Jet Topology and New Jet Counting Algorithms

Stefano Catani†
Theory Division, CERN, CH-1211 Geneva 23, Switzerland

Abstract

A QCD theoretical analysis of jet cross sections is presented. We discuss in detail the JADE (invariant mass-type) and $k_T$-algorithms for multijet cross sections in $e^+e^-$-annihilation, with emphasis on hadronization effects and large perturbative corrections at small values of the jet resolution parameter $y_{cut}$. For the $k_T$-algorithm we present results of calculations which include resummation of leading and next-to-leading logarithms of $y_{cut}$ to all orders in QCD perturbation theory. We also discuss the recently-proposed Geneva algorithm and show that it suffers from some of the same difficulties as the JADE algorithm.

---

†Invited talk at the 17th Workshop of the INFN Eolissatron Project, QCD at 200 TeV, Erice, June 1991.
†On leave of absence from INFN, Sezione di Firenze, Italy.
1. Introduction

A jet can be defined as a large amount of hadronic energy in a small angular region. According to this qualitative definition, the first Evidence for Jet Structure in Hadron Production by $e^+e^-$ Annihilation was reported in 1975 [1]. Since then, hadronic jet production in lepton [2,3] and hadron [4,5] colliders has become a main tool in investigating strong interaction physics and testing Quantum Chromodynamics (QCD).

We are nowadays in a position to use jet cross section data both for

- precise quantitative tests of QCD (measurement of the QCD coupling $\alpha_S$ and scale $\Lambda_{QCD}$ [3], study of QCD coherence [6]);
- looking for breakdown of the standard model and new physics [8].

In order to do that, the qualitative definition of a jet given above is no longer sufficient and must be replaced by a precise quantitative definition. We need a jet algorithm able to specify unambiguously a jet configuration starting from hadrons detected in the final state. A jet algorithm is defined giving

(i) a test variable (energy-angle resolution) $y_{\mu}$;

(ii) a recombination procedure.

The test variable $y_{\mu}$ is needed in order to specify whether or not two hadrons $h_i, h_j$ belong to the same jet, whilst the recombination procedure tells us how jet properties are related to the ones of hadrons belonging to it.

The jet defining conditions (i) and (ii) have to fulfill the requirements of being

i) infrared (IR) and collinear safe

ii) simple to implement in the experimental analysis

iii) simple to implement in the theoretical calculation

iv) subject to small hadronization corrections.

Requirements ii) and iii) are self-evident.

The other requirements follow from the motivation of comparing data with theory, namely QCD. So far, we are able to perform QCD calculations, using perturbation theory, only for the region of small distances in which high energy collisions produce partons (quarks, gluons). At large distances the produced quarks and gluons are confined by the colour force field and are forced to dress themselves up into colourless hadrons. Although it is not possible to describe the hadronization process by perturbation theory, the "preconfinement" property [7] (or "local parton hadron duality" [8]) of QCD implies that hadron jets should maintain the kinematic features of the underlying quarks and gluons. That is because, according to preconfinement, the hadronic flow in the final state follows the partonic flow quite closely, with transfers of momentum and other quantum numbers that are local in phase space.

The way to enhance preconfinement effects (or, equivalently, to reduce hadronization corrections) is to define jet cross sections which at parton level are not very sensitive to large distance physics, namely any small value of parton masses. Therefore jet algorithms must satisfy the property i) above, i.e. jet cross sections at parton level must be finite order by order in perturbation theory in the limiting case of final state massless partons [9, 10]. Obviously, hadronization corrections still affect jet algorithm satisfying the property i). Therefore one should try to minimize hadronization effects for instance using Monte Carlo event generators to compare parton level and hadron level results for different jet algorithms.

Still within the class of jet algorithms fulfilling the requirements i) - iv), one has ample freedom of choosing test variables and recombination procedures. The aim of this paper is not to provide a comprehensive review of existing jet algorithms [11] but to present a QCD inspired critical analysis of them. Our goal is to advocate a QCD-based jet algorithm, i.e. a definition of a jet whose dynamics follows QCD theory as closely as possible.

We start our analysis in Sec. 2, discussing the jet algorithm mostly used at exclusive level, namely the "JADE algorithm" for $e^+e^-$-annihilation [12]. The description of good and bad features of the JADE algorithm performed in Secs. 2 and 3 will provide us with the physical motivations to introduce the QCD inspired jet algorithm which we call $k_t$-algorithm [13]. The $k_t$-algorithm is defined in Sec. 4 where we present our main theoretical results. In Sec. 5 we draw some conclusion. Some results on another jet algorithm, the recently proposed "Geneva algorithm" [14,3], are presented in Appendix A.

2. The JADE algorithm

2.1 Definition

The JADE algorithm is a full exclusive algorithm to define jet cross sections in $e^+e^-$-annihilation. Let us consider a n-hadron final state

$$e^+(p) + e^-(\bar{p}) \rightarrow h_1(p_1) + \cdots + h_n(p_n), \quad p + \bar{p} \equiv Q.$$  

(1)

The JADE algorithm is defined according to the following iterative procedure.

1) Define a resolution parameter $y_{\mu_{\text{cut}}}$.

2) For every pair of hadrons $h_i, h_j$ compute the corresponding "invariant mass" $M_{ij}^2$ and define

$$y_{\mu} = M_{ij}^2 / Q^2.$$  

(2)

3) If $y_{\mu}$ is the smallest value of $y_{\mu}$ computed in 2) and $y_{\mu} < y_{\mu_{\text{cut}}}$ combine $(p_i, p_j)$ in a single jet ("pseudoparticle") $p_{ij}$ according to a recombination prescription.

4) Repeat this procedure from step 2) until all pairs of objects (particles and/or pseudoparticles) have $y_{\mu} > y_{\mu_{\text{cut}}}$. Whatever objects remain at this stage are called jets.

According to the classification given by $y_{\mu}$, $y_{\mu}$ in Sec. 2, the JADE algorithm uses the scaled invariant mass (2) as test variable. In the experimental as well as theoretical analyses several different recombination procedures may be used [12,15]. They differ among themselves in the way in which the "invariant mass" and the 4-momentum $p_{ij}$ of a pseudoparticle are defined in terms of the momenta $(p_i, p_j)$ of the recombed particles and/or pseudoparticles. There are essentially two classes of recombination schemes (Table 1) whose main difference is in the definition of the "invariant mass". In the first class (E-scheme) the invariant mass $M_{ij}$ is the true invariant mass, whilst in the massless recombination schemes (JADE, EC, F, P0) $M_{ij}$ is computed defining the pseudoparticle momenta in such a way to have massless pseudoparticles. The various
schemes in this latter class correspond to various possibilities of rescaling energy and 3-momentum in order to define a massless pseudoparticle.

The JADE algorithm obviously fulfills the requirements (i) and (ii) in Sec. 1. In the following we discuss in detail the points (iii) and (iv).

2.2 Hadronization corrections

Hadronization effects for the JADE algorithm can be investigated using Monte Carlo event generators [16-18]. These Monte Carlo generators include a parton shower, describing parton production according to perturbative QCD, plus hadronization models. Comparing jet cross sections obtained by Monte Carlo simulations both from partons at the end of the QCD shower and from particles after hadronization, one is able to estimate the size of hadronization corrections.

Fig. 1 shows the results of the analysis carried out by the OPAL collaboration [19] at LEP using the JETSET QCD shower model. Similar results have been obtained from other Collaborations at LEP and lower energies as well as using different Monte Carlo generators.

It can be seen from Fig. 1 that hadronization corrections are large for the E-recombination scheme, moderate for the P- and F0-scheme and small for the F0-scheme (the results for the JADE-scheme are similar to the ones for the F0-scheme). The fact that hadronization effects are relevant for most of the recombination schemes may signal that something is going wrong with the JADE algorithm and call for a better theoretical understanding of jet definition.

2.3 Theoretical calculations and renormalization scale problem

Let \( R_n \) represent the n-jet fraction

\[
R_n = \frac{\sigma_{n-jet}}{\sigma_{\nu T}} .
\]

Since the JADE algorithm is IR and collinear safe, \( R_n \) can be computed in QCD perturbation theory. So far the n-jet fractions \((n = 2, 3, 4)\) have been evaluated up to the second order in \( \alpha_s \) and the result can be written as follows

\[
R_n(\alpha_s(\mu), Q^2/\mu^2; \nu_{cut}) = \delta_{n2} + \alpha_s(\mu) C_n^{(1)}(\nu_{cut})
\]

\[
+ \alpha_s(\mu) \left( \frac{C_n^{(2)}(\nu_{cut})}{C_2^{(1)}(\nu_{cut})} - C_n^{(2)}(\nu_{cut}) \beta_0 \ln Q^2/\mu^2 \right) + O(\alpha_s^3(\mu), \alpha_s^2(\mu) \ln^2 Q^2/\mu^2) ,
\]

where \( \delta_{n2} + \alpha_s(\mu) C_n^{(1)}(\nu_{cut}) \) is the n-jet fraction in the \( n_j \)-jet region and \( C_2^{(1)}(\nu_{cut}) = 11C_2 - 2N_f, N_f \) is the number of flavours, \( C_r = N_r + 3 \) is the number of colours and the coupling \( \alpha_s(\mu) \) can be expressed as a function of the QCD scale \( \Lambda_{\overline{MS}} \)

\[
\alpha_s(\mu) = \frac{1}{\ln(\mu^2/\Lambda_{\overline{MS}}^2)} \left[ 1 - \frac{\alpha_s^2(\mu)}{3} \left( \frac{\beta_0}{\beta_0^2} \right) \right] .
\]

In one loop order \( C_2^{(1)} = 0, \) and \( C_2^{(1)}, C_3^{(1)} \) are given by

\[
C_2^{(1)}(\nu_{cut}) = -C_3^{(1)}(\nu_{cut}) = \frac{C_F}{2\pi} \left[ -2 \ln 3 + \frac{\pi^2}{3} - \frac{5}{2} \right]
\]

\[
+ 6\ln(1 + \ln(\nu_{cut}) + 4\nu_{cut} \ln(1 - \nu_{cut}) - 2\ln(1 - \nu_{cut}) - 3(1 - 2\nu_{cut}) \ln(1 - 2\nu_{cut}) + \frac{9}{2} \nu_{cut} \ln(1 - \nu_{cut}),
\]

where

\[
\frac{C_F}{2\pi} = \frac{N_c^2 - 1}{2N_c}, \quad \beta_0 = \frac{\alpha_s}{2\pi} \sum_{n=1}^{\infty} \frac{n}{n^2}. \tag{7}
\]

The two loop functions \( C_n^{(2)}(\nu_{cut}) \) (note that \( C_2^{(2)} \) and \( C_3^{(2)} \) are recombination scheme dependent) have been computed numerically \([15,20]\) using the four-parton matrix elements in Ref. [21].

The two loop expression in eq. (4) depends on the renormalization scale \( \mu^2 \) at which the running coupling \( \alpha_s(\mu) \) is evaluated. This dependence is an artifact of the fixed order perturbative expansion in \( \alpha_s \) because \( R_n(\alpha_s(\mu), Q^2/\mu^2; \nu_{cut}) \) is a renormalization group invariant quantity if computed to all orders in \( \alpha_s \). We have taken into account this property writing explicitly a correction term of order \( \alpha_s^3(\mu) \ln Q^2/\mu^2 \) on the r.h.s. of eq. (4).

In comparing eq. (4) with data, one has to fix the value of \( \mu^2 \). Since the only physical scale in \( \mu^2 \) is the centre of mass energy \( Q^2 \), the "natural" value to be used for \( \mu^2 \) is \( \mu^2 \approx Q^2 \).

The two loop QCD prediction for \( R_4 \) has been compared with data (after correction for detector acceptance and hadronization effects) at \( Z^0 \) energies \([19,22]\) and below \([12,23,24]\). As an illustrative example we show in Fig. 2 the result for \( D_3(\nu_{cut}) = \frac{dR_3(\nu_{cut})}{d\nu_{cut}} \) of the OPAL Collaboration [19].

The main conclusion is that eq. (4) with \( \mu^2 = Q^2 \) agrees with data in the large \( \nu_{cut} \) region \((\nu_{cut} > 0.05)\) whilst a good description of the data in the small \( \nu_{cut} \) region \((\nu_{cut} < 0.05)\) requires the use of very small renormalization scales \( \mu^2 \ll Q^2 \) [19].

The need of such "unnatural" renormalization scale to fit experimental data, already noticed for other quantities \([25,19]\), requires some physical explanation.

A first tentative explanation is that for \( \nu_{cut} \ll 1 \), jet cross section definition involves two very different scales \( Q^2 \) and \( \nu_{cut}^2 \). Therefore one might argue that \( \mu^2 \approx \nu_{cut}^2 \). However that implies that \( R_n \) are functions of \( \alpha_s(\sqrt{\nu_{cut}}) \) and hence for \( \nu_{cut} \ll 1 \) they depend on large distance physics and, eventually, become IR sensitive. It follows that \( \mu^2 \approx \nu_{cut}^2 \) has to be excluded on theoretical grounds.

A second tentative explanation might use the recourse to some optimization procedure. Several prescriptions to fix the value of the renormalization scale \( \mu^2 \) have been proposed \([26-28]\) and are nowadays widely used. Their application to jet cross sections in the JADE algorithm has been investigated in \([29]\). It turns out that the optimized renormalization scale \( \mu_{opt}^2 \) is much smaller than \( Q^2 \) for \( \nu_{cut} > 0.05 \) but one can get \( \mu_{opt}^2 \approx Q^2 \) reducing \( \nu_{cut} \) to very small values (Fig. 3).

Although neither \( \mu^2 \approx \nu_{cut}^2 \) nor \( \mu_{opt}^2 \) have theoretical Justification, one can try to compare the corresponding predictions with data. From the analysis of Ref. [29], we conclude (Fig. 4) that both prescriptions are not able to improve the agreement with data in the relevant region of small \( \nu_{cut} < 0.05 \).

We regard the results described above as the failure of phenomenological attempts to justify the choice of values of \( \mu^2 \) very different from \( Q^2 \) and we think we should offer our theoretical explanation of that.

Let us state more precisely the argument that the physical value for the renormalization scale is \( \mu^2 \approx Q^2 \). In order to compare the \( \alpha_s(\mu) \)-expression (4) with data one has to assume that the perturbative expansion for \( R_n \) is convergent. If \( \mu^2 \) is very
different from $Q^2$, we know that the two loop result for $R_m$ neglects large higher order logarithmic corrections of the type $(\ln^2 Q^2/m^2)^2$ for instance the $\mathcal{O}(a_s^2)(m^2 \ln^2 Q^2/m^2)$ term in eq. (4). Therefore, modulo fortuitous cancellations, when $m^2 \neq Q^2$ one expects the perturbative expansion for $R_m$ is badly convergent and one cannot safely compare the fixed order truncation with data. And vice versa, by studying the $m^2$ dependence of the perturbative series in $a_s(m^2)$ one can test its convergence: if the series is convergent, $R_m$ should be a stable function of $m^2$ for $m^2 \approx Q^2$ and a rapidly varying function for $m^2$ very different from $Q^2$.

It turns out that in second order $dR_m/dm^2$ for $m^2 \approx Q^2$ strongly depends on $y_{cut}$ for small $y_{cut}$ values. This means that the perturbative series in $a_s$ is not convergent for $y_{cut} \ll 1$. The choices $m^2 \approx y_{cut}Q^2$ or $m^2 \approx m^2_{lept}$ are just guesses to improve the convergence of the perturbative expansion. We think the only reliable procedure is to identify the origin of the terms spoiling the convergence and to compute them in higher order.

Large coefficients in the $a_s$,expansion for $R_m$ already appear in one loop order. From eqs. (4), (6) we have ($L = ln 1/y_{cut}$)

$$R_2 \approx 1 - \frac{C_F a_s}{\pi} L^2 , \quad R_3 \approx \frac{C_F a_s}{\pi} L^2 , \quad (y_{cut} \ll 1).$$

(8)

The double logarithmic term in eq. (8) comes from the bremsstrahlung spectrum for emission of a soft (with energy fraction $z = \omega/E \ll 1$) and collinear (with angle $\theta \ll 1$) gluon from the $q \bar{q}$ pair (Fig. 5). In the soft and collinear limit the single gluon emission probability is

$$d\sigma(1) = \frac{C_F a_s}{\pi} \frac{d\sigma}{d\theta^2} \frac{d\theta}{\pi} ,$$

(9)

which, integrated over the phase space for the JADE algorithm ($y_{cut} \approx e^{\theta^2}$), gives

$$\int d\sigma(1) \Theta(e^{\theta^2} - y_{cut}) = \frac{C_F a_s}{\pi} \int_0^\infty d\theta^2 \int_0^{\theta^2} \Theta(e^{\theta^2} - y_{cut}) = \frac{C_F a_s L^2}{2\pi} .$$

(10)

The large contributions $a_sL^2$ are responsible for (i) the unphysical shape of the $n$-jet rates computed in fixed order perturbation theory (for instance in one loop order $R_2 \rightarrow -\infty$ for $y_{cut} \rightarrow 0$) and (ii) the instability of the perturbative expansion with respect to renormalization scale variations [30].

At small values of $y_{cut}$, higher order terms in $a_s$ are enhanced by powers of $\ln^2 y_{cut}$. In this kinematical region the real expansion parameter is the large effective coupling $a_{\text{eff}}L^2$ and therefore any finite-order perturbative calculation cannot give an accurate evaluation of the cross section. The logarithmic terms must be resummed to all orders in $a_{\text{eff}}$ before a reliable prediction can be made. No choice of $m^2$ in a fixed order result can account for such a resummation because changing $m^2$ one can only introduce single logarithmic corrections ($a_sL^2$) in higher orders. A particular choice of $m^2$ can at most simulate the double logarithmic resummation over a limited range of $y_{cut}$ [30,31].

3. The small $y_{cut}$ region

3.1 Exponentiation of Sudakov logarithms

The appearance of large double logarithmic terms is a common feature of any hard process in the semi-inclusive or Sudakov region, where emission of radiation is inhibited by the kinematics [32,33]. In the case of jet cross sections at $y_{cut} \ll 1$, the jet invariant mass is constrained to be so small as to allow only emission of gluons that are soft and collinear with respect to the parton generating the jet. The double logarithmic terms $a_s \ln^2 y_{cut}$ are due to such soft and collinear gluons.

In recent years, following the pioneering calculations [34] of double logarithms, the programme of resummation of Sudakov logarithms has been successfully carried out to next-to-leading order for transverse [35] and longitudinal momentum distributions [36,37] in hadron-initiated processes, as well as for quantities like the energy-energy correlation [38] and the thrust [39] and heavy jet mass distribution [40] in $e^+e^-$ annihilation. All these successful resummations depend on the fact that the large logarithmic corrections to the relevant quantity exponentiate. By this we mean that terms of the form $a_s \ln^m y_{cut}$ with $m \leq 2n$ can be combined into an exponential function of less singular terms, in fact with $m \leq n + 1$. Once exponentisation has been established, the resummation of leading and next-to-leading logarithms to all orders is reduced to a much simpler, finite-order calculation of the relevant exponent.

Let us recall the physical basis for exponentisation starting from the simpler case of Quantum Electrodynamics (QED). In QED photons are not charged and this implies that multiple soft photon emission is an independent process. Thus the probability $d\omega(1,\ldots,n)$ for $n$-soft photon emission factorizes in the product of single photon factors $d\omega(1)$ in eq. (9)

$$d\omega(1,\ldots,n) = \frac{n}{n!} \int d\omega(1) \cdots \int d\omega(n) = \frac{1}{n!} \int d\omega(1) \cdots \int d\omega(n) \Theta(\omega - y_{cut}) .$$

(11)

Starting from the factorized result (11) one can get the corresponding cross sections contribution by integration over the relevant phase-space $\Theta(1,\ldots,n; y_{cut})$. Therefore the factorisation of multiphoton amplitudes leads to the exponentiation of the cross section if and only if the phase space constant factorizes in the soft limit, i.e.

$$\Theta(1,\ldots,n; y_{cut}) \approx \frac{n}{n!} \Theta(1; y_{cut}) ,$$

(12)

$$1 + \sum_{n=1}^\infty \int d\omega(1,\ldots,n) \Theta(1,\ldots,n; y_{cut}) \approx 1 + \sum_{n=1}^\infty \int d\omega(1) \cdots \int d\omega(n) \Theta(1; y_{cut}) = \exp \left[ \int d\omega(1) \Theta(1; y_{cut}) \right] .$$

(13)

The property (11) comes from QED dynamics whilst (12) depends on the kinematical definition of the cross section. The cross section exponentiates in the Sudakov region iff its definition does not induce kinematical correlations among soft photons.

Let us come to the QCD case. The main difference with respect to QED is due to the fact that gluons have colour charge. Therefore they can radiate in cascade and soft gluon emission is no longer an independent process. Strong gluon correlations are enforced by the dynamics, multiple gluon emission is not factorized into single emission contributions and we can only expect some kind of generalized exponentiation.

Nevertheless a simple exponentiation structure is still valid for highly inclusive cross sections like a large class of two jet dominated quantities [34,35,36,37,38]. Let us consider for example the two loop non-factorized contributions in Fig. 6. If the two-jet cross section is defined in such a way that the integration over the soft gluon degrees of freedom is unconstrained, the real (Fig. 6a) and virtual (Fig. 6b) contribution cancel each other in the IR and collinear regions. The only logarithmic term which survives
comes from the ultraviolet region for the virtual diagram. It leads to the running of the
QCD coupling $\alpha_S \to \alpha_S(k_F)$ with an argument as given by the transverse momentum
$k_F$ of the emitted gluon [39].

The cancellation described so far is preserved to higher orders due to unitarity and
gauge invariance properties [40] and leads to the factorisation of the one gluon emission
probability (9) with $\alpha_S \to \alpha_S(k_F)$ \(k_F \sim \epsilon Q\). Exponentiation then follows in the
absence of kinematic correlations among soft partons, related to the actual definition of the
cross section.

3.2 Leading and next-to-leading logarithms for the JADE algorithm

In order to check whether exponentiation works for jet cross sections defined ac-


\[ R_n = \delta_m + \frac{\alpha_S(Q)}{\pi} \left[ A_1^{(n)} L^2 + A_3^{(n)} L + O(1) \right] \]
\[ + \frac{\alpha_S(Q)}{\pi} \left[ B_1^{(n)} L^3 + B_3^{(n)} L^2 + O(L^2) \right] \]

(14)

We have computed the leading \( A_1^{(n)} \), \( B_1^{(n)} \) and next-to-leading \( A_3^{(n)}, B_3^{(n)} \) log-
arithmic coefficients for the JADE algorithm in the E-recombination scheme and the
results [41] are reported in Table 2 (the \( C_F \) contributions for \( B_1^{(n)} \) were first computed by
Brown and Stirling [42]).

One has the surprising result that the leading double logarithmic contributions do
not exponentiate even in the abelian limit \( C_F \to 0 \) [42]. In Table 2 we have split
the \( C_F \) contributions for the two loop coefficient \( B_1^{(n)} \) into two terms. The first one in
parentheses is the term corresponding to the exponentiation and the second one leads to
the violation of exponentiation.

According to our discussion in Sec. 3.1, these results imply that the JADE al-
gorithm leads to non-factorized phase space which, in the E-recombination scheme,
produces attractive kinematic correlations among soft gluons. For the two jet cross
section these correlations appear in the kinematic configuration in which the pair emits
one back-to-back \( (\theta_1, \theta_2 < \epsilon) \) and soft \( (\epsilon, \epsilon_F \ll \epsilon) \) gluons (Fig. 7). The corresponding
factorized phase space is

\[ \epsilon, \epsilon_F < \epsilon_{out}, \]

(15)

but, following the E-recombination procedure, the sub-region

\[ \epsilon, \epsilon_F > \epsilon_F \]  \( \epsilon, \epsilon_F > \epsilon_{out} \)

(16)

has to be subtracted from the two jet region and included in the three jet rate. We see
that the E-scheme gives rise in the region of eqns. (15), (16) to an “anomalous” three jet
configuration in which a jet is given by two very soft partons with a large relative angle.
One may say that the JADE algorithm counts the number of lumps instead of jets. 1

The attractive kinematic correlations induced by the E-recombination scheme im-
ply the following consequences.

1 We owe Nick Brown this nice picture.

- In two loop order, \( R_2 \) and \( R_4 \) decrease and \( R_3 \) increases with respect to the result for
independent emission.

- The kinematic correlations strongly increase with the number of final state par-
ton and therefore even the resummation of the leading ln\( y_{cut} \) terms to all orders appears
hopeless.

- The recombination procedure produces a lot of “anomalous” jets: soft and large
angle jets.

- Since, according to QCD preconfinement, hadronization is supposed to be a local
process in energy and angle, we expect (and do we have) large hadronization corrections.

In order to check whether these features are just a pathology of the E-scheme,
we have evaluated the jet rates (14) for the other recombination schemes. We found
[41] that exponentiation is still violated and that leading and next-to-leading terms are
the same for any massless recombination scheme (JADE, E0, P, P0). The results are
reported in Table 3. Some comments are in order.

Firstly, the next-to-leading coefficients \( E_1^{(n)} \) are equal to the ones for the E-scheme.
This is due to the fact that the kinematic correlations related to the jet definition affect
only the soft gluon phase space region.

Secondly, JADE, E0, P and P0-schemes have the same leading behaviour because the
corresponding pseudoparticles are massless. That provides a first explanation of the
different size of hadronisation corrections for the E-scheme and the massless recombina-
tion schemes (cf. Sec. 2.2).

Finally, soft gluon kinematic correlations for the 2 and 3 jet rates in the mass-
less schemes are smaller than for the E-scheme (compared the second contribution in
parentheses for \( E_1^{(n)} \) in Tables 2 and 3). That supports our explanation of the smaller
hadronisation effects observed for massless schemes. Note also that for massless schemes,
the attractive gluon correlations only reduce \( R_4 \) whilst both \( R_2 \) and \( R_3 \) are enhanced.

On balance, our conclusions about the JADE algorithm in the E-scheme are not
changed very much by our results for the other schemes. The only point is that massless
recombination procedures slightly reduce some bad features of the algorithm.

4. A new clustering algorithm: the \( k_t \)-algorithm

4.1 Definition

We believe that the analysis of the JADE algorithm carried out in Secs. 2 and 3
sufficiently motivates the need of a different jet definition. The main goal is to define a
clustering procedure which destroys soft gluon kinematic correlations in order to make
the theoretical calculation possible for \( y_{cut} \ll 1 \) and to reduce hadronisation effects.

We are looking for a jet algorithm whose dynamics features follow as close as
possible QCD theory. In the naive parton model, hadronic jets have a cylindrical shape.
By this we mean that they are produced with a total energy increasing with the centre
of mass energy and with a limited fixed transverse momentum \( k_t > \) (of order of some
hundred MeV) with respect to the jet axis. The QCD jets have instead an almost conical
shape: parton multiplication due to the fragmentation of the parent parton leads to
hadron jets whose mean transverse momentum increases with the jet energy [32, 33, 43].
Thus it is natural to introduce a jet algorithm\(^5\) in which the transverse momentum replaces the invariant mass of the original JADE algorithm as the jet resolution variable. To emphasize this change of variable, we call the new algorithm the \(k_L\)-algorithm [18].

The \(k_L\)-algorithm is defined in a similar way to the JADE algorithm, following the iterative procedure 1)-4) in Sec. 2.1 and just replacing the invariant mass by the transverse momentum as the jet resolution variable. Therefore, the test variable \(y_{\text{out}}\) at the step 2) becomes

\[
y_{\text{out}} = \frac{2(1 - \cos \theta_{\text{out}})}{Q^2} \min(E_T^2, E_T^t) = 2 \min(c, c_t) (1 - \cos \theta_{\text{out}}) .
\]  

(17)

For the recombinination procedure one can define several schemes (E, E0, P, P0) as in the JADE algorithm.

The main features of the \(k_L\)-algorithm (in any recombinination scheme) we expect are (i) small hadronization effects (positive preliminary results have been already reported at this meeting [3]) and (ii) the possibility of resumming large \(\ln y_{\text{out}}\) terms to all orders in perturbation theory. In fact, since the test variable (17) is diagonal with respect to the parton energies, it is a much more local energy-angle resolution variable than the invariant mass (which depends on both parton energies). For instance, it is trivial to check that the back-to-back configuration in Fig. 7 leads only to two jets in the factorized phase space region \(c_T^2/c_T^t < y_{\text{out}}\).

4.2 Recombiniation in the small \(y_{\text{out}}\) region

The \(n\)-jet rates

\[
R_n(y_{\text{out}}) = \sum_{m=0}^{n} \frac{1}{m!} \int d\sigma_{n+m}^\text{mt} \Theta_{n+m}(y_{\text{out}})
\]

(18)

are obtained by integration of the \(m\)-parton exclusive cross sections \(d\sigma_{n+m}^\text{mt}\) (\(m \geq n\)) over the \(n\)-jet phase space \(\Theta_{n+m}(y_{\text{out}})\) as given by the clustering procedure. We have computed the \(n\)-jet fractions using the coherent branching formalism [44,49,53] to evaluate the exclusive cross sections and working out the \(n\)-jet phase space to next-to-leading accuracy.

Using the \(k_L\)-algorithm we are able to resum the leading \(\mathcal{O}(\alpha_s^2 L^{2n})\) and next-to-leading \(\mathcal{O}(\alpha_s^3 L^n)\) logarithmic contributions to all orders in \(\alpha_s\) for any number of jets\(^6\). The result [13] can be simply stated in terms of the generating function \(\phi(Q, Q_0, u)\) for jet rates (\(Q_0^2 \equiv y_{\text{out}}\))

\[
R_n(y_{\text{out}}) = Q_0^2 (Q^2) \frac{\partial^n}{\partial u^n} \phi(Q, Q_0, u)\big|_{u = 0} .
\]

(19)

Defining the following emission probabilities

\[
\Gamma_e(Q, q) = \frac{2Q}{\pi} \frac{\alpha_s(q)}{q} \left( \log \frac{Q}{q} - \frac{3}{4} \right) ,
\]

\[
\Gamma_s(Q, q) = \frac{2Q}{\pi} \frac{\alpha_s(q)}{q} \left( \log \frac{Q}{q} - \frac{11}{12} \right) ,
\]

\[
\Gamma_t(Q) = \frac{N_f}{3\pi} \frac{\alpha_s(q)}{q} ,
\]

(20)

the generating function \(\phi\) in eq. (19) is given by

\[
\phi(Q, Q_0, u) = u^2 \exp \left( \frac{1}{Q_0} \int_0^u dq \Gamma_e(Q, q) \left( \phi(Q, Q_0, u) - 1 \right) \right) ,
\]

(21)

where the gluon functional \(\phi\) satisfies the following implicit equation

\[
\phi(Q, Q_0, u) = u \exp \left( \int_0^u dq \Gamma_g(Q, q) \left( \phi(Q, Q_0, u) - 1 \right) - \Gamma_t(q) \right)
\]

\[
\cdot \left( 1 + u \int_0^u dq \Gamma_t(q) \exp \left( \int_0^u dq \Gamma_g(q, q') \exp \left( \int_0^u dq' \Gamma_t(q', q'') \right) \right) \right) .
\]

(22)

Solving eq. (22) as a power series in the jet label \(u\), one can compute the jet rates \(R_n\) in eq. (19) for an arbitrary number of jets.

We refer to [13,41] for details of the calculations. Here, we want just to note that we are able to resum leading and next-to-leading \(\ln y_{\text{out}}\) terms because the \(k_L\)-resolution variable in eq. (17), which is diagonal with respect to parton energies, does not induce soft gluon kinematic correlations in the clustering procedure. Moreover, although various recombinination schemes (E, E0, P, P0) can differ significantly at finite \(y_{\text{out}}\), they all give the same leading and next-to-leading logarithms at small \(y_{\text{out}}\) [41]. Therefore the formulas (21),(22) as well as the ones we shall give below are valid for any of the common recombinination schemes.

The \(n\)-jet fractions one gets from eqs. (19)-(22) are (for \(n \leq 4\))

\[
R_0^{(n=1)} = |\Delta_e(Q)|^2 ,
\]

(23)

\[
R_0^{(n=2)} = 2 |\Delta_e(Q)|^2 \int_0^u dq \Gamma_g(Q, q) \Delta_e(q) ,
\]

(24)

\[
R_0^{(n=3)} = 2 |\Delta_e(Q)|^3 \int_0^u dq \int_0^u dq' \Gamma_g(Q, q) \Delta_e(q) \Delta_e(q') ,
\]

(25)

where we have introduced the quark and gluon Sudakov form factors

\[
\Delta_e(Q) = \exp \left( - \int_0^u dq \Gamma_g(Q, q) \right) ,
\]

(26)

\[
\Delta_g(Q) = \exp \left( - \int_0^u dq \Gamma_g(Q, q) + \Gamma_t(q) \right) ,
\]

(27)

which depend implicitly on \(y_{\text{out}}\) via \(Q_0 = Q_0 \sqrt{y_{\text{out}}}\), and we have defined

\[
\Delta_t(q) = |\Delta_t(q)|^2 / |\Delta_e(q)| .
\]

(28)

The relevant two loop coefficients in eq. (14) for the \(k_L\)-algorithm are given in Table 4.

The results in eqs. (23)-(28) are in agreement with the general discussion in Sec. 3.1. We see that a simple exponentiation of \(\ln y_{\text{out}}\) terms is valid for the two jet fraction \(R_1\).
The jet rates $R_n$ with $n > 2$ satisfy simple exponentiation in the pseudo-abelian limit $\Gamma_n, \Gamma_f \to 0$

$$R_n = \frac{\Delta^2(Q)}{(n-2)!} \left[ 2 \int_{q_0}^{Q_0} dq \Gamma_n(Q, q) \right]^{n-2} + (\Gamma_n \text{ and } \Gamma_f \text{ terms}), \quad (29)$$

whilst in the QCD case they follow a generalized exponentiation structure.

5. Discussion and outlook

In this paper we have presented a QCD theory based discussion of jet topology and jet algorithms. We have mainly focused on the small $y_{out}$ Sudakov region. From our analysis we argued that soft gluon kinematic correlations induced by the JADE clustering procedure may be responsible for large hadronization corrections and for bad convergence properties of the QCD perturbative expansion (see also Appendix A for a different jet definition). In Sec. 4 we have discussed a new jet algorithm, the $k_t$-algorithm, which does not suffer from the above problems.

Using the $k_t$-algorithm, large higher order corrections at small values of $y_{out}$ can be easily evaluated. We were able to resum leading and next-to-leading logarithms of $y_{out}$ to all orders in perturbation theory for any number of jets. The results in eqs. (23)-(25) may be combined with the exact fixed order results, after subtraction of the terms that have been exponentiated, to obtain predictions for up to four jets over the full range of $y_{out}$.

A question we have discussed at length is the renormalization scale dependence. The coefficients of the leading and next-to-leading logarithmic resumming terms in eqs. (23)-(25) are renormalization group invariant. Therefore, we cannot properly discuss the renormalization scale dependence in the $k_t$-algorithm until the matching with fixed-order results and a smooth extrapolation between the region of small and large $y_{out}$ have been performed [41]. Anyway, according to our discussion in Sec. 2.3 and to our experience [29,31], the renormalization scale problem is not a problem at all! Once a given quantity has been correctly computed resuming classes of logarithmic higher order corrections, its expansion in this “improved perturbation theory” should be convergent for values of the renormalization scale $\mu$ of the order of the physical scale of the process.

The $k_t$-algorithm can also be used to compute inclusive and differential jet cross sections as well as being applicable to hadron collision processes. Calculations are in progress.

We hope experimental data for jets within the $k_t$-algorithm will be soon available.

Acknowledgements. The results reported in this talk have been obtained in collaboration with Yu.L. Dokshitzer, M. Olsson, G. Turnock and B.R. Webber. Useful discussions with S. Bethke, M. Ciafaloni, G. Marchesini, D.E. Soper, W.J. Stirling and L. Trentadue are gratefully acknowledged. We would especially like to acknowledge many valuable and stimulating conversations with Nick Brown, to whose memory we respectfully dedicate this paper. It is a pleasure to thank Yuri Dokshitzer and Luisa Giarelli for organizing this pleasant and efficient Workshop at the Ettore Majorana Centre for Scientific Culture.

Appendix A - The Geneva algorithm

A different jet algorithm called the Geneva or $E_n$-algorithm [14,3] has been recently proposed to overcome the difficulty of resumming in $y_{out}$ terms in the JADE algorithm. In this clustering algorithm the jet resolution variable $y_{out}$ is defined by

$$y_{out} = \frac{2 E_n E_f}{(E_n + E_f)^2 (1 - \cos \theta_{nf})} = \frac{2 E_n E_f}{(q_n + q_f)^2 (1 - \cos \theta_{nf})} \quad (30)$$

and the particle are recombined according to the $E$-scheme (although other recombination procedure can be used as well).

The test variable (30) is essentially the two parton invariant mass times a rescaling factor $1/(q_n + q_f)^2$. This rescaling factor acts as a repulsive potential countering the attractive soft gluon correlations arising in the JADE algorithm. In order to have a better understanding of this effect we have performed some calculations for the small $y_{out}$ region.

Let us start by considering a pseudo-abelian model in which "gluons" cannot radiate in cascade $gg$ or $q\bar{q}$ real pairs. This model is similar to QED apart from the fact that the coupling constant runs as in QCD. For this model we were able to compute the generating function $\tilde{\phi}(Q, Q_0; u)$ of $n$-jet rates in the Geneva algorithm

$$R_n(y_{out} = Q_0^2/Q^2) = \frac{1}{n!} \left( \frac{\beta}{\mu} \right)^n \tilde{\phi}(Q, Q_0; u) \bigg|_{\mu = 0} \quad . \quad (31)$$

For the generating function $\phi$ to leading and next-to-leading logarithmic accuracy we find [41]

$$\tilde{\phi}(Q, Q_0; u) = u^\gamma \exp \left\{ (u - 1) \left[ \int_{Q_0/q}^{Q/Q_0} \frac{dz}{z} \int_{Q_0/q}^{Q_0/z} \frac{dz}{z} \alpha_S(z) - 3 \int_{Q_0/q}^{Q/Q_0} \frac{dz}{z} \alpha_S(z) \right] \right\} \quad . \quad (32)$$

The generating function (32) leads to a Poisson distribution and hence the corresponding $n$-jet rates exponentiate in this pseudo-abelian limit.

Let us come back to the QCD case. We have evaluated the two jet fraction $R_2$ resuming leading and next-to-leading in $y_{out}$ terms to all orders in $\alpha_S$ and we found the exponentiated expression

$$R_2 = \exp \left\{ - \left[ \int_{Q_0/q}^{Q/Q_0} \frac{dz}{z} \int_{Q_0/q}^{Q_0/z} \frac{dz}{z} \alpha_S(z) - 3 \int_{Q_0/q}^{Q/Q_0} \frac{dz}{z} \alpha_S(z) \right] \right\} \quad . \quad (33)$$

The results (32,33) confirm that the rescaling factor $1/(q_n + q_f)^2$ in eq. (30) is able to cancel the attractive kinematic correlations for the two jet cross section and for any jet rate in the abelian limit. However it looks surprising that the jet resolution variable (30), depending on both parton energies, may lead to a soft gluon phase space which is completely uncorrelated. A full two loop calculation confirms that kinematic correlation are still there. The result [41] for the two loop coefficients of eq. (14) in the Geneva algorithm is given in Table 5.

The abelian $C_F$ contributions to $B_1^{(1)}$ exponentiate, whilst the non-abelian terms $C_F C_A$ differ from those obtained for the JADE algorithm (these latter are reported in Table 5 as the first contribution to $B_1^{(1)}$ inside the parentheses). This difference is due to kinematic configurations like the one in Fig. 8 in which two soft gluons $(1 > q_n > e_f)$ are radiated in cascade $(1 > \theta_f > \theta_f)$ at large angle $(\theta_f > y_{out})$. In the JADE algorithm this configuration leads to a 4-jet event fraction $\{g_1 g_2 g_3 g_4\}$ and to a 3-jet event fraction $\{g_1 g_2 g_3\}$.
event fraction \( \{q, g, (g_1, g_2)\} \). In the Geneva algorithm the rescaling factor \( 1/(e_a + e)^2 \) in the test variable (30) induces a strong effective repulsion between the soft gluons \( g_1 \) and \( g_2 \) to produce a non-local 3-jet recombination \( \{q, (g_1, g_2), g_3\} \). This repulsive correlation enhances the 4-jet cross section and reduces the 3-jet cross section with respect to the JADE algorithm.

We conclude that the clustering procedure of the Geneva algorithm introduces repulsive kinematic correlations among soft gluons, thus leading to analogous problems to the ones arising in the JADE algorithm, viz.

- the resummation of the \( n_{\text{soft}} \) terms for \( K_n \) with \( n \geq 3 \) is (probably) hopeless;
- the jet recombination pattern is non-local in energy-angle and sizeable hadronization corrections are expected (preliminary results presented at this meeting [3] confirm this expectation).

References

S. Bethke, CERN preprint PPE/91-36.
3. S. Bethke, these proceedings.
4. UA2 Collaboration, J. Alitti et al., CERN preprint PPE/90-188;
B.L. Flaugher, FERMILAB-CONF-90/159-E.
5. M. Mangano, these proceedings.
6. V. Khose, these proceedings.
A. Bassetto, M. Ciafaloni and G. Marchesini, Phys. Lett. 83B (1979) 207;
15. Z. Kunst, P. Nason, G. Marchesini and B.R. Webber, in ‘Z Physics at LEP 1’,
CERN 89-08, vol. 1, p. 373.
HERWIG Monte Carlo: G. Abbiendi, I.G. Knowles, G. Marchesini, M.H. Seymour,
L. Stano and B.R. Webber, Cambridge preprint Cavendish-HEP-90/26 (1990),
to be published in Computer Phys. Comm.
L. Lönnblad and U. Pettersson, Lund preprint LU TP 88-15 (1988);
preprint PPE/90-196;
491.
31. S. Catani, G. Turnock and B.R. Webber, CERN preprint TH.6231-91 to be published
in Phys. Lett. B.
34. Yu.L. Dokshitzer and D.I. Dyakonov, Phys. Lett. 84B (1979) 234;

<table>
<thead>
<tr>
<th>scheme</th>
<th>$M_{ij}^2/Q^2$</th>
<th>recombination</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>$(p_i + p_j)^2/Q^2$</td>
<td>$p_{ij} = p_i + p_j$</td>
</tr>
<tr>
<td>JADE</td>
<td>$\epsilon_i = E_i/Q$</td>
<td>$p_{ij} = p_i + p_j$</td>
</tr>
<tr>
<td>E0</td>
<td>$\epsilon_i = E_i/Q$</td>
<td>$E_{ij} = E_i + E_j$</td>
</tr>
<tr>
<td>P</td>
<td>$2\epsilon_i(1 - \cos \theta_{ij})$</td>
<td>$\epsilon_i = E_i/Q$</td>
</tr>
<tr>
<td>P0</td>
<td>$\epsilon_i = E_i/\sum_k E_k$</td>
<td>$p_{ij} = p_i + p_j$</td>
</tr>
</tbody>
</table>

Table 1 - Recombination procedures for the JADE algorithm: massive (E) and massless (JADE, E0, P, P0) recombination schemes.
Table 2 - Leading and next-to-leading two loop coefficients for the JADE algorithm in the E-recombination scheme.

<table>
<thead>
<tr>
<th>n</th>
<th>$A_2^{(2)}$</th>
<th>$A_1^{(2)}$</th>
<th>$B_1^{(2)}$</th>
<th>$B_2^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$-C_F$</td>
<td>$-\frac{3}{2}C_F$</td>
<td>$(\frac{1}{4} - \frac{1}{18})C_F^2$</td>
<td>$-\frac{3}{4}C_F^2 - \pi\beta_0 C_F$</td>
</tr>
<tr>
<td>3</td>
<td>$+C_F$</td>
<td>$-\frac{1}{2}C_F$</td>
<td>$(-1 + \frac{1}{2})C_F^2 - \frac{1}{2}C_F C_A$</td>
<td>$+3C_F^2 + \frac{1}{2}C_F C_A + \frac{1}{2}\pi\beta_0 C_F$</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>$(\frac{1}{4} - \frac{1}{18})C_F^2 + \frac{1}{18}C_F C_A$</td>
<td>$-\frac{3}{4}C_F^2 - \frac{1}{2}C_F C_A - \frac{1}{2}\pi\beta_0 C_F$</td>
</tr>
</tbody>
</table>

Table 3 - Leading and next-to-leading two loop coefficients for the JADE algorithm in the massless recombination schemes.

<table>
<thead>
<tr>
<th>n</th>
<th>$A_2^{(2)}$</th>
<th>$A_1^{(2)}$</th>
<th>$B_1^{(2)}$</th>
<th>$B_2^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$-C_F$</td>
<td>$\frac{3}{2}C_F$</td>
<td>$(\frac{1}{4} + \frac{1}{18})C_F^2$</td>
<td>$-\frac{3}{4}C_F^2 - \pi\beta_0 C_F$</td>
</tr>
<tr>
<td>3</td>
<td>$+C_F$</td>
<td>$-\frac{1}{2}C_F$</td>
<td>$(-1 - \frac{1}{2})C_F^2 - \frac{1}{2}C_F C_A$</td>
<td>$+3C_F^2 + \frac{1}{2}C_F C_A + \frac{1}{2}\pi\beta_0 C_F$</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>$(\frac{1}{4} - \frac{1}{18})C_F^2 + \frac{1}{18}C_F C_A$</td>
<td>$-\frac{3}{4}C_F^2 - \frac{1}{2}C_F C_A - \frac{1}{2}\pi\beta_0 C_F$</td>
</tr>
</tbody>
</table>

Table 4 - Leading and next-to-leading two loop coefficients for the $A_F$ algorithm in any (E, E0, F, F0) recombination scheme.

<table>
<thead>
<tr>
<th>n</th>
<th>$A_2^{(2)}$</th>
<th>$A_1^{(2)}$</th>
<th>$B_1^{(2)}$</th>
<th>$B_2^{(2)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$-C_F$</td>
<td>$\frac{3}{2}C_F$</td>
<td>$\frac{1}{2}C_F$</td>
<td>$-\frac{3}{2}C_F^2 - \pi\beta_0 C_F$</td>
</tr>
<tr>
<td>3</td>
<td>$+C_F$</td>
<td>$-\frac{1}{2}C_F$</td>
<td>$-\frac{3}{2}C_F - \frac{1}{2}C_F C_A$</td>
<td>$+3C_F^2 + \frac{1}{2}C_F C_A + \frac{1}{2}\pi\beta_0 C_F$</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{2}C_F^2 + \frac{1}{2}C_F C_A$</td>
<td>$-\frac{3}{2}C_F^2 - \frac{1}{2}C_F C_A - \frac{1}{2}\pi\beta_0 C_F$</td>
</tr>
</tbody>
</table>

Table 5 - Leading and next-to-leading two loop coefficients for the Geneva algorithm.
Fig. 3

2-jet

3-jet
Fig. 5

Fig. 6