

## Asymmetric jet clustering in deep-inelastic scattering

M. Arratia<sup>1,2,\*</sup> Y. Makris<sup>3,†</sup> D. Neill<sup>4,‡</sup> F. Ringer<sup>5,6,§</sup> and N. Sato<sup>7,||</sup>

<sup>1</sup>Physics Department, University of California, Riverside, California 92521, USA

<sup>2</sup>Thomas Jefferson National Accelerator Facility, Newport News, Virginia 23606, USA

<sup>3</sup>INFN Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy

<sup>4</sup>Theoretical Division, MS B283, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

<sup>5</sup>Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

<sup>6</sup>Physics Department, University of California, Berkeley, California 94720, USA

<sup>7</sup>Theory Center, Jefferson Laboratory, Newport News, Virginia 23606, USA

 (Received 27 August 2020; revised 1 February 2021; accepted 17 June 2021; published 9 August 2021)

We propose a new jet algorithm for deep-inelastic scattering (DIS) that accounts for the target-current asymmetry in the Breit frame. The Centauro algorithm is longitudinally invariant and can cluster jets with near-to Born kinematics, which enables novel studies of transverse-momentum-dependent observables. Furthermore, we show that spherically-invariant algorithms in the Breit frame give access to low-energy jets from current fragmentation. We perform a calculation of the energy spectrum of Centauro jets at next-to-leading logarithmic accuracy and compare to PYTHIA simulations. We furthermore propose novel studies in unpolarized, polarized, and nuclear DIS at the future Electron-Ion Collider.

DOI: [10.1103/PhysRevD.104.034005](https://doi.org/10.1103/PhysRevD.104.034005)

### I. INTRODUCTION

Understanding the structure of nucleons and nuclei in terms of quark and gluons remains an open goal. Jet production in deep inelastic scattering (DIS) provides an excellent tool for this endeavor. The future Electron-Ion Collider (EIC) [1] will produce the first jets in polarized and nuclear DIS, which will enable a rich jet program [2–43].

The HERA jet measurements in DIS targeted gluon-initiated processes by requiring large transverse momentum in the Breit frame [44]. This suppresses the Born configuration,  $\gamma^*q \rightarrow q$ , which has recently been postulated as key to probe transverse-momentum dependent (TMD) PDFs [11–13]. Complementary to semi-inclusive DIS observables, jets avoid nonperturbative TMD fragmentation functions. Moreover, modern jet substructure techniques [45] offer new methods for precise QCD calculations and to control nonperturbative effects, e.g., grooming or a recoil-free axis can be used to minimize hadronization effects or study TMD evolution [46]. These techniques also provide

new ways to connect to lattice QCD calculations, e.g., of the nonperturbative Collins-Soper kernel [47,48].

The Breit frame plays a central role in jet clustering for DIS [49], and it allows for a factorized TMD cross section in terms of the same soft and un-subtracted TMD functions as in Drell-Yan and  $e^+e^- \rightarrow$  dihadron/dijet processes [11–13]. However, the longitudinally-invariant (LI) algorithms commonly used in DIS cannot cluster jets that enclose the beam axis given by the proton/photon direction (see Fig. 1).

In this letter, we introduce a new jet algorithm that is longitudinally invariant but can capture jets close to the Born configuration in the Breit frame. In addition, we use spherically-invariant (SI) algorithms to study the jet energy spectrum and find that they can separate the current and target fragmentation region even for soft jets. Finally, we suggest novel studies of jet energy and TMD observables.

### II. NOTATION AND DIS KINEMATICS

In the Breit frame, the virtual photon momentum is given by:

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

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

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

\*miguel.arratia@ucr.edu  
†yiannis.makris@pv.infn.it  
‡duff.neill@gmail.com  
§fmringer@lbl.gov  
||nsato@jlab.org

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). 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>.

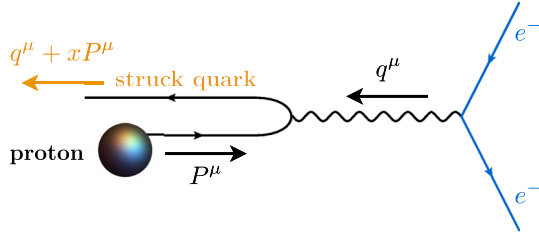


FIG. 1. DIS Born kinematics in the Breit frame.

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

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

The fragmentation of the struck-quark yields a jet that points to the beam direction. The algorithms we introduce below are designed to capture this jet. We define the scaling variable:

$$z_{\text{jet}} = \frac{P \cdot p_{\text{jet}}}{P \cdot q} \Big|_{\text{Breit frame}} z_{\text{jet}} = n \cdot p_{\text{jet}}/Q = p_{\text{jet}}^+/Q. \quad (4)$$

At leading-logarithmic accuracy,  $z_{\text{jet}}$  is the fraction of the struck-quark momentum carried by the jet.

### III. NEW JET ALGORITHMS FOR DIS

The longitudinally-invariant  $k_T$ -type jet algorithms [50–54] use the following distance measure:

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \Delta R_{ij}^2 / R^2, \quad d_{iB} = p_{Ti}^{2p}, \quad (5)$$

where  $\Delta R_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ . Here  $d_{ij}$  is the distance between two particles in the event and  $d_{iB}$  is the beam distance. Since they cluster particles in the rapidity-azimuth ( $y - \phi$ ) plane, they cannot form a jet enclosing the  $\bar{n}^\mu$  direction ( $y = -\infty$ ).

One way to bypass this problem is to use spherically-invariant algorithms. Catani *et al.* first proposed to adapt spherically-invariant algorithms to DIS in Ref. [55]. In this study we consider the  $k_T$ -type algorithms for  $e^+e^-$  collisions [54,56], which have the following distance measure:

$$d_{ij} = \min(E_i^{2p}, E_j^{2p}) \frac{1 - c_{ij}}{1 - c_R}, \quad d_{iB} = E_i^{2p}, \quad (6)$$

where  $c_{ij} = \cos \theta_{ij}$  and  $c_R = \cos R$ . However, these algorithms lack the longitudinal invariance that connects the class of frames related to the Breit frame by  $\hat{z}$  boosts, which is a crucial feature of jet clustering [49]. For example, it is important for multijet events where the parton kinematics is not constrained by  $x_B$  and  $Q^2$ , and

to identify photoproduction or separate the beam remnant from target jets [57].

To solve this issue, we introduce a new jet algorithm that is longitudinally invariant along the Breit frame beam axis but yet captures the struck-quark jet. Recently, Boronati *et al.* [58] proposed a hybrid algorithm that suppresses  $\gamma\gamma$  background in  $e^+e^-$  colliders. In contrast, we suggest a jet algorithm that is asymmetric in the current and target directions, and suggest novel studies for spherically invariant algorithms in DIS.

Starting with the distance measure of the Cambridge/Aachen (C/A) algorithm for  $e^+e^-$  (i.e., Eq. (6) for  $p = 0$ ), we write the numerator in Eq. (6) in terms of the unit vectors along the directions of particles  $i$  and  $j$ ,

$$1 - c_{ij} = 1 - \hat{n}_i \cdot \hat{n}_j = 1 - s_i s_j \cos \Delta\phi_{ij} - c_i c_j, \quad (7)$$

with  $c_i = \cos \theta_i$  and  $s_i = \sin \theta_i$ . Expanding in the very current limit (i.e.,  $\bar{\theta}_i \equiv \pi - \theta_i \ll 1$ ) we find:

$$1 - c_{ij} \simeq \frac{1}{2}(\bar{\theta}_i - \bar{\theta}_j)^2 + \bar{\theta}_i \bar{\theta}_j (1 - \cos \Delta\phi_{ij}). \quad (8)$$

We then introduce the replacements:

$$\bar{\theta}_i \rightarrow f_i = f(\bar{\eta}_i), \quad \bar{\eta}_i \equiv -\frac{2Q}{\bar{n} \cdot q} \frac{p_i^\perp}{n \cdot p_i}, \quad (9)$$

where the function  $f$  must satisfy:  $f(x) = x + \mathcal{O}(x^2)$ , and  $p_i^\perp$  is the transverse momentum in the Breit frame. The term  $\bar{\eta}_i$  (which in the Breit frame is  $2p_i^\perp/(n \cdot p_i)$ ) introduces an asymmetry: in the current region the distance between particles is given by their separation in  $\bar{\eta}$ , which decreases as particles become closer in angle. In contrast, in the target region  $\bar{\eta}$  diverges and thus prevents jets from enclosing the proton beam direction, like the anti- $k_T$ (LI) algorithm. We thus introduce the following distance measure:

$$d_{ij} = [(\Delta f_{ij})^2 + 2f_i f_j (1 - \cos \Delta\phi_{ij})] / R^2, \quad d_{iB} = 1 \quad (10)$$

which defines a new class of algorithms, which we call Centauro algorithms. Two relevant choices [59] for the function  $f$  are

$$f(x) = x, \quad f(x) = \sinh^{-1}(x). \quad (11)$$

The Centauro algorithm is invariant along the  $\hat{z}$  direction, but in the current hemisphere it matches the spherically-invariant algorithms [see Eq. (6)]. This feature is largely independent of the choice of  $f$  [60].

#### IV. COMPARISON TO PREVIOUS BREIT FRAME OBSERVABLES

It is worthwhile to examine the relationship between our inclusive jet algorithm, and other observables traditionally used to examine the physics of jet production in the Breit frame. Historically, much attention has been devoted to exclusive event shapes, Refs. [39,41,61–63], rather than inclusive jet production. In the event shape, one selects for an explicit geometrical configuration of energy flow in the event, with the understanding that the majority of events will have this configuration due to the enhancement of that region of phase-space due to collinear or soft singularities, a classic example being a two-jet configuration in  $e^+e^-$  with energy flowing parallel to an axis. For two-jet event shapes, whose control variable we generically call  $\tau$ , takes on small values in the two jet configuration, while large values correspond to multijet configurations. In the Breit frame, a corresponding configuration is to replace one of the jets with the initial state beam and keeping a single final state jet, with all events divided into two hemispheres. Implicitly, one then forces every event to have essentially a single final state jet, whose invariant mass is large or small depending on the value of the event shape.

Centauro and anti- $k_T$ (SI), like any inclusive jet algorithm, will not select any specific event geometry with its control variable  $R$ , the jet radius. Events clustered with large  $R$  may have only one or two jets found, though the actual event display may look “multijet” to the eye. However, jets at threshold,  $z_{\text{jet}} \rightarrow 1$ , can act as an implicit event shape: a single jet is now carrying the bulk of the final state’s total invariant mass, making it kinematically impossible for multiple other hard jets to be found in the event. When  $R \sim 1$ , this is essentially the observable defined in Ref. [61]. In that reference, all events in the Breit frame were divided into two hemispheres, with one hemisphere encompassing the energetic final state jet(s). The final-state

hemisphere energy loss was defined, which quantified the amount of energy of the final state that leaked out of the hemisphere containing the struck quark. If we take  $R \sim 1$  and  $z_{\text{jet}} \rightarrow 1$ , the Centauro algorithm will generically find a “hemisphere” containing the struck quark, and  $1 - z_{\text{jet}}$  also measures the energy that leaks out of this “hemisphere.” Part of the advantage of the inclusive algorithms proposed in this paper versus the previously used event shapes is that while it can act as an implicit event shape at large  $R$  and at threshold, thus capturing much the same physics as an event shape, one can also interpolate into the fragmentation regime at small radius  $R$  or generic values of  $z_{\text{jet}}$ : the physics that one can probe for is much richer, since one has not locked into representing events as single final state jet configuration. Thus if one wants to study the transition between nonperturbative hadron formation through fragmentation to perturbative jet production, this is easily accomplished by examining the evolution of the energy spectrum’s change as a function of  $R$ . Such a study would be illuminating, for instance, in differentiating hadron and jet formation in a cold medium versus the vacuum. A corresponding study in event shapes is, at best, a very round-about means to capturing this physics, or at worst, simply not possible.

#### V. SIMULATION RESULTS AND APPLICATIONS

Throughout this paper we analyze DIS events with  $Q > 10$  GeV simulated in PYTHIA8 [64] with 10 and 100 GeV electron and proton beam energies respectively [65]. We exclude neutrinos and particles with  $|\eta| > 4$  or  $p_T < 200$  MeV in the laboratory frame. We use FastJet [54] to cluster jets in the laboratory frame with the anti- $k_T$ (LI), anti- $k_T$ (SI), and Centauro algorithms [67]; Fig. 2 illustrates the resulting jet clustering for an exemplary PYTHIA8 event. The anti- $k_T$ (LI) algorithm clusters the particles from the fragmentation of the struck-quark into four different

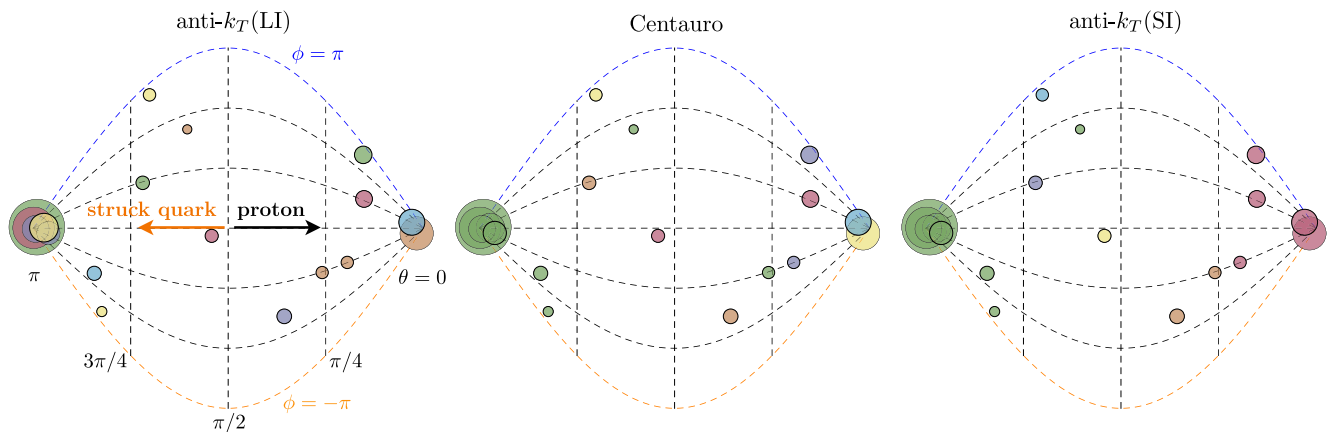


FIG. 2. Jet clustering in the Breit frame using the longitudinally-invariant anti- $k_T$ (LI), Centauro, and spherically-invariant anti- $k_T$ (SI) algorithms in a DIS event simulated with PYTHIA8. Each particle is illustrated as a disk with area proportional to its energy and the position corresponds to the direction of its momentum projected onto the unfolded sphere about the hard-scattering vertex. The vertical dashed lines correspond to constant  $\theta$  and curved lines to constant  $\phi$ . All the particles clustered into a given jet are colored the same.

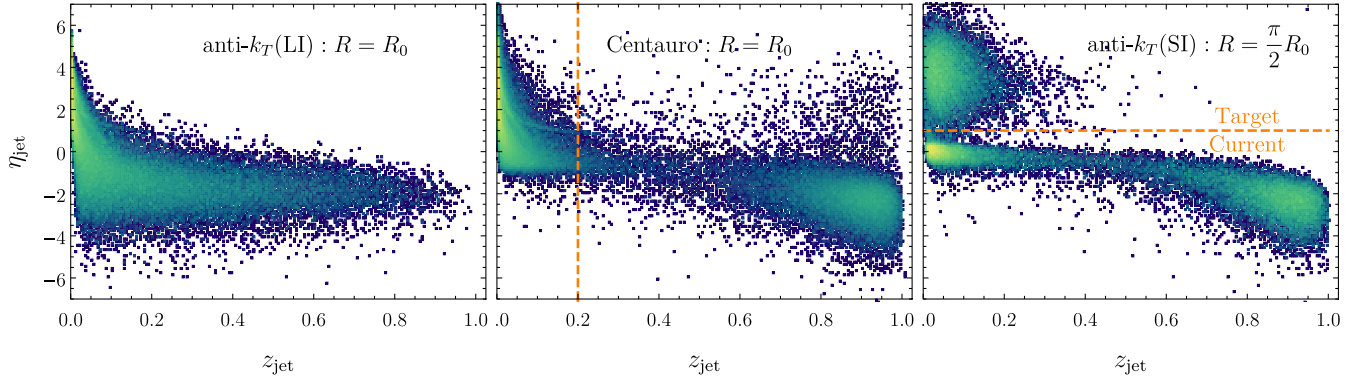


FIG. 3. The distribution of jets in the Breit frame in terms of their pseudorapidity  $\eta_{\text{jet}}$  and momentum fraction  $z_{\text{jet}}$ . The left, center, and right panels correspond to jets identified with the anti- $k_T$ (LI), Centauro, and anti- $k_T$ (SI) algorithms, respectively. The dashed lines indicate the separation of jets in the current and target fragmentation region.

jets [68]. In contrast, the anti- $k_T$ (SI) and Centauro algorithms cluster all of these particles into a single jet with  $z_{\text{jet}} \sim 1$ . The Centauro algorithm mimics the features of the anti- $k_T$ (SI) in the current direction and the anti- $k_T$ (LI) in the target direction.

Furthermore, with the use of Centauro and anti- $k_T$ (SI) jets it is also possible to suppress the target fragmentation with a cut on  $z_{\text{jet}} \sim 0.2-0.7$ , as shown in Fig. 3 (center panel). This allows for direct studies of quark TMD observables. For the anti- $k_T$ (SI) [69] algorithm, a cut on  $\eta_{\text{jet}} < 1$  separates current and target regions (right panel of Fig. 3). This reveals the full  $z_{\text{jet}}$  spectrum, which cannot be accessed with hadron measurements due to the contamination from the target fragmentation region [70,71]. For comparison we also show the result for the anti- $k_T$ (LI) algorithm in the left panel of Fig. 3.

Fig. 4 shows the  $z_{\text{jet}}$  and  $\eta_{\text{jet}}$  distributions of inclusive jets as described above. While in the current region ( $\eta_{\text{jet}} < 0$ ), the Centauro and anti- $k_T$ (SI) algorithms result in a peak at large  $z_{\text{jet}} \sim 1$ , the anti- $k_T$ (LI) algorithm separates that jet into several and yields a peak at small- $z_{\text{jet}}$ . The two peaks at  $z_{\text{jet}} \sim 1$  and  $z_{\text{jet}} \sim 0$  correspond to current and mid rapidity jets. The intermediate  $z_{\text{jet}}$  region is described in terms of jet functions and DGLAP evolution [72–75]. The large- $z_{\text{jet}}$  jets probe the threshold region [76], whereas the small- $z_{\text{jet}}$  region is related to soft fragmentation in  $e^+e^-$  collisions [77–81] and small- $x$  physics [82–84]. In Fig. 5, we show also the energy spectrum for jets that results from a perturbative calculation supplemented with a nonperturbative shape function. As detailed in the Appendix A, the spectrum results from the calculation of the factorization formula:

$$d\sigma = \sigma_0 H(Q^2, \mu^2) \int_{z_{\text{jet}}}^1 \frac{dz}{z} B_q(x_B, Q^2(1-z), \mu^2) \times D_q\left(\frac{z_{\text{jet}}}{z}, QR, \mu^2\right),$$

where the formula is differential in  $x_B$ ,  $Q^2$  and  $z_{\text{jet}}$ . The function  $B_q$  is the quark beam function of refs. [85,86] and  $D_q$  is the quark fragmentation function to a jet at the endpoint from Ref. [76]. The resummation formula at the

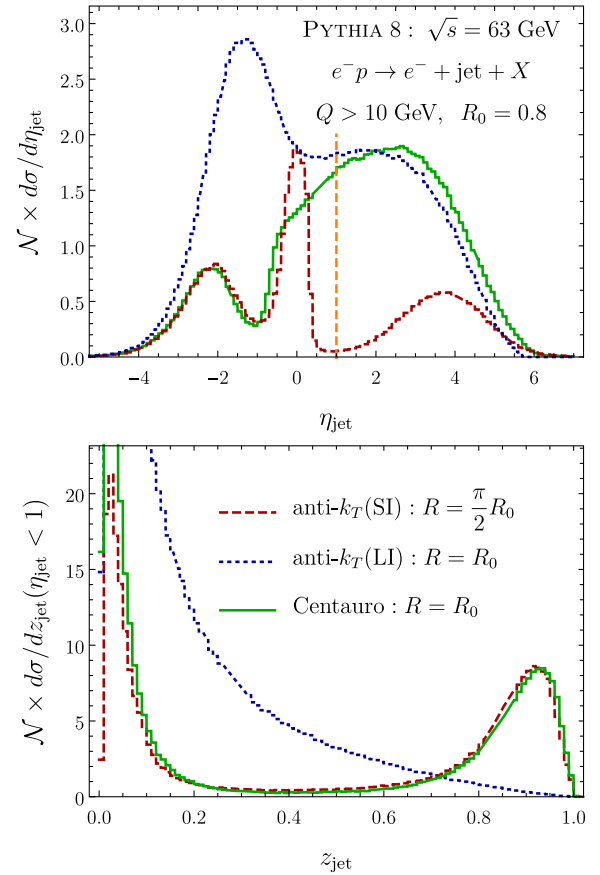


FIG. 4. Pseudorapidity (top panel) and momentum fraction  $z_{\text{jet}}$  (bottom panel) of jets clustered with anti- $k_T$ (LI), anti- $k_T$ (LI) and Centauro algorithms in the Breit frame. Here  $\mathcal{N}$  is an overall normalization constant chosen to improve readability and is the same for all curves in a graph.



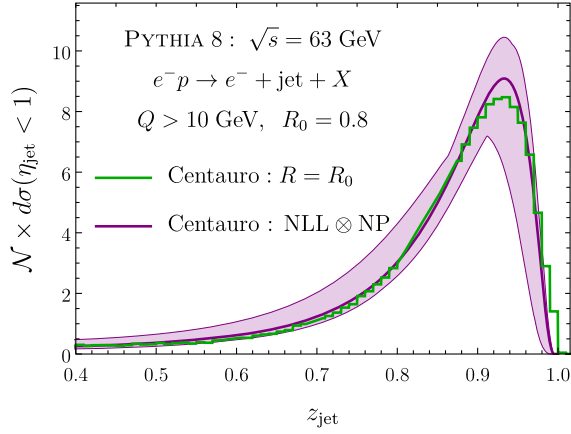


FIG. 5. The NLL prediction for the Centauro algorithm including threshold effects to NLL accuracy, full DGLAP to LL, as well as a nonperturbative shape function.

end-point can be derived by combining the methods developed in Refs. [74–76,85,87,88], valid to next-to-leading logarithm (NLL) including non-global effects of Refs. [89,90]. We also matched to the full leading order DGLAP evolution in the moderate  $z_{\text{jet}}$  region. Exploiting the sum-rule for the jets which demands conservation of the final state momentum, we can normalize to the leading order DIS cross section. The PDFs were obtained from Refs. [91]. The NLL uncertainty band is obtained from the envelope of varying each low scale of the renormalization group evolution by a factor of two, as well as all nonperturbative shape function scales and cutoffs for the Landau pole.

We propose a measurement of  $z_{\text{jet}}$  at HERA, which has not been done before, and the future EIC, where preliminary experimental projections are provided in Appendix B. The high- $z_{\text{jet}}$  region corresponds to jets with high- $p_T$  in the laboratory frame that can be measured with high precision and with an accuracy limited by the jet energy scale uncertainty, which reached 1% at HERA [44]. The measurement of the small- $z_{\text{jet}}$  region will be challenging because these jets correspond to jet  $p_T$  up to a few GeV in the laboratory frame [92], a region that can be limited by calorimeter noise and resolution. These issues could be bypassed by defining jets with charged particles only, which would require the inclusion of track-based jet functions on the theory side [93,94].

We also propose to use Centauro jets to study quark TMDs by measuring  $q_T = p_{\text{jet}}^\perp / z_{\text{jet}}$ . Figure 6 shows that the  $q_T$  spectrum for  $z_{\text{jet}} > 0.5$  peaks at  $q_T < Q_{\text{min}}/4$ , which is ideal for TMD phenomenology. With the polarized proton beams available at the EIC, this observable would provide clean access to the Siverts PDFs. Figure 6 also shows the  $q_T$  distribution for anti- $k_T$ (SI) jets for  $z_{\text{jet}} > 0.5$  and  $0 < z_{\text{jet}} < 0.5$ . Note the latter is only possible since we can suppress the target fragmentation by requiring  $\eta_{\text{jet}} < 1$ . While for  $z_{\text{jet}} > 0.5$  we find similar result as the Centauro

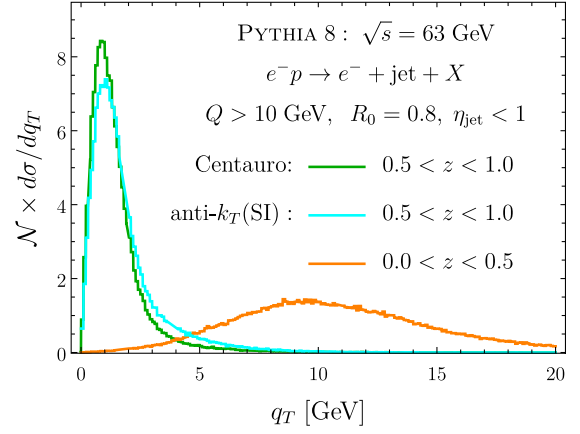


FIG. 6. The  $q_T = p_{\text{jet}}^\perp / z_{\text{jet}}$  spectrum for Centauro and anti- $k_T$ (SI) jets with  $\eta_{\text{jet}} < 1.0$ .

jets, for  $z_{\text{jet}} < 0.5$  the spectrum peaks at  $q_T \sim Q$ . Novel theoretical techniques are necessary to describe this kinematic  $q_T$  region of midrapidity jets.

In addition, the longitudinal invariance and ability to measure jets close to Born kinematics makes the Centauro algorithm an attractive option to: (i) extract the strong-coupling constant from the rates of  $n$ -jets [44]; (ii) enable “tag-and-probe” studies of nuclei [8]; (iii) identify the background for gluon helicity and Siverts PDF studies [9,27], (iv) study jet substructure and event shape observables in DIS Breit frame, (v) probe TMD evolution observables that can be related to lattice QCD. We leave those studies for future work.

## VI. CONCLUSIONS

We have proposed a new jet clustering approach tailored to the study of energetic jets with low transverse momentum in DIS that relies on spherically-invariant algorithms and a new longitudinally-invariant algorithm that is asymmetric in the current and target directions, which we call Centauro. The Centauro algorithm enables novel studies of transverse-momentum-dependent observables in the Breit frame. Furthermore, we find that spherically symmetric  $k_T$ -type algorithms yield clean access to the soft jet fragmentation region, which also reveals a new  $q_T$  regime where  $q_T \sim Q$ . The new jet algorithms introduced here are relevant for the studies of jet energy spectra, jet substructure, quark TMDs and spin physics, and cold-nuclear matter effects. All these studies will be central for the jet physics program of the future Electron-Ion Collider.

## ACKNOWLEDGMENTS

We thank Anselm Vossen, Renee Fatemi, and Fernando Torres-Acosta and Benjamin Nachman for insightful comments on our manuscript. M. A. and N. S. are supported through DOE Contract No. DE-AC05-06OR23177 under which JSA operates the Thomas Jefferson National

Accelerator Facility. 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. D. N. was supported by the U.S. DOE under Contract DE-AC52-06NA25396 at LANL and through the LANL/LDRD Program. F. R. was supported by LDRD funding from Berkeley Lab provided by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 as well as the National Science Foundation under Grant No. ACI-1550228.

## APPENDIX A: FACTORIZATION AND THRESHOLD RESUMMATION

In this appendix, we give the necessary technical details behind the theory prediction for the jet energy spectrum using the Centauro algorithm. We work to next-to-leading logarithmic order, where we resum all logarithms found in the cross section of order  $\alpha_s \times L$ , with  $L = \ln(1 - z_{\text{jet}})$  or  $\ln R$ ,  $R$  the jet radius and  $z_{\text{jet}}$  is the momentum fraction of the event carried by the jet. This includes logarithms of the jet radius outside the threshold limit where  $1 - z_{\text{jet}} \ll 1$ , which is formally a leading logarithmic resummation within strictly collinear factorization. The relevant details for the theory of semi-inclusive jet production and the jet function's threshold factorization can be found in Refs. [74–76,87], while the necessary details for the treat of the initial state for the threshold jet can be adapted from Refs. [85,88]. See also Refs. [95,96]. For the reader unfamiliar with the technology of resummation via soft-collinear effective field theory, we suggest Ref. [97]. The resummation of the jet spectrum is accomplished via the factorization formulas:

$$d\sigma = \tilde{\sigma}_0 \sum_{i,j} \int_{x_B}^1 \frac{dx}{x} \int_{z_{\text{jet}}}^1 \frac{dz}{z} C_{ij}(x, z, Q^2, \mu^2) f_{i/P}\left(\frac{x_B}{x}, \mu^2\right) \times D_j\left(\frac{z_{\text{jet}}}{z}, QR, \mu^2\right), \quad (\text{A1})$$

$$d\sigma = \sigma_0 H(Q^2, \mu^2) \int_{z_{\text{jet}}}^1 \frac{dz}{z} B_q(x_B, Q^2(1-z), \mu^2) \times D_q\left(\frac{z_{\text{jet}}}{z}, QR, \mu^2\right). \quad (\text{A2})$$

The sum is over the flavor indices of QCD, and  $C$  is the matching coefficient for collinear factorization in semi-inclusive DIS,  $H$  is the current matching squared for DIS processes,  $f$  is the parton distribution function, and  $D, B$  are the quark or anti-quark fragmentation function to a jet and the inclusive beam function. Equation (A1) is valid when  $1 - z_{\text{jet}} \sim 1$ , while Eq. (A2) controls the region  $1 - z_{\text{jet}} \ll 1$ . Finally,  $\tilde{\sigma}_0$  and  $\sigma_0$  are the Born-level cross sections for each factorization. The two formulas are related via the operator product expansion for the beam function:

$$B_i(x_B, Q^2(1-z), \mu^2) = \sum_j \int_{x_B}^1 \frac{dx}{x} \mathcal{I}_{ij}(x, Q^2(1-z), \mu^2) \times f_{j/P}\left(\frac{x_B}{x}, \mu^2\right) + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}^2}{Q^2(1-z)}\right), \quad (\text{A3})$$

with  $\Lambda_{\text{QCD}}$  the scale of confinement. The functions  $\mathcal{I}_{ij}$  are the matching coefficients of the beam functions  $B_i$  onto the parton distribution functions  $f_{j/P}$ , and can be calculated perturbatively. Moreover,  $D$  receives its own factorization in the endpoint region. Running each function to its natural scale (for detailed discussion, see Ref. [98]), evaluating them at their tree-level expressions, and factoring the PDFs, while using a nonperturbative shape function gives:

$$d\sigma = \sigma_0 \left( \sum_{q'} f_{q'/P}(x_B, \mu_F^2) \right) \left( \sum_j \int_{z_{\text{jet}}}^1 \frac{dz}{z} U_{qj}(z_{\text{jet}}/z; \mu_H^2, \mu_J^2) \times \left[ \mathcal{S}_j \otimes \frac{d\mathcal{R}_j}{dz} \right] (z, QR; \mu_H^2, \mu_J^2, \mu_F^2, \mu_{cs}^2) \right), \quad (\text{A4})$$

$$\mu_F^2 \sim Q^2(1 - z_{\text{jet}}), \quad \mu_H^2 \sim Q^2, \quad \mu_J^2 \sim Q^2 R^2, \quad \mu_{cs}^2 \sim Q^2 R^2 (1 - z_{\text{jet}})^2. \quad (\text{A5})$$

We factorize the beam function from the PDFs at the beam function scale, which we take as  $\mu_B^2 = \mu_F^2 \sim Q^2(1 - z_{\text{jet}})$ .  $\mathcal{R}_j$  is the NLL radiator function (described below for the case of quark jets), and  $U_{ij}$  is the *full* NLO DGALP evolution evolved from the scale  $\mu_H$  to the jet scale  $\mu_J$ . This is so that the formula is valid to leading log when  $z_{\text{jet}} \sim 0.5$ , away from the end point, but will have the two loop cusp at the endpoint, so that it is still NLL valid at the endpoint. For NLL, several simplifications occur: note that at the factorization scale for the PDFs, we only probe the quark and anti-quark content of the proton, so we restrict the PDF flavor sum accordingly. Moreover, the initial hard parton that generates the jets will also be a quark or anti-quark, and since these jets have identical jet functions, this restricts us to the singlet sector of the DGLAP evolution.  $\mathcal{S}_j$  is a nonperturbative shape function which we define as:

$$\left[ \mathcal{S}_j \otimes \frac{d\mathcal{R}_j}{dz} \right] (z, QR; \mu_H^2, \mu_J^2, \mu_F^2, \mu_{cs}^2) = \int_z^1 \frac{dz'}{z'} \mathcal{S}(z', d_j) \frac{d\mathcal{R}_j}{dz} \left( \frac{z}{z'}, QR; \mu_H^2, \mu_J^2, \mu_F^2, \mu_{cs}^2 \right), \quad (\text{A6})$$

$$\mathcal{S}(z, d) = \frac{(1-z) \exp(-\frac{1}{d}(1-z))}{d(d-(1+d)e^{-\frac{1}{d}})}, \quad d_j \sim \frac{\Lambda_j}{QR}. \quad (\text{A7})$$

$\Lambda_j \sim 400 \text{ MeV} \sim \Lambda_{\text{QCD}}$  should be of the order of the confinement scale, and we have normalized the integral of  $S$  to be one when integrated between zero and one. Note that we allow  $\Lambda_j$  to be different for quark or gluon jets, but in practice we take it to be the same.

The cumulative radiator function has the form:

$$\eta_{\text{tot}} = -\eta(\mu_H, \mu_F) + \eta(\mu_J, \mu_{\text{cs}}), \quad (\text{A8})$$

$$K_{\text{tot}} = -\frac{1}{2}K(\mu_J, \mu_{\text{cs}}) + K(\mu_H, \mu_F) - \frac{1}{2}\eta(\mu_J, \mu_{\text{cs}}) \ln\left(\frac{\mu_{\text{cs}}^2}{Q^2 R^2}\right) + \eta(\mu_H, \mu_F) \ln\left(\frac{\mu_F^2}{Q^2}\right) + \omega_B, \quad (\text{A9})$$

$$\mathcal{R}(z, QR; \mu_H^2, \mu_J^2, \mu_F^2, \mu_{\text{cs}}^2) = (1-z)^{\eta_{\text{tot}}} \frac{\exp(K_{\text{tot}} - \gamma_E \eta_{\text{tot}})}{\Gamma(1 + \eta_{\text{tot}})} \mathcal{R}^{\text{NGL}}(\mu_J^2, \mu_{\text{cs}}^2). \quad (\text{A10})$$

This defines the radiator as the cumulative distribution, which we then differentiate *after* scale setting.  $\gamma_E$  is the Euler-Gamma constant,  $\Gamma$  is the gamma function, and the functions  $K$  and  $\eta$  are integrals over the cusp-anomalous dimension  $\Gamma_{\text{cusp}}$  (given to two loops in Ref. [99]), while  $\omega_B$  is the integral over the noncusp components of the beam function's anomalous dimension:

$$K(\mu_f, \mu_i) = \int_{\mu_i^2}^{\mu_f^2} \frac{d\mu^2}{\mu^2} \left\{ \Gamma_{\text{cusp}}(\mu^2) \ln\left(\frac{\mu^2}{\mu_i^2}\right) \right\},$$

$$\eta(\mu_f, \mu_i) = \int_{\mu_i^2}^{\mu_f^2} \frac{d\mu^2}{\mu^2} \Gamma_{\text{cusp}}(\mu^2), \quad \omega_B = \int_{\mu_F^2}^{\mu_H^2} \frac{d\mu^2}{\mu^2} \gamma_B(\mu^2), \quad (\text{A11})$$

$$\Gamma_{\text{cusp}}(\mu^2) = C_F \frac{\alpha_s(\mu^2)}{\pi} \left( 1 + \frac{\alpha_s(\mu^2)}{4\pi} \left( C_A \left( \frac{67}{9} - \frac{\pi^2}{3} \right) - \frac{10}{9} n_f \right) \right) + \dots, \quad (\text{A12})$$

$$\gamma_B(\mu^2) = C_F \frac{3\alpha_s(\mu^2)}{2\pi} + \dots. \quad (\text{A13})$$

Where we give the anomalous dimensions explicitly for a quark jet. We also have the nonglobal contribution, which we take as given by the hemisphere distribution for the anti- $k_T$  type algorithms given here:

$$\mathcal{R}^{\text{NGL}}(\mu_f^2, \mu_i^2) = \exp\left(-C_F C_A \frac{\pi^2}{3} t^2 \frac{1 + 0.85 C_A t^2}{1 + (0.86 C_A t)^{1.33}}\right),$$

$$t = \frac{1}{\beta_0} \ln \frac{\alpha_s(\mu_i^2)}{\alpha_s(\mu_f^2)}, \quad (\text{A14})$$

where  $\beta_0 = \frac{11}{3} C_A - \frac{2}{3} n_f$ . We use the two-loop running of the strong coupling constant with  $\alpha_s(M_Z^2) = 0.1187$ , and regulate the Landau pole via the prescription:

$$\mu_F^2 \sim Q^2(1 - z_{\text{jet}}) + z_{\text{jet}}^2 m^2,$$

$$\mu_{\text{cs}}^2 \sim Q^2 R^2(1 - z_{\text{jet}})^2 + z_{\text{jet}}^2 m^2, \quad m = 0.5 \text{ GeV}. \quad (\text{A15})$$

The final cross section we plot in Fig. 5 is given by:

$$\frac{d\sigma}{dz_{\text{jet}}} = \int_{Q_0^2}^s dQ^2 \int_0^1 dx_B \Theta\left(\frac{1-x_B}{x_B} Q^2 - Q^2 R^2\right) \times \Theta\left(1 - \frac{Q^2}{x_B s}\right) \frac{d\sigma}{dx_B dz_{\text{jet}} dQ^2}. \quad (\text{A16})$$

Where we express the DIS born cross section  $\sigma_0$  in terms of  $s$ ,  $Q^2$  and  $x_B$ .  $Q_0$  is the minimum hard momentum transfer probed in the measurement. The  $\Theta$  function in the integral just states that the invariant mass of the jet must be less than the total invariant mass of the hadronic final state, which is given by  $(q + P_h)^2 = \frac{1-x_B}{x_B} Q^2$ , up to ‘‘target mass corrections.’’ We note that we have very little sensitivity to the constraint on  $QR$  toward the endpoint  $x_B \rightarrow 1$ .

## APPENDIX B: PROJECTIONS

As shown by the fast-detector simulations presented in Ref. [100], the current specifications for future EIC detectors yield a jet-energy resolution that ranges from 15% to 7% for jet energy from 15 GeV to 100 GeV, for anti- $k_T$  jets reconstructed in the laboratory frame based on particle-flow objects as implemented in DELPHES [101]. We estimate a similar performance for jets reconstructed in the Breit frame with the Centauro algorithm.

In Figure 7 we show a projection of statistical and systematic uncertainties of the jet-energy spectrum

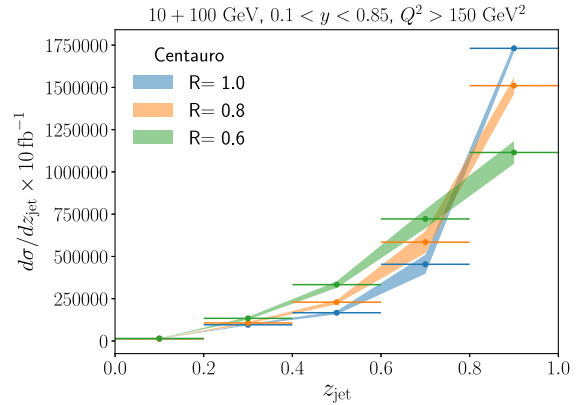


FIG. 7. EIC projection of the jet-energy spectrum for Centauro jets with different  $R$  parameters, assuming an integrated luminosity of  $10 \text{ fb}^{-1}$  and a jet-energy-scale uncertainty of  $\pm 2\%$ .

for Centauro jets at the future EIC. With an integrated luminosity of  $10 \text{ fb}^{-1}$ , the statistical uncertainty is expected to be negligible. Previous measurements of jets in DIS at HERA suggest that the dominant source of uncertainty will be associated with the jet-energy-scale (JES) calibration, which reached 1% for the HERA experiments [44]. We estimate the JES uncertainty will be worse for the jet-

energy spectrum measurement we propose, as the cross-calibration based on electron-jet balance in events close to Born kinematics will not be available (as for this case it represents the signal channel). We thus estimate a JES uncertainty of 2%, which we consider a conservative estimate. The resulting correlated uncertainty on the jet-energy spectrum is shown as a band in Fig. 7.

- 
- [1] A. Accardi *et al.*, *Eur. Phys. J. A* **52**, 268 (2016).
- [2] M. Arratia, Y. Furletova, T. Hobbs, F. Olness, and S. J. Sekula, *Phys. Rev. D* **103**, 074023 (2021).
- [3] I. Borsa, D. de Florian, and I. Pedron, *Phys. Rev. Lett.* **125**, 082001 (2020).
- [4] G. Peccini, L. Moriggi, and M. Machado, *Phys. Rev. C* **102**, 034903 (2020).
- [5] V. Guzey and M. Klasen, *J. High Energy Phys.* 05 (2020) 074.
- [6] V. Guzey and M. Klasen, *Phys. Rev. C* **102**, 065201 (2020).
- [7] Z.-B. Kang, K. Lee, and F. Zhao, *Phys. Lett. B* **809**, 135756 (2020).
- [8] M. Arratia, Y. Song, F. Ringer, and B. Jacak, *Phys. Rev. C* **101**, 065204 (2020).
- [9] B. Page, X. Chu, and E. Aschenauer, *Phys. Rev. D* **101**, 072003 (2020).
- [10] X. Li *et al.*, *EPJ Web Conf.* **235**, 04002 (2020).
- [11] D. Gutierrez-Reyes, I. Scimemi, W. J. Waalewijn, and L. Zoppi, *J. High Energy Phys.* 10 (2019) 031.
- [12] D. Gutierrez-Reyes, Y. Makris, V. Vaidya, I. Scimemi, and L. Zoppi, *J. High Energy Phys.* 08 (2019) 161.
- [13] D. Gutierrez-Reyes, I. Scimemi, W. J. Waalewijn, and L. Zoppi, *Phys. Rev. Lett.* **121**, 162001 (2018).
- [14] Y.-Y. Zhang, G.-Y. Qin, and X.-N. Wang, *Phys. Rev. D* **100**, 074031 (2019).
- [15] E.-C. Aschenauer, K. Lee, B. Page, and F. Ringer, *Phys. Rev. D* **101**, 054028 (2020).
- [16] Y. Hatta, N. Mueller, T. Ueda, and F. Yuan, *Phys. Lett. B* **802**, 135211 (2020).
- [17] H. Mäntysaari, N. Mueller, and B. Schenke, *Phys. Rev. D* **99**, 074004 (2019).
- [18] U. D'Alesio, F. Murgia, C. Pisano, and P. Tael, *Phys. Rev. D* **100**, 094016 (2019).
- [19] R. Kishore, A. Mukherjee, and S. Rajesh, *Phys. Rev. D* **101**, 054003 (2020).
- [20] D. Kang and T. Maji, *Proc. Sci.*, LC2019 (2019) 061 [arXiv:1912.10656].
- [21] K. Roy and R. Venugopalan, *Phys. Rev. D* **101**, 034028 (2020).
- [22] F. Salazar and B. Schenke, *Phys. Rev. D* **100**, 034007 (2019).
- [23] R. Boughezal, F. Petriello, and H. Xing, *Phys. Rev. D* **98**, 054031 (2018).
- [24] M. Klasen and K. Kovařík, *Phys. Rev. D* **97**, 114013 (2018).
- [25] A. Dumitru, V. Skokov, and T. Ullrich, *Phys. Rev. C* **99**, 015204 (2019).
- [26] X. Liu, F. Ringer, W. Vogelsang, and F. Yuan, *Phys. Rev. Lett.* **122**, 192003 (2019).
- [27] L. Zheng, E. Aschenauer, J. Lee, B.-W. Xiao, and Z.-B. Yin, *Phys. Rev. D* **98**, 034011 (2018).
- [28] M. D. Sievert and I. Vitev, *Phys. Rev. D* **98**, 094010 (2018).
- [29] M. Klasen, K. Kovarik, and J. Potthoff, *Phys. Rev. D* **95**, 094013 (2017).
- [30] P. Hinderer, M. Schlegel, and W. Vogelsang, *Phys. Rev. D* **96**, 014002 (2017).
- [31] X. Chu, E.-C. Aschenauer, J.-H. Lee, and L. Zheng, *Phys. Rev. D* **96**, 074035 (2017).
- [32] G. Abelof, R. Boughezal, X. Liu, and F. Petriello, *Phys. Lett. B* **763**, 52 (2016).
- [33] Y. Hatta, B.-W. Xiao, and F. Yuan, *Phys. Rev. Lett.* **116**, 202301 (2016).
- [34] A. Dumitru and V. Skokov, *Phys. Rev. D* **94**, 014030 (2016).
- [35] D. Boer, P. J. Mulders, C. Pisano, and J. Zhou, *J. High Energy Phys.* 08 (2016) 001.
- [36] A. Dumitru, T. Lappi, and V. Skokov, *Phys. Rev. Lett.* **115**, 252301 (2015).
- [37] P. Hinderer, M. Schlegel, and W. Vogelsang, *Phys. Rev. D* **92**, 014001 (2015); **93**, 119903(E) (2016).
- [38] T. Altinoluk, N. Armesto, G. Beuf, and A. H. Rezaeian, *Phys. Lett. B* **758**, 373 (2016).
- [39] D. Kang, C. Lee, and I. W. Stewart, *Phys. Rev. D* **88**, 054004 (2013).
- [40] C. Pisano, D. Boer, S. J. Brodsky, M. G. A. Buffing, and P. J. Mulders, *J. High Energy Phys.* 10 (2013) 024.
- [41] Z.-B. Kang, S. Mantry, and J.-W. Qiu, *Phys. Rev. D* **86**, 114011 (2012).
- [42] Z.-B. Kang, A. Metz, J.-W. Qiu, and J. Zhou, *Phys. Rev. D* **84**, 034046 (2011).
- [43] D. Boer, S. J. Brodsky, P. J. Mulders, and C. Pisano, *Phys. Rev. Lett.* **106**, 132001 (2011).
- [44] P. Newman and M. Wing, *Rev. Mod. Phys.* **86**, 1037 (2014).
- [45] A. J. Larkoski, I. Moulton, and B. Nachman, *Phys. Rep.* **841**, 1 (2020).
- [46] Y. Makris, D. Neill, and V. Vaidya, *J. High Energy Phys.* 07 (2018) 167.
- [47] M. A. Ebert, I. W. Stewart, and Y. Zhao, *Phys. Rev. D* **99**, 034505 (2019).



- [48] P. Shanahan, M. Wagman, and Y. Zhao, *Phys. Rev. D* **102**, 014511 (2020).
- [49] B. Webber, *J. Phys. G* **19**, 1567 (1993).
- [50] S. D. Ellis and D. E. Soper, *Phys. Rev. D* **48**, 3160 (1993).
- [51] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *J. High Energy Phys.* **08** (1997) 001.
- [52] M. Wobisch and T. Wengler, in *Hadronization Corrections to Jet Cross-Sections in Deep Inelastic Scattering* (1998), pp. 270–279, <https://inspirehep.net/literature/484872>.
- [53] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [54] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012).
- [55] S. Catani, Y. L. Dokshitzer, and B. Webber, *Phys. Lett. B* **285**, 291 (1992).
- [56] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. Webber, *Phys. Lett. B* **269**, 432 (1991).
- [57] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, *Nucl. Phys.* **B406**, 187 (1993).
- [58] M. Boronat, J. Fuster, I. Garcia, E. Ros, and M. Vos, *Phys. Lett. B* **750**, 95 (2015).
- [59] Different choices can be optimized to match jets found by other algorithms, for example we found that jets clustered in the target region are better matched to the jets clustered with the anti- $k_T$  (L) algorithm.
- [60] For the analysis in this section we started with the choice  $p = 0$  in Eq. (6). Analogous algorithms can be obtained for the choices  $p = +1$  and  $p = -1$ .
- [61] M. Dasgupta and B. R. Webber, *J. High Energy Phys.* **10** (1998) 001.
- [62] V. Antonelli, M. Dasgupta, and G. P. Salam, *J. High Energy Phys.* **02** (2000) 001.
- [63] M. Dasgupta and G. P. Salam, *Eur. Phys. J. C* **24**, 213 (2002).
- [64] T. Sjostrand, S. Mrenna, and P. Z. Skands, *Comput. Phys. Commun.* **178**, 852 (2008).
- [65] In addition, we verified that our conclusions are the same when using the DIRE shower of Ref. [66].
- [66] S. Höche and S. Prestel, *Eur. Phys. J. C* **75**, 461 (2015).
- [67] The Centauro algorithm is available as part of the official release of the FastJet Contrib package, since fjcontrib-1.045, see <https://fastjet.hepforge.org/contrib/>.
- [68] Note that this cannot be corrected for by choosing a larger jet radius due to the exponential increase of the rapidity in the target and current region.
- [69] Note that we include a factor of  $\pi/2$  for the jet radius of the anti- $k_T$ (SI) algorithm to account for the difference in the denominator of the clustering metric in Eq. (6) compared to Eq. (5).
- [70] M. Boglione, J. Collins, L. Gamberg, J. O. Gonzalez-Hernandez, T. C. Rogers, and N. Sato, *Phys. Lett. B* **766**, 245 (2017).
- [71] E. C. Aschenauer, I. Borsa, R. Sassot, and C. Van Hulse, *Phys. Rev. D* **99**, 094004 (2019).
- [72] M. Dasgupta, F. Dreyer, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2015) 039.
- [73] T. Kaufmann, A. Mukherjee, and W. Vogelsang, *Phys. Rev. D* **92**, 054015 (2015); **101**, 079901(E) (2020).
- [74] Z.-B. Kang, F. Ringer, and I. Vitev, *J. High Energy Phys.* **10** (2016) 125.
- [75] L. Dai, C. Kim, and A. K. Leibovich, *Phys. Rev. D* **94**, 114023 (2016).
- [76] L. Dai, C. Kim, and A. K. Leibovich, *Phys. Rev. D* **95**, 074003 (2017).
- [77] A. H. Mueller, *Phys. Lett.* **104B**, 161 (1981).
- [78] A. Bassetto, M. Ciafaloni, G. Marchesini, and A. H. Mueller, *Nucl. Phys.* **B207**, 189 (1982).
- [79] Y. Dokshitzer, G. Marchesini, and G. Salam, *Phys. Lett. B* **634**, 504 (2006).
- [80] C.-H. Kom, A. Vogt, and K. Yeats, *J. High Energy Phys.* **10** (2012) 033.
- [81] D. P. Anderle, T. Kaufmann, M. Stratmann, and F. Ringer, *Phys. Rev. D* **95**, 054003 (2017).
- [82] B. Basso and G. Korchemsky, *Nucl. Phys.* **B775**, 1 (2007).
- [83] Y. Hatta, *J. High Energy Phys.* **11** (2008) 057.
- [84] D. Neill and F. Ringer, *J. High Energy Phys.* **06** (2020) 086.
- [85] S. Fleming, A. K. Leibovich, and T. Mehen, *Phys. Rev. D* **74**, 114004 (2006).
- [86] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, *Phys. Rev. D* **81**, 094035 (2010).
- [87] X. Liu, S.-O. Moch, and F. Ringer, *Phys. Rev. Lett.* **119**, 212001 (2017).
- [88] G. Lusterians, J. K. Michel, and F. J. Tackmann, [arXiv: 1908.00985](https://arxiv.org/abs/1908.00985).
- [89] M. Dasgupta and G. Salam, *Phys. Lett. B* **512**, 323 (2001).
- [90] A. Banfi, G. Marchesini, and G. Smye, *J. High Energy Phys.* **08** (2002) 006.
- [91] D. Clark, E. Godat, and F. Olness, *Comput. Phys. Commun.* **216**, 126 (2017).
- [92] This depends on  $Q$ , and a high  $Q$  is preferred to reach the lowest  $z_{\text{jet}}$ .
- [93] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, *Phys. Rev. D* **88**, 034030 (2013).
- [94] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, *Phys. Rev. Lett.* **111**, 102002 (2013).
- [95] M. Cacciari and S. Catani, *Nucl. Phys.* **B617**, 253 (2001).
- [96] D. P. Anderle, F. Ringer, and W. Vogelsang, *Phys. Rev. D* **87**, 034014 (2013).
- [97] T. Becher, A. Broggio, and A. Ferroglia, *Lect. Notes Phys.* **896** (2015).
- [98] L. G. Almeida, S. D. Ellis, C. Lee, G. Sterman, I. Sung, and J. R. Walsh, *J. High Energy Phys.* **04** (2014) 174.
- [99] G. Korchemsky and A. Radyushkin, *Nucl. Phys.* **B283**, 342 (1987).
- [100] M. Arratia, Z.-B. Kang, A. Prokudin, and F. Ringer, *Phys. Rev. D* **102**, 074015 (2020).
- [101] J. de Favereau *et al.* (DELPHES 3 Collaboration), *J. High Energy Phys.* **02** (2014) 057.