The Catchment Area of Jets

Matteo Cacciari, Gavin P. Salam

LPTHE
UPMC Université Paris 6,
Université Paris Diderot – Paris 7,
CNRS UMR 7589, Paris, France

Gregory Soyez

Brookhaven National Laboratory, Upton, NY 11973, USA

Abstract

The area of a jet is a measure of its susceptibility to radiation, like pileup or underlying event (UE), that on average, in the jet's neighbourhood, is uniform in rapidity and azimuth. In this article we establish a theoretical grounding for the discussion of jet areas, introducing two main definitions, passive and active areas, which respectively characterise the sensitivity to pointlike or diffuse pileup and UE radiation. We investigate the properties of jet areas for three standard jet algorithms, k_t, Cambridge/Aachen and SISCone. Passive areas for single-particle jets are equal to the naive geometrical expectation πR^2, but acquire an anomalous dimension at higher orders in the coupling, calculated here at leading order. The more physically relevant active areas differ from πR^2 even for single-particle jets, substantially so in the case of the cone algorithms like SISCone with a Tevatron Run-II split–merge procedure. We compare our results with direct measures of areas in parton-shower Monte Carlo simulations and find good agreement with the main features of the analytical predictions. We furthermore justify the use of jet areas to subtract the contamination from pileup.

Contents

1 Introduction

For nearly three decades now, jets [1] have represented the principal tool for accessing information about an event's partonic hard-scattering structure and kinematics. As a result of this, much work has been carried out on understanding the properties of jets in a range of collider contexts, addressing issues such as jet substructure [2, 3, 4], the correlations between multi-jet production

and the hard colour-structures present in an event [5, 6], and perturbative threshold corrections to jet production [7].

One issue that has been largely neglected, but that is highly relevant in a hadron collider context such as Tevatron or LHC, is that of the modification of jet kinematics by non-perturbative effects associated with the proton beams. These effects, often referred to as a whole as the "underlying event" (UE), are rather poorly understood. However, from tuned underlying event models [8, 9] one consistently finds that a principal effect of the underlying event is to add a rather large amount of transverse momentum, fairly uniformly throughout the event — according to models $3 - 5 \, \mathrm{GeV}$ per unit rapidity at the Tevatron, $10 - 15 \, \mathrm{GeV}$ per unit rapidity at LHC (see $e.g.$ [10]), with substantial fluctuations in the amount from one event to another. This is an order of magnitude larger than the normal scale for non-perturbative effects in an e^+e^- or DIS context, which amount to $\sim 0.5 \, \mathrm{GeV}$ per unit rapidity (with respect to the $q\bar{q}$ axis) [11, 12]. Thus for jets with transverse momenta, p_t, of several tens of GeV, the UE can be as significant as the perturbative corrections to the jet transverse momentum $\alpha_s p_t$. A related issue is that of pileup (PU), minimum-bias collisions that occur in the same bunch crossing as the main event, with the potential of further adding up to $100 \, \mathrm{GeV}$ of soft radiation per unit rapidity.

Given the large momentum scales associated with the underlying event and pileup, it is important to develop tools for understanding how they both affect jets. Of the two, PU is conceptually simpler because the particles it adds are entirely uncorrelated with those of the hard scatter. In contrast UE particles cannot entirely be disentangled from those due to the hard scatter (neither theoretically, nor in practice). Nevertheless it is useful, and probably not too poor an approximation, to treat the UE as largely independent from the hard scatter, just like PU.

The UE and PU can affect a jet in two ways. Firstly, particles from the UE and PU may be clustered into the jet, increasing its transverse momentum. Secondly, the presence of the UE/PU particles can modify the way in which the particles from the hard scatter get clustered into jets.

To study the first of these effects, we shall introduce the concept of the 'jet area'. The logic that motivates this particular quantity is that UE/PU particles are distributed uniformly in rapidity and azimuth, at least after averaging over many events. Therefore a good measure of a jet's susceptibility to contamination by the UE and PU should be given by the extent of the region (on the rapidity-azimuth cylinder) over which it is liable to capture UE and pileup particles, $i.e.$ the jet's "catchment area".

One might naively think of this area as the area of the surface covering all the particles that make up the jet. However, a moment's reflection shows that this area is actually zero, the jet being made of pointlike particles. An alternative obvious definition would be to take the area of the convex hull surrounding all the particles in a jet. However this definition also fails to satisfy basic sensibleness requirements: for example, if two jets share an irregular boundary it is possible that their convex hulls will overlap and the assignment of area to one or other of the jets becomes ambiguous.

Two main definitions of the jet area can be introduced without such shortcomings. A first one, the *passive area*, involves scanning a single infinitely soft particle (a "ghost") over the rapidity-azimuth cylinder and determining the region over which it is clustered within a given jet. It can be understood as a measure of the susceptibility of the jet to contamination from an UE with pointlike structure. It is discussed in detail in section 2. A second definition, studied in section 3, the *active area*, involves adding a dense coverage of infinitely soft "ghosts" and counting how many are clustered inside a given jet. It can be understood as a measure of the susceptibility of the jet to contamination from an UE with uniform, diffuse structure. The passive and active areas differ because in the latter case the ghosts can cluster among themselves as well as with the actual event

3

particles, thus playing a more active role in the clustering. We will find that the numerical value of the jet area can differ according to the kind of particles which make up the jet. The simplest case to study will be that of jets with a single hard particle. We shall then consider how the jet's area changes when its main hard parton emits an additional soft gluon, which effectively causes the jet area to acquire an anomalous dimension. We will also study "pure ghost jets," those exclusively made up of ghost particles.

In section 4 we shall then consider how the presence of UE/PU affects the clustering into jets of particles from the hard scatter, an effect that we refer to as the *back-reaction* of the UE/PU on the jet structure. Again we shall examine this in two limits: pointlike and perfectly diffuse UE/PU.

As well as making approximations for the structure of the UE/PU it will also be necessary to make simplifying approximations for the jet structure. Thus we will work in an energy ordered limit, in which the perturbative event consists of one very hard particle, plus an optional additional soft perturbative particle. This will be equivalent to making a leading (single) logarithmic approximation, truncated at its first non-trivial order in α_s. For some purposes we shall also assume a small jet-radius parameter R (*i.e.* the small-cone approximation of [13]).

Despite these many simplifications, the calculations that we present will be seen to provide considerable insight into the mechanisms at play in jet clustering in events with UE/PU. In particular they will help highlight characteristic analytical structures that are common to a range of rather different jet algorithms, as well as significant jet-algorithm-dependent differences in the quantitative impact of these analytical structures. Of the algorithms that we will consider, two are based on sequential recombination (the inclusive k_t [14] and Cambridge/Aachen [15] algorithms), while the third is a modern, infrared-safe stable-cone algorithm (SISCone [16]) with a Tevatron Run-II type split–merge procedure to resolve overlapping stable cones. A brief description of the three jet algorithms is given in Appendix A.

To help reinforce the connection between our analytical results and the (simulated) real world, we shall close the article in section 5 with Herwig [17] and Pythia [18] Monte Carlo studies. We shall also give the foundations of the use of the area concept for performing PU subtractions [19].

2   Passive Area

Suppose we have an event composed of a set of particles $\{p_i\}$ which are clustered into a set of jets $\{J_i\}$ by some infrared-safe jet algorithm.

Imagine then adding to the $\{p_i\}$ a single *infinitely soft* ghost particle g at rapidity y and azimuth ϕ, and repeating the jet-finding. As long as the jet algorithm is infrared safe, the set of jets $\{J_i\}$ is not changed by this operation: their kinematics and hard particle content will remain the same, the only possible differences being the presence of g in one of the jets or the appearance of a new jet containing only g.

The passive area of the jet J can then either be defined as a scalar

$$a(J) \equiv \int dy\, d\phi\, f(g(y,\phi), J) \qquad\qquad f(g, J) = \left\{ \begin{array}{ll} 1 & g \in J \\ 0 & g \notin J \end{array} \right. , \qquad (1)$$

which corresponds to the area of the region where g is clustered with J, or as a 4-vector,

$$a_\mu(J) \equiv \int dy\, d\phi\, f_\mu(g(y,\phi), J) \qquad\qquad f_\mu(g, J) = \left\{ \begin{array}{ll} g_\mu/g_t & g \in J \\ 0 & g \notin J \end{array} \right. , \qquad (2)$$

where g_t is the ghost transverse momentum. For a jet with a small area $a(J) \ll 1$, the 4-vector area has the properties that its transverse component satisfies $a_t(J) = a(J)$, it is roughly massless and

it points in the direction of J. For larger jets, $a(J) \sim 1$, the 4-vector area acquires a mass and may not point in the same direction as J. We shall restrict our attention here to scalar areas because of their greater simplicity. Nearly all results for scalar areas carry over to 4-vector areas, modulo corrections suppressed by powers of the jet radius (usually accompanied by a small coefficient).[1]

2.1 Areas for 1-particle configurations

Consider this definition in the context of three different jet algorithms: inclusive k_t, inclusive Cambridge/Aachen, and the seedless infrared-safe cone algorithm, SISCone. The definitions of all three algorithms are summarised in appendix A.

Each of these jet algorithms contains a parameter R which essentially controls the reach of the jet algorithm in y and ϕ. Given an event made of a single particle p_1, the passive area of the jet J_1 containing it is $a(J_1) = \pi R^2$ for all three algorithms.

2.2 Areas for 2-particle configurations

Let us now consider what happens to the passive jet areas in the presence of additional soft perturbative radiation. We add a particle p_2 such that the transverse momenta are strongly ordered,

$$p_{t1} \gg p_{t2} \gg \Lambda_{QCD} \gg g_t , \qquad (4)$$

and p_1 and p_2 are separated by a geometrical distance $\Delta_{12} = (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2$ in the y-ϕ plane. Subsequently we shall integrate over Δ_{12} and p_{t2}. Note that g_t has been taken to be infinitesimal compared to all physical scales to ensure that the presence of the ghost particle does not affect the real jets.

For $\Delta_{12} = 0$ collinear safety ensures that the passive area is still equal to πR^2 for all three algorithms. However, as one increases Δ_{12}, each algorithm behaves differently.

2.2.1 k_t

Let us first consider the behaviour of the k_t jet algorithm, a sequential recombination algorithm, which has 2-particle distance measure $d_{ij} = \min(k_{ti}^2, k_{tj}^2)\Delta_{ij}^2/R^2$ and beam-particle distance $d_{iB} = k_{ti}^2$ [14]. Taking $\Delta_{12} \sim \Delta_{1g} \sim \Delta_{2g} \sim R$ and exploiting the strong ordering (4) one has

$$d_{1B} \gg d_{2B} \sim d_{12} \gg d_{g1} \sim d_{g2} \sim d_{gB} . \qquad (5)$$

From this ordering of the k_t distances, one sees that the ghost always undergoes its clustering before either of the perturbative particles p_1 and p_2. Specifically, if at least one of Δ_{1g} and Δ_{2g} is smaller than R, the ghost clusters with the particle that is geometrically closer.

If both Δ_{1g} and Δ_{2g} are greater than R the ghost clusters with the beam and will not belong to any of the perturbative jets.

---

[1] The above definitions apply to jet algorithms in which each gluon is assigned at most to one jet. For a more general jet algorithm (such as the "Optimal" jet finder of [20] or those which perform $3 \to 2$ recombination like ARCLUS [21]), then one may define the 4-vector area as

$$a_\mu(J) = \lim_{g_t \to 0} \frac{1}{g_t} \int dy \, d\phi \, (J_\mu(\{p_i\}, g(y, \phi)) - J_\mu(\{p_i\})) , \qquad (3)$$

where $J_\mu(\{p_i\}, g(y, \phi))$ is the 4-momentum of the jet as found on the full set of event particles $\{p_i\}$ plus the ghost, while $J_\mu(\{p_i\})$ is the jet-momentum as found just on the event particles.
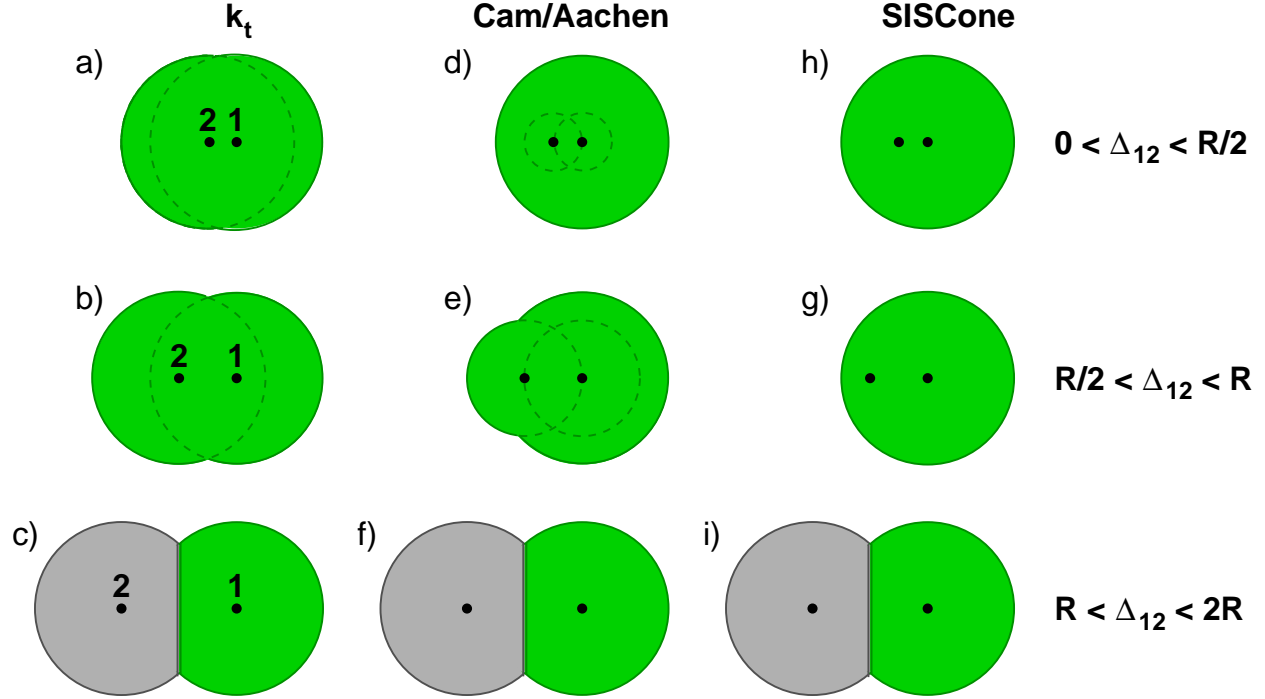
Figure 1: Schematic representation of the passive area of a jet containing one hard particle "1" and a softer one "2" for various separations between them and different jet algorithms. Different shadings represent distinct jets.

Let us now consider various cases. If $\Delta_{12} < R$, (fig. 1a,b) the particles p_1 and p_2 will eventually end up in the same jet. The ghost will therefore belong to the jet irrespectively of having been clustered first with p_1 or p_2. The area of the jet will then be given by union of two circles of radius R, one centred on each of the two perturbative particles,

$$a_{k_t,R}(\Delta_{12}) = u\left(\frac{\Delta_{12}}{R}\right) \pi R^2\,, \qquad \text{for } \Delta_{12} < R\,, \tag{6}$$

where

$$u(x) = \frac{1}{\pi}\left[x\sqrt{1 - \frac{x^2}{4}} + 2\left(\pi - \arccos\left(\frac{x}{2}\right)\right)\right]\,, \tag{7}$$

represents the area, divided by π, of the union of two circles of radius one whose centres are separated by x.

The next case we consider is $R < \Delta_{12} < 2R$, fig. 1c. In this case p_1 and p_2 will never be able to cluster together. Hence, they form different jets, and the ghost will belong to the jet of the closer of p_1 or p_2. The two jets will each have area

$$a_{k_t,R}(\Delta_{12}) = \frac{u(\Delta_{12}/R)}{2} \pi R^2\,, \qquad \text{for } R < \Delta_{12} < 2R\,. \tag{8}$$

Finally, for $\Delta_{12} > 2R$ the two jets formed by p_1 and p_2 each have area πR^2.

The three cases derived above are summarised in table 1 and illustrated in fig. 2.

2.2.2 Cambridge/Aachen

Now we consider the behaviour of the Cambridge/Aachen jet algorithm [15], also a sequential recombination algorithm, defined by the 2-particle distance measure $d_{ij} = \Delta_{ij}^2/R^2$ and beam-particle distance $d_{iB} = 1$. Because in this case the distance measure does not involve the transverse momentum of the particles, the ghost only clusters first if $\min(\Delta_{1g}, \Delta_{2g}) < \Delta_{12}$. Otherwise, p_1 and p_2 cluster first into the jet J, and then J captures the ghost if $\Delta_{Jg} \simeq \Delta_{1g} < R$.

The region $\Delta_{12} < R$ now needs to be separated into two parts, $\Delta_{12} < R/2$ and $R/2 < \Delta_{12} < R$.

If the ghost clusters first, then it must have been contained in either of the dashed circles depicted in figs. 1d,e. If $\Delta_{12} < R/2$ both these circles are contained in a circle of radius R centred on p_1 (fig. 1d), and so the jet area is πR^2.

If $R/2 < \Delta_{12} < R$ (fig. 1e) the circle of radius Δ_{12} centred on p_2 protrudes, and adds to the area of the final jet, so that

$$a_{\mathrm{Cam},R}(\Delta_{12}) = w\left(\frac{\Delta_{12}}{R}\right)\pi R^2\,, \qquad \text{for } R/2 < \Delta_{12} < R\,, \tag{9}$$

where

$$w(x) = \frac{1}{\pi}\left[\pi - \arccos\left(\frac{1}{2x}\right) + \sqrt{x^2 - \frac{1}{4}} + x^2 \arccos\left(\frac{1}{2x^2} - 1\right)\right]\,. \tag{10}$$

For $\Delta_{12} > R$ a Cambridge/Aachen jet has the same area as the k_t jet, *cf.* fig. 1f. As with the k_t algorithm, the above results are summarised in table 1 and fig. 2. The latter in particular illustrates the significant difference between the k_t and Cambridge/Aachen areas for $\Delta_{12} \sim R/2$, caused by the different order of recombinations in the two algorithms.

2.2.3 SISCone

Modern cone jet algorithms identify stable cones and then apply a split/merge procedure to over-lapping stable cones. The arguments that follow are identical for midpoint and seedless cone jet algorithms with a Tevatron run II type split–merge procedure [22]. For higher orders, or more realistic events, it is mandatory to use an infrared-safe seedless variant, a main example of which is SISCone [16].

For $\Delta_{12} < R$ only a single stable cone is found, centred on p_1. Any ghost within distance R of p_1 will therefore belong to this jet, so its area will be πR^2, *cf.* fig. 1h,g.

For $R < \Delta_{12} < 2R$ two stable cones are found, centred on p_1 and p_2. The split/merge procedure will then always split them, because the fraction of overlapping energy is zero. Any ghost falling within either of the two cones will be assigned to the closer of p_1 and p_2 (see fig. 1i). The area of the hard jet will therefore be the same as for the k_t and Cambridge algorithms.

Again these results are summarised in table 1 and fig. 2 and one notices the large differences relative to the other algorithms at $R \lesssim 1$ and the striking feature that the cone algorithm only ever has negative corrections to the passive area for these energy-ordered two-particle configurations.

2.3 Area scaling violation

Given that jet passive areas are modified by the presence of a soft particle in the neighbourhood of a jet, one expects the average jet area to acquire a logarithmic dependence on the jet transverse momentum when the jets acquire a sub-structure as a consequence of radiative emission of gluons.To determine its coefficient we shall work in the approximation of small jet radii, motivated by the

	k_t	$a_{\mathrm{JA},R}(\Delta_{12})/\pi R^2$ Cambridge/Aachen	SISCone
$0 < \Delta_{12} < R/2$	$u(\Delta_{12}/R)$	1	1
$R/2 < \Delta_{12} < R$	$u(\Delta_{12}/R)$	$w(\Delta_{12}/R)$	1
$R < \Delta_{12} < 2R$	$u(\Delta_{12}/R)$ / 2	$u(\Delta_{12}/R)$ / 2	$u(\Delta_{12}/R)$ / 2
$\Delta_{12} > 2R$	1	1	1

Table 1: Summary of the passive areas for the three jet algorithms for the hard jet in an event containing one hard and one soft particle, separated by a y-ϕ distance Δ_{12}. The functions u and w are defined in eqs. (7) and (10).
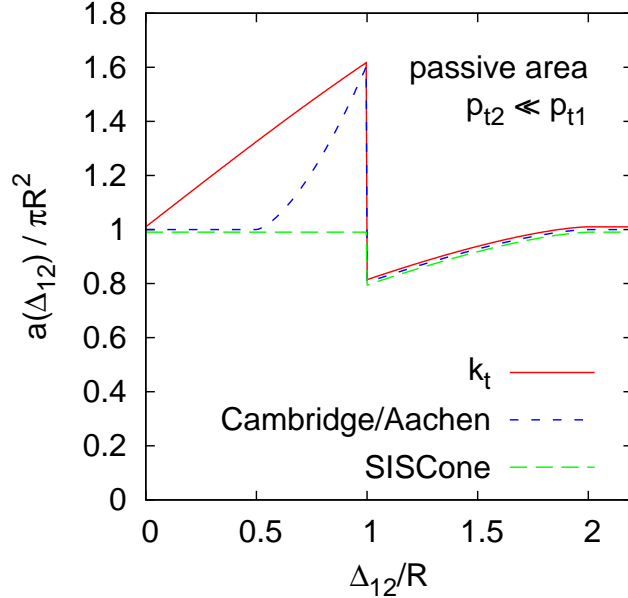

Figure 2: Plot of the passive areas of the hard jet as a function of the distance between the hard and the soft particle, as given in table 1.

observation that corrections for finite R, proportional to powers of R, are usually accompanied by a small coefficient [13, 10].

In the small-angle limit, the QCD matrix element for the emission of the perturbative soft gluon, p_2, is

$$\frac{dP}{dp_{t2}\,d\Delta_{12}} = C_1 \frac{2\alpha_s(p_{t2}\Delta_{12})}{\pi} \frac{1}{\Delta_{12}} \frac{1}{p_{t2}}, \tag{11}$$

where C_1 is C_F or C_A according to whether the hard particle p_1 is a quark or a gluon. We make use of the fact that Δ_{12} is just the angle between the two particles (to within a longitudinal boost-dependent factor), and will assume that R is sufficiently small that the small-angle approximation is legitimate. The scale of the coupling is taken to be the transverse momentum of p_2 relative to p_1.

At order α_s the mean jet area with a given jet algorithm (JA) can be written

$$\langle a_{\mathrm{JA},R}\rangle = a_{\mathrm{JA},R}(0) + \langle\Delta a_{\mathrm{JA},R}\rangle = \pi R^2 + \langle\Delta a_{\mathrm{JA},R}\rangle, \tag{12}$$

where we explicitly isolate the $\mathcal{O}(\alpha_s)$ higher-order contribution,

$$\langle \Delta a_{\mathrm{JA},R} \rangle \simeq \int_0^{2R} d\Delta_{12} \int_{Q_0/\Delta_{12}}^{p_{t1}} dp_{t2} \frac{dP}{dp_{t2}\, d\Delta_{12}} (a_{\mathrm{JA},R}(\Delta_{12}) - a_{\mathrm{JA},R}(0)) \,, \tag{13}$$

with the $-a_{\mathrm{JA},R}(0)$ term accounting for virtual corrections. $\langle \cdots \rangle$ represents therefore an average over perturbative emission. Note that because of the $1/p_{t2}$ soft divergence for the emission of p_{t2}, eq. (13) diverges unless one specifies a lower limit on p_{t2} — accordingly we have to introduce an infrared cutoff Q_0/Δ_{12} on the p_t of the emitted gluon. This value results from requiring that the transverse momentum of p_2 relative to p_1, *i.e.* $p_{t2}\Delta_{12}$, be larger than Q_0. The fact that we need to place a lower limit on p_{t2} means that jet areas are infrared unsafe[2] — they cannot be calculated order by order in perturbative QCD and for real-life jets they will depend on the details of non-perturbative effects (hadronisation). One can account for this to some extent by leaving Q_0 as a free parameter and examining the dependence of the perturbative result on Q_0.

As concerns the finiteness of the Δ_{12} integration, all the jet algorithms we consider have the property that

$$\lim_{\Delta_{12}\to 0} a_{\mathrm{JA},R}(\Delta_{12}) = a_{\mathrm{JA},R}(0) = \pi R^2 \,; \tag{14a}$$

$$a_{\mathrm{JA},R}(\Delta_{12}) = \pi R^2 \quad \text{for } \Delta_{12} > 2R \tag{14b}$$

so that the integral converges for $\Delta_{12} \to 0$, and we can place the upper limit at $\Delta_{12} = 2R$.

After evaluating eq. (13), with the replacement $\Delta_{12} \to R$, both in the lower limit of the p_{t2} integral and the argument of the coupling, we obtain

$$\langle \Delta a_{\mathrm{JA},R} \rangle = d_{\mathrm{JA},R} \frac{2\alpha_s C_1}{\pi} \ln \frac{R p_{t1}}{Q_0} \,, \qquad d_{\mathrm{JA},R} = \int_0^{2R} \frac{d\theta}{\theta} (a_{\mathrm{JA},R}(\theta) - \pi R^2) \,, \tag{15}$$

in a fixed coupling approximation, and

$$\langle \Delta a_{\mathrm{JA},R} \rangle = d_{\mathrm{JA},R} \frac{C_1}{\pi b_0} \ln \frac{\alpha_s(Q_0)}{\alpha_s(R p_{t1})} \,, \tag{16}$$

with a one-loop running coupling, where $b_0 = \frac{11C_A - 2n_f}{12\pi}$. The approximation $\Delta_{12} \sim R$ in the argument of the running coupling in the integrand corresponds to neglecting terms of $\mathcal{O}(\alpha_s)$ without any enhancements from logarithms of R or p_{t1}/Q_0.

The results for $d_{\mathrm{JA},R}$ are

$$d_{k_t,R} = \left(\frac{\sqrt{3}}{8} + \frac{\pi}{3} + \xi \right) R^2 \simeq 0.5638\, \pi R^2 \,, \tag{17a}$$

$$d_{\mathrm{Cam},R} = \left(\frac{\sqrt{3}}{8} + \frac{\pi}{3} - 2\xi \right) R^2 \simeq 0.07918\, \pi R^2 \,, \tag{17b}$$

$$d_{\mathrm{SISCone},R} = \left(-\frac{\sqrt{3}}{8} + \frac{\pi}{6} - \xi \right) R^2 \simeq -0.06378\, \pi R^2 \,, \tag{17c}$$

where

$$\xi \equiv \frac{\psi'(1/6) + \psi'(1/3) - \psi'(2/3) - \psi'(5/6)}{48\sqrt{3}} \simeq 0.507471 \,, \tag{18}$$

---

[2]An exception is for the SISCone algorithm with a cut, $p_{t,min}$, on the minimum transverse momentum of protojets that enter the split–merge procedure procedure — in that situation Q_0 is effectively replaced by $p_{t,min}$.

with $\psi'(x) = d^2/dx^2(\ln \Gamma(x))$. One notes that the coefficient for the k_t algorithm is non-negligible, given that it is multiplied by the quantity $2\alpha_s C_1/\pi \ln Rp_{t1}/Q_0$ in eq. (15) (or its running coupling analogue), which can be of order 1. In contrast the coefficients for Cambridge/Aachen and SISCone are much smaller and similar (the latter being however of opposite sign). Thus k_t areas will depend significantly more on the jet p_t than will those for the other algorithms.

The fluctuation of the areas can be calculated in a similar way. Let us define

$$\langle \sigma_{\text{JA},R}^2 \rangle = \langle a_{\text{JA},R}^2 \rangle - \langle a_{\text{JA},R} \rangle^2 = \sigma_{\text{JA},R}^2(0) + \langle \Delta\sigma_{\text{JA},R}^2 \rangle , \qquad \sigma_{\text{JA},R}^2(0) = 0 , \qquad (19)$$

where we have introduced $\sigma_{\text{JA},R}^2(0)$, despite its being null, so as to facilitate comparison with later results. We then evaluate

$$\langle \Delta\sigma_{\text{JA},R}^2 \rangle = \langle \Delta a_{\text{JA},R}^2 \rangle - \langle \Delta a_{\text{JA},R} \rangle^2 \simeq \langle \Delta a_{\text{JA},R}^2 \rangle , \qquad (20)$$

where we neglect $\langle \Delta a_{\text{JA},R} \rangle^2$ since it is of $\mathcal{O}\left(\alpha_s^2 \ln^2(Rp_{t1}/Q_0)\right)$. The calculation of $\langle \Delta a_{\text{JA},R}^2 \rangle$ proceeds much as for $\langle \Delta a_{\text{JA},R} \rangle$ and gives

$$\langle \Delta a_{\text{JA},R}^2 \rangle = s_{\text{JA},R}^2 \frac{C_1}{\pi b_0} \ln \frac{\alpha_s(Q_0)}{\alpha_s(Rp_{t1})} , \qquad s_{\text{JA},R}^2 = \int_0^{2R} \frac{d\theta}{\theta} (a_{\text{JA},R}(\theta) - \pi R^2)^2 \qquad (21)$$

for running coupling. The results are

$$s_{k_t,R}^2 = \left(\frac{\sqrt{3}\pi}{4} - \frac{19}{64} - \frac{15\zeta(3)}{8} + 2\pi\xi \right) R^4 \simeq (0.4499\,\pi R^2)^2 , \qquad (22a)$$

$$s_{\text{Cam},R}^2 = \left(\frac{\sqrt{3}\pi}{6} - \frac{3}{64} - \frac{\pi^2}{9} - \frac{13\zeta(3)}{12} + \frac{4\pi}{3}\xi \right) R^4 \simeq (0.2438\,\pi R^2)^2 , \qquad (22b)$$

$$s_{\text{SISCone},R}^2 = \left(\frac{\sqrt{3}\pi}{12} - \frac{15}{64} - \frac{\pi^2}{18} - \frac{13\zeta(3)}{24} + \frac{2\pi}{3}\xi \right) R^4 \simeq (0.09142\,\pi R^2)^2 . \qquad (22c)$$

We have a hierarchy between algorithms that is similar to that observed for the average area scaling violations, though now the coefficient for Cambridge/Aachen is more intermediate between the other two.

2.4 n-particle properties and the Voronoi area

The only algorithm for which one can make any statement about the passive area for a general n-particle configuration is the k_t algorithm.

Because of the k_t distance measure, the single ghost will cluster with one of the event particles before any other clustering takes place. One can determine the region in which the ghost will cluster with a given particle, and this is a definition of the area $a_{k_t,R}(p_i)$ of a particle p_i. Since the ghost-particle clustering will occur before any particle-particle clustering, the jet area will be the sum of the areas of all its constituent particles:

$$a_{k_t,R}(J) = \sum_{p_i \in J} a_{k_t,R}(p_i) . \qquad (23)$$

Can anything be said about the area of a particle? The ghost will cluster with the event particle to which it is closest, as long it is within a distance R. There exists a geometrical construction known as the Voronoi diagram, which subdivides the plane with a set of vertices into cells around
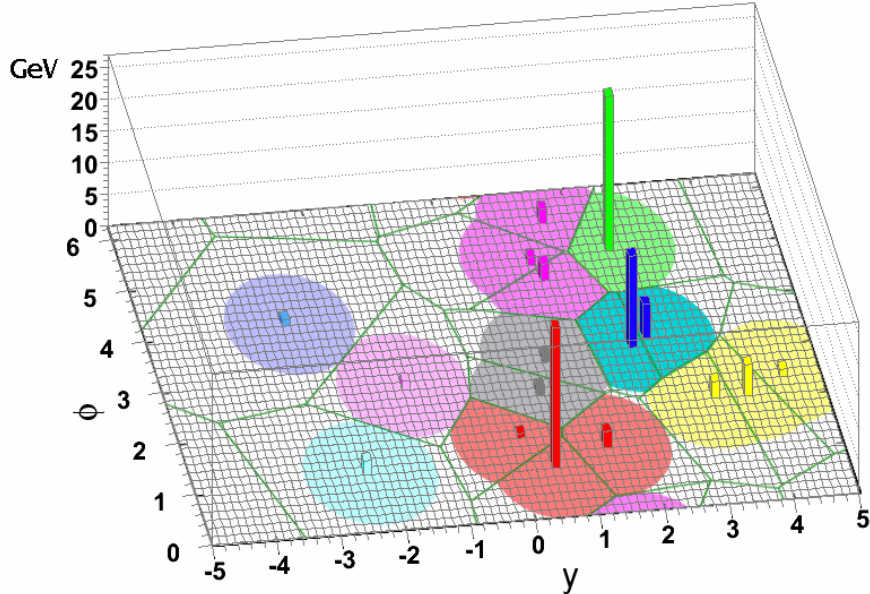
Figure 3: The passive area of jets in a parton-level event generated by Herwig and clustered with the k_t algorithm with $R = 1$. The towers represent calorimeter cells containing the particles, the straight (green) lines are the edges of the Voronoi cells and the shaded regions are the areas of the jets.

each vertices.[3] Each cell has the property that all points in the cell have as their closest vertex the cell's vertex. Thus the Voronoi cell is remarkably similar to the region in which a ghost will cluster with a particle. The only difference arises because of the limitation that the ghost should be within a distance R of the particle — this causes the area of particle i to be the area of its Voronoi cell $\mathcal{V}_i$ intersected with a circle of radius R, $\mathcal{C}_{i,R}$, centred on the particle. This leads us to define a Voronoi area for a particle, $a_R^{\mathcal{V}}(p_i)$,

$$a_R^{\mathcal{V}}(p_i) \equiv \text{area}(\mathcal{V}_i \cap \mathcal{C}_{i,R}) . \tag{24}$$

Thus given a set of momenta, the passive area of a k_t jet can be directly determined from the Voronoi diagram of the event,[4] using eq. (23) and the relation

$$a_{k_t,R}(p_i) = a_R^{\mathcal{V}}(p_i) . \tag{25}$$

It is quite straightforward to see that this result holds for the 2-particle case, because the Voronoi diagram there consists of a single line, equidistant between the two points. It divides the plane into two half-planes, each of which is the Voronoi cell of one of the particles (this is best seen in fig. 1c). The intersection of the halfplane with the circle of radius R centred on the particle has area $\frac{1}{2}\pi R^2 u(\Delta_{12}/R)$, and this immediately gives us the results eqs. (6), (8) according to whether the particles cluster into a single jet or not.

The Voronoi construction of the k_t-algorithm passive area is illustrated for a more complex event in fig. 3. One sees both the Voronoi cells and how their intersection with circles of radius $R = 1$ gives the area of the particles making up those jets.

---

[3]It is this same geometrical construction that was used to obtain a nearest neighbour graph that allowed k_t jet clustering to be carried out in $N \ln N$ time [23].

[4]Strictly speaking it should be the Voronoi diagram on a $y - \phi$ cylinder, however this is just a technical detail.
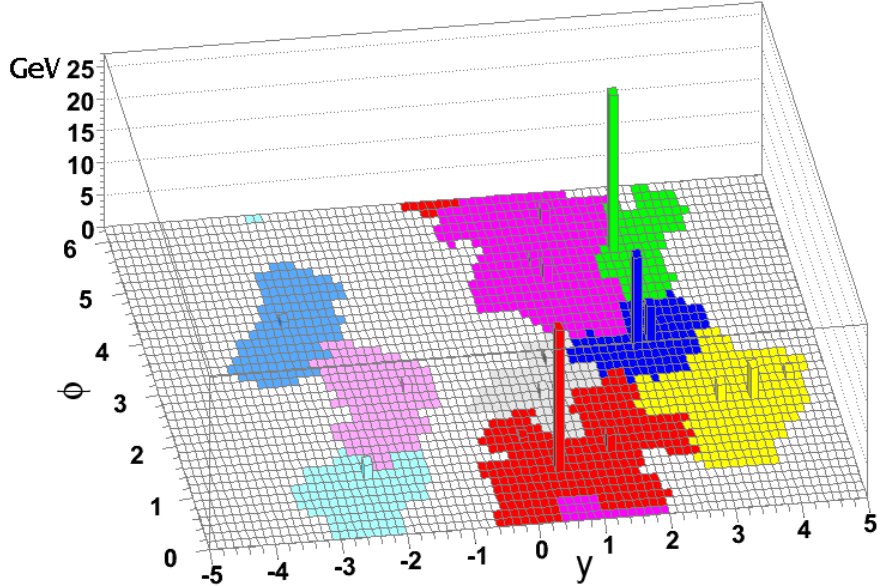
Figure 4: Active area for the same event as in figure 3, once again clustered with the k_t algorithm and $R = 1$. Only the areas of the hard jets have been shaded — the pure 'ghost' jets are not shown.

Note that it is not possible to write passive areas for jet algorithms other than k_t in the form eq. (23). One can however introduce a new type of area for a generic algorithm, a *Voronoi area*, in the form

$$a^{\mathcal{V}}_{\mathrm{JA},R}(J) = \sum_{p_i \in J} a^{\mathcal{V}}_{R}(p_i) \,. \tag{26}$$

While for algorithms other than k_t (for which, as we have seen, $a_{k_t,R}(J) = a^{\mathcal{V}}_{k_t,R}(J)$), this area is not in general related to the clustering of any specific kind of background radiation, it can nevertheless be a useful tool, because its numerical evaluation is efficient [24, 23] and as we shall discuss later (sec. 3.4), for dense events its value coincides with both passive and active area definitions.

3   Active Area

To define an active area, as for the passive area, we start with an event composed of a set of particles $\{p_i\}$ which are clustered into a set of jets $\{J_i\}$ by some infrared-safe jet algorithm. However, instead of adding a single soft ghost particle, we now add a dense coverage of ghost particles, $\{g_i\}$, randomly distributed in rapidity and azimuth, and each with an infinitesimal transverse momentum.[5] The clustering is then repeated including the set of particles plus ghosts.

During the clustering the ghosts may cluster both with each other and with the hard particles. This more 'active' participation in the clustering is the origin of the name that we give to the area to be defined shortly. It contrasts with the definition of section 2, in which the single ghost acted more as a passive spectator, and in particular could not cluster with any other ghosts (there weren't any).

---

[5]In most cases the distribution of those transverse momenta will be irrelevant, at least in the limit in which the density of ghosts is sufficiently high.

Because of the infrared safety of any proper jet algorithm, even the addition of many ghosts does not change the momenta of the final jets $\{J_i\}$. However these jets do contain extra particles, ghosts, and we use the number of ghosts in a jet as a measure of its area. Specifically, if the number of ghosts per unit area (on the rapidity-azimuth cylinder) is ν_g and $\mathcal{N}_g(J)$ is the number of ghosts contained in jet J, then the (scalar) active area of a jet, given the specific set of ghosts $\{g_i\}$ is

$$A(J \,|\, \{g_i\}) = \frac{\mathcal{N}_g(J)}{\nu_g} \,. \tag{27}$$

An example of jet areas obtained in this way is shown in figure 4. One notes that the boundaries of the jets are rather ragged. Clustering with a different set of ghosts would lead to different boundaries. These differences propagate through to the number of ghosts clustered within a given jet, even in the limit $\nu_g \to \infty$, resulting in a different area each time. To obtain a unique answer for active area of a given jet one must therefore average over many sets of ghosts, in addition to taking the limit of infinite ghost density,

$$A(J) = \lim_{\nu_g \to \infty} \langle A(J \,|\, \{g_i\}) \rangle_g \,. \tag{28}$$

Note that as one takes $\nu_g \to \infty$, the ghost transverse momentum density, $\nu_g \langle g_t \rangle$, is to be kept infinitesimal.

The active area should bear a close resemblance to the average susceptibility of the jet to a high density of soft radiation (*e.g.* minimum-bias pileup), since the many soft particles will cluster between each other and into jets much in the same way as will the ghosts.

One may also define the standard deviation $\Sigma(J)$ of the distribution of a jet's active area across many ghost ensembles,

$$\Sigma^2(J) = \lim_{\nu_g \to \infty} \langle A^2(J \,|\, \{g_i\}) \rangle_g - A^2(J) \,. \tag{29}$$

This provides a measure of the variability of a given jet's contamination from (say) pileup and is closely connected with the momentum resolution that can be obtained with a given jet algorithm.

A feature that arises when adding many ghosts to an event is that some of the final jets contain nothing but ghost particles. They did not appear in the original list of $\{J_i\}$ and we refer to them as pure ghost jets. These 'ghost' jets (not shown in fig. 4), fill all of the 'empty' area, at least in jet algorithms for which all particles are clustered into jets. They will be similar to the jets formed from purely soft radiation in events with minimum-bias pileup, and so are interesting to study in their own right.

We can also define a 4-vector version of the active area (in analogy with the 4-vector passive area). It is given by

$$A_\mu(J \,|\, \{g_i\}) = \frac{1}{\nu_g \langle g_t \rangle} \sum_{g_i \in J} g_{\mu i} \,, \qquad A_\mu(J) = \lim_{\nu_g \to \infty} \langle A_\mu(J \,|\, \{g_i\}) \rangle_g \,. \tag{30}$$

The sum of the $g_{\mu i}$ is to be understood as carried out in the same recombination scheme as used in the jet clustering.[6]

---

[6]Though we do not give the details it is simple to extend the 4-vector active area definition to hold also for a general IR safe jet algorithm, in analogy with the extension of the passive area definition in eq. (3).
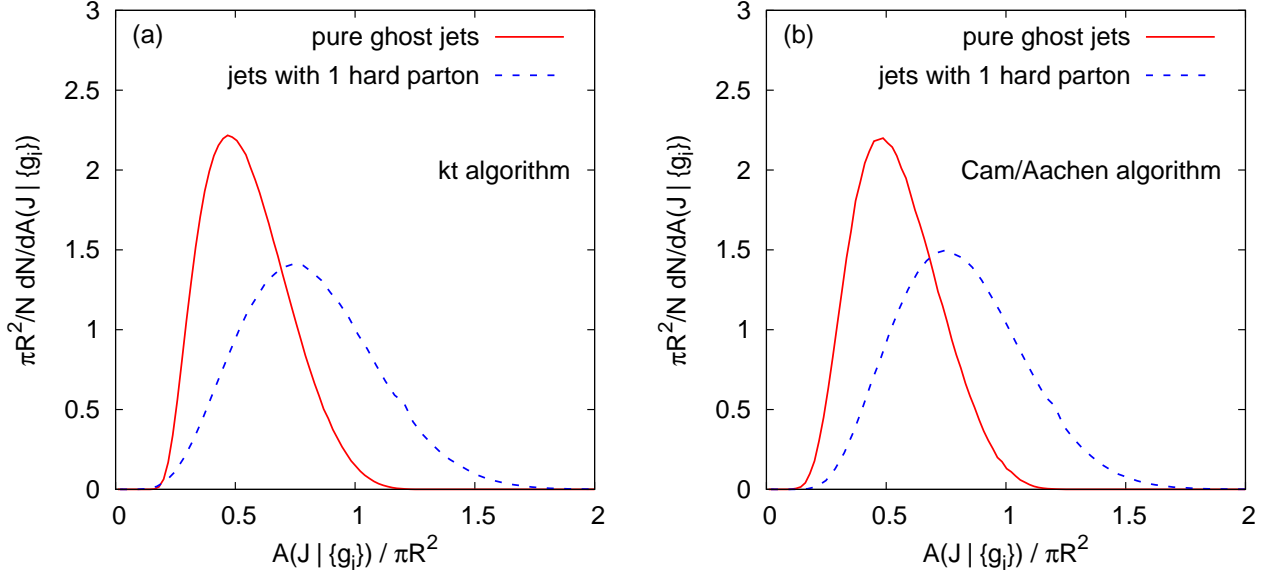
Figure 5: Distribution of active areas for pure ghost jets and jets with a single hard particle: (a) k_t algorithm, (b) Cambridge/Aachen algorithm.

3.1 Areas for 1-particle configurations and for ghost jets

3.1.1 k_t and Cambridge/Aachen

The active area for the k_t and Cambridge/Aachen algorithms is most readily studied numerically, by directly clustering large numbers of ghost particles, possibly together with one or more hard particles. This is feasible because of the availability of fast computational methods for carrying out the clustering in these algorithms, implemented in the FastJet package [23]. Typically we add ghosts with a density ν_g of ~ 100 per unit area,[7] in the rapidity region $|y| < y_{max} = 6$, and study jets in the region $|y| < y_{max} - R$. This leads to about 7500 ghost particles, which can be clustered in about $0.1\,$s on a $3.4\,$GHz processor. Each ghost is given a transverse momentum $\sim 10^{-100}\,$GeV and the one hard particle that we study has a transverse momentum of $100\,$GeV. The results are insensitive to the values chosen as long as their ratio is sufficiently large. We investigate in Appendix B how the distribution of the "1-hard-parton" jet area gets modified when the transverse momentum of the parton is progressively reduced below the scale of a generic set of soft particles.

Figure 5 shows the distribution of values of $A(J\,|\,\{g_i\})$ for pure ghost jets and jets with one hard particle. The distribution is obtained over a large ensemble of sets of ghosts. Let us concentrate initially on the case with a hard particle. Firstly the average active area, eq. (28) differs noticeably from the passive result of πR^2:

$$A_{k_t,R}(\text{one-particle-jet}) \simeq 0.812\,\pi R^2\,, \tag{31a}$$

$$A_{\text{Cam},R}(\text{one-particle-jet}) \simeq 0.814\,\pi R^2\,. \tag{31b}$$

Secondly, the distributions of the area in fig. 5 are rather broad. The randomness in the initial distribution of ghosts propagates all the way into the shape of the final jet and hence its area.

---

[7]They are placed on a randomly scattered grid, in order to limit the impact of the finite density, *i.e.* one effectively carries out quasi Monte Carlo integration of the ghost ensembles, so that that finite density effects ought to vanish as $\nu_g^{-3/4}$, rather than $\nu_g^{-1/2}$ as would be obtained with completely random placement.

14

This occurs because the k_t and Cambridge/Aachen algorithms flexibly adapt themselves to local structure (a good property when trying to reconstruct perturbative showering), and once a random perturbation has formed in the density of ghosts this seeds further growth of the soft part of the jet. The standard deviations of the resulting distributions are

$$\Sigma_{k_t,R}(\text{one-particle-jet}) \simeq 0.277 \, \pi R^2 \,, \tag{32a}$$

$$\Sigma_{\text{Cam},R}(\text{one-particle-jet}) \simeq 0.261 \, \pi R^2 \,. \tag{32b}$$

Figure 5 also shows the distribution of areas for pure ghost jets. One sees that pure ghost jets typically have a smaller area than hard-particle jets:[8]

$$A_{k_t,R}(\text{ghost-jet}) \simeq 0.554 \, \pi R^2 \,, \tag{34a}$$

$$A_{\text{Cam},R}(\text{ghost-jet}) \simeq 0.551 \, \pi R^2 \,, \tag{34b}$$

and the standard deviations are

$$\Sigma_{k_t,R}(\text{ghost-jet}) \simeq 0.174 \, \pi R^2 \,, \tag{35a}$$

$$\Sigma_{\text{Cam},R}(\text{ghost-jet}) \simeq 0.176 \, \pi R^2 \,. \tag{35b}$$

The fact that pure ghost jets are smaller than hard jets has an implication for certain physics studies: one expects jets made of soft 'junk' (minimum bias, pileup, thermal noise in heavy ions) to have area properties similar to ghost jets; since they are smaller on average than true hard jets, the hard jets will emerge from the junk somewhat more clearly than if both had the same area.

3.1.2 SISCone

The SISCone algorithm is unique among the algorithms studied here in that its active area is amenable to analytical treatment, at least in some cases.

We recall that a modern cone algorithm starts by finding all stable cones. One stable cone is centred on the single hard particle. Additionally, there will be a large number of other stable cones, of order of the number of ghost particles added to the event [16]. In the limit of an infinite number of ghosts, all cones that can be drawn in the rapidity-azimuth plane and that do not overlap with the hard particle will be stable. Many of these stable cones will still overlap with the cone centred on the hard particle, as long as they do not contain the hard particle itself (see figure 6, left).

Next, the SISCone algorithm involves a split–merge procedure. One defines $\tilde{p}_t$ for a jet as the scalar sum of the transverse momenta of its constituents. During the split–merge step, SISCone finds the stable cone with the highest $\tilde{p}_t$, and then the next hardest stable cone that overlaps with it. To decide whether these two cones (protojets) are to be merged or split, it determines the fraction

---

[8]Obtaining these values actually requires going beyond the ghost density and the rapidity range previously mentioned. In fact, when going to higher accuracy one notices the presence of small edge and finite-density effects, $\mathcal{O}(R/(y_{max} - R))$ and $\mathcal{O}(1/(\nu_g R^2))$ to some given power. Choosing the ghost area sufficiently small to ensure that finite-density effects are limited to the fourth decimal (in practice this means $1/(\nu_g R^2) < 0.01$) and extrapolating to infinite y_{max} one finds

$$A_{k_t}(\text{ghost-jet}) \simeq (0.5535 \pm 0.0005) \, \pi R^2 \,, \tag{33a}$$

$$A_{\text{Cam}}(\text{ghost-jet}) \simeq (0.5505 \pm 0.0005) \, \pi R^2 \,, \tag{33b}$$

with a conservative estimate of the residual uncertainty. This points to a small but statistically significant difference between the two algorithms.
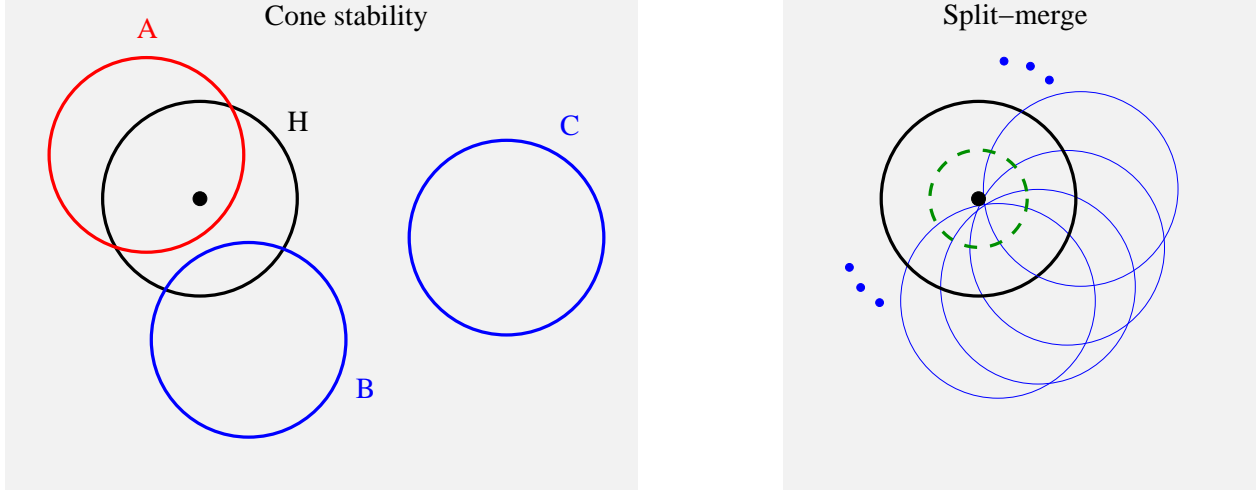
Figure 6: Left: the hard particle with the stable cone (H) centred on it, an example of a cone (A) that is unstable because it also contains the hard particle, and of two cones (B) and (C) that contain just ghost particles and are therefore stable. Right: some of the stable ghost cones (thin blue circles) that have the largest possible overlap with (H), together with the boundary of the hard jet after the split–merge procedure (dashed green line). In both diagrams, the grey background represents the uniform coverage of ghosts.

of the softer cone's $\tilde{p}_t$ that is shared with the harder cone. If this fraction is smaller than some value f (the overlap parameter of the cone algorithm), the two protojets are split: particles that are shared between them are assigned to the protojet to whose centre[9] they are closer. Otherwise they are merged into a single protojet. This procedure is repeated until the hard protojet no longer has any overlap with other protojets. At this point it is called a jet, and the split–merge procedure continues on the remaining protojets (without affecting the area of the hard jet).

The maximum possible overlap fraction,[10] f_{max}, between the hard protojet and a ghost protojet occurs in the situation depicted in figure 6 (right), $i.e.$ when the ghost protojet's centre is just outside the edge of the original hard stable cone (H). It is given by $f_{\mathrm{max}} = 2 - u(1) = \frac{2}{3} - \frac{\sqrt{3}}{2\pi} \approx 0.391$. This means that for a split–merge parameter $f > f_{\mathrm{max}}$ (commonly used values are $f = 0.5$ and $f = 0.75$) every overlap between the hard protojet and a pure-ghost stable cone will lead to a splitting. Since these pure-ghost stable cones are centred at distances $d > R$ from the hard particle, these splittings will reduce the hard jet to a circle of radius $R/2$ (the dashed green line in the right hand part of figure 6). The active area of the hard jet is thus

$$A_{\mathrm{SISCone},R}(\text{one-particle-jet}) = \frac{\pi R^2}{4} \,. \tag{36}$$

This result has been verified numerically using the same technique employed above for k_t and Cambridge/Aachen.

Note that this area differs considerably from the passive area, πR^2. This shows that the cone area is very sensitive to the structure of the event, and it certainly does not always coincide with

---

[9]This centre is given by the sum of momenta in the protojet before the split–merge operation.

[10]The fraction of momentum coincides with the fraction of area because the ghosts have uniform transverse momentum density.

the naive geometrical expectation πR^2, contrary to assumptions sometimes made in the literature (see for example [25]).

We further note that in contrast to k_t and Cambridge/Aachen algorithms, the SISCone algorithm always has the same active area for a single hard particle, independently of fluctuations of an infinitely dense set of ghosts, *i.e.*

$$\Sigma_{\mathrm{SISCone},R}(\text{one-particle-jet}) = 0 \,. \tag{37}$$

SISCone ghost-jet areas. While the area of a hard particle jet could be treated analytically, this is not the case for the pure-ghost jet area. Furthermore, numerical investigations reveal that the pure-ghost area distribution has a much more complicated behaviour than for k_t or Cambridge/Aachen. One aspect is that the distribution of pure-ghost jet areas is sensitive to the fine details of how the ghosts are distributed (density and transverse momentum fluctuations). Another is that it depends significantly on the details of the split–merge procedure. Figure 7a shows the distribution of areas of ghost jets for SISCone, for different values of the split–merge overlap threshold f. One sees, for example, that for smaller values of f there are occasional rather large ghost jets, whereas for $f \gtrsim 0.6$ nearly all ghost jets have very small areas.

One of the characteristics of SISCone that differs from previous cone algorithms is the specific ordering and comparison variable used to determine splitting and merging. As explained above, the choice made in SISCone was $\tilde{p}_t$, the scalar sum of transverse momenta of all particles in a jet. Previous cone algorithms used either the vector sum of constituent transverse momenta, p_t (an infrared unsafe choice), or the transverse energy $E_t = E \sin\theta$ (in a 4-vector recombination scheme). With both of these choices of variable, split–merge thresholds $f \lesssim 0.55$ can lead to the formation of 'monster' ghost jets, which can even wrap around the whole cylindrical phase space. For $f = 0.45$ this is a quite frequent occurrence, as illustrated in figure 7b, where one sees a substantial number of jets occupying the whole of the phase space (*i.e.* an area $4\pi y_{\max} \simeq 24\pi R^2$). Monster jets can be formed also with the $\tilde{p}_t$ choice, though it is a somewhat rarer occurrence — happening in 'only' $\sim 5\%$ of events.[11]

We have observed the formation of such monster jets also from normal pileup momenta simulated with Pythia [18], indicating that this disturbing characteristic is not merely an artefact related to our particular choice of ghosts. This indicates that a proper choice of the split–merge variable and threshold is critical in high-luminosity environments. The results from Figure 7a, suggest that if one wants to avoid monster jets, one has to choose a large enough value for f. Our recommendation is to adopt $f = 0.75$ as a default value for the split–merge threshold (together with the use of the $\tilde{p}_t$ variable, already the default in SISCone, for reasons related to infrared safety and longitudinal boost invariance).

3.2 Areas for 2-particle configurations

In this section we study the same problem described in section 2.2, *i.e.* the area of jets containing two particles, a hard one and a softer (but still "perturbative") one, *i.e.* eq. (4), but now for active areas. As before, the results will then serve as an input in understanding the dependence of the active area on the jet's transverse momentum when accounting for perturbative radiation.

---

[11]This figure is not immediately deducible from fig. 7b, which shows results normalised to the total number of ghost jets, rather than to the number of events.
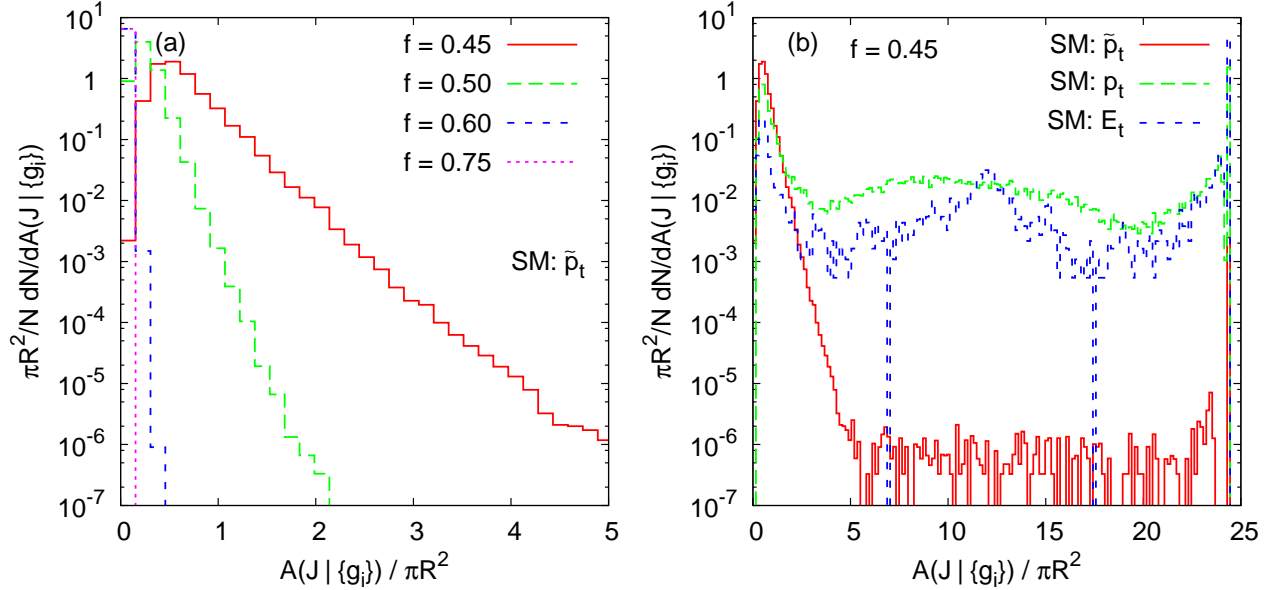
Figure 7: Distribution of pure-ghost jet areas for the SISCone algorithm, (a) with different values of the split–merge parameter f and (b) different choices for the scale used in the split–merge procedure. Ghosts are placed on a grid up to $|y| < 6$, with an average spacing of 0.2×0.2 in y, ϕ, and a random displacement of up to ± 0.1 in each direction, and transverse momentum values that are uniform modulo a $\pm 5\%$ random rescaling for each ghost. We consider all jets with $y < 5$. All jet definitions use $R = 1$ and multiple passes.

3.2.1 k_t and Cambridge/Aachen

As was the case for the active area of a jet containing a single hard particle, we again have to resort to numerical analyses to study that of jets with two energy-ordered particles. We define $A_{\mathrm{JA},R}(\Delta_{12})$ to be the active area for the energy-ordered two particle configuration already discussed in section 2.2.

Additionally since we have a distribution of areas for the single-particle active area case, it becomes of interest to study also $\Sigma_{\mathrm{JA},R}(\Delta_{12})$ the standard deviation of the distribution of areas obtained for the two-particle configuration.

The results are shown in figure 8: the active areas can be seen to be consistently smaller than the passive ones, as was the case for the 1-particle area, but retain the same dependence on the angular separation between the two particles. Among the various features, one can also observe that the active area does not quite reach the single-particle value ($\simeq 0.81\pi R^2$) at $\Delta_{12} = 2R$, but only somewhat beyond $2R$. This contrasts with the behaviour of the passive area. The figure also shows results for the cone area, discussed in the following subsection.

3.2.2 SISCone

In the case of the SISCone algorithm it is possible to find an analytical result for the two-particle active area, in an extension of what was done for one-particle case.[12]

---

[12]This is only possible for configurations with strong energy ordering between all particles — as soon as 2 or more particles have commensurate transverse momenta then the cone's split–merge procedure will include 'merge' steps, whose effects on the active area are currently beyond analytical understanding.
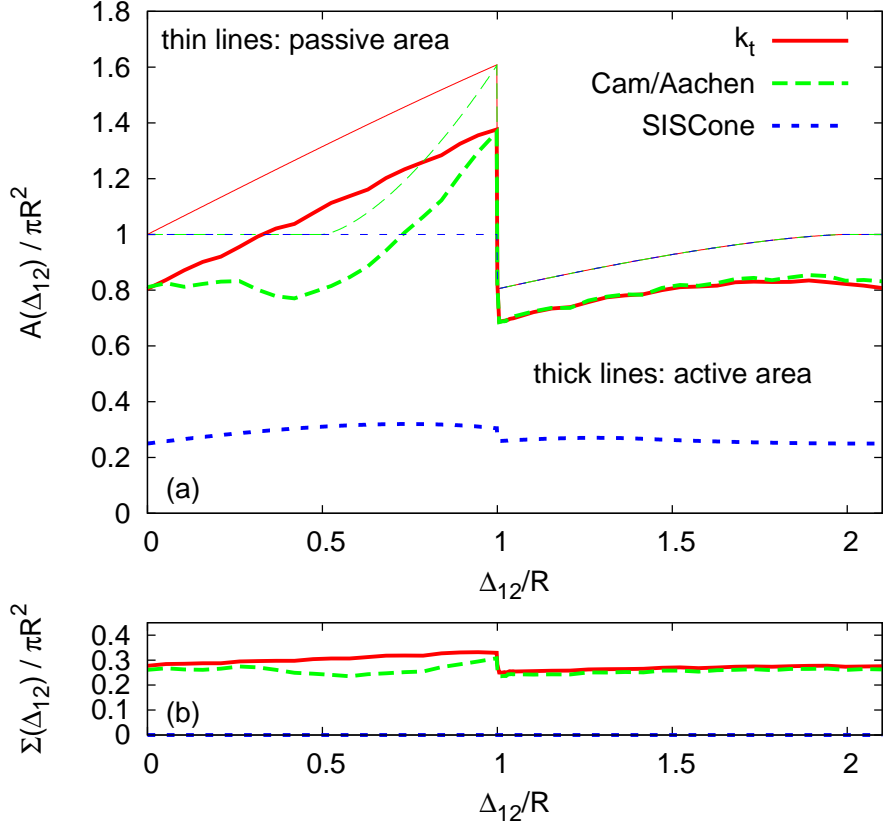
Figure 8: (a) Active areas (divided by πR^2) for the three jet algorithms as a function of the separation between a hard and a softer particle. For comparison we also include the passive areas, previously shown in fig. 2. (b) The corresponding standard deviations.

The stable-cone search will find one or two "hard" stable cones: the first centred on the hard particle and the second centred on the soft one, present only for $\Delta_{12} > R$. The pure-ghost stable cones will be centred at all positions distant by more than R from both p_1 and p_2, *i.e.* outside the two circles centred on p_1 and p_2.

We shall consider the active area of the jet centred on the hard particle p_1. When $\Delta_{12} > R$, the jet centred on the soft particle has the same area.

As in section 3.1.2, the split–merge procedure first deals with the pure-ghost protojets overlapping with the hard stable cone. For $f > f_{\max}$, this again only leads to splittings. Depending of the value of Δ_{12}, different situations are found as shown on figure 9. If the y-ϕ coordinates of the particles are $p_1 \equiv (\Delta, 0)$ and $p_2 \equiv (0, 0)$, the geometrical objects that are present are: the circle centred on p_1 with radius $R/2$; the tangents to this circle at y-ϕ coordinates $(\Delta_{12}/4, \pm\frac{1}{2}\sqrt{R^2 - (\Delta_{12}/2)^2})$; and, for $\Delta_{12} < R$, the ellipse of eccentricity Δ_{12}/R whose foci are p_1 and p_2 (given by the equation, $\Delta_{a1} + \Delta_{a2} = R$, where a is a point on the ellipse). For $\Delta_{12} < R$ the area of the jet is given by the union of the ellipse, the circle and the regions between the ellipse, the circle and each of the tangents (fig. 9a). For $R < \Delta_{12} < \sqrt{2}R$ it is given by the circle plus the region between the circle, the two tangents and the line equidistant from p_1 and p_2 (fig. 9b). For $\sqrt{2}R < \Delta_{12} < 2R$ it is given by the circle plus the region between the circle and two tangents, up to the intersection of the two tangents (fig. 9c). Finally, for $\Delta \geq 2R$, the area is $\pi R^2/4$.
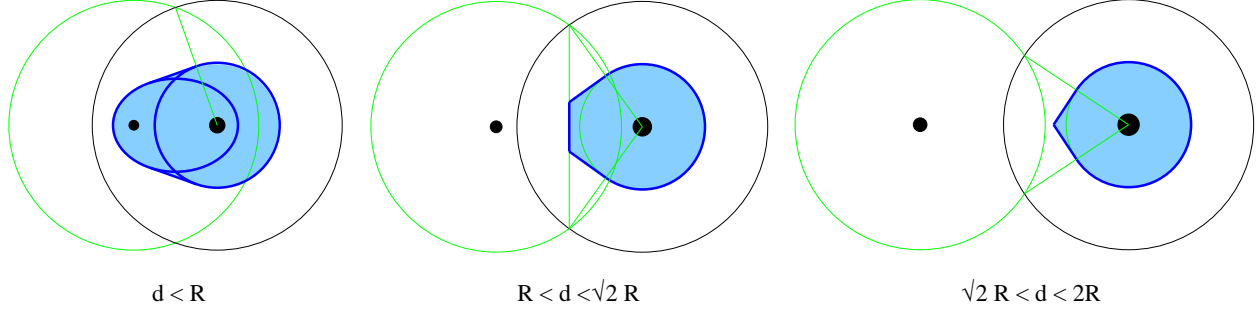
d < R   R < d < √2 R   √2 R < d < 2R

Figure 9: Picture of the active jet area for 2-particle configurations in the case of the SISCone jet algorithm. The black points represent the hard (big dot) and soft (small dot) particles, the black circle is the hard stable cone. The final hard jet is represented by the shaded area. The left (a) (centre (b), right (c)) plot corresponds to $\Delta_{12} < R$ ($R < \Delta_{12} < \sqrt{2}R$, $\sqrt{2}R < \Delta_{12} < 2R$).
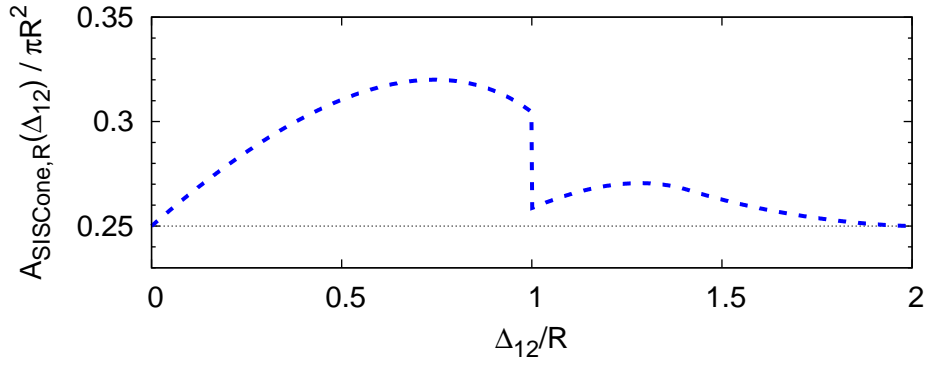

Figure 10: Active area of the hardest jet as a function of the distance between the hard and soft particle for the SISCone algorithm, *cf.* eq. (38).

An analytic computation of the active area gives

$$
\frac{A_{\mathrm{SISCone},R}(\Delta_{12})}{\pi R^2} = \frac{1}{4}\left[1 - \frac{1}{\pi}\arccos\left(\frac{x}{2}\right)\right] \tag{38}
$$
$$
+ \begin{cases} \frac{x}{2\pi}\sqrt{1 - \frac{x^2}{4}} + \frac{1}{4\pi}\sqrt{1 - x^2}\arccos\left(\frac{x}{2 - x^2}\right) & \Delta_{12} < R \\ \frac{x}{2\pi}\sqrt{1 - \frac{x^2}{4}} - \frac{x}{8\pi}\frac{1}{\sqrt{1 - \frac{x^2}{4}}} & R < \Delta_{12} \leq \sqrt{2}R \\ \frac{1}{2\pi x}\sqrt{1 - \frac{x^2}{4}} & \sqrt{2}R < \Delta_{12} \leq 2R \end{cases} .
$$

with $x = \Delta_{12}/R$, while for $\Delta_{12} > 2R$, one recovers the result $\pi R^2/4$. The SISCone active area is plotted in figure 10 and it is compared to the results for k_t and Cambridge/Aachen in figure 8. One notes that the SISCone result is both qualitatively and quantitatively much further from the passive result than was the case for k_t and Cambridge/Aachen.

The main reason why the 2-point active area is larger than the 1-point active area (whereas we saw the opposite behaviour for the passive areas) is that the presence of the 2-particle configuration causes a number of the pure-ghost cones that were present in the 1-particle case to now contain the second particle and therefore be unstable. Since these pure ghost cones are responsible for reducing the jet area relative to the passive result (during the split–merge step), their absence causes the

20

active area to be 'less-reduced' than in the 1-particle case.

3.3 Area scaling violation

We can write the order α_s contribution to the average active area in a manner similar to the passive area case, eq. (13):

$$\langle A_{\text{JA},R}\rangle = A_{\text{JA},R}(0) + \langle \Delta A_{\text{JA},R}\rangle\,, \tag{39}$$

where we have used the relation $A_{\text{JA},R}(\text{one-particle-jet}) \equiv A_{\text{JA},R}(0)$ and where

$$\langle \Delta A_{\text{JA},R}\rangle \simeq \int_0 d\Delta_{12} \int_{Q_0/\Delta_{12}}^{p_{t1}} dp_{t2} \frac{dP}{dp_{t2}\,d\Delta_{12}} (A_{\text{JA},R}(\Delta_{12}) - A_{\text{JA},R}(0))\,. \tag{40}$$

Note that compared to eq. (13) we have removed the explicit upper limit at $2R$ on the Δ_{12} integral since, for active areas of sequential recombination algorithms, $(A_{\text{JA},R}(\Delta_{12}) - A_{\text{JA},R}(0))$ may be non zero even for $\Delta_{12} > 2R$. Note also the notation for averages: we use $\langle \cdots \rangle$ to refer to an average over perturbative emissions, while $\langle \cdots \rangle_g$, implicitly contained in $A_{\text{JA},R}$ (see eq. (28)), refers to an average over ghost ensembles. We now proceed as in section 2.3, and write

$$\langle \Delta A_{\text{JA},R}\rangle \simeq D_{\text{JA},R}\frac{C_1}{\pi b_0} \ln \frac{\alpha_s(Q_0)}{\alpha_s(Rp_{t1})}\,, \qquad D_{\text{JA},R} = \int_0 \frac{d\theta}{\theta}(A_{\text{JA},R}(\theta) - A_{\text{JA},R}(0))\,, \tag{41}$$

where for brevity we have given just the running-coupling result. The coefficients for the anomalous dimension of the active area for the various algorithms are

$$D_{k_t,R} \simeq 0.52\,\pi R^2\,, \tag{42a}$$

$$D_{\text{Cam},R} \simeq 0.08\,\pi R^2\,, \tag{42b}$$

$$D_{\text{SISCone},R} \simeq 0.1246\,\pi R^2\,, \tag{42c}$$

where the SISCone result has been obtained by integrating the analytical result, eq. (38), while the results for k_t and Cambridge/Aachen have been obtained both by integrating the 2-point active-area results shown in fig. 8 and by a direct Monte Carlo evaluation of eq. (41). Note that while the coefficients for k_t and Cambridge/Aachen are only slightly different from their passive counterparts, the one for SISCone has the opposite sign relative to the passive one.

The treatment of higher-order fluctuations of active areas is more complex than that for the passive ones, where the one-particle area was a constant. We can separate the fluctuations of active areas into two components, one (described above) that is the just the one-particle result and the other, $\langle \Delta\Sigma^2_{\text{JA},R}\rangle$, accounting for their modification in the presence of perturbative radiation:

$$\langle \Sigma^2_{\text{JA},R}\rangle = \Sigma^2_{\text{JA},R}(0) + \langle \Delta\Sigma^2_{\text{JA},R}\rangle\,, \tag{43}$$

where $\Sigma_{\text{JA},R}(0)$ is given by eqs. (32a,b) for the k_t and the Cambridge/Aachen algorithms respectively, and it is equal to zero for the SISCone algorithm. The perturbative modification $\langle \Delta\Sigma^2_{\text{JA},R}\rangle$ is itself now driven by two mechanisms: the fact that the second particle causes the average area to change, and that it also causes modifications of the fluctuations associated with the sampling over

many ghost sets. We therefore write

$$\langle \Delta \Sigma_{\mathrm{JA},R}^2 \rangle \simeq S_{\mathrm{JA},R}^2 \frac{C_1}{\pi b_0} \ln \frac{\alpha_s(Q_0)}{\alpha_s(R p_{t1})} \,, \tag{44}$$

$$S_{\mathrm{JA},R}^2 = \int_0 \frac{d\theta}{\theta} \left[(A_{\mathrm{JA},R}(\theta) - A_{\mathrm{JA},R}(0))^2 + \Sigma_{\mathrm{JA},R}^2(\theta) - \Sigma_{\mathrm{JA},R}^2(0) \right] \tag{45}$$

$$= \int_0 \frac{d\theta}{\theta} (A_{\mathrm{JA},R}^2(\theta) - A_{\mathrm{JA},R}^2(0)) - 2 A_{\mathrm{JA},R}(0) D_{\mathrm{JA},R} \,, \tag{46}$$

where as usual we neglect contributions that are not enhanced by any logarithm or that are higher-order in α_s. The details of how to obtain these results are given in Appendix C.

The coefficient $S_{\mathrm{JA},R}^2$ can be determined only numerically for the k_t and Cambridge/Aachen algorithms, while for the SISCone algorithm the result can be deduced from eq. (38) together with the knowledge that $\Sigma_{\mathrm{SISCone},R}(\theta) \equiv 0$:

$$S_{k_t,R}^2 \simeq (0.41 \, \pi R^2)^2 \,, \tag{47a}$$

$$S_{\mathrm{Cam},R}^2 \simeq (0.19 \, \pi R^2)^2 \,, \tag{47b}$$

$$S_{\mathrm{SISCone},R}^2 \simeq (0.0738 \, \pi R^2)^2 \,. \tag{47c}$$

Again, both the values and their ordering are similar to what we have obtained for the passive areas (see eq. (22)).

3.4 n-particle properties and large-n behaviour

3.4.1 k_t algorithm

As for the passive area, the k_t algorithm's active area has the property that it can be expressed as a sum of individual particle areas:

$$A_{k_t,R}(J) = \sum_{p_i \in J} A_{k_t,R}(p_i) \,. \tag{48}$$

This is the case because the presence of the momentum scale k_t in the distance measure means that all ghosts cluster among themselves and with the hard particles, before any of the hard particles start clustering between themselves. However in contrast to the passive-area situation, there is no known simple geometrical construction for the individual particle area.

3.4.2 Equivalence of all areas for large n

The existence of different area definitions is linked to the ambiguity in assigning 'empty space' to any particular jet. In the presence of a sufficiently large number of particles n, one expects this ambiguity to vanish because real particles fill up more and more of the empty space and thus help to sharpen the boundaries between jets. Thus in the limit of many particles, all sensible area definitions should give identical results for a given jet.

To quantify this statement, we examine (a bound on) the scaling law for the relation between the density of particles and the magnitude of the potential differences between different area definitions. We consider the limit of 'dense' coverage, defined as follows: take square tiles of some size and use them to cover the rapidity–azimuth cylinder (up to some maximal rapidity). Define λ as the smallest value of tile edge-length such that all tiles contain at least one particle. In an event consisting of

uniformly distributed particles, λ is of the same order of magnitude as the typical interparticle distance. The event is considered to be dense if $\lambda \ll R$.

Now let us define a boundary tile of a jet to be a tile that contains at least one particle of that jet and also contains a particle from another jet or has an adjacent tile containing one or more particle(s) belonging to a different jet. We expect that the difference between different jet-area definitions cannot be significantly larger than the total area of the boundary tiles for a jet.

The number of boundary tiles for jets produced by a given jet algorithm (of radius R) may scale in a non-trivial manner with the inter-particle spacing λ, since the boundary may well have a fractal structure. We therefore parametrise the average number of boundary tiles for a jet, $N_{b,\mathrm{JA},R}$ as

$$ N_{b,\mathrm{JA},R} \sim \left(\frac{R}{\lambda} \right)^{\zeta_{\mathrm{JA}}} , \tag{49} $$

where the fractal dimension $\zeta_{\mathrm{JA}} = 1$ would correspond to a smooth boundary. The total area of these boundary tiles gives an upper limit on the ambiguity of the jet area

$$ \langle |a_{\mathrm{JA},R} - A_{\mathrm{JA},R}| \rangle \lesssim N_{b,\mathrm{JA},R}\, \lambda^2 \sim R^{\zeta_{\mathrm{JA}}} \lambda^{2-\zeta_{\mathrm{JA}}} , \tag{50} $$

and similarly for the difference between active or passive and Voronoi areas. As long as $\zeta_{\mathrm{JA}} < 2$ the differences between various area definitions are guaranteed to vanish in the infinitely dense limit, $\lambda \to 0$. We note that $\zeta_{\mathrm{JA}} = 2$ corresponds to a situation in which the boundary itself behaves like an area, $i.e.$ occupies the same order of magnitude of space as the jet itself. This would be visible in plots representing jet active areas (such as fig. 4), in the form of finely intertwined jets. We have seen no evidence for this and therefore believe that $1 \leq \zeta_{\mathrm{JA}} < 2$ for all three jet algorithms considered here.

In practice we expect the difference between any two area definitions to vanish much more rapidly than eq. (50) as $\lambda \to 0$, since the upper bound will only be saturated if, for every tile, the difference between two area definitions has the same sign. This seems highly unlikely. If instead differences in the area for each tile are uncorrelated (but each of order λ^2) then one would expect to see

$$ \langle |a_{\mathrm{JA},R} - A_{\mathrm{JA},R}| \rangle \sim \sqrt{N_{b,\mathrm{JA},R}}\, \lambda^2 \sim R^{\zeta_{\mathrm{JA}}/2} \lambda^{2-\zeta_{\mathrm{JA}}/2} . \tag{51} $$

We have measured the fractal dimension for the k_t and Cambridge/Aachen algorithms and find $\zeta_{k_t} \simeq \zeta_{\mathrm{Cam}} \simeq 1.20 - 1.25$.[13] Note that any measurement of the fractal dimension of jet algorithms in real data would be severely complicated by additional structure added to jets by QCD branching, itself also approximately fractal in nature.

The fact that active and passive (or Voronoi) areas all give the same result in dense events has practical applications in real-life situations where an event is populated by a very large number of particles (heavy ion collisions being an example). In this case it will be possible to choose the area type which is fastest to compute (for instance the Voronoi area) and use the results in place of the active or passive one.

4   Back reaction

So far we have considered how a set of infinitely soft ghosts clusters with a hard jet, examining also cases where the jet has some finitely-soft substructure. This infinitely-soft approximation for the

---

[13]This has been measured on pure ghost jets, because their higher multiplicity facilitates the extraction of a reliable result, however we strongly suspect that it holds also for single-particle jets.

ghosts is not only adequate, but also necessary from the point of view of properly defining jet areas. However if we are to understand the impact of minimum-bias (MB) and underlying-event radiation on jet finding — the original motivation for studying areas — then we should take into account the fact that these contributions provide a dense background of particles at some small but *finite* soft scale $\sim \Lambda_{QCD}$.

This has two main consequences. The first (more trivial) one is that the minimum-bias or underlying event can provide an alternative, dynamic infrared cutoff in the p_{t2} integration in equations such as eq. (13): assuming that the density of MB transverse momentum per unit area is given by ρ then one can expect that when $p_{t2} \ll \pi R^2 \rho$, the presence of p_2 will no longer affect the clustering of the ghosts (*i.e.* the MB particles). In this case the p_{t2} integral will then acquire an effective infrared cutoff $\sim \pi R^2 \rho$ and in expressions such as eqs. (15,16), Q_0 will be replaced by $\pi R^3 \rho$. Note that neither with infinitely nor finitely-soft ghosts do we claim control over the coefficient in front of this cutoff, though we do have confidence in the prediction of its R dependence for small R.[14]

The second consequence of the finite softness of the MB contribution is that the addition of the MB particles can modify the set of non-MB particles that make it into a jet. We call this 'back-reaction' and it is the main subject of this section.

That back reaction should happen is quite intuitive as concerns non-MB particles whose softness is commensurate with the MB scale. However the extent to which it occurs depends significantly on the jet algorithm. Furthermore for some algorithms it can also occur (rarely) even for non-MB particles that are much harder than the MB scale.

As with the studies of areas, there are two limits that can usefully be examined for back-reaction: the illustrative and mathematically simpler (but less physical) case in which the MB is pointlike and the more realistic case with diffuse MB radiation.

4.1   Back reaction from pointlike minimum-bias

Let us first calculate back reaction in the case of pointlike minimum bias. We will consider minimum-bias particles with transverse momentum p_{tm} distributed uniformly on the y–ϕ cylinder with density $\nu_m \ll 1$. We use a subscript m rather than g to differentiate them from ghost particles, the key distinction being that p_{tm} is small but finite, where p_{tg} is infinitesimal.

Let us consider the situation in which a particle p_1, with large transverse momentum, $p_{t1} \gg p_{tm}$, has emitted a soft particle p_2 on a scale commensurate with the minimum-bias particles, $p_{t2} \sim p_{tm}$. We shall calculate the probability that p_2 was part of the jet in the absence of the minimum-bias particle, but is *lost* from it when the minimum-bias particle is added. This can be written

$$\frac{dP_{\mathrm{JA},R}^{(L)}}{dp_{t2}} = \int d\phi_m dy_m \nu_m \int d\Delta_{12} \frac{dP}{dp_{t2}\, d\Delta_{12}} H_{\mathrm{JA},R}(p_2 \in J_1)\, H_{\mathrm{JA},R}(p_2 \notin J_1 | p_m)\,, \qquad (52)$$

where $H_{\mathrm{JA},R}(p_2 \in J_1)$ is 1 (0) if, in the absence of p_m, p_2 is inside (outside) the jet that contains p_1. Similarly, $H_{\mathrm{JA},R}(p_2 \notin J_1 | p_m)$ is 1 (0) if, in the presence of p_m, p_2 is inside (outside) the jet that contains p_1. One can also define the probability for p_2 to not be part of the jet in the absence of the minimum-bias particle, but to be *gained* by the jet when the minimum-bias particle is added,

$$\frac{dP_{\mathrm{JA},R}^{(G)}}{dp_{t2}} = \int d\phi_m dy_m \nu_m \int d\Delta_{12} \frac{dP}{dp_{t2}\, d\Delta_{12}} H_{\mathrm{JA},R}(p_2 \notin J_1)\, H_{\mathrm{JA},R}(p_2 \in J_1 | p_m)\,. \qquad (53)$$

---

[14]The coefficient can actually be determined quite straightforwardly, however since its impact is of the same order as other effects that we neglect (*i.e.* free of any logarithmic enhancements) we leave its determination to future work.

24

It is convenient to factor out the particle production probability as follows, in the small R limit,

$$\frac{dP_{\mathrm{JA},R}^{(L)}}{dp_{t2}} = \Delta_{12} \frac{dP}{dp_{t2}\, d\Delta_{12}}\bigg|_{\Delta_{12}=R} \nu_m\, b_{\mathrm{JA},R}^{(L)}(p_{t2}/p_{tm})\,, \tag{54}$$

where $b_{\mathrm{JA},R}^{(L)}(p_{t2}/p_{tm})$ can be thought of as the effective 'back-reaction area' over which the minimum-bias particle causes a loss of jet contents, given a $d\Delta_{12}/\Delta_{12}$ angular distribution for the jet contents:[15]

$$b_{\mathrm{JA},R}^{(L)}(p_{t2}/p_{tm}) = \int d\phi_m dy_m \int \frac{d\Delta_{12}}{\Delta_{12}} H_{\mathrm{JA},R}(p_2 \in J_1)\, H_{\mathrm{JA},R}(p_2 \notin J_1 | p_m)\,. \tag{55}$$

One can similarly define an effective back-reaction area for gain,

$$\frac{dP_{\mathrm{JA},R}^{(G)}}{dp_{t2}} = \Delta_{12} \frac{dP}{dp_{t2}\, d\Delta_{12}}\bigg|_{\Delta_{12}=R} \nu_m\, b_{\mathrm{JA},R}^{(G)}(p_{t2}/p_{tm})\,, \tag{56}$$

$$b_{\mathrm{JA},R}^{(G)}(p_{t2}/p_{tm}) = \int d\phi_m dy_m \int \frac{d\Delta_{12}}{\Delta_{12}} H_{\mathrm{JA},R}(p_2 \notin J_1)\, H_{\mathrm{JA},R}(p_2 \in J_1 | p_m)\,. \tag{57}$$

For sequential-recombination algorithms the H functions in eqs. (55,57) translate to a series of Θ-functions, *e.g.*

$$H_{k_t,R}(p_2 \in J_1) H_{k_t,R}(p_2 \notin J_1 | p_m) = \Theta(R - \Delta_{12})\Theta(\Delta_{1(2+m)} - R)\times$$
$$\times\, \Theta(\Delta_{1m} - \min(1, p_{t2}/p_{tm})\Delta_{2m})\Theta(\Delta_{12} - \min(1, p_{tm}/p_{t2})\Delta_{2m})\Theta(R - \Delta_{2m})\,, \tag{58}$$

for the k_t algorithm and

$$H_{\mathrm{Cam},R}(p_2 \in J_1) H_{\mathrm{Cam},R}(p_2 \notin J_1 | p_m) = \Theta(R - \Delta_{12})\Theta(\Delta_{1(2+m)} - R)\times$$
$$\times\, \Theta(\Delta_{1m} - \Delta_{2m})\Theta(\Delta_{12} - \Delta_{2m})\Theta(R - \Delta_{2m})\,, \tag{59}$$

for Cambridge/Aachen, where $\Delta_{1(2+m)}$ is the distance between p_1 and the recombined $p_2 + p_m$. Evaluating integrals with the above Θ-functions is rather tedious, but one can usefully consider the limit $p_{tm} \ll p_{t2} \ll p_{t1}$. This of physical interest because it relates to the probability that the minimum-bias particle induces changes in jet momentum that are much larger than p_{tm}, and a number of simplifications occur in this limit. Since

$$\Delta_{1(2+m)} = \left|\vec{\Delta}_{12} + \frac{p_{tm}}{p_{t2}}\vec{\Delta}_{2m}\right| = \Delta_{12} + \frac{p_{tm}}{p_{t2}}\frac{\vec{\Delta}_{12} \cdot \vec{\Delta}_{2m}}{\Delta_{12}} + \mathcal{O}\left(\frac{p_{tm}^2}{p_{t2}^2}R\right)\,, \tag{60}$$

p_2 must be close to the edge of the jet in order for it to be pulled out by p_m, $|\Delta_{12} - R| \ll 1$. Without loss of generality, we can set $y_1 = \phi_1 = \phi_2 = 0$, so that $\Delta_{12} = y_2 \simeq R$, and

$$\Delta_{1(2+m)} = y_2 + \frac{p_{tm}}{p_{t2}}(y_m - R) + \mathcal{O}\left(\frac{p_{tm}^2}{p_{t2}^2}R\right)\,. \tag{61}$$

We can then carry out the integrations over ϕ_m and y_2 straightforwardly, leading to following the result for loss at high p_{t2},

$$b_{k_t,R}^{(L)}(p_{t2}/p_{tm} \gg 1) \simeq b_{\mathrm{Cam},R}^{(L)}(p_{t2}/p_{tm} \gg 1) \simeq \int_R^{2R} dy_m \frac{(y_m - R)}{R}\frac{p_{tm}}{p_{t2}} 2\sqrt{R^2 - (y_m - R)^2} = \frac{2}{3}\frac{p_{tm}}{p_{t2}}R^2\,. \tag{62}$$

---

[15]Strictly speaking it is the integral over area of the probability of causing a loss of jet contents.
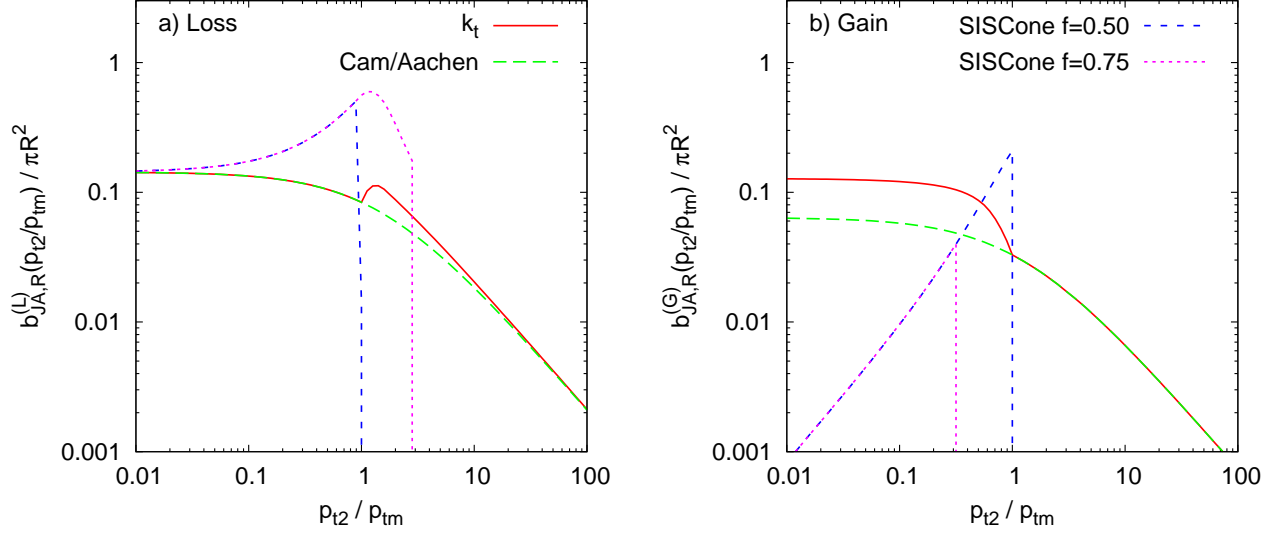
Figure 11: The effective area for back-reaction as a function of the ratio of the soft perturbative scale p_{t2} and the point-like minimum-bias scale p_{tm}, showing separately the loss (a) and gain (b) components for four jet definitions.

In a similar manner one obtains for the gain,

$$b_{k_t,R}^{(G)}(p_{t2}/p_{tm} \gg 1) \simeq b_{\text{Cam},R}^{(G)}(p_{t2}/p_{tm} \gg 1) \simeq \int_{R/2}^{R} dy_m \, \frac{(R-y_m)}{R} \frac{p_{tm}}{p_{t2}} 2\sqrt{R^2 - (R-y_m)^2} =$$
$$= \left(\frac{2}{3} - \frac{\sqrt{3}}{4} \right) \frac{p_{tm}}{p_{t2}} R^2 . \quad (63)$$

The results for general p_{t2}/p_{tm}, determined numerically, are shown in figure 11. The SISCone results are included and have the property that ,

$$b_{\text{SISCone},R}^{(L)}(p_{t2}/p_{tm}) = 0 \quad \text{for} \quad \frac{p_{t2}}{p_{tm}} > \frac{f}{1-f} , \qquad (64a)$$

$$b_{\text{SISCone},R}^{(G)}(p_{t2}/p_{tm}) = 0 \quad \text{for} \quad \frac{p_{t2}}{p_{tm}} > \frac{1-f}{f} . \qquad (64b)$$

i.e. at high p_{t2}, point-like minimum bias never induces back-reaction in the cone algorithm,[16] in contrast to the situation with the sequential recombination algorithms, for which back-reaction occurs with a suppressed, but non-zero probability $\sim p_{tm}/p_{t2}$. On the other hand, for $p_{t2} \sim p_{tm}$ back-reaction is more likely with the cone algorithm — the effective area over which the MB particle can cause a change in jet contents is $\sim 0.5\pi R^2$, to be compared to $\sim 0.1\pi R^2$ for the k_t and Cambridge/Aachen algorithms.

One may use the results eqs. (62)–(64) to determine the average change in jet-momentum due to back reaction. Because of the logarithmic spectrum of emissions $dP/(dp_{t2} \, d\Delta_{12})$, one finds that

---

[16]For large but finite p_{t1}, back-reaction can actually occur beyond the above limits, but only with probability $\sim p_{tm}/p_{t1}$.

it receives contributions from the whole logarithmic region $p_{tm} < p_{t2} < p_{t1}$,

$$\langle \Delta p_{t,\mathrm{JA},R}^{(G-L)} \rangle \simeq \int_{p_{tm}}^{p_{t1}} dp_{t2} p_{t2} \left[\frac{dP_{\mathrm{JA},R}^{(G)}}{dp_{t2}} - \frac{dP_{\mathrm{JA},R}^{(L)}}{dp_{t2}} \right] = \beta_{\mathrm{JA},R}\, \rho \cdot \frac{C_1}{\pi b_0} \ln \frac{\alpha_s(p_{tm}R)}{\alpha_s(p_{t1}R)}, \qquad (65)$$

(evaluated for fixed coupling), where $\rho = \nu_m p_{tm}$ corresponds to the average transverse momentum of minimum-bias radiation per unit area and

$$\beta_{\mathrm{JA},R} = \lim_{p_{t2} \to \infty} \frac{p_{t2}}{p_{tm}} \left(b_{\mathrm{JA},R}^{(G)}(p_{t2}/p_{tm}) - b_{\mathrm{JA},R}^{(L)}(p_{t2}/p_{tm}) \right). \qquad (66)$$

The structure of the correction in eq. (65) is very similar to that for the actual contamination from minimum bias, $\rho \langle \Delta a_{\mathrm{JA},R} \rangle$, with $\langle \Delta a_{\mathrm{JA},R} \rangle$ as determined in section 2.3: notably, for fixed coupling, the average back-reaction scales with the logarithm of the jet p_t. The coefficients $\beta_{\mathrm{JA},R}$,

$$\beta_{k_t,R} = \beta_{\mathrm{Cam},R} = -\frac{\sqrt{3}}{4} R^2 \simeq -0.1378 \pi R^2, \qquad (67a)$$

$$\beta_{\mathrm{SISCone},R} = 0. \qquad (67b)$$

can be directly compared to the results for the $d_{\mathrm{JA},R}$ there. The values are relatively small, similar in particular to what one observes for the Cambridge/Aachen algorithm, though of opposite sign.

Though the average change in jet momentum, both from scaling violations of the area and from back-reaction, have a similar analytical structure, it is worth bearing in mind that these similar analytical structures come about quite differently in the two cases. Regarding area scaling violations, a significant fraction of jets, $\sim \alpha_s \ln p_{t1}/p_{tm}$, are subject to a change in area $\sim R^2$ (cf. section 2.3), and a consequent modification of the minimum-bias contamination by a modest amount $\sim p_{tm}$. In contrast the average back reaction effect $\sim \alpha_s p_{tm} \ln p_{t1}/p_{tm}$, is due to large modifications of the jet momentum $\sim p_{t2}$ with $p_{tm} \ll p_{t2} \ll p_{t1}$ occurring rarely, with a differential probability that depends on p_{t2} as $\sim \alpha_s dp_{t2} p_{tm}/p_{t2}^2$. One consequence of this is that the mean square change in transverse momentum due to back-reaction is dominated by very rare modifications of jets in which $p_{t2} \sim p_{t1}$, giving

$$\left\langle \left(\Delta p_{t,\mathrm{JA},R}^{(G,L)} \right)^2 \right\rangle \sim \alpha_s p_{t1} p_{tm} \nu_m. \qquad (68)$$

Note that the coefficient of this dispersion is non-zero even for the SISCone algorithm, due to the residual probability $\sim p_{tm}/p_{t1}$ that it has, for finite p_{t1}, for any change in structure with $p_{tm} \ll p_{t2} \lesssim p_{t1}$.

4.2   Back reaction from diffuse MB

Let us now examine back reaction in the case where the minimum-bias radiation has a uniform diffuse structure, consisting of a high density of minimum-bias particles, $\nu_m \gg 1$. The relation between back reaction in the point-like and diffuse cases is rather similar to the relation between passive and active areas: key analytical features remain unaffected by the change in the structure of the minimum bias, but certain specific coefficients change.

We define the probability for loss in the presence of diffuse MB as

$$\frac{dP_{\mathrm{JA},R}^{(L)}}{dp_{t2}} = \lim_{\substack{\nu_m \to \infty \\ \rho = \nu_m \langle p_{tm} \rangle \text{ fixed}}} \left\langle \int d\Delta_{12} \frac{dP}{dp_{t2}\, d\Delta_{12}} H_{\mathrm{JA},R}(p_2 \in J_1)\, H_{\mathrm{JA},R}(p_2 \notin J_1 | \rho) \right\rangle_{\mathrm{MB}}, \qquad (69)$$

27

where the average is performed over the ensemble of MB configurations, and now $H_{\mathrm{JA},R}(p_2 \notin J_1 | \rho)$ is 1 (0) if p_2 is outside (inside) the jet containing p_1 in the presence of the specific minimum-bias instance. A similar equation holds for the gain probability.

Then, as with eq. (54) we factorise this,

$$\frac{dP_{\mathrm{JA},R}^{(L)}}{dp_{t2}} = \Delta_{12} \left.\frac{dP}{dp_{t2}\, d\Delta_{12}}\right|_{\Delta_{12}=R} B_{\mathrm{JA},R}^{(L)}(p_{t2}/\rho)\,, \tag{70}$$

into one piece related to the probability for perturbative emission and a second piece $B_{\mathrm{JA},R}^{(L)}$ that is the diffuse analogue of the effective back-reaction area $b_{\mathrm{JA},R}^{(L)}$ in the point-like case. Note however that it is not so obvious exactly what geometrical area $B_{\mathrm{JA},R}^{(L)}$ actually corresponds to.[17] Similarly, we introduce $dP_{\mathrm{JA},R}^{(G)}/dp_{t2}$ and $B_{JA,R}^{(G)}$ corresponding to the gain in the presence of a diffuse MB.

The only case for which we have analytical results for $B_{\mathrm{JA},R}^{(L,G)}(p_{t2}/\rho)$ is the SISCone algorithm, for which (with $f > f_{\max} \simeq 0.391$)

$$B_{\mathrm{SISCone},R}^{(L)}(p_{t2}/\rho) = B_{\mathrm{SISCone},R}^{(G)}(p_{t2}/\rho) = 0\,. \tag{71}$$

This is a consequence of the facts (a) that the addition of a uniform background of MB particles has no effect on the stability (or instability) of a specific cone, (b) that for $p_{t2} \ll p_{t1}$ the split–merge step is immaterial to the jet finding if p_2 is within the cone around p_1, and (c) that for p_2 outside the cone around p_1, the maximal possible overlap is of p_2's stable cone with that of p_1 is $f_{\max}$ and if $f > f_{\max}$ then the two cones will always be split, ensuring that p_2 remains in a jet distinct from p_1. We believe that real-life corrections to the zero in eq. (71) are proportional to the standard deviation of MB transverse-momentum density from point to point in the event.

Numerical results for the $B_{\mathrm{JA},R}^{(L,G)}(p_{t2}/\rho)$ for the k_t and Cambridge/Aachen algorithms are given in in figure 12 and compared to the results in the point-like case. One sees that the general functional form is rather similar though the normalisations are somewhat smaller (by a factor of 2 for loss, a factor ~ 10 for gain). The asymptotic large-p_{t2} behaviours are observed to be

$$B_{k_t,R}^{(L)}(p_{t2}/\rho) \simeq B_{\mathrm{Cam},R}^{(L)}(p_{t2}/\rho) \simeq 0.11\, \pi R^2\, \frac{\rho}{p_{t2}} \tag{72a}$$

$$B_{k_t,R}^{(G)}(p_{t2}/\rho) \simeq B_{\mathrm{Cam},R}^{(G)}(p_{t2}/\rho) \simeq 0.013\, \pi R^2\, \frac{\rho}{p_{t2}} \tag{72b}$$

As in the point-like case we can calculate the mean change in jet transverse momentum due to back reaction and we obtain

$$\langle \Delta p_{t,\mathrm{JA},R}^{(G-L)} \rangle \simeq \int_{p_{tm}}^{p_{t1}} dp_{t2} p_{t2} \left[\frac{dP_{\mathrm{JA},R}^{(G)}}{dp_{t2}} - \frac{dP_{\mathrm{JA},R}^{(L)}}{dp_{t2}} \right] = \mathcal{B}_{\mathrm{JA},R}\, \rho \cdot \frac{C_1}{\pi b_0} \ln \frac{\alpha_s(\rho R^3)}{\alpha_s(p_{t1}R)}\,, \tag{73}$$

with

$$\mathcal{B}_{\mathrm{JA},R} = \lim_{p_{t2}\to\infty} \frac{p_{t2}}{\rho} \left(B_{\mathrm{JA},R}^{(G)}(p_{t2}/\rho) - B_{\mathrm{JA},R}^{(L)}(p_{t2}/\rho) \right)\,. \tag{74}$$

Even though $b_{\mathrm{JA},R}^{(L,G)}$ had p_{t2}/p_{tm} as its argument and $B_{\mathrm{JA},R}^{(L,G)}$ has p_{t2}/ρ, the final expressions for the average back-reaction in the point-like and diffuse cases, eqs. (65), (73), have almost identical

---

[17]A related issue is that the precise choice of normalisation of $B_{\mathrm{JA},R}^{(L)}$ is somewhat arbitrary — our specific choice is intended to provide a meaningful connection with $b_{\mathrm{JA},R}^{(L)}$ in the large p_{t2} limit.
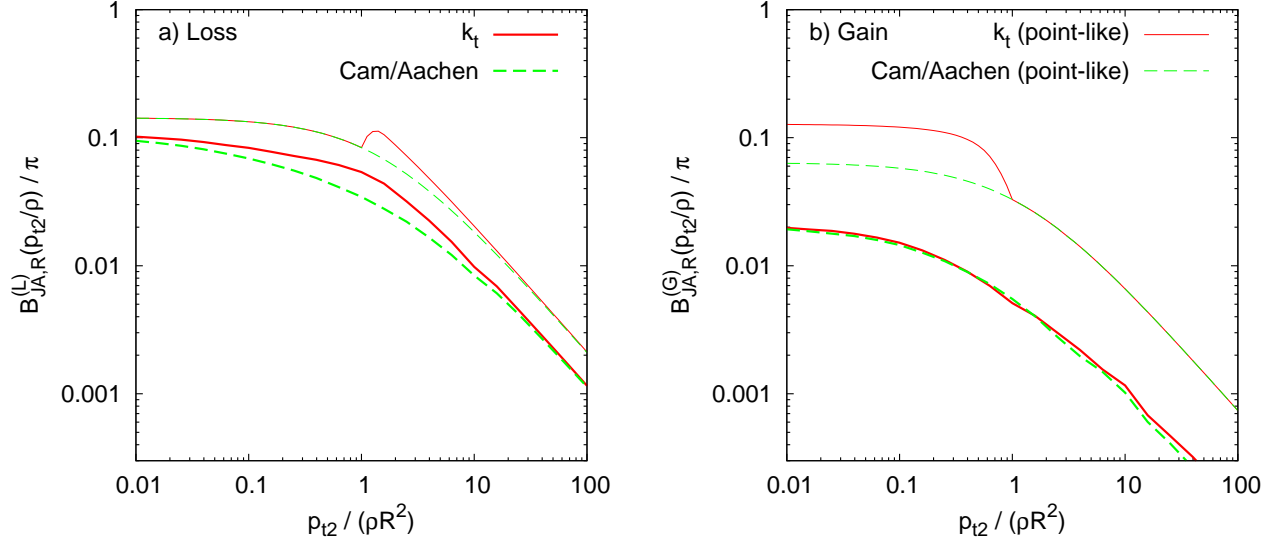
Figure 12: Numerical results for the diffuse effective back-reaction 'area', $B_{\mathrm{JA},R}^{(L,G)}(p_{t2}/\rho)$, for the k_t and Cambridge/Aachen algorithms, with the point-like results $b_{\mathrm{JA},1}^{(L,G)}(p_{t2}/p_{tm})$ shown for comparison also. Results obtained for $R = 1$ and verified also for $R = 0.7$.

forms — in particular the overall scale appearing in each is ρ and the only difference appears in the denominator for the argument of the logarithm. The coefficients are slightly smaller,

$$\mathcal{B}_{k_t,R} = \mathcal{B}_{\mathrm{Cam},R} \simeq -0.10\pi R^2 \,, \tag{75a}$$

$$\mathcal{B}_{\mathrm{SISCone},R} = 0 \,, \tag{75b}$$

and will again translate to modest effects compared to the overall minimum-bias contamination in the jets. Note also however that in general the scaling with R of $b_{k_t,R}^{(L/G)}(p_{t2}/p_{tm})$ and $B_{k_t,R}^{(L/G)}(p_{t2}/\rho)$ is subtly different. The former truly behaves like an area, in that $b_{k_t,R}^{(L/G)}(p_{t2}/p_{tm})/R^2$ is R-independent; the latter instead has the property that it is $B_{k_t,R}^{(L/G)}(R^2 p_{t2}/\rho)$ that is R-independent.

5 Areas in (simulated) real life

In this section we shall examine the properties of jet areas in the context of realistic events, as simulated with Herwig [17] and Pythia [18]. There are two purposes to doing so. Firstly we wish to illustrate the extent to which the simple single-gluon emission arguments of the previous section hold once one includes full parton-showering and hadronisation. Secondly, jet areas have the potential to play an important role in the estimation and subtraction of underlying event and pileup contamination, as discussed in [19]. The study of jet areas in realistic events can help to highlight some of the issues that arise in such a procedure.

5.1 Jet area distributions and anomalous dimension

Let us start with an investigation of the distribution of the areas of hard jets with $p_t \gtrsim 1\,\mathrm{TeV}$ in simulated LHC dijet events, fig. 13. The area distributions are shown for various 'levels' in the Monte Carlo: just after parton showering, after hadronisation, both with and without an UE, and
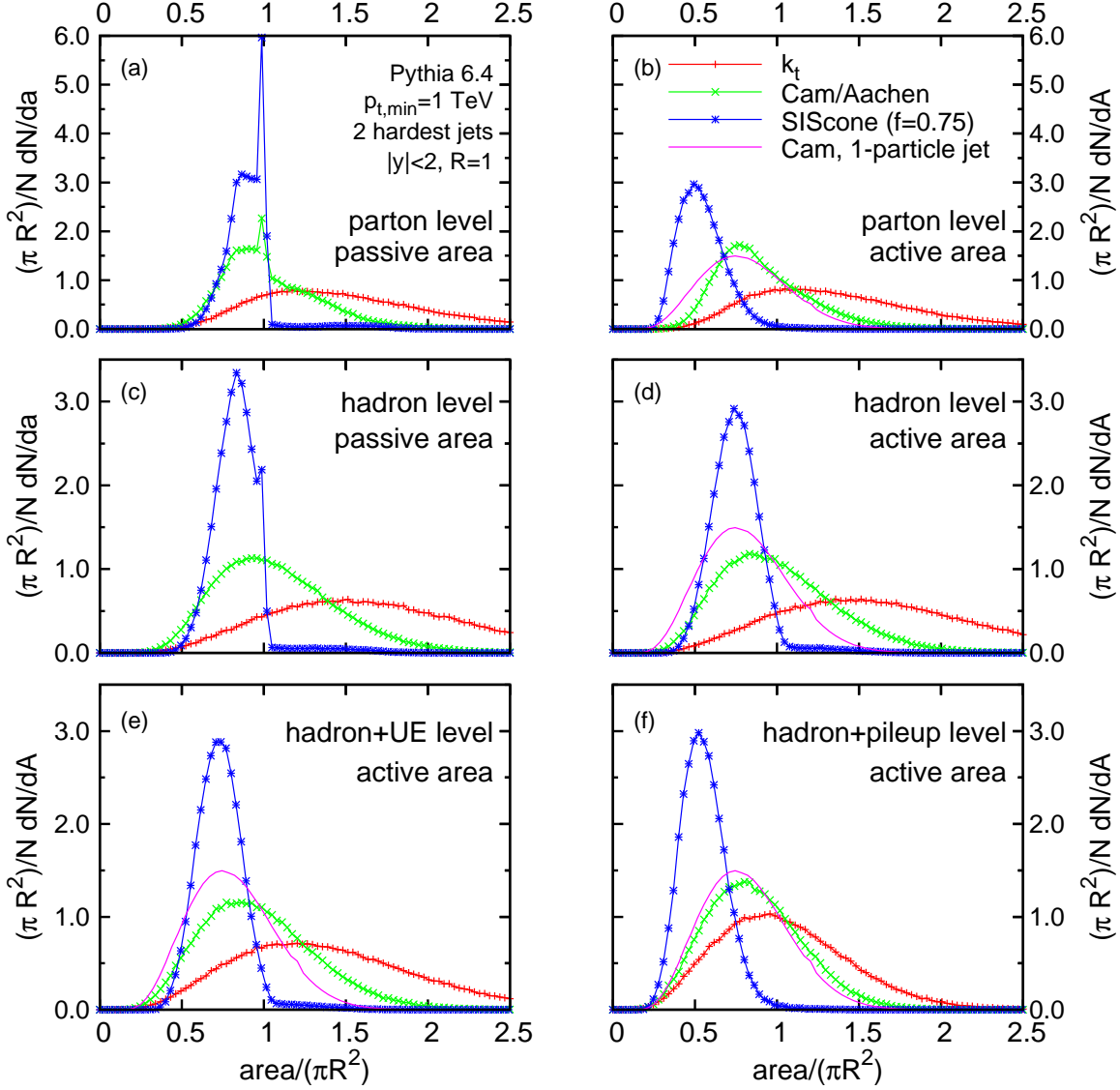
Figure 13: Distribution of active and passive areas of the two hardest jets in a range of simulated LHC dijet events, with a minimum p_t of $1\,\mathrm{TeV}$ in the Pythia [18] event generation (version 6.4, default tune). Only jets with $|y| < 2$ have been included in the histogram; 'parton' indicates parton-level, 'hadron' indicates hadron-level with the UE switched off, 'UE' corresponds to hadron-level with UE switched on and in the 'pileup' case the UE-level event is supplemented with additional minimum-bias events corresponding to $0.25\,\mathrm{mb}^{-1}$ per bunch crossing (about 25 simultaneous inter-actions). For all jet algorithms we use $R = 1$. A ghost area $\simeq 0.02$ was used throughout except for SISCone, where the ghost area was roughly 0.1 in the active area cases. Note that "area" in these plots corresponds to $a_{\mathrm{JA},R}(J_i)$ and $A_{\mathrm{JA},R}(J_i)$ respectively for the passive and active areas. As such the latter has been averaged over ghost ensembles (see eq. (28)) and the dispersion is a consequence of the event and jet structure. In contrast, the Cam/Aachen 1-particle jet result corresponds to $\pi R^2/N dN/dA(\text{1-particle-jet}|\{g_i\})$ and serves to illustrate how the impact of variability in the event structure in real events has consequences rather similar to that of the variability of ghost-particle ensembles in the theoretical arguments of the previous sections.

finally with a substantial pileup contribution (high-luminosity LHC, $\sim 25\ pp$ interactions per bunch crossing). We examine both passive and active areas.

The passive areas distributions at parton-shower level, fig. 13a, are those that are most amenable to understanding in terms of our analytical results. Firstly one notes that the SISCone and Cambridge/Aachen algorithms have a clear peak at $a = \pi R^2$. These two algorithms both have the property (cf. section 2.2) that the area is not affected by moderately collinear ($\Delta_{12} < R$ for SIS-Cone, $\Delta_{12} < R/2$ for Cambridge/Aachen) soft particle emission. Thus it is possible, even in the presence of parton showering (which is mostly collinear), for the passive area to remain πR^2. For the cone algorithm, the other main structure is a "shoulder" region $0.8 \lesssim a/(\pi R^2) \lesssim 1$, which coincides nicely with the range of values that can be induced by 2-particle configurations (cf. fig. 2). A similar shoulder exists for the Cambridge algorithm, which however additionally has jets with $a > \pi R^2$ — again the range of values spanned, up to $a \simeq 1.6\pi R^2$, is rather similar to what would be expected from the two-particle calculations. Further broadening of these distributions at the edges is attributable to parton-level states with more than two particles. In contrast the parton-level passive area distribution for the k_t algorithm seems less directly related to the 2-particle calculations of section 2.2. This can be understood by observing that the k_t passive area is modified even by rather collinear emissions, and the multiple collinear emissions present in parton showers add together to cause substantial broadening of the area distribution.

At parton shower level, there are relatively few particles in the event and there is no obvious boundary to the jet — the ghosts that we add provide a way of assigning that empty area. It is therefore not surprising to see significant differences between the two ways of adding ghosts, i.e. the passive and active area distributions. This is most marked for SISCone, as is to be expected from the results of section 3.1.2, which showed that for a 1-particle jet the active area is $\pi R^2/4$. There is a trace of this result in fig. 13b, where one sees that the SISCone distribution now extends down to $A = \pi R^2/4$. There is however no peak structure, presumably because even a highly collinear emission gives a slight modification of the area, cf. fig. 10 (the same argument given for the absence of a peak for the passive k_t area). For the Cambridge/Aachen and k_t algorithms, there is less difference between active and passive area distributions, again as expected.

As one moves to events with more particles, for example hadron level, fig. 13c and d, the particles themselves start to give a clearer outline to the jets. Thus the passive and active distributions are far more similar. This is less so for SISCone than the others, as is to be expected, for which one still sees a trace of the peak at $a = \pi R^2$.

For events with many particles, for example with the UE (fig. 13e) and then pileup (fig. 13f) added, the difference between passive and active area distributions is so small that we show only the active area result. These last two plots are the most relevant to phenomenology. A feature of both is that the dispersion is smallest for SISCone (but significantly different from zero) and largest for k_t, precisely as one would expect from eqs. (43) and (47). Another feature is how for Cambridge/Aachen the dispersion is essentially that associated with $\Sigma(0)$, the distribution being very similar to the 1-parton active area result, shown as the solid line. This similarity is strongest when there is pileup: the logarithmic enhancement that enters in eq. (44) is reduced because the pileup introduces a large value for $Q_0 \sim 20\,\text{GeV}$. In this case even the k_t algorithm starts to have a distribution of areas that resembles the 1-parton active area result, and for SISCone one once again sees signs of the lower limit at $A = \pi R^2/4$, the one-particle active area result.

While the plots of fig. 13 provide an illustration of many of the features that we have discussed in the earlier sections of this paper, the fact that we have not systematically calculated area distributions means that the discussion necessarily remains qualitative. For quantitative checks, one should instead examine the mean active area and its dispersion and compare them to the results of
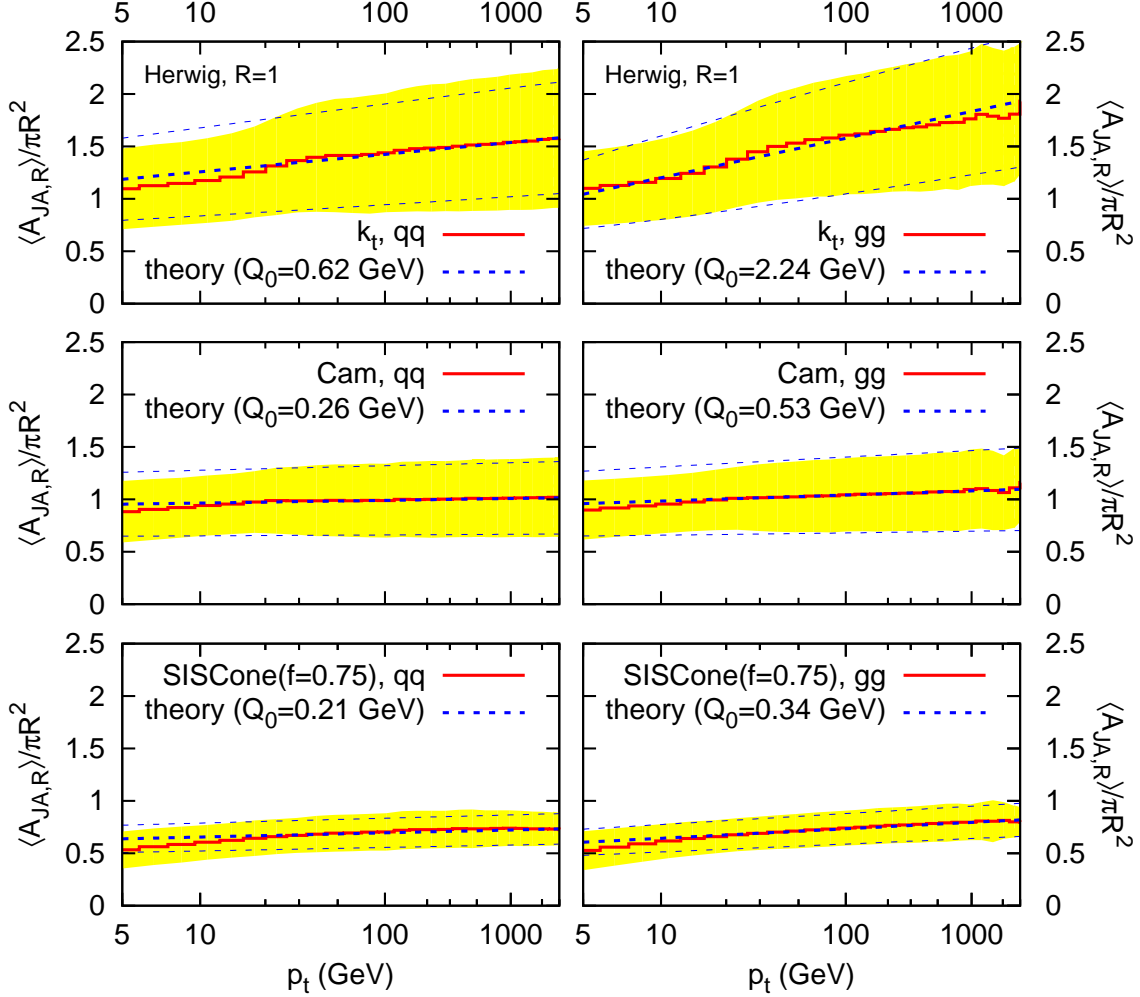
Figure 14: The mean (solid line) and standard deviation (band) of the active area for each of the two hardest jets in $qq \to qq$ and $gg \to gg$ events simulated with Herwig 6.5, as a function of the jet transverse momentum. The theory curves (thick dashed for mean, thin dashed for standard deviation) correspond to eqs. (39)–(47) with a 1-loop coupling ($\Lambda_{QCD} = 0.2\,\mathrm{GeV}$, $n_f = 5$), and Q_0 fitted (for the mean values). The areas have been obtained with ghosts of size 0.02 for k_t and Cambridge/Aachen, and 0.1 for SISCone. For all algorithms, $R = 1$. Note that the horizontal scale is uniform in $\ln \ln p_t/\Lambda$. Shown for pp collisions with $\sqrt{s} = 14\,\mathrm{TeV}$.

section 3. This is done in fig. 14, separately for $qq \to qq$ and for $gg \to gg$ scattering, as a function of the jet transverse momentum. The horizontal scale has been chosen uniform in $\ln \ln p_t / \Lambda$ so that our predictions correspond to straight lines.

The predictions of section 3 set the slope of those lines for each algorithm, and the agreement is reasonable in all cases. The infrared cutoff scale Q_0 is not predicted, and may differ both between quark and gluon jets and between algorithms. The values for Q_0 have therefore been fitted, using the results for the mean, and are consistent with a general non-perturbative origin (modulo issues in SISCone discussed below). The standard deviation (indicated by the band for the Monte Carlo simulation and the thin dashed lines for the theory result) is then entirely predicted, and also agrees remarkably well. Thus, overall, our simple analytical calculations provide a surprisingly successful picture of the mean and dispersions for various algorithms, over a range of jet transverse momenta. This is all the more striking considering that the calculation is based on just the first term in a series $\alpha_s^n \ln^n p_t$, and is in part based on a small angle approximation.

Some remarks are due regarding the Q_0 values. Two clear patterns emerge: it is largest for k_t, smallest for SISCone, and systematically larger for gluon jets than for quark jets. In the case of SISCone with quark jets, the value is uncomfortably close to the value of $\Lambda_{QCD} = 0.2\,\mathrm{GeV}$ used in the one-loop coupling. It may be that this low value is an artefact whose origin lies in the finite density both of actual event particles and of ghosts (the latter due to speed limitations in SISCone): when one has a limited density of particles and/or ghosts, the measured area may be intermediate between the passive and "ideal" active areas; because SISCone has such a large difference between passive and active areas (a factor of 4 for the 1-particle results), the finite density effects can be significant, and so Q_0 may be taking an extreme value in order to compensate for this.

A final comment concerns the choice of event generator. Here we have shown results from Herwig (v. 6.5). Pythia with the original shower (the default choice in v. 6.4) is known to have difficulties reproducing anomalous dimensions associated with soft large angle radiation [26, 27], whereas the new shower [28] mostly resolves these issues [27]. Similar concerns are potentially relevant here too. We actually find that both Pythia showers give results similar to Herwig's (and ours) in all cases except the k_t algorithm, for which both Pythia showers give a slope a factor of two smaller than expected. This suggests that an experimental measure of the p_t dependence of jet areas might provide some non-trivial constraints on parton-showering dynamics.

5.2   Jet areas and pileup subtraction

One of the main uses of measured jet areas is in the subtraction of pileup [19]. As we have seen there are two contributions to the modification of a jet's p_t in the presence of pileup: the pileup particles can end up in the jet, and assuming a roughly uniform distribution of pileup (as in the case in the limit of many simultaneous minimum-bias collisions), the resulting change in jet p_t will be proportional to the jet's area, by definition. Of course the pileup is not exactly uniform in a given event and this must also be accounted for. The only other modification of the jet's p_t comes from the back reaction. This gives us the following equation for the modification of a jet's transverse momentum in the presence of pileup, originally stated in [19],

$$\Delta p_t = A\rho \pm \sigma\sqrt{A} + \Delta p_t^B \,, \tag{76}$$

where, as before, ρ is the mean amount of transverse momentum per unit area that has been added to the event by pileup; σ measures the fluctuations of the pileup from point-to-point within the event (defined as the standard deviation of the distribution of pileup across many squares of area 1); and Δp_t^B is the net change in transverse momentum due to back reaction. An important point
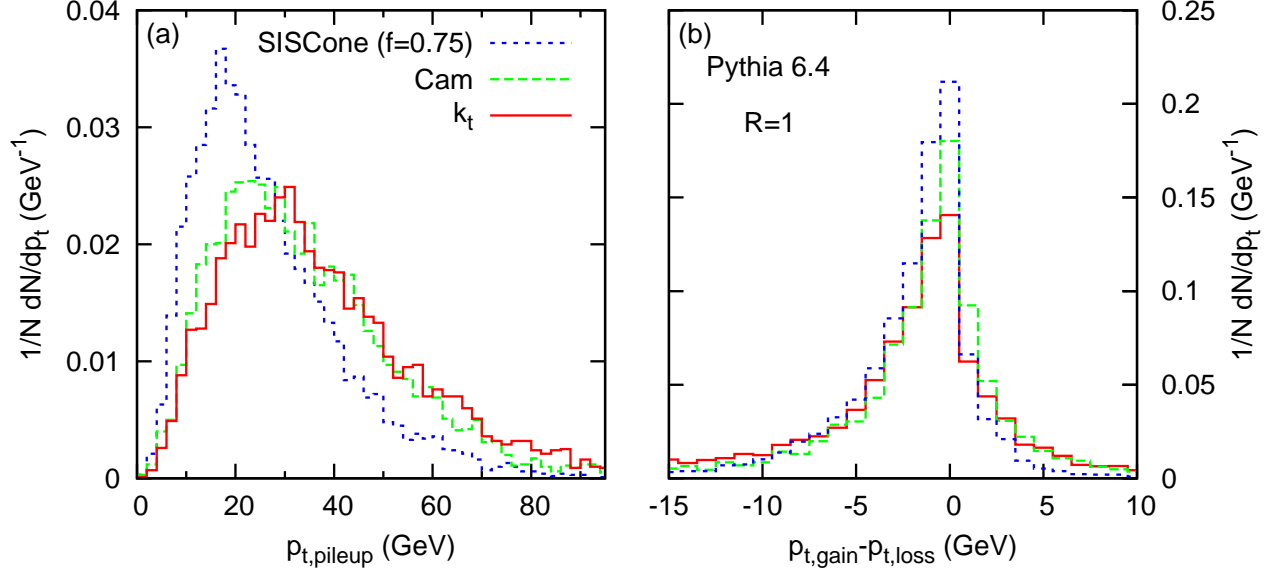
Figure 15: Both plots refer to events generated with Pythia 6.4 in which the two hardest jets have $p_t > 200\,\text{GeV}$ and are both situated at $|y| < 2$. Plot (a) shows the distribution of transverse momentum of the pileup particles that entered each of the jets; plot (b) shows the back reaction on the jet, *i.e.* the component of the net change in transverse momentum of the jet that is associated with reassignment of the non-pileup particles in the presence of pileup. The pileup was generated with Pythia 6.4. A jet in an event with pileup is matched to the jet in the event without pileup with which it overlaps most, the overlap being defined as the transverse component of the sum of the momenta of all constituents that are present in both jets (*i.e.* the method introduced in [10]).

here is that A is the area of the jet *after* the pileup has been added — the fact that the area is not an IR safe quantity means that it is not the same before and after pileup, and it is in the latter case that it correctly measures how much pileup will have entered the jet.

The fluctuations in the jet p_t due to the direct pileup contribution can be written

$$\langle \Delta p_t^2 \rangle_{\text{PU}} - \langle \Delta p_t \rangle_{\text{PU}}^2 \simeq \langle \Sigma_{\text{JA},R}^2 \rangle \langle \rho \rangle^2 + \langle A_{\text{JA},R} \rangle \langle \sigma^2 \rangle + \langle A_{\text{JA},R} \rangle^2 (\langle \rho^2 \rangle - \langle \rho \rangle^2), \qquad (77)$$

where the area-related averages are those for events including the pileup (*i.e.* $Q_0 \sim R^2 \rho$ in calculations of anomalous dimensions), and we separately account both for point to point fluctuations of the pileup within an event (second term), as well as fluctuations in the overall level of pileup from event to event (third term).

The direct pileup contribution to a sample of jets is shown in fig. 15a corresponding to about 25 pp interactions per bunch crossing (high-luminosity LHC), as simulated with Pythia tune A. One observes that SISCone has smaller average pileup contamination and smaller fluctuations in this contamination, which is consistent with our analytical calculations. More surprisingly, at least at first sight, k_t has only slightly larger and broader pileup contamination than Cambridge/Aachen, despite its larger area. However its area is generally larger only because of the anomalous dimension: since the logarithm that appears in the anomalous dimension with pileup, $\log p_t/\rho + \mathcal{O}(1)$, is not very large, the difference between k_t and Cambridge/Aachen is essentially absent. This is visible also in fig. 13f (the somewhat larger difference there is a consequence of the larger jet p_t).

The reader may have observed that we omitted the back-reaction term in eq. (77). The tail of the distribution for $\Delta p_t^B \sim p_{t,1}$ has the property $dP/d\Delta p_t^B \sim \alpha_s \rho/(\Delta p_t^B)^2$, so the standard
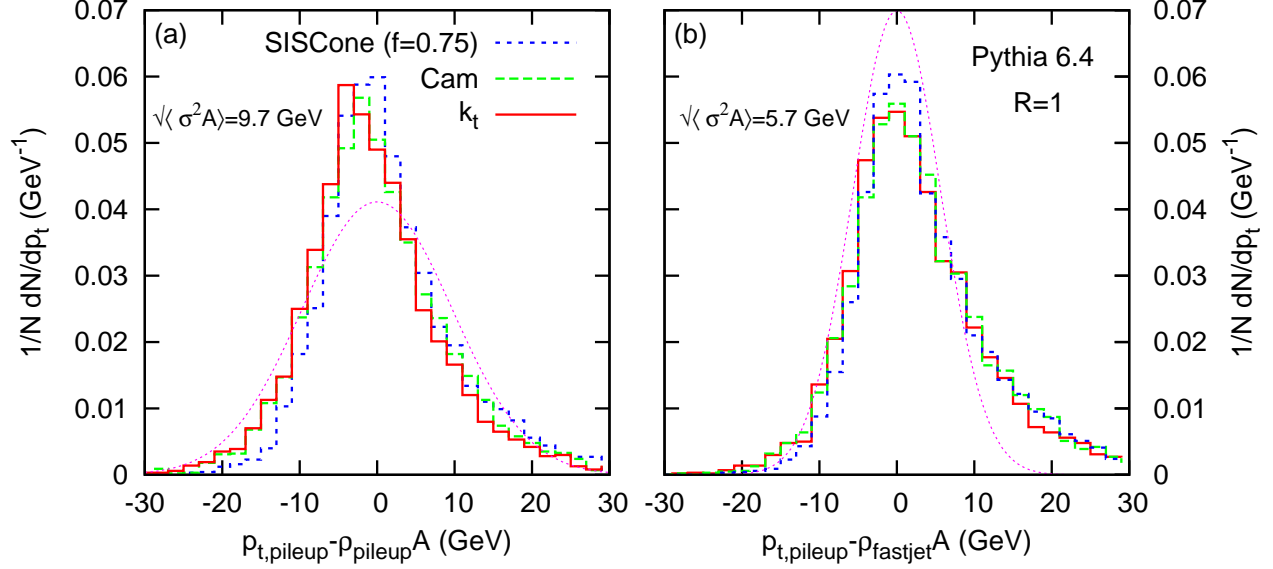
34

Figure 16: For the same events as in fig. 15, the transverse momentum contamination from pileup, minus its estimate $A\rho$ with ρ estimated event-by-event and A obtained jet-by-jet (strictly for A we have used the transverse component of the 4-vector area). It is compared to a Gaussian distribution with standard deviation $\sqrt{\langle\sigma^2 A\rangle} \simeq \sqrt{\langle\sigma^2\rangle\langle A\rangle}$. In (a) ρ and σ have been estimated by splitting the y-ϕ cylinder into squares of area 1 up to $|y| = 3.5$, calculating the sum of the scalar p_t of the pileup particles in each of them; ρ and σ are then given by the average and standard deviation of the contents of the squares. In (b) ρ and σ have been evaluated by the procedure of [19], $i.e.$ based on median and related percentiles of the distribution of p_t/A for all jets up to $|y| = 3.5$, carried out on the sample of jets obtained with the k_t algorithm with $R = 0.5$.

deviation is dominated by large $\Delta p_t^B \sim p_{t1}$, giving a contribution to to eq. (77) of order $\alpha_s\rho p_{t1}$. For large p_{t1} this is parametrically bigger than the direct pileup contributions, however it is dominated by extremely rare events (those in which a hard particle in the jet is just near the edge of the jet), a fraction of $\sim \alpha_s\rho/p_{t1}$ of all events. Consequently it provides a poor estimate of the typical back-reaction.

The actual distribution for the back reaction is shown in fig. 15b, and one sees that the change in momentum is nearly always limited to a few GeV, centred close to zero. It is slightly skewed towards loss, which is consistent with the larger loss than gain probabilities found in section 4.[18] Actual pileup is not perfectly uniform and for this reason the back-reaction with SISCone is not exactly zero as would have been expected from section 4.2 for diffuse radiation. Nevertheless the gain in particular is smaller than for other algorithms, which happens to coincide with what we would expect from fig. 11 for loss and gain with pointlike pileup.

Let us finally come to the question of subtraction of pileup. The equation proposed in [19] was

$$p_{tj}^{(\text{sub})} = p_{tj} - A_j\rho. \tag{78}$$

There we estimated ρ from the event itself and effectively subtracted both UE and pileup. Here let us estimate ρ purely from the pileup particles (with Monte Carlo simulations this is possible) and

---

[18]Actually, for the k_t and Cambridge/Aachen algorithms, one can verify that the high-Δp_t tails of the distributions are consistent with the $dP^{(L/G)}/d\Delta p_t \simeq \mathcal{B}^{(L/G)} C_1 \frac{2\alpha_s}{\pi} \frac{\rho}{(\Delta p_t)^2}$ expectation from section 4.2.

compare the difference between the p_t added by the pileup and the p_t removed by the our subtraction procedure. The distribution of this difference is shown in fig. 16. The two plots correspond to two different ways of measuring ρ (either directly from the scalar sum of p_t's of pileup particles, left, or with the median-based method of [19], which is applicable also in data). In both cases the distributions are clearly centred on zero, indicating a successful subtraction of the pileup.

Of the fluctuation terms in eq. (77), only the middle $\langle A_{\mathrm{JA},R} \rangle \langle \sigma^2 \rangle$ piece should remain after subtraction. This is consistent with the distributions being much narrower than those in fig. 15a. This is an important point as it means that, in addition to subtracting the average contamination due to pileup, we also reduce its effect on the p_t fluctuations. For example, if one wants to look at the top mass spectrum, the corresponding peak will appear narrower (and with a better signal-to-background ratio) after subtraction than before. As a quantitative test, in fig. 16, we show also a Gaussian of width $\langle A_{\mathrm{JA},R} \rangle \langle \sigma^2 \rangle$. Here σ has been estimated in each event again either directly or with the methods of [19]. Whereas the ρ values obtained with the two methods are similar, those for σ differ noticeably. This may in part be due to an occasional, hard pointlike contribution in the minimum-bias events, causing the fluctuations measured by σ^2 to be non-Gaussian. This would explain why the statistically more standard evaluation of σ shown in Fig. 16a leads to slightly too broad a distribution. In contrast the evaluation of [19], in Fig. 16b, is designed specifically to be insensitive to the pointlike component, on the grounds that one wants a measure not influenced by the main hard scatter. It provides a reasonable estimate of the lower half of the distribution, but undershoots the (asymmetric) upper tail, which one can probably ascribe to the rare pointlike component of the minimum bias collisions.

6   Conclusions

The concept of a jet area is one that initially seems rather intuitive. Yet, as we have seen in this article, there is a wealth of physics associated with it, both in terms of how one defines it and as concerns the interplay between a jet's internal structure and its area.

Our guiding principle in defining jet areas has been that they should provide a measure of a jet's susceptibility to additional underlying event and pileup radiation, in the limit in which these are infinitely soft. Two opposite hypotheses for the structure of such radiation, pointlike or diffuse, lead to two definitions of the area, respectively passive or active, both calculated in practice with the help of infinitely soft "ghost" particles. The two definitions can be used with any infrared safe jet algorithm, and we have applied them both to sequential recombination algorithms (k_t and Cambridge/Aachen) and to a stable-cone with split–merge algorithm, SISCone.

The area of a jet may depend on its substructure. For the simplest case of a jet made of a single hard particle, the passive area coincides with one's naive expectation of πR^2 for all jet algorithms considered here. The active area — that related to the more realistic, diffuse picture for UE and pileup — instead differs from πR^2, the most surprising result being that for SISCone, whose active area is $\pi R^2/4$, a consequence of the split–merge dynamics (the widely used midpoint algorithm behaves similarly). Thus the widespread assumption that cone-based algorithms automatically have an area of πR^2 is, in many cases, unjustified.

Real jets of course consist of more than a single particle. The first level of substructure involves the addition of a soft particle to the neighbourhood of the jet. This modifies the jet area for *all* of the algorithms considered, again including SISCone. The effects can be summarised in terms of an anomalous dimension which encodes how the jet's average area depends on its transverse momentum. We have calculated this to leading order and seen that it agrees remarkably well with

	a(1PJ)	A(1PJ)	σ(1PJ)	Σ(1PJ)	d	D	s	S	A(GJ)	Σ(GJ)
k_t	1	0.81	0	0.28	0.56	0.52	0.45	0.41	0.55	0.17
Cam/Aachen	1	0.81	0	0.26	0.08	0.08	0.24	0.19	0.55	0.18
SISCone	1	1/4	0	0	-0.06	0.12	0.09	0.07	—	—

Table 2: A summary of main results: passive (a, eq. (1)) and active (A, eq. (28)) areas for 1-particle jets (1PJ), the magnitude of the passive/active area fluctuations (σ, Σ, eqs. (19, 29)), followed by the coefficients of the respective anomalous dimensions (d, D, eqs. (16, 41); s, S, eqs. (21, 44)) in the presence of perturbative QCD radiation. The last two entries are the active area and its fluctuations for pure ghost jets (GJ), in the cases where the ghost jets are insensitive to the detailed ghost distribution. All results are normalised to πR^2, and rounded to the second decimal figure. Anomalous dimensions multiply powers of $\alpha_s^n \ln^n p_t/Q_0$ that for typical jet transverse momenta sum to something of order 1. Active-area and anomalous-dimension results hold only in the small-R limit, though finite-R corrections are small.

the p_t dependence of measures of the jet areas in hadron-level Monte Carlo simulations.

A summary of our main results for active and passive areas, their fluctuations and their anomalous dimensions is given in table 2. As is visible also in figure 14 for a broad range of p_t, there is a hierarchy in the areas of the jet algorithms, $\langle A_{\text{SISCone},R} \rangle \lesssim \langle A_{\text{Cam},R} \rangle \lesssim \langle A_{k_t,R} \rangle$. A likely consequence is that SISCone jets will be the least affected by UE contamination, though the extent to which this holds depends on the precise balance between pointlike and diffuse components of the UE. The above hierarchy might also suggest an explanation for the opposite hierarchy in the size of hadronisation corrections, observed in Monte Carlo studies for the different jet algorithms in [10]. There the k_t algorithm was seen to be least affected, which now appears natural, since a larger jet is less likely to lose momentum through hadronisation.

Jet areas, as well as being of interest from the point of view of understanding contamination from UE and pileup, are also the basis of recently developed methods for the jet-by-jet subtraction of this contamination [19]. The significant fluctuations of the area from one jet to the next mean that it is important to take into account the area of each individual jet, rather than assume some typical mean value. Among results here of relevance to the subtraction procedure, we highlight the demonstration that all areas (passive, active) are identical for highly populated events, which is important in ensuring that the subtraction procedure is free of significant ambiguities; we also remark on the calculation of back-reaction of pileup on jet structure, which is found to be a small effect.

There are many avenues for potential further study of jet areas. The calculations presented here have extracted only the leading logarithmic part of the first non-trivial order in α_s, and usually we have concentrated on properties of the mean area. The Monte Carlo results in section 5 also suggest interesting structures in the distributions of areas, and these merit further investigation. Another question for future work is that of the transition between passive and active areas if one considers a continuous transformation of the ghosts from pointlike to diffuse. Since real UE and pileup contamination is neither fully pointlike, nor fully diffuse, this question is of particular relevance for the stable-cone type algorithms, for which there is a large difference between the passive and active areas.

Finally, an understanding of the behaviour of jet areas can play a key role in the choice of parameters for jet algorithms, as well as in the design of new algorithms. One example of this concerns SISCone, for which we saw in section 3.1.2 that a split–merge overlap threshold $f \simeq 0.5$ can lead to the formation of "monster jets," whereas a choice of $f \simeq 0.75$, eliminates the problem, suggesting

that the latter is a more appropriate default value. Another example is given in a companion paper [29], where we show how to build an infrared and collinear safe jet algorithm, dubbed anti-k_t, whose passive and active areas for single-particle jets are both πR^2, and for which the area anomalous dimension is zero to all orders in α_s. These are precisely the properties that one imagines for an ideal cone algorithm — but it is only once one has the tools for a quantitative discussion of jet areas that one may establish whether a given algorithm actually has these properties.

Acknowledgements

A   Definitions of the three jet algorithms

Throughout this paper, we have considered three jet algorithms: k_t [14], Cambridge/Aachen [15] and SISCone [16].

The first and the second of these are sequential recombination algorithms. They introduce a distance d_{ij} between each pair of particles and a beam distance d_{iB} for each particle. At each step, the smallest distance is computed. If it involves a pair of particles, they are recombined using a E-scheme sum of four-momenta (*i.e.* direct addition of the four-momenta), otherwise the particle is clustered with the beam and called a 'jet'. The definition of the distance for the k_t algorithm is

$$d_{ij} = \min(k_{ti}^2, k_{tj}^2)\frac{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}{R^2}, \qquad d_{iB} = k_{ti}^2, \tag{79}$$

where $\Delta y_{ij} = y_i - y_j$ and $\Delta \phi_{ij} = \phi_i - \phi_j$. For the Cambridge/Aachen algorithm, the distance measures are

$$d_{ij} = \frac{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}{R^2}, \qquad d_{iB} = 1. \tag{80}$$

The SISCone jet algorithm is an infrared- and collinear-safe implementation of a modern cone algorithm. It first finds all stable cones of radius R, a stable cone being a circle in the (y, ϕ) plane such that the E-scheme sum of the momenta inside the cone points in the same direction as the centre of the cone. It then runs a Tevatron run II type [22] split–merge procedure to deal with overlapping stable cones. The stable cones are ordered in $\tilde{p}_t$, the scalar sum of the p_t of a cone's constituents, to produce the initial list of protojets. One takes the hardest protojet and finds the next hardest one that overlaps with it (its scaled transverse momentum being $\tilde{p}_{t,j}$). If the shared $\tilde{p}_t$ is larger than $f\tilde{p}_{t,j}$ (f is the split–merge overlap threshold parameter for the algorithm), they are merged, otherwise, the common particles are attributed to the protojet with the closer centre. If no overlaps are found, the protojet is added to the list of jets and removed from the list of protojets. This is repeated until no protojets remain.

B   Transition from one-particle jet to soft jet.

Having observed the different properties of the area of pure ghost jets and jets containing a hard particle, one may wonder what happens to the area of a jet containing a "trigger" particle whose
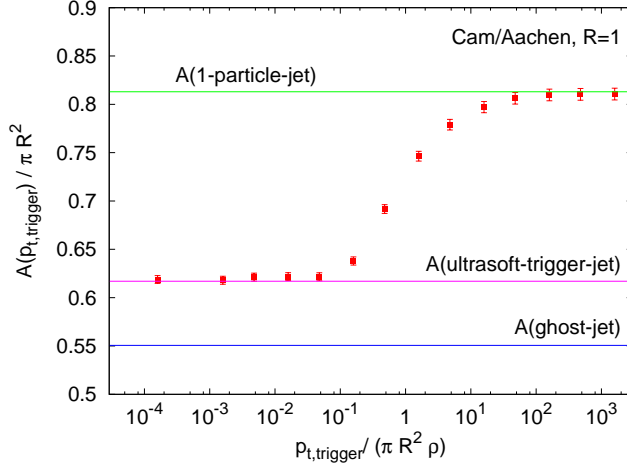
Figure 17: Area of the jet containing a trigger particle when it is immersed in a bath of soft particles with transverse momentum momentum density ρ; shown as a function of the trigger particle transverse momentum.

transverse momentum p_t is progressively reduced, until it becomes much smaller (ultrasoft) than the momentum scale of a soft background, $\pi R^2 \rho$ where, in the notation of section 5, ρ is the transverse momentum density per unit area of the soft background.[19]

Figure 17 shows, as one expects, that there are two asymptotic regions. At large p_t, the area tends to that of a jet with an infinitely hard particle, eq. (31). At small p_t, it tends not to the pure ghost jet area, eq. (34), as one might naively expect, but rather to a value which can be predicted as

$$A(\text{ultrasoft-trigger-jet}) = \frac{\int dA\, A^2\, dN/dA}{\int dA\, A\, dN/dA}\,, \tag{81}$$

where dN/dA is the distribution of the number of pure ghost jets with a given area, and corresponds to the solid curves depicted in figure 5. This equation can be understood in the following way: when the momentum of the trigger particle becomes negligible compared to that of the soft background, it does not influence the size of its own jet. However, the likelihood of the trigger particle being found in a soft-background jet is proportional to that soft-background jet's area and one gets an extra factor of A in the integrand of the numerator of eq. (81). One is able to use the pure-ghost-jet area distribution in eq. (81), because it coincides with that of soft background jets.

Figure 17 helps to illustrate the point that a diffuse background of particles such as pileup (represented here by the ghosts) provides an effective cutoff scale, below which a particle will not have any effect on the jet's area. Numerically, the effective cutoff does indeed coincide with the soft-transverse momentum scale $\pi R^2 \rho$.

C   Fluctuations of the active area

In this appendix, we derive the results eqs. (44)–(46) for the fluctuation coefficient $S^2_{\text{JA},R}$ in the case of active areas. This is slightly more technical than for passive areas as we also have to deal with averages over ghosts distributions. Let us briefly recall our notation: $\langle \cdots \rangle$ represents an average

---

[19]Imagine the trigger particle being a B meson, which you could tag through a secondary vertex — you could then recognise its presence in any jet, regardless of its transverse momentum relative to the background.

over perturbative emission, while $\langle\cdots\rangle_g$ is an average over ghosts ensembles. When a quantity is evaluated for a specific ghost ensemble $\{g_i\}$, we will explicitly state so with notation of the form $A_{\mathrm{JA},R}(\cdots|\{g_i\})$. From section 3 we have the implicit notation that a quantity specified without any mention of ghosts, such as $A_{\mathrm{JA},R}(0)$ is already averaged over ghost ensembles.

At order α_s, the mean active area can then be written (using $A_{\mathrm{JA},R}(\Delta = 0) = A_{\mathrm{JA},R}(\text{one particle})$)

$$
\langle A_{\mathrm{JA},R}\rangle \equiv \langle\langle A_{\mathrm{JA},R}(\cdots|\{g_i\})\rangle\rangle_g \tag{82}
$$

$$
\simeq \left\langle A_{\mathrm{JA},R}(0\,|\,\{g_i\}) + \int dP\,[A_{\mathrm{JA},R}(\Delta\,|\,\{g_i\}) - A_{\mathrm{JA},R}(0\,|\,\{g_i\})]\right\rangle_g \tag{83}
$$

$$
\simeq A_{\mathrm{JA},R}(0) + D_{\mathrm{JA},R}\frac{C_1}{\pi b_0}\log\left(\frac{\alpha_s(Q_0)}{\alpha_s(Rp_{t,1})}\right), \tag{84}
$$

with

$$
D_{\mathrm{JA},R} = \int\frac{d\Delta}{\Delta}\,(\langle A_{\mathrm{JA},R}(\Delta\,|\,\{g_i\})\rangle_g - \langle A_{\mathrm{JA},R}(0\,|\,\{g_i\})\rangle_g). \tag{85}
$$

Here, we have taken into account both the corrections due to the radiation of a soft particle and the average over the distribution of the ghosts. Note that we have used the shorthand dP to represent eq. (11), $dP/(dp_{t2}d\Delta)$, times $dp_{t2}d\Delta$, and that "$\cdots$" in eq. (82) represents all possible perturbative states.

For the corresponding fluctuations, the derivation goes along the same line

$$
\langle\Sigma^2_{\mathrm{JA},R}\rangle = \langle\langle A^2_{\mathrm{JA},R}(\cdots|\{g_i\})\rangle\rangle_g - \langle A_{\mathrm{JA},R}\rangle^2 \tag{86}
$$

$$
= \left\langle A^2_{\mathrm{JA},R}(0\,|\,\{g_i\}) + \int dP\,[A^2_{\mathrm{JA},R}(\Delta\,|\,\{g_i\}) - A^2_{\mathrm{JA},R}(0\,|\,\{g_i\})]\right\rangle_g
$$

$$
- \left\langle A_{\mathrm{JA},R}(0\,|\,\{g_i\}) + \int dP\,[A_{\mathrm{JA},R}(\Delta\,|\,\{g_i\}) - A_{\mathrm{JA},R}(0\,|\,\{g_i\})]\right\rangle_g^2. \tag{87}
$$

Neglecting the terms proportional to α_s^2, we can write $\left\langle\Sigma^2_{\mathrm{JA},R}\right\rangle = \Sigma^2_{\mathrm{JA},R}(0) + \left\langle\Delta\Sigma^2_{\mathrm{JA},R}\right\rangle$, where

$$
\Sigma^2_{\mathrm{JA},R}(\Delta) = \left\langle A^2_{\mathrm{JA},R}(\Delta\,|\,\{g_i\})\right\rangle_g - \left\langle A_{\mathrm{JA},R}(\Delta\,|\,\{g_i\})\right\rangle_g^2, \tag{88}
$$

which for $\Delta = 0$ is the leading order result. We can also write

$$
\left\langle\Delta\Sigma^2_{\mathrm{JA},R}\right\rangle = S^2_{\mathrm{JA},R}\frac{C_1}{\pi b_0}\log\left(\frac{\alpha_s(Q_0)}{\alpha_s(Rp_{t,1})}\right). \tag{89}
$$

Using straightforward algebra one obtains

$$
S^2_{\mathrm{JA},R} = \int\frac{d\Delta}{\Delta}\left\{\langle[A^2_{\mathrm{JA},R}(\Delta\,|\,\{g_i\}) - A^2_{\mathrm{JA},R}(0\,|\,\{g_i\})]\rangle_g\right.
$$

$$
\left. -2\langle A_{\mathrm{JA},R}(0\,|\,\{g_i\})\rangle_g\langle[A_{\mathrm{JA},R}(\Delta\,|\,\{g_i\}) - A_{\mathrm{JA},R}(0\,|\,\{g_i\})]\rangle_g\right\} \tag{90}
$$

$$
= \int\frac{d\Delta}{\Delta}\langle[A^2_{\mathrm{JA},R}(\Delta\,|\,\{g_i\}) - A^2_{\mathrm{JA},R}(0\,|\,\{g_i\})]\rangle_g - 2\langle A_{\mathrm{JA},R}(0\,|\,\{g_i\})\rangle_g D_{\mathrm{JA},R} \tag{91}
$$

$$
= \int\frac{d\Delta}{\Delta}\left[\langle A_{\mathrm{JA},R}(\Delta\,|\,\{g_i\}) - A_{\mathrm{JA},R}(0\,|\,\{g_i\})\rangle_g^2 + \Sigma^2_{\mathrm{JA},R}(\Delta) - \Sigma^2_{\mathrm{JA},R}(0)\right]. \tag{92}
$$

The second equality is a direct rewriting of the first. One gets to the last line by rearranging the different terms of the first one. The last two lines correspond exactly to eqs. (46) and (45) of section 3.3.

40

References

[1] G. Sterman and S. Weinberg, Phys. Rev. Lett. **39** (1977) 1436.

[2] for a review, see M. Dasgupta and G. P. Salam, J. Phys. G **30** (2004) R143 [arXiv:hep-ph/0312283].

[3] M. H. Seymour, Nucl. Phys. B **513** (1998) 269 [arXiv:hep-ph/9707338]; A. Banfi, G. P. Salam and G. Zanderighi, JHEP **0408** (2004) 062 [arXiv:hep-ph/0407287].

[4] R. Perez-Ramos, F. Arleo and B. Machet, arXiv:0712.2212 [hep-ph].

[5] R. B. Appleby and M. H. Seymour, JHEP **0212** (2002) 063 [arXiv:hep-ph/0211426]; JHEP **0309** (2003) 056 [arXiv:hep-ph/0308086].

[6] A. Banfi and M. Dasgupta, Phys. Lett. B **628** (2005) 49 [arXiv:hep-ph/0508159].

[7] see e.g. N. Kidonakis and J. F. Owens, Phys. Rev. D **63** (2001) 054019 [hep-ph/0007268]; D. de Florian and W. Vogelsang, Phys. Rev. D **76** (2007) 074031 [arXiv:0704.1677 [hep-ph]].

[8] R. Field [CDF Run II Collaboration],

[9] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, Z. Phys. C **72** (1996) 637 [arXiv:hep-ph/9601371].

[10] M. Dasgupta, L. Magnea and G. P. Salam, arXiv:0712.3014 [hep-ph].

[11] B. R. Webber, Phys. Lett. B **339** (1994) 148 [arXiv:hep-ph/9408222].

[12] M. Dasgupta and G. P. Salam, J. Phys. G **30** (2004) R143 [arXiv:hep-ph/0312283].

[13] D. de Florian and W. Vogelsang, Phys. Rev. D **76** (2007) 074031 [arXiv:0704.1677 [hep-ph]].

[14] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. B **406** (1993) 187 and refs. therein; S. D. Ellis and D. E. Soper, Phys. Rev. D **48** (1993) 3160 [hep-ph/9305266].

[15] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **9708**, 001 (1997) [hep-ph/9707323]; M. Wobisch and T. Wengler, hep-ph/9907280.

[16] G. P. Salam and G. Soyez, JHEP **0705** (2007) 086 [arXiv:0704.0292],

[17] G. Corcella *et al.*, arXiv:hep-ph/0210213.

[18] T. Sjostrand *et al.*, Comput. Phys. Commun. **135** (2001) 238 [hep-ph/0010017]; T. Sjostrand *et al.*, hep-ph/0308153.

[19] M. Cacciari and G. P. Salam, Phys. Lett. B **659** (2008) 119 [arXiv:0707.1378 [hep-ph]].

[20] D. Y. Grigoriev, E. Jankowski and F. V. Tkachov, Phys. Rev. Lett. **91**, 061801 (2003) [hep-ph/0301185]; Comput. Phys. Commun. **155**, 42 (2003) [hep-ph/0301226].

[21] L. Lonnblad, Z. Phys. C **58** (1993) 471.

[22] G. C. Blazey *et al.*, hep-ex/0005012.

[23] M. Cacciari and G. P. Salam, Phys. Lett. B **641** (2006) 57 [hep-ph/0512210].

[24] S. J. Fortune, Algorithmica (1987) 153–174.

[25] A. Bhatti *et al.*, Nucl. Instrum. Meth. A **566** (2006) 375 [arXiv:hep-ex/0510047].

[26] G. P. Salam and D. Wicke, JHEP **0105**, 061 (2001) [arXiv:hep-ph/0102343].

[27] A. Banfi, G. Corcella and M. Dasgupta, JHEP **0703**, 050 (2007) [arXiv:hep-ph/0612282].

[28] T. Sjostrand and P. Z. Skands, Eur. Phys. J. C **39** (2005) 129 [arXiv:hep-ph/0408302].

[29] M. Cacciari, G. P. Salam and G. Soyez, arXiv:0802.1189.