

## MIT Open Access Articles

*The first calculation of fractional jets*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Bertolini, Daniele, Jesse Thaler, and Jonathan R. Walsh. "The First Calculation of Fractional Jets." J. High Energ. Phys. 2015, no. 5 (May 2015).

**As Published:** [http://dx.doi.org/10.1007/JHEP05\(2015\)008](http://dx.doi.org/10.1007/JHEP05(2015)008)

**Publisher:** Springer-Verlag

**Persistent URL:** <http://hdl.handle.net/1721.1/97247>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution



# The first calculation of fractional jets

Daniele Bertolini,<sup>a,b</sup> Jesse Thaler<sup>c</sup> and Jonathan R. Walsh<sup>a,b</sup>

<sup>a</sup>*Ernest Orlando Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, U.S.A.*

<sup>b</sup>*Center for Theoretical Physics, University of California, Berkeley, CA 94720, U.S.A.*

<sup>c</sup>*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*

*E-mail:* [dbertolini@lbl.gov](mailto:dbertolini@lbl.gov), [jthaler@mit.edu](mailto:jthaler@mit.edu), [jwalsh@lbl.gov](mailto:jwalsh@lbl.gov)

**ABSTRACT:** In collider physics, jet algorithms are a ubiquitous tool for clustering particles into discrete jet objects. Event shapes offer an alternative way to characterize jets, and one can define a jet multiplicity event shape, which can take on fractional values, using the framework of “jets without jets”. In this paper, we perform the first analytic studies of fractional jet multiplicity  $\tilde{N}_{\text{jet}}$  in the context of  $e^+e^-$  collisions. We use fixed-order QCD to understand the  $\tilde{N}_{\text{jet}}$  cross section at order  $\alpha_s^2$ , and we introduce a candidate factorization theorem to capture certain higher-order effects. The resulting distributions have a hybrid jet algorithm/event shape behavior which agrees with parton shower Monte Carlo generators. The  $\tilde{N}_{\text{jet}}$  observable does not satisfy ordinary soft-collinear factorization, and the  $\tilde{N}_{\text{jet}}$  cross section exhibits a number of unique features, including the absence of collinear logarithms and the presence of soft logarithms that are purely non-global. Additionally, we find novel divergences connected to the energy sharing between emissions, which are reminiscent of rapidity divergences encountered in other applications. Given these interesting properties of fractional jet multiplicity, we advocate for future measurements and calculations of  $\tilde{N}_{\text{jet}}$  at hadron colliders like the LHC.

**KEYWORDS:** QCD Phenomenology, Jets

**ARXIV EPRINT:** [1501.01965](https://arxiv.org/abs/1501.01965)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Aspects of fractional jets</b>	<b>4</b>
2.1	Hybrid jet algorithm/event shape behavior	5
2.2	Soft and collinear limits of fractional jets	7
2.3	Non-global yet local structure	10
<b>3</b>	<b>Rapidity-like divergences</b>	<b>11</b>
<b>4</b>	<b>Calculating the near-integer behavior</b>	<b>14</b>
4.1	Singular contributions using splitting functions	15
4.2	Nonsingular contributions from EVENT2	19
4.3	Complete results to order $\alpha_s^2$	22
<b>5</b>	<b>Towards a factorization theorem</b>	<b>23</b>
5.1	The cross section at $\tilde{N}_{\text{jet}} = 2$	23
5.2	Soft-collinear non-factorization	24
5.3	Introducing collinear functions	25
5.4	A “local” factorization theorem	26
5.5	Complete results	27
<b>6</b>	<b>Looking towards the LHC</b>	<b>28</b>
<b>7</b>	<b>Monte Carlo comparisons</b>	<b>31</b>
<b>8</b>	<b>Conclusions</b>	<b>33</b>
<b>A</b>	<b>Results for non-Abelian contributions</b>	<b>35</b>
<b>B</b>	<b>Properties of distributions</b>	<b>35</b>
B.1	One-sided distributions and their convolutions	36
B.2	Two-sided distributions and their convolutions	37
<b>C</b>	<b>Quark and gluon collinear functions</b>	<b>38</b>

---

# 1 Introduction

For almost forty years, we have known that high energy particle collisions can produce jets [1, 2]. The term “jets” has two related but different meanings: “jet formation” is the physical process by which quarks and gluons shower and fragment into collimated sprays of hadrons [3], while “jet algorithms” are an analysis technique used to cluster those hadrons into proxies for the underlying quarks and gluons [4]. Jet algorithms are a powerful way to categorize and organize collision events [5, 6], but event shapes (and jet shapes) offer a more sensitive probe of jet formation itself [7]. Indeed, though the observation of three-jet structure in  $e^+e^-$  collisions strongly hinted at the existence of gluons [8, 9], an unambiguous discovery at PETRA [10–13] required the use of event shapes like thrust [14] and oblateness [11].

Recently, the distinction between jet algorithms and event shapes was blurred through the “jets without jets” framework, in which standard jet-based observables are mapped into jet-like event shapes [15]. These observables incorporate a transverse momentum threshold  $p_{T\text{cut}}$  and a jet radius  $R$  just like traditional jet algorithms, but they behave more like event shapes because they involve an inclusive sum over particles in an event and do not uniquely assign hadrons to jet objects. A particularly interesting jets-without-jets observable is jet multiplicity,

$$\tilde{N}_{\text{jet}}(p_{T\text{cut}}, R) = \sum_{i \in \text{event}} \frac{p_{Ti}}{p_{Ti,R}} \Theta(p_{Ti,R} - p_{T\text{cut}}), \tag{1.1}$$

where  $p_{Ti,R}$  is the transverse momentum contained in a cone of radius  $R$  around particle  $i$ . By design, this observable is highly correlated with the standard jet count (for the same  $p_{T\text{cut}}$  and  $R$  values). Crucially,  $\tilde{N}_{\text{jet}}$  can yield fractional values, offering a new probe of jet formation.

In this paper, we initiate the analytic study of fractional jet multiplicity. For simplicity, we treat the case of  $e^+e^- \rightarrow$  jets, though we briefly mention how to adapt our calculational techniques to collisions at the Large Hadron Collider (LHC). With two or three partons in the final state ( $e^+e^- \rightarrow q\bar{q}$  or  $q\bar{q}g$ ),  $\tilde{N}_{\text{jet}}$  always yields an integer value. Fractional jets only start appearing with four or more partons (e.g.  $e^+e^- \rightarrow q\bar{q}gg$ ), so our analytic calculations will start at  $\mathcal{O}(\alpha_s^2)$ . This is a general feature of fractional jets: non-integer values only appear with three or more particles in a given jet region.<sup>1</sup> Our main technical results will be obtained using fixed-order perturbative QCD, though we will also discuss connections to factorization properties in soft-collinear effective theory (SCET) [16–20].

We will focus on the behavior of  $\tilde{N}_{\text{jet}}$  in the vicinity of dijet configurations,

$$\Delta_{2+} = \tilde{N}_{\text{jet}} - 2, \quad \Delta_{2-} = 2 - \tilde{N}_{\text{jet}}, \tag{1.2}$$

though we do present some results for  $\Delta_{3-} = 3 - \tilde{N}_{\text{jet}}$  as well. This near-integer behavior exhibits a number of curious analytic features, which are already visible at  $\mathcal{O}(\alpha_s^2)$ .

- *Hybrid jet algorithm/event shape behavior.* Jet algorithms have a finite cross section at (integer) jet multiplicities whereas event shapes typically have form factors that

---

<sup>1</sup>For  $e^+e^-$  collisions in the rest frame, momentum conservation prevents all particles from being in the same jet region. Hence four partons are required for non-integer values.

suppress the cross section at singular values. The  $\tilde{N}_{\text{jet}}$  distribution has both types of behavior. Even though the Born-level process  $e^+e^- \rightarrow q\bar{q}$  gives the integer value  $\tilde{N}_{\text{jet}} = 2$ , one might naively expect the corresponding spike at  $\Delta_{2\pm} = 0$  to be completely smeared out by multiple emissions that generate finite values of  $\Delta_{2\pm}$ . Instead, the spike at  $\Delta_{2\pm} = 0$  is robust, as there is a finite region of the many-body phase space that still gives rise to integer values of  $\tilde{N}_{\text{jet}}$ . At the same time, the appearance of logarithms in  $\Delta_{2\pm}$  at every perturbative order generates a shoulder at finite values of  $\Delta_{2\pm}$ , which is suppressed as  $\Delta_{2\pm} \rightarrow 0$ . Thus, the cross section at exact-integer values of  $\tilde{N}_{\text{jet}}$  behave like a jet algorithm while near-integer values behave like an event shape (see section 2.1).

- *Cancellation of single- and double-soft divergences.* The first non-trivial contributions to non-integer behavior of  $\tilde{N}_{\text{jet}}$  arise from configurations where three partons are within a mutual radius of  $2R$  but not within  $R$  (see section 2.2). For  $\Delta_{2+}$ , this three-parton phase space has singularities when one or two of the partons goes soft. These divergences arise because the observable receives parametrically equivalent contributions from the single- and double-soft regions, which are not regulated in dimensional regularization. Interestingly, these divergences are structurally similar to rapidity divergences, and we will introduce the analog of rapidity regularization [21, 22] to see that the single-soft and double-soft divergences do indeed cancel (see section 3). We note that soft emissions contribute in a non-linear way, and therefore  $\tilde{N}_{\text{jet}}$  is a non-additive observable.
- *No collinear logarithms.* Typical event shapes have singularities in both the soft and collinear limit, giving rise to both soft and collinear logarithms. The resulting double-logarithmic structure appears as Sudakov form factors in the cross section. By contrast, collinear emissions do not generate logarithms of  $\tilde{N}_{\text{jet}}$ , and only soft logarithms appear in the  $\tilde{N}_{\text{jet}}$  distribution (see section 2.2).<sup>2</sup> Thus, the suppression in the  $\Delta_{2\pm} \rightarrow 0$  limit is only single-logarithmic.
- *Non-global yet local structure.* The  $\Delta_{2\pm}$  cross section does not satisfy ordinary soft-collinear factorization, because the coefficients of the soft logarithms depend on the collinear structure of the jets (despite the absence of collinear logarithms). Furthermore, the soft logarithms in  $\tilde{N}_{\text{jet}}$  are purely non-global [26, 27], in the sense that they arise from soft emissions in a restricted region of phase space (see section 2.3).<sup>3</sup> These features would seem to preclude any standard factorization theorem, especially given the non-additive nature of  $\tilde{N}_{\text{jet}}$ . In order to change the value of  $\tilde{N}_{\text{jet}}$ , however, soft emissions must lie within  $\lesssim 2R$  from the hard core of the jet. Thus, color coherence ensures that the dependence on  $\Delta_{2\pm}$ , albeit non-global, is local to each jet region. We will present a candidate factorization theorem that exploits this local structure (see section 5).

---

<sup>2</sup>Mass-dropped observables [23–25] have the reverse behavior of only having collinear logarithms.

<sup>3</sup>As explained in section 2.3, our use of the term “non-global” can refer to both non-Abelian and Abelian correlations.

In addition to analytic studies, we will test  $\tilde{N}_{\text{jet}}$  using high-statistics Monte Carlo samples from PYTHIA 8 [28] and HERWIG++ [29]. Within theoretical uncertainties, the Monte Carlo results confirm our analytic understanding.

Given its many features and potential applications,  $\tilde{N}_{\text{jet}}$  would be very interesting to measure at the LHC. As mentioned above,  $\tilde{N}_{\text{jet}}$  is a purely non-global observable, with the near-integer behavior determined only by soft and not by collinear divergences. To our knowledge, it is the only jet or event shape observable with this behavior. As such, it is a unique probe of soft physics, and measurements of  $\tilde{N}_{\text{jet}}$  can be used to test color coherence, underlying event models, and pileup mitigation strategies. Furthermore,  $\tilde{N}_{\text{jet}}$  is useful basis to compare parton shower predictions for jet substructure, and we present an initial comparison in this work. For new physics searches involving high-multiplicity final states, fractional  $\tilde{N}_{\text{jet}}$  values interpolate between different jet multiplicities, obviating the need for exclusive jet bins. This interpolation also makes for an interesting version of the classic “staircase” plots for vector boson plus  $N$  jet production [30–38]. Finally, for the growing field of matrix element/parton shower matching/merging [39–50],  $\tilde{N}_{\text{jet}}$  has a continuous distribution unlike standard jet algorithms and a huge dynamic range compared to standard event shapes, so  $\tilde{N}_{\text{jet}}$  can be used to test whether matching/merging procedures achieve a smooth interpolation, even in the soft regime. We note, of course, that theoretical calculations of  $\tilde{N}_{\text{jet}}$  for hadronic collisions must contend with additional effects such as the underlying event and pileup contamination. Moreover, it will be non-trivial to include non-perturbative effects, power-suppressed terms, and higher-order perturbative effects such as the resummation of non-global logarithms [26, 27, 51–54]. Although a detailed study of  $\tilde{N}_{\text{jet}}$  for the LHC is beyond the scope of this paper, we briefly discuss some of these issues and potential solutions in section 6.

The rest of this paper is organized as follows. In section 2, we review the basic physics behind  $\tilde{N}_{\text{jet}}$  and explain the kinematic regimes that give rise to fractional jets. In section 3, we discuss the structure of rapidity-like divergences and how they appear in the  $\tilde{N}_{\text{jet}}$  calculation. In section 4, we perform fixed-order calculations of  $\Delta_{2\pm}$  at  $\mathcal{O}(\alpha_s^2)$ , using both the full  $e^+e^- \rightarrow 4$  parton matrix element as well as a  $1 \rightarrow 3$  splitting function approximation. We then present a candidate factorization theorem for  $\Delta_{2\pm}$  in section 5, which includes a renormalization-group-independent “collinear function”. In section 6, we briefly discuss how to extend our results to the LHC. We compare our analytic calculations to PYTHIA 8 and HERWIG++ in section 7, and we conclude in section 8. The appendices contain further calculational results and details.

## 2 Aspects of fractional jets

Since we will be looking at electron-positron collisions, it is more natural to work with a variant of  $\tilde{N}_{\text{jet}}$  based on energies and angles (instead of transverse momenta and azimuthal-rapidity distances):

$$\tilde{N}_{\text{jet}}(E_{\text{cut}}, R) = \sum_{i \in \text{event}} \frac{E_i}{E_{i,R}} \Theta(E_{i,R} - E_{\text{cut}}), \tag{2.1}$$

where  $E_i$  is the energy of particle  $i$ ,

$$E_{i,R} = \sum_{j \in \text{event}} E_j \Theta(R - \theta_{ij}), \quad (2.2)$$

and  $\theta_{ij}$  is the opening angle between particles  $i$  and  $j$ .

For particles whose angular separation is always larger than  $R$ ,  $E_i/E_{i,R}$  reduces to 1, and  $\tilde{N}_{\text{jet}}$  simply counts the number of hard particles above the energy threshold  $E_{\text{cut}}$ . Because  $\tilde{N}_{\text{jet}}$  is an infrared/collinear (IRC) safe observable,  $\tilde{N}_{\text{jet}}$  always takes on integer values in the extreme soft and collinear limit. We will use the notation

$$\Delta_{n\pm} = \pm(\tilde{N}_{\text{jet}} - n) \quad (2.3)$$

to indicate the near-integer behavior, with  $\Delta_{n+}$  ( $\Delta_{n-}$ ) parameterizing the  $\tilde{N}_{\text{jet}}$  behavior just above (below)  $n$ . Our calculations will focus on  $\Delta_{2\pm}$  and  $\Delta_{3-}$  in  $e^+e^-$  collisions.

## 2.1 Hybrid jet algorithm/event shape behavior

The  $\tilde{N}_{\text{jet}}$  distribution is peaked at integer values, with substantial support in the near-integer regime. These different parts of the  $\tilde{N}_{\text{jet}}$  distribution can be qualitatively studied with Monte Carlo generators.<sup>4</sup> We generate events for  $e^+e^- \rightarrow$  hadrons at a center-of-mass energy of  $Q = 500$  GeV in PYTHIA 8.183 [28] and HERWIG++ 2.7.1 [29], including showering and hadronization. In figure 1a, we plot the distribution of  $\tilde{N}_{\text{jet}}$  across a wide range of values, showing the circus-tent-like peak and fall-off behavior of the cross section.<sup>5</sup> In figure 1b, we focus on the distribution in the range  $2 \lesssim \tilde{N}_{\text{jet}} \lesssim 3$ , which is the region we aim to quantitatively describe in this paper. In figure 2, we plot the  $\tilde{N}_{\text{jet}}$  distribution in a “trptych” form that shows in more detail the near-integer behavior in  $\Delta_{2-}$ ,  $\Delta_{2+}$ , and  $\Delta_{3-}$ , in particular the logarithmic enhancement as  $\Delta_{n\pm} \rightarrow 0$ .

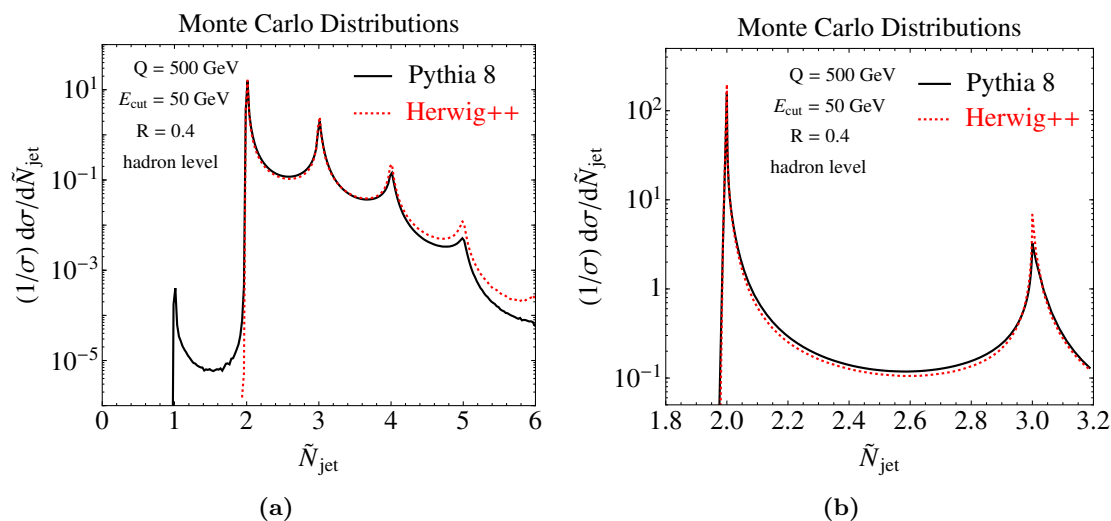
As mentioned in the introduction, the cross section at exact integer values has a different behavior than at near-integer values, a feature related to the hybrid jet algorithm/event shape nature of the observable. Like a jet algorithm, the cross section at integer values  $\tilde{N}_{\text{jet}} = n$  has a non-zero value. Like an event shape, the non-integer cross section is suppressed by all-orders emissions as  $\tilde{N}_{\text{jet}} \rightarrow n$ . The behavior can be seen in figure 3, where we plot the distribution in the very near vicinity of  $\tilde{N}_{\text{jet}} = 2$ .

The reason why integer values  $\tilde{N}_{\text{jet}} = n$  have a finite cross section is that they receive contributions from regions of phase space with non-zero measure. This can be seen easily at  $\mathcal{O}(\alpha_s)$  for  $e^+e^- \rightarrow q\bar{q}g$ , where the entire cross section lies at  $\tilde{N}_{\text{jet}} = 2$  or 3. More generally, any collection of particles within a mutual radius of  $R$  will give an integer contribution to

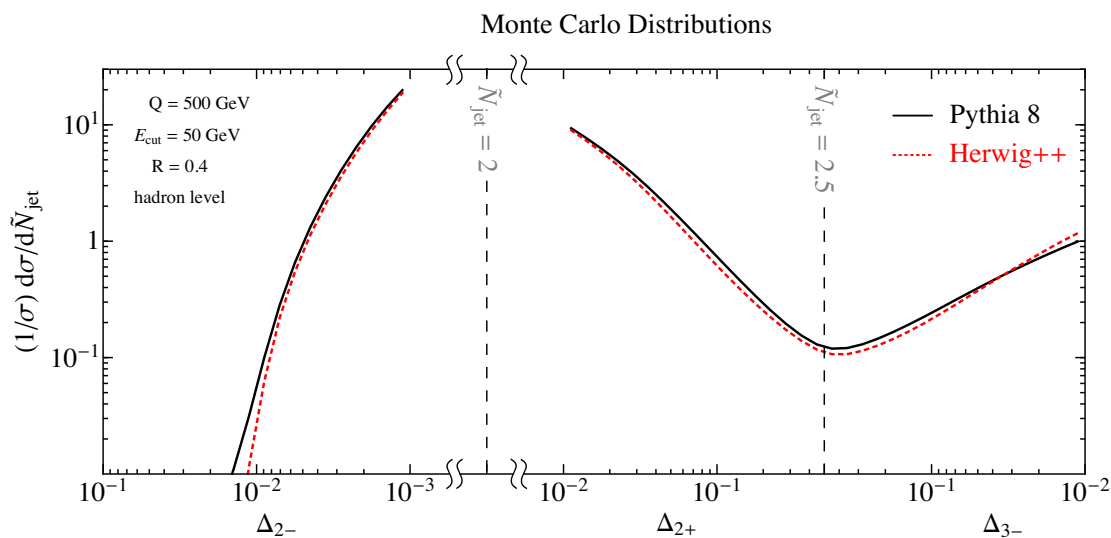
---

<sup>4</sup>The original  $\tilde{N}_{\text{jet}}$  variable in eq. (1.1) is available through the JETSWITHOUTJETS package, an add-on to FASTJET 3 [55] contained in the FASTJET CONTRIB library (<http://fastjet.hepforge.org/contrib/>). The variant in eq. (2.1) is available from the authors upon request.

<sup>5</sup>The peak at  $\tilde{N}_{\text{jet}} = 1$  in PYTHIA 8 seems to arise from events with  $\tau$  leptons produced in hadron decays, where the resulting neutrinos carry away a substantial fraction of the jet momentum. The same feature is not visible in HERWIG++, nor is it visible when hadronization effects are turned off in PYTHIA 8.

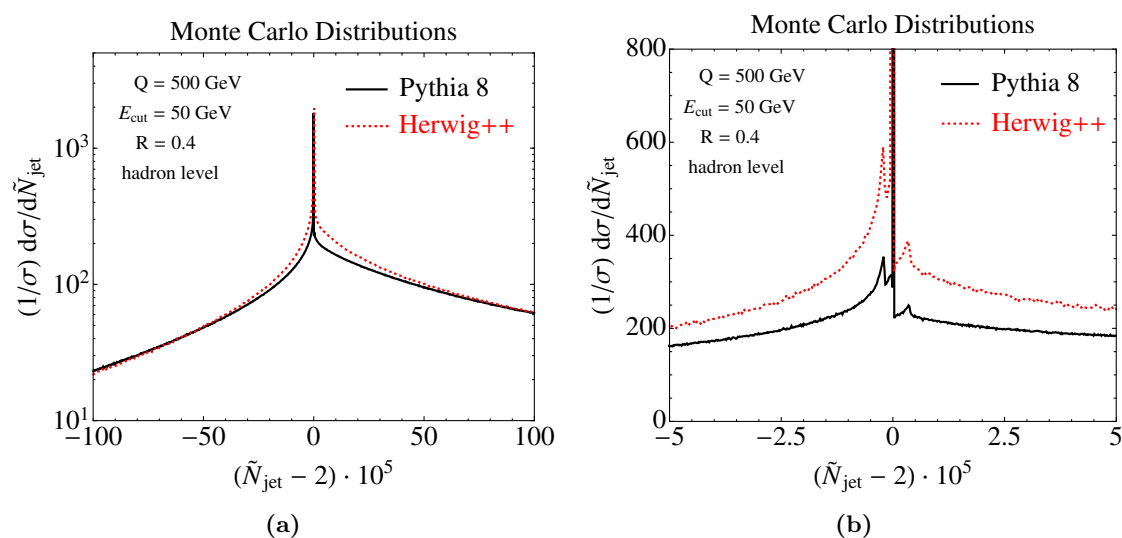


**Figure 1.** The  $\tilde{N}_{\text{jet}}$  distribution in  $e^+e^-$  collisions from PYTHIA 8 and HERWIG++, over a wide range of multiplicities (left plot) and for  $2 \lesssim \tilde{N}_{\text{jet}} \lesssim 3$  (right plot). The distribution peaks at integer values, and near-integer values are enhanced relative to intermediate values between integers. It is this enhancement that we will understand quantitatively through our calculations.



**Figure 2.** The same  $\tilde{N}_{\text{jet}}$  distributions as figure 1, now plotted in “triptych” form to show the near-integer behavior in  $\Delta_{2-}$ ,  $\Delta_{2+}$ , and  $\Delta_{3-}$ . Note that the  $\Delta_{2-}$  and  $\Delta_{3-}$  axes run backwards. The  $\Delta_{2-}$  and  $\Delta_{2+}$  distributions interface at  $\tilde{N}_{\text{jet}} = 2$  (where  $\Delta_{2\pm} = 0$ ), indicated by the left-hand dashed vertical line. Since we do not plot  $\Delta_{2\pm}$  all the way to 0, we put a gap over the region around  $\tilde{N}_{\text{jet}} = 2$ . The  $\Delta_{2+}$  and  $\Delta_{3-}$  distributions are connected at  $\tilde{N}_{\text{jet}} = 2.5$  (where  $\Delta_{2+} = \Delta_{3-} = 0.5$ ), indicated by the right-hand dashed vertical line.





**Figure 3.** The same  $\tilde{N}_{\text{jet}}$  distributions as figure 1, now plotted in the very near vicinity of  $\tilde{N}_{\text{jet}} = 2$ . Note the scale on the  $x$ -axis and the fact that the  $y$ -axis is logarithmic on the left plot but linear on the right one. At  $\tilde{N}_{\text{jet}} = 2$ , there is a spike in the distribution from the jet algorithm behavior of the fractional jet multiplicity. At non-integer values, the continuous distribution is more similar to event shapes. We show the near-integer values in two different ranges, and for the closer range (right plot) one can see bumps in the Monte Carlo distributions from multiple emission and hadronization effects.

$\tilde{N}_{\text{jet}}$ .<sup>6</sup> This means that the differential distribution at integer values has the form

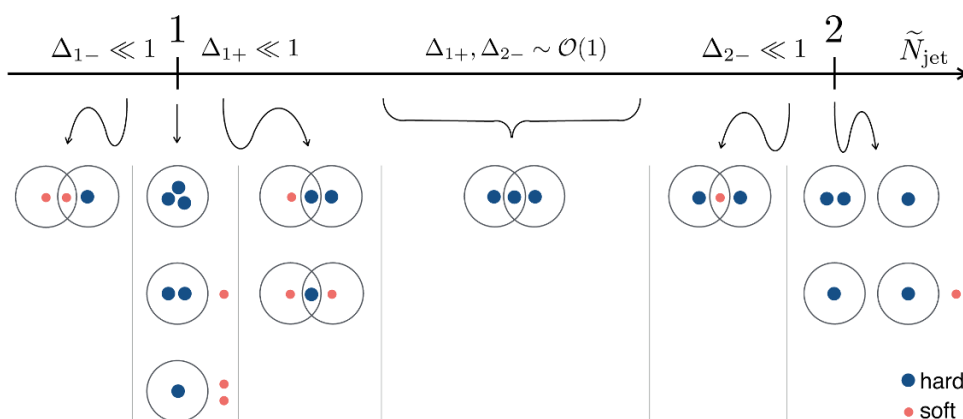
$$\frac{d\sigma}{d\tilde{N}_{\text{jet}}}(\tilde{N}_{\text{jet}} = n) \propto \delta(n). \quad (2.4)$$

In contrast, the distribution for near-integer values  $\Delta_{n\pm} > 0$  is dominated by logarithms of  $\Delta_{n\pm}$ . We will later show that the logarithms of  $\Delta_{n\pm}$  arise from soft emissions, and the most important terms scale single logarithmically as  $(\alpha_s \ln \Delta_{n\pm})^k$ . These logarithms combine at all orders to suppress the cross section as  $\Delta_{n\pm} \rightarrow 0$ , leading to the disjoint behavior at integer  $\tilde{N}_{\text{jet}}$ . Note that single-logarithmic suppression is not as strong as the more familiar double-logarithmic suppression, so there are no Sudakov peaks in the  $\Delta_{n\pm}$  distributions.

## 2.2 Soft and collinear limits of fractional jets

To understand the leading near-integer behavior of  $\tilde{N}_{\text{jet}}$ , we can focus on the soft and collinear limits. As discussed further in section 5, the all-orders structure of the cross section does not satisfy standard soft-collinear factorization, but a soft-collinear analysis is still fruitful at fixed order. The essential physics appears already in the three-parton phase space, where the near-integer behavior is dominated by soft physics. We will focus on a

<sup>6</sup>Regions within mutual radius  $R$  are bounded by curves of constant width  $R$ . A circle of radius  $R/2$  is the simplest example, but there are more general examples like the Reuleaux triangle or the Canadian loonie.



**Figure 4.** Schematic representation of the different phase space configurations with three merged partons. The numbered line shows the corresponding range of  $\tilde{N}_{\text{jet}}$ . Here, we are considering just a single jet region; an  $e^+e^- \rightarrow$  dijet event has two such jet regions. Circles represent cones of radius  $R$ , large blue dots represent energetic partons, and small red dots represent soft partons with  $z < z_{\text{cut}}$ . The near-integer regions  $\Delta_{1-}$ ,  $\Delta_{1+}$ , and  $\Delta_{2-}$  get contributions from single- and/or double-soft emissions.

single jet region here; to describe  $e^+e^- \rightarrow$  dijets, we sum over the contributions to  $\tilde{N}_{\text{jet}}$  from both jet regions (see eq. (5.2)).

We call a group of particles *merged* if their contribution to  $\tilde{N}_{\text{jet}}$  is connected, such that removing a subset of particles affects the contribution to  $\tilde{N}_{\text{jet}}$  from other particles. For a single isolated particle, its contribution to  $\tilde{N}_{\text{jet}}$  is 1. For a merged pair of particles with separation less than  $R$ , one still obtains an integer value of  $\tilde{N}_{\text{jet}}$ :

$$\tilde{N}_{\text{jet}} = \frac{E_1}{E_1 + E_2} + \frac{E_2}{E_1 + E_2} = 1. \tag{2.5}$$

Because a single soft/collinear emission does not contribute to the value of  $\tilde{N}_{\text{jet}}$ , it is not linear in soft and collinear momenta in the singular region of phase space, and hence it is a non-additive observable.

Now consider three merged partons. As shown in figure 4, there are various different phase space configurations that lead to different values of  $\tilde{N}_{\text{jet}}$ . To achieve non-integer values, one needs a phase space configuration with

$$\theta_{12} > R, \quad \theta_{13}, \theta_{23} < R, \tag{2.6}$$

or permutations of the parton labels. If all  $E_i > E_{\text{cut}}$ , then the contribution to the jet multiplicity is

$$\tilde{N}_{\text{jet}} = \frac{E_1}{E_1 + E_3} + \frac{E_2}{E_2 + E_3} + \frac{E_3}{E_1 + E_2 + E_3} \in (1, 2). \tag{2.7}$$

In terms of energy fractions

$$z_i \equiv \frac{E_i}{E_1 + E_2 + E_3}, \quad z_1 + z_2 + z_3 = 1, \quad z_{\text{cut}} \equiv \frac{E_{\text{cut}}}{E_1 + E_2 + E_3}, \tag{2.8}$$

which removes the overall energy scale of the jet as a degree of freedom, the observable takes the value

$$\tilde{N}_{\text{jet}} = \frac{z_1}{z_1 + z_3} + \frac{z_2}{z_2 + z_3} + z_3. \quad (2.9)$$

For this merged triplet, near-integer behavior is obtained when one or two of the partons goes soft, as shown in figure 4. In these soft limits, the observable may depend on a single soft momentum or a product of soft momenta. For example, if particle 1 is soft,

$$\tilde{N}_{\text{jet}} \xrightarrow{z_1 \ll 1} 1 + z_1 \frac{z_2(2 - z_2)}{1 - z_2} + \mathcal{O}(z_1^2). \quad (2.10)$$

Or, if particles 1 and 2 are both soft,

$$\tilde{N}_{\text{jet}} \xrightarrow{z_{1,2} \ll 1} 1 + 2z_1z_2 + \mathcal{O}(z_{1,2}^3). \quad (2.11)$$

The first case is typical of event shape observables, as the near-integer behavior is linear in the soft particle's energy. The second case, however, is unusual — it goes as the *product* of soft momenta, another demonstration of the non-additive nature of  $\tilde{N}_{\text{jet}}$ . This feature creates a novel structure in the perturbative series, reminiscent of rapidity divergences in SCET<sub>II</sub> [21, 56]. We discuss this further in section 3, using a toy observable  $\Delta = z_1z_2$  that exhibits the same analytic features.

With more emissions, near-integer behavior requires the soft limit of one or more particles.<sup>7</sup> As mentioned above, an arbitrary collection of energetic particles will yield an integer  $\tilde{N}_{\text{jet}}$  if the particles can be grouped into merged clusters where each particle in the cluster is within  $R$  of all other particles in the cluster (such that each cluster has  $\tilde{N}_{\text{jet}} = 1$ ). Non-integer values are obtained when this is not satisfied (as in the merged triplet example above), though generically the resulting values are far from integers. Near-integer values are obtained by starting from an integer  $\tilde{N}_{\text{jet}}$  configuration and then adding any number of soft particles which are not within the mutual radius  $R$  of the cluster. The contribution of these soft particles to  $\tilde{N}_{\text{jet}}$  will be suppressed by

$$\frac{E_{\text{soft}}}{E_{\text{soft},R}}, \quad (2.12)$$

and the deviation of  $\tilde{N}_{\text{jet}}$  from integer values is similarly suppressed by this ratio as long as  $E_{\text{soft},R} > E_{\text{cut}}$ . In this way, the near-integer behavior is determined by soft emissions in the vicinity of hard clusters, and soft emissions will generate logarithms of  $\Delta_{n\pm}$ .

By contrast, collinear splittings cannot generate near-integer behavior of  $\tilde{N}_{\text{jet}}$  and hence do not generate logarithms of  $\Delta_{n\pm}$ . For collinear splittings of angle  $R_c$ , the only effect on  $\tilde{N}_{\text{jet}}$  comes from particles who are within  $R$  of either of the two daughters or the parent, but not within  $R$  of all three. For small  $R_c$ , this is a power-suppressed region of phase space and not logarithmically enhanced. Said another way, the effect on  $\tilde{N}_{\text{jet}}$  from a small-angle splitting is not smooth, as it either preserves the value of  $\tilde{N}_{\text{jet}}$  or discretely changes it by

---

<sup>7</sup>We are neglecting special configurations of energetic particles where the value of  $\tilde{N}_{\text{jet}}$  happens to be near-integer, since those regions of phase space are power suppressed.

including or excluding particles from the various  $E_{i,R}$  terms; this behavior cannot generate logarithms of  $\Delta_{n\pm}$ . So unlike standard jet shapes (like jet mass) which depend on both the softness and collinearity of a splitting, the near-integer  $\tilde{N}_{\text{jet}}$  value depends only on the softness of emissions.

### 2.3 Non-global yet local structure

We have argued that soft emissions contribute to near-integer values of  $\tilde{N}_{\text{jet}}$ , but this is only the case if the soft emissions lie within a restricted phase space of size  $\lesssim 2R$  around the jet. Wide-angle soft radiation does not contribute to  $\tilde{N}_{\text{jet}}$ , since those emissions yield  $E_{i,R} < E_{\text{cut}}$ .<sup>8</sup> This angular restriction on soft radiation produces *non-global* logarithms of  $\Delta_{n\pm}$ , which are logarithms that arise from emissions in a restricted angular region of phase space [26, 27].

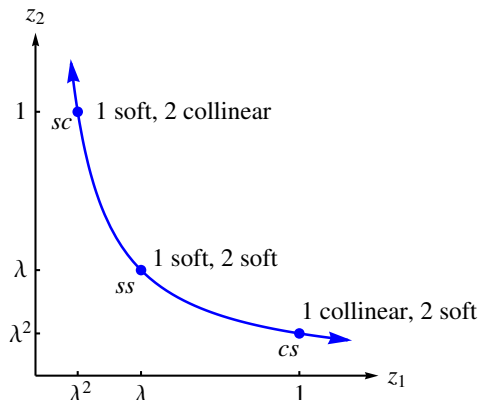
At leading order, non-global logarithms are often associated with correlated soft emissions from non-Abelian matrix elements and are therefore proportional to  $C_F C_A$ . For the case of fractional jet multiplicity, however, the measurement itself introduces a correlation between different emissions, and this effect appears for both non-Abelian and Abelian matrix elements. As we discuss in detail in section 4.1, the allowed phase space for a soft emission to change  $\tilde{N}_{\text{jet}}$  depends on the phase space of other soft emissions and on the phase space of the hard partons, similarly to what happens for clustering logarithms [57–60]. In this regard, all soft logarithms of  $\tilde{N}_{\text{jet}}$ , including those proportional to  $C_F^2$ , can be considered non-global.

Moreover, the fact that the allowed phase space for soft emissions to change  $\tilde{N}_{\text{jet}}$  depends on the phase space of the hard partons means that one cannot consider how  $\tilde{N}_{\text{jet}}$  depends on soft emissions without also considering collinear emissions, and vice versa. Because the contributions to  $\tilde{N}_{\text{jet}}$  from soft and collinear emissions are inextricably linked, we will show in section 5.2 that  $\tilde{N}_{\text{jet}}$  does not obey standard soft-collinear factorization [19, 61–63]. As mentioned above, collinear emissions by themselves do not generate logarithms of  $\tilde{N}_{\text{jet}}$ , but they do alter the allowed phase space for soft radiation within a jet region. Thus, collinear emissions will modify the coefficients of soft logarithms of  $\tilde{N}_{\text{jet}}$ , which is a sign of collinear-soft non-factorization.

While the  $\Delta_{n\pm}$  dependence is non-global, it is also local to each jet. Emissions affecting  $\tilde{N}_{\text{jet}}$  are restricted to an angular region of  $\lesssim 2R$  around each jet direction. Additional emissions away from all jets can also create their own jets, changing  $\tilde{N}_{\text{jet}}$  by an integer or near-integer amount, but an emission far from all jets cannot change  $\tilde{N}_{\text{jet}}$  by a small amount. Because of color coherence, this implies that the contribution to  $\tilde{N}_{\text{jet}}$  from a given jet is, to leading power, independent of all other jets in the event.<sup>9</sup> The  $\tilde{N}_{\text{jet}}$  distribution is therefore a sum of contributions from jet regions that are to a good approximation independently determined. We will formalize this “local” factorization structure in section 5.4, and see how it could simplify LHC calculations in section 6.

<sup>8</sup>As discussed in section 5.4, wide-angle soft radiation does affect the cross section at a fixed value of  $\tilde{N}_{\text{jet}}$ , just not the  $\tilde{N}_{\text{jet}}$  value itself.

<sup>9</sup>Color coherence states that emissions collinear to a given jet and well-separated from all other jets are only sensitive to the kinematics and color of that jet and the *anti*-color of the jet, i.e., the color of all other jets in the event [64].



**Figure 5.** The single- and double-soft modes depicted in the  $z_1$ - $z_2$  plane for the simplified observable  $\Delta = z_1 z_2$  (similar to  $\Delta_{2+}$ ). We assign the power counting  $\Delta \sim \lambda^2$ . In the case with only one soft particle, the soft energy scales as  $\lambda^2$ , whereas in the case with two soft particles, the soft energies scale as  $\lambda$ . These modes are connected by the hyperbola shown, which can be thought of as tracing out the rapidity in the energies of the two partons (see eq. (3.1)). In our calculation, we encounter divergences in this rapidity variable for each mode depicted. This picture has a strong analog in the divergences in physical rapidity of single particles in SCET<sub>II</sub>.

### 3 Rapidity-like divergences

As mentioned in section 2.2, the calculation of  $\tilde{N}_{\text{jet}}$  at  $\mathcal{O}(\alpha_s^2)$  features divergences not regulated by dimensional regularization (dim reg) when the calculation is divided into unique soft limits. These divergences have a strong resemblance to rapidity divergences in SCET<sub>II</sub>, where the large rapidities of soft and collinear modes in the effective theory generate divergent integrals not regulated by dim reg [21, 22, 56]. The similar divergences appearing in the  $\tilde{N}_{\text{jet}}$  calculation are not from physical rapidities, but instead originate from the *energy sharing* between particles and are easily cast in terms of a “rapidity” of this energy sharing:

$$y = \frac{1}{2} \ln \frac{z_1}{z_2}. \tag{3.1}$$

In this section, we show how these divergences arise in the  $\tilde{N}_{\text{jet}}$  calculation and how we can adapt standard rapidity regulators to our case.

Instead of the complete  $\tilde{N}_{\text{jet}}$  calculation, consider the simplified observable

$$\Delta = z_1 z_2, \tag{3.2}$$

where particles 1 and 2 are particles that may be soft.  $\Delta$  has the same soft scaling properties as the near-integer behavior of  $\tilde{N}_{\text{jet}}$  (see eqs. (2.10) and (2.11)), and the other complications from the complete  $\tilde{N}_{\text{jet}}$  calculation are irrelevant for the discussion here. Like  $\Delta_{2+}$  to be calculated in section 4.1,  $\Delta$  receives contributions from both single-soft and double-soft regions of phase space. Also like  $\Delta_{2+}$ ,  $\Delta$  is a non-additive observable that does not get a linear contribution from each soft emission.

The interplay between the single- and double-soft limits is illustrated in figure 5. In order to have a consistent power counting for  $\Delta$ , we need to consider two different scalings

of the soft modes. Let  $\lambda \ll 1$  be the power counting parameter, with  $\Delta \sim \lambda^2$ . The phase space regions with the same parametric contribution to  $\Delta$  are:

$$\begin{aligned}
 & 1 \text{ soft, } 2 \text{ collinear (sc): } z_1 \sim \lambda^2, \quad z_2 \sim 1, \\
 & \quad 1 \text{ soft, } 2 \text{ soft (ss): } z_1 \sim \lambda, \quad z_2 \sim \lambda, \\
 & 1 \text{ collinear, } 2 \text{ soft (cs): } z_1 \sim 1, \quad z_2 \sim \lambda^2.
 \end{aligned} \tag{3.3}$$

Thus, there are two types of soft modes — single-soft modes scale as  $\lambda^2$ , whereas double-soft modes scale as  $\lambda$  — and there is a unique soft power counting within each sector. This relative scaling is the same as ultrasoft modes in SCET<sub>I</sub> (which scale as  $\lambda^2$ ) and soft modes in SCET<sub>II</sub> (which scale as  $\lambda$ ). This indicates that the calculation has contributions to the observable from soft modes in both SCET<sub>I</sub> and SCET<sub>II</sub>. The “energy rapidity” variable  $y$  in eq. (3.1) separates these phase space regions by measuring the relative energy sharing between particles 1 and 2. Like an ordinary rapidity, the range is  $y \in (-\infty, \infty)$ . The *sc* region has  $z_1 \ll z_2$  ( $y \rightarrow -\infty$ ), the *ss* region has  $z_1 \simeq z_2$  ( $|y| \simeq 1$ ), and the *cs* region has  $z_1 \gg z_2$  ( $y \rightarrow \infty$ ). Each of the *sc/cs/ss* sectors gives an independent contribution to the  $\Delta$  distribution at  $\mathcal{O}(\alpha_s^2)$  which should be properly summed (i.e. one has to remove potential double-counting). As we will now see, unbounded integrals in  $y$  produce divergences, not regulated by dim reg.

In full QCD with dim reg in  $d = 4 - 2\epsilon$  dimensions, the soft divergences of  $\Delta$  are encapsulated by the integral

$$\begin{aligned}
 I_{\text{full}}(\Delta) &= \int_0^1 \frac{dz_1}{z_1} \frac{dz_2}{z_2} (z_1 z_2)^{-2\epsilon} \delta(\Delta - z_1 z_2) \\
 &= \Delta^{-1-2\epsilon} \int_{-\infty}^{\infty} dy \Theta\left(-\frac{1}{2} \ln(1/\Delta) < y < \frac{1}{2} \ln(1/\Delta)\right) \\
 &= \frac{1}{4\epsilon^2} \delta(\Delta) - \mathcal{L}_1(\Delta) \Theta(\Delta < 1) + \mathcal{O}(\epsilon),
 \end{aligned} \tag{3.4}$$

where  $\mathcal{L}_1$  is a logarithmic plus-function defined in eq. (B.3). If we take soft limits of the phase space, where the boundaries scale as  $[0, 1] \rightarrow [0, \infty]$ , we get:

$$\begin{aligned}
 1 \text{ soft: } I_{sc}(\Delta) &= \int_0^\infty \frac{dz_1}{z_1} \int_0^1 \frac{dz_2}{z_2} (z_1 z_2)^{-2\epsilon} \delta(\Delta - z_1 z_2) \\
 &= \Delta^{-1-2\epsilon} \int_{-\infty}^\infty dy \Theta\left(-\frac{1}{2} \ln(1/\Delta) < y\right), \\
 2 \text{ soft: } I_{cs}(\Delta) &= \int_0^1 \frac{dz_1}{z_1} \int_0^\infty \frac{dz_2}{z_2} (z_1 z_2)^{-2\epsilon} \delta(\Delta - z_1 z_2) \\
 &= \Delta^{-1-2\epsilon} \int_{-\infty}^\infty dy \Theta\left(y < \frac{1}{2} \ln(1/\Delta)\right), \\
 1, 2 \text{ soft: } I_{ss}(\Delta) &= \int_0^\infty \frac{dz_1}{z_1} \int_0^\infty \frac{dz_2}{z_2} (z_1 z_2)^{-2\epsilon} \delta(\Delta - z_1 z_2) \\
 &= \Delta^{-1-2\epsilon} \int_{-\infty}^\infty dy.
 \end{aligned} \tag{3.5}$$

Each integral is divergent and unregulated, but the  $\Delta < 1$  regime of full QCD is reproduced by the combination<sup>10</sup>

$$I_{\text{full}}(\Delta) = [I_{sc}(\Delta) + I_{cs}(\Delta) - I_{ss}(\Delta)]\Theta(\Delta < 1). \quad (3.6)$$

This result is as expected, since the double-soft limit should remove the double-counting of the single-soft limits, even if each contribution is not well-defined individually. Note that the unregulated divergences appear only in  $y$ .

This structure of divergences in  $y$  is identical to those in physical rapidities of single particles in SCET<sub>II</sub>. These divergences may be handled in a number of ways, and we will use the rapidity renormalization group (rapidity RG) [21, 22]. In the rapidity RG, divergences are regulated analogously to dim reg, using a scale  $\nu$  and infinitesimal parameter  $\eta$  that correspond to the usual  $\mu$  and  $\epsilon$  in dim reg. At one loop, these regulator factors are [21, 22]

$$\begin{aligned} 1 \text{ soft: } R_{\eta}^{sc} &= \left(\frac{\nu}{E_J}\right)^{\eta} z_2^{-\eta} = \left(\frac{\nu}{E_J}\right)^{\eta} s^{-\eta/2} e^{y\eta}, \\ 2 \text{ soft: } R_{\eta}^{cs} &= \left(\frac{\nu}{E_J}\right)^{\eta} z_1^{-\eta} = \left(\frac{\nu}{E_J}\right)^{\eta} s^{-\eta/2} e^{-y\eta}, \\ 1, 2 \text{ soft: } R_{\eta}^{ss} &= \left(\frac{\nu}{E_J}\right)^{\eta} |z_1 - z_2|^{-\eta} = \left(\frac{\nu}{E_J}\right)^{\eta} s^{-\eta/2} |2 \sinh y|^{-\eta}, \end{aligned} \quad (3.7)$$

where  $s = z_1 z_2$ , and  $E_J$  is the total jet energy such that  $z_i = E_i/E_J$ . These regulators give well-defined terms for each soft limit,

$$\begin{aligned} 1 \text{ soft: } I_{sc}(\Delta) &= \int_0^{\infty} \frac{dz_1}{z_1} \int_0^1 \frac{dz_2}{z_2} (z_1 z_2)^{-2\epsilon} \delta(\Delta - z_1 z_2) R_{\eta}^{sc} \\ &= -\left(\frac{\nu}{E_J}\right)^{\eta} \Delta^{-1-2\epsilon} \frac{1}{\eta}, \\ 2 \text{ soft: } I_{cs}(\Delta) &= \int_0^1 \frac{dz_1}{z_1} \int_0^{\infty} \frac{dz_2}{z_2} (z_1 z_2)^{-2\epsilon} \delta(\Delta - z_1 z_2) R_{\eta}^{cs} \\ &= -\left(\frac{\nu}{E_J}\right)^{\eta} \Delta^{-1-2\epsilon} \frac{1}{\eta}, \\ 1, 2 \text{ soft: } I_{ss}(\Delta) &= \int_0^{\infty} \frac{dz_1}{z_1} \int_0^{\infty} \frac{dz_2}{z_2} (z_1 z_2)^{-2\epsilon} \delta(\Delta - z_1 z_2) R_{\eta}^{ss} \\ &= \left(\frac{\nu}{E_J}\right)^{\eta} \Delta^{-1-2\epsilon-\eta/2} \left[\frac{2}{\eta} + \mathcal{O}(\eta)\right]. \end{aligned} \quad (3.8)$$

Now the sum of contributions is independent of the rapidity regulator and matches the full QCD results for  $\Delta < 1$ :

$$[I_{sc}(\Delta) + I_{cs}(\Delta) + I_{ss}(\Delta)]\Theta(\Delta < 1) = \frac{1}{4\epsilon^2} \delta(\Delta) - \mathcal{L}_1(\Delta) \Theta(\Delta < 1) + \mathcal{O}(\epsilon, \eta) = I_{\text{full}}(\Delta). \quad (3.9)$$

Comparing eq. (3.9) to eq. (3.6), we see that with the rapidity regulator, the full result is reproduced by the *sum* of single- and double-soft contributions, instead of the difference.

---

<sup>10</sup>Note that outside this toy example, only the logarithms of  $\Delta$  in full QCD will be reproduced by the soft limits, instead of the full  $\Delta < 1$  result.

This happens because the single- and double-soft limits produce canceling poles in the large positive and negative rapidity regions. For example, the large rapidity limit is allowed in both the  $sc$  and  $ss$  contributions, where the rapidity regulators scale as

$$\begin{aligned} 1 \text{ soft: } R_\eta^{sc}(y \rightarrow \infty) &\sim e^{y\eta}, \\ 1, 2 \text{ soft: } R_\eta^{ss}(y \rightarrow \infty) &\sim e^{-y\eta}, \end{aligned} \tag{3.10}$$

which lead to canceling poles:

$$\begin{aligned} 1 \text{ soft: } \int_0^\infty dy e^{y\eta} &= -\frac{1}{\eta}, \\ 1, 2 \text{ soft: } \int_0^\infty dy e^{-y\eta} &= \frac{1}{\eta}. \end{aligned} \tag{3.11}$$

Thus the cancellation of the large rapidity regimes happens at the level of the sum of contributions, instead of the difference.

Not only does the rapidity RG make the various soft contributions well defined, but it also separates the double-soft limit from the zero-bin subtraction (needed to remove double counting [65]) of the single-soft limits.<sup>11</sup> Because the rapidity regulators in the single- and double-soft limits are *not* related by scaling in rapidities, the double-soft contribution is not the zero-bin of the single-soft limit. In fact, the zero-bin limit is scaleless; the rapidity regulator is unchanged in the zero-bin limit but now all rapidities are allowed, leading to the scaleless integral [22]

$$\int_{-\infty}^{\infty} dy e^{\pm y\eta} = 0. \tag{3.12}$$

In SCET<sub>II</sub> applications where the collinear and soft contributions to the distribution can be factorized into separated jet/beam and soft functions, this factorization allows these rapidity divergences to be divided into the collinear and soft functions and the resulting logarithms to be resummed using standard techniques (see, e.g., [21, 22]).

The features of rapidity divergences in this simple example of  $\Delta$  repeat themselves in the calculation of near-integer behavior of  $\tilde{N}_{\text{jet}}$  (specifically,  $\Delta_{2+}$ ) in section 4.1 below.

## 4 Calculating the near-integer behavior

As seen in figure 4, fractional values of  $\tilde{N}_{\text{jet}}$  arise when radiation in a jet region extends beyond a radius  $R$ . In particular, near-integer values of  $\tilde{N}_{\text{jet}}$  come from soft radiation beyond  $R$ , and the deviation from integer values grows as emissions become harder. In this section, we study the leading-order near-integer behavior of  $\tilde{N}_{\text{jet}}$  in  $e^+e^- \rightarrow$  jets at a center-of-mass energy of  $Q$ , which first occurs for  $e^+e^- \rightarrow 4$  partons.

Let  $\Delta_{2\pm}$  be the jet multiplicity near the 2-jet peak as defined in eq. (2.3); we take the power counting  $\Delta_{2\pm} \sim \lambda^2$ , with  $\lambda \ll 1$ . We will first derive an analytic expression for the cross section at leading order in  $\lambda$  (i.e. the singular contributions), including a discussion of

---

<sup>11</sup>In this case, the zero-bin limit is the overlap of the single- and double-soft limits, and is tantamount to removing the restriction on  $y$  in the single-soft limits in eq. (3.5).



the aforementioned rapidity-like divergences. We will then calculate the full  $\mathcal{O}(\alpha_s^2)$  result using the Monte Carlo program EVENT2 [66, 67], which allows us to include non-singular terms as well as cross check our results for the singular contributions. We will also calculate the  $\mathcal{O}(\alpha_s^2)$  contributions to  $\Delta_{3-}$ , though this is not the focus of our studies.

### 4.1 Singular contributions using splitting functions

The cross section for  $\Delta_{2\pm}$  can be written as

$$\frac{d\sigma}{d\Delta_{2\pm}} = \int d\Phi_4 \mathcal{T}(e^+e^- \rightarrow 4 \text{ partons}) \mathcal{F}(\Delta_{2\pm}, \Phi_4), \quad (4.1)$$

where  $\Phi_4$  represents four-body phase space,  $\mathcal{T}$  is the matrix element for  $e^+e^- \rightarrow 4$  partons, and  $\mathcal{F}(\Delta_{2\pm}, \Phi_4)$  is the measurement function, which projects out the slice of phase space corresponding to a constant value of  $\Delta_{2\pm}$ . The allowed values of  $\tilde{N}_{\text{jet}}$  for four partons are:<sup>12</sup>

$$\text{Four isolated partons: } \tilde{N}_{\text{jet}} = 4, \quad (4.2)$$

$$\text{Two isolated partons and one merged pair: } \tilde{N}_{\text{jet}} = 3, \quad (4.3)$$

$$\text{One isolated parton and one merged triplet: } \tilde{N}_{\text{jet}} \in (2 - (E_{\text{cut}}/Q)^2, 3), \quad (4.4)$$

$$\text{Two merged pairs: } \tilde{N}_{\text{jet}} = 2. \quad (4.5)$$

The only non-integer behavior is obtained for the merged triplet in eq. (4.4), which requires three particles to be within an angular distance  $\lesssim 2R$ . Thus, for sufficiently small  $R$ , we can use the matrix element in the limit where three partons are collinear, allowing us to take advantage of collinear factorization [66–70]:<sup>13</sup>

$$\mathcal{T}(e^+e^- \rightarrow 4 \text{ partons}) \simeq \mathcal{T}(e^+e^- \rightarrow q\bar{q}) \cdot \sum_k \mathcal{T}_k^{\text{coll}}(1 \rightarrow 3) \quad (4.6)$$

$$= \mathcal{T}(e^+e^- \rightarrow q\bar{q}) \cdot \frac{(8\pi\alpha_s\mu^{2\epsilon})^2}{s_{123}^2} \sum_k \langle \hat{P}_{1 \rightarrow 3}^k \rangle, \quad (4.7)$$

where  $k$  labels one of the parton channels  $q \rightarrow ggq$ ,  $\bar{q} \rightarrow gg\bar{q}$ ,  $q \rightarrow q'q'q$ , or  $\bar{q} \rightarrow q'\bar{q}'\bar{q}$ ,  $s_{123}$  is the squared invariant mass of the three parton system, and  $\langle \hat{P}_{1 \rightarrow 3}^k \rangle$  is the spin-averaged  $1 \rightarrow 3$  splitting function [76, 77]. The factorization in eq. (4.6) implies that the relevant phase space for the collinear splitting is 3-body, meaning we decompose the 4-body phase space of the whole event into the 2-body  $q\bar{q}$  system of the hard interaction and the 3-body phase space of the  $1 \rightarrow 3$  splitting. In the collinear limit, the 4-body phase space factorizes as  $d\Phi_4 = d\Phi_2(q\bar{q}) \times ds_{123}/(2\pi) \times d\Phi_3(1 \rightarrow 3)$ .

The collinear matrix elements can be further decomposed according to their color structure. The channel  $q \rightarrow ggq$  (and  $\bar{q} \rightarrow gg\bar{q}$ ) can be written as a sum of Abelian (“ab”)

<sup>12</sup>Here we assume that each isolated parton and group of partons are sufficiently energetic to pass the energy cut  $E_{\text{cut}}$ . If this does not hold, then the integer values of  $\tilde{N}_{\text{jet}} = 3$  or 4 may be reduced to 2 or 3.

<sup>13</sup>We use spin-averaged splitting functions, which are sufficient since the value of  $\tilde{N}_{\text{jet}}$  for the collinear system is independent of its orientation relative to other jets in the event (see, e.g., [71]). This  $1 \rightarrow 3$  splitting function approach also appears in ref. [72] for calculating the jet substructure observable planar flow [73–75].

and non-Abelian (“nab”) contributions:

$$\mathcal{T}^{\text{coll}}(q \rightarrow gqg) = C_F^2 \mathcal{T}_{\text{ab}}^{\text{coll}}(q \rightarrow gqg) + C_F C_A \mathcal{T}_{\text{nab}}^{\text{coll}}(q \rightarrow gqg), \quad (4.8)$$

while the  $q \rightarrow q' \bar{q}' q$  (and  $\bar{q} \rightarrow q' \bar{q}' \bar{q}$ ) channels will give contributions proportional to the  $C_F T_R$  color structure.<sup>14</sup> At  $\mathcal{O}(\alpha_s^2)$  the cross section for  $\Delta_{2\pm}$  is obtained by summing over all channels and color structures. In the rest of this subsection we will discuss the calculation of the Abelian piece of the cross section; results for the other color structures can be found in appendix A.

The Abelian contribution comes from a  $q \rightarrow gqg$  or  $\bar{q} \rightarrow gq\bar{q}$  splitting. The two channels give identical contributions and we will focus on the  $q \rightarrow gqg$  case for definiteness, labeling the final state as:

$$g_1 g_2 q_3. \quad (4.9)$$

As discussed in section 2.2, at leading power we can take the soft limit of our observable, which corresponds to either *one* or *both* gluons becoming soft. As in the toy calculation in section 3, we will label the limit where gluon 1 (gluon 2) is soft as *sc* (*cs*) and the double-soft limit as *ss*. We again use the energy sharing variables  $z_i = E_i/E_J$  and scaled veto variable  $z_{\text{cut}} = E_{\text{cut}}/E_J$  in eq. (2.8), where here  $E_J$  is the energy of one jet region (with  $E_J \simeq Q/2$  for a collision at center-of-mass energy  $Q$ ). We generally assume  $z_{\text{cut}} \ll 1$ , which allows us to neglect power-suppressed contributions from soft quarks, and if some  $z_i < z_{\text{cut}}$  then that particle may be treated as soft.

In tables 1 and 2 we collect contributions to  $\Delta_{2\pm}$  in the relevant soft limits and regions of phase space. At this order, we also get a contribution to  $\Delta_{3-} = 3 - \tilde{N}_{\text{jet}}$ , and we will carry out the calculation for  $\Delta_{3-}$  as well.

We now gather the relevant pieces to evaluate eq. (4.1) in the collinear limit. There are three relevant regions of phase space:

$$\mathcal{R}_A = \Theta(\theta_{13} > R) \Theta(\theta_{23} < R) \Theta(\theta_{12} < R), \quad (4.10)$$

$$\mathcal{R}_B = \mathcal{R}_A(1 \leftrightarrow 2), \quad (4.11)$$

$$\mathcal{R}_C = \Theta(\theta_{13} < R) \Theta(\theta_{23} < R) \Theta(\theta_{12} > R), \quad (4.12)$$

where  $\theta_{ij}$  labels the opening angle between partons  $i$  and  $j$ . In table 1 we show a schematic representation of the angular phase space regions and soft limits which are explicitly listed in table 2. From table 2, the measurement functions are given by


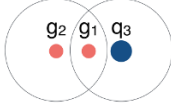


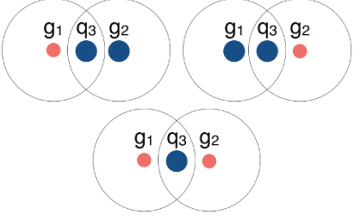


$$\mathcal{F}(\Delta_{2-}, \Phi_4) = \mathcal{F}_{\Delta_{2-}}^{ss}, \quad (4.13)$$

$$\mathcal{F}(\Delta_{2+}, \Phi_4) = \mathcal{F}_{\Delta_{2+}}^{sc} + \mathcal{F}_{\Delta_{2+}}^{cs} + \mathcal{F}_{\Delta_{2+}}^{ss}, \quad (4.14)$$

$$\mathcal{F}(\Delta_{3-}, \Phi_4) = \mathcal{F}_{\Delta_{3-}}^{sc} + \mathcal{F}_{\Delta_{3-}}^{cs}, \quad (4.15)$$

---

<sup>14</sup>For same flavor final state quarks  $q' = q$ , this matrix element contains also terms proportional to  $C_F^2$  and  $C_F C_A$ . However, these terms do not contribute to the cross section at leading power.

Observable	$\mathcal{R}_A$	$\mathcal{R}_B$	$\mathcal{R}_C$
$\Delta_{2-}$			—
$\Delta_{2+}$			
$\Delta_{3-}$			—

**Table 1.** Representation of the phase space configurations contributing to the near-integer jet multiplicities  $\Delta_{2-}$ ,  $\Delta_{2+}$ , and  $\Delta_{3-}$ . Compared to figure 4, the value of  $\tilde{N}_{\text{jet}}$  is 1 unit higher, because the event contains an additional isolated parton (not shown). For each observable, we show the corresponding contributions from different angular regions and soft limits. Circles represent cones of radius  $R$ , large blue dots represent energetic partons, small red dots soft partons with  $z < z_{\text{cut}}$ . The angular regions  $\mathcal{R}_A$ ,  $\mathcal{R}_B$ , and  $\mathcal{R}_C$  are defined in eq. (4.10).

where

$$\mathcal{F}_{\Delta_{2-}}^{ss} = \delta(\Delta_{2-} - z_1 z_2) \Theta(z_1 + z_2 < z_{\text{cut}}) \mathcal{R}_A + (1 \leftrightarrow 2), \quad (4.16)$$

$$\begin{aligned} \mathcal{F}_{\Delta_{2+}}^{sc} &= \delta[\Delta_{2+} - z_1(1/z_2 - z_2)] \Theta(z_2 > z_{\text{cut}}) \mathcal{R}_A \\ &+ \delta[\Delta_{2+} - z_1 z_2 (2 - z_2)/(1 - z_2)] \mathcal{R}_C, \end{aligned} \quad (4.17)$$

$$\mathcal{F}_{\Delta_{2+}}^{cs} = \mathcal{F}_{\Delta_{2+}}^{sc} (1 \leftrightarrow 2), \quad (4.18)$$

$$\mathcal{F}_{\Delta_{2+}}^{ss} = \delta(\Delta_{2+} - 2z_1 z_2) \mathcal{R}_C, \quad (4.19)$$

$$\mathcal{F}_{\Delta_{3-}}^{sc} = \delta(\Delta_{3-} - z_1[1 - z_2(1 - z_2)]/[z_2(1 - z_2)]) \Theta(z_2 > z_{\text{cut}}) \mathcal{R}_B, \quad (4.20)$$

$$\mathcal{F}_{\Delta_{3-}}^{cs} = \mathcal{F}_{\Delta_{3-}}^{sc} (1 \leftrightarrow 2). \quad (4.21)$$

The single- and double-soft limits for the  $1 \rightarrow 3$  matrix elements are

$$\mathcal{T}_{\text{ab}}(q \rightarrow g_1 g_2 q_3) \simeq \begin{cases} \mathcal{T}_{\text{ab}}^{sc} = \frac{4(4\pi\alpha_s\mu^{2\epsilon})^2}{E_J^4} \frac{2 - 2z_2 + z_2^2(1 - \epsilon)}{\theta_{13}^2 \theta_{23}^2 z_1^2 z_2^2 (1 - z_2)}, & (1 \text{ soft, } 2 \text{ collinear}) \\ \mathcal{T}_{\text{ab}}^{cs} = \mathcal{T}_{\text{ab}}^{sc} (1 \leftrightarrow 2), & (1 \text{ collinear, } 2 \text{ soft}) \\ \mathcal{T}_{\text{ab}}^{ss} = \frac{8(4\pi\alpha_s\mu^{2\epsilon})^2}{E_J^4} \frac{1}{\theta_{13}^2 \theta_{23}^2 z_1^2 z_2^2}, & (1 \text{ soft, } 2 \text{ soft}) \end{cases} \quad (4.22)$$

where we have included a symmetry factor of  $1/2!$  for identical gluons. Notice that in the double-soft limit, the Abelian matrix element simply reduces to the product of two eikonal

Observable	Region	Expression	Limit	Cuts
$\Delta_{2-}$	$\mathcal{R}_A$	$z_1 z_2$	1, 2 soft	$z_1 + z_2 < z_{\text{cut}}$
$\Delta_{2-}$	$\mathcal{R}_B$	$z_1 z_2$	1, 2 soft	$z_1 + z_2 < z_{\text{cut}}$
$\Delta_{2+}$	$\mathcal{R}_A$	$z_1(1 - z_2^2)/z_2$	1 soft	$z_2 > z_{\text{cut}}$
$\Delta_{2+}$	$\mathcal{R}_B$	$z_2(1 - z_1^2)/z_1$	2 soft	$z_1 > z_{\text{cut}}$
$\Delta_{2+}$	$\mathcal{R}_C$	$z_1 z_2(2 - z_2)/(1 - z_2)$	1 soft	–
$\Delta_{2+}$	$\mathcal{R}_C$	$z_1 z_2(2 - z_1)/(1 - z_1)$	2 soft	–
$\Delta_{2+}$	$\mathcal{R}_C$	$2z_1 z_2$	1, 2 soft	–
$\Delta_{3-}$	$\mathcal{R}_A$	$z_2[1 - z_1(1 - z_1)]/[z_1(1 - z_1)]$	2 soft	$z_1 > z_{\text{cut}}$
$\Delta_{3-}$	$\mathcal{R}_B$	$z_1[1 - z_2(1 - z_2)]/[z_2(1 - z_2)]$	1 soft	$z_2 > z_{\text{cut}}$

**Table 2.** Near-integer behavior of  $\tilde{N}_{\text{jet}}$ , shown for various phase space regions as depicted in table 1. In each case, the expression for the observable is given along with the relevant limits and phase space cuts.

factors. In the collinear regime, the 3-body phase space can be written as [78]

$$d\Phi_3^{\text{coll}} = \frac{E_J^{4-4\epsilon}}{(4\pi)^{4-2\epsilon}\Gamma(1-2\epsilon)} d\Phi_\Omega d\Phi_z, \tag{4.23}$$

where the angular phase space  $d\Phi_\Omega$  is given by

$$d\Phi_\Omega = d\theta_{13}^2 d\theta_{23}^2 d\theta_{12}^2 \Theta(\Delta) \Delta^{-1/2-\epsilon}, \quad \Delta = 4\theta_{13}^2 \theta_{23}^2 - (\theta_{12}^2 - \theta_{23}^2 - \theta_{13}^2)^2, \tag{4.24}$$

and the energy phase space  $d\Phi_z$  is given by

$$d\Phi_z \simeq \begin{cases} d\Phi_z^{sc} = \Theta(z_1 < \infty)\Theta(z_2 < 1)[z_1 z_2(1 - z_2)]^{1-2\epsilon} dz_1 dz_2, & (1 \text{ soft, } 2 \text{ collinear}) \\ d\Phi_z^{cs} = d\Phi_z^{sc}(1 \leftrightarrow 2), & (1 \text{ collinear, } 2 \text{ soft}) \\ d\Phi_z^{ss} = \Theta(z_1 < \infty)\Theta(z_2 < \infty)(z_1 z_2)^{1-2\epsilon} dz_1 dz_2. & (1 \text{ soft, } 2 \text{ soft}) \end{cases} \tag{4.25}$$

As discussed in section 3, for certain contributions to the observable, these energy integrals give rise to rapidity divergences which require special regularization to make each term well defined. We use the rapidity regulator to do this, and the sum of all contributions is regulator independent. Specifically, the contribution to  $\Delta_{2+}$  from  $\mathcal{R}_C$  is the only one with both single- and double-soft limits (see tables 1 and 2), and hence the only contribution which requires this extra regulator. While  $\Delta_{2-}$  does receive contributions from the double-soft region of phase space, the constraints on the gluon energies from  $z_{\text{cut}}$  implies that both gluons must be soft, and hence the single-soft limits do not contribute. The limit on the gluon energies also implies there is a kinematic limit on the observable,  $\Delta_{2-} < (z_{\text{cut}}/2)^2$ .

By combining the measurement function with the proper limits of the matrix elements, we get an analytic expression for the Abelian contribution to the  $\Delta_{2\pm}$  and  $\Delta_{3-}$  distributions

at leading power in  $\lambda$ . Including the identical contribution from  $\bar{q} \rightarrow \bar{q}gg$  yields

$$\left[ \frac{d\sigma}{d\Delta_{2-}} \right]_{\text{ab}} = C_F^2 \left( \frac{\alpha_s}{\pi} \right)^2 \mathcal{I}_\Omega \{ -2\mathcal{L}_1(\Delta_{2-}) + 4 \ln z_{\text{cut}} \mathcal{L}_0(\Delta_{2-}) \} \Theta(\Delta_{2-} < z_{\text{cut}}^2/4), \quad (4.26)$$

$$\left[ \frac{d\sigma}{d\Delta_{2+}} \right]_{\text{ab}} = C_F^2 \left( \frac{\alpha_s}{\pi} \right)^2 \mathcal{I}_\Omega \left\{ -\frac{14}{5} \mathcal{L}_1(\Delta_{2+}) + \left( -\frac{57}{10} + \frac{14}{5} \ln 2 - 2 \ln z_{\text{cut}} \right) \mathcal{L}_0(\Delta_{2+}) \right\}, \quad (4.27)$$

$$\left[ \frac{d\sigma}{d\Delta_{3-}} \right]_{\text{ab}} = C_F^2 \left( \frac{\alpha_s}{\pi} \right)^2 \mathcal{I}_\Omega \left( -\frac{3}{2} - 2 \ln z_{\text{cut}} \right) \mathcal{L}_0(\Delta_{3-}), \quad (4.28)$$

where the distributions  $\mathcal{L}_n$  are the usual logarithmic plus distributions, defined in appendix B. The integral over the angular phase space is given by<sup>15</sup>

$$\mathcal{I}_\Omega = \frac{4}{\pi} \int \frac{d\Phi_\Omega}{\theta_{13}^2 \theta_{23}^2} \mathcal{R}_A = \frac{5\pi^2}{54}. \quad (4.29)$$

Eqs. (4.26), (4.27), and (4.28) represent the leading order distributions in the small  $R$  and  $z_{\text{cut}}$  limits. Notice that at this order there is no dependence on  $R$ , as the  $1/\theta_{ij}^2$  factors in the denominator of the angular integral provide a logarithmic scaling which can be used to explicitly remove the  $R$  dependence. Since we are focused on the non-integer behavior of  $\tilde{N}_{\text{jet}}$ , we have suppressed contributions proportional to  $\delta(\Delta_{n\pm})$  in our results (see, however, eq. (5.5)).

The calculation of the  $C_F C_A$  and  $C_F T_R n_f$  contributions are very similar to the Abelian case, albeit more involved as the matrix elements do not simply factorize into separate angular and energy functions as in eq. (4.22). In appendix A, we give the full set of results for the different color structures.

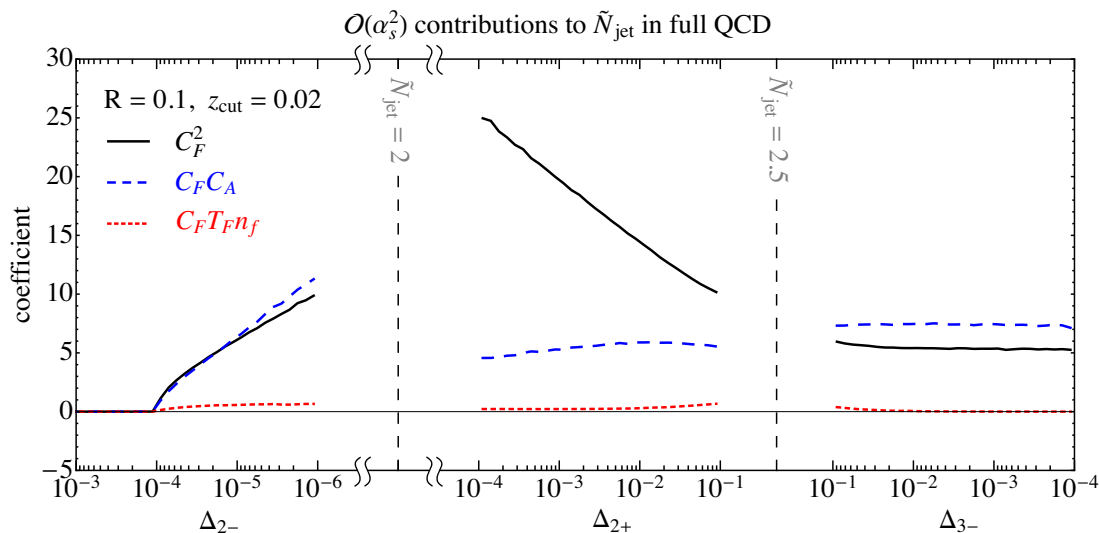
In the  $\mathcal{O}(\alpha_s^2)$  calculation performed above, double logarithms of  $\Delta_{n\pm}$  (appearing as  $\mathcal{L}_1(\Delta_{n\pm})$ ) arose when both emitted gluons became soft, and single logarithms ( $\mathcal{L}_0(\Delta_{n\pm})$ ) arose when a single gluon became soft. This correspondence holds true with more soft emissions: each soft emission results in a single logarithm of  $\Delta_{n\pm}$  (in certain regions of phase space). Therefore, in general the largest logarithm at  $\mathcal{O}(\alpha_s^k)$  appears when all emissions are soft gluons, and the contribution to the cross section is of the form  $\alpha_s^k \mathcal{L}_{k-1}(\Delta_{n\pm})$ .

## 4.2 Nonsingular contributions from Event2

To go beyond leading power in  $\lambda$ , we need to incorporate the  $\mathcal{O}(\alpha_s^2)$  matrix elements from full QCD. For fixed-order calculations of  $e^+e^- \rightarrow$  partons at low multiplicity, the program EVENT2 is a particularly useful tool [66, 67]. EVENT2 performs next-to-leading order calculations of  $e^+e^- \rightarrow 2$  and 3 partons, so it includes the needed  $e^+e^- \rightarrow 4$  parton tree-level matrix elements. Conveniently, it allows for the decomposition of its results order-by-order in  $\alpha_s$  and by color structure. Crucially, EVENT2 can probe the far infrared regions of phase space, meaning the (singular) logarithmic terms of  $\Delta_{2\pm}$  and  $\Delta_{3-}$  are enhanced. This allows us to perform a robust comparison and cross check of our splitting

---

<sup>15</sup>We have written the common angular coefficient  $\mathcal{I}_\Omega$  in terms of the phase space region  $\mathcal{R}_A$ . The same angular coefficient appears for regions  $\mathcal{R}_A$ ,  $\mathcal{R}_B$ , and  $\mathcal{R}_C$  due to an unexpected symmetry in the angular integrals. This symmetry is only present for certain angular integrals in the non-Abelian case.



**Figure 6.** The  $\Delta_{2\pm}$  and  $\Delta_{3-}$  distributions extracted from EVENT2. Shown are the separate  $C_F^2$ ,  $C_F C_A$ , and  $C_F T_F n_f$  contributions to the cross section, plotted as the coefficient of  $(\alpha_s/2\pi)^2 C$ , where  $C$  is the relevant color factor. Note that the  $\Delta_{2-}$  and  $\Delta_{3-}$  axes run backwards.

function calculations above. Having verified the singular logarithmic contributions to  $\Delta_{2\pm}$  and  $\Delta_{3-}$ , we can then extract the nonsingular  $\mathcal{O}(\alpha_s^2)$  contributions directly from EVENT2.

In figure 6, we show the  $\mathcal{O}(\alpha_s^2)$  contributions to  $\Delta_{2\pm}$  and  $\Delta_{3-}$  extracted from EVENT2. We plot the coefficients of the  $(\alpha_s/2\pi)^2 C$  terms in the cross section as a function of  $\ln \Delta_{n\pm}$ , where  $C$  is the relevant color factor ( $C_F^2$ ,  $C_F C_A$ , or  $C_F T_F n_f$ ). To enhance the logarithmic contributions and minimize the power corrections in  $R$  and  $z_{\text{cut}}$ , we choose the small values  $R = 0.1$  and  $z_{\text{cut}} = 0.02$ . Plotted this way, double logarithms ( $\mathcal{L}_1(\Delta_{n\pm})$ ) appear as lines of constant non-zero slope,<sup>16</sup> single logarithms ( $\mathcal{L}_0(\Delta_{n\pm})$ ) appear as lines with zero slope and non-zero offset, and nonsingular contributions vanish as  $\Delta_{n\pm} \rightarrow 0$ . It is clear that logarithmically-enhanced terms are indeed present in the full QCD result from EVENT2.

To make sure the logarithmic behavior from EVENT2 matches our analytic calculations in section 4.1 and appendix A, we can extract the nonsingular contribution to the cross section, which are the residual fixed-order terms after the logarithmic contributions are subtracted:

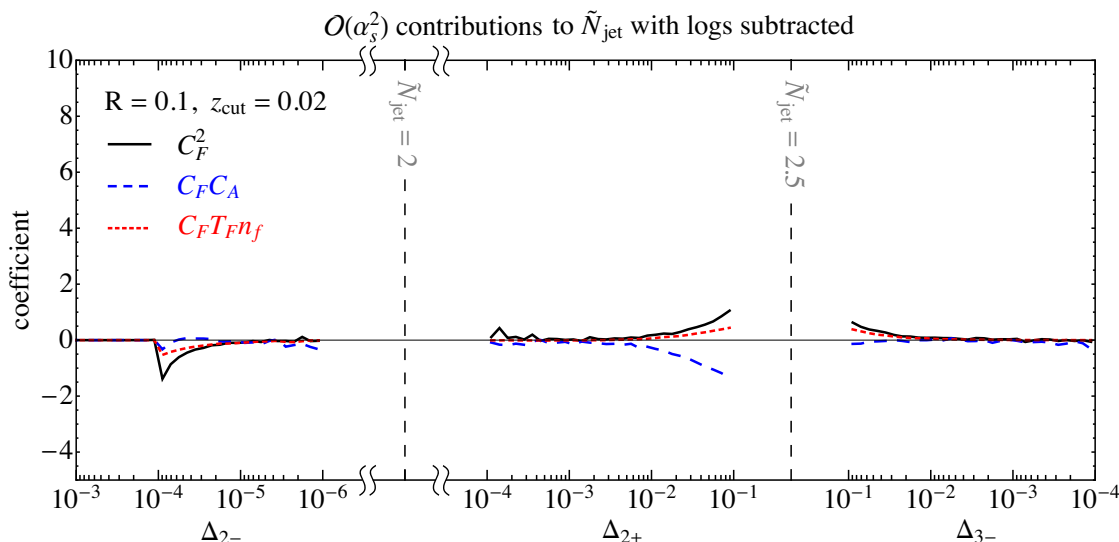
$$\frac{d\sigma_{\text{ns}}}{d\Delta_{n\pm}} = \frac{d\sigma_{\text{full}}}{d\Delta_{n\pm}} - \frac{d\sigma_{\text{sing}}}{d\Delta_{n\pm}}. \tag{4.30}$$

These are shown in figure 7, again separated by color structure, and confirm that our splitting function calculation, which includes the leading contributions in a small  $R$  and  $z_{\text{cut}}$  expansion, captures the leading-order near-integer behavior of  $\tilde{N}_{\text{jet}}$  correctly.<sup>17</sup>

For  $\tilde{N}_{\text{jet}} < 2$ , the singular terms tend to dominate over the nonsingular ones. The reason is that the  $\Delta_{2-}$  distribution has a kinematic endpoint at  $(z_{\text{cut}}/2)^2 \ll 1$  at this order, so the distribution only has support in the singular region of phase space.

<sup>16</sup>Note that these are not Sudakov double logarithms, since they appear at  $\mathcal{O}(\alpha_s^2)$ , not  $\mathcal{O}(\alpha_s)$ .

<sup>17</sup>Power suppressed terms of the form  $\mathcal{O}(R^2, z_{\text{cut}}) \mathcal{L}_{0,1}(\Delta_{n\pm})$  are generically present in the nonsingular terms, though their contributions are negligible in figure 7 since we have chosen both  $R$  and  $z_{\text{cut}}$  small.



**Figure 7.** Same as figure 6, but subtracting our calculations for the singular contributions in each region. The residuals vanish in the logarithmic ( $\Delta_{n\pm} \rightarrow 0$ ) regime, indicating that the only remaining terms are nonsingular.

For  $2 < \tilde{N}_{\text{jet}} < 3$ , there is an ambiguity in determining the nonsingular terms in the cross section, related to the fact that singular terms exist at multiple points in the  $\tilde{N}_{\text{jet}}$  spectrum. In this range, there are singular terms from  $\Delta_{2+}$  that dominate near  $\tilde{N}_{\text{jet}} = 2$  and singular terms from  $\Delta_{3-}$  that dominate near  $\tilde{N}_{\text{jet}} = 3$ . In figure 7, we defined the nonsingular term by removing the singular terms from the nearest singular point in the  $\tilde{N}_{\text{jet}}$  distribution, using  $\tilde{N}_{\text{jet}} = 2.5$  as an arbitrary dividing line.<sup>18</sup> Alternatively, we could define the nonsingular term by removing *both* sets of singular terms over the entire spectrum.<sup>19</sup> This is a valid approach as well, since the singular  $\Delta_{2+}$  and  $\Delta_{3-}$  distributions are governed by different soft limits of phase space, hence there is no double-counting by including both singularities.

Since we are mainly interested in describing the behavior in the vicinity of  $\tilde{N}_{\text{jet}} = 2$ , though, for the rest of the paper we will simply adopt the nonsingular definition in eq. (4.30), even in the vicinity of  $\tilde{N}_{\text{jet}} = 3$ :

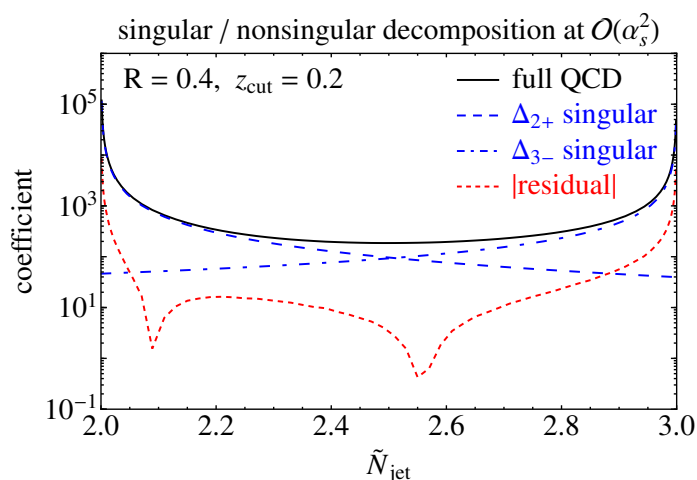
$$\frac{d\sigma_{\text{ns}}}{d\tilde{N}_{\text{jet}}}(2 < \tilde{N}_{\text{jet}} < 3) = \frac{d\sigma_{\text{ns}}}{d\Delta_{2+}} = \frac{d\sigma_{\text{full}}}{d\tilde{N}_{\text{jet}}} - \frac{d\sigma_{\text{sing}}}{d\Delta_{2+}}. \quad (4.31)$$

<sup>18</sup>This corresponds to the nonsingular definition

$$\frac{d\sigma_{\text{ns}}}{d\tilde{N}_{\text{jet}}}(2 < \tilde{N}_{\text{jet}} < 3) = \frac{d\sigma_{\text{full}}}{d\tilde{N}_{\text{jet}}} - \left[ \frac{d\sigma_{\text{sing}}}{d\Delta_{2+}} \Theta(\tilde{N}_{\text{jet}} < 2.5) + \frac{d\sigma_{\text{sing}}}{d\Delta_{3-}} \Theta(\tilde{N}_{\text{jet}} > 2.5) \right].$$

<sup>19</sup>This would correspond to the nonsingular definition

$$\frac{d\sigma_{\text{ns}}}{d\tilde{N}_{\text{jet}}}(2 < \tilde{N}_{\text{jet}} < 3) = \frac{d\sigma_{\text{full}}}{d\tilde{N}_{\text{jet}}} - \left[ \frac{d\sigma_{\text{sing}}}{d\Delta_{2+}} + \frac{d\sigma_{\text{sing}}}{d\Delta_{3-}} \right].$$



**Figure 8.** The  $2 < \tilde{N}_{\text{jet}} < 3$  distribution in full QCD (black, solid), decomposed into singular contributions from  $\Delta_{2+}$  and  $\Delta_{3-}$  (blue, dashed/dot-dashed) and residual nonsingular contributions (red, dotted). The coefficient of  $(\alpha_s/2\pi)^2$  is plotted for each contribution. Note that the sum of singular contributions is numerically dominant over the entire range.

That said, this non-singular term turns out to be dominated by the singular  $\Delta_{3-}$  piece. Writing the nonsingular term as

$$\frac{d\sigma_{\text{ns}}}{d\Delta_{2\pm}} = \frac{d\sigma_{\text{sing}}}{d\Delta_{3-}} + \frac{d\sigma_{\text{res}}}{d\Delta_{2\pm}}, \quad (4.32)$$

the residual term is quite small, even in the vicinity of  $\tilde{N}_{\text{jet}} = 2.5$ . This is shown in figure 8, where the full QCD result between  $2 < \tilde{N}_{\text{jet}} < 3$  is decomposed into the  $\Delta_{2+}$  singular,  $\Delta_{3-}$  singular, and residual terms for  $R = 0.4$  and  $z_{\text{cut}} = 0.2$ . The fact that the  $\Delta_{2+}$  nonsingular term is nearly saturated by the  $\Delta_{3-}$  singular term suggests that higher-order logarithmic terms can play an important role in determining the shape of the  $\tilde{N}_{\text{jet}}$  distribution, even for large deviations from integer values of the observable.

### 4.3 Complete results to order $\alpha_s^2$

Summarizing the results from sections 4.1 and 4.2, the full  $\mathcal{O}(\alpha_s^2)$  result for the near-integer  $\Delta_{2\pm}$  distributions can be written as

$$\frac{d\sigma}{d\Delta_{2\pm}} = \underbrace{\left[ \frac{d\sigma}{d\Delta_{2\pm}} \right]_{\text{ab}} + \left[ \frac{d\sigma}{d\Delta_{2\pm}} \right]_{\text{nab}} + \left[ \frac{d\sigma}{d\Delta_{2\pm}} \right]_{C_F T_R n_f}}_{\text{singular}} + \frac{d\sigma_{\text{ns}}}{d\Delta_{2\pm}}. \quad (4.33)$$

The first three terms in eq. (4.33) are the singular contributions to the cross section. The Abelian terms are given in eqs. (4.26) and (4.27), while the contributions of non-Abelian and  $C_F T_R n_f$  color structures can be found in appendix A. The last term in eq. (4.33) is the nonsingular contribution that we extract numerically from EVENT2 as discussed above.



As discussed in eq. (4.31), the entirety of the  $\Delta_{3-}$  singular distribution in eq. (4.28) has been absorbed into the  $\Delta_{2+}$  nonsingular distribution. Thus, when we plot  $\Delta_{3-}$  in section 7, we are in effect plotting the prediction for  $\Delta_{3-} = 1 - \Delta_{2+}$ .

## 5 Towards a factorization theorem

In the previous section, we have seen that soft and collinear matrix elements govern the near-integer behavior of  $\tilde{N}_{\text{jet}}$  in fixed-order QCD. In this section, we explore the all-orders behavior at and near  $\tilde{N}_{\text{jet}} = 2$ , building a candidate factorization theorem to describe the logarithmically-enhanced contributions to the cross section.

Near  $\tilde{N}_{\text{jet}} = 2$ , energetic collinear modes must be confined to two jet regions of diameter  $R$ , such that no two energetic particles are more than  $R$  apart. This suggests that the collinear radiation may be described by jet functions. Additional radiation must be sufficiently soft so as not to create an additional jet, meaning there is a local veto of size  $z_{\text{cut}}$  outside of the primary jets. This suggests that additional radiation may be described by a soft function. However, while one might hope that a factorization theorem of the form

$$\sigma(\Delta_{2\pm}) \sim \sigma_0 H_{q\bar{q}} \left[ J_q(\Delta_{2\pm}) \otimes J_{\bar{q}}(\Delta_{2\pm}) \otimes S_{q\bar{q}}(\Delta_{2\pm}) \right] \quad (5.1)$$

would hold, it fails on several fronts. First, this standard factorization picture is challenged by both the non-additivity of the observable and soft-collinear non-factorization (discussed in section 2.3 and examined in more detail below). Additionally, the logarithms of  $\Delta_{2\pm}$  are non-global, and may not be summed with the above factorization into jet and soft functions. Furthermore, the cross section at  $\tilde{N}_{\text{jet}} = 2$  ( $\Delta_{2\pm} = 0$ ) behaves similarly to a dijet cross section from a standard discrete jet algorithm, for which a convolution structure does not apply.

We will discuss these issues below, en route to a candidate “local” factorization theorem which applies in the small  $R$  limit. This factorization theorem will take the form

$$\frac{d\sigma}{d\Delta_{2\pm}} = \sigma(\tilde{N}_{\text{jet}} = 2) \left[ C_q(\Delta_{2\pm}) \otimes C_{\bar{q}}(\Delta_{2\pm}) \right], \quad (5.2)$$

where  $\sigma(\tilde{N}_{\text{jet}} = 2)$  is the cross section exactly at  $\tilde{N}_{\text{jet}} = 2$  (see section 5.1) and  $C_{q,\bar{q}}$  are “collinear functions” (see section 5.3). This form satisfies a number of plausibility checks, but a formal proof of its validity is beyond the scope of the present paper.

### 5.1 The cross section at $\tilde{N}_{\text{jet}} = 2$

We begin by considering the cross section at exactly  $\tilde{N}_{\text{jet}} = 2$ . At  $\mathcal{O}(\alpha_s)$ , the constraints to obtain  $\tilde{N}_{\text{jet}} = 2$  are the same as for obtaining two jets from a discrete jet algorithm, namely that either two of the three partons must lie within a mutual radius of  $R$  or one of the partons must be below  $E_{\text{cut}}$ . At leading power, the phase space restrictions on the  $q\bar{q}g$  final state are

$$\Theta(\theta_{qg} < R) + \Theta(\theta_{\bar{q}g} < R) + \Theta(\theta_{qg} > R)\Theta(\theta_{\bar{q}g} > R)\Theta(z_g < z_{\text{cut}}). \quad (5.3)$$

At higher orders, the phase space restrictions differ from discrete jet algorithms (which also differ among themselves), and the  $\tilde{N}_{\text{jet}} = 2$  cross section has unique features. Consider an  $\mathcal{O}(\alpha_s^2)$  configuration contributing to the near-integer behavior, as discussed in the section 4.1. A concrete example is the  $\Delta_{2+}\mathcal{R}_C$  configuration in table 1, which has a collinear  $q \rightarrow qg$  splitting with another soft gluon (labeled  $s$ ) emitted such that

$$\theta_{qg} < R, \quad \theta_{qs} > R, \quad \theta_{gs} < R. \tag{5.4}$$

This event will give a non-integer contribution to  $\tilde{N}_{\text{jet}}$ , and its contribution to the cross section is of the form

$$\sim \alpha_s^2 \left[ \frac{1}{\epsilon^2} \delta(\tilde{N}_{\text{jet}} - 2) + \zeta \delta(\tilde{N}_{\text{jet}} - 2) + \kappa_0 \mathcal{L}_0(\Delta_{2+}) + \kappa_1 \mathcal{L}_1(\Delta_{2+}) \right]. \tag{5.5}$$

The plus distribution terms were already calculated in eq. (4.27). The  $\delta(\Delta_{2+})$  terms were suppressed in our previous discussion (since we were focused on the non-integer behavior), but both the  $1/\epsilon^2$  term and the  $\zeta$  term follow directly from the calculations in section 4.1.

The structure of eq. (5.5) reflects the hybrid jet algorithm/event shape behavior of  $\tilde{N}_{\text{jet}}$ . The plus distribution terms have no support at  $\tilde{N}_{\text{jet}} = 2$ , since the plus function prescription removes any divergence there.<sup>20</sup> This feature is similar to other event shape variables with a singular limit, such as thrust, where there is zero cross section in the singular limit at higher orders. By contrast, the delta function terms behave more like a jet algorithm, as anticipated in eq. (2.4), with a finite cross section at exactly  $\tilde{N}_{\text{jet}} = 2$ . The IR pole from the real radiation at  $\tilde{N}_{\text{jet}} = 2$  will cancel divergences from virtual matrix elements, such that the  $\tilde{N}_{\text{jet}} = 2$  cross section is IR finite.<sup>21</sup> The additional finite  $\zeta$  term at  $\tilde{N}_{\text{jet}} = 2$  implies that the non-global structure in the near-integer part of the cross section is also contributing at  $\tilde{N}_{\text{jet}} = 2$ .

## 5.2 Soft-collinear non-factorization

The non-global nature of the near-integer behavior of  $\tilde{N}_{\text{jet}}$  is suggestive of a non-standard picture of factorization. In fact, the  $\tilde{N}_{\text{jet}}$  observable does not obey soft-collinear factorization, meaning separate soft and collinear functions cannot be easily (or usefully) defined. This was foreshadowed in section 2.3 and can be illustrated with a simple example at  $\mathcal{O}(\alpha_s^2)$ .

Consider the same phase space configuration as eq. (5.4), which has a single soft gluon and a pair of collinear (energetic) partons. The soft gluon is in the region of phase space where it is in the jet region of only one of the collinear partons, and the value of the observable is dependent on this fact. This implies that the observable receives contributions that intrinsically depend on the soft and collinear modes in a non-factorizable

<sup>20</sup>To see this, we can use the definition of the plus function in eq. (B.1). Integrating  $[q(x)]_+$  against  $\delta(x)$  gives the value at 0,

$$[q(0)]_+ = \int dx [q(x)]_+ \delta(x) = 0. \tag{5.6}$$

<sup>21</sup>It is straightforward to see that at  $\mathcal{O}(\alpha_s^2)$  the cross section at  $\tilde{N}_{\text{jet}} = 2$  must be IR finite. Since non-integer values of  $\tilde{N}_{\text{jet}}$ , as well as  $\tilde{N}_{\text{jet}} = 3$  and 4, have IR finite cross sections, and since the inclusive cross section is IR finite, the remaining piece, the cross section at  $\tilde{N}_{\text{jet}} = 2$ , must be as well.

way; the measurement function cannot be separated into separate soft and collinear pieces that do not depend on each other (see figure 4). Therefore, soft-collinear factorization does not hold. We note that, unlike other cases where soft-collinear factorization is not straightforward (see, e.g., [63, 79]), here the *leading* non-integer behavior does not factorize.

Because the non-global structure of  $\Delta_{2\pm}$  feeds into the  $\tilde{N}_{\text{jet}} = 2$  cross section through the  $\zeta$  term in eq. (5.5), the same non-factorization is also true of the integer value. Of course, for the exact  $\tilde{N}_{\text{jet}} = 2$  case, non-factorization of the soft and collinear modes happens only at  $\mathcal{O}(\alpha_s^2)$ . Therefore, one can write the  $\tilde{N}_{\text{jet}} = 2$  cross section as the sum of terms, one with global contributions that can consistently be resummed and one with non-factorizing contributions. Such a form is

$$\begin{aligned} \sigma(\tilde{N}_{\text{jet}} = 2) &= \sigma_0 H_{q\bar{q}}(Q, \mu) J_q(Q, R, z_{\text{cut}}, \mu) J_{\bar{q}}(Q, R, z_{\text{cut}}, \mu) S_{q\bar{q}}(R, z_{\text{cut}}, \mu) \\ &+ \sigma_2^{\text{non-fac}}(Q, R, z_{\text{cut}}, \mu). \end{aligned} \tag{5.7}$$

The factorized part of the cross section is similar to dijet (or, generally, exclusive jet) cross sections. In  $e^+e^-$  collisions, such cross sections are known to contain non-global logarithms that spoil a standard effective theory picture of the dynamics. These non-global logarithms span the jet and soft functions, affecting the RG evolution in nontrivial ways. Similar to the non-factorizing terms, the non-global contributions start at  $\mathcal{O}(\alpha_s^2)$ .

### 5.3 Introducing collinear functions

Let us now try to simultaneously describe the near-integer and exact-integer behavior of  $\tilde{N}_{\text{jet}}$ . We have established that the logarithms of  $\Delta_{2\pm}$  are purely non-global, and there is no standard soft-collinear factorization. As mentioned in section 2.3, however, the contributions from each jet region are independent, meaning they can be separated:

$$\frac{d\sigma}{d\Delta_{2\pm}} \sim \sigma_0 C_q(\Delta_{2\pm}) \otimes C_{\bar{q}}(\Delta_{2\pm}), \tag{5.8}$$

where  $\sigma_0$  is an overall prefactor and  $\otimes$  refers to the standard convolution in eq. (B.6). Here, we have introduced “collinear functions”  $C_q$  and  $C_{\bar{q}}$  for the two jet regions, which give the contribution to  $\Delta_{2\pm}$  for each jet region separately.

The collinear functions contain only the singular terms and have the general form

$$C_{q,\bar{q}}(\Delta_{2\pm}) = \delta(\Delta_{2\pm}) + \sum_{n=2}^{\infty} \left(\frac{\alpha_s}{\pi}\right)^n \sum_{k=-1}^{n-1} \left[ \kappa_{k,+}^{(n)} \mathcal{L}_k(\Delta_{2+}) + \kappa_{k,-}^{(n)} \mathcal{L}_k(\Delta_{2-}) \right], \tag{5.9}$$

where  $\kappa_{k,\pm}^{(n)}$  are coefficients which in general depend on  $z_{\text{cut}}$  (except for the leading coefficient  $\kappa_{n-1,\pm}^{(n)}$ ). Recall that  $\mathcal{L}_{-1}(x) = \delta(x)$ , and these delta functions are needed to describe the  $\tilde{N}_{\text{jet}} = 2$  cross section. Since  $\Delta_{2\pm}$  describe different regimes of the same observable ( $\tilde{N}_{\text{jet}}$  just above/below 2), each term in the expansion has support for one term or the other.

The convolution between the collinear functions in eq. (5.8) mixes the distributions for  $\Delta_{2+}$  and  $\Delta_{2-}$ . Therefore, we require not only the usual convolutions between distributions of either  $\Delta_{2+}$  or  $\Delta_{2-}$  (which we refer to as convolutions of one-sided distributions, e.g.

$\mathcal{L}_0(\Delta_{2+}) \otimes \mathcal{L}_0(\Delta_{2+})$ ), but also convolutions between distributions of  $\Delta_{2+}$  and  $\Delta_{2-}$  (which we refer to as convolutions of two-sided distributions, e.g.  $\mathcal{L}_0(\Delta_{2+}) \otimes \mathcal{L}_0(\Delta_{2-})$ ). Details about the definitions of one- and two-sided distributions and convolutions between them are presented in appendix B.

We can now reinterpret the calculation in section 4.1 directly as a calculation of  $C_q$  and  $C_{\bar{q}}$  to  $\mathcal{O}(\alpha_s^2)$  (see appendix C). The convolution between  $C_q$  and  $C_{\bar{q}}$  in eq. (5.8) then gives part of the higher-order  $\mathcal{O}(\alpha_s^4)$  cross section. In addition, we can estimate the  $\mathcal{O}(\alpha_s^4)$  structure of an individual collinear function by performing naïve (Abelian) exponentiation to capture some of the multiple emission terms. That is, we assume that each of the collinear functions in eq. (5.9) has the form

$$C_{q,\bar{q}}(\Delta_{2\pm}) = \delta(\Delta_{2\pm}) + \left(\frac{\alpha_s(\mu)}{\pi}\right)^2 \left[ \mathcal{K}_+(\Delta_{2+}) + \mathcal{K}_-(\Delta_{2-}) \right] + \frac{1}{2} \left(\frac{\alpha_s(\mu)}{\pi}\right)^4 \left[ \mathcal{K}_+(\Delta_{2+}) + \mathcal{K}_-(\Delta_{2-}) \right] \otimes \left[ \mathcal{K}_+(\Delta_{2+}) + \mathcal{K}_-(\Delta_{2-}) \right], \quad (5.10)$$

where

$$\mathcal{K}_+(\Delta_{2+}) = \kappa_{0,+}^{(2)} \mathcal{L}_0(\Delta_{2+}) + \kappa_{1,+}^{(2)} \mathcal{L}_1(\Delta_{2+}), \quad (5.11)$$

$$\mathcal{K}_-(\Delta_{2-}) = \kappa_{0,-}^{(2)} \mathcal{L}_0(\Delta_{2-}) + \kappa_{1,-}^{(2)} \mathcal{L}_1(\Delta_{2-}). \quad (5.12)$$

The convolution terms at  $\mathcal{O}(\alpha_s^4)$  are those coming from naïve Abelian exponentiation in one jet region. Note that this does not fully capture the correct higher-order terms (for example, we do not get any term at  $\mathcal{O}(\alpha_s^3)$  from this exponentiation), but we will see in section 7 that it is enough to reproduce some of the higher-order effects observed in parton shower Monte Carlo generators. In defining eq. (5.10), we have used the fact that one can absorb corrections to the  $\tilde{N}_{\text{jet}} = 2$  cross section into the  $\sigma_0$  prefactor (see further discussion below).

#### 5.4 A “local” factorization theorem

Via eq. (5.8), we can capture the impact on  $\tilde{N}_{\text{jet}}$  of soft and collinear emissions within the two jet regions. But soft emissions away from the jet regions can still give logarithmically-enhanced contributions to the cross section, even if they do not change the value of  $\tilde{N}_{\text{jet}}$ . For example, a soft gluon well-separated from the jets with energy fraction less than  $z_{\text{cut}}$  will not change  $\tilde{N}_{\text{jet}}$ , but it will contribute to the cross section. For  $e^+e^- \rightarrow q\bar{q}g$ , this wide-angle soft radiation is part of the  $\tilde{N}_{\text{jet}} = 2$  phase space. Going to higher orders, there is a contribution to both the exact-integer and near-integer cross sections from wide-angle soft radiation, and up to power corrections for small  $R$ , both contributions are identical.

This logic implies that the  $\tilde{N}_{\text{jet}} = 2$  cross section should *multiply* the near-integer contributions, leading to the candidate factorization theorem from eq. (5.2), repeated for convenience:

$$\frac{d\sigma}{d\Delta_{2\pm}} = \sigma(\tilde{N}_{\text{jet}} = 2) \left[ C_q(\Delta_{2\pm}) \otimes C_{\bar{q}}(\Delta_{2\pm}) \right]. \quad (5.13)$$

This is the same structure as eq. (5.8), but we have identified  $\sigma_0$  with  $\sigma(\tilde{N}_{\text{jet}} = 2)$ . As discussed in eq. (5.7),  $\sigma(\tilde{N}_{\text{jet}} = 2)$  itself has its own quasi-factorization theorem. If this

candidate factorization theorem is indeed true, then the definition of  $C_{q,\bar{q}}(\Delta_{2\pm})$  in eq. (5.10) should be revised such that the coefficient of the  $\delta(\Delta_{2\pm})$  piece is always 1 (i.e. the sum over  $k$  should start at 0 instead of  $-1$ ).

We stress here that eq. (5.2) is only a candidate factorization theorem, and we have not proven that such a form exists. In particular, we do not have an operator definition of the collinear functions  $C_{q,\bar{q}}$ , and without such a definition, the extraction of  $C_{q,\bar{q}}$  is ambiguous from fixed-order calculations alone. The reason is that the structure of  $C_q(\Delta_{2\pm}) \otimes C_{\bar{q}}(\Delta_{2\pm})$  is single-logarithmic order-by-order, without any definite relation between coefficients, so higher-order terms can absorb corrections from lower-order ones. For example, the leading nontrivial terms in  $\sigma(\tilde{N}_{\text{jet}} = 2)$  (the  $\mathcal{O}(\alpha_s)$  term) and  $C_q(\Delta_{2\pm}) \otimes C_{\bar{q}}(\Delta_{2\pm})$  (the  $\mathcal{O}(\alpha_s^2 \mathcal{L}_1)$  terms) will contribute to  $\kappa_{1,\pm}^{(3)}$  (the  $\mathcal{O}(\alpha_s^3) \mathcal{L}_1(\Delta_{2\pm})$  terms), but cannot unambiguously determine that coefficient.

Despite these limitations, our candidate factorization theorem does describe important effects (like convolutions between the two jet regions) that go beyond a simple perturbative, log series expansion, which is why we will use eq. (5.2) in our comparison studies in section 7. One thing we can say unambiguously is that if the factorization theorem in eq. (5.2) is valid, then the collinear functions must be RG independent. The reason is that the prefactor  $\sigma(\tilde{N}_{\text{jet}} = 2)$  is itself a cross section, so it must be RG independent, and therefore  $C_q \otimes C_{\bar{q}}$  must be RG independent. Similarly, the modes that contribute to  $C_q$  and  $C_{\bar{q}}$  are completely disjoint, so there is no possibility that RG-scale-dependence could cancel between  $C_q$  and  $C_{\bar{q}}$ . This does not rule out, however, a further factorization of the collinear functions.

### 5.5 Complete results

Let us summarize our final prediction for the  $\Delta_{2\pm}$  distributions using the candidate factorization theorem in eq. (5.2). Beyond the calculation summarized in section 4.3, we can include two higher-order effects.

First, as described in section 5.3, the collinear function approach allows us to capture  $\mathcal{O}(\alpha_s^4)$  terms coming from convolutions. There are two types of convolutions: convolutions between  $C_q$  and  $C_{\bar{q}}$  given by the factorization theorem, and convolutions within an individual collinear function coming from naïve Abelian exponentiation. Though the naïve exponentiation can in principle give results to all orders in  $\alpha_s$ , we will truncate the convolutions to order  $\mathcal{O}(\alpha_s^4)$ , especially since there are known  $\mathcal{O}(\alpha_s^3)$  terms missed by this approach.

Second, we can include running coupling effects by evaluating  $\alpha_s(\mu)$  at an energy scale  $\mu = Q\sqrt{\Delta_{2\pm}}$ . To see why this is the correct scale, note that in section 4.1, the fixed-order expansion in dim reg had a prefactor of

$$\left(\frac{\mu^2}{E_J^2 \Delta_{2\pm}}\right)^{2\epsilon}, \tag{5.14}$$

where  $E_J \simeq Q/2$  is the jet energy. This implies that at higher orders, the fixed-order calculation generates terms like  $\ln(\mu^2/E_J^2 \Delta_{2\pm})$  where  $\mu$  is the renormalization scale, suggesting that  $Q\sqrt{\Delta_{2\pm}}$  is a relevant running coupling energy scale.

As in section 4.3, the nonsingular contributions to the cross section enter at fixed  $\mathcal{O}(\alpha_s^2)$  and can be simply included additively. Like in section 4.3, we absorb the entirety of the  $\Delta_{3-}$  singular distribution into the  $\Delta_{2+}$  nonsingular distribution. To try to capture some higher-order effects in  $\Delta_{3-}$ , though, we make use of the nonsingular decomposition in eq. (4.32), repeated for convenience:

$$\frac{d\sigma_{\text{ns}}}{d\Delta_{2+}} = \frac{d\sigma_{\text{sing}}}{d\Delta_{3-}} + \frac{d\sigma_{\text{res}}}{d\Delta_{2+}}. \quad (5.15)$$

We evaluate the singular  $\Delta_{3-}$  piece at the scale  $\mu = Q\sqrt{\Delta_{3-}}$  and the small  $\Delta_{2+}$  residual term at a fixed scale  $\mu = Q$ . While this approach misses out on genuine  $\mathcal{O}(\alpha_s^3)$  fixed-order corrections from 3-jet events, they have an endpoint at  $\Delta_{3-} = (z_{\text{cut}}/2)^2$  analogous to eq. (4.4), so we will simply not make a prediction for  $\Delta_{3-} < (z_{\text{cut}}/2)^2$ .

Finally, while the singular terms for  $\Delta_{2\pm}$  come with an overall prefactor of  $\sigma(\tilde{N}_{\text{jet}} = 2)$ , the nonsingular terms do not. However, since the nonsingular terms start at  $\mathcal{O}(\alpha_s^2)$ , we can multiply them by  $\sigma(\tilde{N}_{\text{jet}} = 2)/\sigma_0$ , which introduces corrections beyond the order to which we are working. This allows us to factor out a global  $\sigma(\tilde{N}_{\text{jet}} = 2)$  from our final predictions.

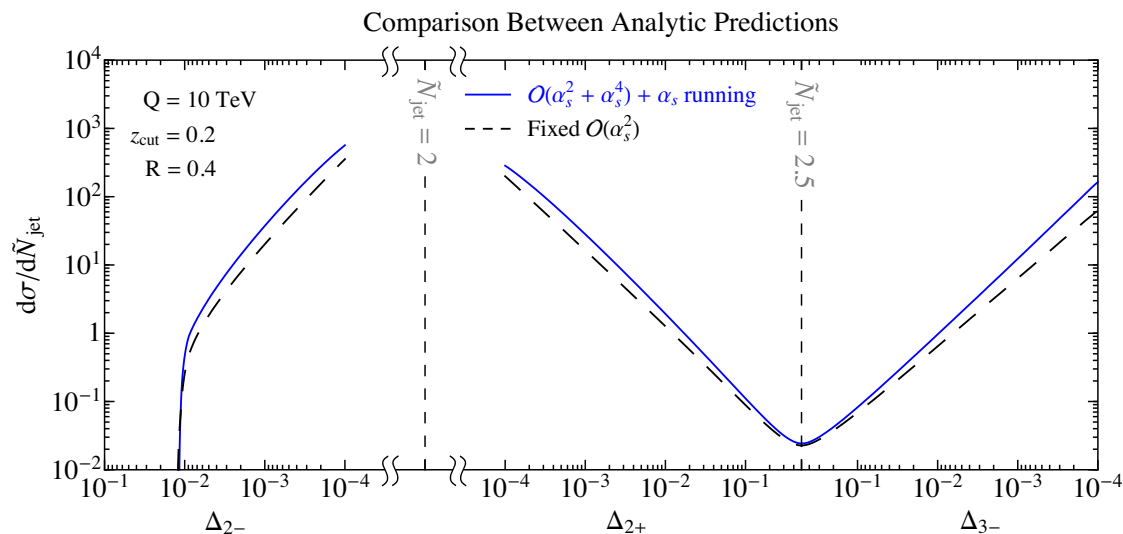
Putting these pieces together, our final analytic prediction for the  $\Delta_{2\pm}$  distribution is:

$$\begin{aligned} \frac{d\sigma}{d\Delta_{2\pm}} &= \sigma(\tilde{N}_{\text{jet}} = 2) \left\{ C_q(\Delta_{2\pm}) \otimes C_{\bar{q}}(\Delta_{2\pm}) + \frac{1}{\sigma_0} \frac{d\sigma_{\text{ns}}}{d\Delta_{2\pm}} \right\} \\ &= \sigma(\tilde{N}_{\text{jet}} = 2) \left\{ \delta(\Delta_{2\pm}) + 2 \left( \frac{\alpha_s(Q\sqrt{\Delta_{2\pm}})}{\pi} \right)^2 \left[ \mathcal{K}_+(\Delta_{2+}) + \mathcal{K}_-(\Delta_{2-}) \right] \right. \\ &\quad \left. + 2 \left( \frac{\alpha_s(Q\sqrt{\Delta_{2\pm}})}{\pi} \right)^4 \left[ \mathcal{K}_+(\Delta_{2+}) + \mathcal{K}_-(\Delta_{2-}) \right] \otimes \left[ \mathcal{K}_+(\Delta_{2+}) + \mathcal{K}_-(\Delta_{2-}) \right] \right. \\ &\quad \left. + \frac{1}{\sigma_0} \frac{d\sigma_{\text{ns}}}{d\Delta_{2\pm}} \right\}. \end{aligned} \quad (5.16)$$

In figure 9, we compare the pure fixed  $\mathcal{O}(\alpha_s^2)$  distributions for  $\Delta_{2\pm}$  and  $\Delta_{3-}$  from eq. (4.33) with the final prediction from eq. (5.16). At small values of  $\Delta_{n\pm}$ , the dominant differences come from the higher-order logarithms included in our final prediction, while at larger values of  $\Delta_{n\pm}$  the running effects from our  $\Delta_{n\pm}$ -dependent scale choice give most of the difference from the fixed-order prediction. Since we do not have a candidate factorization theorem for the  $\tilde{N}_{\text{jet}} = 3$  region of phase space, the  $\Delta_{3-}$  distribution is really just eq. (5.16) evaluated at  $\Delta_{3-} = 1 - \Delta_{2+}$ .

## 6 Looking towards the LHC

Given the interesting features of  $\tilde{N}_{\text{jet}}$  and the wide-ranging and robust measurements of jet substructure at the LHC [80–82], a natural consideration is the  $\tilde{N}_{\text{jet}}$  distribution in hadronic collisions. In this case, it is more convenient to use the original definition of  $\tilde{N}_{\text{jet}}$  from eq. (1.1) based on transverse momenta and rapidity-azimuth distances (instead of



**Figure 9.** A comparison between analytic predictions for the  $\tilde{N}_{\text{jet}}$  distribution. Shown in dashed black are the fixed  $\mathcal{O}(\alpha_s^2)$  calculations from eq. (4.33). Shown in solid blue are our final predictions from eq. (5.16), which include an estimate of higher-order  $\mathcal{O}(\alpha_s^4)$  terms as well as  $\Delta_{2\pm}$ -dependent running of  $\alpha_s$ . Note that the  $\Delta_{2-}$  and  $\Delta_{3-}$  axes run backwards.

energies and angles):

$$\tilde{N}_{\text{jet}}(p_{T\text{cut}}, R) = \sum_{i \in \text{Event}} \frac{p_{Ti}}{p_{Ti,R}} \Theta(p_{Ti,R} - p_{T\text{cut}}), \quad (6.1)$$

where  $p_{Ti,R} = \sum_j p_{Tj} \Theta(R - R_{ij})$  and  $R_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}$ .

At first glance, the calculation of  $\tilde{N}_{\text{jet}}$  at the LHC would seem to be much more difficult than at an  $e^+e^-$  collider. After all,  $\tilde{N}_{\text{jet}}$  depends sensitively on soft radiation, and soft QCD is notoriously complicated at a hadron collider. However,  $\tilde{N}_{\text{jet}}$  only depends on soft radiation in a region of size  $\lesssim 2R$  around energetic partons. To the extent that we can exploit color coherence at small  $R$ , we can make predictions for  $\tilde{N}_{\text{jet}}$  by simply knowing the collinear functions for quark- and gluon-initiated jets,  $C_q$  and  $C_g$ . In particular, the  $1 \rightarrow 3$  splitting function formalism that we used in section 4.1 to calculate the collinear function for quarks in  $e^+e^-$  may also be used to calculate the quark and gluon collinear functions for LHC processes.

Conveniently, when switching between the  $e^+e^-$  and hadronic definitions of  $\tilde{N}_{\text{jet}}$ , the collinear functions  $C_{q,g}$  only differ by terms of  $\mathcal{O}(R)$ . For small jet radii, one can therefore neglect those corrections and use the  $e^+e^-$  collinear functions calculated in this paper also for the hadronic case, though one has to be careful to match the  $z_{\text{cut}}$  value to the outgoing parton momentum. To see why this is the case, notice that

$$z_i = \frac{E_i}{E_J} \simeq \frac{p_{Ti}}{p_{TJ}} [1 + \mathcal{O}(R)], \quad z_{\text{cut}} = \frac{E_{\text{cut}}}{E_J} \simeq \frac{p_{T\text{cut}}}{p_{TJ}} [1 + \mathcal{O}(R)], \quad (6.2)$$

where  $p_{TJ}$  is the scalar sum of transverse momenta in a jet region. Thus, the energy integrals in  $C_{q,g}$  would differ at most by  $\mathcal{O}(R)$  terms. The angular phase space regions in

eq. (4.10) would be written as constraints on  $R_{ij}$  instead of  $\theta_{ij}$ , but notice that

$$R_{ij} = \theta_{ij}(\cosh y_i \cosh y_j)^{1/2} \simeq \theta_{ij} \cosh y_J [1 + \mathcal{O}(R)], \quad (6.3)$$

where  $y_J$  is some characteristic rapidity associated with the jet region, for example the rapidity of the summed jet region momenta. At leading power, the logarithmic scaling of the angular integrals can be used to remove any dependence on  $\cosh y_J$ , so that the differences in the angular integrals are again at most an  $\mathcal{O}(R)$  effect.

In order to extend the candidate dijet factorization formula in eq. (5.2) to dijet events from hadronic collisions, we have to sum over all relevant partonic channels:

$$\frac{d\sigma}{d\Delta_{2\pm}} = \sum_{k,\ell} \sigma_{h_1 h_2 \rightarrow k\ell}(\tilde{N}_{\text{jet}} = 2) C_k(\Delta_{2\pm}) \otimes C_\ell(\Delta_{2\pm}), \quad (6.4)$$

where  $k, \ell = q, \bar{q},$  or  $g$ , and  $h_{1,2}$  represents the colliding hadrons. In general,  $\sigma_{h_1 h_2 \rightarrow k\ell}$ ,  $C_k$ , and  $C_\ell$  depend on the outgoing parton kinematics, though we have suppressed that dependence in eq. (6.4) for readability. Note that the incoming beams only create additional high- $p_T$  jets through hard, perturbative emissions. We can write the total rate  $\sigma_{h_1 h_2 \rightarrow k\ell}(\tilde{N}_{\text{jet}} = 2)$  schematically as

$$\sigma_{h_1 h_2 \rightarrow k\ell}(\tilde{N}_{\text{jet}} = 2) = \sum_{a,b} \int_0^1 dx_a dx_b f_{a/h_1}(x_a) f_{b/h_2}(x_b) \sigma_{ab \rightarrow k\ell}(\tilde{N}_{\text{jet}} = 2), \quad (6.5)$$

where  $f_{a/h_1}$  and  $f_{b/h_2}$  are parton distribution functions for partons  $a$  and  $b$  carrying momentum fractions  $x_{a,b}$  of the initial hadrons  $h_{1,2}$ , and  $\sigma_{ab \rightarrow k\ell}(\tilde{N}_{\text{jet}} = 2)$  is the total rate at  $\tilde{N}_{\text{jet}} = 2$  for the partonic channel  $ab \rightarrow k\ell$ . As in the  $e^+e^-$  case, this cross section can have large logarithms of  $R$  and  $p_{T\text{cut}}$  that require resummation to obtain reliable predictions.

Compared to the  $e^+e^-$  case, the main new ingredient is the collinear function  $C_g(\Delta_{2\pm})$  for a gluon-initiated jet. For completeness, we have calculated the gluon collinear function using the same approach as section 4.1, and we present both  $C_q$  and  $C_g$  in appendix C. The structure in eq. (6.4) can be easily extended to handle the  $\Delta_{n\pm}$  cross section for  $n$ -jet processes by using  $n$  collinear functions multiplied by the  $\tilde{N}_{\text{jet}} = n$  cross section, appropriately summed over the various partonic channels.

At a hadron collider, there is also a new potential source of non-global logarithms from initial state partons. In general, non-global logarithms of  $\Delta_{n\pm}$  only appear in the collinear functions  $C_{q,g}$  for each final state jet, and are associated with correlated emissions from final state partons. For events with well-separated jets, the collinear functions are universal and independent of other jets in the event. However, there are also non-global logarithms of  $z_{\text{cut}}$  and  $R$  which appear both in the collinear functions and in the exact integer  $\tilde{N}_{\text{jet}}$  cross section (e.g.  $\sigma_{h_1 h_2 \rightarrow k\ell}(\tilde{N}_{\text{jet}} = 2)$ , see eq. (5.7)). For the exact integer cross section at a hadron collider, one also has to include correlated emissions from initial state partons, which may introduce super-leading logarithms [83–85] at sufficiently high  $\alpha_s$  order. Ideally, one would want to understand the resummation of non-global effects in the collinear functions and the exact integer cross section to achieve accurate predictions.



Because a detailed study for the LHC is beyond the scope of this paper, we will not present any results for hadronic collisions. An important effect to account for in future LHC predictions and comparisons with measurements is hadronization, which we have not considered here. Additionally, for hadronic collisions, effects from the underlying event and initial state radiation are not present in the  $e^+e^-$  case. To lowest order in  $R$ , we expect those effects to be captured by color coherence, and one could imagine using the techniques of ref. [86] to understand the  $R$  dependence. A similar concern is pileup contamination, though the closed-form nature of  $\tilde{N}_{\text{jet}}$  makes it well-suited to analytic studies using area subtraction [87–89]. One could further mitigate the impact of pileup by using a version of  $\tilde{N}_{\text{jet}}$  that includes jet trimming [90] in closed form [15], and we expect the collinear function for the trimmed  $\tilde{N}_{\text{jet}}$  version could be calculated using the same techniques used here.

## 7 Monte Carlo comparisons

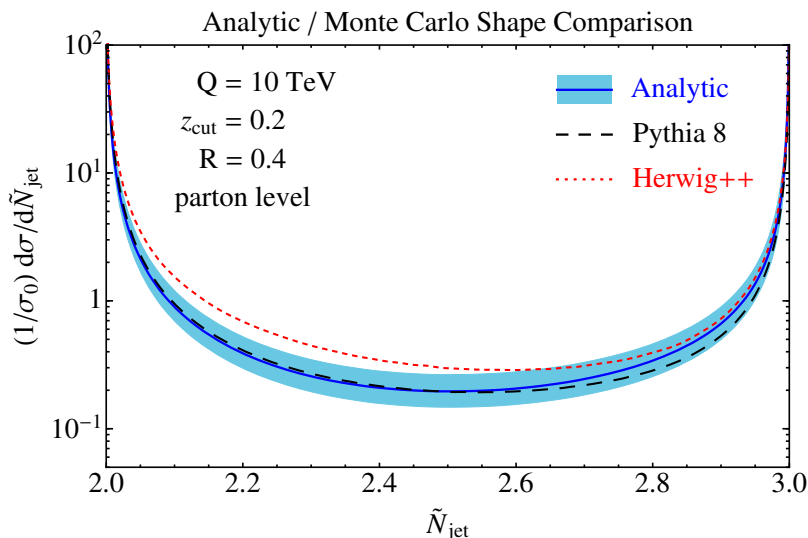
We now compare our analytic prediction in eq. (5.16) with parton shower Monte Carlo generators. As discussed in section 5.4, the rate  $\sigma(\tilde{N}_{\text{jet}} = 2)$  only enters as an overall normalization factor, and it in principle requires resummation of logarithms of  $R$  and  $z_{\text{cut}}$  to obtain a reliable prediction. Since we are mainly interested in describing only the near-integer (and not the exact-integer) behavior, we will divide out by an overall normalization factor and perform only a shape comparison. In all of the plots below, we normalize the cross section to the region  $\Delta_{2\pm} \in (10^{-4}, 10^{-2})$ .

To make an apples-to-apples comparison of our analytic results to PYTHIA 8 and HERWIG++, we turn off hadronization in the Monte Carlo generators. In principle, the collinear functions  $C_q$  and  $C_{\bar{q}}$  should get hadronization corrections, but the non-factorizing nature of  $\tilde{N}_{\text{jet}}$  means that we cannot adopt a standard shape function analysis [91, 92]. We run at a sufficiently high energy to allow for a comparison over a wide logarithmic range in  $\Delta_{2\pm}$ . The large collision energy we choose,  $Q = 10$  TeV, also mitigates the effect of low energy cutoffs ( $\simeq 1$  GeV) on the parton shower.

Because the singular terms in our calculation are numerically dominant for small  $\Delta_{2\pm}$ , higher-order logarithms may be important in determining the shape of the distribution in this regime. Some of the  $\mathcal{O}(\alpha_s^4)$  effects are captured by the convolution structure in eq. (5.16), but they are not fully reliable. Therefore, we include an uncertainty in our predictions derived from the addition of  $\mathcal{O}(\alpha_s^3 \mathcal{L}_2(\Delta_{2\pm}))$  and  $\mathcal{O}(\alpha_s^4 \mathcal{L}_3(\Delta_{2\pm}))$  terms with unknown coefficients that we vary. These terms represent the leading logarithms at each order, and are an estimate of missing higher-order terms. The form we use to determine the uncertainty is

$$\frac{d\sigma}{d\Delta_{2\pm}} \supset \left(\frac{\alpha_s}{\pi}\right)^3 \kappa_{2,\pm}^{(3)} \mathcal{L}_2(\Delta_{2\pm}) + \left(\frac{\alpha_s}{\pi}\right)^4 \kappa_{3,\pm}^{(4)} \mathcal{L}_3(\Delta_{2\pm}), \quad (7.1)$$

and we vary independently the coefficients in the ranges  $\kappa_{2,\pm}^{(3)} \in (-5, 5)$  and  $\kappa_{3,\pm}^{(4)} \in (-15, 15)$ . These coefficients are of similar size as the leading coefficients  $\kappa_{1,\pm}^{(2)}$  in our calculation, and the  $\mathcal{O}(\alpha_s^4)$  term generated by the naïve exponentiation in eq. (5.16) has

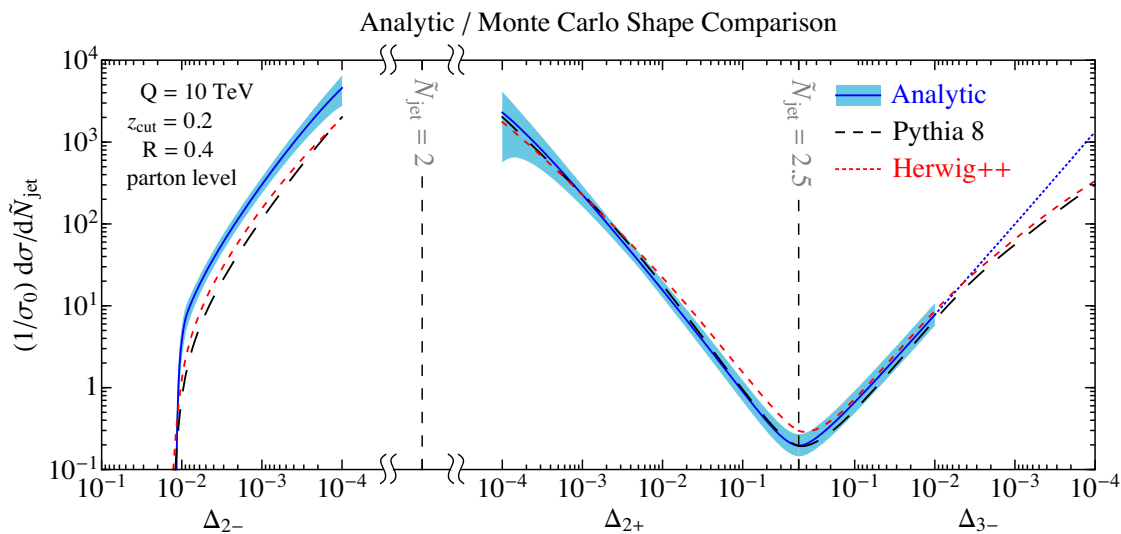


**Figure 10.** A shape comparison between our analytic predictions for the  $2 < \tilde{N}_{\text{jet}} < 3$  distribution with the Monte Carlo generators PYTHIA 8 and HERWIG++. Our calculation is summarized in eq. (5.16) with uncertainties given in eq. (7.1). To make an apples-to-apples comparison, we include showering but not hadronization in the Monte Carlo samples. All cross sections are normalized to have the same value in the  $\Delta_{2+} \in (10^{-4}, 10^{-2})$  window.

$\kappa_{3,\pm}^{(4)} \sim \mathcal{O}(10)$ . The uncertainty band is given by the envelope of these variations, including effects both on the cross section normalization (where we just add eq. (7.1)) and on the cross section shape (where we add eq. (7.1) and readjust the normalization in the  $\Delta_{2+} \in (10^{-4}, 10^{-2})$  window).

In figure 10, we show the shape comparison between our result and PYTHIA 8/HERWIG++ over the range  $2 < \tilde{N}_{\text{jet}} < 3$ . The singular cross section for  $\Delta_{3-}$  is included as part of the nonsingular correction to the  $\Delta_{2+}$  distribution (see eq. (5.15)). Overall, we find good agreement with the Monte Carlo generators within uncertainties. It is amusing that there is such close agreement with PYTHIA 8 in the  $\tilde{N}_{\text{jet}} \rightarrow 2$  region and with HERWIG++ in the  $\tilde{N}_{\text{jet}} \rightarrow 3$  region, with the analytic result effectively interpolating between the two. Given that the  $\Delta_{2+}$  and  $\Delta_{3-}$  parts of the distribution are dominated by different phase space configurations (see table 1), it is possible that we are seeing the impact of different shower ordering variables in PYTHIA 8 ( $p_{\perp}$ -ordered) versus HERWIG++ (angular-ordered). Of course, the theoretical uncertainties in our calculation are too large to make a definitive statement.

In figure 11, we compare distributions of  $\Delta_{2\pm}$  over the range  $(10^{-4}, 1)$ , using the triptych format to also see the  $\Delta_{3-}$  region. Again, our analytic prediction reproduces the shape of the PYTHIA 8 and HERWIG++ distributions remarkably well over the whole range. The Monte Carlo distributions, which include higher-order logarithms of  $\Delta_{2\pm}$  through multiple emissions, generally lie within the uncertainty band of our prediction for  $\tilde{N}_{\text{jet}} > 2$ . This indicates that we have made a reasonable estimate of the higher-order corrections not included in our calculation. Because we normalize to the window  $\Delta_{2+} \in (10^{-4}, 10^{-2})$ , it is



**Figure 11.** Same distributions as figure 10, but in triptych form to highlight the near integer behavior. Because we normalize only to the  $\Delta_{2+} \in (10^{-4}, 10^{-2})$  window, the normalization differences in the  $\Delta_{2-}$  region are accentuated; the shape agreement is much better. For the  $\Delta_{3-}$  distribution, we extend our central prediction (with a dotted blue line) to  $\Delta_{3-} < (z_{\text{cut}}/2)^2$ , where genuine 3-jet events, whose contribution we have not calculated, contribute to the observable. Note that the  $\Delta_{2-}$  and  $\Delta_{3-}$  axes run backwards.

not surprising that there is a normalization discrepancy in  $\Delta_{2-}$ , though one can see that the shape agreement for  $\tilde{N}_{\text{jet}} < 2$  is excellent. We suspect that the normalization issues for  $\tilde{N}_{\text{jet}} < 2$  may be due to the treatment of correlated soft radiation (and angular ordering) in the Monte Carlo generators, since the  $\Delta_{2-}$  distribution is dominated by double-soft emissions with a  $C_F C_A$  color structure. The overall shape agreement between PYTHIA 8 and HERWIG++ is quite surprising, given that they exhibit a factor of 3 difference (not shown) in their predicted value of  $\sigma(\tilde{N}_{\text{jet}} = 2)$ , suggesting that the shape of the non-integer distributions is more robust than the cross section at integer values.

## 8 Conclusions

In this paper, we studied the analytic properties of  $\tilde{N}_{\text{jet}}$ , a jets-without-jets event shape that can return a fractional value of jet multiplicity. Focusing on  $e^+e^- \rightarrow$  jets, we calculated the distribution of  $\tilde{N}_{\text{jet}}$  in the vicinity of a dijet configuration at  $\mathcal{O}(\alpha_s^2)$ . A fractional number of jets requires at least three collimated partons, such that for  $e^+e^- \rightarrow$  jets, the first non-trivial contribution requires at least two emissions. The singular parts of this emission structure can be captured using  $1 \rightarrow 3$  splitting functions, and we validated this approach (and included nonsingular contributions) using the fixed-order code EVENT2. To partially capture higher-order effects out to  $\mathcal{O}(\alpha_s^4)$ , we included convolutions from different phase space regions and running coupling effects. We found very good agreement between our calculation and Monte Carlo distributions from both PYTHIA 8 and HERWIG++.

Fractional jet multiplicity exhibits unique analytic features that are not shared by other jet observables. At  $\mathcal{O}(\alpha_s^2)$ , we showed how rapidity-like divergences, related to the energy sharing between emissions, appear when one or two partons become soft, and we explained how to regulate them. Beyond the fixed-order result, we proposed a candidate local factorization theorem and used it to predict a hybrid jet algorithm/event shape behavior for the  $\tilde{N}_{\text{jet}}$  distribution in the vicinity of an integer. At exact-integer values,  $\tilde{N}_{\text{jet}}$  behaves much like a standard jet algorithm, yielding spikes in the  $\tilde{N}_{\text{jet}}$  cross section. The near-integer behavior of  $\tilde{N}_{\text{jet}}$  is more characteristic of event shapes, where towers of higher-order logarithms give rise to (single-logarithmic) suppression of the singular phase space, yielding shoulders in the  $\tilde{N}_{\text{jet}}$  cross section. Finally, as opposed to typical event shapes, collinear emissions do not generate logarithms of  $\tilde{N}_{\text{jet}}$ , so the shape of the near-integer distribution is entirely determined by soft logarithms. These soft logarithms are purely non-global, as near-integer values force soft partons to lie in a restricted angular region of phase space and correlate different emissions (even if generated from an Abelian matrix element).

Beyond our present understanding of  $\tilde{N}_{\text{jet}}$ , there are three key directions to pursue. The first is to extend our calculations to  $\mathcal{O}(\alpha_s^3)$ . Though the differential cross section shapes were largely within our uncertainty estimates, the normalization differences seen between our analytic calculations and Monte Carlo generators suggest that higher-order terms might be relevant. While we were able to estimate some  $\mathcal{O}(\alpha_s^4)$  effects through convolutions, there are genuine  $\mathcal{O}(\alpha_s^3)$  effects that may get a phase space enhancement in the merged jet region (relative to the  $\mathcal{O}(\alpha_s^2)$  phase space) to partially overcome the  $\alpha_s$  suppression. The second is to attempt an understanding of logarithmic resummation. Because of the non-global nature of the observable, standard renormalization group methods will not work, but there may be a way to exploit the rapidity-like divergences seen in section 3 or the techniques introduced in refs. [93, 94]. The third is to perform a detailed study for hadron colliders like the LHC, as discussed in section 6, including the important effect of hadronization. Fractional jets offer a more nuanced understanding of jet formation than is possible with standard jet algorithms, and a measurement of  $\tilde{N}_{\text{jet}}$  seems feasible given the increasingly sophisticated approach to jet physics at the LHC.

## Acknowledgments

We thank Andrew Larkoski for helpful discussions. D.B. and J.R.W. are supported by the Office of Science, Office of High Energy Physics, of the U.S. Department of Energy (DOE) under Contract No. DE-AC02-05CH11231, and J.T. is supported under grant Contract No. DE-SC00012567. J.T. is also supported by the DOE Early Career research program DE-FG02-11ER-41741 and by a Sloan Research Fellowship from the Alfred P. Sloan Foundation. This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the DOE under Contract No. DE-AC02-05CH11231.

## A Results for non-Abelian contributions

Following the Abelian results in section 4.1, here we show the singular  $C_F C_A$  and  $C_F T_{Rn_f}$  contributions to the  $\Delta_{2\pm}$  and  $\Delta_{3-}$  cross sections. The  $C_F C_A$  terms are

$$\left[ \frac{d\sigma}{d\Delta_{2-}} \right]_{\text{nab}} = \left( \frac{\alpha_s}{\pi} \right)^2 C_F C_A \left\{ -\frac{14}{5} \mathcal{I}_\Omega \mathcal{L}_1(\Delta_{2-}) + \left[ \frac{28}{5} \mathcal{I}_\Omega \ln z_{\text{cut}} + \mathcal{I}_\Omega^{(a)} \right] \mathcal{L}_0(\Delta_{2-}) \right\} \Theta(\Delta_{2-} < z_{\text{cut}}^2/4), \quad (\text{A.1})$$

$$\left[ \frac{d\sigma}{d\Delta_{2+}} \right]_{\text{nab}} = \left( \frac{\alpha_s}{\pi} \right)^2 C_F C_A \left\{ \frac{2}{5} \mathcal{I}_\Omega \mathcal{L}_1(\Delta_{2+}) + \left[ \left( -\frac{3}{2} - \frac{2}{5} \ln 2 - \frac{14}{5} \ln z_{\text{cut}} \right) \mathcal{I}_\Omega + \mathcal{I}_\Omega^{(b)} \right] \mathcal{L}_0(\Delta_{2+}) \right\}, \quad (\text{A.2})$$

$$\left[ \frac{d\sigma}{d\Delta_{3-}} \right]_{\text{nab}} = \left( \frac{\alpha_s}{\pi} \right)^2 C_F C_A \mathcal{I}_\Omega \left( -\frac{21}{10} - \frac{14}{5} \ln z_{\text{cut}} \right) \mathcal{L}_0(\Delta_{3-}), \quad (\text{A.3})$$

where  $\mathcal{I}_\Omega$  is defined in eq. (4.29), and the remaining angular integrals can only be done numerically:

$$\mathcal{I}_\Omega^{(a)} = -2.44393, \quad \mathcal{I}_\Omega^{(b)} = -0.035397. \quad (\text{A.4})$$

The  $C_F T_{Rn_f}$  terms are

$$\left[ \frac{d\sigma}{d\Delta_{2-}} \right]_{C_F T_{Rn_f}} = \left( \frac{\alpha_s}{\pi} \right)^2 C_F T_{Rn_f} \mathcal{I}_\Omega^{(c)} \mathcal{L}_0(\Delta_{2-}) \Theta(\Delta_{2-} < z_{\text{cut}}^2/4), \quad (\text{A.5})$$

$$\left[ \frac{d\sigma}{d\Delta_{2+}} \right]_{C_F T_{Rn_f}} = \left( \frac{\alpha_s}{\pi} \right)^2 C_F T_{Rn_f} \mathcal{I}_\Omega^{(d)} \mathcal{L}_0(\Delta_{2+}), \quad (\text{A.6})$$

with the angular integrals

$$\mathcal{I}_\Omega^{(c)} = 0.724689, \quad \mathcal{I}_\Omega^{(d)} = 0.251525. \quad (\text{A.7})$$

The  $C_F T_{Rn_f}$  contribution to  $\Delta_{3-}$  is purely power suppressed.

## B Properties of distributions

In this appendix we collect useful formulae for the convolution of one-sided distributions, and then discuss two-sided distributions and their convolutions. The relations given here are straightforwardly derived from the results and techniques in ref. [95].

Throughout this paper, we use the standard plus distribution notation. For a function  $q$ ,

$$[q(x)]_+ = \lim_{\beta \rightarrow 0} \frac{d}{dx} [\Theta(x - \beta) Q(x)], \quad Q(x) = \int_1^x dx' q(x'). \quad (\text{B.1})$$

This distribution has a boundary at 1, so that

$$\int_{-\infty}^1 dx [q(x)]_+ = 0. \quad (\text{B.2})$$

We also use the shorthand

$$\mathcal{L}_n(x) = \left[ \frac{\ln^n x}{x} \Theta(x) \right]_+ \quad \text{for } n \text{ an integer } \geq 0, \quad \mathcal{L}_{-1}(x) = \delta(x), \quad (\text{B.3})$$

as well as the distribution

$$[\Theta(x) x^{-1+\beta}]_+^{[\infty]} = \frac{1}{\beta} \delta(x) + \sum_{n=0}^{\infty} \frac{1}{n!} \beta^n \mathcal{L}_n(x). \quad (\text{B.4})$$

The  $[\infty]$  notation on the plus distribution indicates that the boundary is at  $\infty$  (instead of 1), so that

$$\int_{-\infty}^{\infty} [\Theta(x) x^{-1+\beta}]_+^{[\infty]} = 0. \quad (\text{B.5})$$

### B.1 One-sided distributions and their convolutions

For many applications, one often makes use of convolutions between plus distributions, especially  $\mathcal{L}_n$ . The convolution is defined as

$$(f \otimes g)(x) \equiv \int_{-\infty}^{\infty} dx' dx'' \delta(x - x' - x'') f(x') g(x''). \quad (\text{B.6})$$

We can take the Fourier transform  $\mathcal{F}$  to make the convolution multiplicative, and also use the Fourier transform and its inverse to determine a convolution:

$$\mathcal{F}\{f \otimes g\} = \mathcal{F}\{f\} \mathcal{F}\{g\} \quad \Rightarrow \quad f \otimes g = \mathcal{F}^{-1}\{\mathcal{F}\{f\} \mathcal{F}\{g\}\}. \quad (\text{B.7})$$

It is straightforward to determine the convolution between two plus functions  $\mathcal{L}_k$  and  $\mathcal{L}_n$  by relating it to eq. (B.4):

$$(\mathcal{L}_k \otimes \mathcal{L}_n)(x) = \text{the } \mathcal{O}\left(\frac{\alpha^k \beta^n}{k! n!}\right) \text{ coefficient of } \left([\Theta(x) x^{-1+\alpha}]_+^{[\infty]}\right) \otimes \left([\Theta(x) x^{-1+\beta}]_+^{[\infty]}\right). \quad (\text{B.8})$$

Using the Fourier transform

$$\mathcal{F}\{[\Theta(x) x^{-1+\beta}]_+^{[\infty]}\} = \Gamma[\beta] (-is)^{-\beta}, \quad (\text{B.9})$$

with  $s$  as the conjugate variable, it is straightforward to show

$$\begin{aligned} (\mathcal{L}_k \otimes \mathcal{L}_n)(x) &= \text{the } \mathcal{O}\left(\frac{\alpha^k \beta^n}{k! n!}\right) \text{ coefficient of } \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} [\Theta(x) x^{-1+\alpha+\beta}]_+^{[\infty]} \quad (\text{B.10}) \\ &= \text{the } \mathcal{O}\left(\frac{\alpha^k \beta^n}{k! n!}\right) \text{ coefficient of } \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \left[ \frac{1}{\alpha+\beta} \delta(x) + \sum_{m=0}^{\infty} \frac{1}{m!} (\alpha+\beta)^m \mathcal{L}_m(x) \right]. \end{aligned}$$

This approach is similar to the one used in ref. [95] to give the general form of the convolution  $(\mathcal{L}_k \otimes \mathcal{L}_n)(x)$ , where an equivalent result to eq. (B.10) is given.

We give the first few convolutions here for convenience:

$$\begin{aligned}
 \mathcal{L}_0(x) \otimes \mathcal{L}_0(x) &= 2\mathcal{L}_1(x) - \frac{\pi^2}{6}\delta(x), \\
 \mathcal{L}_0(x) \otimes \mathcal{L}_1(x) &= \frac{3}{2}\mathcal{L}_2(x) - \frac{\pi^2}{6}\mathcal{L}_0(x) + \zeta_3\delta(x), \\
 \mathcal{L}_1(x) \otimes \mathcal{L}_1(x) &= \mathcal{L}_3(x) - \frac{\pi^2}{3}\mathcal{L}_1(x) + 2\zeta_3\mathcal{L}_0(x) - \frac{\pi^4}{360}\delta(x).
 \end{aligned}
 \tag{B.11}$$

These distributions only have support for  $x > 0$ , and appear in many applications.

## B.2 Two-sided distributions and their convolutions

In this work, we encountered observables whose cross sections have support for all  $x$  and behave like distributions as  $x \rightarrow 0^\pm$ . We will refer to them as *two-sided distributions* (one may think of the usual distributions as one-sided), and we now discuss convolutions of them.

Consider an observable  $x$  with singular behavior for  $x \rightarrow 0^\pm$ . The fixed-order singular behavior of  $x$  is described by distributions  $\mathcal{L}_n(x_\pm)$ , where  $x_+ = x$  for  $x \geq 0$  and  $x_- = -x$  for  $x \leq 0$ , so that

$$\begin{aligned}
 \mathcal{L}_n(x_+) &= \mathcal{L}_n(x) = \left[ \frac{\ln^n x}{x} \Theta(x) \right]_+, \\
 \mathcal{L}_n(x_-) &= \mathcal{L}_n(-x) = \left[ \frac{\ln^n(-x)}{-x} \Theta(-x) \right]_+.
 \end{aligned}
 \tag{B.12}$$

Convolutions between various  $\mathcal{L}_n(x_\pm)$  will mix the  $x_\pm$  distributions.

First, we note that convolutions between two  $x_+$  distributions or two  $x_-$  distributions are effectively one-sided, meaning the above results can be used without modification. Convolutions between an  $x_+$  and an  $x_-$  distribution are the novel ones we derive here, adapting the technique in appendix B.1 to find the general form of two-sided convolutions.

Using the Fourier transforms of the  $x_\pm$  distributions

$$\mathcal{F}\{[\Theta(x_\pm) x_\pm^{-1+\beta}]_+^{[\infty]}\} = \Gamma(\beta)(\mp is)^{-\beta},
 \tag{B.13}$$

we have

$$\begin{aligned}
 \mathcal{F}\left\{\left([\Theta(x_+) x_+^{-1+\alpha}]_+^{[\infty]}\right) \otimes \left([\Theta(x_-) x_-^{-1+\beta}]_+^{[\infty]}\right)\right\} &= \Gamma(\alpha)\Gamma(\beta)(-is)^{-\alpha}(is)^{-\beta} \\
 &= \frac{\Gamma(b)\Gamma(1-(a+b))}{\Gamma(1-a)} \Gamma(a+b)(-is)^{-(\alpha+\beta)} + \frac{\Gamma(a)\Gamma(1-(a+b))}{\Gamma(1-b)} \Gamma(a+b)(is)^{-(\alpha+\beta)}.
 \end{aligned}
 \tag{B.14}$$

We can easily Fourier invert the right-hand side given eq. (B.13). This implies

$$\begin{aligned}
 & (\mathcal{L}_k(x_+) \otimes \mathcal{L}_n(x_-))(x_{\pm}) \\
 &= \text{the } \mathcal{O}\left(\frac{\alpha^k}{k!} \frac{\beta^n}{n!}\right) \text{ coefficient of} \tag{B.15} \\
 & \quad \frac{\Gamma(b)\Gamma(1-(a+b))}{\Gamma(1-a)} [\Theta(x_+) x_+^{-1+\alpha+\beta}]_+^{[\infty]} + \frac{\Gamma(a)\Gamma(1-(a+b))}{\Gamma(1-b)} [\Theta(x_-) x_-^{-1+\alpha+\beta}]_+^{[\infty]}
 \end{aligned}$$

$$\begin{aligned}
 &= \text{the } \mathcal{O}\left(\frac{\alpha^k}{k!} \frac{\beta^n}{n!}\right) \text{ coefficient of} \\
 & \quad \frac{\Gamma(b)\Gamma(1-(a+b))}{\Gamma(1-a)} \left[ \frac{1}{\alpha+\beta} \delta(x_+) + \sum_{m=0}^{\infty} \frac{1}{m!} (\alpha+\beta)^m \mathcal{L}_m(x_+) \right] \\
 & \quad + \frac{\Gamma(a)\Gamma(1-(a+b))}{\Gamma(1-b)} \left[ \frac{1}{\alpha+\beta} \delta(x_-) + \sum_{m=0}^{\infty} \frac{1}{m!} (\alpha+\beta)^m \mathcal{L}_m(x_-) \right]. \tag{B.16}
 \end{aligned}$$

Using this relation, the first few nontrivial convolutions for two-sided distributions are

$$\begin{aligned}
 \mathcal{L}_0(x_+) \otimes \mathcal{L}_0(x_-) &= \mathcal{L}_1(x_+) + \mathcal{L}_1(x_-) + \frac{\pi^2}{3} \delta(x), \\
 \mathcal{L}_1(x_+) \otimes \mathcal{L}_0(x_-) &= \mathcal{L}_2(x_+) + \frac{\pi^2}{6} \mathcal{L}_0(x_+) \\
 & \quad + \frac{1}{2} \mathcal{L}_2(x_-) + \frac{\pi^2}{6} \mathcal{L}_0(x_-) + \zeta_3 \delta(x), \\
 \mathcal{L}_0(x_+) \otimes \mathcal{L}_1(x_-) &= \frac{1}{2} \mathcal{L}_2(x_+) + \frac{\pi^2}{6} \mathcal{L}_0(x_+) \\
 & \quad + \mathcal{L}_2(x_-) + \frac{\pi^2}{6} \mathcal{L}_0(x_-) + \zeta_3 \delta(x), \\
 \mathcal{L}_1(x_+) \otimes \mathcal{L}_1(x_-) &= \frac{1}{2} \mathcal{L}_3(x_+) + \frac{\pi^2}{3} \mathcal{L}_1(x_+) + \zeta_3 \mathcal{L}_0(x_+) \\
 & \quad + \frac{1}{2} \mathcal{L}_3(x_-) + \frac{\pi^2}{3} \mathcal{L}_1(x_-) + \zeta_3 \mathcal{L}_0(x_-) + \frac{7\pi^4}{180} \delta(x). \tag{B.17}
 \end{aligned}$$

## C Quark and gluon collinear functions

We summarize here results for the collinear functions for quark- and gluon-initiated jets through  $\mathcal{O}(\alpha_s^2)$ . We express our results for dijet observables ( $\Delta_{2\pm}$  and  $\Delta_{3-}$ ), but they can be applied to the more general  $n$ -jet case of  $\Delta_{n\pm}$  as well. The quark collinear function can be derived from the calculations in section 4.1 and appendix A. The gluon collinear function calculation proceeds in analogy with the quark case, and there is a strong relation between the quark and gluon results for nearly all coefficients.



To  $\mathcal{O}(\alpha_s^2)$ , the quark and gluon collinear functions contribute to  $\Delta_{2\pm}$  and  $\Delta_{3-}$ , with

$$\begin{aligned}
 C_i(\Delta_{2-}) &= \delta(\Delta_{2-}) + \left(\frac{\alpha_s}{\pi}\right)^2 \mathcal{K}_{2-}^{(i)}(\Delta_{2-}), \\
 C_i(\Delta_{2+}) &= \delta(\Delta_{2+}) + \left(\frac{\alpha_s}{\pi}\right)^2 \mathcal{K}_{2+}^{(i)}(\Delta_{2+}), \\
 C_i(\Delta_{3-}) &= \left(\frac{\alpha_s}{\pi}\right)^2 \mathcal{K}_{3-}^{(i)}(\Delta_{3-}),
 \end{aligned} \tag{C.1}$$

where  $i = q, g$  for the different parton types, and the subscripts on  $\mathcal{K}^{(i)}$  indicate the contribution of an individual jet region to  $\tilde{N}_{\text{jet}}$ . The  $\mathcal{O}(\alpha_s^2)$  terms are given by

$$\begin{aligned}
 \mathcal{K}_{2-}^{(q)}(\Delta_{2-}) &= \left\{ \left( -C_F^2 - \frac{7}{5} C_F C_A \right) \mathcal{I}_\Omega \mathcal{L}_1(\Delta_{2-}) + \left( 2 \ln z_{\text{cut}} C_F^2 + \frac{14}{5} \ln z_{\text{cut}} C_F C_A \right) \mathcal{I}_\Omega \mathcal{L}_0(\Delta_{2-}) \right. \\
 &\quad \left. + \left( \frac{1}{2} \mathcal{I}_\Omega^{(a)} C_F C_A + \frac{1}{2} \mathcal{I}_\Omega^{(c)} C_F T_R n_f \right) \mathcal{L}_0(\Delta_{2-}) \right\} \Theta(\Delta_{2-} < z_{\text{cut}}^2/4), \\
 \mathcal{K}_{2-}^{(g)}(\Delta_{2-}) &= \left\{ \left( -\frac{12}{5} C_A^2 \right) \mathcal{I}_\Omega \mathcal{L}_1(\Delta_{2-}) + \left( \frac{24}{5} \ln z_{\text{cut}} C_A^2 \right) \mathcal{I}_\Omega \mathcal{L}_0(\Delta_{2-}) \right. \\
 &\quad \left. + \left( \frac{1}{2} \mathcal{I}_\Omega^{(a)} C_A^2 + \frac{1}{2} \mathcal{I}_\Omega^{(c)} C_A T_R n_f \right) \mathcal{L}_0(\Delta_{2-}) \right\} \Theta(\Delta_{2-} < z_{\text{cut}}^2/4),
 \end{aligned} \tag{C.2}$$

and

$$\begin{aligned}
 \mathcal{K}_{2+}^{(q)}(\Delta_{2+}) &= \left( -\frac{7}{5} C_F^2 + \frac{1}{5} C_F C_A \right) \mathcal{I}_\Omega \mathcal{L}_1(\Delta_{2+}) \\
 &\quad + \left[ \left( -\frac{57}{20} + \frac{7}{5} \ln 2 - \ln z_{\text{cut}} \right) C_F^2 + \left( -\frac{3}{4} - \frac{1}{5} \ln 2 - \frac{7}{5} \ln z_{\text{cut}} \right) C_F C_A \right] \mathcal{I}_\Omega \mathcal{L}_0(\Delta_{2+}) \\
 &\quad + \left( \frac{1}{2} \mathcal{I}_\Omega^{(b)} C_F C_A + \frac{1}{2} \mathcal{I}_\Omega^{(d)} C_F T_R n_f \right) \mathcal{L}_0(\Delta_{2+}), \\
 \mathcal{K}_{2+}^{(g)}(\Delta_{2+}) &= \left( -\frac{6}{5} C_A^2 \right) \mathcal{I}_\Omega \mathcal{L}_1(\Delta_{2+}) \\
 &\quad + \left[ \frac{6}{5} \left( -\frac{11}{3} + \ln 2 - 2 \ln z_{\text{cut}} \right) C_A^2 + \frac{2}{3} C_F T_R n_f + \frac{2}{3} \cdot \frac{7}{5} C_A T_R n_f \right] \mathcal{I}_\Omega \mathcal{L}_0(\Delta_{2+}) \\
 &\quad + \left( \frac{1}{2} \mathcal{I}_\Omega^{(b)} C_A^2 + \frac{1}{2} \mathcal{I}_\Omega^{(d)} C_A T_R n_f \right) \mathcal{L}_0(\Delta_{2+}),
 \end{aligned} \tag{C.3}$$

and

$$\begin{aligned}
 \mathcal{K}_{3-}^{(q)}(\Delta_{3-}) &= \left[ \left( -\frac{3}{4} - \ln z_{\text{cut}} \right) C_F^2 + \left( -\frac{3}{4} - \ln z_{\text{cut}} \right) \frac{7}{5} C_F C_A \right] \mathcal{I}_\Omega \mathcal{L}_0(\Delta_{3-}), \\
 \mathcal{K}_{3-}^{(g)}(\Delta_{3-}) &= \left[ \frac{6}{5} \left( -\frac{11}{6} - 2 \ln z_{\text{cut}} \right) C_A^2 + \frac{2}{3} \cdot \frac{7}{5} C_F T_R n_f - \frac{2}{3} \cdot \frac{1}{5} C_A T_R n_f \right] \mathcal{I}_\Omega \mathcal{L}_0(\Delta_{3-}).
 \end{aligned} \tag{C.4}$$

The angular integral factor  $\mathcal{I}_\Omega$  is defined in eq. (4.29) and  $\mathcal{I}_\Omega^{(a,b,c,d)}$  are given in eqs. (A.4) and (A.7).

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] J.D. Bjorken and S.J. Brodsky, *Statistical Model for electron-Positron Annihilation Into Hadrons*, *Phys. Rev. D* **1** (1970) 1416 [[INSPIRE](#)].
- [2] G. Hanson et al., *Evidence for Jet Structure in Hadron Production by  $e^+e^-$  Annihilation*, *Phys. Rev. Lett.* **35** (1975) 1609 [[INSPIRE](#)].
- [3] R.D. Field and R.P. Feynman, *A Parametrization of the Properties of Quark Jets*, *Nucl. Phys. B* **136** (1978) 1 [[INSPIRE](#)].
- [4] G.F. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, *Phys. Rev. Lett.* **39** (1977) 1436 [[INSPIRE](#)].
- [5] S.D. Ellis, J. Huston, K. Hatakeyama, P. Loch and M. Tonnesmann, *Jets in hadron-hadron collisions*, *Prog. Part. Nucl. Phys.* **60** (2008) 484 [[arXiv:0712.2447](#)] [[INSPIRE](#)].
- [6] G.P. Salam, *Towards Jetography*, *Eur. Phys. J. C* **67** (2010) 637 [[arXiv:0906.1833](#)] [[INSPIRE](#)].
- [7] M. Dasgupta and G.P. Salam, *Event shapes in  $e^+e^-$  annihilation and deep inelastic scattering*, *J. Phys. G* **30** (2004) R143 [[hep-ph/0312283](#)] [[INSPIRE](#)].
- [8] J.R. Ellis, M.K. Gaillard and G.G. Ross, *Search for Gluons in  $e^+e^-$  Annihilation*, *Nucl. Phys. B* **111** (1976) 253 [[INSPIRE](#)].
- [9] B. Wiik, *First Results from PETRA*, *Conf. Proc. C* **7906181** (1979) 113 [[INSPIRE](#)].
- [10] TASSO collaboration, R. Brandelik et al., *Evidence for Planar Events in  $e^+e^-$  Annihilation at High-Energies*, *Phys. Lett. B* **86** (1979) 243 [[INSPIRE](#)].
- [11] MARK-J collaboration, D.P. Barber et al., *Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies*, *Phys. Rev. Lett.* **43** (1979) 830 [[INSPIRE](#)].
- [12] PLUTO collaboration, C. Berger et al., *Evidence for Gluon Bremsstrahlung in  $e^+e^-$  Annihilations at High-Energies*, *Phys. Lett. B* **86** (1979) 418 [[INSPIRE](#)].
- [13] JADE collaboration, W. Bartel et al., *Observation of Planar Three Jet Events in  $e^+e^-$  Annihilation and Evidence for Gluon Bremsstrahlung*, *Phys. Lett. B* **91** (1980) 142 [[INSPIRE](#)].
- [14] E. Farhi, *Quantum chromodynamics test for jets*, *Phys. Rev. Lett.* **39** (1977) 1587.
- [15] D. Bertolini, T. Chan and J. Thaler, *Jet Observables Without Jet Algorithms*, *JHEP* **04** (2014) 013 [[arXiv:1310.7584](#)] [[INSPIRE](#)].
- [16] C.W. Bauer, S. Fleming and M.E. Luke, *Summing Sudakov logarithms in  $B \rightarrow X_s \gamma$  in effective field theory*, *Phys. Rev. D* **63** (2000) 014006 [[hep-ph/0005275](#)] [[INSPIRE](#)].
- [17] C.W. Bauer, S. Fleming, D. Pirjol and I.W. Stewart, *An Effective field theory for collinear and soft gluons: Heavy to light decays*, *Phys. Rev. D* **63** (2001) 114020 [[hep-ph/0011336](#)] [[INSPIRE](#)].
- [18] C.W. Bauer and I.W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett. B* **516** (2001) 134 [[hep-ph/0107001](#)] [[INSPIRE](#)].
- [19] C.W. Bauer, D. Pirjol and I.W. Stewart, *Soft collinear factorization in effective field theory*, *Phys. Rev. D* **65** (2002) 054022 [[hep-ph/0109045](#)] [[INSPIRE](#)].

- [20] C.W. Bauer, S. Fleming, D. Pirjol, I.Z. Rothstein and I.W. Stewart, *Hard scattering factorization from effective field theory*, *Phys. Rev. D* **66** (2002) 014017 [[hep-ph/0202088](#)] [[INSPIRE](#)].
- [21] J.-y. Chiu, A. Jain, D. Neill and I.Z. Rothstein, *The Rapidity Renormalization Group*, *Phys. Rev. Lett.* **108** (2012) 151601 [[arXiv:1104.0881](#)] [[INSPIRE](#)].
- [22] J.-Y. Chiu, A. Jain, D. Neill and I.Z. Rothstein, *A Formalism for the Systematic Treatment of Rapidity Logarithms in Quantum Field Theory*, *JHEP* **05** (2012) 084 [[arXiv:1202.0814](#)] [[INSPIRE](#)].
- [23] J.M. Butterworth, A.R. Davison, M. Rubin and G.P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001 [[arXiv:0802.2470](#)] [[INSPIRE](#)].
- [24] M. Dasgupta, A. Fregoso, S. Marzani and G.P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029 [[arXiv:1307.0007](#)] [[INSPIRE](#)].
- [25] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *JHEP* **05** (2014) 146 [[arXiv:1402.2657](#)] [[INSPIRE](#)].
- [26] M. Dasgupta and G.P. Salam, *Resummation of nonglobal QCD observables*, *Phys. Lett. B* **512** (2001) 323 [[hep-ph/0104277](#)] [[INSPIRE](#)].
- [27] M. Dasgupta and G.P. Salam, *Accounting for coherence in interjet  $E(t)$  flow: a case study*, *JHEP* **03** (2002) 017 [[hep-ph/0203009](#)] [[INSPIRE](#)].
- [28] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [[arXiv:0710.3820](#)] [[INSPIRE](#)].
- [29] M. Bahr et al., *HERWIG++ Physics and Manual*, *Eur. Phys. J. C* **58** (2008) 639 [[arXiv:0803.0883](#)] [[INSPIRE](#)].
- [30] S.D. Ellis, R. Kleiss and W.J. Stirling, *W's, Z's and Jets*, *Phys. Lett. B* **154** (1985) 435 [[INSPIRE](#)].
- [31] F.A. Berends, W.T. Giele, H. Kuijf, R. Kleiss and W.J. Stirling, *Multi-jet production in W, Z events at  $p\bar{p}$  colliders*, *Phys. Lett. B* **224** (1989) 237 [[INSPIRE](#)].
- [32] F.A. Berends, H. Kuijf, B. Tausk and W.T. Giele, *On the production of a W and jets at hadron colliders*, *Nucl. Phys. B* **357** (1991) 32 [[INSPIRE](#)].
- [33] CDF collaboration, *Measurement of inclusive jet cross-sections in  $Z/\gamma^* \rightarrow e^+e^- + jets$  production in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Rev. Lett.* **100** (2008) 102001 [[arXiv:0711.3717](#)] [[INSPIRE](#)].
- [34] D0 collaboration, V.M. Abazov et al., *Measurement of  $Z/\gamma^*+jet+X$  angular distributions in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV*, *Phys. Lett. B* **682** (2010) 370 [[arXiv:0907.4286](#)] [[INSPIRE](#)].
- [35] ATLAS collaboration, *Measurement of the production cross section of jets in association with a Z boson in pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *JHEP* **07** (2013) 032 [[arXiv:1304.7098](#)] [[INSPIRE](#)].
- [36] CMS collaboration, *Differential cross section measurements for the production of a W boson in association with jets in proton-proton collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Lett. B* **741** (2015) 12 [[arXiv:1406.7533](#)] [[INSPIRE](#)].
- [37] CMS collaboration, *Measurements of jet multiplicity and differential production cross sections of  $Z + jets$  events in proton-proton collisions at  $\sqrt{s} = 7$  TeV*, *Phys. Rev. D* **91** (2015) 052008 [[arXiv:1408.3104](#)] [[INSPIRE](#)].

- [38] ATLAS collaboration, *Measurements of the  $W$  production cross sections in association with jets with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 82 [[arXiv:1409.8639](#)] [[INSPIRE](#)].
- [39] M. Bengtsson and T. Sjöstrand, *Coherent Parton Showers Versus Matrix Elements: Implications of PETRA — PEP Data*, *Phys. Lett. B* **185** (1987) 435 [[INSPIRE](#)].
- [40] S. Catani, F. Krauss, R. Kuhn and B.R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063 [[hep-ph/0109231](#)] [[INSPIRE](#)].
- [41] L. Lönnblad, *Correcting the color dipole cascade model with fixed order matrix elements*, *JHEP* **05** (2002) 046 [[hep-ph/0112284](#)] [[INSPIRE](#)].
- [42] S. Mrenna and P. Richardson, *Matching matrix elements and parton showers with HERWIG and PYTHIA*, *JHEP* **05** (2004) 040 [[hep-ph/0312274](#)] [[INSPIRE](#)].
- [43] S. Hoeche et al., *Matching parton showers and matrix elements*, [hep-ph/0602031](#) [[INSPIRE](#)].
- [44] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J. C* **53** (2008) 473 [[arXiv:0706.2569](#)] [[INSPIRE](#)].
- [45] W.T. Giele, D.A. Kosower and P.Z. Skands, *A simple shower and matching algorithm*, *Phys. Rev. D* **78** (2008) 014026 [[arXiv:0707.3652](#)] [[INSPIRE](#)].
- [46] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [[hep-ph/0204244](#)] [[INSPIRE](#)].
- [47] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [[hep-ph/0409146](#)] [[INSPIRE](#)].
- [48] N. Lavesson and L. Lönnblad, *Extending CKKW-merging to One-Loop Matrix Elements*, *JHEP* **12** (2008) 070 [[arXiv:0811.2912](#)] [[INSPIRE](#)].
- [49] K. Hamilton and P. Nason, *Improving NLO-parton shower matched simulations with higher order matrix elements*, *JHEP* **06** (2010) 039 [[arXiv:1004.1764](#)] [[INSPIRE](#)].
- [50] S. Alioli et al., *Matching Fully Differential NNLO Calculations and Parton Showers*, *JHEP* **06** (2014) 089 [[arXiv:1311.0286](#)] [[INSPIRE](#)].
- [51] A. Banfi, G. Marchesini and G. Smye, *Away from jet energy flow*, *JHEP* **08** (2002) 006 [[hep-ph/0206076](#)] [[INSPIRE](#)].
- [52] R.B. Appleby and G.P. Salam, *Theory and phenomenology of nonglobal logarithms*, [hep-ph/0305232](#) [[INSPIRE](#)].
- [53] M. Dasgupta, L. Magnea and G.P. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, *JHEP* **02** (2008) 055 [[arXiv:0712.3014](#)] [[INSPIRE](#)].
- [54] M. Dasgupta, F. Dreyer, G.P. Salam and G. Soyez, *Small-radius jets to all orders in QCD*, *JHEP* **04** (2015) 039 [[arXiv:1411.5182](#)] [[INSPIRE](#)].
- [55] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J. C* **72** (2012) 1896 [[arXiv:1111.6097](#)] [[INSPIRE](#)].
- [56] J.-y. Chiu, A. Fuhrer, A.H. Hoang, R. Kelley and A.V. Manohar, *Soft-Collinear Factorization and Zero-Bin Subtractions*, *Phys. Rev. D* **79** (2009) 053007 [[arXiv:0901.1332](#)] [[INSPIRE](#)].
- [57] A. Banfi and M. Dasgupta, *Problems in resumming interjet energy flows with  $k_t$  clustering*, *Phys. Lett. B* **628** (2005) 49 [[hep-ph/0508159](#)] [[INSPIRE](#)].
- [58] Y. Delenda, R. Appleby, M. Dasgupta and A. Banfi, *On QCD resummation with  $k(t)$  clustering*, *JHEP* **12** (2006) 044 [[hep-ph/0610242](#)] [[INSPIRE](#)].

- [59] K. Khelifa-Kerfa, *Non-global logs and clustering impact on jet mass with a jet veto distribution*, *JHEP* **02** (2012) 072 [[arXiv:1111.2016](#)] [[INSPIRE](#)].
- [60] R. Kelley, J.R. Walsh and S. Zuberi, *Abelian non-global logarithms from soft gluon clustering*, *JHEP* **09** (2012) 117 [[arXiv:1202.2361](#)] [[INSPIRE](#)].
- [61] C.W. Bauer, A. Hornig and F.J. Tackmann, *Factorization for generic jet production*, *Phys. Rev.* **D 79** (2009) 114013 [[arXiv:0808.2191](#)] [[INSPIRE](#)].
- [62] S.D. Ellis, C.K. Vermilion, J.R. Walsh, A. Hornig and C. Lee, *Jet shapes and jet algorithms in SCET*, *JHEP* **11** (2010) 101 [[arXiv:1001.0014](#)] [[INSPIRE](#)].
- [63] J.R. Walsh and S. Zuberi, *Factorization Constraints on Jet Substructure*, [arXiv:1110.5333](#) [[INSPIRE](#)].
- [64] R.K. Ellis, W.J. Stirling and B. Webber, *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Vol. 8: QCD and collider physics*, Cambridge University Press, Cambridge U.K. (1996).
- [65] A.V. Manohar and I.W. Stewart, *The Zero-Bin and Mode Factorization in Quantum Field Theory*, *Phys. Rev.* **D 76** (2007) 074002 [[hep-ph/0605001](#)] [[INSPIRE](#)].
- [66] S. Catani and M.H. Seymour, *The Dipole formalism for the calculation of QCD jet cross-sections at next-to-leading order*, *Phys. Lett.* **B 378** (1996) 287 [[hep-ph/9602277](#)] [[INSPIRE](#)].
- [67] S. Catani and M.H. Seymour, *A General algorithm for calculating jet cross-sections in NLO QCD*, *Nucl. Phys.* **B 485** (1997) 291 [[hep-ph/9605323](#)] [[INSPIRE](#)].
- [68] R.K. Ellis, H. Georgi, M. Machacek, H.D. Politzer and G.G. Ross, *Factorization and the Parton Model in QCD*, *Phys. Lett.* **B 78** (1978) 281 [[INSPIRE](#)].
- [69] R.K. Ellis, H. Georgi, M. Machacek, H.D. Politzer and G.G. Ross, *Perturbation Theory and the Parton Model in QCD*, *Nucl. Phys.* **B 152** (1979) 285 [[INSPIRE](#)].
- [70] J.M. Campbell and E.W.N. Glover, *Double unresolved approximations to multiparton scattering amplitudes*, *Nucl. Phys.* **B 527** (1998) 264 [[hep-ph/9710255](#)] [[INSPIRE](#)].
- [71] S. Alioli and J.R. Walsh, *Jet Veto Clustering Logarithms Beyond Leading Order*, *JHEP* **03** (2014) 119 [[arXiv:1311.5234](#)] [[INSPIRE](#)].
- [72] M. Field, G. Gur-Ari, D.A. Kosower, L. Mannelli and G. Perez, *Three-Prong Distribution of Massive Narrow QCD Jets*, *Phys. Rev.* **D 87** (2013) 094013 [[arXiv:1212.2106](#)] [[INSPIRE](#)].
- [73] J. Thaler and L.-T. Wang, *Strategies to identify boosted tops*, *JHEP* **07** (2008) 092 [[arXiv:0806.0023](#)] [[INSPIRE](#)].
- [74] L.G. Almeida et al., *Substructure of high- $p_T$  Jets at the LHC*, *Phys. Rev.* **D 79** (2009) 074017 [[arXiv:0807.0234](#)] [[INSPIRE](#)].
- [75] L.G. Almeida, S.J. Lee, G. Perez, I. Sung and J. Virzi, *Top Jets at the LHC*, *Phys. Rev.* **D 79** (2009) 074012 [[arXiv:0810.0934](#)] [[INSPIRE](#)].
- [76] S. Catani and M. Grazzini, *Collinear factorization and splitting functions for next-to-next-to-leading order QCD calculations*, *Phys. Lett.* **B 446** (1999) 143 [[hep-ph/9810389](#)] [[INSPIRE](#)].
- [77] S. Catani and M. Grazzini, *Infrared factorization of tree level QCD amplitudes at the next-to-next-to-leading order and beyond*, *Nucl. Phys.* **B 570** (2000) 287 [[hep-ph/9908523](#)] [[INSPIRE](#)].

- [78] A. Gehrmann-De Ridder and E.W.N. Glover, *A Complete  $O(\alpha \alpha_s)$  calculation of the photon + 1 jet rate in  $e^+e^-$  annihilation*, *Nucl. Phys. B* **517** (1998) 269 [[hep-ph/9707224](#)] [[INSPIRE](#)].
- [79] F.J. Tackmann, J.R. Walsh and S. Zuberi, *Resummation Properties of Jet Vetoes at the LHC*, *Phys. Rev. D* **86** (2012) 053011 [[arXiv:1206.4312](#)] [[INSPIRE](#)].
- [80] A. Abdesselam et al., *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur. Phys. J. C* **71** (2011) 1661 [[arXiv:1012.5412](#)] [[INSPIRE](#)].
- [81] A. Altheimer et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J. Phys. G* **39** (2012) 063001 [[arXiv:1201.0008](#)] [[INSPIRE](#)].
- [82] A. Altheimer et al., *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012*, *Eur. Phys. J. C* **74** (2014) 2792 [[arXiv:1311.2708](#)] [[INSPIRE](#)].
- [83] J.R. Forshaw, A. Kyrieleis and M.H. Seymour, *Super-leading logarithms in non-global observables in QCD*, *JHEP* **08** (2006) 059 [[hep-ph/0604094](#)] [[INSPIRE](#)].
- [84] J.R. Forshaw, A. Kyrieleis and M.H. Seymour, *Super-leading logarithms in non-global observables in QCD: Colour basis independent calculation*, *JHEP* **09** (2008) 128 [[arXiv:0808.1269](#)] [[INSPIRE](#)].
- [85] J. Keates and M.H. Seymour, *Super-leading logarithms in non-global observables in QCD: Fixed order calculation*, *JHEP* **04** (2009) 040 [[arXiv:0902.0477](#)] [[INSPIRE](#)].
- [86] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *Dissecting Soft Radiation with Factorization*, *Phys. Rev. Lett.* **114** (2015) 092001 [[arXiv:1405.6722](#)] [[INSPIRE](#)].
- [87] M. Cacciari and G.P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett. B* **659** (2008) 119 [[arXiv:0707.1378](#)] [[INSPIRE](#)].
- [88] M. Cacciari, G.P. Salam and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005 [[arXiv:0802.1188](#)] [[INSPIRE](#)].
- [89] G. Soyez, G.P. Salam, J. Kim, S. Dutta and M. Cacciari, *Pileup subtraction for jet shapes*, *Phys. Rev. Lett.* **110** (2013) 162001 [[arXiv:1211.2811](#)] [[INSPIRE](#)].
- [90] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084 [[arXiv:0912.1342](#)] [[INSPIRE](#)].
- [91] G.P. Korchemsky and G.F. Sterman, *Power corrections to event shapes and factorization*, *Nucl. Phys. B* **555** (1999) 335 [[hep-ph/9902341](#)] [[INSPIRE](#)].
- [92] G.P. Korchemsky and S. Tafat, *On power corrections to the event shape distributions in QCD*, *JHEP* **10** (2000) 010 [[hep-ph/0007005](#)] [[INSPIRE](#)].
- [93] A.J. Larkoski, I. Moult and D. Neill, *Non-Global Logarithms, Factorization, and the Soft Substructure of Jets*, [arXiv:1501.04596](#).
- [94] S. Caron-Huot, *Resummation of non-global logarithms and the BFKL equation*, [arXiv:1501.03754](#).
- [95] Z. Ligeti, I.W. Stewart and F.J. Tackmann, *Treating the b quark distribution function with reliable uncertainties*, *Phys. Rev. D* **78** (2008) 114014 [[arXiv:0807.1926](#)] [[INSPIRE](#)].