

# Quark/gluon jet separation in the photoproduction region with a neural network algorithm

Giuseppe Barbagli  
Università di Firenze and C.N.R.  
Giulio D'Agostini and Daniela Monaldi  
Università "La Sapienza" and I.N.F.N., Roma

**Abstract:** A neural network approach has been used to separate quark and gluon jets at HERA in the photoproduction region. The network was trained to learn the structural features of the jets. The jet recognition was used to enhance the contents of a sample of Photon-Gluon Fusion events by reducing the background of QCD Compton events.

## 1 Introduction

There is a strong interest in separating Photon-Gluon Fusion (PGF) and QCD Compton scattering (QCDC) in direct photoproduction at HERA or, at least, in obtaining enriched samples of events for either reaction. Two jet photoproduction has been proposed as a tool to determine the gluon density of the proton from PGF [1,2], while QCD Compton scattering could yield information about the sum of the quark densities. Moreover, both processes could provide interesting information about hadronization schemes, especially if one is able to distinguish, within the two classes, the quark from the antiquark or the quark from the gluon.

In the region of interest, the distributions of the kinematic variables of the final state partons do not present meaningful separation and a jet identification based on the differential hard cross-sections is not efficient, hence, one has to fully exploit the different jet topologies and match several features in order to distinguish them. Neural network methods have been extensively used to solve the problems of pattern recognition, especially when the decision criteria are complex and subtle. They are inspired by the actual modeling of biological cognition, i.e., they are made up of a set of processing elements, called nodes or *neurons*, linked to each other by *synapses* with well defined strengths, or *weights*. Neurons perform simple operations: they calculate the weighted sum of inputs and transmit it to the other nodes through a nonlinear threshold function. Neural networks exhibit high intrinsic parallelism (suitable for hardware implementation) and fault-tolerance. They perform learning, recognition and classification functions, and allow one to solve problems, such as classification patterns, recognition

where  $T$  is a parameter, called *temperature*, which controls the steepness of the function at the origin. All the weights and the thresholds are optimized with the following iterative learning procedure. The weights are initially set to small random numbers and updated regularly as the iterations proceed. The network is exposed to a series of patterns; for each of them the value of every output node,  $y_i$ , is evaluated and compared to the class flag  $t_i$  (*target*), then a fraction of the deviation  $\delta_i$  is back-propagated as a feedback. The targets have values 0 or 1. For the nodes of the last layer the deviation is:

$$\delta_i = (y_i - t_i) f'(s_i^a), \quad (6)$$

where  $f'$  is the derivative of the function  $f$  with respect to  $s$ . For the nodes of the intermediate layer it is:

$$\delta_j = \sum_{i=1}^N w_{ij}^a \delta_i f'(s_j^b). \quad (7)$$

The weights at each node are then modified adding a feedback term, given by the *learning rate*,  $\lambda$ , times the deviation times the last input contribution and a balance term, to reduce oscillations, given by  $\alpha$  times the previous correction. Both of these two parameters have values between 0 and 1. The updating of the weights is governed by the equation:

$$\Delta w_{ij}^a = -\lambda \delta_i h_j + \alpha \Delta w_{ij}^{a \text{ previous}}, \quad (8)$$

for the weights connecting the nodes of last layer (output nodes) to the nodes of the last intermediate layer, and:

$$\Delta w_{jk}^b = -\lambda \delta_j x_k + \alpha \Delta w_{jk}^{b \text{ previous}}, \quad (9)$$

for the weights connecting the intermediate layer with the inputs.

A multi-layer perceptron is able to draw, in the multidimensional input variable space, decision regions limited by arbitrarily complex boundaries. The only limitations to the capability of separating classes come from the number of nodes and from the intrinsic overlap of the classes in the multidimensional space of the input variables (Bayes limit). We have written a program to implement the previously described algorithm. We have carefully investigated the behaviour of the network and its relation to the parameters and to the number of hidden layers and nodes, in order to choose a suitable tool to apply to the analysis of the jet features.

## 3 Simulation and selection of events

We have applied the network algorithm to the identification of gluon and quark jets, relying only on their topological differences. We simulated photoproduction at HERA through QCDC and PGF with PYTHIA 5.5 [7], including initial and final state parton showers, with a  $p_t$  cut of 10 GeV/c chosen to obtain only hard jets in the final state. The cross-sections of the two processes are 1.4 nb and 3.4 nb, respectively. Jets were reconstructed using the LUCCEL algorithm (from JETSET 7.3 [8]) with cell size  $[\delta\eta, \delta\phi] = [0.2, 0.3]$ , jet cone  $\sqrt{\Delta\eta^2 + \Delta\phi^2} \leq 1$  and jet transverse energy  $E_T^{\text{jet}} > 7 \text{ GeV}$  ( $\eta$  is the pseudorapidity and  $\phi$  the azimuthal angle). Finally, only events are selected that have two well separated and coplanar jets ( $\pi - 0.25 < |\phi_{\text{jet1}} - \phi_{\text{jet2}}| < \pi + 0.25$ ) at least  $\sim 8^\circ$  from the beam pipe ( $|\eta_{\text{jet}}| \leq 2.7$ ). With these cuts, approximately 40% of the generated events survived.

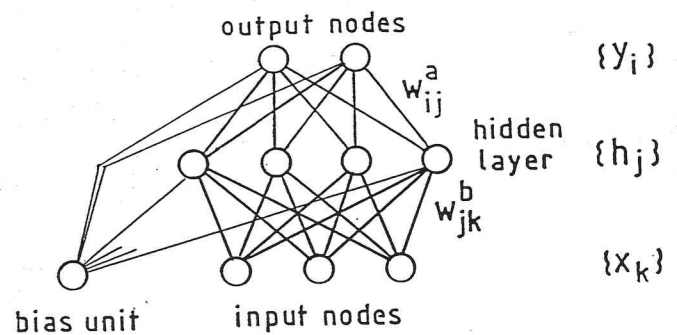


Figure 1: Structure of a two layer perceptron.

of images and sounds, optimization etc., which are simple for the human brain but are non-trivial for traditional computers.

We built a neural network and we taught it to recognize quark and gluon jets, by exposing it to a large sample of simulated events. After that, we evaluated the efficiency of the network as a classifier and we analyzed photoproduction events with two jets in the final state in order to tag the underlying process as QCDC or PGF.

## 2 Description of a 2 layer perceptron

Perceptron [3,4,5] is an old neural network algorithm, suitable for classification problems. It is organized (Fig. 1) according to layers: a set of inputs, one or more intermediate (*hidden*) layers of processing nodes and one or more output nodes. Each layer is connected to the following layer via links of given weights,  $w_{ij}$ . Let us have  $M$  input nodes, one hidden layer of  $M_h$  nodes, and  $N$  output nodes. The  $x_k$  are the inputs,  $h_j$  are the values of the nodes of the hidden layer and  $y_i$  are the values of the output nodes. The  $w_{ij}^a$  are the weights connecting the output nodes with the hidden layer and  $w_{jk}^b$  are the weights connecting the hidden layer with the input nodes. Then one has:

$$y_i = f(s_i^a), \quad (1)$$

where  $f$  is a nonlinear threshold function, and

$$s_i^a = \sum_{j=1}^{M_h} w_{ij}^a h_j - \theta_i^a = \sum_{j=0}^{M_h} w_{ij}^a h_j \quad (2)$$

$$h_j = f(s_j^b) \quad (3)$$

$$s_j^b = \sum_{k=1}^M w_{jk}^b x_k - \theta_j^b = \sum_{k=0}^M w_{jk}^b x_k. \quad (4)$$

The  $\theta_i$  are characteristic thresholds that are adjusted as the processing is carried on, being regarded as particular weights linking the nodes to external *bias* input sources at fixed values. A valid choice of  $f$ , having values in the range  $[0,1]$ , is:

$$f(s) = \frac{1}{2} [1 + \tanh(\frac{s}{T})], \quad (5)$$

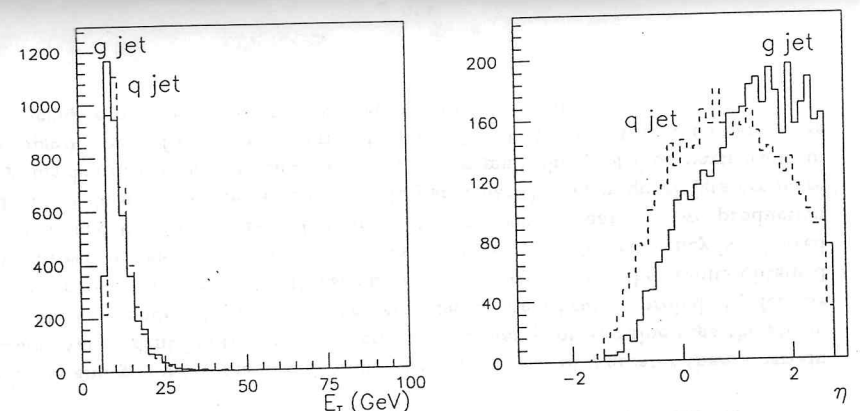


Figure 2: Transverse energy and the rapidity for jets produced in QCDC and PGF.

## 4 Choice of input variables and learning procedure

In the high  $p_t$  photoproduction region there is no obvious separation between quarks and gluons based on kinematic variables (see Fig.2): both kinds of jets have the same energy distribution and only the angular one differs slightly. However, the different fragmentation properties are reflected in different jet structures that often permit one to identify visually the gluon jet, in a two jet configuration, as the one with higher multiplicity, softer spectrum and broader spatial development. The recognition of the jet flavour based on these features requires a multidimensional analysis combining information coming from several physical quantities: this is a typical task required of a neural network.

The necessity of giving the most detailed information on the jet has to be balanced with the requirement of a reasonably sized network. As a compromise, we used the detailed information on some leading particles, together with some global variables which quantify the features of the jet. After some attempts, we chose Fodor's moments, defined as [9]:

$$F_{\alpha\beta} = \sum_{i=1}^N (\frac{p_{Ti}}{E_T})^\alpha \eta_i^\beta \quad (10)$$

where the sum is over all the particles in the jet and  $p_T$  and  $\eta$  are the transverse momentum and the pseudorapidity, both defined with respect to the jet axis. We found that Fodor's moments are more sensitive to different jet structure in the Hard Scattering Center of Mass frame (HSCM), than in the laboratory frame. Since the two processes are essentially at  $Q^2 \approx 0$ , the HSCM is related to the laboratory frame by a boost along the beam axis. This has been calculated using HSCM rapidity, estimated by the average pseudorapidity of the two jets.

To summarize, the patterns presented to the network for learning and testing consisted of 22 variables, all of them calculated in the HSCM:

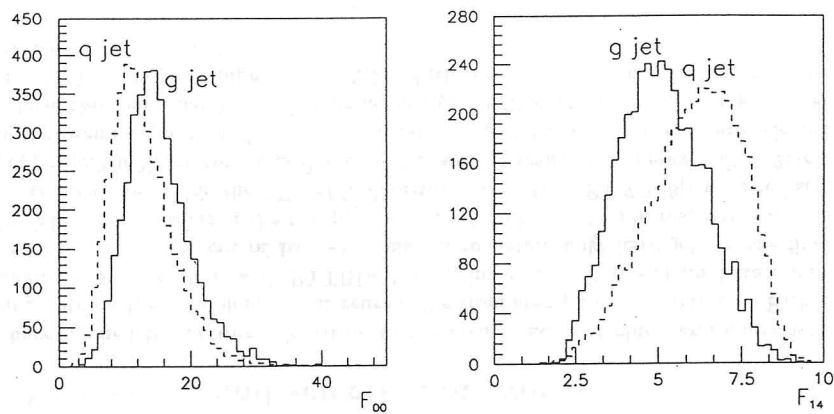


Figure 3: Some Fodor's moments in the Hard Scattering Center of Mass frame.

- the transverse energy of the jet  $E_T^j$ ,
- the pseudorapidity of the jet  $\eta^j$ ,
- the azimuth of the jet  $\phi^j$ ,
- $(E_T^c/E_T^j, \Delta\eta_c, \Delta\phi_c)$  for the cells in the  $\eta-\phi$  grid with the largest fraction  $E_T^c/E_T^j$ , where the differences are taken with respect to the jet axis,
- the Fodor's moments  $F_{00}$  (multiplicity),  $F_{10}$ ,  $F_{14}$  and  $F_{26}$ .

The network consisted of 22 inputs, one hidden layer and only one output node. After some tuning, we chose the following parameters:

- hidden layer size: 46 nodes,
- learning rate:  $\lambda = 2 \times 10^{-4}$  for the first 200 learning sequences, reduced to  $0.5 \times 10^{-4}$  for the following ones,
- balance coefficient:  $\alpha = 0.5$ ,
- temperature:  $T = 1$ .

The network was exposed to 154,000 quark jets (from QCDC and PGF) and 154,000 gluon jets (from QCDC) randomly mixed. The training was subdivided into sequences; each sequence consisted of 2000 trainings: 1000 with quark jets and 1000 with gluon jets. After each learning cycle, the capability of the network to separate classes was tested on an independent sample of 4000 quark jets and 4000 gluon jets. After a steep rise during the first 20,000 trainings, the recognition efficiency continued growing slowly toward a final stable value.

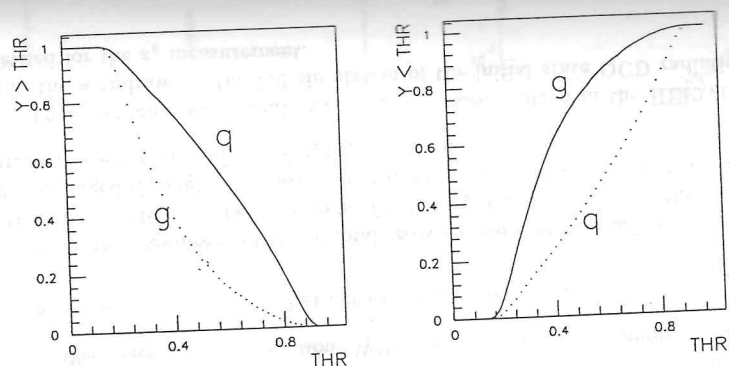


Figure 5: Identification efficiencies and misidentification probabilities as function of the threshold.

| NETWORK DECISION | GENERATED PROCESS |      |
|------------------|-------------------|------|
|                  | PGF               | QCDC |
| qq               | 37%               | 18%  |
| gq               | 51%               | 65%  |
| gg               | 12%               | 17%  |

Table 2: Event classification results.

## 6 Separation of Photon-Gluon Fusion Events and QCD Compton events

A classification of the two jet events on the basis of jet identification has been used in order to tag the underlying physical process and to enhance samples of PGF events. Tab. 2 shows the probabilities of finding a combination of two jets as they are recognized by the neural network. The probability to be classified as "quark-quark" is, for a PGF event, twice as high as for a QCDC one<sup>1</sup>. In Fig. 6, the distributions of the reconstructed momentum fraction carried by the incoming parton[2] before and after the neural network quark-quark selection. are presented.

An attempt has been made to recognize the underlying process from the global event shape instead of from the individual jet identification. This should at least give the same result and, perhaps, increase the efficiency due to long range correlations which are lost in the present approach. Unfortunately, the problem becomes technically more difficult because of the larger number of input variables and more study is required.

<sup>1</sup>Note that the probability of finding two quark jets in the same event is not the product of that of finding a quark in the single jet, but it is given by a conditioned probability.

| NETWORK DECISION | GENERATED PARTONS |                     |      |     |     |
|------------------|-------------------|---------------------|------|-----|-----|
|                  | GLUON             | QUARK               |      |     |     |
|                  |                   | $\langle q \rangle$ | QCDC | PGF |     |
| g                | 73%               | 32%                 | 26%  | 34% | 42% |
| q                | 27%               | 67%                 | 74%  | 66% | 58% |

Table 1: Jet classification results

## 5 Results on quark-gluon jet identification

With the final set of weights, the output variable distribution for the two classes of jets achieves a much larger separation than is obtainable with the most sensitive of the Fodor's moments, as can be seen in Fig. 4. The ideal separation would yield  $y = 1$  for

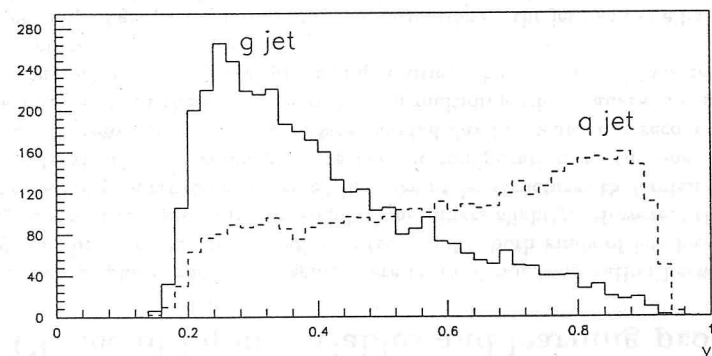


Figure 4: Distribution of the output node variable for quark and gluon jets.

quark jets and  $y = 0$  for gluon jets. Thus, the requirement of  $y$  greater than a certain threshold value identifies the quark, whereas  $y$  less than another value tags the gluon. Fig. 5 shows the identification efficiency and the misidentification probability for the two classes as a function of the threshold. In Tab. 1 we present the results obtained with a threshold value of 0.5 for both classes. One sees that 73% efficiency is achieved for gluons and 67% for quarks. It is interesting to remark that the two production mechanisms do not give the same results. In particular the heavy quark jets are more easily misidentified as gluons because of the higher multiplicity of final particles from heavy mesons and of the more spherical topology, due to the fact that the quark mass is not negligible with respect to the jet energy in the HSCM frame.

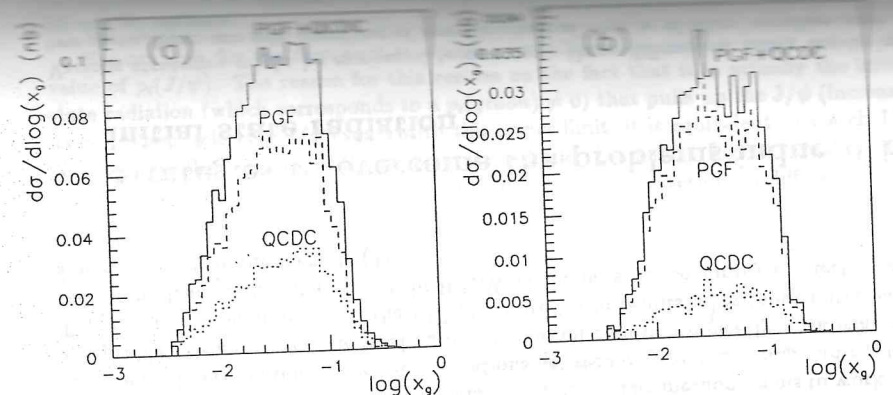


Figure 6: Reconstructed fractional momentum of the parton extracted from the proton, for PGF and QCDC events before (a) and after (b) the requirement of two quark jet identification.

## 7 Conclusions

A neural network approach has been used to separate quark and gluon jets in the photoproduction region. The efficiencies were 67% and 73% for quark and gluon jets, respectively. The requirement of two identified quark jets has allowed an enhancement of samples of Photon-Gluon Fusion events over the background QCD Compton scattering up to 5:1.

## References

- [1] G.Barbagli and G.D'Agostini, Proc. of the DESY Workshop *Physics at HERA*, Hamburg, 1987, R.Peccei ed., pag. 135.
- [2] G. D'Agostini and D. Monaldi, Z. Phys. C 48(1990)467.
- [3] R.P. Lippmann, IEEE ASP Magazine 4(1987)4.
- [4] T.Khanna, *Foundations of Neural Networks*, Addison-Wesley (1990).
- [5] *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 1 Foundations* (D. Rumelhart, J. McClelland and the PDP Research Group editors), MIT Press (1986).
- [6] L.Lönnblad, C.Peterson, T. Rönngvaldsson, LU-TP 90-8, May 1990.
- [7] H.-U.Bengtsson and T.Sjöstrand, Computer Phys. Comm. 46(1987)43
- [8] T.Sjöstrand and M.Bengtsson, Computer Phys. Comm. 43(1987)367
- [9] Z.Fodor, Phys. Rev. D 41(1990)1726