CERN-TH.6038/91

ITP-Rep. Budapest 483

**Combined Neural Network-QCD classifier
for quark and gluon jet separation**

I. Csabai

Department of Atomic Physics

Eötvös University, Budapest, H-1088 Puskin u. 5-7

F. Czakó

Department of Surface Physics

Eötvös University, Budapest, H-1088 Múzeum krt. 6-8

Z. Fodor *

CERN, Theory Division, CH-1211, Geneve 23

ABSTRACT

In this paper we show a possible quark/gluon jet discrimination method using a neural network, combined with QCD, in electron-positron annihilation. The network has been trained by quark and gluon jets of the same energy, therefore only the internal structure of the jets has been considered. The input data correspond to a total basis, in the sense that all the longitudinal and transverse momenta of the outgoing hadrons with respect to the jet axis (excluding soft particles) have been used in the analysis. The neural network alone has reached $\approx 71\%$ identification accuracy for jets of the same energy. On the other hand the QCD matrix element for definite jet energies also gives the probability of a jet being a gluon or quark one. By the combination of the neural network, and the QCD second order matrix element, we have reached 92% accuracy in identifying jets.

* On leave of absence from Institute for Theoretical Physics, Eötvös University, Budapest, H-1088 Puskin u. 5-7

CERN-TH.6038/91

ITP-Rep. Budapest No. 483

March 1991

1. INTRODUCTION

Quantum Chromo Dynamics (QCD) has been proposed as a theory for the strong interactions [1]. The fundamental particles in QCD are the quarks and gluons. In experimental situations (e.g. in electron-positron annihilation), due to the low Q^2 hadronization processes, hadrons are observed. These fragmentation processes have been described only by phenomenological models. Uncertainties in these models mean a major obstacle to the detailed testing of QCD. To understand several phenomena the separation of quark jets from gluon ones would be required. Among numerous examples we quote one here. There has been remarkable progress in three-gluon vertex studies [2]. Nevertheless the effect in those analyses (the difference between the Abelian model and QCD in terms of the variables used) is approximately 20%. This small difference is not sensitive enough to establish the existence of the three gluon vertex unambiguously, since this 20% is the same order of magnitude as a typical higher order contribution in QCD. Unfortunately, only the lowest order Feynman graphs have been calculated up to now for the relevant processes. (There is no hope to go behind this level in the near future.) Thus, this type [2] of experimental results could be seriously affected by the unknown higher order terms. As it has already been emphasized, the solution to this problem is the identification of the parton ancestor of the jets [3].

Gluons carry a stronger color charge than quarks, and one expects this feature to lead to differences in their fragmentation, namely higher multiplicity, softer hadron spectrum and broader p_t distribution. To date, intense theoretical study has been devoted to calculate variables sensitive to the differences between quark and gluon jets [4]. On the other hand there, are several theoretical papers which provide methods to analyse the experimental data and so to see the above mentioned differences between quark and gluon jets [5,6].

Large amount of data has been accumulated on the fragmentation of quark and antiquark jets in e^+e^- annihilation. Far less is known experimentally on the fragmentation of high energy gluon jets. The experimental problem in these studies is that in e^+e^- annihilation quark jets occur dominantly in the two-jet ($q\bar{q}$) topology, but gluon jets appear only in three- or more-jet topologies ($e^+e^- \rightarrow q\bar{q}g, \gamma \rightarrow ggg$). There is thus a large kinematic difference superimposed on any dynamical difference. In $p\bar{p}$ collisions two-jet topologies are a mixture of $q\bar{q}, qg$ and gg events. Due to these problems most gluon jet fragmentation studies have led to inconclusive results.

One can study symmetric three-jet events and compare the results with the results of two-jet events produced at $\frac{2}{3}$ center of mass (CM) energies [7]. The MARK II Collaboration has studied inclusive charged particle distributions in terms of the fractional momentum. Their study apparently favours a softer gluon fragmentation. On the other hand the TASSO Collaboration has repeated these studies but their results do not confirm the MARK II result.

There are several approaches based on studies of jets with different energies [8]. The JADE collaboration has reported that particles coming from the least energetic jet (the most probable appears to be a gluon jet) in three-jet events tend to carry on the average larger p_t . In the studies of the CELLO Collaboration, they have compared the least energetic jets in three-jet events at $W=35$ GeV to two-jet events at $W=14$ GeV (thus a gluon enriched sample of jets has been compared to a quark enriched sample), they report $\frac{\langle p_{t(35)} \rangle}{\langle p_{t(14)} \rangle} = 1.03 \pm 0.03 \pm 0.04$ not seeing any increase in p_t for gluon jets. The AMY Collaboration has studied different regions of the three-parton phase space with the same jet energies but slightly different quark contents. They have seen differences in the concentration of energy near the jet core, in the rapidity of the leading particle and in the distribution of the energy-energy correlation for quark and gluon jets. (This type of study compares jets in drastically different jet environments [6]. They have compared jets in "nearly collinear" configurations with jets in "nearly symmetric" three-jet configurations. The difference between the quark contents of these two samples is very small, a few %.)

In spite of the remarkable amount of experimental effort, no unambiguous differences between quark and gluon-induced jets have been established so far.

A recent proposal to identify the parton ancestor of jets has been a neural network method [9,10]. The method is capable of separating gluon and quark jets originating from Monte Carlo generated e^+e^- events with $\approx 85\%$ accuracy. The result does not depend on the Monte Carlo model used. The network has access to the energy of the jets, thus might exploit the fact that gluon jets are usually less energetic.

The primary goal of this work is to present a possible method to discriminate quark and gluon jets by using a neural network combined with QCD in electron-positron annihilation. (Some early results have been presented previously [11].) The training of the network has been accomplished by quark and gluon jets of the same energy, thus only the internal structure of a jet has been considered (The network cannot utilise the fact that gluon jets are less energetic). The input data, based on

our Monte Carlo event generator, correspond to a total basis, in the sense that all the longitudinal and transverse momenta of outgoing hadrons with respect to the jet axis (excluding soft particles) have been used in the analysis. On the other hand, the QCD matrix element for definite jet energies also gives the probability of a jet being a gluon or quark. By the combination of the neural network, and the QCD second order matrix element, we show methods to separate quark induced and gluon induced jets with accuracy superior to all previous attempts.

This sort of study (e.g. [9,10] and also the ones presented in this paper) are remarkably different from experiments at least in one respect. In the Monte Carlo analysis it is easy to determine the parton ancestor of a jet since the event generator is under control. One can record the momenta of the partonic state and only then fragment the partons. By comparing the momenta of the partons and the jets it is easy to assign partons to jets. In the present Monte Carlo study we also know the parton ancestor of the jets therefore the network can be trained by well-defined quark and gluon jets. In real experimental situations the parton ancestor is unknown (actually the problem is to determine it) therefore this information may not be included in the training set. Experimentally one cannot train the network with known quark and gluon jets.

The discussion is organised as follows: Section 2 describes the neural network algorithm applied. Section 3 contains the properties of our Monte Carlo generator, fragmentation parameters, jet finding algorithm and the specific features of the NN used. In Section 4 we give methods based on kinematic variables to discriminate quark jets from gluon jets and critically discuss the previously proposed NN analysis. Section 5 contains the details of NN studies used in the present paper and also the combined (QCD and NN) method. The conclusions can be found in Section 6.

2. Neural Network Algorithm

The new interdisciplinary method of Neural networks (NN), besides their theoretical significance, have proven successful on various fields of science and provide an alternative way for data processing. In this paper we use one of these NN methods to separate quark jets from gluon jets by studying their fragmentation properties (the longitudinal and transversal momenta of the outgoing hadrons), thus solve a discrimination problem.

In general, the application of neural networks consists of three distinct epochs:

learning, testing and generalisation. In discrimination problems the learning phase corresponds to the construction of the discrimination function. This is realized by providing input patterns (namely: multidimensional vectors) to the NN. In the testing phase one tests the performance of the discrimination function produced by the network with data that has been never seen by the NN. Generalisation, the active phase of the network yields predictions for unknown samples.

With the present model learning and testing is realized in a supervised way: jets (produced by an electron-positron Monte Carlo generator) simultaneously with the information on their established quark or gluon origin are provided to the NN. Hence the network is enabled to react based on correct information.

We encourage experimentalists to perform the generalisation and classify the experimental data.

2.1 The Learning Vector Quantization method (LVQ)

LVQ is a modified supervised-learning version of the original self-organizing Kohonen model that has been introduced and described in detail in the author's book [12]. The model aims to create a topologically correct internal representation of the arriving inputs. This behaviour may be utilized to perform classification of the primary data.

A version of the self-organizing model, the Learning Vector Quantization (LVQ) [12] has already proven to be superior to the widespread Backpropagation (BP) [13] model in hard statistical pattern recognition problems [14].

In the r category, d dimensional LVQ discrimination method for each c category we have (n_c) number of d dimensional reference vectors or "reference centers" $(\mathbf{k}_i^{(c)}, i = 1, \dots, n_c; c = 1, \dots, r)$. The learning period begins with the initialization of the reference vectors (e.g. the first n_c input data for each category are equal to $\mathbf{k}_i^{(c)}$). Then an iterative adaptation process follows. In this process we move the reference vectors in the discretized time t according to the incoming data $\mathbf{x}^{(c)}(t)$. (Here the incoming data $\mathbf{x}(t)$ produced at t belongs to the category c .) These "movements" are organised in a way to obtain an optimal discrimination with the nearest neighbour method.

(In our case we have an $r = 2$ category problem, since we want to discriminate quark jets and gluon jets. $c = 1$ for quark jets and $c = 2$ for gluon jets. The input multidimensional vectors contain the longitudinal and transverse momentum fractions

with respect to the jet axis, carried by the outgoing hadrons. The first $n_{quark/gluon}$ input vectors are used for the initialization.)

The learning algorithm is the following.

Let $\mathbf{k}_w^{(m)}(t)$ be the reference vector closest to a given input $\mathbf{x}^{(c)}(t)$ in some (e.g. Euclidean) metric at time t :

$$\rho \left[\mathbf{k}_w^{(m)}(t) - \mathbf{x}^{(c)}(t) \right] = \min_{i,s} \left(\rho \left[\mathbf{k}_i^{(s)}(t) - \mathbf{x}^{(c)}(t) \right] \right), \quad (1)$$

where ρ is the distance and the minimum is with respect to all reference vectors in all the categories. If $\mathbf{x}^{(c)}(t)$ belongs to the same category as the closest reference vector ($c = m$) than this reference vector is updated in the following way:

$$\mathbf{k}_w^{(m)}(t+1) = \mathbf{k}_w^{(m)}(t) + \alpha(t) \left[\mathbf{x}^{(c)}(t) - \mathbf{k}_w^{(m)}(t) \right], \quad (2)$$

if they are not identical ($m \neq c$)

$$\mathbf{k}_w^{(m)}(t+1) = \mathbf{k}_w^{(m)}(t) - \alpha(t) \left[\mathbf{x}^{(c)}(t) - \mathbf{k}_w^{(m)}(t) \right], \quad (3)$$

for all other reference vectors

$$\mathbf{k}_i^{(m)}(t+1) = \mathbf{k}_i^{(m)}(t), \quad (4)$$

where $0 < \alpha(t) < 1$ and $\alpha(t)$ decreases monotonously with the discrete time.

As a result of the adaptive learning process the reference vectors are moved to map the multi-dimensional distributions of the different categories, thus all the clusters in this multi-dimensional space are supposed to be represented by certain reference vectors.

In the test phase one provides inputs of known origin. The network selects the closest reference vector to the given input and then the known category of the input and the category associated with the selected reference vector are compared. In this way one can measure the classification performance of the network.

In the generalizing phase category predictions for inputs of unknown class are yielded. (In the present case this means the discrimination between quark and gluon jets in the experimental data.)

These ideas are illustrated in Fig.1. The discrimination problem is the real quark vs. gluon jet separation. We have plotted jets having 30 GeV energy in two dimensions (transverse momentum fractions of the first vs. the third particles). The

naive approach results in a straight line discriminating curve (Fig.1.a). Following the above method (e.g. having two quark and two gluon reference vectors) the created discriminating curve is not a straight line, it has two bends (Fig.1.b), consequently the performance increases. By adding new reference vectors the discriminating curve becomes more and more sophisticated, consequently the performance of the discrimination increases.

3. The e^+e^- Monte Carlo events

To study the properties of quark and gluon jets we have used Monte Carlo data generated by JETSET 6.3 [15]. The partonic configuration is calculated from perturbative QCD in the form of second order matrix element for finite jet resolution parameters y_{min} . We have changed the QCD part of JETSET 6.3 to get rid of the GKS [16] approximation, by the method of the MARK J collaboration (for more details see [17]). In our Monte Carlo analysis we have used the the fragmentation parameters fitted to LEP data given by OPAL [18].

It is worth mentioning two things.

- It is well known that the parton shower option gives a better description of the final hadronic state in general [e.g. 18]. Particularly its predictions for $n \geq 4$ jet rates and for global observables are in a good agreement with the experimental data.
- On the other hand in our work we want to study and compare separate jets, (all parton ancestors are well separated) thus the fixed order perturbative results are applicable in this region of the phase space. (The leading-log approximation used in parton shower models is valid close to the soft and collinear configurations, but does not necessarily give relevant answer for well separated parton states. For example the three-jet distribution in parton shower models is readjusted to reproduce the fixed order matrix element result. It is easy to understand the good agreement between the parton shower model and the data. One of the most important reasons is that the production close to the soft and collinear region is highly favoured.) As it has been emphasized in [18] the second order matrix element model provides a reasonable description of some regions of the data distributions which are dominated by three-jet production.

Another possibility could have been to study Monte Carlo data generated by different (e.g. HERWIG 3.4 or ARIADNE) e^+e^- generators [14] (models based on parton shower approach) but as mentioned above the second order matrix element

method with fragmentation parameters fitted to the LEP data is perhaps the best available solution for the well separated three-jet event studies. (On the other hand there are some hints in [10] that the NN trained by JETSET data can recognize HERWIG or ARIADNE data pretty well. The differences seen by their network are not really relevant.)

3.1 Jet samples

2 000 000 hadronic events at the Z^0 resonance have been used. In the hadronic final states not only the charged, but also the neutral particles have been considered.

After generating an event we have used the LUCCLUS cluster algorithm [12] to determine the number, the energies and the directions of the jets. The method used in LUCCLUS is to form clusters of particles which are close together in phase space where closeness is defined as the relative p_t of two particles. All particles which are closer together than a definite cut (d_{join}) are clustered into a cluster or jet. We have used in our treatment $d_{join} = 3.5$ GeV. Of course the jet number, as determined by the cluster algorithm, does not always coincide with the parton number. Events, where it does not coincide are called background. Fortunately this background is negligible at this energy and d_{join} parameter for three-jet events. Our monte Carlo studies show that $\approx 99\%$ of the three jet events originated from three-parton events and there is a very small 0.6% and 0.4% background from four- and two-parton events respectively. Soft particles obtaining energy less than 5% of the total jet energy have been rejected.

We have analysed the following data samples:

- a. symmetric three-jet events, where all the three jets have approximately the same energy of about 30 GeV
- b. non-symmetric three-jet events

(In both cases the partonic configuration contained three partons)

A quark (antiquark or gluon) jet is defined as the jet closest to the quark (antiquark or gluon) parton, with the closeness defined as the relative angle between them. (This closeness definition gives exactly the same parton-jet correspondence as the application of the above mentioned p_t closeness definition e.g. used in [10].) According to the fixed order matrix element Monte Carlo approach we have only used events where the jets are well separated. The minimum angle between two jets is 35° . In addition we have required that no jet has less than 5% of the total center of mass

energy.

- c. In order to compare our results, particularly the results obtained by imposing a cut on the energy of the jets, with those of [10,11] we have also studied the JETSET 7.2 Parton Shower option with the default values of the generator. In this analysis the same cluster algorithm has been used as above. The d_{join} cut has been set to 2.5 GeV, with an additional constraint requiring exactly or at least three jets. No jets could have less than 5% of E_{CM} . A quark jet is defined as the jet closest to a quark with the closeness defined in the same way as in the clustering. Jets not assigned to a quark are defined as gluon jets.

3.2 Configuration of the neural network

Test results have shown decreasing fluctuations and higher performance in the success rate when the number of the reference vectors has been increased (Fig.2.a,b), therefore we fixed their number in all the following investigations to 50, a computationally reasonable, but large enough number.

Reference vectors have been associated in equal number with quark and gluon categories. Initial values for the vectors were picked out from the input data namely the first 25 quark and gluon data were chosen. The distance between the reference vectors and the input was measured by the sum of the absolute coordinate differences weighted by $1/w_i$. ($w_i = (\sigma_i^{(q)} + \sigma_i^{(g)})/2$, where $\sigma_i^{(q)}$ is the square-root of the variance of the i -th coordinate of the input data for quarks and gluons respectively.) We have used the ordinary Euclidean metric with the above-mentioned weighting but we have not seen any increase in the performance.

The accuracy of the prediction monotonously increased with the number of learning steps and seemed to saturate after about 8000 steps (Fig.2.a.) 4000 test patterns from each category have been used in the test period to determine the network's performance.

The structure of the input information we have used is the following

- a. the jet label (whether a particular jet was of quark or gluon origin),
- b. the energy (E_{jet}) and the invariant mass square (Q_{jet}^2) of the jet,
- c. the longitudinal p_L/p_{jet} and transverse p_T/p_{jet} momenta fractions of all the outgoing particles. Previously mentioned soft particles have been rejected. Particles were listed according to decreasing absolute momenta.

4. Quark/gluon jet separation based on kinematical variables

In this section we review the possible discrimination methods based on the kinematical variables of the outgoing jets.

In the following the performance of a given method will be presented in the form e.g. 84%(76/88). This notation means that the method gives 76% and 88% identification success rates for gluon and quark jets respectively, while the weighted (for all the quark and gluon jets used in the analysis) averaged identification rate is 84%.

For a three-jet event $0 < x_q, x_{\bar{q}}, x_g < 1$ and $x_q + x_{\bar{q}} + x_g = 2$ (here $x_q, x_{\bar{q}}, x_g$ denote the $2E_{jet}/E_{CM}$ type energy fractions of the quark, antiquark and gluon jets respectively), thus there are only two independent variables, e.g. the energy fractions of the quark (x_q) and the antiquark ($x_{\bar{q}}$). The possible kinematical configurations are: $x_q < 1$; $x_{\bar{q}} < 1$; $x_q + x_{\bar{q}} > 1$. Due to the bremsstrahlung-like process of the gluon emission - three-jet production - most of the final states have $x_q, x_{\bar{q}} \approx 1$. This two-dimensional probability density distribution (as a function of x_q and $x_{\bar{q}}$) is plotted on Fig. 3. The upper triangle in $x_q, x_{\bar{q}}$ plain [(1,0); (1,1); (0,1)] contains the above mentioned kinematically allowed final states. Clearly the upper right corner of this triangle is highly favoured.

4.1 Selection based on energy cut

As gluons are produced in bremsstrahlung-like processes, they usually have smaller energy than the primary quark jets. Thus the most transparent way to separate gluon jets from quark jets is to impose a cut on jet energies (e.g. jets having energy larger than a given cut are called quark jets, others are called gluon jets). It is straightforward to show that the best discrimination can be obtained at the jet energy where the percentage of the quark and gluon jets are equal (50%) in the jet sample. Thus we have calculated the percentage of the quark and gluon jets in our three-jet sample, at different jet energies. The result is plotted in Fig. 4. It can be seen that the number of quark and gluon jets at ≈ 23 GeV is equal. Imposing a cut on our three-jet sample at 23 GeV 77% of the gluon jets are detected as gluons and 93% of the quark jets are detected as quark, thus the performance of this analysis is 88%(77/93).

In Fig 5.a. the result of the $E_{CM}/4$ energy cut is presented in a $x_q - x_{\bar{q}}$ plot. As already mentioned the kinematically allowed region is the upper triangle. The dotted area gives those $x_q, x_{\bar{q}}$ values where the method identifies all the three-jets correctly (100% performance). In the shaded area the energy cut method recognizes the q and $\bar{q}$ correctly, but it does not classify the gluon jet correctly because its energy is larger

than $E_{CM}/4$. In this region of the phase space the performance is 67%. In the plain white area the approach misidentifies the gluon and one of the quark jets but the other quark is correctly determined (33% performance). Since most of the events are in the dotted area (see Fig.3.) it is easy to understand the high success rate of this type of analysis.

We have checked this simple energy cut method for the multijet sample of [10]. (JETSET 7.2, minimal angle between jets 40° , $d_{join} = 2.5$ GeV, requiring at least three jets, no jet has less than 5% of the total E_{CM} , 5400 hadronic events) They report 84% success rate with a feedforward neural network with one hidden layer. Surprisingly, the simple energy cut leads to a higher 85% result. Presumably their network has learnt that jets with small energy are usually gluons, while jets with larger energy are quarks.

4.2 The jet with the smallest energy

Another possibility is to select the jet with the smallest energy as the gluon jet. This approach gives 88%(82/92). This discrimination method needs more information than the previous one, since the detailed description of the whole event (the energies of all the three jets) is needed to select the jet with the smallest energy. Surprisingly this additional information does not increase significantly the performance of the analysis (actually there is a small 0.7% increase).

In Fig 5.b we have plotted the different regions with different success rate. (There is no region with 67% success rate because if one misidentifies the jet with the smallest energy as the gluon jet then automatically misidentifies another jet as well.)

4.3 Probabilities given by QCD

The underlying perturbative QCD matrix element gives the $p_g(x_a | x_b, x_c)$ probability of a jet (in a three-jet event) with a given $x_a = 2E_a/E_{CM}$ energy fraction being a gluon one (the two other jets have x_b and x_c energy fractions). For instance in first order of α_s this probability is clearly:

$$p_g(x_a | x_b, x_c) = \frac{x_b^2 + x_c^2}{(1 - x_b)(1 - x_c)} \times \left(\frac{x_a^2 + x_b^2}{(1 - x_a)(1 - x_b)} + \frac{x_b^2 + x_c^2}{(1 - x_b)(1 - x_c)} + \frac{x_c^2 + x_a^2}{(1 - x_c)(1 - x_a)} \right)^{-1} \quad (5)$$

One can apply this probability (5) to identify jets, e.g. if $p_g > 1/2$ then a jet is said to be gluon, otherwise quark. This identification method yields 87%(77/92) performance for hadronic three-jet events.

For a given p the equation of the discriminating curve in first order of the strong coupling constant is the following (if $p_g > p$ then a jet is said to be a gluon jet):

$$x_b^3 + x_c^3 + (3p+1)(x_b+x_c)x_bx_c - (6p+1)(x_b^2+x_c^2) + 12p(x_b+x_c-x_bx_c) - 8p = 0 \quad (6)$$

(The second order curves are very close to these. The distance - measuring it in x_q - is smaller than a few %, not affecting the performance of the jet separation.) As has been emphasized in [6] the $p = \text{const}$ curves in the x_b, x_c plane are almost parallel with the $[(1,0),(0,1)]$ line. Thus the jets in one of these jet samples ($p = \text{const}$) have almost the same energy. This is the reason why the methods in Sect. 4.1 and 4.3 give almost the same success rate.

Similarly to the energy cut and the minimal energy jet method, it is straightforward to determine the regions where the performance of the analysis is 100-67-33%. The complete result is plotted in Fig 5.c. The shaded and plotted regions denote the same as in Fig. 5.a.

5. Neural network results

In Section 4 we have seen that kinematical variables contain a lot of information on the fact whether a jet is gluon or a quark induced. It is clear that a NN realizes this simple fact and this large kinematic difference is overemphasized on any dynamical one. The network is not able to find the underlying information, cannot see the structure of the jets. This is the reason why [10] cannot exceed the simple energy cut performance. In order to utilize the jet structure in the identification of jets a more sophisticated approach is needed.

In this Section we will use NN to separate quark jets from gluon jets. First we study the whole event sample, then choose the symmetric three-jet events, finally we analyse the jet discrimination performance in definite energy regions and combine it with the kinematic informations.

5.1 Study of the whole three-jet sample by NN

Our first approach was to teach the NN with the data set including energy, invariant mass square and all the transverse and longitudinal momenta of the six most

energetic particles. The performance has been quite independent of the learning steps and of the number of particles in the input data.

The obtained success rate of 88% does not exceed the result in Section 4 where no NN has been used, only the energy as a cut was applied. It seems that the NN has picked out energy as the most important selection factor rather than the hidden information contained in other details. As a conclusion the failure of this type of brute force approach has been confirmed since even providing the network the whole data set turned out to be of no use.

Thus a different approach, based on a separate study of kinematic properties and internal structure of jets is needed.

5.2 Study of symmetric three-jet events

As we have seen in 5.1 Section there has been no improvement in the quark/gluon discrimination if one uses the structure of the jets. The performance remains the same as it was with the energy cut. The dynamical information has remained hidden.

One of the possible best ways to get rid of these problems, and really analyse the internal properties of the jets is to study symmetric three-jet events. In this case real dynamical differences appear, since the quark and gluon jets are of the same energy. There is another advantage of the symmetric three-jet event studies, namely the available phase space for fragmentation is the same for all the three partons. (It has been shown in [20] that the Monte Carlo fragmentation depends on the environment in which the parton is imbedded in.)

Nevertheless there are two far from trivial problems with symmetric three-jet events.

- There is a serious difficulty with statistics. Even the nearly symmetric event selection where the angles are between 100° and 140° (this was the choice of the TASSO Collaboration) cuts the number of the multihadronic events by two orders of magnitude.
- Moreover, there are still kinematic configurations where the quark jet has 40% higher energy than gluon jet.

Thus we have studied nearly symmetric three-jet events, and selected a sample where the average energy of the gluons and quarks are equal. The nearly symmetric three-jet selection criteria for our case is that our sample must contain jets with energy between 30 GeV and 31 GeV.

We have tested the performance of the network by varying both the number of reference vectors and the learning steps. We have also checked the success rate dependence on the number of particles encountered. To avoid the computer load of excess dimensions we tested how excluding the softest particles from the input data affected the success rate. We found the success rate to saturate after 5-6 input dimensions (Fig.6). Increasing the number of the input space dimensions the success rate is slightly decreasing, because not all the jets have multiplicity large enough. In our analysis we have used the six most energetic particles for our input data. (Sect. 5.3.) The network's performance vs. the number of learning steps is shown in Fig.2. After 8000 learning steps the NN seems to saturate at a maximum accuracy of 79-80%.

5.3 Non symmetric three-jet events studies with NN within energy classes

As we have seen in the studies of the symmetric three-jet event, (where the jet energy is irrelevant information) the jet structure contains a lot of information on the parton ancestor of a jet. (At least at the level of Monte Carlo events.) We have tried to exploit this information to compare jets of the same energy.

We have neglected energy from the NN input data and sorted the jets into ten energy classes ranging from 5 to 50 GeV by 5 GeV, with each class having 8000 jets. Learning and testing were performed separately for the energy classes.

Thus the NN was able to utilize only the information contained in the Q^2 and in the momenta fractions of the particles in the jets. The reference vectors were expected to represent and find the fine details hidden in the data. In this study we have modified the nearest neighbour method. A given jet is said to be quark jet if the ratio of its distance from the nearest quark reference vector and its distance from the nearest gluon reference vector is smaller than an energy dependent parameter. These parameters were chosen in that way to obtain approximately the same success rate for quark and gluon jets.

The above procedure resulted in a success rate of 71% as a weighted average over all energy categories. We have also analysed the performance of the network trained on a given energy and tested on another one. The results are summarized in Table 1. It can be seen that the success rate is high for quark jets if the test run jet sample has smaller energy than the training sample, while the success rate is high for gluon jets if it is higher. This fact is in good agreement with the expectation that gluon

jets are softer than quark jets. A high energy quark jet with high multiplicity is recognized by the neural network as a gluon jet. Close to the soft and collinear region the background is high, thus the performance of the NN is small.

The obtained success rate seemed to be low but the information yielded by the NN could be considered to be independent of the QCD categorization based on energy values. Therefore to improve the performance one would try to combine the information gained by the NN and the underlying QCD matrix element. The NN approach is a nearest neighbour method with a discrete output, thus the first step is to extend it in a continuous way, in order to obtain probability values for a given jet being of quark or gluon origin.

It is straightforward to define the R variable for any jet given by $\mathbf{x}$:

$$R = \frac{\rho(\mathbf{x} - \mathbf{k}_w^{(g)}) - \rho(\mathbf{x} - \mathbf{k}_w^{(q)})}{\rho(\mathbf{x} - \mathbf{k}_w^{(g)}) + \rho(\mathbf{x} - \mathbf{k}_w^{(q)})} \quad (7)$$

where $\mathbf{k}_w^{(g/q)}$ is the nearest quark or gluon reference vector to the input $\mathbf{x}$ (as in eq.1). R falls within $[-1, 1]$.

We calculated R values for the gluon and quark jet test samples separately. $f_q(R)$ and $f_g(R)$ denote the gained density functions for 30 GeV energy jets. These functions are shown in Fig.7. These functions can be used to obtain a probabilistic value for a given input with a certain R value. The input jet is expected to be a quark jet with a probability of $p_q^{(NN)}$ and a gluon jet with probability $p_g^{(NN)}$.

$$p_q^{NN}(R) = \frac{f_q(R)}{f_q(R) + f_g(R)} \quad , \quad p_g^{(NN)}(R) = \frac{f_g(R)}{f_q(R) + f_g(R)} \quad (8)$$

respectively. These probabilities are plotted in Fig.8. as a function of R .

Provided the $p_g^{(QCD)}$ by (5) and $p_g^{(NN)}$ by (8) probabilities are independent one easily gets the $\tilde{p}_g$ combined probability for an input to be a quark jet by:

$$\tilde{p}_g = \frac{p_g^{(QCD)} p_g^{(NN)}}{1 - p_g^{(QCD)} - p_g^{(NN)} + 2p_g^{(QCD)} p_g^{(NN)}} \quad (9)$$

(This probability may be combined with any other probability of being a quark jet.) The above combination of the probabilities with a $\tilde{p}_g = 1/2$ cut has led to 92% success rate. This means that the best way is to combine the information of the fragmentation properties and the underlying QCD matrix element.

The same result could be reached by using the quark (gluon) probability at given energy (Fig. 4.), instead of $p_g^{(QCD)}$, thus information on the two other jets is not necessary. This fact is in good agreement with the remark about the constant energy and $p = \text{const}$ jet samples mentioned in Sect. 4.3.

We have tried to give these NN and QCD informations to another neural network, but we have not seen any further improvement in the success rate.

The increase from 87% success rate (only QCD matrix element approach) to the present 92% accuracy does not seem to be a very significant change, however this result means that the misidentification rate is 50% higher in the QCD method (e.g. Section 4.1 or [10]) compared to the combined one.

We have plotted the distribution of $p_{q/g}^{(QCD)}$, and $\tilde{p}_{q/g}$ for quark and gluon induced jets respectively (Fig. 9.). It is quite transparent that the method gives $p \approx 0$ or 1 for most of the quark and gluon jets respectively, thus the separation is pretty effective. As it can be seen there are fewer gluon jets with a predicted probability of being a gluon ≈ 1 . The reason of this is that the $E_{jet} > 0.05 \cdot E_{CM}$ condition rejects jets with small energy, and these jets are the most probable gluon jets.

An interesting phenomenon has also been found. We have tried to teach the NN without sorting the input data according to energy classes but excluding the energy from the input data. This procedure alone yielded 82% prediction rate, higher than the values in Table 1. It is straightforward to apply the previously mentioned method for this sample in order to get a $p_{q/g}^{(NN)}$ distribution for this sample. The combination of the predictions of the NN and QCD matrix elements resulted in only 88%. (Despite the fact that the NN performance has been much higher than in the previous case, in energy classes, this combined method has not been able to reach the success rate of the previous combined method.) This surprising result can be explained by the fact that though the NN has not seen the energy of the jets, nevertheless the network has been able to conclude it from the hidden information in the jet's structure consequently the two probabilities have not been independent. Even these hidden kinematic differences have been superimposed on the dynamical ones in this case.

In our problem no theoretical Bayesian limit could be estimated. Due to the overwhelming number of dimensions (14), no Monte Carlo estimation process can be furnished within reasonable computer time. It does not seem to be very promising to study e.g. a two-dimensional projection of it, since the information as a whole is certainly contained in all of the dimensions. (The performance of the network

significantly has increased with the number particles in the data set.)

6. Conclusions

We have used several methods to separate quark jets from gluon ones. Some of these methods have used a conventional kinematic approach while others a non-conventional simulated NN approach based on the LVQ algorithm. The LVQ and QCD results have been combined resulting in a discrimination rate never seen before.

The conclusions can be summarized as follows:

- i) The kinematic separation based on three possibilities (selection of the minimal jet, information contained by $p^{(QCD)}$ or simple energy cut) resulted $\approx 87-88\%$ identification success rate.
- ii) The NN alone could reach $\approx 88\%$. Thus the performance of the method based on kinematic variables and the naive NN is the same. This is the reason why the success rate in [9,10] with NN is not higher than that with the kinematic cut for the same event sample. Presumably what the neural network does is not more than realizing this simple kinematic information.
- iii) by excluding the jet energy information from the data provided to the NN, we have investigated two cases
 - a. teaching the network with the full available set of input data,
 - b. dividing the data into ten energy categories (equidistantly from 5 to 50 GeV) and processing them independently.

The results we found the most satisfactory were yielded by the combination of the probabilities given by the QCD matrix element and the latter (b) case. This method has led to 92% identification accuracy.

The NN algorithm combined with the QCD method is very easy to implement in parallel hardware, thus offers a very promising way for on-line triggering in high energy experiments.

Acknowledgments

The authors would like to thank the constructive discussions to F. Csikor, T. Geszti, M. Jacob, G. Pócsik and T. Sjöstrand. Transputer cards were supplied by the Physics Institutes of Eötvös University.

REFERENCES

1. H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Phys. Lett.* **47B**, 365 (1973)
D.J. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1343 (1973)
- H. D. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973)
2. M. Jacob, CERN-TH 90-5821, Rapporteur talk given at the 25th ICHEP Singapore, 1990, experimental details:
B. Adeva et al. (L3 Coll.) *Phys. Lett.* **B248** 227 (1990)
M.Z. Akrawy et al. (OPAL Coll.) CERN-PPE/90-97
P. Abreu et al. (DELPHI Coll.) CERN-PPE/90-174
3. Z. Fodor, *Phys. Rev.* **D40** 267 (1989); **D40** 3590 (1989)
S. Bethke, A. Ricker and P. Zerwas, PITHA 90/14 (unpublished)
4. A.H.Mueller, *Nucl. Phys.* **B241** 141 (1984)
J.B. Gaffney and A.H. Mueller, *Nucl. Phys.* **B250** 109 (1985)
G.Sh. Dzhaparidze, *Z. Phys.* **C32** 59 (1986)
5. H.P. Nilles and K.H. Streng, *Phys. Rev.* **D23** 1944 (1981)
O. Nachtmann, *Z. Phys.* **C16** 257 (1983)
L.M. Jones, *Phys. Rev.* **D39** 2550 (1989); **D42**, 811 (1990)
Z. Fodor, *Phys. Rev.* **D41** 1726 (1990)
6. Z. Fodor CERN-TH 6005/91
7. HRS Collaboration, M. Derrick et al., *Phys. Lett.* **165B** 449 (1985)
MARK II Collaboration, A. Petersen et al., *Phys. Rev. Lett.* **55** 1954 (1985)
TASSO Collaboration, W. Braunschweig et al. *Z. Phys.* **C45** 1 (1989)
8. JADE Collaboration, W. Bartel et al., *Phys. Lett.* **123B** 460 (1983)
CELLO Collaboration, contributed paper to the International Symposium on Lepton and Photon Interactions at High energies, Hamburg, 1987
Y.K. Kim et al. (AMY Collaboration), KEK 90-79 preprint, presented at the XXV International Conference of High Energy Physics, Singapore, 1990
9. L. Lönblad, C. Peterson and T. Rönngvaldsson, *Phys. Rev. Lett.* **65**, 1321 (1990),
10. L. Lönblad, C. Peterson and T. Rönngvaldsson, *Nucl. Phys.* **B349** 675 (1991)
11. I. Csabai, F. Czako, and Z. Fodor, ITP-Budapest Report No. 477
12. T. Kohonen, *Self-Organization and Associative Memory*, 2nd Edition (Springer, Berlin 1988)
13. D. E. Rumelhart and J.L. McClelland, *Parallel Distributed Processing*, (MIT Press, London 1988)
14. Gy. Barna and K.Kaski, *J. Phys. A.* **22**, 5143 (1989)
15. T. Sjöstrand and M. Bengtsson, *Comp. Phys. Comm.* **43** 367 (1987)
16. F.Guthbrod, G.Kramer and G.Schierholz, *Z. Phys.* **C21** 235 (1984)
17. F. Csikor, M. Dhina and G. Pócsik, *Modern Phys. Lett.* **A12**, 1177 (1988)
18. OPAL Coll., M. Z. Akrawy et al., *Z. Phys.* **C47**, 505 (1990)
19. G. Marchesini and B. R. Webber, *Nucl. Phys.* **B310**, 461 (1988); L. Lönblad, LU TP 89-10 (unpublished)
20. W. de Boer, CERN-PPE/90-161-IEKP-KA/90-06, preprint, November 1990

Table 1. The performance of the network for quark and gluon jets trained on a given energy (vertical line bold faced energy values in GeV) and tested on another one (horizontal line). Along the diagonal the training and testing energies are the same. The first number is the success rate (percentage) for gluon jets, the second number is the success rate for quark jets.

Table 1.

	10	15	20	25	30	35	40	45
10	64/68	65/56	44/51	31/58	14/62	7/67	2/76	0/89
15	32/76	65/68	80/64	90/41	98/36	98/23	99/14	99/3
20	19/92	59/77	73/74	87/56	96/50	98/35	99/21	99/6
25	12/94	42/90	59/87	72/70	84/68	90/53	98/39	99/17
30	1/99	17/97	40/94	58/85	79/78	88/65	95/50	99/23
35	0/99	8/99	23/98	50/91	66/85	82/77	89/63	99/33
40	0/99	4/99	13/99	33/95	40/94	61/89	79/81	95/54
45	0/99	6/98	15/99	38/94	52/88	67/86	71/73	58/47

Figure captions

Fig. 1. Discrimination of 30 GeV quark and gluon jets (see text) in two dimension, (circles are gluons, crosses are quarks, large circles are the gluon reference vectors, large crosses are the quark reference vectors) a. naive approach; b. LVQ result with two quark and two gluon reference vectors

Fig. 2. The success rate of the LVQ method for 30 GeV jets vs. the number of learning steps; a. LVQ with 50 reference cells; b. LVQ with 4 reference cells

Fig. 3. Two-dimensional density distribution as a function of x_q and x_g

Fig. 4. The quark and gluon contents of the jet samples with different energy at the Z^0 resonance, as a function of the energy of the jet samples

Fig. 5. The results of the different quark gluon jet separation methods based on kinematical variables. The $[(0,1),(1,1),(1,0)]$ triangle is the allowed phase space in terms of x_q and x_g for three-jet events. In the dotted area the performance of the analysis is 100%, while in the shaded and plain white areas it is 67% and 33% respectively. a. for the $E_{CM}/4$ energy cut b. choosing the jet with the smallest energy as the gluon jet c. QCD prediction based on eq (6), $p = 1/2$

Fig. 6. The success rate of the neural network for 30 GeV jets as a function of the number of the input dimensions (each particle in the input set corresponds to two input dimensions: p_t/p_{jet} and p_t/p_{jet})

Fig. 7. Quark and gluon jet density functions in R

Fig. 8. Probability of being a quark or a gluon jet calculated from eq (8)

Fig. 9. The distribution of $p^{(QCD)}$ and the NN-QCD combined probability for quark jets and for gluon jets

FIGURE 1a

FIGURE 2a

FIGURE 1b

FIGURE 2b

FIGURE 3

FIGURE 4

FIGURE 5a

FIGURE 5b

FIGURE 5c

FIGURE 6