Factorization and resummation for groomed multi-prong jet shapes

Andrew J. Larkoski,[a] Ian Moult[b,c] and Duff Neill[d]

[a] *Physics Department, Reed College,*
Portland, OR 97202, U.S.A.
[b] *Berkeley Center for Theoretical Physics, University of California,*
Berkeley, CA 94720, U.S.A.
[c] *Theoretical Physics Group, Lawrence Berkeley National Laboratory,*
Berkeley, CA 94720, U.S.A.
[d] *Theoretical Division, Los Alamos National Laboratory,*
Los Alamos, NM 87545, U.S.A.

E-mail: larkoski@reed.edu, ianmoult@lbl.gov, duff.neill@gmail.com

Abstract: Observables which distinguish boosted topologies from QCD jets are playing an increasingly important role at the Large Hadron Collider (LHC). These observables are often used in conjunction with jet grooming algorithms, which reduce contamination from both theoretical and experimental sources. In this paper we derive factorization formulae for groomed multi-prong substructure observables, focusing in particular on the groomed D_2 observable, which is used to identify boosted hadronic decays of electroweak bosons at the LHC. Our factorization formulae allow systematically improvable calculations of the perturbative D_2 distribution and the resummation of logarithmically enhanced terms in all regions of phase space using renormalization group evolution. They include a novel factorization for the production of a soft subjet in the presence of a grooming algorithm, in which clustering effects enter directly into the hard matching. We use these factorization formulae to draw robust conclusions of experimental relevance regarding the universality of the D_2 distribution in both e^+e^- and pp collisions. In particular, we show that the only process dependence is carried by the relative quark vs. gluon jet fraction in the sample, no non-global logarithms from event-wide correlations are present in the distribution, hadronization corrections are controlled by the perturbative mass of the jet, and all global color correlations are completely removed by grooming, making groomed D_2 a theoretically clean QCD observable even in the LHC environment. We compute all ingredients to one-loop accuracy, and present numerical results at next-to-leading logarithmic accuracy for e^+e^- collisions, comparing with parton shower Monte Carlo simulations. Results for pp collisions, as relevant for phenomenology at the LHC, are presented in a companion paper [1].

Keywords: Jets

ArXiv ePrint: 1710.00014

Contents

1   Introduction

The efficient and robust identification of hadronically-decaying boosted electroweak bosons is a problem of fundamental importance at Run 2 and into the future of the Large Hadron Collider (LHC) physics programme. The identification of boosted hadronically-decaying $H/W/Z$ bosons requires two key ingredients: observables for discriminating multi-prong jet substructure, and a method for removing contamination within the jets. There has been substantial progress in both of these aspects of jet physics over the past few years.[1,2] One of the most powerful observables for discrimination of two-prong substructure is D_2 [14, 15], based on the n-point energy correlation functions [16]. D_2 parametrically separates jets

---

[1]For a recent review of theory and machine learning aspects of jet substructure, see [2].

[2]A summary of studies from the LHC using jet substructure can be found at https://twiki.cern.ch/twiki/bin/view/AtlasPublic and http://cms-results.web.cern.ch/cms-results/public-results/publications/. For the Run 2 experimental status, see the performance documentation in [3–13].

with one- and two-prongs, and has been used extensively during Run 2 by ATLAS [17–26]. Because of the high-luminosity environment of the LHC, the robust measurement of the substructure of jets requires methods for removing contamination from underlying event, pile-up, or other sources that are mostly uncorrelated with the hard scattering. Of these so-called jet grooming techniques, the modified mass drop (mMDT) [27, 28] and soft drop [29] groomers are theoretically most well-understood. Soft drop groomed mass distributions of jets initiated both by light QCD partons and by hadronically decaying Z bosons were measured recently by CMS [30].

Many theory advancements in understanding these observables and techniques have been made recently. The groomed jet mass and several related single prong observables have been calculated, both for QCD jets and electroweak bosons, up to next-to-leading logarithm (NLL) [27–29, 31], and very recently for top jets [32]. Refs. [33, 34] presented the first jet substructure calculation to next-to-next-to-leading logarithmic accuracy (NNLL) matched to fixed-order at $\mathcal{O}(\alpha_s^2)$ for the mMDT or soft drop groomed jet mass. A number of two-prong observables, both ungroomed [35] and groomed [36, 37], have been computed at leading-logarithmic accuracy (LL), and the observable D_2 was calculated for QCD jets and boosted Z bosons produced in e^+e^- collisions to next-to-leading logarithmic accuracy (NLL) [15]. These calculations have helped to put the understanding of jet substructure observables on firmer theoretical footing, and have inspired a number of new jet substructure techniques currently used at the LHC.

In this paper, we present the first theory calculations to NLL accuracy for groomed jets on which two-prong observables like D_2 are measured, and present a systematically improvable framework for their calculation. This is achieved using factorization formulae for groomed two-prong jet observables derived using the techniques of soft-collinear effective theory (SCET) [38–42]. These allow for the resummation of logarithmically enhanced terms in all regions of phase space using the renormalization group evolution of field theoretic operators, significantly simplifying higher order calculations, and also provide operator definitions of non-perturbative effects. We will illustrate our framework by performing a calculation of the D_2 observable on jets produced in e^+e^- collisions, although our approach is more general, and could be applied to related observables, such as N_2 [43]. In a companion paper [1] we present D_2 distributions for mMDT/soft drop groomed jets produced in pp collisions for processes of phenomenological relevance for jet substructure at the LHC.

Schematically, the soft drop groomer steps through the clustering history of a jet and removes those branches in the jet which fail the requirement

$$\frac{\min[E_i, E_j]}{E_i + E_j} > z_{\text{cut}} \left(\frac{\theta_{ij}}{R} \right)^\beta . \tag{1.1}$$

Here, E_i is the energy of branch i, θ_{ij} is the angle between branches i and j, and R is the jet radius. z_{cut} and β are parameters of soft drop; in this paper, we will only consider $\beta = 0$, for which soft drop coincides with the mMDT groomer (Due to their equivalence we will use their names interchangeably.). Typical values of z_{cut} are $z_{\text{cut}} \approx 0.1$. When eq. (1.1) is satisfied, the procedure terminates. On jets that have been groomed with soft drop, we

will measure both the jet mass and the observable D_2.[3] We will work in the formal limit

$$\frac{m_J^2}{E_J^2} \ll z_{\text{cut}} \ll 1\,, \tag{1.2}$$

where m_J is the groomed jet mass and E_J is the jet energy. This limit is both relevant for phenomenology and vital for theoretical simplicity. Eq. (1.2) is satisfied for electroweak scale masses, TeV scale jet energies, and $z_{\text{cut}} = 0.1$. By working in the regime of eq. (1.2), only collinear emissions pass the soft drop requirement, no non-global logarithms arising from color-connections to the rest of the event [46] contribute to the shape of the distribution, and corrections from hadronization are significantly reduced. This will enable us to make precise and robust predictions for the distribution of D_2 as measured on these jets.

In fact, the factorization formulae that we derive for the mMDT/soft dropped D_2 exhibit a universality, and the resulting D_2 distribution is largely independent of the jet energy. From the factorization formulae, we will show that:

- the leading non-perturbative corrections are suppressed by powers of the groomed jet mass, and are independent of the jet energy;

- the endpoint of the distribution is set by the grooming parameter z_{cut}, and is independent of the jet mass and energy;

- perturbative power corrections are suppressed by $m_J^2/(z_{\text{cut}} E_J^2) \ll 1$ throughout the distribution;

- the quark or gluon flavor of a jet is well-defined at leading power in $z_{\text{cut}} \ll 1$.

These properties imply that the full mMDT/soft drop D_2 is only very weakly dependent on the jet mass m_J and the jet energy E_J. This is a desirable property, especially for experimental applications. The explicit design of observables that are independent of mass and p_T (or energy) cuts has been a subject of recent research [43, 47]. Due to the final property, quark and gluon jet distributions can be individually extracted from D_2 distributions of jets produced in different processes.

The outline of this paper is as follows. In section 2, we define the mMDT and soft drop groomers appropriate for both jets produced in e^+e^- and pp collisions, and define the energy correlation functions and the ratio observable D_2. In section 3, we will review the previously published factorization formulae for D_2 and groomed jet mass and derive new factorization formulae for the groomed D_2 cross section in e^+e^- collisions. Collinear factorization enables us to use all of the results from e^+e^- collisions to formulate the factorized cross section for the groomed D_2 observable for jets produced in pp collisions in section 4. From the form of the factorization formula, we are immediately able to make all-orders statements regarding properties of the distribution. We will review those properties that are identical to those of the groomed jet mass and discuss in some detail reduced hadronization corrections for the D_2 distribution in section 5. In section 6, we present numerical results, comparing

---

[3]Without a mass cut, D_2 is not infrared and collinear safe, but is Sudakov safe [44, 45]. It can therefore be calculated in resummed perturbation theory.

our NLL predictions in e^+e^- collisions to parton shower Monte Carlo. A comparison of robust qualitative features of the distribution derived from the factorization formula for pp are compared with parton shower Monte Carlo predictions in section 7. We conclude in section 8. Technical details and calculations are presented in appendices.

2 Observables

In this section, we review the definitions of the observables and grooming procedures that will be studied in this paper. As mentioned in the introduction, we will restrict our focus to grooming with mMDT [27, 28] or equivalently, soft drop with angular exponent $\beta = 0$ [29]. Definitions of mMDT and energy correlation functions will be presented for jets produced in both e^+e^- and pp collisions.

2.1 Modified mass drop/soft drop grooming

The modified mass drop/ soft drop groomer proceeds in the following way. The identified jet is reclustered with the Cambridge/Aachen algorithm [48–50], which orders emissions in the jet by their relative angle. Then, starting at the widest angle, the two branches following from each splitting in the jet are required to satisfy an energy fraction constraint. For branches i and j in a jet produced in e^+e^- collisions, this requirement is

$$\frac{\min[E_i, E_j]}{E_i + E_j} > z_{\text{cut}} \,, \tag{2.1}$$

where E_i is the energy of branch i. For a jet from pp collisions, the requirement is

$$\frac{\min[p_{Ti}, p_{Tj}]}{p_{Ti} + p_{Tj}} > z_{\text{cut}} \,, \tag{2.2}$$

where p_{Ti} is the transverse momentum with respect to the proton beam of branch i. If these requirements are not satisfied, the softer (lower energy/transverse momentum) branch is removed from the jet, and the procedure iterates to the next splitting on the harder branch. The process terminates when the two branches in a splitting in the jet satisfy eq. (2.1) or eq. (2.2), as appropriate. The cut parameter z_{cut} is typically taken to be $z_{\text{cut}} \sim 0.1$. In this paper, we will formally assume that $z_{\text{cut}} \ll 1$, so that emissions that fail these requirements are necessarily soft. Once a jet has been groomed, any observable can be measured on its remaining constituents.

2.2 Energy correlation functions and D_2

For powerful identification of hadronic decays of electroweak bosons, we use the energy correlation functions [16].[4] The n-point energy correlation function is sensitive to radiation about $n-1$ hard cores in the jet. Therefore, for this application, we use the two- and three-point energy correlation functions, which are sensitive to radiation about one or two hard

---

[4]Another common approach is to use the N-subjettiness ratio observables [51, 52]. However, these are poorly behaved in perturbation theory [53].

cores in the jet. For a set of particles $\{i\}$, in a jet J, the energy correlation functions in e^+e^- collisions are defined as

$$e_2^{(\alpha)}\Big|_{e^+e^-} = \frac{1}{E_J^2} \sum_{i<j\in J} E_i E_j \left(\frac{2p_i \cdot p_j}{E_i E_j}\right)^{\alpha/2} ,\qquad(2.3)$$

$$e_3^{(\alpha)}\Big|_{e^+e^-} = \frac{1}{E_J^3} \sum_{i<j<k\in J} E_i E_j E_k \left(\frac{2p_i \cdot p_j}{E_i E_j}\frac{2p_i \cdot p_k}{E_i E_k}\frac{2p_j \cdot p_k}{E_j E_k}\right)^{\alpha/2} .\qquad(2.4)$$

For jets produced in pp collisions, the energy correlation functions are simply modified as

$$e_2^{(\alpha)}\Big|_{pp} = \frac{1}{p_{TJ}^2} \sum_{i<j\in J} p_{Ti} p_{Tj} R_{ij}^\alpha ,\qquad(2.5)$$

$$e_3^{(\alpha)}\Big|_{pp} = \frac{1}{p_{TJ}^3} \sum_{i<j<k\in J} p_{Ti} p_{Tj} p_{Tk} R_{ij}^\alpha R_{ik}^\alpha R_{jk}^\alpha .\qquad(2.6)$$

Here R_{ij} is the distance between particles i and j in the pseudorapidity-azimuth angle plane. For jets that are central ($p_{TJ} \sim E_J$) and if all emissions in the jet are collinear, the definitions of the two- and three-point energy correlation functions for jets in pp collisions are equivalent to those for e^+e^- collisions.

The angular exponent α in the definition of the energy correlation functions is a parameter that controls sensitivity to wide-angle emissions. For jets that consist of massless particles, the two-point energy correlation function in e^+e^- collisions reduces to a function of the jet mass, m_J, if $\alpha = 2$:

$$e_2^{(2)}\Big|_{e^+e^-} = \frac{m_J^2}{E_J^2} .\qquad(2.7)$$

It has been shown that the optimal observable, formed from $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$, for discrimination of boosted hadronic decays of electroweak bosons (with a hard two-prong substructure) from QCD jets (typically with a single hard prong) is a particular ratio of the energy correlation functions called $D_2^{(\alpha)}$ [14, 15]. $D_2^{(\alpha)}$ is defined as

$$D_2^{(\alpha)} = \frac{e_3^{(\alpha)}}{(e_2^{(\alpha)})^3} .\qquad(2.8)$$

In this paper, for jets which have been groomed with mMDT, we subsequently measure $D_2^{(\alpha)}$. For brevity, we will often denote $D_2^{(\alpha)}$ generically as D_2.

2.3 Phase space and behavior of groomed D_2

In this section, we use the power counting techniques of [14] to review the phase space structure of a jet on which the observables $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$ are measured. We then discuss how this is modified when the jet is groomed with mMDT or soft drop. In particular, we emphasize the parametric features of the groomed D_2 distribution that will be reproduced by our calculation, as well as demonstrating that D_2 remains a powerful discriminant even after grooming has been applied.

We begin with a brief review of the structure of the $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$ phase space without grooming, as was considered in detail in [14, 15]. With no grooming the dominant emissions in a one-prong jet with $e_2^{(\alpha)}, e_3^{(\alpha)} \ll 1$ are either soft (low energy) or collinear to the jet core. These contributions to the value of $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$ scale like

$$
\begin{aligned}
e_2^{(\alpha)} &\sim \theta_{cc}^\alpha + z_s \,, \\
e_3^{(\alpha)} &\sim \theta_{cc}^{3\alpha} + \theta_{cc}^\alpha z_s + z_s^2 \,,
\end{aligned}
\tag{2.9}
$$

where z_s is the characteristic energy fraction of soft emissions and θ_{cc} is the characteristic angle of collinear emissions. Depending on the assumptions made about the relative scaling of z_s and θ_{cc}, one finds the upper and lower boundaries of the phase space for one-prong jets. The upper boundary is the so-called soft haze region where

$$
e_3^{(\alpha)} \sim (e_2^{(\alpha)})^2 \,,
\tag{2.10}
$$

while the lower boundary corresponds to the scaling

$$
e_3^{(\alpha)} \sim (e_2^{(\alpha)})^3 \,.
\tag{2.11}
$$

Therefore, up to order-1 coefficients, one-prong jets have a measured value of $e_3^{(\alpha)}$ that lies between $(e_2^{(\alpha)})^3$ and $(e_2^{(\alpha)})^2$.

To determine the region of phase space where a two-prong jet lives, it is sufficient to consider a jet with two hard, collinear prongs, with all other radiation at much lower energy. In this case, the scaling of the contributions to $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$ are

$$
\begin{aligned}
e_2^{(\alpha)} &\sim \theta_{ab}^\alpha \,, \\
e_3^{(\alpha)} &\sim \theta_{cc}^\alpha \theta_{ab}^{2\alpha} + \theta_{ab}^{3\alpha} z_{cs} + \theta_{ab}^\alpha z_s \,.
\end{aligned}
\tag{2.12}
$$

Here, θ_{ab} is the angle between the hard prongs, θ_{cc} is the characteristic angular size of each of the hard prongs individually, z_{cs} is the energy fraction of collinear-soft radiation emitted from the dipole of the two hard prongs, and z_s is the energy fraction of soft radiation emitted at large angles. The requirement that the hard prongs are well-defined restricts the energy fraction of the collinear-soft radiation to be small:

$$
z_{cs} \sim \frac{e_3^{(\alpha)}}{(e_2^{(\alpha)})^3} \ll 1 \,.
\tag{2.13}
$$

That is, two-prong jets have measured values of $e_3^{(\alpha)}$ that are much smaller than $(e_2^{(\alpha)})^3$. The one- and two-prong regions of phase space in the $(e_2^{(\alpha)}, e_3^{(\alpha)})$ plane are illustrated in figure 1a.

We now consider how this phase space is modified in the presence of grooming. First, we consider the parametric scaling of contributions in the one-prong region of phase space. We assume that there exists radiation in the groomed jet whose energy fraction is set by z_{cut} at a characteristic angle θ_{sc} from the jet core, in addition to the collinear modes. The
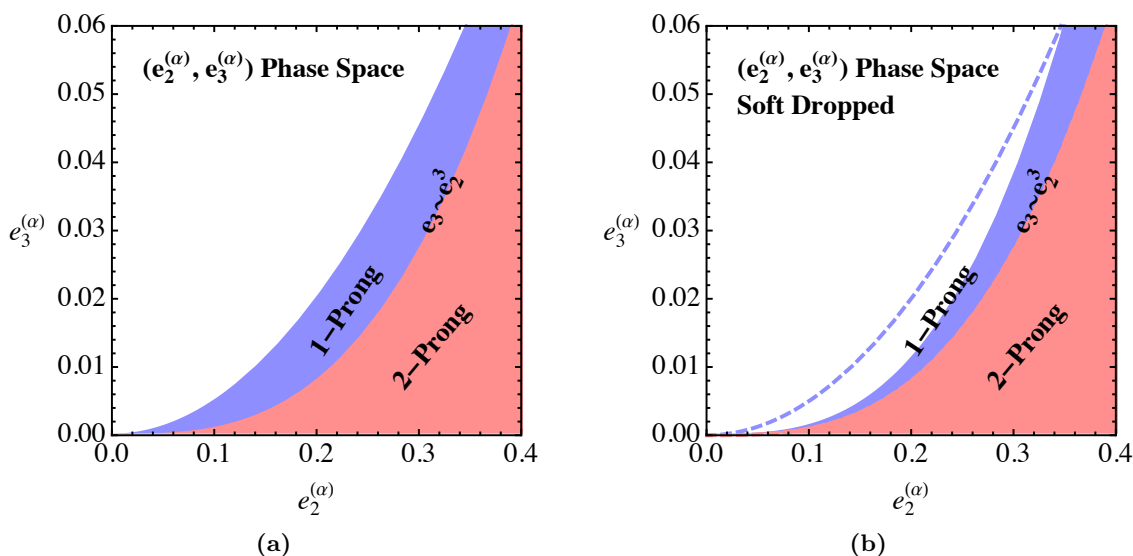
Figure 1. The phase space for one- and two-prong jets on which the energy correlation functions $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$ are measured. (a) Phase space for ungroomed jets, with the absolute upper boundary scaling like $e_3^{(\alpha)} \sim (e_2^{(\alpha)})^2$. (b) Phase space for mMDT/soft drop groomed jets, where the two-prong region is unchanged, while the absolute upper boundary now scales like $e_3^{(\alpha)} \sim (e_2^{(\alpha)})^3/z_{\rm cut}$.

soft modes (which had energy fraction z_s) have been removed by the mMDT/soft drop procedure. The contributions to the observables are then

$$
\begin{aligned}
e_2^{(\alpha)} &\sim z_{\rm cut}\theta_{sc}^\alpha + \theta_{cc}^\alpha \,, \\
e_3^{(\alpha)} &\sim \theta_{cc}^{3\alpha} + \theta_{cc}^\alpha \theta_{sc}^{2\alpha} z_{\rm cut} + \theta_{sc}^{3\alpha} z_{\rm cut}^2 \,.
\end{aligned}
\tag{2.14}
$$

From this we can determine the boundaries of the phase space by imposing different relationships between the characteristic angles and energy fractions. As in the ungroomed case, the lower bound of this phase space region occurs when collinear emissions dominate the value of $e_2^{(\alpha)}$:

$$
e_2^{(\alpha)} \sim \theta_{cc}^\alpha \gg z_{\rm cut}\theta_{sc}^\alpha, \qquad e_3^{(\alpha)} \sim (e_2^{(\alpha)})^3 \,.
\tag{2.15}
$$

This is the same lower boundary as with ungroomed jets. This implies in particular, that D_2 remains a powerful discriminant, even after grooming has been applied.

The upper boundary of this region is more interesting. Assuming that the contributions to $e_2^{(\alpha)}$ are democratic

$$
e_2^{(\alpha)} \sim z_{\rm cut}\theta_{sc}^\alpha + \theta_{cc}^\alpha \,,
\tag{2.16}
$$

we find the characteristic angular scale of the radiation sensitive to $z_{\rm cut}$ to be

$$
\theta_{sc}^\alpha \sim \frac{e_2^{(\alpha)}}{z_{\rm cut}} \,.
\tag{2.17}
$$

Using this scaling, for $e_3^{(\alpha)}$ for groomed jets, we then find

$$
e_3^{(\alpha)} \sim \frac{(e_2^{(\alpha)})^3}{z_{\rm cut}} \,.
\tag{2.18}
$$

We refer to the region of phase space near this upper boundary as "collinear-soft haze". Note that, assuming that two-prong jets have two hard prongs, their phase space is unchanged from the ungroomed case.

This is quite interesting. Since we formally assume $z_{\text{cut}} \ll 1$, there is a separation of the collinear-soft haze region from the lower boundary of groomed one-prong jets. However, both boundaries have the same cubic relationship between $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$. The phase space for groomed jets is illustrated in figure 1b. Grooming the jet removes the region of phase space where $e_3^{(\alpha)} \sim (e_2^{(\alpha)})^2$. From this analysis, it is straightforward to determine the maximal value of D_2 with and without grooming. For the ungroomed jet, the maximal value of D_2 is when

$$D_2^{(\alpha)}\Big|_{\text{max}} \sim \frac{e_3^{(\alpha)}}{(e_2^{(\alpha)})^3}\Bigg|_{e_3^{(\alpha)} \sim (e_2^{(\alpha)})^2} \sim \frac{1}{2e_2^{(\alpha)}}\,. \tag{2.19}$$

With a more careful analysis (discussed in ref. [14]), one can derive the factor of $1/2$ in the location of the endpoint. Note that for $\alpha = 2$ this maximum value is sensitive to both the energy and mass of the jet:

$$D_2^{(2)}\Big|_{\text{max}} \sim \frac{E_J^2}{2m_J^2}\,. \tag{2.20}$$

The endpoint of the D_2 distribution formally increases without bound as the energy of the jet increases, for a fixed mass cut.[5] On the other hand, when the jet is groomed, we find

$$D_2^{(\alpha)}\Big|_{\text{max, soft drop}} \sim \frac{e_3^{(\alpha)}}{(e_2^{(\alpha)})^3}\Bigg|_{e_3^{(\alpha)} \sim (e_2^{(\alpha)})^3/z_{\text{cut}}} \sim \frac{1}{2z_{\text{cut}}}\,. \tag{2.21}$$

Therefore, when a jet is groomed with mMDT or soft drop, the endpoint of the D_2 distribution is independent of both the jet mass and energy. This property will be one part of the reason why the groomed D_2 distribution is incredibly robust to changes in energy and/or mass cuts.

To demonstrate that this scaling is satisfied in simulation, in figure 2 we plot the distribution of jets in the $(e_2^{(\alpha)}, e_3^{(\alpha)})$ phase space plane as simulated in parton shower Monte Carlo. The details of the Monte Carlo simulation will be described in section 6. Here we use the angular exponent $\alpha = 1$ and impose an upper cut on the mass of $m_J < 100\,\text{GeV}$, which more clearly illustrates the phase space regions. The same general features are present for other values of α. On these jets we then measure $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$, either before grooming or after mMDT/soft drop grooming, with $z_{\text{cut}} = 0.1$. In figure 2a, we show the ungroomed phase space, and jets populate up to the curve where $e_3^{(\alpha)} \sim (e_2^{(\alpha)})^2$. Once grooming is applied, however, jets only populate up to the curve $e_3^{(\alpha)} \sim (e_2^{(\alpha)})^3/z_{\text{cut}}$, as illustrated in figure 2b. This demonstrates that our parametric scaling analysis of the phase space is satisfied by parton shower simulation. More detailed tests will be provided in section 6, when we study the structure of the D_2 distribution in our analytic calculation, and in parton shower Monte Carlo.

---

[5]Of course, this isn't quite true because there is a characteristic mass scale of QCD. Even perturbatively this isn't true because the Sudakov factor will exponentially suppress low masses.
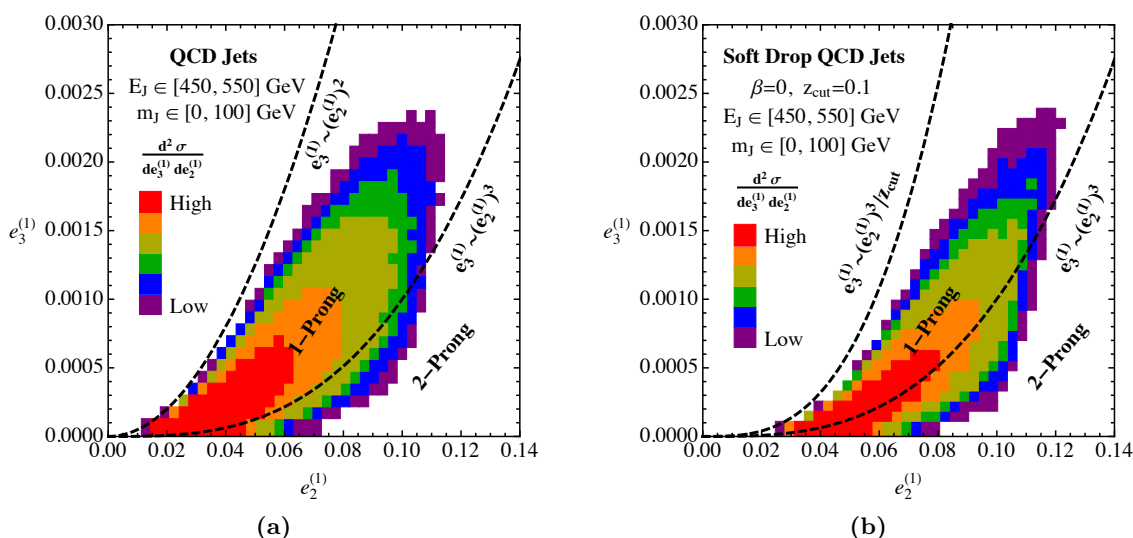
Figure 2. Illustration of the population of jets from $1\,\mathrm{TeV}$ e^+e^- collisions in the $(e_2^{(\alpha)}, e_3^{(\alpha)})$ phase space plane as simulated in parton shower Monte Carlo. Here, $\alpha = 1$ and the mass of the jet is restricted as $m_J < 100\,\mathrm{GeV}$. (a) Ungroomed jets, that extend up to $e_3^{(\alpha)} \sim (e_2^{(\alpha)})^2$. (b) mMDT/soft drop groomed jets, with $z_{\mathrm{cut}} = 0.1$, that extend up to $e_3^{(\alpha)} \sim (e_2^{(\alpha)})^3/z_{\mathrm{cut}}$.

The location of the endpoint for the groomed D_2 distribution also has important consequences for its calculation. In particular, for the ungroomed D_2 distribution, the endpoint is at $\sim 1/e_2^{(\alpha)}$. In the limit $e_2^{(\alpha)} \ll 1$, this is formally large, and can be neglected. This is what was done in ref. [15]. However, for the groomed D_2 distribution, the endpoint of the distribution is at $\sim 1/z_{\mathrm{cut}}$. Since we assume $z_{\mathrm{cut}} \gg e_2^{(\alpha)}$, we must compute the matrix element in this region of the phase space, and match it to our resummed calculation, to accurately predict the endpoint of the distribution.

3 Factorized cross section in e^+e^- collisions

In this section we present factorization formulae for mMDT/soft drop groomed D_2. These allow for a systematically improvable calculation of the D_2 distribution, and the resummation of logarithmically enhanced terms in all regions of phase space to be resummed by renormalization group evolution. The factorization formulae are presented in the language of SCET [38–41], an effective field theory describing soft and collinear radiation in the presence of a hard scattering. For the case of jet substructure observables, where multiple hierarchies are present within the jet, extensions of SCET are required. These have been developed in refs. [15, 54–56], and were discussed in detail in the context of the D_2 observable in ref. [15]. In this section we will restrict ourselves to giving physical descriptions of the functions appearing in the factorization formulae. Field theoretic definitions, and one-loop calculations, are given in the appendices.

Our approach to deriving the factorization formulae will closely follow the techniques used to study the groomed jet mass and the ungroomed D_2 observable. In particular, we

will begin from the factorization for the groomed jet mass cross section, and then perform a refactorization of the resolved substructure. Because of this, we begin in section 3.1 with a review of the factorization formulae in both these cases. We then present the factorization for groomed D_2 in section 3.2. Throughout this section, we will restrict ourselves to the case of e^+e^- collisions for simplicity. In section 4, we will then show that the grooming algorithm allows us to trivially extend this factorization formula to the case of pp collisions.

3.1 Review of known results

Factorization formulae are known for both the groomed jet mass [33, 34] and for the ungroomed D_2 observable [15]. Since our factorization for the groomed D_2 observable will rely heavily on ingredients from both these analyses, we begin by reviewing the essential ingredients of the factorization formulae for these two cases. The discussion will be brief, and more details can be found in the respective papers.

3.1.1 Groomed jet mass factorization formula

A factorization formulae was presented in SCET for the soft dropped two-point energy correlation functions $e_2^{(\alpha)}$, and was used to calculate the distribution to NNLL order [33, 34]. Throughout this section we will always take the soft drop parameter $\beta = 0$. The case $\beta > 0$ follows an identical logic, and is discussed in detail in refs. [33, 34].

The factorization formula is valid in the limit $e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1$. It can be derived through a multi-stage matching procedure from the standard SCET involving a global soft function and jet functions. The first stage of the matching is a soft and collinear factorization, with the soft virtuality set by $Q z_{\text{cut}}$, and the collinear virtuality set by $e_{2,R}^{(\alpha)}$:

$$\frac{d^2\sigma}{de_{2,L}^{(\alpha)} de_{2,R}^{(\alpha)}} = H(Q^2) S(z_{\text{cut}}) \left[J(e_{2,L}^{(\alpha)}, z_{\text{cut}}) \right] \left[J(e_{2,R}^{(\alpha)}, z_{\text{cut}}) \right] . \tag{3.1}$$

The soft drop grooming has isolated the jet dynamics from the rest of the event, due to the angular ordering of the algorithm. However, this factorization still contains large logarithms within the collinear sector. These can be resummed by refactorizing into a collinear-soft function, which allows for the resummation of all logarithms of $e_{2,R}^{(\alpha)}$. The final factorization formula for measuring $e_2^{(\alpha)}$ in each of the groomed hemispheres in e^+e^- collisions is given by

$$\frac{d^2\sigma}{de_{2,L}^{(\alpha)} de_{2,R}^{(\alpha)}} = H(Q^2) S(z_{\text{cut}}) \left[J(e_{2,L}^{(\alpha)}) \otimes S_c(e_{2,L}^{(\alpha)} z_{\text{cut}}) \right] \left[J(e_{2,R}^{(\alpha)}) \otimes S_c(e_{2,R}^{(\alpha)} z_{\text{cut}}) \right] . \tag{3.2}$$

The physical interpretation of the functions entering this factorization formula are as follows (field theoretic definitions can be found in [33, 34]):

- $H(Q^2)$ is the standard hard function, describing in this case the production of two back to back jets in an e^+e^- collision.

- $S(z_{\text{cut}})$ is the global soft function. It describes wide angle soft radiation, which is removed by the groomer. It is therefore independent of the observable, and depends just on z_{cut}.
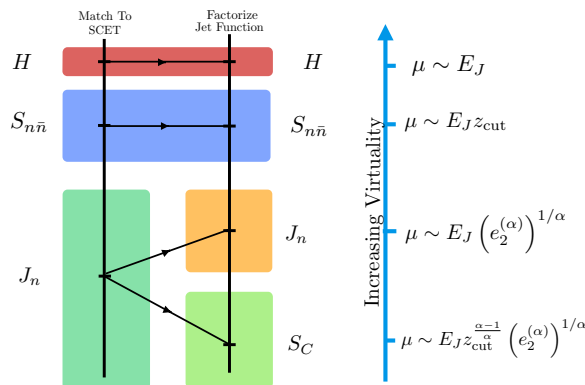
Figure 3. An illustration of the multistage matching procedure for the factorization of the soft dropped energy correlation function $e_2^{(\alpha)}$. For simplicity, we have taken the soft drop angular exponent, $\beta = 0$.

- $J(e_2^{(\alpha)})$ is a jet function describing collinear radiation. Since this radiation is energetic, it is not affected by the groomer, so that this function does not depend on z_{cut}.

- $S(e_2^{(\alpha)} z_{\text{cut}}^{\alpha-1})$ describes collinear soft radiation, which contributes to the observable, but is sensitive to the groomer. It can be shown that it depends only on the scales z_{cut} and $e_2^{(\alpha)}$ through the combination $e_2^{(\alpha)} z_{\text{cut}}^{\alpha-1}$, as indicated by the argument of the function.

The multi-stage matching procedure is shown in figure 3, which also shows the virtualities of the modes contributing to the factorization formula. The results for all functions appearing in the factorization formula of eq. (3.2) allowing for resummation up to NNLL were computed in refs. [33, 34]

3.1.2 Ungroomed D_2 factorization formula

In ref. [15] a factorization formula was presented for the D_2 observable. For a two-prong substructure observable such as D_2, multiple kinematic regimes with distinct hierarchies exist, each of which contribute to a different region of the multi-dimensional phase space discussed in section 2.3. The approach taken in ref. [15] was to identify all parametric regions of phase space where hierarchies occur, and to develop distinct effective field theories describing each of these regions. The different effective field theories can then be pieced together to give a complete description of the entire phase space region.

In ref. [15], three phase space regions were required to provide a description of the D_2 observable:

- Soft Haze: the jet does not have a resolved substructure, and is formed from unresolved soft and collinear radiation, as in figure 4a. Here the factorization formula involves multi-differential jet and soft functions as developed in refs. [57–59].

- Collinear Subjets: the jet is formed of two subjets with small opening angle, and large energies, as shown in figure 4b. The factorization formula in this region of phase space
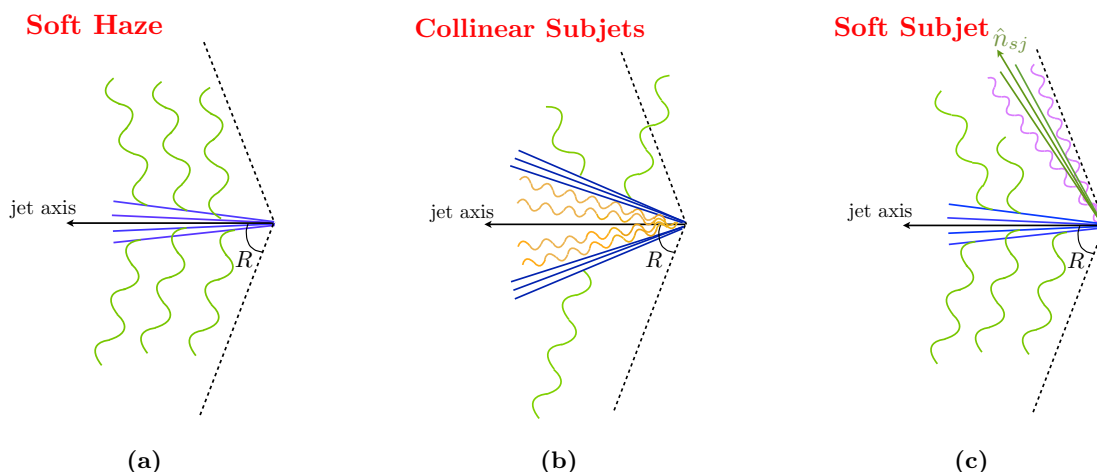
Figure 4. Regions of interest for studying the two-prong substructure of a jet. (a) Soft haze region in which no subjets are resolved. (b) Collinear subjets with comparable energy and a small opening angle. (c) Soft subjet carrying a small fraction of the total energy, and at a wide angle from the hard subjet.

is formulated in the SCET_+ theory of ref. [54]. In addition to the standard soft and collinear modes of SCET, it involves collinear-soft modes emitted from the dipole formed by the two subjets. In this factorization formula, the modes describing the radiation within the subjets are not sensitive to the presence of the jet boundary.

- Soft Subjet: the jet is formed of a single highly-energetic subjet, and a wide-angle subjet with energy fraction $z_{sj} \ll 1$, as shown in figure 4c. The effective field theory description of this region of phase space was first presented in ref. [55]. Its complexity arises due to the fact that the soft subjet is sensitive to the presence of the jet boundary.

A smooth transition between the collinear subjets and soft subjet regions of phase space was achieved using a zero bin-like procedure to remove any overlap. A similar approach was advocated in ref. [56].

3.2 Groomed D_2 factorization formula

Having reviewed the factorization formula for the soft dropped energy correlation functions, as well as for the D_2 observable, we can now combine these two approaches to provide a factorized description for groomed D_2. This will be accomplished by refactorizing an analogous parent effective theory to the expression eq. (3.1) in the different parametric regions:

$$\frac{d^4\sigma}{de_{2,L}^{(\alpha)}de_{3,L}^{(\alpha)}de_{2,R}^{(\alpha)}de_{3,R}^{(\alpha)}} = H(Q^2)S(z_{\text{cut}})\left[J(e_{2,L}^{(\alpha)}, e_{3,L}^{(\alpha)}, z_{\text{cut}})\right]\left[J(e_{2,R}^{(\alpha)}, e_{3,R}^{(\alpha)}, z_{\text{cut}})\right] . \quad (3.3)$$

The merging and region analysis will be similar to that performed in ref. [15] for D_2 without soft drop. Before giving a detailed discussion of each of the factorized expressions

in the different phase space regions, we give a brief overview of the different regions of phase space that can contribute and the dynamics occurring in each region, as well as comparing them to the three phase space regions which contributed to ungroomed D_2, as shown in figure 4.

To describe the D_2 distribution of a jet on which the soft drop grooming algorithm has been applied, we will similarly need three regions of phase space. Note, however, that since all the factorizations will appear as refactorizations of the eq. (3.3), all components of the factorized expression which contribute to D_2 will be collinear in nature. This will significantly simplify the analysis. In particular, the wide-angle soft subjet region of phase space is completely removed from contributing to the observables by the soft drop algorithm. In the soft subjet region of phase space, we would have $e_2^{(\alpha)} \sim z_{sj}$. However, by assumption, we take $e_2^{(\alpha)} \ll z_{\text{cut}}$, and therefore, the wide angle soft subjet is removed by the soft drop algorithm. This region of phase space will instead be replaced by a collinear-soft subjet which has characteristic energy fraction $z_{cs} \sim z_{\text{cut}}$. The effective field theory description for this hierarchy is new, and will be described in section 3.2.3.

The three phase space regions that will contribute to the D_2 observable as measured on a soft dropped jet are shown schematically in figure 5. A brief description of each of the different phase space regions is as follows:

- Collinear-Soft Haze: the jet does not have a resolved substructure. It is formed entirely from unresolved collinear-soft radiation. This is shown schematically in figure 5a.

- Collinear Subjets: as shown in figure 5b, in the collinear subjets region of phase space, the jet consists of two subjets of approximately equal energies, and a small opening angle, surrounded by collinear-soft radiation.

- Collinear-Soft Subjets: the jet is formed of two subjets, of parametrically different energies, with the softer jet energy set by z_{cut}, but where the opening angle between the jets is still assumed to be small. Unlike the previous two phase space regions, because z_{cut} sets the energy of the soft jet, there is no additional collinear-soft radiation at a wider angle than the soft subjet. This is shown schematically in figure 5c.

It is interesting to contrast the different phase space regions for the D_2 observable with and without the soft drop grooming algorithm applied, as shown in figures 4 and 5. These configurations are similar, with the exception that the wide angle radiation is removed by the soft drop algorithm, so that only collinear-soft radiation remains. Importantly, this radiation is boosted along the direction of the jet. It is therefore not sensitive to the directions of other jets in the event, all of which appear boosted in the opposite direction, and it is also not sensitive to the radius of the jet. This will lead to a large degree of universality for the soft dropped D_2 distributions, and simplify their calculation in the presence of additional jets.

We now discuss each of the phase space regions in figure 5 in detail, and present factorization formulae describing the radiation in these different regions of phase space. These
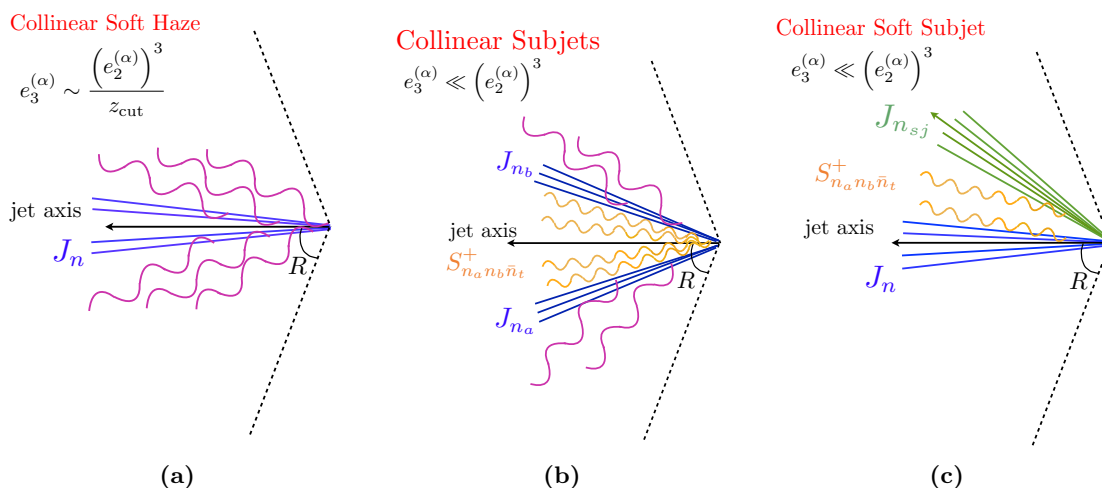
Figure 5. Regions of interest for studying the two-prong substructure of a jet on which the soft drop grooming algorithm has been applied. (a) Collinear-soft haze region in which no subjets are resolved. (b) Collinear subjets with comparable energy and a small opening angle. (c) Collinear-soft subjet carrying a small fraction of the total energy, with $z_{cs} \sim z_{\mathrm{cut}}$.

factorization formulae will allow for the radiation at each hierarchical scale to be described by a different function, allowing for large logarithms in the perturbative calculation to be resummed. A complete description of the groomed D_2 distribution can then be obtained by merging these different factorization formulae. We will discuss how this is done in sections 3.2.4 and 3.2.5.

3.2.1 Unresolved substructure: collinear-soft haze

We begin by discussing the factorization in the region of phase space where the jet has no resolved subjets. The factorization formula in this region of phase space will follow almost identically the soft-haze factorization formula of ref. [15], except that the soft function will be replaced by a boosted collinear-soft function due to the implementation of the soft drop algorithm. Recall that in this region of phase space we have collinear modes and collinear-soft modes, and the power counting for the observables is

$$
\begin{aligned}
e_2^{(\alpha)} &\sim z_{\mathrm{cut}}\theta_{sc}^{\alpha} + \theta_{cc}^{\alpha}, \\
e_3^{(\alpha)} &\sim \theta_{cc}^{\alpha}\theta_{sc}^{2\alpha}z_{\mathrm{cut}} + \theta_{sc}^{3\alpha}z_{\mathrm{cut}}^2 \,.
\end{aligned}
\tag{3.4}
$$

In this region of phase space, $e_3^{(\alpha)}$ can be expressed as the sum of two contributions. One is just the three-point energy correlation function $e_3^{(\alpha)}$ measured in the collinear-soft function and the other is a product of two-point energy correlation functions with different exponents as measured in the jet function and the collinear-soft function. The factorization formula
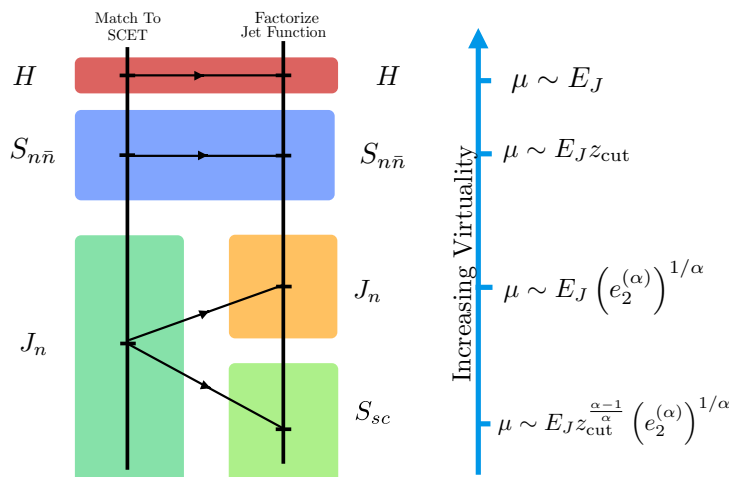
Figure 6. An illustration of the multistage matching procedure and relevant scales for the collinear-soft haze region of phase space. Because the jet and collinear-soft functions are not factorized, this is the same structure as with just measuring $e_2^{(\alpha)}$ on a groomed jet.

in this region of phase space is then given by

$$\frac{d^2\sigma}{de_2^{(\alpha)}de_3^{(\alpha)}} = HS(z_{\text{cut}}) \int de_2^c\, de_2^{sc}\, de_2^{(2\alpha)sc}\, de_3^{sc}\, \delta(e_3^{(\alpha)} - e_3^{sc} - e_2^c \cdot e_2^{(2\alpha)sc})\delta(e_2^{(\alpha)} - e_2^{sc} - e_2^c)$$
$$\times J(e_2^c)S_{sc}(e_2^{sc}, e_2^{(2\alpha)sc}, e_3^{sc}, z_{\text{cut}})\,. \quad (3.5)$$

For brevity, we only write the $e_2^{(\alpha)}$ and $e_3^{(\alpha)}$ dependence of a single hemisphere; including both hemispheres is trivial. Since the $e_3^{(\alpha)}$ observable is first non-zero with two emissions, this factorization formula first gives a non-trivial contribution at NNLL′ order, i.e., it requires the two-loop matrix elements (and the product of two one-loop matrix elements).

A brief description of the different functions entering the factorization formula in the collinear soft haze region is as follows

- $H(Q^2)$ is the hard function describing the underlying hard process, namely $e^+e^- \to q\bar{q}$.

- $S(z_{\text{cut}})$ is the global soft function, describing radiation which has been removed by the soft drop procedure. It depends only on z_{cut}, and not on the observables $e_2^{(\alpha)}$, or $e_3^{(\alpha)}$.

- $J(e_2^c)$ describes the collinear dynamics within the jet. It is independent of the soft drop algorithm. It contributes to the $e_3^{(\alpha)}$ observable only through the product form entering eq. (3.5).

- $S_{sc}(e_2^{sc}, e_2^{(2\alpha)sc}, e_3^{sc}, z_{\text{cut}})$ describes the soft collinear radiation within the jet. It is sensitive both to the soft drop criterion as well as contributing to the $e_2^{(\alpha)}$, $e_2^{(2\alpha)}$, and $e_3^{(\alpha)}$ observables. Here, $e_2^{(2\alpha)}$ means that the angular exponent is 2α.

An illustration of the multi-stage matching for this factorization theorem is illustrated in figure 6.

A similar factorization formula was proposed in ref. [15] for describing the unresolved region of phase space for the D_2 observable without the soft drop algorithm. This region also first contributed to the observable at NNLL′ order, and was therefore not considered. This was because the endpoint of the ungroomed distribution is $1/e_2^{(\alpha)} \gg 1$, and therefore the distribution has a smooth long tail, which can be well-approximated by simply extending the factorization formulae from the two-prong region of phase space. However, in the case that the soft drop algorithm is applied, it was shown in section 2.3 that the D_2 distribution has an upper boundary at $D_2^{\max} = 1/(2z_{\text{cut}})$. This endpoint feature is not described by the factorization formulae in the two-prong region of phase space, as it is expanded away. Matrix elements in the collinear-soft haze region of phase space are required to describe this kinematic feature. We will therefore compute the fixed-order matrix elements at $\mathcal{O}(\alpha_s^2)$ and match within the effective theory.

The most convenient way to calculate the D_2 distribution in the soft haze region to $\mathcal{O}(\alpha_s^2)$ is to integrate the appropriate $1 \to 3$ splitting functions, as described in appendix D. This is equivalent to calculating the D_2 distribution in the parent theory of eq. (3.3). One can explicitly check that one reproduces the matrix elements of the collinear-soft haze factorization when two of the emissions in the $1 \to 3$ splitting functions are taken to be soft, and when two are taken to be collinear and one is soft. These contributions reproduce the two-loop soft function, and the convolution between the one-loop jet and one-loop soft functions within the factorization formula of eq. (3.5).

A critical feature of the $1 \to 3$ splitting functions, as shown in appendix D, and by extension, also the collinear-soft haze factorization formula, is that at $\mathcal{O}(\alpha_s^2)$ all the $e_2^{(\alpha)}$ dependence explicitly scales out of the matrix element when we scale $e_3^{(\alpha)}$ to the D_2 ratio. Thus the $e_2^{(\alpha)}$ dependence merely becomes a multiplicative factor to the shape of the D_2 distribution in the collinear-soft haze region. This implies to N³LL logarithmic counting in the $e_2^{(\alpha)}$ logarithms, that the $e_2^{(\alpha)}$ spectrum is simply *multiplicative* to the normalized D_2 distribution. This is consistent with the arguments given in section 2.3 about the endpoints of the groomed and ungroomed D_2 distributions. The endpoint of the ungroomed D_2 distribution is set by the value of $e_2^{(\alpha)}$ at fixed order, so the functional dependence of the ungroomed D_2 spectrum is highly nontrivial. One would have to convolve the Sudakov resummation of the $e_2^{(\alpha)}$ spectrum with the ungroomed D_2 distribution as a function of $e_2^{(\alpha)}$ in order to accurately describe even the *normalized* endpoint of the ungroomed D_2 distribution. The grooming procedure decouples the shape of the endpoint from the value of $e_2^{(\alpha)}$, significantly simplifying the calculation of the D_2 distribution at large values of D_2. We explain in more detail the importance of these observations when considering the matching between resolved and unresolved limits in section 3.2.5.

As a check of the splitting function integration, we also compute the D_2 distribution with EVENT2 [60] and then match to the factorization formulae for the two-prong phase space regions.

3.2.2 Resolved substructure: collinear limit

Here, we will determine the factorization formula in the limit when the jet has two relatively hard collinear subjets. To derive this factorization formula, we must return to the parent theory of eq. (3.3) (for brevity, we just focus on one hemisphere):

$$\frac{d\sigma}{de_2^{(\alpha)}de_3^{(\alpha)}} = H(Q^2)S(z_{\text{cut}})J\big(e_2^{(\alpha)}, e_3^{(\alpha)}, z_{\text{cut}}\big). \tag{3.6}$$

Now, on this soft dropped jet on which we have measured $e_2^{(\alpha)}$, we additionally measure $e_3^{(\alpha)}$, with the assumption that $e_3^{(\alpha)} \ll (e_2^{(\alpha)})^3$. In this limit, and using the mode decomposition outlined in ref. [15], we can factorize the jet function into a hard, collinear splitting:

$$J\big(e_2^{(\alpha)}, e_3^{(\alpha)}, z_{\text{cut}}\big) \to H_2(z, e_2^{(\alpha)})J_1(e_3^{(\alpha)}) \otimes J_2(e_3^{(\alpha)}) \otimes C_s(e_3^{(\alpha)}, z_{\text{cut}}). \tag{3.7}$$

Here, z is the momentum fraction of one of the subjets, and $H_2(z, e_2^{(\alpha)})$ is a function that depends on $e_2^{(\alpha)}$ that describes the hard, collinear splitting.[6] $J_1(e_3^{(\alpha)})$ and $J_2(e_3^{(\alpha)})$ are the jet functions that describe the collinear radiation off of the two hard prongs in the splitting. $C_s(e_3^{(\alpha)}, z_{\text{cut}})$ is the collinear-soft function that describes relatively soft radiation emitted off of the dipole formed by the two hard prongs. In contrast to the ungroomed D_2 distribution, there is no global soft contribution (and thus for e^+e^-, no two-eikonal line soft function depending on $e_3^{(\alpha)}$), as the jet has already been isolated by the grooming procedure. The factorization formula in the two-prong collinear limit is then:

$$\frac{d^3\sigma}{dz\,de_2^{(\alpha)}\,de_3^{(\alpha)}} = H(Q^2)S(z_{\text{cut}})H_2(z, e_2^{(\alpha)})J_1(e_3^{(\alpha)}) \otimes J_2(e_3^{(\alpha)}) \otimes C_s(e_3^{(\alpha)}, z_{\text{cut}}). \tag{3.8}$$

We could stop with this factorization, and begin calculating the resummation of D_2; however, it is worthwhile to further analyze the structure of the collinear-soft function.

As $e_3^{(\alpha)} \to 0$, this forces the energy of the soft modes within $C_s(e_3^{(\alpha)}, z_{\text{cut}})$ to zero. All emissions generated off of the eikonal lines within the collinear-soft function can only contribute to the observable by being clustered with one of the legs of the hard prongs, before the legs themselves are clustered together. Otherwise, the emission will be at too low an energy scale and too wide of an angle to be included in the groomed jet. Correspondingly, emissions that do contribute to $e_3^{(\alpha)}$ from $C_s(e_3^{(\alpha)})$ cannot be emitted at too wide of an angle. If these emissions are not first clustered with one of the two hard prongs, then they are necessarily groomed away. Therefore, we separate out two angular regions of the collinear soft function and write:

$$C_s(e_3^{(\alpha)}, z_{\text{cut}}) \to C_s(e_3^{(\alpha)}, \theta < \theta_{ab})S_c(z_{\text{cut}}, \theta > \theta_{ab}) \otimes C_s^{\text{NG}}(e_3^{(\alpha)}, z_{\text{cut}}). \tag{3.9}$$

Again, we emphasize that the constraint $\theta < \theta_{ab}$ or $\theta > \theta_{ab}$ is schematic; the precise constraint will depend on the detailed clustering history, however it is purely geometrical

---

[6]While the momentum fraction of the subjets is not well-defined in the unresolved region, we may use a combination of energy correlation functions with different angular exponents to give a definition to z outside the two prong region; see ref. [15].
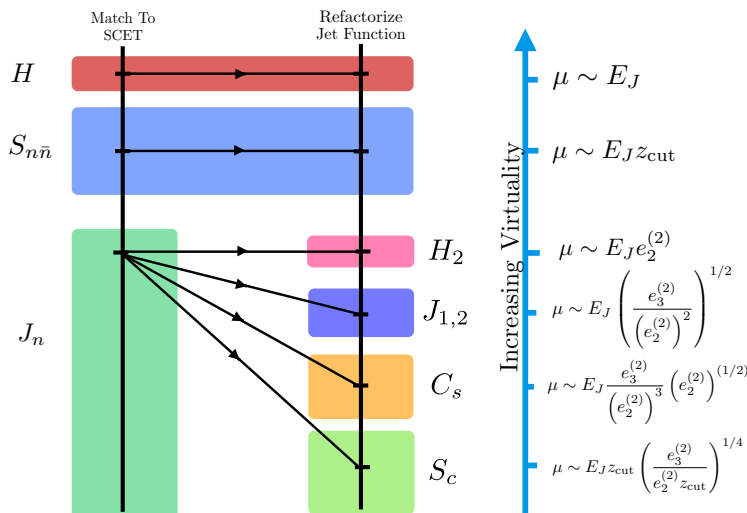
Figure 7. A schematic of the multistage matching procedure in the collinear limit of resolved substructure. The function incorporating non-global collinear effects is not shown, but is discussed in the text.

in its implementation. The last function is independent of the renormalization group, and encodes *local-to-the-jet* non-global correlations. The presence of the hard splitting with an opening angle set by $e_2^{(\alpha)}$ implies that an effective jet area is created within the two prong region. This will lead to non-global correlations between emissions that are groomed away, but emit into this opening angle, and the emissions which come off of the primary hard legs.

An illustration of the origin of these non-global logarithms (NGLs) is illustrated in figure 8. In this figure, a hard collinear quark and collinear gluon (denoted by the curly curve with a line through it) sets the mass of the jet, and then their dipole emits a soft-collinear gluon. This soft-collinear gluon has sufficiently low energy and fails soft drop, but re-emits another soft-collinear gluon that is clustered into the hard collinear particles. Such a re-emission is non-global in origin, as it is simultaneously sensitive to the infrared scales z_{cut} and $e_3^{(\alpha)}$. However, all particles in this picture are collinear, as the jet was already isolated from the rest of the event in the first stage of matching. Therefore the resulting NGLs only depend on the fact that the jet was initiated by a hard quark (in general, they depend on the flavor structure of the splitting). Because of this universality, these non-global logarithms are significantly less worrying than more standard NGLs that occur in (ungroomed) jet mass distributions. Techniques have been developed for systematic calculation of NGLs [55, 61–68], and these NGLs associated with the soft drop procedure have interesting features not previously encountered due to the clustering history. We have performed some preliminary estimations of these NGLs, and find their numerical effect is small, well within our uncertainties for the purely global (Sudakov) contributions. At leading logarithmic accuracy in the large-N_c limit, they can be computed using an extension of the Monte Carlo algorithm of Dasgupta and Salam [46], which is described in appendix E.
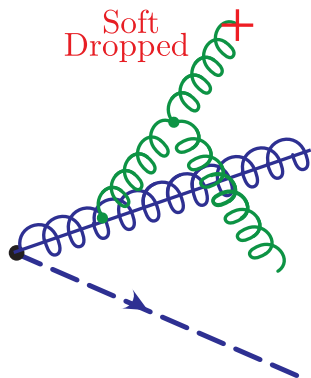
Figure 8. Schematic illustration of a configuration of emissions that contributes to collinear non-global logarithms. The soft-collinear gluon emission off of the hard-collinear gluon is groomed away, but not before it re-emitted into the ungroomed region between the two hard prongs.

With these replacements, the factorization formula now becomes

$$\frac{d^3\sigma}{dz\,de_2^{(\alpha)}\,de_3^{(\alpha)}} = H(Q^2)S(z_{\rm cut})S_c(z_{\rm cut},\theta > \theta_{ab})H_2(e_2^{(\alpha)}) \tag{3.10}$$

$$\times J_1(e_3^{(\alpha)}) \otimes J_2(e_3^{(\alpha)}) \otimes C_s(e_3^{(\alpha)},\theta < \theta_{ab}) \otimes C_s^{\rm NG}(e_3^{(\alpha)},z_{\rm cut})\,,$$

where the functions are as follows

- $H(Q^2)$ is the hard function, in this case for $e^+e^- \to$ dijets.

- $S(z_{\rm cut})$ is the soft function describing wide angle soft radiation which has been soft dropped.

- $S_c(z_{\rm cut},\theta > \theta_{ab})$ describes collinear soft radiation at $\theta > \theta_{ab}$.

- $H_2(e_2^{(\alpha)})$ is a hard function describing the production of the two collinear subjets.

- $J_{1,2}(e_3^{(\alpha)})$ are the jet functions for the collinear subjets.

- $C_s(e_3^{(\alpha)},\theta < \theta_{ab})$ describes collinear soft radiation emitted from the dipole formed from the subjets at $\theta < \theta_{ab}$.

- $C_s^{\rm NG}(e_3^{(\alpha)},z_{\rm cut})$ describes the entanglement between the groomed soft-collinear emissions and the two-prong region.

The hierarchy of scales of the functions in this region of phase space is illustrated in figure 7. The calculations of the functions in this factorization formula to one-loop accuracy are presented in appendix A. There, we also demonstrate the consistency of this factorization formula by showing that the sum of anomalous dimensions is indeed 0.
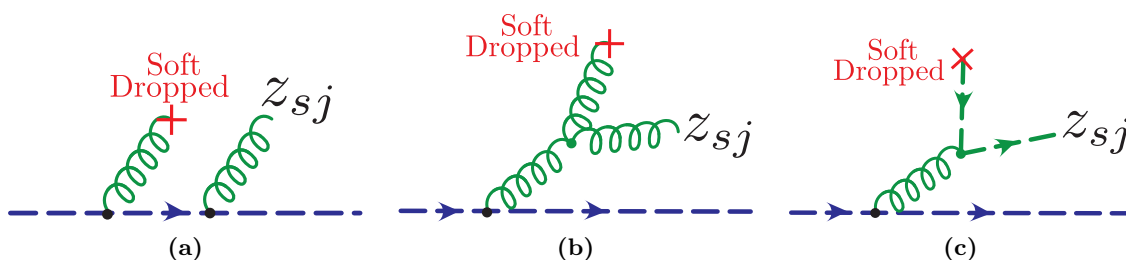
Figure 9. Structure of emissions at one-loop order that contribute to the hard matching function in the soft subjet factorization formula for groomed jets. The emission at a wider angle will be groomed away, while the other emission sets the mass of the jet. (a) corresponds to the Abelian emission of gluons, and will contribute proportional to C_F; (b) corresponds to non-Abelian gluon emissions, and will contribute proportional to C_A; and (c) corresponds to a gluon splitting to quarks, which contributes proportional to n_F.

3.2.3 Resolved substructure: soft limit

When the jet has two subjets whose energies are hierarchically separated we can determine the form of the appropriate factorization formula in the same manner as in the previous section. As in that case, we start with the parent jet function from eq. (3.3):

$$J\big(e_2^{(\alpha)}, de_3^{(\alpha)}, z_{\rm cut}\big) \to HC_{s2}(z, e_2^{(\alpha)}, e_3^{(\alpha)}, z_{\rm cut}) \otimes J_{sc}(e_3^{(\alpha)}) \otimes J_2(e_3^{(\alpha)})\,. \tag{3.11}$$

This preliminary factorization has removed the collinear subjet contributions, but has not distentangled all the soft scales. This requires a matching procedure that cannot be implemented at the level of the amplitude, but must be performed at the amplitude-squared.[7] The matching procedure is complicated by essentially the same physics that determines the NGLs encountered in the collinear-subjets of the resolved region, but now we must take the energy scale of one of the legs to be just above $z_{\rm cut}$, and hence not parametrically separated from the soft-collinear emissions which are groomed away just below $z_{\rm cut}$. Thus we write

$$HC_{s2}(z, e_2^{(\alpha)}, e_3^{(\alpha)}, z_{\rm cut}) \to H_2^{sj}(z, e_2^{(\alpha)}, z_{\rm cut}) C_s(e_3^{(\alpha)}, \theta < \theta_{ab}) \otimes C_s^{sj-{\rm NG}}(e_3^{(\alpha)}, z_{\rm cut})\,. \tag{3.12}$$

Where $C_s^{sj-{\rm NG}}(e_3^{(\alpha)}, z_{\rm cut})$ denotes hard matching contributions where additional Wilson lines and jet functions are introduced to capture the non-global correlations. The function $C_s(e_3^{(\alpha)}, \theta < \theta_{ab})$ is the same as found in eq. (3.10). As it stands, this factorization is sufficient to resum all large global (Sudakov) logarithms, and to leading logarithmic accuracy in the NGLs, the function $C_s^{sj-{\rm NG}}(e_3^{(\alpha)}, z_{\rm cut})$ is identical to that found in eq. (3.10).[8]

There are a few things to note about this factorization formula. First, there is a jet function $J_{sc}(e_3^{(\alpha)})$ that describes collinear radiation off of the soft subjet in the larger jet. To leading power, this subjet is always a gluon and is identical to the $z \to 0$ limit of the

---

[7]This is similar in spirit to ref. [62].

[8]At higher orders we would have to keep track of the color correlations between multiple directions at the soft subjet scale being integrated out and groomed away, and the soft emissions into the opening angle of the $1 \to 2$ splitting.

corresponding gluon jet function in the factorization formula in the case of hard, collinear subjets. Additionally, there is the identical collinear-soft function as in the hard collinear subjets factorization formula. Because we assume that $D_2^{(\alpha)} \ll z_{\text{cut}}$, emissions that set the value of $D_2^{(\alpha)}$ must be at parametrically lower energies than either of the subjets. Therefore, the soft drop constraint on these emissions is just a geometric constraint that enforces the emissions to first cluster with one of the hard subjets. This geometrical constraint is necessarily independent of the energy of the subjets of the larger jet, and therefore this collinear-soft function is identical to that which appears in eq. (3.10).

The novel part of this factorization formula is the hard matching function, $H_2^{sj}(z, e_2^{(\alpha)}, z_{\text{cut}})$ that describes the production of the soft subjet. This function has now two contributions relevant for a NLL resummation, or one-loop calculation. First, there are the standard virtual contributions, which just correspond to the $z \to 0$ limit of the corresponding matching coefficient in eq. (3.10). There is, however, a new contribution to the matching function in this factorization formula. Because we apply soft drop, it is possible that there is an initial emission in the jet that fails soft drop, and so does not seed the production of a soft subjet. However, a secondary emission could then pass soft drop, and produce the soft subjet. These different configurations are shown in figure 9.

The calculation of this two emission contribution to the hard matching function $H_2^{sj}(z, e_2^{(\alpha)}, z_{\text{cut}})$ is presented in appendix B.1, but we will describe its features here. We must consider all possible pairs of soft emissions which are reclustered in such a way that the first angular-ordered emission fails soft drop, while the second passes. This is the reason why we explicitly show the z_{cut} dependence in this function. Note that the constraint that one emission fails soft drop while the other passes eliminates the collinear singularity when the emissions become close in angle. If the two emissions are sufficiently close in angle compared to their collective angle to the hard jet core, then they will be clustered together first, which is forbidden by assumption. This implies that the contribution to this hard function from the emission of a soft quark-anti-quark pair does not contribute to NLL order. The emission of soft gluons will contribute at NLL order. Figure 9 shows a schematic picture of these two-emission contributions to the hard matching function.

We therefore find that the complete factorization formula for a soft dropped groomed jet with a soft subjet is

$$\frac{d^3\sigma}{dz\,de_2^{(\alpha)}\,de_3^{(\alpha)}} = H(Q^2)S(z_{\text{cut}})H_2^{sj}(z, e_2^{(\alpha)}, z_{\text{cut}})C_s(e_3^{(\alpha)}, \theta < \theta_{ab}) \otimes J_{sc}(e_3^{(\alpha)}) \otimes J(e_3^{(\alpha)}) \otimes C_s^{sj,\text{NG}}(e_3^{(\alpha)}, z_{\text{cut}}).$$

(3.13)

A brief review of the different functions entering the factorization is as follows

- $H(Q^2)$ is the hard function, in this case for $e^+e^- \to$ dijets.

- $H_2^{sj}(z, e_2^{(\alpha)}, z_{\text{cut}})$ is a hard function describing the production of the soft subjet.

- $J(e_3^{(\alpha)})$ is the jet function for the energetic jet.

- $J_{sc}(e_3^{(\alpha)})$ is the jet function for the soft jet.

- $C_s(e_3^{(\alpha)}, \theta < \theta_{ab})$ is the collinear-soft function describing the radiation entering the dipole off of the primary eikonal lines.
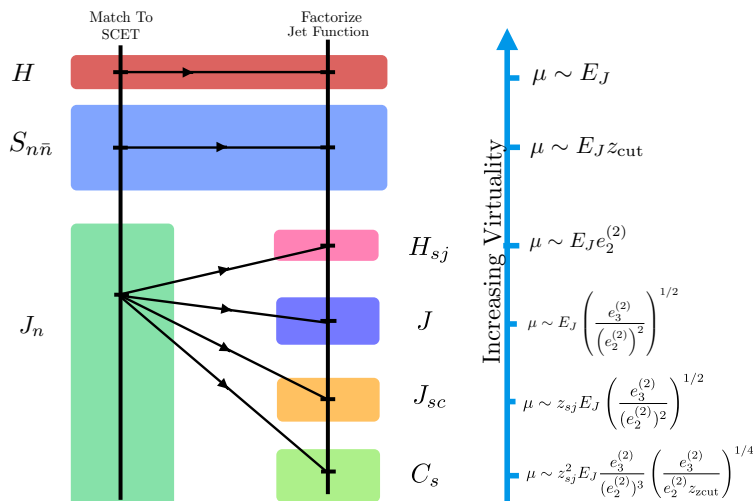
Figure 10. A schematic of the multistage matching procedure in the soft-collinear subjet region of phase space. The function incorporating non-global collinear effects is not shown, but is discussed in the text.

- $C_s^{sj,\text{NG}}(e_3^{(\alpha)}, z_{\text{cut}})$ describes the non-global correlations arising from groomed soft-collinear emissions.

- $S(z_{\text{cut}})$ is the soft function describing wide angle soft radiation which has been soft dropped.

For future use, we record the virtualities of the different modes, which are also shown in figure 10:

$$
\mu_{sj} \sim z_{sj} E_J \left(\frac{e_3^{(\alpha)}}{(e_2^{(\alpha)})^2} \right)^{1/\alpha} , \qquad \mu_{cs} \sim z_{sj}^2 E_J \frac{e_3^{(\alpha)}}{(e_2^{(\alpha)})^3} \left(\frac{e_3^{(\alpha)}}{e_2^{(\alpha)} z_{\text{cut}}} \right)^{1/(2\alpha)} . \tag{3.14}
$$

The scalings of these modes will play an important role when studying the behavior of non-perturbative power corrections. Note that while the functions in the factorization theorems in the collinear and soft subjet regions are similar, the hierarchy of scales is different. These differences are present in figures 7 and 10, and are related to the different sensitivities to infrared scales in the two factorization theorems. For example, the $J_{sc}(e_3^{(\alpha)})$ and $C_s^{sj,\text{NG}}(e_3^{(\alpha)}, z_{\text{cut}})$ functions in figure 10 are explicitly sensitive to the soft energy of the subjet, while no such dependence exists in the functions of the hard, collinear subjet factorization.

3.2.4 Merging collinear and soft resolved limits

To perform a complete calculation, we must merge our description of the different resolved regions. We merge between the soft-collinear subjet and collinear subjet region by

subtracting their overlap. This gives:

$$\frac{d^3\sigma}{dz\,de_2^{(\alpha)}\,de_3^{(\alpha)}} = H(Q^2)S(z_{\text{cut}})\Bigg[S_c(z_{\text{cut}},\theta>\theta_{ab})C_s^{\text{NG}}(e_3^{(\alpha)},z_{\text{cut}})\Big[H_2(z,e_2^{(\alpha)})-H_2(e_2^{(\alpha)},z\to z_{\text{cut}})\Big]$$

$$+H_2^{sj}(z,e_2^{(\alpha)},z_{\text{cut}})C_s^{sj,\text{NG}}(e_3^{(\alpha)},z_{\text{cut}})\Bigg]\otimes C_s(e_3^{(\alpha)},\theta<\theta_{ab})\otimes J_{sc}(e_3^{(\alpha)})\otimes J(e_3^{(\alpha)}). \qquad (3.15)$$

However, to NLL accuracy, including NGLs, we can show that the collinear factorization suffices to capture all large logarithms with the appropriate scale setting. First we note that the tree-level results of the subtracted hard matching of the collinear factorization and the soft-collinear subjet agree. Then all one must check is that the resummation in the collinear sector arising from running $S_c(z_{\text{cut}},\theta>\theta_{ab})$ naturally merges with the resummation in the soft-collinear sector of the function $H_2^{sj}(z,e_2^{(\alpha)},z_{\text{cut}})$. For this to happen, the natural renormalization scales of product:

$$S_c(z_{\text{cut}},\theta>\theta_{ab})H_2(z,e_2^{(\alpha)}), \qquad (3.16)$$

must merge to the natural renormalization scale found in $H_2^{sj}(z,e_2^{(\alpha)},z_{\text{cut}})$ when $z\to z_{\text{cut}}$. This is accomplished so long as we use the transverse momentum of the collinear splitting as the renormalization scale for the collinear hard splitting function. We then compare the scales take from appendix A.10 (for simplicity, we take $\alpha=2$):

$$\mu_{H_2}^2 = \frac{z(1-z)e_2^{(2)}Q^2}{4}, \qquad (3.17)$$

$$\mu_{S_c}^2 = \frac{z_{\text{cut}}^2 e_2^{(2)}Q^2}{4z(1-z)}. \qquad (3.18)$$

In the limit $z\to z_{\text{cut}}$, we have:

$$\mu_{H_2}^2 \to \mu_{S_c}^2, \qquad (3.19)$$

showing that the two scales merge.

Finally, we note that the sum of the anomalous dimensions in eq. (A.62) gives eq. (B.7) in the limit $1-z_q\to z_{\text{cut}}$, that is, the collinear subjets approach the soft-collinear region. That this must be the case stems from the purely geometrical character of the soft drop constraint. Regardless of the relative energy scales between the emissions that sets the $1\to 2$ splitting, and the emission which fails soft drop, once all additional emissions are required to fail on their own, whether or not it can contribute to D_2 depends on whether it is clustered into the hard splitting, that is, the angular structure of the emissions. Indeed, we exploit this fact to simplifiy the calculation of the collinear-soft subjet matching presented in appendix B. Thus to NLL accuracy, the merging of the collinear and soft resolved limits is accomplished by simply running the splitting scale to the transverse momentum of the collinear splitting that sets $e_2^{(2)}$, and only using the collinear factorization formula for the resummation. Note that the analogous simplification could not be made in the case of the ungroomed D_2 distribution, mainly due to the presence of boundary soft modes in the soft-subjet factorization.

3.2.5 Matching resolved and unresolved limits

In this section we discuss how the factorization formulae in the resolved and unresolved limits can be merged to provide a complete description of the entire D_2 distribution. We consider two distinct merging schemes: one using profile functions [69, 70] to turn off the resummation at the endpoint of the distribution, and a second using only canonical scales for the resummation at all values of D_2, never turning off the resummation. We scale set at the level of the cumulative distribution, and take the derivative for the differential cross section. When using profiles, we retain all the logarithms of μ over the natural scale of the function in the matching, jet, and soft-collinear/collinear-soft functions to order $\mathcal{O}(\alpha_s)$.[9] Thus when we turn off the resummation by taking all scales to be the factorization scale, we are left with the singular terms of the fixed-order D_2 distribution. When resummation is fully turned on by the profile function, this trivializes the contribution from the factorized functions. The specific profile used and the canonical scale choices are summarized in appendix A.10.

For all distributions (quark, gluon, and signal), the use of canonical scales in the resummation gives a resummed distribution that completely over-shoots the singular terms of the fixed order result throughout the range of D_2, leading to an unphysical endpoint of the distribution much greater than $1/(2z_{\rm cut})$. If one were to additively match and normalize, the resulting curve would be equivalent to just the normalized resummed canonical prediction, with completely unphysical behavior in the large D_2 region, and with a peak much too low due to the broad tail. Thus we adopt a strategy of multiplicative matching for the distribution

$$\frac{d\sigma}{de_2^{(2)}de_3^{(2)}}\bigg|_{\rm matched} = \frac{d\sigma}{de_2^{(2)}de_3^{(2)}}\bigg|_{\rm resum} \left(\frac{\frac{d\sigma}{de_2^{(2)}de_3^{(2)}}\big|_{\rm fo}}{\frac{d\sigma}{de_2^{(2)}de_3^{(2)}}\big|_{\rm resum\ fo}}\right). \qquad (3.20)$$

Here "fo" stands for the fixed-order distribution, which is determined from the $1 \to 3$ splitting functions (as discussed in appendix D) or from EVENT2 [60]. Thus, regardless of whether we use canonical scale choices in the resummed distribution or profiles to turn off the resummation, the distribution will always terminate at the physical value $1/(2z_{\rm cut})$.

The only subtlety is if the singular distribution has a zero in the physical range of D_2. This occurs in some cases, and we are then forced to only use distributions where the resummation is turned off via profiles before this zero is reached. We find this to be the case generically for the signal distributions if $z_{\rm cut} < 0.2$, and for the quark distribution if $z_{\rm cut} \leq 0.05$.

It is worthwhile to understand how the merging interplays with the resummation of the $e_2^{(\alpha)}$ spectrum, and the counting of logs of D_2 versus logs of $e_2^{(\alpha)}$. As can be directly seen from appendix D, the D_2 spectrum in the large D_2 region, which is controlled by the factorization in the soft-haze region of section 3.2.1, is *independent* of the value of $e_2^{(\alpha)}$. Since the D_2 spectrum at leading order is set by the two-loop matrix elements in the soft-haze region, we may write

$$\frac{d\sigma}{de_2^{(\alpha)}dD_2} = \frac{d\sigma}{de_2^{(\alpha)}}\bigg|_{\rm N^3LL} \times \left[F(D_2, z_{\rm cut})\right]_+ + O\left(\alpha_s^4 \log e_2^{(\alpha)}\right), \qquad (3.21)$$

---

[9]If we had also retained the constants, this would be equivalent to NLL′.

where $F(D_2, z_{\text{cut}})$ reproduces the fixed order spectrum in D_2. The subscript N³LL indicates that this expression is valid up to N³LL order. Although the fixed order distribution diverges at $D_2 = 0$, the plus distribution ensures that the singularity at $D_2 = 0$ is formally cancelled by the appropriate virtual corrections, so that we have

$$\int_0^\infty dD_2 \big[F(D_2, z_{\text{cut}})\big]_+ = 1\,. \tag{3.22}$$

The factor multiplying the fixed order D_2 spectrum is simply the groomed $e_2^{(\alpha)}$ spectrum to N³LL accuracy. Once we match the resummed D_2 spectrum to the fixed-order D_2 spectrum, we replace

$$\big[F(D_2, z_{\text{cut}})\big]_+ \to F^{\text{matched}}(D_2, z_{\text{cut}}, Qe_2^{(\alpha)})\,. \tag{3.23}$$

The matched function satisfies the properties:

$$\int_0^\infty dD_2 F^{\text{matched}}(D_2, z_{\text{cut}}, Qe_2^{(\alpha)}) = 1\,, \tag{3.24}$$

$$F^{\text{matched}}(D_2, z_{\text{cut}}, Qe_2^{(\alpha)}) \to F(D_2, z_{\text{cut}}) \text{ as } D_2 \to \frac{1}{2z_{\text{cut}}}\,. \tag{3.25}$$

The resummation ensures the integrability of the matched distribution, and the matching ensures that in the region of validity of the soft-haze factorization, we reproduce the soft-haze spectrum. Thus we may simply replace the plus-distribution for the fixed order result within eq. (3.21)

$$\frac{d\sigma}{de_2^{(\alpha)} dD_2} = \frac{d\sigma}{de_2^{(\alpha)}}\bigg|_{\text{N}^3\text{LL}} \times F^{\text{matched}}(D_2, z_{\text{cut}}, Qe_2^{(\alpha)}) + O\Big(\alpha_s^4 \log e_2^{(\alpha)}\Big)\,. \tag{3.26}$$

This result is still valid to the same logarithmic accuracy in the log counting for the $e_2^{(\alpha)}$ spectrum, while maintaining the correct sum rules on the D_2 variable, and gives the correct shape of the end-point of the distribution where the soft-haze factorization applies. Since the resummation of the groomed $e_2^{(\alpha)}$ spectrum is multiplictative to the D_2 spectrum, we can correctly predict the shape of the distribution for both large and small D_2 without resumming any logs of $e_2^{(\alpha)}$ to at least N⁴LL accuracy, which is far beyond the practically achievable accuracy.

We stress that such a simple matching procedure does not work when considering the ungroomed D_2 distribution. For the ungroomed distribution in the soft haze region, we are forced to write

$$\frac{d\sigma}{de_2^{(\alpha)} dD_2} = \frac{d\sigma}{de_2^{(\alpha)}}\bigg|_{\text{N}^3\text{LL}} \otimes \big[F(D_2, e_2^{(\alpha)})\big]_+ + O\Big(\alpha_s^4 \log e_2^{(\alpha)}\Big)\,, \tag{3.27}$$

where we now have a *convolution* in the $e_2^{(\alpha)}$ variable, denoted by the $\otimes$. We may still replace the fixed order distribution in D_2 with the matched distribution, both normalized to obey the correct D_2 sum rule, but now we must perform a convolution in $e_2^{(\alpha)}$! Given that the endpoint of the ungroomed distribution behaves as the inverse of $e_2^{(\alpha)}$, performing such a convolution would be daunting and computationally expensive, since at each $e_2^{(\alpha)}$ value, one would need to calculate the full matched and normalized D_2 distribution.
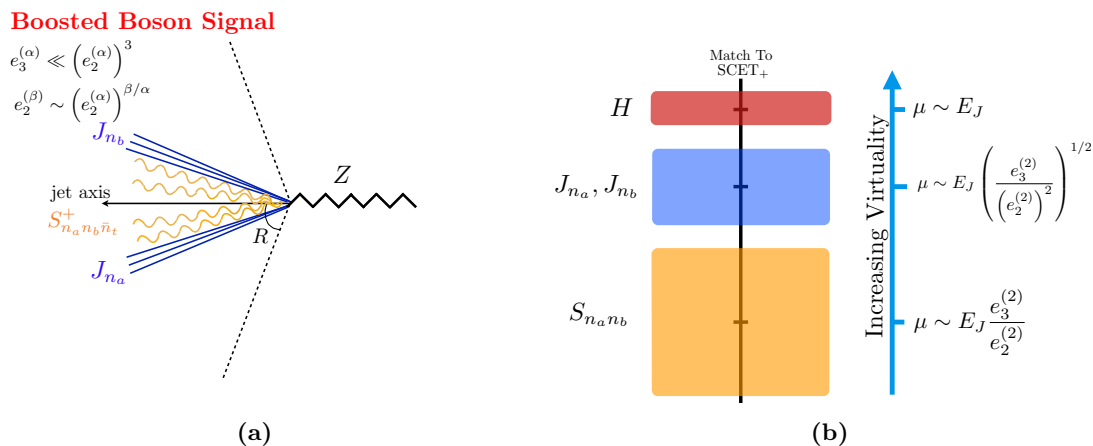
Figure 11. The collinear subjets configuration for a boosted color singlet decay is shown in (a). The structure of factorization formula is shown in (b). Figure from [15].

3.2.6 Signal jets

In this section, we give the effective field theory description for a groomed hadronically-decaying color singlet, which we take for concreteness to be a Z boson:

$$\frac{d^3\sigma}{dz\,de_2^{(\alpha)}de_3^{(\alpha)}} = H(Q^2)H_2^{Z\to q\bar{q}}(e_2^{(\alpha)}, m_Z^2)J_a(e_3^{(\alpha)}) \otimes J_b(e_3^{(\alpha)}) \otimes S(e_3^{(\alpha)}, z_{\text{cut}}). \quad (3.28)$$

A brief description of the functions appearing in eq. (3.28) is as follows:

- $H(Q^2)$ is the hard function describing the production of the on-shell Z boson.

- $H_2^{Z\to q\bar{q}}(z, e_2^{(\alpha)}, m_Z^2)$ is a hard function describing the decay of the Z boson into a $q\bar{q}$ pair.

- $J_a(e_3^{(\alpha)})$, $J_b(e_3^{(\alpha)})$ are the jet functions describing the two collinear subjets.

- $S(e_3^{(\alpha)}, z_{\text{cut}})$ is the collinear-soft function describing the radiation emitted from the $q\bar{q}$ dipole.

The factorization formula, and the region of phase space it describes is shown schematically in figure 11. One-loop calculations are given in appendix C.

The factorization formula of eq. (3.28) is valid when $e_3^{(\alpha)} \ll (e_2^{(\alpha)})^3$. The structuring of the radiation within $S(e_3^{(\alpha)}, z_{\text{cut}})$ is similar to the collinear-soft function in eq. (3.9), and can also be refactorized similarily, except that there is no third Wilson line corresponding to the hard recoil direction of the jet. When $e_3^{(\alpha)} \gtrsim (e_2^{(\alpha)})^3$, one must match to the full $Z \to q\bar{q}g$ matrix element.

4 Factorized cross section in pp collisions

In this section we will discuss the extension of the e^+e^- factorization formulae of section 3.2 to pp collisions. In particular, we show that for phenomenologically relevant parameters for

the jet mass and p_T, the assumptions of the factorization formula hold, and no new ingredients are required to extend the factorization formula to pp. The only process dependence is carried by the quark, anti-quark and gluon fractions of the process. This will follow straightforwardly from the universality of collinear factorization and the fact that all the factorization formulae of section 3.2 were obtained through a refactorization of the jet or collinear-soft functions. For concreteness, in this section we will consider the factorization for the groomed D_2 observable in $pp \to Z + j$. We identify the highest p_T jet satisfying $|\eta_J| < \eta_{\max}$, groom it with the mMDT/soft drop algorithm, and then measure D_2 on the groomed jet. It is important to emphasize that here we are completely inclusive over additional hadronic activity throughout the event. We do not need to apply any form of veto on out-of-jet radiation, as is sometimes imposed to study ungroomed jet mass (for example, see ref. [71]).

Since our factorization formula for the groomed D_2 observable is obtained as a refactorization of the cross section for mMDT/soft drop groomed $e_2^{(\alpha)}$ observable, we begin by summarizing its factorization in pp collisions. In ref. [34] it was shown that in the region where the factorization formula applies, namely $e_2^{(\alpha)} \ll z_{\mathrm{cut}} \ll 1$, the cross section can be written as

$$\frac{d\sigma^{pp}}{de_2^{(\alpha)}} = \sum_{k=q,\bar{q},g} D_k(p_T^{\min}, y_{\max}, z_{\mathrm{cut}}, R) S_{C,k}(z_{\mathrm{cut}}, e_2^{(\alpha)}) \otimes J_k(e_2^{(\alpha)}) \,. \tag{4.1}$$

Here, it is important to emphasize that since we are inclusive over hadronic activity in the event, a strict factorization into jet and soft functions does not apply. Indeed, it is clear that this must be the case, since the number of jets in the event is not fixed. Nevertheless, eq. (4.1) shows that all dependence on the rest of the event can be absorbed into a process dependent normalization factor D_k, which does not depend on the $e_2^{(\alpha)}$ observable. In general, D_k depends on the minimum p_T cut, the jet radius R, rapidity cuts, parton distributions, z_{cut}, etc. The $e_2^{(\alpha)}$ observable is set by universal collinear physics described by the convolution between the collinear-soft function and the jet function. Since these are collinear matrix elements, they depend only on the collinear dynamics of the particular jet in question, and are independent of other jets in the event. In particular, global color correlations are absent.

The D_k functions depend on the parton flavor, which must be summed over, an added complication of jets in pp collisions. While parton flavor is not in general an IRC safe quantity, due to the fact that soft partons can radiate flavor into or out of the jet, it was shown in ref. [34] that the parton flavor can be defined on soft dropped jets in the limit $e_2^{(\beta)} \ll z_{\mathrm{cut}} \ll 1$, where the factorization formula applies. On a soft dropped jet, we can define the flavor of the jet as

$$f_J = \sum_{i \in J_{\mathrm{SD}}} f_i \,, \tag{4.2}$$

where $f_q = 1$, $f_{\bar{q}} = -1$, $f_g = 0$, and J_{SD} indicates the constituents of the jet after the soft drop algorithm has been applied. If $f_J = \pm 1$, then the jet is defined as quark type, while if $f_J = 0$, the jet is defined as gluon type. In the normalized distribution, the D_k can therefore be interpreted as quark, anti-quark, and gluon jet fractions in the event sample

under consideration, and can easily be extracted from fixed-order Monte Carlo codes, such as MCFM [72–74].

The factorization of D_2 in pp collisions now follows trivially from combining the factorization formula of eq. (4.1) for soft dropped $e_2^{(\alpha)}$ with $e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1$ with the factorization formulae derived for $e^+ e^-$ in section 3.2. To proceed, starting from eq. (4.1), we refactorize the jet and collinear-soft functions, as appropriate. This also implies that the same process dependent functions D_k also appear in the expression of the cross section of D_2. We can then write

$$\frac{d^2 \sigma^{pp,\text{coll}}}{de_2^{(\alpha)} de_3^{(\alpha)}} = \sum_{k=q,\bar{q},g} D_k \left[H_2(e_2^{(\alpha)}) C_s(e_3^{(\alpha)}) \otimes J_1(e_3^{(\alpha)}) \otimes J_2(e_3^{(\alpha)}) \otimes S_c(e_3^{(\alpha)}, z_{\text{cut}}) \right] , \quad (4.3)$$

in the collinear subjets region of phase space,

$$\frac{d^2 \sigma^{pp,\text{c-soft}}}{de_2^{(\alpha)} de_3^{(\alpha)}} = \sum_{k=q,\bar{q},g} D_k \left[H_2^{sj}(e_2^{(\alpha)}, z_{\text{cut}}) C_s(e_3^{(\alpha)}) \otimes J_{sc}(e_3^{(\alpha)}) \otimes J(e_3^{(\alpha)}) \right] , \quad (4.4)$$

in the soft subjet region of phase space, and

$$\frac{d^3 \sigma^{pp,\text{cs haze}}}{de_2^{(\alpha)} de_2^{(2\alpha)} de_3^{(\alpha)}} = \sum_{k=q,\bar{q},g} D_k \left[J(e_2^{(\alpha)}) \otimes S_{sc}(e_2^{(\alpha)}, e_2^{(2\alpha)}, e_3^{(\alpha)}, z_{\text{cut}}) \right] , \quad (4.5)$$

in the collinear-soft haze region of phase space.

Importantly, since the same D_k factor appears in each of the factorization formulae in the different regions of phase space, we can then perform the marginalization separately over the different factorization formulae. We can therefore write, for the normalized distribution when summed over the factorization formulae:

$$\frac{d\sigma^{pp,\text{norm}}}{dD_2^{(\alpha)}} = \sum_{k=q,\bar{q},g} \kappa_k \frac{d\sigma_k^{pp,\text{norm}}}{dD_2^{(\alpha)}} , \quad (4.6)$$

where the κ_k can be interpreted as the fraction of jets in the sample with flavor k.

5 Consequences of factorization formulae

Given the factorization formulae developed in the previous sections, there are several fascinating consequences that immediately follow. Several of these have been noted before (see ref. [34]), and are consequences of the fact that mMDT or soft drop removes soft, wide angle radiation in a jet from contributing to the observables of interest. Here, we will briefly mention these general properties of mMDT and soft drop grooming, and discuss in some detail features that are new to measuring D_2 on these groomed jets.

The absence of soft, wide angle radiation in the jet eliminates event-wide color correlations and NGLs of the groomed jet observables to all orders in α_s. With the relative scaling that we have assumed between the two-point energy correlation function and z_{cut}, $e_2^{(\alpha)} \ll z_{\text{cut}} \ll 1$, all radiation that remains in the jet after grooming must be collinear. Assuming collinear factorization, this then implies that the shape of the mass distribution

– 28 –

is independent of the process that created that jet, up to the relative fraction of quark and gluon jets in the sample. The quark and gluon groomed jet fractions are well-defined to leading power in $e_2^{(\alpha)}$ and $z_{\rm cut}$, and can be determined from fixed-order codes. Because the measurement of D_2 is more differential than just the groomed jet mass, all of these properties continue to hold in that case.

Additionally, the mMDT groomed D_2 distribution enjoys other properties that actually make its perturbative distribution more well-defined and robust than the jet mass. Because the cut on the groomed jet mass can be tuned to satisfy $e_2^{(2)}E_J^2 = m_J^2 \ll z_{\rm cut}E_J^2 \ll E_J^2$, perturbative power corrections to the D_2 distribution can formally be made arbitrarily small. Additionally, because soft, wide angle emissions do not contribute to the groomed observables, non-perturbative corrections are suppressed by powers of the ratio of $\Lambda_{\rm QCD}$ to the groomed jet mass. These make D_2 a good candidate for QCD studies at the LHC, and therefore we will discuss these points in some detail.

5.1 Universality of the shape of the D_2 distribution

Typically, resummation is only important in a restricted region of the distribution of a particular observable. For example, soft and collinear emissions dominate the hadronic final state of e^+e^- collisions when an appropriately chosen event shape, such as thrust [75], is small. In the case of soft drop groomed jet mass, radiation in the jet is constrained to be collinear if $m_J^2 \ll z_{\rm cut}E_J^2 \ll E_J^2$; however, this is not the whole allowed phase space. There are regions where $m_J^2 \gtrsim z_{\rm cut}E_J^2$, which are vital to describe correctly to claim a precision description of the distribution.

The entire mMDT/soft drop groomed D_2 distribution, however, enjoys a universality. First, requiring the groomed jet mass to satisfy $m_J^2 \ll z_{\rm cut}E_J^2 \ll E_J^2$, all radiation that remains in the jet is collinear. At this stage, both $z_{\rm cut}$ and m_J are fixed. Then, with this configuration, we measure D_2 on the remaining constituents of the jet. All remaining emissions in the jet are necessarily collinear, and so any measured value of D_2 of these groomed jets is well-described just by resummation. Perturbative power corrections beyond the resummation (non-singular contributions) are small, and can be made arbitrarily small in perturbation theory by going further into the regime where $m_J^2 \ll z_{\rm cut}E_J^2 \ll E_J^2$. Note that this property requires that we restrict the jet mass appropriately and then measure D_2, an observable which resolves further substructure of the jet.

For applications to the LHC, it is interesting to briefly consider the values of the jet mass and p_T for which our factorization formula, and therefore this universality, holds. Observables such as D_2 are used at the LHC both to identify hadronically decaying $W/Z/H$ bosons, as well as to search for new light particles with $m \lesssim m_Z$ which decaying hadronically [76]. For many of these searches, the bulk of the data is for $p_T > 500\,{\rm GeV}$, and extends up to approximately $p_T \sim 1000\,{\rm GeV}$. Using the condition $e_2^{(2)} \ll z_{\rm cut} \ll 1$, with $z_{\rm cut} = 0.1$, we expect that our factorization will begin to break down around $p_T = 500\,{\rm GeV}$ for $m_J \sim m_Z$, if the value of $z_{\rm cut} = 0.1$ is fixed. For lower values of p_T, one will become sensitive again to global color correlations from emissions with energy fraction greater than $z_{\rm cut}$, which do not fail the soft drop criteria, and can contribute to the observable. Taking as a concrete example a bin from $p_T = 600 - 800\,{\rm GeV}$ in which the D_2 observable has been

measured by ATLAS [77], for a jet mass of $m_J \sim m_Z$, this has $e_2^{(2)} \lesssim 0.02$. For $z_{\rm cut} = 0.1$ the assumptions of our factorization formula safely hold. For lighter particles the p_T range can be extended, or alternatively, the expansion parameter is smaller. We therefore find that our factorization applies for most of the p_T range of phenomenological interest, and therefore so do our conclusions regarding the universality of the distribution. We believe that this understanding of universality derived from the factorization formula is one of the most important outcomes of our analysis.

5.2 Hadronization corrections suppressed by perturbative jet mass

The dominant non-perturbative corrections to a factorization formula arise from modes whose virtualities approach $\Lambda_{\rm QCD}$. A simple estimate of the size or importance of these non-perturbative effects follows from determining the value of the observable at which the lowest virtuality mode becomes comparable to $\Lambda_{\rm QCD}$. The mode with the lowest virtuality often corresponds to soft, wide angle emissions. So, by grooming them away with mMDT or soft drop, we can significantly reduce the effect of non-perturbative corrections and render the perturbative distribution more robust.

To see this for $D_2^{(\alpha)}$ on groomed jets, we first review the size of non-perturbative corrections in the ungroomed case. For concreteness, we will focus on the non-perturbative corrections to the collinear subjets factorization formula. In the ungroomed case the lowest virtuality mode is that of soft, wide-angle radiation; see section 3.1.2. Its virtuality was identified in ref. [15] and is

$$\mu_S \simeq \frac{e_3^{(\alpha)}}{e_2^{(\alpha)}} E_J \simeq (e_2^{(\alpha)})^2 D_2^{(\alpha)} E_J \,, \tag{5.1}$$

where we have expressed $e_3^{(\alpha)}$ in terms of $e_2^{(\alpha)}$ and $D_2^{(\alpha)}$. Setting $\mu_S = \Lambda_{\rm QCD}$, we find that non-perturbative effects dominate when

$$D_2^{(\alpha)}\Big|_{\rm np} \simeq \frac{\Lambda_{\rm QCD}}{(e_2^{(\alpha)})^2 E_J} \,. \tag{5.2}$$

If we take $\alpha = 2$ for concreteness, this can be rewritten in terms of the jet mass and energy as

$$D_2^{(2)}\Big|_{\rm np} \simeq \frac{\Lambda_{\rm QCD} E_J^3}{m_J^4} \,. \tag{5.3}$$

Therefore, perhaps surprisingly, as the jet energy increases for a fixed jet mass, non-perturbative corrections increase significantly. If we assume that $E_J = 500\,{\rm GeV}$, $m_J = 100\,{\rm GeV}$, and take $\Lambda_{\rm QCD} = 1\,{\rm GeV}$, then $D_2^{(2)}\Big|_{\rm np} \simeq 1$. That is, we expect non-perturbative physics to dominate right at the boundary between where one- and two-prong jets live in the $D_2^{(\alpha)}$ distribution.

Now, let's do the same analysis but for the mMDT/soft drop D_2 cross section. The lowest virtuality mode that appears in any factorization formula is the collinear-soft radiation of the collinear-soft subjet factorization formula; see section 3.2.3. The virtuality of

this mode is

$$\mu_{cs} \sim E_J \frac{z_{\text{cut}}^2 e_3^{(\alpha)}}{(e_2^{(\alpha)})^3} \left(\frac{e_2^{(\alpha)}}{z_{\text{cut}}} \right)^{1/\alpha} = E_J z_{\text{cut}}^2 D_2^{(\alpha)} \left(\frac{e_2^{(\alpha)}}{z_{\text{cut}}} \right)^{1/\alpha} . \tag{5.4}$$

Setting $\mu_{cs} = \Lambda_{\text{QCD}}$, non-perturbative effects dominate this mode when

$$\left. D_2^{(\alpha)} \right|_{\text{np},cs} \simeq \frac{1}{(e_2^{(\alpha)})^{1/\alpha} z_{\text{cut}}^{2-1/\alpha}} \left(\frac{\Lambda_{\text{QCD}}}{E_J} \right) . \tag{5.5}$$

As before, taking $\alpha = 2$ for concreteness, we find that non-perturbative effects dominate when

$$\left. D_2^{(2)} \right|_{\text{np},cs} \simeq \frac{\Lambda_{\text{QCD}}}{z_{\text{cut}}^{3/2} m_J} . \tag{5.6}$$

This result is quite remarkable. Without grooming, non-perturbative effects for $D_2^{(2)}$ become larger, for a fixed jet mass cut, as the energy of the jet is increased. However, by grooming the jet with mMDT or soft drop, non-perturbative corrections are *independent of the jet energy!* Physically this arises since after grooming the jet behaves loosely like a boosted event shape, and it is the jet mass that sets the scale. As long as the mass cut on the jet is perturbative, hadronization corrections are highly suppressed. Importantly, the distribution is perturbative well below $D_2^{(2)} \sim 1$, into the region where two-prong jets live. Taking the numerical values of $z_{\text{cut}} = 0.1$, $\Lambda_{\text{QCD}} = 1\,\text{GeV}$, and $m_J = m_Z$, we find that dominant non-perturbative correction arises from the soft dropped soft subjet region of phase space, and we can estimate that non-pertubative effects becomes important at $D_2^{(2)} \sim 0.35$. Non-perturbative corrections for the other regions of phase space in the factorization formulae are further suppressed, and so are ignored. A more detailed study of non-perturbative effects for the D_2 distribution is performed in a companion paper [1].

Combined with the fact that the distribution terminates at (as discussed in section 2.3)

$$\left. D_2^{(\alpha)} \right|_{\text{max}} \sim \frac{1}{2z_{\text{cut}}} , \tag{5.7}$$

this implies that, for a fixed mass cut, the full distribution, including non-perturbative effects, of the mMDT/soft drop groomed $D_2^{(\alpha)}$ is largely independent of the jet energy! Unlike the ungroomed $D_2^{(\alpha)}$ distribution, which had both an upper endpoint and location of non-perturbative corrections that depended on the jet energy, the groomed $D_2^{(\alpha)}$ distribution has endpoints and non-perturbative corrections that are independent of the jet energy. We will demonstrate in sections 6 and 7 that both the NLL calculation of the distribution as well as the Monte Carlo simulation respect this prediction.

5.3 Grooming efficiency for signal jets

While we have focused on the properties of the mMDT/soft drop D_2 distribution for background (QCD) jets, jet grooming can have an effect on the signal distribution as well. For an unpolarized boosted Z boson that decays to a $q\bar{q}$ pair, the distribution of the energy fraction z of the quark, say, is approximately flat:

$$\frac{d\sigma}{dz} \simeq \Theta(1-z)\Theta(z) . \tag{5.8}$$

This implies that when the boosted Z jet is groomed, a fraction $2z_{\text{cut}}$ of the jets will have one prong removed by grooming. For these jets that lose one prong, they will also typically fail the mass cut, as well as no longer have a clear two-prong structure. Of course, for $z_{\text{cut}} \ll 1$, this is formally a small effect, but practically, if $z_{\text{cut}} \simeq 0.1$, then about 20% of the Z jets could have a prong removed. This effect could have a large effect on the signal D_2 distribution.

While at leading-order the distribution of the energy fraction z is approximately flat, when all-orders effects are included the regions with $z \to 0$ and $z \to 1$ are suppressed by a Sudakov factor. When $z \to 0$, for example, there is of course no divergence in the leading-order Z decay matrix element. However, a gluon emitted off of the soft decay product will itself necessarily be soft, and result in a divergence at fixed-order. When all-orders effects are included, these soft gluon divergences arrange themselves into a Sudakov factor that suppresses the probability for a decay product to only carry a small fraction of the energy of the Z. At double logarithmic accuracy (DLA), this Sudakov factor is

$$\frac{d\sigma^{\text{DLA}}}{dz} \simeq \Theta(1-z)\Theta(z) \exp\left[-\frac{\alpha_s}{2\pi}C_F\left(\log^2 z + \log^2(1-z)\right)\right], \qquad (5.9)$$

which can be derived from the Z boson decay matrix element at next-to-leading order. This Sudakov factor pushes decay products of the Z to have more equal energies, and reduces the fraction of Z jets that have a subjet that is removed by the jet groomer. That is, due to all-orders effects, hadronic decays of Z bosons can look more two-prong-like than their fixed-order description would suggest.

In our analytic calculations for the prediction of the D_2 distribution on groomed signal jets, we include this resummation to NLL accuracy. The suppression of the $z \to 0$ and $z \to 1$ regions will be much larger than that suggested by the simple Sudakov factor that exists at DLA accuracy. Nevertheless, even at DLA accuracy, this suppression is non-trivial. With $z_{\text{cut}} = 0.1$ and using the distribution of eq. (5.9), only about 15% of Z jets fail soft drop, as compared to 20% using eq. (5.8).

6 NLL predictions in e^+e^- collisions

In this section we use our factorization formulae to provide numerical results for the D_2 distribution in e^+e^- collisions. In section 6.1 we compare our result, expanded to fixed order, with the fixed order code EVENT2 to ensure that we reproduce the singular behavior of the D_2 distribution. In section 6.2 we compare our resummed results, matched to fixed order, with parton shower Monte Carlo.

In all of the plots that follow, we set the angular exponent in the definition of D_2 to $\alpha = 2$. We do this for two reasons. First, when $\alpha = 2$, the two-point energy correlation function $e_2^{(\alpha)}$ is proportional to the squared mass of the jet. Therefore, it is easy to enforce a mass cut on jets of fixed energy. Additionally, from the analysis of non-perturbative effects in section 5.2, with $\alpha = 2$, for a given mass cut, there is no dependence on jet energy of the size of non-perturbative corrections. A study of mMDT groomed D_2 with exponent $\alpha \neq 2$ is beyond the scope of this paper, though we expect that many of the features we observe to apply to groomed two-pronged observables more generally.
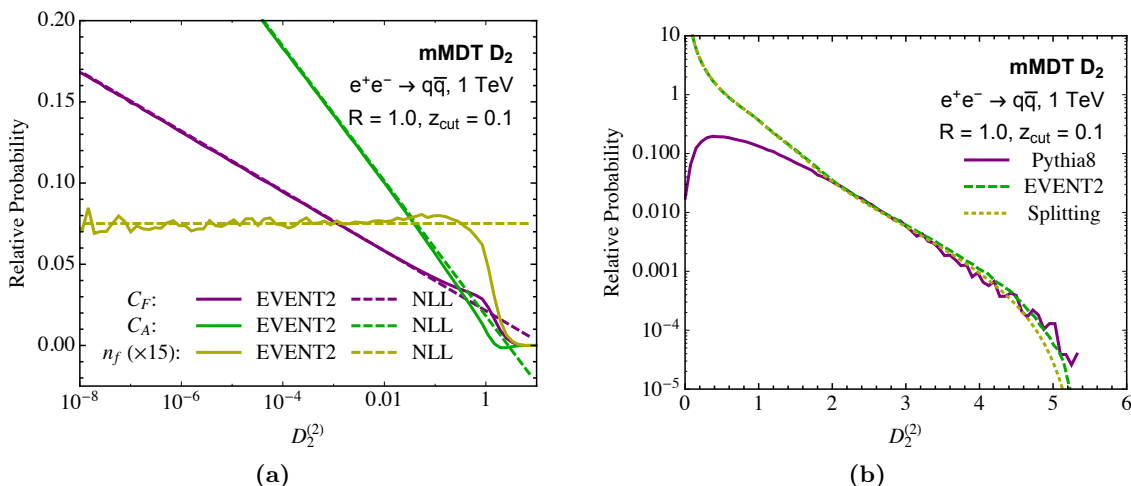
Figure 12. (a) A comparison of the NLL prediction expanded to fixed order with EVENT2. The singular behavior of the D_2 distribution is reproduced by our factorization. (b) Comparison of the end-point value found by fixed-order codes and Monte Carlo event generators. Note the robustness of the endpoint set by z_{cut} even in the prescence of multiple emissions.

6.1 Singular results and comparison with EVENT2

To verify that our factorization reproduces the singular behavior of the D_2 distribution as $D_2 \to 0$, we can compare the results of our factorization formula, expanded to α_s^2, with the fixed order generator EVENT2 [60]. In figure 12a we show the result of EVENT2 in each of the color channels, compared with the expansion of our NLL formula. We see that at small values of D_2, our NLL formula captures the singular structure of the EVENT2 distribution, as is required. Here we consider $e^+e^- \to$ dijets at 1 TeV with $z_{\text{cut}} = 0.1$, but we have found similar agreement for other values of the z_{cut} parameter, while verifying the independence on the center-of-mass energy and jet mass bin. This verifies the consistency of our factorization to $\mathcal{O}(\alpha_s^2)$. Due to the complexity of our factorization, this is a highly non-trivial check, and gives us confidence that have correctly incorporated all modes in the effective theory.

In figure 12b we show a linear plot of the D_2 distribution, comparing EVENT2 [60], PYTHIA 8.226 [78, 79], and a calculation using the $1 \to 3$ splitting functions that is discussed in detail in appendix D. The details of the PYTHIA 8.226 result will be discussed in section 6.2. This figure illustrates two important points. First, as described in section 3, our factorization formula for the D_2 observable isolates the collinear physics. If we did not want to resum the small D_2 behavior, then this shows that the fixed order result can be computed, up to power corrections, using the $1 \to 3$ splitting function. This is seen by the excellent agreement between the result of EVENT2 and the result computed using the $1 \to 3$ splitting functions, shown in figure 12b. Second, our factorization formulae describe both the small D_2 region, where there is a resolved substructure, as well as the large D_2 region, where the substructure is unresolved. A correct description of the unresolved region of phase space, with the collinear-soft haze factorization of section 3.2.1, is required to describe the correct endpoint of the distribution, which occurs at $1/(2z_{\text{cut}})$. In the

collinear-soft haze factorization, we do not need to resum logarithms of D_2, and therefore we can simply compute to fixed order, which is equivalent to a fixed order calculation using the $1 \to 3$ splitting function. In figure 12b, we see first of all that all three curves reproduce well the expected $1/(2z_{\text{cut}})$ endpoint, and second, that the calculation based on the $1 \to 3$ splitting function describes relatively well the distribution at large values of D_2, and in particular, the approach to the endpoint.[10] This is important, since it illustrates that already at LO one can have a reasonable description of the endpoint of the distribution, and that the phase space of the observable is already reasonably well filled out. In section 6.2 we will further study this in the matched distributions for different values of z_{cut}.
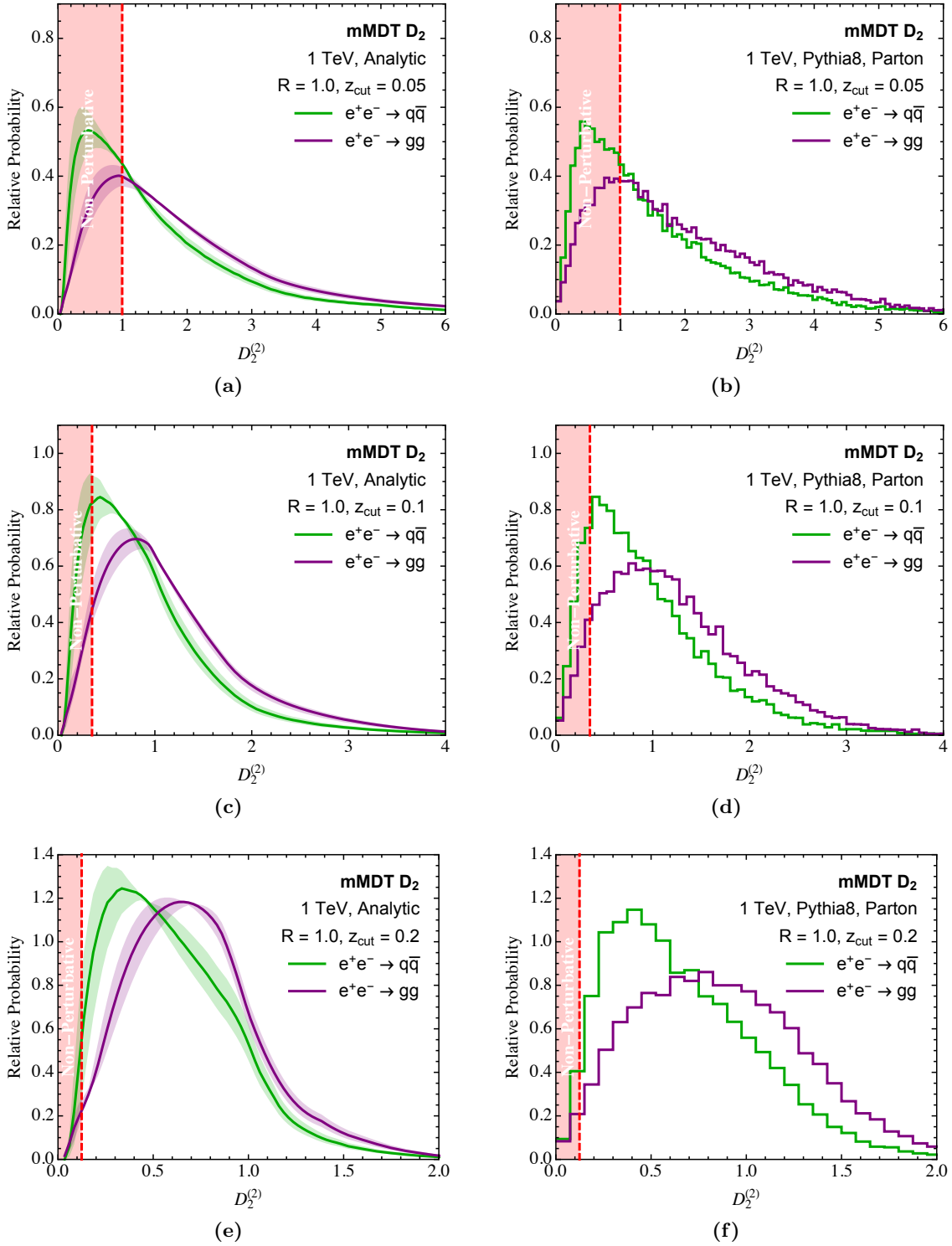
6.2 Comparison with parton shower Monte Carlo

Having shown that we reproduce the singular structure of the D_2 distribution, in this section we compare our NLL resummed predictions multiplicatively matched to the leading order (LO) EVENT2 or $1 \to 3$ splitting functions with parton shower Monte Carlo. For QCD jets, we consider both $e^+e^- \to q\bar{q}$, as well as $e^+e^- \to gg$, generated through an off shell Higgs, while for signal, we consider $e^+e^- \to ZZ$ events with both Zs decaying hadronically. The events were generated with MADGRAPH5 2.5.5 [80], and showered with PYTHIA 8.226 [78, 79]. We also verified that similar results are obtained with VINCIA [81–85], although for simplicity we do not show distributions from VINCIA. Throughout this section we use FASTJET 3.1.2 [86] and the EnergyCorrelator FASTJET CONTRIB [86, 87] for jet clustering and analysis. All jets are clustered using the e^+e^- anti-k_T metric [86, 88] using the WTA recombination scheme [89, 90], with an energy metric.

In figure 13 we show comparisons of our analytic predictions (on the left) with parton shower Monte Carlo results at parton level (on the right). Results are shown for both quark and gluon jets. We also highlight the region where non-perturbative effects from hadronization will have a significant impact on the distribution, as will be discussed shortly. The distributions are shown for three different values of the z_{cut} parameter, namely $z_{\text{cut}} = 0.05, 0.1, 0.2$. Overall, good agreement between the analytic calculation and the parton shower Monte Carlo is observed, and differences between quarks and gluons, as well as the behavior as a function of z_{cut} are well reproduced. In particular, due to the inclusion of the fixed order corrections, the correct endpoint of the distribution is obtained in the analytic calculation. This is crucial for obtaining agreement of the distributions. We also note that, as observed for a wide range of groomed jet observables, the mMDT groomed D_2 distributions for quarks and gluons have a relatively small difference, as compared to its ungroomed counterpart.

It is interesting to consider the behavior as a function of z_{cut}, and in particular the differences between the parton shower and analytic results as the value of z_{cut} is increased, corresponding to a more aggressive grooming. In figure 13 we see that for smaller values

---

[10]The precise behavior of the $1 \to 3$ splitting function calculation and EVENT2 at the endpoint becomes sensitive to the binning used in this region, since the distribution is rapidly vanishing. One must trade accuracy of reproducing the endpoint for numerical stability of the bins. We have found that using smaller binning always improves the agreement between EVENT2 and the $1 \to 3$ splitting functions, at the expense of having to run longer to achieve adequate accuracy and precision.

Figure 13. A comparison of the analytic $D_2^{(2)}$ distributions for both quark and gluons (left column) with parton shower Monte Carlo at parton level (right column), for different values of the z_{cut} parameter. A mass cut of $m_J \in [80, 100]\,\text{GeV}$ has been applied. Good agreement in the shape is observed, particularly for smaller values of z_{cut}.
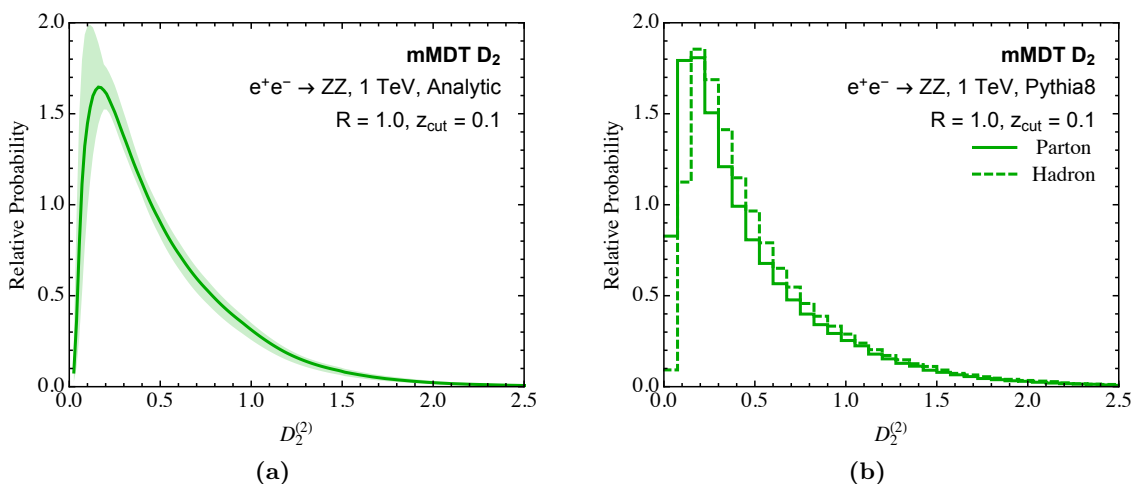
Figure 14. A comparison of the analytic $D_2^{(2)}$ distribution for signal (Z) jets (left) and the parton shower Monte Carlo distribution (right). A mass cut of $m_J \in [80, 100]\,\mathrm{GeV}$ has been applied. Good agreement is observed.

of z_{cut}, where the grooming has a smaller effect, better agreement is observed between the analytic and parton shower results. We believe that this arises primarily due to two effects. First, since the endpoint of the distribution scales as $1/z_{\mathrm{cut}}$, for smaller values of z_{cut}, there is a less rapid transition between the resummation and fixed order regime, ensuring that there are well separated resummation and fixed orders regions. Secondly, we observe that for more aggressive grooming the LO fixed order prediction undershoots the Monte Carlo distribution at large values of D_2. Since we are considering normalized distributions, this then translates into a large difference in the peak region, as seen most clearly for $z_{\mathrm{cut}} = 0.2$. We believe that this could be remedied by including higher order perturbative corrections, which would fill out the phase space better, more similar to the parton shower. Indeed, the LO predictions are known to undershoot in other grooming or D_2 studies [34, 35, 53, 91, 92]. It would be very interesting to include higher order corrections, but this is beyond the scope of the current paper. Although our focus in this paper is primarily on the calculation of groomed D_2 for QCD jets, in figure 14 we show a comparison of our analytic results with parton shower Monte Carlo for a hadronically decaying boosted Z. Excellent agreement is observed.

Finally, it is important to address the impact of hadronization corrections to the distribution. Although it was shown in section 5.2 that hadronization corrections are suppressed significantly by the grooming procedure, they still have a non-negligible impact on the D_2 distribution. In figure 15 we show the effect of hadronization on the groomed D_2 spectrum in parton shower Monte Carlo for both quark and gluon jets, and for the different values of z_{cut} considered above. In figure 14b the effect of hadronization is shown for the boosted Z distribution. We see that in all cases, the effect of hadronization is quite minor, and does not dominate the shape of the distribution. It can be included in the analytic calculation using a model shape function [69, 93–96], although we will not pursue this further in this paper.
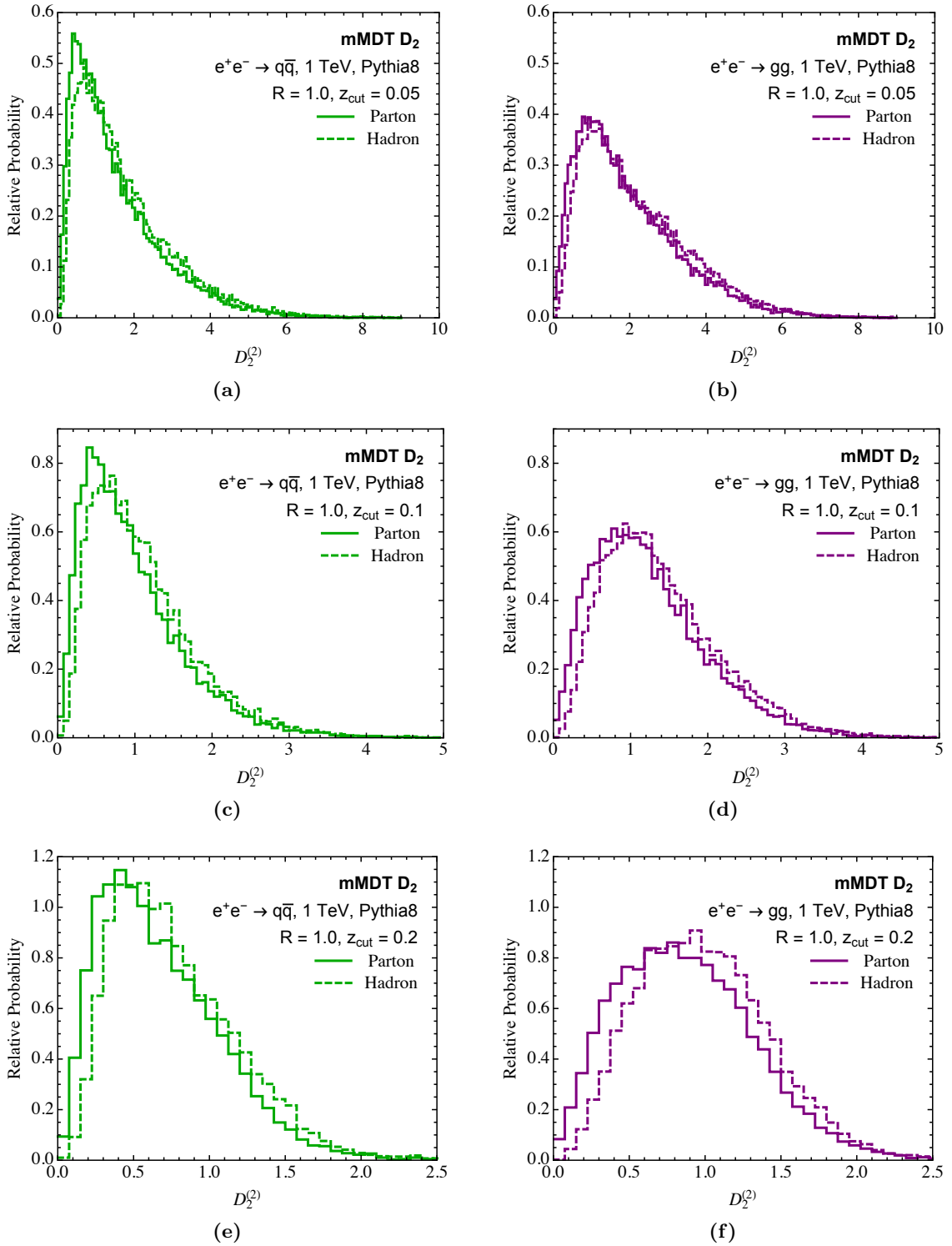
Figure 15. The effect of hadronization on the $D_2^{(2)}$ distribution for quark (left column) and gluon (right column) jets. Non-perturbative corrections are suppressed by the grooming procedure. A mass cut of $m_J \in [80, 100]\,\text{GeV}$ has been applied.

A detailed study of the non-perturbative corrections for the D_2 distribution at the LHC, and their incorporation through shape functions, are presented in a companion paper [1].

7 Monte Carlo results in pp collisions

Analytic predictions of the D_2 distribution for pp processes of interest are presented in a companion paper [1]. In this section we perform a Monte Carlo study demonstrating the consequences of the power counting and factorization analysis presented earlier. For convenience, we recall here the major predictions of our analysis. We emphasize that these are robust predictions of the factorization formula, combined with the power counting analysis, which hold within the region of validity of the factorization formula, namely $e_2^{(\alpha)} \ll z_{\rm cut}$. They should therefore be independent of the details of the hadronization model or details of the perturbative shower. The non-trivial predictions are:

- The endpoint of the D_2 distribution is fixed as $1/(2z_{\rm cut})$. This is independent of the jet mass, jet energy, hadronization, or the angular exponent used to define the energy correlation functions.

- The scale at which hadronization corrections become important is independent of the p_T of the jet. It depends only on $\Lambda_{\rm QCD}$, the jet mass, and the value of the $z_{\rm cut}$ parameter.

- The distributions depend only the quark vs. gluon fraction of the jets in the event, but are otherwise process independent. We have previously argued that the soft drop procedure also reduces the dependence of the distribution on the parton flavor.

We will see that each of these predictions is well reproduced by Monte Carlo parton shower in pp collisions. The parton-level samples in this section were generated at the 13 TeV LHC with MADGRAPH5 2.5.5 [80] and showered with PYTHIA 8.226 [78, 79] with default settings. Jets were clustered with the anti-k_T algorithm [88] in FASTJET 3.1.2 [86] and the EnergyCorrelator and SoftDrop FASTJET CONTRIBS [86, 87] for jet analysis.

7.1 Signal distributions

We first show the distributions for signal jets. The signal jets are clustered from $pp \to ZZ$ events in which one Z boson is forced to decay to neutrinos, and the other to hadrons. In figure 16, we plot the groomed jet $D_2^{(2)}$ distributions of the hadronically-decaying Z boson. In these plots, the mMDT groomer (soft drop with $\beta = 0$) is used, with the parameter $z_{\rm cut} = 0.1$. A mass cut of 80 GeV $< m_J <$ 100 GeV is also imposed on the groomed jet. First, in figure 16a, we plot the $D_2^{(2)}$ distribution for various p_T cuts on the jets. The signal distributions are stable with respect to p_T, which might be expected as the Z boson is a color-singlet and has an intrinsic, Lorentz-invariant energy scale, m_Z. Importantly, the grooming appears to successfully remove contamination radiation at wide angles in the jet that would distort the distribution, and potentially become more important at larger p_T.

In figure 16b, we fix the transverse momentum range to $p_T > 500$ GeV, and vary the clustered jet radius from $R = 0.6$ to $R = 1.0$. Essentially no effect on the distribution is
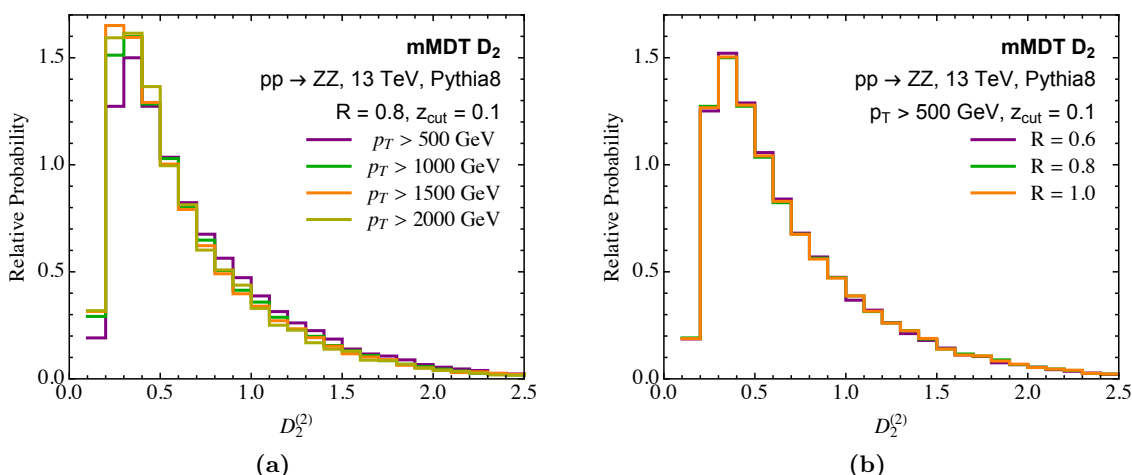
Figure 16. Distributions of $D_2^{(2)}$ measured on mMDT groomed hadronically-decaying Z jets with $z_{\rm cut} = 0.1$ from the process $pp \to ZZ$ at the 13 TeV LHC. The groomed jet mass is restricted to lie in the range 80 GeV $< m_J <$ 100 GeV. (a) Distributions of $D_2^{(2)}$ for various p_T cuts, with the jet radius fixed to $R = 0.8$. (b) Distributions of $D_2^{(2)}$ for various jet radii R, with the jet p_T cut fixed to $p_T > 500$ GeV.

observed in changing the jet radius, corroborating the performance of the mMDT groomer in removing contamination radiation. Note also that with this p_T range and these jet radii, the Z boson decay products are well-contained within the jet. For this p_T range, the angular scale of the Z boson decay products R_Z is approximately

$$R_Z \simeq \frac{2m_Z}{p_T} \simeq 0.4 \,. \tag{7.1}$$

7.2 Background distributions

It may have been expected that the signal distributions were robust under changes of jet parameters, both due to grooming as well as the intrinsic mass scale. Here, we will show that mMDT/soft drop grooming also renders the background distributions extremely robust to jet parameters. We study jets produced in two processes for our background: $pp \to Zj$ and $pp \to Hj$. The production of the Z or H bosons in association with the jet enables a handle on the quark and gluon jet fractions. Because soft drop grooming formally makes quark and gluon jet definitions infrared and collinear safe, we could in principle extract the individual quark and gluon jet distributions of $D_2^{(2)}$ from these two samples; however, since separate quark and gluon distributions were studied in the context of e^+e^- in section 6, here we will focus only on the mixed distributions. To easily isolate the hadronic jet in these events, we force the Z boson to decay to neutrinos and the H boson to decay to photons. As with the signal events, a mass cut of 80 GeV $< m_J <$ 100 GeV is imposed on the groomed jet.

In figure 17, we plot the mMDT $D_2^{(2)}$ distributions for various jet p_T cuts, for both the $pp \to Zj$ and $pp \to Hj$ samples. As predicted from our factorization formula, the background distributions are very weakly dependent on the jet p_T. This is a consequence of the facts that the endpoint of the distributions are fixed at $D_2^{\max} = 1/(2z_{\rm cut}) = 5$ and that non-perturbative effects become important at a scale set by the jet mass, and not the jet p_T.

(a)   **(b)**
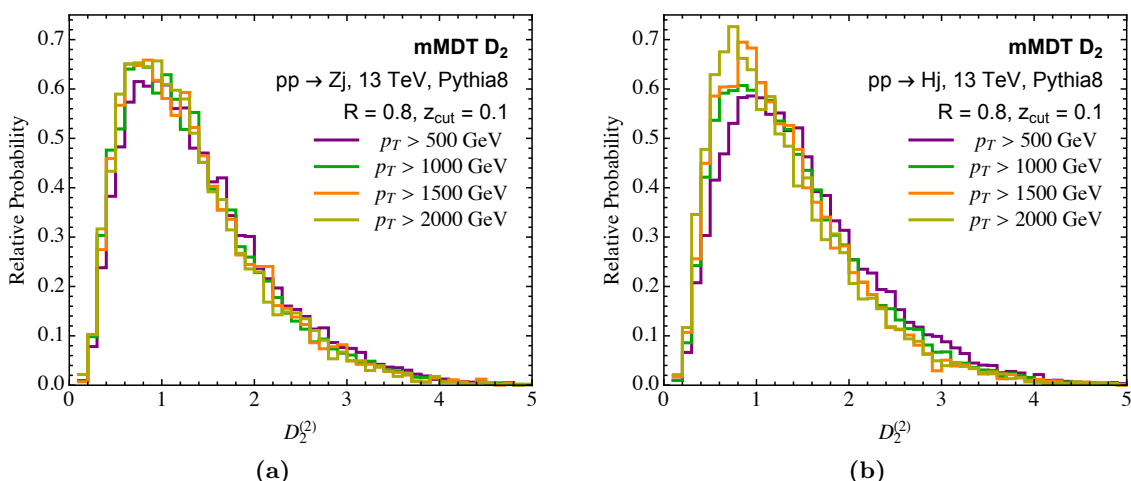
Figure 17. Distributions of $D_2^{(2)}$ measured on mMDT groomed QCD jets with $z_{\mathrm{cut}} = 0.1$ from the processes $pp \to Zj$ (left) and $pp \to Hj$ (right) at the 13 TeV LHC. The groomed jet mass is restricted to lie in the range 80 GeV $< m_J <$ 100 GeV. Various p_T cuts are shown, with the jet radius fixed to $R = 0.8$.

(a)   **(b)**
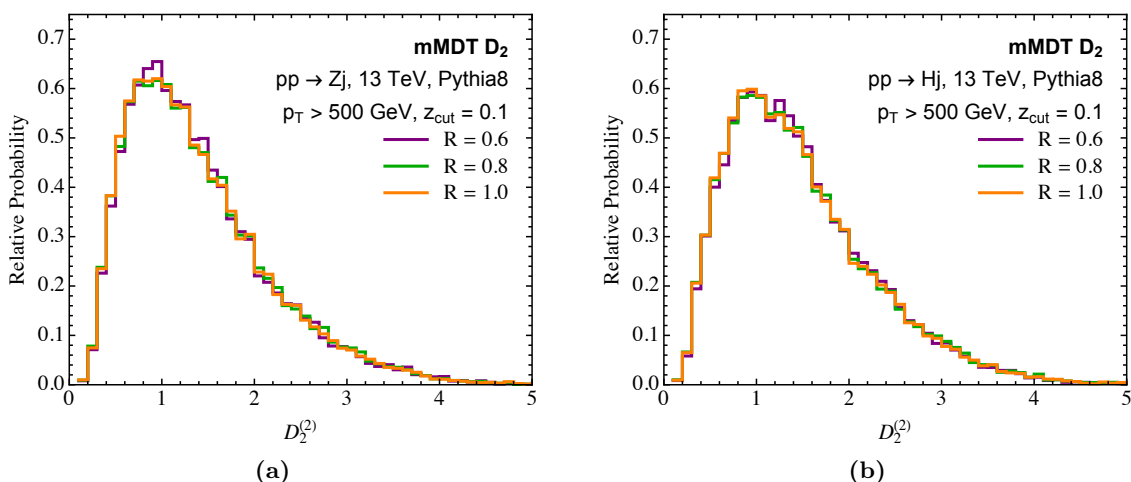
Figure 18. Distributions of $D_2^{(2)}$ measured on mMDT groomed QCD jets with $z_{\mathrm{cut}} = 0.1$ from the processes $pp \to Zj$ (left) and $pp \to Hj$ (right) at the 13 TeV LHC. The groomed jet mass is restricted to lie in the range 80 GeV $< m_J <$ 100 GeV. Various jet radii R are shown, with the jet p_T cut fixed to $p_T > 500$ GeV.

Additionally, the distributions in the Z and H samples are very similar, demonstrating that quark vs. gluon flavor effects are small. This is a consequence of both the constraints on the threshold and endpoint kinematics as well as the mixture of quark and gluon jets in the two samples. The constraints on the bounds of the distribution were derived independent of jet flavor, and so the distribution has a relatively weak dependence on jet flavor.

Plots of the groomed $D_2^{(2)}$ distributions on the background jet samples with different jet radii are shown in figure 18. Here, the jet p_T cut is fixed to $p_T > 500$ GeV, while the

jet radius ranges from $R = 0.6$ to $R = 1.0$. As observed with the signal distributions, there is extremely weak dependence on the jet radius, demonstrating that mMDT/soft drop is efficient at removing wide-angle radiation in the jet that would be sensitive to the jet radius. After grooming, all radiation in the jet is collinear, and the relevant angular scales are set by the ratio of the groomed mass to the jet p_T. By construction, this angular scale is always less than the jet radius (see eq. (7.1)), and so the jet radius is never relevant.

8   Conclusions

In this paper we have performed a detailed study of the factorization properties of groomed multi-prong observables, focusing in particular on the D_2 observable with mMDT or soft drop $\beta = 0$ grooming. We derived factorization formulae which describe the observable to all orders in α_s, and allow us to make powerful statements of phenomenological relevance about the behavior of the groomed D_2 observable. Most interesting are the fixed endpoint of the distribution at $1/(2z_{\rm cut})$, and the independence of non-perturbative corrections on the jet energy scale. Combined, these imply a remarkable robustness of the groomed D_2 observable, which is important for jet substructure applications.

We have introduced factorization formulae describing each region of phase space relevant for groomed boosted boson discrimination. Some of these factorization formulae follow by combining those which previously existed in the literature, however some are new. In particular, we derived a novel factorization describing the production of a soft subjet with energy the scale of the soft drop parameter $z_{\rm cut}$. This factorization has the interesting property that clustering effects enter into the hard matching coefficient for the production of the soft subjet. We computed the functions entering the factorization at one-loop, and showed renormalization group consistency of the factorization formulae. While we have focused on applying these factorization formulae to the particular case of soft dropped D_2, we believe that they will be more generally applicable for describing groomed jet substructure observables, particularly those based on the energy correlation functions. This includes, for example, the N_2 [43] observable used by CMS [76, 97], or more ambitiously, energy correlation based observables for boosted top tagging [16, 43, 98].

We performed a numerical study for $e^+e^- \to$ dijets at NLL order, considering both the case of $e^+e^- \to Z \to q\bar{q}$ and $e^+e^- \to H \to gg$, allowing us to understand the differences between the D_2 distributions for quark and gluon jets. Non-perturbative effects were found to be small, and good agreement was found between the predictions of the Monte Carlo parton shower, and the analytic calculation. Since the D_2 observable probes multi-particle splittings, it may prove useful for testing Monte Carlo generators that implement $1 \to 3$ [99, 100] or $2 \to 4$ splittings [101].

Perhaps most interestingly, we have shown that due to the grooming procedure, our calculations extend straightforwardly to proton-proton collisions at the LHC. This allows for experimentally realistic jet substructure observables currently used at the LHC to be calculated with theoretical precision. In this paper we performed a Monte Carlo study, showing that the features predicted by the factorization formulae are reproduced by Monte Carlo parton shower generators. In particular, we focused on the robustness of the groomed

D_2 distribution as a function of the jet p_T, the scaling of non-perturbative hadronization effects, and the partonic content of the jet. We also showed that the groomed D_2 distribution is process independent up to the quark-gluon fraction of the jet. These are experimentally desirable features, which can be derived from a first principles theoretical description, and make groomed D_2 a promising observable for QCD studies at the LHC, as well as putting its theoretical understanding as a jet substructure tagger on firm theoretical footing. Analytic results for D_2 distributions for relevant processes at the LHC are presented in a companion publication [1].

Acknowledgments

A   Ingredients for collinear subjets

In this appendix we present the one-loop calculation of all the functions appearing in the collinear subjets factorization formula.

A.1   Kinematics and notation

We begin by briefly describe the kinematics and notation. We follow closely [15]. We let Q be the center of mass energy of the e^+e^- collisions. The energy deposited in each hemisphere is therefore $Q/2$, and the four-momenta of the hemispheres are

$$p_{\text{hemisphere}_1} = \left(\frac{Q}{2}, \vec{p}_1 \right) , \qquad p_{\text{hemisphere}_2} = \left(\frac{Q}{2}, -\vec{p}_1 \right) , \tag{A.1}$$

so that we have $s = Q^2$. For the kinematics of the subjets, we will use the following notation

Subjet a,b momenta:	$p_a, p_b \,,$	(A.2)
Subjet a,b spatial directions:	$\hat{n}_a, \hat{n}_b \,,$	(A.3)
Thrust axis:	$\hat{n} = \dfrac{\hat{n}_a + \hat{n}_b}{|\hat{n}_a + \hat{n}_b|} \,,$	(A.4)
Light-cone vectors:	$n = (1, \hat{n}), \bar{n} = (1, -\hat{n}),$	
	$n_{a,b} = (1, \hat{n}_{a,b}), \bar{n}_{a,b} = (1, -\hat{n}_{a,b}) \,.$	(A.5)

These satisfy the relations

$$n \cdot n_a = n \cdot n_b = \frac{n_a \cdot n_b}{4}, \qquad \bar{n} \cdot n_a = \bar{n} \cdot n_b = 2, \tag{A.6}$$

$$n_{\perp a,b} \cdot \bar{n}_{\perp a,b} = -n_{\perp a,b} \cdot n_{\perp a,b} = \hat{n}_{\perp a,b} \cdot \hat{n}_{\perp a,b} = \frac{n_a \cdot n_b}{2}. \tag{A.7}$$

For a particle in the collinear sector a or b, we have

$$p_a \sim \frac{1}{2}(\bar{n} \cdot p_a)n_a, \quad p_b \sim \frac{1}{2}(\bar{n} \cdot p_b)n_b, \qquad p_a^0 \sim \frac{1}{2}(\bar{n} \cdot p_a), \quad p_b^0 \sim \frac{1}{2}(\bar{n} \cdot p_b). \tag{A.8}$$

We will label the energy fractions carried in each subjet by

$$z_{a,b} = \frac{2p_{a,b}^0}{Q} = \frac{\bar{n} \cdot p_{a,b}}{Q}, \tag{A.9}$$

where the second relation is true to leading power.

We can now compute the leading power expressions for the observables in these kinematics. The value of $e_2^{(\alpha)}$ is given by

$$\begin{aligned}
e_2^{(\alpha)} &= \frac{1}{E_J^2} E_a E_b \left(\frac{2p_a \cdot p_b}{E_a E_b} \right)^{\alpha/2} \\
&= 2^{\alpha/2} z_a z_b \, (n_a \cdot n_b)^{\alpha/2}. \tag{A.10}
\end{aligned}$$

For three emissions, with momenta k_1, k_2, k_3, the expression for the three point energy correlation function is

$$e_3^{(\alpha)} = \frac{1}{E_J^3} k_1^0 k_2^0 k_3^0 \left(\frac{2k_1 \cdot k_2}{k_1^0 k_2^0} \right)^{\alpha/2} \left(\frac{2k_1 \cdot k_3}{k_1^0 k_3^0} \right)^{\alpha/2} \left(\frac{2k_2 \cdot k_3}{k_2^0 k_3^0} \right)^{\alpha/2}. \tag{A.11}$$

For an emission collinear with one of the subjets, where we have the splitting $p_{a,b} \to k_1 + k_2$, we have

$$e_3^{(\alpha)} \Big|_{k_1, k_2 \| n_a} = 2^{5\alpha/2} z_b (n_a \cdot n_b)^\alpha \left(\frac{k_1 \cdot k_2}{Q^2} \right)^{\frac{\alpha}{2}} \left(\frac{\bar{n}_a \cdot k_1}{Q} \right)^{1 - \frac{\alpha}{2}} \left(\frac{\bar{n}_a \cdot k_2}{Q} \right)^{1 - \frac{\alpha}{2}}, \tag{A.12}$$

$$e_3^{(\alpha)} \Big|_{k_1, k_2 \| n_b} = 2^{5\alpha/2} z_a (n_a \cdot n_b)^\alpha \left(\frac{k_1 \cdot k_2}{Q^2} \right)^{\frac{\alpha}{2}} \left(\frac{\bar{n}_b \cdot k_1}{Q} \right)^{1 - \frac{\alpha}{2}} \left(\frac{\bar{n}_b \cdot k_2}{Q} \right)^{1 - \frac{\alpha}{2}}. \tag{A.13}$$

For a third collinear-soft emission k off of the $p_{a,b}$ partons, we have

$$e_3^{(\alpha)} \Big|_{k \to \text{c-soft}} = 2^{3\alpha/2 + 1} z_a z_b \, (n_a \cdot n_b)^{\alpha/2} \left(\frac{\bar{n} \cdot k}{2Q} \right)^{1 - \alpha} \left(\frac{n_a \cdot k}{Q} \right)^{\alpha/2} \left(\frac{n_b \cdot k}{Q} \right)^{\alpha/2}. \tag{A.14}$$

The factorization formula in this region of phase space, which we repeat here for convenience, is:

$$\frac{d^3\sigma}{dz \, de_2^{(\alpha)} \, de_3^{(\alpha)}} = H(Q^2) S(z_{\text{cut}}) S_c(z_{\text{cut}}, \theta > \theta_{ab}) H_2(z, e_2^{(\alpha)}) \tag{A.15}$$

$$\cdot J_1(e_3^{(\alpha)}) \otimes J_2(e_3^{(\alpha)}) \otimes C_s(e_3^{(\alpha)}, \theta < \theta_{ab}) \otimes C_s^{\text{NG}}(e_3^{(\alpha)}, z_{\text{cut}}).$$

The convolutions over $e_3^{(\alpha)}$ can be turned into products by Laplace transforming with respect to the variable $e_3^{(\alpha)}$. We denote the Laplace transformed variables and functions with a tilde. When Laplace transformed, the cross section can be written as

$$\frac{d^2\sigma}{dz\,de_2^{(\alpha)}\,d\tilde{e}_3^{(\alpha)}} = H(Q^2)S(z_{\rm cut})S_c(z_{\rm cut},\theta>\theta_{ab})H_2(z,e_2^{(\alpha)}) \tag{A.16}$$

$$\cdot \tilde{J}_1(\tilde{e}_3^{(\alpha)})\tilde{J}_2(\tilde{e}_3^{(\alpha)})\tilde{C}_s(\tilde{e}_3^{(\alpha)},\theta<\theta_{ab})C_s^{\rm NG}(\tilde{e}_3^{(\alpha)},z_{\rm cut})\,.$$

In this appendix, we will present the expressions for these Laplace transformed functions.

A.2 Matrix element definitions

We give the matrix element definitions of all functions in the factorization formula in eqs. (A.15) and (3.13), and eq. (3.28). The factorization formulae presented in this paper are formulated in the language of SCET [38–41]. We refer readers unfamiliar with SCET to the reviews [102, 103]. The jet functions are identical to [15], so we only give the soft matrix elements. They are defined in terms of soft Wilson lines

$$S_q = \mathbf{P}\exp\left(ig\int\limits_0^\infty ds\, q\cdot A(x+sq)\right)\,. \tag{A.17}$$

Explicitly, we have

$$C_s(\tilde{e}_3^{(\alpha)},\theta<\theta_{ab}) = \frac{1}{N_f}{\rm tr}\left[\mathbf{T}_f\langle0|T\{S_aS_bS_{\bar{n}}\}\exp\left(-\tilde{e}_3^{(\alpha)}\mathbf{E_3}^{(\alpha)}\big|_{{\rm SD}a,b}\right)\bar{T}\{S_aS_bS_{\bar{n}}\}|0\rangle\mathbf{T}_f^\dagger\right], \tag{A.18}$$

$$S_c(z_{\rm cut},\theta>\theta_{ab}) = \frac{1}{N_f}{\rm tr}\left[\mathbf{T}_f\langle0|T\{S_aS_bS_{\bar{n}}\}\Theta_{\rm SD}(a,b,z_{\rm cut})\bar{T}\{S_aS_bS_{\bar{n}}\}|0\rangle\mathbf{T}_f^\dagger\right], \tag{A.19}$$

$$C_s^{\rm NG}(\tilde{e}_3^{(\alpha)},z_{\rm cut}) = \frac{1}{N_f}$$

$$\times\frac{{\rm tr}\left[\mathbf{T}_f\langle0|T\{S_aS_bS_{\bar{n}}\}\Theta_{\rm SD}(a,b,z_{\rm cut})\exp\left(-\tilde{e}_3^{(\alpha)}\mathbf{E_3}^{(\alpha)}\big|_{{\rm SD}a,b}\right)\bar{T}\{S_aS_bS_{\bar{n}}\}|0\rangle\mathbf{T}_f^\dagger\right]}{C_s(\tilde{e}_3^{(\alpha)},\theta<\theta_{ab})S_c(z_{\rm cut},\theta>\theta_{ab})}\,. \tag{A.20}$$

Here a,b are the light-cone directions of the subjets, the operator $\Theta_{\rm SD}(a,b,z_{\rm cut})$ imposes an energy cut of $Qz_{\rm cut}/2$ on any emission that is not clustered into legs a,b before they themselves are clustered, and $\mathbf{E_3}^{(\alpha)}\big|_{{\rm SD}a,b}$ returns only the energy-correlation function computed on all momenta that are clustered into a or b *before* a and b are clustered. This works out to be a purely geometrical constraint, for both the soft-collinear and collinear-soft functions, since all emissions are at a scale where they cannot pass soft drop on their own. $\mathbf{T}_f$ are color matrices contracted into the Wilson lines, and in general depend upon the flavor structure of the splitting, as does the representation of the Wilson lines, that is, whether the $1\to2$ subjet splitting was $g\to gg$, $q\to qg$, or $g\to q\bar{q}$.

Throughout these appendices we will refer to the perturbative order of the calculation with a superscript (L), where L denotes the loop order of the calculations. For example, for a function F, we have

$$F = F^{(T)} + F^{(1)} + F^{(2)} + \dots\,, \tag{A.21}$$

where $F^{(L)}$ denotes the L-th loop correction to the function F. The superscript (T) stands for "tree".

A.3 Hard function

The hard function for $e^+e^- \to$ dijets can be found in [54, 104–106]. To $\mathcal{O}(\alpha_s)$ it is given by

$$H(Q^2) = 1 + \frac{\alpha_s C_F}{2\pi}\left(-\log^2\frac{\mu^2}{Q^2} - 3\log\frac{\mu^2}{Q^2} - 8 + \frac{7}{6}\pi^2\right), \tag{A.22}$$

and its anomalous dimension is given by

$$\gamma_H = \frac{\alpha_s C_F}{\pi}\left(-2\log\frac{\mu^2}{Q^2} - 3\right). \tag{A.23}$$

For notational simplicity, throughout these appendices we do not explicitly write the arguments of the anomalous dimensions.

A.4 Hard splitting function

The hard splitting function describing the $q \to qg$ splitting into two collinear subjets was first derived in the SCET$_+$ context in [54] using results from [107, 108], while the $g \to gg$ splitting and $g \to q\bar{q}$ was derived in [15], where all the coefficients were given in terms of the variables most useful for the current study. Taking as an example initial quarks splitting to $\mathcal{O}(\alpha_s)$, the hard splitting function is given by

$$H_2^{q \to qg}(e_2^{(2)}, z_q, \mu) = \frac{\alpha_s C_F}{2\pi}\frac{1}{e_2^{(2)}}\frac{1 + z_q^2}{1 - z_q} \tag{A.24}$$

$$\times \left\{ 1 + \frac{\alpha_s}{2\pi}\left[\left(\frac{C_A}{2} - C_F\right)\left(2\log\left(\frac{Q^2}{\mu^2}\frac{e_2^{(2)}}{4}\right)\log z_q + \log^2 z_q + 2\mathrm{Li}_2(1 - z_q)\right)\right.\right.$$

$$-\frac{C_A}{2}\left(\log^2\left(\frac{Q^2}{\mu^2}\frac{e_2^{(2)}}{4}\right) - \frac{7\pi^2}{6} + 2\log\left(\frac{Q^2}{\mu^2}\frac{e_2^{(\alpha)}}{4}\right)\log(1 - z_q) + \log^2(1 - z_q) + 2\mathrm{Li}_2(z_q)\right)$$

$$\left.\left. + (C_A - C_F)\frac{1 - z_q}{1 + z_q^2}\right]\right\}, $$

and its anomalous dimension is

$$\gamma_{H_2}^{q \to qg} = -2\frac{\alpha_s}{\pi}\left(\frac{C_A}{2} - C_F\right)\log z_q - \frac{\alpha_s}{\pi}C_A\log\frac{4\mu^2}{e_2^{(2)}Q^2} + \frac{\alpha_s}{\pi}C_A\log(1 - z_q) - \frac{\alpha_s}{2\pi}\left(\frac{11}{3}C_A - \frac{2}{3}n_f\right). \tag{A.25}$$

A.5 Jet functions

The jet functions in the collinear subjets region of phase space are identical to the case that no grooming is applied, and therefore are given in [15]. To $\mathcal{O}(\alpha_s)$, the quark jet function in the direction n_a is given by

$$\tilde{J}_{q,n_a}^{(1)}(Q_J, \tilde{e}_3^{(2)}) = \frac{\alpha_s}{2\pi}C_F\left[\frac{2}{\epsilon^2} + \frac{3}{2\epsilon} + \frac{2}{\epsilon}\log\left(16(n_a \cdot n_b)^2(1 - z_q)\tilde{e}_3^{(2)}\frac{\mu^2}{Q^2}\right)\right. \tag{A.26}$$

$$\left. + \log^2\left(16(n_a \cdot n_b)^2(1 - z_q)\tilde{e}_3^{(2)}\frac{\mu^2}{Q^2}\right) + \frac{3}{2}\log\left(16(n_a \cdot n_b)^2(1 - z_q)\tilde{e}_3^{(2)}\frac{\mu^2}{Q^2}\right) - \frac{\pi^2}{3} + \frac{7}{2}\right],$$

and its anomalous dimension is

$$\gamma_{J_q} = \frac{\alpha_s}{2\pi} C_F \left[4 \log \left(4 \frac{\left(e_2^{(2)}\right)^2}{z_q^2(1-z_q)} \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + 3 \right] . \tag{A.27}$$

The gluon jet function is given by

$$\tilde{J}_{g,n_a}^{(1)}(Q_J, \tilde{e}_3^{(2)}) = \frac{\alpha_s}{2\pi} C_A \left[\frac{2}{\epsilon^2} + \frac{2}{\epsilon} \log \left(16(n_a \cdot n_b)^2 z_q \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + \frac{1}{\epsilon} \frac{11C_A - 2n_f}{6C_A} \right.$$
$$+ \log^2 \left(16(n_a \cdot n_b)^2 z_q \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + \left(\frac{11}{6} - \frac{n_f}{3C_A} \right) \log \left(16(n_a \cdot n_b)^2 z_q \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right)$$
$$\left. + \frac{67}{18} - \frac{\pi^2}{3} - \frac{5n_f}{9C_A} \right] , \tag{A.28}$$

and its anomalous dimension is

$$\gamma_{J_g} = \frac{\alpha_s}{2\pi} C_A \left[4 \log \left(4 \frac{\left(e_2^{(2)}\right)^2}{z_q(1-z_q)^2} \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + \frac{11C_A - 2n_f}{3C_A} \right] . \tag{A.29}$$

A.6 Wide-angle soft function

The wide angle soft function in the collinear subjets region of phase space does not resolve the collinear splitting, and is therefore identical to that for the soft dropped mass in e^+e^-, which was derived in [34]. At $\mathcal{O}(\alpha_s)$ it can be written as a sum over dipoles, and is given by

$$S(z_{\text{cut}}) = 1 + \frac{\alpha_s C_F}{2\pi} \left(\log^2 \frac{\mu^2}{z_{\text{cut}}^2 Q^2} - \frac{\pi^2}{6} \right) . \tag{A.30}$$

Its anomalous dimension is

$$\gamma_S = 2 \frac{\alpha_s C_F}{\pi} \log \frac{\mu^2}{z_{\text{cut}}^2 Q^2} . \tag{A.31}$$

A.7 Collinear-soft function

The collinear-soft function, $C_s(e_3^{(\alpha)}, \theta < \theta_{ab})$ is new, and we therefore calculate it to $\mathcal{O}(\alpha_s)$ in this appendix. At $\mathcal{O}(\alpha_s)$, it can be written as a sum over dipoles

$$C_s^{(1)}(e_3^{(\alpha)}, \theta < \theta_{ab}) = \frac{1}{N_f} \text{tr} \left[\mathbf{T}_f^\dagger \left(C_{s,ab}(e_3^{(\alpha)}, \theta < \theta_{ab}) + C_{s,a\bar{n}}(e_3^{(\alpha)}, \theta < \theta_{ab}) + C_{s,b\bar{n}}(e_3^{(\alpha)}, \theta < \theta_{ab}) \right) \mathbf{T}_f \right], \tag{A.32}$$

where the contribution from a given dipole is given by

$$C_{s,ij}(e_3^{(\alpha)}, \theta < \theta_{ab}) = -2g^2 \mu^{2\epsilon} \mathbf{T}_i \cdot \mathbf{T}_j \int \frac{d^d k}{(2\pi)^{d-1}} \frac{n_i \cdot n_j}{(n_i \cdot k)(n_j \cdot k)} \delta(k^2) \Theta_{\text{SD}} \delta_{e_3^{(2)}} . \tag{A.33}$$

The soft drop constraint in this region of phase space is given by

$$\Theta_{\text{SD}} = \Theta \left(n_a \cdot n_b - \frac{n_b \cdot k}{\bar{n} \cdot k} \right) \Theta \left(\frac{n_a \cdot k}{\bar{n} \cdot k} - \frac{n_b \cdot k}{\bar{n} \cdot k} \right) + \Theta \left(n_a \cdot n_b - \frac{n_a \cdot k}{\bar{n} \cdot k} \right) \Theta \left(\frac{n_b \cdot k}{\bar{n} \cdot k} - \frac{n_a \cdot k}{\bar{n} \cdot k} \right) . \tag{A.34}$$

The soft drop constraint is easily understood: the emission is to be clustered into direction a or direction b of the splitting before the two legs are themselves clustered. The measurement function is

$$\delta_{e_3^{(2)}} = \delta\left(e_3^{(2)} - 16 e_2^{(2)} \frac{(n_a \cdot k)(n_b \cdot k)}{Q \bar{n} \cdot k}\right). \tag{A.35}$$

We may take the light-cone direction of the parent jet to be $n = \frac{1}{2}(n_a + n_b)$, conjugate to $\bar{n}$. Decomposing k into the light-cone basis formed by n-$\bar{n}$, we have

$$n_a \cdot k = \frac{n_a \cdot n_b}{4} \bar{n} \cdot k + n \cdot k + n_a \cdot k_\perp, \tag{A.36}$$

$$n_b \cdot k = \frac{n_a \cdot n_b}{4} \bar{n} \cdot k + n \cdot k - n_a \cdot k_\perp. \tag{A.37}$$

Here we have used $n_a \cdot \bar{n} = n_b \cdot \bar{n} \approx 2$. We can write the dot product in the transverse plane as

$$n_a \cdot k_\perp = -\sqrt{n_a \cdot n_b}\, k_\perp \cos\phi. \tag{A.38}$$

Here we choose a fixed but arbitrary direction in the transverse plane for the projection of the direction n_a, defining the angle ϕ. Solving the on-shell condition and rescaling, we have

$$\bar{n} \cdot k \to Qz, \qquad k_\perp \to Q z \sqrt{n_a \cdot n_b}\, k_\perp, \tag{A.39}$$

and the Lorentz invariants become

$$n_a \cdot k = \frac{Q}{4} n_a \cdot n_b\, z \left(1 + 4k_\perp^2 - 4k_\perp \cos\phi\right) = \frac{Q}{4} n_a \cdot n_b\, z(1 + 4\vec{k}^2 - 4\vec{k} \cdot \hat{n}), \tag{A.40}$$

$$n_b \cdot k = \frac{Q}{4} n_a \cdot n_b\, z \left(1 + 4k_\perp^2 + 4k_\perp \cos\phi\right) = \frac{Q}{4} n_a \cdot n_b\, z(1 + 4\vec{k}^2 + 4\vec{k} \cdot \hat{n}). \tag{A.41}$$

We can now see that the integration can be efficiently represented by an integral in the transverse plane. The integration measure, having solved the on-shell conditions with these rescalings, is then given by

$$\int \frac{d^d k}{(2\pi)^{d-1}} \delta(k^2) = \frac{1}{2\pi} \left(Q^2 n_a \cdot n_b\right)^{1-\epsilon} \int_0^\infty dz\, z^{1-2\epsilon} \int \frac{d^{2-2\epsilon}\vec{k}}{(2\pi)^{2-2\epsilon}}. \tag{A.42}$$

We can now perform a shift $\vec{k} \to \vec{k} + \frac{1}{2}\hat{n}$, to get the compact form

$$n_a \cdot k = Q n_a \cdot n_b\, z \vec{k}^2, \tag{A.43}$$

$$n_b \cdot k = Q n_a \cdot n_b\, z(\vec{k} + \hat{n})^2. \tag{A.44}$$

The measurement function then becomes

$$\delta_{e_3^{(2)}} = \delta\left(e_3^{(2)} - \frac{e_2^{(2)}(n_a \cdot n_b)^2}{16} z \vec{k}^2 (\vec{k} + \hat{n})^2\right), \tag{A.45}$$

and the soft drop condition becomes

$$\Theta_{\mathrm{SD}} = \Theta\left(1 - \vec{k}^2\right)\Theta\left((\vec{k} + \hat{n})^2 - \vec{k}^2\right) + \Theta\left(1 - (\vec{k} + \hat{n})^2\right)\Theta\left(\vec{k}^2 - (\vec{k} + \hat{n})^2\right). \tag{A.46}$$

We note that the reduction of the collinear-soft function to integrals over a transverse plane is not entirely unexpected, given the duality between time-like and space-like soft processes, see refs. [109–111].

A.7.1 Soft integrals in the transverse plane

We solve the measurement constraint using the z integral, and go to Laplace space. We note that the roles of a and b are interchangeable (this amounts to mapping $\vec{k} \to -\vec{k} - \hat{n}$), so that we have

$$C_{s,ab}(\tilde{e}_3^{(2)}, \theta < \theta_{ab}) = -C_\epsilon^{\text{in}} \mathbf{T}_a \cdot \mathbf{T}_b \int \frac{d^{2-2\epsilon}k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right)\Theta\left((\vec{k}+\hat{n})^2 - \vec{k}^2\right)}{(\vec{k}^2(\vec{k}+\hat{n})^2)^{1-2\epsilon}}, \tag{A.47}$$

$$C_{s,a\bar{n}}(\tilde{e}_3^{(2)}, \theta < \theta_{ab}) = -C_\epsilon^{\text{in}} \mathbf{T}_a \cdot \mathbf{T}_{\bar{n}} \int \frac{d^{2-2\epsilon}k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right)\Theta\left((\vec{k}+\hat{n})^2 - \vec{k}^2\right)}{(\vec{k}^2(\vec{k}+\hat{n})^2)^{1-2\epsilon}}\left((\vec{k}^2) + (\vec{k}+\hat{n})^2)\right),$$

$$C_{s,b\bar{n}}(\tilde{e}_3^{(2)}, \theta < \theta_{ab}) = -C_\epsilon^{\text{in}} \mathbf{T}_b \cdot \mathbf{T}_{\bar{n}} \int \frac{d^{2-2\epsilon}k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right)\Theta\left((\vec{k}+\hat{n})^2 - \vec{k}^2\right)}{(\vec{k}^2(\vec{k}+\hat{n})^2)^{1-2\epsilon}}\left((\vec{k}^2) + (\vec{k}+\hat{n})^2)\right).$$

Here we have introduced the prefactor

$$C_\epsilon^{\text{in}} = 8\alpha_s \left(\frac{\mu}{Q} e_2^{(2)} \tilde{e}_3^{(2)} (n_a \cdot n_b)^{3/2}\right)^{2\epsilon} \Gamma(-2\epsilon). \tag{A.48}$$

A.7.2 Isolating divergent contributions

We note that all integrals in eq. (A.47) only have a divergence at $\vec{k}^2 = 0$, so that we may add and subtract to each color structure the integral

$$-C_\epsilon^{\text{in}} \int \frac{d^{2-2\epsilon}k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right)}{(\vec{k}^2)^{1-2\epsilon}} = -\frac{C_\epsilon^{\text{in}}}{(4\pi)^{1-\epsilon}\epsilon\,\Gamma(1-\epsilon)}. \tag{A.49}$$

The resulting subtracted integrals are all finite, and can be evaluated numerically as a Taylor series in ϵ in terms of two-dimensional integrals. We have found an analytic expression for all the divergent contributions. For instance, in the a-b dipole, we first perform the integral over the magnitude of the vector in the transverse plane, which after some algebra, gives the angular integral

$$\int \frac{d^{2-2\epsilon}k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right)\Theta\left((\vec{k}+\hat{n})^2 - \vec{k}^2\right)}{(\vec{k}^2(\vec{k}+\hat{n})^2)^{1-2\epsilon}} - \int \frac{d^{2-2\epsilon}\vec{k}}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right)}{(\vec{k}^2)^{1-2\epsilon}}$$

$$= \frac{1}{2\pi^2}\left(\int_0^{\frac{\pi}{3}} d\phi\,\phi\cot\phi + \frac{1}{2}\int_{\frac{\pi}{3}}^\pi d\phi\Big\{(\pi-\phi)\cot\phi - \log(1-\cos\phi) - \log 2\Big\}\right) + O(\epsilon)$$

$$= -\frac{\text{Cl}_2(\frac{\pi}{3})}{4\pi^2} + O(\epsilon). \tag{A.50}$$

Here $\text{Cl}_2(x)$ is the Clausen function, which has a value of approximately 1.01494 at its maximum $x = \frac{\pi}{3}$. The other dipoles are handled similarly. Putting in the appropriate factors for $\overline{\text{MS}}$ scheme, we find

$$C_{s,ab}(\tilde{e}_3^{(2)}, \theta < \theta_{ab}) = \frac{\alpha_s}{\pi}\mathbf{T}_a \cdot \mathbf{T}_b\left(\frac{1}{\epsilon^2} + \frac{2}{\epsilon}\left(L_{\theta<\theta_{ab}} - 2\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi}\right) + 2L_{\theta<\theta_{ab}}^2 - 8\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi}L_{\theta<\theta_{ab}}\right) + O(\epsilon^0),$$

$$C_{s,a\bar{n}}(\tilde{e}_3^{(2)}, \theta < \theta_{ab}) = \frac{\alpha_s}{\pi} \mathbf{T}_a \cdot \mathbf{T}_{\bar{n}} \left(\frac{1}{\epsilon^2} + \frac{2}{\epsilon} \left(L_{\theta < \theta_{ab}} + 2 \frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} \right) + 2L_{\theta < \theta_{ab}}^2 + 8 \frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} L_{\theta < \theta_{ab}} \right) + O(\epsilon^0),$$

$$C_{s,b\bar{n}}(\tilde{e}_3^{(2)}, \theta < \theta_{ab}) = \frac{\alpha_s}{\pi} \mathbf{T}_b \cdot \mathbf{T}_{\bar{n}} \left(\frac{1}{\epsilon^2} + \frac{2}{\epsilon} \left(L_{\theta < \theta_{ab}} + 2 \frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} \right) + 2L_{\theta < \theta_{ab}}^2 + 8 \frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} L_{\theta < \theta_{ab}} \right) + O(\epsilon^0),$$

$$\tag{A.51}$$

where the logarithm is defined as

$$L_{\theta < \theta_{ab}} = \log \left(\frac{16\mu (e_2^{(2)})^{5/2} \tilde{e}_3^{(2)} e^{\gamma_E}}{(z(1-z))^{3/2} Q} \right). \tag{A.52}$$

From these expressions it is straightforward to extract the anomalous dimension and perform the renormalization group evolution using standard techniques.

A.8 Soft-collinear function

Now we compute how radiation that is not clustered into the two hard prongs gets groomed. Such radiation is described by the function $S_c(z_{\mathrm{cut}}, \theta > \theta_{ab})$. At one-loop order, the contribution from the generic dipole i, j is given by

$$S_c^{(1)}(z_{\mathrm{cut}}, \theta > \theta_{ab}) = \frac{1}{N_f}$$
$$\times \mathrm{tr} \left[\mathbf{T}_f^\dagger \left(S_{c,ab}(z_{\mathrm{cut}}, \theta > \theta_{ab}) + S_{c,a\bar{n}}(z_{\mathrm{cut}}, \theta > \theta_{ab}) + S_{c,b\bar{n}}(z_{\mathrm{cut}}, \theta > \theta_{ab}) \right) \mathbf{T}_f^\dagger \right], \quad \text{(A.53)}$$

$$S_{c,ij}(z_{\mathrm{cut}}, \theta > \theta_{ab}) = -2g^2 \mu^{2\epsilon} \mathbf{T}_i \cdot \mathbf{T}_j \int \frac{d^d k}{(2\pi)^{d-1}} \frac{n_i \cdot n_j}{(n_i \cdot k)(n_j \cdot k)} \delta(k^2) (1 - \Theta_{\mathrm{SD}}) \Theta_{z_{\mathrm{cut}}}, \quad \text{(A.54)}$$

where Θ_{SD} was defined in eq. (A.34), and the constraint on the energy fraction of the groomed radiation is given by

$$\Theta_{z_{\mathrm{cut}}} = \Theta \left(z_{\mathrm{cut}} - \frac{\bar{n} \cdot k}{Q} \right). \tag{A.55}$$

We use the same coordinate system as above, and arrive at the following representation for each dipole

$$S_{c,ab}(z_{\mathrm{cut}}, \theta > \theta_{ab}) = -C_\epsilon^{\mathrm{out}} \mathbf{T}_a \cdot \mathbf{T}_b \int \frac{d^{2-2\epsilon} k}{(2\pi)^{2-2\epsilon}} \left(1 - \Theta\left(1 - \vec{k}^2\right) \Theta\left((\vec{k} + \hat{n})^2 - \vec{k}^2\right) \right) \frac{1}{\vec{k}^2 (\vec{k} + \hat{n})^2},$$

$$S_{c,a\bar{n}}(z_{\mathrm{cut}}, \theta > \theta_{ab}) = C_\epsilon^{\mathrm{out}} \mathbf{T}_a \cdot \mathbf{T}_{\bar{n}} \int \frac{d^{2-2\epsilon} k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right) \Theta\left((\vec{k} + \hat{n})^2 - \vec{k}^2\right)}{\vec{k}^2 (\vec{k} + \hat{n})^2} \left((\vec{k}^2) + (\vec{k} + \hat{n})^2) \right),$$

$$S_{c,b\bar{n}}(z_{\mathrm{cut}}, \theta > \theta_{ab}) = C_\epsilon^{\mathrm{out}} \mathbf{T}_b \cdot \mathbf{T}_{\bar{n}} \int \frac{d^{2-2\epsilon} k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^2\right) \Theta\left((\vec{k} + \hat{n})^2 - \vec{k}^2\right)}{\vec{k}^2 (\vec{k} + \hat{n})^2} \left((\vec{k}^2) + (\vec{k} + \hat{n})^2) \right),$$

$$\tag{A.56}$$

where we have defined the constant

$$C_\epsilon^{\mathrm{out}} = -\frac{4\alpha_s}{\epsilon} \left(\frac{\mu}{z_{\mathrm{cut}} Q (n_a \cdot n_b)^{1/2}} \right)^{2\epsilon}. \tag{A.57}$$

To isolate the $\vec{k} = 0$ divergence, we now add and subtract the integral

$$C_\epsilon^{\text{out}} \int \frac{d^{2-2\epsilon}k}{(2\pi)^{2-2\epsilon}} \frac{\Theta\left(1 - \vec{k}^{\,2}\right)}{\vec{k}^{\,2}} = \frac{2C_\epsilon^{\text{out}}}{(4\pi)^{1-\epsilon}\,\Gamma(1-\epsilon)} \,. \tag{A.58}$$

We then find that the divergences and μ-dependent logs have the structure

$$S_{c,ab}(z_{\text{cut}}, \theta > \theta_{ab}) = -2\frac{\alpha_s}{\pi}\mathbf{T}_a \cdot \mathbf{T}_b \left(\frac{4}{\epsilon}\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} + 8\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi}L_{\theta>\theta_{ab}} \right) + O(\epsilon^0), \tag{A.59}$$

$$S_{c,a\bar{n}}(z_{\text{cut}}, \theta > \theta_{ab}) = -2\frac{\alpha_s}{\pi}\mathbf{T}_a \cdot \mathbf{T}_{\bar{n}} \left(\frac{1}{\epsilon^2} + \frac{2}{\epsilon}\left(L_{\theta>\theta_{ab}} - 2\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right) - 2L_{\theta>\theta_{ab}}^2 + 8\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi}L_{\theta>\theta_{ab}} \right) + O(\epsilon^0),$$

$$S_{c,b\bar{n}}(z_{\text{cut}}, \theta > \theta_{ab}) = -2\frac{\alpha_s}{\pi}\mathbf{T}_b \cdot \mathbf{T}_{\bar{n}} \left(\frac{1}{\epsilon^2} + \frac{2}{\epsilon}\left(L_{\theta>\theta_{ab}} - 2\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right) - 2L_{\theta>\theta_{ab}}^2 + 8\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi}L_{\theta>\theta_{ab}} \right) + O(\epsilon^0),$$

where the logarithm is defined as

$$L_{\theta>\theta_{ab}} = \log\left(\frac{\mu(z(1-z))^{1/2}}{z_{\text{cut}}\left(e_2^{(2)}\right)^{1/2}Q} \right) \,. \tag{A.60}$$

It is again straightforward to extract the anomalous dimensions from the above results, and perform the renormalization group evolution.

A.9 Anomalous dimensions

Here we show that the anomalous dimensions sum to zero, showing the consistency of the factorization formula at the one-loop level. We label the anomalous dimension's flavor structure to distinguish different channels. For $q \to qg$ splitting, the anomalous dimensions calculated above are given by

$$\gamma_H^q = \frac{\alpha_s C_F}{\pi}\left(-2\log\frac{\mu^2}{Q^2} - 3 \right) \,, \tag{A.61}$$

$$\gamma_{H_2}^{q\to qg} = -2\frac{\alpha_s}{\pi}\left(\frac{C_A}{2} - C_F \right)\log z_q - \frac{\alpha_s}{\pi}C_A\log\frac{4\mu^2}{e_2^{(2)}Q^2}$$

$$\qquad + \frac{\alpha_s}{\pi}C_A\log(1 - z_q) - \frac{\alpha_s}{2\pi}\left(\frac{11}{3}C_A - \frac{2}{3}n_f \right) \,, \tag{A.62}$$

$$\gamma_{J_q} = \frac{\alpha_s}{2\pi}C_F\left[4\log\left(4\frac{\left(e_2^{(2)}\right)^2}{z_q^2(1-z_q)}\tilde{e}_3^{(2)}\frac{\mu^2}{Q^2} \right) + 3 \right] \,, \tag{A.63}$$

$$\gamma_{J_g} = \frac{\alpha_s}{2\pi}C_A\left[4\log\left(4\frac{\left(e_2^{(2)}\right)^2}{z_q(1-z_q)^2}\tilde{e}_3^{(2)}\frac{\mu^2}{Q^2} \right) + \frac{11C_A - 2n_f}{3C_A} \right] \,, \tag{A.64}$$

$$\gamma_S^q = 2\frac{\alpha_s C_F}{\pi}\log\frac{\mu^2}{z_{\text{cut}}^2 Q^2} \,, \tag{A.65}$$

$$\gamma_{C_s}^{q \to qg} = -\frac{\alpha_s}{\pi} C_A \left[\log \left(\frac{4(e_2^{(2)})^5 (\tilde{e}_3^{(2)})^2 \mu^2}{z_q^3 (1-z_q)^3 Q^2} \right) - \frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right]$$

$$- \frac{\alpha_s}{\pi} C_F \left[\log \left(\frac{4(e_2^{(2)})^5 (\tilde{e}_3^{(2)})^2 \mu^2}{z_q^3 (1-z_q)^3 Q^2} \right) + \frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right] , \tag{A.66}$$

$$\gamma_{S_c}^{q \to qg} = -\frac{\alpha_s}{\pi} C_A \left(\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right) - \frac{\alpha_s}{\pi} C_F \left[\log \left(\frac{4 z_q (1-z_q) \mu^2}{z_{\text{cut}}^2 e_2^{(2)} Q^2} \right) - \frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right] . \tag{A.67}$$

For $g \to gg$ splitting, we have

$$\gamma_H^q = \frac{\alpha_s C_A}{\pi} \left(-2 \log \frac{\mu^2}{Q^2} - \beta_0 \right) , \tag{A.68}$$

$$\gamma_{H_2}^{g \to gg} = -\frac{\alpha_s}{\pi} \left[-C_A \log \left(\frac{4\mu^2}{e_2^{(2)} z(1-z) Q^2} \right) + \frac{11}{6} C_A - \frac{2}{6} n_f \right] , \tag{A.69}$$

$$\gamma_S^g = \gamma_S^q \Big|_{C_F \to C_A} , \tag{A.70}$$

$$\gamma_{C_s}^{g \to gg} = \gamma_{C_s}^{q \to qg} \Big|_{C_F \to C_A} , \tag{A.71}$$

$$\gamma_{S_c}^{g \to gg} = \gamma_{S_c}^{q \to qg} \Big|_{C_F \to C_A} . \tag{A.72}$$

One can deduce the anomalous dimension structure for $g \to q\bar{q}$ using the appropriate color generators in the collinear-soft matrix elements, and the matching for the splitting given in [15]. To achieve NLL accuracy, one must include the contribution from the two-loop cusp anomalous dimension to the coefficient multiplying logarithmic terms in each of the anomalous dimensions. For $e^+ e^- \to$ hadrons events in which we divide the event into hemispheres, groom each hemisphere, and then measure $e_2^{(2)}$ and $e_3^{(2)}$ on each hemisphere, the anomalous dimensions must satisfy

$$0 = \frac{\gamma_H + \gamma_S}{2} + \gamma_{H_2} + \gamma_{J_q} + \gamma_{J_g} + \gamma_{C_s} + \gamma_{S_c} . \tag{A.73}$$

One can verify that indeed this is satisfied, demonstrating consistency of the factorization at one-loop, or NLL accuracy. Due to the highly non-trivial combinations of scales appearing in the different functions, this provides a strong cross-check of our calculation and factorization formula.

A.10 Description of resummation, scale choices and profiles

Since we are resumming to NLL, the contribution to the cross section from each factorized function is given by the formula

$$F(\mu) = F(\mu_f) \exp \left(\int_{\mu_f}^{\mu} \frac{d\mu'}{\mu'} \gamma^F(\alpha_s(\mu'), \mu') \right) , \tag{A.74}$$

where γ^F is the appropriate anomalous dimension for the considered function, and μ_f is the scale where we run the function to. For canonical scale setting, μ_f is where all large

logarithms are minimized. To perform the resummation, we substitute in eq. (A.15) the resummed expression for each function. In general, so that we may make use of profiles to control the precise value of the resummation scale where needed, and so that we can match to the fixed order result, we keep all terms in the renormalized functions that explicitly contribute to the anomalous dimensions. Thus, when we turn off resummation, we will recover explicitly the differential cross section. When we resum we always scale set in the cumulative distribution. That is, we exponentiate all anomalous dimensions in Laplace space, perform the inverse Laplace transform with generic endpoints to the RG evolution, and integrate to get the cumulative distribution. *Then* we set all scales to their canonical values (which are given below), and take the derivative to get the differential distribution. For a detailed discussion of how to implement the resummation procedure in SCET, see [112].

The canonical scales, given as functions of D_2, are

$$\mu_H = Q \,, \tag{A.75}$$

$$\mu_{H_2} = \frac{1}{2} Q \sqrt{e_2^{(2)} z(1-z)} \,, \tag{A.76}$$

$$\mu_{C_s} = D_2 \sqrt{e_2^{(2)} (z(1-z))^3} \frac{Q}{2} \,, \tag{A.77}$$

$$\mu_{S_c} = z_{\text{cut}} \frac{Q}{2} \sqrt{\frac{e_2^{(2)}}{z(1-z)}} \,, \tag{A.78}$$

$$\mu_{J_a} = \sqrt{D_2 \, e_2^{(2)} z(1-z)^2} \frac{Q}{2} \,, \tag{A.79}$$

$$\mu_{J_b} = \sqrt{D_2 \, e_2^{(2)} z^2(1-z)} \frac{Q}{2} \,, \tag{A.80}$$

$$\mu_S = z_{\text{cut}} Q \,, \tag{A.81}$$

$$\mu = \sqrt{e_2^{(2)}} \frac{Q}{2} . \tag{A.82}$$

When performing the resummation, we take the wide-angle soft scale as the common scale where we factorize. When assessing resummation uncertainties, we vary all scales up and down by a factor of two. To handle the Landau pole in the running coupling, we smoothly "freeze out" the running coupling as a function of its scale at 1 GeV, so that it is simply a constant function below this value, and vary this freeze out scale up and down by 0.5 GeV. In general we have very little sensitivity to the freezing scale of the running coupling. To achieve NLL accuracy, we also promote the coefficient of the logarithmic terms in the anomalous dimensions in appendix A.9 to two-loop accuracy, as given by the perturbative expansion of the cusp anomalous dimension, and the running of the coupling (including when integrating the anomalous dimensions).

To turn off the resummation, we use the simple profile

$$p(x; t, s) = \Theta(x - t) \left[1 - \exp\left(-s(x - t)^2\right) \right] \,. \tag{A.83}$$

This function is zero below t, and asymptotes to one as $x - t \gg \frac{1}{s}$. Using this function, the scale choice for a function F in the factorization formula can be profiled as

$$\mu_F^{\text{profile}}(D_2) = \mu_F \left(1 - p(D_2; t, s)\right) + \mu \cdot p(D_2; t, s) \,, \tag{A.84}$$
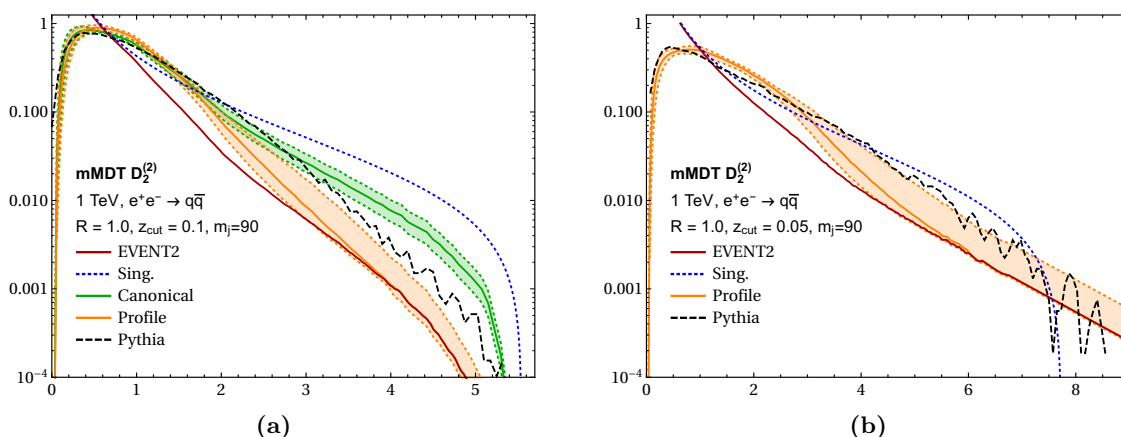
Figure 19. Distributions of $D_2^{(2)}$ measured on mMDT groomed quark jets from the process $e^+e^- \to h$ at the 1 TeV, for various matching schemes, compared against parton-level PYTHIA, the fixed distribution from Event2, and the singular part of the distribution. The groomed jet mass is fixed at 90 GeV. (a) Distributions of $D_2^{(2)}$ at $z_{cut} = 0.1$. (b) Distributions of $D_2^{(2)}$ at $z_{cut} = 0.05$.

where μ is the common factorization scale in eq. (A.82), and μ_F is the canonical choice. We chose as default value for the profile transition points

$$s = 4z_{cut}, \qquad t = z_{cut}. \tag{A.85}$$

This choice ensures that canonical resummation effectively dominates at $D_2 \sim 1$, and that the resummation is turned off at the endpoint $D_2 = 1/(2z_{cut})$. As part of the uncertainty estimate (in addition to scale variations of the canonical resummation scales), these choices were varied by 50%. We found that these profiles were sufficient to numerically cancel any zero that developed by dividing by the singular result in eq. (3.20), when the singular result developed a zero in the physical range. A more sophisticated profile would ensure that such a cancellation would happen exactly, but we found negligible differences between the matched result and the fixed-ordered cross section even in a small neighborhood about the zero, given our numerical accuracy and sampling of the matched spectra.

The differences between the canonical scales setting and the profile scale setting is illustrated in figure 19a. The profile scale setting shifts the peak of the distribution to larger values of $D_2^{(2)}$, and the canonical scheme is more consistent with the MC distribution found from PYTHIA. In figure 19b we give an example where the profile scale setting must be used when performing multiplicative matching, due to the singularity that would otherwise be induced by not turning off the resummation. Again, the profiled distribution has its peak shifted to larger values of the distribution when compared against PYTHIA. For the background distributions considered in this paper, only the quark jet with $z_{cut} = 0.05$ needed a profile scheme.

B   Ingredients for collinear-soft subjets

In this appendix, we present the calculations of the functions in the collinear-soft subjets factorization formula. The notation and formulation of the calculations will be identical to that presented in appendix A. The factorization formula is

$$\frac{d^3\sigma}{dz de_2^{(\alpha)} de_3^{(\alpha)}} = H(Q^2)S(z_{\text{cut}})H_2^{sj}(z, e_2^{(\alpha)}, z_{\text{cut}})C_s(e_3^{(\alpha)}, \theta < \theta_{12}) \otimes J_{sc}(e_3^{(\alpha)}) \otimes J(e_3^{(\alpha)}). \quad \text{(B.1)}$$

The convolutions can be removed by Laplace transforming in $e_3^{(2)}$, after which we find

$$\frac{d^3\sigma}{dz d\tilde{e}_2^{(\alpha)} d\tilde{e}_3^{(\alpha)}} = H(Q^2)S(z_{\text{cut}})H_2^{sj}(z, e_2^{(\alpha)}, z_{\text{cut}})\tilde{C}_s(\tilde{e}_3^{(\alpha)}, \theta < \theta_{12})\tilde{J}_{sc}(\tilde{e}_3^{(\alpha)})\tilde{J}(\tilde{e}_3^{(\alpha)}). \quad \text{(B.2)}$$

We again denote the Laplace transformed variables and functions with a tilde. The calculation of the low scale functions has been presented in the previous section, so we will focus on the hard function calculation.

B.1   Hard function calculation

The hard function for the collinear-soft subjet region of phase space has two parts at one-loop: there is the pure virtual term, which is familiar, and also the term with two emissions, where one emission is removed by the soft drop groomer. We focus on the case where $\alpha = 2$. The tree level integrand is given by integrating over the square of the tree-level soft current for the n and $\bar{n}$ dipole

$$H_{cs}^{(T)}(z, e_2^{(2)}, z_{\text{cut}}) = \int \frac{d^d k_1}{(2\pi)^d} 2\pi\delta(k_1^2)\delta\left(z - \frac{\bar{n} \cdot k_1}{Q}\right)\delta\left(e_2^{(2)} - \frac{n \cdot k_1}{Q}\right)|\mathcal{M}_{n\bar{n}}^R(k_1)|^2,$$
$$= H_2^{(T)}(z, e_2^{(2)}, z_{\text{cut}})\Big|_{z\to 0}, \quad \text{(B.3)}$$

and we can see that it exactly matches the soft limit of the splitting function. The one-loop contribution to the hard function can be calculated from

$$H_{cs}^{(1)}(z, e_2^{(2)}, z_{\text{cut}}) = \int \frac{d^d k_1}{(2\pi)^d} 2\pi\delta(k_1^2)\delta\left(z - \frac{\bar{n} \cdot k_1}{Q}\right)\delta\left(e_2^{(2)} - \frac{n \cdot k_1}{Q}\right)\Bigg\{|\mathcal{M}_{n\bar{n}}^{RV}(k_1)|^2$$
$$+ 2\int \frac{d^d k_2}{(2\pi)^d} 2\pi\delta(k_2^2)|\mathcal{M}_{n\bar{n}}^{RR}(k_1, k_2)|^2(1 - \Theta_{\text{SD}})\Theta_{z_{\text{cut}}}\Bigg\}. \quad \text{(B.4)}$$

The matrix elements are given by squaring the one soft emission plus virtual current, and two soft emission currents, where the hard directions are given by the n and $\bar{n}$ directions. The double real emission is obtained from [113]. The soft drop geometrical constraints are again given by eq. (A.34), with the substitutions $n_a \to n$, $n_b \to \frac{k_1}{\bar{n}\cdot k_1}$, and $k \to k_2$. The constraint on the momentum fraction of k_2 is given by eq. (A.55), with $k \to k_2$. Critically, we note that the soft drop conditions are *purely geometrical*: we can rephrase the soft drop constraint of eq. (A.34) in terms of angles, as is done in eq. (E.3). Thus regardless of how one power counts the relative momentum fractions of emissions k_1 and k_2, the exact same

soft drop condition applies to an emission which is not clustered. Then all that can change is the expansion of the matrix element $|\mathcal{M}_{n\bar{n}}^{RR}(k_1, k_2)|^2$, whether we take it to be strongly-ordered in the energy of the emissions, or not. If we strongly order, then we reproduce the calculation of appendix A.8, once we factor out the tree-level result for the k_1 emission that forms a leg of the hard $1 \to 2$ splitting and take the limit that $z \to 0$. With this observation we may write eq. (B.4) as

$$
\begin{aligned}
H_{cs}^{(1)}(z, e_2^{(2)}, z_{\text{cut}}) = &\int \frac{d^d k_1}{(2\pi)^d} 2\pi \delta(k_1^2) \delta\left(z - \frac{\bar{n} \cdot k_1}{Q}\right) \delta\left(e_2^{(2)} - \frac{n \cdot k_1}{Q}\right) \left\{ |\mathcal{M}_{n\bar{n}}^{RV}(k_1)|^2 \right. \\
&+ 2 \int \frac{d^d k_2}{(2\pi)^d} 2\pi \delta(k_2^2) \left(|\mathcal{M}_{n\bar{n}}^{RR}(k_1, k_2)|^2 \Big|_{\text{s.o.: } k_1^0 \gg k_2^0} \right) (1 - \Theta_{\text{SD}}) \Theta_{z_{\text{cut}}} \\
+ 2 \int \frac{d^d k_2}{(2\pi)^d} 2\pi \delta(k_2^2) &\left. \left(|\mathcal{M}_{n\bar{n}}^{RR}(k_1, k_2)|^2 - |\mathcal{M}_{n\bar{n}}^{RR}(k_1, k_2)|^2 \Big|_{\text{s.o.: } k_1^0 \gg k_2^0} \right) (1 - \Theta_{\text{SD}}) \Theta_{z_{\text{cut}}} \right\} \\
= H_{cs}^{(T)}(z, e_2^{(2)}, z_{\text{cut}}) &\left(S_c^{(1)}(z_{\text{cut}}, \Theta > \Theta_{ab}) + H_2^{(1)}(z, e_2^{(2)}) \Big|_{z_g \to 0} + \text{const} \right), \quad \text{(B.5)}
\end{aligned}
$$

where the subscript "s.o." denotes strongly ordered. Here $S_c^{(1)}(z_{\text{cut}}, \theta > \theta_{ab})$ is given by eq. (A.53), and corresponds to the strongly ordered contribution, and the virtual correction is fully captured by taking the energy fraction of the gluon to zero ($z_g \to 0$) in the collinear-splitting function calculation. The residual constant is formally given as

$$
\begin{aligned}
\text{const} = &\frac{2}{H_{cs}^{(T)}(z, e_2^{(2)}, z_{\text{cut}})} \int \frac{d^d k_1}{(2\pi)^d} \pi \delta(k_1^2) \delta\left(z - \frac{\bar{n} \cdot k_1}{Q}\right) \delta\left(e_2^{(2)} - \frac{n \cdot k_1}{Q}\right) \quad \text{(B.6)} \\
&\times \int \frac{d^d k_2}{(2\pi)^d} 2\pi \delta(k_2^2) \left(|\mathcal{M}_{n\bar{n}}^{RR}(k_1, k_2)|^2 - |\mathcal{M}_{n\bar{n}}^{RR}(k_1, k_2)|^2 \Big|_{\text{s.o.: } k_1^0 \gg k_2^0} \right) (1 - \Theta_{\text{SD}}) \Theta_{z_{\text{cut}}}.
\end{aligned}
$$

We label this as a constant, since this contribution is easily found to be ultraviolet and infrared finite. We may therefore set $\epsilon = 0$ in its calculation, and it will give no contribution to the anomalous dimension. We emphasize that the ability to simplify the matching in the collinear-soft subjet region is purely because the soft drop constraint is geometrical, and thus is not sensitive to the exact relative power counting of the two emissions.

B.2 Hard function anomalous dimension

We give the collinear-soft subjet anomalous dimensions for both quark and gluon initiated jets. We have

$$
\begin{aligned}
\gamma_{H_{cs}}^{q \to gq} = &-\frac{\alpha_s}{\pi} C_A \log \frac{4\mu^2}{e_2^{(2)}(1 - z_q) Q^2} - \frac{\alpha_s}{2\pi} \left(\frac{11}{3} C_A - \frac{2}{3} n_f \right) \\
&- \frac{\alpha_s}{\pi} C_A \left(\frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right) - \frac{\alpha_s}{\pi} C_F \left[\log\left(\frac{4(1 - z_q)\mu^2}{z_{\text{cut}}^2 e_2^{(2)} Q^2} \right) - \frac{\text{Cl}_2(\frac{\pi}{3})}{\pi} \right], \quad \text{(B.7)}
\end{aligned}
$$

$$
\gamma_{H_{cs}}^{g \to gg} = -\frac{\alpha_s}{\pi} C_A \log \frac{4\mu^2}{e_2^{(2)} z Q^2} - \frac{\alpha_s}{2\pi} \left(\frac{11}{3} C_A - \frac{2}{3} n_f \right) - \frac{\alpha_s}{\pi} C_A \log \frac{4z\mu^2}{z_{\text{cut}}^2 e_2^{(2)} Q^2}. \quad \text{(B.8)}
$$

We note that we do not have large logs of z or $1 - z_q$ over z_{cut}, since in this region of phase space these scales are parametrically the same.

C Ingredients for signal factorization

In this appendix, we present the results for the signal factorization formula. The factorization formula is

$$\frac{d^2\sigma}{dz\,de_2^{(\alpha)}de_3^{(\alpha)}} = H(Q^2)H_2^{Z\to q\bar{q}}(z, e_2^{(\alpha)}, m_Z^2)J_1(e_3^{(\alpha)}) \otimes J_2(e_3^{(\alpha)}) \otimes S(e_3^{(\alpha)}, z_{\rm cut}). \qquad (\text{C.1})$$

Laplace transforming in $e_3^{(\alpha)}$ removes the convolutions, and factoring out the global contributions, we have

$$\frac{d^2\sigma}{dz\,de_2^{(\alpha)}d\tilde{e}_3^{(\alpha)}} = H(Q^2)H_2^{Z\to q\bar{q}}(z, e_2^{(\alpha)}, m_Z^2)\tilde{J}_1(\tilde{e}_3^{(\alpha)})\tilde{J}_2(\tilde{e}_3^{(\alpha)})\tilde{S}(\tilde{e}_3^{(\alpha)}, z_{\rm cut}) \qquad (\text{C.2})$$

$$= H(Q^2)H_2^{Z\to q\bar{q}}(z, e_2^{(\alpha)}, m_Z^2)\tilde{J}_1(\tilde{e}_3^{(\alpha)})\tilde{J}_2(\tilde{e}_3^{(\alpha)})C_s(\tilde{e}_3^{(\alpha)}, \theta < \theta_{ab})S_c(z_{\rm cut}, \theta > \theta_{ab})C_s^{\rm NG}(\tilde{e}_3^{(\alpha)}, z_{\rm cut}).$$

Results in this appendix will be expressed in Laplace space. Note that the anomalous dimension of the hard function $H(Q^2)$ is zero in QCD because the production of the Z boson occurs in the QCD vacuum.

C.1 Soft matrix elements

The soft matrix elements are simplified relative to the earlier case of QCD splittings, since we have no Wilson line in the direction of the recoil. We therefore have

$$C_s(\tilde{e}_3^{(\alpha)}, \theta < \theta_{ab}) = \frac{1}{N}{\rm tr}\Big[\langle 0|T\{S_aS_b\}\exp\Big(-\tilde{e}_3^{(\alpha)}\mathbf{E_3}^{(\alpha)}\big|_{{\rm SD}a,b}\Big)\bar{T}\{S_aS_b\}|0\rangle\Big], \qquad (\text{C.3})$$

$$S_c(z_{\rm cut}, \theta > \theta_{ab}) = \frac{1}{N}{\rm tr}\Big[\langle 0|T\{S_aS_b\}\Theta_{\rm SD}(a, b, z_{\rm cut})\bar{T}\{S_aS_b\}|0\rangle\Big], \qquad (\text{C.4})$$

$$C_s^{\rm NG}(\tilde{e}_3^{(\alpha)}, z_{\rm cut}) = \frac{1}{N}\frac{{\rm tr}\Big[\langle 0|T\{S_aS_b\}\Theta_{\rm SD}(a, b, z_{\rm cut})\exp\Big(-\tilde{e}_3^{(\alpha)}\mathbf{E_3}^{(\alpha)}\big|_{{\rm SD}a,b}\Big)\bar{T}\{S_aS_b\}|0\rangle\Big]}{C_s(\tilde{e}_3^{(\alpha)}, \theta < \theta_{ab})S_c(z_{\rm cut}, \theta > \theta_{ab})}. \qquad (\text{C.5})$$

As before, S_q are soft Wilson lines, as defined in eq. (A.17).

C.2 Hard decay function

The anomalous dimension of the hard decay function, $H_2^{Z\to q\bar{q}}(e_2^{(\alpha)}, m_Z^2)$, is identical to the anomalous dimension for the hard function of $e^+e^- \to q\bar{q}$, at $Q^2 = m_Z^2$. We therefore have

$$\gamma_{H_2} = \frac{\alpha_s C_F}{\pi}\left(-2\log\frac{\mu^2}{m_Z^2} - 3\right). \qquad (\text{C.6})$$

Identifying the mass of the jet with the two-point energy correlation function as

$$m_Z^2 = e_2^{(2)}E_J^2 = \frac{e_2^{(2)}Q^2}{4}, \qquad (\text{C.7})$$

this anomalous dimension can also be expressed as

$$\gamma_{H_2} = \frac{\alpha_s C_F}{\pi}\left(-2\log\frac{4\mu^2}{e_2^{(2)}Q^2} - 3\right). \qquad (\text{C.8})$$

The tree-level matrix element for Z boson decay to quarks is well-known and was first calculated in ref. [114].

C.3 Jet functions

The subjets of the decay of a Z boson are quarks. The jet functions have the corresponding anomalous dimensions

$$\gamma_{J_1} = \frac{\alpha_s C_F}{2\pi} \left[4 \log \left(4 \frac{\left(e_2^{(2)}\right)^2}{z_a^2 z_b} \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + 3 \right] , \tag{C.9}$$

$$\gamma_{J_2} = \frac{\alpha_s C_F}{2\pi} \left[4 \log \left(4 \frac{\left(e_2^{(2)}\right)^2}{z_a z_b^2} \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + 3 \right] . \tag{C.10}$$

Here z_a and z_b are the energy fractions of the two subjets, with $z_a + z_b = 1$.

C.4 Collinear-soft function

The soft radiation from the dipole of the subjets is constrained by the same restrictions as for background jets. In this case, however, there is only one non-trivial dipole formed from the subjets with color factor $\mathbf{T}_a \cdot \mathbf{T}_b = -C_F$. Therefore, the anomalous dimension of this function is just

$$\gamma_{C_s} = -2 \frac{\alpha_s C_F}{\pi} \left[\log \left(\frac{4(e_2^{(2)})^5 (\tilde{e}_3^{(2)})^2 \mu^2}{z_a^3 z_b^3 Q^2} \right) - \frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} \right] . \tag{C.11}$$

C.5 Soft-collinear function

As with the collinear-soft function, there is only one contribution to the soft-collinear function, namely from emissions that are emitted at a wide angle. We find that the anomalous dimension of the soft-collinear function is

$$\gamma_{S_c} = -2 \frac{\alpha_s C_F}{\pi} \left(\frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} \right) . \tag{C.12}$$

C.6 Consistency

The anomalous dimensions of the functions in the factorization formula are

$$\gamma_{H_2} = \frac{\alpha_s C_F}{\pi} \left(-2 \log \frac{4\mu^2}{e_2^{(2)} Q^2} - 3 \right) , \tag{C.13}$$

$$\gamma_{J_1} = \frac{\alpha_s C_F}{2\pi} \left[4 \log \left(4 \frac{\left(e_2^{(2)}\right)^2}{z_a^2 z_b} \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + 3 \right] , \tag{C.14}$$

$$\gamma_{J_2} = \frac{\alpha_s C_F}{2\pi} \left[4 \log \left(4 \frac{\left(e_2^{(2)}\right)^2}{z_a z_b^2} \tilde{e}_3^{(2)} \frac{\mu^2}{Q^2} \right) + 3 \right] , \tag{C.15}$$

$$\gamma_{C_s} = -2 \frac{\alpha_s C_F}{\pi} \left[\log \left(\frac{4(e_2^{(2)})^5 (\tilde{e}_3^{(2)})^2 \mu^2}{z_a^3 z_b^3 Q^2} \right) - \frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} \right] , \tag{C.16}$$

$$\gamma_{S_c} = -2 \frac{\alpha_s C_F}{\pi} \left(\frac{\mathrm{Cl}_2(\frac{\pi}{3})}{\pi} \right) . \tag{C.17}$$

It is straightforward to verify that the sum of anomalous dimensions is 0, namely

$$0 = \gamma_{H_2} + \gamma_{J_1} + \gamma_{J_2} + \gamma_{C_s} + \gamma_{S_c}. \tag{C.18}$$

This demonstrates that the factorization formula is consistent at one-loop or NLL accuracy.

D $1 \to 3$ splitting function integration

In this appendix we describe the calculation of the D_2 distribution using the $1 \to 3$ splitting functions. The splitting function calculation on the soft dropped jet is given by

$$\frac{d\sigma}{de_2 dD_2} \propto (2\pi)^{(3-2\epsilon)} \int [d^d k_a]_+ [d^d k_b]_+ [d^d k_c]_+ \delta(2E_J - \bar{n}\cdot k_a - \bar{n}\cdot k_b - \bar{n}\cdot k_c) \delta^{(d-2)}\Big(k_{a\perp} + k_{b\perp} + k_{c\perp}\Big)$$

$$\times \delta\left(e_2 - \frac{1}{E_J^2}((2k_a\cdot k_b) + (2k_a\cdot k_c) + (2k_b\cdot k_c))\right) \delta\left(D_2 - \frac{8(2k_a\cdot k_b)(2k_a\cdot k_c)(2k_b\cdot k_c)}{E_J^3 (e_2^{(2)})^3 \bar{n}\cdot k_a \bar{n}\cdot k_b \bar{n}\cdot k_c}\right)$$

$$\times \mathrm{SD}\Big(z_{\mathrm{cut}}, k_a, k_b, k_c\Big) \Theta_J\Big(R;, k_a, k_b, k_c\Big) \mathrm{Split}\Big(k_a, k_b, k_c; \bar{n}\Big). \tag{D.1}$$

The soft drop and jet region constraints are defined as

$$\Theta_J\Big(R;, k_a, k_b, k_c\Big) = \Theta\left(\tan^2\frac{R}{2} - \frac{n\cdot k_a}{\bar{n}\cdot k_a}\right)\Theta\left(\tan^2\frac{R}{2} - \frac{n\cdot k_b}{\bar{n}\cdot k_b}\right)\Theta\left(\tan^2\frac{R}{2} - \frac{n\cdot k_c}{\bar{n}\cdot k_c}\right), \tag{D.2}$$

$$\mathrm{SD}\Big(z_{\mathrm{cut}}, k_a, k_b, k_c\Big) = \tag{D.3}$$

$$\times \left[\Theta\left(\frac{1}{2E_J}\min(\bar{n}\cdot(k_b+k_c), \bar{n}\cdot k_a) - z_{\mathrm{cut}}\right)\Theta\left(\frac{k_a\cdot k_b}{\bar{n}\cdot k_a \bar{n}\cdot k_b} - \frac{k_b\cdot k_c}{\bar{n}\cdot k_b \bar{n}\cdot k_c}\right)\Theta\left(\frac{k_a\cdot k_c}{\bar{n}\cdot k_a \bar{n}\cdot k_c} - \frac{k_b\cdot k_c}{\bar{n}\cdot k_b \bar{n}\cdot k_c}\right)\right.$$

$$+\Theta\left(\frac{1}{2E_J}\min(\bar{n}\cdot(k_a+k_b), \bar{n}\cdot k_c) - z_{\mathrm{cut}}\right)\left(\frac{k_a\cdot k_c}{\bar{n}\cdot k_a \bar{n}\cdot k_c} - \frac{k_a\cdot k_b}{\bar{n}\cdot k_a \bar{n}\cdot k_b}\right)\left(\frac{k_b\cdot k_c}{\bar{n}\cdot k_b \bar{n}\cdot k_c} - \frac{k_a\cdot k_b}{\bar{n}\cdot k_a \bar{n}\cdot k_b}\right)$$

$$\left.+\Theta\left(\frac{1}{2E_J}\min(\bar{n}\cdot(k_a+k_c), \bar{n}\cdot k_b) - z_{\mathrm{cut}}\right)\Theta\left(\frac{k_a\cdot k_b}{\bar{n}\cdot k_a \bar{n}\cdot k_b} - \frac{k_a\cdot k_c}{\bar{n}\cdot k_a \bar{n}\cdot k_c}\right)\Theta\left(\frac{k_b\cdot k_c}{\bar{n}\cdot k_b \bar{n}\cdot k_c} - \frac{k_a\cdot k_c}{\bar{n}\cdot k_a \bar{n}\cdot k_c}\right)\right].$$

We note that the jet region constraint is that of a simple cone algorithm, where we have drawn a cone of radius R about the direction n. Further, we have demanded that the total transverse momentum of all emissions inside this cone is zero. Because of the soft drop constraint, we could actually take the effective cone radius to be infinite, that is, the jet be the whole sphere. However, in our numerical integration of the splitting functions, we keep the jet region constraint so that we can explicitly test the R and $e_2^{(2)}$ independence of the fixed order result.

D.1 Solving observable constraints

To integrate the above phase space, we express the kinematic variables as

$$\bar{n}\cdot k_a = 2E_J z_a, \qquad n\cdot k_a = \frac{k_{a\perp}^2}{2E_J z_a}, \qquad \bar{n}\cdot k_b = 2E_J z_b, \quad n\cdot k_b = \frac{k_{b\perp}^2}{2E_J z_b},$$

$$\bar{n}\cdot k_c = 2E_J(1 - z_a - z_b), \qquad n\cdot k_c = \frac{k_{c\perp}^2}{2E_J(1 - z_a - z_b)}, \qquad k_{c\perp} = -k_{a\perp} - k_{b\perp}. \quad \text{(D.4)}$$

We can then shift $k_{a\perp}$ as

$$k_{a\perp} \to k_{a\perp} - \frac{z_b}{1 - z_a} k_{b\perp}. \tag{D.5}$$

This allows us to solve the $e_2^{(2)}$ delta function constraint as

$$k_{a\perp}^2 = \frac{z_a(1-z_a-z_b)}{z_b(1-z_b)}\left(e_2^{(2)}E_J^2 z_b(1-z_b) - k_{b\perp}^2\right), \qquad k_{b\perp}^2 < e_2^{(2)}E_J^2 z_b(1-z_b)\,. \quad (D.6)$$

We again rescale $k_{b\perp}^2$ as

$$k_{b\perp}^2 = e_2^{(2)}E_J^2 z_b(1-z_b)k^2\,. \tag{D.7}$$

Our left over variables are z_a, z_b, k^2 and ϕ, where ϕ is a relative angle in the transverse plane. The Lorentz invariants have the functional form

$$k_a\cdot k_b = E_J^2 e_2^{(2)}\left(\frac{k^2(z_a-(1-z_a)z_b+z_b^2)+z_b(1-z_a-z_b)-2\sqrt{k^2(1-k^2)z_az_b(1-z_a-z_b)}\cos\phi}{2(1-z_b)}\right),$$

$$k_b\cdot k_c = E_J^2 e_2^{(2)}\left(\frac{k^2(1-z_b-z_a(1+z_b))+z_az_b+2\sqrt{k^2(1-k^2)z_az_b(1-z_a-z_b)}\cos\phi}{2(1-z_b)}\right),$$

$$k_a\cdot k_c = E_J^2 e_2^{(2)}\left(\frac{(1-k^2)(1-z_b)}{2}\right). \tag{D.8}$$

It is convenient to introduce the variables

$$s_{ij} = \frac{k_i\cdot k_j}{E_J^2 e_2^{(2)}}\,, \qquad a\le i<j\le c\,. \tag{D.9}$$

Further we use the fact that if we rescale the invariants $k_i\cdot k_j$ in the splitting function by a factor λ, this simply induces an overall factor of λ^{-2}

$$\mathrm{Split}\left(k_a,k_b,k_c;\bar{n}\right) \to \frac{1}{\lambda^2}\mathrm{Split}\left(k_a,k_b,k_c;\bar{n}\right). \tag{D.10}$$

The final phase space then has the form

$$\frac{d\sigma}{de_2^{(2)}dD_2} \propto \frac{g^4}{e_2^{(2)}}(2\pi)^{-6+4\epsilon}\left(\frac{2\pi^{1-\epsilon}}{\Gamma(1-\epsilon)}\right)\left(\frac{\mu^2}{e_2^{(2)}E_J^2}\right)^\epsilon\left(\frac{\mu^2}{e_2^{(2)}E_J^2}\right)^\epsilon$$

$$\cdot\int_0^1 dz_a\int_0^1 dz_b\int_0^1 dk^2\int_0^\pi d\phi\,\sin^{-2\epsilon}\phi\,\Theta(1-z_a-z_b)$$

$$\cdot\left(z_az_b(1-z_a-z_b)k^2(1-k^2)\right)^{-\epsilon}\delta\left(D_2 - \frac{8s_{ab}s_{bc}s_{ac}}{z_az_b(1-z_a-z_b)}\right)$$

$$\cdot\mathrm{SD}\left(z_{\mathrm{cut}},z_a,z_b,k^2,\phi\right)\Theta_J\left(R;e_2^{(2)},z_a,z_b,k^2,\phi\right)\mathrm{Split}\left(z_a,z_b,k^2,\phi\right). \quad (D.11)$$

Note that the only dependence on $e_2^{(2)}$ enters through the jet constraints. Had we expanded the jet region constraints, we would have formally found the cross section to depend on $e_2^{(2)}$ only via an overall scaling. That is, the shape of the differential spectrum is independent of $e_2^{(2)}$. Retaining the jet region constraints we can explicitly test the independence on the jet radius at all values of D_2 by varying either $e_2^{(2)}$ or R. At finite D_2, the cross section is finite, and so we can set the dimensional regularization parameter to zero, and numerically compute the result.

D.2 Numerical integration

To perform the numerical integration, we first rescale z_b as

$$z_b \rightarrow (1 - z_a)z_b\,. \tag{D.12}$$

This gives the integral

$$
\begin{aligned}
\frac{d\sigma}{de_2^{(2)}dD_2} &\propto \frac{2g^4}{e_2^{(2)}}(2\pi)^{-6+4\epsilon}\left(\frac{2\pi^{1-\epsilon}}{\Gamma(1-\epsilon)}\right)\left(\frac{\mu^2}{e_2^{(2)}E_J^2}\right)^\epsilon \left(\frac{\mu^2}{e_2^{(2)}E_J^2}\right)^\epsilon \int_0^1 dz_a \int_0^1 dz_b \int_0^1 dk^2 \int_0^\pi d\phi \sin^{-2\epsilon}\phi \\
&\quad \cdot \left(z_a(1-z_a)^2 z_b(1-z_b)k^2(1-k^2)\right)^{-\epsilon}\delta\left(D_2 - \frac{8s_{ab}s_{bc}s_{ac}}{z_a(1-z_a)^2 z_b(1-z_b)}\right) \\
&\quad \cdot \mathrm{SD}\left(z_{\mathrm{cut}},z_a,(1-z_a)z_b,k^2,\phi\right)\Theta_J\left(R;e_2^{(2)},z_a,(1-z_a)z_b,k^2,\phi\right)(1-z_a) \\
&\quad \cdot \mathrm{Split}\left(z_a,(1-z_a)z_b,k^2,\phi\right).
\end{aligned}
\tag{D.13}
$$

We can then transform to the rapidity-like variables

$$z_a = \frac{1}{2}\left[1 + \tanh\left(\frac{y_a}{\delta}\right)\right]\,,\quad z_b = \frac{1}{2}\left[1 + \tanh\left(\frac{y_b}{\delta}\right)\right]\,,\quad k^2 = \frac{1}{2}\left[1 + \tanh\left(\frac{y_{k^2}}{\delta}\right)\right]\,. \tag{D.14}$$

This transformation has the Jacobian

$$\int_0^1 dz_a \int_0^1 dz_b \int_0^1 dk^2 = \int_{-\infty}^\infty dy_a \int_{-\infty}^\infty dy_b \int_{-\infty}^\infty dy_{k^2}\, J(y_a, y_b, y_{k^2})\,, \tag{D.15}$$

with

$$J(y_a, y_b, y_{k^2}) = \frac{1}{16\delta^3}\mathrm{sech}^2\left(\frac{y_a}{\delta}\right)\mathrm{sech}^2\left(\frac{y_b}{\delta}\right)\mathrm{sech}^2\left(\frac{y_{k^2}}{\delta}\right)\,. \tag{D.16}$$

The numerical integral is then performed at $\epsilon = 0$. To perform the integral, we randomly sample y_a, y_b, and y_{k^2} uniformly on an interval $[-y_{\max}, y_{\max}]$, while ϕ is uniformly sampled on $[0, 2\pi]$. The variable δ simply controls how often one samples the region where $D_2 \sim 1$. The value of $y_{\max}$ sets the minimal D_2 that we can integrate down to. For each generated phase space point we compute the value of corresponding value of D_2 and the weight

$$
\begin{aligned}
w(y_a, y_b, y_{k^2}, \phi) &= \mathrm{SD}\left(z_{\mathrm{cut}}, z_a, (1-z_a)z_b, k^2, \phi\right)\Theta_J\left(R; e_2^{(2)}, z_a, (1-z_a)z_b, k^2, \phi\right) \\
&\quad \cdot (1-z_a)\mathrm{Split}\left(z_a, (1-z_a)z_b, k^2, \phi\right).
\end{aligned}
\tag{D.17}
$$

Then a histogram H_{D_2}, indexed by D_2, is updated according to

$$H_{D_2} \rightarrow H_{D_2} + J(y_a, y_b, y_{k^2})\frac{w(y_a, y_b, y_{k^2}, \phi)}{\Delta D_2}\,, \tag{D.18}$$

where ΔD_2 is the width of bin at position D_2. We then divide all bins in the histogram by the total number of phase space points sampled. In the limit of infinitely narrow bins, infinite statistics, and $y_{\max} \rightarrow \infty$, the final histogram H_{D_2} is proportional to the differential spectrum

$$\frac{d\sigma}{de_2^{(2)}dD_2} \propto H_{D_2}\,. \tag{D.19}$$

Aside from the transformation to the rapidity like variables y_a, y_b and y_{k^2}, which serve to smooth out the soft and collinear singularities, no further importance sampling is performed. To actually fix the normalization of the histogram, the simplest procedure is to compare to the analytic predictions in the singular region. For a fixed bin size, we can fit for the singular behavior of the histogram at a specific soft drop parameter z_{cut}. Then we take the ratio to the analytic result in the limit $D_2 \to 0$. This selects for the ratio of the double-logarithmic terms, and gives the relative normalization of the histogram to the singular result. This normalization is the same for all other values of z_{cut} when using the same bin spacing, so that for all other z_{cut} we can then *test* that the fixed order result is reproducing the analytic double and single logarithmic structure at small D_2.

D.3 Singular cross section

To find the singular cross section for gluon splitting, $g \to gg$, we sum the collinear-soft and jet function contributions. This gives

$$\frac{d\sigma}{de_2^{(2)} dD_2} = \frac{\alpha_s}{\pi} \int_0^1 dz_a \Theta\Big(\min(z_a, 1 - z_a) - z_{\text{cut}}\Big) \frac{P_{g \to gg}(z_a)}{e_2^{(2)}} \tag{D.20}$$
$$\cdot \left(\frac{\alpha_s C_A}{\pi D_2}\right) \left(-\log D_2 - \frac{\beta_0}{C_A} - \frac{3}{2}\log\left(z_a(1 - z_a)\right)\right).$$

Performing a similar calculation for $q \to qg$ gives

$$\left.\frac{d\sigma}{de_2^{(2)} dD_2}\right|_{C_A} = \frac{\alpha_s}{\pi} \int_0^1 dz_a \Theta\Big(\min(z_a, 1 - z_a) - z_{\text{cut}}\Big) \frac{P_{q \to qg}(z_a)}{e_2^{(2)}} \tag{D.21}$$
$$\cdot \left(\frac{\alpha_s C_A}{\pi D_2}\right) \left(-\frac{1}{2}\log D_2 - \frac{\beta_0}{2C_A} - \log(1 - z_a) - \frac{1}{2}\log z_a + \frac{\text{Cl}_2(\frac{\pi}{3})}{\pi}\right),$$

$$\left.\frac{d\sigma}{de_2^{(2)} dD_2}\right|_{C_F} = \frac{\alpha_s}{\pi} \int_0^1 dz_a \Theta\Big(\min(z_a, 1 - z_a) - z_{\text{cut}}\Big) \frac{P_{q \to qg}(z_a)}{e_2^{(2)}} \tag{D.22}$$
$$\cdot \left(\frac{\alpha_s C_F}{\pi D_2}\right) \left(-\frac{1}{2}\log D_2 - \frac{3}{8} - \frac{1}{2}\log(1 - z_a) - \log z_a - \frac{\text{Cl}_2(\frac{\pi}{3})}{\pi}\right).$$

E Collinear non-global logarithms

In this appendix we give the appropriate modification of the Dasgupta-Salam Monte Carlo algorithm [46] for computing NGLs in the large-N_c limit, changed to account for the phase space constraints of the soft drop procedure. The origin of these NGLs was discussed in section 3.2.2, and the lowest order diagram that contributes was shown in figure 8. Throughout this section, we define the eikonal factor as

$$W_{xy}(j) = \frac{x \cdot y}{(x \cdot j)(j \cdot y)}. \tag{E.1}$$

Here the round bracket is defined as $(i \cdot j) \equiv 1 - \cos\theta_{ij}$.

We start the algorithm with a list of initial dipoles, $\mathcal{D}_{\text{init}}$, that depends on the flavor structure of the $1 \rightarrow 2$ splitting, and we generate a list of active radiating dipoles $\mathcal{D}$ and a list of emissions E that fail soft drop. That is, they are not clustered into the legs a and b of the hard $1 \rightarrow 2$ splitting. We introduce the phase space constraint given this list of emissions

$$C_{ab}\Big(j, E\Big) = \Theta\Big(\theta_{ab} - \theta_{aj}\Big)\Theta\Big(\theta_{bj} - \theta_{aj}\Big)\prod_{i \in E}\Theta\Big(\theta_{ij} - \theta_{aj}\Big) + \Theta\Big(\theta_{ab} - \theta_{bj}\Big)\Theta\Big(\theta_{aj} - \theta_{bj}\Big)\prod_{i \in E}\Theta\Big(\theta_{ij} - \theta_{bj}\Big). \tag{E.2}$$

Here θ_{xy} is the angle between x and y. In other words, we test emission j to see if it is closer in angle to either direction a or b than any other emission in the list E. If it is, it will pass soft drop by virtue of being clustered into the hard splitting. Finally, we introduce the one emission phase space that sets the resummation of the Sudakov or global logarithms

$$C_{ab}(j) = \Theta\Big(\theta_{ab} - \theta_{aj}\Big)\Theta\Big(\theta_{bj} - \theta_{aj}\Big) + \Theta\Big(\theta_{ab} - \theta_{bj}\Big)\Theta\Big(\theta_{aj} - \theta_{bj}\Big). \tag{E.3}$$

Though written explicitly in terms of angles, this is the same phase space as eq. (A.34), since the Lorentz products can be simplified to an angular constraint once we factor out and cancel the energy scales.

The idea of the algorithm is that in the leading logarithmic approximation for the NGLs, every emission not clustered into the a-b dipole is formally at an energy scale below Qz_{cut}, and the emissions are ordered in energy. We therefore generate the emissions at each step according to a distribution

$$\tilde{F}_{\mathcal{D}}(j) = \sum_{i \in \mathcal{D}} W_{x_i y_i}(j) - \sum_{i \in \mathcal{D}_{\text{init}}} C_{ab}(j) W_{a_i b_i}(j). \tag{E.4}$$

Here $\mathcal{D}$ is the set of decaying dipoles, and $\mathcal{D}_{\text{init}}$ the set of initial dipoles. After generating the emission, we update the list of real emissions E and the dipole list $\mathcal{D}$, so long as the emission qualifies as a real emission, rather than a virtual process. The virtual subtraction is implemented via a veto algorithm following the original Dasgupta-Salam algorithm. We allow the initial dipoles to decay until we have an emission that is clustered into the a-b dipole, before it is clustered into any other emission, that is, $C_{ab}(j, E) = 1$. This replaces the hemisphere condition to start a new event. Once this condition is met, we end the event and book the histogram. Thus, for each event we must track all the real emissions that have been created so far, and check for each new emission whether it clusters into the a-b dipole rather than any other emission generated so far.[11] The virtual subtraction condition, that is whether or not we treat the emission as a virtual correction and so reweight the event, is triggered when $C_{ab}(j) > 0$ and j was emitted from a dipole containing an initial leg. We take the high scale for the NGL evolution to be μ_{S_c}, and evolve down to the scale μ_{C_s}, see appendix A.10.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

---

[11]The clustering constraint produced by the soft drop procedure introduces a non-Markovian evolution. Whether or not we terminate the dipole evolution depends on the complete emission history up to that point. However, the generation of emissions still proceeds in a Markovian fashion, as the decay of each dipole is independent and universal.

References

[1] A.J. Larkoski, I. Moult and D. Neill, *Analytic boosted boson discrimination at the large hadron collider*, arXiv:1708.06760 [INSPIRE].

[2] A.J. Larkoski, I. Moult and B. Nachman, *Jet substructure at the large hadron collider: a review of recent advances in theory and machine learning*, arXiv:1709.04464 [INSPIRE].

[3] ATLAS collaboration, *Performance of top quark and W boson tagging in run 2 with ATLAS*, ATLAS-CONF-2017-064, CERN, Geneva Switzerland, (2017).

[4] ATLAS collaboration, *Boosted Higgs ($\to b\bar{b}$) boson identification with the ATLAS detector at $\sqrt{s} = 13\,TeV$*, ATLAS-CONF-2016-039, CERN, Geneva Switzerland, (2016).

[5] ATLAS collaboration, *Quark versus gluon jet tagging using charged particle multiplicity with the ATLAS detector*, ATL-PHYS-PUB-2017-009, CERN, Geneva Switzerland, (2017).

[6] ATLAS collaboration, *Jet mass reconstruction with the ATLAS detector in early run 2 data*, ATLAS-CONF-2016-035, CERN, Geneva Switzerland, (2016).

[7] ATLAS collaboration, *Jet reclustering and close-by effects in ATLAS run II*, ATLAS-CONF-2017-062, CERN, Geneva Switzerland, (2017).

[8] ATLAS collaboration, *In-situ measurements of the ATLAS large-radius jet response in 13 TeV pp collisions*, ATLAS-CONF-2017-063, CERN, Geneva Switzerland, (2017).

[9] CMS collaboration, *W/Z/Top tagging for run 2 in CMS*, CMS-DP-2015-043, CERN, Geneva Switzerland, (2015).

[10] CMS collaboration, *Top tagging with new approaches*, CMS-PAS-JME-15-002, CERN, Geneva Switzerland, (2015).

[11] CMS collaboration, *Performance of quark/gluon discrimination in 13 TeV data*, CMS-DP-2016-070, CERN, Geneva Switzerland, (2016).

[12] CMS collaboration, *b-tagging in boosted topologies*, CMS-DP-2015-038, CERN, Geneva Switzerland, (2015).

[13] CMS collaboration, *Jet algorithms performance in 13 TeV data*, CMS-PAS-JME-16-003, CERN, Geneva Switzerland, (2016).

[14] A.J. Larkoski, I. Moult and D. Neill, *Power counting to better jet observables*, *JHEP* **12** (2014) 009 [arXiv:1409.6298] [INSPIRE].

[15] A.J. Larkoski, I. Moult and D. Neill, *Analytic boosted boson discrimination*, *JHEP* **05** (2016) 117 [arXiv:1507.03018] [INSPIRE].

[16] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy correlation functions for jet substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [INSPIRE].

[17] ATLAS collaboration, *Performance of jet substructure techniques in early $\sqrt{s} = 13\,TeV$ pp collisions with the ATLAS detector*, ATLAS-CONF-2015-035, CERN, Geneva Switzerland, (2015).

[18] ATLAS collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8\,TeV$*, *Eur. Phys. J.* **C 76** (2016) 154 [arXiv:1510.05821] [INSPIRE].

[19] ATLAS collaboration, *Searches for heavy diboson resonances in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, *JHEP* **09** (2016) 173 [arXiv:1606.04833] [INSPIRE].

[20] ATLAS collaboration, *Search for heavy resonances decaying to a Z boson and a photon in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, *Phys. Lett.* **B 764** (2017) 11 [arXiv:1607.06363] [INSPIRE].

[21] ATLAS collaboration, *Search for dark matter produced in association with a hadronically decaying vector boson in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, *Phys. Lett.* **B 763** (2016) 251 [arXiv:1608.02372] [INSPIRE].

[22] ATLAS collaboration, *Search for pair production of heavy vector-like quarks decaying to high-p_T W bosons and b quarks in the lepton-plus-jets final state in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, *JHEP* **10** (2017) 141 [arXiv:1707.03347] [INSPIRE].

[23] ATLAS collaboration, *Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the* $q\bar{q}^{(')}b\bar{b}$ *final state in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, *Phys. Lett.* **B 774** (2017) 494 [arXiv:1707.06958] [INSPIRE].

[24] ATLAS collaboration, *Search for diboson resonances with boson-tagged jets in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, *Phys. Lett.* **B 777** (2018) 91 [arXiv:1708.04445] [INSPIRE].

[25] ATLAS collaboration, *Searches for heavy ZZ and ZW resonances in the* $\ell\ell qq$ *and* $\nu\nu qq$ *final states in pp collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, arXiv:1708.09638 [INSPIRE].

[26] ATLAS collaboration, *A search for resonances decaying into a Higgs boson and a new particle X in the* $XH \to qqbb$ *final state with the ATLAS detector*, *Phys. Lett.* **B 779** (2018) 24 [arXiv:1709.06783] [INSPIRE].

[27] M. Dasgupta, A. Fregoso, S. Marzani and G.P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029 [arXiv:1307.0007] [INSPIRE].

[28] M. Dasgupta, A. Fregoso, S. Marzani and A. Powling, *Jet substructure with analytical methods*, *Eur. Phys. J.* **C 73** (2013) 2623 [arXiv:1307.0013] [INSPIRE].

[29] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *JHEP* **05** (2014) 146 [arXiv:1402.2657] [INSPIRE].

[30] CMS collaboration, *V tagging observables and correlations*, CMS-PAS-JME-14-002, CERN, Geneva Switzerland, (2014).

[31] M. Dasgupta, A. Powling and A. Siodmok, *On jet substructure methods for signal jets*, *JHEP* **08** (2015) 079 [arXiv:1503.01088] [INSPIRE].

[32] A.H. Hoang, S. Mantry, A. Pathak and I.W. Stewart, *Extracting a short distance top mass with light grooming*, arXiv:1708.02586 [INSPIRE].

[33] C. Frye, A.J. Larkoski, M.D. Schwartz and K. Yan, *Precision physics with pile-up insensitive observables*, arXiv:1603.06375 [INSPIRE].

[34] C. Frye, A.J. Larkoski, M.D. Schwartz and K. Yan, *Factorization for groomed jet substructure beyond the next-to-leading logarithm*, *JHEP* **07** (2016) 064 [arXiv:1603.09338] [INSPIRE].

[35] M. Dasgupta, L. Schunk and G. Soyez, *Jet shapes for boosted jet two-prong decays from first-principles*, *JHEP* **04** (2016) 166 [arXiv:1512.00516] [INSPIRE].

[36] M. Dasgupta, A. Powling, L. Schunk and G. Soyez, *Improved jet substructure methods: Y-splitter and variants with grooming*, *JHEP* **12** (2016) 079 [arXiv:1609.07149] [INSPIRE].

[37] G.P. Salam, L. Schunk and G. Soyez, *Dichroic subjettiness ratios to distinguish colour flows in boosted boson tagging*, *JHEP* **03** (2017) 022 [arXiv:1612.03917] [INSPIRE].

[38] 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].

[39] C.W. Bauer and I.W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett.* **B 516** (2001) 134 [hep-ph/0107001] [INSPIRE].

[40] 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].

[41] 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].

[42] I.Z. Rothstein and I.W. Stewart, *An effective field theory for forward scattering and factorization violation*, *JHEP* **08** (2016) 025 [arXiv:1601.04695] [INSPIRE].

[43] I. Moult, L. Necib and J. Thaler, *New angles on energy correlation functions*, *JHEP* **12** (2016) 153 [arXiv:1609.07483] [INSPIRE].

[44] A.J. Larkoski and J. Thaler, *Unsafe but calculable: ratios of angularities in perturbative QCD*, *JHEP* **09** (2013) 137 [arXiv:1307.1699] [INSPIRE].

[45] A.J. Larkoski, S. Marzani and J. Thaler, *Sudakov safety in perturbative QCD*, *Phys. Rev.* **D 91** (2015) 111501 [arXiv:1502.01719] [INSPIRE].

[46] M. Dasgupta and G.P. Salam, *Resummation of nonglobal QCD observables*, *Phys. Lett.* **B 512** (2001) 323 [hep-ph/0104277] [INSPIRE].

[47] J. Dolen, P. Harris, S. Marzani, S. Rappoccio and N. Tran, *Thinking outside the ROCs: designing decorrelated taggers (DDT) for jet substructure*, *JHEP* **05** (2016) 156 [arXiv:1603.00027] [INSPIRE].

[48] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [hep-ph/9707323] [INSPIRE].

[49] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Monte Carlo generators for HERA physics. Proceedings, Workshop*, Hamburg Germany, (1998)–(1999), pg. 270 [hep-ph/9907280] [INSPIRE].

[50] M. Wobisch, *Measurement and QCD analysis of jet cross-sections in deep inelastic positron proton collisions at $\sqrt{s} = 300\,GeV$*, Germany, (2000).

[51] J. Thaler and K. Van Tilburg, *Identifying boosted objects with N-subjettiness*, *JHEP* **03** (2011) 015 [arXiv:1011.2268] [INSPIRE].

[52] J. Thaler and K. Van Tilburg, *Maximizing boosted top identification by minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [arXiv:1108.2701] [INSPIRE].

[53] A.J. Larkoski and I. Moult, *The singular behavior of jet substructure observables*, *Phys. Rev.* **D 93** (2016) 014017 [arXiv:1510.08459] [INSPIRE].

[54] C.W. Bauer, F.J. Tackmann, J.R. Walsh and S. Zuberi, *Factorization and resummation for dijet invariant mass spectra*, *Phys. Rev.* **D 85** (2012) 074006 [arXiv:1106.6047] [INSPIRE].

[55] A.J. Larkoski, I. Moult and D. Neill, *Non-global logarithms, factorization and the soft substructure of jets*, *JHEP* **09** (2015) 143 [arXiv:1501.04596] [INSPIRE].

[56] P. Pietrulewicz, F.J. Tackmann and W.J. Waalewijn, *Factorization and resummation for generic hierarchies between jets*, *JHEP* **08** (2016) 002 [arXiv:1601.05088] [INSPIRE].

[57] A.J. Larkoski, I. Moult and D. Neill, *Toward multi-differential cross sections: measuring two angularities on a single jet*, *JHEP* **09** (2014) 046 [arXiv:1401.4458] [INSPIRE].

[58] M. Procura, W.J. Waalewijn and L. Zeune, *Resummation of double-differential cross sections and fully-unintegrated parton distribution functions*, *JHEP* **02** (2015) 117 [arXiv:1410.6483] [INSPIRE].

[59] G. Lustermans, W.J. Waalewijn and L. Zeune, *Joint transverse momentum and threshold resummation beyond NLL*, *Phys. Lett.* **B 762** (2016) 447 [arXiv:1605.02740] [INSPIRE].

[60] S. Catani and M.H. Seymour, *A general algorithm for calculating jet cross-sections in NLO QCD*, *Nucl. Phys.* **B 485** (1997) 291 [*Erratum ibid.* **B 510** (1998) 503] [hep-ph/9605323] [INSPIRE].

[61] S. Caron-Huot, *Resummation of non-global logarithms and the BFKL equation*, arXiv:1501.03754 [INSPIRE].

[62] D. Neill, *The edge of jets and subleading non-global logs*, arXiv:1508.07568 [INSPIRE].

[63] T. Becher, M. Neubert, L. Rothen and D.Y. Shao, *Effective field theory for jet processes*, *Phys. Rev. Lett.* **116** (2016) 192001 [arXiv:1508.06645] [INSPIRE].

[64] T. Becher, M. Neubert, L. Rothen and D.Y. Shao, *Factorization and resummation for jet processes*, *JHEP* **11** (2016) 019 [*Erratum ibid.* **05** (2017) 154] [arXiv:1605.02737] [INSPIRE].

[65] A.J. Larkoski, I. Moult and D. Neill, *The analytic structure of non-global logarithms: convergence of the dressed gluon expansion*, *JHEP* **11** (2016) 089 [arXiv:1609.04011] [INSPIRE].

[66] D. Neill, *The asymptotic form of non-global logarithms, black disc saturation and gluonic deserts*, *JHEP* **01** (2017) 109 [arXiv:1610.02031] [INSPIRE].

[67] T. Becher, B.D. Pecjak and D.Y. Shao, *Factorization for the light-jet mass and hemisphere soft function*, *JHEP* **12** (2016) 018 [arXiv:1610.01608] [INSPIRE].

[68] T. Becher, R. Rahn and D.Y. Shao, *Non-global and rapidity logarithms in narrow jet broadening*, *JHEP* **10** (2017) 030 [arXiv:1708.04516] [INSPIRE].

[69] 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].

[70] R. Abbate, M. Fickinger, A.H. Hoang, V. Mateu and I.W. Stewart, *Thrust at N^3LL with power corrections and a precision global fit for $\alpha_s(m_Z)$*, *Phys. Rev.* **D 83** (2011) 074021 [arXiv:1006.3080] [INSPIRE].

[71] T.T. Jouttenus, I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *Jet mass spectra in Higgs boson plus one jet at next-to-next-to-leading logarithmic order*, *Phys. Rev.* **D 88** (2013) 054031 [arXiv:1302.0846] [INSPIRE].

[72] J.M. Campbell and R.K. Ellis, *An update on vector boson pair production at hadron colliders*, *Phys. Rev.* **D 60** (1999) 113006 [hep-ph/9905386] [INSPIRE].

[73] J.M. Campbell and R.K. Ellis, *MCFM for the Tevatron and the LHC*, *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10 [arXiv:1007.3492] [INSPIRE].

[74] J.M. Campbell, R.K. Ellis and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **07** (2011) 018 [arXiv:1105.0020] [INSPIRE].

[75] E. Farhi, *A QCD test for jets*, *Phys. Rev. Lett.* **39** (1977) 1587 [INSPIRE].

[76] CMS collaboration, *Search for light vector resonances decaying to a quark pair produced in association with a jet in proton-proton collisions at $\sqrt{s} = 13\,TeV$*, CMS-PAS-EXO-17-001, CERN, Geneva Switzerland, (2017).

[77] ATLAS collaboration, *Performance of jet substructure techniques in early $\sqrt{s} = 13\,TeV$ pp collisions with the ATLAS detector*, ATLAS-CONF-2015-035, CERN, Geneva Switzerland, (2015).

[78] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[79] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[80] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [arXiv:1405.0301] [INSPIRE].

[81] 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].

[82] W.T. Giele, D.A. Kosower and P.Z. Skands, *Higher-order corrections to timelike jets*, *Phys. Rev.* **D 84** (2011) 054003 [arXiv:1102.2126] [INSPIRE].

[83] A. Gehrmann-De Ridder, M. Ritzmann and P.Z. Skands, *Timelike dipole-antenna showers with massive fermions*, *Phys. Rev.* **D 85** (2012) 014013 [arXiv:1108.6172] [INSPIRE].

[84] M. Ritzmann, D.A. Kosower and P. Skands, *Antenna showers with hadronic initial states*, *Phys. Lett.* **B 718** (2013) 1345 [arXiv:1210.6345] [INSPIRE].

[85] L. Hartgring, E. Laenen and P. Skands, *Antenna showers with one-loop matrix elements*, *JHEP* **10** (2013) 127 [arXiv:1303.4974] [INSPIRE].

[86] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[87] *Fastjet contrib webpage*, http://fastjet.hepforge.org/contrib/.

[88] M. Cacciari, G.P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[89] A.J. Larkoski, D. Neill and J. Thaler, *Jet shapes with the broadening axis*, *JHEP* **04** (2014) 017 [arXiv:1401.2158] [INSPIRE].

[90] A.J. Larkoski and J. Thaler, *Aspects of jets at 100 TeV*, *Phys. Rev.* **D 90** (2014) 034010 [arXiv:1406.7011] [INSPIRE].

[91] S. Marzani, L. Schunk and G. Soyez, *A study of jet mass distributions with grooming*, *JHEP* **07** (2017) 132 [arXiv:1704.02210] [INSPIRE].

[92] CMS collaboration, *Measurement of the differential jet production cross section with respect to jet mass and transverse momentum in dijet events from pp collisions at $\sqrt{s} = 13\,TeV$*, CMS-PAS-SMP-16-010, CERN, Geneva Switzerland, (2016).

[93] G.P. Korchemsky and G.F. Sterman, *Power corrections to event shapes and factorization*, *Nucl. Phys.* **B 555** (1999) 335 [hep-ph/9902341] [INSPIRE].

[94] G.P. Korchemsky and S. Tafat, *On power corrections to the event shape distributions in QCD*, *JHEP* **10** (2000) 010 [hep-ph/0007005] [INSPIRE].

[95] S.W. Bosch, B.O. Lange, M. Neubert and G. Paz, *Factorization and shape function effects in inclusive B meson decays*, *Nucl. Phys.* **B 699** (2004) 335 [hep-ph/0402094] [INSPIRE].

[96] A.H. Hoang and I.W. Stewart, *Designing gapped soft functions for jet production*, *Phys. Lett.* **B 660** (2008) 483 [arXiv:0709.3519] [INSPIRE].

[97] CMS collaboration, *Inclusive search for the Standard Model Higgs boson produced in pp collisions at $\sqrt{s} = 13\,TeV$ using $H \to b\bar{b}$ decays*, CMS-PAS-HIG-17-010, CERN, Geneva Switzerland, (2017).

[98] A.J. Larkoski, I. Moult and D. Neill, *Building a better boosted top tagger*, *Phys. Rev.* **D 91** (2015) 034035 [arXiv:1411.0665] [INSPIRE].

[99] S. Höche and S. Prestel, *Triple collinear emissions in parton showers*, *Phys. Rev.* **D 96** (2017) 074017 [arXiv:1705.00742] [INSPIRE].

[100] S. Höche, F. Krauss and S. Prestel, *Implementing NLO DGLAP evolution in parton showers*, *JHEP* **10** (2017) 093 [arXiv:1705.00982] [INSPIRE].

[101] H.T. Li and P. Skands, *A framework for second-order parton showers*, *Phys. Lett.* **B 771** (2017) 59 [arXiv:1611.00013] [INSPIRE].

[102] I.W. Stewart and C.W. Bauer, *Lectures on the soft-collinear effective theory*, http://ocw.mit.edu/courses/physics/8-851-effective-field-theory-spring-2013/lecture-notes/MIT8_851S13_scetnotes.pdf.

[103] T. Becher, A. Broggio and A. Ferroglia, *Introduction to soft-collinear effective theory*, *Lect. Notes Phys.* **896** (2015) pp.1–206 [arXiv:1410.1892] [INSPIRE].

[104] C.W. Bauer, C. Lee, A.V. Manohar and M.B. Wise, *Enhanced nonperturbative effects in Z decays to hadrons*, *Phys. Rev.* **D 70** (2004) 034014 [hep-ph/0309278] [INSPIRE].

[105] A.V. Manohar, *Deep inelastic scattering as $x \to 1$ using soft collinear effective theory*, *Phys. Rev.* **D 68** (2003) 114019 [hep-ph/0309176] [INSPIRE].

[106] 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].

[107] R.K. Ellis, D.A. Ross and A.E. Terrano, *The perturbative calculation of jet structure in e^+e^- annihilation*, *Nucl. Phys.* **B 178** (1981) 421 [INSPIRE].

[108] D.A. Kosower and P. Uwer, *One loop splitting amplitudes in gauge theory*, *Nucl. Phys.* **B 563** (1999) 477 [hep-ph/9903515] [INSPIRE].

[109] G. Marchesini and A.H. Mueller, *BFKL dynamics in jet evolution*, *Phys. Lett.* **B 575** (2003) 37 [hep-ph/0308284] [INSPIRE].

[110] Y. Hatta, *Relating e^+e^- annihilation to high energy scattering at weak and strong coupling*, *JHEP* **11** (2008) 057 [arXiv:0810.0889] [INSPIRE].

[111] E. Avsar, Y. Hatta and T. Matsuo, *Soft gluons away from jets: distribution and correlation*, *JHEP* **06** (2009) 011 [`arXiv:0903.4285`] [INSPIRE].

[112] L.G. Almeida, S.D. Ellis, C. Lee, G. Sterman, I. Sung and J.R. Walsh, *Comparing and counting logs in direct and effective methods of QCD resummation*, *JHEP* **04** (2014) 174 [`arXiv:1401.4460`] [INSPIRE].

[113] S. Catani and M. Grazzini, *The soft gluon current at one loop order*, *Nucl. Phys.* **B 591** (2000) 435 [`hep-ph/0007142`] [INSPIRE].

[114] G. Altarelli, R.K. Ellis and G. Martinelli, *Large perturbative corrections to the Drell-Yan process in QCD*, *Nucl. Phys.* **B 157** (1979) 461 [INSPIRE].