

Systematics of quark/gluon tagging

Marat Freytsis,^a Philippe Gras,^b Deepak Kar,^c Lief Lönnblad,^d Simon Plätzer,^{e,f} Andrzej Siódmok,^g Peter Skands,^h Davison Soper,^a Gregory Soyez,^{i,†} and Jesse Thaler^{j,†}

^a*Institute of Theoretical Science, University of Oregon, Eugene, OR 97403-5203, USA*

^b*CEA/IRFU Saclay, France*

^c*School of Physics, University of the Witwatersrand, Johannesburg, Wits 2050, South Africa*

^d*Department of Astronomy and Theoretical Physics, Lund University, Sweden*

^e*Institute for Particle Physics Phenomenology, University of Durham, Durham DH1 3LE, UK*

^f*Particle Physics Group, School of Physics and Astronomy, University of Manchester, Manchester M13 9PL, UK*

^g*CERN, TH Department, CH1211 Geneva, Switzerland*

^h*School of Physics and Astronomy, Monash University, VIC-3800, Australia*

ⁱ*IPhT, CEA Saclay, CNRS UMR 3681, F-91191 Gif-sur-Yvette, France*

^j*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*

E-mail: freytsis@uoregon.edu, philippe.gras@cern.ch, deepak.kar@cern.ch,
leif.lonnblad@thep.lu.se, simon.platzer@durham.ac.uk, andrzej@cern.ch,
peter.skands@monash.edu, soper@uoregon.edu, soyez@lpthe.jussieu.fr,
jthaler@mit.edu

ABSTRACT: By measuring the substructure of a jet, one can assign it a “quark” or “gluon” tag. In the eikonal (double-logarithmic) limit, quark/gluon discrimination is determined solely by the color factor of the initiating parton (C_F versus C_A). In this paper, we confront the challenges faced when going beyond this leading-order understanding, using both parton shower generators and first-principles calculations to assess the impact of higher-order perturbative and nonperturbative physics. Working in the idealized context of electron-positron collisions, where one can define a proxy for quark and gluon jets based on the Lorentz structure of the production vertex, we find a fascinating interplay between perturbative shower effects and nonperturbative hadronization effects. Turning to proton-proton collisions, we highlight a core set of measurements that would constrain current uncertainties in quark/gluon tagging and improve the overall modeling of jets at the Large Hadron Collider.

[†]Coordinators for an initial version of this study appearing in Ref. [1].

Contents

1	Overview	1
2	What is a quark/gluon jet?	5
3	Quantifying tagging performance	7
3.1	Generalized angularities	7
3.2	Classifier separation	10
4	Analytic quark/gluon predictions	12
4.1	Casimir scaling at LL	12
4.2	NLL resummation	14
4.3	Nonperturbative shape function	15
5	Idealized quark/gluon discrimination	17
5.1	Baseline analysis	18
5.2	Parameter dependence	22
5.3	Impact of generator settings	24
5.4	Testing the shape function	30
6	Quark/gluon tagging at the LHC	30
6.1	Defining enriched samples	30
6.2	Baseline analysis	32
6.3	Parameter dependence	32
7	Summary and recommendations	32

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 [2–9], and there is a growing catalog of quark/gluon studies at the Large Hadron Collider (LHC) [10–15].¹

¹For an incomplete selection of earlier proposed discriminants, see Refs. [16–21].

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 $C_A = 3$, and quark/gluon tagging performance is simply a function of C_A/C_F . As we will see, quark/gluon discrimination performance is highly sensitive to subleading perturbative effects beyond the eikonal limit, such as $g \rightarrow q\bar{q}$ splittings and color coherence, as well as to nonperturbative 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 paper is to highlight these uncertainties in quark/gluon tagging, using both parton shower generators and first-principles calculations. We start in the idealized context of electron-positron collisions, where one can study final-state quark/gluon radiation patterns in the absence of initial-state complications. Here, we find modest differences in the predicted cross sections for quark/gluon discriminants, which then translate to large differences in the predicted quark/gluon separation power. Motivated by these uncertainties, we propose a set of LHC measurements that should help improve the modeling of jets in general and quark/gluon tagging in particular. A summary and outline of this paper follows.

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. 2, one can indeed create a well-defined quark/gluon tagging procedure based unambiguous hadron-level measurements. In this way, even if what one means by “quark” or “gluon” is based on a naive or ambiguous concept (like Born-level cross sections or eikonal limits), quark/gluon discrimination is still a well-defined technique for enhancing desired signals over unwanted backgrounds.

In order to quantify quark/gluon discrimination power, there are a wide range of possible quark/gluon discriminants and a similarly large range of performance metrics, both discussed in Sec. 3. As a concrete set of discriminants, we consider the generalized angularities λ_β^κ [8] (see also [22–25]),

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \theta_i^\beta, \quad (1.1)$$

with the notation to be explained in Sec. 3.1. We consider five different (κ, β) working points, which roughly map onto five variables in common use in the literature:

$$\begin{array}{ccccc} (0, 0) & (2, 0) & (1, 0.5) & (1, 1) & (1, 2) \\ \text{multiplicity} & p_T^D & \text{LHA} & \text{width} & \text{mass} \end{array} \quad (1.2)$$

Here, multiplicity is the hadron multiplicity within the jet, p_T^D was defined in Refs. [5, 6], LHA refers to the “Les Houches Angularity” (named after the workshop venue where this study was initiated [1]), width is closely related to jet broadening [26–28], and mass is closely related

to jet thrust [29]. To quantify discrimination performance, we focus on classifier separation (a default output of TMVA [30]):

$$\Delta = \frac{1}{2} \int d\lambda \frac{(p_q(\lambda) - p_g(\lambda))^2}{p_q(\lambda) + p_g(\lambda)}, \quad (1.3)$$

where p_q (p_g) is the probability distribution for λ in a generated quark jet (gluon jet) sample. This and other potential performance metrics are discussed in Sec. 3.2.

(Need to add refs here. –jdt) To gain a baseline analytic understanding, we use resummed calculations in Sec. 4 to provide a first-order approximation for quark/gluon radiation patterns. For $\kappa = 1$, the generalized angularities are infrared and collinear (IRC) safe, and therefore calculable in (resummed) perturbation theory. At leading-logarithmic (LL) accuracy, the IRC safe angularities satisfy a property called Casimir scaling, and the resulting classifier separation Δ is a universal function of C_A/C_F , independent of the value of β . At present, the distributions for generalized angularities are known to next-to-LL (NLL) accuracy [7, 8], though without the resummation of non-global logarithms [] nor the resummation of jet radius logs []. These NLL calculations are effectively at parton-level, so to obtain hadron-level distributions, we estimate the impact of nonperturbative effects using shape functions [].

To gain a more realistic understanding with a full hadronization model, we use parton shower generators in Sec. 5 to predict quark/gluon separation power. In an idealized setup with e^+e^- collisions, we can use the following processes as proxies for quark and gluon jets:

$$\text{“quark jets”} : e^+e^- \rightarrow (\gamma/Z)^* \rightarrow u\bar{u}, \quad (1.4)$$

$$\text{“gluon jets”} : e^+e^- \rightarrow h^* \rightarrow gg, \quad (1.5)$$

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 define truth-level quarks and gluons labels without reference to the final state.² We compare seven different parton shower generators both before hadronization (“parton level”) and after hadronization (“hadron level”):

- PYTHIA 8.205 [31, 32] **(Go to latest version? –jdt)**,
- HERWIG++ 2.7.1 [33, 34] and HERWIG 7 [35] **(Are we going to do one or both? –jdt)**,
- SHERPA 2.1.1 [36], **(Go to Sherpa 2.2? –jdt)**
- VINCIA 1.201 [37], **(Go to Vincia 2? –jdt)**
- DEDUCTOR 1.0.2 [38] (with hadronization performed by PYTHIA 8.212),³

²Of course, the quantum numbers of the color singlet operator are not measurable event by event. The idea here is to have a fundamental definition of “quark” and “gluon” that does not reference QCD partons directly.

³Note that this DEDUCTOR plus PYTHIA combination has not yet been tuned to data.

- ARIADNE 5.0. β [39],⁴
- DIRE [41] (with hadronization performed by (**what?** **-jdt**)).

The analysis code used for this study is available as a RIVET routine [42], which can be downloaded from <https://github.com/gsoyez/lh2015-qg> under MC_LHQG_EE.cc.

As we will see, the differences between these generators arise from physics at the interface between perturbative showering and nonperturbative fragmentation. One might think that the largest differences between generators would appear for IRC unsafe observables like multiplicity and p_T^D , where nonperturbative 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. 5.2, we study these differences as a function of collision energy Q , jet radius R , and strong coupling constant α_s , showing that the generators have somewhat different discrimination trends. In Sec. 5.3, 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 most of these 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 [43–46]. In Sec. 6, 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 find that one can already learn a lot from hadron-level measurements, without trying to isolate “pure” quark or gluon samples. In particular, we advocate measuring the generalized angularities on quark/gluon enriched samples:

$$\text{“quark enriched”} : \quad pp \rightarrow Z + \text{jet}, \quad (1.6)$$

$$\text{“gluon enriched”} : \quad pp \rightarrow \text{dijets}, \quad (1.7)$$

where “enriched” means that the Born-level process contributing to these channels is dominated by the corresponding jet flavor. By making judicious kinematic cuts, we could further flavor-enrich these samples [47], though we will not pursue that in this paper for simplicity.

We present our final recommendations and conclusions in Sec. 7. The main take home message from this study is that, contrary to the standard lore, existing 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.

⁴This version of ARIADNE is not yet public, but available from the author on request. For e^+e^- collisions, the physics is the same as in ARIADNE 4 [40].

What is a Quark Jet?

From lunch/dinner discussions



Figure 1. Original slide from the June 10, 2015 summary report of the quark/gluon Les Houches subgroup [1].

2 What is a quark/gluon jet?

As part of the 2015 Les Houches workshop on “Physics at TeV Colliders”, an attempt was made to define exactly what is meant by a “quark jet” or “gluon jet” (see Fig. 1). Here are some suggested options for defining a quark jet, in (approximate) 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 definition assumes that the parton shower history is meaningful, though, 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 nonperturbative hadronization.

- **A maximum- p_T quark parton within a jet in a final state parton shower.** This “max- p_T ” prescription is a variant on the initiating parton prescription above (see further discussion in [48]). It differs from the initiating parton by a calculable amount in a leading logarithmic shower [49] and is based on the same (naive) assumption that the parton shower history is meaningful.
- **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 Ref. [50]. 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 Ref. [47]. 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, especially for observables involving jet grooming []. 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. *The main challenge with this definition is how to determine the criterion that corresponds to successful quark enrichment.* For that, we have to rely to some degree on the other less well-defined notions of what a quark jet is.

To better understand this last definition, consider a quark/gluon discriminant λ . Since λ can be measured on any jet, one can unambiguously determine the cross section $d\sigma/d\lambda$ for any jet sample of interest. But measuring λ does not directly lead to the probability that the jet is a quark jet, nor to the probability distribution $p_q(\lambda)$ for λ within a quark jet sample. Rather, the process of measuring λ must be followed by a separate process of interpreting how the value of λ should be used as part of an analysis.

For example, the user could choose that small λ jets should be tagged as “quark-like” while large λ jets should be tagged as “gluon-like”. Alternatively, the user might combine λ with other discriminant variables as part of a more sophisticated classification scheme. The key point is that one first measures hadron-level discriminant variables on a final state of interest, and only later does one interpret exactly what those discriminants accomplish (which could be different depending on this physics goals of a specific analysis). Typically, one might use a Born-level or eikonal analysis to define which regions of phase space should be associated with “quarks” or “gluons”, but even if these phase space regions are based on naive or ambiguous logic, λ itself is a well-defined discriminant variable.

In Sec. 5, we will consider the generalized angularities λ_β^κ as our discriminant variables and we will assess the degree to which the measured values of λ_β^κ agree with a quark/gluon interpretation based on Born-level production modes. This is clearly an idealization, though one that makes some sense in the context of e^+e^- collisions, since the truth-level “quark” and “gluon” labels are defined by the Lorentz structure of the production vertex. In Sec. 6, we will recommend that the LHC experiments perform measurements of λ_β^κ in well-defined hadron-level final states, without necessarily attempting to determine separate $p_q(\lambda_\beta^\kappa)$ and $p_g(\lambda_\beta^\kappa)$ distributions. Eventually, one would want to use these hadron-level measurements to infer something about parton-level quark/gluon radiation patterns. Even without that interpretation step, though, direct measurements of $d\sigma/d\lambda_\beta^\kappa$ would provide valuable information for parton shower tuning. This in turn would help λ_β^κ become a more robust and powerful discriminant in searches for new physics beyond the standard model.

3 Quantifying tagging performance

3.1 Generalized angularities

A wide variety of quark/gluon discriminants have been proposed (see Ref. [3] for an extensive catalog), but here we limit ourselves to a two-parameter family of generalized angularities [8],



Figure 2. Two-parameter family of generalized angularities, adapted from Ref. [8]. The dots correspond to the five benchmark angularities used in this study. The horizontal line at $\kappa = 1$ corresponds to the IRC safe angularities, $e_\beta = \lambda_\beta^1$.

shown in Fig. 2. These are defined as (repeating Eq. (1.1) for convenience)

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \theta_i^\beta, \quad (3.1)$$

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 [25] (see also [24, 51–54]), and we will sometimes use the shorthand

$$e_\beta \equiv \lambda_\beta^1. \quad (3.2)$$

For general $\kappa \neq 1$, there are quasi-perturbative techniques based on generalized fragmentation functions [8] (see also [4, 55–57]). 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 (at the expense of losing some particle-level information).

For our e^+e^- study, we cluster jets with FASTJET 3.1.3 [58] using the ee -variant of the anti- k_t algorithm [59], with $|\vec{p}|$ -ordered winner-take-all recombination [25, 60, 61] to determine the jet axis \hat{n} . Unlike standard E -scheme recombination [62], the winner-take-all scheme yields a jet axis \hat{n} that does not necessarily align with the jet three-momentum \vec{p} ; this turns out to be a desirable feature for avoiding soft recoil effects [7, 25, 26, 63, 64]. We define

$$z_i \equiv \frac{E_i}{E_{\text{jet}}}, \quad \theta_i \equiv \frac{\Omega_{i\hat{n}}}{R}, \quad (3.3)$$

where E_i is the particle energy, $\Omega_{i\hat{n}}$ is the opening angle to the jet axis, and R is the jet radius (taken to be $R = 0.6$ by default).

For our pp study, we use the standard pp version of anti- k_t with p_T -ordered winner-take-all recombination, defining

$$z_i \equiv \frac{p_{Ti}}{\sum_{j \in \text{jet}} p_{Tj}}, \quad \theta_i \equiv \frac{R_{i\hat{n}}}{R}, \quad (3.4)$$

where p_{Ti} is the particle transverse momentum and $R_{i\hat{n}}$ is the rapidity-azimuth distance to the jet axis. In addition to directly measuring the angularities, we also test the impact of jet grooming using the modified mass drop tagger (mMDT) with $\mu = 1$ [] (equivalently, soft drop declustering with $\beta = 0$ []). This grooming procedure is based on reclustering a jet with the Cambridge/Aachen algorithm [] to form an angular-ordered tree. This tree is then declustered, removing the softer branch until

$$\frac{\min[p_{T1}, p_{T2}]}{p_{T1} + p_{T2}} > z_{\text{cut}}, \quad (3.5)$$

where 1 and 2 label the two branches of a splitting. By applying the mMDT procedure, we can test whether quark/gluon discrimination performance is adversely affected by removing soft radiation from a jet. For concreteness, we set $z_{\text{cut}} = 0.1$ to match the studies in Refs. [].

By adjusting κ and β in the angularities, one can probe different aspects of the jet fragmentation. We consider five benchmark values for (κ, β) indicated by the black dots in Fig. 2:

$$\begin{aligned} (0, 0) &= \text{hadron multiplicity}, \\ (2, 0) &\Rightarrow p_T^D \text{ [5, 6] (specifically } \lambda_0^2 = (p_T^D)^2), \\ (1, 0.5) &= \text{Les Houches Angularity (LHA)}, \\ (1, 1) &= \text{width or broadening [26–28]}, \\ (1, 2) &\Rightarrow \text{mass or thrust [29] (specifically } \lambda_2^1 \simeq m_{\text{jet}}^2/E_{\text{jet}}^2). \end{aligned} \quad (3.6)$$

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. [7]. Most of the results in this paper are shown in terms of the LHA; results for the other four benchmark values are available in the `arXiv` source files, where each figure in this paper corresponds to multipage file.

For the IRC safe case of $\kappa = 1$, there is an alternative version of the angularities based on energy correlation functions [7] (see also [64, 65]),

$$\text{ecf}_\beta = \sum_{i < j \in \text{jet}} z_i z_j \theta_{ij}^\beta \simeq \lambda_\beta^1, \quad (3.7)$$

where equality holds in the extreme eikonal limit.⁵ 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

⁵This equality also relies on using a recoil-free axis choice \hat{n} to define θ_i . Amusingly, $\lim_{\beta \rightarrow 0} \text{ecf}_\beta = (1 - \lambda_0^2)/2$ (i.e. $\kappa = 2, \beta = 0$), such that the $\beta \rightarrow 0$ limit of the IRC safe energy correlation functions corresponds to the IRC unsafe p_T^D .



Figure 3. Alternative metrics for discrimination power. (a) A ROC curve, showing the gluon rejection rate at fixed 20% and 50% quark efficiency. While not shown in this paper, some of this ROC information is available from the `arXiv` source files. (b) Mutual information $I(T; A)$ as a function of the quark fraction f , showing the relationship to classifier separation $\Delta \equiv I_{\frac{1}{2}}''$ and other information-theoretic quantities. The mutual information at $f = 1/2$ (i.e. $I_{\frac{1}{2}}$) is also available from the `arXiv` source files.

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 [3] and 2-subjettiness [66, 67] can improve quark/gluon discrimination.

3.2 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 (repeating Eq. (1.3) for convenience),

$$\Delta = \frac{1}{2} \int d\lambda \frac{(p_q(\lambda) - p_g(\lambda))^2}{p_q(\lambda) + p_g(\lambda)}, \quad (3.8)$$

where p_q (p_g) is the probability distribution for the quark jet (gluon jet) 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, shown in Fig. 3a. 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. There are various ways to turn the ROC curve into a single number, and in the `arXiv` source

files, every figure for Δ is part of a multipage file that also has results for

$$\{g_{20}^{\text{rej}}, g_{50}^{\text{rej}}\} : \quad \text{Gluon rejection rate at } \{20\%, 50\%\} \text{ quark efficiency;} \quad (3.9)$$

$$\{q_{20}^{\text{rej}}, q_{50}^{\text{rej}}\} : \quad \text{Quark rejection rate at } \{20\%, 50\%\} \text{ gluon efficiency;} \quad (3.10)$$

$$s^{\text{rej}} : \quad \text{Symmetric rejection rate at } s^{\text{rej}} \text{ efficiency.} \quad (3.11)$$

Since we are more interested in understanding the relative performance between parton show-ers rather than the absolute performance, we will not show full ROC curves in this paper, 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 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 Ref. [8]). 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

$$I(T; \Lambda) = \int d\lambda \left(f p_q(\lambda) \log_2 \frac{p_q(\lambda)}{p_{\text{tot}}(\lambda)} + (1 - f) p_g(\lambda) \log_2 \frac{p_g(\lambda)}{p_{\text{tot}}(\lambda)} \right), \quad (3.12)$$

where $T = \{q, g\}$ is the set of truth labels, $\Lambda = \{\lambda\}$ is the (continuous) set of discriminant values, and

$$p_{\text{tot}}(\lambda) = f p_q(\lambda) + (1 - f) p_g(\lambda). \quad (3.13)$$

The choice $f = \frac{1}{2}$ was used in Ref. [8] and is also available in the **arXiv** source files,

$$I(T; A)|_{f=\frac{1}{2}} \equiv I_{\frac{1}{2}}, \quad (3.14)$$

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

$$-\frac{\log 2}{4} \frac{\partial^2 I(T; \Lambda)}{\partial f^2} \Big|_{f=\frac{1}{2}} \equiv I''_{\frac{1}{2}} = \Delta. \quad (3.15)$$

More broadly, the dependence of $I(T; A)$ on f can be related to other concepts in statistics, as visualized in Fig. 3b. At $f = 0$ and $f = 1$, the mutual information itself is zero, but the derivatives are:

$$\frac{\partial I}{\partial f} \Big|_{f=0} \equiv I'_0 = \int d\lambda p_q(\lambda) \log_2 \frac{p_q(\lambda)}{p_g(\lambda)}, \quad -\log 2 \frac{\partial^2 I}{\partial f^2} \Big|_{f=0} \equiv I''_0 = \int d\lambda \frac{p_q(\lambda)^2}{p_g(\lambda)}, \quad (3.16)$$

$$-\frac{\partial I}{\partial f} \Big|_{f=1} \equiv I'_1 = \int d\lambda p_g(\lambda) \log_2 \frac{p_g(\lambda)}{p_q(\lambda)}, \quad -\log 2 \frac{\partial^2 I}{\partial f^2} \Big|_{f=1} \equiv I''_1 = \int d\lambda \frac{p_g(\lambda)^2}{p_q(\lambda)}. \quad (3.17)$$

The first derivative is sometimes called relative entropy and the second derivative is sometimes called discrimination significance. Unlike classifier separation, these later metrics do not treat quark and gluon distributions symmetrically.

One advantage of Δ over $I(T; \Lambda)$ is that the integrand in Eq. (3.8) is easier to interpret, since it tracks the fractional difference between the signal and background at a given value of λ .⁶ Specifically, by plotting

$$\frac{d\Delta}{d\lambda} = \frac{1}{2} \frac{(p_q(\lambda) - p_g(\lambda))^2}{p_q(\lambda) + p_g(\lambda)}, \quad (3.18)$$

one can easily identify which regions of phase space contribute the most to quark/gluon discrimination. One can then ask whether or not the regions exhibiting the most separation power are under sufficient theoretical control, including both the size of perturbative uncertainties and the impact of nonperturbative corrections.

4 Analytic quark/gluon predictions

(Need to gather references for this section. –jdt)

For the IRC safe angularities with $\kappa = 1$ (namely $e_{0.5}$, e_1 , and e_2 from Eq. (3.6)), we can use analytic calculations to get a baseline expectation for the degree of quark/gluon separation. At LL accuracy, a jet effectively consists of a single gluon emission from a hard quark/gluon, with a suitable Sudakov form factor coming from vetoing additional radiation. At this order, the strong coupling constant is fixed and only the leading splitting function is used. In particular, the IRC safe angularities at LL order satisfy a property called Casimir scaling (reviewed below), such that the discrimination power is independent of β .

At NLL order, the jet is described by multiple gluon emissions from a hard quark/gluon, including the effects of α_s running and subleading terms in the splitting function, but neglecting matrix element corrections and energy-momentum conservation. For the IRC safe angularities, the NLL-accurate distributions were calculated in Ref. [1] (see also [2]). For the generalized angularities, generalized fragmentation functions [3] were used to extend the NLL calculation beyond the IRC safe regime [4], though we will not use that technique in the present paper. To date, the impact of soft nonperturbative physics has not been included in the distributions for the IRC safe angularities, but we will do so below.

4.1 Casimir scaling at LL

As shown in Refs. [1], the IRC safe angularities satisfy Casimir scaling at LL accuracy, which implies that the quark/gluon discrimination performance only depends on the color factor

⁶Another 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 [8].

ratio C_A/C_F . To see this, let us first introduce the notation for the cumulative distribution $\Sigma(\lambda)$, which is defined by

$$\Sigma(\lambda) = \int_0^\lambda d\lambda' p(\lambda'), \quad p(\lambda) = \frac{d\Sigma}{d\lambda}. \quad (4.1)$$

If an observable satisfies Casimir scaling, then the quark and gluon cumulative distributions can be written as

$$\Sigma_q(\lambda) = e^{-C_F r(\lambda)}, \quad \Sigma_g(\lambda) = e^{-C_A r(\lambda)}, \quad (4.2)$$

where $r(\lambda)$ is a monotonically decreasing function of λ . Here, the only difference between the quark and gluon distributions is in the color factors $C_F = 4/3$ versus $C_A = 3$.

At LL accuracy, the distributions for the IRC safe angularities take precisely this form [1],

$$\Sigma_i^{\text{LL}}(e_\beta) = \exp \left[-\frac{\alpha_s C_i}{\pi \beta} \log^2 e_\beta \right]. \quad (4.3)$$

This LL result can be understood from the fact that quarks and gluons have the same splitting function up to an overall multiplicative color factor

$$P_i(z) \simeq \frac{2C_i}{z}, \quad (4.4)$$

and therefore the Sudakov form factor (which is the only thing that appears in Eq. (4.3)) differs only by the color factor in the exponent.

Observables that satisfy Casimir scaling have universal ROC curves [1] and universal truth overlaps [1], which are independent of the precise functional form of $r(\lambda)$ and only depend on the ratio C_A/C_F . We can derive the same universality for classifier separation. Using

$$p_i(\lambda) = -C_i r'(\lambda) e^{-C_i r(\lambda)} \quad (4.5)$$

and the change of variables $u = e^{-C_F r(\lambda)}$, we have

$$\Delta = -\frac{1}{2} \int d\lambda r'(\lambda) \frac{(C_F e^{-C_F r(\lambda)} - C_A e^{-C_A r(\lambda)})^2}{C_F e^{-C_F r(\lambda)} + C_A e^{-C_A r(\lambda)}}, \quad (4.6)$$

$$= \frac{1}{2} \int_0^1 \frac{du}{u} \frac{(u - (C_A/C_F)u^{C_A/C_F})^2}{u + (C_A/C_F)u^{C_A/C_F}}, \quad (4.7)$$

$$= 2 \left({}_2F_1 \left[1, \frac{C_F}{C_A - C_F}; \frac{C_A}{C_A - C_F}; -\frac{C_A}{C_F} \right] - \frac{1}{2} \right), \quad (4.8)$$

where ${}_2F_1[a, b; c; z]$ is the hypergeometric function.⁷ For the case of QCD with $C_A/C_F = 9/4$,

$$\Delta_{\text{QCD}} \simeq 0.1286. \quad (4.9)$$

⁷An alternative way to derive this result is taking the result for $I(T; A)$ in Ref. [8] and using Eq. (3.15) to extract Δ .

We mark this benchmark value on the subsequent plots for reference.⁸ **(We don't yet, but we should consider it. -jdt)**

Going beyond LL accuracy, Casimir scaling is typically violated, and Δ depends on the precise observable in question. Thus, all of the differences we see in our subsequent studies are effects that are truly higher-order or nonperturbative.

4.2 NLL resummation

Going to NLL accuracy is straightforward, using the formalism of [9]. The cumulative distribution takes the form

$$\Sigma(e_\beta) = \frac{e^{-\gamma_E R'(e_\beta)}}{\Gamma(1 + R'(e_\beta))} e^{-R(e_\beta)}, \quad (4.10)$$

where $R(e_\beta)$ is known as the radiator function, γ_E is the Euler-Mascheroni constant, Γ is the gamma function, and primes indicate logarithmic derivatives:

$$R'(e_\beta) = -\frac{\partial}{\partial \log e_\beta} R(e_\beta). \quad (4.11)$$

The $e^{-R(e_\beta)}$ factor in Eq. (4.10) is just the Sudakov form factor, and the prefactor containing $R'(e_\beta)$ captures the effect of multiple emissions on the e_β distribution.

For the IRC safe angularities, the radiator function is [10]

$$R(e_\beta) = C_i \int_0^1 \frac{d\theta}{\theta} \int_0^1 dz p_i(z) \frac{\alpha_s(k_t)}{\pi} \Theta(z\theta^\beta - e_\beta), \quad (4.12)$$

and the strong coupling constant is evaluated using the **(what scheme? -jdt)** at the k_t emission scale

$$k_t = z\theta R E_J. \quad (4.13)$$

The reduced splitting functions (i.e. splitting functions summed over all allowed $1 \rightarrow 2$ processes) are

$$p_q(z) = \frac{1 + (1-z)^2}{z}, \quad p_g(z) = 2\frac{1-z}{z} + z(1-z) + \frac{T_R n_f}{C_A} (z^2 + (1-z)^2), \quad (4.14)$$

where $T_R = 1/2$ and we take the number of light quark flavors to be $n_f = 5$.

(I think Andrew's formula actually expands this out to only have NLL pieces. Need to ask him to write up the details. -jdt)

(Andrew actually gave me NGL information, so we should include that. NGLs Casimir scale, though, so it shouldn't have a major impact on the discrimination power. -jdt)

⁸In large N_c QCD where $C_A/C_F \rightarrow 2$, $\Delta \rightarrow \ln 3 - 1 \simeq 0.0986$, so quark/gluon separation is expected to be harder as $N_c \rightarrow \infty$

⁹Here, NLL refers only to global logarithms; non-global logarithms [11] require a more sophisticated treatment [12]. **(I think I might have NGLs from Andrew, need to look. -jdt)** We also assume that the jet radius R is order 1 so we can ignore $\log R$ resummation.

4.3 Nonperturbative shape function

The quark/gluon studies in Refs. [] used solely the distributions as calculated in Eq. (4.10). As we will see, nonperturbative hadronization has a big effect in our parton shower studies, so we would like to include the corresponding effect in our analytic results.

The IRC safe angularities are additive observables, meaning that at leading power in the small e_β limit, one can decompose them into separate contributions from perturbative and nonperturbative modes:

$$e_\beta = e_\beta^{(\text{pert})} + e_\beta^{(\text{NP})}. \quad (4.15)$$

One can then convolve the perturbative distributions for $e_\beta^{(\text{pert})}$ with a nonperturbative shape function F that describes the distribution of $\epsilon \equiv e_\beta^{(\text{NP})}$:

$$\frac{d\sigma}{de_\beta} = \int d\epsilon \hat{\sigma}(e_\beta - \epsilon) F(\epsilon), \quad (4.16)$$

where $\hat{\sigma} \equiv d\sigma/de_\beta$ refers to the perturbative distribution. The shape function prescription gives sensible results in the small e_β limit, but it breaks down at large values of e_β , since the convolution in Eq. (4.16) can yield e_β values that extend beyond the physical range. To address this, we need to smoothly turn off the nonperturbative shift as e_β approaches the physical endpoint e_β^{max} . There is no unique way to do this, but we find sensible results using

$$\frac{d\sigma}{de_\beta} = \int d\epsilon \hat{\sigma}\left(e_\beta - \epsilon + \epsilon \frac{e_\beta}{e_\beta^{\text{max}}}\right) F(\epsilon) \left(1 + \frac{\epsilon}{e_\beta^{\text{max}}}\right), \quad (4.17)$$

where the last factor ensures that the total cross section is unmodified **(Need to check this. It comes from Jacobian after performing the perturbative integration –jdt)**. For simplicity, we take $e_\beta^{\text{max}} = 1$ for all of our distributions.¹⁰

The shape function F has to be extracted from data, but we can use simple parametrizations that account for some aspects of its known behavior []:

$$F(\epsilon) = \frac{\pi\epsilon}{2\epsilon_0^2} \exp\left[-\frac{\pi}{4} \frac{\epsilon^2}{\epsilon_0^2}\right], \quad F_{\text{alt}}(\epsilon) = \frac{4\epsilon}{\epsilon_0^2} \exp\left[-\frac{2\epsilon}{\epsilon_0}\right] \quad (4.18)$$

Both of these functions go linearly to zero at $\epsilon \rightarrow 0$, fall exponentially as $\epsilon \rightarrow \infty$, are normalized by $\int d\epsilon F(\epsilon) = 1$, and have an expectation value $\langle \epsilon \rangle = \int d\epsilon \epsilon F(\epsilon) = \epsilon_0$. The parameter ϵ_0 can be interpreted as the average shift of the perturbative distribution from nonperturbative effects. The second form was used in Ref. [], but we prefer the

Following the work of Refs. [] (see also []), one can estimate ϵ_0 as a function of β . Nonperturbative modes have $k_t \simeq \Lambda_{\text{QCD}}$, which implies the relationship

$$z\theta \simeq \frac{\Lambda_{\text{QCD}}}{R E_J}. \quad (4.19)$$

¹⁰**(Add footnote about actual maximum value. –jdt)**

Appealing to local parton-hadron duality, we can estimate the average contribution to $z\theta^\beta$ from nonperturbative soft gluon emissions as

$$\begin{aligned}\epsilon_0 &= \frac{\Omega_0}{R E_J} \int_0^1 \frac{dz}{z} \int_0^1 \frac{d\theta}{\theta} z \theta^\beta \delta\left(z\theta - \frac{\Xi_0}{R E_J}\right) \\ &= \frac{1}{\beta-1} \frac{\Omega_0}{R E_J} \left(1 - \left(\frac{\Xi_0}{R E_J}\right)^{\beta-1}\right),\end{aligned}\tag{4.20}$$

where we are using the soft gluon matrix element to determine the phase space integration, and Ω_0 and Ξ_0 are nonperturbative parameters that are both of order Λ_{QCD} . This estimate of ϵ_0 can also be understood by considering two types of nonperturbative modes:

$$\text{NP Soft : } z \simeq \frac{\Lambda_{\text{QCD}}}{R E_J}, \quad \theta \simeq 1,\tag{4.21}$$

$$\text{NP Collinear : } z \simeq 1, \quad \theta \simeq \frac{\Lambda_{\text{QCD}}}{R E_J}.\tag{4.22}$$

These contribute to the angularities as

$$e_\beta^{(\text{NP})} \simeq \frac{\Lambda_{\text{QCD}}}{R E_J} + \left(\frac{\Lambda_{\text{QCD}}}{R E_J}\right)^\beta.\tag{4.23}$$

We see that soft modes dominate for $\beta > 1$, collinear modes dominate for $\beta < 1$, and soft and collinear modes are equally important for $\beta = 1$. This behavior is indeed encoded in Eq. (4.20), which smoothly interpolates between these regimes.

Comparing quarks and gluons, we expect the overall size of the nonperturbative shift Ω_0 should scale proportional to the Casimir factors, as in Eq. (4.4). The scaling of Ξ_0 is less clear, but for simplicity, we assume that it also obeys Casimir scaling:

$$\frac{\Omega_0^g}{\Omega_0^q} = \frac{\Xi_0^g}{\Xi_0^q} = \frac{C_A}{C_F}.\tag{4.24}$$

By tying Ω_0 and Ξ_0 together, this has the effect of reducing the phase space for nonperturbative emissions from gluons, which is particularly important for $\beta < 1$. Ideally, one would like a more rigorous justification for the assumptions in Eq. (4.24) (as well as the convolution structure in Eq. (4.17)).

To test whether Eq. (4.20) is a plausible estimate for nonperturbative corrections, we take the parton shower generators studied in the next section and study how the average value of e_β shifts as hadronization is turned off and on, and use that to estimate ϵ_0 . We emphasize that a hadronization model in a parton shower is not the same as a shape function in an analytic calculation, so one has to be careful drawing conclusions about the size of ϵ_0 from a study like this. In particular, effects that are captured by $F(\epsilon)$ in an analytic calculation could either be part of the perturbative showering or nonperturbative hadronization in a generator. That said, we expect that that scaling of the e_β shift as a function of E_J , R , and β should be roughly the same.

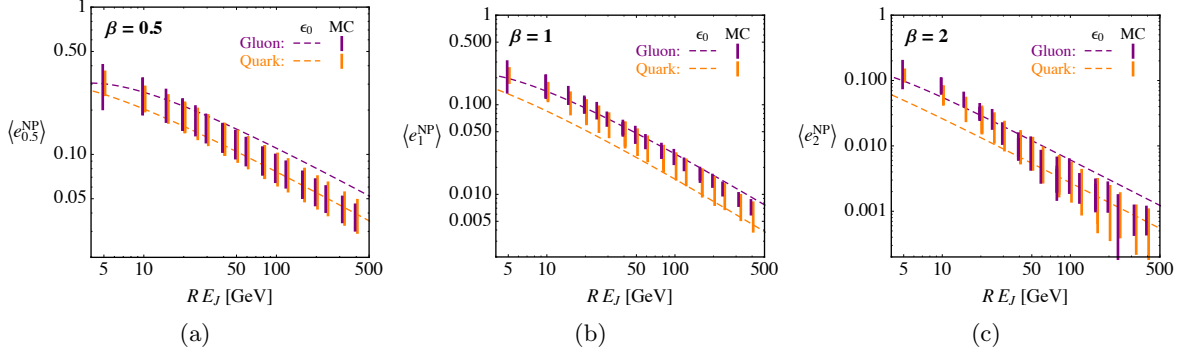


Figure 4. Nonperturbative shifts to the IRC safe angularities as a function of RE_J for (a) $\beta = 0.5$, (b) $\beta = 1$, and (c) $\beta = 2$. The vertical bands correspond to the range of shifts seen by turning hadronization off and on in the parton shower generators. The dashed line corresponds to the best fit to Ω_0 and Ξ_0 , assuming the functional form for ϵ_0 in Eq. (4.20) and the assumption of Casimir scaling in Eq. (4.24). While the RE_J trend agrees, the hadronization corrections (as implemented in the parton showers) do not appear to exhibit Casimir scaling. **(Gregory recommends showing ratio as well. -jdt)**

In Fig. 4, we show the size of the e_β shift as a function of RE_J for the three benchmark β values, where the band indicates the minimum and maximum shifts seen among PYTHIA, HERWIG, SHERPA, VINCIA, and ARIADNE.¹¹ We then compare to the expected shift from Eq. (4.20) with

$$\Omega_0^i = C_i \times 0.21 \text{ GeV}, \quad \Xi_0^i = C_i \times 0.34 \text{ GeV}, \quad (4.25)$$

where these values are obtained by doing a (logarithmic) fit to all of the parton shower shift values. While the parton shower trends with RE_J and β roughly agree with, there is no evidence for Casimir scaling, most likely because parton showers already achieve some degree of Casimir scaling through multiple (perturbative) soft gluon emissions from the shower. Despite this caveat, we use the extracted values from Eq. (4.25) for our baseline distributions below. As an alternative, we also test a variant without Casimir scaling with

$$\Omega_0^i = 0.40 \text{ GeV}, \quad \Xi_0^i = 0.64 \text{ GeV}, \quad (4.26)$$

where again these values are estimated by fitting to the parton shower e_β shifts.

5 Idealized quark/gluon discrimination

(I haven't re-editted this section yet. -jdt)

Our parton shower studies are based on the idealized case of discriminating quark and gluon jets in e^+e^- collisions. As we will see, this example demonstrates the importance of final state evolution for quark/gluon discrimination, separate from initial state complications

¹¹We verified that the overall conclusions do not change when considering the separate scaling of R and E_J .

arising in pp collisions. The analysis code used for this study is available as a RIVET routine [42], which can be downloaded from <https://github.com/gsoyez/lh2015-qg> under MC_LHQQ_EE.cc.

To define the truth-level jet flavor, we use a simple definition: a quark jet is a jet produced by a parton shower event generator in $e^+e^- \rightarrow (\gamma/Z)^* \rightarrow u\bar{u}$ hard scattering, while a gluon jet is a jet produced in $e^+e^- \rightarrow h^* \rightarrow gg$. Of course, an $e^+e^- \rightarrow u\bar{u}$ event can become a $e^+e^- \rightarrow u\bar{u}g$ event after one step of shower evolution, just as $e^+e^- \rightarrow gg$ can become $e^+e^- \rightarrow gu\bar{u}$. This illustrates the inescapable ambiguity in defining jet flavor.¹² To partially mitigate the effect of wide-angle emissions, we restrict our analysis to jets that satisfy

$$\frac{E_{\text{jet}}}{Q/2} > 0.8, \quad (5.1)$$

with up to two jets studied per event. There is also the ambiguity of which parton shower to use, and we investigate this ambiguity by looking at results from several event generators: PYTHIA 8.205 [31, 32], HERWIG++ 2.7.1 [33, 34], SHERPA 2.1.1 [36], VINCIA 1.201 [37], DEDUCTOR 1.0.2 [38] (with hadronization performed by PYTHIA), and ARIADNE 5.0. β [39].

5.1 Baseline analysis

In Fig. 5, 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 six 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. 5a, there is relatively little variation between the generators, which is not surprising since most of these programs have been tuned to match LEP data (though LEP never measured the LHA itself). Turning to the gluon sample in Fig. 5b, we see somewhat larger variations between the generators; this is expected since there is no data to directly constrain $e^+e^- \rightarrow gg$. In Fig. 5c, we plot the integrand of classifier separation, $d\Delta/d\lambda$ from Eq. (3.18). This 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 [10]. The predicted discrimination power from the other four generators are intermediate between these extremes.

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. We should say at the outset that it is nearly impossible to do a true apples-to-apples comparison of parton-level results, since these generators are interfaced to different hadronization models, and only the hadron-level comparison is physically meaningful. In particular, the crossover between the perturbative and nonperturbative regions is ambiguous and each of these showers has a different effective shower cutoff scale,

¹²In an e^+e^- context, our definition at least respects the Lorentz structure of the production vertex, so in that sense it is a fundamental definition that does not reference (ambiguous) quark or gluon partons directly.

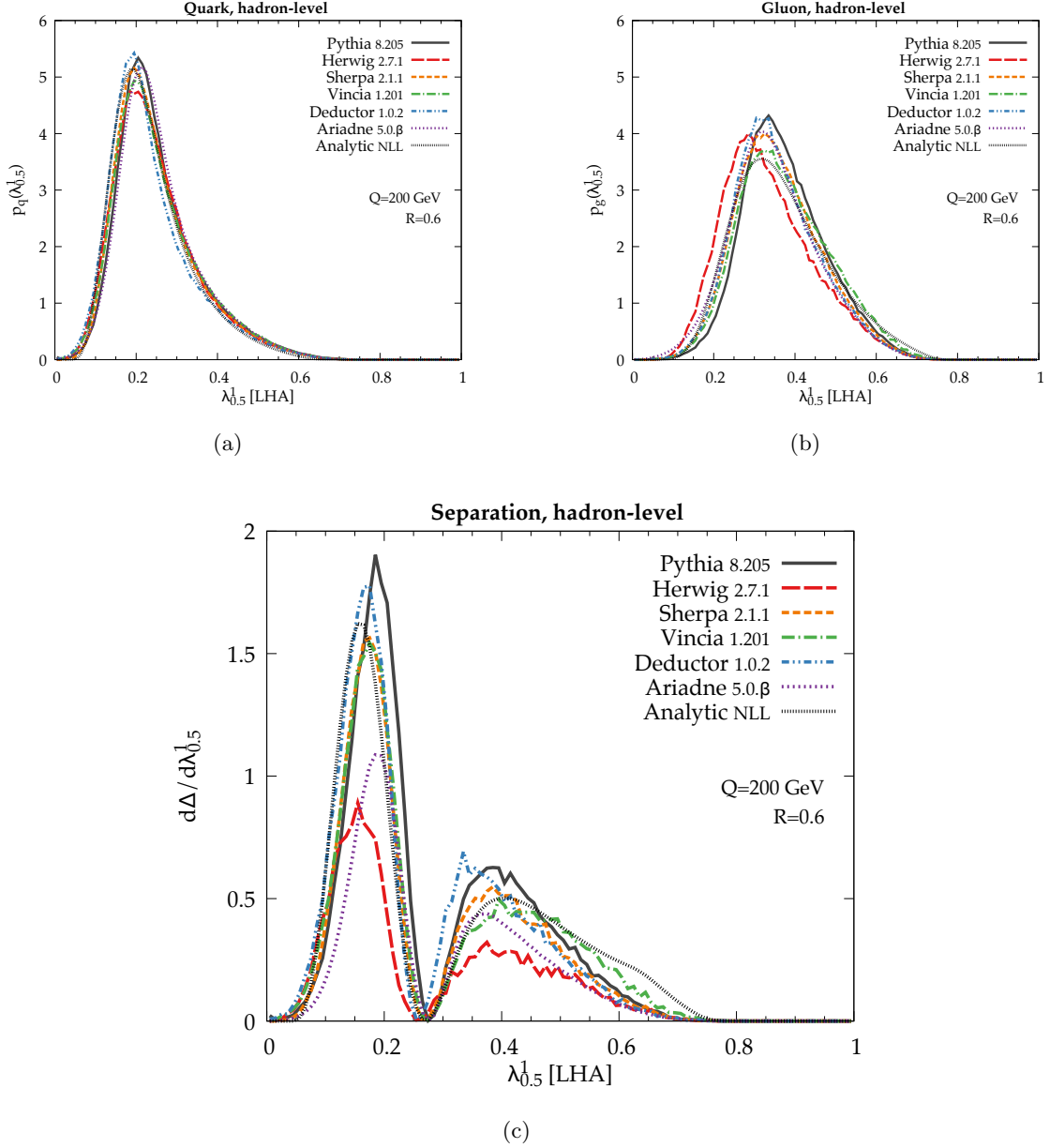


Figure 5. Hadron-level distributions of the LHA for (a) the $e^+e^- \rightarrow u\bar{u}$ (“quark jet”) sample, (b) the $e^+e^- \rightarrow gg$ (“gluon jet”) sample, and (c) the classifier separation integrand in Eq. (3.18). Six parton shower generators—PYTHIA 8.205, HERWIG++ 2.7.1, SHERPA 2.1.1, VINCIA 1.201, DEDUCTOR 1.0.2, and ARIADNE 5.0.β—are run at their baseline settings with center-of-mass energy $Q = 200$ GeV and jet radius $R = 0.6$.

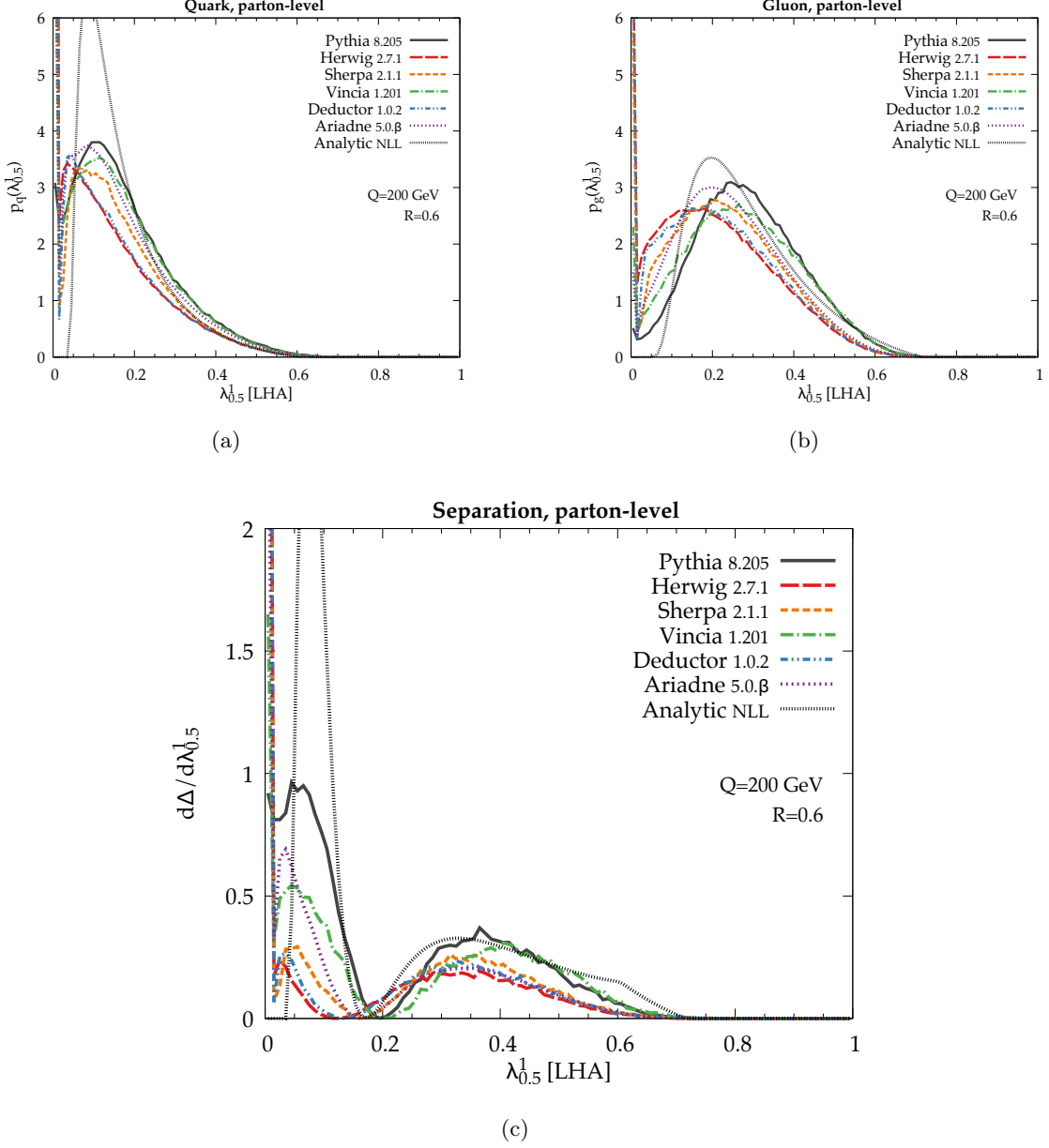


Figure 6. Same as Fig. 5, but at the parton level. Note that HERWIG, SHERPA, and DEDUCTOR all have cross section spikes at $\lambda_{0.5}^1 = 0$ that extend above the plotted range.

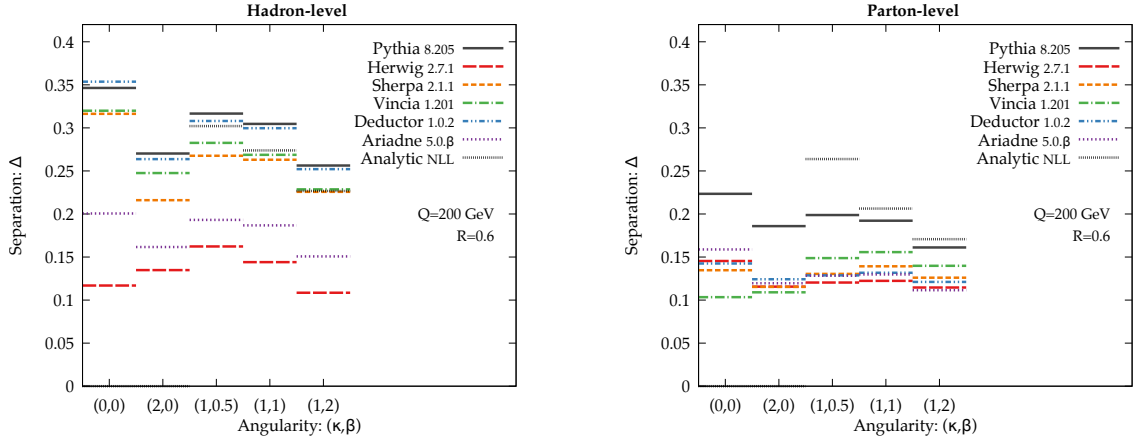


Figure 7. Classifier separation Δ for the five benchmark angularities in Eq. (3.6), determined from the various generators at (a) hadron level and (b) parton level. The first two columns correspond to IRC unsafe distributions (multiplicity and p_T^D), while the last three columns are the IRC safe angularities. The LHA (i.e. $\kappa = 1$, $\beta = 1/2$) is shown in the middle column. For results in terms of ROC values, see Fig. ??.

resulting in different amounts of radiation being generated in the showering versus hadronization steps.¹³

With that caveat in mind, we show parton-level results in Fig. 6. One immediately notices that three of the generators—HERWIG, SHERPA, and DEDUCTOR—yield a large population of events where the perturbative shower generates no emissions. This gives $\lambda_{0.5}^1 = 0$ 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. As mentioned above, some of this difference can be explained simply by the different shower cutoff scales used in each generator, but it probably also reflects a difference in how semi-perturbative gluon splittings are treated. Since Fig. 5a 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 nonperturbative hadronization.

To summarize the overall discrimination power, we integrate Eq. (3.18) to obtain the value of classifier separation Δ for the LHA. This is shown in Fig. 7, which also includes the four other benchmark angularities from Eq. (3.6). There is a rather large spread in predicted discrimination power between the generators, especially at hadron level in Fig. 7a. While such differences might be expected for IRC unsafe angularities (multiplicity and p_T^D) which depend on nonperturbative modeling, these differences persist even for the IRC safe

¹³In general, generators based on string hadronization tend to use a lower shower cutoff scale (~ 0.5 MeV) compared to those based on cluster hadronization (~ 1 GeV).

angularities at parton level (see Fig. 7b).¹⁴ This suggests a more fundamental difference between the generators that is already present in the perturbative shower

For the IRC safe angularities with $\kappa = 1$, there is a generic trend seen by all of the hadron-level generators that discrimination power decreases as β increases. This trend agrees with the study performed in Ref. [7], but disagrees with the ATLAS study in Ref. [10], which found flat (or even increasing) discrimination power with increasing β . Understanding this β trend will therefore be crucial for understanding quark/gluon radiation patterns.

5.2 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:

$$\begin{aligned} \text{Collision Energy : } Q &= \{50, 100, \mathbf{200}, 400, 800\} \text{ GeV}, \\ \text{Jet Radius : } R &= \{0.2, 0.4, \mathbf{0.6}, 0.8, 1.0\}, \\ \text{Strong Coupling : } \alpha_s/\alpha_{s0} &= \{0.8, 0.9, \mathbf{1.0}, 1.1, 1.2\}, \end{aligned} \tag{5.2}$$

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. 8, at both the hadron level and parton level. There are number of surprising features in these plots. Perhaps the most obvious (and seen already in Fig. 7) is that even for the IRC safe angularities, the effect of hadronization is rather large, both on the absolute scale of discrimination and the trends. The main exception to this is HERWIG, which does not exhibit as much of an effect from hadronization, though an effect is still present.

The next surprising feature is that the parton-level trends for sweeping α_s do not necessarily correspond to those for sweeping Q and R . According to the perturbative next-to-leading-logarithmic (NLL) logic in Ref. [7], quark/gluon discrimination should depend on α_s evaluated at the scale $QR/2$, with larger values of $\alpha_s(QR/2)$ leading to improved discrimination power. Indeed, PYTHIA, HERWIG, and ARIADNE do show improved performance with larger α_s . However, larger values of Q and R correspond to smaller values of α_s , so the NLL logic would predict that increasing Q or R should lead to worse discrimination power. Instead, all of the generators show the opposite trend.

One reason to expect quark/gluon discrimination to improve as higher energies is that that phase space available for shower evolution increases as Q increases. The scale μ of the shower splitting is $\mu_0^2 < \mu^2 < Q^2$, where $\mu_0 = \mathcal{O}(\text{GeV})$ is the shower cutoff scale. With more range for shower evolution at higher Q , there is a greater possibility to see that a quark jet is different from a gluon jet. Similarly, larger values of R allow for more emissions within a jet,

¹⁴It is interesting that four of the generators—HERWIG, SHERPA, DEDUCTOR, and ARIADNE—have a comparatively narrow spread in predicted discrimination power at parton level, though this spread increases dramatically at hadron level.

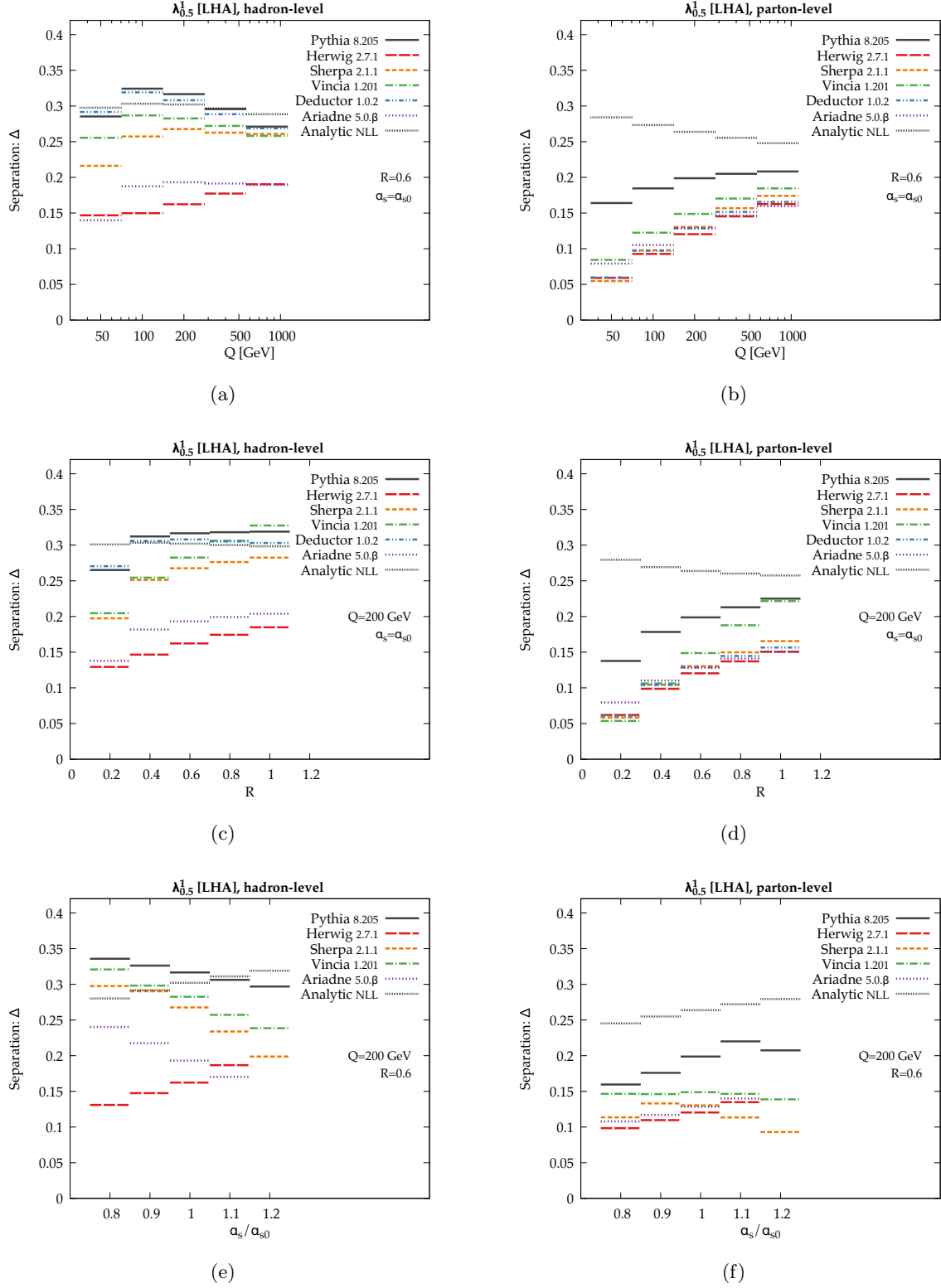


Figure 8. Classifier separation Δ for the LHA, sweeping the collision energy Q (top row), jet radius R (middle row), and coupling constant α_s/α_{s0} (bottom row). Results are shown at hadron level (left column) and parton level (right column). For results in terms of ROC values, see Figs. ?? and ??.

and from scaling symmetry, one expects that parton-level discrimination power should depend on the combination QR .¹⁵ By contrast, the NLL logic says that quark/gluon discrimination should be dominated by the leading emission(s) in a jet, and since α_s is smaller at higher values of QR , those leading emissions are more similar between quarks and gluons. Given these two different but equally plausible logics, both of which probably play some role in the complete story, this motivates experimental tests of quark/gluon separation as a function of Q and R .

For many of the generators, going from parton-level to hadron-level reverses or flattens the Q and α_s trends, though the R trends are more stable. The study in Ref. [7] did not include nonperturbative hadronization corrections, so we do not yet have a theoretical expectation for the impact of hadronization. In future work, we plan to follow Ref. [51] and include nonperturbative hadronization corrections via shape functions [69–75] as well as test the impact of imposing a hard IR cutoff scale. With confusions already at parton level, though, further perturbative calculations beyond NLL accuracy are also needed. For example, 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. By contrast, parton showers (while formally only LL accurate) allow every emission to reradiate, which might be driving the Q and R discrimination trends.

5.3 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. [76]. In Fig. 9, we consider the following PYTHIA variations:

- PYTHIA: NO $g \rightarrow q\bar{q}$. While the dominant gluon splitting in the parton shower is $g \rightarrow gg$, PYTHIA—and every other shower in this study—also generates the subleading

¹⁵At small values of R , one has to worry about the flavor purity of a jet sample, since scale evolution can change the leading parton flavor [49, 68]. Similarly, the restriction in Eq. (5.1) can impose a non-trivial bias on the jet flavor at small R .

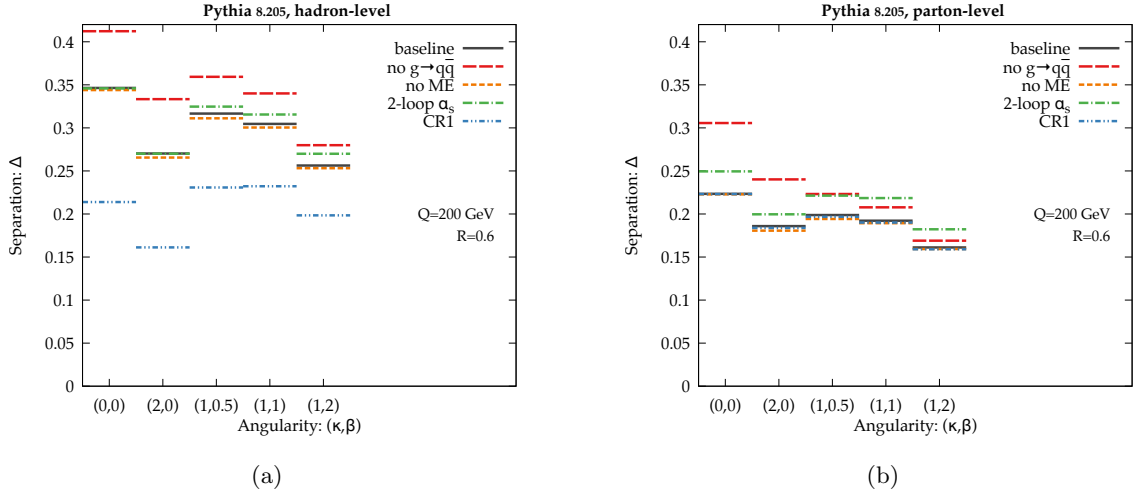


Figure 9. Settings variations for PYTHIA 8.205. Shown are (a) hadron-level and (b) parton-level results for the classifier separation Δ derived from the five benchmark angularities.

$g \rightarrow q\bar{q}$ splittings by default. This variation turns off $g \rightarrow q\bar{q}$, which makes gluon jets look more gluon-like, thereby increasing the separation power.

- **PYTHIA: NO ME.** The first emission in PYTHIA is improved by applying a matrix element correction [77], but this variation turns those corrections off, showing the impact of non-singular terms. No matrix element correction is available for $h^* \rightarrow gg$, though, so the true impact of these corrections might be larger than the relatively small effect seen for this variation.
- **PYTHIA: 2-LOOP α_s .** 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: 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. Compared to the baseline, this variation uses an alternative “SU(3)”-based color reconnection model [78] (i.e. `ColourReconnection:mode = 1`). No attempts were made to retune any of the other hadronization parameters (as would normally be mandated in a tuning context), so this change simply illustrates the effect of switching on this reconnection model with default parameters, leaving all other parameters unchanged. At parton level, this variation has no effect as expected. At hadron level, this variation considerably decreases quark/gluon separation compared to the baseline.

The most surprising PYTHIA effect is the large potential impact of the color reconnection model, which is also important for the HERWIG generator described next.

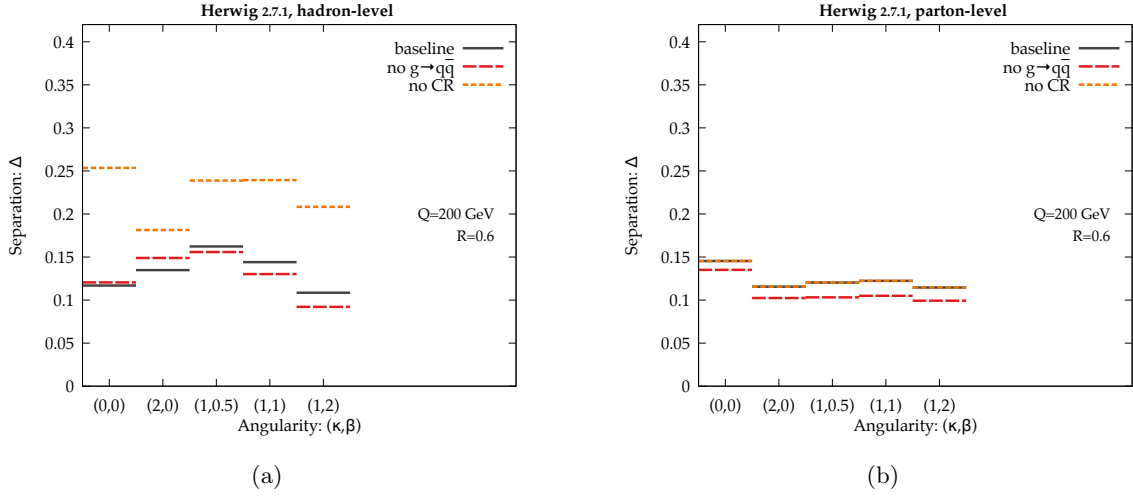


Figure 10. Same as Fig. 9, but for HERWIG 2.7.1.

Our HERWIG++ baseline uses version 2.7.1, with improved modeling of underlying event [79] and the most recent UE-EE-5-MRST tune [80], which is able to describe the double-parton scattering cross section [81] and underlying event data from $\sqrt{s} = 300$ GeV to $\sqrt{s} = 7$ TeV. In Fig. 10, we consider the following HERWIG variations:

- HERWIG: NO $g \rightarrow q\bar{q}$. Turning off $g \rightarrow q\bar{q}$ splittings in HERWIG has the reverse behavior as seen in PYTHIA, leading to slightly worse discrimination power, though the effect is modest.
- HERWIG: NO CR. The variation turns off color reconnections in HERWIG. This has no effect at parton level, as expected. At hadron level, 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.

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. In the future, we also plan to test the default angular-ordered HERWIG shower against an alternative dipole shower [82].

Our SHERPA baseline uses matrix element corrections for the first two emissions ($N_{\text{jet}} = 2$) with CKKW-style matching [83]. In Fig. 11, we consider the following SHERPA variations:

- SHERPA: NO $g \rightarrow q\bar{q}$. Turning off $g \rightarrow q\bar{q}$ splittings in SHERPA has a negligible effect at parton level, but it leads to a large jump in discrimination power at hadron level, again due to an interplay between the perturbative shower and nonperturbative hadronization.

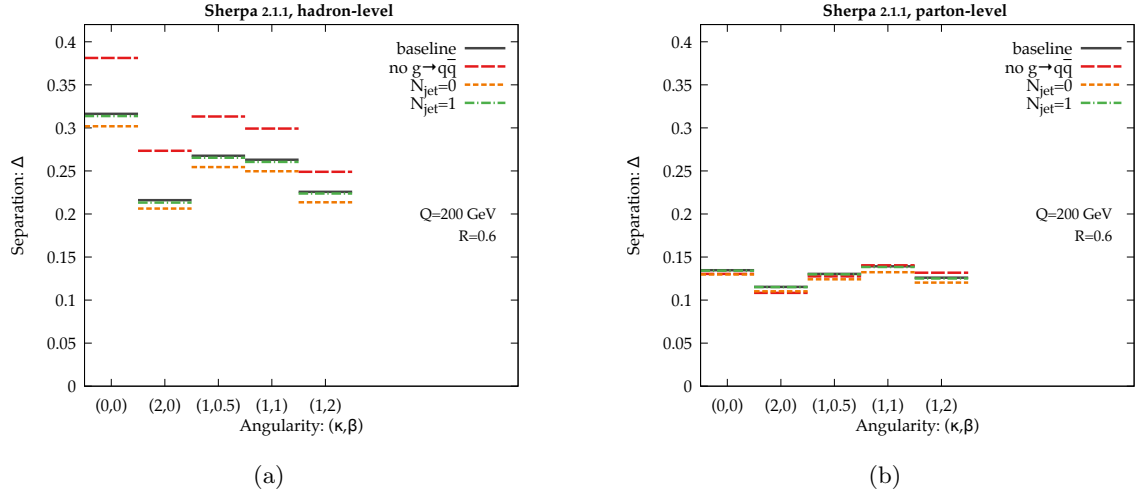


Figure 11. Same as Fig. 9, but for SHERPA 2.1.1.

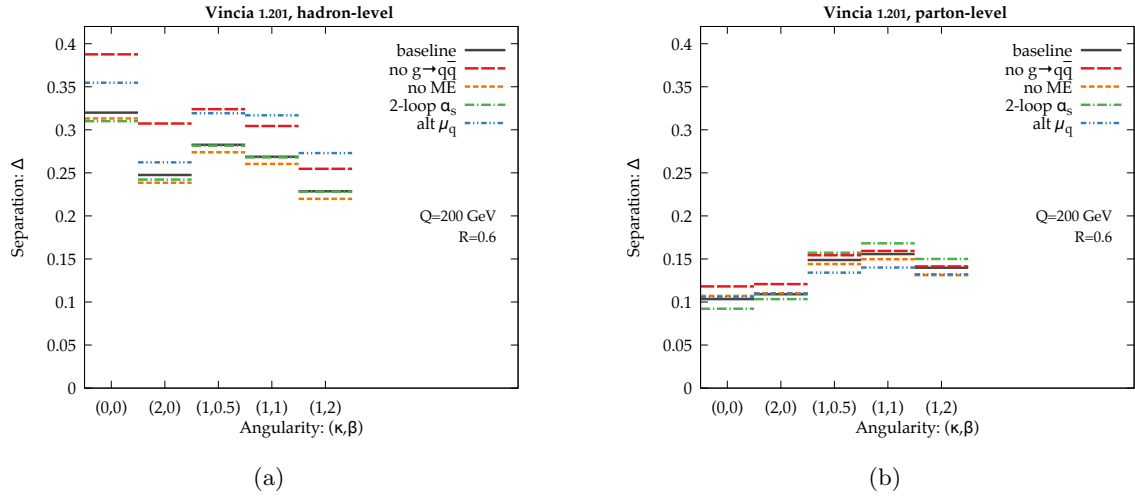


Figure 12. Same as Fig. 9, but for VINCIA 1.201.

- SHERPA: $N_{\text{jet}} = 1$. This variation only performs CKKW matching for the first emission, leading to negligible changes in the discrimination performance.
- SHERPA: $N_{\text{jet}} = 0$. Turning off all matrix element corrections in SHERPA slightly decreases the predicted quark/gluon discrimination power, in agreement with the behavior of PYTHIA.

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 $g \rightarrow q\bar{q}$ splittings motivates further investigations into the shower/hadronization interface.

Our VINCIA baseline is based on the “jeppsson5” tune. While VINCIA has NLO matrix elements for $e^+e^- \rightarrow q\bar{q}$, it does not have them for $e^+e^- \rightarrow gg$, so we will use LO VINCIA throughout. In Fig. 12, we consider the following VINCIA variations:

- VINCIA: NO $g \rightarrow q\bar{q}$. This variation turns off $g \rightarrow q\bar{q}$, leading to the expected increase in separation power as seen in PYTHIA.
- VINCIA: NO ME. By default, each $2 \rightarrow 3$ antenna in VINCIA has an associated matrix element correction factor. Since the antenna are already rather close to the true matrix elements, turning off these matrix elements has a modest effect on quark/gluon discrimination power.
- VINCIA: 2-LOOP α_s . 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: ALT μ_q . 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 nonperturbative effects.

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 DEDUCTOR baseline uses leading color plus (LC+) showering, which includes some subleading color structures. We find that switching from LC+ to LC showering at parton level has a negligible impact on quark/gluon discrimination power. When DEDUCTOR interfaces with the default tune of PYTHIA 8.212 for hadronization, only leading color is used in the showering, such that partons with their LC color information can be directly passed to the Lund string model. No DEDUCTOR variations are shown here, but we anticipate studying the effect of $g \rightarrow q\bar{q}$ splitting in future work.

Our ARIADNE baseline is based on a beta release of version 5. In Fig. 13, we consider the following ARIADNE variation:

- ARIADNE: NO $g \rightarrow q\bar{q}$. This variation turns off $g \rightarrow q\bar{q}$, leading to modest change in separation power, similar in magnitude to HERWIG.
- ARIADNE: NO SWING. Swing refers to color reconnections performed during the perturbative cascade, where dipoles in the same color state are allowed to reconnect in a way which prefers low-mass dipoles [39, 84]. Turning off swing has an effect already at parton level, which is amplified at hadron level, leading to improved quark/gluon separation.

Like for PYTHIA and HERWIG, color reconnections play a surprisingly important role in ARIADNE.

Finally in Fig. 14, ([analytic here](#). -jdt)

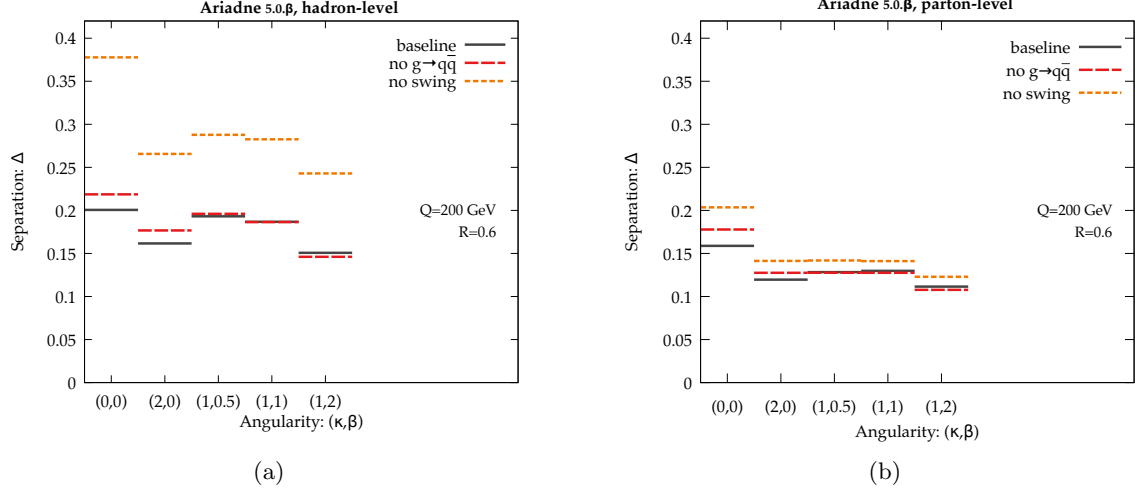


Figure 13. Same as Fig. 9, but for ARIADNE 5.0.β.

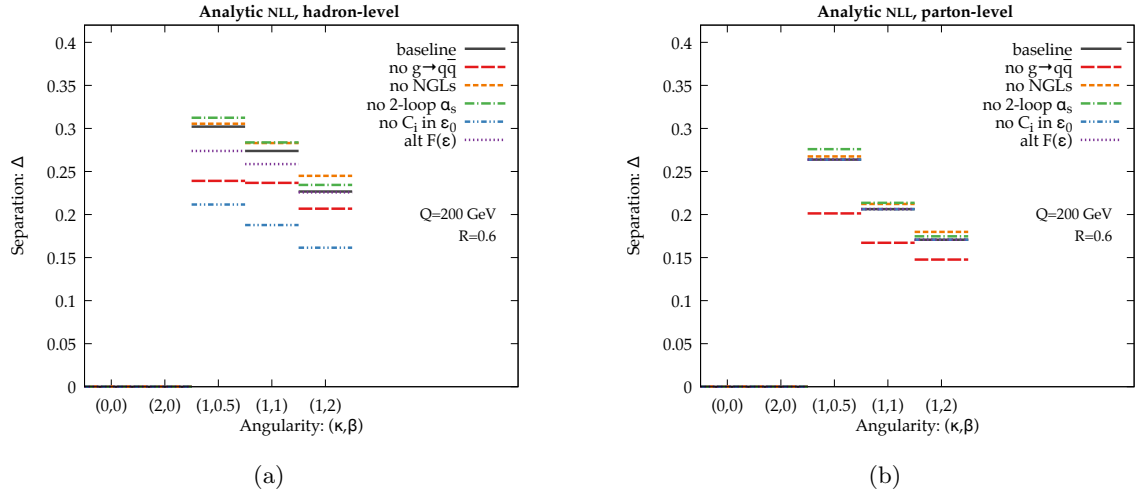


Figure 14. Same as Fig. 9, but for the analytic NLL calculation.

5.4 Testing the shape function

(Copied from previous file, need to think about. –jdt)

In Fig. 5, we saw that the analytic calculations have a noticeably broader distribution than the parton shower results. This is a consequence of two things. First, we are not including matrix element corrections nor energy-momentum conservation, such that the analytic calculations are exploring more of the phase space than they should. Second, we are using a shape function with a substantial tail (despite being exponentially suppressed). We have decided not to try to artificially fix these issues by changing the functional form of the shape function, since we have no analytic control over that choice. Despite these shape differences, the overall nonperturbative shift predicted by the shape function is in agreement with our expectations. In Fig. ??, we show the average values of e_β , before and after hadronization. (I think this is an important plot to make. I’m not sure how to do it. –jdt)

6 Quark/gluon tagging at the LHC

It is clear from our e^+e^- study that quark/gluon radiation patterns face considerable theoretical uncertainties, as seen from 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. In this section, we perform an example analysis for pp collisions that highlights the kind of information one can gain about quark/gluon radiation patterns, despite the additional complications faced by hadronic collisions.

6.1 Defining enriched samples

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. 15, the Born-level jet in $W/Z/\gamma + \text{jet}$ 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 $H + \text{jet}$ is more than 60% gluon enriched, with that fraction decreasing as the jet p_T increases. More sophisticated enrichment procedures are described in Ref. [47].

In principle, one 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. [10]). In the spirit of Sec. 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

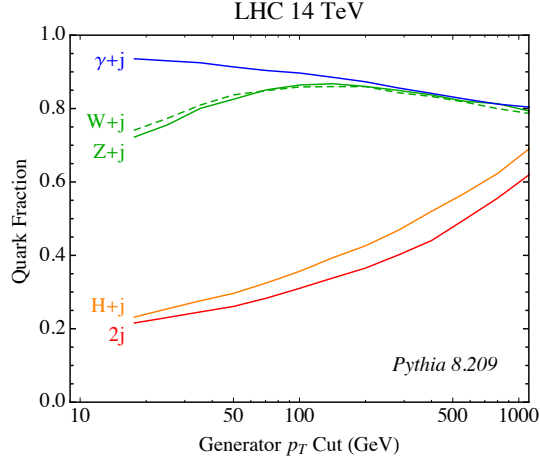


Figure 15. Quark fraction of jets at parton level, as defined by the Born-level parton flavor.

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). Similarly, one could test for differences within a jet sample, such as comparing central jets versus forward jets in dijet production. This process-based strategy is also helpful for sidestepping the known process dependence 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.

In this study, we will simply ask the question whether one can tell “the jets in Z plus jet” (quark-enriched) apart from “the jets in dijets” (gluon-enriched). **(Need to think this through. –jdt)**

(Old language –jdt) 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 range of generators (and generator modifications) would help in understanding how to improve quark/gluon modeling. The five benchmark angularities in Eq. (3.6) probe both the perturbative and nonperturbative 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. [15], though for parton shower tuning it is important to have measurements not only of jet shape averages but also of the full jet shape probability distributions.

(Old language –jdt) Finally, we think it would be useful to perform LHC jet shape measurements after jet grooming (see e.g. [85–88]). 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. [89, 90], 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 [90] remove soft radiation from jets, they tend to reduce the process dependence in quark/gluon radiation patterns, and may therefore yield a more robust theoretical definition for quark and gluon jets.

6.2 Baseline analysis

In Fig. ??, we show LHA distributions for quark-enriched and gluon-enriched samples. (Need plot to say more. –jdt)

6.3 Parameter dependence

(Do the same sweeps –jdt)

7 Summary and recommendations

(I haven’t edited this section yet. –jdt)

By measuring the substructure of jets, one can gain valuable information about the quark/gluon composition of a jet sample. The challenge we have identified in this study is that the precise radiation patterns 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 a useful guide. Therefore, LHC measurements are the best near-term strategy to constrain quark/gluon radiation patterns and enable quark/gluon discrimination to become a robust experimental 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 studied here 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 be most useful for parton shower tuning, but even detector-level measurements compared to detector-simulated parton showers would help spot troubling trends. For the IRC safe angularities in particular, studying the β dependence would help separate information about collinear and soft radiation patterns, especially given the fact that the β trends seen in the parton shower generators here disagree with those seen in Ref. [10].

If possible, it would be interesting to study the LHA ($\beta = 1/2$) on archival LEP data, since this angularity probes the core of jets in a new way, distinct from broadening-like ($\beta = 1$) or thrust-like ($\beta = 2$) observables. Among the IRC safe angularities studied here, LHA has the best predicted discrimination power, making it (and other $0 < \beta < 1$ angularities) a well-motivated target for future lepton collider measurements. Similarly, it would be worthwhile to improve our analytic understanding of the LHA. From Fig. 5c, we see that the LHA has discrimination power both at small values of $\lambda_{0.5}^1$ (where nonperturbative corrections play an

important role) as well as at larger values of $\lambda_{0.5}^1$ (where fixed-order corrections are important), so one must go beyond an NLL understanding to understand the quark/gluon performance of the LHA.

The key lesson to parton shower authors is that, contrary to some standard lore, existing LEP measurements *do 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 [15]. However, A14 has not been tested on LEP event shapes, suggesting that a global tuning strategy is needed. In addition, it is worth mentioning that similar quark/gluon studies have been carried out in deep inelastic electron-proton scattering [91], which offer an intermediate step between pp and e^+e^- collisions, and this ep data could also be valuable for parton shower tuning.

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 differences between generators came from turning on and off the $g \rightarrow q\bar{q}$ splitting process. While PYTHIA, SHERPA, VINCIA, and ARIADNE suggest that (unphysically) turning off $g \rightarrow q\bar{q}$ would improve quark/gluon separation, HERWIG (and the analytic calculation in Ref. [7]) suggests the opposite conclusion. Beyond quark/gluon discrimination, it would be helpful to identify other contexts where $g \rightarrow q\bar{q}$ 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. We also saw large changes from PYTHIA: CR1 and ARIADNE: NO SWING, suggesting that one should revisit color reconnection physics when tuning to LEP data.
- *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 , though, probably also leads to too much radiation within gluon jets.

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 the LHC.

Acknowledgments

We thank L'École de Physique des Houches and the organizers of the 2015 Les Houches workshop on “Physics at TeV Colliders” for a stimulating environment while this work was being initiated. We benefitted from helpful discussions and feedback from Samuel Bein, Andy Buckley, Jon Butterworth, Mario Campanelli, Andrew Larkoski, Peter Loch, Ben Nachman, Zoltan Nagy, Chris Pollard, Salvatore Rappoccio, Gavin Salam, Alexander Schmidt, Frank Tackmann, and Wouter Waalewijn. DK acknowledges support from the Science Faculty Research Council, University of Witwatersrand. The work of LL, SP, and AS is supported in part by the MCnetITN FP7 Marie Curie Initial Training Network, contract PITN-GA-2012-315877. LL is also support by the Swedish Research Council (contracts 621-2012-2283 and 621-2013-4287). SP acknowledges support from a FP7 Marie Curie Intra European Fellowship under Grant Agreement PIEF-GA-2013-628739. The work of DS is supported by the U.S. Department of Energy (DOE) under Grant DE-SC-0011640. The work of GS is supported by the Paris-Saclay IDEX under the IDEOPTIMALJE grant. The work of JT is supported by the DOE under cooperative research agreement DE-SC-00012567, by the DOE Early Career research program DE-SC-0006389, and by a Sloan Research Fellowship from the Alfred P. Sloan Foundation.

References

- [1] J. R. Andersen et al., *Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report*, in *9th Les Houches Workshop on Physics at TeV Colliders (PhysTeV 2015) Les Houches, France, June 1-19, 2015*, 2016. [arXiv:1605.04692](#).
- [2] J. Gallicchio and M. D. Schwartz, *Quark and Gluon Tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001, [[arXiv:1106.3076](#)].
- [3] J. Gallicchio and M. D. Schwartz, *Quark and Gluon Jet Substructure*, *JHEP* **04** (2013) 090, [[arXiv:1211.7038](#)].
- [4] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, *Jet Charge at the LHC*, *Phys. Rev. Lett.* **110** (2013), no. 21 212001, [[arXiv:1209.2421](#)].
- [5] F. Pandolfi and D. Del Re, *Search for the Standard Model Higgs Boson in the $H \rightarrow ZZ \rightarrow llqq$ Decay Channel at CMS*. PhD thesis, Zurich, ETH, 2012.
- [6] **CMS** Collaboration, S. Chatrchyan et al., *Search for a Higgs boson in the decay channel H to $ZZ^{(*)}$ to $q \bar{q} \ell^+ \ell^-$ in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **04** (2012) 036, [[arXiv:1202.1416](#)].

- [7] A. J. Larkoski, G. P. Salam, and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108, [[arXiv:1305.0007](#)].
- [8] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, *Gaining (Mutual) Information about Quark/Gluon Discrimination*, *JHEP* **11** (2014) 129, [[arXiv:1408.3122](#)].
- [9] B. Bhattacharjee, S. Mukhopadhyay, M. M. Nojiri, Y. Sakaki, and B. R. Webber, *Associated jet and subjet rates in light-quark and gluon jet discrimination*, *JHEP* **04** (2015) 131, [[arXiv:1501.04794](#)].
- [10] **ATLAS** Collaboration, G. Aad et al., *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C74** (2014), no. 8 3023, [[arXiv:1405.6583](#)].
- [11] **ATLAS** Collaboration, G. Aad et al., *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C75** (2015) 17, [[arXiv:1406.0076](#)].
- [12] **CMS** Collaboration, V. Khachatryan et al., *Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C75** (2015), no. 2 66, [[arXiv:1410.3153](#)].
- [13] **ATLAS** Collaboration, G. Aad et al., *Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **12** (2015) 055, [[arXiv:1506.00962](#)].
- [14] **CMS** Collaboration, V. Khachatryan et al., *Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$* , *Phys. Rev.* **D92** (2015), no. 3 032008, [[arXiv:1506.01010](#)].
- [15] **ATLAS** Collaboration, G. Aad et al., *Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector*, [arXiv:1602.00988](#).
- [16] H. P. Nilles and K. H. Streng, *Quark - Gluon Separation in Three Jet Events*, *Phys. Rev.* **D23** (1981) 1944.
- [17] L. M. Jones, *Tests for Determining the Parton Ancestor of a Hadron Jet*, *Phys. Rev.* **D39** (1989) 2550.
- [18] Z. Fodor, *How to See the Differences Between Quark and Gluon Jets*, *Phys. Rev.* **D41** (1990) 1726.
- [19] L. Jones, *TOWARDS A SYSTEMATIC JET CLASSIFICATION*, *Phys. Rev.* **D42** (1990) 811–814.
- [20] L. Lonnblad, C. Peterson, and T. Rognvaldsson, *Using neural networks to identify jets*, *Nucl. Phys.* **B349** (1991) 675–702.
- [21] J. Pumplin, *How to tell quark jets from gluon jets*, *Phys. Rev.* **D44** (1991) 2025–2032.
- [22] C. F. Berger, T. Kucs, and G. F. Sterman, *Event shape / energy flow correlations*, *Phys. Rev.* **D68** (2003) 014012, [[hep-ph/0303051](#)].
- [23] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, and J. Virzi, *Substructure of high- p_T Jets at the LHC*, *Phys. Rev.* **D79** (2009) 074017, [[arXiv:0807.0234](#)].

- [24] S. D. Ellis, C. K. Vermilion, J. R. Walsh, A. Hornig, and C. Lee, *Jet Shapes and Jet Algorithms in SCET*, *JHEP* **11** (2010) 101, [[arXiv:1001.0014](#)].
- [25] A. J. Larkoski, D. Neill, and J. Thaler, *Jet Shapes with the Broadening Axis*, *JHEP* **04** (2014) 017, [[arXiv:1401.2158](#)].
- [26] S. Catani, G. Turnock, and B. R. Webber, *Jet broadening measures in e^+e^- annihilation*, *Phys. Lett.* **B295** (1992) 269–276.
- [27] P. E. L. Rakow and B. R. Webber, *Transverse Momentum Moments of Hadron Distributions in QCD Jets*, *Nucl. Phys.* **B191** (1981) 63.
- [28] R. K. Ellis and B. R. Webber, *QCD Jet Broadening in Hadron Hadron Collisions*, *Conf. Proc.* **C860623** (1986) 74.
- [29] E. Farhi, *A QCD Test for Jets*, *Phys. Rev. Lett.* **39** (1977) 1587–1588.
- [30] 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, *TMVA - Toolkit for Multivariate Data Analysis*, *ArXiv Physics e-prints* (Mar., 2007) [[physics/0703039](#)].
- [31] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [[hep-ph/0603175](#)].
- [32] T. Sjostrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [[arXiv:1410.3012](#)].
- [33] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J.* **C58** (2008) 639–707, [[arXiv:0803.0883](#)].
- [34] J. Bellm et al., *Herwig++ 2.7 Release Note*, [arXiv:1310.6877](#).
- [35] J. Bellm et al., *Herwig 7.0 / Herwig++ 3.0 Release Note*, [arXiv:1512.01178](#).
- [36] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, *JHEP* **02** (2009) 007, [[arXiv:0811.4622](#)].
- [37] W. T. Giele, L. Hartgring, D. A. Kosower, E. Laenen, A. J. Larkoski, J. J. Lopez-Villarejo, M. Ritzmann, and P. Skands, *The Vincia Parton Shower*, *PoS DIS2013* (2013) 165, [[arXiv:1307.1060](#)].
- [38] Z. Nagy and D. E. Soper, *A parton shower based on factorization of the quantum density matrix*, *JHEP* **06** (2014) 097, [[arXiv:1401.6364](#)].
- [39] C. Flensburg, G. Gustafson, and L. Lonnblad, *Inclusive and Exclusive Observables from Dipoles in High Energy Collisions*, *JHEP* **08** (2011) 103, [[arXiv:1103.4321](#)].
- [40] L. Lonnblad, *ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model*, *Comput. Phys. Commun.* **71** (1992) 15–31.
- [41] S. Hche and S. Prestel, *The midpoint between dipole and parton showers*, *Eur. Phys. J.* **C75** (2015), no. 9 461, [[arXiv:1506.05057](#)].

- [42] A. Buckley, J. Butterworth, L. Lonnblad, D. Grellscheid, H. Hoeth, J. Monk, H. Schulz, and F. Siegert, *Rivet user manual*, *Comput. Phys. Commun.* **184** (2013) 2803–2819, [[arXiv:1003.0694](#)].
- [43] **ALEPH** Collaboration, A. Heister et al., *Studies of QCD at e^+e^- centre-of-mass energies between 91-GeV and 209-GeV*, *Eur. Phys. J.* **C35** (2004) 457–486.
- [44] **DELPHI** Collaboration, J. Abdallah et al., *A Study of the energy evolution of event shape distributions and their means with the DELPHI detector at LEP*, *Eur. Phys. J.* **C29** (2003) 285–312, [[hep-ex/0307048](#)].
- [45] **L3** Collaboration, P. Achard et al., *Studies of hadronic event structure in e^+e^- annihilation from 30-GeV to 209-GeV with the L3 detector*, *Phys. Rept.* **399** (2004) 71–174, [[hep-ex/0406049](#)].
- [46] **OPAL** Collaboration, G. Abbiendi et al., *Measurement of event shape distributions and moments in $e^+e^- \rightarrow \bar{q}q$ hadrons at 91-GeV - 209-GeV and a determination of $\alpha(s)$* , *Eur. Phys. J.* **C40** (2005) 287–316, [[hep-ex/0503051](#)].
- [47] J. Gallicchio and M. D. Schwartz, *Pure Samples of Quark and Gluon Jets at the LHC*, *JHEP* **10** (2011) 103, [[arXiv:1104.1175](#)].
- [48] A. Buckley and C. Pollard, *QCD-aware partonic jet clustering for truth-jet flavour labelling*, [arXiv:1507.00508](#).
- [49] M. Dasgupta, F. Dreyer, G. P. Salam, and G. Soyez, *Small-radius jets to all orders in QCD*, *JHEP* **04** (2015) 039, [[arXiv:1411.5182](#)].
- [50] A. Banfi, G. P. Salam, and G. Zanderighi, *Infrared safe definition of jet flavor*, *Eur. Phys. J.* **C47** (2006) 113–124, [[hep-ph/0601139](#)].
- [51] A. J. Larkoski and J. Thaler, *Unsafe but Calculable: Ratios of Angularities in Perturbative QCD*, *JHEP* **09** (2013) 137, [[arXiv:1307.1699](#)].
- [52] A. J. Larkoski, I. Moult, and D. Neill, *Toward Multi-Differential Cross Sections: Measuring Two Angularities on a Single Jet*, *JHEP* **09** (2014) 046, [[arXiv:1401.4458](#)].
- [53] M. Procura, W. J. Waalewijn, and L. Zeune, *Resummation of Double-Differential Cross Sections and Fully-Unintegrated Parton Distribution Functions*, *JHEP* **02** (2015) 117, [[arXiv:1410.6483](#)].
- [54] A. Hornig, Y. Makris, and T. Mehen, *Dijet Event Shapes at the LHC in SCET*, [arXiv:1601.01319](#).
- [55] W. J. Waalewijn, *Calculating the Charge of a Jet*, *Phys. Rev.* **D86** (2012) 094030, [[arXiv:1209.3019](#)].
- [56] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, *Calculating Track-Based Observables for the LHC*, *Phys. Rev. Lett.* **111** (2013) 102002, [[arXiv:1303.6637](#)].
- [57] H.-M. Chang, M. Procura, J. Thaler, and W. J. Waalewijn, *Calculating Track Thrust with Track Functions*, *Phys. Rev.* **D88** (2013) 034030, [[arXiv:1306.6630](#)].
- [58] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) 1896, [[arXiv:1111.6097](#)].

- [59] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- $k(t)$ jet clustering algorithm*, *JHEP* **04** (2008) 063, [[arXiv:0802.1189](#)].
- [60] D. Bertolini, T. Chan, and J. Thaler, *Jet Observables Without Jet Algorithms*, *JHEP* **04** (2014) 013, [[arXiv:1310.7584](#)].
- [61] G. Salam, *E_t^∞ Scheme*, *Unpublished*.
- [62] G. C. Blazey et al., *Run II jet physics*, 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](#).
- [63] Y. L. Dokshitzer, A. Lucenti, G. Marchesini, and G. P. Salam, *On the QCD analysis of jet broadening*, *JHEP* **01** (1998) 011, [[hep-ph/9801324](#)].
- [64] A. Banfi, G. P. Salam, and G. Zanderighi, *Principles of general final-state resummation and automated implementation*, *JHEP* **03** (2005) 073, [[hep-ph/0407286](#)].
- [65] M. Jankowiak and A. J. Larkoski, *Jet Substructure Without Trees*, *JHEP* **06** (2011) 057, [[arXiv:1104.1646](#)].
- [66] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N -subjettiness*, *JHEP* **03** (2011) 015, [[arXiv:1011.2268](#)].
- [67] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N -subjettiness*, *JHEP* **02** (2012) 093, [[arXiv:1108.2701](#)].
- [68] M. Dasgupta, F. A. Dreyer, G. P. Salam, and G. Soyez, *Inclusive jet spectrum for small-radius jets*, [arXiv:1602.01110](#).
- [69] A. V. Manohar and M. B. Wise, *Power suppressed corrections to hadronic event shapes*, *Phys. Lett.* **B344** (1995) 407–412, [[hep-ph/9406392](#)].
- [70] Y. L. Dokshitzer and B. R. Webber, *Calculation of power corrections to hadronic event shapes*, *Phys. Lett.* **B352** (1995) 451–455, [[hep-ph/9504219](#)].
- [71] G. P. Korchemsky and G. F. Sterman, *Power corrections to event shapes and factorization*, *Nucl. Phys.* **B555** (1999) 335–351, [[hep-ph/9902341](#)].
- [72] G. P. Korchemsky and S. Tafat, *On power corrections to the event shape distributions in QCD*, *JHEP* **10** (2000) 010, [[hep-ph/0007005](#)].
- [73] G. P. Salam and D. Wicke, *Hadron masses and power corrections to event shapes*, *JHEP* **05** (2001) 061, [[hep-ph/0102343](#)].
- [74] C. Lee and G. F. Sterman, *Momentum Flow Correlations from Event Shapes: Factorized Soft Gluons and Soft-Collinear Effective Theory*, *Phys. Rev.* **D75** (2007) 014022, [[hep-ph/0611061](#)].
- [75] V. Mateu, I. W. Stewart, and J. Thaler, *Power Corrections to Event Shapes with Mass-Dependent Operators*, *Phys. Rev.* **D87** (2013), no. 1 014025, [[arXiv:1209.3781](#)].
- [76] P. Skands, S. Carrazza, and J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, *Eur. Phys. J.* **C74** (2014), no. 8 3024, [[arXiv:1404.5630](#)].
- [77] G. Miu and T. Sjostrand, *W production in an improved parton shower approach*, *Phys. Lett.* **B449** (1999) 313–320, [[hep-ph/9812455](#)].
- [78] J. R. Christiansen and P. Z. Skands, *String Formation Beyond Leading Colour*, *JHEP* **08** (2015) 003, [[arXiv:1505.01681](#)].

- [79] S. Gieseke, C. Rohr, and A. Siodmok, *Colour reconnections in Herwig++*, *Eur. Phys. J.* **C72** (2012) 2225, [[arXiv:1206.0041](#)].
- [80] M. H. Seymour and A. Siodmok, *Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data*, *JHEP* **10** (2013) 113, [[arXiv:1307.5015](#)].
- [81] M. Bahr, M. Myska, M. H. Seymour, and A. Siodmok, *Extracting $\sigma_{effective}$ from the CDF $\gamma+3jets$ measurement*, *JHEP* **03** (2013) 129, [[arXiv:1302.4325](#)].
- [82] S. Platzer and S. Gieseke, *Dipole Showers and Automated NLO Matching in Herwig++*, *Eur. Phys. J.* **C72** (2012) 2187, [[arXiv:1109.6256](#)].
- [83] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD matrix elements + parton showers*, *JHEP* **11** (2001) 063, [[hep-ph/0109231](#)].
- [84] C. Bierlich, G. Gustafson, L. Lnnblad, and A. Tarasov, *Effects of Overlapping Strings in pp Collisions*, *JHEP* **03** (2015) 148, [[arXiv:1412.6259](#)].
- [85] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001, [[arXiv:0802.2470](#)].
- [86] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev.* **D80** (2009) 051501, [[arXiv:0903.5081](#)].
- [87] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys. Rev.* **D81** (2010) 094023, [[arXiv:0912.0033](#)].
- [88] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084, [[arXiv:0912.1342](#)].
- [89] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029, [[arXiv:1307.0007](#)].
- [90] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [[arXiv:1402.2657](#)].
- [91] **ZEUS** Collaboration, S. Chekanov et al., *Substructure dependence of jet cross sections at HERA and determination of $\alpha(s)$* , *Nucl. Phys.* **B700** (2004) 3–50, [[hep-ex/0405065](#)].