

# Infrared-safe definition of jet flavour

A. Banfi<sup>1,2,3,a</sup>, G.P. Salam<sup>4</sup>, G. Zanderighi<sup>5,6</sup>

<sup>1</sup> DAMTP, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK

<sup>2</sup> Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK

<sup>3</sup> University of Milano–Bicocca and INFN Sezione di Milano, Italy

<sup>4</sup> LPTHE: Université Pierre et Marie Curie – Paris 6; Université Denis Diderot – Paris 7; CNRS; 75252 Paris 75005, France

<sup>5</sup> Theory Group, Fermilab, P.O. Box 500, Batavia, IL, USA

<sup>6</sup> Theory Division, Physics Department, CERN, 1211 Geneva 23, Switzerland

Received: 19 January 2006 / Revised version: 10 March 2006 /

Published online: 19 May 2006 – © Springer-Verlag / Società Italiana di Fisica 2006

**Abstract.** It is common, in both theoretical and experimental studies, to separately discuss quark and gluon jets. However, even at parton level, widely-used jet algorithms fail to provide an infrared-safe way of making this distinction. We examine the origin of the problem and propose a solution in terms of a new “flavour- $k_t$ ” algorithm. As well as being of conceptual interest this can be a powerful tool when combining fixed-order calculations with multi-jet resummations and parton showers. It also has applications to studies of heavy-quark jets.

## 1 Introduction

A search through the SPIRES database reveals over 350 articles whose titles contain the expressions “quark jet(s)” or “gluon jet(s)”<sup>1</sup> The idea of quark and gluon jets appears so intuitive that it hardly seems necessary to examine the question of what it means. Yet, when going beyond leading-order perturbative QCD, the concept of quark and gluon jets is only meaningful once a procedure has been defined to classify an ensemble of partons into a set of jets, each with a well-defined flavour – a flavour that is insensitive to the addition of extra soft or collinear branchings. To our knowledge the theoretical question of how to do this has not been addressed in the literature.

In contrast there has been much experimental work on the identification of jet flavour. Some studies attempt to extract separate unbiased samples of quark and gluon-induced jets so as to examine the differences between them. For example in three-jet ( $q\bar{q}g$ ) events in  $e^+e^-$ ,  $b$ -tagging and knowledge of leading-order QCD kinematic distributions can be used to determine which is the gluon jet [1]. While successful, such a procedure applies only to  $e^+e^- \rightarrow q\bar{q}g$  events and cannot be readily generalised to more complex events, e.g. in hadron–hadron collisions. Another kind of study exploits known differences between typical quark

and gluon jets (notably their different distributions for the jet profile and subjet multiplicity) in order to tag the two kinds of jets [2]. This can be done in any process and is quite promising as an operative discriminator in real events, but does not provide an infrared-safe definition of the flavour: it defines a quark jet as being one that is sufficiently narrow and with a sufficiently small subjet multiplicity – so in the limit of there being only soft and collinear radiation, a gluon jet will always be misclassified as a quark jet. This means that the definition has no sense at fixed perturbative order.

Other experimental studies are carried out in the context of processes in which each jet originates from a quark (e.g. in  $e^+e^- \rightarrow q\bar{q}$ ) and try to identify the specific flavour or charge of that quark, generally using the hardest hadron in the jet (e.g. [3, 4]), or a momentum-weighted sum of charges (e.g. [5, 6]). Such approaches are collinear unsafe because the momentum of the leading quark can be significantly modified by collinear splittings. In practice this, together with hadronisation, means that the measured flavour, or charge, has only a limited correlation with that of the original quark. This correlation cannot be calculated theoretically, though it can be modelled and/or measured, as was done for the determinations of forward–backward asymmetries [3, 5].

In this article we will restrict our attention to the most basic theoretical question about jet flavour – that of how to define it for partonic events. As well as being of intrinsic interest, this is becoming of increasing practical importance as the study of QCD is extended to multi-jet ensembles (by jets we mean both incoming and outgoing ones): in studies of  $e^+e^- \rightarrow$  jets one knows that the basic two-jet Born

<sup>a</sup> e-mail: andrea.banfi@mib.infn.it

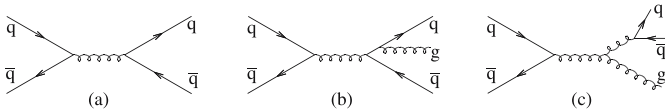
<sup>1</sup> It would be excessive to cite them all here. Instead the reader may examine them by submitting the spires search “FIND T jet# AND (T quark OR T gluon)” and then extracting from the results only those articles whose title contains “quark jet(s)” or “gluon jet(s)”.

configuration consists of quark jets; but for jet production at hadron colliders, the Born configuration involves two incoming and two outgoing jets and many flavour channels are possible:  $q\bar{q} \rightarrow q\bar{q}$ ,  $q\bar{q} \rightarrow gg$ ,  $gg \rightarrow gg$ , etc. The ability to assign flavours to the jets is especially useful when combining fixed-order predictions with all-order calculations (be it for parton showers as in [7] or for analytical resummations [8–10]). This is because all-order calculations are carried out for a fixed Born configuration, with a single flavour channel at a time, while fixed-order calculations implicitly sum over all flavour channels and can at best be split up a posteriori to match onto the individual flavour channels of the all-order calculation.

As a concrete example, consider the calculation of higher-order corrections to the process  $q\bar{q} \rightarrow q\bar{q}$ , Fig. 1a. An all-order calculation treats the addition of any number of soft/collinear gluons and extra  $q\bar{q}$  pairs implicitly, leaving the underlying  $2 \rightarrow 2$  flavours unchanged. When trying to supplement this with results of a fixed-order calculation one encounters the problem that higher-order contributions cannot be uniquely assigned to any given  $2 \rightarrow 2$  flavour channel – the  $\mathcal{O}(\alpha_s)$  corrections to  $q\bar{q} \rightarrow q\bar{q}$  include e.g. a  $q\bar{q} \rightarrow q\bar{q} \rightarrow q\bar{q}g$  piece, but a fixed-order calculation gives only the squared sum of all  $q\bar{q} \rightarrow q\bar{q}g$  diagrams, among them  $q\bar{q} \rightarrow q\bar{q} \rightarrow q\bar{q}g$  and  $q\bar{q} \rightarrow gg \rightarrow q\bar{q}g$ , illustrated in Fig. 1b and c respectively. There can exist no unambiguous procedure for separating the  $q\bar{q} \rightarrow q\bar{q}g$  contribution into its different underlying channels, both because the different channels are not individually gauge invariant and because they interfere when squaring the amplitude.

One therefore needs a prescription to assign  $q\bar{q} \rightarrow q\bar{q}g$  either to the  $q\bar{q} \rightarrow q\bar{q}$  or the  $q\bar{q} \rightarrow gg$  underlying Born  $2 \rightarrow 2$  process (or else to declare it irreducibly  $2 \rightarrow 3$ -like), it only being in the  $q\bar{q} \rightarrow q\bar{q}$  case that one needs to put it together with the  $q\bar{q} \rightarrow q\bar{q}$  all-order calculation. This reclassification of a  $2 \rightarrow 3$  event as a  $2 \rightarrow 2$  event is similar conceptually to what is done in a normal jet algorithm, except that not only should the momenta of the resulting  $2 \rightarrow 2$  configuration be infrared and collinear safe, but so should the flavours. Accordingly we call it a jet-flavour algorithm.

An obvious approach to defining jet flavours at the perturbative level would be to start with an existing jet algorithm, such as the  $k_t$ -clustering [11–13] or cone [14] algorithm, that defines jets such that each particle belongs to at most one jet. One can then determine the net flavour content of each of the jets, as the total number of quarks minus antiquarks for each quark flavour. Jets with no net flavour are identified as gluon jets, those with (minus) one unit of net flavour are (anti) quark jets, while those with more than



**Fig. 1.** **a** Specific  $q\bar{q} \rightarrow q\bar{q}$  flavour channel for a  $2 \rightarrow 2$  parton scattering process; **b** higher-order diagram that can be seen as a correction to **a**; **c** higher-order diagram that can be seen as a correction to the process  $q\bar{q} \rightarrow gg$ , but with the same final-state partons as **b**

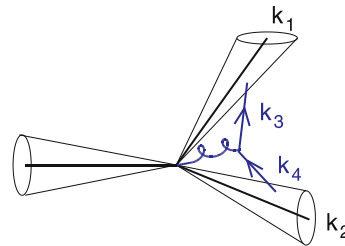
one unit of flavour (or both a flavour and a different anti-flavour) cannot be identified with a single QCD parton.

Applied to the  $k_t$  or cone algorithms, this procedure yields a jet flavour that is infrared (IR) safe at (relative) order  $\alpha_s$  discussed in our example above. However at (relative) order  $\alpha_s^2$  a large-angle soft gluon can split into a widely separated soft  $q\bar{q}$  pair and the  $q$  and  $\bar{q}$  may end up being clustered into different jets, “polluting” the flavour of those jets; see Fig. 2. Because this happens for arbitrarily soft gluons branching to quarks, the resulting jet flavours are infrared unsafe from order  $\alpha_s^2$  onwards. We are not aware of this problem having been discussed previously in the literature, though there do exist statements that are suggestive of IR safety issues when discussing flavour [15].

In Sect. 2 we shall discuss IR flavour unsafety with respect to the  $k_t$  (or “Durham”) algorithm in  $e^+e^-$  [11]. There we shall recall that the  $k_t$  closeness measure is specifically related to the divergences of QCD matrix elements when producing soft and collinear gluons. However there are no divergences for the production of soft quarks and, as we shall see, it is the use for quarks of a distance measure designed for gluons that leads to the infrared unsafety of jet flavour in the  $k_t$  algorithm. By taking into account the absence of a soft-quark divergence when designing the jet-clustering distance measure, one can eliminate the infrared divergence of the jet flavour.

The essence of the modification to the  $k_t$  distance is that instead of the  $\min(E_i^2, E_j^2)$  factor that appears usually, one needs to use  $\max(E_i^2, E_j^2)$  when the softer of  $i, j$  is a quark. In Sect. 3 we will examine how this can be extended to processes with incoming hadrons. There the added difficulty is the need for a particle-beam distance measure. Traditionally this involves only one dimensionful scale, related to the squared transverse momentum  $k_{t_i}^2$  of the particle. There is a sense in which this can be understood as  $\min(k_{t_i}^2, k_{t_B}^2)$ , where  $k_{t_B}^2$  is some transverse scale associated with the beam that is larger than all  $k_{t_i}^2$  and so could up to now be ignored. In order to obtain a sensible jet-flavour algorithm we shall however need to consider also  $\max(k_{t_i}^2, k_{t_B}^2)$  and therefore in Sect. 3 we shall investigate how to construct sensible “beam scales”.

As well as explaining how to build jet algorithms that provide an infrared-safe jet flavour, we shall also examine how they fare in practice. In  $e^+e^-$  it will be possible to carry out tests both with an NLO code (which explicitly reveals the IR unsafety of flavour in traditional jet algo-



**Fig. 2.** A large-angle soft gluon splitting to a large-angle soft  $q\bar{q}$  pair ( $k_3, k_4$ ) with the  $q$  and  $\bar{q}$  then clustered into different jets ( $k_1, k_2$ )

gorithms) and with parton-shower Monte Carlo codes. For hadron–hadron collisions only parton-shower Monte Carlo tests will be possible because none of the currently available NLO codes provides access to the final-state parton flavour information.

## 2 The $k_t$ -flavour algorithm for $e^+e^-$

The aim of clustering algorithms is to recombine particles into jets in a manner that approximates the inverse of the nearly probabilistic picture of ordered QCD branching. Since, however, the branching itself is a quantum mechanical process, there is no unique way of inverting it for a given final ensemble of particles. What can at most be done is to design it to work correctly in limits in which the QCD branching behaves classically, e.g. when a given particle is emitted as if from a single identifiable parent. The design of good jet algorithms is therefore more a craft than a deductive science. Nevertheless certain general principles will help us identify how to extend existing jet algorithms to deal properly with flavour.

Let us start by considering the most widespread clustering algorithm, the standard  $e^+e^-$  Durham (or  $k_t$ ) algorithm [11].

1. Introduce a distance measure  $y_{ij}^{(D)}$  between every pair of partons  $i, j$ :

$$y_{ij}^{(D)} = \frac{2 \min(E_i^2, E_j^2)}{Q^2} (1 - \cos \theta_{ij}), \quad (1)$$

where  $E_i$  is the energy of particle  $i$ ,  $\theta_{ij}$  is the angle between particles  $i$  and  $j$  and  $Q$  is the centre of mass energy.

2. Find the specific  $i$  and  $j$  that correspond to the smallest  $y_{ij}^{(D)}$  and recombine them according to some recombination scheme (we shall here use the  $E$  scheme, which sums the four-momenta).
3. Repeat the procedure until all  $y_{ij}^{(D)} > y_{\text{cut}}$  (or, alternatively, until one reaches a predetermined number of jets).

The defining characteristic of such clustering algorithms is the distance measure, because it determines the order in which emissions are recombined.<sup>2</sup> It is closely related to the divergences of the QCD matrix elements – for a gluon  $j$  that is soft and collinear to a gluon  $i$  the product of phase-space and matrix element for a parent gluon to branch to  $i$  and  $j$  is

$$[dk_j] |M_{g \rightarrow g_i g_j}^2(k_j)| \simeq \frac{\alpha_s C_A}{\pi} \frac{dE_j}{E_j} \frac{d\theta_{ij}^2}{\theta_{ij}^2}, \quad (E_j \ll E_i, \theta_{ij} \ll 1). \quad (2)$$

Thus with the distance measure (1), two particles are deemed to be close when either of the parameters in which

the matrix element has a divergence,  $E_j = \min(E_i, E_j)$  or  $\theta_{ij}$ , is small. This is a key characteristic of a good distance measure because where there is a strong divergence there will be many splittings that are independent of the “hard” properties of the event – such splittings should be undone (recombined) at the early stages of the clustering to leave at the end only well-separated hard pseudo-jets.

A second key characteristic of a distance measure can be understood by examining the Jade algorithm [17], which is identical to (and predates) the  $k_t$  algorithm, except that its distance measure is

$$y_{ij}^{(J)} = \frac{2E_i E_j}{Q^2} (1 - \cos \theta_{ij}). \quad (3)$$

Again, for  $E_j \ll E_i$ ,  $\theta_{ij} \ll 1$ , the distance  $y_{ij}^{(J)}$  becomes smaller when either  $E_j$  or  $\theta_{ij}$  is reduced, i.e. whenever the matrix-element divergence is made stronger. However it also becomes smaller when  $E_i$  is reduced, even though a modification of  $E_i$  has no effect on the divergence of the matrix element in (2). The undesirable consequence of this is that the Jade algorithm strongly “prefers” to recombine pairs of soft particles at large relative angle, instead of combining the individual soft particles with any collinear but harder neighbours, and so “pulls” particles out of their natural jet.

From this brief discussion, one can see that the distance measure should satisfy two main characteristics: (a) two particles should be considered close when there is a corresponding divergence in their matrix elements,<sup>3</sup> and (b) the measure should not inadvertently introduce “spurious” extra closeness for a variation of the momenta that does not lead to any extra divergence (see however discussion below (6)).

For generic hadron-level jet studies the Durham measure (1) is a good choice because the majority of emissions are gluons – the correct matrix element to consider in the design of the measure is that for soft gluon emission (be it from a quark or a gluon) and it always has both a soft (energy) and collinear (angular) divergence. For flavour algorithms one should remember that the matrix elements for  $g \rightarrow q\bar{q}$  or  $q \rightarrow qg$  (with a soft quark) have no soft divergence, but just the collinear divergence,

$$[dk_j] |M_{g \rightarrow q_i \bar{q}_j}^2(k_j)| \simeq \frac{\alpha_s T_R}{2\pi} \frac{dE_j}{E_i} \frac{d\theta_{ij}^2}{\theta_{ij}^2}, \quad (E_j \ll E_i, \theta_{ij} \ll 1) \quad (4)$$

(note the index  $i$  in the energy denominator), and analogously for  $q \rightarrow g_i q_j$ . With the  $y_{ij}^{(D)}$  measure, (1), a branching that produces a soft quark,  $E_j \ll E_i$ , has the same closeness as in the case of the gluon – however this closeness is now spurious because, in contrast to the gluon-emission case, there is no divergence for  $E_j \rightarrow 0$ . The replacement of

<sup>2</sup> There exist also jet algorithms in which the measure that determines the order of recombination differs from that defining the stopping point for recombination, e.g. the Cambridge and Aachen algorithms [16].

<sup>3</sup> This discussion is somewhat of an oversimplification – for example the angular ordered Durham algorithm [16] retains only the angular part of the closeness measure and nonetheless behaves sensibly.

the  $E_j$  denominator in the gluon-emission case, (2), with  $E_i$  in the “soft-quark” emission case, (4), suggests that the closeness measure for soft  $g \rightarrow q\bar{q}$  branching should become  $2 \max(E_i^2, E_j^2)/Q^2(1 - \cos \theta_{ij})$ . A similar argument holds in the case of  $q \rightarrow g_i q_j$  with  $E_j \ll E_i$ . Thus we should use a distance measure that depends on the flavours of the particles being considered:

$$y_{ij}^{(F)} = \frac{2(1 - \cos \theta_{ij})}{Q^2} \times \begin{cases} \max(E_i^2, E_j^2), & \text{softer of } i, j \text{ is flavoured,} \\ \min(E_i^2, E_j^2), & \text{softer of } i, j \text{ is flavourless,} \end{cases} \quad (5)$$

where the softer of  $i, j$  is the one with the smaller energy and where we use the terms flavoured and flavourless rather than quark-like and gluon-like so as to allow also for situations with diquarks or other multi-flavoured objects. With (5) soft-quark “emission” leads to no smaller a distance measure than non-soft-quark emission, in accord with the absence of a soft divergence for quark emission. Furthermore if a quark is to recombine with a harder particle it will favour one that is not too hard, in accord with the presence of  $\max(E_i, E_j)$  in the denominator of (4), which implies that the harder the parent, the less likely it is that it will produce a quark of a given softness.

With such a distance measure, for configurations as in Fig. 2 the soft  $q$  and  $\bar{q}$  will have similar energies,  $E_3 \sim E_4 \ll Q$ . Thus  $y_{13} \sim y_{14} \sim y_{23} \sim y_{24} \sim 1$ , whereas  $y_{34} \sim E_3^2/Q^2 \ll 1$ . So independently of the precise (large) angles of the soft  $q\bar{q}$  pair, 3 and 4, it is that soft pair that will recombine first to give a gluon-like pseudo-jet  $g$ . This will have  $y_{1g} \sim y_{2g} \sim E_2^2/Q^2$  and now the soft gluon pseudo-jet will recombine with either 1 or 2 (which one depends on the angles) and the net flavour of the hard particles will remain unchanged. Therefore, at order  $\alpha_s^2$ , our new measure correctly eliminates the soft flavour-changing divergence that exists for the plain Durham algorithm.

Sometimes in the above algorithm a quark can be recombined with another quark or with an antiquark of a different flavour. This can happen for example if there are two large-angle  $q\bar{q}$  pairs. As long as the resulting “doubly-flavoured” object is treated in the same way as a quark in the definition of  $y_{ij}^{(F)}$ , the algorithm will remain infrared safe, because in the subsequent clustering steps there will be a strong preference for recombining the multiply-flavoured object with other objects of similar softness, until all soft large-angle multiply-flavoured objects combine between themselves to produce gluon-like objects (these then recombine normally with the hard partons).

One may wish to avoid the appearance of multiply-flavour pseudo-jets altogether, since they cannot be associated with QCD partons. This can be achieved by vetoing any recombination that would lead to a multiply-flavoured object, i.e. by replacing step 2 with

2. (bland) Find the specific  $i$  and  $j$  that correspond to the smallest  $y_{ij}^{(F)}$  among those combinations of

$i$  and  $j$  whose net flavour corresponds either to an (anti)quark or a gluon, and recombine them.

We call this a “bland” variant of the jet-flavour algorithm, since “excessively-flavoured” clusterings are forbidden. We note that a blandness requirement on clusterings has been discussed also in [7] (though a simple “bland” Durham algorithm with the standard  $y_{ij}^{(D)}$  remains infrared unsafe).

An interesting question is that of how much freedom exists in the definition of the distance measure for a flavour algorithm. Returning to the analysis of Fig. 2 the main requirement for infrared flavour safety is that the soft fermions 3 and 4 should recombine between themselves before recombining with harder particles. This property is maintained for the following class of distance measures:<sup>4</sup>

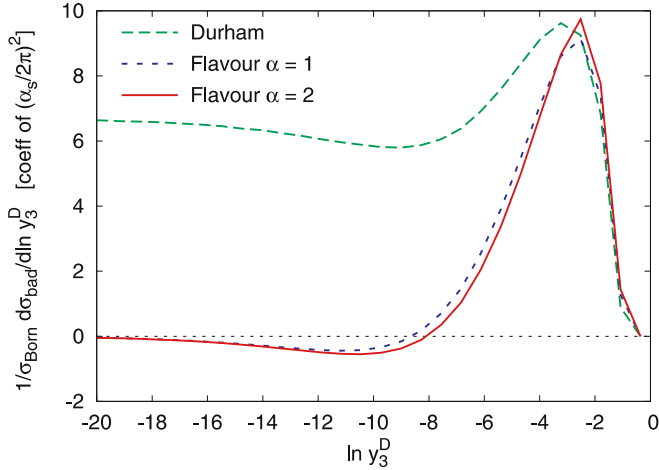
$$y_{ij}^{(F,\alpha)} = \frac{2(1 - \cos \theta_{ij})}{Q^2} \times \begin{cases} [\min(E_i, E_j)]^{2-\alpha} [\max(E_i, E_j)]^\alpha, & \text{softer of } i, j \text{ is flavoured,} \\ \min(E_i^2, E_j^2), & \text{softer of } i, j \text{ is flavourless,} \end{cases} \quad (6)$$

where  $\alpha$  is a continuous parameter in the range  $0 < \alpha \leq 2$  (so far we have implicitly discussed  $\alpha = 2$ ). Above, we stated the requirement that the distance measure should not introduce “spurious” extra closeness for a variation of the momenta that does not lead to any extra divergence. Here though, for  $\alpha < 2$  such a spurious extra closeness is present. Infrared flavour safety is nevertheless preserved, because the extra closeness is weaker than that that arises in the case of a divergence, i.e. for a soft gluon  $j$ ,  $y_{ij}$  vanishes as  $E_j^2$ , whereas for a soft quark  $j$  it only vanishes as  $E_j^{2-2\alpha}$ .

Naively it would seem that  $\alpha = 2$  should give the best identification of flavour. However there are situations where a hard quark loses energy through multiple collinear gluon emission and thus becomes a relatively soft quark. In principle there are no large ratios between the quark energy and the softest of the harder gluons it has emitted. However if that gluon is a bit harder than the quark, a value of  $\alpha < 2$  can make it easier for them to recombine. Accordingly below we shall examine both  $\alpha = 1$  and  $\alpha = 2$ .

An order-by-order rigorous test of jet-flavour algorithms can be obtained by numerically investigating the infrared safety of the jet flavour in fixed-order calculations. For example, one generates events  $e^+e^- \rightarrow q\bar{q}$  together with higher orders and clusters them to two jets. With a jet algorithm that provides a good reconstruction of the flavour, one expects that each of the two jets should have net flavour corresponding to an (anti)quark. Sometimes this does not happen — for example each of the two jets may have no net flavour, i.e. be gluon-like. This is legitimate in events in which there has been a hard branching (there is not a unique clustering to two jets), but for an infrared-safe flavour jet algorithm, the probability of this

<sup>4</sup> We consider only those that reduce to the Durham algorithm for purely gluonic ensembles of particles.



**Fig. 3.** NLO differential cross section for  $e^+e^- \rightarrow q\bar{q}$  events that after jet clustering have their flavour badly identified, i.e. identified as consisting of two gluon jets (that is, each of zero net flavour) or two jets each of net flavour larger than 1; the coefficient of  $(\alpha_s/2\pi)^2$ , as generated with Event2 [18], is plotted as a function of the Durham  $y_3$  three-jet resolution threshold; results are shown for the Durham and flavour algorithms (for two values of  $\alpha$ )

happening should vanish in the limit in which there are only soft and collinear emissions.

To measure the hardness of a given event we use  $y_3^D$ , the threshold value of the Durham jet-resolution below which the event is clustered to three jets or more.<sup>5</sup> Figure 3 shows the differential cross section at next-to-leading order (NLO, order  $\alpha_s^2$ ) for producing events in which the flavour of the two jets is badly identified. It has been obtained with Event2 [18], to our knowledge the only NLO code that provides information on the flavour of the final-state partons.<sup>6</sup> One sees that for the Durham algorithm the differential cross section for events whose jet flavour does not corresponds to  $q\bar{q}$  goes to a constant as  $\ln y_3^D$  goes to  $-\infty$ . This is the sign of the infrared unsafety of flavour identification in the Durham jet algorithm. In contrast, in our flavour algorithms (for both values of  $\alpha$ ) the corresponding cross section vanishes for  $\ln y_3^D \rightarrow -\infty$ . Detailed examination of the events with badly identified flavour at small  $y_3^D$  reveals that one of the (anti)quarks has lost nearly all of its energy to a hard splitting and goes into the same hemisphere as the other quark, i.e. identification of the event as consisting of two gluon jets is actually legitimate. Such configurations appear at order  $\alpha_s$  where their cross section

is  $d\sigma_1^{\text{bad}}/d\ln y_3 \sim \alpha_s \sqrt{y_3^D}$ . At NLO, Sudakov suppression of an extra soft gluon leads to a contribution

$$\begin{aligned} \frac{d\sigma_2^{\text{bad}}}{d\ln y_3} &\simeq -\frac{\alpha_s}{2\pi} \left( \frac{C_A}{2} + \frac{C_F}{4} \right) \ln^2 y_3^D \cdot \frac{d\sigma_1^{\text{bad}}}{d\ln y_3} \\ &\sim \alpha_s^2 \sqrt{y_3^D} \ln^2 y_3^D, \end{aligned} \quad (7)$$

which is found to be consistent with the observed numerical results, thus confirming the interpretation given above for the origin of the small fraction of  $g\bar{g}$ -like events at small  $y_3^D$ .

Given that one of the possible applications of jet-flavour algorithms is in the merging of matrix-element and parton-shower calculations, we also wish to examine how flavour algorithms behave for Monte Carlo generated parton-level ensembles of quarks and gluons. This is interesting for various other reasons too: Monte Carlo generators produce multiple soft and collinear gluon emissions and  $g \rightarrow q\bar{q}$  splittings, so they are more likely to “stress-test” a jet-flavour algorithm; also we can study a much wider variety of processes with them – for example one can simulate a fake  $e^+e^- \rightarrow g\bar{g}$  to examine jet-flavour algorithms in a simple gluonic context; one can also easily use them for studies of hadron-hadron events (next section) where currently none of the NLO programs gives direct access to information on the flavour of the outgoing partons.

While Monte Carlo event generators provide considerable flexibility, it can be difficult to interpret their results. For example infrared unsafety of the flavour in fixed-order programs manifests itself as a non-vanishing probability of misidentification of the flavour as  $y_3^D \rightarrow 0$ . With an event generator one is instead likely to see this probability vanishing with an anomalous dimension, e.g.  $(y_3^D)^{c\alpha_s}$  where  $c$  is some coefficient (assuming, for the purpose of the discussion, fixed coupling).

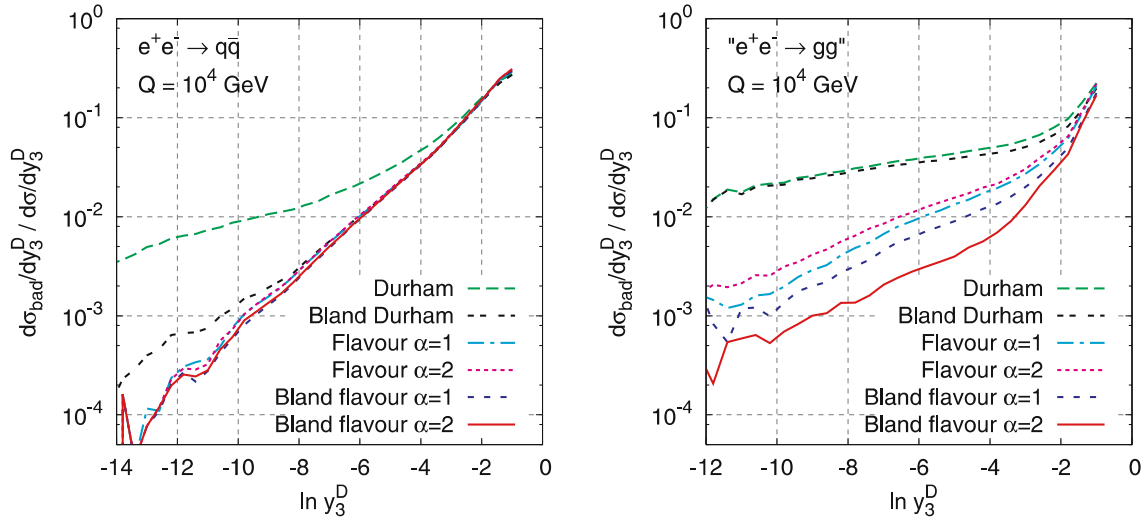
For an infrared-safe jet flavour one expects that for the clustered jets to have a different flavour from the Born channel there should have been a hard branching, as in the discussion above for the NLO  $e^+e^-$  calculation. This would lead to flavour misidentification vanishing as  $(y_3^D)^d$  where  $d$  is some pure number (above,  $d = 1/2$ ). This too may however be modified by an anomalous dimension, becoming for example  $(y_3^D)^{d+e\alpha_s}$  where  $e$  is some further pure number.<sup>7</sup>

In the presence of anomalous dimensions it is difficult to establish from Monte Carlo events exactly which functional form one is seeing. Yet another complication is that Monte Carlo event generators often do not contain the full structure of soft large-angle divergences, so that in any case the anomalous dimensions observed may not correspond to the true ones.

<sup>5</sup> Any other global event-shape-like variable that measures the departure from two jets could equally well have been used – the only requirement is that for consistency in comparing the flavour behaviour of different jet algorithms one always use a common measure for determining the hardness of the event.

<sup>6</sup> In the default version of Event2 there were subtraction terms that had contributions from final states with different flavours – for our studies here we split those subtraction terms so that each one corresponded to a unique set of final-state flavours.

<sup>7</sup> One kind of diagram that leads to flavour misidentification is that in which a hard quark loses most of its momentum by repeated gluon emission, and ends up in the opposite jet. This is similar to non-singlet small- $x$  quark production in parton distribution functions, known to be enhanced by an all-order double logarithmic series [19]. Such a series might also appear in the jet-flavour case, leading to a more complex modification of the naive  $(y_3^D)^d$  behaviour than stated in the main text.



**Fig. 4.** Fraction of events (generated by Herwig [20] at parton level) whose flavour is badly identified by various jet algorithms, shown as a function of the Durham  $y_3^D$  jet resolution threshold; a large value of  $Q$  has been chosen for illustrational purposes, so as to provide a correspondingly large range in  $y_3^D$ ; the *left-hand plot* shows results for  $e^+e^- \rightarrow q\bar{q}$ , while the *right-hand plot* shows a fake “ $e^+e^- \rightarrow gg$ ” process as generated by Herwig (code = 107)

Despite these complications, for an infrared-safe jet-flavour algorithm one expects flavour misidentification to vanish visibly faster as  $y_3^D \rightarrow 0$  than for the infrared unsafe case. This signal can be made clearer by going to large  $Q$  so as to have access to a large range in  $y_3^D$  (note though that a large value of  $Q$  also “stresses” the jet-flavour algorithm, since it increases the phase space for extra soft  $q\bar{q}$  production). Fig. 4 (left) shows the fraction of events, for each  $y_3^D$  value, where the flavour has been misidentified in various jet algorithms. It has been generated for  $Q = 10^4$  GeV, using Herwig [20] (chosen because it provides default access also to a fake  $e^+e^- \rightarrow gg$  reaction, code 107).

One sees clearly different  $y_3^D$  dependences for the Durham versus the flavour jet algorithms, with the flavour jet algorithm misidentification vanishing considerably more rapidly (actually as  $\sqrt{y_3^D}$ ). Here all the flavour algorithms behave similarly. Note also that the bland Durham algorithm works considerably better than the plain Durham algorithm and only at very small  $y_3^D$  values does one see it doing worse than the flavour algorithms: for the bland algorithm to generate a wrong-flavour event there must be a soft  $q\bar{q}$  pair of the same flavour as the hard  $q\bar{q}$ , and additionally the directions of the soft  $q\bar{q}$  must be such as to lead to jets with net gluon flavour rather than diquark flavour.

This situation changes in the right-hand plot of Fig. 4, where we consider fake  $e^+e^- \rightarrow gg$  events. Here the bland Durham algorithm behaves almost identically to the normal Durham algorithm. This is expected, since a soft  $q\bar{q}$  pair encounters no blandness problems when contaminating the flavour of gluon jets. The flavour algorithms all work systematically better than the Durham-based algorithms, clearly vanishing faster with  $y_3^D$ . One sees differences in normalisation between the different flavour algorithms and the blandness requirement provides a non-

negligible advantage, especially for  $\alpha = 2$ . This implies that the flavour misidentification involves more than one  $q\bar{q}$  pair. Nevertheless, the algorithm remains infrared safe even for multiple soft or collinear  $q\bar{q}$  pairs, as discussed above<sup>8</sup> (see also the appendix for a more general outline of the discussion of IR safety).

### 3 Jet-flavour algorithms for hadron–hadron collisions

For hadron–hadron collisions (and DIS) the  $k_t$  jet algorithm is similar to that described in Sect. 2, with a few modifications in the definition of the distances [12, 13]. Given that there is no unique hard scale  $Q$ , instead of examining dimensionless  $y_{ij}$ ’s one looks at dimensionful  $d_{ij}$ ’s. These need to be invariant under longitudinal boosts and the most widespread convention is to take

$$d_{ij} = \min(k_{ti}^2, k_{tj}^2)(\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2), \quad (8)$$

where  $\Delta\eta_{ij} = \eta_i - \eta_j$ ,  $\Delta\phi_{ij} = \phi_i - \phi_j$  and  $k_{ti}$ ,  $\eta_i$  and  $\phi_i$  are respectively the transverse momentum, rapidity and azimuth of particle  $i$ , with respect to the beam. A particle  $i$  can also recombine with the beam and here too one needs a distance measure, usually taken to be

$$d_{iB} = k_{ti}^2. \quad (9)$$

It is the smallest of the  $d_{iB}$  and the  $d_{ij}$  that determines which recombination takes place. If it is  $d_{iB}$  that is smallest at a given step, then  $i$  recombines with the beam (or

<sup>8</sup> Note though that for a fixed degree of softness, the presence of multiple  $q\bar{q}$  pairs, spread densely in rapidity from large-angles all the way to the hard-fragmentation region can lead to a systematic worsening of the flavour identification.

else gets called a jet, in the “inclusive” version of the algorithm).

The modification of the  $d_{ij}$  needed to obtain a flavour-safe jet algorithm is directly analogous to that used for the  $e^+e^-$  algorithm:

$$d_{ij}^{(F)} = (\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2) \times \begin{cases} \max(k_{ti}^2, k_{tj}^2), & \text{softer of } i, j \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tj}^2), & \text{softer of } i, j \text{ is flavourless,} \end{cases} \quad (10)$$

where by “softer” we now mean that having lower  $k_t$  and where temporarily, for simplicity, we consider only the case  $\alpha = 2$ .

It is less obvious how to modify the beam distance. The problem is that  $d_{iB}$  involves just a single scale,  $k_{ti}^2$ , and so there is no “minimum” that one can replace with a “maximum”. However one could imagine that  $d_{iB}$  is actually the minimum of  $k_{ti}^2$  and some transverse scale associated with the beam,  $k_{tB}^2$ , which has never been explicitly needed so far because it was always larger than any of the  $k_{ti}^2$ . The analogue of (10) would then be to take

$$d_{iB}^{(F)} = \begin{cases} \max(k_{ti}^2, k_{tB}^2), & i \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tB}^2), & i \text{ is flavourless.} \end{cases} \quad (11)$$

The question that remains is how to define  $k_{tB}$ .

A first issue is that we will want to identify the flavour of each of the incoming beams. So whereas for the normal  $k_t$  algorithm one recombines particles with “the beams”, here we will need to specify which of the two beams a particle recombines with. Therefore we will need to define  $k_{tB}$  for the beam moving towards positive rapidities (right) and  $k_{t\bar{B}}$  for the other beam.

In line with the DGLAP idea [21] of logarithmic ordering, such that harder emissions are at successively larger angles with respect to the beam that produced them, it makes sense for the beam hardness to be a function of rapidity,  $k_{tB}(\eta)$ . In the definition of  $d_{iB}$ , (11), one would then use  $k_{tB}(\eta_i)$ . For the right-moving (positive rapidity) beam, one scale that appears naturally is (with  $\Theta(0) \equiv 1/2$ ),

$$P_{t,\text{right}}(\eta) = \sum_i k_{ti} \Theta(\eta_i - \eta), \quad (12)$$

i.e. the beam scale should be at least as hard as all emissions that have already occurred from that beam (i.e. all emissions that are at larger rapidity). Another scale that arises is

$$P_{\alpha,\text{left}}(\eta) = \sum_i k_{ti} e^{\eta_i - \eta} \Theta(\eta - \eta_i). \quad (13)$$

When one performs a Sudakov decomposition of all momenta  $k_i = \alpha_i P + \beta_i \bar{P} + k_{ti}$  ( $P = (1, 0, 0, 1)$  and  $\bar{P} = (1, 0, 0, -1)$ ), in the massless approximation, this scale is just the sum of the  $\alpha_i = k_{ti} e^{\eta_i}$  components of all particles that are still to be emitted by this beam (i.e. are at smaller rapidity). It is equivalent to the light-cone momentum still left in the beam. This scale depends on the reference frame,

but can be transformed into a boost invariant, local “transverse” hardness by multiplying it by  $e^{-\eta}$ , giving<sup>9</sup>

$$P_{t\alpha,\text{left}}(\eta) = \sum_i k_{ti} e^{\eta_i - \eta} \Theta(\eta - \eta_i). \quad (14)$$

By adding the two measures,  $P_{t,\text{right}}(\eta)$  and  $P_{t\alpha,\text{left}}(\eta)$  for the beam scale, one obtains an overall beam hardness measure,

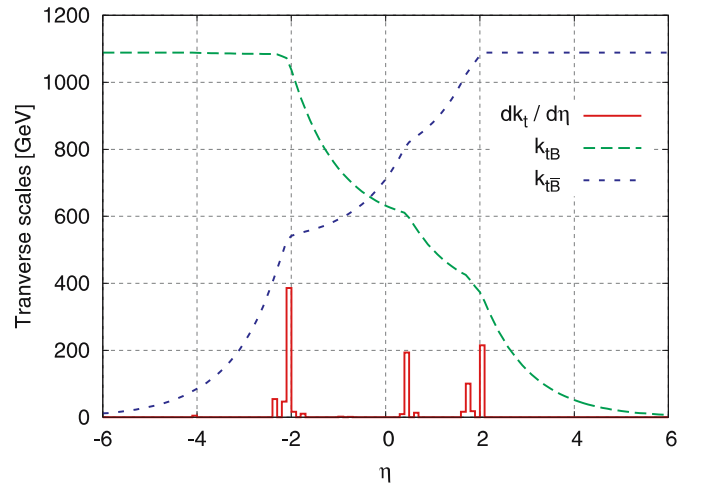
$$k_{tB}(\eta) = \sum_i k_{ti} (\Theta(\eta_i - \eta) + \Theta(\eta - \eta_i) e^{\eta_i - \eta}), \quad (15)$$

that takes into account both emissions that have already occurred at a certain rapidity (in the picture of ordering of emissions) and those that will occur further on. Similarly one defines a scale for the other beam

$$k_{t\bar{B}}(\eta) = \sum_i k_{ti} (\Theta(\eta - \eta_i) + \Theta(\eta_i - \eta) e^{\eta - \eta_i}). \quad (16)$$

In the same way that one updates the  $d_{ij}$  and  $d_{iB}$  after each clustering, one should update also the  $k_{tB}$  and  $k_{t\bar{B}}$ .

To illustrate the properties of  $k_{tB}$  and  $k_{t\bar{B}}$ , Fig. 5 shows these two quantities for a typical multi-jet LHC event (represented as a histogram of total transverse momentum per bin of rapidity). Towards positive rapidities,  $k_{tB}(\eta)$  decreases as  $e^{-\eta}$ , while  $k_{t\bar{B}}(\eta)$  approaches a constant, so that as is natural, positive-rapidity particles combine with  $B$ , while negative-rapidity particles combine with  $\bar{B}$ . At the point where  $k_{tB}$  and  $k_{t\bar{B}}$  cross, they are of the same order



**Fig. 5.** Plot of  $k_{tB}$  and  $k_{t\bar{B}}$  for a multi-jet parton-level LHC event, generated by Herwig; also shown is the histogram of the rapidity distribution of transverse momenta

<sup>9</sup> Another way of seeing how this scale arises naturally is to recall that in the non-longitudinally invariant version of the  $k_t$  algorithm for DIS and hadron-hadron collisions [22], the beam distance is  $d_{iB} = 2E_i^2(1 - \cos\theta_{iB})$ . Replacing  $E_i$  with the effective beam energy  $\frac{1}{2}P_{\alpha,\text{left}}$  (i.e. taking the larger of  $E_i$  and the effective beam energy) and taking the small-angle limit gives precisely  $P_{t\alpha,\text{left}}^2$ .

of magnitude as the total transverse momentum in the event, i.e. its overall hardness. Note also that  $k_{tB}(\eta)$  and  $k_{t\bar{B}}(\eta)$  are always at least as hard as the hardest emission at rapidity  $\eta$ .

Let us now summarise the jet-flavour algorithm for hadron–hadron collisions.

1. Introduce a distance measure  $d_{ij}^{(F)}$  between every pair of partons  $i, j$ :

$$d_{ij}^{(F,\alpha)} = (\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2) \times \begin{cases} \max(k_{ti}, k_{tj})^\alpha \min(k_{ti}, k_{tj})^{2-\alpha}, \\ \text{softer of } i, j \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tj}^2), \\ \text{softer of } i, j \text{ is flavourless,} \end{cases} \quad (17)$$

as well as distances to the two beams,

$$d_{iB}^{(F,\alpha)} = \begin{cases} \max(k_{ti}, k_{tB}(\eta_i))^\alpha \min(k_{ti}, k_{tB}(\eta_i))^{2-\alpha}, \\ i \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tB}^2(\eta_i)), \\ i \text{ is flavourless,} \end{cases} \quad (18)$$

and an analogous definition of  $d_{i\bar{B}}^{(F,\alpha)}$  involving  $k_{t\bar{B}}(\eta_i)$  instead of  $k_{tB}(\eta_i)$  (both defined as in (15) and (16)).<sup>10</sup> As in Sect. 2 we have introduced a class of measures, parametrised by  $0 < \alpha \leq 2$ .

2. Identify the smallest of the distance measures. If it is a  $d_{ij}^{(F,\alpha)}$ , recombine  $i$  and  $j$ ; if it is a  $d_{iB}^{(F,\alpha)}$  ( $d_{i\bar{B}}^{(F,\alpha)}$ ) declare  $i$  to be part of beam  $B$  ( $\bar{B}$ ) and eliminate  $i$ ; in the case where the  $d_{iB}^{(F,\alpha)}$  and  $d_{i\bar{B}}^{(F,\alpha)}$  are equal (which will occur if  $i$  is a gluon), recombine with the beam that has the smaller  $k_{tB}(\eta_i)$ ,  $k_{t\bar{B}}(\eta_i)$ .
3. Repeat the procedure until all the distances are larger than some  $d_{\text{cut}}$ , or, alternatively, until one reaches a predetermined number of jets.<sup>11,12</sup>

<sup>10</sup> The beam distances in (15) and (16) have been constructed by considering situations with just massless partons. However, their definition can be extended to cases with massive particles in the final state by replacing  $k_{ti}$  with  $\sqrt{k_{ti}^2 + m_i^2}$ . Notice that any heavy non-QCD particles should also be included in the sums (15) and (16), even if they do not enter the clustering. In DIS, in the Breit frame,  $k_{tB}(\eta)$  should include an additional contribution related to the virtual photon, given by  $Q(\Theta(\eta)e^{-\eta} + \Theta(-\eta))$ , while  $k_{t\bar{B}}(\eta)$  should have an additional contribution  $Q(\Theta(\eta) + \Theta(-\eta)e^{\eta})$ , where  $Q$  is the photon virtuality.

<sup>11</sup> Yet another possibility is to introduce separate measures for the ordering of recombinations and for the point where recombination comes to a stop, as in the Cambridge and Aachen algorithms [16].

<sup>12</sup> In light of recent work that relates the  $k_t$  algorithm to a geometrical nearest neighbour problem [23] to reduce its computational complexity to  $N \ln N$ , it is worth commenting that the simultaneous use here of both  $\min(k_{ti}^2, k_{tj}^2)$  and  $\max(k_{ti}^2, k_{tj}^2)$

In the ‘‘bland’’ variant of the algorithm one considers only those  $d_{ij}$  for which the product of the recombination would have at most one flavour. Similarly one considers only a subset of the  $d_{iB}$  – in this case the blandness requirement is imposed on the flavour of the parton entering the hard interaction, or equivalently on the difference between the flavour of the incoming hadron and the flavour contained in the outgoing beam jet.

The infrared safety of this algorithm follows from the same arguments that were used in the  $e^+e^-$  context. The beam scales simply ensure that  $q\bar{q}$  pairs that are soft but separated by  $\Delta\eta^2 + \Delta\phi^2 > 1$  recombine with each other before recombining with the beam. This eliminates the potentially dangerous situation that would otherwise occur, in which first the  $q$  recombines with one beam and then the  $\bar{q}$  recombines with the other beam. Therefore it is not just the flavours of the outgoing jets that are infrared and collinear safe, but also those of the incoming beam jets (the determination of the beam-jet flavours of course also requires knowledge of the incoming parton flavours).

A concrete demonstration of the infrared safety of the hadron–hadron algorithms, analogous to Fig. 3 for  $e^+e^-$ , is not possible with currently available tools, because none of the higher-order NLO jet codes [24, 25] provide direct access to information about final-state flavour. Even if they did, there would be an additional complication compared to  $e^+e^-$ . In  $e^+e^-$  at Born level, there is only one flavour channel, i.e.  $e^+e^- \rightarrow q\bar{q}$ . Therefore one could identify flavour infrared unsafety by examining, for example, the three-jet NLO cross section for jets classified as  $gg$ . In hadron–hadron collisions all flavour channels are present at Born order, therefore to verify the infrared safety of, say, the  $gg \rightarrow gg$  channel one must supplement the NLO  $2+3$  jet calculation with the  $gg \rightarrow gg$  Born contribution and its two loop corrections, i.e. one must carry out a NNLO  $2+2$  jet calculation, which is beyond today’s technology. Fortunately an alternative method exists for verifying the IR safety of flavour identification using just a NLO  $2+3$  jet calculation, namely by examining the cross section for doubly-flavoured jets, since these do not appear at Born level but are infrared unsafe in the plain  $k_t$  algorithm. We hope that flavour information will soon become available in  $2+3$  jet NLO codes, making it possible to demonstrate this explicitly.

In the absence of any way of obtaining a fixed-order illustration of the infrared safety of the flavour algorithms, we resort to investigations of reconstruction of the flavour in parton-level Monte Carlo events. This is achieved by comparing, event-by-event, the flavours in the hard  $2 \rightarrow 2$  partonic scattering with those of the beam and outgoing jets after clustering of the event to  $2+2$  jets.

Since the normal  $k_t$  algorithm does not usually distinguish between the two beams, we extend it (both normal

invalidates the Lemma of [23] that was central in making the connection with a nearest neighbour problem. It is therefore not clear whether it would be possible to write the flavour algorithm such that its complexity goes as  $N \ln N$ . The implementation that we use has a complexity that scales roughly as  $N^2$ .



and bland variants) such that a particle destined to recombine with “the beams” is assigned to that with the smaller of  $k_{tB}(\eta_i)$  and  $k_{t\bar{B}}(\eta_i)$ .

The proportion of events where the original and reconstructed  $2 \rightarrow 2$  flavours do not match is shown in Fig. 6, as a function of  $y_3^{kt} = d_3^{kt}/(E_{t1} + E_{t2})^2$ . Here  $d_3^{kt}$  is the threshold value of  $d_{\text{cut}}$  below which the event is clustered to three or more jets in the standard exclusive longitudinally invariant  $k_t$  algorithm [12];  $E_{t1}$  and  $E_{t2}$  are the transverse energies of the two last jets to be recombined with the beam if there is no  $d_{\text{cut}}$  [27] (equivalently the two hardest jets when running the inclusive  $k_t$  algorithm [13]). We consider simulated LHC events and require the hardest jet to have a transverse energy larger than 1 TeV and the two hardest jets to have  $|\eta| < 1$ .

Three representative channels,  $qq \rightarrow qq$  (including  $q\bar{q} \rightarrow q\bar{q}$ ),  $q\bar{q} \rightarrow gg$  and  $gg \rightarrow gg$  are shown in Fig. 6, as obtained with Herwig [20]. The standard parton showering in Pythia [28] gives similar results (with a slightly higher normalisation). We also illustrate the  $gg \rightarrow gg$  channel using the recently developed transverse-momentum ordered shower in Pythia [26]. In all cases one sees that the rate of flavour misidentification falls significantly more rapidly towards small  $y_3^{kt}$  for the flavour algorithms than for the normal  $k_t$  algorithm or its bland variant.<sup>13</sup> This is similar to what was observed for  $e^+e^-$  in Sect. 2 and is a sign of the infrared safety of the flavour algorithms.

One notes that for all algorithms the fall-off is less rapid in the hadron-hadron case than in  $e^+e^-$ . This is natural given the increased number of jets and therefore of sources of radiation which can lead to extra flavour in the final state. Another difference compared to  $e^+e^-$  is that now the  $\alpha = 1$  flavour algorithms sometimes fare better than the  $\alpha = 2$  case. This is not systematic and also depends on the Monte Carlo program used to generate events (compare Figs. 6c and d). The overall normalisation of the curves also depends on the Monte Carlo program used and one sees that Pythia with transverse-momentum ordered showers produces parton-level final states in which it is systematically harder to cluster back to the original flavour.

## 4 Outlook

We have shown in this article that it is possible to define parton-level jets in a manner that ensures that their flavour is infrared safe. The key ingredient in doing this

<sup>13</sup> It is interesting to note that the bland  $k_t$  algorithm sometimes behaves worse than the normal  $k_t$  algorithm (e.g. for  $q\bar{q} \rightarrow gg$ ). To see why this happens, consider a beam corresponding to incoming  $u$  flavour, together with a soft collinear  $u\bar{u}$  pair. In the normal  $k_t$  algorithm, the  $u$  and  $\bar{u}$  can recombine with the beam in any order. In the bland variant the  $\bar{u}$  is prevented from recombining first (because the parton entering the reaction would then implicitly have  $uu$  flavour) and if it has the lower  $k_t^2$  it will instead try to recombine with the other (wrong) beam. Therefore the bland algorithm actually has an extra source of infrared-collinear flavour unsafety relative to the plain  $k_t$  algorithm.

was a modification of the  $k_t$  distance measure, inspired by the different structures of divergences that appear in quark production and gluon production. In the case of hadron-hadron collisions it was also necessary to introduce the concept of a hardness associated with the beam at any given point in rapidity. Where possible, explicit NLO verifications confirm the infrared safety of the new “flavour” jet algorithms. Parton-level Monte Carlo studies also indicate a significant improvement in the identification of flavour relative to the  $k_t$  algorithm.

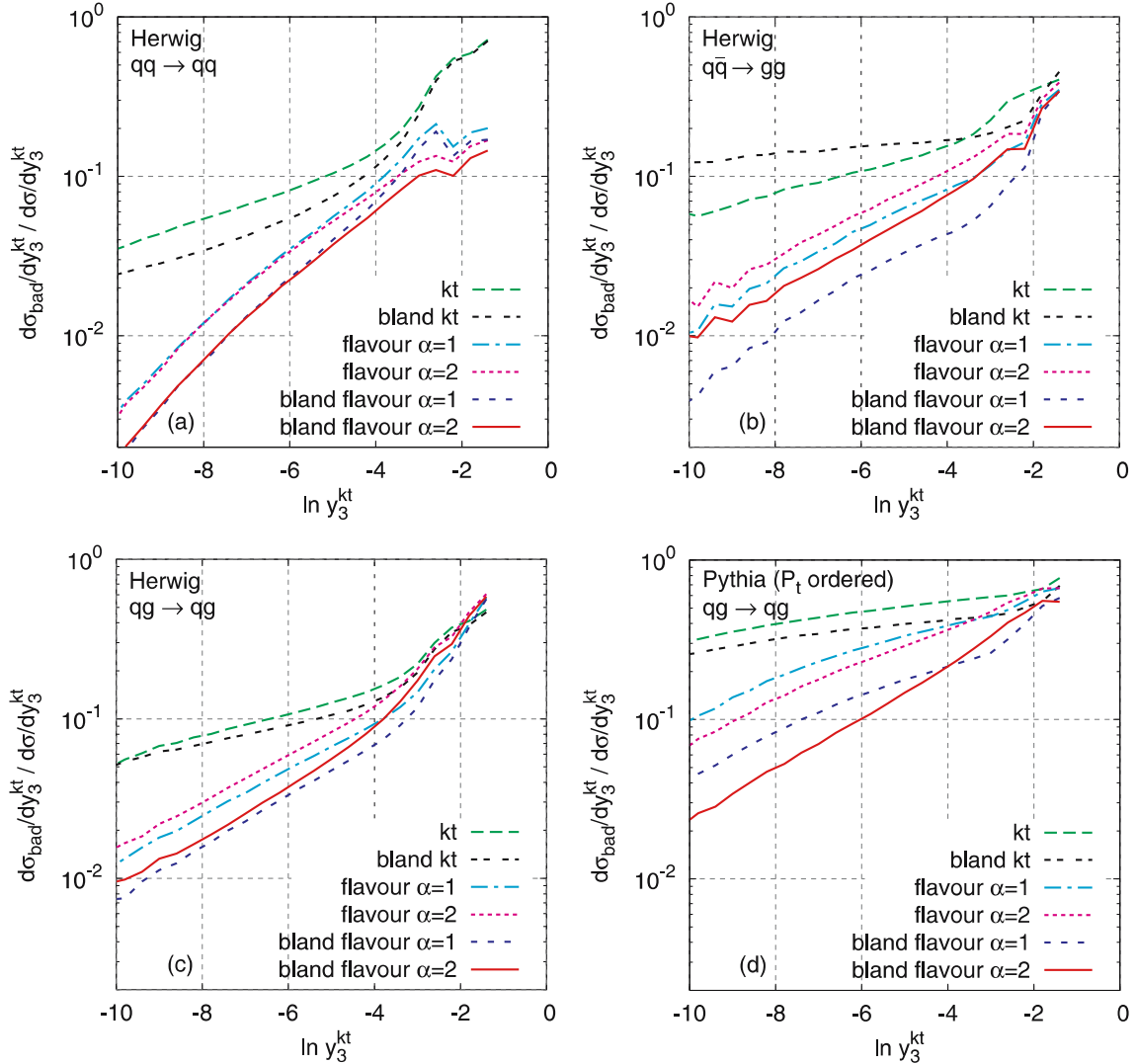
To make use of our new algorithms to accurately study jet flavour, it is necessary to have access to information about the flavour of final-state partons in NLO jet codes. Currently however, most NLO jet codes have been designed assuming that the user has no need for information about final-state parton flavour (an exception is Event2 [18]). In light of the developments presented here, we look forward to flavour information being made available in the future (e.g. [29]).

Our original motivation for studying the problem of jet flavour was the need to accurately combine resummed predictions for hadron-collider dijet event shapes [10, 27] with corresponding fixed-order predictions [24, 25]. Another simple flavour-related study would be the investigation of how the relative fractions of quark and gluon jets at hadron colliders are modified by NLO corrections and how they vary with jet transverse momentum. Apart from its intrinsic interest, such information could be of relevance also to the tuning of Monte Carlo event generators and studies of hadron multiplicities in jets, both of which are sensitive to the proportions of quark and gluon jets.

One drawback of the algorithms presented here is that, when considering light flavours, they can only be applied to partonic and not hadronic events. This is because at each recombination they require knowledge of which objects are flavoured (quark-like) rather than flavourless (gluon-like) and that information is not present in hadronic final states. It would be interesting to find a jet algorithm based purely on particle momenta, that nevertheless provides a good infrared-safe determination of the flavour at parton level. It is not clear to what extent this is possible.<sup>14</sup>

There is nevertheless one hadron-level context in which this article’s flavour algorithms could be used directly, that is for heavy-quark jets [30]. Currently a heavy-quark jet is defined as a jet containing one or more heavy quarks (or heavy-quark hadrons). The fraction of jets of transverse momentum  $p_t$  containing a heavy quark of mass  $m_Q$  is enhanced by terms  $\alpha_s^n \ln^{2n-1} p_t/m_Q$  for  $p_t \gg m_Q$ , due to the large multiplicity  $\sim \alpha_s^n \ln^{2n} p_t/m_Q$  of gluons above scale  $m_Q$ , combined with the possibility that they split collinearly  $g \rightarrow Q\bar{Q}$ , responsible for a further factor

<sup>14</sup> A candidate for a jet-flavour algorithm that does not use flavour information during the recombination sequence might be the JADE algorithm [17]. The flavour of its jets can be shown to be infrared safe in  $e^+e^-$  at  $\mathcal{O}(\alpha_s^2)$ . However it has numerous other drawbacks, which we suspect are part of the reason why in Monte Carlo studies we find that its flavour identification properties are only a little better than those of the (flavour) infrared unsafe Durham algorithm.



**Fig. 6.** The proportion of Monte Carlo events in which the flavour of one or more incoming or outgoing reconstructed parton-level jets differs from the flavour in the corresponding parton in the original hard event; shown as a function of  $\ln y_3^{k_t}$  for three channels (in the case of  $qg \rightarrow qg$  for both Herwig [20] and a recently developed parton shower algorithm in Pythia [26]); LHC kinematics are used and the events selected are those where the hardest  $k_t$ -algorithm jet has a transverse momentum greater than 1 TeV and where the two hardest jets have  $|\eta| < 1$ . The range of most common values of  $y_3^{k_t}$  depends on the subprocess but is typically roughly  $-8 \lesssim \ln y_3^{k_t} \lesssim -3$

$\alpha_s \ln p_t/m_Q$  [31–33]. Therefore, at high  $p_t$  the majority of so-called heavy-quark jets are not jets induced by a heavy quark, but rather jets in which a heavy quark has appeared from the internal branching in the jet. This implies that the current definition of heavy-quark jet will lead to large QCD backgrounds in searches for new particles which aim to tag an “intrinsic” heavy-quark jet among the decay products of the new particle.

An alternative approach to the study of heavy-quark jets would be to consider the net heavy flavour of jets,<sup>15</sup> i.e. the number of heavy-quark hadrons minus heavy an-

tiquark hadrons in a given jet.<sup>16</sup> With the cone or  $k_t$  algorithms such a definition would eliminate nearly all the final-state logarithmically enhanced terms, leaving just  $\alpha_s^n \ln^{n-1} p_t/m_Q$  contributions (involving a final-state BFKL-type resummation [35, 36]). These remaining terms come from the same diagrams that led to the infrared unsafety of light flavour of a jet. They can therefore be eliminated altogether by applying our flavour jet algorithm with the minor modification that every occurrence of “flavour” is to be replaced with “heavy flavour”. In this way it becomes possible to give meaning to a concept of

<sup>15</sup> A study that goes partially in this direction is the recent investigation of angular correlations between  $b\bar{b}$  pairs [34].

<sup>16</sup> In an event with just two heavy hadrons one need not know which one is quark-like and which antiquark-like – it suffices to know that if combined they give zero net heavy flavour.

intrinsic heavy flavour, i.e. heavy flavour that originates exclusively from the heavy-flavour component of parton distribution functions, from hard QCD flavour “creation” (e.g.  $gg \rightarrow Q\bar{Q}$ ) and from the decay of other heavier particles. We look forward to future phenomenological investigation of this concept.

*Acknowledgements.* All three of us are grateful to the Milano Bicocca theoretical physics group for hospitality during the initial stages of this work. GPS also wishes to thank Fermilab, the Cavendish Laboratory and CERN for hospitality. We thank Yuri Dokshitzer for several discussions and suggestions, Matteo Cacciari and Giuseppe Marchesini for comments, Bryan Webber for bringing to our attention the original use of the “bland Durham” algorithm in CKKW matching, Michael Seymour for assistance with Event2, Peter Skands for help with Pythia, and William Kilgore and Zoltan Nagy for information about the accessibility of flavour information in hadron-collider NLO jet codes.

This work was supported in part by grant ANR-05-JCJC-0046-01 from the French Agence Nationale de la Recherche.

## Appendix

The arguments for the infrared safety of our jet-flavour algorithms, as discussed in Sect. 2, applied only to the case of one or two extra soft  $q\bar{q}$  pairs. Here we give an outline of a general all-order discussion of infrared and collinear safety of the flavour. It will be framed in the context of  $e^+e^-$  collisions, and then in closing we will briefly mention hadronic collisions.

For a general discussion of the infrared and collinear safety of flavour one needs to examine all divergent cases in which flavour is either produced or moved from one part of the event to another. Production of flavour arises from gluon splitting. This has just a collinear divergence; additionally the gluon itself has soft and collinear divergences with respect to other quarks and gluons. Flavour can “move” during the branching process when a quark recoils due to emission of a gluon of similar hardness to the quark. This has no divergences, but there may be divergences associated with the original production of the quark itself. Flavour can also move during the jet-clustering procedure whenever a quark recombines with a parton that is not collinear to it and whose momentum is of the same order of magnitude as (or larger than) the quark.

Let us first consider flavour production by collinear splitting of a gluon. The Durham algorithm always recombines collinear particles into the same jet. Since in  $g \rightarrow q\bar{q}$  splitting there is no soft divergence, the  $q$  and  $\bar{q}$  have commensurate hardnesses. Therefore the “flavour” distance measure (5) is of the same order of magnitude as the Durham distance measure and so the  $q\bar{q}$  from a collinear splitting of a gluon will end up in the same jet also in the flavour algorithm, leaving the jet flavour unchanged as is required for IRC safety of the flavour.

Next we consider non-collinear splitting of a gluon into  $q\bar{q}$ . This has divergences when the original gluon is

collinear to some other parton and/or soft. If the gluon itself is collinear to some other parton  $a$ , angle  $\theta_{ag} \ll 1$ , then the gluon splitting to  $q\bar{q}$  is strongly suppressed unless  $\theta_{q\bar{q}} \sim \theta_{ag}$ , i.e. non-collinear splitting is not possible from a gluon that is collinear to some other parton. This is the basis of the widely-used angular ordering approximation. Therefore a  $q\bar{q}$  produced from a collinear (and optionally soft) gluon will always recombine, in the flavour algorithm as in the Durham algorithm, ensuring the safety of the flavour of any resulting jet.

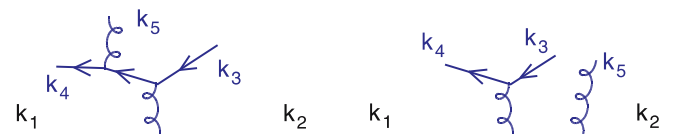
This leaves the case discussed already in the main text, in which a large-angle  $q\bar{q}$  pair is produced from a large-angle soft gluon. We have already presented the arguments that explain the IR unsafety of the Durham algorithm in this case and the IR safety of the flavour algorithms.

In generalising the analysis to higher orders one needs also to examine potential “motion” of the soft large-angle  $q$  and  $\bar{q}$ . It will be useful to introduce the compact notation  $y_{1\{2\dots n\}}$  for the set of distance measures  $y_{12}, y_{13}, \dots, y_{1n}$ .

Firstly the quark (or anti-quark) can itself emit a large-angle gluon of similar softness ( $k_5$ ), Fig. 7 (left). This will change the direction of the quark ( $k_4$ ). In the Durham algorithm, each of the  $y_{\{12\}\{345\}}$  is of the order of the soft gluon  $k_5^2/Q^2$ , and the recombination sequence depends significantly on the angles. In particular the emission of  $k_5$  from the quark may have moved it further away from the antiquark making it more likely that the soft  $q\bar{q}$  end up in different jets. In contrast, in the flavour algorithm  $y_{\{12\}\{34\}}$  are of order 1, whereas  $Y_{\{1234\}5}$  and  $y_{34}$  are of order of the soft gluon  $k_5^2/Q^2$ . Therefore 3, 4 and 5 will all recombine together first, or 5 will recombine with the hard jets and then 3 and 4 will recombine together. In both cases the flavour of the soft quarks is neutralised.

The analysis of the right-hand diagram of Fig. 7 is largely similar as long as  $k_5$  is at large angles and of the same hardness as  $k_3$  and  $k_4$ . The additional issue is that now  $k_5$  has a collinear divergence with respect to  $k_2$ . One might generally worry that semi-hard radiation collinear to  $k_2$  might pull  $k_3$  far away from its original direction. This could happen if  $k_5$  is collinear to  $k_2$  and if  $k_3$  and  $k_5$  recombine, with  $E_5 \gg E_3$  such that the recombination product ends up collinear to  $k_2$ . However if  $E_5 \gg E_3$  then  $y_{35} \gg y_{34}$  and the  $k_3$ - $k_4$  will recombine first, neutralising the flavour. Note that if  $E_5 \sim E_3$  and  $k_5$  is collinear to  $k_2$  then the  $k_2$ - $k_5$  recombination will occur first, leaving the usual (safe) configuration consisting of a soft  $q\bar{q}$  pair.

One can straightforwardly extend this analysis to multiple  $q\bar{q}$  pairs and multiple gluons. The originally soft large-angle quark can be dragged further and further towards



**Fig. 7.** Configurations in which flavour “moves” during branching and clustering, discussed in the text with regards to infrared and collinear safety

the hard jet in ensembles with multiple gluons of similar  $k_t$ 's but successively larger (but not strongly ordered) energies. However, given any fixed number of recombinations, in the soft limit the resulting quark-like object always has energy  $\ll Q$  and will recombine with the soft antiquark rather than with the hard particles.

A final comment concerns hadron–hadron collisions. There, the beam jets have a hardness  $k_{tB}(\eta)$ , which is of the same order of magnitude as any hard final-state jets that might have been emitted at the rapidity  $\eta$ . Therefore there is no difference from the point of view of IRC safety between recombination into final-state jets and into beam jets and all the arguments given here apply equally well in the hadron–hadron context.

## References

1. DELPHI Collaboration, P. Abreu et al., *Z. Phys. C* **70**, 179 (1996); ALEPH Collaboration, D. Buskulic et al., *Phys. Lett. B* **384**, 353 (1996); OPAL Collaboration, G. Abbiendi et al., *Phys. Rev. D* **69**, 032002 (2004) [hep-ex/0310048] and references therein
2. ZEUS Collaboration, S. Chekanov et al., *Nucl. Phys. B* **700**, 3 (2004) [hep-ex/0405065]
3. OPAL Collaboration, K. Ackerstaff et al., *Z. Phys. C* **76**, 387 (1997) [hep-ex/9707019]
4. ZEUS Collaboration, S. Chekanov et al., *Phys. Lett. B* **553**, 141 (2003) [hep-ex/0211025]
5. OPAL Collaboration, K. Ackerstaff et al., *Z. Phys. C* **75**, 385 (1997); ALEPH Collaboration, R. Barate et al., *Phys. Lett. B* **426**, 217 (1998); L3 Collaboration, M. Acciarri et al., *Phys. Lett. B* **439**, 225 (1998); DELPHI Collaboration, P. Abreu et al., *Eur. Phys. J. C* **9**, 367 (1999)
6. CDF Collaboration, F. Abe et al., *Phys. Rev. D* **60** (1999) 072003 [hep-ex/9903011]
7. S. Catani, F. Krauss, R. Kuhn, B.R. Webber, *JHEP* **0111**, 063 (2001) [hep-ph/0109231]
8. N. Kidonakis, G. Oderda, G. Sterman, *Nucl. Phys. B* **531**, 365 (1998) [hep-ph/9803241]
9. R. Bonciani, S. Catani, M.L. Mangano, P. Nason, *Phys. Lett. B* **575**, 268 (2003) [hep-ph/0307035]
10. A. Banfi, G.P. Salam, G. Zanderighi, *JHEP* **0503**, 073 (2005) [hep-ph/0407286]
11. S. Catani, Y.L. Dokshitzer, M. Olsson, G. Turnock, B.R. Webber, *Phys. Lett. B* **269**, 432 (1991)
12. S. Catani, Y.L. Dokshitzer, M.H. Seymour, B.R. Webber, *Nucl. Phys. B* **406**, 187 (1993)
13. S.D. Ellis, D.E. Soper, *Phys. Rev. D* **48**, 3160 (1993) [hep-ph/9305266]
14. See e.g. CDF Collaboration, F. Abe et al., *Phys. Rev. D* **45**, 1448 (1992); G.C. Blazey et al., hep-ex/0005012
15. Z. Nagy, D.E. Soper, *JHEP* **0510**, 024 (2005) [hep-ph/0503053]
16. Y.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, *JHEP* **9708**, 001 (1997) [hep-ph/9707323]; M. Wobisch, Ph.D. thesis, RWTH Aachen, 1999, DESY-THESIS-2000-049
17. JADE Collaboration, W. Bartel et al., *Z. Phys. C* **33**, 23 (1986); JADE Collaboration, S. Bethke et al., *Phys. Lett. B* **213**, 235 (1988)
18. S. Catani, M.H. Seymour, *Nucl. Phys. B* **485**, 291 (1997) [Erratum-ibid. B **510**, 503 (1997)] [hep-ph/9605323]; *Phys. Lett. B* **378**, 287 (1996) [hep-ph/9602277]
19. J. Bartels, B.I. Ermolaev, M.G. Ryskin, *Z. Phys. C* **70**, 273 (1996) [hep-ph/9507271], and references therein
20. G. Marchesini, B.R. Webber, G. Abbiendi, I.G. Knowles, M.H. Seymour, L. Stanco, *Comput. Phys. Commun.* **67**, 465 (1992); G. Corcella et al., *JHEP* **0101**, 010 (2001) [hep-ph/0011363]
21. V.N. Gribov, L.N. Lipatov, *Sov. J. Nucl. Phys.* **15**, 438 (1972); G. Altarelli, G. Parisi, *Nucl. Phys. B* **126**, 298 (1977); Y.L. Dokshitzer, *Sov. Phys. JETP* **46**, 641 (1977)
22. S. Catani, Y.L. Dokshitzer, B.R. Webber, *Phys. Lett. B* **285**, 291 (1992)
23. M. Cacciari, G.P. Salam, hep-ph/0512210
24. W.B. Kilgore, W.T. Giele, *Phys. Rev. D* **55**, 7183 (1997) [hep-ph/9610433]
25. Z. Nagy, *Phys. Rev. Lett.* **88**, 122003 (2002) [hep-ph/0110315]; *Phys. Rev. D* **68**, 094002 (2003) [hep-ph/0307268]
26. T. Sjostrand, P.Z. Skands, *Eur. Phys. J. C* **39**, 129 (2005) [hep-ph/0408302]
27. A. Banfi, G.P. Salam, G. Zanderighi, *JHEP* **0408**, 062 (2004) [hep-ph/0407287]
28. T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, E. Norrbin, *Comput. Phys. Commun.* **135**, 238 (2001) [hep-ph/0010017]
29. Z. Nagy, private communication
30. G. Grunberg, Y.J. Ng, S.H.H. Tye, *Phys. Rev. D* **21**, 62 (1980); for recent work see S. Frixione, M.L. Mangano, *Nucl. Phys. B* **483**, 321 (1997) [hep-ph/9605270]; D0 Collaboration, B. Abbott et al., *Phys. Rev. Lett.* **85**, 5068 (2000) [hep-ex/0008021]
31. A.H. Mueller, P. Nason, *Nucl. Phys. B* **266**, 265 (1986)
32. M.L. Mangano, P. Nason, *Phys. Lett. B* **285**, 160 (1992)
33. M.H. Seymour, *Z. Phys. C* **63**, 99 (1994)
34. CDF Collaboration, D. Acosta et al., *Phys. Rev. D* **71**, 092001 (2005) [hep-ex/0412006]
35. G. Marchesini, A.H. Mueller, *Phys. Lett. B* **575**, 37 (2003) [hep-ph/0308284]
36. G. Marchesini, E. Onofri, *JHEP* **0407**, 031 (2004) [hep-ph/0404242]