Preprint typeset in JHEP style. - PAPER VERSION Cavendish–HEP–97/06

Better Jet Clustering Algorithms

Yu.L. Dokshitzer*
INFN Sezione di Milano, Via Celoria 16, 20133 Milan, Italy
E-mail: yuri@mi.infn.it

G.D. Leder, S. Moretti, B.R. Webber
Cavendish Laboratory, University of Cambridge,
Madingley Road, Cambridge CB3 0HE, U.K.
E-mail: leder@hep.phy.cam.ac.uk, *etc.*

Abstract: We investigate modifications to the $k_\perp$-clustering jet algorithm which preserve the advantages of the original Durham algorithm while reducing non-perturbative corrections and providing better resolution of jet substructure. We find that a simple change in the sequence of clustering (combining smaller-angle pairs first), together with the 'freezing' of soft resolved jets, has beneficial effects.

Keywords: QCD, jets, NLO computations, LEP physics, phenomenological models.

*Permanent address: St Petersburg Nuclear Physics Institute, Gatchina, St Petersburg 188350, Russian Federation.

1 Introduction

Jet clustering algorithms have become an indispensable tool for the analysis of hadronic final states in e^+e^- annihilation. More recently they have also started to be applied to other types of particle collisions. Clustering algorithms have permitted a wide range of important tests of QCD and will be of continuing value in more refined studies and in searches for new physics. It therefore remains worthwhile to think of ways in which the existing algorithms can be modified to improve their theoretical properties and phenomenological performance.

In the present paper we propose improvements which can be applied to any iterative clustering algorithm of the basic JADE type [1], although we mainly discuss them with reference to the so-called Durham or $k_\perp$ variant of that algorithm [2]. By the JADE type of algorithm we mean an exclusive iterative pairwise clustering scheme, in which jets are constructed out of primary objects, the latter being hadrons or calorimeter cells in the real experimental case and partons in the perturbative theoretical calculation. The term *exclusive* means that each primary object is assigned to a unique jet and each final state has a unique jet multiplicity, for a given value of the jet resolution parameter $y_{\rm cut}$.

Algorithms of the JADE type have two basic components:

1) a test variable y_{ij}, and

2) a combination procedure.

The test variable is used to decide whether the objects i and j should be combined, according to whether $y_{ij} < y_{\rm cut}$. It is also used to choose which objects to consider next for combination, namely the pair with the smallest value of y_{ij}. In the original JADE algorithm, $y_{ij} = M_{ij}^2/Q^2$ where Q is the hard scale (i.e. the centre-of-mass or visible energy in e^+e^- annihilation) and

$$M_{ij}^2 = 2E_iE_j(1-\cos\theta_{ij}) \;, \tag{1.1}$$

which is essentially the invariant mass-squared of the pair. The combination procedure specifies the properties of the new object formed by combining i and j, for example that its four-momentum should be simply the sum $p_{ij} = p_i + p_j$ (the so-called E scheme, which we shall adopt in the present paper). The clustering procedure is repeated until no objects can be combined further (all $y_{ij} > y_{\rm cut}$), at which stage all objects are defined as jets.

A crucial point is that the *identical* algorithm should be applicable to real experimental data and to the partons that appear in the perturbative calculation. It follows that an essential feature of the algorithm must be *infrared safety*, i.e. insensitivity to the emission of arbitrarily soft and/or collinear particles. Otherwise, the presence of massless partons in the perturbative calculation would lead to divergent results. Beyond this fundamental feature, the ideal algorithm should lead to a close correspondence between the theoretical (partonic) and actual (hadronic) jet characteristics and multiplicities after clustering, over the widest possible range of values of $y_{\rm cut}$. That is, it should have the smallest possible *hadronization corrections*, this being the name given to empirical adjustments that are applied to the theoretical predictions before comparing them with experiment. One would also like the algorithm to have good theoretical properties, in particular resummability of large terms to all orders in perturbation theory and a reduced renormalization-scale dependence in fixed order.

While the original JADE algorithm satisfies the basic criteria of infrared safety, its theoretical properties turn out to be surprisingly complicated. This is because the invariant mass is not the most relevant variable for the evolution of QCD jets, and consequently the multijet phase space has a non-factorizing structure when expressed in term of the JADE test variable [3, 4]. As a result the range of theoretical predictions that can be compared with experiment using this algorithm is limited, and the hadronization corrections are not optimally small.

A basic weakness of the JADE algorithm is illustrated by the way it deals with the parton-level 'seagull diagram' (Fig. 1), in the phase-space region where the two gluons g_3 and g_4 are soft and almost collinear with the quark and antiquark [3]. We then have

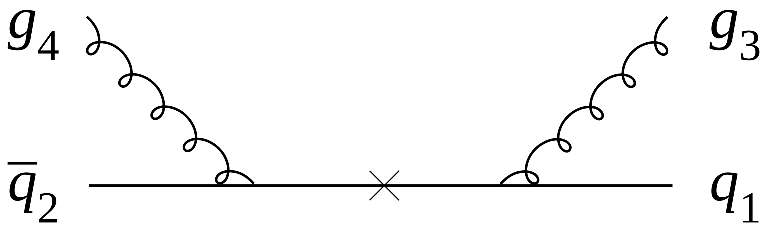

Figure 1: The 'seagull diagram'.

$$y_{13} \sim x_3\theta_{13}^2 \ , \quad y_{24} \sim x_4\theta_{24}^2 \ , \quad y_{34} \sim x_3 x_4 \ , \tag{1.2}$$

where $x_i = 2E_i/Q$. Since all these are of the same order in small quantities, there is an important subregion in which y_{34} is the smallest. In that subregion the two gluons will be combined first, making a 'phantom' gluon jet with a resultant momentum at a large angle to the quark and antiquark, where there are in fact no particles.

The shortcomings of the JADE algorithm are greatly alleviated by the Durham modification [2, 5, 6], which consists of replacing M_{ij}^2 in the definition of the test variable by $k_{\perp ij}^2$, where

$$k_{\perp ij}^2 = 2\min\{E_i, E_j\}^2(1-\cos\theta_{ij}) \ , \tag{1.3}$$

which is essentially the relative transverse momentum-squared of i and j. This choice of test variable reflects the more fundamental role of the transverse momentum in setting the scale of jet evolution, as the argument of the running coupling, and in defining the boundary between perturbative and non-perturbative physics. In the treatment of the seagull diagram, Eqs. (1.2) become

$$y_{13} \sim x_3^2\theta_{13}^2 \ , \quad y_{24} \sim x_4^2\theta_{24}^2 \ , \quad y_{34} \sim \min\{x_3^2, x_4^2\} \ , \tag{1.4}$$

and so the two gluons cannot be combined first in this region of phase space. As a result, one finds that the phase space now has simple factorization properties and the available range of predictions is increased. In particular, the terms in the perturbation series that involve leading and next-to-leading powers of $\ln y_{\rm cut}$ can be identified and resummed to all orders [5, 6], thereby improving the reliability of the predictions at small values of $y_{\rm cut}$. Furthermore, hadronization corrections, estimated according to the best available Monte Carlo models [7, 8], are found to be reduced [9, 10].

The good features of the Durham algorithm at low $y_{\rm cut}$ lead one to hope that it should be possible to probe the interface between perturbative and non-perturbative QCD by studying jet properties as a function of $y_{\rm cut}$ in the relevant region, say $y_{\rm cut} < 10^{-4}$ at $Q = M_Z$, corresponding to $k_\perp < 1$ GeV. One would like, for example, to see whether the low-energy behaviour of the effective strong coupling could be studied in this way [11, 12]. However, that hope is dashed by the observation [13, 14] that the hadronization corrections, although reduced, are still substantial at $y_{\rm cut} \sim 10^{-3}$, corresponding to $k_\perp \sim 3$ GeV, well above the values at which non-perturbative effects would normally be expected.

In the following Section we discuss the defects of the Durham algorithm which lead to this situation, and the steps that can be taken to remedy them. We define modified algorithms and study the associated non-perturbative effects that would be expected in simple hadronization models. In Sect. 3 we compute the jet fractions for these new algorithms to next-to-leading order in perturbation theory and study their properties. Sect. 4 is devoted to studies of the mean jet multiplicity, first in resummed perturbation theory and then using the HERWIG event generator [8]. Comparisons show clearly the benefits of the new algorithms at small values of $y_{\rm cut}$, beyond the reach of fixed-order perturbation theory. Finally, in Sect. 5 we briefly summarize our results and conclusions.

2 Jet algorithms: defects and cures

2.1 'Junk-jet' formation

The basic problem that we have to address is as follows. With decreasing $y_{\rm cut}$, all jet algorithms inevitably start to search for jets amongst the hadrons with low transverse momenta and to form spurious 'junk-jets' from them, pretending that these are legitimate, resolvable gluon jets. As we shall see below, the original JADE algorithm, with its test variable related to invariant mass, accumulates the first junk-jet at $y_{\rm cut} \sim \lambda/Q$, where $\lambda \sim 0.5$ GeV is a soft scale set by the mean non-perturbative transverse momentum. The Durham algorithm eventually falls into the same trap, although at lower $y_{\rm cut}$.

One might expect a $k_\perp$-based algorithm to avoid resolving junk-jets as long as $y_{\rm cut}$ is kept above λ^2/Q^2. The Durham algorithm, however, does not respect such a natural expectation. According to the clustering procedure adopted, one usually starts from the softest particle in a jet (call it hadron #1) and merges it with the one nearest in *angle*, to minimize the relative $k_\perp$. Thus hadron #1 gets clustered not with the leading hadron in the jet but, typically, with the softest among the hadrons which happen to lie on the same side of the jet axis in the transverse plane. Half of them do, on average. The typical transverse momentum of the largest junk-jet, and hence the $y_{\rm cut}$ value at which it is resolved, is therefore enhanced by a factor proportional to $N(Q)$, the soft hadron multiplicity in a jet at scale Q, which is proportional to $\ln^p Q$, with p depending on the model of soft physics. As a result, junk-jets start to appear at $y_{\rm cut} \propto \ln^{2p} Q(\lambda/Q)^2$. Taking account of fluctuations, such a misinterpretation proves to be a substantial effect.

To understand these features in more detail, we can examine the action of the algorithms on simple models of the non-perturbative hadronization process. Consider for example the simplest possible string or 'tube' model, in which the hadronization of a back-to-back two-parton system yields a distribution of hadrons which is uniform in rapidity η and strongly damped in transverse momentum p_t, both variables being defined with respect to the original parton directions. If the particle density in (η, p_{tx}, p_{ty})-space is given by $\rho(p_t)$, we have (neglecting hadron masses)

$$Q = E_{\rm cm} = \int d\eta d^2 p_t \rho(p_t) p_t \cosh\eta = 2\lambda \sinh Y \tag{2.1}$$

where $Y \sim \ln(Q/\lambda)$ is the maximum value of $|\eta|$, and

$$\lambda \equiv \int d^2 p_t \rho(p_t) p_t = N \langle p_t \rangle / 2Y \tag{2.2}$$

where the total hadron multiplicity in this model is $N \sim (2\lambda/\langle p_t \rangle) \ln(Q/\lambda)$.

Ideally, we would like hadronization to affect the jet structure of events as little as possible, and therefore we would prefer the tube to remain classified as a two-jet final state down to the smallest possible values of the resolution $y_{\rm cut}$. Nevertheless any algorithm will eventually resolve a third (junk-) jet at some value $y_{\rm cut} = y_3$; the algorithm should be designed to make this as small as possible.

The action of various algorithms in resolving a third jet inside the tube is illustrated in Fig. 2. In the JADE algorithm, the resolution measure is related to the invariant mass, and therefore jets tend to be resolved as slices of the tube. In particular, an extra third jet is resolved when $y_{\rm cut} \sim M_{\rm jet}^2/Q^2$, where $M_{\rm jet}$ is the mass of the jet in either hemisphere, given by

$$M_{\rm jet}^2 = E_{\rm jet}^2 - p_{\rm jet}^2 = 2\lambda^2 \cosh Y \sim \lambda Q \; , \tag{2.3}$$

so that

$$\langle y_3 \rangle^J \sim \frac{\lambda}{Q} \; . \tag{2.4}$$

This is shown by the dashed curve in Fig. 3, to be compared with data from a Monte Carlo simulation of the tube model (squares). For this simulation, we simply generated a number N of massless four-momenta with an exponential transverse momentum distribution and a uniform rapidity distribution in the interval $-Y < \eta < Y$, Y being given by Eq. (2.1) and N by Eq. (2.2). As illustrative values, we have taken $\lambda = 0.5$ GeV and $\langle p_t \rangle = 0.3$ GeV.

In the original Durham (D) algorithm, to resolve a third jet one has to subdivide one of the jets into two parts with relative transverse momentum P_t, where

$$P_t^2 \sim y_{\rm cut} Q^2 \; . \tag{2.5}$$

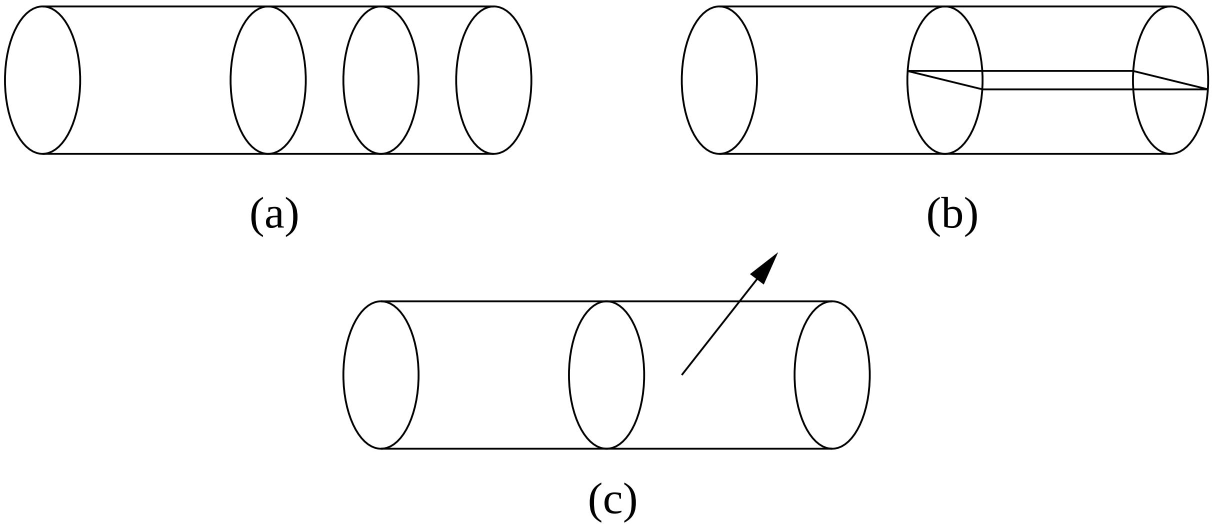

Figure 2: Resolving a third jet in the final state of the tube model: (a) JADE, (b) Durham, (c) angular-ordered Durham algorithm.

The largest value of $y_{\rm cut}$ at which this can be achieved occurs when one half of the tube is divided axially into two half-cylinders, as illustrated in Fig. 2(b), giving

$$P_t \sim \int_0^Y \int d^2p_t |p_{tx}| \rho(p_t) \sim \frac{2}{\pi}\lambda Y \tag{2.6}$$

and hence

$$\langle y_3 \rangle^D \sim \left(\frac{2\lambda \ln(Q/\lambda)}{\pi Q}\right)^2 . \tag{2.7}$$

This is shown by the solid curve in Fig. 3, which agrees well enough with the Monte Carlo tube model data (circles). We see the expected great improvement relative to the JADE algorithm, due to the power-suppression factor of $1/Q^2$ rather than $1/Q$. However, the presence of the log-squared enhancement factor means that the coefficient of $1/Q^2$ is far larger than $\mathcal{O}(\lambda^2)$, the order of magnitude that one might hope to be achievable with an optimal jet algorithm.

An alternative way of estimating non-perturbative contributions to $\langle y_3 \rangle$ has been proposed in Ref. [15]. At lowest order in perturbation theory, for any infrared-safe jet algorithm, this quantity is proportional to $\alpha_{\rm S}$. In higher orders it is given by a power series in $\alpha_{\rm S}(Q)$, where the argument of the coupling is set by the only available hard-scattering scale $Q = E_{\rm cm}$. Now although the perturbative predictions may be expressed in terms of $\alpha_{\rm S}(Q)$, one cannot avoid sensitivity to the region of low momenta $k \ll Q$ inside integrals that contribute to those predictions. This sensitivity makes the perturbation series in $\alpha_{\rm S}(Q)$ strongly divergent at high orders, leading to power-behaved ambiguities.

In the 'dispersive approach' of Ref. [15] these so-called renormalon ambiguities are resolved by assuming the existence of a universal low-energy effective strong coupling $\alpha_{\rm eff}(k)$ with sensible analyticity properties. This leads to genuine power-behaved non-perturbative contributions, which may be parametrized in terms of moments of $\delta\alpha_{\rm eff}(k)$,

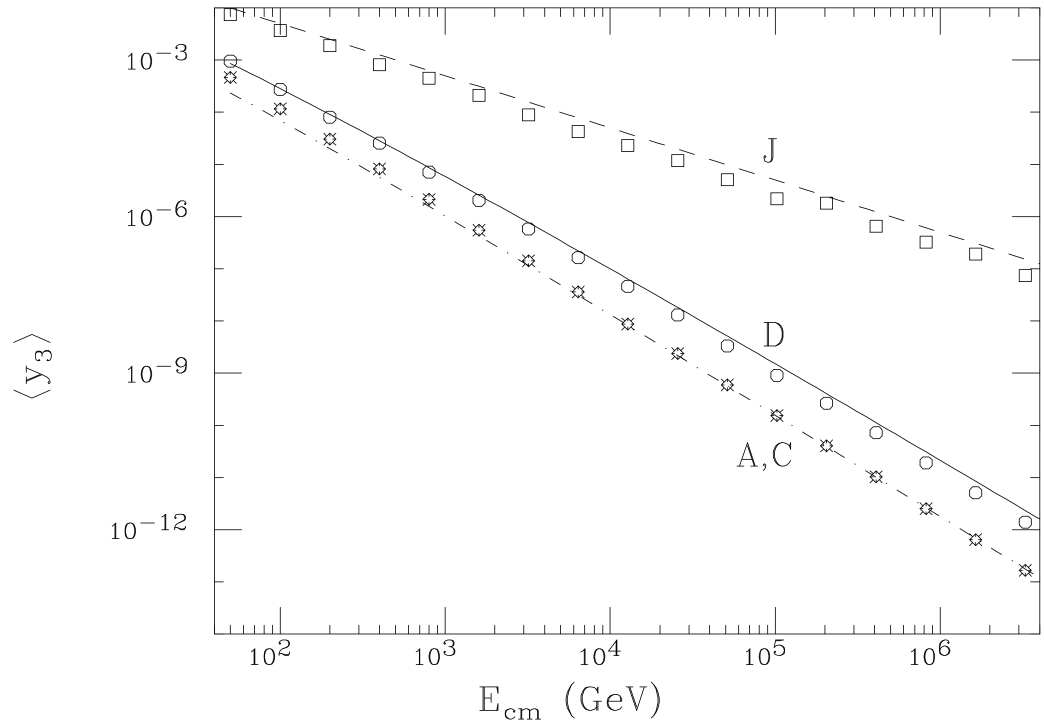

Figure 3: Mean values of the three-jet resolution in the 'tube' hadronization model, for the J (square symbols), D (circles) and A, C (stars) jet clustering schemes. The corresponding curves show the approximate formulae discussed in the text.

the discrepancy between the effective coupling and its perturbative expansion in terms of $\alpha_S(Q)$ (up to the order included in the perturbative contribution).

As was observed in Ref. [15], the expectation based on the dispersive approach is that $\langle y_3 \rangle$ should have a leading non-perturbative contribution proportional to $1/Q$ in the case of the JADE algorithm and proportional to $\ln Q/Q^2$ for the Durham algorithm. Thus the predicted power behaviour agrees in each case with that expected from the simple tube model, although the logarithmic enhancement of the non-perturbative contribution for the Durham algorithm is not as large as in the tube model.

In Sect. 4 we shall present the results of studies of hadronization effects in a more realistic model.

2.2 Angular-ordered Durham (A) algorithm

We now show that a simple modification of the Durham algorithm suffices to delay the onset of junk-jet formation, which results mainly from a non-optimal sequence of clustering, rather than from a poor definition of the test variable (as was the case for the JADE algorithm). The key to alleviating the problem is to notice that the most general definition of a clustering algorithm in fact involves *three components* rather than the two mentioned above, viz.

0) an ordering variable v_{ij},

1) a test variable y_{ij}, and

2) a combination procedure.

In other words, we now distinguish between the variable v_{ij}, used to decide which pair of objects to test first, and the variable y_{ij} to be compared with the resolution parameter $y_{\rm cut}$. The algorithm then operates as follows. One considers first the pair of objects (ij) with the smallest value of v_{ij}. If $y_{ij} < y_{\rm cut}$, they are combined. Otherwise the pair with the next smallest value of v_{ij} is considered, and so on until either a $y_{ij} < y_{\rm cut}$ is found or, if not, clustering has finished and all remaining objects are defined as jets.

From this viewpoint we see that JADE-type ($v_{ij} \equiv y_{ij}$) algorithms, including the Durham variant, are likely to produce the *largest* number of jets for a given definition of the test variable and combination procedure. Suppose that at some stage of clustering the pair (ij) with the smallest value of y_{ij} does not have the smallest value of v_{ij}. We can assume that $y_{ij} < y_{\rm cut}$, since otherwise clustering has already finished. If the pair (kl) with the smallest value of v_{kl} has $y_{kl} < y_{\rm cut}$, they will be combined. Otherwise, the pair with the next smallest value will be considered, and so on until v_{ij} is reached, whereupon i and j will be combined. Thus at least the (ij) clustering performed by the JADE-type algorithm is available at any stage. However, we cannot make a rigorous argument that the JADE-type jet multiplicity is an upper bound on that for $v_{ij} \neq y_{ij}$ algorithms, because the sequence of clustering will be different in general.

Once one separates the functions of the ordering and test variables, it is natural to adopt the *relative angle* for ordering purposes, thus taking into account the angular-ordering property of QCD [16]. Recall that angular ordering corresponds to the fact that a soft gluon emitted at some angle to the jet axis cannot resolve the colours of jet constituents at smaller angles. For a correct correspondence with the colour structure of the jet, one should therefore combine the constituents at smaller angles first. We shall refer to this choice of ordering variable, combined with the Durham test variable and combination procedure, as the *angular-ordered Durham* (A) algorithm. One starts with a table of the energies E_i of primary objects and their relative angles θ_{ij}, or equivalently the quantities

$$v_{ij} = 2(1 - \cos\theta_{ij}) \ , \tag{2.8}$$

all defined in the e^+e^- c.m. frame. The A algorithm is then as follows:

Step 1. Select the pair of objects (ij) with the minimal value of the ordering variable, v_{ij}.

Step 2. Inspect the value of the test variable,

$$y_{ij} = \min\{E_i, E_j\}^2 \, v_{ij} \ .$$

- If $y_{ij} < y_{\rm cut}$, then update the table by deleting i and j, introducing a new particle (ij) with 4-momentum $p_{ij} = p_i + p_j$, and recomputing the relevant values of the ordering variable; then go to Step 1.

- If $y_{ij} \geq y_{\rm cut}$, then consider the pair (if any) with the next smallest value of the ordering variable and repeat Step 2. If no such pair exists, then clustering is finished and all remaining objects are jets.

The earlier discussion of the logarithmic enhancement of non-perturbative effects in the Durham algorithm showed that it arises because spurious junk-jets can be formed by combining soft, large-angle (low-rapidity) hadronization products. The A algorithm avoids this by choosing small-angle combinations preferentially. Then the two real jets in the final state of the tube model are built up from the high-rapidity ends of the tube, and the first junk-jet to be resolved will be the wide-angle hadron or cluster with the largest transverse momentum,

$$\langle y_3 \rangle^A \sim \frac{\left\langle p_{t,\max}^2 \right\rangle}{Q^2}\,, \tag{2.9}$$

as represented schematically in Fig. 2(c). The relationship between $\left\langle p_{t,\max}^2 \right\rangle$ and $\langle p_t \rangle^2$ depends on the multiplicity N and the form of the p_t-distribution. For an exponential p_t-distribution we have

$$\left\langle p_{t,\max}^2 \right\rangle \sim \left(\langle p_t \rangle \ln N\right)^2 \tag{2.10}$$

and therefore

$$\langle y_3 \rangle^A \sim \left(\frac{\lambda}{Q} \ln\ln(Q/\lambda)\right)^2\,. \tag{2.11}$$

This is shown by the dot-dashed curve in Fig. 3, which again agrees quite well with the Monte Carlo data (stars).

2.3 Mis-clustering

It is difficult to see how the value of $\langle y_3 \rangle$ could be reduced much further by refinement of the A algorithm, since fluctuations in p_t will always lead to a multiplicity-dependent enhancement of $\langle p_{t,\max} \rangle$ relative to the intrinsic non-perturbative scale set by $\langle p_t \rangle_{\rm NP}$. There is however a remaining weakness of the algorithm that, while having no perceptible effect on $\langle y_3 \rangle$, can be expected to show up more in studies of multi-jet rates and the internal structure of jets. It has to do with the wrong assignment of soft wide-angle radiation to *already resolved* jets.

Consider the parton-level application of the algorithm to the basic splitting $q_1 \to q_1 + g_2$ accompanied by a soft large-angle gluon g_3 (Fig. 4), such that

$$E_1 \gg E_2 \gg E_3\,, \qquad \theta_{12} \ll \theta_{13} \approx \theta_{23}\,.$$

According to the angular ordering property of QCD, the soft gluon g_3 is radiated *coherently* and the corresponding radiation intensity is proportional to the total colour

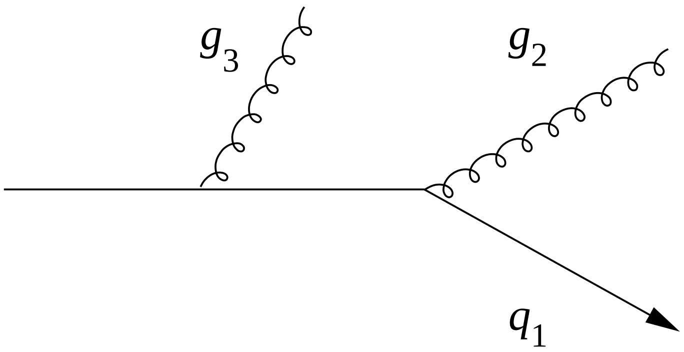

Figure 4: Parton branching with soft, large-angle gluon emission.

charge of the relatively narrow system $q_1 + g_2$, which is just that of the initial quark. However, in *half* of the events the A (or D) algorithm will erroneously assign g_3 to the gluon jet g_2 because the latter happens to lie a little closer in angle to g_3 than the quark.

At first sight, this does not seem to do any harm. The three partons are going to be identified as a 1-, 2-, or 3-jet system depending on the chosen value of y_{cut}, and the corresponding probabilities of the 1-, 2-, 3-jet configurations remain unaffected by mis-clustering. This is true as long as one is concerned only with the jet fractions and not with the internal structure of the jets. However, after having resolved this system as *two* jets at some resolution y_1 one may attempt to choose another yet smaller $y_2 \ll y_1$ and to study the internal sub-jet structure by examining the history of merging. In this case the results will be misleading because one will find that the gluon sub-jet is artificially overpopulated. The same mis-clustering will occur when all three partons in our example are gluons, although in this case the practical damage may be less pronounced.

2.4 Cambridge (C) algorithm

The modification we propose in order to reduce the mis-clustering effect consists of 'freezing' the *softer* of two resolved objects to prevent it from attracting any extra partners from among the remaining objects with larger emission angles. The corresponding *Cambridge* (C) algorithm is thus defined as follows. As before, one starts with a table of the energies E_i of primary objects and their relative angles as given by the ordering variable $v_{ij} = 2(1 - \cos\theta_{ij})$.

Step 0. If only one object remains in the table, then store this as a jet and stop.
Step 1. Otherwise, select the pair of objects (ij) having the minimal value of the ordering variable, v_{ij}. Order the pair such that $E_i \leq E_j$.
Step 2. Inspect the value of the test variable,

$$y_{ij} = E_i^2\, v_{ij} \ .$$

- If $y_{ij} < y_{\text{cut}}$, then update the table by deleting i and j, introducing a new particle (ij) with 4-momentum $p_{ij} = p_i + p_j$, and recomputing the relevant values of the ordering variable.

- If $y_{ij} \geq y_{\text{cut}}$, then store i as a jet and delete it from the table.

Step 3. Go to Step 0.

This procedure suffices to prevent mis-clustering in the dominant region of phase space where, in the above example, the gluon g_2 is softer than the quark q_1. It will still cluster the partons incorrectly in the subleading configuration where the gluon (call it g_1 now since $E_1 > E_2$) is accidentally harder than the quark (q_2). After having resolved them, it will erroneously assign the soft gluon g_3 to g_1 (the more energetic one), spoiling the association between the clustering and the colour factor. A similar mistake will occur in the case of a resolved $g \to q_1\bar{q}_2$ splitting plus a soft wide-angle gluon g_3: the algorithm will inevitably cluster g_3 with the quark or the antiquark, which is incorrect because the source of g_3 is the coherent octet colour of the pair, not a triplet. However, the fraction of events of these types should be small, and it is hard to see how such occasional mistakes can be avoided without distinguishing between different species of partons.

As already mentioned, the additional 'soft freezing' step in the Cambridge algorithm does not affect the quantity $\langle y_3 \rangle$ significantly. This is shown by Fig. 3, where the points for the A and C algorithms coincide within the plotting resolution.

We do, however, expect the internal jet structure resulting from the A and C algorithms to be different. The aim of the freezing is to prevent soft jets from acquiring particles that do not belong to them. A good diagnostic quantity for this is the mean number of particles in the first junk-jet to be resolved in the tube model, i.e. the third (softest) jet when $y_{\text{cut}} = y_3$. In the optimal algorithm this quantity, $\langle n_3 \rangle$, should be as small as possible.

Our results on $\langle n_3 \rangle$ for the various algorithms are presented in Fig. 5. As could be guessed from Fig. 2, the values in the JADE and original Durham algorithms are quite similar, both being somewhat below one quarter of the total multiplicity N. As shown by the curve, $\langle n_3 \rangle^D$ is in fact about $0.22N$, with $\langle n_3 \rangle^J$ rising slightly *less* rapidly. Therefore in this respect the unmodified Durham algorithm performs slightly *worse* than JADE asymptotically. The angular-ordered Durham algorithm performs a good deal better, but the population of the junk-jet still grows steadily with increasing energy. Introducing the soft freezing step as well (C algorithm) kills the growth of the spurious jet almost completely, and the value of $\langle n_3 \rangle^C$ remains near 2 at all energies.

Taking Figs. 3 and 5 together, we see that the Cambridge algorithm performs best in limiting both the amount of junk-jet formation and the growth of junk-jets once they have formed.

In real hard processes like e^+e^- annihilation, jet algorithms have to operate on final states containing both genuine hard jets and spurious junk-jets, and should ideally suppress the latter as far as possible while remaining sensitive to the former. We shall study these issues in more depth in later Sections, but a few qualitative observations can be made here. First, we would expect the angular ordering in the A and C algorithms to reduce the number of jets at a given value of y_{cut}, relative to the D algorithm, because, as remarked earlier, separating the ordering and test variables generally increases the

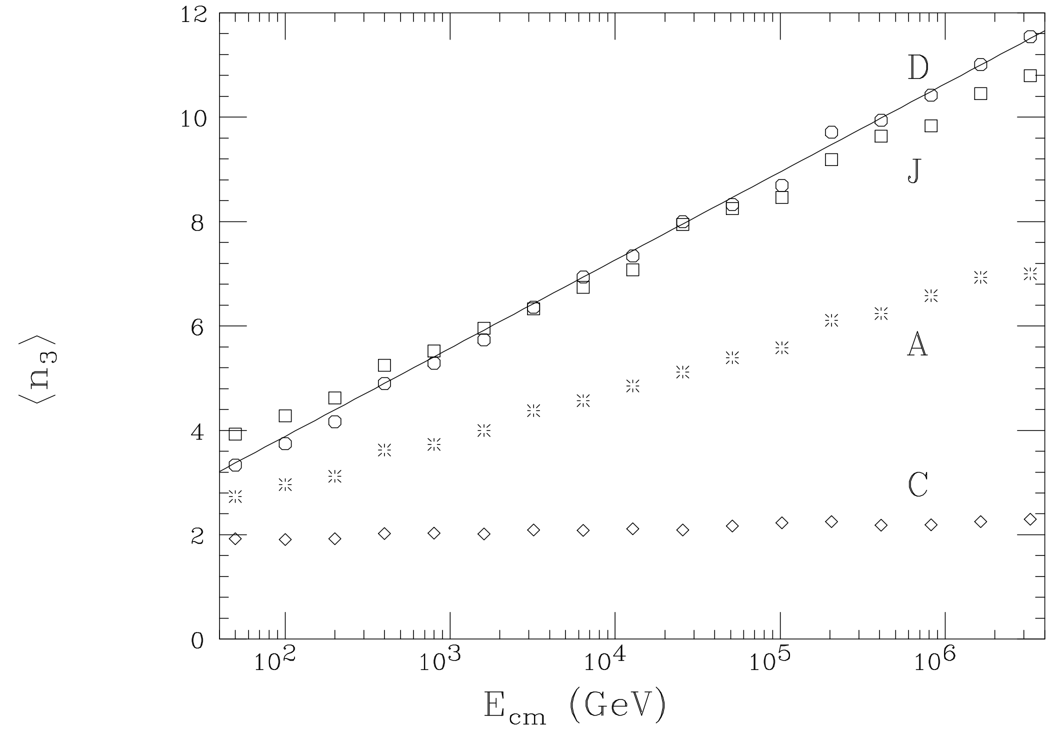

Figure 5: Mean number of particles in the third jet at $y_{\rm cut} = y_3$ in the 'tube' hadronization model, for the J (square symbols), D (circles), A (stars) and C (diamonds) jet clustering schemes. The curve shows 0.22 times the total number of particles.

number of combinations attempted. This effect should become more pronounced as the number of objects increases, i.e. with increasing energy and/or decreasing $y_{\rm cut}$. On the other hand, the soft freezing step in the C algorithm tends to enhance the number of jets, relative to the A algorithm, by forbidding combination with frozen jets. Again this difference should be more pronounced for a larger number of objects.

It should be emphasised that neither the angular-ordering nor the soft freezing modifications affect the good properties of the Durham algorithm with respect to the resummation of large logarithmic terms at small values of $y_{\rm cut}$. The leading and next-to-leading logarithms of $y_{\rm cut}$ come from kinematic regions where the sequence of clustering is not affected by the extra steps in the modified algorithms. Thus the only adjustment to resummed predictions will be in the subleading 'remainder' terms, as will be discussed in more detail in Sec. 4.

In terms of computing, the modified algorithms are slightly less convenient than the original JADE-type (including Durham) ones, because of the new distinction between the ordering and test variables. In a JADE-type algorithm, the sequence of clustering is independent of the value of $y_{\rm cut}$. Therefore the clustering can be done once and for all, the sequence can be stored, and all questions about jet multiplicity and substructure can be answered subsequently without much further computation. This is the strategy used in the `KTCLUS` package [17]. Once the ordering and test variables are separated, however, the clustering sequence depends in general on the value of $y_{\rm cut}$ and all computations involve a complete reclustering up to that value.

For example, in the original Durham algorithm the value of y_3 for a given event is defined precisely by the point in the clustering sequence at which three objects become two, independently of $y_{\rm cut}$. In the modified algorithms, the calculation of y_3 involves repeatedly reclustering the event with different values of $y_{\rm cut}$, until the value at which three objects become two is found with adequate precision, e.g. by a binary search procedure. The extra computation might have caused difficulties when the JADE algorithm was originally proposed, but advances in computing over the past ten years mean it is no longer a problem.

3 Fixed-order results

It is common in the specialized literature to define the n-jet fraction $f_n(y)$ by[1]

$$f_n(y) = \frac{\sigma_n(y)}{\sum_m \sigma_m(y)} = \frac{\sigma_n(y)}{\sigma_{\rm tot}}, \tag{3.1}$$

where y is the jet resolution parameter. If $\sigma_{\rm tot}$ identifies the *total* hadronic cross section $\sigma_{\rm tot} = \sigma_0(1+\alpha_{\rm S}/\pi+...)$, σ_0 being the lowest-order Born cross section, then the constraint $\sum_n f_n(y) = 1$ applies.

For the choice $\mu = Q = E_{\rm cm}$ of the renormalization scale, one can conveniently write the n-jet fractions (e.g., for the cases $n = 2, 3$ and 4) in the following form [10]:

$$f_2(y) = 1 - \left(\frac{\alpha_{\rm S}}{2\pi}\right) A(y) + \left(\frac{\alpha_{\rm S}}{2\pi}\right)^2 (2A(y) - B(y) - C(y)) + ..., \tag{3.2}$$

$$f_3(y) = \left(\frac{\alpha_{\rm S}}{2\pi}\right) A(y) + \left(\frac{\alpha_{\rm S}}{2\pi}\right)^2 (B(y) - 2A(y)) + ..., \tag{3.3}$$

$$f_4(y) = \left(\frac{\alpha_{\rm S}}{2\pi}\right)^2 C(y) + ..., \tag{3.4}$$

where the coupling constant $\alpha_{\rm S}$ and the functions $A(y), B(y)$ and $C(y)$ are defined in some renormalization scheme (in what follows we shall make use of the $\overline{\rm MS}$ scheme). The terms of order $\mathcal{O}(\alpha_{\rm S}^2)$ involving $A(y)$ take account of the normalization to $\sigma_{\rm tot}$ rather than to σ_0.

For the original Durham jet algorithm, several studies were carried out in Ref. [10], together with a parametrization of the auxiliary functions $A(y), B(y)$ and $C(y)$. There, analytic fits of the form

$$F(y) = \sum_{n=0}^{4} k_n \left(\ln\frac{1}{y}\right)^n \tag{3.5}$$

[1] From now on, in order to simplify the notation, we shall often use y to represent $y_{\rm cut}$.

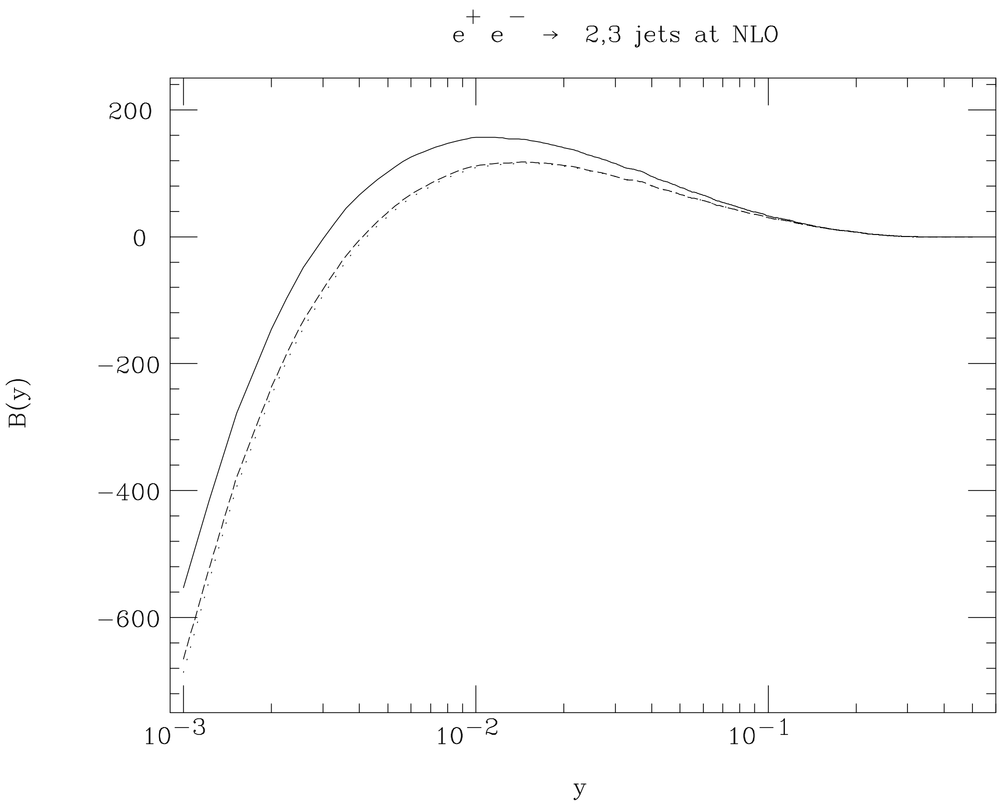

Figure 6: The parton level $B(y)$ function entering in the two- and three-jet fractions (3.2)–(3.3) at NLO in the D (continuous line), A (dashed line) and C (dotted line) schemes.

were performed for $F = A, B$ and C. The coefficients k_n were tabulated for values of y in the range 0.01–0.33 in the case of $A(y)$ and $B(y)$, whereas for $C(y)$ the interval was 0.01–0.1.

In this Section we present results similar to those of Ref. [10] for the case of the two modifications of the original Durham clustering scheme that we are proposing, the 'angular-ordered Durham' and 'Cambridge' schemes, and we compare the performances of all three algorithms in predicting several results of phenomenological relevance in QCD analyses. In particular, in order to match the advances on the experimental side in resolving jets at very small values of y, in our studies we have extended the range of the jet resolution parameter down to the minimum figure of 0.001. However, in this respect, we have to stress that at very small values of y (i.e., around 0.001) the perturbative calculations become unreliable, because of the growing five-jet contribution, which of course does not appear in our $\mathcal{O}\left(\alpha_S^2\right)$ computation. In other words, for small y's, terms of the form $\mathcal{O}\left(\alpha_S^n \ln^m y\right)$ with $m \leq 2n$ become large and have to be resummed [18][2]. In

[2]We look forward to the availability in numerical form (see, e.g., Ref. [19]) of the complete $\mathcal{O}\left(\alpha_S^3\right)$

producing the numerical results presented in this Section we have made use of the Monte Carlo (MC) program EERAD [22], which implements the complete next-to-leading-order (NLO) corrections to the $e^+e^- \to 2$ jet and $e^+e^- \to 3$ jet rates, as well as the leading-order (LO) contributions to $e^+e^- \to 4$ jets. For the purpose of illustration, the total c.m. energy Q is always chosen to be the mass of the Z-boson and no QED initial-state-radiation (ISR) has been included. The numerical values are $M_Z = 91.17$ GeV for the mass and $\Gamma_Z = 2.516$ GeV for the width (as we use the $e^+e^- \to Z, \gamma^* \to \ldots$ annihilation rates).

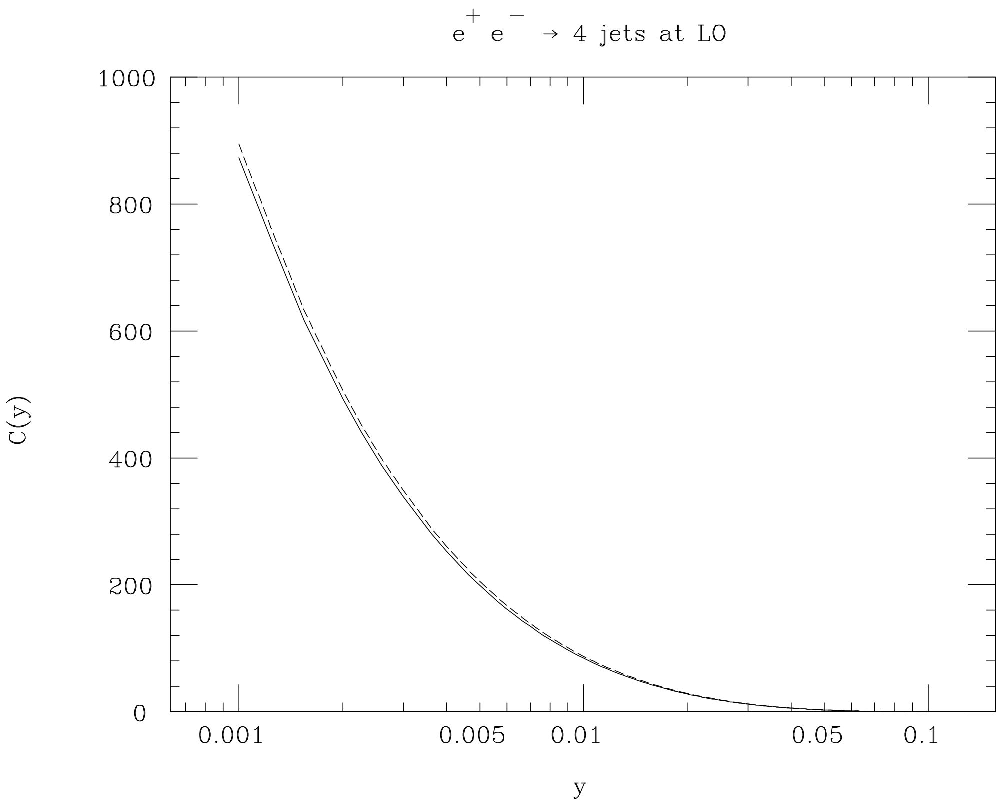

Figure 7: The parton level $C(y)$ function entering in the four-jet fraction (3.4) at LO in the D, A (continuous line) and C (dashed line) schemes.

Up to the order $\mathcal{O}(\alpha_s^2)$, differences in the jet fractions (3.2)–(3.4) as predicted by the three jet algorithms D, A and C can only occur when calculating the NLO contributions to the two- and three-jet rates and the LO ones to the four-jet cross section. More precisely, only the functions $B(y)$ and $C(y)$ can vary depending on the scheme adopted. In particular, $B(y)$ turns out to be different for all three jet algorithms, whereas $C(y)$ is the same for the D and A schemes but differs for the C scheme. This can easily be

results, toward which contributions have been made by two different groups [20, 21].

understood if one recalls that, in comparison to the D algorithm, the A algorithm only involves a modification in the combination scheme (i.e., the 'angular ordering') used in the n-parton $\to$ $(n-m)$-jet transitions (with $n \geq 4$ and $1 \leq m \leq n-1$), whereas the C algorithm also affects the n-parton $\to$ n-jet rates for $n \geq 4$ (i.e., via the 'soft freezing' procedure). Indeed, note that for $n = 3$ partons, kinematical constraints impose that, on the one hand, the two closest particles are also those for which y is minimal and, on the other hand, the identification of the softest of the three partons as a jet implies that the remaining two particles are naturally the most energetic and far apart. This clearly implies that the $A(y)$ function is the same for all three algorithms. For the above reasons then, we only plot the distributions for the $B(y)$ and $C(y)$ functions. This is done in Figs. 6 and 7, respectively.

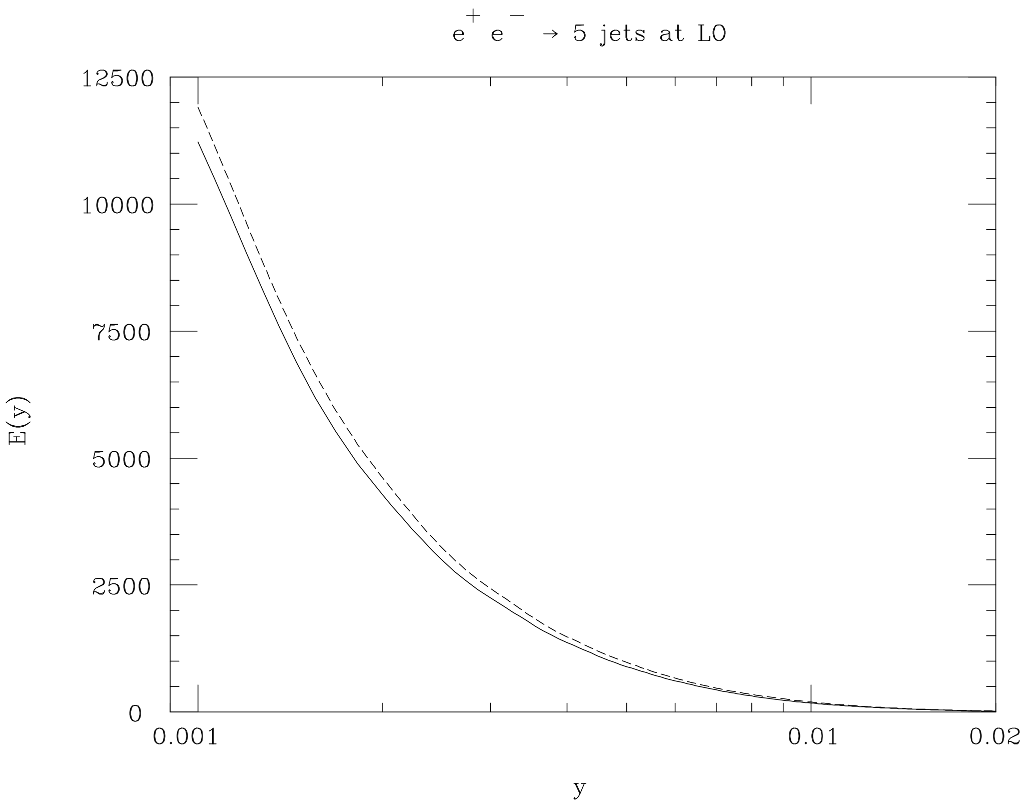

Figure 8: The parton level $E(y)$ function entering in the five-jet fraction at LO in the D, A (continuous line) and C (dashed line) schemes.

Whereas at LO all the contributions to the three-jet cross section come from the parton process $e^+e^- \to q\bar{q}g$, at NLO contributions to the three-jet rates are due to two sources. First, the real emission diagrams for the processes $e^+e^- \to q\bar{q}gg$ and $e^+e^- \to q\bar{q}Q\bar{Q}$, in which one of the partons is unresolved. This can happen when one has either two

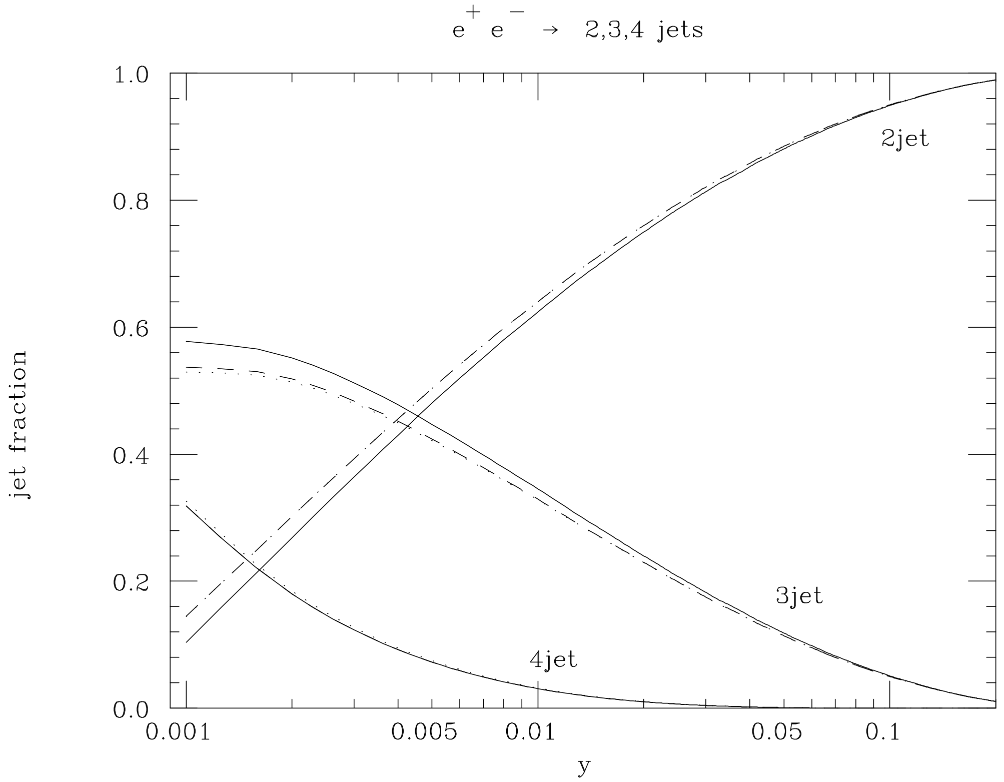

Figure 9: The $\mathcal{O}(\alpha_s^2)$ two-, three- and four-jet fractions at parton level in the D (continuous lines), A (dashed lines) and C (dotted lines) schemes.

collinear partons within one jet or one soft parton outside the jet cones. Both these contributions are (in general, positively) divergent. Thanks to the Bloch-Nordsieck [23] and Kinoshita-Lee-Nauenberg [24] theorems (see also Ref. [25]), these collinear and soft singularities are cancelled at the same order in α_s by the divergent contributions (generally negative) provided by the second source, namely the virtual loop graphs. Therefore, after renormalizing the coupling constant α_s a finite three-jet cross section is obtained. The function $B(y)$ accounts for the above-mentioned three- and four-parton contributions. The LO four-jet cross section at a given y-value is due to four-parton events for which all the measures y_{ij} are greater than the resolution parameter. These contributions are contained in the $C(y)$ term.

From Fig. 6, one can appreciate that, whereas differences between the two proposed modifications of the Durham algorithm are negligible, those between these two and the original one can amount to several tens of percent in a sizable part of the experimentally relevant y-range, especially for $y \lesssim 0.01$. Let us then investigate their behaviour. It is clear that $B(y)$ for the A and C algorithms is consistently smaller (both in magnitude and sign) than for the D one. This turns out to produce a smaller three-jet fraction, compare to Eq. (3.3), where the $B(y)$ term enters with a positive sign (the leading piece proportional to $A(y)$ clearly dominates). This is rather natural since, as discussed in Sect. 2.2, the D algorithm represents a sort of upper bound in jet multiplicity to its variation A. The 'angular ordering' procedure tends to combine more pairs of partons than the original D clustering does, as the former would pair the two closest particles even when their y-value is not the smallest of the entire event. If these two partons yield the minimum y, then the two combination schemes coincide. Conversely, this inevitably produces an enhancement of the lowest order jet fraction, that is the two-jet one, see Eq. (3.2), where $B(y)$ is indeed preceded by a minus sign and the $C(y)$ term is unaffected by the A prescriptions. The additional step of 'soft freezing' implemented in the C algorithm tends to enhance the final jet multiplicity of the original D scheme, by preventing the softer particle in a resolved jet pair from attracting the remaining particles (at large angle) into unresolved parton pairs, which would then be merged together, producing a lower number of final jets. In this respect, therefore, the yield of the clustering procedure of the original D algorithm acts, in terms of final jet multiplicity, as a lower bound on that produced by the implementation of the 'soft freezing' step. This can be seen from Fig. 7, for the auxiliary function $C(y)$ to which the four-jet rate (3.4) is proportional, though the effect remains at the level of $2-4\%$ over the considered range in y. The increase of the four-jet rate is compensated for by a further decrease (with respect to the A scheme) of the C-scheme three-jet fraction, as can be seen in the $B(y)$ term from Fig. 6 and Eq. (3.3). However, contrary to the 'angular ordering' case, when 'soft freezing' is adopted the two-jet rates are much less affected, as both the two terms $B(y)$ and $C(y)$ vary and in opposite directions.

As an additional exercise, in order to show that, as previously discussed, the effects of 'soft freezing' do become more evident as the final state multiplicity increases, we plot in Fig. 8 the $(\alpha_{\rm S}/2\pi)^3$ coefficient of the five-jet fraction at LO (which we call $E(y)$), as obtained from the programs already employed in Refs. [26, 27]. Here, differences between, on the one hand, the D and A schemes and, on the other hand, the C one, can be quite large, more than twice those obtained in the case of the $C(y)$ function (compare to Fig. 7). For example, at $y = 0.001$ these amount to 6% and become relatively larger with y. At the upper end of the interval considered in Fig. 8, $y = 0.02$, they are at the level of $\approx 15\%$.

In summary, at the order $\mathcal{O}\left(\alpha_{\rm S}^2\right)$, one could conclude the following: whereas 'angular ordering' tends to shift a part of the three-jet rate into the two-jet fraction without affecting the four-jet cross section, 'soft freezing' enhances the fraction of four-jet events and reduces the three-jet one, with a small effect on the two-jet rate. Note that the A scheme only implements the first step, so that the corresponding effect is more promptly visible. In contrast, as the C scheme adopts both these procedures, in this case the two effects tend to mix together in the final rates. However, in this case one expects the 'angular ordering' effects to be dominant over those due to the 'soft freezing', as can be argued by comparing Figs. 6 and 7. These effects are summarised in Fig. 9, where the two-, three- and four-jet fractions as defined via eqs. (3.2)–(3.4) are plotted. As we do not possess the complete $\mathcal{O}\left(\alpha_{\rm S}^3\right)$ result, note that we are unable to present similar results

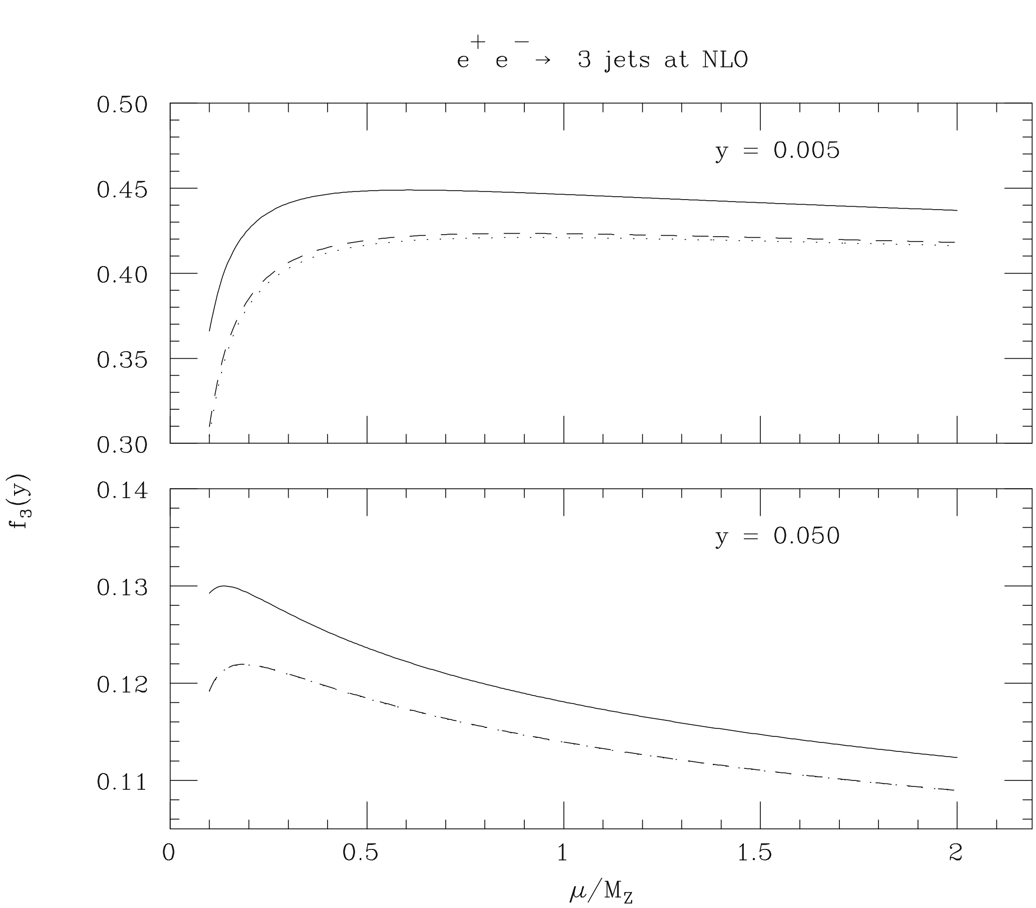

Figure 10: The parton level three-jet fraction at NLO as a function of the renormalization scale μ at $y = 0.005$ and $y = 0.050$, for the D (continuous lines), A (dashed lines) and C (dotted lines) schemes.

for the five-jet fraction. The value adopted in Fig. 9 for $\alpha_{\rm S}$ is 0.120, as given in Ref. [28].

In the remainder of this Section we focus our attention on the three-jet fraction. In particular, we would like to point out that, since the algorithms D, A and C have identical three-jet rates at LO, it is evident from Fig. 9 and Eq. (3.3) that the NLO corrections to $f_3(y)$ contained in the coefficient function $B(y)$ are significantly smaller in the two new algorithms. From the point of view of the convergence of the QCD asymptotic expansion, and hence of the reliability of the theoretical predictions, this should represent an improvement. In the sense that, provided one adopts a suitable value of the scale parameter μ regulating the corrections (see Ref. [29] for discussions of the best choice of the subtraction scale), it might not be unreasonable to argue that jet algorithms having smaller NLO terms may also have smaller higher-order corrections. Clearly, this can be assessed only when the latter are computed. However, some hints on this point can be deduced from studying the scale dependence of the complete NLO results. In fact, we know that the $\mathcal{O}\left(\alpha_{\rm S}^3\right)$ corrections are guaranteed to cancel the μ-dependence of the

$\mathcal{O}(\alpha_S^2)$ three-jet fraction up to the order $\mathcal{O}(\alpha_S^4)$.

The μ-dependence of the three-jet rate is simply introduced in Eq. (3.3) by the two substitutions

$$\alpha_S \to \alpha_S(\mu), \qquad\qquad B(y) \to B(y) - A(y)\beta_0 \ln\frac{Q}{\mu}, \tag{3.6}$$

where β_0 is the first coefficient of the QCD β-function, that is $\beta_0 = 11 - 2N_f/3$, with N_f the number of flavours active at the energy scale μ. Our results are shown in Fig. 10, where the value of $f_3(y)$ is plotted as a function of the scale parameter $\mu/Q \equiv \mu/M_Z$, over the range between 1/10 and 2, at two fixed values of the jet resolution parameter. Note that for the strong coupling constant we have used the two-loop expression, with $N_f = 5$ active flavours and $\Lambda^{(5)}_{\overline{\rm MS}} = 250$ MeV (yielding $\alpha_S(M_Z) = 0.120$, to match the rates given in Fig. 9).

Although the structure of the QCD perturbative expansion does not prescribe which value should be adopted for the scale μ, an obvious requirement is that it should be of the order of the energy scale involved in the problem (i.e., the c.m. energy Q). Such a choice prevents the appearance of large terms of the form $(\alpha_S \ln(\mu/Q))^n$ in the perturbative series. Therefore, it is reassuring to recognise that in Fig. 10 the A and C curves are more stable than the D ones around the value $\mu/Q = 1$, especially at smaller y-values, where the differences in $f_3(y)$ among the three schemes are larger. For example, over the range $1/2 \le \mu/Q \le 2$ the variation of $f_3(y)$ between its maximum and minimum values at $y = 0.005$ is $\approx 2.7\%$ for the D scheme and $\approx 1.3\%$ for A and C. For $y = 0.050$ differences between the three algorithms are smaller, as $f_3(y)$ varies by $\approx 10\%$ in the first scheme and $\approx 8.7\%$ in the other two.

F	Algorithm	y-range	k_0	k_1	k_2	k_3	k_4
A	D, A, C	$0.001 - 0.20$	0.843	-2.177	1.237	-0.0287	0.00312
B	D	$0.001 - 0.20$	49.287	-64.104	14.639	7.955	-1.529
B	A	$0.001 - 0.20$	26.869	-33.941	2.930	9.053	-1.573
B	C	$0.001 - 0.20$	28.970	-37.056	4.448	8.800	-1.569
C	D,A	$0.001 - 0.10$	2.370	-12.736	15.431	-7.070	1.121
C	C	$0.001 - 0.10$	-1.405	-7.776	13.201	-6.705	1.111

Table 1: Parametrization of the three- and four-jet QCD functions A, B and C as polynomials $\sum_n k_n(\ln(1/y))^n$, for the Durham algorithm and its variants. The range of validity in y is given for each case.

We conclude this Section by presenting a polynomial fit of the form (3.5) to the A, B and C functions in the D, A and C schemes. The range of validity of our parametrization extends from $y = 0.001$ (for all three auxiliary functions) up to $y = 0.2$ for A and B and $y = 0.1$ for C. This should allow for an easy comparison between data and perturbative predictions over the y-range normally adopted in phenomenological studies (see, e.g., Ref. [14]). The values of the coefficients k_n, with $n = 0, ...4$, are reported in Tab. 1.

4 Mean number of jets

In Sect. 2 we studied the mean value of y_3, the value of $y_{\rm cut}$ at which three jets are just resolved, as a measure of jet algorithm performance, using a simple 'tube' model to estimate hadronization effects. That model had no perturbative contribution to y_3, and so we could explore the region of low $y_{\rm cut}$ simply by increasing the energy (see Fig. 3). In the presence of perturbative contributions, $\langle y_3\rangle$ is not so small (about 0.02 at $Q = M_Z$). Therefore for more realistic investigation of the interplay between perturbative and non-perturbative effects, we choose here to investigate a different quantity, namely the mean number of jets, $n_{\rm jets}$. This may be written in terms of the individual n-jet fractions (3.1) as

$$n_{\rm jets} \equiv \mathcal{N}(y) = \sum_{n=1}^{\infty} n f_n(y). \tag{4.1}$$

The mean number of jets has the advantage that it can be studied as a function of the jet resolution y, down to arbitrarily low values, at fixed energy. Furthermore, its perturbative behaviour at very low values of y can be computed including resummation of leading and next-to-leading logarithmic terms to all orders [13]. This behaviour can be matched with the fixed-order results of the previous section, to give reliable predictions (apart from non-perturbative contributions) throughout the whole range of $y_{\rm cut}$. Non-perturbative contributions can then be estimated by comparing the perturbative results with those of a Monte Carlo event generator such as HERWIG [8].

4.1 Resummed predictions

Using the notation of the previous Section, we have from Eqs. (3.2)-(3.4)

$$\mathcal{N}(y) = 2 + \left(\frac{\alpha_{\rm S}}{2\pi}\right) A(y) + \left(\frac{\alpha_{\rm S}}{2\pi}\right)^2 (B(y) + 2C(y) - 2A(y)) + \ldots\,. \tag{4.2}$$

The behaviour of the first-order coefficient $A(y)$ at small y is of the form

$$A(y) = C_F \left(\ln^2 y + 3\ln y + r(y)\right)\ , \tag{4.3}$$

where $C_F = 4/3$ and the non-logarithmic contribution is [5, 6]

$$r(y) = 6\ln 2 + \frac{5}{2} - \frac{\pi^2}{6} + 4\left(\ln(1+\sqrt{2}) - 2\sqrt{2}\right)\sqrt{y} - 3.7y\ln y + \mathcal{O}(y)\ . \tag{4.4}$$

For the second-order coefficient the expected behaviour, for all three versions of the algorithm, is

$$F(y) \equiv B(y) + 2C(y) - 2A(y) = C_F\left[\frac{1}{12}C_A\ln^4 y - \frac{1}{9}(C_A - N_f)\ln^3 y + \mathcal{O}\left(\ln^2 y\right)\right]\ , \tag{4.5}$$

where $C_A = 3$. The terms of order $\ln^2 y$ depend in general on the version of the algorithm (D, A or C). To find them, we made a fit of the form (3.5), restricted now to the region $0.001 < y < 0.02$, with the coefficients k_3 and k_4 fixed at the values prescribed by

Algorithm	k_0	k_1	k_2	k_3	k_4
D	58.334	−37.546	13.420	−0.2963	0.3333
A	88.817	−45.218	11.554	−0.2963	0.3333
C	113.20	−58.455	13.412	−0.2963	0.3333

Table 2: Parametrization of the second-order coefficient in the average number of jets as a polynomial $\sum_n k_n(\ln(1/y))^n$, for the Durham algorithm and its variants. The range of validity is $y < 0.02$.

Eq. (4.5) and k_0, k_1, k_2 as free parameters. The results are given in Tab. 2. Using these fits for $y < 0.01$ and those in Tab. 1 for $0.01 < y < 0.1$, we obtain simple, smooth parametrizations of the second-order coefficient $F(y)$ over the whole range of $y < 0.1$.

To obtain resummed perturbative predictions for the mean number of jets, we now proceed as in Ref. [13]. To next-to-leading logarithmic (NLL) accuracy, the resummed results are independent of the version of the algorithm. Therefore the only modifications to the predictions for the original D algorithm come from the matching to the fixed-order results given above. We simply subtract the first- and second-order terms of the NLL resummed result and substitute the corresponding exact terms. Denoting by $\mathcal{N}_q$ the NLL multiplicity in a quark jet, given in [13], we obtain

$$\mathcal{N}(y) = 2\mathcal{N}_q(y) + C_F\left(\frac{\alpha_{\rm S}}{2\pi}\right) r(y) + \left(\frac{\alpha_{\rm S}}{2\pi}\right)^2 (F(y) - 2F_q(y)) \tag{4.6}$$

where F_q is the second-order coefficient in $\mathcal{N}_q$, given in [30]:

$$F_q(y) = C_F\left\{\frac{1}{24}C_A\ln^4 y - \frac{1}{18}(C_A - N_f)\ln^3 y + \frac{N_f}{9}\left(1-\frac{C_F}{C_A}\right)\left[\left(4\frac{C_F}{C_A}-1\right)\frac{N_f}{C_A}-1\right]\ln^2 y\right\}\,. \tag{4.7}$$

The resulting predictions for the three algorithms, using $\alpha_{\rm S} = 0.12$ as in the previous Section, are shown in Fig. 11.

4.2 Monte Carlo studies

We have used the Monte Carlo event generator HERWIG [8] (version 5.9 [31]) to study non-perturbative corrections to the above parton-level predictions. The program generates a shower by coherent branching of the partons involved down to a fixed transverse momentum scale ~ 1 GeV, and then converts these partons into hadrons using a cluster hadronization model. The jet algorithms can be run both on the results of the parton shower and on the final-state hadrons: the difference between these is customarily used as an estimate of the hadronization correction.

Fig. 12 shows the results for various jet algorithms at $Q = M_Z$. On the right are the Durham algorithm and the two variants of it proposed in this paper, with angular ordering (A) and with angular ordering and soft freezing (C). On the left, for comparison, are the JADE algorithm and its corresponding variants, in which the same ordering and freezing steps are introduced, but the JADE test variable (1.1) replaces (1.3).

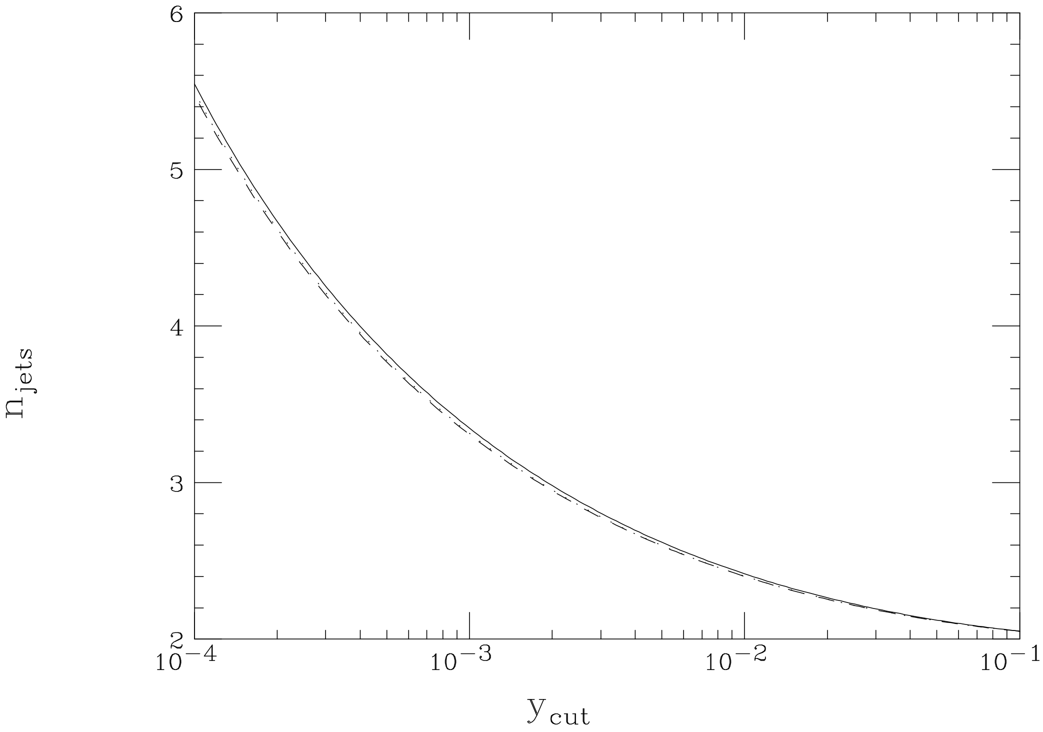

Figure 11: Resummed predictions for the mean number of jets at parton level in the D (continuous lines), A (dashed lines) and C (dotted lines) schemes.

Since the aim was to study the specific process $e^+e^- \to$ hadrons at a scale equal to the full centre-of-mass energy, HERWIG was set not to produce any initial-state radiation (`ZMXISR`=0). Each data point was calculated using 2000 events. A different pair of random number seeds was used for each value of $y_{\rm cut}$ in order for the results to be uncorrelated. The statistical errors are smaller than the size of the points.

At the parton level, the effects of the modifications to the basic JADE and Durham algorithms are scarcely visible in Fig. 12. Looking in more detail at the Durham results in comparison with the resummed predictions (Figs. 13-15), we see that in each case HERWIG at the parton level (squares) somewhat overestimates the number of jets, corresponding to a larger effective value of $\alpha_{\rm S}$. The dot-dashed curves correspond to $\alpha_{\rm S} = 0.120$, as in Fig. 11, whereas the dashed ones, which lie close to the Monte Carlo results, are for $\alpha_{\rm S} = 0.126$. The value $\alpha_{\rm S} = 0.114$ (dotted) corresponds to interpreting the HERWIG input parameter `QCDLAM`, for which we used the default value of 0.18 GeV, as the NLO scale parameter $\Lambda^{(5)}_{\overline{\rm MS}}$. Such an interpretation is only justified in a small region of phase space (see [32]) which we would not expect to be dominant, so a larger effective value is not surprising.

Turning to the hadron-level results, we see that the differences between the various algorithms are much more evident than at parton level. In particular, the jet multiplicity is reduced at $y_{\rm cut} \lesssim 10^{-2}$ in the A and C algorithms, resulting in substantially smaller hadronization corrections for $y_{\rm cut} \lesssim 10^{-3}$. At larger values, the hadron-level points even fall below the parton-level ones, suggesting a negative hadronization effect or possibly a positive non-perturbative contribution at the HERWIG parton level, which could account for the apparently high effective $\alpha_{\rm S}$. The presence of cut-offs and kinematic boundaries in the parton shower could indeed give rise to non-perturbative parton-level contributions. This illustrates the potential danger of interpreting the hadron–parton difference as an estimate of non-perturbative effects and simply adding it to the resummed prediction. In the case of the A and C algorithms, our results suggest this would lead to an overestimate of $\alpha_{\rm S}$.

Comparing the hadron-level results in Fig. 12, it can be seen that the angular ordering criterion alone is not enough to guarantee low hadronization corrections: it is also necessary to use the Durham test variable. In the 'seagull diagram' (Fig. 1) there is a significant fraction of phase space where the JADE test variable for the two gluons is less than $y_{\rm cut}$, whereas those for either gluon and its associated quark are both greater than $y_{\rm cut}$. In this region it will still be the two gluons that are combined even though this is the last of the three possibilities to be considered. As a consequence, the angular-ordered JADE algorithm still leads to 'phantom jets' in regions where there are no particles, enhancing the jet multiplicity and hadronization effects.

It is interesting to see that the angular-ordered JADE algorithm with soft freezing performs even worse in this context than its equivalent without freezing. This is because the freezing step increases the mean number of jets at the hadron level whilst having little effect at the parton level.

Fig. 16 shows the equivalent results at a typical LEP2 energy of 172 GeV. It is useful to reduce the hadronization corrections at this energy since that may help in discriminating between the process $e^+e^- \to WW \to 4$ jets, used to measure the W mass, and the background from $e^+e^- \to Z^0/\gamma \to 4$ jets. Here the results have been continued down to $y_{\rm cut} = 2\times10^{-5}$, the point at which non-perturbative corrections might be expected to set in at this energy.

The results at 172 GeV are similar to those at 90 GeV, except that they exhibit a lower multiplicity for all six algorithms at a given $y_{\rm cut}$, owing to the running of $\alpha_{\rm S}$, and that the parton- and hadron-level curves diverge at a smaller value of $y_{\rm cut}$, owing to the power-decrease in the hadronization corrections.

5 Discussion and conclusions

In this paper we have identified some deficiencies of the Durham (D) jet clustering algorithm and their causes. They are due mainly to the tendency of the algorithm to assemble 'junk-jets' from soft, wide-angle hadrons or partons, and to mis-cluster particles into already-resolved soft, wide-angle jets. To suppress junk-jet formation, we propose the separation of the ordering and resolution testing steps of the algorithm. Using the relative angle as the ordering variable and the $k_\perp$-resolution as the test variable, we obtain the angular-ordered Durham (A) algorithm. To reduce mis-clustering, we propose 'soft freezing', i.e. forbidding further clustering with the softer jet when a pair of jets are resolvable but have the smallest angle. Together with angular ordering, this gives the Cambridge (C) algorithm.

As measures of performance of jet algorithms, we studied two quantities in a simple 'tube' model of hadronization: $\langle y_3 \rangle$, the mean value of the resolution at which three jets are just resolved, and $\langle n_3 \rangle$, the mean multiplicity in the just-resolved third jet. Since the tube model has no real multijet production, in an optimal algorithm these quantities should both be as small as possible. We found that $\langle y_3 \rangle$ decreases much more rapidly with increasing energy in the Durham algorithm than in the original JADE scheme. Nevertheless the Durham value is enhanced by junk-jet formation, and $\langle y_3 \rangle$ is therefore reduced significantly by the angular-ordering step in the A and C schemes. The quantity $\langle n_3 \rangle$, on the other hand, is quite similar in the JADE and Durham schemes, smaller but still increasing in the A scheme, and essentially flat in the C scheme. Thus the two measures provide neatly complementary evidence of the benefits of the two new features of the C algorithm.

Next we presented full NLO perturbative calculations of the jet fractions in the new A and C algorithms, and compared the results with those for the original D scheme. The main difference is that the angular ordering in the new algorithms shifts a part of the three-jet rate into the two-jet fraction, while the extra 'soft freezing' in the C scheme moves some of the three-jet rate into the four-jet one. This is understandable because angular ordering tends to permit more clustering, while soft freezing prevents some clustering. The beneficial results include a significant reduction in the sensitivity to the renormalization scale.

We also extended the parametrization of the fixed-order results for the D scheme in Ref. [10] down to $y_{\rm cut} = 10^{-3}$ and provided similar parametrizations for the A and C schemes.

For a more realistic estimate of hadronization effects than that obtained from the tube model, we investigated the mean number of jets as a function of jet resolution in the region of low $y_{\rm cut}$. We combined the fixed-order results for this quantity with those obtained by resummation of leading and next-to-leading logarithms of $y_{\rm cut}$ to all orders, thereby obtaining reliable perturbative predictions for all $y_{\rm cut} < 10^{-1}$. The resummed terms are the same for the D, A and C algorithms; only the subleading logarithms and non-logarithmic parts had to be fitted differently in each case. This extended the results of Ref. [13] for the jet multiplicity in the D scheme to include the A and C schemes.

Finally, we used the Monte Carlo program HERWIG to generate simulated e^+e^- annihilation events, and compared the parton- and hadron-level jet multiplicities for the Durham algorithm and its proposed modifications with the perturbative predictions and with each other. We also investigated the equivalent modifications to the JADE algorithm, but these either showed no benefit or performed worse than the original.

We found that the mean number of jets at the HERWIG parton level was similar in the D, A and C schemes, and close to that predicted perturbatively, albeit for a rather high effective value of α_S. The HERWIG hadron-level values were more different, being lowest for the A scheme, slightly higher for the C scheme, and much higher for the original D scheme at small $y_{\rm cut}$. Down to $y_{\rm cut} \sim 10^{-3}$ at $E_{\rm cm} = M_Z$ (10^{-4} at $E_{\rm cm} = 172$ GeV), the C-scheme hadron-level values lie close to the perturbative ones, if the HERWIG input parameter `QCDLAM` is interpreted as the QCD scale parameter $\Lambda^{(5)}_{\overline{\rm MS}}$. In addition, the hadronization effects are greatly reduced below $y_{\rm cut} \sim 10^{-3}$, suggesting that the C algorithm will be particularly useful for exploring the interface between perturbative and non-perturbative dynamics [12].

A Fortran subroutine to perform jet clustering according to the C algorithm, `CAMJET`, may be obtained via the World-Wide Web at
`http://www.hep.phy.cam.ac.uk/theory/webber/camjet/camjet.html`.

Acknowledgments

We have benefited from valuable conversations on this topic with S. Bethke, S. Catani and M.H. Seymour. We are grateful to Nigel Glover for his help in using the program EERAD. Yu.L.D. thanks the Cavendish Laboratory and B.R.W. thanks the CERN Theory Division and the St Petersburg Nuclear Physics Institute for hospitality while part of this work was carried out.

This research was supported in part by the U.K. Particle Physics and Astronomy Research Council and by the EC Programme "Training and Mobility of Researchers", Network "Hadronic Physics with High Energy Electromagnetic Probes", contract ERB FMRX-CT96-0008.

References

[1] JADE Collaboration, W. Bartel et al., *Phys. Lett.* **B123** (1983) 460; *Z. Physik* **C33** (1986) 23.

[2] Yu.L. Dokshitzer, contribution cited in Report of the Hard QCD Working Group, Proc. Workshop on Jet Studies at LEP and HERA, Durham, December 1990, *J. Phys.* **G17** (1991) 1537.

[3] N. Brown and W.J. Stirling, *Phys. Lett.* **B252** (1990) 657.

[4] S. Catani, in "QCD at 200-TeV", Proc. 17th Workshop of the INFN Eloisatron Project, Erice, June 1991.

[5] S. Catani, Yu.L. Dokshitzer, M. Olsson, G. Turnock and B.R. Webber, *Phys. Lett.* **B269** (1991) 432.

[6] N. Brown and W.J. Stirling, *Z. Physik* **C53** (1992) 629.

[7] T. Sjöstrand, *Comp. Phys. Commun.* **39** (1984) 347;
M. Bengtsson and T. Sjöstrand, *Comp. Phys. Commun.* **43** (1987) 367.

[8] G. Marchesini, B.R. Webber, G. Abbiendi, I.G. Knowles, M.H. Seymour and L. Stanco, *Comp. Phys. Commun.* **67** (1992) 465.

[9] OPAL Collaboration, M.Z. Akrawy et al., *Z. Physik* **C49** (1991) 375.

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

[11] Yu.L. Dokshitzer, V.A. Khoze and S.I. Troyan, *Phys. Rev.* **D53** (1996) 89.

[12] Yu.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, Cambridge preprint in preparation.

[13] S. Catani, Yu.L. Dokshitzer, F. Fiorani and B.R. Webber, *Nucl. Phys.* **B377** (1992) 445.

[14] OPAL Collaboration, P.D. Acton et al., *Z. Physik* **C59** (1993) 1.

[15] Yu.L. Dokshitzer, G. Marchesini and B.R. Webber, *Nucl. Phys.* **B469** (1996) 93.

[16] A.H. Mueller, *Phys. Lett.* **B104** (1981) 161;
B.I. Ermolaev and V.S. Fadin, *JETP Lett.* **33** (1981) 285.

[17] M.H. Seymour, *Z. Physik* **C62** (1994) 127 and private communications;
`KTCLUS` program available at `http://surya11.cern.ch/users/seymour/ktclus/`.

[18] S. Catani, G. Turnock, B.R. Webber and L. Trentadue, preprint Cavendish-HEP-90/16, August 1990 (unpublished); *Nucl. Phys.* **B407** (1993) 3;
G. Dissertori and M. Schmelling, *Phys. Lett.* **B361** (1995) 167.

[19] A. Signer, preprint SLAC-PUB-7531, submitted to *Comp. Phys. Commun.*

[20] Z. Bern, L. Dixon, D.A. Kosower and S. Weinzier, *Nucl. Phys.* **B489** (1997) 3;
Z. Bern, L. Dixon, D.A. Kosower, preprint SLAC-PUB-7529, June 1996, `hep-ph/9606378`;
A. Signer and L. Dixon, *Phys. Rev. Lett.* **78** (1997) 811;
A. Signer, SLAC preprint SLAC-PUB-7490, May 1997, `hep-ph/9705218`;
L. Dixon and A. Signer, preprint SLAC-PUB-7528, June 1997, `hep-ph/9706285`.

[21] E.W.N. Glover and D.J. Miller, *Phys. Lett.* **B396** (1997) 257;
J.M. Campbell, E.W.N. Glover and D.J. Miller, preprint DTP/97/44, RAL TR 97-027, June 1997, `hep-ph/9706297`.

[22] W.T. Giele and E.W.N. Glover, *Phys. Rev.* **D46** (1992) 1980.

[23] F. Bloch and A. Nordsieck, *Phys. Rev.* **52** (1937) 54.

[24] T. Kinoshita, *J. Math. Phys.* **3** (1962) 650;
T.D. Lee and M. Nauenberg, *Phys. Rev.* **133** (1964) 1549.

[25] G. Sterman, *Phys. Rev.* **D17** (1978) 2789.

[26] A. Ballestrero and E. Maina, *Phys. Lett.* **B323** (1994) 53.

[27] A. Ballestrero, E. Maina and S. Moretti, Proceedings of the "XXIXth Rencontres de Moriond: QCD and High Energy Hadronic Interactions", Méribel, Savoie, France, 16/26 March 1994, ed. by J. Trân Thanh Vân, ed. Frontières, Gif-sur-Yvette, 1994, 367.

[28] The LEP Electroweak Working Group and the SLD Heavy Flavor Group, preprints CERN-PPE-96-183, December 1996.

[29] G. Grunberg, *Phys. Lett.* **B95** (1980) 70;
S.J. Brodsky, G.P. Lepage and P.B. Mackenzie, *Phys. Rev.* **D28** (1983) 228;
P.M. Stevenson, *Nucl. Phys.* **B231** (1984) 65;
H.D. Politzer, *Nucl. Phys.* **B194** (1982) 493.

[30] S. Catani, Yu.L. Dokshitzer and B.R. Webber, *Phys. Lett.* **B322** (1994) 263.

[31] G. Marchesini, B.R. Webber, G. Abbiendi, I.G. Knowles, M.H. Seymour and L. Stanco, `hep-ph/9607393`.

[32] S. Catani, G. Marchesini and B.R. Webber, *Nucl. Phys.* **B349** (1991) 635.

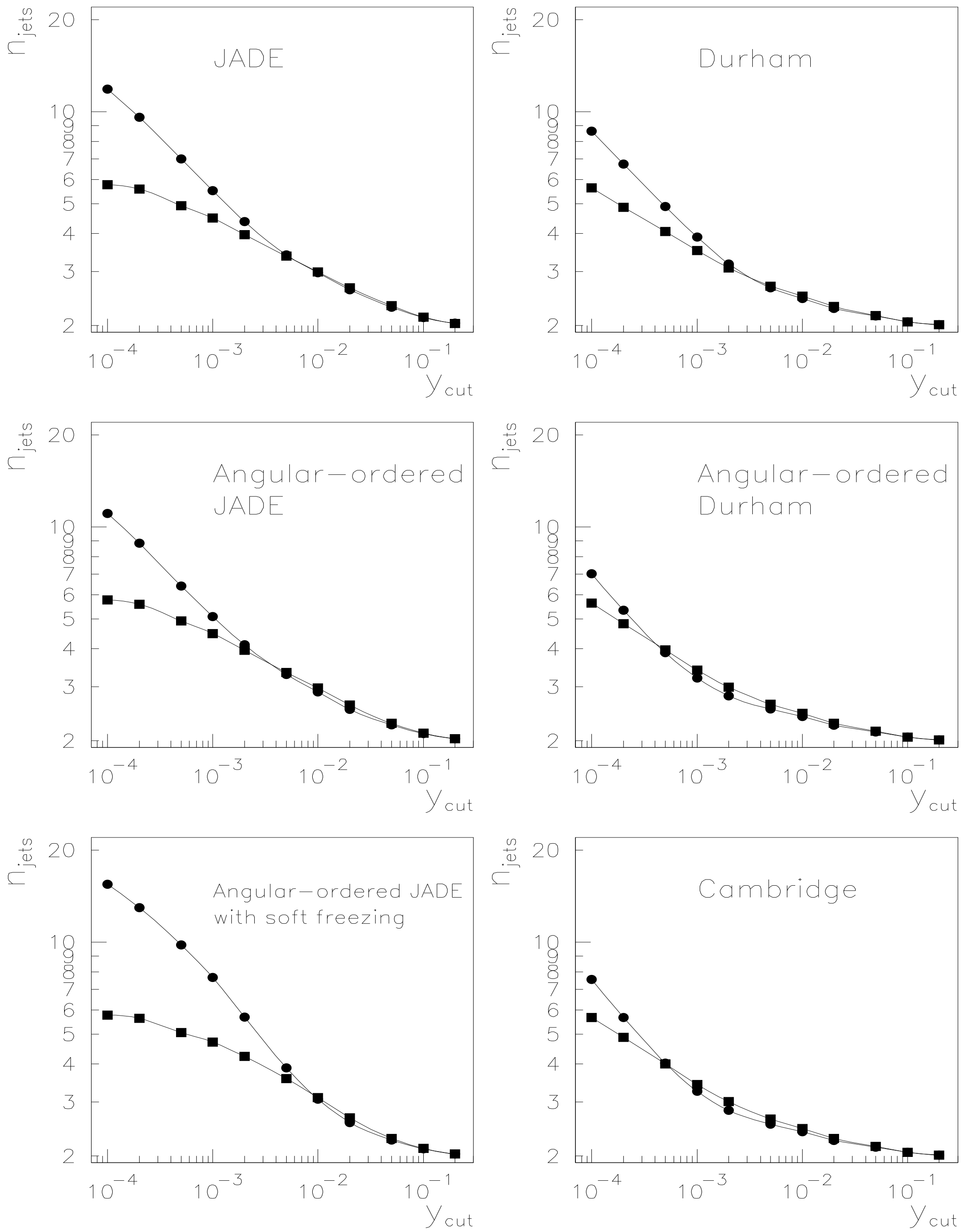

Figure 12: Parton-level (squares) and hadron-level (circles) results from HERWIG on the mean number of jets at $Q = M_Z$ for various jet algorithms as labelled.

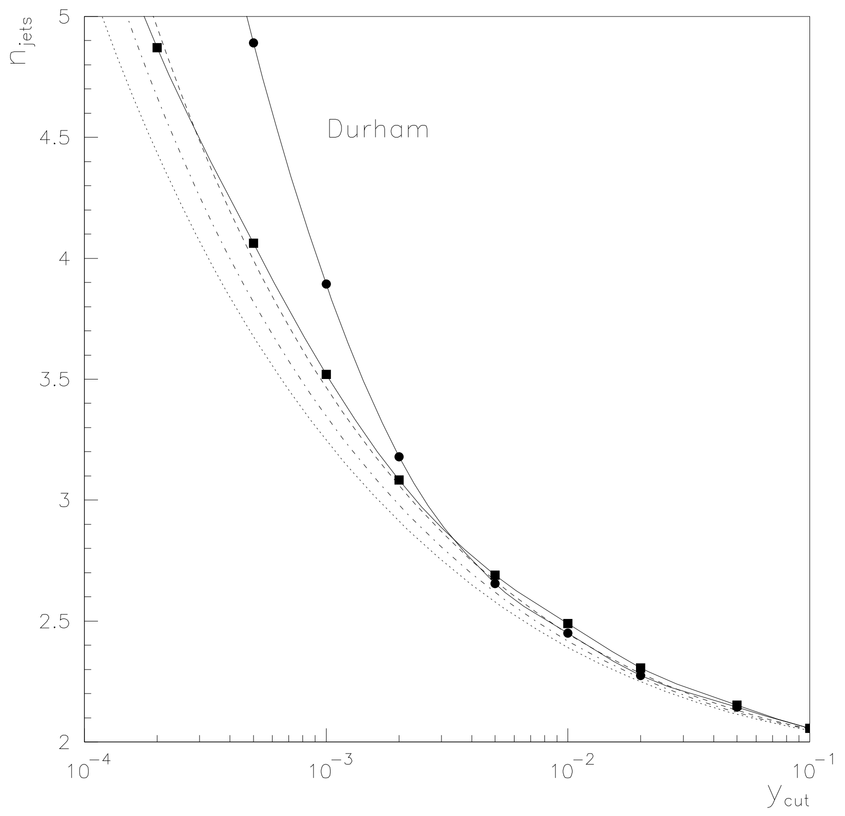

Figure 13: Results on the mean number of jets at $Q = M_Z$ for the original Durham algorithm. HERWIG: squares, parton level; circles, hadron level. Resummed: dashed, $\alpha_{\rm S} = 0.126$; dot-dashed, $\alpha_{\rm S} = 0.120$; dotted, $\alpha_{\rm S} = 0.114$.

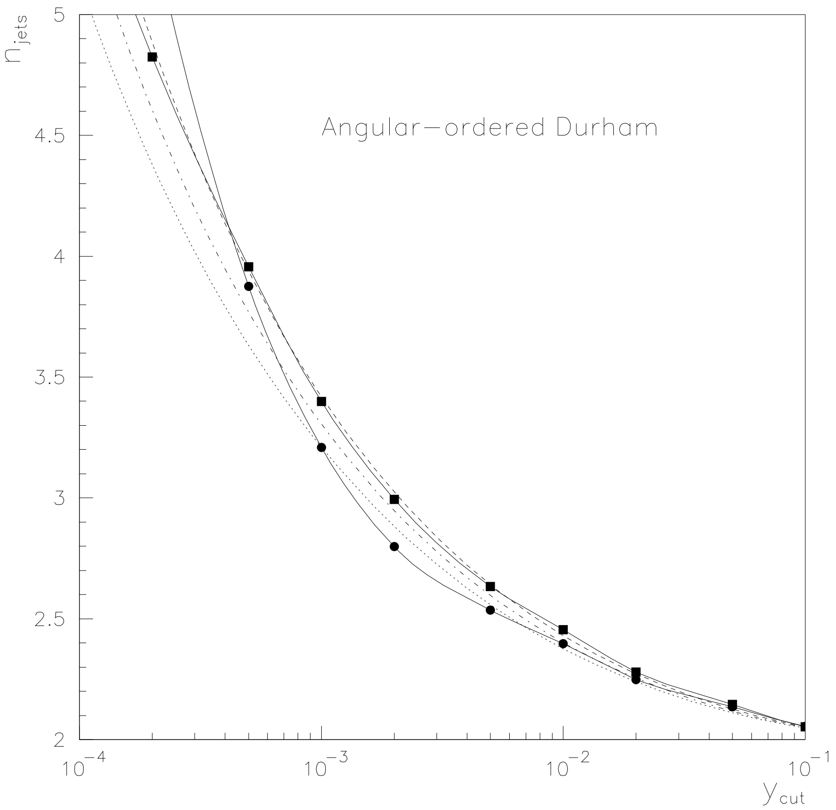

Figure 14: Results on the mean number of jets at $Q = M_Z$ for the angular-ordered Durham algorithm. Points and curves as in Fig. 13.

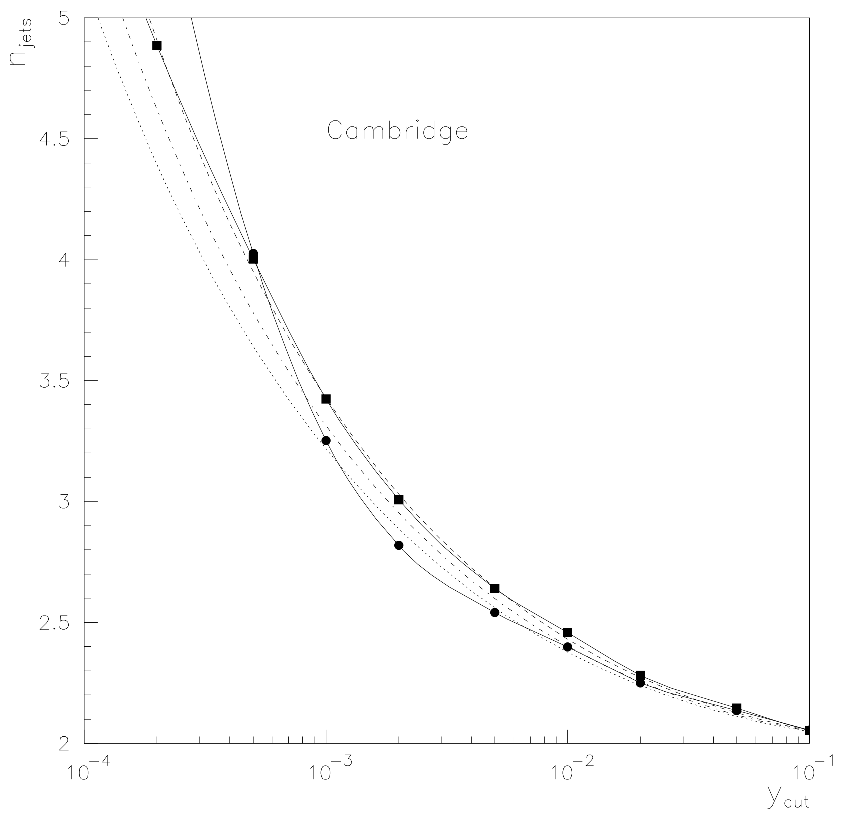

Figure 15: Results on the mean number of jets at $Q = M_Z$ for the Cambridge algorithm. Points and curves as in Fig. 13.

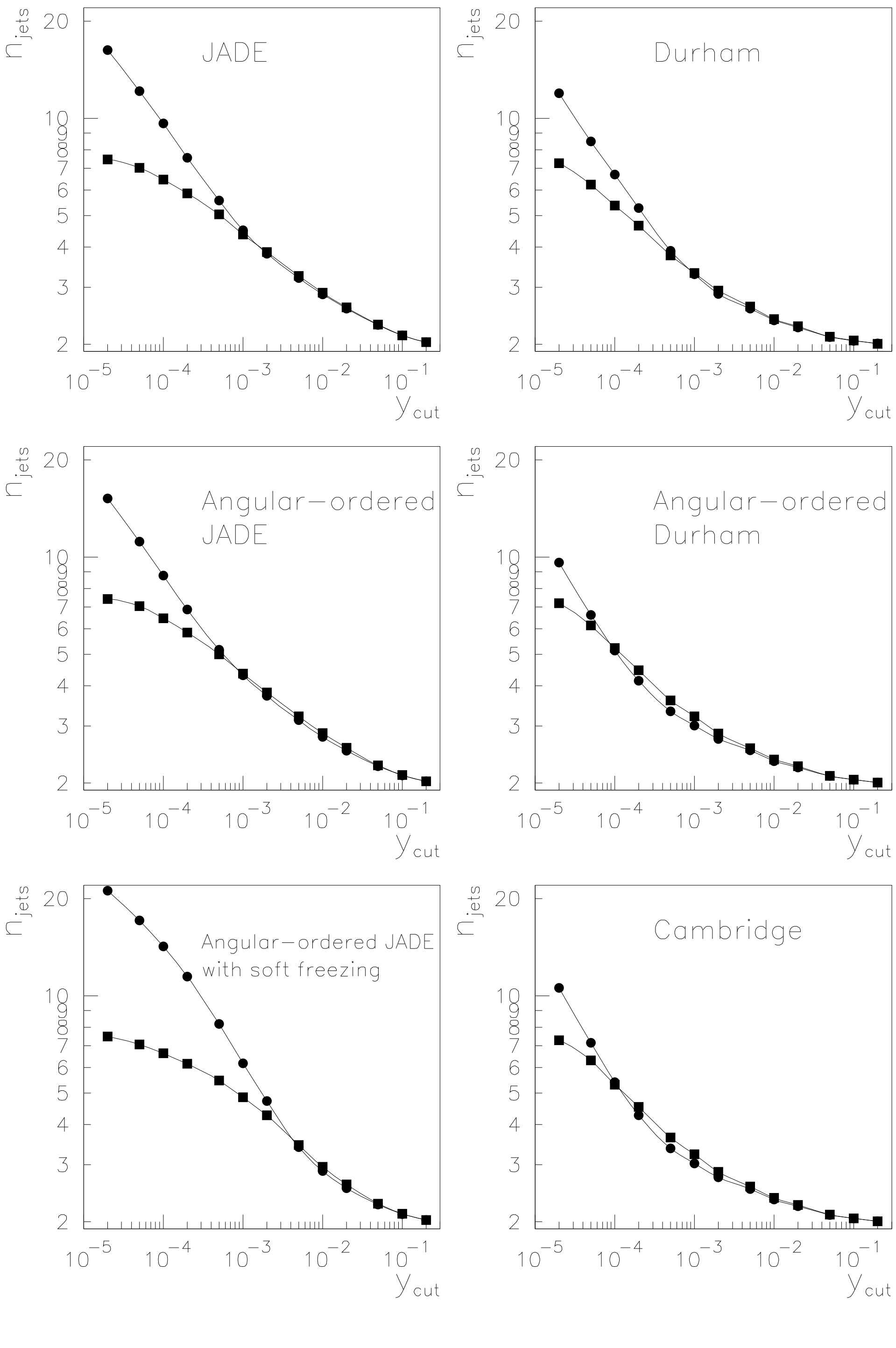

Figure 16: Parton- and hadron-level HERWIG results on the mean number of jets, as in Fig. 12 but at $Q = 172$ GeV.