DEUTSCHES ELEKTRONEN-SYNCHROTRON

DESY 93-190
December 1993

Study of Jet Reconstruction Algorithms for HERA *ep* Collider Events

V. Hedberg, C. Jacobsson, L. Jönsson
Dept. of Physics, University of Lund, Sweden

G. Ingelman
Deutsches Elektronen-Synchrotron DESY, Hamburg
and
Dept. of Radiation Sciences, Uppsala University, Sweden

ISSN 0418-9833

NOTKESTRASSE 85 - 22603 HAMBURG

To be sure that your preprints are promptly included in the HIGH ENERGY PHYSICS INDEX, send them to (if possible by air mail):

DESY **Bibliothek** **Notkestraße 85** **22603 Hamburg** **Germany**	**DESY-IfH** **Bibliothek** **Platanenallee 6** **15738 Zeuthen** **Germany**

DESY 93-190
December 1993

ISSN 0418-9833

Study of Jet Reconstruction Algorithms for HERA *ep* collider events

V. Hedberg[a], G. Ingelman[b,c], C. Jacobsson[a], L. Jönsson[a]

[a] Dept. of Physics, Lund University, Sölvegatan 14, S-223 62 Lund, Sweden
[b] Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, D-22603 Hamburg, Germany
[c] Dept. of Radiation Sciences, Uppsala University, Box 535, S-751 21 Uppsala, Sweden

Abstract

Three commonly used jet algorithms and two new jet algorithms have been studied with respect to their performance in the new experimental situation of deep inelastic scattering at the HERA electron-proton collider. Their ability to reconstruct properties of the underlying parton level subprocesses was investigated. This relates to first order QCD matrix elements and higher order parton emissions as described by parton showers. A new method was devised to determine suitable values for the resolution parameters of the algorithms and assess their reconstruction quality. The JADE algorithm, which is frequently used in e^+e^- annihilation, is found to perform less well compared to other algorithms.

1 Introduction

High energy quarks and gluons emerging from hard scattering processes are not directly observable due to the confinement of colour charges in quantum chromodynamics (QCD). Instead they are revealed through the appearence of so-called jets, i.e. collimated flows of hadrons, that can be reconstructed from the observed particles. Although this transformation, or hadronisation, process from partons (i.e. quarks and gluons) to hadrons is not understood from first principles it is well described by phenomenological models. Due to the complexity of the process these models require elaborate Monte Carlo simulation programs for the calculations of detailed final state properties. In all models (e.g. independent, string or cluster hadronisation) the kinematic properties of a jet are strongly correlated with those of the parton from which it evolved. This is not only true in the case of independent fragmentation, where one expects a direct relation, but also for models where, as a matter of principle, one cannot unambigously associate a given final hadron with a specific parton.

The basic, or lowest order, parton level process must be corrected for higher order perturbative QCD effects causing the emission of extra partons on a short space-time scale before hadronisation. These can either be calculated by exact matrix elements, which has so far only been made to low orders in the strong coupling α_s, or approximately by so-called parton showers based on the iteration (to arbitrary order in the coupling α_s) of the basic quark (q) and gluon (g) radiation processes $q \to qg$, $g \to gg$ and $g \to q\bar{q}$. Due to the bremsstrahlung nature of this radiation most partons will be soft (low energy) or collinear with the emitting parton. Hence they will not give rise to separately observable jets, but rather to a softening of the hadron momenta and broadening of the jet corresponding to the original parton. At a reduced rate, however, hard emission at large angles do occur which gives rise to observable extra jets. In order to establish a correlation between a jet and a single parton one therefore has to correctly account for both the parton emission process and the hadronisation process. Two jets will be individually observed in a detector only if they are separated by a distance which is at least as large as the lateral dimensions of the jets. On the parton level this corresponds to fairly high parton momenta and a certain angular separation.

Ever since the first observation of jet structure in e^+e^- collisions the question about how to define jets has been controversial. The picture of a jet as a collimated flow of particles implies that a minimum energy has to be available for the jet-formation. In particular, the particle momenta along the parton momentum direction must be much larger than the transverse momentum fluctuations induced in the hadronisation process. As the parton momentum increases, the jets tend to become more pencil-like which simplifies their identification. Experimentally it is found that 2-jet systems, e.g. in $e^+e^- \to q\bar{q}$, can only be observed as containing two visible jets if the invariant mass is at least 10 GeV [1]. This sets a lower limit on the invariant mass of systems to which jet algorithms should be applied.

Different types of jet-finding algorithms have been developed to provide a way of reconstructing jets but since the definition of a jet is not unique, the algorithms all include a parameter to specify the desired resolution. The danger in blindly relying upon a jet-algorithm lies in the fact that most algorithms will find jets, even if the event is completely spherical and contains no collimated flows of particles. Unphysical results may also be produced with an exagerated jet resolution that splits a well collimated particle flow from

a single parton into two jets, or a too course jet resolution that joins two separately detectable clusters of particles from different partons into a single jet. The outcome of the jet reconstruction is thus a consequence of what algorithm has been used and how the resolution parameter has been set.

Clearly, one needs proper guidance in the choice of jet reconstruction algorithm and the value of its jet resolution parameter. Unfortunately, there is no unique 'best choice' since it depends on the purpose for which the jets are being reconstructed. If one only wants to demonstrate the occurance of multi-jet events the choice is not so critical as when quantitative tests of QCD are to be made, such as measuring the strong coupling constant α_s. Since the latter requires a well defined renormalisation scheme to be fully significant one must invoke exact QCD matrix elements rather than leading logarithm parton shower models. A jet algorithm which is defined such that it can be imposed on analytical matrix element calculations is therefore necessary to enable a direct comparison or to fit data to the theory. Such investigations based on jets are the most interesting, but also the most challenging with respect to the jet reconstruction. It has to be tuned so as to reconstruct only the hard parton emission calculable with exact matrix elements that are only available in leading and next-to-leading order in α_s. The sensitivity to higher order emissions, which can only be calculated approximately, should be minimized by effectively merging the outcome of such extra, softer parton emissions with the few harder jets that can be associated with the matrix element calculation.

Due to the complexity of the event structure and the jet algorithms, these problems can in practice only be investigated through Monte Carlo simulation of complete events based on the theories of electroweak interaction and QCD for the hard processes and a model for the hadronization. To study the physics one may first use exact QCD matrix elements only and neglect the approximate higher order emissions. This is also an advantage for our methodological study of jet finding algorithms, since it is then possible to formulate unambigous quality measures of how well the reconstructed jets correspond to a well defined underlying parton level system. With higher orders giving many but softer partons, that may or may not give observable jets, this becomes an ill-defined problem. Our approach is therefore to assess the quality of jet reconstruction algorithms applied to Monte Carlo generated events with first order QCD effects only. The best performance of the algorithms in reconstructing the partons in such matrix element events will consequently determine proper values of the resolution parameters. Once these values are specified, the effects of higher order emissions simulated through parton showers are added to investigate how the quality of the reconstruction is modified.

Previous studies [2] have mostly investigated the quality of reconstruction by applying jet algorithms to the multi-parton state generated by parton showers and to the same events after hadronization. The test of the algorithm is then restricted to a comparison between the reconstructed parton-jets and the reconstructed hadron-jets. Hence, an algorithm which makes the same mistakes on the parton and hadron level will appear better than it really is. With this type of analysis one is mainly determining how sensitive an algorithm is to the hadronization process and not how well it will reconstruct the hard matrix element.

A further motivation for this new study of jet reconstruction at HERA is that some recently proposed algorithms need to be investigated and compared with the old ones. We have therefore performed a systematic study of the five jet-algorithms: JADE [3], LUCLUS [4], LUCELL [5], ARCLUS [6] and K_T [7]. Their ability to reconstruct jets in HERA

ep collider events is examined with the purpose of extracting information on hard parton emission processes. In section 2 we discuss the characteristics of event and jet topologies at HERA. Section 3 describes the Monte Carlo event generation and the five jet-algorithms are defined and briefly discussed in section 4. In section 5 we present the main results starting with the general behaviour of the algorithms and how to measure their performance, followed by results on the reconstruction of the hard subsystem and the parton four-momenta. Finally, the effects of higher order parton emissions are investigated. In section 6 we give the conclusions of this study.

2 Event characteristics at HERA

At HERA 30 GeV electrons collide with 820 GeV protons giving 314 GeV in cms energy. The four-momenta p_e, P, q, p_ℓ represent the incoming electron, incoming proton, the exchanged boson and the scattered lepton (cf. Fig. 1a). The basic kinematic variables for deep inelastic scattering (DIS) are then given by

$$Q^2 \equiv -q^2 = -(p_e - p_\ell)^2, \quad x \equiv \frac{Q^2}{2P\cdot q}, \quad y \equiv \frac{P\cdot q}{P\cdot p_e}, \quad W^2 \equiv (q+P)^2 = Q^2\frac{1-x}{x} + m_p^2 \quad (1)$$

which specify the momentum transfer squared, the Bjorken-x and y scaling variables, and the invariant mass squared of the hadronic system. Only two of these variables are independent variables.

The major difference between reconstructing jets in e^+e^--events compared with ep-events is due to the presence of the spectator jet in the latter case. This jet, which goes in the proton-beam direction, originates from that part of the proton which does not take part in the hard scattering process. The spectator-jet is in most analysis of no interest and the fact that a major part of this jet disappears into the beam pipe does usually not cause any problems. However, if part of it enter the detector it may upset or even spoil the reconstruction of the jets associated with the hard scattering, henceforth called current jets.

The energy of the spectator jet is usually very high at HERA. Only events with high x-values will have current-jets with similar energy as the spectator-jet. This is easily understood with the help of the naive quark parton model (QPM) in which x is the fraction of the proton momentum carried by the struck quark and the energy of the spectator jet is given by $(1-x)E_p$. Since the cross section is large at small x (and consequently also small Q^2) most HERA events have $x \ll 0.1$ and, with naive QPM kinematics, a spectator jet with energy larger than 700 GeV. Thus, the normal situation is that most of the energy goes into the spectator jet.

In e^+e^--collisions the total centre of mass energy ($\sqrt{s}$) is available in the hard scattering to produce jets. The invariant mass of the hadronic final state, W, is here equal to $\sqrt{s}$, such that the phase space for QCD jet production is directly given by W. This differs from ep-collisions where the initial momentum fraction of the scattered parton regulates how much energy that is available for hard QCD jet production and how much that goes into the spectator jet.

The leading order process (zero-order in QCD) in Fig. 1a contains only one current-jet. In this note these type of events will be called 1 + 1 jet events to account for both the

current jet and the spectator jet. The squared invariant mass of the hadronic final state can in a massless approximation be written as:

$$W^2 = 2(1-x)E_pE_q(1-\cos\theta_q) \tag{2}$$

where E_q is the energy of the struck quark and θ_q is its scattering angle with respect to the proton beam direction, see Fig. 2a. We notice that the separation between the spectator jet and the current jet is directly correlated to W^2. In addition, E_q, θ_q and the azimuthal angle, ϕ_q, of the struck quark, can be calculated from Q^2, y and the azimuthal angle of the scattered electron, ϕ_e, using the following relations:

$$\begin{aligned} E_q &= yE_e + \tfrac{Q^2(1-y)}{4E_{eV}} \\ \cos\theta_q &= 1 - \tfrac{2E_e y}{E_q} \\ \phi_q &= \phi_e + \pi \end{aligned} \tag{3}$$

While the 4-momentum of the current jet in the QPM case can be calculated directly from Q^2 and y, the first order QCD processes have three more degrees of freedom to describe the internal structure of the 2+1 jet system. The two possible processes in this case are the QCD-Compton process where a gluon is emitted from the quark (see Fig. 1b), and the boson-gluon fusion process where a gluon from the proton interacts with the virtual photon to give a quark-antiquark pair (Fig. 1c). The mass of the total hadronic system in a boson-gluon fusion event can (when neglecting parton masses) be written as a sum of three terms:

$$W^2 = m^2_{pq} + m^2_{p\bar{q}} + \hat{s} \tag{4}$$

Here $m_{pq/\bar{q}}$ is the invariant mass of the spectator-jet and the quark/antiquark-jet while $\sqrt{\hat{s}}$ is the invariant mass of the two current-jets. Obviously the expression above is also true for a QCD-Compton event if $m^2_{p\bar{q}}$ is changed to m^2_{pg} which then denotes the invariant mass of the spectator-jet and the gluon-jet.

With ξ denoting the momentum fraction of the parton entering the first order QCD process, the energy of the spectator jet is $(1-\xi)E_p$. With massless partons one obtains:

$$\begin{aligned} m^2_{pq} &= 2(1-\xi)E_pE_q(1-\cos\theta_{pq}) \\ m^2_{p\bar{q}} &= 2(1-\xi)E_pE_{\bar{q}}(1-\cos\theta_{p\bar{q}}) \\ \hat{s} &= 2E_qE_{\bar{q}}(1-\cos\theta_{q\bar{q}}) \end{aligned} \tag{5}$$

where E_q and $E_{\bar{q}}$ are the energies of the quark- and antiquark-jet respectively and θ_{pq}, $\theta_{p\bar{q}}$ and $\theta_{q\bar{q}}$ are the angles between the jets. Although ξ can have any value in the range $x < \xi < 1$, small x values are still correlated with small ξ such that typically $(1-\xi)E_p \gg E_q$ and $E_{\bar{q}}$. Therefore, W^2 will be dominated by the two terms m^2_{pq} and $m^2_{p\bar{q}}$ involving the spectator, cf. Fig. 2b and 2c. This means that high values of W do not neccessarily correspond to high values of the hard subsystem mass $\sqrt{\hat{s}}$. In event-samples with $2+1$ jets produced at HERA it is therefore not enough to study the jet-reconstruction as a function of W. Also $\sqrt{\hat{s}}$ has to be considered as demonstrated more explicitly in Fig. 3. In this scatterplot no strong correlation between $\sqrt{\hat{s}}$ and W can be seen. In conclusion, while the energies and the separation of the spectator jet and the current jet in a $1+1$ jet event are correlated with W, the invariant mass of the hard subsystem in a $2+1$ jet event is not. Therefore, one cannot use W in ep collisions the same way as one uses W in e^+e^- collisions.

3 Event generation

To study the performance of the jet algorithms it is preferable to apply them to Monte Carlo generated events. The available knowledge about the underlying parton process can then be used to judge the correctness of the jet assignment. We use the program LEPTO 6.1 [8] which is known to reproduce the DIS data from previous lepton-nucleon scattering at fixed target energies [9] and also to give a fair description of the limited data presently available from HERA [10].

LEPTO simulates the basic deep inelastic scattering at the parton level and we have chosen to consider the dominating neutral current process. In addition to the leading order QPM process, $\gamma^*q \to q$, also the first order (α_s) processes of gluon radiation, i.e. QCD-Compton, $\gamma^*q \to qg$, and boson-gluon fusion, $\gamma^*g \to q\bar{q}$, are included based on the QCD matrix elements (ME). In accordance with our desire to have a well defined underlying parton level process for our test of the jet reconstruction algorithms, we have switched off higher order QCD parton emission processes. To check the effects of such higher order corrections we later include them through the parton shower (PS) option available in the program. The QCD matrix elements have the usual divergences for soft and colinear emission, which are avoided by a requirement of a minimum invariant mass m_{ij} of the resulting parton pairs. This is normally implemented by the cut $m^2_{ij} \geq y_{cut}W^2$ (see PARL(8) and PARL(9) in [8]). With standard values of y_{cut} in the region 0.015 to 0.0025, this gives hard scattering subsystem masses $\sqrt{\hat{s}} = m_{ij}$ down to a few GeV at small W. As discussed above, jet structures cannot be observed at such small masses and we therefore instead used a fixed cut $m_{ij} \geq 10\ GeV$. By not letting the cut scale with W we avoid introducing a bias in the relation between $\sqrt{\hat{s}}$ and W. This relation, see Fig. 3, is of interest for our investigation of how the jet resolution parameter should change with varying ep scattering kinematics (in particular W) in order to get an optimal jet reconstruction.

Detector effects on the generated events have been neglected, except for a requirement of particles to be within the overall acceptance of the detector, i.e. have a minimum angle of 4° to the beam line. Since HERA covers a wide range in x and Q^2, with large variation in the event topologies, one must investigate the performance of the jet algorithms as a function of this basic scattering kinematics. We have therefore chosen to generate six event-samples (A-F) at fixed x and Q^2, see Fig. 4, representing a kinematic region where these overall event variables can be reasonably well measured by the detector.

4 Description of jet algorithms

We have investigated five different jet algorithms in order to find the one best suited for reconstructing multi-jet events at HERA. All five algorithms use a set of 4-vectors,

representing the 4-momenta of the particles in an event, as input and the 4-vectors of the reconstructed jets are provided as output. The algorithms contain a cut-off parameter which determines the resolution of the jet reconstruction. If, however, the resolution depends on the kinematic region in which the algorithm is applied, a suitable scaling variable has to be found to account for this kinematic dependence. The six specific event-samples mentioned above have been used to study whether there is such a dependence and, if so, find the relevant scaling variable. In this investigation we have initially not included any scaling variables in the jet algorithms. Scaling is not needed for an event-sample with fixed x and Q^2 and different optimal cut-off values for event-samples at different x and Q^2 values will thus indicate the necessity of scaling.

4.1 Jade

The jet reconstruction procedure of the JADE algorithm [3] starts by calculating the distance parameter m_{ij} for all pairs of particles according to the expression:

$$m_{ij}^2 = 2E_iE_j(1-\cos\theta_{ij}) \tag{6}$$

where E_i and E_j are the energies of particles i and j and θ_{ij} the angle between them. The parameter m_{ij} is the invariant mass of particle i and j under the approximation that both i and j are massless. The pair with the smallest m_{ij} is combined into a cluster. This combination can be made according to different schemes [11, 12] where the natural choice of adding their four-vectors is made in the following so as to conserve energy and momentum and obtaining massive jets. The procedure is repeated until all remaining pairs have an invariant mass exceeding a preselected cut-off value, $m_{ij}^2 > y_{cut}M^2$, determined by a resolution parameter y_{cut} and a reference mass M (in this study initially set to 1). The final clusters represent the jets of the event.

4.2 Luclus

The LUCLUS algorithm [4] uses a similar clustering scheme as JADE but the distance parameter is

$$d_{join}^2 = \frac{2(|\bar{P}_i||\bar{P}_j| - \bar{P}_i\cdot\bar{P}_j)\bar{P}_i\cdot\bar{P}_j}{(\bar{P}_i+\bar{P}_j)^2} \tag{7}$$

where $\bar{P}_i$ and $\bar{P}_j$ are the 3-momenta of two particles or clusters. For small angles between $\bar{P}_i$ and $\bar{P}_j$ one can interpret d_{join} as the transverse momentum of i (or j) with respect to the direction given by the sum of $\bar{P}_i$ and $\bar{P}_j$. An additional feature of LUCLUS, which we have used, is the option to reassign particles to new clusters as the clustering procedure progresses.

4.3 Lucell (ΔR-cone)

Another approach to jet-reconstruction is to form cones around localised flows of energy. This was originally developed for $p\bar{p}$ colliders by the UA1 collaboration [13] and we use the equivalent algorithm LUCELL in [5]. The total solid angle coverage of the detector is subdivided into equal cells in pseudorapidity ($\eta = -\ln\tan\theta/2$) and azimuthal angle (ϕ).

The deposited transverse energy ($E_\perp = E\sin\theta$) is summed up in each cell and all cells with $E_\perp$ greater than a preselected value are taken as possible initiators of jets. Taking these initiators in falling $E_\perp$ sequence, the algorithm checks if the total transverse energy in a cone around the initiator cell exceeds a minimum $E_{\perp min}$. If so, these cells define a jet and are removed from further considerations. The jet energy is given by the cells and its direction by their energy-weighted center. The cone size is given by $\Delta R = \sqrt{\Delta\eta^2+\Delta\phi^2}$ which in our study is set to 1 and the minimum initiator energy is 0.3 GeV. Within this scheme all particles are not necessarily assigned to jets.

4.4 Arclus

The ARCLUS algorithm [6] is closely related to the colour dipole CDM [14], where the emission of a gluon from a $q\bar{q}$ pair is treated as radiation from a colour dipole formed by the quark and the antiquark. A major difference between ARCLUS on one hand and JADE, K_T and LUCLUS on the other is that ARCLUS considers all possible combinations of three particles instead of two. The idea is to regard each combination of three clusters (i,j and k) as two clusters forming a dipole which radiates off a third cluster.

Since ARCLUS is working with three particles/clusters instead of two, ARCLUS calculates a parameter called $P_\perp$ three times. To start with

$$P_{\perp k}^2 = m_{ijk}^2(1-x_i+\frac{m_i^2-(m_k+m_j)^2}{m_{ijk}^2})(1-x_j+\frac{m_j^2-(m_k+m_i)^2}{m_{ijk}^2}) \tag{8}$$

is calculated, with $x_i = 2E_i/m_{ijk}$ where E_i is the energy of particle/cluster i in the center-of-mass system of particle i, j and k and m_{ijk} is the invariant mass of these particles. The masses m_i, m_j and m_k of the three individual particles are usually set to zero in which case equation (8) is reduced to

$$P_{\perp k}^2 = m_{ijk}^2(1-x_i)(1-x_j) = \frac{m_{ik}m_{jk}}{m_{ijk}} \tag{9}$$

where m_{ik} is the mass of particles i and k and m_{jk} is the mass of particles j and k.

Since the $P_{\perp k}^2$ defined by equation (9) is calculated from three Lorentz invariant masses it is obvious that $P_{\perp k}^2$ is also Lorentz invariant. What is perhaps less obvious is that for small energies of k (in the cms of i, j and k) or for small angles of k with respect to i or j the two clusters i and j will be allmost back-to-back and $P_{\perp k}$ is then the transverse momentum of k with respect to the direction of i and j.

In the algorithm not only $P_{\perp k}$ is used but also the two other possible $P_\perp$-values ($P_{\perp i}$ and $P_{\perp j}$) are calculated and this is then repeated for all combinations of three particles. The smallest $P_\perp$ is compared to a resolution parameter $P_{\perp min}$ and if it is smaller than this parameter the three particles/clusters which gave the smallest $P_\perp$ are replaced by two new clusters, n and m, in the following way:

The two massless clusters n and m are required to have a combined invariant mass (m_{nm}) which is equal to the invariant mass of the original three clusters (m_{ijk}) and momentum vectors, $\bar{P}_n$ and $\bar{P}_m$, that are equal in magnitude but oppositely directed

along a fixed axis [1]. In the c.m.s. of the three original clusters $\bar{P}_i + \bar{P}_j + \bar{P}_k = \bar{0}$ and thus $m_{ijk} = E_i + E_j + E_k$. Since the two new clusters n and m also have $\bar{P}_n + \bar{P}_m = \bar{0}$ and $m_{nm} = E_n + E_m$ the requirements of $\bar{P}_n = -\bar{P}_m$ and $m_{nm} = m_{ijk}$ assures momentum- and energy-conservation in the clustering-process (provided the original hadrons in the event are assumed to be massless).

The two resulting clusters are boosted back to the laboratory system and the procedure is repeated until all combinations have a $P_\perp$ larger than the specified resolution scale $P_{\perp min}$. The remaining clusters are the jets. It can be noted that with this procedure no hadron will be uniquely assigned to one specific jet but its momentum is shared by two jets and that the minimum number of jets found by ARCLUS is two, i.e. in our case $1+1$ jets.

4.5 K_T

A two step procedure, similar to the one we have proposed in an earlier work [15] for the JADE algorithm, is used in the K_T algorithm [7] for DIS processes, where the first step comprises preclustering of hadrons into a beam jet and so-called final state macro jets. A second step aims at resolving jet structures within the system of macro jets.

In the preclustering, a variable K_{Tij} is calculated for all pairs of particles/clusters i and j according to the scheme:

$$K^2_{Tij} = 2(1 - \cos\theta_{ij}) min(E_i^2, E_j^2) \tag{10}$$

where θ_{ij} is the angle between the momentum vectors of i and j and E_i and E_j are the corresponding energies. In addition, a second variable K_{Tip} is calculated for all particles/clusters i according to:

$$K^2_{Tip} = 2(1 - \cos\theta_{ip}) E_i^2 \tag{11}$$

where θ_{ip} is the angle between i and the proton direction.

One specific feature of the K_T algorithm for DIS processes is that all quantities (E_i, E_j, θ_{ij} and θ_{ip}) are calculated in the Breit frame, i.e. where the exchanged vector boson has four-vector q =(0,0,0,-Q). The boson and proton are here both along the z-axis, but longitudinally boosted compared to their c.ms. system. Consequently, for small angles and for $E_i \leq E_j$, K_{Tij} can be regarded as the transverse momentum, in the Breit frame, of i with respect to the direction of j while K_{Tip} can be regarded as the transverse momentum of i with respect to the proton direction.

At this stage a new variable ($E_\perp$) is introduced in the algorithm and

$$y_{ij} = \frac{K^2_{Tij}}{E^2_\perp} \qquad y_{ip} = \frac{K^2_{Tip}}{E^2_\perp} \tag{12}$$

are calculated. The algorithm proceeds by checking if the smallest y-value obtained is less than one. If this is the case and the y-value is of the type y_{ij} then i and j are combined to a new cluster, preserving both energy and momentum. If on the other hand, the smallest

[1] The direction of this axis is given by $\theta_i = \frac{x_j^2}{x_i^2+x_j^2}(\pi - \psi)$ where θ_i is the polar angle between the axis and particle i and ψ is the angle between particles i and j. This axis is choosen so that the sum of the transverse momentum squared of particle i and j with respect to this axis is minimized.

y-value is of the type y_{ip} then i is regarded as a part of the spectator jet and is removed from further treatment. The procedure is repeated until all y-values are larger than one. In other words, the algorithm checks if $min(K^2_{Tij}, K^2_{Tip}) < E^2_\perp$ in which case either two clusters are combined or one cluster is assigned to the spectator-jet. One can therefore regard $E^2_\perp$ as the resolution parameter in the algorithm's first step which tries to separate the spectator-jet from the macro-jets.

In the algorithm proposed by [7], the algorithm then continues with a second step in which a parameter Q_0 is introduced and the variable $y_{cut} = Q_0^2/E_\perp^2$ is calculated ($E_\perp^2$ remains the same as for step one). All particles in the event which were not assigned to the spectator jet in step one are used again. For every pair of particles the y_{ij}-value is calculated and the pair giving the smallest y-value is compared with the y_{cut} value. If $y_{ij} < y_{cut}$ the hadrons are combined into a cluster. When all y-values exceed the y_{cut}-value the procedure is stopped and the remaining clusters are the final jets. In fact, what is done in this step, is to test if $K^2_{Tij} < Q_0^2$ in which case the clusters i and j are combined to a new cluster. The parameter Q_0 is thus a second resolution parameter which is used to resolve the macro-jets.

It can be added that for our purposes and with the method of chosing $E^2_\perp$ that we have developed (this method will be described below), the second step is not needed and the jets found in the first step can be used directly. We therefore do not discuss the Q_0-parameter further.

4.6 Theoretical considerations

The choice of jet algorithm depends on what physics analysis to be made with the reconstructed jets. In order to have a proper comparison with theoretical results the jet definition should be applicable also in the theoretical calculations. The invariant mass cut used in JADE is suitable for fixed order matrix element calculations where it is often used as the basic cut-off against divergences. This works theoretically well as long as the resolution y_{cut} is not too small. At small y_{cut}, however, higher order corrections become enhanced by powers of $\ln y_{cut}$ and certain resummation techniques are then necessary. In particular, the large logarithmic corrections to the relevant quantity may then exponentiate. As discussed in [16], the K_T algorithm is preferable under such circumstances since its use of transverse momentum to resolve jets is suggested by the coherence properties of soft QCD emission processes in order to preserve the exponentiation. The other algorithms do not have this property (it has been suggested [6] that it might hold for ARCLUS). The distance measure d_{join} and ΔR used in LUCLUS and LUCELL, as well as the complexity of ARCLUS make these algorithms unsuitable for analytical QCD calculations. However, in applications where the theory can be considered to be treated well enough by a Monte Carlo program that generates complete final states, the choice of algorithm can be made more freely since practically any jet definition can then be applied to the theoretical events in the same way as to the data. For a study of some of these jet algorithms applied to e^+e^- annihilation see ref. [12], where also the corresponding next-to-leading order QCD and hadronization corrections are investigated. For DIS, the importance of factorization and the special role of the Breit system have been emphasized and discussed for the JADE- and K_T-type of algorithms [17].

5 Performance of jet algorithms

5.1 General behaviour

As discussed, the proton remnant continues in the direction of the incoming proton and produces a jet which to a large extent escapes down the beam pipe without detection. Some fringe particles might, however, hit the detector and must be treated correctly by the jet algorithm. This can be achieved by using the method of inserting a pseudoparticle in the beampipe according to the following recipe.

The missing longitudinal momentum carried by the spectator-jet is determined from the known longitudinal momentum of the initial state and the measured longitudinal momentum of the final state. The lost fraction of the spectator jet is represented by a pseudoparticle carrying the missing longitudinal momentum and moving along the proton beam direction. This pseudoparticle is treated like any other particle in the jet reconstruction procedure (for details see [18]).

A good starting point in the jet reconstruction procedure is to separate the spectator-jet from the hard subsystem. If this separation is not optimized, the reconstruction of the jets from the hard subprocess can be highly distorted by the very energetic spectator jet. In the first step we therefore concentrate on the reconstruction of the spectator jet and not on the jet structure of the hard subsystem. A relevant measure of how well the spectator jet has been reconstructed is the reconstruction of the invariant mass, $\sqrt{\hat{s}}$, of the hard subsystem.

As shown for Luclus in Fig. 5 the distribution of the difference between the generated $\sqrt{\hat{s}}$ and the reconstructed invariant mass $\sqrt{\hat{s}_{rec}}$ exhibits different shapes depending on how the cut-off value of the resolution parameter is chosen. At low cut-off values, the reconstructed invariant mass has a tendency of being too high and there is a tail on the negative side of the distribution, Fig. 5a. This tail indicates that particles from the spectator jet have been included into the hard subsystem by the jet algorithm. At a somewhat higher cut-off value, Fig. 5b, the negative tail decreases and there is a narrow peak centered at zero but also an additional bump at positive values. This bump corresponds to events where some particles or an entire jet, related to the hard subprocess, have been assigned to the spectator jet resulting in a too low $\sqrt{\hat{s}_{rec}}$. Finally, at high cut-off values, Fig. 5c, the negative tail is even less pronounced and the bump at positive values has grown. This is a general behaviour of all the algorithms studied. The ideal distribution would of course be a narrow peak centered at zero with no tail on the negative side and no bump at positive values.

In order to approach this ideal situation one can use the fact that there is a correlation between the value of $\sqrt{\hat{s}_{rec}}$ and $\sqrt{\hat{s}} - \sqrt{\hat{s}_{rec}}$ as is demonstrated in Fig. 6. A lower mass-cut in $\sqrt{\hat{s}_{rec}}$ of for example 15 GeV remove the bump at positive values entirely and a high enough cut-off value for the resolution parameter surpresses the tail. Such a mass-cut also corresponds to the lower limit above which clear jet structures might be observed. However, the effect of the mass-cut is a decrease of the event sample and since an increasing number of events will populate the bump as the algorithm cut-off value increases, the remaining sample will decrease with increasing cut-off value. The precision of the reconstruction will, however, improve. The events that are lost by the mass-cut are those where the spectator jet is not clearly separated from the hard subsystem, i.e. where one of the jets from the subsystem goes in the very forward direction. These events are

almost impossible to reconstruct in an accurate way and the loss is therefore difficult to avoid.

In this study we have selected three observables which give information about the reconstruction of $\sqrt{\hat{s}}$. Two of these observables, the full width half maximum ($fwhm$) and the root mean square (rms) describe the width of the $\sqrt{\hat{s}} - \sqrt{\hat{s}_{rec}}$ distribution. (The latter is more sensitive to contributions from the tail.) The third observable is the mean value which was used to measure systematic shifts of the distribution.

The five jet-algorithms were applied to an event sample using ten different cut-off values of the resolution parameter for each algorithm. (The cut-off values used in event samples C and D are given as an example in Table 1.) For each value, the fraction of 2+1-jet events surviving the mass-cut and the three observables describing the $\sqrt{\hat{s}} - \sqrt{\hat{s}_{rec}}$ distribution were measured. These observables were then plotted as a function of the fraction of events remaining after the 15 GeV mass-cut. This way of presenting the results is motivated by the fact that in most analysis one has to find a compromize between the statistics needed and the required precision. Another motivation is that the properties of the different algorithms can be directly compared despite their different separation criteria and cut-off values. The investigation was repeated for all six event-samples to see how the quality of the reconstruction changed and if a universal cut-off value in the algorithms could be found for the entire kinematic range studied.

event-sample C	X	IX	$VIII$	VII	VI	V	IV	III	II	I
JADE	1300	1100	900	700	500	400	300	200	100	30
LUCLUS	6.0	5.0	3.8	3.4	3.0	2.6	2.2	1.8	1.4	0.7
LUCELL	6.0	5.0	4.0	3.0	2.5	2.0	1.5	1.2	0.7	0.2
K_T	50.	40.	30.	20.	10.	5.0	3.0	2.0	1.5	1.0
ARCLUS	84.	64.	48.	32.	24.	14.	6.0	2.0	1.0	0.2
event-sample D	x	ix	$viii$	vii	vi	v	iv	iii	ii	i
JADE	95000	80000	20000	5000	1000	500	200	100	50	20
LUCLUS	18.	10.	7.0	5.0	3.8	3.4	3.0	2.6	2.2	1.8
LUCELL	10.	8.0	6.0	5.0	4.0	3.5	3.0	2.5	2.0	1.5
K_T	140.	80.	50.	40.	30.	20.	10.	5.0	3.0	2.0
ARCLUS	140.	100.	80.	60.	40.	32.	24.	16.	8.0	4.0

Table 1: *Jet algorithm cut-off values used for event-samples C and D. The roman numbers in the head of the table are used to correlate the cut-off values for an algorithm with the points plotted in Figs. 8-10. Increasing numbers correspond to increasing cut-off values, i.e. decreasing fraction of remaining 2 + 1-jet events.*

Results from samples with equal W (A,C,E and B,D,F respectively) have a similar correlation between the three reconstruction observables and the cut-off values of the resolution parameters. A possible explanation for this is that events of equal W have similar event topologies. This can be realized from studying Fig. 7a and b where the polar angle with respect to the proton beam direction of the two current jets are plotted against each other. Low and high W-values are examplified by the event-samples C and D. At low W (Fig. 7a) both jets are most likely found in the forward region, i.e. the polar angles are small. This is also shown in Fig. 2c which describes a typical C-event. The high W-events (Fig. 7b) have most often a jet in the forward direction and the other jet in the

backward direction with a θ-value of around 160°. This is examplified in Fig. 2b which shows a typical D-event. At the parton level, ignoring initial state parton showers, the relationship between the θ-values of the jets can be expressed in a massless approximation by

$$2yE_e = E_{\perp 1} \tan\frac{\theta_1}{2} + E_{\perp 2} \tan\frac{\theta_2}{2} \qquad (13)$$

where $E_{\perp i}$ is the transverse energy of jet i. At low x ($x < 0.1$) we can to a good approximation assume that $y \approx W^2/s$ (see Fig. 4) and equation (13) then gives the energy and angular correlations between the two jets and W. Taking the average values of $E_{\perp i}$ for both jets ($\approx 10\ GeV$), we have from (13) calculated the correlation curve for the polar angles and depicted it in the scatter plots 7a and b. Since the jet-angles are mainly correlated to W we present the results below only for two samples (C and D) with different W. Results from the other four event-samples will only be given in the form of summary tables.

5.2 Reconstruction of the hard scattering subsystem

As already mentioned, we have chosen three observables ($fwhm$, rms and the $mean$-value of the $\sqrt{\hat{s}} - \sqrt{\hat{s}_{rec}}$ distribution) to describe how well the invariant mass of the hard subsystem is reconstructed by different algorithms. Fig. 8 and 9 show, for event-samples C and D respectively, the three observables versus the remaining fraction of events as the resolution parameters are varied in the five jet algorithms. Each point in these figures corresponds to a specific cut-off value given in Table 1.

Comparing the results of the five jet algorithms we note that JADE reconstructs $\sqrt{\hat{s}}$ less well than the other four algorithms. In almost all respects it gives worse results, i.e. higher values of $fwhm$ and rms (at least at high W), and a $mean$-value different from zero. ARCLUS seems to perform somewhat better than the other algorithms at low W-values.

Another observation is that the reconstruction of $\sqrt{\hat{s}}$ seems to be better at low (Fig. 8) rather than at high values of W (Fig. 9). The variation of the three reconstruction observables is also much smaller at low W than at high W, making the choice of cut-off values less critical at low W-values.

According to Fig. 8 and 9 the quality of the $\sqrt{\hat{s}}$-reconstruction generally improves with increasing cut-off values and consequently an unambigous cut-off value cannot be found. However, at least for high W we clearly observe that the curves level off which means that the quality in the reconstruction does not improve significantly with increasing cut-off value; only the statistics is reduced further. This behaviour therefore gives an indication about how to make a reasonable choice of cut-off values. The curves describing the $mean$-value (Fig. 9c) level off at somewhat higher cut-off values than what is the case for the curves describing $fwhm$ (Fig. 9a). Our choice of cut-off values is thus mainly determined by the observed behaviour of the $mean$-value. This method of choosing the cut-off values is of course qualitative and, considering the dependence of rms, governed by the importance of good statistics versus good quality. For all algorithms, the curves describing the $mean$-value start to level off at a remaining fraction of events of about 0.5 which corresponds to the following cut-off values: $m^2_{ij} = 2000\ GeV^2$ in JADE, $d^2_{join} = 4.0\ GeV$ in LUCLUS, $E_{\perp min} = 3.5\ GeV$ in LUCELL, $E^2_\perp = 20\ GeV^2$ in K_T and $P^2_{\perp min} = 25\ GeV^2$ in ARCLUS.

The selected cut-off values are indicated in Fig. 8 and 9 by arrows at the very top of the figures. As can be seen in Fig. 8 the cut-off values seem reasonable even for events at low values of W, except for the JADE algorithm, where the remaining fraction of events is almost zero. The same is true for the other two low W-samples, A and E.

Even if the behaviour of the reconstruction observables at low W indicate that the cut-off values can be set quite low without a deterioration of the reconstruction capabilities, we will see later when we study the number of jets as a function of the cut-off value, that JADE requires a m^2_{ij} of about 200 GeV^2. This is a factor of ten lower than the value needed at high W which also correspond to the ratio of high to low W^2 (79200 GeV^2 to 7920 GeV^2). Thus, we find that the well-established use of W^2 as the scaling variable in JADE ($y_{cut} = m^2_{ij}/W^2$) is reasonable for reconstruction of the spectator jet giving a $y_{cut} = 0.025$. The final cut-off values chosen for the five algorithms are given in Table 2.

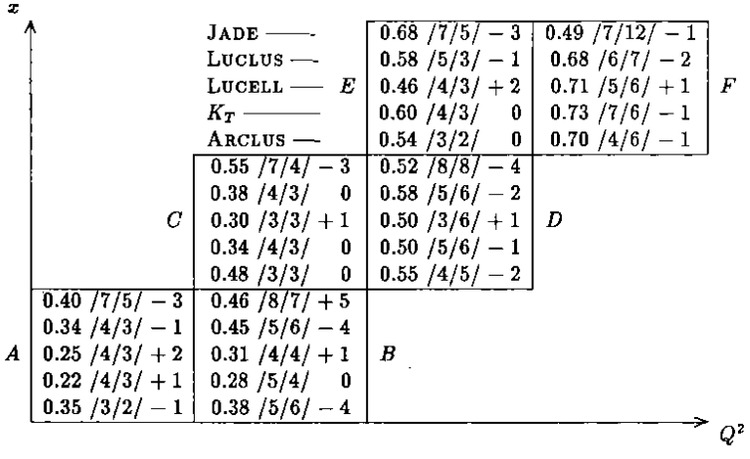
JADE	LUCLUS	LUCELL	K_T	ARCLUS
0.025	4.0	3.5	20.	25.
m^2_{ij}/W^2	d^2_{join}	$E_{\perp min}$	$E^2_\perp$	$P^2_{\perp min}$

Table 2: *The jet algorithm cut-off values chosen according to the method described in the text.*

Using the cut-off values of Table 2, we obtain the reconstruction observables $fwhm$, rms and $mean$ given in Table 3 for all event-samples $A - F$. We conclude that event-samples with equal W have a similar $\sqrt{\hat{s}}$-reconstruction but that for samples with different W values a difference is observed mainly in the rms values. This is due to the more pronounced tail in the $\sqrt{\hat{s}} - \sqrt{\hat{s}_{rec}}$ distribution at high W. The reason for this tail is not so much due to the angular distributions of the jets shown in Fig. 7a and b, but to the energy distributions of the jets closest to the spectator jet, Fig. 7c and d. At high W the most likely jet-energy for the low angle jet is only a few GeV while at low W this energy is 10-15 GeV. We know from Fig. 3 that we have similar $\hat{s}$ distributions for all values of W and consequently, large opening angles between the two current jets, as is true for high W values, in general corresponds to small jet energies. This typically results in a reconstruction of $\sqrt{\hat{s}}$ which is better at low W than at high W.

5.3 Reconstruction of the parton four-momenta

We now turn to the reconstruction of individual jets within the hard subsystem. Our first requirement is that the jet multiplicity, which in our study should be $2 + 1$ jets, is correctly reproduced. How the number of $2 + 1$ jet events is affected by the cut-off values is shown in Fig. 10 where the fraction of $2 + 1$ jet events is plotted against the remaining fraction of events, i.e. the purity of the $2 + 1$ jet sample versus the efficiency of the selection. All algorithms, except JADE at high W, reach a probability of almost 100% to reconstruct the events as $2 + 1$ jet events at the chosen cut-off values of Table 2. The reason JADE performes less well is that the cut-off values of the resolution parameter is scaled with W^2 in order to give a good reconstruction of the spectator jet. As pointed out in the previous section such a scaling results in m^2_{ij} values of 2000 GeV^2 at high W^2. This is clearly too large to resolve jets of the hard subsystem with invariant masses squared around 225 GeV^2. The two jets will thus be merged and one gets $1 + 1$ jet events.



		JADE ——	0.68 /7/5/ − 3	0.49 /7/12/ − 1
		LUCLUS —	0.58 /5/3/ − 1	0.68 /6/7/ − 2
		LUCELL — E	0.46 /4/3/ + 2	0.71 /5/6/ + 1 F
		K_T ———	0.60 /4/3/ 0	0.73 /7/6/ − 1
		ARCLUS —	0.54 /3/2/ 0	0.70 /4/6/ − 1
	0.55 /7/4/ − 3		0.52 /8/8/ − 4	
	0.38 /4/3/ 0		0.58 /5/6/ − 2	
	C 0.30 /3/3/ + 1		0.50 /3/6/ + 1	D
	0.34 /4/3/ 0		0.50 /5/6/ − 1	
	0.48 /3/3/ 0		0.55 /4/5/ − 2	
0.40 /7/5/ − 3	0.46 /8/7/ + 5			
0.34 /4/3/ − 1	0.45 /5/6/ − 4			
A 0.25 /4/3/ + 2	0.31 /4/4/ + 1	B		
0.22 /4/3/ + 1	0.28 /5/4/ 0			
0.35 /3/2/ − 1	0.38 /5/6/ − 4			

Table 3: *The remaining fraction of events and /fwhm/rms/mean in GeV for the five jet algorithms in event-sample A,B,C,D,E and F.*

So even if W^2 is a good scaling variable for JADE in order to obtain a universal cut-off value for the reconstruction of $\hat{s}$, the hard subsystem is not resolved correctly. We have therefore proposed [15] a two step procedure to reconstruct the jet structure of the hard subsystem. In the first step we reconstruct the spectator jet using W^2 as a scale. The spectator jet is then removed and JADE is applied a second time to the remaining hard subsystem. A suitable scale in the second step is the reconstructed invariant mass squared, $\hat{s}_{rec}$. The results of the second step, with $y_{cut} = 0.25$ using $\hat{s}_{rec}$ as the scale, is that almost 100% of the events will be reconstructed as 2+1 jet events, compared to only 30% without a second step, at high values of W^2, see Fig. 10b.

With the modified JADE algorithm we have five algorithms tuned to give reasonable reconstructions of $\sqrt{\hat{s}}$ and the correct jet multiplicity. The next step is a comparison of the jets and the parton four-momenta using the chosen cut-off values presented in Table 2 and applying the 15 GeV mass-cut as before. In addition we require the polar angles of the two reconstructed current jets (θ) to be within 10° and 160°. This improves the reconstruction slightly since it forces most of the jet energy flow to be inside the detector acceptance region. Each jet was assigned to the closest parton from the hard subprocess, which was defined in terms of the three-momentum difference, $\mid \bar{P}_{jet} - \bar{P}_{parton} \mid$, in order to account for both energy scale and direction. The relative error of the reconstructed parton energy as well as the absolute error of the θ-angle and the azimuthal angle ϕ was calculated. As before, the results of event-samples C and D are typical for the behaviour of the algorithms for high and low W-values and are shown in Fig. 11 and 12 respectively. Not surprisingly, these results indicate a behaviour of the jet algorithms which is consistent with that found in the previous chapter reconstructing $\hat{s}$. JADE has problem especially at low values of W and the other algorithms give comparable results.

5.4 Effects of higher order parton emissions

To complete this comparison of jet algorithms we investigate the effect on the reconstruction quality from higher order QCD parton emission. These emissions are included by adding parton showers (in leading log approximation) after the matrix element treatment in LEPTO 6.1 [8]. Within a given event, these additional emissions are associated with a smaller momentum transfer scale compared to the one in the matrix element. The number of emitted partons can be large, depending on the energy (or virtuality) scale of the parton initiating the shower and on the cut-off applied to stop the QCD parton branching processes, but most of them are neither energetic enough nor sufficiently separated in phase space to give rise to observable jets. The main effect is therefore a broadening of the energy flow of the jets given by the matrix element. At a suppressed rate there will, however, be hard emissions resulting in events with a jet multiplicity of 3+1 or more. The exact rate depends, of course, on the jet resolution.

The aim of the jet reconstruction should be related to the physics issue investigated. If the purpose is to study the first order matrix element processes, i.e. to correlate a jet with a parton of the matrix element, events with more than $2+1$ reconstructed jets are problematic. To simply reject them could introduce a bias in the remaining sample and would also lower the statistics of the sample. Alternatively, one can reassign them as 2+1 jet events by merging jets, through an increased cut-off in the jet algorithm, until only the wanted jet multiplicity is obtained. Although this corresponds to a merging of partons, in order to recover the hardest emissions governed by the matrix element calculation, the exact kinematics of the matrix elements will not be obtained since the remaining jets are affected by higher order emissions.

These issues are illustrated in Fig. 13, where LUCLUS has been applied to events with parton shower emissions included. The reconstruction of $\sqrt{\hat{s}}$ is poor for events with more than $2+1$ jets. This is caused by extra jets from the initial state parton emission which results in a larger energy-momentum fraction being taken from the proton. Compared to the matrix element 2+1 jet case, energy will be taken from the spectator causing an increased invariant mass of the observable hard scattering system including such extra jets. The reconstruction improves considerably when these events are forced to give $2+1$ jets using a higher cut-off. Naturally, the best reconstruction is still obtained from the $2+1$ jet sample acquired with the original cut-off value of 4.0 GeV^2, i.e. when no jets from higher order QCD are present and can disturb the matrix element kinematics.

In the following study of the influence of higher order QCD effects we have chosen to keep the cut-off values of Table 2, i.e. values obtained from events based on matrix elements. This is appropriate given our main purpose of reconstructing jets for investigations of the QCD matrix elements. In this spirit we also keep only 2+1 jet events in order to suppress distortions from higher order hard emissions giving manifest deviations from the leading order QCD matrix elements. One should realise, however, that in doing so the remaining sample may have some biases, e.g. towards smaller $\hat{s}$, which must be controlled by Monte Carlo studies before physics conclusions can be safely made. With this caveat, the results are shown in Fig. 14 and 15 and should be compared with the ones in Fig. 11 and 12 which show the results when no parton showers were included. The broadening of the jets, caused by the parton showers, increases the danger of misassigning particles. This leads to a more pronounced tail towards too large $\hat{s}_{rec}$ as well as a shift in the reconstructed parton energy and polar angle and increased rms spread. The larger

effect noticed at higher W values is most likely dominated by an increased smearing due to the enhanced initial state radiation at the higher Q^2-values associated with larger W.

In the reconstruction of the first order matrix element subsystem and its jets, we have so far only considered events with an invariant mass of the resulting parton pairs $m_{ij} > 10\ GeV$. Below this cut-off the matrix elements increase without bound for $m_{ij} \to 0$ corresponding to the usual soft and collinear divergences for bremsstrahlung processes of massless quanta. The Sudakov form factor suppression built into the parton shower makes it more suitable to simulate emissions in this region. The class of events with no hard emission from the matrix element treatment, called q-events [8], but with possible softer radiation from parton showers also have to be considered. In general these events will be reconstructed as 1+1 jets and give a small $\sqrt{\hat{s}}$. However, with some fluctuations and given their large cross section compared to the genuine 2+1 jet cross section one cannot neglect them *a priori*. Table 4 shows the results of a generation of about 50 000 events. The QCD-Compton and the boson-gluon fusion events, denoted qg and $q\bar{q}$ respectively, are generated with QCD matrix elements for $m_{ij} > 10\ GeV$, whereas below that limit zeroth order events, denoted q-events, are generated. Parton showers are added to all types of events. As can be seen, q-events dominate the inclusive sample, but the fraction of events surviving the different criteria for $2+1$ jet events is very small. The criteria are *(i)* a reconstructed subsystem mass of at least 15 GeV, *(ii)* a reconstruction of exactly $2+1$ jets and *(iii)* the polar angles of the two current jets in $10° \leq \theta \leq 160°$. In event sample C the amount of surviving q-events can be neglected, but in event sample D the amount is comparable to the real 2+1 jet events.

event-sample:	C			D		
event-type:	q	qg	$q\bar{q}$	q	qg	$q\bar{q}$
initial sample	39608	5336	9414	32094	9032	7556
$\sqrt{\hat{s}} > 15\ GeV$	892	1890	4396	4492	5590	6421
$n_{jet} = 2+1$	787	1600	3787	3772	3518	3144
$10° < \theta_{jet} < 160°$	181	921	2000	1602	2154	2000
$\lvert\eta_{jet1} - \eta_{jet2}\rvert < 2$	181	920	1997	274	1673	1233

Table 4: *Reduction of event statistics due to the cuts applied to the three event types (described in the text) to obtain a clean $2+1$-jet sample.*

The difference in the rate of fake 2+1 jets in sample C and D is due to their different jet topologies. The current quark direction is, in QPM kinematics, in the backward direction for the large W in sample D, see eq. (3) and Fig. 2a, whereas it is in the forward direction for the lower W in sample C. The additional jets in q-events from parton showers mainly come from the initial state radiation along the incoming parton, i.e. they emerge along the proton in the forward direction. Such emissions, which is more abundant at large W, will therefore be better separated from the current jet in sample D and have a larger chance of being reconstructed as a jet in that case. This fake jet will thus have a large opening angle with respect to the scattered quark and consequently the invariant mass of this pair will be large and might survive the mass-cut of 15 GeV. However, the pseudo-rapidity distributions of the three event-classes turn out to have different shapes such that a cut in the difference between the jet pseudorapidities, $|\eta_{jet1} - \eta_{jet2}| < 2$, would greatly improve the ratio of qg- and $q\bar{q}$-events to q-events in sample D and would not

be harmful in kinematic regions (like C) where such an improvement is not necessary. Fig. 16 demonstrates this for events that passed the $2+1$ jet criteria of Table 4.

6 Conclusions

The ability to reconstruct the jet structures in HERA events has been investigated in a comparison of three commonly used and two new algorithms. In particular an attempt has been made to optimize the jet reconstruction to reproduce the properties of partons generated according to the first order α_s matrix elements. This was achieved by using a Monte Carlo simulation based solely on matrix elements from which the multiplicities and properties of the partons were well defined. In addition we required the invariant mass of any two partons to be larger than 10 GeV to ensure reasonably clear jet structures. After optimization of the algorithms the influence due to parton showers was studied.

The reconstruction of jets at HERA contains the well known problem of separating the spectator jet from the jets of the hard subsystem in the event. In order to perform a systematic study of this problem we have investigated how the reconstructed invariant mass of the hard subsystem, $\sqrt{\hat{s}}$, depends on the resolution parameter of the algorithm. To avoid events where the hard subsystem is partly merged with the spectator jet we have found that a cut on the reconstructed invariant mass of the hard subsystem of 15 GeV is efficient. The fraction of events which are removed by this cut depends on the cut-off value of the resolution parameter in the algorithm. Since both the quality of the $\sqrt{\hat{s}}$ reconstruction and the number of events surviving the mass cut are dependent on the resolution parameter their variation with the cut-off value can be used to compare different jet algorithms and to get a hint on appropriate choice of cut-off value for a specific algorithm.

In contrast to what might be intuitively expected we find that jets are in general better reconstructed at low than at high invariant masses of the hadronic final state (W). The explanation of this comes from the fact that the $\sqrt{\hat{s}}$-distributions at high and low W are similar but the event topologies are different. At high W the opening angle between the jets is in general larger than at low W and consequently the jet energies must be smaller at high W compared to low W in order to obtain similar $\sqrt{\hat{s}}$-values. The low energy of the jet closest to the spectator jet at high W makes a correct reconstruction difficult.

One of the investigated algorithms (JADE) gives consistently worse performance than the other four (LUCLUS,LUCELL,ARCLUS and K_t). In order to find one single cut-off value which gives reasonable reconstruction of $\sqrt{\hat{s}}$ at various W, the invariant mass m_{ij}, used as a distance parameter in JADE, has to be scaled with W^2. For high W this implies that the invariant mass will take values which are much larger than the typical invariant mass of the hard subsystem and thus the algorithm, although it works well in separating the hard subsystem from the spectator jet with this scaling variable, will not resolve the individual jets of the hard subsystem. We therefore propose to use JADE in two steps where the scale in the second step should be the reconstructed $\hat{s}$. In spite of this improvement the JADE algorithm reconstructs the parton four momenta less well than the other four algorithms which give similar results.

The effect of higher order parton emissions has been studied by adding parton showers to the partons generated by first order matrix elements. As expected the inclusion of

parton showers deteriorates the quality of the reconstruction somewhat but it does not change our overall conclusions from the matrix element study. Finally, we note that the background from QPM-type events where initial state radiation fakes an additional jet can not be neglected, especially not at high W. In this case we have found that a cut in the measured pseudorapidity-difference between the two reconstructed jets greatly improves the signal to background ratio.

Acknowledgements: We want to thank M. Nyberg for many valuable discussions. R. Nisius, L. Lönnblad, M. Seymour and Yu. Dokshitzer are acknowledged for useful comments to various parts of the content. The Lund group wants to thank the DESY directorate for its kind hospitality.

References

[1] G. Hansson et.al., Phys. Rev. Lett. 35 (1975) 1609
Ch. Berger et.al., PLUTO Collaboration, Phys. Lett. B78 (1978) 176
R. Brandelid et.al., TASSO Collaboration, Phys. Lett. B83 (1979) 261

[2] For previous studies of jet reconstruction at *ep* colliders, see e.g.
P. Burrows, G. Ingelman, E. Ros, Z. Phys. C39 (1988) 257
M. Fleischer et al., in [19] vol. 1, p. 303

[3] W. Bartel et al., JADE Collaboration, Z. Physik C33 (1986) 23
S. Bethke et al., JADE Collaboration, Phys. Lett. B213 (1988) 235

[4] T. Sjöstrand, Computer Phys. Comm. 28 (1983) 227

[5] T. Sjöstrand, Computer Phys. Comm. 39 (1986) 347, *ibid.* 43 (1987) 367

[6] L. Lönnblad, Z. Phys. C58 (1993) 471

[7] S. Catani, Yu.L. Dokshitzer, B.R. Webber, Phys. Lett. 285B (1992) 291

[8] G. Ingelman, LEPTO version 6.1, in [19] vol. 3, p. 1366

[9] M. Arneodo et al., (EMC), Z. Phys. C36 (1987) 527, Phys. Lett. 150B (1985) 458
G.T. Jones et al., (WA21), Z. Phys. C25 (1985) 121, *ibid.* C27 (1985) 43
D. Allasia et al., (WA25), Z. Phys. C24 (1984) 119

[10] T. Ahmed et al., H1 collaboration, Phys. Lett. B298 (1993) 469
I. Abt et al., H1 collaboration, DESY 93-137
M. Derrick et al., ZEUS collaboration, Z. Phys. C59 (1993) 231
M. Derrick et al., ZEUS collaboration, Phys. Lett. B306 (1993) 158

[11] S. Bethke, in proc. 'QCD – 20 years later', Aachen 1992, World Scientific 1993, p. 43
S. Bethke, J. Phys. G: Nucl. Part. Phys. 17 (1991) 1455

[12] S. Bethke, Z. Kunszt, D.E. Soper, W.J. Stirling, Nucl. Phys. B370 (1992) 310

[13] G. Arnison, et al., UA1 collaboration, Phys. Lett. B123 (1983) 115, *ibid.* B132 (1983) 223

[14] B. Andersson, G. Gustafson, L. Lönnblad, U. Pettersson, Z. Phys. C43 (1989) 621

[15] V. Hedberg, G. Ingelman, C. Jacobsson, L. Jönsson, in [19] vol. 1, p. 331

[16] S. Catani et al., Phys. Lett. B269 (1991) 432

[17] B.R Webber, J. Phys. G: Nucl. Part. Phys. 19 (1993) 1567

[18] D. Graudenz, N. Magnussen, in [19] vol. 1, p. 261

[19] Proceedings 'Physics at HERA', Eds. W. Buchmüller, G. Ingelman, DESY Hamburg 1992, vol. 1–3

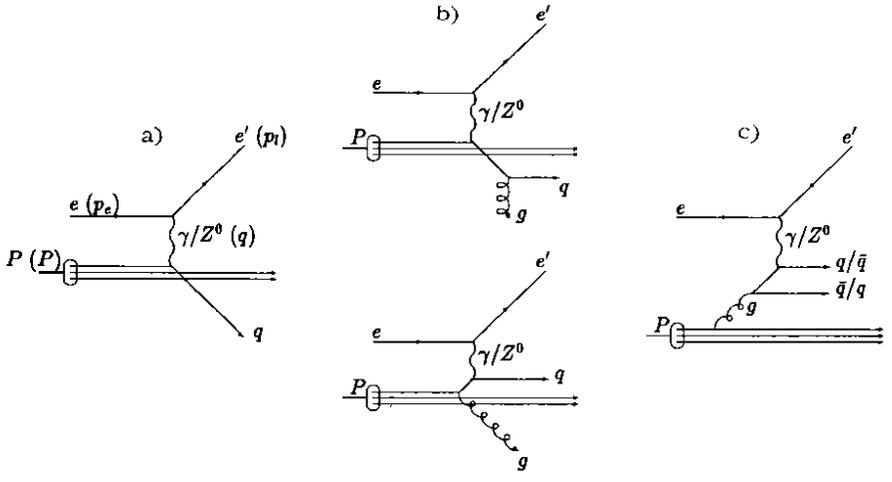


Figure 1: *Deep inelastic ep scattering at HERA with (a) the leading order QPM process and the first order QCD processes (b) QCD Compton and (c) boson-gluon fusion (BGF).*

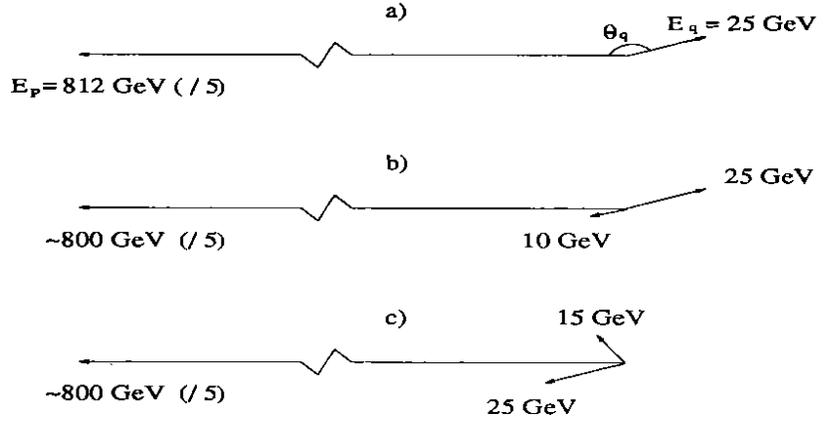


Figure 2:
a) The parton momentum-vectors in a QPM event with $x = 0.01$ and $y = 0.9$
b) A 'typical' boson-gluon fusion event with $x = 0.01$ and $y = 0.9$
c) A 'typical' boson-gluon fusion event with $x = 0.01$ and $y = 0.09$

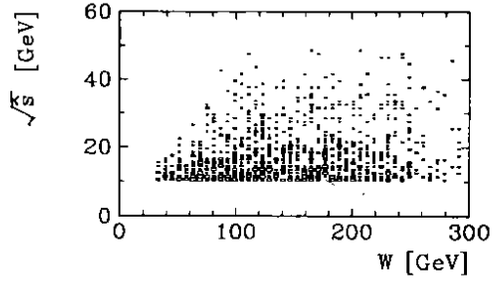


Figure 3: *Scatter plot of the invariant mass of the current-jets ($\sqrt{\hat{s}}$) and the total hadronic mass (W) for a sample of ME events.*

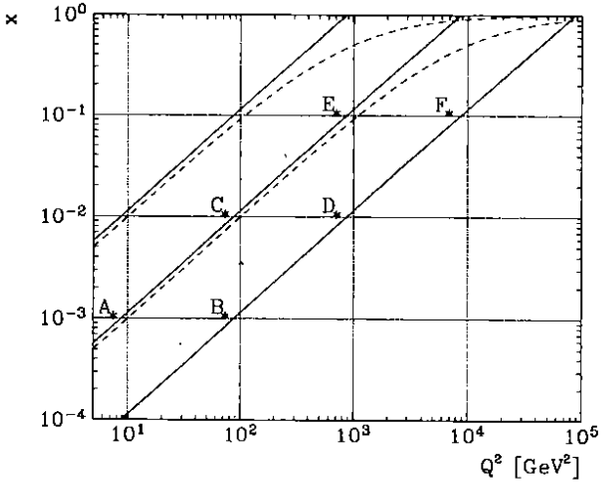


Figure 4: *The x- and Q^2-values of the generated six event-samples A-F. The diagonal full lines represent contsant y (= 0.01, 0.1, 1, top to bottom) and the dashed lines constant W^2 (= 10^3, 10^4 GeV^2).*

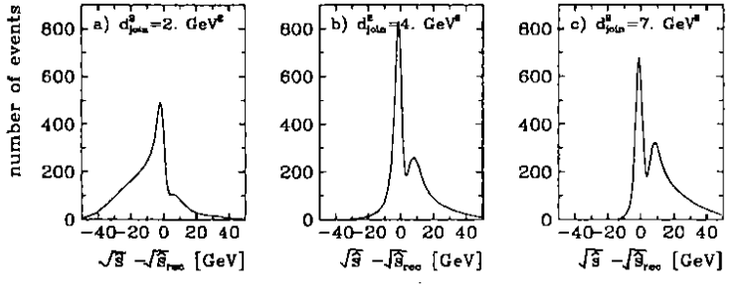


Figure 5: *The difference between the generated and reconstructed invariant mass ($\sqrt{\hat{s}}$) of the current jets, using* LUCLUS *with three values of the resolution parameter d^2_{join}.*

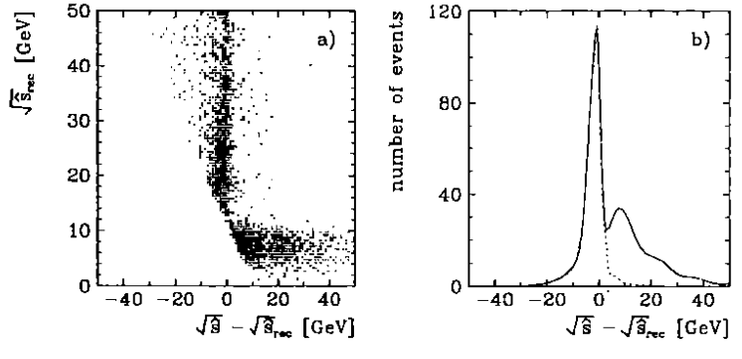


Figure 6: *a) Scatter plot of $\sqrt{\hat{s}}$ versus $\sqrt{\hat{s}} - \sqrt{\hat{s}_{rec}}$. b) The distribution of $\sqrt{\hat{s}} - \sqrt{\hat{s}_{rec}}$ before (full line) and after (dashed line) a mass-cut of $\sqrt{\hat{s}_{rec}} > 15$ GeV.*

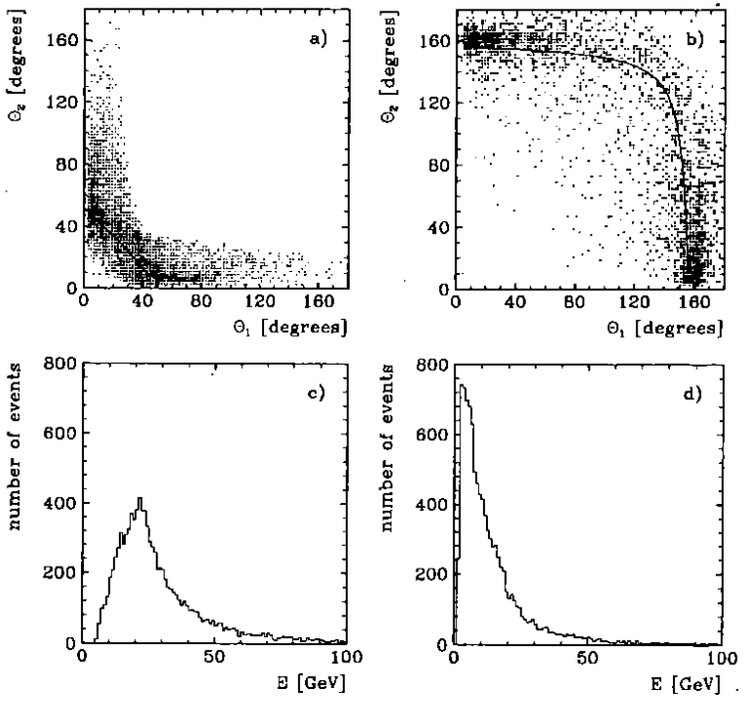


Figure 7: *a) The angles of the two partons of the first order QCD matrix element with respect to the proton direction for event-sample C (W^2 = 7920 GeV^2) and b) for event-sample D (W^2 = 79200 GeV^2). c) Energy of the parton closest to the spectator jet for event-sample C and d) for event-sample D.*

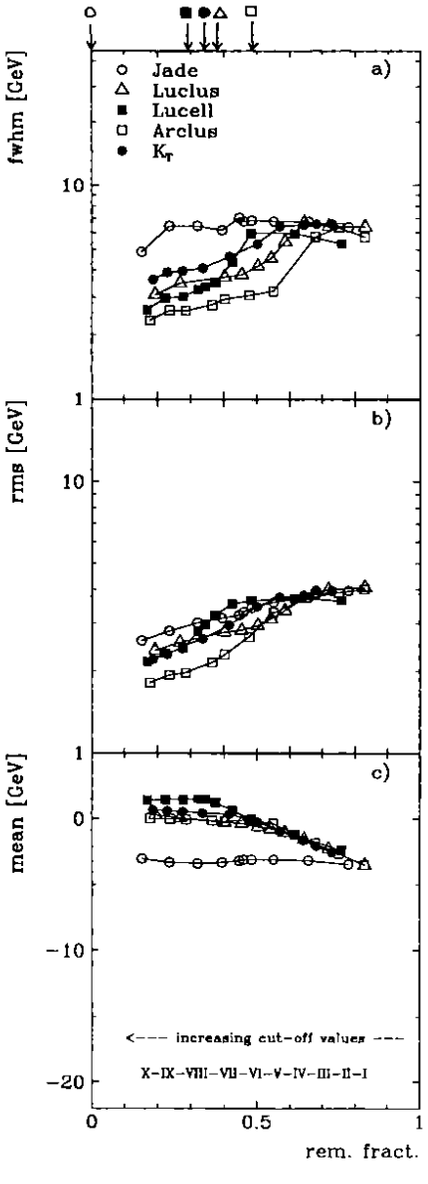


Figure 8:

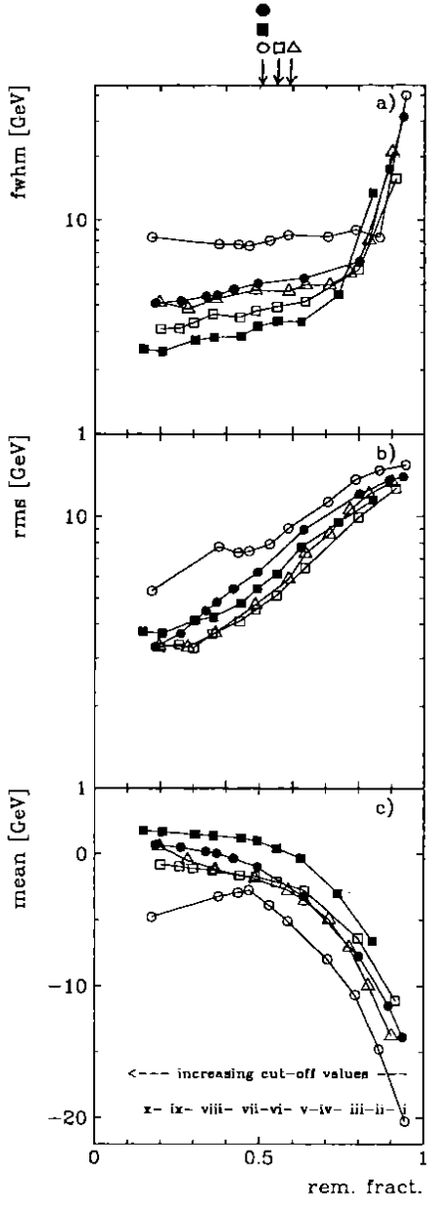


Figure 9:

Quality measures (y-axis) of reconstructed $\sqrt{\hat{s}}$ (invariant mass of hard scattering system) versus remaining fraction of the event-sample (x-axis) obtained with the indicated jet algorithms applied to event-samples C (fig.8) and D (fig.9). The points correspond to the cut-off values specified in Table 1 and the arrows on top indicate chosen cut-off values as described in the text.

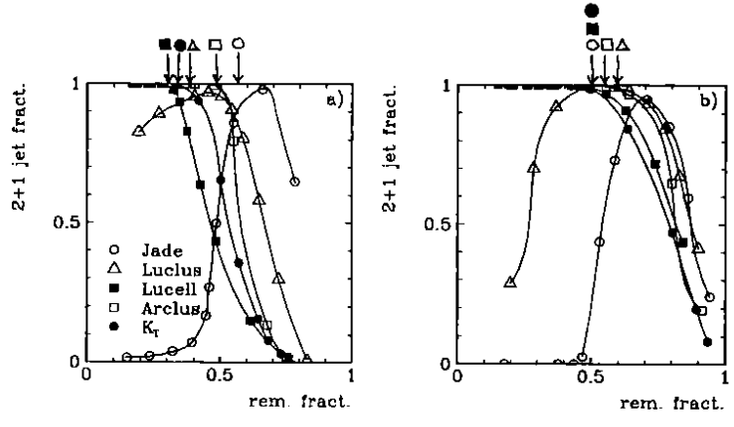
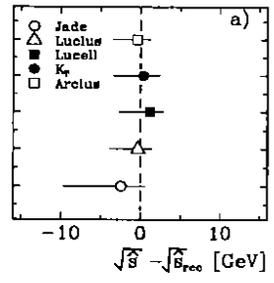
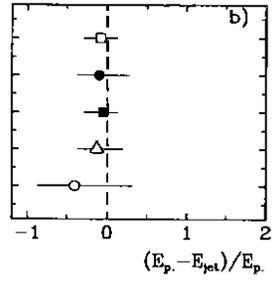
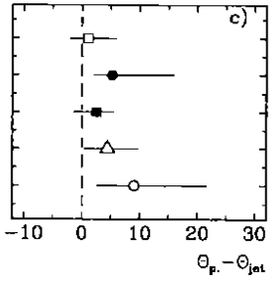


Figure 10: *The fraction $2+1$ jet results versus 'remaining fraction' in event samples a) C and b) D, i.e. the purity of the $2+1$-jet sample versus the efficiency of the selection. The points correspond to the cut-off values specified in Table 1, increasing from right to left. The chosen cut-off values, given in Table 2, are indicated by arrows on top of the figures.*







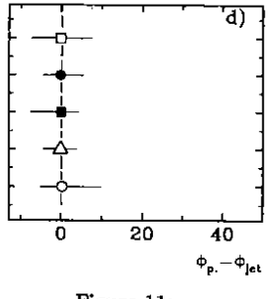
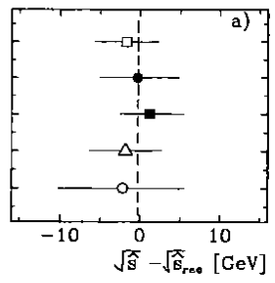
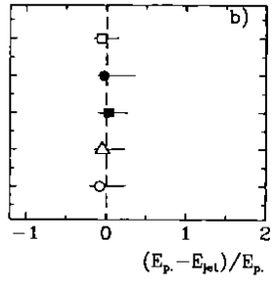
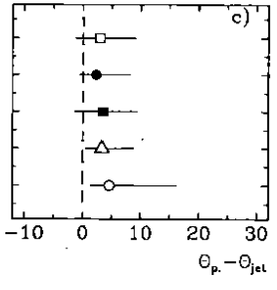


Figure 11:







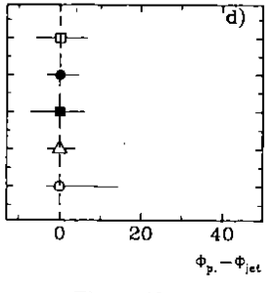


Figure 12:

Quality of reconstructed $\sqrt{\hat{s}}$, jet energy and angles in event-samples *C* (fig.11) and *D* (fig.12) shown by the *mean* value (data point), full-width-half-maximum (left error bar) and root-mean-square (right error bar) of the distributions for the given quantity (x-axis) obtained with the indicated jet algorithms.

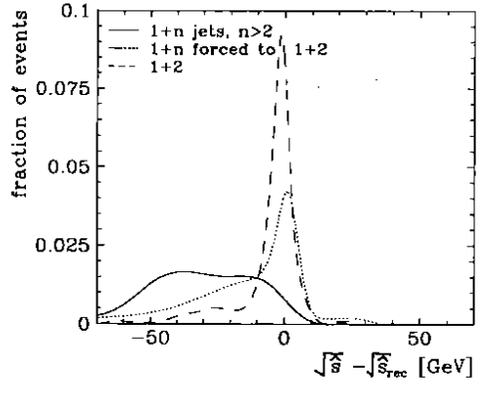


Figure 13: *Quality of reconstructed $\sqrt{\hat{s}}$ using* LUCLUS *on an event-sample (kinematic region D) with higher order QCD included through parton showers. The curves are for the specified jet multiplicities with cut-off 4 GeV^2 and an increased cut-off to avoid multijet reconstruction (dotted).*

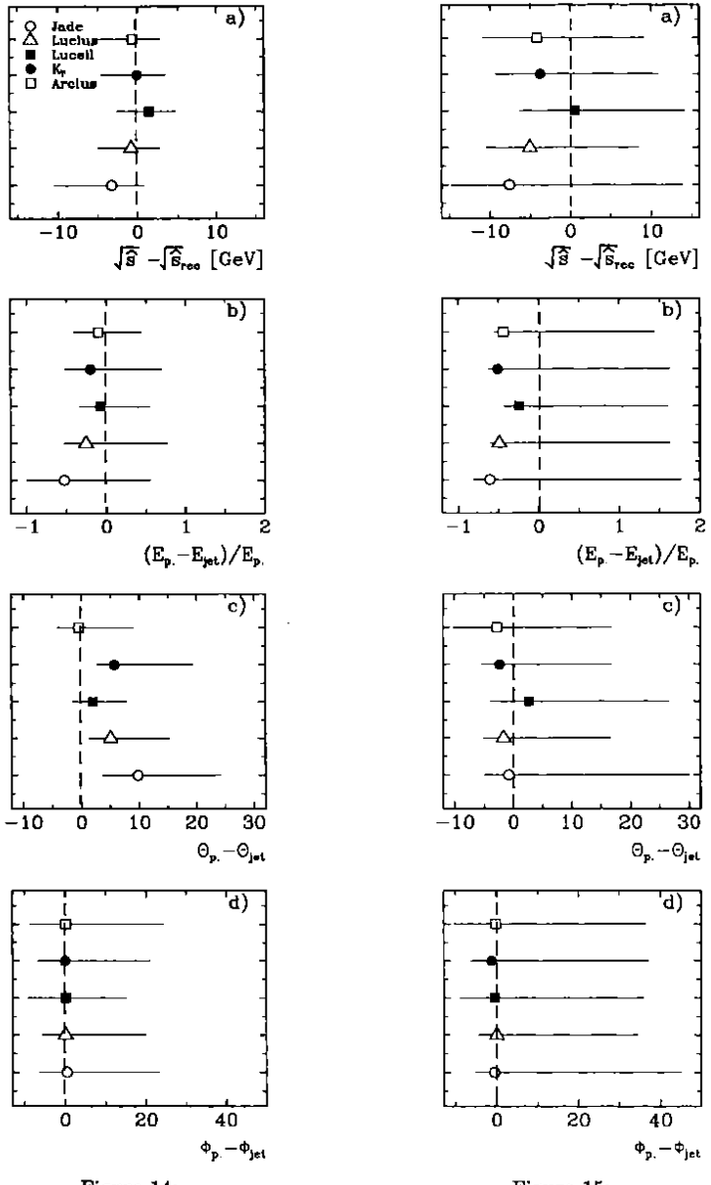


Figure 14:

Figure 15:

Quality of reconstructed $\sqrt{\hat{s}}$, jet energy and angles in event-samples C (fig.14) and D (fig.15) as in fig.11 and 12 but with higher order QCD parton emission included through parton showers.

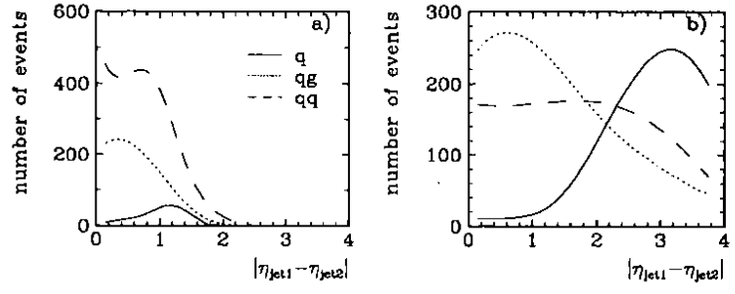


Figure 16: *Difference in pseudorapidity between the two jets reconstructed in q- qg- and $q\bar{q}$-events for event-samples a) C and b) D.*