


## Revisiting the role of grooming in DIS

Y. Makris<sup>\*</sup>*INFN Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy*
 (Received 19 January 2021; accepted 15 February 2021; published 3 March 2021)

We introduce a novel grooming procedure, which is an extension of the modified MassDrop tagging algorithm, tailored to the needs of deep inelastic scattering (DIS). The new algorithm, which grooms the event as a whole, takes advantage of the natural separation of current and target fragmentation in the Breit frame, in order to eliminate radiation in the beam and central rapidity regions. We study the groomed invariant mass in DIS and within soft-collinear effective theory we construct a factorization theorem for the cross section in the back-to-back limit. In this limit we show that, up to a normalization factor, the cross section does not depend on the incoming hadronic matrix element and we propose this measurement at HERA and the future electron-ion collider (EIC) as a probe to hadronization, precision QCD, and cold nuclear matter effects. We also give an event based definition of the winner-take-all axis and comment on possible applications.

DOI: [10.1103/PhysRevD.103.054005](https://doi.org/10.1103/PhysRevD.103.054005)

### I. INTRODUCTION

Some of the main goals of the future Electron-Ion-Collider [1] are the study of hadronization, QCD dynamics, and 3D-imaging of the nucleon and nuclei. Among others, event shapes, jet, and jet substructure observables are proposed to achieve these goals. These proposals are often motivated by the fact that event shapes and jet observables have already been applied successfully in various aspects of QCD phenomenology. However, the applicability and the necessity for such observables strongly depends on the experiment and the objective in consideration. For example, while for the clean environment of lepton-lepton colliders, event shape measurements can be directly compared to theoretical predictions, in hadronic colliders such comparison is spoiled by the contamination from underlying event (UE) and soft multi-parton interactions. A remedy to this problem is to use jet and jet substructure observables which are often extension of event shapes but limited only to the particles within the jets. Nonetheless, soft and uniformly distributed radiation can still significantly modify jet substructure measurements. To this end, theorists and experimentalists often employ grooming techniques [2–8] to remove wide angle soft radiation, isolating this way the energetic core of the jet which is usually found near the jet axis. Groomed jets are shown to

be mostly insensitive to the contamination from UE in hadronic colliders and thus constitute a robust framework for comparison between theory and experiment. In the past few years there has been a significant theoretical progress [9–26] in our understanding of groomed observables. These studies were further complemented by experimental results [27–32] both in  $pp$  and  $AA$  collisions.

Groomed jet substructure observables have also been considered for the EIC. However, the role and the applicability of grooming at EIC is still a new and unexplored subject. While at EIC contamination from UE is not going to be an issue, there are many other motivations for the use of grooming techniques. Some of those are: (i) constructing observables free from nonglobal-logarithms, (ii) mitigation of hadronization corrections, (iii) phenomenological handle on soft radiation, and (iv) as a dial for nonperturbative contributions. Despite the many motivations for the use of grooming, at EIC we anticipate jets with rather low particle multiplicities. In this case it is unclear what the effect of grooming will be and how groomed jet observables should be interpreted. In this paper we introduce an event-level grooming procedure which restores the role of grooming at EIC. To achieve that we modify a widely used grooming algorithm, the modified MassDrop tagging (mMDT) [4]. By creating an angular ordering, measuring the angle from the direction of the photon in the Breit frame, the proposed grooming algorithm removes radiation away from the struck quark direction and close to proton remnants. Then, groomed events at EIC can be treated in an equivalent manner to groomed jets.

Beyond the motivations mentioned above, the event-grooming algorithm we discuss in this paper can be used for eliminating beam remnants and initial state radiation

---

<sup>\*</sup>yiannis.makris@pv.infn.it

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

which can be a significant source of uncertainty to global event observables at EIC and HERA. As we discuss later using the example of groomed invariant mass, the shape of the groomed distribution is insensitive to the initial state hadronic matrix element and the kinematic variables of the hard process. Furthermore, the groomed distributions are also insensitive to the activity in the forward region where (due to detector acceptance) constraints have been implemented at HERA and are expected to be implemented at the future electron-ion-collider (EIC) as well.

This paper is organized as follows. In Sec. II we briefly introduce our notation, review the DIS kinematics in the Breit frame, and we introduce the new grooming algorithm. Then we discuss the application of our algorithm to event shape observables using as an example the groomed invariant mass. We formulate a factorization theorem in the back-to-back limit within the soft and collinear effective theory (SCET) [33–37], we compute the next-to- (NLL) and next-to-next-to-leading logarithmic (NNLL) accuracy distributions, and compare against the partonic shower of PYTHIA 8. We close this section with a brief discussion on hadronization effects. In Sec. III we give an event based definition of the winner-take-all (WTA) axis and we discuss possible applications. We finally conclude in Sec. IV. In the Appendix we have collected fixed order results of various elements of the factorized cross section, as well as the leading order full-QCD cross section.

## II. THE GROOMING ALGORITHM

In this paper we work explicitly in the Breit frame, where exist a clean geometrical separation of target and current fragmentation.<sup>1</sup> To avoid contributions from the resolved photon events we will be considering only events with large photon virtuality, i.e.,  $Q = \sqrt{-q^2} \gg 1$  GeV, where  $q^\mu$  is the four-momentum of the virtual photon. In the Breit frame we have,

$$q^\mu = \frac{Q}{2}(\bar{n}^\mu - n^\mu) = Q(0, 0, 0, -1), \quad (1)$$

where  $n^\mu \equiv (1, 0, 0, +1)$  and  $\bar{n}^\mu \equiv (1, 0, 0, -1)$ . The proton momentum (up to mass corrections) is,

$$P^\mu \simeq \frac{Q}{2x_B} n^\mu = \frac{Q}{2x_B}(1, 0, 0, +1), \quad (2)$$

with  $x_B$  the standard Bjorken variable,  $x_B \equiv Q^2/(2q \cdot P)$ . At Born level, the struck quark back-scatters against the proton and has momentum ( $x \simeq x_B$ ):

<sup>1</sup>However, all of the kinematic variables we introduce are either Lorentz invariant or boost invariant along the direction of the proton in Breit frame. Thus, the grooming algorithm we develop will also be boost invariant along that same direction, e.g., in the  $\gamma^*p$  center of mass frame.

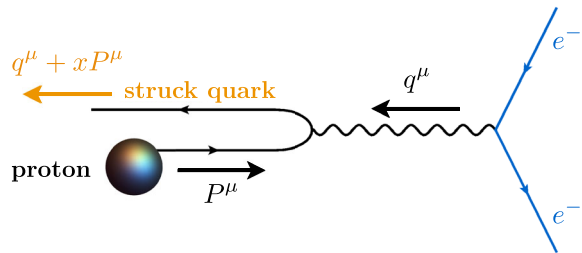


FIG. 1. Diagrammatic illustration of the Born level process in deep inelastic quark scattering in the Breit frame.

$$p_q^\mu = xP^\mu + q^\mu \simeq (Q/2)\bar{n}^\mu. \quad (3)$$

The fragmentation of the struck-quark leads to collimated (jet-like) radiation pointing to the opposite of the beam direction as illustrated in Fig. 1. On the other hand, initial state radiation and beam remnants are moving in the opposite direction close to the proton’s direction of motion. It is this feature of the Breit frame, which leads to clean separation of target and current fragmentation, that we utilize in the new grooming algorithm.

Throughout this paper we will use the Lorentz invariant momentum fraction  $z_i$ ,

$$z_i = \frac{P \cdot p_i}{P \cdot q} \xrightarrow{\text{Breit frame}} z_i = n \cdot p_i / Q = p_i^+ / Q. \quad (4)$$

where  $p_i^\mu$  is the four-momentum of the particle  $i$ . Here and in the rest of this paper we use the standard notation,  $p^+ \equiv n \cdot p$  and  $p^- \equiv \bar{n} \cdot p$ . Note that from conservation of momentum we have

$$\sum_i z_i = 1, \quad (5)$$

where the sum extends over all final state particles in the event, excluding the scattered lepton or its decay products. The struck quark fragments, found close the direction of the virtual photon share the largest fraction of the lightcone momenta  $p^+$  and thus satisfy  $z_i \sim 1$ . Then, the constraint in Eq. (5) requires that soft radiation and particles close to the direction of the target hadron are described by parametrically smaller values of the momentum fraction  $z_i \ll 1$ .

### A. Clustering and declustering

We aim to remove radiation in the region close to the beam, but not in the opposite direction where the struck quark fragments are found. Therefore, the grooming procedure we introduce it is necessarily asymmetric under  $\hat{z} \rightarrow -\hat{z}$ . Following a similar procedure to mMDT and SoftDrop, all particles in the event (excluding the scattered lepton or its decay products) are organized in a clustering tree based on their angular separation and their distance from the direction of the virtual photon. To do exactly that

we use the Centauro measure introduced in Ref. [38] in the context of jet algorithms for DIS in the Breit frame. The clustering procedure is:

- (1) For every pair of particles  $\{i, j\}$  in the event, we calculate the Centauro measure,

$$d_{ij} = (\Delta\bar{\eta}_{ij})^2 + 2\bar{\eta}_i\bar{\eta}_j(1 - \cos\Delta\phi_{ij}), \quad (6)$$

where

$$\bar{\eta}_i \equiv 2\sqrt{1 + \frac{q \cdot p_i}{x_B P \cdot p_i} \xrightarrow{\text{Breit frame}} \frac{2p_i^\perp}{p_i^+}}, \quad (7)$$

and  $\Delta\bar{\eta}_{ij} = \bar{\eta}_i - \bar{\eta}_j$ . Note that in the very backward limit (close to the photon axis)  $\bar{\eta}_i$  is up to power-corrections the angle of the particle  $i$  and the photon axis.

- (2) We find the minimum of all  $d_{ij}$  and we merge the particles  $i$  and  $j$  into a new “branch” by adding their momenta and removing the merged particles  $i$  and  $j$  from the list.
- (3) Repeat until all particles in the event are merged together.

Keeping the history of all mergings, maps the event into an angular ordering tree where nearby particles are merged together in early stages. Then, clusters of particles near to the direction of the photon are combined first. Last in the tree, are clustered particles near the target hadron beam. With this angular ordering in hand we may now start the declustering process. The procedure is similar to SoftDrop, with some modifications. We open the tree back up in the reverse order of clustering. At each stage of the declustering, we have two branches available, label them  $i$  and  $j$ . We require:

$$\frac{\min(z_i, z_j)}{z_i + z_j} > z_{\text{cut}}, \quad (8)$$

where  $z_{\text{cut}}$  is the grooming parameter, and  $z_i$  is the momentum fraction variable of the branch  $i$  as defined in Eq. (4). If the two branches fail this requirement, the branch with smaller  $z_i$  is removed from the event, and we decluster the other branch, once again testing Eq. (8). The pruning continues until we have a branch that when declustered passes the condition in Eq. (8).

We demonstrate the result of grooming in Fig. 2, where we visualize a sample of events as simulated in PYTHIA 8 and groomed according to the procedure described above. In this figure, each particle is illustrated as a disk, with area proportional to its energy, projected onto the unfolded sphere about the hard-scattering vertex. The vertical dashed lines correspond to constant  $\theta$  and curved lines to constant  $\phi$ . Grayed out disks correspond to particles that have been “groomed away” where green disks constitute the groomed event. Characteristic events are event 1 and 2, where the wide-angle particles are groomed away while the energetic and collinear particles (close to direction of the struck quark) pass grooming. However, for (relatively rare) events where an energetic cluster of particles exist in central rapidity regions (see for example event 3 in Fig. 2), grooming will have small effect at central rapidities. This is due to the fact that such events contain central rapidity jets with relatively large momentum fraction  $z_i$  and capable of passing the momentum-fraction condition in Eq. (8). The deferment event radiation patterns can be identified by considering measurements of event shapes which we discuss next.

## B. Groomed invariant mass: A case study

In the case of multiple final-state hadrons, it is hard to study and understand the collective behavior and

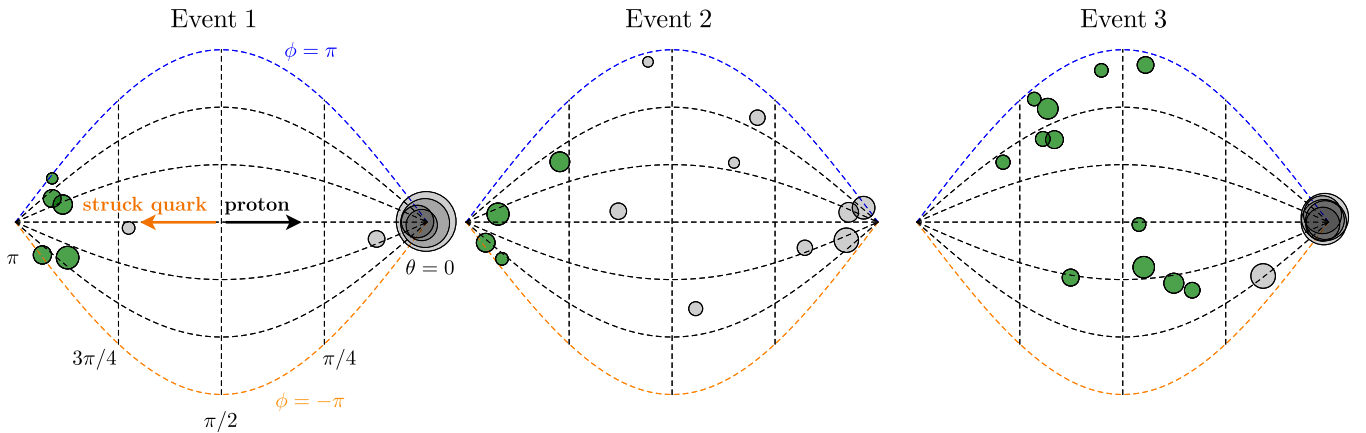


FIG. 2. Visualization of three PYTHIA 8 events at  $\sqrt{s} = 63$  GeV and  $Q \sim 10$  GeV before and after grooming. The particles in this events are represented by disks on the unfolded sphere. Green disks represent particles that pass grooming where grayed-out particles are removed from the event by the grooming procedure. For the grooming parameter we use here  $z_{\text{cut}} = 0.1$

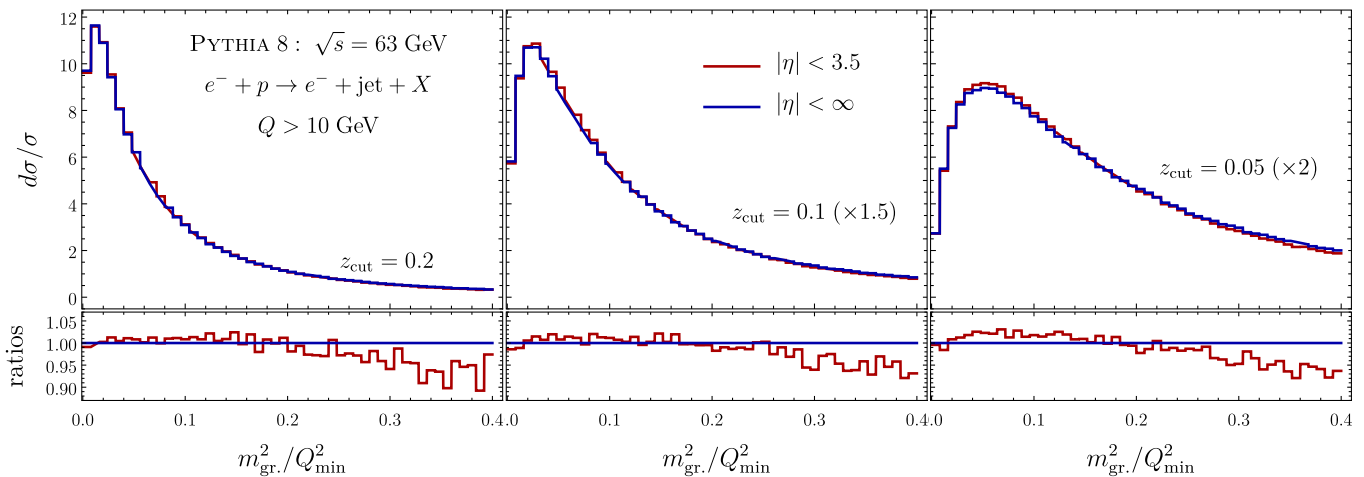


FIG. 3. PYTHIA 8 groomed invariant mass (GIM) spectrum for EIC kinematics. We present three different values of the grooming parameter  $z_{\text{cut}} = 0.2, 0.1,$  and  $0.05$  from left to right. We compare the GIM distributions before and after imposing rapidity cutoff at  $|\eta| < 3.5$  in the Laboratory frame.

correlations between those hadrons by simply looking at individual particle properties such as four-momenta and charge. It is therefore useful to have a single variable, the value of which, characterizes the energy flow and distribution of momenta in space. Event shape variables play exactly this role. For the past several decades, there has been many event shape observables proposed and studied extensively, and they have played a central part in our understanding of quantum chromodynamics (QCD) with plethora of applications in QCD phenomenology. Some important examples of such applications are: (i) determination of the QCD fundamental constants such as the strong coupling [39–46] and the heavy quark masses [47–51], (ii) testing perturbative QCD [39,52–58], (iii) studying nonperturbative effects and hadronization [59–62], and (iv) tuning Monte-Carlo event generators [63–66].

The role of event shape studies in this work is twofold. First, we are considering groomed event shapes as a possible measurement in a future analysis at HERA and at the EIC, and second as a theoretical tool for quantitatively investigating the various features of the grooming algorithm that we introduced in this work. For our purposes we will focus on the groomed invariant mass (GIM) which we denote with  $m_{\text{gr}}^2$  and is given by the invariant mass squared of the groomed event,

$$m_{\text{gr}}^2 \equiv \left( \sum_i p_i^\mu \right)^2, \quad (9)$$

where the sum runs over all particles that pass grooming, excluding the scattered lepton or its decay products. Note that if the sum included all hadrons in the event then this definition would correspond to the photon-proton center-of-mass energy,  $W^2 \equiv (P^\mu + q^\mu)^2$ , which can be uniquely determined from the kinematics of the hard process.

However, GIM depends on the distribution of radiation in the event and is a measure of how well collimated is the radiation that passes grooming. For example, in the back-to-back limit where radiation is distributed in two pencil-like “jets” (from which one jet is the beam, see for example event 1 in Fig. 2), we have  $m_{\text{gr}}^2 \ll Q^2$ . On the other hand, in a dijet configuration (i.e., two final state jets, see for example event 3 in Fig. 2) we have  $m_{\text{gr}}^2 \sim Q^2$ . We will study the spectrum of the GIM in different kinematic regimes and for a range of values of the grooming parameter,  $z_{\text{cut}}$ . For these studies we are employing Monte-Carlo event generators as well as effective field theory (EFT) techniques.

We first consider PYTHIA 8 simulations for potential EIC beam energies: 10 GeV electron on a 100 GeV proton ( $\sqrt{s} \simeq 63$  GeV). After imposing the rapidity cuts on the final state particles in the Laboratory frame we perform the transformation in the Breit frame where we construct the observable and impose the grooming algorithm. We only focus on events with photon virtuality  $Q > 10$  GeV and we test the effect of cutoff on the pseudo-rapidities,  $|\eta| < 3.5$ . The results are shown in Fig. 3. The GIM distributions are shown for three values of the grooming parameter  $z_{\text{cut}} = 0.2, 0.1,$  and  $0.05$  from left to right respectively. In each plot we show the results for  $|\eta| < 3.5$  (red) and no rapidity cuts (blue). To improve readability, the distributions are normalized and the horizontal axis has been re-scaled by the square of the minimum value of  $Q$  included in the analysis, in this case  $Q_{\text{min}}^2 = 100$  GeV<sup>2</sup>. It is very clear from these plots that the distributions shift to smaller GIM values with larger  $z_{\text{cut}}$ . This is expected due to the additive property of the observable. The correction to the distributions due to the rapidity cutoff at  $|\eta| < 3.5$  and for all values of  $z_{\text{cut}}$  is found to be 1%-2% in the back-to-back limit and about 5% in the tail of the

distribution.<sup>2</sup> This demonstrates that the groomed events are practically insensitive to rapidity cutoffs due to effective removal of initial state radiation and beam remnants in the forward region. This property of the grooming algorithm could be proven particularly useful in measurements of global observables such as thrust, N-jettiness, angularities, etc., which could suffer from large systematic uncertainties due to detector acceptance in the forward region.

### C. Factorization

One of the powerful aspects of grooming algorithms, such as mMDT and SoftDrop, is that the corresponding

$$\left. \frac{d\sigma}{dx dQ^2 dm_{\text{gr}}^2} \right|_{\substack{m_{\text{gr}}^2 \ll Q^2 \\ z_{\text{cut}} \ll 1}} = \frac{\sigma_0(Q, y)}{Q^2} \left( C^{(1)}(z_{\text{cut}}, m_{\text{gr}}^2) \frac{\alpha_s(\mu)}{4\pi} \Theta \left( \frac{1-x}{x} - \frac{m_{\text{gr}}^2}{Q^2} \right) \sum_{i=q, \bar{q}} Q_i^2 f_{i/P}(x, \mu) + \mathcal{O}(\alpha_s^2) \right), \quad (10)$$

where  $Q_i$  is the charge of the quark flavor  $i$  and  $f_{i/P}$  are the corresponding proton PDFs. The coefficient  $C^{(1)}$  is extracted from expanding the full QCD cross section in the region  $m_{\text{gr}}^2 \ll Q^2$  and  $z_{\text{cut}} \ll 1$  keeping only the leading terms. Doing so we find,

$$C^{(1)}(m_{\text{gr}}^2, z_{\text{cut}}) = -4C_F \frac{Q^2}{m_{\text{gr}}^2} \left( \frac{3}{4} + \ln z_{\text{cut}} \right). \quad (11)$$

Evaluating the coefficient for typical values of  $m_{\text{gr}}^2/Q^2$  and  $z_{\text{cut}}$  we find  $C^{(1)} \sim 50$ – $200$ . Comparing this to the expansion parameter,  $\alpha_s(20 \text{ GeV})/(4\pi) \sim 0.01$ , we find that the convergence of the perturbative expansion in this region of phase-space might be slow or even unstable and therefore potentially resulting in large theoretical uncertainties. Thus, it is crucial that we perform the resummation of these logarithms to all orders in perturbation theory. To achieve this we factorize the cross section into various matrix elements each depending on a single physical scale (or multiple scales of similar size). The renormalization group equations satisfied by these matrix elements will then allow us to establish a systematic framework for the resummation of large logarithms up to the desired logarithmic accuracy.

Here we will limit our analysis in the case where  $m_{\text{gr}}^2 \ll Q^2$  and  $z_{\text{cut}} \ll 1$ , and we construct the factorized cross section for two possible hierarchies,

$$\begin{aligned} \text{region 1: } & 1 \gg z_{\text{cut}} \gg m_{\text{gr}}^2/Q^2, \\ \text{region 2: } & 1 \gg z_{\text{cut}} \sim m_{\text{gr}}^2/Q^2. \end{aligned} \quad (12)$$

<sup>2</sup>The effects are anticipated to be larger in the tail of the distribution since this region is populated by uniformly distributed radiation and dijet configurations which are more sensitive to radiation in the forward region.

<sup>3</sup>In principle small hadronization correction are also anticipated here.

groomed jet mass distributions can be calculated in perturbation theory at very high precision. The mMDT jet mass in  $e^+e^-$  has been calculated at fixed order up to NNLO [15,67–69] and up to NNNLL [70] in resummation.

Similarly the GIM distribution can be calculated perturbatively in the regime  $m_{\text{gr}}^2 \gg \Lambda_{\text{QCD}}$ , by matching onto the parton distribution functions.<sup>3</sup> The order  $\mathcal{O}(\alpha_s)$  result of this matching procedure is given in the Appendix. However, in the case where  $m_{\text{gr}}^2/Q^2 \ll 1$  and/or  $z_{\text{cut}} \ll 1$ , logarithms of these quantities can induce enhancements of higher order terms and potentially ruin the convergence of the perturbative expansion. To observe this enhancement it is sufficient to look at the LO result,

Although both regions are present in the GIM spectrum, region 1 is more relevant for larger values of  $z_{\text{cut}}$  where region 2 better describes distributions with milder grooming. This is due to the fact that as we decrease  $z_{\text{cut}}$ , more radiation passes grooming, which then leads to larger values of  $m_{\text{gr}}^2$ . This can be observed in Fig. 3.

#### 1. Region 1

To establish the factorized form of the cross section we work in soft collinear effective theory (SCET). We first identify the degrees of freedom relevant to our observable in region 1 and then factorize the cross section accordingly using the EFT techniques. We will not show the details of such a factorization procedure here but we will simply show the final result. Details for factorization within SCET in DIS can be found in Refs. [20,71]. For the discussion that follows it is instructive to introduce a dimensionless variable,  $\lambda$ ,

$$\lambda = \frac{m_{\text{gr}}^2}{Q^2}, \quad (13)$$

which will play the role of expansion and power-counting parameter of the EFT.

An important element of the grooming algorithm is that it organizes radiation based on the angular separation and distribution in space. It is then critical to think of the relevant modes the same way. The collinear radiation from the fragmentation of the struck quark is found close to the photon axis and therefore will merge into the clustering tree at the very early stages of the clustering process. Also energetic particles along this direction are anticipated to have large  $p^+$  component, leading to momentum fractions  $z_a \sim 1$  and as a result will pass the energy-fraction condition, thus contributing to the GIM measurement. We refer to these modes as “ $\bar{n}$ -collinear”. At relatively wider angles,

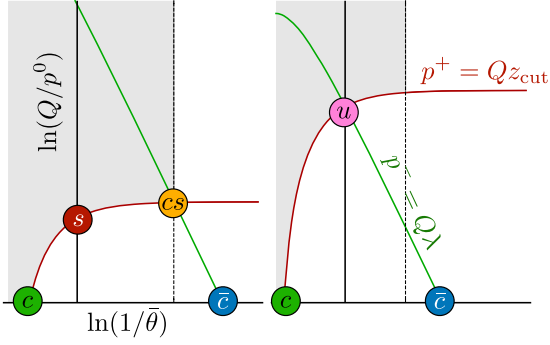


FIG. 4. The modes contributing to the factorization theorem of the groomed-invariant-mass spectrum in region 1 (top-panel) and region 2 (bottom panel). The modes are represented by locations on the energy and polar-angle plane, here  $\bar{\theta} \equiv \pi - \theta$ . The shaded area correspond to regions of phase-space failing grooming.

but still close to the direction of the virtual photon we can find radiation energetic just enough to pass the grooming procedure, i.e.,  $z_a \sim z_{\text{cut}}$  and thus can also contribute to the measurement of GIM. We refer to these modes as “collinear-soft” [72]. In the other side of the event (close to the beam direction) we find initial state radiation and beam remnants from the incoming target-hadron. While this radiation is in principle very energetic, its directionality suggests that it will be merged to the clustering tree last and thus tested first against the energy momentum-fraction condition. With  $z_a \ll 1$  it is most certain that particles from

this region of phase-space will not pass grooming. We refer to these modes as “ $n$ -collinear”. Finally, radiation in the mid-rapidity region can only be soft, otherwise will lead to large values of GIM. Therefore, soft radiation is constrained to fail grooming, so it does not induces contributions to the measurement beyond the hierarchy described in region 1. The directionality, the energy fraction (based on if they pass or fail grooming), and the GIM measurement, fix the  $p^+$  and  $p^-$  component of each mode. The  $p^\perp$  component is then fixed by the on-shell condition:  $(p^\perp)^2 \sim p^+ p^-$ . Thus, for the modes we discussed here we have the following momenta scaling,

$$\begin{aligned} n\text{-collinear: } p_n &\sim Q(z_{\text{cut}}, 1, \sqrt{z_{\text{cut}}}), \\ \text{soft: } p_s &\sim Q z_{\text{cut}}(1, 1, 1), \\ \text{collinear-soft: } p_{cs} &\sim Q(\lambda, z_{\text{cut}}, \sqrt{z_{\text{cut}}\lambda}), \\ \bar{n}\text{-collinear: } p_{\bar{n}} &\sim Q(1, \lambda, \sqrt{\lambda}), \end{aligned} \quad (14)$$

where we used the standard notation  $p = (p^+, p^-, p^\perp)$ . The modes are shown on the  $\ln(Q/p^0) - \ln(1/\bar{\theta})$  plane in Fig. 4 (left panel), where  $\bar{\theta} = \pi - \theta$ .

Considering these modes and following Ref. [71] we can factorize the cross section into the EFT matrix elements and a hard function (which is derived by the matching of the EFT operators onto the QCD current). The factorized cross section reads,

$$\begin{aligned} \frac{d\sigma}{dx dQ^2 dm_{\text{gr}}^2} &= H(Q, y, \mu) S(Q z_{\text{cut}}, \mu) \sum_f \mathcal{B}_{f/P}(x, Q^2 z_{\text{cut}}, \mu) \int de_{\bar{n}} de_{cs} \delta(m_{\text{gr}}^2 - e_{\bar{n}} - e_{cs}) J(e_{\bar{n}}, \mu^2) \mathcal{C}(e_{cs} z_{\text{cut}}, \mu^2) \\ &\times \left[ 1 + \mathcal{O}\left(z_{\text{cut}}, \frac{m_{\text{gr}}^2}{z_{\text{cut}} Q^2}\right) \right], \end{aligned} \quad (15)$$

where  $H$  is the hard function,  $J$  is the quark thrust jet function,  $\mathcal{C}$  is the SoftDrop collinear-soft function for jet-mass measurement,  $\mathcal{B}_{f/P}$  is the quark (flavor  $f$ ) groomed beam function, and finally  $S$  is the global soft function which up to clustering effects, in pure dimensional-regulator, is a scaleless function and thus its contribution to the factorization theorem starts at NNLL.

The beam function [73] describes the initial-state and  $n$ -collinear radiation constrained to be groomed away. For perturbative values of  $Q\sqrt{z_{\text{cut}}}$  and up to power corrections, the beam function can be matched onto the collinear parton-distribution-functions (PDFs) [74],

$$\begin{aligned} \mathcal{B}_{f/P}(x, Q^2 z_{\text{cut}}, \mu) &= \sum_i \int_x^1 \frac{dz}{z} \mathcal{I}_{f/i}\left(\frac{x}{z}, Q^2 z_{\text{cut}}, \mu\right) \\ &\times f_{i/P}(z, \mu) \left[ 1 + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}^2}{Q^2 z_{\text{cut}}}\right) \right]. \end{aligned} \quad (16)$$

The perturbative result of the matching coefficients  $\mathcal{I}_{f/i}$  up to NLO is the same as for the integrated beam-thrust measurement which we included in the Appendix. Beyond NLO, where two or more final state partons are present, we need to consider clustering effects. The sum over  $f$  in Eq. (15) runs over all quark and antiquark flavours relevant to the scale we are working, where the sum over  $i$  in Eq. (16) also includes the gluon.

Note that the cross section in Eq. (15) is differential in  $x$  and  $Q^2$  as well as in  $m_{\text{gr}}^2$ . In order to match experimental measurements which are binned over various ranges of  $x$  and  $Q^2$  one needs to perform the integration numerically. Here we propose an alternative approach that allow us to compare with experimental measurements without the need for integration in  $x$  and  $Q^2$  by simply looking at the normalized cross section instead. This relies on the observation that the cross section dependence on  $x$  and  $Q^2$  is isolated in the hard, beam, and soft functions.

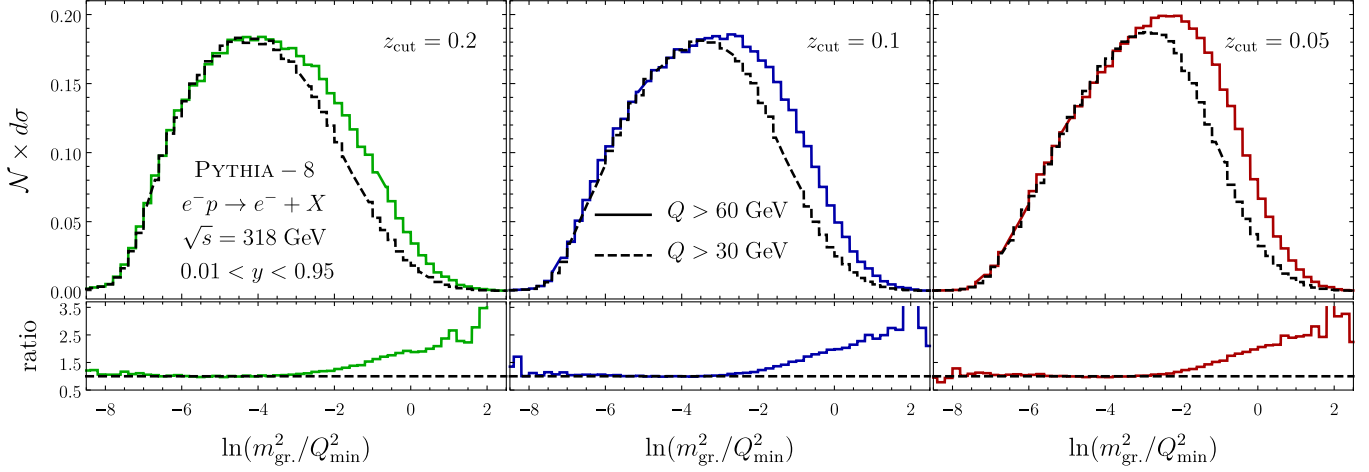


FIG. 5. GIM for two values of  $Q > 30$  and  $60$  GeV for HERA beam energies:  $27.5$  GeV electron on  $920$  GeV proton. We find that for small values of the GIM spectrum the normalized distributions merge. This confirms what is anticipated from the form of the factorized cross section as discussed in the main text.

These functions do not depend on the GIM measurement. Therefore, the shape of the GIM distribution in region 1 is determined by the jet and collinear-soft function which are

independent  $x$  and  $Q^2$ . This is tested using PYTHIA simulations as shown in Fig. 5.

The normalized cross section is then defined as follows,

$$\frac{d\sigma^{\text{norm.}}}{dm_{\text{gr.}}^2}(m_{\text{gr.}}^2, e_{\text{min}}, e_{\text{max}}) \equiv \left[ \int_{e_{\text{min}}}^{e_{\text{max}}} dm_{\text{gr.}}^2 \int_{x, Q^2} \frac{d\sigma}{dx dQ^2 dm_{\text{gr.}}^2} \right]^{-1} \times \int_{x, Q^2} \frac{d\sigma}{dx dQ^2 dm_{\text{gr.}}^2}, \quad (17)$$

where the integration over  $x$  and  $Q^2$  is determined based on the experimental choice of binning. Expanding in terms of the factorized cross section we find

$$\frac{d\sigma^{\text{norm.}}}{dm_{\text{gr.}}^2}(m_{\text{gr.}}^2, e_{\text{min}}, e_{\text{max}}) = \mathcal{N}(e_{\text{min}}, e_{\text{max}}, \mu) \int de_{\bar{n}} de_{cs} \delta(m_{\text{gr.}}^2 - e_{\bar{n}} - e_{cs}) J(e_{\bar{n}}, \mu^2) \mathcal{C}(e_{cs} z_{\text{cut}}, \mu^2), \quad (18)$$

where

$$\mathcal{N}^{-1}(e_{\text{min}}, e_{\text{max}}, \mu) = \int_{e_{\text{min}}}^{e_{\text{max}}} dm_{\text{gr.}}^2 \int de_{\bar{n}} de_{cs} \delta(m_{\text{gr.}}^2 - e_{\bar{n}} - e_{cs}) J(e_{\bar{n}}, \mu^2) \mathcal{C}(e_{cs} z_{\text{cut}}, \mu^2). \quad (19)$$

The moralized cross section can now be calculated without considering the integration over the variables  $x$  and  $Q^2$ . There are some obvious practical advantages adapting this approach. First, the calculation of the normalized cross section can be achieved significantly easier and avoids systematic uncertainties associated with PDF errors. Beyond these, considering the normalized cross section can help reduce theoretical uncertainties since it allow us to push the perturbative calculation at the same order as has been achieved for mMDT jet invariant mass (i.e., NNNLL). The latter statement relies on the fact that the Centauro measure in the collinear limit (along the photon's direction) is, up to power corrections, equal to the  $C/A(\text{SI})$  measure.

We point out that, although the expression in Eq. (19) does not explicitly depend on  $Q$  the factorization only holds

for  $m_{\text{gr.}}^2 \ll Q^2 z_{\text{cut}}$  and thus the region of  $m_{\text{gr.}}^2$  for which we can use this result depends on the lowest value  $Q$  which we are considering for a particular analysis. Therefore, we require that we normalize only within that same region,  $m_{\text{gr.}}^2 < e_{\text{max}} \ll Q_{\text{min}}^2 z_{\text{cut}}$ , or within the region where power corrections to the factorization theorem remain small.

## 2. Region 2

Going beyond region 1 the shape of the cross section depends on both  $Q^2$  and  $x$ . For example in region 2, which is described by the hierarchy of scales:  $1 \gg z_{\text{cut}} \sim \lambda$ , the soft-collinear and soft radiation merge into a single mode which we will refer to us ultra-soft (u-soft). The scaling of the momenta of the relevant modes is,

$$\begin{aligned}
n\text{-collinear: } p_n &\sim Q(z_{\text{cut}}, 1, \sqrt{z_{\text{cut}}}), \\
\text{ultra-soft: } p_{us} &\sim Qz_{\text{cut}}(1, 1, 1) \sim Q\lambda(1, 1, 1), \\
\bar{n}\text{-collinear: } p_{\bar{n}} &\sim Q(1, \lambda, \sqrt{\lambda}).
\end{aligned} \tag{20}$$

These modes are shown on the  $\ln(Q/p^0) - \ln(1/\bar{\theta})$  plane in Fig. 4 (right panel). In addition the veto on mid-rapidity soft radiation that pass grooming has to be lifted since u-soft modes may now contribute to the measurement. The resulting factorization theorem is,

$$\begin{aligned}
\frac{d\sigma}{dx dQ^2 dm_{\text{gr}}^2} &= H(Q, y, \mu) \sum_f \mathcal{B}_{f/P}(x, Q^2 z_{\text{cut}}, \mu) \int de_{\bar{n}} de_{us} \delta(m_{\text{gr}}^2 - e_{\bar{n}} - e_{us}) J(e_{\bar{n}}, \mu^2) \mathcal{U}(e_{us}, Q, z_{\text{cut}}, \mu^2) \\
&\times \left[ 1 + \mathcal{O}\left(z_{\text{cut}}, \frac{m_{\text{gr}}^2}{Q^2}\right) \right].
\end{aligned} \tag{21}$$

In the above equation the hard, beam, and jet functions are the same as in region 1, but the product of soft and collinear-soft functions has been replaced by the u-soft function,  $\mathcal{U}$ , that now depends explicitly on the hard scale  $Q$ , the grooming parameter,  $z_{\text{cut}}$ , and the GIM measurement. At fixed order in the strong coupling expansion one can relate the two expressions within the hierarchy of region 1 up to power corrections,

$$\mathcal{U}(m_{\text{gr}}^2, Q, z_{\text{cut}}, \mu^2)|_{m_{\text{gr}}^2/Q^2 \ll z_{\text{cut}}} = S(Qz_{\text{cut}}, \mu) \mathcal{C}(m_{\text{gr}}^2, z_{\text{cut}}, \mu^2) \left[ 1 + \mathcal{O}\left(\frac{m_{\text{gr}}^2}{Q^2 z_{\text{cut}}}\right) \right]. \tag{22}$$

We evaluated the u-soft function at NLO and showed that the power corrections vanish at this order. Therefore we have that at NLO the u-soft function is simply given by the collinear-soft function of region 1,

$$\mathcal{U}(m_{\text{gr}}^2, Q, z_{\text{cut}}, \mu^2)|_{\text{NLO}} = \mathcal{C}(m_{\text{gr}}^2, z_{\text{cut}}, \mu^2)|_{\text{NLO}}. \tag{23}$$

With this, we now have all the necessary ingredients to perform the resummation of all large logarithms in region 1 and region 2 up to NLL.

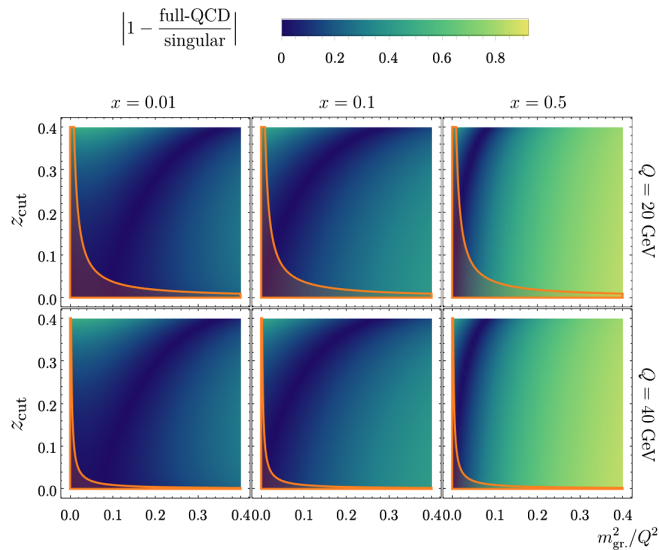


FIG. 6. The relative size of power corrections as a function of  $m_{\text{gr}}^2$  and  $z_{\text{cut}}$  for different values of  $x - Q^2$ . The orange shaded area denotes the nonperturbative region for which  $m_{\text{gr}}^2, z_{\text{cut}} \leq 1.5 \text{ GeV}^2$ .

### 3. Power corrections

Before we start considering numerical results is important to discuss the role of power corrections to the factorization theorems we analyzed in this section. These corrections rise due to the fact that we have effectively expanded the QCD amplitudes in particular kinematic regimes, specifically the ones shown in Eq. (12). One can estimate the size of these corrections order-by-order in perturbation theory by subtracting the singular limit in Eq. (10) from the full QCD result in Eq. (A29). For different values of  $x$  and  $Q^2$  and as a function of  $z_{\text{cut}}$  and  $m_{\text{gr}}^2/Q^2$  we plot the relative corrections in Fig. 6. We find that for  $x \lesssim 0.1$  the power corrections decrease with smaller  $x$  and remain small over a wide range of  $z_{\text{cut}}$  and  $m_{\text{gr}}^2/Q^2$ . However, the corrections seem to increase significantly in the region  $x \sim 1$ . This increase is due to the fact that at large  $x$  and at large invariant mass the cross section is suppressed by the constraint,

$$m_{\text{gr}}^2 \leq W^2 \simeq \frac{1-x}{x} Q^2. \tag{24}$$

The nonperturbative regime is denoted with the orange shading in Fig. 6 which corresponds to the most IR scale of the problem to be of order 1 GeV (in this case  $\mu_{\text{IR}}^2 \sim m_{\text{gr}}^2, z_{\text{cut}} \sim 1 \text{ GeV}$ ).

Such corrections can in principle be included in the resummed result order-by-order in the strong coupling expansion, if the full QCD result is known either analytically or numerically. This can be done in a matching procedure where we add the fixed order corrections to the resummed spectrum after turning off evolution in the region



$m_{\text{gr.}}^2 \gtrsim Q^2$ . This way one can obtain a reasonable prediction in the whole range of the GIM spectrum. Note however that, in-contrast to the standard groomed observables,<sup>4</sup> the observable we discuss here depends on the grooming parameters in the whole GIM spectrum,  $0 < m_{\text{gr.}}^2/Q^2 < (1-x)/x$ . Thus, logarithmic enhancements for  $z_{\text{cut}} \ll 1$  remain relevant through-out the whole spectrum and as a result resummation of these logarithms is important for even large values of GIM. For this reason in this work we will focus in the resummation regions 1 and 2 and we will not pursue the task of matching at the tail of the distribution.

#### D. Resummed results

The various elements of the factorization have renormalization scale,  $\mu$ , dependence. The resummed cross section, where all large logarithms are resummed up to a particular logarithmic accuracy, is obtained by evaluating the elements of factorization at their canonical scales and then use renormalization group (RG) equations to evolve them up to a common scale. The RG equation satisfied by each of the relevant functions is,

$$\frac{d}{d \ln \mu} G(\mu) = \gamma_G(\mu) G(\mu), \quad (25)$$

where  $G$  can be any of the functions that appear in the factorization theorems in Eqs. (15) and (21).<sup>5</sup> The anomalous dimensions  $\gamma_G(\mu)$  are usually written in terms of the cusp anomalous dimension,  $\gamma_{\text{cusp}}[\alpha_s]$ , and a noncusp term,  $\gamma_G[\alpha_s]$ ,

$$\gamma_G(\mu) = d_G \gamma_{\text{cusp}}[\alpha_s] \ln\left(\frac{\mu}{m_G}\right) + \gamma_G[\alpha_s], \quad (26)$$

where  $d_G$  is a number,  $m_G$  has dimensions of mass and both  $d_G$  and  $m_G$  depend on the function  $G$ . The solution of the RG equation in Laplace space can be written as a product of the evolution kernel  $U_G$  and the function  $G$ , evaluated at some initial scale  $\mu_0$ ,

$$G(\mu) = G(\mu_0) \times \exp\left(\int_{\mu_0}^{\mu} d \ln \mu' \gamma_G(\mu')\right) \quad (27)$$

$$\equiv G(\mu_0) \times U_G(\mu, \mu_0). \quad (28)$$

<sup>4</sup>The standard groomed jet-mass observable does not depend on the grooming algorithm at sufficiently large values of the jet-mass, creating a cusp in the distribution where the transition occurs. The same we would have observed if we studied the groomed 1-jettiness for example instead of the GIM.

<sup>5</sup>For functions that depend on the GIM measurement (i.e.,  $J$ ,  $\mathcal{C}$ , and  $\mathcal{U}$ ) the above equation applies to the Laplace transformed expressions (with respect to the variable  $e$ ), otherwise the lhs can be expressed as a convolution between  $\gamma_G$  and  $G$ .

The kernel  $U_G(\mu, \mu_0)$  is computed at NLL and NNLL accuracy in terms of  $d_G$ ,  $m_G$ ,  $\gamma_{\text{cusp}}[\alpha_s]$  and  $\gamma_G[\alpha_s]$  in multiple references in literature (see for example Refs. [43,75,76]). The initial scale  $\mu_0$  for each function is usually chosen to be the canonical scale, i.e., the scale that minimizes the logarithms in the fixed order expansion. For the functions we are considering here, the canonical scales are

$$\begin{aligned} \mu_H &= Q, & \mu_B &= Q\sqrt{z_{\text{cut}}}, & \mu_S &= Qz_{\text{cut}} \\ \mu_J^2 &= m_{\text{gr.}}^2, & \mu_C^2 &= \mu_U^2 = m_{\text{gr.}}^2 z_{\text{cut}}. \end{aligned} \quad (29)$$

For small values of the invariant mass the jet and collinear-soft scales become nonperturbative and the perturbative expansion breaks. For this reason, at small invariant masses, we freeze these scales before reaching the Landau pole using the following prescription:

$$\mu_G \rightarrow g(\mu_G, a, b) = \frac{(1 + (a\mu_G)^b)^{1/b}}{a}, \quad (30)$$

where we chose  $a = 2 \text{ GeV}^{-1}$  and  $b = 5$ .

Regarding the factorization in region 1, we collect from the literature all the necessary ingredients for the construction of the NLL cross section in the Appendix. To achieve this accuracy we need the  $\mathcal{O}(\alpha_s^2)$  cusp terms and the  $\mathcal{O}(\alpha_s)$  noncusp terms of the anomalous dimensions. For the NNLL predictions we need the  $\mathcal{O}(\alpha_s^2)$  noncusp anomalous dimensions, which we also have with the exception of the soft and beam functions. However, the normalized cross section can still be calculated at NNLL with what is already known, since it is independent of the soft and beam functions. For region 2 we only have ingredients sufficient for a NLL calculation which in this case is identical to region 1. We compare our NLL numerical results (region 1 and 2) for the normalized distributions and compare against the partonic shower of PYTHIA 8 in Fig. 7. We give the NNLL normalized distributions (region 1) including theoretical uncertainties in 8.

We first consider the comparison of NLL cross section to PYTHIA's partonic shower. This serves a consistency check of our formalism and we discuss hadronization effects later in this section. To make meaningful comparison of our perturbative result with PYTHIA, we require relatively large values of  $Q$  where a perturbative regime exist. Particularly we choose  $Q_{\text{min}} = 20 \text{ GeV}$  and for this reason we consider the potential EIC kinematics with beam energies: 18 GeV electron on 275 GeV proton ( $\sqrt{s} = 140.7 \text{ GeV}$ ). For the normalization of the cross section the lower boundary is fixed such that the softest scale of the problem (that is the collinear-soft scale,  $\mu_C$ ) remains within the perturbative regime. This leads to the constraint:

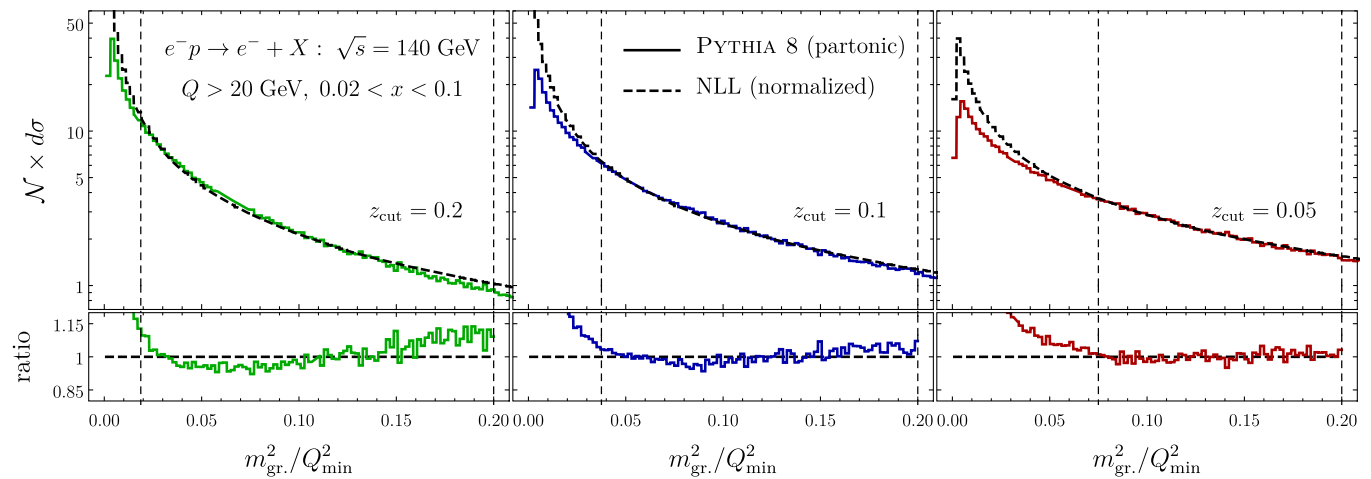


FIG. 7. Comparing the NLL predictions (dashed) against the partonic shower PYTHIA 8 (solid) for three different values of the grooming parameter  $z_{\text{cut}} = 0.2, 0.1,$  and  $0.05$ .

$$e_{\min} = \frac{\Lambda_{\text{NP}}^2}{z_{\text{cut}}}, \quad (31)$$

where  $\Lambda_{\text{NP}}$  is a nonperturbative scale of order  $\gtrsim 1$  GeV. In this paper we chose  $\Lambda_{\text{NP}}^2 = 1.5$  GeV<sup>2</sup>. The upper boundary should in general be chosen such that power corrections to the factorization theorem in Eq. (15) remain small. For the purpose of this comparison we can extend the upper limit as long as we remain well within region 1 and 2,  $m_{\text{gr}}^2 \ll Q_{\text{min}}^2$ . Guided by the size of power corrections in Fig. 6 here we chose,

$$e_{\max} = 0.2. \quad (32)$$

For three different values of the grooming parameter  $z_{\text{cut}} = 0.2, 0.1,$  and  $0.05$  the comparison with PYTHIA 8 is shown in Fig. 7. The values  $e_{\max}$  and  $e_{\min}$  are shown with

dashed vertical lines. We find excellent agreement of our NLL resummed result with the simulation. However, we see that the perturbative regime shrinks as we decrease  $z_{\text{cut}}$ . This effect will be exaggerated if we consider smaller values of  $Q$  where the nonperturbative boundary,  $e_{\min}$ , moves to the right toward larger  $m_{\text{gr}}^2/Q_{\text{min}}^2$  values, further shrinking this way the perturbative regime. This should motivate further studies in both directions: in the perturbative tail of the distribution, to the right of  $e_{\max}$ , and in the nonperturbative regime, to the left of  $e_{\min}$  where  $m_{\text{gr}}^2 z_{\text{cut}} \sim \Lambda_{\text{NP}}^2$ .

In Fig. 8 we show the NNLL normalized distribution in region 1. We have also estimated the theoretical uncertainty by multiplying the three scales  $\mu_J, \mu_C,$  and  $\mu$  by a factor of 2 and 1/2 separately. The largest deviations from the central curve are collected into the envelope that represents the theoretical uncertainty. We find that within the perturbative

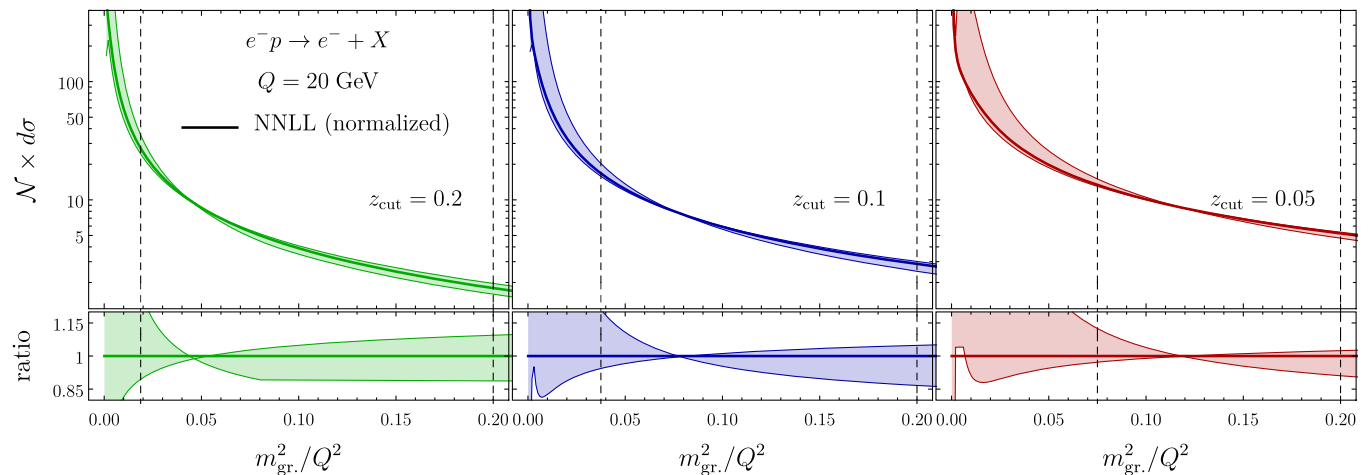


FIG. 8. NNLL prediction including theoretical uncertainties. The theory band is constructed by varying all scales in the factorization theorem by a factor 2 and 1/2.

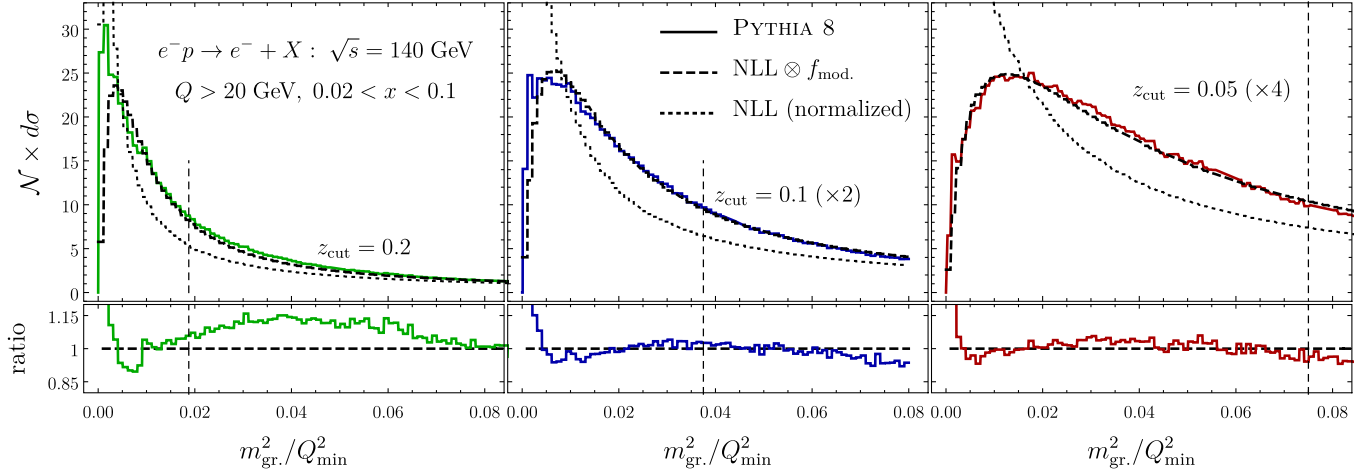


FIG. 9. Hadronic PYTHIA simulations for GIM distribution compared against NLL convolved with shape function  $f_{\text{mod.}}$  for EIC ( $\sqrt{s} = 140$  GeV) kinematics. To emphasize the role of the shape function we also included the NLL distribution without including the model for hadronization.

regime the uncertainty varies between 10%–15%. Note that the size of this error bands is sensitive to the value of the nonperturbative scale  $\Lambda_{\text{NP}}$  which sets the lower limit of integration. As this value decreases we incorporate more of the nonperturbative regime into the normalization factor resulting in large theoretical uncertainties. This is also apparent from the fact that the theoretical uncertainty below the  $e_{\text{min}}$  increases rapidly as shown in Fig. 8.

### E. Hadronization and nonperturbative corrections

In the previous section, as a consistency test of our factorization we compared our resummed predictions against PYTHIA 8 partonic shower. In region 1 ( $1 \gg z_{\text{cut}} \gg m_{\text{gr.}}^2/Q^2$ ), comparing our resummed distributions against PYTHIA’s hadronic spectrum we find significant deviation. In this section we perform a proof-of-concept exercise where we incorporate hadronization effects through a nonperturbative shape function.

Equipped with a factorization theorem and precise predictions of the perturbative spectrum, one can approach hadronization effects in a systematic and rigorous manner. A field-theoretic description and leading power approach to hadronization effects in groomed jet mass is given in Ref. [19]. The formalism is also applicable here since the collinear-soft function that appears in this work, and which dominates the nonperturbative contributions, is the same as in the SoftDrop ( $\beta = 0$ ) groomed jet-mass. In the formalism of Ref. [19], two regions of the invariant mass are identified,

$$\begin{aligned} \text{SDNP: } m_{\text{gr.}, z_{\text{cut}}}^2 &\lesssim \Lambda_{\text{QCD}}^2, \\ \text{SDOE: } m_{\text{gr.}, z_{\text{cut}}}^2 &\gg \Lambda_{\text{QCD}}^2. \end{aligned} \quad (33)$$

While in the region SDNP hadronization corrections are encoded into a universal shape function, in the perturbative

region SDOE the hadronization corrections are described by two nonperturbative matrix elements associated with the “shift” and “boundary” effects. In our study for simplicity we focus only on the nonperturbative region SDNP which includes the peak of the distribution and where the hadronization effects are more apparent. Particularly we incorporate a shape function for describing the hadronization corrections on the GIM spectrum and we use the parameterization discussed in ref. [76],

$$\frac{d\sigma_{\text{had.}}}{dx dQ^2 dm_{\text{gr.}}^2} = \int d\epsilon \frac{d\sigma}{dx dQ^2 dm_{\text{gr.}}^2} \left( m_{\text{gr.}}^2 - \frac{\epsilon^2}{z_{\text{cut}}} \right) f_{\text{mod.}}(\epsilon), \quad (34)$$

where

$$f_{\text{mod.}}(\epsilon) = N_{\text{mod.}} \frac{4\epsilon}{\Omega^2} \exp\left(\frac{2\epsilon}{\Omega}\right), \quad (35)$$

is the model shape function which depends on two nonperturbative parameters: the mean of the shape function  $\Omega$  and the normalization  $N_{\text{mod.}}$ . For the numerical implementation we choose  $\Omega = 1$  GeV and  $N_{\text{mod.}} = 1.1$ .

We compare the analytic predictions against PYTHIA simulation in Fig. 9. We denote the end of SDNP region with a vertical dashed line which is located at

$$m_{\text{gr.}}^2 = \frac{\Lambda_{\text{NP}}^2}{z_{\text{cut}}}. \quad (36)$$

The hadronic and partonic spectra of the simulation are normalized to the tail of the distributions ( $m_{\text{gr.}}^2 \sim Q^2$ ) where nonperturbative effects are anticipated to have small effect. Then the NLL prediction is normalized to the region  $e_{\text{min}} < m_{\text{gr.}}^2 < e_{\text{max}}$  as described above. The normalized NLL

distribution is then convolved to the shape function as in Eq. (34). We find that normalizing the distribution to the tail is necessary for the  $N_{\text{mod.}}$  to be independent of the grooming parameter,  $z_{\text{cut}}$ . Within the SDNP region the agreement with the simulation is within 5% to 10%, and since in this region the shape of the cross section is independent of the incoming hadronic matrix element it offers a unique opportunity for detail hadronization studies. Significant deviation is found for  $m_{\text{gr.}}^2 \lesssim 1 \text{ GeV}^2$  where hadron mass effects are anticipated to play a significant role.

### III. AN EVENT-BASED WINNER-TAKE-ALL AXIS

In the spirit of event-based definition of grooming, in this section we extend the discussion to the recoil-free Winner-Take-All (WTA) axis. We will not discuss in detail applications of such an axis-finding procedure but we point to some opportunities in the context of TMD physics and 3D-imaging of the nucleon and nuclei.

The WTA axis [77,78] was introduced in the context of jets as an alternative recombination scheme which results to a recoil free jet axis. In contrast to the standard jet axis (SJA), which is defined by the sum of all particles within the jet, WTA axis follows the direction of the most energetic branch during the jet-clustering procedure. The resulting axis is then insensitive to soft radiation within the jet and thus free from underlying event contamination.

In DIS and lepton colliders the WTA axis was proposed [79–81] as a probe to quark TMD-PDFs and TMD-FFs. In this case the advantage of using this particular axis instead of the SJA is that in contrast to SJA, a recoil-free axis is insensitive to nonglobal logarithms (NGLs) which they could induce large systematic uncertainties. It was also discussed in refs. [20,80] that a recoil free axis is in general less sensitive to hadronization corrections compared to the SJA and thus permit for a cleaner extraction of initial and final state TMDs. In Ref. [82] it was also discussed the insensitivity of the WTA based observables to tracks versus all-particle measurements.

Here we discuss the construction of a WTA axis using an event level clustering procedure. In practice our approach differs from the standard definition of WTA axis since here one does not construct any jets and rather it applies an axis-finding algorithm directly to the event. The algorithm reads as follows,

- (1) For every pair of particles  $\{i, j\}$  we calculate the Centauro measure,

$$d_{ij} = (\Delta\vec{\eta}_{ij})^2 + 2\vec{\eta}_i \cdot \vec{\eta}_j (1 - \cos \Delta\phi_{ij}), \quad (37)$$

- (2) We find the minimum of all  $d_{ij}$  and we merge the particles  $i$  and  $j$  into a new “branch”. The momentum of the new branch is given by the magnitude of the sum of the momenta of particles/branches  $i$  and  $j$

but is aligned along the direction of the one with the highest energy fraction  $z$ , defined in Eq. (4).

- (3) Repeat until all particles in the event are merged together.

When all particles are merged together the direction of momenta of the final single branch is the WTA axis as we define it in this work.

In the back-to-back limit the finding algorithm we propose here returns a similar axis as if one reconstructed the WTA-axis of the leading jet for  $R \sim 1$  (where  $R$  is the jet radius) using the conventional definition. However, in multiple-jet configurations or for  $R \ll 1$  the jet based definition it returns one axis per jet while the event-based algorithm returns a unique axis.

In the Breit frame and in the back-to-back limit the angle between the WTA axis and the virtual photon is sensitive to the universal TMD-PDFs and the rapidity anomalous dimension. Note that this measurement involves the back-to-back soft function, the same matrix element that appears in the conventional TMD processes. This ensures the universality of the TMDPDFs relevant for this process to the ones that appear in conventional SIDIS and Drell-Yan, in contrast to other proposed observables that involve lepton-jet correlations in the Laboratory frame and involve soft functions that are sensitive to the jet kinematics [83,84]. The details for the theoretical treatment of the proposed observable is discussed in Sec. 2.3 of Ref. [81] and the formalism is directly applicable here.

### IV. SUMMARY AND OUTLOOK

In this paper we discussed a novel grooming procedure for deep inelastic scattering (DIS) events. The corresponding algorithm we propose here is an extension of modified MassDrop Tagger (mMDT). Some of the main differences between our proposal and mMDT are: (in our proposal) (i) clustering is done using Centauro measure instead of the  $C/A$  algorithm, (ii) the algorithm takes as an input the DIS event in the Breit frame rather than a jet, and (iii) the grooming condition uses the light-cone component of the momenta rather than the energy or transverse momentum. With these modifications, we can effectively remove radiation close to the direction of the proton which usually consists of beam remnants and initial state radiation. In practice, it also removes soft particles, isolating this way the energetic final state radiation which comes from the fragments of the struck quark in a leading order DIS process.<sup>6</sup> We demonstrate the applicability of this grooming procedure by considering an event shape measurement. Particularly, we investigate the groomed invariant mass (GIM),  $m_{\text{gr.}}^2$ , which is a measure of how well collimated are the particles that pass grooming. In the back-to-back

<sup>6</sup>For recent developments on event-level groomed event shapes in hadronic colliders see [85].

limit (i.e.,  $m_{\text{gr}}^2 \ll Q^2$ ), this observable is (up to power-corrections) directly related to the groomed event shape 1-jettiness. We show that the GIM spectrum is insensitive to rapidity cutoffs in the Laboratory frame usually imposed due to detector acceptance. We derive factorization theorems within SCET for the back-to-back limit considering two different hierarchies: region 1 where  $1 \gg z_{\text{cut}} \gg m_{\text{gr}}^2/Q^2$  and region 2 where  $1 \gg z_{\text{cut}} \sim m_{\text{gr}}^2/Q^2$ . We use the factorized forms to evaluate the resummed cross section at NLL and compare against the partonic shower of PYTHIA 8. We find excellent agreement within the perturbative regime. At NNLL we show that the theoretical uncertainty for the normalized cross section is about 10%. Furthermore we discuss hadronization effects and we show that these effects (for  $m_{\text{gr}}^2 z_{\text{cut}} \lesssim \Lambda_{\text{QCD}}$ ) can be well described with a shape function convolved with the resummed spectrum.

Although in this work we only consider the measurement of GIM, one may also consider a plethora of other event shape measurements (e.g., N-jettiness and angularities) or even novel event-level generalizations of jet substructure measurements. An interesting example is groomed jet radius  $R_g = d_{ij}$ , which corresponds to the geometric separation of the two final branches,  $i$  and  $j$ , that stopped the pruning. In the case of jet grooming this observable has been studied extensively both experimentally [32] and theoretically [6,86,87]. In DIS and using the proposed algorithm we can generalize this observable to event level grooming.

Groomed jet substructure observables such as the energy fraction sharing  $z_g$  and the groomed radius  $R_g$  are particularly sensitive to the parton shower evolution of the jet. They are therefore great observables to nuclear medium induced modifications to the evolution of jets. The CMS [88], ALICE [89], and STAR [90] collaborations have already measured the modification to the  $z_g$  spectrum in heavy-ion collisions, although, the interpretation of these modification remains a subject of debate [91–94]. The future Electron-Ion-Collider (EIC) will offer a unique opportunity for the study of nuclear effects in  $eA$  collisions with unprecedented accuracy. These observables in the context of groomed jets have been discussed in Ref. [95]. We propose the observables,  $z_g$  and  $R_g$ , adapted to event grooming in DIS (proposed in this work) as a probe to cold nuclear matter effects at the EIC.

Beyond event shapes and jet substructure, the use of groomed jets as probes to TMDs is an other interesting concept that has been gaining significant attention recently [11,14,20], particularly in the context of EIC. Generalization of this observables to event level grooming it is also possible. One can consider the angle between the virtual photon and the groomed axis (as a probe to TMDPDFs) or the transverse momentum of an identified hadron with respect to the groomed axis (as a probe to TMDFFs and TMD evolution).<sup>7</sup>

Finally we give an event based definition of the winner-take-all (WTA) axis using the Centauro measure to cluster the whole event. We emphasize that initial state universal TMDs (such as unpolarized TMDPDFs and quark Siver's function) can be accessed by measuring the photon transverse momentum with respect to the WTA axis, similarly as discussed above for the groomed jet axis. The theoretical framework for such a measurement has been discussed in Refs. [80,81]. Furthermore, three dimensional fragmentation can also be accessed through measurements of the transverse momentum of an identified hadron with respect to the WTA axis [79].

## ACKNOWLEDGMENTS

We thank Miguel Arratia and Duff Neill for feedback on the manuscript. Y. M. is supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 754496—FELLINI.

## APPENDIX: FIXED ORDER INGREDIENTS

Here we collect from the literature all the fixed order ingredients we used for the construction of the resummed cross section. It is organized in five subsections: hard, beam, jet, collinear-soft, and ultra-soft functions. The soft function which is scaleless at this order is not discussed. At the end of the Appendix we also give the leading order full-QCD cross section for the groomed invariant mass.

### 1. Hard functions

The relevant DIS hard function given in terms of the photon's virtuality is

$$H(Q, y, \mu) = \sigma_0(Q, y) \left\{ 1 + a_s \left[ -2\gamma_{\text{cusp}}^0 \ln^2\left(\frac{\mu}{Q}\right) + \gamma_H^0 \ln\left(\frac{\mu}{Q}\right) + C_F \left(-16 + \frac{\pi^2}{3}\right) \right] + \mathcal{O}(\alpha_s^2) \right\}, \quad (\text{A1})$$

where

$$\sigma_0(Q, y) = 2\pi\alpha_e^2 \frac{1 + (1-y)^2}{Q^4} \quad (\text{A2})$$

<sup>7</sup>Groomed axis in the context of event grooming in DIS is defined by the axis pointing along the direction of the total three-momenta of given by the sum of all particles that passed grooming.

The hard function satisfies the following RGE:

$$\frac{d}{d \ln \mu} H(Q, y, \mu) = \gamma_H(Q, \mu) H(Q, y, \mu), \quad (\text{A3})$$

with  $\gamma_H$  the hard function anomalous dimension. Following the standard notation we may decompose the anomalous dimension into a term proportional to the cusp anomalous dimension,  $\Gamma_{\text{cusp}}$ , and a noncusp terms and they have the following,

$$\gamma_H(Q, \mu) = -4\gamma_{\text{cusp}}[a_s] \ln\left(\frac{\mu}{Q}\right) + \gamma_H[a_s], \quad (\text{A4})$$

where for the expansion in the strong coupling of all anomalous dimensions we are using the following notation

$$\gamma_G[a_s] = \sum_{n=0}^{\infty} a_s^{n+1} \gamma_G^n, \quad (\text{A5})$$

where the subscript  $G$  is a placeholder index for a generic function. For up to the NNLL cross section we need the following coefficients,

$$\begin{aligned} \gamma_{\text{cusp}}^0 &= 4C_F, \\ \gamma_{\text{cusp}}^1 &= 4C_F \left[ C_A \left( \frac{67}{9} - \frac{\pi^2}{3} \right) - \frac{20}{9} T_F n_f \right], \\ \gamma_{\text{cusp}}^2 &= 4C_F \left[ C_A^2 \left( \frac{245}{6} - \frac{134\pi^2}{27} + \frac{11\pi^4}{45} + \frac{22\zeta_3}{3} \right) + C_A T_F n_f \left( -\frac{418}{27} + \frac{40\pi^2}{27} - \frac{56\zeta_3}{3} \right) \right. \\ &\quad \left. + C_F T_F n_f \left( -\frac{55}{3} + 16\zeta_3 \right) - \frac{16}{27} T_F^2 n_f^2 \right], \end{aligned} \quad (\text{A6})$$

and

$$\begin{aligned} \gamma_H^0 &= -12C_F, \\ \gamma_H^1 &= C_F \left[ C_A \left( -\frac{164}{9} + 104\zeta_3 \right) + C_F (-6 + 8\pi^2 - 96\zeta_3) + \beta_0 \left( -\frac{130}{9} - 2\pi^2 \right) \right]. \end{aligned} \quad (\text{A7})$$

## 2. Groomed beam function

At NLO there is only one final state gluon and therefore no clustering effects enter at this order. The beam function is then given by the *cumulant* beam-thrust beam function with the replacement  $t_{\text{cut}} \rightarrow Q^2 z_{\text{cut}}$ . This result can be found in Ref. [74]

$$\mathcal{B}_{q/P}(x, Q^2 z_{\text{cut}}, \mu) = \sum_i \int_x^1 \frac{dz}{z} \mathcal{I}_{qi} \left( \frac{x}{z}, Q^2 z_{\text{cut}}, \mu \right) f_{i/P}(z, \mu), \quad (\text{A8})$$

where the short distance matching coefficients are

$$\mathcal{I}_{qi}(x, Q^2 z_{\text{cut}}, \mu) = \delta_{qi} \delta(1-x) + a_s \left\{ \frac{1}{2} \delta_{qi} \delta(1-x) \left[ \gamma_{\text{cusp}}^0 \ln \left( \frac{\mu^2}{Q^2 z_{\text{cut}}} \right) + \gamma_B^0 \ln \left( \frac{\mu^2}{Q^2 z_{\text{cut}}} \right) - \ln \left( \frac{\mu^2}{Q^2 z_{\text{cut}}} \right) P_{qi}(x) + \mathcal{I}_{qi}^{(1)}(x) \right] \right\}. \quad (\text{A9})$$

The one-loop QCD splitting kernels are,

$$P_{qq}(x) = 2C_F \left( \frac{1+x^2}{1-x} \right)_+, \quad P_{qg}(x) = 2T_F(1-2x(1-x)), \quad (\text{A10})$$

and the nonsingular terms are,

$$\begin{aligned}\mathcal{I}_{qq}^{(1)}(x) &= 2C_F \left[ (1+x^2) \left( \frac{\ln(1-x)}{1-x} \right)_+ - \frac{\pi^2}{6} \delta(1-x) + 1-x - \frac{1+x^2}{1-x} \ln(x) \right], \\ \mathcal{I}_{qg}^{(1)}(x) &= 2T_F + P_{qg}(x) \left[ \ln\left(\frac{1-x}{x}\right) - 1 \right].\end{aligned}\quad (\text{A11})$$

The beam function satisfies the following RGE,

$$\frac{d}{d \ln \mu} \mathcal{B}_{q/P}(x, Q^2 z_{\text{cut}}, \mu) = \gamma_B(Q^2 z_{\text{cut}}, \mu) \mathcal{B}_{q/P}(x, Q^2 z_{\text{cut}}, \mu), \quad (\text{A12})$$

where

$$\gamma_B(Q^2 z_{\text{cut}}, \mu) = 2\gamma_{\text{cusp}}[a_s] \ln\left(\frac{\mu^2}{Q^2 z_{\text{cut}}}\right) + \gamma_B[a_s], \quad (\text{A13})$$

with

$$\gamma_B^0 = 6C_F, \quad (\text{A14})$$

As discussed earlier the two-loop beam anomalous dimension,  $\gamma_B^1$ , is unknown and requires a nontrivial computation considering clustering effect. However the combination,  $\gamma_B^1 + \gamma_S^1$ , we can obtain from consistency relations,

$$\gamma_B^1 + \gamma_S^1 = C_F \left[ C_A(-7.73 - 80\zeta_3) + C_F(20 - 4\pi^2 + 48\zeta_3) + \beta_0 \left( 18.71 + \frac{2\pi^2}{3} \right) \right]. \quad (\text{A15})$$

Note for resumming all logarithms  $\ln(z_{\text{cut}})$  at NNLL requires the knowledge one of  $\gamma_B^1$  or  $\gamma_S^1$  separately.

### 3. Jet function

The same jet function appears in the factorization theorem for both regions 1 and 2. In fact the jet function that appears in these factorization theorems is the same as the jet function that appears the jet-mass measurements [96] and has been calculate up to three-loop accuracy. Here we collect the one-loop fixed order results and the two-loop anomalous dimension. In terms of the variable  $u$  Laplace conjugate of the invariant mass (in literature usually denoted with  $s$ ) we have,

$$J(u, \mu^2) = 1 + a_s \left[ \frac{1}{2} \gamma_{\text{cusp}}^0 \ln^2(u\mu^2) + \frac{1}{2} \gamma_J^0 \ln(u\mu^2) + C_F \left( 7 - 2\frac{\pi^2}{3} \right) \right], \quad (\text{A16})$$

and satisfies the following RGE:

$$\frac{d}{d \ln \mu} J(u, \mu^2) = \gamma_J(u, \mu^2) J(u, \mu^2), \quad (\text{A17})$$

where

$$\gamma_J(u, \mu^2) = -2\gamma_{\text{cusp}}[a_s] \ln(u\mu^2) + \gamma_J[a_s], \quad (\text{A18})$$

with

$$\begin{aligned}\gamma_J^0 &= 6C_F, \\ \gamma_J^1 &= C_F \left[ C_A \left( \frac{146}{9} - 80\zeta_3 \right) + C_F(3 - 4\pi^2 + 48\zeta_3) + \beta_0 \left( \frac{121}{9} + \frac{2\pi^2}{3} \right) \right].\end{aligned}\quad (\text{A19})$$

#### 4. Collinear-soft function

The collinear-soft function that appears in our analysis is the same function that also appears in the groomed jet-mass measurements for SoftDrop grooming for  $\beta = 0$ . This is true since in the collinear limit the Centauro measure and the C/A clustering measure are converge. The one-loop result for the this collinear-soft function is given in Eqs. (E.4) and (E.5) of Ref. [76] (for our case  $\alpha = 2$  and  $\beta = 0$ ),

$$\mathcal{C}(m_{cs}^2, z_{\text{cut}}, \mu) = \frac{1}{N_C} \text{tr} \left\langle \text{T}[Y_n^\dagger W_l] \delta \left( m_{cs}^2 - \Theta_{\text{gr.}} \frac{Q^2 e_2^{(2)}}{4} \right) \bar{\text{T}}[W_l^\dagger Y_n] \right\rangle = \frac{Q^2}{4} \mathcal{S}_C(z_{\text{cut}} e_2^{(2)}), \quad (\text{A20})$$

where  $\Theta_{\text{gr.}}$  is a particle level selection function for particles that pass grooming. Taking the Laplace transform of the one-loop result with respect to  $m_{cs}^2$  we have,

$$\mathcal{C}(u, z_{\text{cut}}, \mu) = 1 - a_s \frac{\gamma_{\text{cusp}}^0}{2} \ln^2 \left( \frac{\mu^2 u}{z_{\text{cut}}} \right). \quad (\text{A21})$$

This is the collinear-soft function at one loop for the groomed invariant mass measurement relevant to region 1. It satisfies the following RG,

$$\frac{d}{d \ln \mu} \mathcal{C}(u, z_{\text{cut}}, \mu) = \gamma_C(u, z_{\text{cut}}, \mu) \mathcal{C}(u, z_{\text{cut}}, \mu), \quad (\text{A22})$$

where

$$\gamma_C(u, z_{\text{cut}}, \mu) = -2\gamma_{\text{cusp}} \ln \left( \frac{\mu^2 u}{z_{\text{cut}}} \right) + \gamma_C[a_s], \quad (\text{A23})$$

with

$$\begin{aligned} \gamma_C^0 &= 0, \\ \gamma_C^1 &= C_F \left[ C_A(9.73 + 56\zeta_3) - C_F 17.00 + \beta_0 \left( -17.71 + \frac{2\pi^2}{3} \right) \right]. \end{aligned} \quad (\text{A24})$$

#### 5. Ultra-soft function

We give the operator definition of the collinear ultra-soft function,  $\mathcal{U}$ , That appears in the factorization theorem for region 2. We then continue with the one loop calculation and show that at this order gives the same result as the collinear-soft function. The operator definition of ultra-soft function is written in terms of the SCET<sub>1</sub> ultra-soft Wilson-lines (hence the name ultra-soft function) along the  $n$  and  $\bar{n}$  directions,

$$\mathcal{U}(m_{us}^2, Q, z_{\text{cut}}, \mu) = \frac{1}{Q N_C} \text{tr} \left\langle \text{T}[Y_n^\dagger Y_{\bar{n}}] \delta \left( \frac{m_{us}^2}{Q} - \hat{\mathcal{M}}_{\text{gr.}}^- \right) \bar{\text{T}}[Y_n^\dagger Y_{\bar{n}}] \right\rangle, \quad (\text{A25})$$

where  $\hat{\mathcal{M}}_{\text{gr.}}^-$  its is the operator that returns the  $\bar{n} \cdot p$  component of the total momenta of the particles that pass grooming. We can cont give a close form for this operator at all orders in perturbation theory since it depends on the number of final state particles, but it can be defined at the Feynman diagram level. At leading order there are no virtual or real gluons and all Wilson-lines are set to the identity operator and thus we get,

$$\mathcal{U}(m_{us}^2, Q, z_{\text{cut}}, \mu) = \delta(m_{us}^2) + \mathcal{O}(\alpha_s), \quad (\text{A26})$$

At next-to-leading order we need to consider the diagrams shown in Fig. 10. Diagram (a) corresponds to the virtual gluon exchange between the Wilson-lines in the  $n$  and  $\bar{n}$  directions. In pure dimensional regularization this yields a scaleless integral and thus vanishes in our scheme. Diagram (b) involves one real gluon exchange. The contribution of that gluon with momentum,  $k^\mu$ , to the invariant mass can be obtained by considering the sum of all radiation that pass grooming,

$$m_{\text{gr.}}^2 = \left( \sum_i p_i^\mu \right) = (p_{\bar{n}}^\mu + k^\mu)^2 = p_{\bar{n}}^2 + Qk^-. \quad (\text{A27})$$



Assuming that the gluon passes grooming there are two terms relevant to the invariant mass measurement: the contribution from the  $\bar{n}$ -collinear radiation, and the interference term  $Qk^-$  which involves the virtual gluon. Therefore, the diagram (b) contribution to the one-loop u-soft function is given by

$$d_{(b)} = g^2 C_F \left( \frac{e^{\gamma_E} \mu^2}{4\pi} \right)^\epsilon \int \frac{d^d k}{(2\pi)^{d-1}} \frac{1}{k^- k^+} \delta(k^2) [\delta(m_{us}^2 - Qk^-) \Theta(k^+ - Qz_{\text{cut}}) + \delta(m_{us}^2) \Theta(Qz_{\text{cut}} - k^+)]. \quad (\text{A28})$$

The first term in the square brackets corresponds to the case where the ultra-soft gluon passes grooming and the second term in the case where it fails. The second term also yield a scaleless integral and thus ignore in our calculation. Performing the trivial integration and multiplying with a factor if two (for the mirror diagram) we find that the u-soft and collinear-soft functions are identical at this order.

## 6. Fixed order QCD at $\mathcal{O}(\alpha_s)$

In this section we give the main aspects and result of the  $\mathcal{O}(\alpha_s)$  calculation of the GIM cross section in QCD. For this task we will closely follow the derivation and notation of Ref. [97] which was suited for the calculation of 1-jettiness in DIS but also very adaptable to our calculation. The differential cross section can be written in terms of the hadronic tensor,

$$\frac{d\sigma}{dx dQ^2 dm_{\text{gr}}^2} = \frac{\alpha^2}{2Q^4} ([1 + (1-y)^2] W_G(x, Q^2, m_{\text{gr}}^2) + 2x^2 [6 - 6y + y^2] W_P(x, Q^2, m_{\text{gr}}^2)), \quad (\text{A29})$$

where

$$W_G(x, Q^2, m_{\text{gr}}^2) = -g_{\mu\nu} W^{\mu\nu}(x, Q^2, m_{\text{gr}}^2), \quad W_P(x, Q^2, m_{\text{gr}}^2) = \frac{P_\mu P_\nu}{Q^2} W^{\mu\nu}(x, Q^2, m_{\text{gr}}^2), \quad (\text{A30})$$

which we refer to as the  $G$  and  $P$  projections of the hadronic tensor. For groomed invariant mass measurement the hadronic tensor is written in terms of the QCD quark current operator,  $J^\mu(r)$ , as follows:

$$W^{\mu\nu}(x, Q^2, m_{\text{gr}}^2) = \int d^4 r \exp(iq \cdot r) \langle P | J^{\mu\dagger}(r) \delta(m_{\text{gr}}^2 - \hat{\mathcal{M}}_{\text{gr}}) J^\nu(0) | P \rangle. \quad (\text{A31})$$

The operator  $\hat{\mathcal{M}}_{\text{gr}}$  projects the value of the hadronic GIM for a given final state. Our goal in this section is to give explicit results for  $W_G$  and  $W_P$ .

The hadronic tensor can be matched onto the collinear parton distribution functions in an operator product expansion (OPE),

$$W_j(x, Q^2, m_{\text{gr}}^2) = \sum_{i \in \{q, \bar{q}, g\}} \int_x^1 \frac{d\xi}{\xi} f_{i/P}(\xi, \mu) w_j^i\left(\frac{x}{\xi}, Q^2, m_{\text{gr}}^2, \mu\right) \times \left[ 1 + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}}{\sqrt{m_{\text{gr}}^2}}\right) \right] \quad (\text{A32})$$

where  $w_j^i$  are the perturbative matching coefficients calculable order-by-order in perturbation theory. We now discuss the calculation of the  $\mathcal{O}(\alpha_s)$  (leading order)  $w_j^i$  functions. We consider only finite values of the GIM (i.e.,  $m_{\text{gr}}^2 > 0$ ) and thus drop any regulator dependence and only consider the real emissions. We therefore have two final state partons that contribute to the measurement. Their momenta (working in the Breit frame) are given by,

$$\begin{aligned} p_1^\mu &= Q(1-v) \frac{\bar{n}^\mu}{2} + \frac{1-x}{x} Qv \frac{n^\mu}{2} - p_\perp^\mu \\ p_2^\mu &= Qv \frac{\bar{n}^\mu}{2} + \frac{1-x}{x} Q(1-v) \frac{n^\mu}{2} + p_\perp^\mu \end{aligned} \quad (\text{A33})$$

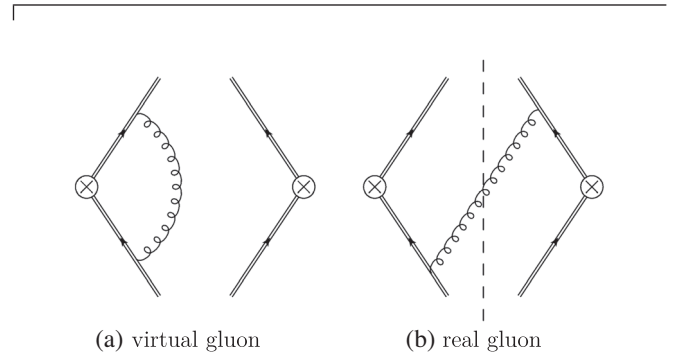


FIG. 10. Diagrams contributing at next-to-leading order in the perturbative calculation of the u-soft function. Mirror diagrams not shown in this figure.

and  $p_{\perp}^{\mu}$  can be obtained from the on-shell condition  $p_1^2 = p_2^2 = 0$ . Thus for fixed values of  $x - Q^2$  the only free parameter is  $v$  defined as  $v \equiv p_2^+ / Q$ . The grooming algorithm will first cluster these two particles and in the declustering process the grooming condition reads,

$$\min[v, 1 - v] > z_{\text{cut}}. \quad (\text{A34})$$

Therefore if this condition is satisfied both particles will pass grooming and contribute to the measurement,

$$m_{\text{gr.}}^2 = (p_1 + p_2)^2 = \frac{1-x}{x} Q^2. \quad (\text{A35})$$

Otherwise, if the condition is not satisfied only one particle pass grooming yielding  $m_{\text{gr.}}^2 = p_i^2 = 0$  and thus does not contribute to the finite GIM spectrum. We then have for the matching coefficients,

$$w_{\mu\nu}^i(x, Q^2, m_{\text{gr.}}^2, \mu)|_{\mathcal{O}(\alpha_s)} = -\frac{1}{16\pi} \delta\left(m_{\text{gr.}}^2 - \frac{1-x}{x} Q^2\right) \int_{z_{\text{cut}}}^{1-z_{\text{cut}}} dv \mathcal{M}_v^{\text{real}}(v, Q^2, \mu) (\mathcal{M}_\mu^{\text{real}}(v, Q^2, \mu))^* \quad (\text{A36})$$

with  $\mathcal{M}_v^{\text{real}}$  the QCD amplitude from real emission in the process  $\gamma^* i \rightarrow 12$ . The projected matching coefficients that appear in Eq. (A32) are given by

$$w_G^i(x, Q^2, m_{\text{gr.}}^2, \mu) = -g^{\mu\nu} w_{\mu\nu}^i(x, Q^2, m_{\text{gr.}}^2, \mu), \quad w_P^i(x, Q^2, m_{\text{gr.}}^2, \mu) = \frac{P^\mu P^\nu}{Q^2} w_{\mu\nu}^i(x, Q^2, m_{\text{gr.}}^2, \mu), \quad (\text{A37})$$

the equivalent projections of the amplitude squared and the integration over  $v$  can be done using the expressions in Eqs. (B.8–B.10) of ref. [97]. Finally the integration over  $\xi$  in Eq. (A32) can be trivially performed using the measurement delta function in Eq. (A36). Therefore the  $G$  and  $P$  projections of the hadronic tensor are

$$W_j(x, Q^2, m_{\text{gr.}}^2) = \alpha_s(\mu) \frac{2x}{\xi Q^2} \Theta(x < \xi < 1) \sum_{i \in \{q, \bar{q}, g\}} Q_i^2 f_{i/p}(\xi, \mu) h_j^i\left(\frac{x}{\xi}, z_{\text{cut}}\right) \quad (\text{A38})$$

with  $\xi = x(1 + m_{\text{gr.}}^2/Q^2)$  and

$$\begin{aligned} h_G^q(x, z_{\text{cut}}) &= C_F \frac{1}{2(1-x)} \left( (1-4x)(1-2z_{\text{cut}}) + 2(1+x^2) \ln \frac{1-z_{\text{cut}}}{z_{\text{cut}}} \right) \\ h_P^q(x, z_{\text{cut}}) &= C_F \frac{1}{4x} (1-2z_{\text{cut}}) \\ h_G^g(x, z_{\text{cut}}) &= T_F (-2(1-2z_{\text{cut}}) + 4(1-2x+2x^2) \tanh^{-1}(1-2z_{\text{cut}})) \\ h_P^g(x, z_{\text{cut}}) &= T_F \frac{1-x}{x} (1-2z_{\text{cut}}) \end{aligned} \quad (\text{A39})$$

and  $h_{G/P}^{\bar{q}} = h_{G/P}^q$ . For quarks and antiquarks the charge  $Q_i^2$  is simply the electric charge of that parton and for gluons we have,

$$Q_g^2 \equiv \sum_f Q_f^2 \quad (\text{A40})$$

where the sum runs over the quark flavors only.

- [1] A. Accardi *et al.*, *Eur. Phys. J. A* **52**, 268 (2016).
- [2] D. Krohn, J. Thaler, and L.-T. Wang, *J. High Energy Phys.* **02** (2010) 084.
- [3] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev. D* **81**, 094023 (2010).
- [4] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *J. High Energy Phys.* **09** (2013) 029.
- [5] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **75**, 59 (2015).
- [6] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *J. High Energy Phys.* **05** (2014) 146.
- [7] C. Frye, A. J. Larkoski, J. Thaler, and K. Zhou, *J. High Energy Phys.* **09** (2017) 083.
- [8] F. A. Dreyer, L. Necib, G. Soyez, and J. Thaler, *J. High Energy Phys.* **06** (2018) 093.
- [9] A. J. Larkoski, I. Moulton, and D. Neill, *J. High Energy Phys.* **02** (2018) 144.
- [10] A. J. Larkoski, I. Moulton, and D. Neill, [arXiv:1708.06760](https://arxiv.org/abs/1708.06760).
- [11] Y. Makris, D. Neill, and V. Vaidya, *J. High Energy Phys.* **07** (2018) 167.
- [12] J. Baron, S. Marzani, and V. Theeuwes, *J. High Energy Phys.* **08** (2018) 105; **05** (2019) 56(E).
- [13] Z.-B. Kang, K. Lee, X. Liu, and F. Ringer, *Phys. Lett. B* **793**, 41 (2019).
- [14] Y. Makris and V. Vaidya, *J. High Energy Phys.* **10** (2018) 019.
- [15] A. Kardos, G. Somogyi, and Z. Trócsányi, *Phys. Lett. B* **786**, 313 (2018).
- [16] J. Chay and C. Kim, *J. Korean Phys. Soc.* **74**, 439 (2019).
- [17] D. Napoletano and G. Soyez, *J. High Energy Phys.* **12** (2018) 031.
- [18] C. Lee, P. Shrivastava, and V. Vaidya, *J. High Energy Phys.* **09** (2019) 045.
- [19] A. H. Hoang, S. Mantry, A. Pathak, and I. W. Stewart, *J. High Energy Phys.* **12** (2019) 002.
- [20] D. Gutierrez-Reyes, Y. Makris, V. Vaidya, I. Scimemi, and L. Zoppi, *J. High Energy Phys.* **08** (2019) 161.
- [21] S. Marzani, D. Reichelt, S. Schumann, G. Soyez, and V. Theeuwes, *J. High Energy Phys.* **11** (2019) 179.
- [22] Y. Mehtar-Tani, A. Soto-Ontoso, and K. Tywoniuk, *Phys. Rev. D* **101**, 034004 (2020).
- [23] A. Kardos, A. J. Larkoski, and Z. Trócsányi, *Phys. Rev. D* **101**, 114034 (2020).
- [24] A. J. Larkoski, *J. High Energy Phys.* **09** (2020) 072.
- [25] A. Lifson, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **10** (2020) 170.
- [26] P. Cal, K. Lee, F. Ringer, and W. J. Waalewijn, *J. High Energy Phys.* **11** (2020) 012.
- [27] M. Aaboud *et al.* (ATLAS Collaboration), *Phys. Rev. Lett.* **121**, 092001 (2018).
- [28] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **10** (2018) 161.
- [29] A. M. Sirunyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **11** (2018) 113.
- [30] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. D* **101**, 052007 (2020).
- [31] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. Lett.* **124**, 222002 (2020).
- [32] J. Adam *et al.* (STAR Collaboration), *Phys. Lett. B* **811**, 135846 (2020).
- [33] C. W. Bauer, S. Fleming, and M. E. Luke, *Phys. Rev. D* **63**, 014006 (2000).
- [34] C. W. Bauer, S. Fleming, D. Pirjol, and I. W. Stewart, *Phys. Rev. D* **63**, 114020 (2001).
- [35] C. W. Bauer and I. W. Stewart, *Phys. Lett. B* **516**, 134 (2001).
- [36] C. W. Bauer, D. Pirjol, and I. W. Stewart, *Phys. Rev. D* **65**, 054022 (2002).
- [37] M. Beneke, A. Chapovsky, M. Diehl, and T. Feldmann, *Nucl. Phys.* **B643**, 431 (2002).
- [38] M. Arratia, Y. Makris, D. Neill, F. Ringer, and N. Sato, [arXiv:2006.10751](https://arxiv.org/abs/2006.10751) [*Phys. Rev. Lett.* (to be published)].
- [39] P. Abreu *et al.* (DELPHI Collaboration), *Phys. Lett. B* **456**, 322 (1999).
- [40] P. Movilla Fernandez, O. Biebel, S. Bethke, S. Kluth, and P. Pfeifenschneider (JADE Collaboration), *Eur. Phys. J. C* **1**, 461 (1998).
- [41] G. Abbiendi *et al.* (OPAL Collaboration), *Eur. Phys. J. C* **40**, 287 (2005).
- [42] R. Davison and B. Webber, *Eur. Phys. J. C* **59**, 13 (2009).
- [43] R. Abbate, M. Fickinger, A. H. Hoang, V. Mateu, and I. W. Stewart, *Phys. Rev. D* **83**, 074021 (2011).
- [44] R. Abbate, M. Fickinger, A. H. Hoang, V. Mateu, and I. W. Stewart, *Phys. Rev. D* **86**, 094002 (2012).
- [45] T. Gehrmann, G. Luisoni, and P. F. Monni, *Eur. Phys. J. C* **73**, 2265 (2013).
- [46] A. H. Hoang, D. W. Kolodrubetz, V. Mateu, and I. W. Stewart, *Phys. Rev. D* **91**, 094018 (2015).
- [47] R. Akers *et al.* (OPAL Collaboration), *Z. Phys. C* **65**, 31 (1995).
- [48] G. Rodrigo, A. Santamaria, and M. S. Bilenky, *Phys. Rev. Lett.* **79**, 193 (1997).
- [49] P. Abreu *et al.* (DELPHI Collaboration), *Phys. Lett. B* **418**, 430 (1998).
- [50] R. Barate *et al.* (ALEPH Collaboration), *Eur. Phys. J. C* **18**, 1 (2000).
- [51] G. Abbiendi *et al.* (OPAL Collaboration), *Eur. Phys. J. C* **21**, 411 (2001).
- [52] D. Schlatter *et al.*, *Phys. Rev. Lett.* **49**, 521 (1982).
- [53] B. Adeva *et al.*, *Phys. Rev. Lett.* **50**, 2051 (1983).
- [54] S. Catani, G. Turnock, and B. Webber, *Phys. Lett. B* **295**, 269 (1992).
- [55] J. Abdallah *et al.* (DELPHI Collaboration), *Eur. Phys. J. C* **29**, 285 (2003).
- [56] P. Achard *et al.* (L3 Collaboration), *Phys. Rep.* **399**, 71 (2004).
- [57] A. Gehrmann-De Ridder, T. Gehrmann, E. Glover, and G. Heinrich, *J. High Energy Phys.* **05** (2009) 106.
- [58] S. Weinzierl, *J. High Energy Phys.* **06** (2009) 041.
- [59] C. Lee and G. F. Sterman, *eConf C0601121*, A001 (2006), [arXiv:hep-ph/0603066](https://arxiv.org/abs/hep-ph/0603066).
- [60] C. Lee and G. F. Sterman, *Phys. Rev. D* **75**, 014022 (2007).
- [61] T. Gehrmann, M. Jaquier, and G. Luisoni, *Eur. Phys. J. C* **67**, 57 (2010).
- [62] T. Becher and G. Bell, *Phys. Rev. Lett.* **112**, 182002 (2014).
- [63] M. Akrawy *et al.* (OPAL Collaboration), *Z. Phys. C* **47**, 505 (1990).
- [64] K. Wraight and P. Skands, *Eur. Phys. J. C* **71**, 1628 (2011).
- [65] P. Skands, S. Carrazza, and J. Rojo, *Eur. Phys. J. C* **74**, 3024 (2014).

- [66] M. Bahr *et al.*, *Eur. Phys. J. C* **58**, 639 (2008).
- [67] A. Kardos, G. Somogyi, and Z. Trócsányi, *Proc. Sci., LL2016* (2016) 021.
- [68] V. Del Duca, C. Duhr, A. Kardos, G. Somogyi, Z. Szőr, Z. Trócsányi, and Z. Tulipánt, *Phys. Rev. D* **94**, 074019 (2016).
- [69] V. Del Duca, C. Duhr, A. Kardos, G. Somogyi, and Z. Trócsányi, *Phys. Rev. Lett.* **117**, 152004 (2016).
- [70] A. Kardos, A. J. Larkoski, and Z. Trócsányi, *Phys. Lett. B* **809**, 135704 (2020).
- [71] D. Kang, C. Lee, and I. W. Stewart, *Phys. Rev. D* **88**, 054004 (2013).
- [72] C. W. Bauer, F. J. Tackmann, J. R. Walsh, and S. Zuberi, *Phys. Rev. D* **85**, 074006 (2012).
- [73] S. Fleming, A. K. Leibovich, and T. Mehen, *Phys. Rev. D* **74**, 114004 (2006).
- [74] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, *Phys. Rev. D* **81**, 094035 (2010).
- [75] Z. Ligeti, I. W. Stewart, and F. J. Tackmann, *Phys. Rev. D* **78**, 114014 (2008).
- [76] C. Frye, A. J. Larkoski, M. D. Schwartz, and K. Yan, *J. High Energy Phys.* **07** (2016) 064.
- [77] D. Bertolini, T. Chan, and J. Thaler, *J. High Energy Phys.* **04** (2014) 013.
- [78] G. Salam (to be published).
- [79] D. Neill, I. Scimemi, and W. J. Waalewijn, *J. High Energy Phys.* **04** (2017) 020.
- [80] D. Gutierrez-Reyes, I. Scimemi, W. J. Waalewijn, and L. Zoppi, *Phys. Rev. Lett.* **121**, 162001 (2018).
- [81] D. Gutierrez-Reyes, I. Scimemi, W. J. Waalewijn, and L. Zoppi, *J. High Energy Phys.* **10** (2019) 031.
- [82] Y.-T. Chien, R. Rahn, S. Schrijnder van Velzen, D. Y. Shao, W. J. Waalewijn, and B. Wu, [arXiv:2005.12279](https://arxiv.org/abs/2005.12279).
- [83] X. Liu, F. Ringer, W. Vogelsang, and F. Yuan, *Phys. Rev. Lett.* **122**, 192003 (2019).
- [84] X. Liu, F. Ringer, W. Vogelsang, and F. Yuan, *Phys. Rev. D* **102**, 094022 (2020).
- [85] J. Baron, D. Reichelt, S. Schumann, N. Schwanemann, and V. Theeuwes, [arXiv:2012.09574](https://arxiv.org/abs/2012.09574).
- [86] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 005.
- [87] Z.-B. Kang, K. Lee, X. Liu, D. Neill, and F. Ringer, *J. High Energy Phys.* **02** (2020) 054.
- [88] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **120**, 142302 (2018).
- [89] S. Acharya *et al.* (ALICE Collaboration), *Phys. Lett. B* **802**, 135227 (2020).
- [90] K. Kauder (STAR Collaboration), *Nucl. Part. Phys. Proc.* **289–290**, 137 (2017).
- [91] Y. Mehtar-Tani and K. Tywoniuk, *J. High Energy Phys.* **04** (2017) 125.
- [92] Y.-T. Chien and I. Vitev, *Phys. Rev. Lett.* **119**, 112301 (2017).
- [93] G. Milhano, U. A. Wiedemann, and K. C. Zapp, *Phys. Lett. B* **779**, 409 (2018).
- [94] N.-B. Chang, S. Cao, and G.-Y. Qin, *Phys. Lett. B* **781**, 423 (2018).
- [95] M. Arratia, Y. Song, F. Ringer, and B. V. Jacak, *Phys. Rev. C* **101**, 065204 (2020).
- [96] C. W. Bauer and A. V. Manohar, *Phys. Rev. D* **70**, 034024 (2004).
- [97] D. Kang, C. Lee, and I. W. Stewart, *J. High Energy Phys.* **11** (2014) 132.