An essential part of high-energy hadronic collisions is the soft hadronic activity that underlies the primary hard interaction. It includes 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 $Z + \text{jet}$ and $H + \text{jet}$ events is directly sensitive to these effects, and we use a QCD factorization theorem to predict its 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. We prove that 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, but this degeneracy is broken by dependence on the jet $p_T$. 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 $gg \rightarrow Zg$, 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.

This is done using the dependence of the jet mass spectrum and its first moment on the jet radius $R$, jet momentum $p_T$, jet rapidity $y_J$, and participating partons. We will not consider factorization-breaking effects here (see, e.g., Ref. [21]). We consider the jet mass spectrum in exclusive $pp \rightarrow Z + 1\text{jet}$ and $pp \rightarrow H + 1\text{jet}$ events. The factorization formula for $m_J \ll p_T$ that includes sources 1 and 2 is given by [22–24]

$$
\frac{d\sigma}{dm_J^2d\Phi_2} = \sum_{k,a,b} H_k(\Phi_2) \int dk_S dk_B (I_{k,a}I_{k,b} \otimes f_a f_b)(k_B) \times J_{k_f}(m_J^2 - 2p_T^2k_S) S_k(k_S, p^\text{cut} - k_B, y_J, R).
$$

(1)

Here, $\Phi_2 = \{p_T, y_J, Y\}$, $Y$ is the rapidity of the $Z/H + \text{jet}$ system, $\kappa$ denotes the partonic channel, and $k_S$ and $k_B$ account for soft contributions to the jet mass $m_J^2$ and jet veto $p^\text{cut}$ (which vetoes additional jets). The $H_\kappa(\Phi_2)$ contains the perturbative matrix elements for the hard process, and $I_{k,a}I_{k,b} \otimes f_a f_b$ describes perturbative collinear initial-state radiation convolved with the parton distribution functions. For the normalized jet mass spectrum, the dependence on $p^\text{cut}$ largely drops out [24]. As a result, the shape of the jet mass spectrum is determined by the jet function $J_{k_f}$, describing energetic final-state radiation, and by the soft function $S_k$. See also Refs. [25,26].

The soft function $S_k$ describes the primary initial and final-state soft radiation. It depends on the jet through $y_J$ and $R$ but not $p_T$, and can be factorized as [27–29].
For our numerical studies, we consider both PYTHIA8 and MSTW2008LO \[16\] and HERWIG++ 2.7 \[35,36\] with its default tune UE-EE-5-MRST \[18\]. Both in quark and gluon channels, with the leading jet mass spectrum. 

\begin{align}
S_x(k_S, k_B, y_J, R) &= \int dk S^\text{pert}_x(k_S - k, k_B, y_J, R) \\
&\quad \times F_x(k, y_J, R) \left[1 + \mathcal{O}(\Lambda_{\text{QCD}}/k_B)\right], \tag{2}
\end{align}

where $S^\text{pert}_x$ contains the perturbative soft contributions. $F_x$ is a normalized nonperturbative shape function which encodes the smearing effect that the hadronization has on the soft momentum $k_S$. For $k_S \gg \Lambda_{\text{QCD}}$, the full $F_x(k)$ 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 $k_S \gg \Lambda_{\text{QCD}}$, $S_x$ can be expanded,

\begin{align}
S_x(k_S, y_J, R) &= S^\text{pert}_x(k_S - \Omega_x(R), y_J, R) \\
&\quad + \mathcal{O}(\Lambda^2_{\text{QCD}}/k^2_S, \alpha_s \Lambda_{\text{QCD}}/k^2_S), \tag{3}
\end{align}

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

Factorization also underlies the Monte Carlo description of the primary collision, where $H$ corresponds to the hard matrix element, while $I$, $J$, and $S$ are described by 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 $S_x$. Monte Carlo programs include MPI (source 3), which are not 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 line) and including both hadronization and MPI (blue dashed line). Equation (3) predicts that for $m_J^2 > \Lambda^2_{\text{QCD}}p_T^J$ the nonperturbative corrections shift the tail of the jet mass spectrum by

$$m_J^2 = (m_J^2)^\text{pert} + 2p_T^J\Omega_x(R). \tag{4}$$

We can regard the partonic result from PYTHIA8 as the baseline purely perturbative result. Choosing $\Omega = 2.4$ GeV for $qg \to Zg$ and $\Omega = 2.7$ GeV for $qg \to Zg$ 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 this shift. For hadronization, Eqs. (1), (2) predict a convolution with a nonperturbative function,

$$\frac{d\sigma_x}{dm_J^2} = \int dk \frac{d\sigma_x^{\text{partonic}}}{dm_J^2} \left( (m_J^2)^\text{pert} - 2p_T^Jk \right) F_x(k). \tag{5}$$

With the above $\Omega$’s, this convolution gives the red solid curves in Fig. 1, yielding excellent agreement with the hadronization + MPI result over the full range of the jet mass spectrum. [Here, $F_x(k) = (4k/\Omega_x^2)e^{-k/\Omega_x}$; the simplest ansatz that satisfies the required properties: normalization, vanishing at $k = 0$, falling off exponentially for $k \to \infty$, and having a first moment $\Omega_x$. Fixing the value of $\Omega_x$ from the tail, we find similar levels of agreement across all values of $p_T^J$, $y_J$, $R$, for all parton channels, and for different jet veto cuts (including no jet veto.)] 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 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 program contributions for soft ISR, hadronization, and MPI.
We consider the first moment in \( m^2 \),
\[
M_1 = \frac{1}{\sigma} \int dm^2 m^2 \frac{d\sigma}{dm^2},
\]
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, giving
\[
M_1 = M^\text{pert}_{1k}(p_T^j, y_j, R) + 2 p_T^{\mu} \Omega_\kappa(R).
\]  
(7)  

Here, \( M^\text{pert}_{1k}(p_T^j, y_j, R) \) contains all perturbative contributions, while \( \Omega_\kappa(R) \) encodes the shift due to nonperturbative effects.

For \( pp \to H/Z + \text{jet} \), \( \Omega_\kappa(R) \) is given by the vacuum matrix element of lightlike soft Wilson lines \( Y_a, \ Y_b, \) and \( Y_j \equiv Y(y_j, \phi_j) \) along the beam and jet directions,
\[
\Omega_\kappa(R) = \int_0^1 dr \int_0^\infty dy \int_0^{2\pi} d\phi f(r, y, y_j, \phi - \phi_j, R) 
\times \langle 0 | \hat{T} [Y_a^j Y_j^i Y_b^j] \hat{\xi}_T(r, y, \phi) \hat{T} | Y_a Y_b Y_j] | 0 \rangle.
\]  
(8)  

Here, the rapidity \( y \), azimuthal angle \( \phi \), 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, giving the \( \kappa \) dependence of \( \Omega_\kappa \). The jet mass measurement function is \( f(r, y, \phi, R) = (\cosh y - r \cos \phi) \delta[b(y, \phi, r) < R]\), where \( b(y, \phi, r) \) specifies the jet boundary. The matrix element involves the energy flow operator \( \hat{\xi}_T(r, y, \phi)|X\rangle = \sum_i \delta(x_m m^2) \delta(r - r) \delta(y - y_i) \delta(\phi - \phi_i)|X\rangle \).

From Eq. (8), 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 \( \phi_j \) (see the Supplemental Material [44]).

Expanding Eq. (8) for small \( R \), we find [44,45]
\[
\Omega_\kappa(R) = \frac{R}{2} \Omega_\kappa^{(1)} + \frac{R^3}{8} \Omega_\kappa^{(3)} + \frac{R^5}{32} \Omega_\kappa^{(5)} + O \left( \frac{R}{2} \right)^7,
\]  
(9)  

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. [46] from a QCD hadronization model. Our operator definition implies a universality for the linear \( R \) nonperturbative parameter in Eq. (9). For \( R \to 0 \), the beam Wilson line fuse into a Wilson line in the conjugate representation to the jet, \( Y_a Y_b \to Y_j \). The result is given by (see the Supplemental Material [44])
\[
\Omega_\kappa^{(1)} = \int_0^1 dr \langle 0 | \hat{T} Y_j^i Y_j^i | \hat{\xi}_T(r') T | Y_j Y_j | 0 \rangle,
\]  
(10)  

which only depends on whether the jet is a quark or gluon jet. For quarks, we can compare this to thrust in deep-inelastic scattering (DIS) [47], where precisely this parameter \( \Omega_\kappa^{(1)} \) appears [48].

Consider next \( M^\text{pert}_{1k} \) in Eq. (7). Dimensional analysis and the kinematical bound \( m_J \lesssim p_T^j R \) imply that \( M^\text{pert}_{1k} \) scales like \( (p_T^j R)^2 \). Resummation modifies the leading \( R \) dependence to \( R^{2-\gamma_s} \), where \( \gamma_s \sim \alpha_s > 0 \). The soft function contains a contribution due to interference between ISR from the two beams (see the Supplemental Material [44]),
\[
S^\text{pert}_\kappa(k_S) \sim \frac{\alpha_s C_k R}{\frac{1}{\mu}} \frac{1}{k_S^2}.
\]  
(11)  

The extra \( R^2 \) for soft ISR causes it to contribute to \( M^\text{pert}_{1k} \) as \( (p_T^j R)^2 R^k \) with the color factors
\[
C_{qg-gq} = C_{gg-qg} = C_\lambda = \frac{3}{2},
\]
\[
C_{qg-qg} = C_F - \frac{C_\lambda}{2} = \frac{1}{6}.
\]  
(12)  

The above factorization results can be compared to PYTHIA8 and HERWIG++, where we find that the dependence of \( M_1 \) on \( p_T^j, y_j, \kappa \), is well described by

![FIG. 2](color online) The \( R \) dependence of \( \Omega^{\text{had}}_\kappa(R) \) extracted from \( M_1 \) in PYTHIA8 (left panel) and HERWIG++ (right panel), shown as dots, triangles, and squares for different channels. The fit using Eq. (9) (shown by lines) demonstrates the agreement with factorization. The small-\( R \) behavior only depends on whether the jet is initiated by a quark (blue dashed line) or gluon (orange solid and green dotted lines).
\begin{equation}
M_1 = M_{1x}^{\text{partonic}}(p_T^f, y_J, R) + 2 p_T^f [\Omega_k^{\text{had}}(R) + 2 p_T^f [\Upsilon_{\text{MPI}}(y_J, R) + \Omega_k^{\text{MPI}}(y_J, R)]].
\end{equation}

Here, \(M_{1x}^{\text{partonic}}\) is the partonic contribution, \(\Omega_k^{\text{had}}\) is defined by partonic \(\rightarrow\) hadronic, and \(\Upsilon_{\text{MPI}}\) by partonic \(\rightarrow\) partonic + MPI. The small remainder from hadronization of the MPI, \(\Omega_k^{\text{MPI}}\), is defined to ensure the sum of terms yields the full partonic \(\rightarrow\) hadronic + MPI. Note that hadronization and MPI contributions are each individually described by shifts to \(M_1\). Also, the independence of \(\Omega_k\) to \(y_J\) and \(\phi_J\) is observed in both \textsc{Pythia}8 and \textsc{Herwig}++ (see the Supplemental Material [44]). Equation (13) contains MPI contributions with no analog in Eq. (7).

The hadronization \(\Omega_k^{\text{had}}(R)/(R/2)\) from \textsc{Pythia}8 and \textsc{Herwig}++ is shown in Fig. 2 for different channels. For \(R \ll 1\), \(\Omega_k^{\text{had}}(R)\) is linear in \(R\) and has the same slope for the two channels involving gluon jets, as predicted by factorization. For \textsc{Pythia}8, all channels differ for large \(R\) and can be fit to the factorization form in Eq. (9). For the quark jet, we extract \(\Omega_k^{(1)} = 1.2\) GeV and for gluon jets \(\Omega_k^{(1)} = 2.2\) 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 \textsc{Herwig}++, the dependence on higher powers of \(R\) is much weaker, and \(\Omega_k^{(1)} \approx \Omega_k^{(1)}\). The full set of fit coefficients is in the Supplemental Material [44].

In Fig. 3, we compare our perturbative next-to-leading logarithmic (NLL) and next-to-next-to-leading logarithmic (NNLL) factorization predictions [24] for \(M_{1x}^{\text{partonic}}\) to the corresponding \(M_{1x}^{\text{partonic}}\) from \textsc{Pythia}8 and \textsc{Herwig}++ as a function of \(R\), dividing by the leading \(R^2\) dependence. The \(R^4\) contribution from soft ISR only enters at NNLL and is seen in the rise at large \(R\) for \(qg \rightarrow Zq\) (left panel). This effect is partially modeled by soft emissions in the parton shower, which explains the similar \(R^4\) contribution for \(qg \rightarrow Zq\) in \textsc{Pythia}8 and \textsc{Herwig}++. For \(q\bar{q} \rightarrow Zq\) (right panel), Eqs. (11) and (12) predict the \(R^4\) contribution from soft ISR to be negative, which we observe at NNLL. This negative interference effect is not captured by these Monte Carlo programs.

The apparent ambiguity between \(R^4\) contributions from soft ISR and MPI can be resolved through their \(p_T^f\) dependence. In Fig. 4, we show the \(R^4\) component \(c_4^k\) of the partonic moment, obtained by fitting

\begin{equation}
\frac{M_{1x}^{\text{part}}}{2 p_T^f R^2} = c_4^k R^{-c_4} + c_4^k R^2.
\end{equation}

and also the MPI contribution to the moment, \(\Upsilon_{\text{MPI}}/R^2 \sim R^2\). The differences between various tunes for \(c_4^k\) and \(\Upsilon_{\text{MPI}}\) reflects their apparent ambiguity, whereas their sum agrees much better. The \(p_T^f\) dependence clearly resolves the ambiguity: \(c_4^k \sim p_T^f\) as predicted by factorization, whereas \(\Upsilon_{\text{MPI}}\) is independent of \(p_T^f\). As shown in the Supplemental Material [44], the channel dependence could...
also be used to separate soft ISR from MPI: $c^e$ depends on the color channel as in Eq. (12), whereas $\gamma^\text{MPI}$ is channel independent. Also, the $\gamma_j$ 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 $pp \rightarrow H/\tilde{Z} + \text{jet}$. 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 Carlo programs 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 $p_T^f$ and on the partonic process. Hence, separately measuring quark and gluon channels in Drell-Yan events and in different bins of $p_T^f$ 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-SC0011090, the DFG Emmy-Noether Grant No. TA 867/1-1, and the Marie Curie International Incoming Fellowship PIIF-GA-2012-328913 within the 7th European Community Framework Program. We thank the Erwin Schrödinger Institute for hospitality while portions of this work were completed.