

Dissecting Soft Radiation with Factorization

Iain W. Stewart,¹ Frank J. Tackmann,² and Wouter J. Waalewijn^{3,4}

¹*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

²*Theory Group, Deutsches Elektronen-Synchrotron (DESY), D-22607 Hamburg, Germany*

³*Nikhef, Theory Group, Science Park 105, 1098 XG, Amsterdam, The Netherlands*

⁴*ITFA, University of Amsterdam, Science Park 904, 1018 XE, Amsterdam, The Netherlands*

An essential part of high-energy hadronic collisions is the soft hadronic activity that underlies the primary hard interaction. It can receive contributions from soft radiation from the primary hard partons, secondary multiple parton interactions (MPI), and factorization violating effects. The invariant mass spectrum of the leading jet in Z +jet and H +jet events is directly sensitive to these effects. We use a QCD factorization theorem to predict the dependence on the jet radius R , jet p_T , jet rapidity, and partonic process for both the perturbative and nonperturbative components of primary soft radiation. The nonperturbative contributions involve only odd powers of R , and the linear R term is universal for quark and gluon jets. The hadronization model in PYTHIA8 agrees well with these properties. The perturbative soft initial state radiation (ISR) has a contribution that depends on the jet area in the same way as the underlying event. This degeneracy is broken by the jet p_T dependence. The size of this soft ISR contribution is proportional to the color state of the initial partons, yielding the same positive contribution for $gg \rightarrow Hg$ and $gq \rightarrow Zq$, but a negative interference contribution for $q\bar{q} \rightarrow Zg$. Hence, measuring these dependencies allows one to separate hadronization, soft ISR, and MPI contributions in the data.

Soft hadronic activity plays a role in practically all but the most inclusive measurements at the LHC. It is often an important yet hard-to-quantify source of uncertainty, so improving its theoretical understanding is vital. One can consider four conceptually different sources for the effects that are experimentally associated with soft hadronic activity and the underlying event (UE):

1. Perturbative soft radiation from the primary hard partons within factorization
2. Nonperturbative soft effects within factorization associated with hadronization
3. Multiple parton interactions (MPI) at lower scales in the same proton-proton collision
4. Factorization breaking contributions

For any given observable, the question is how much of each of these sources is required to describe the data. For example, it is known that including higher-order perturbative corrections (source 1) in parton-shower Monte Carlo can give a nontrivial contribution to traditional UE measurements [1, 2].

Traditionally, the UE activity is measured in regions of phase space away from hard jets [2–12]. These results are used to tune the MPI models that are used to describe the UE in Monte Carlo programs [13–18]. These models are then extrapolated into the jet region, where they are used to describe various jet observables, including the jet mass spectrum in dijet and Drell-Yan events [19, 20], which is an important benchmark jet observable at the LHC.

In this paper, we directly consider the jet region and give a field-theoretic description of primary soft effects

(sources 1 and 2). Using the jet mass spectrum and its first moment, we show how to cleanly distinguish sources 1, 2, and 3 via their characteristic dependence on the jet radius R , jet momentum p_T^J , and participating partons. We will not consider factorization breaking effects here (see e.g. Ref. [21]).

We consider the jet mass spectrum in exclusive $pp \rightarrow Z$ +1-jet and $pp \rightarrow H$ +1-jet events. The factorization formula for $m_J \ll p_T^J$ that includes sources 1 and 2 is given by [22–24]

$$\frac{d\sigma}{dm_J^2 d\Phi_2} = \sum_{\kappa, a, b} H_\kappa(\Phi_2) \int dk_S dk_B (\mathcal{I}_{\kappa_a a} \mathcal{I}_{\kappa_b b} \otimes f_a f_b)(k_B) \times J_{\kappa_J}(m_J^2 - 2p_T^J k_S) S_\kappa(k_S, p^{\text{cut}} - k_B, y_J, R). \quad (1)$$

Here, $\Phi_2 = \{p_T^J, y_J, Y\}$, where Y is the rapidity of the Z/H +jet system, and κ denotes the partonic channel. The $H_\kappa(\Phi_2)$ contains the perturbative matrix elements for the hard process, the $\mathcal{I}_{\kappa_a a}$ describe perturbative collinear initial-state radiation from the incoming primary partons, and the f_a are the parton distribution functions. For the normalized jet mass spectrum, the dependence on p^{cut} , which vetoes additional jets, largely drops out. As a result, the shape of the jet mass spectrum is determined by the jet function J_{κ_J} , describing energetic final-state radiation from the outgoing primary parton, and by the soft function S_κ . See also Refs. [25, 26].

The soft function S_κ describes the primary initial and final state soft radiation. It depends on the jet through y_J and R but not p_T^J , and can be factorized as [27–29]

$$S_\kappa(k_S, k_B, y_J, R) = \int dk S_\kappa^{\text{pert}}(k_S - k, k_B, y_J, R) \times F_\kappa(k, y_J, R) [1 + \mathcal{O}(\Lambda_{\text{QCD}}/k_B)], \quad (2)$$

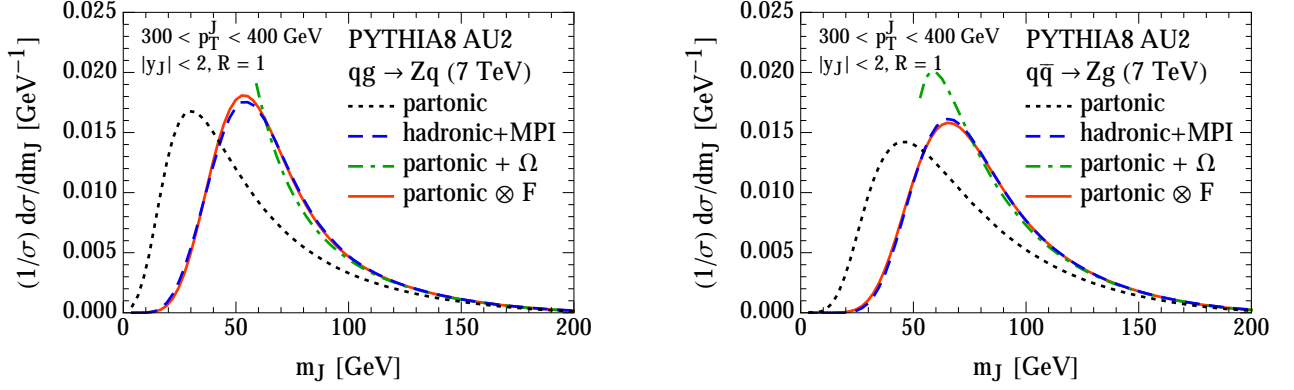


FIG. 1. For the jet mass spectrum in PYTHIA8, the change from partonic to hadronization+MPI is described by a simple shift in the tail, and a simple convolution everywhere, for both quark jets (left panel) and gluon jets (right panel).

where S_κ^{pert} contains the perturbative soft contributions. F_κ is a nonperturbative shape function which is normalized and encodes the smearing effect that the hadronization has on the distribution of the soft momentum k_S . For $k_S \sim \Lambda_{\text{QCD}}$, the full $F_\kappa(k)$ is required and shifts the peak region of the jet mass spectrum to higher jet masses.

In the perturbative tail of the jet mass spectrum, where $k_S \gg \Lambda_{\text{QCD}}$, S_κ can be expanded,

$$S_\kappa(k_S, y_J, R) = S_\kappa^{\text{pert}}[k_S - \Omega_\kappa(R), y_J, R] + \mathcal{O}(\Lambda_{\text{QCD}}^2/k_S^3, \alpha_s \Lambda_{\text{QCD}}/k_S^2), \quad (3)$$

where $\Omega_\kappa(R) = \int dk k F_\kappa(k) \sim \Lambda_{\text{QCD}}$ is a nonperturbative parameter. In this region, factorization predicts a shift in the jet mass spectrum which is described by the single parameter $\Omega_\kappa(R)$. Below, we use the field-theoretic definition of Ω_κ to quantify its R -dependence and prove that it is independent of y_J . The above treatment of hadronization provides an excellent description of these effects in both B -meson decays and e^+e^- event shapes [30, 31], and will also do so in the present case.

Factorization also underlies the Monte Carlo description of the primary collision, where H corresponds to the hard matrix element, while \mathcal{I} , J , and S are described by the initial-state and final-state parton showers, and F corresponds to the hadronization models. The standard parton shower paradigm does not completely capture interference effects between wide-angle soft emissions from different primary partons that appear at $\mathcal{O}(\alpha_s)$ in S_κ . Monte Carlos include MPI (source 3), which are not included in Eq. (1). See Ref. [32] for a recent discussion. For our numerical studies, we consider both PYTHIA8 [33, 34] with the ATLAS underlying event tune AU2-MSTW2008LO [16] and HERWIG++ 2.7 [35, 36] with its default underlying event tune UE-EE-5-MRST [18]. Both give a reasonable description of the CMS jet mass spectrum in Z +jet events [20]. We also compare to the PYTHIA8 default tune 4C.

We consider exclusive Z/H +jet events at $E_{\text{cm}} = 7$ TeV in both quark and gluon channels, with the leading jet

within a certain range of p_T^J and y_J , and we veto additional jets with $p_T^J > 50$ GeV. The jets are defined using anti- k_T [37, 38]. In Fig. 1, we show the jet mass spectrum for quark and gluon jets with $R = 1$ after parton showering (black dotted) and including both hadronization and MPI (blue dashed). Using Eq. (3) with Eq. (1) predicts that for $m_J^2 \gg \Lambda_{\text{QCD}} p_T^J$ the nonperturbative corrections shift the tail of the jet mass spectrum by

$$m_J^2 = (m_J^2)^{\text{pert}} + 2p_T^J \Omega_\kappa(R). \quad (4)$$

We can regard the partonic result from PYTHIA8 as the baseline purely perturbative result. Shifting it in this way with $\Omega = 2.4$ GeV for $qg \rightarrow Zq$ and $\Omega = 2.7$ GeV for $q\bar{q} \rightarrow Zg$ yields the green dot-dashed curves in Fig. 1. We see that the effect of both hadronization and MPI in the tail is well captured by a shift, so $\Omega_\kappa \rightarrow \Omega_\kappa^{\text{had}} + \Upsilon_\kappa^{\text{MPI}}$. For hadronization, combining Eq. (2) with Eq. (1) predicts a convolution with a nonperturbative function,

$$\frac{d\sigma_\kappa}{dm_J^2} = \int dk \frac{d\sigma_\kappa^{\text{partonic}}}{dm_J^2} (m_J^2 - 2p_T^J k) F_\kappa(k). \quad (5)$$

The result of this convolution, shown by the red solid curves in Fig. 1, yields excellent agreement with the hadronization+MPI result over the full range of the jet mass spectrum!¹ Both hadronization and MPI populate the jet region with a smooth background of soft particles, which can explain why the MPI effect is reproduced alongside the hadronization by a convolution of the form Eq. (5). This apparent degeneracy motivates us to determine the calculable behavior of the jet mass spectrum

¹ Here, $F_\kappa(k) = (4k/\Omega_\kappa^2) e^{-2k/\Omega_\kappa}$; the simplest ansatz that satisfies the required properties of being normalized, vanishing at $k = 0$, falling off exponentially for $k \rightarrow \infty$, and having a first moment Ω_κ with the above values. Fixing the value of Ω_κ from the tail, we find similar levels of agreement across all values of p_T^J , y_J , R , for all partonic channels, as well as for different jet veto cuts (including no jet veto).

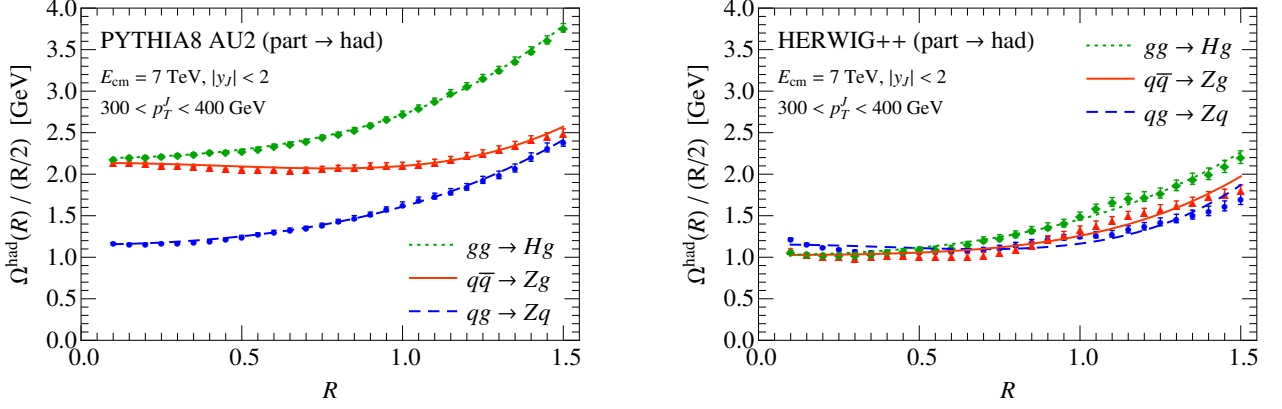


FIG. 2. The R dependence of $\Omega_\kappa^{\text{had}}(R)$ extracted from the first jet mass moment in PYTHIA8 (left panel) and HERWIG++ (right panel). It follows the form in Eq. (10) predicted from factorization. The small- R behavior only depends on whether the jet is initiated by a quark (blue dashed) or gluon (orange solid and green dotted).

due to primary perturbative and nonperturbative soft radiation within factorization, study its dependence on p_T^J , y_J , and R , and compare these results to Monte Carlo contributions for soft ISR, hadronization, and MPI.

We consider the first moment in m_J^2 ,

$$M_1 = \frac{1}{\sigma} \int dm_J^2 m_J^2 \frac{d\sigma}{dm_J^2}, \quad (6)$$

which tracks the shift observed in Fig. 1. Taking the first moment of Eq. (1) combined with Eqs. (2) and (3), we can compute the dependence of primary soft radiation on p_T^J , y_J , R , and partonic channel,

$$M_1 = M_{1\kappa}^{\text{pert}}(p_T^J, y_J, R) + 2p_T^J \Omega_\kappa(R). \quad (7)$$

Here, $M_{1\kappa}^{\text{pert}}(p_T^J, y_J, R)$ contains all perturbative contributions, while $\Omega_\kappa(R)$ encodes the shift due to nonperturbative effects.

To describe the results for M_1 from PYTHIA8 and HERWIG++, including their partonic, hadronization, and MPI contributions, we write

$$M_1 = M_{1\kappa}^{\text{partonic}}(p_T^J, y_J, R) + 2p_T^J \Omega_\kappa^{\text{had}}(R) + 2p_T^J \left[\Upsilon^{\text{MPI}}(y_J, R) + \Omega_\kappa^{\text{MPI}}(y_J, R) \right]. \quad (8)$$

Here, $M_{1\kappa}^{\text{partonic}}$ is the partonic contribution, $\Omega_\kappa^{\text{had}}$ is given by partonic \rightarrow hadronic, Υ^{MPI} is defined by partonic \rightarrow partonic+MPI, and $\Omega_\kappa^{\text{MPI}}$ is the small remainder which ensures these contributions add up to the full partonic \rightarrow hadronic+MPI. Equation (8) encodes the dependence on p_T^J , y_J , κ observed in PYTHIA8 and HERWIG++ and the fact that their hadronization and MPI are each individually described by shifts to M_1 .

For $pp \rightarrow H/Z + \text{jet}$, the nonperturbative parameter $\Omega_\kappa(R)$ in Eq. (7) is given by the vacuum matrix element of lightlike soft Wilson lines Y_a , Y_b , and $Y_J \equiv Y_J(y_J, \phi_J)$ along the beam and jet directions,

$$\Omega_\kappa(R) = \int_0^1 dr \int_{-\infty}^{\infty} dy \int_0^{2\pi} d\phi f(r, y - y_J, \phi - \phi_J, R)$$

$$\times \langle 0 | \bar{T}[Y_J^\dagger Y_b^\dagger Y_a^\dagger] \hat{\mathcal{E}}_T(r, y, \phi) T[Y_a Y_b Y_J] | 0 \rangle. \quad (9)$$

Here, the rapidity y , azimuthal angle ϕ , and transverse velocity $r = p_T/m_T$ are measured with respect to the beam axis. The color representation of the Wilson lines depends on the partonic channel, which induces the κ dependence of Ω_κ . The jet mass measurement function is $f(r, y, \phi, R) = (\cosh y - r \cos \phi) \theta[b(y, \phi, r) < R^2]$ where $b(y, \phi, r)$ specifies the jet boundary. The matrix element involves the energy flow operator [39, 40] $\hat{\mathcal{E}}_T(r, y, \phi) | X \rangle = \sum_{i \in X} m_{Ti} \delta(r - r_i) \delta(y - y_i) \delta(\phi - \phi_i) | X \rangle$. From Eq. (9) it follows immediately that $\Omega_\kappa(R)$ is independent of p_T^J . Using invariance under boosts and rotations, we can prove that it is also independent of y_J and ϕ_J [41]. This agrees with the observed behaviour of $\Omega_\kappa^{\text{had}}(R)$ in PYTHIA8 and HERWIG++ [41].

Expanding Eq. (9) for small R , we find [41, 42]

$$\Omega_\kappa(R) = \frac{R}{2} \Omega_\kappa^{(1)} + \frac{R^3}{8} \Omega_\kappa^{(3)} + \frac{R^5}{32} \Omega_\kappa^{(5)} + \mathcal{O}\left[\left(\frac{R}{2}\right)^7\right], \quad (10)$$

where the $\Omega_\kappa^{(i)}$ are R independent and only odd powers of R occur. This R -scaling of our nonperturbative operator for jet mass agrees with that found in Ref. [43] from a QCD hadronization model. Our operator definition implies a universality for the nonperturbative parameter in Eq. (10) linear in R . Since the $Y_a Y_b$ Wilson lines fuse for $R \rightarrow 0$, it is given by [41]

$$\Omega_\kappa^{(1)} = \int_0^1 dr' \langle 0 | \bar{T}[Y_J^\dagger Y_J^\dagger] \hat{\mathcal{E}}_\perp(r') T[Y_J Y_J] | 0 \rangle, \quad (11)$$

which only depends on whether the jet is a quark or gluon jet. For quarks, we can compare this to thrust in DIS [45] where precisely this parameter $\Omega_q^{(1)}$ appears [44].

To compare the hadronization in PYTHIA8 and HERWIG++ with these predictions, we show $\Omega_\kappa^{\text{had}}(R)/(R/2)$ for different channels in Fig. 2. For $R \ll 1$, $\Omega_\kappa^{\text{had}}(R)$ is linear in R and has the same slope for the two channels involving gluon jets (as predicted). For PYTHIA8,

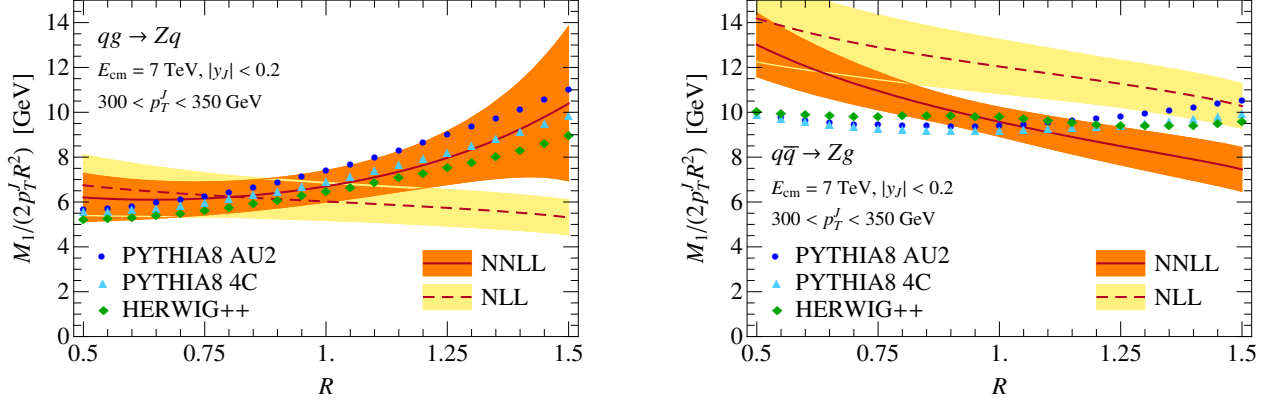


FIG. 3. R dependence of the perturbative jet mass moment $M_{1\kappa}^{\text{pert}}$ at NLL and NNLL and the partonic jet mass moment $M_{1\kappa}^{\text{partonic}}$ in PYTHIA8 (tune AU2 and 4C) and HERWIG++ for $qg \rightarrow Zq$ (left panel) and $q\bar{q} \rightarrow Zg$ (right panel). The soft ISR contribution $\sim R^4$ is well modelled by the Monte Carlos for $qg \rightarrow Zq$, but not for the destructive interference in $q\bar{q} \rightarrow Zg$.

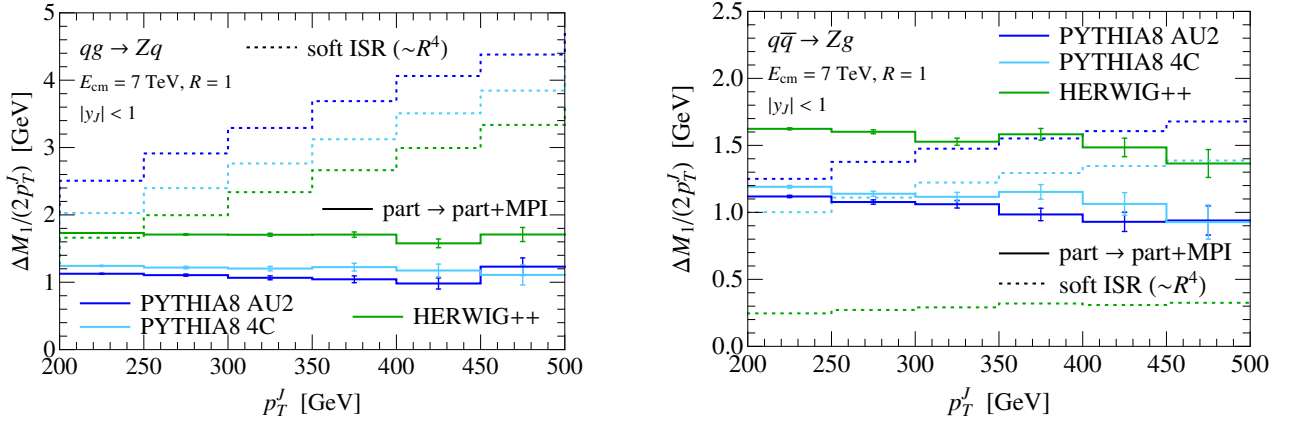


FIG. 4. p_T^J dependence of the $\sim R^4$ contributions to the jet mass moment in PYTHIA8 and HERWIG++ from MPI, Υ^{MPI} , and soft ISR [extracted using Eq. (14)] for $qg \rightarrow Zq$ (left panel) and $q\bar{q} \rightarrow Zg$ (right panel). These contributions are traded for each other in the different Monte Carlos and tunes, but can be separated by their p_T^J and channel dependence.

all channels differ for large R and can be fit to the form in Eq. (10). For the quark jet we extract $\Omega_q^{(1)} = 1.2 \text{ GeV}$ and the gluon jets $\Omega_g^{(1)} = 2.2 \text{ GeV}$. For $qg \rightarrow Zq$ and $gg \rightarrow Hg$ the R dependence is strong enough that an additional R^2 contribution is disfavored in the fit. For HERWIG++, the dependence on higher powers of R is much weaker, and surprisingly $\Omega_g^{(1)} \approx \Omega_q^{(1)}$. The full set of coefficients are given in [41].

We now turn to the perturbative moment $M_{1\kappa}^{\text{pert}}$. Dimensional analysis and the kinematical bound $m_J \lesssim p_T^J R$ imply that $M_{1\kappa}^{\text{pert}}$ scales like $(p_T^J R)^2$. Resummation modifies the leading R dependence to $R^{2-\gamma_\kappa}$, where $\gamma_\kappa \sim \alpha_s > 0$. The soft function contains a contribution due to interference between ISR from the two beams [41],

$$S_\kappa^{\text{pert}}(k_S) \supset \frac{\alpha_s C_\kappa}{\pi} R^2 \frac{1}{\mu} \left(\frac{\mu}{k_S} \right)_+, \quad (12)$$

which contributes to $M_{1\kappa}^{\text{pert}}$ as $(p_T^J)^2 R^4$ with color factor

$$C_{qg \rightarrow q} = C_{gg \rightarrow g} = \frac{C_A}{2} = \frac{3}{2},$$

$$C_{q\bar{q} \rightarrow g} = C_F - \frac{C_A}{2} = -\frac{1}{6}. \quad (13)$$

In Fig. 3, we compare our NLL and NNLL predictions [24] for $M_{1\kappa}^{\text{pert}}$ to the corresponding $M_{1\kappa}^{\text{partonic}}$ from PYTHIA8 and HERWIG++ as a function of R , dividing by the leading R^2 dependence. The effect of γ_κ is seen in the slight negative slope at NLL, and is also present in the Monte Carlos. The R^4 contribution only enters at NNLL and is seen in the rise at large R for $qg \rightarrow Zq$ (left panel). This soft ISR effect is partially modelled by soft emissions in the parton shower, which explains the similar R^4 contribution for $qg \rightarrow Zq$ in PYTHIA8 and HERWIG++. For $q\bar{q} \rightarrow Zg$ (right panel) Eqs. (12) and (13) predict the R^4 contribution from soft ISR to be negative, which we observe at NNLL. This negative interference effect is not captured by the Monte Carlos.

In Fig. 4, we show the p_T^J dependence of the R^4 component, c_4^κ , of the partonic moment, obtained by fitting

$$\frac{M_{1\kappa}^{\text{part}}}{2p_T^J R^2} = c_2^\kappa R^{-\gamma_\kappa} + c_4^\kappa R^2. \quad (14)$$

The MPI contribution to the moment, $\Upsilon^{\text{MPI}}/R^2 \sim R^2$, is shown as well. The apparent ambiguity between these different R^4 contributions is demonstrated by the notable differences between the various tunes for c_4^κ and Υ^{MPI} , while their sum $c_4^\kappa + \Upsilon^{\text{MPI}}$ is much closer. However, it is clearly resolved by the p_T^J dependence: $c_4^\kappa \sim p_T^J$ (as predicted), whereas Υ^{MPI} is independent of p_T^J . Although the destructive soft interference for $q\bar{q} \rightarrow Zg$ is not modelled by the Monte Carlos, the value of c_4^κ is significantly smaller than in $qg \rightarrow Zq$, in particular for HERWIG++. The channel dependence provides an additional handle to separate soft ISR from MPI: c_4^κ depends on the color channel as in Eq. (13), whereas Υ^{MPI} is channel independent. As shown in Ref. [41], the y_J dependence of soft ISR is quite different between HERWIG++ and PYTHIA8.

To conclude, we have used QCD factorization to predict the properties of the perturbative and nonperturbative components of primary soft radiation for jet mass in $pp \rightarrow H/Z + \text{jet}$. We have shown that the nonperturbative soft effects involve odd powers of R and are universal for quark and gluon jets for $R \ll 1$. Hadronization models in Monte Carlos agree with these predictions. The perturbative soft radiation has a contribution that scales like R^4 , just like the contribution from MPI. These components depend differently on p_T^J and on the partonic process. Hence, separately measuring quark and gluon channels in Drell-Yan and in different bins of p_T^J provides the possibility to clearly distinguish between MPI and primary soft radiation.

We thank Jesse Thaler and Simon Plätzer for helpful conversations. This work was supported in part by the Office of Nuclear Physics of the U.S. Department of Energy under Grant No. DE-FG02-94ER40818, the DFG Emmy-Noether Grant No. TA 867/1-1, and the Marie Curie Fellowship PIIF-GA-2012-328913. We thank the Erwin Schrödinger Institute for hospitality while portions of this work were completed.

[1] M. Cacciari, G. P. Salam, and S. Sapeta, JHEP, **1004**, 065 (2010), arXiv:0912.4926 [hep-ph].
[2] S. Chatrchyan *et al.* (CMS Collaboration), Eur. Phys. J. C, **72**, 2080 (2012), arXiv:1204.1411 [hep-ex].
[3] T. Affolder *et al.* (CDF Collaboration), Phys. Rev. D, **65**, 092002 (2002).
[4] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D, **70**, 072002 (2004), hep-ex/0404004.
[5] R. Field and R. C. Group (CDF Collaboration), (2005), hep-ph/0510198.
[6] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D, **82**, 034001 (2010), arXiv:1003.3146 [hep-ex].
[7] R. Field, (2010), arXiv:1010.3558 [hep-ph].
[8] G. Aad *et al.* (ATLAS Collaboration), (2010), ATL-PHYS-PUB-2010-014.
[9] G. Aad *et al.* (ATLAS Collaboration), Phys. Rev. D, **83**,

112001 (2011), arXiv:1012.0791 [hep-ex].
[10] G. Aad *et al.* (ATLAS Collaboration), (2011), ATL-PHYS-PUB-2011-009.
[11] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C, **71**, 1636 (2011), arXiv:1103.1816 [hep-ex].
[12] S. Chatrchyan *et al.* (CMS Collaboration), JHEP, **1109**, 109 (2011), arXiv:1107.0330 [hep-ex].
[13] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, and J. E. von Seggern, Eur. Phys. J. C, **65**, 331 (2010), arXiv:0907.2973 [hep-ph].
[14] P. Z. Skands, Phys. Rev. D, **82**, 074018 (2010), arXiv:1005.3457 [hep-ph].
[15] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, *et al.*, Phys. Rept., **504**, 145 (2011), arXiv:1101.2599 [hep-ph].
[16] G. Aad *et al.* (ATLAS Collaboration), (2012), ATL-PHYS-PUB-2012-003.
[17] S. Gieseke, C. Rohr, and A. Siodmok, Eur. Phys. J. C, **72**, 2225 (2012), arXiv:1206.0041 [hep-ph].
[18] M. H. Seymour and A. Siodmok, JHEP, **1310**, 113 (2013), arXiv:1307.5015 [hep-ph].
[19] G. Aad *et al.* (ATLAS), JHEP, **1205**, 128 (2012), arXiv:1203.4606 [hep-ex].
[20] S. Chatrchyan *et al.* (CMS Collaboration), JHEP, **1305**, 090 (2013), arXiv:1303.4811 [hep-ex].
[21] J. R. Forshaw, M. H. Seymour, and A. Siodmok, JHEP, **1211**, 066 (2012), arXiv:1206.6363 [hep-ph].
[22] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, Phys. Rev. Lett., **105**, 092002 (2010), arXiv:1004.2489 [hep-ph].
[23] T. T. Jouttenus, I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, Phys. Rev. D, **83**, 114030 (2011), arXiv:1102.4344 [hep-ph].
[24] T. T. Jouttenus, I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, Phys. Rev. D, **88**, 054031 (2013), arXiv:1302.0846 [hep-ph].
[25] M. Dasgupta, K. Khelifa-Kerfa, S. Marzani, and M. Spannowsky, JHEP, **1210**, 126 (2012), arXiv:1207.1640 [hep-ph].
[26] Y.-T. Chien, R. Kelley, M. D. Schwartz, and H. X. Zhu, Phys. Rev. D, **87**, 014010 (2013), arXiv:1208.0010 [hep-ph].
[27] G. P. Korchemsky and G. Sterman, Nucl. Phys. B, **555**, 335 (1999), hep-ph/9902341.
[28] A. H. Hoang and I. W. Stewart, Phys. Lett. B, **660**, 483 (2008), arXiv:0709.3519 [hep-ph].
[29] Z. Ligeti, I. W. Stewart, and F. J. Tackmann, Phys. Rev. D, **78**, 114014 (2008), arXiv:0807.1926 [hep-ph].
[30] F. U. Bernlochner *et al.* (SIMBA), PoS, **ICHEP2012**, 370 (2013), arXiv:1303.0958 [hep-ph].
[31] R. Abbate, M. Fickinger, A. H. Hoang, V. Mateu, and I. W. Stewart, Phys. Rev. D, **83**, 074021 (2011), arXiv:1006.3080 [hep-ph].
[32] J. R. Gaunt, (2014), arXiv:1405.2080 [hep-ph].
[33] T. Sjöstrand, S. Mrenna, and P. Skands, JHEP, **05**, 026 (2006), hep-ph/0603175.
[34] T. Sjöstrand, S. Mrenna, and P. Skands, Comput. Phys. Commun., **178**, 852 (2008), arXiv:0710.3820 [hep-ph].
[35] M. Bahr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton, *et al.*, Eur. Phys. J. C, **58**, 639 (2008), arXiv:0803.0883 [hep-ph].
[36] J. Bellm, S. Gieseke, D. Grellscheid, A. Papaefstathiou, S. Platzer, *et al.*, (2013), arXiv:1310.6877 [hep-ph].
[37] M. Cacciari, G. P. Salam, and G. Soyez, JHEP, **04**, 063

- (2008), arXiv:0802.1189 [hep-ph].
- [38] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. C, **72**, 1896 (2012), arXiv:1111.6097 [hep-ph].
 - [39] C. Lee and G. F. Sterman, Phys. Rev. D, **75**, 014022 (2007), hep-ph/0611061.
 - [40] V. Mateu, I. W. Stewart, and J. Thaler, Phys. Rev., **D87**, 014025 (2013), arXiv:1209.3781 [hep-ph].
 - [41] See Supplemental Material at the end of this preprint.
 - [42] C. Marcantonini and I. W. Stewart, Phys. Rev., **D79**, 065028 (2009), arXiv:0809.1093 [hep-ph].
 - [43] M. Dasgupta, L. Magnea, and G. P. Salam, JHEP, **02**, 055 (2008), arXiv:0712.3014 [hep-ph].
 - [44] D. Kang, C. Lee, and I. W. Stewart, Phys. Rev. D, **88**, 054004 (2013), arXiv:1303.6952 [hep-ph].
 - [45] M. Dasgupta and G. P. Salam, JHEP, **0208**, 032 (2002), hep-ph/0208073 [hep-ph].

SUPPLEMENTAL MATERIAL

Nonperturbative corrections

The leading hadronization effects in the first jet mass moment and in the tail of the jet mass spectrum are described by the parameter $\Omega_\kappa(R)$, which is defined by

$$\Omega_\kappa(R) = \int_0^1 dr \int_{-\infty}^{\infty} dy \int_0^{2\pi} d\phi f(r, y - y_J, \phi - \phi_J, R) \times \langle 0 | \bar{T}[Y_J^\dagger Y_b^\dagger Y_a^\dagger] \hat{\mathcal{E}}_T(r, y, \phi) T[Y_a Y_b Y_J] | 0 \rangle. \quad (15)$$

Here, Y_a and Y_b are incoming soft Wilson lines along the beam directions in the color representation of the incoming primary hard partons. $Y_J \equiv Y_J(y_J, \phi_J)$ is an outgoing soft Wilson line along the jet direction in the color representation of the outgoing hard parton. The color contractions between the Wilson lines are suppressed in Eq. (15), but are normalized such that $\langle 0 | Y_J^\dagger Y_b^\dagger Y_a^\dagger Y_a Y_b Y_J | 0 \rangle = 1$. In Fig. 6 we show that $\Omega_\kappa^{\text{had}}(R)$ in PYTHIA8 and HERWIG++ is independent of p_T^J over the large range of p_T^J considered. (The $qg \rightarrow Zq$ and $gg \rightarrow Hg$ channels in HERWIG++ have a small downward trend in p_T^J .)

For $e^+e^- \rightarrow$ dijets, a boost allows one to derive important universality properties of the corresponding Ω parameter [39]. Here for Eq. (15), boosting by $-y_J$ and rotating by $-\phi_J$, as illustrated in Fig. 5, the Wilson lines and energy flow operator transform as

$$\begin{aligned} Y_{a,b} &\rightarrow Y_{a,b}, \\ Y_J(y_J, \phi_J) &\rightarrow Y_J(0, 0), \\ \hat{\mathcal{E}}_T(r, y, \phi) &\rightarrow \hat{\mathcal{E}}_T(r, y - y_J, \phi - \phi_J). \end{aligned} \quad (16)$$

Changing variables $y \rightarrow y + y_J$ and $\phi \rightarrow \phi + \phi_J$ then yields an expression depending only on y and ϕ , which thus shows that $\Omega_\kappa(R)$ in Eq. (15) is independent of y_J and ϕ_J . We therefore set $y_J = \phi_J = 0$ in the following. Note that unlike for $e^+e^- \rightarrow$ dijets, the matrix element is not independent of y and ϕ , so these dependencies in the measurement $f(r, y, \phi)$ do not generically decouple. In Fig. 7 we show that $\Omega_\kappa^{\text{had}}$ obtained from Monte Carlos does not depend on the jet rapidity y_J . The behaviour with p_T^J and y_J shown in these plots does not depend on the value of R . In Figs. 6 and 7 we see again that in PYTHIA8 the overall size of $\Omega_\kappa^{\text{had}}(R)$ depends on the channel, being larger for the channels with a gluon jet. In contrast, $\Omega_\kappa^{\text{had}}(R)$ in HERWIG++ is smaller and of similar sizes for all channels.

To discuss the R dependence of $\Omega_\kappa(R)$, we switch to coordinates $\{y', \phi', r'\}$ measured with respect to the jet axis. This gives

$$\Omega_\kappa(R) = \int_0^1 dr' \int_{-\infty}^{\infty} dy' \int_0^{2\pi} d\phi' f(r', y', \phi', R)$$

$$\times \langle 0 | \bar{T}[Y_J^\dagger Y_b^\dagger(0, \pi) Y_a^\dagger(0, 0)] \hat{\mathcal{E}}_\perp(r', y', \phi') \times T[Y_a(0, 0) Y_b(0, \pi) Y_J] | 0 \rangle, \quad (17)$$

where the incoming beam Wilson lines $Y_{a,b}$ point in the $(y', \phi') = (0, 0)$ and $(0, \pi)$ directions, and $r' = p_\perp/m_\perp$. The measurement function in the original coordinates in Eq. (15) is given by

$$f(r, y, \phi, R) = (\cosh y - r \cos \phi) \theta[b(y, \phi, r) < R^2], \quad (18)$$

where we use $b(y, \phi, r) = 2(\cosh y - r \cos \phi)$ to define the jet boundary. In the primed coordinates it takes the form

$$\begin{aligned} f(r', y', \phi', R) &= e^{-y'} \theta \left[e^{2y'} - 2r'^2 \cos^2 \phi' + 1 \right. \\ &\quad \left. - \frac{2}{R^2} \sqrt{4 + R^4(r'^4 \cos^4 \phi' - r'^2 \cos^2 \phi')} \right]. \end{aligned} \quad (19)$$

Boosting along the *jet axis* by $\ln(R/2)$ as in Fig. 8, the Wilson lines and energy flow operator transform as

$$\begin{aligned} Y_a(0, 0) &\rightarrow Y_a(\ln \frac{R}{2}, 0), \\ Y_b(0, \pi) &\rightarrow Y_b(\ln \frac{R}{2}, \pi), \\ Y_J &\rightarrow Y_J, \\ \hat{\mathcal{E}}_\perp(r', y', \phi') &\rightarrow \hat{\mathcal{E}}_\perp(r', y' + \ln \frac{R}{2}, \phi'). \end{aligned} \quad (20)$$

In these coordinates, the beam Wilson lines are an angle $\theta = 4 \tan^{-1}(R/2) \simeq 2R$ apart.

We can now expand the result in R . To leading order in R , the measurement in Eq. (19) becomes

$$f(r', y', \phi', R) = e^{-y'} \theta[y' + \ln(R/2)] [1 + \mathcal{O}(R^2)]. \quad (21)$$

For the leading term in the $R \rightarrow 0$ limit, the beam Wilson lines fuse

$$Y_a(\ln \frac{R}{2}, 0) Y_b(\ln \frac{R}{2}, \pi) = Y_{\bar{J}}(-\infty, 0) + \mathcal{O}(R^2), \quad (22)$$

where $Y_{\bar{J}}$ is an incoming Wilson line along the direction opposite to the jet and in the appropriate conjugate color representation that forms a color singlet with the outgoing jet Wilson line Y_J . Since we now have two Wilson lines along the jet axis, we can boost along the jet axis to eliminate the y' dependence. Integrating over ϕ' then yields the result in Eq. (10), namely

$$\Omega_\kappa(R) = \frac{R}{2} \Omega_\kappa^{(1)} + \frac{R^3}{8} \Omega_\kappa^{(3)} + \frac{R^5}{32} \Omega_\kappa^{(5)} + \mathcal{O}\left[\left(\frac{R}{2}\right)^7\right], \quad (23)$$

where the coefficient of the leading term comes from integrating the measurement function over y' ,

$$\int_{-\infty}^{\infty} dy' e^{-y'} \theta[y' + \ln(R/2)] = \frac{R}{2}. \quad (24)$$

The leading nonperturbative parameter in Eq. (23) is given the universal matrix element

$$\Omega_\kappa^{(1)} = c_e \int_0^1 dr' g_e(r') \langle 0 | \bar{T}[Y_J^\dagger Y_{\bar{J}}^\dagger] \hat{\mathcal{E}}_\perp(r') T[Y_{\bar{J}} Y_J] | 0 \rangle. \quad (25)$$

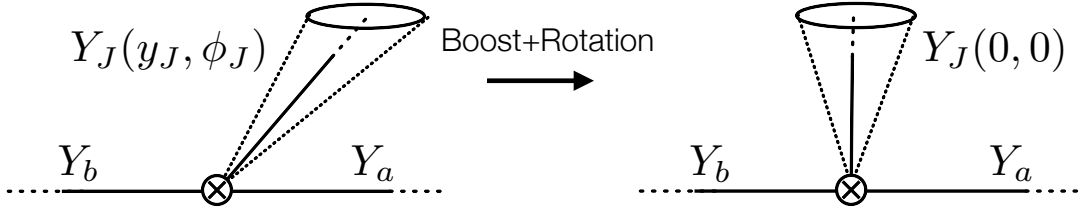


FIG. 5. Boost by $-y_J$ along the beam direction and rotation by $-\phi_J$ around the beam direction used to show that Ω_κ is independent of y_J and ϕ_J .

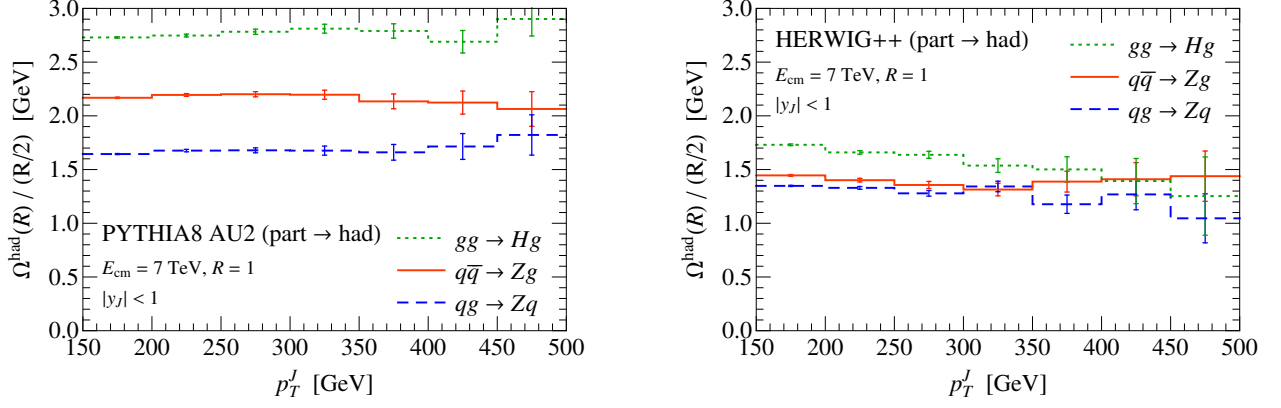


FIG. 6. p_T^J dependence of $\Omega_\kappa^{\text{had}}(R)$ for PYTHIA8 (left panel) and HERWIG++ (right panel).

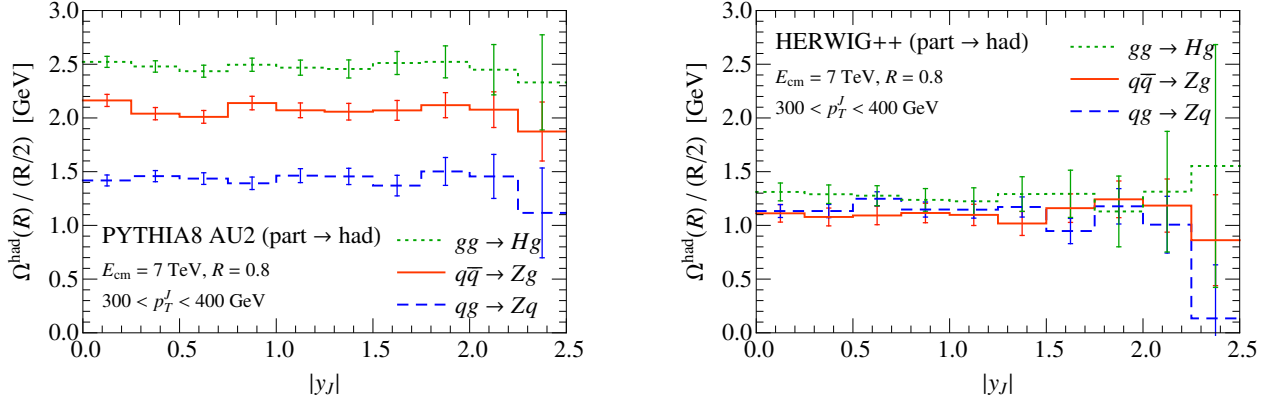


FIG. 7. Jet rapidity dependence of $\Omega_\kappa^{\text{had}}(R)$ for PYTHIA8 (left panel) and HERWIG++ (right panel).

It depends on the color representation of the Wilson line (quark vs. gluon) but not the full original color configuration. To extend our result to a more general jet measurement e , we included the parameters c_e and $g_e(r')$, which in our case simply are given by $c_e = g_e(r') = 1$. In general c_e is the coefficient for the observable e [39] obtained here by integrating over our y' variable, and g_e encodes the hadron mass effects [40].

The expansions in Eqs. (22) and (21) can be carried out to higher orders in R , using Ref. [42] to expand the Wilson lines about the \bar{J} direction, and lead to new nonperturbative matrix elements, collectively denoted as $\Omega_\kappa^{(3,5)}$ in Eq. (23). Terms with an odd number of gauge field components that are transverse to the jet direction vanish due to parity invariance. Together with the overall

	κ	$\Omega_\kappa^{(1)}$	$\Omega_\kappa^{(3)}$	$\Omega_\kappa^{(5)}$	[GeV]
PYTHIA8 AU2	$qg \rightarrow q$	1.2	1.5	1.3	
PYTHIA8 4C	$qg \rightarrow q$	1.1	0.7	2.0	
HERWIG++	$qg \rightarrow q$	1.2	-0.9	4.0	
PYTHIA8 AU2	$q\bar{q} \rightarrow g$	2.1	-0.9	3.0	
PYTHIA8 4C	$q\bar{q} \rightarrow g$	2.1	-1.4	3.4	
HERWIG++	$q\bar{q} \rightarrow g$	1.0	0.3	2.4	
PYTHIA8 AU2	$gg \rightarrow g$	2.2	1.5	2.4	
PYTHIA8 4C	$gg \rightarrow g$	2.1	0.4	3.0	
HERWIG++	$gg \rightarrow g$	1.0	1.3	1.5	

TABLE I. Fit coefficients for $\Omega_\kappa(R)$ in Eq. (10) for different Monte Carlos and tunes which give the lines shown in Fig. 2.

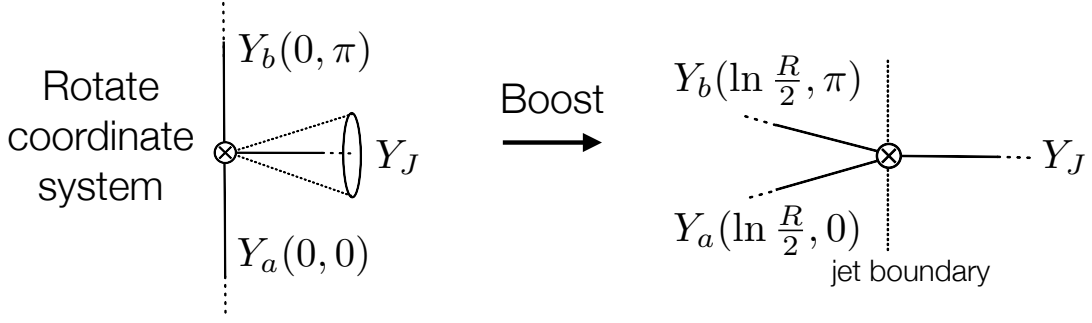


FIG. 8. Changing coordinates and boosting by $\ln(R/2)$ along the jet direction in order to expand around small R .

factor of R , this implies that $\Omega_\kappa(R)$ only contains odd powers of R . The coefficients of the fits shown in Fig. 2 are given in Table I. The leading coefficient in R , $\Omega_\kappa^{(1)}$, is the same for quark and gluon jets, while the higher coefficients are quite different for all three channels. The

higher coefficients $\Omega_\kappa^{(3)}$ and $\Omega_\kappa^{(5)}$ strongly depend on the Monte Carlo and tune. They are also correlated so their separation is not well constrained by the fit. The fact that all the coefficients are of similar size confirms that $R/2$ is indeed the appropriate expansion parameter.

As an illustration of the utility of the operator formulation, we give explicit results for some $\Omega_\kappa^{(1,3)}$ s. For $\Omega_{qg \rightarrow q}^{(1)}$ and $\Omega_{qg \rightarrow q}^{(3)}$ we have

$$\begin{aligned} \Omega_{qg \rightarrow q}^{(1)} &= \Omega_q^{(1)} = \int_0^1 dr \frac{1}{N_c} \langle 0 | \text{tr} \{ Y_J^\dagger Y_J \hat{\mathcal{E}}_\perp(r) Y_J^\dagger Y_J \} | 0 \rangle, \\ \Omega_{qg \rightarrow q}^{(3)} &= \int_0^1 dr \frac{(-1)}{N_c} \langle 0 | \text{tr} \left\{ \left[Y_J^\dagger Y_J \hat{\mathcal{E}}_\perp(r) Y_J^\dagger Y_J, \frac{1}{\bar{n}_J \cdot i\partial} g n_J \cdot \mathcal{B}_J + \frac{2}{(\bar{n}_J \cdot i\partial)^2} i\partial_\perp \cdot g \mathcal{B}_{J\perp} - \frac{1}{\bar{n}_J \cdot i\partial} \left[\frac{1}{\bar{n}_J \cdot i\partial} g \mathcal{B}_{J\perp}^\mu, g \mathcal{B}_{J\perp\mu} \right] \right] \right. \\ &\quad + \left[\left[Y_J^\dagger Y_J \hat{\mathcal{E}}_\perp(r) Y_J^\dagger Y_J, \frac{1}{\bar{n}_J \cdot i\partial} g \mathcal{B}_{J\perp}^\mu \right], \frac{1}{\bar{n}_J \cdot i\partial} g \mathcal{B}_{J\perp\mu} \right] + \frac{1}{C_F} \left[\left[Y_J^\dagger Y_J \hat{\mathcal{E}}_\perp(r) Y_J^\dagger Y_J, \frac{1}{\bar{n}_J \cdot i\partial} g \mathcal{B}_{J\perp}^A \right], \frac{1}{\bar{n}_J \cdot i\partial} g \mathcal{B}_{J\perp\mu}^A \right] \\ &\quad \left. - \frac{(1-r^2)}{2} Y_J^\dagger Y_J \hat{\mathcal{E}}_\perp(r) Y_J^\dagger Y_J \right\} | 0 \rangle, \end{aligned} \quad (26)$$

where the Wilson lines Y_J , $Y_{\bar{J}}$, and $\mathcal{B}_J^\nu = \mathcal{B}_J^{\nu A} T^A = \frac{1}{g} [Y_J^\dagger i D^\nu Y_J]$ are all in the fundamental representation, and tr is a trace over 3 and $\bar{3}$ color indices. The path for $Y_J^\dagger Y_J$ is $[-\infty, 0]$ along \bar{n}_J , then $[0, \infty]$ along n_J . The measurement is normalized such that $\hat{\mathcal{E}}_\perp(r) = 2\pi \hat{\mathcal{E}}_\perp(r, 0, 0)$, which is equal to $\hat{\mathcal{E}}_\perp(r, y = 0)$ of Ref. [40]. In Eq. (26) the inverse derivatives $1/(\bar{n}_J \cdot i\partial)$ only act on the fields they are next to, and the fields on the right (left) side of the measurement $\hat{\mathcal{E}}_\perp(r)$ are (anti) time-ordered. For $\Omega_{q\bar{q} \rightarrow g}^{(1)}$, $\Omega_{gg \rightarrow g}^{(1)}$, and $\Omega_{q\bar{q} \rightarrow g}^{(3)}$ we have

$$\begin{aligned} \Omega_{q\bar{q} \rightarrow g}^{(1)} &= \Omega_g^{(1)} = \int_0^1 dr \frac{1}{N_c^2 - 1} \langle 0 | \text{Tr} \{ \mathcal{Y}_J^T \mathcal{Y}_J \hat{\mathcal{E}}_\perp(r) \mathcal{Y}_J^T \mathcal{Y}_J \} | 0 \rangle = \Omega_{gg \rightarrow g}^{(1)}, \\ \Omega_{q\bar{q} \rightarrow g}^{(3)} &= \int_0^1 dr \frac{(-1)}{N_c^2 - 1} \langle 0 | \text{Tr} \left\{ \left[\mathcal{Y}_J^T \mathcal{Y}_J \hat{\mathcal{E}}_\perp(r) \mathcal{Y}_J^T \mathcal{Y}_J, \frac{1}{\bar{n}_J \cdot i\partial} g n_J \cdot \tilde{\mathcal{B}}_J + \frac{2}{(\bar{n}_J \cdot i\partial)^2} i\partial_\perp \cdot g \tilde{\mathcal{B}}_{J\perp} + \frac{1}{\bar{n}_J \cdot i\partial} \left[\frac{1}{\bar{n}_J \cdot i\partial} g \tilde{\mathcal{B}}_{J\perp}^\mu, g \tilde{\mathcal{B}}_{J\perp\mu} \right] \right] \right. \\ &\quad + \left[\left[\mathcal{Y}_J^T \mathcal{Y}_J \hat{\mathcal{E}}_\perp(r) \mathcal{Y}_J^T \mathcal{Y}_J, \frac{1}{\bar{n}_J \cdot i\partial} g \tilde{\mathcal{B}}_{J\perp}^\mu \right], \frac{1}{\bar{n}_J \cdot i\partial} g \tilde{\mathcal{B}}_{J\perp\mu} \right] + \frac{1}{T_F N_c} \left[\left[\mathcal{Y}_J^T \mathcal{Y}_J \hat{\mathcal{E}}_\perp(r) \mathcal{Y}_J^T \mathcal{Y}_J, \frac{1}{\bar{n}_J \cdot i\partial} g \tilde{\mathcal{B}}_{J\perp}^A \right], \frac{1}{\bar{n}_J \cdot i\partial} g \tilde{\mathcal{B}}_{J\perp\mu}^A \right] \\ &\quad \left. - \frac{(1-r^2)}{2} \mathcal{Y}_J^T \mathcal{Y}_J \hat{\mathcal{E}}_\perp(r) \mathcal{Y}_J^T \mathcal{Y}_J \right\} | 0 \rangle, \end{aligned} \quad (27)$$

where the Wilson lines \mathcal{Y}_J and $\mathcal{Y}_{\bar{J}}$ are in the adjoint representation, the gluon fields $\tilde{\mathcal{B}}_J^{ab} = -if^{Cab} \mathcal{B}_J^C$ and $\bar{\mathcal{B}}_J^{ab} = d^{Cab} \mathcal{B}_J^C$ are matrices, and Tr is a trace over adjoint color indices. When Tr acts on the term with $\tilde{\mathcal{B}}_J^A = \mathcal{B}_J^A$ it simply contracts these color vectors to the appropriate sides of the color matrix $\mathcal{Y}_J^T \mathcal{Y}_J \hat{\mathcal{E}}_\perp(r) \mathcal{Y}_J^T \mathcal{Y}_J$.

Soft function contribution to ISR

At $\mathcal{O}(\alpha_s)$ the soft function contains the following term [23]

$$S_\kappa^{\text{pert}}(k_S) \supset [I_0(\alpha, \beta) + I_0(\beta, \alpha)] \frac{\alpha_s C_\kappa}{\pi} \frac{1}{\mu} \left(\frac{\mu}{k_S} \right)_+. \quad (28)$$

The color factor for this interference of soft ISR from the two beams is given by the color charge of the two incoming partons, $C_\kappa = -\mathbf{T}_a \cdot \mathbf{T}_b$. For the processes we consider, this is simply a number given in Eq. (13), but in general this is a matrix in color space. The I_0 in Eq. (28) is given by the following integral

$$I_0(\alpha, \beta) = \frac{1}{\pi} \int_{-\pi}^{\pi} d\phi \int dy \theta(e^{y_J - y} - \sqrt{\beta/\alpha}) \times \theta(1/\alpha - 1 - e^{2(y_J - y)} + 2e^{y_J - y} \cos \phi). \quad (29)$$

with parameters

$$\begin{aligned} \alpha &= (1 - \tanh y_J)/(2\rho), \\ \beta &= (1 + \tanh y_J)/(2\rho). \end{aligned} \quad (30)$$

Here, $\rho(R, y_J)$ controls the jet size, which is chosen such that the jet area in (y, ϕ) space equals πR^2 [23]. The total integral in Eq. (29) is an area in (y, ϕ) space, where the second theta function restricts the integral to the jet and the first theta function reduces to $\theta(y < 0)$ and $\theta(y > 0)$

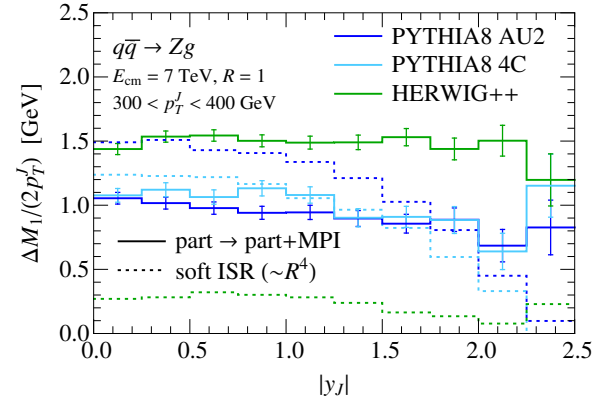
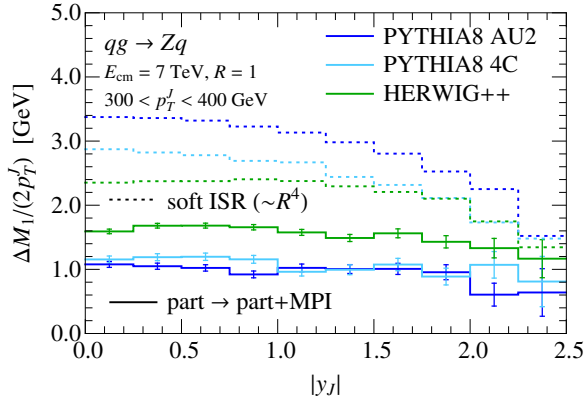
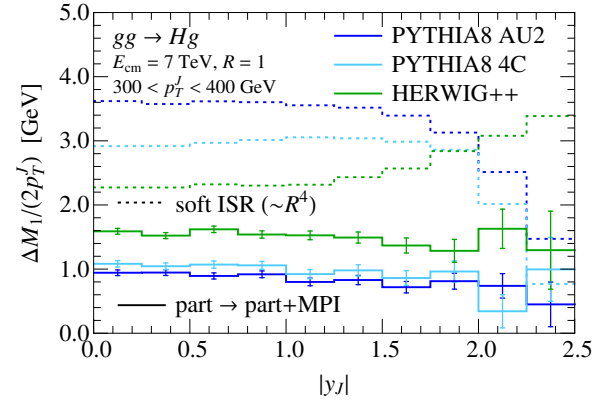
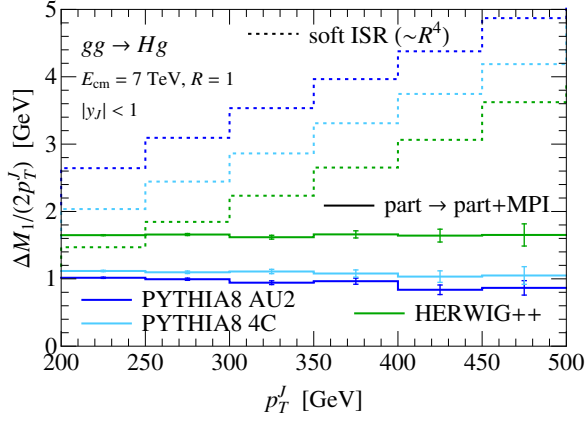
for $I_0(\alpha, \beta)$ and $I_0(\beta, \alpha)$, respectively. Therefore, including the overall $1/\pi$ factor,

$$I_0(\alpha, \beta) + I_0(\beta, \alpha) = R^2, \quad (31)$$

which yields the R^2 dependence shown in Eq. (12).

The p_T dependence of the MPI and soft ISR contributions to the jet mass moment is discussed in Fig. 4. In Fig. 9 we show in addition the y_J dependence in the same way. The y_J dependence of the MPI is essentially flat, except for perhaps a small reduction at large rapidities. Since soft ISR emissions are constant in rapidity, one would expect the soft ISR contribution to the moment to be independent of the jet rapidity at central rapidities. This agrees well with what is observed in HERWIG++ for $|y_J| \lesssim 1.5$, while for larger y_J the soft ISR contribution reduces. As already observed before, PYTHIA8 has a larger soft ISR and smaller MPI contribution than HERWIG++. In addition, the y_J dependence of the soft ISR differs noticeably between PYTHIA8 and HERWIG++. Hence, measurements of the y_J rapidity dependence can also provide constraints on the modelling of soft ISR in the Monte Carlos.

For completeness we have included the analogs of Fig. 4 and Fig. 9 for the $gg \rightarrow Hg$ channel in Fig. 10. Note that the size of the R^4 contribution from soft ISR for this channel is very similar to $qg \rightarrow Zq$ (at central rapidities). This might be surprising since this is a purely gluonic process, but it is in agreement with the prediction from the color factors in Eq. (13).

FIG. 9. Same as Fig. 4 but for the y_J dependence.FIG. 10. Same as Fig. 4 and Fig. 9 but for the $gg \rightarrow Hg$ process.