of the new electron finder to the non-uniformities of the calorimeter has been checked with a high Q^2 MC sample, in which the scattering angles of the electron span a larger range. As shown in Fig. 10 the efficiency is very weakly depended on the scattering angle while the purity remains everywhere very high.

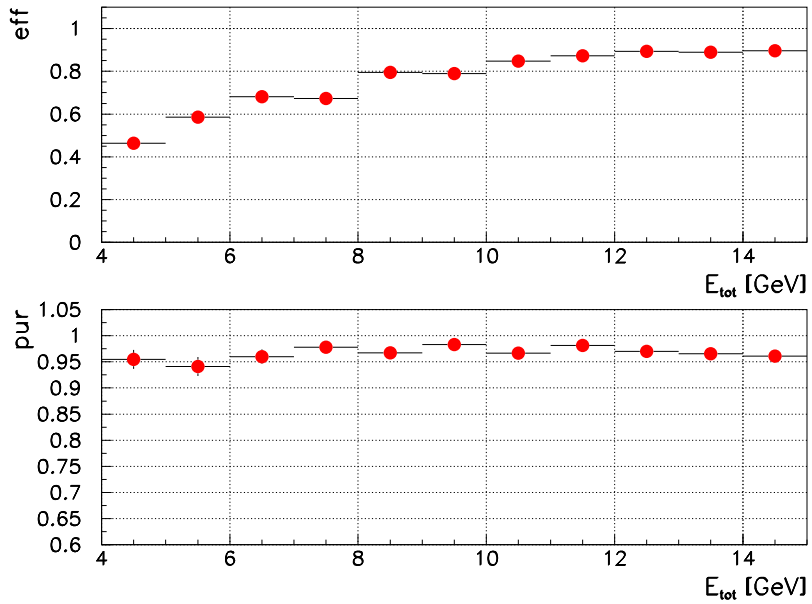

Figure 9: Efficiency and purity to identify the scattered lepton as a function of reconstructed energy for a MC DIS $Q^2 > 2.2$ GeV^2 high y event sample.

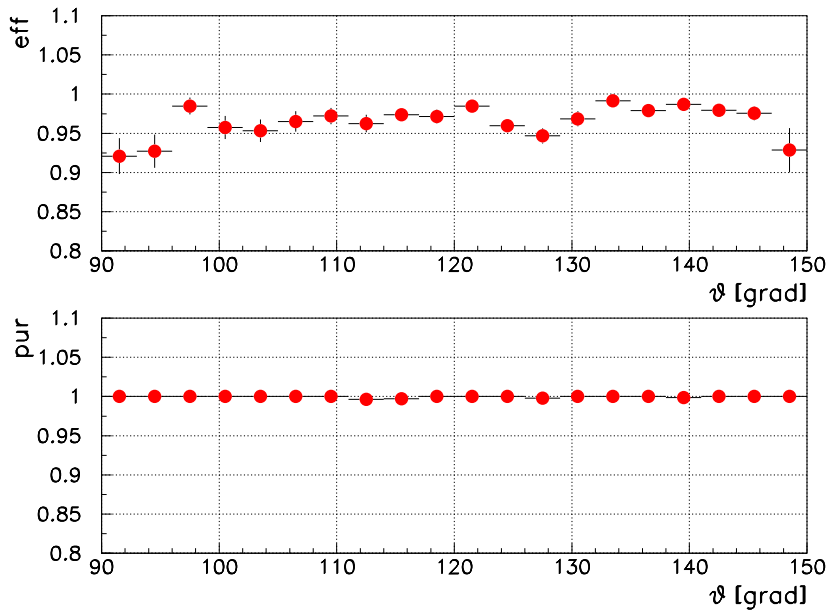

Figure 10: Efficiency and purity to identify the scattered lepton as a function of reconstructed scattering angle for a MC DIS $Q^2 > 300$ GeV^2 event sample. No significant inefficiencies can be observed at $\theta \approx 130°$, which is where the super-crack between BCAL-RCAL is located.

8 Conclusions

We present an extended method (compared to [2]) to identify electromagnetic particles based upon their showering properties in the segmented calorimeter of the ZEUS detector using a feed-forward neural network. The new features are the preprocessing of the energy distribution of individual calorimeter showers with a rescaling of the input variables after a PCA transformation before training the neural network. The preprocessing consists of 3 part :

1. to reduce the effect of non-projectivity a procedure is applied which alters the position of each shower cell in space according to its actual depth in the calorimeter relative to the vertex,
2. the obtained discrete representation of the shower profile is expanded in terms of rotationally invariant moments of two orthogonal function sets (the Zernike and the Legendre functions),
3. the distributions of the fixed number of moments chosen as input variables are transformed to their main axis and rescaled by their according variances.

This results in the following improvements :

- the full granularity of the calorimeter is used,
- the non-projective regions like the ones between the calorimeter parts are properly handled

- and the performance in identifying electromagnetic objects is increased considerable at low energies ($\approx$ 5 GeV).

A network trained on input distributions which had not been rescaled shows a deteriorated performance. Thus the rescaling procedure helps in converging towards the global minimum of the error function of the neural network.

Acknowledgments

This work has been pursued in the framework of the ZEUS collaboration and we would like to acknowledge the motivating interest of our colleagues. We would like to thank Prof. Halina Abramowicz for her constant support and the fruitful discussions. We are indebted to Prof. David Horn for his many helpful suggestions.

References

[1] ZEUS Collaboration, DESY PRC 93-05 (1993).

[2] H. Abramowicz, A. Caldwell, and R. Sinkus, NIM A **365**, 508 (1995).

[3] M. Derrick. et al., Nucl. Inst. Meth. A **309**, 77 (1991).

[4] R. Sinkus, NIM A **361**, 290 (1995).

[5] F. Zernike, Physica **1**, 689 (1934).

[6] H. Wind, *Formulae and Methods in Experimental Data Evaluation* (R.K. Bock, 1984), Vol. 3, pp. K1–K16.

[7] ZEUS Collaboration; M. Derrick et al., Phys. Lett. B **293**, 465 (1992).

[8] A. Andresen. et al., Nucl. Inst. Meth. A **309**, 101 (1991).

[9] M. Stone, Journal of Royal Statistical Society, Series B **36**, 111 (1974).