1. SYSTEMATICS OF QUARK/GLUON TAGGING[1,2,3]

By measuring the substructure of a jet, one can assign it a "quark" or "gluon" tag. In the eikonal limit, quark/gluon discrimination is determined solely by the color factor of the initiating parton (C_F versus C_A). In this section, we confront the challenges faced when going beyond this leading-order understanding, using parton shower generators to assess the impact of higher-order perturbative and non-perturbative physics. Working in the idealized context of electron-positron collisions, where there exists an unambiguous definition of quark and gluon jets, we find a fascinating interplay between perturbative shower effects and non-perturbative hadronization effects.

1.1 Overview

Jets are a robust tool for studying short-distance collisions involving quarks and gluons. With a suitable jet definition, one can connect jet measurements made on clusters of hadrons to perturbative calculations made on clusters of partons. More ambitiously, one can try to tag jets with a suitably-defined flavor label, thereby enhancing the fraction of, say, quark-tagged jets over gluon-tagged jets. This is relevant for searches for physics beyond the standard model, where signals of interest are often dominated by quarks while the corresponding backgrounds are dominated by gluons. A wide variety of quark/gluon discriminants have been proposed [1, 2, 3, 4, 5, 6, 7], and there is a growing catalog of quark/gluon studies at the Large Hadron Collider (LHC) [8, 9, 10, 11, 12, 13].

In order to achieve robust quark/gluon tagging, though, one needs theoretical and experimental control over quark/gluon radiation patterns. At the level of eikonal partons, a quark radiates proportional to its $C_F = 4/3$ color factor while a gluon radiates proportional to $C_A = 3$. In this section, we demonstrate that quark/gluon discrimination performance is highly sensitive to subleading perturbative effects beyond the eikonal limit, such as $g \to q\bar{q}$ splittings and color coherence, as well as to non-perturbative effects such as color reconnection and hadronization. While these effects are modeled (to differing degrees) in parton shower generators, they are relatively unconstrained by existing collider measurements, especially in the gluon channel. The goal of this section is to highlight these uncertainties, which then suggests a set of future LHC measurements that will improve the modeling of jets in general and quark/gluon tagging in particular.

A common misconception about quark/gluon tagging is that it is an intrinsically ill-defined problem. Of course, quark and gluon partons carry color while jets are composed of color-singlet hadrons, so the labels "quark" and "gluon" are fundamentally ambiguous. But this is philosophically no different from the fact that a "jet" is fundamentally ambiguous and one must therefore always specify a concrete jet finding procedure. As discussed in Sec. 1.2, the proper way to think about quark/gluon discrimination is in the context of unambiguous hadron-level measurements. In that sense, quark/gluon tagging is a well-defined technique for enhancing desired signals over undesired backgrounds.

There are a wide range of possible quark/gluon discriminants and a similarly large range of ways to quantify discrimination power. As a concrete set of discriminants, we consider the generalized angularities λ_β^κ [7] (see also [14, 15, 16, 17]). As discussed further in Sec. 1.3, we consider five different (κ, β) working points, which roughly map onto five variables in common use in the literature:

$$\text{multiplicity,} \quad p_T^D, \quad \text{LHA,} \quad \text{width,} \quad \text{mass.} \tag{1}$$

Here, multiplicity is the hadron multiplicity within the jet, p_T^D was defined in Refs. [4, 5], LHA refers to the "Les Houches Angularity" (named after the venue of this workshop), width is closely related to

---

jet broadening [18, 19, 20], and mass is closely related to jet thrust [21]. To quantify discrimination performance, we focus on classifier separation (a default output of TMVA [22]):

$$\Delta = \frac{1}{2} \int d\lambda \, \frac{\left(p_q(\lambda) - p_g(\lambda)\right)^2}{p_q(\lambda) + p_g(\lambda)}, \tag{2}$$

where p_q (p_g) is the probability distribution for a quark-tagged (gluon-tagged) sample. Other potential performance metrics are discussed in Sec. 1.4.

We begin our parton shower generator predictions for quark/gluon discrimination in Sec. 1.5, using an idealized setup with e^+e^- collisions. Here, we can use the following processes as unambiguous proxies for quark and gluon jets:

$$\text{"quark jets"}: \quad e^+e^- \to (\gamma/Z)^* \to u\bar{u}, \tag{3}$$
$$\text{"gluon jets"}: \quad e^+e^- \to h^* \to gg, \tag{4}$$

where h is the Higgs boson. These processes are physically distinguishable by the quantum numbers of the associated color singlet production operator, giving a way to label quarks and gluons without reference to the final state. We compare four different parton shower generators both before hadronization ("parton level") and after hadronization ("hadron level"):

- PYTHIA 8.205 [23],
- VINCIA 1.201 [24],
- HERWIG++ 2.7.1 [25],
- SHERPA 2.1.1 [26].

We also show parton-only results for a new generator:

- DEDUCTOR 1.0.2 [27].

In the future, we plan to make DIRE [28] predictions about quark/gluon discrimination as well as investigate predictions from analytic resummation [6, 7].

As we will see, the differences between these generators arise from physics at the interface between perturbative showering and non-perturbative fragmentation. One might think that the largest differences between generators would appear for infrared-and-collinear (IRC) unsafe observables like multiplicity and p_T^D, where non-perturbative hadronization plays an important role. Surprisingly, comparably-sized differences are also seen for the IRC safe angularities, indicating that these generators have different behavior even at the level of the perturbative final state shower. In Sec. 1.6, we study these differences as a function of collisions energy Q, jet radius R, and strong coupling constant α_s, showing that the generators have somewhat different discrimination trends. In Sec. 1.7, we compare the default parton shower configurations to physically-motivated changes, showing that modest changes to the shower/hadronization parameters can give rather large differences in quark/gluon separation power.

At the end of the day, most of the disagreement between generators is due to gluon radiation patterns. This is not so surprising, since all of the generators have been tuned to reproduce distributions from e^+e^- colliders, and quark (but less so gluon) radiation patterns are highly constrained by event shape measurements at LEP [29, 30, 31, 32]. In Sec. 1.8, we suggest a possible analysis strategy at the LHC to specifically constrain gluon radiation patterns. At a hadron collider, the distinction between quark jets and gluon jets is rather subtle, since radiation patterns depend on color connections between the measured final state jets and the unmeasured initial state partons. That said, we suspect that much can be learned from hadron-level measurements, even without isolating "pure" quark or gluon samples.

We present our final recommendations and conclusions in Sec. 1.9. The main take home message from this study is that, contrary to the standard lore, measurements at e^+e^- colliders are insufficient to constrain uncertainties in the final state shower. Therefore, gluon-enriched measurements at the LHC will be crucial to achieve robust quark/gluon discrimination.

1.2 What is a Quark/Gluon Jet?

As part of the 2015 Les Houches workshop, an attempt was made to define exactly what is meant by a "quark jet" or "gluon jet". Here are some suggested options for defining a quark jet, in order from most ill-defined to most well-defined. Related statement can be made for gluon jets.

A quark jet is...

- **A quark parton.** This definition (incorrectly) assumes that there is a one-to-one map between a jet and its initiating parton. Because it neglects the important role of additional radiation in determining the structure of a jet, we immediately dismiss this definition.

- **A Born-level quark parton.** This definition at least acknowledges the importance of radiative corrections to jet production, but it leaves open the question of how exactly to define the underlying Born-level process from an observed final state. (For one answer valid at the parton level, see flavored jet algorithms below.)

- **An initiating quark parton in a final state parton shower.** We suspect that this is the definition most LHC experimentalists have in mind. This is sometimes referred to as the "max-p_T" prescription (see further discussion in [33]), though it assumes that the parton shower history is meaningful, which may not be the case beyond the strongly-ordered or leading-logarithmic approximations. Because the parton shower is semi-classical, this definition neglects the impact of genuinely quantum radiative corrections as well as non-perturbative hadronization.

- **An eikonal line with baryon number 1/3 and carrying triplet color charge.** This is another semi-classical definition that attempts to use a well-defined limit of QCD to define quarks in terms of light-like Wilson lines. Philosophically, this is similar to the parton shower picture, with a similar concern about how to extrapolate this definition away from the strict eikonal limit.

- **A parton-level jet object that has been quark-tagged using an IRC safe flavored jet algorithm.** This is the strategy adopted in [34]. While this definition neglects the impact of hadronization, it does allow for the calculation of quark jet cross sections at all perturbative orders, including quantum corrections.

The unifying theme in the above definitions is that they try to identify a quark as an object unto itself, without reference to the specific final state of interest. However, it is well-known that a "quark" in one process may not look like a "quark" in other process, due to color correlations with the rest of the event, especially the initial state in pp collisions. The next definition attempts to deal with the process dependence in defining quarks.

- **A quark operator appearing in a hard matrix element in the context of a factorization theorem.** This is similar to the attitude taken in [35]. In the context of a well-defined cross section measurement, one can (sometimes) go to a limit of phase space where the hard production of short-distance quarks and gluons factorizes from the subsequent long-distance fragmentation. This yields a nice (gauge-covariant) operator definition of a quark jet. That said, even if a factorization theorem does exist for the measurement of interest, this definition is potentially ambiguous beyond leading power.

The definition we adopt for this study is inspired by the idea that one should think about quark/gluon tagging in the context of a specific measurement, but it tries to avoid relying on the presence of a factorization theorem.

- **A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion).** Here, the goal is to *tag* a phase space region as being quark-like, rather than try to determine a truth definition of a quark. This definition has the advantage of being explicitly tied to hadronic final states and to the discriminant variables of interest. Of course, the main challenge with this definition is how to determine the criterion that corresponds to successful quark enrichment. For that, we will have to rely to some degree on the other less well-defined notions of what a quark jet is.
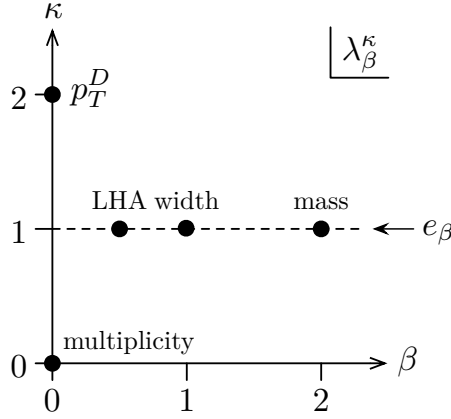
Fig. 1: Two parameter family of generalized angularities, adapted from [7]. The dots correspond to the five benchmark angularities used in thiss study. The horizontal line at $\kappa = 1$ corresponds to the IRC safe angularities.

It is worth emphasizing that this last definition separates the measurement process from the interpretation process. That is, one first measures hadron-level discriminant variables on a final state of interest, and only later does one interpret exactly what that discrimination accomplishes (either as quark/gluon separation or as some more sophisticated classification). In Sec. 1.9, we will recommend that the LHC experiments perform measurements of the generalized angularities λ_β^κ in well-defined hadron-level final states; it is then a separate interpretation step to infer something about parton-level quark/gluon radiation patterns from these hadron-level measurements.

1.3 Generalized Angularities

A wide variety of quark/gluon discriminants have been proposed (see [2] for an extensive catalog), but here we limit ourselves to a two-parameter family of generalized angularities [7], shown in Fig. 1. These are defined as

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \theta_i^\beta, \tag{5}$$

where i runs over the jet constituents, $z_i \in [0, 1]$ is a momentum fraction, and $\theta_i \in [0, 1]$ is a (normalized) angle to the jet axis. The parameters $\kappa \geq 0$ and $\beta \geq 0$ determine the momentum and angle weighting, respectively. For $\kappa = 1$, the generalized angularities are IRC safe and hence calculable in perturbation theory [17] (see also [16, 36, 37, 38, 39]). For general $\kappa \neq 1$, there are quasi-perturbative techniques based on generalized fragmentation functions [7] (see also [3, 40, 41, 42]). In our parton shower studies, we determine λ_β^κ using all constituents of a jet, though one could also consider using charged-particle-only angularities to improve robustness to pileup.

For our e^+e^- study, we cluster jets using the ee-variant of the anti-k_t algorithm [43], with $|\vec{p}|$-ordered winner-take-all recombination [17, 44, 45] to determine the jet axis $\hat{n}$. Unlike standard E-scheme recombination [46], the winner-take-all scheme yields a jet axis $\hat{n}$ that does not necessarily align with the jet momentum $\vec{p}$; this turns out to be a desirable feature for avoiding soft recoil effects [6, 17, 18, 47, 48]. We define

$$z_i \equiv \frac{E_i}{E_{\text{jet}}}, \qquad \theta_i \equiv \frac{\Omega_{i\hat{n}}}{R}, \tag{6}$$

where $\Omega_{i\hat{n}}$ is the opening angle to the jet axis and R is the jet radius (taken to be $R = 0.6$ by default). To translate our ee study to an eventual pp study (left to future work), one would use the standard pp version

of anti-k_t with p_T-ordered winner-take-all recombination, defining

$$z_i \equiv \frac{p_{Ti}}{\sum_{j \in \text{jet}} p_{Tj}}, \qquad \theta_i \equiv \frac{R_{i\hat{n}}}{R}, \tag{7}$$

where $R_{i\hat{n}}$ is the rapidity-azimuth distance to the jet axis.

By adjusting κ and β, one can probe different aspects of the jet fragmentation. We consider five benchmark values for (κ, β) indicated by the black dots in Fig. 1:

$$\begin{aligned}
(0,0) &= \text{hadron multiplicity}, \\
(2,0) &\Rightarrow p_T^D \text{ [4, 5] (specifically } \lambda_0^2 = (p_T^D)^2), \\
(1,0.5) &= \text{Les Houches Angularity (LHA)}, \\
(1,1) &= \text{width or broadening [18, 19, 20]}, \\
(1,2) &\Rightarrow \text{mass or thrust [21] (specifically } \lambda_2^1 \simeq m_{\text{jet}}^2 / E_{\text{jet}}).
\end{aligned} \tag{8}$$

Except for the LHA, these angularities (or their close cousins) have already been used in quark/gluon discrimination studies. The LHA has been included to have an IRC safe angularity that weights energies more heavily than angles, similar in spirit to the $\beta = 0.2$ value advocated in Ref. [6].

For the IRC safe case of $\kappa = 1$, there is an alternative version of the angularities based on energy correlation functions [6] (see also [48, 49]

$$\text{ecf}_\beta = \sum_{i < j \in \text{jet}} z_i z_j \theta_{ij}^\beta \simeq \lambda_\beta^1, \tag{9}$$

where equality holds in the extreme eikonal limit.[4] For the $e^+ e^-$ case, the pairwise angle θ_{ij} is typically normalized to the jet radius as $\theta_{ij} \equiv \Omega_{ij}/R$. To avoid a proliferation of curves, we will not show any results for ecf_β. We will also neglect quark/gluon discriminants that take into account azimuthal asymmetries within the jet, though observables like the covariance tensor [2] and 2-subjettiness [50, 51] can improve quark/gluon discrimination.

1.4 Classifier Separation

Since we will be testing many parton shower variants, we need a way to quantify quark/gluon separation power in a robust way that can easily be summarized by a single number. For that purpose we use classifier separation,

$$\Delta = \frac{1}{2} \int d\lambda \, \frac{\left(p_q(\lambda) - p_g(\lambda)\right)^2}{p_q(\lambda) + p_g(\lambda)}, \tag{10}$$

where p_q (p_g) is the probability distribution for the quark-tagged (gluon-tagged) sample as a function of the classifier λ. Here, $\Delta = 0$ corresponds to no discrimination power and $\Delta = 1$ corresponds to perfect discrimination power.

A more common way to talk about discrimination power is in terms of receiver operating characteristic (ROC) curves. At a point (q,g) on the ROC curve, where $q, g \in [0, 1]$, one can define a selection that yields q efficiency for quarks and g mistag rate for gluons, or equivalently, a $(1 - g)$ efficiency for gluons for a $(1 - q)$ mistag rate for quarks. To turn the ROC curve into a single number, it is common to report the gluon rejection rate at, say, 50% quark efficiency. Since we are more interested in understanding the relative performance between parton showers rather than the absolute performance, we will not show ROC curves here, though they can be easily derived from the p_q and p_g distributions. If one observable has an everywhere better ROC curve than another (i.e. it is Pareto optimal), then it will also

---

[4]This equality also relies on using a recoil-free axis choice $\hat{n}$ to define θ_i. Amusingly, $\lim_{\beta \to 0} \text{ecf}_\beta = (1 - \lambda_0^2)/2$ (i.e. $\kappa = 2$, $\beta = 0$), such that the $\beta \to 0$ limit of the IRC safe energy correlation functions corresponds to the IRC unsafe p_T^D.

have a larger Δ value. The converse is not true, however, since depending on the desired working point, a "bad" discriminant as measured by Δ might still be "good" by another metric. In that sense, Δ contains less information than the full ROC curve.

An alternative way to quantify discrimination power is through mutual information, which counts the number of "bits" of information gained from measuring a discriminant variable (see [7]). Given a sample with quark fraction $f \in [0, 1]$ and gluon fraction $(1 - f)$, the mutual information with the truth (a.k.a. the truth overlap) is

$$I(T; \Lambda) = \int d\lambda \left(f \, p_q(\lambda) \log_2 \frac{p_q(\lambda)}{p_{\text{tot}}(\lambda)} + (1 - f) \, p_g(\lambda) \log_2 \frac{p_g(\lambda)}{p_{\text{tot}}(\lambda)} \right), \tag{11}$$

where

$$p_{\text{tot}}(\lambda) = f \, p_q(\lambda) + (1 - f) \, p_g(\lambda). \tag{12}$$

The choice $f = 1/2$ was used in Ref. [7], though other f choices are plausible. Though we will not use mutual information in this study, it is amusing to note that the second derivative of $I(T; \Lambda)$ with respect to f is related to classifier separation as

$$-\frac{\log 2}{4} \frac{\partial^2 I(T; \Lambda)}{\partial f^2} \bigg|_{f=\frac{1}{2}} \equiv \Delta. \tag{13}$$

One advantage of Δ over $I(T; \Lambda)$ is that the integrand in Eq. (10) is easier to plot and interpret, since it tracks the fractional difference between the signal and background at a given value of λ.[5]

1.5 Idealized Quark/Gluon Discrimination

Our studies are based on the idealized case of discriminating $e^+e^- \to (\gamma/Z)^* \to u\bar{u}$ (quark-tagged) from $e^+e^- \to h^* \to gg$ (gluon-tagged). This example demonstrates the importance of final state evolution for quark/gluon discrimination, separate from initial state complications arising from pp collisions. Here, the flavor of the outgoing jets is fixed by the Lorentz structure of the production vertex. Of course, even in the quark-tagged sample, there can be "gluon jets", as in $e^+e^- \to u\bar{u}g$ where a hard gluon recoils against a nearly collinear $u\bar{u}$. This reflects the intrinsic ambiguity in defining jet flavor. To resolve that ambiguity, we rely on the production vertex alone to define the truth-level jet flavor.

In Fig. 2, we show hadron-level distributions of the LHA (i.e. $\lambda_{0.5}^1$) in the quark sample (p_q) and gluon sample (p_g), comparing the baseline settings of four different parton shower generators with a center-of-mass collision energy of $Q = 200$ GeV and jet radius $R = 0.6$. In the quark sample in Fig. 2a, there is relatively little variation between the generators, which is not surprising since all four programs have been tuned to match LEP data (though LEP never measured the LHA itself). Turning to the gluon sample in Fig. 2b, we see somewhat larger variations between the generators; this is expected since there is no data to directly constrain $e^+e^- \to gg$. We plot the integrand of classifier separation in Fig. 2c:

$$\frac{1}{2} \frac{\left(p_q(\lambda) - p_g(\lambda) \right)^2}{p_q(\lambda) + p_g(\lambda)}. \tag{14}$$

This quantity shows where in the LHA phase space the actual discrimination power lies, with large values of the integrand corresponding to places where the quark and gluon distributions are most dissimilar. Now we see considerable differences between the generators, reproducing the well-known fact that PYTHIA is more optimistic about quark/gluon separation compared to HERWIG. The prediction discrimination power from VINCIA and SHERPA are intermediate between these extremes and similar to each other.

---

[5]Another advantage of Δ over $I(T; \Lambda)$ arises when trying to assign statistical uncertainties to finite Monte Carlo samples. Since Δ is defined as a simple integral, one can use standard error propagation to assign uncertainties to Δ. By contrast, because of the logarithms in the $I(T; \Lambda)$ integrand, one has to be careful about a potential binning bias [7].
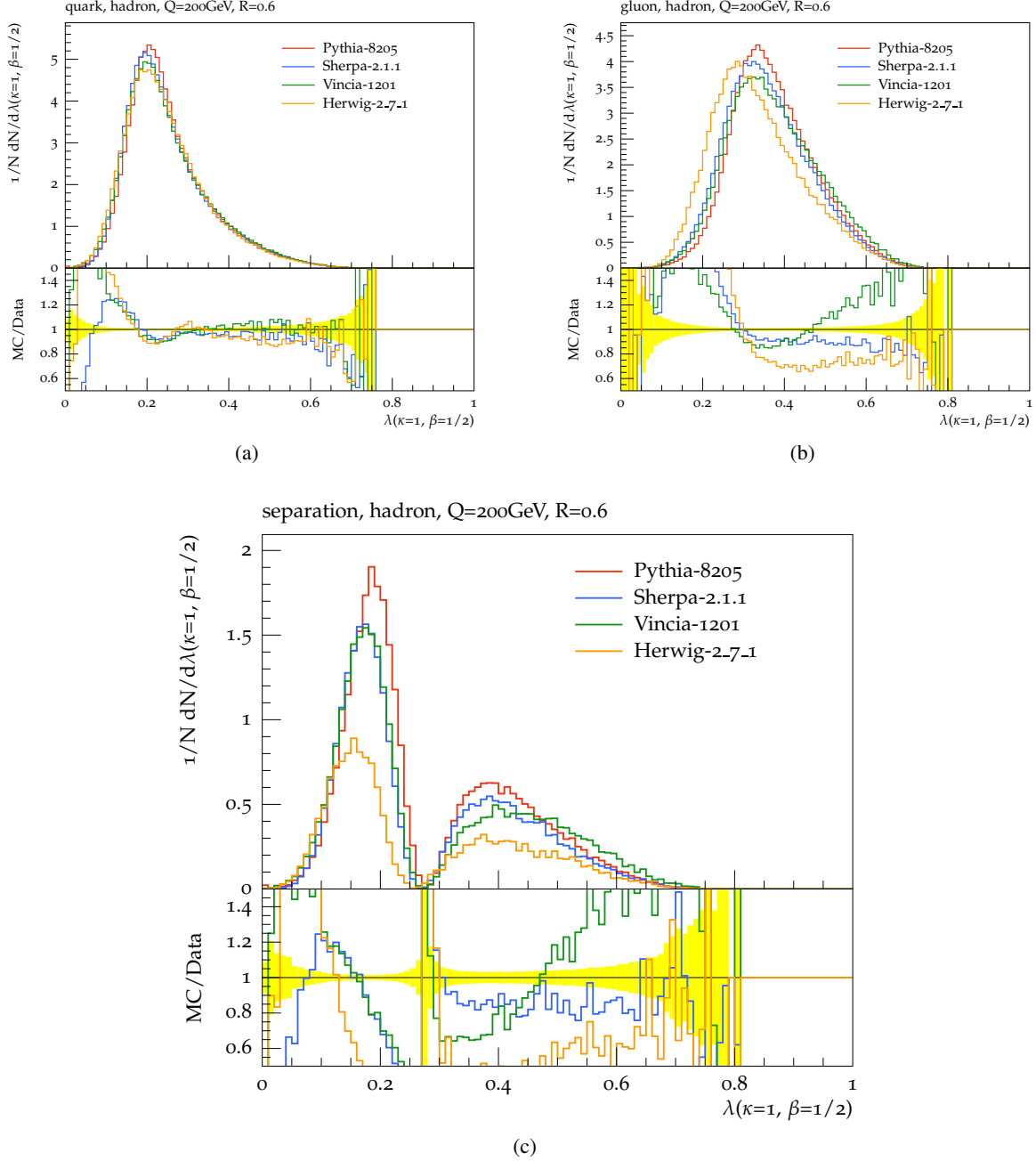
Fig. 2: Hadron-level distributions of the LHA for (a) the $e^+e^- \to u\bar{u}$ quark-tagged sample, (b) the $e^+e^- \to gg$ gluon-tagged sample, and (c) the classifier separation integrand in Eq. (14). Four parton shower generators—PYTHIA 8.205, VINCIA 1.201, HERWIG++ 2.7.1, SHERPA 2.1.1—are run at their baseline settings with center-of-mass energy $Q = 200$ GeV and jet radius $R = 0.6$. **(Is there a way to make the font size bigger in the yoda plots? Need to say ratio to Pythia. –jdt)**
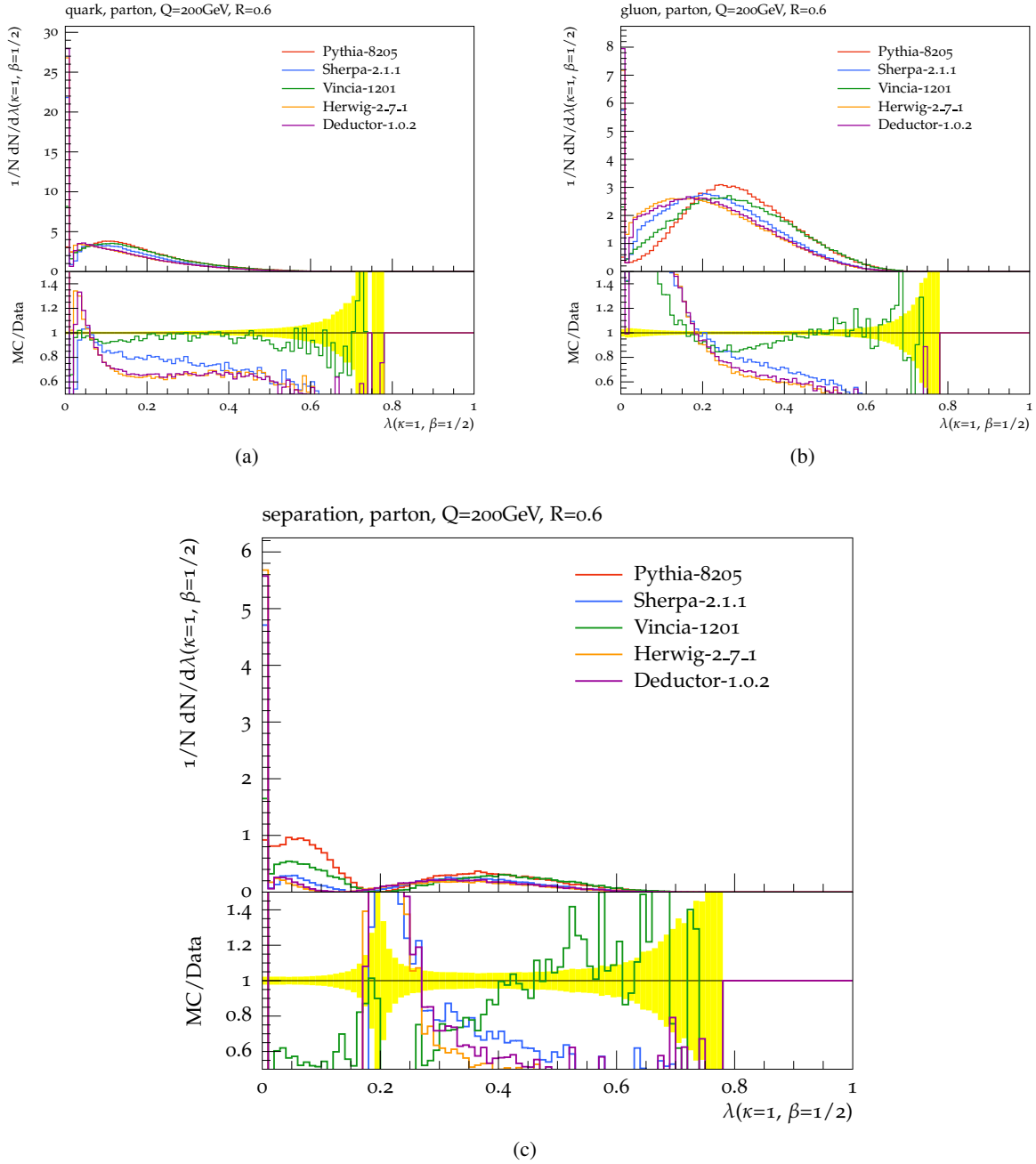
quark, parton, Q=200GeV, R=0.6

gluon, parton, Q=200GeV, R=0.6

(a)

(b)

separation, parton, Q=200GeV, R=0.6

(c)

Fig. 3: Same as Fig. 2, but at the parton level. Here, we have also included predictions from DEDUCTOR 1.0.2, which tend to agree with HERWIG++ 2.7.1.
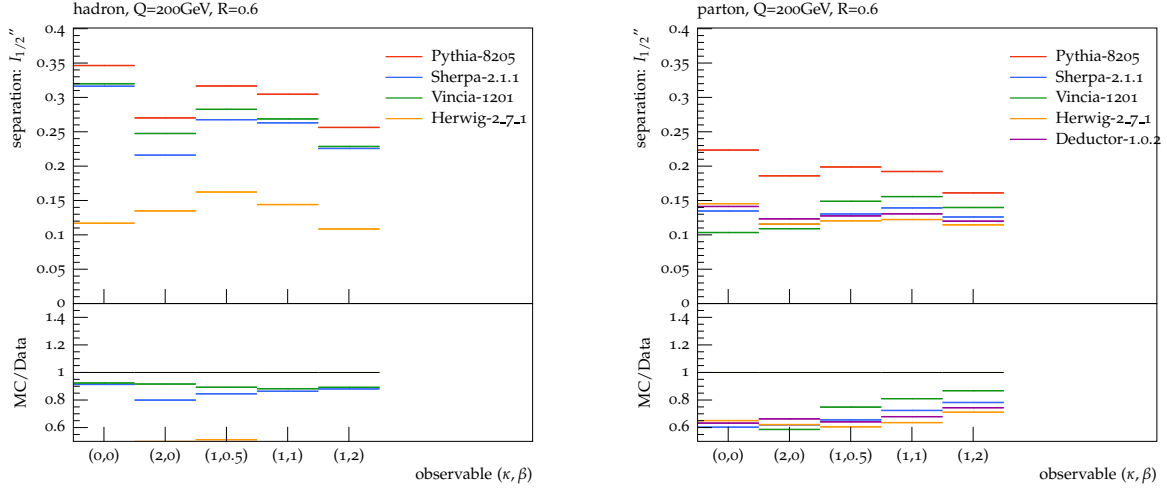
Fig. 4: Classifier separation Δ for the five different benchmark angularities in Eq. (8), determined from the various generators at (a) hadron level and (b) parton level. The first two columns correspond to IRC unsafe distributions (multiplicity and p_T^D), while the last three columns are the IRC safe angularities. The LHA is shown in the middle column.

One might expect that the differences between generators are due simply to their having different hadronization models. It seems, however, that the differences already appear at the parton level prior to hadronization, as shown in Fig. 3. The first thing one notices is that three of the generators—HERWIG, SHERPA, and DEDUCTOR—yield a large population of events where the perturbative shower generates no emissions, giving $\lambda_{0.5}^1 = 0$ such that non-zero values of the LHA are generated only by the hadronization model. By contrast, PYTHIA and VINCIA give overall larger values of the LHA from the perturbative shower alone. Since Fig. 2a shows that all generators give similar distributions for quark jets after hadronization, we conclude that understanding quark/gluon discrimination is a challenge at the interface between perturbative showering and non-perturbative hadronization.

To summarize the overall discrimination power, we integrate Eq. (14) to obtain the value of classifier separation Δ for the LHA. This is shown in Fig. 4, which also includes the four other benchmark angularities from Eq. (8). There is a rather large spread in predicted discrimination power between the generators, especially at hadron level in Fig. 4a. While such differences might be expected for IRC unsafe angularities (multiplicity and p_T^D) which depend on non-perturbative modeling, these differences persist even for the IRC safe angularities at parton level (see Fig. 4b). This suggests a more fundamental difference between the generators that is already present in the perturbative shower.

1.6 Parameter Dependence

Given the large absolute differences in discrimination power seen above, we next want to check if the parton shower generators exhibit similar or dissimilar trends as parameters are varied. We perform three parameter sweeps, using the boldface values below as defaults:

$$\text{Collision Energy} : Q = \{50, 100, \mathbf{200}, 400, 800\} \text{ GeV},$$
$$\text{Jet Radius} : R = \{0.2, 0.4, \mathbf{0.6}, 0.8, 1.0\}, \tag{15}$$
$$\text{Strong Coupling} : \alpha_s/\alpha_{s0} = \{0.8, 0.9, \mathbf{1.0}, 1.1, 1.2\},$$

where α_{s0} is the default value of the strong coupling, which is different between the generators (and sometimes different between different aspects of the same generator).
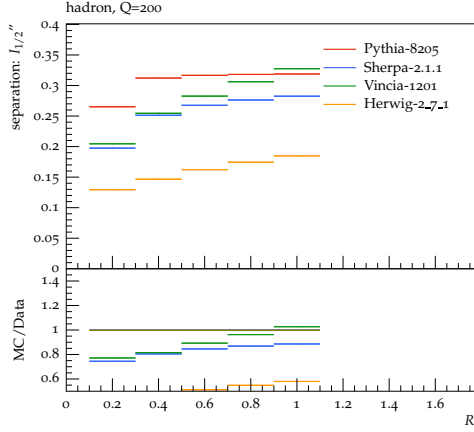
The resulting values of Δ for the LHA are shown in Fig. 5, at both the hadron level and parton level. There are number of surprising features in these plots. First, even for the IRC safe angularities, the effect of hadronization is rather large, both on the absolute scale of discrimination and the trends. The
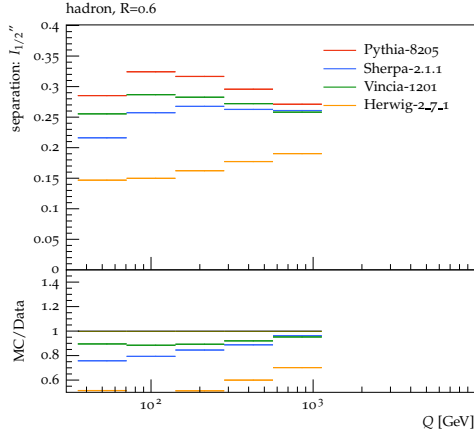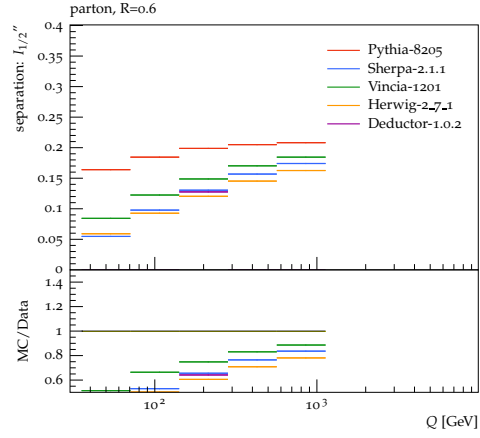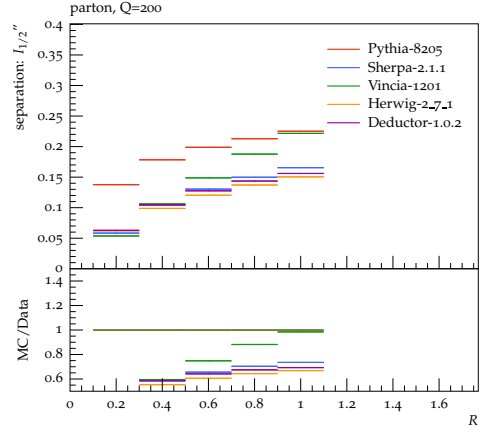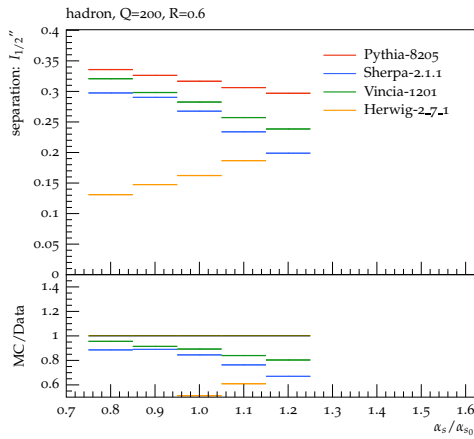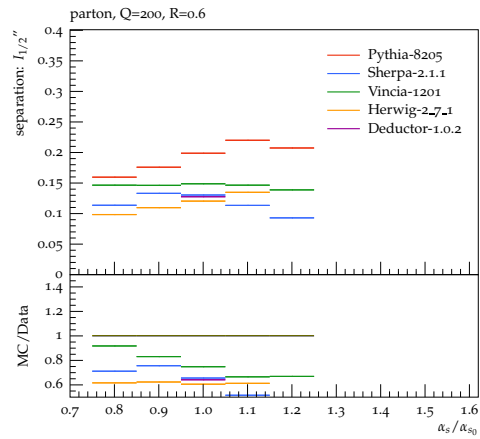
Fig. 5: Classifier separation Δ for the LHA, sweeping the collision energy Q (top row), jet radius R (middle row), and coupling constant α_s/α_{s0} (bottom row). Results are shown at hadron level (left column) and parton level (right column). **(We need to label the plots as saying that they are for the LHA. –jdt)**
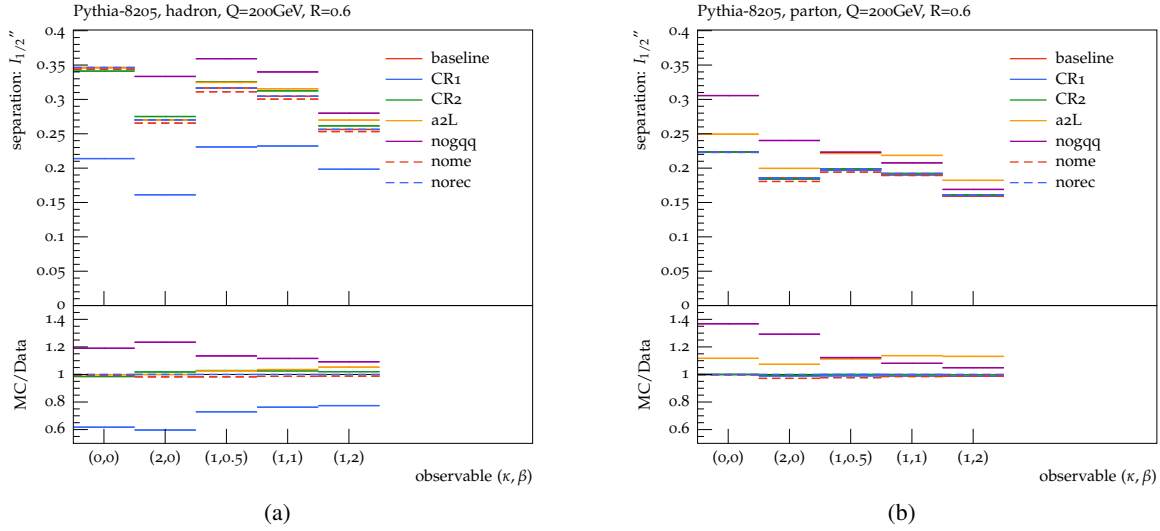
Fig. 6: Settings variations for PYTHIA 8.205. Shown are (a) hadron-level and (b) parton-level results for the classifier separation Δ derived from the LHA.

main exception to that trend is HERWIG, which does not exhibit much of an effect from hadronization. Second, the trends for sweeping Q and sweeping α_s seem contradictory. According to the perturbative logic in Ref. [6] at next-to-leading-logarithmic (NLL) accuracy, larger values of α_s should lead to improved discrimination power (which is seen for HERWIG but not the other generators). But larger values of Q correspond to smaller values of α_s, so increasing Q was expected to lead to worse discrimination power (seen by none of the generators at parton level, and somewhat present for all the generators *except* HERWIG at hadron level). Third, the trend for sweeping R is again backwards from the logic in Ref. [6], where larger jet radii were expected to yield worse separation power (seen by none of the generators).

We suspect that part of the issue is that Ref. [6] did not include non-perturbative hadronization corrections, which could potentially reverse the perturbative trends. That said, some of these confusions are present already at parton level, so further study is needed to assess whether additional perturbative effects beyond NLL could explain these features. At NLL accuracy, one does account for the fact that a jet can contain multiple perturbative emissions, but those emissions are treated as if they themselves do not radiate. Perhaps the fact that parton showers allow every emission to reradiate might explain these surprising quark/gluon discrimination trends.

1.7 Impact of Generator Settings

Formally, parton shower generators are only accurate to modified leading logarithmic (MLL) accuracy, though they include physically important effects like energy/momentum conservation and matrix element corrections that go beyond MLL. We can assess the impact of these higher-order effects by changing the baseline parameter settings in each parton shower generator.

Because each generator is different, we cannot always make the same changes for each generator. Similarly, the spread in discrimination power shown below should *not* be seen as representing the intrinsic uncertainties in the shower, since many of these changes we explore are not physically motivated. The goal of these plots is to demonstrate possible areas where small parameter changes could have a large impact on quark/gluon discrimination. Ultimately, collider data and higher-order calculations will be essential for understanding the origin of quark/gluon differences. In all cases, we show both hadron-level and parton-level results, even if a setting is only expected to have an impact at the hadron level.

Our PYTHIA baseline is based on the Monash 2013 tune, with parameters described in Ref. [52]. In Fig. 6, we consider the following PYTHIA variations: **(We should probably give these nicer names.**
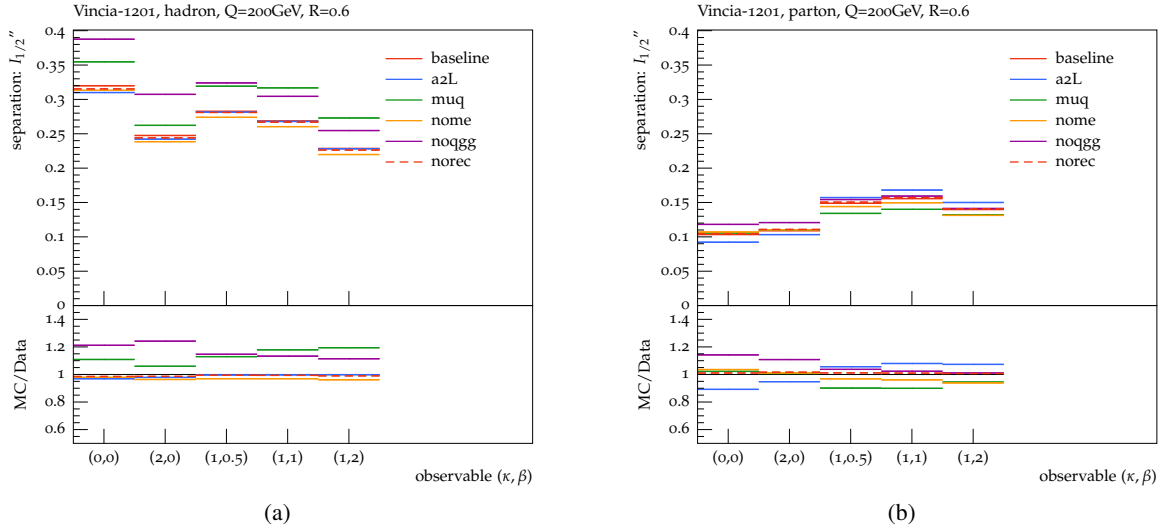
Fig. 7: Same as Fig. 6, but for VINCIA 1.201.

Also, I tried to put these in a more logical order than alphabetical, and we should probably use the same order in the figures. –jdt)

- PYTHIA-NOGQQ. While the dominant gluon splitting in the parton shower is $g \to gg$, PYTHIA also generates the subleading $g \to q\bar{q}$ splittings by default. This variation turns off $g \to q\bar{q}$, which makes gluon jets look more gluon-like, thereby increasing the separation power.

- PYTHIA-A2L. The default PYTHIA setting is to use 1-loop running for α_s. This variation turns on 2-loop running for α_s, which has a small (beneficial) effect at parton level which is washed out at hadron level.

- PYTHIA-NOME. The first emission in PYTHIA is improved by applying a matrix element correction, but this variation turns those corrections off. This has relatively little effect in our study, however, since there is no matrix element correction available for $h^* \to gg$.

- PYTHIA-CR1. Often, one thinks of color reconnection as being primarily important for hadron colliders, but even at a lepton collider, color reconnection will change the Lund strings used for hadronization. This variation uses an alternative color reconnection model compared to the default **(Is is alternative or just off? –jdt)**. At parton level, this variation has no effect as expected. At hadron level, this variation considerably decreases quark/gluon separation compared to the baseline.

- PYTHIA-CR2. **(This variation seems to not be useful, right? Remove? –jdt)**

- PYTHIA-NOREC. **(This setting seems to not be useful, right? Remove? –jdt)**

The most surprising PYTHIA effect is related to color reconnection, which will also turn out to be important for the HERWIG generator below.

Our VINCIA baseline is based on the "jeppsson5" tune. While VINCIA has NLO matrix elements for $e^+e^- \to q\bar{q}$, it does not have them for $e^+e^- \to gg$, so we will use LO VINCIA throughout. In Fig. 7, we consider the following VINCIA variations:

- VINCIA-NOGQQ. Like for PYTHIA, this variation turns off $g \to q\bar{q}$, leading to the expected increase in separation power.

- VINCIA-A2L. Like for PYTHIA, this variation switches from 1-loop to 2-loop α_s running, yielding a modest parton-level difference and almost no hadron-level difference.

- VINCIA-NOME. By default, each $2 \to 3$ antenna in VINCIA has an associated matrix element correction factor. Since the antenna are already rather close to the true matrix elements, this has a modest effect on quark/gluon discrimination power.
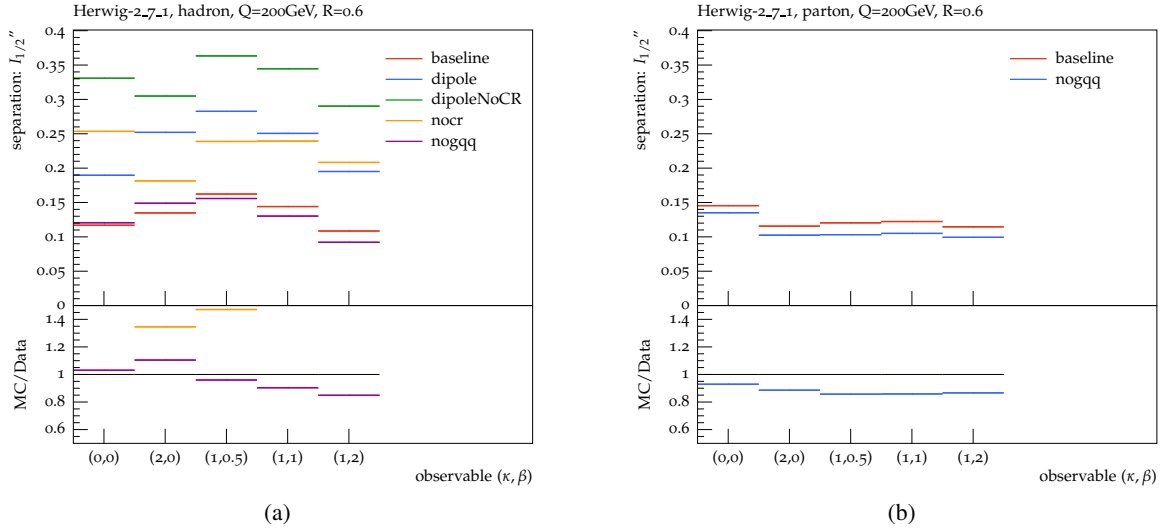
Fig. 8: Same as Fig. 6, but for HERWIG 2.7.1. **(We are missing some Herwig parton level curves. –jdt)**

- VINCIA-MUQ. By default, the α_s scale used in VINCIA is determined by the transverse momentum of the corresponding antenna. In this variation, the α_s scale is set to half the invariant mass of the mother antenna. This slightly decreases the discrimination power at hadron level, but increases the discrimination power at parton level, again showing the complicated interplay of perturbative and non-perturbative effects.

- VINCIA-NOREC. **(This setting seems to not be useful, right? Remove? –jdt)**

Since VINCIA and PYTHIA share the same underlying engine, it is not surprising that they exhibit similar behaviors as parameters are changed. The biggest surprise is the way that changing the α_s scale for the antenna can lead to different trends at parton and hadron level.

Our baseline HERWIG uses the default settings from version 2.7.1. In Fig. 8, we consider the following HERWIG variations:

- HERWIG-NOGQQ. Turning off $g \to q\bar{q}$ splittings in HERWIG has the reverse behavior as seen in PYTHIA and VINCIA, leading to slightly worse discrimination power, though the effect is modest.

- HERWIG-NOCR. Like PYTHIA, HERWIG has the option to turn off color reconnections. Unlike for PYTHIA, this variation for HERWIG gives a rather dramatic improvement in quark/gluon discrimination power. We think this arises since color reconnection in HERWIG allows any color/anticolor pair to reconnect, even if they arose from an initially color octet configuration. By turning off color reconnection, the gluons look more octet-like, explaining the improvement seen.

- HERWIG-DIPOLE. The default HERWIG shower is based on angular ordering, but this variation uses a diplole shower instead. Since this shower is closer in spirit to PYTHIA, one expects it to yield improved quark/gluon discrimination, which it does.

- HERWIG-DIPOLENOCR. By switching to the dipole shower without color reconnections, one can make HERWIG look rather similar to PYTHIA in terms of predicted discrimination performance. Therefore, LHC data will be essential to understand which shower structure is closer to nature.

The importance of color reconnections in HERWIG is a big surprise from this study, motivating future detailed studies into which color reconnection models are most realistic when compared to data.

Finally, our baseline SHERPA implementation uses matrix element corrections for the first two emissions with CKKW-style matching [53]. In Fig. 9, we consider the following SHERPA variations:

- SHERPA-NOQQ. Turning off $g \to q\bar{q}$ splittings in SHERPA has a negligible effect at parton level, but it leads to a large jump in discrimination performance at hadron level, again due to an interplay the perturbative shower and non-perturbative hadronization
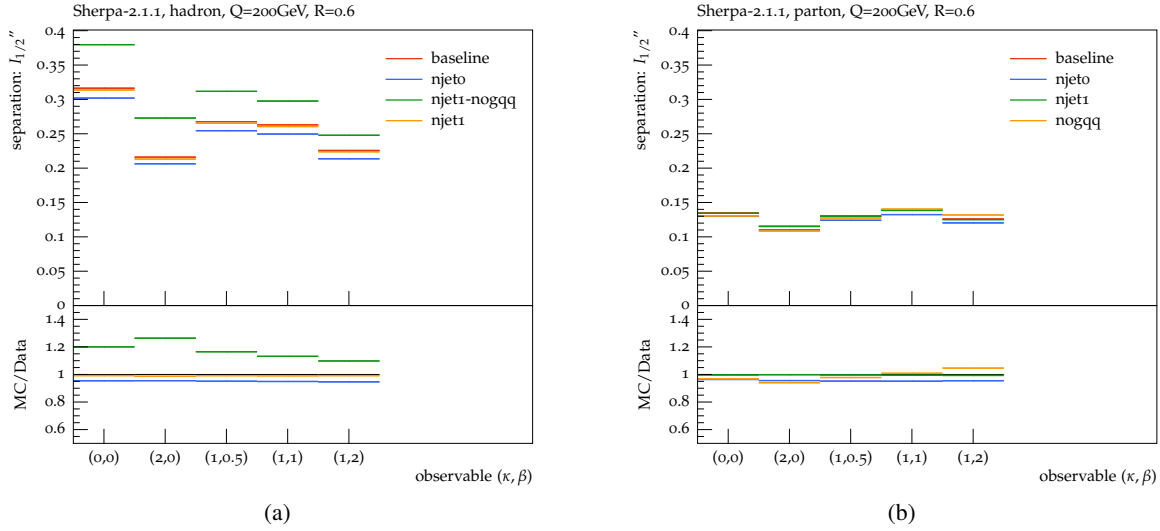
Fig. 9: Same as Fig. 6, but for SHERPA 2.1.1. **(Confused about the naming scheme differences for Sherpa at parton level and hadron level. –jdt)**

- SHERPA-NJET1. This variation only performs CKKW matching for the first emission, leading to negligible changes in the discrimination performance.
- SHERPA-NJET0. Turning off all matrix element corrections in SHERPA slightly decreases the predicted quark/gluon discrimination power, in agreement with the behavior in PYTHIA and VINCIA.

Within SHERPA, matrix element corrections appear to have a very small effect at parton level. The large changes seen at hadron level from turning off $g \to q\bar{q}$ splittings motivates further investigations into the shower/hadronization interface.

1.8 Looking Towards the LHC

It is clear from our e^+e^- study that quark/gluon radiation patterns face considerable theoretical uncertainties, as seen in the differing behaviors of parton shower generators. This is true even accounting only for final state physics effects, so additional initial state complications can only increase the uncertainties faced in pp collisions. Beyond just the application to quark/gluon tagging, this is an important challenge for any analysis that uses jets. For example, a proper experimental determination of jet energy scale corrections requires robust parton shower tools that correctly model effects like out-of-cone radiation. Eventually, one would like to perform improved analytic calculations to address these radiation pattern uncertainties. In the near term, though, measurements from the LHC will be essential for improving the parton shower modeling of jets.

At the LHC, there is no way to isolate pure samples of quark or gluon jets, but one can isolate quark/gluon-enriched samples, as defined by the flavor label of the jet in the corresponding Born-level partonic process. As shown in Fig. 10, the Born-level jet in $W/Z/\gamma$ + jet is more than 70% quark enriched over the entire jet p_T range of interest. For jets softer than around 200 GeV, the Born-level jet in dijets or H + jet is more than 60% gluon enriched, with that fraction decreasing as the jet p_T increases. More sophisticated enrichment procedures are described in Ref. [35].

In principle, could could try to "diagonalize" some combination of vector boson plus jet and dijet samples in order to define separate quark or gluon samples (see e.g. [8]). In the spirit of Sec. 1.2, though, we think it is more beneficial for the LHC experiments to perform process-specific measurements without trying to directly determine their quark and gluon composition. For example, instead of quark/gluon separation, one can ask the more well-defined question about whether one can tell "the jet in Z plus jets" (quark-enriched) apart from "the jet in dijets" (gluon-enriched). This process-based strategy is also
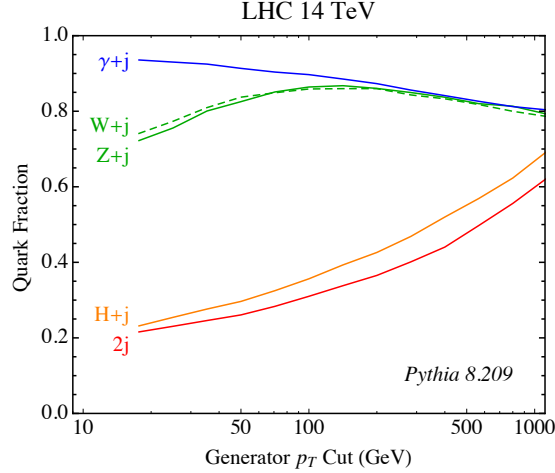
Fig. 10: Quark fraction of jets at parton level, as defined by the generator-level parton flavor.

helpful for sidestepping the known process dependance of defining quarks and gluons at the LHC, where color correlations to the rest of the event impede a universal definition of quark and gluon jets.

Already, one would learn a lot from unfolded measurements at the LHC of the generalized angularities in a variety of quark/gluon enriched samples. Even detector-level measurement comparing a wide variety of generators (and generator modifications) would help in understanding how to improve quark/gluon modeling. The five benchmark angularities in Eq. (8) probe both the perturbative and non-perturbative structure of jets and are therefore a good starting point for a more comprehensive quark/gluon jet shape analysis. In this spirit, we are encouraged by the track multiplicity study of Ref. [13], though for parton shower tuning is it is important to have measurements not only of jet shape averages but also of the full jet shape probability distributions.

Finally, we think it would be useful to perform LHC jet shape measurements after jet grooming (see e.g. [54, 55, 56, 57]). Often jet grooming is described as a pileup mitigation strategy, but even independent of pileup, grooming modifies the observed jet radiation patterns in ways that are interesting from the quark/gluon perspective. Using techniques from Refs. [59, 58], one can calculate the quark/gluon discrimination power for angularities after jet grooming. Those calculations generically predict that groomed jet shapes should have reduced quark/gluon discrimination power compared to ungroomed jet shapes. That said, because techniques like soft drop [58] remove soft radiation from jets, they tend to reduce the process dependence in quark/gluon radiation patterns, and may therefore yield a better theoretical definition for quark and gluon jets.

1.9 Summary and Recommendations

By measuring the substructure of jets, it is possible to gain information about the quark/gluon composition of a jet sample. The challenge we have identified in this study is that the precise radiation pattern of quark and gluon jets is poorly understood, in the sense that parton shower generators give rather different predictions for absolute quark/gluon discrimination power as well as relative trends as a function of the jet kinematics. At the moment, analytic calculations are not at a level of accuracy where they can offer any useful guide. Therefore, LHC data is the best hope to constrain quark/gluon radiation patterns and enable quark/gluon discrimination to be a robust tool.

In terms of specific measurements that should be highest priority for ATLAS and CMS, our study has not revealed a silver bullet. Rather, all of the generalized angularities studies showed similar levels of disagreement between generators, so a systematic LHC study of even one observable is likely to offer crucial new information. What is essential is to make measurements at multiple jet p_T scales with

multiple jet radii R in multiple different jet samples. Unfolded distributions would likely be most useful for parton shower tuning, but even detector-level measurements compared to detector-simulated parton showers would help spot troubling trends.

The key lesson to parton shower authors is that, contrary to some standard lore, LEP data *does not* constrain all of the relevant aspects of the final state parton shower. While we have extensive information about quark jet radiation patterns from LEP event shapes, gluon jet radiation patterns are largely unconstrained. This has important implications for parton shower tuning strategies, since LHC data can and should be used to adjust final state shower parameters. For example, the ATLAS A14 tune of PYTHIA has a 10% lower value of α_s in the final state shower compared to the Monash tune, which yields better agreement with charge particle multiplicity distributions [13]. However, A14 has not been tested on LEP event shapes, suggesting that a global tuning strategy is needed.

Based on this study, we have identified three aspects of the final state parton shower that deserve closer scrutiny.

- *Gluon splitting to quarks*. Some of the largest difference between generators came from turning on and off the $g \to q\bar{q}$ process. While PYTHIA, VINCIA, and SHERPA suggest that (unphysically) turning off $g \to q\bar{q}$ would improve quark/gluon separation, HERWIG (and the analytic calculation in Ref. [6]) suggests the opposite conclusion. Beyond quark/gluon discrimination, it would be helpful to think of other contexts where $g \to q\bar{q}$ might play an important role.

- *Color reconnection in the final state*. Color reconnection is often thought of as an issue mainly at hadron colliders, but we have seen that it can have an impact in e^+e^- collisions as well. This is particularly the case with the default color reconnection model in HERWIG, since it allows the reconnection of color/anticolor lines even if they originally came from an octet configuration.

- *Reconsidering α_s defaults*: In the context of parton shower tuning, the value of α_s used internally within a code need not match the world average value, since higher-order effects not captured by the shower can often be mimicked by adjusting α_s. That said, one has to be careful whether a value of α_s tuned for one process is really appropriate for another. For example, PYTHIA uses a relatively large value of α_s in its final state shower, which allows it to match LEP event shape data. The same value of α_s probably also leads to too much radiation within gluon jets, though.

Finally, we want to emphasize that despite the uncertainties currently present in parton shower generators, parton showers in particular (and QCD resummation techniques more generally) will be essential for understanding quark/gluon discrimination. Fixed-order QCD calculations cannot reliably probe the very soft and very collinear structure of jets, which is precisely where valuable information about quark/gluon radiation patterns reside. Given the ubiquity and value of parton shower generators, improving the understanding of quark/gluon discrimination will assist every jet study at LHC.

Acknowledgments

References

[1] J. Gallicchio and M. D. Schwartz, *Phys. Rev. Lett.* **107** (2011) 172001, [1106.3076].

[2] J. Gallicchio and M. D. Schwartz, *JHEP* **04** (2013) 090, [1211.7038].

[3] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, *Phys. Rev. Lett.* **110** (2013), no. 21 212001, [1209.2421].

[4] F. Pandolfi and D. Del Re,, *Search for the Standard Model Higgs Boson in the $H \to ZZ \to llqq$ Decay Channel at CMS*. PhD thesis, Zurich, ETH, 2012.

[5] S. Chatrchyan *et. al.*,, **CMS** Collaboration *JHEP* **04** (2012) 036, [`1202.1416`].

[6] A. J. Larkoski, G. P. Salam, and J. Thaler, *JHEP* **06** (2013) 108, [`1305.0007`].

[7] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, *JHEP* **11** (2014) 129, [`1408.3122`].

[8] G. Aad *et. al.*,, **ATLAS** Collaboration *Eur. Phys. J.* **C74** (2014), no. 8 3023, [`1405.6583`].

[9] G. Aad *et. al.*,, **ATLAS** Collaboration *Eur. Phys. J.* **C75** (2015) 17, [`1406.0076`].

[10] V. Khachatryan *et. al.*,, **CMS** Collaboration *Eur. Phys. J.* **C75** (2015), no. 2 66, [`1410.3153`].

[11] G. Aad *et. al.*,, **ATLAS** Collaboration *JHEP* **12** (2015) 055, [`1506.00962`].

[12] V. Khachatryan *et. al.*,, **CMS** Collaboration *Phys. Rev.* **D92** (2015), no. 3 032008, [`1506.01010`].

[13] G. Aad *et. al.*,, **ATLAS** Collaboration `1602.00988`.

[14] C. F. Berger, T. Kucs, and G. F. Sterman, *Phys. Rev.* **D68** (2003) 014012, [`hep-ph/0303051`].

[15] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, and J. Virzi, *Phys. Rev.* **D79** (2009) 074017, [`0807.0234`].

[16] S. D. Ellis, C. K. Vermilion, J. R. Walsh, A. Hornig, and C. Lee, *JHEP* **11** (2010) 101, [`1001.0014`].

[17] A. J. Larkoski, D. Neill, and J. Thaler, *JHEP* **04** (2014) 017, [`1401.2158`].

[18] S. Catani, G. Turnock, and B. R. Webber, *Phys. Lett.* **B295** (1992) 269–276.

[19] P. E. L. Rakow and B. R. Webber, *Nucl. Phys.* **B191** (1981) 63.

[20] R. K. Ellis and B. R. Webber, *Conf. Proc.* **C860623** (1986) 74.

[21] E. Farhi, *Phys. Rev. Lett.* **39** (1977) 1587–1588.

[22] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay, Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla, *ArXiv Physics e-prints* (Mar., 2007) [`physics/0703039`].

[23] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *Comput. Phys. Commun.* **191** (2015) 159–177, [`1410.3012`].

[24] W. T. Giele, L. Hartgring, D. A. Kosower, E. Laenen, A. J. Larkoski, J. J. Lopez-Villarejo, M. Ritzmann, and P. Skands, *PoS* **DIS2013** (2013) 165, [`1307.1060`].

[25] J. Bellm *et. al.*, `1310.6877`.

[26] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *JHEP* **02** (2009) 007, [`0811.4622`].

[27] Z. Nagy and D. E. Soper, *JHEP* **06** (2014) 097, [`1401.6364`].

[28] S. Höche and S. Prestel, *Eur. Phys. J.* **C75** (2015), no. 9 461, [`1506.05057`].

[29] A. Heister *et. al.*,, **ALEPH** Collaboration *Eur. Phys. J.* **C35** (2004) 457–486.

[30] J. Abdallah *et. al.*,, **DELPHI** Collaboration *Eur. Phys. J.* **C29** (2003) 285–312, [`hep-ex/0307048`].

[31] P. Achard *et. al.*,, **L3** Collaboration *Phys. Rept.* **399** (2004) 71–174, [`hep-ex/0406049`].

[32] G. Abbiendi *et. al.*,, **OPAL** Collaboration *Eur. Phys. J.* **C40** (2005) 287–316, [`hep-ex/0503051`].

[33] A. Buckley and C. Pollard, `1507.00508`.

[34] A. Banfi, G. P. Salam, and G. Zanderighi, *Eur. Phys. J.* **C47** (2006) 113–124, [`hep-ph/0601139`].

[35] J. Gallicchio and M. D. Schwartz, *JHEP* **10** (2011) 103, [`1104.1175`].

[36] A. J. Larkoski and J. Thaler, *JHEP* **09** (2013) 137, [`1307.1699`].

[37] A. J. Larkoski, I. Moult, and D. Neill, *JHEP* **09** (2014) 046, [`1401.4458`].

[38] M. Procura, W. J. Waalewijn, and L. Zeune, *JHEP* **02** (2015) 117, [`1410.6483`].

[39] A. Hornig, Y. Makris, and T. Mehen, `1601.01319`.

[40] W. J. Waalewijn, *Phys. Rev.* **D86** (2012) 094030, [`1209.3019`].

[41] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, *Phys. Rev. Lett.* **111** (2013) 102002, [`1303.6637`].

[42] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, *Phys. Rev.* **D88** (2013) 034030, [`1306.6630`].

[43] M. Cacciari, G. P. Salam, and G. Soyez, *JHEP* **04** (2008) 063, [`0802.1189`].

[44] D. Bertolini, T. Chan, and J. Thaler, *JHEP* **04** (2014) 013, [`1310.7584`].

[45] G. Salam, *Unpublished*.

[46] G. C. Blazey *et. al.*, in *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pp. 47–77, 2000. `hep-ex/0005012`.

[47] Y. L. Dokshitzer, A. Lucenti, G. Marchesini, and G. P. Salam, *JHEP* **01** (1998) 011, [`hep-ph/9801324`].

[48] A. Banfi, G. P. Salam, and G. Zanderighi, *JHEP* **03** (2005) 073, [`hep-ph/0407286`].

[49] M. Jankowiak and A. J. Larkoski, *JHEP* **06** (2011) 057, [`1104.1646`].

[50] J. Thaler and K. Van Tilburg, *JHEP* **03** (2011) 015, [`1011.2268`].

[51] J. Thaler and K. Van Tilburg, *JHEP* **02** (2012) 093, [`1108.2701`].

[52] P. Skands, S. Carrazza, and J. Rojo, *Eur. Phys. J.* **C74** (2014), no. 8 3024, [`1404.5630`].

[53] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *JHEP* **11** (2001) 063, [`hep-ph/0109231`].

[54] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Phys. Rev. Lett.* **100** (2008) 242001, [`0802.2470`].

[55] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev.* **D80** (2009) 051501, [`0903.5081`].

[56] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Phys. Rev.* **D81** (2010) 094023, [`0912.0033`].

[57] D. Krohn, J. Thaler, and L.-T. Wang, *JHEP* **02** (2010) 084, [`0912.1342`].

[58] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *JHEP* **05** (2014) 146, [`1402.2657`].

[59] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *JHEP* **09** (2013) 029, [`1307.0007`].