Gaining (Mutual) Information about Quark/Gluon Discrimination

Larkoski, A.J.; Thaler, J.; Waalewijn, W.J.

[Link to publication](#)

Gaining (mutual) information about quark/gluon discrimination

Andrew J. Larkoski,[a] **Jesse Thaler**[a] **and Wouter J. Waalewijn**[b,c]

[a]*Center for Theoretical Physics, Massachusetts Institute of Technology,*
Cambridge, MA 02139, U.S.A.

[b]*Nikhef, Theory Group,*
Science Park 105, 1098 XG, Amsterdam, The Netherlands

[c]*ITFA, University of Amsterdam,*
Science Park 904, 1018 XE, Amsterdam, The Netherlands

E-mail: larkoski@mit.edu, jthaler@mit.edu, wouterw@nikhef.nl

ABSTRACT: Discriminating quark jets from gluon jets is an important but challenging problem in jet substructure. In this paper, we use the concept of mutual information to illuminate the physics of quark/gluon tagging. Ideal quark/gluon separation requires only one bit of truth information, so even if two discriminant variables are largely uncorrelated, they can still share the same "truth overlap". Mutual information can be used to diagnose such situations, and thus determine which discriminant variables are redundant and which can be combined to improve performance. Using both parton showers and analytic resummation, we study a two-parameter family of generalized angularities, which includes familiar infrared and collinear (IRC) safe observables like thrust and broadening, as well as IRC unsafe variants like p_T^D and hadron multiplicity. At leading-logarithmic (LL) order, the bulk of these variables exhibit Casimir scaling, such that their truth overlap is a universal function of the color factor ratio C_A/C_F. Only at next-to-leading-logarithmic (NLL) order can one see a difference in quark/gluon performance. For the IRC safe angularities, we show that the quark/gluon performance can be improved by combining angularities with complementary angular exponents. Interestingly, LL order, NLL order, PYTHIA 8, and HERWIG++ all exhibit similar correlations between observables, but there are significant differences in the predicted quark/gluon discrimination power. For the IRC unsafe angularities, we show that the mutual information can be calculated analytically with the help of a nonperturbative "weighted-energy function", providing evidence for the complementarity of safe and unsafe observables for quark/gluon discrimination.

KEYWORDS: Jets, QCD Phenomenology

ARXIV EPRINT: 1408.3122

Contents

1 Introduction

Jets are collimated sprays of hadrons that act as proxies for short-distance quarks and gluons. Because quarks and gluons have different color charges, they have different showering and fragmentation patterns, and one can exploit this information to discriminate quark-initiated jets from gluon-initiated jets on a statistical basis. Quark/gluon discrimination is one of the key goals of the jet substructure community [1–3]. A number of quark/gluon tagging methods have been pursued [4–9], with corresponding performance studies [10–12] at the Large Hadron Collider (LHC).

Two seemingly conflicting themes have emerged from these quark/gluon discrimination studies (as well as from other tagging studies). An optimistic theme is that tagging performance can be substantially improved by combining multiple jet substructure observables, as advocated in refs. [4, 5]. A more pessimistic theme is that even if two discriminant variables are largely uncorrelated, their joint performance may not be much better than their individual performance. These dueling themes can be seen by comparing the results of recent tagging studies of boosted W bosons [13, 14]. Since quark/gluon discrimination has so many potential physics applications, it is essential to understand why both of these themes can be true.

To achieve this goal, we pursue a twofold approach in this paper. First, we use the concept of "mutual information" to illuminate the statistical aspects of quark/gluon tagging. Mutual information characterizes the correlations between variables by counting the number of shared bits of information.[1] Ideal quark/gluon tagging requires just one bit of "truth" information (e.g. $0 =$ quark and $1 =$ gluon), so even if a variable has many bits of total information, those bits may or may not have much overlap with the truth. For an observable a and an equal admixture of quarks and gluons, the mutual information with the truth is

$$I(T;A) = \int \mathrm{d}a \left(\frac{p_q(a)}{2} \log_2 \frac{p_q(a)}{p_{\text{tot}}(a)} + \frac{p_g(a)}{2} \log_2 \frac{p_g(a)}{p_{\text{tot}}(a)} \right), \tag{1.1}$$

where p_q (p_g) is the probability distribution for quarks (gluons), and $p_{\text{tot}} = (p_q + p_g)/2$. Since $0 \leq I(T;A) \leq 1$, we will sometimes refer to it as the "truth overlap". In essence, the conflicting themes above can be traced to the difference between the total information (measured by e.g. the Shannon entropy) and relevant information for quark/gluon discrimination (measured by $I(T;A)$).

Second, we will introduce a two-parameter family of discriminant variables to illuminate the physics aspects of quark/gluon tagging. We will call them "generalized angularities", which depend not only on an angular exponent $\beta \geq 0$, but also on an energy weighting factor $\kappa \geq 0$. They are defined as

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \left(\frac{R_i}{R_0} \right)^\beta, \tag{1.2}$$

---

[1]To our knowledge, the only use of mutual information in the particle physics literature is ref. [15], though it has been discussed recently in ref. [16] as a robust measure of correlations. Elsewhere in high energy physics, mutual information is used in the study of (holographic) entanglement entropy (see e.g. [17–20]).
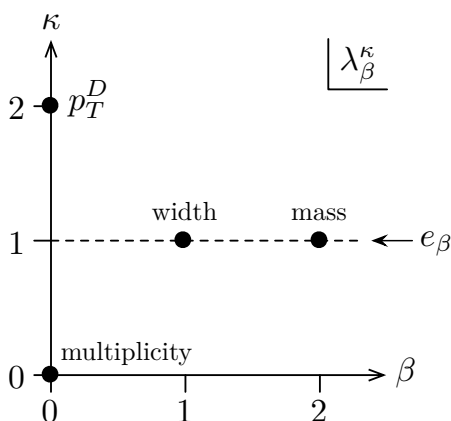
Figure 1. Visualization of the space of observables λ^κ_β, which includes several well-known jet observables used in quark/gluon discrimination: the line $\kappa = 1$ corresponds to the IRC safe angularities e_β, the origin $(\beta, \kappa) = (0,0)$ to multiplicity, and $(0,2)$ to p^D_T. Here, "width" at $(1,1)$ refers also to broadening and girth, and "mass" at $(2,1)$ refers to jet-mass-squared divided by energy (i.e. thrust).

where z_i is the momentum fraction of particle i, R_i is its rapidity/azimuth angle to a suitable axis,[2] and R_0 is the jet radius. These variables are infrared and collinear (IRC) safe only for $\kappa = 1$. As shown in figure 1, certain values of (β, κ) correspond to well-known observables: $(0,0)$ is particle multiplicity, $(0,2)$ is p^D_T [7, 8, 11], and the line $(\beta, 1)$ are the (recoil-free) angularities e_β [22, 28–30], including broadening/width/girth at $\beta = 1$ [24, 31, 32] and thrust at $\beta = 2$ [33].[3] We will present analytic calculations and parton shower simulations to understand the quark/gluon discrimination power of the λ^κ_β variables.

Our analytic calculations build on previous work calculating the substructure of quark and gluon jets [9, 29, 30, 34–36] and well as calculating angularities in e^+e^- event shapes [22, 28, 37]. For the IRC safe angularities ($\kappa = 1$), we will be able to analytically study the correlations between two angularities e_α and e_β up to next-to-leading logarithmic (NLL) accuracy. For the IRC unsafe angularities ($\kappa \neq 1$), we can use the techniques developed in refs. [6, 38–40] to introduce a new nonperturbative object called the "weighted-energy function". For $\beta = 0$, the discrimination power depends on the details of this object and only the dependence on the jet p_T and radius R_0 is calculable. However, for $\kappa\beta \gtrsim 0.5$, just the first (logarithmic) moment of this function enters at NLL. As long as these moments are sufficiently small, we can predict the quark/gluon tagging performance for an individual λ^κ_β to NLL accuracy, as well as study correlations between two IRC unsafe angularities to NLL.

---

[2]To have recoil-free observables, we use the winner-take-all axis [21–23]. The winner-take-all axis always coincides with one of the particles in the jet, so there is guaranteed to be at least one particle with $R_i = 0$. For $\beta = 0$, we define $\lambda^\kappa_0 = \sum_i z^\kappa_i$. Though we will not discuss the issue of recoil [9, 22, 24–27] in much detail, our analytic results require using recoil-free instead of recoil-sensitive angularities.

[3]The recoil-free angularities are sometimes denoted as $\tau^{(\beta)}$ [9]. The generalized angularities also have the honorific notation of ϕ^κ_β.

Some of the results in this paper are well-known to experts, though perhaps not in the language we use here. On the statistical side, tagging performance is typically shown in terms of receiver operating characteristic (ROC) curves, which show the background mistag rate for a given signal efficiency. As discussed in appendix A.1, ROC curves and mutual information are related to each other, and while ROC curves are perhaps more intuitive, mutual information has a closed form analytic definition and also has a nice visualization in terms of Venn diagrams. Furthermore, in appendix A.2, we prove that if one observable has a better ROC curve than another, then it also has a larger truth overlap, showing that $I(T; A)$ is a robust measure of tagging performance. On the physics side, λ_β^κ is only a subset of the possible quark/gluon discriminants (see ref. [5] for a catalog). Our goal is not to be exhaustive, but rather explain why different values of κ and β are sensitive to different properties of quarks and gluons. To maximize quark/gluon performance when combining variables, clearly one wants to pick variables that are sensitive to different physical effects.

Before presenting our results, we wish to make some general remarks about the definition of "quark jets" and "gluon jets". For most jet algorithms, quark and gluon jets are only well-defined at lowest order in α_s (see, however, ref. [41]). Defining quark and gluon jets to all orders in perturbation theory is a subtle and challenging problem, and becomes even more so when non-perturbative effects are taken into account. That said, quarks and gluons are well-defined if you are (deep) in the resummation region, i.e. in the limit of energetic, narrow, well-separated jets. In this regime quark/gluon radiation patterns are universal (including non-perturbative effects [42]), although at NNLL soft interference effects start playing a role, introducing a dependence on the color structure of the whole event. For the analytic predictions in this paper, we will adopt a pragmatic definition: a quark (gluon) jet is what results from the showering of a quark (gluon) parton. This was also the strategy used in ref. [9], and is sufficient to NLL accuracy. An alternative approach is to avoid trying to directly tag quarks and gluons, and instead use event categories as a well-defined proxy for jet flavor. For example, at the LHC, one can achieve a quark-enriched sample by looking at $\gamma/W/Z$ plus jet events and a gluon-enriched sample from dijet events, with further enrichment possible through judicious kinematic selections [43]. As long as one accounts for the corresponding dilution factor, the jet-based mutual information techniques in this paper will work equally well on event categories.

In section 2, we review the definition of mutual information, and use it to emphasize why the joint tagging performance of two observables is a separate concept from the correlations between two observables. In section 3, we show that for observables with "Casimir scaling" behavior, the quark/gluon truth overlap is a universal function of the color factor ratio C_A/C_F. In section 4, we discuss the general features of the λ_β^κ variables and show parton shower results for their mutual information. We then turn to analytic calculations, treating the IRC safe case of the angularities e_β in section 5 and the more general IRC unsafe case of λ_β^κ in section 6. We conclude in section 7.

2 Mutual information

2.1 Definition

Mutual information is a measure of the shared information content of two observables. For continuous distributions of variables a and b, the mutual information is (see e.g. [44])

$$I(A;B) = \int \mathrm{d}a \, \mathrm{d}b \, p(a,b) \log_2 \frac{p(a,b)}{p(a)p(b)}, \tag{2.1}$$

where $p(a,b)$ is the joint probability distribution (normalized to have unit integral), and $p(a) \equiv \int db \, p(a,b)$ and $p(b) \equiv \int da \, p(a,b)$ are the marginal probability distributions.[4] Here, we are using base 2 logarithms such that $I(A;B) = 1$ corresponds to one binary bit of shared information.

In order to visualize mutual information, it is helpful to rewrite $I(A;B)$ in terms of Shannon entropies H:

$$I(A;B) = H(A) + H(B) - H(A,B). \tag{2.2}$$

Strictly speaking, the entropy (unlike the mutual information) is not well-defined for continuous observables, though it can be made sensible by binning the distributions. For discrete-valued observables, we have

$$H(A) = -\sum_{a \in A} p(a) \log_2 p(a), \qquad H(A,B) = -\sum_{a \in A}\sum_{b \in B} p(a,b) \log_2 p(a,b), \tag{2.3}$$

such that H "counts" the number of bits of information carried by the corresponding variables. The entropies satisfy the same inequalities familiar from set theory[5]

$$0 \le H(A) \le H(A,B) \le H(A) + H(B). \tag{2.6}$$

Thus, mutual information falls in the range

$$0 \le I(A;B) \le \min\{H(A), H(B)\}. \tag{2.7}$$

As shown in figure 2a, $I(A;B)$ can be interpreted as the "area" of the intersection $A \cap B$ in information space, and it useful for quantifying the degree of correlation between two variables, with low values corresponding less correlated variables.

---

[4]The typical notation in the information theory literature is lower case symbols (a) to denote the observable and upper case symbols (A) to indicate the set of possible values, such that $a \in A$. We will keep this notation for a generic observable, but switch to all lower case in section 4 when we specialize to λ_β^κ.

[5]To derive these inequalities, consider any binned distributions $p(a)$ and $p(a,b)$. The relation $H(A) \ge 0$ follows from $0 \le p(a) \le 1$. The relation $H(A,B) \ge H(A)$ can be derived by noting that $p(a,b) \le p(a)$ for any b, and therefore

$$H(A,B) = -\sum_{a,b} p(a,b) \log_2 p(a,b) \ge -\sum_{a,b} p(a,b) \log_2 p(a) = -\sum_{a} p(a) \log_2 p(a) = H(A). \tag{2.4}$$

The relation $H(A) + H(B) \ge H(A,B)$ follows from $-\log_2 x \ge (1-x)/\ln 2$, and therefore

$$H(A) + H(B) - H(A,B) = -\sum_{a,b} p(a,b) \log_2 \frac{p(a)p(b)}{p(a,b)} \ge \frac{1}{\ln 2}\sum_{a,b} p(a,b)\left(1 - \frac{p(a)p(b)}{p(a,b)}\right) = 0. \tag{2.5}$$
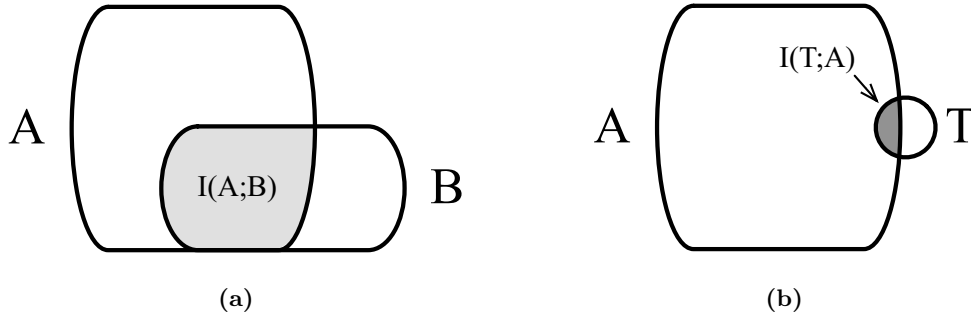
Figure 2. Left: the mutual information $I(A;B)$ between observables A and B is visualized as the area of the shaded overlap region in information space. In keeping with the set-theoretic relation in eq. (2.6), the region labelled A has area $H(A)$, the region labelled B has area $H(B)$, the union $A \cup B$ has area $H(A, B)$, and the intersection $A \cap B$ has area $I(A;B)$. Right: as a special case, we can consider the mutual information $I(T;A)$ between observable A and the truth T (i.e. the truth overlap).

2.2 Single variable discrimination

For a single variable a, we can quantify how well it performs as a signal/background discriminant by calculating how much mutual information it shares with the truth T. Consider an event sample with signal fraction f and background fraction $(1 - f)$, and let $t = 0$ for signal events and $t = 1$ for background events. Because t is a discrete variable, it has a well-defined Shannon entropy

$$H(T) = -f \log_2 f - (1 - f) \log_2(1 - f), \tag{2.8}$$

which is the number of available "truth bits". The most intuitive choice is $f = 1/2$ which yields $H(T) = 1$, corresponding to one bit of truth (i.e. signal = 0 vs. background = 1).

Without knowing the truth information, the measured a distribution would be

$$p_{\text{tot}}(a) = f \, p_0(a) + (1 - f) \, p_1(a), \tag{2.9}$$

where p_0 (p_1) is the normalized a probability distribution for signal (background) events. With the addition of truth information, the joint probability distribution is

$$p(t, a) = \delta_{t0} \, f \, p_0(a) + \delta_{t1} \, (1 - f) \, p_1(a), \tag{2.10}$$

so the mutual information between A and the truth T (i.e. the truth overlap) is

$$\begin{aligned} I(T;A) &= H(T) + H(A) - H(T, A) \\ &= \int \mathrm{d}a \left(f \, p_0(a) \log_2 \frac{p_0(a)}{p_{\text{tot}}(a)} + (1 - f) \, p_1(a) \log_2 \frac{p_1(a)}{p_{\text{tot}}(a)} \right). \end{aligned} \tag{2.11}$$

By eq. (2.7) and shown in figure 2b,

$$0 \leq I(T;A) \leq H(T). \tag{2.12}$$

The mutual information $I(T; A)$ quantifies how well the variable a can separate signal and background, with $I(T; A) = 0$ corresponding to no discrimination power and $I(T; A) = H(T)$ corresponding to full discrimination power.

In this paper, we will focus almost exclusively on $f = 1/2$ (see eq. (1.1)), such that $0 \leq I(T; A) \leq 1$. As explained in appendix A.1, by varying f, $I(T; A)$ can be related to the ROC curves typically used to quantify discrimination power. Unlike ROC curves, there is a closed-form expression for the mutual information, which makes it better suited for analytical studies. To justify that $I(T; A)$ is robust measure of tagging performance, we show in appendix A.2 that if one observable is Pareto optimal with respect to another (i.e. its ROC curve is everywhere improved), then the corresponding truth overlap is strictly larger. When eq. (2.11) is applied on a sample of events generated by a Monte Carlo program, the finite sample size requires one to replace the integral by a sum over bins. Care is needed to avoid biasing the mutual information from binning, which is discussed in detail in appendix A.3.

2.3 Pairwise correlations in discrimination

Given two variables a and b, $I(T; A)$ and $I(T; B)$ quantify how well each performs individually as signal/background discriminants. Similarly, we can assess how well a and b perform as joint discriminant variables by calculating

$$I(T; A, B) = H(T) + H(A, B) - H(T, A, B). \tag{2.13}$$

Unlike $I(A; B)$, which only tests whether or not a and b or correlated, $I(T; A, B)$ tests whether the (lack of) correlations between a and b is useful for signal/background discrimination. Note that

$$\max\{I(T, A), I(T; B)\} \leq I(T; A, B), \tag{2.14}$$

such that a and b always have the same or better joint discrimination than either variable individually. If $I(T; A, B) = \max\{I(T, A), I(T; B)\}$, then there is no gain in quark/gluon discrimination in considering both a and b, and one of the two variables is redundant, at least for this purpose.

To highlight the difference between $I(A; B)$ and $I(T; A, B)$, consider figure 3 which shows two variables with a low degree of correlation (i.e. $I(A; B)$ is relatively small). In the left example, a and b are both decent discriminant variables individually (i.e. A and B both have overlap with the truth T), but they have considerably improved joint discrimination power (i.e. $I(T; A, B)$ is larger than both $I(T; A)$ and $I(T; B)$). In the right example, despite the fact that a and b are largely uncorrelated (as measured by $I(A; B)$), there is no gain in discrimination power by considering a and b jointly (i.e. $I(T; A, B) = I(T; A) = I(T; B)$.)

In the jet substructure literature, there are known examples of both situations in figure 3. The left example is the ideal case (see e.g. [13]), where two variables a and b give complementary information for discrimination. The right example is the more puzzling case (see e.g. [14]), where two variables exhibit comparable discrimination power, a low degree of correlation, yet little gain in performance when considered jointly. Using mutual
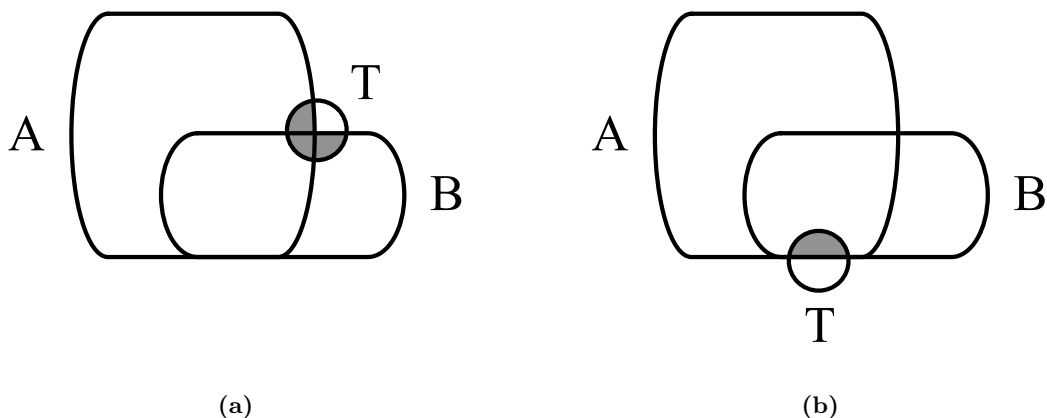
Figure 3. The mutual information $I(T; A, B)$ between observables A, B, and the truth T are shown as the area of their respective intersection. Though the mutual information between A and B (i.e. their correlation) is the same for both figures, their overlap with the truth differs. In the left figure, the mutual information with the truth (shaded) is complementary, such that combining A and B increases the truth overlap. This is not so in the right figure.

information, it is straightforward to diagnose and visualize this situation, helping to identify redundant variables. (Of course, redundant but largely uncorrelated variables can still be helpful for other purposes, such as calibration.)

3 Quark/gluon discrimination from Casimir scaling

As a simple example of using mutual information, consider an observable a that satisfies the property of "Casimir scaling". For such an observable, the normalized cumulative distribution for quarks (Σ_q) and gluons (Σ_g) can be written as

$$\Sigma_q(a) = e^{-C_F\, r(a)}, \qquad \Sigma_g(a) = e^{-C_A\, r(a)}, \tag{3.1}$$

where $r(a)$ is any monotonically decreasing function of a, $C_F = 4/3$ is the color factor for quarks, and $C_A = 3$ is the color factor for gluons. As we will see in section 5, the angularities e_β obey Casimir scaling at leading-logarithmic (LL) order. In section 6.2, we will even find that the generalized angularities λ_β^κ with $\kappa\beta \gtrsim 0.5$ obey Casimir scaling at LL as well.

As discussed in ref. [9], any observable that exhibits Casimir scaling has a ROC curve of

$$\mathrm{ROC} = x^{C_A/C_F} = x^{9/4}, \tag{3.2}$$

where x is the quark jet efficiency and x^{C_A/C_F} is the gluon jet mistag rate. This result follows from making a cut $a < a_{\mathrm{cut}}$, which keeps a fraction $\Sigma_q(a_{\mathrm{cut}})$ of quarks and a fraction $\Sigma_g(a_{\mathrm{cut}}) = (\Sigma_q(a_{\mathrm{cut}}))^{C_A/C_F}$ of gluons. Since (approximate) Casimir scaling is so ubiquitous among quark/gluon discriminants, this explains why so many discriminant variables have such similar performance. To improve performance, one has to probe the jet beyond just its overall color charge C_i.

We can understand this same feature from the point of view of mutual information by showing that the truth overlap $I(T; A)$ is a universal function of C_A/C_F for observables that exhibit Casimir scaling.[6] The probability distribution for a is just the derivative of the cumulative distribution with respect to a:

$$p_q(a) = \frac{\mathrm{d}}{\mathrm{d}a} \Sigma_q(a) = -C_F \, r'(a) \, e^{-C_F r(a)}, \tag{3.3}$$

$$p_g(a) = \frac{\mathrm{d}}{\mathrm{d}a} \Sigma_g(a) = -C_A \, r'(a) \, e^{-C_A r(a)}, \tag{3.4}$$

If f is the fraction of quark jets in the sample and $(1 - f)$ is the fraction of gluon jets, the total probability distribution is

$$p_{\mathrm{tot}}(a) = f \, p_q(a) + (1 - f) \, p_g(a), \tag{3.5}$$

and the truth overlap is

$$
\begin{aligned}
I(T; A) &= f \int \mathrm{d}a \, p_q(a) \log_2 \frac{p_q(a)}{p_{\mathrm{tot}}(a)} + (1 - f) \int \mathrm{d}a \, p_g(a) \log_2 \frac{p_g(a)}{p_{\mathrm{tot}}(a)} \\
&= f \int \mathrm{d}a \left(C_F \, r'(a) \, e^{-C_F r(a)} \right) \log_2 \left(f + (1 - f)\frac{C_A}{C_F} e^{-(C_A - C_F)r(a)} \right) \\
&\quad + (1 - f) \int \mathrm{d}a \left(C_A \, r'(a) \, e^{-C_A r(a)} \right) \log_2 \left(f \frac{C_F}{C_A} e^{-(C_F - C_A)r(a)} + (1 - f) \right). \tag{3.6}
\end{aligned}
$$

By making the change of variables

$$u \equiv e^{-C_F \, r(a)}, \tag{3.7}$$

all dependence on the distribution $r(a)$ can be removed and the integrals can be evaluated exactly. We find

$$
\begin{aligned}
I(T; A) = \frac{1}{\ln 2} \Bigg[&f \frac{(C_A - C_F)^2}{C_F C_A} \left(1 - {}_2F_1 \Big(1, \frac{C_F}{C_A - C_F}; \frac{C_A}{C_A - C_F}; \frac{(f - 1)C_A}{f C_F} \Big) \right) \\
&- f \ln \left(\frac{C_A}{C_F} - f \frac{C_A - C_F}{C_F} \right) - (1 - f) \ln \left(1 - f \frac{C_A - C_F}{C_A} \right) \Bigg], \tag{3.8}
\end{aligned}
$$

where ${}_2F_1(a, b; c; z)$ is the hypergeometric function and ln is the natural logarithm. As advertised, for any observable exhibiting Casimir scaling, $I(T; A)$ is a universal function of C_A/C_F.

Setting the quark fraction f equal to $1/2$ and $C_A/C_F = 9/4$ for QCD, the mutual information for quark/gluon discrimination is

$$I(T; A)_{f=1/2} \simeq 0.103. \tag{3.9}$$

This will be the baseline value to which all observables will be compared. Note that $I(T, A)$ is quite far from 1 (i.e. a full truth bit), demonstrating the inherent challenge of quark/gluon tagging.

---

[6]This result can also be derived from the relationship between the ROC curve and mutual information presented in appendix A.1.

4 Generalized angularities

Our analytic studies of quark/gluon separation will focus on the generalized angularities λ_β^κ defined in eq. (1.2), repeated for convenience:

$$\lambda_\beta^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \theta_i^\beta \,. \tag{4.1}$$

Here z_i is the energy fraction and $\theta_i = R_i/R_0$ the angular fraction with respect to the jet radius R_0, such that $0 \leq z_i, \theta_i \leq 1$. We measure the angles R_i with respect to the recoil-free winner-take-all axis [21–23] and we use a jet algorithm that centers the jet on the winner-take-all axis, such that $\theta_i \leq 1$ is strictly enforced. For the IRC safe angularities e_β, it is known that a recoil-free axis improves quark/gluon discrimination power [9]. For the generalized angularities λ_β^κ, a recoil-free axis is crucial for the calculations with $\beta \lesssim \kappa$, since it ensures that λ_β^κ measures the radiation pattern around the initiating hard quark or gluon and not the displacement (i.e. recoil) of the hard parton away from the jet axis.

These variables are effective quark/gluon discriminants because they probe the angular and energetic structure of jets, both of which are sensitive to the differing color factors between quarks and gluons, among other effects. Large β emphasizes wide-angle radiation whereas small β emphasizes collinear radiation. Large κ emphasizes harder hadrons, whereas small κ emphasizes softer hadrons. For reference, we highlight the $\kappa = 1$ and $\beta = 0$ cases:

$$e_\beta \equiv \lambda_\beta^1 = \sum_{i \in \text{jet}} z_i \theta_i^\beta, \tag{4.2}$$

$$\lambda_0^\kappa = \sum_{i \in \text{jet}} z_i^\kappa. \tag{4.3}$$

While $\lambda_0^1 = 1$ is a trivial observable, we can expand around $\kappa = 1$ to find

$$\lim_{\kappa \to 1} \lambda_0^\kappa = 1 + \sum_{i \in \text{jet}} (\kappa - 1) z_i \ln z_i, \tag{4.4}$$

so when we present studies for λ_0^1, we really mean $\lim_{\kappa \to 1} \lambda_0^\kappa$, which is effectively the same as the observable $\sum_{i \in \text{jet}} z_i \ln z_i$.

To get a feel for the performance of the various λ_β^κ, we can use parton shower simulations to estimate their quark/gluon truth overlap. We generate an equal admixture of quark and gluon jets (i.e. $f = 1/2$) from the processes $qq \to qq$ and $gg \to gg$ using PYTHIA 8.183 [45, 46] and HERWIG++ 2.6.3 [47, 48] at the 8 TeV LHC.[7] The transverse momenta of the jets is required to be $p_T > 400$ GeV with a jet radius of $R_0 = 0.6$. To avoid any effects from recoil [9, 24–27], we identify jets using 1-jettiness [49, 50] as a jet finder [51], taking the winner-take-all axis [21–23] as the jet center. This style of jet finding always returns one perfectly circular jet cone, and FASTJET 3.0.3 [52] code is available from the Nsubjettiness package through the FASTJET contrib project (http://fastjet.hepforge.org/contrib/).[8]

---

[7]The choice of 8 TeV allows us to use the same event sample and event selection as ref. [9]. Results at 14 TeV are qualitatively similar.

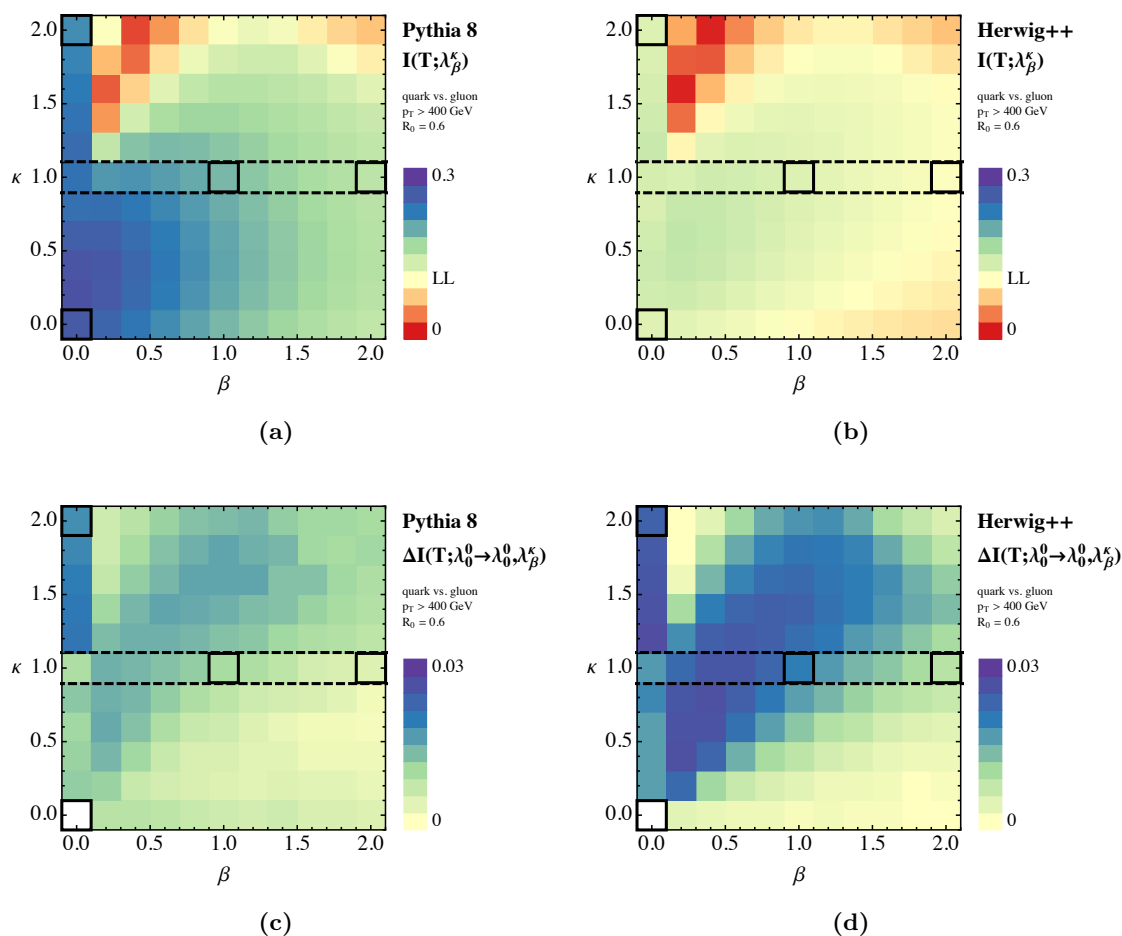[8]We thank T.J. Wilkason for providing a beta version of his code.

Figure 4. Parton shower study of quark/gluon discrimination for PYTHIA 8 (left) and HER-WIG++ (right). Top: quark/gluon discrimination power of λ_β^κ as characterized by the truth overlap $I(T; \lambda_\beta^\kappa)$. Bottom: improvement in discrimination power from supplementing multiplicity with λ_β^κ, $\Delta I(T; \lambda_0^0 \to \lambda_0^0, \lambda_\beta^\kappa) \equiv I(T; \lambda_0^0, \lambda_\beta^\kappa) - I(T; \lambda_0^0)$. The small solid boxes correspond to the dots indicated in figure 1, the wide dashed box indicates the IRC safe angularities e_β, and "LL" in light yellow indicates the result from Casimir scaling (i.e. $I(T; \lambda_\beta^\kappa) \simeq 0.1$ from eq. (3.9)).

In figures 4a and 4b, we show the truth overlap $I(T; \lambda_\beta^\kappa)$ from eq. (2.11) for different choices of λ_β^κ.[9] Confirming the results of refs. [4, 5], one of the best single discriminant variables is λ_0^0 (i.e. hadron multiplicity). In figures 4c and 4d, we show the truth gain

$$\Delta I(T; \lambda_0^0 \to \lambda_0^0, \lambda_\beta^\kappa) \equiv I(T; \lambda_0^0, \lambda_\beta^\kappa) - I(T; \lambda_0^0), \tag{4.5}$$

which is a measure of the information gain by using a second λ_β^κ in addition to λ_0^0. We see that observables like $\lambda_1^1 \equiv e_1$ (i.e. width) and λ_0^2 (i.e. p_T^D) do add additional information, in agreement with LHC performance studies [10–12].

---

[9]As discussed in appendix A.3, there is an important subtlety in calculating mutual information for binned samples with finite statistics. To avoid sample size artifacts, we use the same number of events to estimate $p_q(a)$, $p_g(a)$, and $p_{\text{tot}}(a)$.

Of course, these parton shower results should be taken as just illustrative, especially since it is known that PYTHIA 8 typically overestimates the quark/gluon separation power [12]. The differences between PYTHIA 8 and HERWIG++ are quite striking, but the origin of the disagreement is not known at present. For this reason, we want to calculate $I(T; \lambda_\beta^\kappa)$ from first principles to predict which observable (or combination of observables) has the best discrimination power, which is the subject of the next sections.

5 IRC safe angularities

We start our analytic studies with the IRC safe limit $\kappa = 1$, corresponding to the recoil-free angularities $e_\beta \equiv \lambda_\beta^1$. For all $\beta > 0$, these are IRC safe. To the order of accuracy of our calculations, e_β are identical to the energy correlation functions $C_1^{(\beta)}$ [9]. The case $\beta = 1$ is also known as width (or broadening or girth) and $\beta = 2$ is known as thrust (which is related to mass-squared at a fixed jet energy).

It was observed in ref. [9] that the recoil-free angularities are good quark/gluon discriminants, with better performance at fixed β than the traditional recoil-sensitive angularities (i.e. angularities measured with respect to the jet momentum axis). The discrimination power of e_β increased as the angular exponent β decreases towards zero, and we will verify this behavior from the mutual information viewpoint. In addition, using the double differential cross sections from refs. [53, 54], we can study the correlations between different angularities e_α and e_β to show how using additional information can improve tagging performance.

5.1 Truth overlap for one angularity

The properties of e_β are particularly simple at LL accuracy.[10] The normalized cumulative distribution of the angularity e_β was computed in, e.g. ref. [9]:

$$\Sigma_i(e_\beta) = \exp\left(-\frac{\alpha_s}{\pi} \frac{C_i}{\beta} \ln^2 e_\beta \right). \tag{5.2}$$

Here, C_i is the color of the jet: $C_F = 4/3$ for quarks and $C_A = 3$ for gluons. This distribution satisfies the Casimir scaling property of eq. (3.1), and therefore the truth overlap $I(T; e_\beta)$ is given by the formula in eq. (3.8), independent of β.

To determine the β-dependence of $I(T; e_\beta)$, we have to go to next-to-leading logarithmic (NLL) accuracy, as in ref. [9].[11] We use the NLL distributions for the recoil-free angularities computed in ref. [22] (which are identical to the NLL resummation of the energy correlation functions from ref. [26]) and compute the mutual information of the angularities with truth. For $\beta > 1$, our NLL distributions correspond to the calculations for

---

[10]We define logarithmic accuracy through the cumulative distribution of the observable of interest. For an observable e, the cumulative distribution has the expansion

$$\ln \Sigma(e) = \alpha_s \ln^2 e + \alpha_s \ln e + \alpha_s + \mathcal{O}(\alpha_s^2). \tag{5.1}$$

We define "LL" as keeping the leading terms in this expansion with the scaling $\alpha_s \ln^2 e \sim 1$.

[11]We define "NLL" as the leading terms in the expansion of eq. (5.1) with the scaling $\alpha_s \ln e \sim 1$.
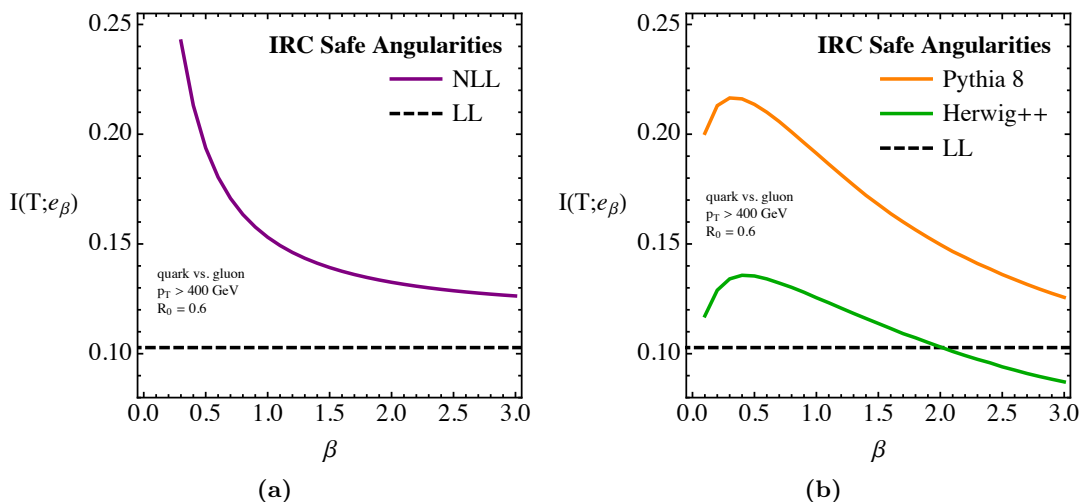
Figure 5. The quark/gluon truth overlap for an individual IRC safe angularity e_β as a function of angular exponent β. The transverse momentum of the jets is $p_T > 400\,\text{GeV}$ and the jet radius is 0.6. Left: comparing the NLL truth overlap to the baseline LL result. Right: comparing the PYTHIA 8 and HERWIG++ samples.

(recoil-sensitive) angularities performed in ref. [30]. For reference, the cumulative distributions are given in appendix B.1, and we determine $I(T; e_\beta)$ through numeric integration.

The truth overlap $I(T; e_\beta)$ as a function of the angular exponent β is shown in figure 5 for $f = 1/2$. The left plot is from the NLL calculation and the right plot shows PYTHIA 8 and HERWIG++, using the same event generation settings as in section 4 (i.e. $p_T > 400\,\text{GeV}$ and $R_0 = 0.6$). The LL result from Casimir scaling is plotted for reference. We see that $I(T; e_\beta)$ increases significantly as β decreases, showing that the quark/gluon discrimination improves. As discussed in ref. [9], the qualitative β-dependence is the same at NLL compared to the two parton shower programs, but there are significant numerical differences. Part of that is because the NLL result is lacking effects like nonperturbative power corrections which modify the quark/gluon discrimination power. The large difference between PYTHIA 8 and HERWIG++ has been seen in other contexts [12], and the underlying reason is as-of-yet unknown.

5.2 Truth overlap for two angularities

We now turn to a study of the quark/gluon discrimination power of two angularities. This will highlight the analytic benefits of using mutual information (instead of ROC curves) to study correlated observables. Constructing the ROC curve for more than a single observable is a formidable challenge because contours of constant signal/background significance can be non-trivial functions of the observables. Typically, the procedure for determining the discrimination power is to use a multivariate analysis (MVA) such as a boosted decision tree. In contrast, mutual information is defined by simply integrating over the joint probability distribution, so all correlations between observables are automatically taken into account.

At LL accuracy, the double differential cross section of two angularities was computed in ref. [53]. For angularities e_α and e_β with different angular exponents $\alpha > \beta$, the double cumulative distribution is

$$\Sigma_i(e_\alpha, e_\beta) = \exp\left[-\frac{\alpha_s}{\pi} C_i \left(\frac{\ln^2 e_\beta}{\beta} + \frac{\ln^2 \frac{e_\alpha}{e_\beta}}{\alpha - \beta} \right) \right]. \tag{5.3}$$

At this order, the angularities satisfy the inequalities

$$e_\beta > e_\alpha, \qquad (e_\alpha)^\beta > (e_\beta)^\alpha. \tag{5.4}$$

While the LL distribution does exhibit Casimir scaling, it does so for a multivariate exponential function, so the analysis of section 3 does not apply. The double differential cross section is defined by differentiating

$$\frac{\mathrm{d}^2 \sigma_i}{\mathrm{d}e_\alpha \, \mathrm{d}e_\beta} = \left(\frac{\partial^2}{\partial e_\alpha \, \partial e_\beta} \Sigma_i(e_\alpha, e_\beta) \right) \Theta_0(e_\alpha, e_\beta), \tag{5.5}$$

with explicit expressions in appendix B.2. The function Θ_0 enforces the phase space restrictions in eq. (5.4), and has to be *outside* of the derivatives.

At NLL accuracy, the (conjectured) double differential cross section was determined in ref. [54] by interpolating between effective theories at the $e_\beta = e_\alpha$ and $(e_\alpha)^\beta = (e_\beta)^\alpha$ boundaries of phase space. The NLL expression is given in appendix B.2 for reference, and an equivalent derivation using Soft-Collinear Effective Theory (SCET) [56–59] is given in appendix C.2.

In figure 6 we show the truth overlap $I(T; e_\alpha, e_\beta)$, comparing the LL expression, the NLL expression, PYTHIA 8, and HERWIG++. As before we have set the quark fraction $f = 0.5$, and the diagonal entries correspond to the single observable values from figure 5. From the baseline LL value of a single angularity in eq. (3.9) (i.e. $I(T; e_\beta) \simeq 0.1$) the truth overlap can be increased noticably even at LL. For example, for angularities e_2 and $e_{0.5}$, the joint truth overlap is greater than 0.12 at LL. At NLL, the discrimination power uniformly rises, as expected from figure 5. Because our NLL expressions do not account for the nonperturbative region of phase space, one should be cautious interpreting the results for $\beta \lesssim 0.5$. Turning to the parton showers, they give quite different prediction for $I(T; e_\alpha, e_\beta)$, with PYTHIA 8 even more optimistic than the NLL result and HERWIG++ closer to the LL result.

The large numerical differences between these methods highlights the considerable theoretical uncertainties present in quark/gluon discrimination. It is important to note that these large differences are not present when trying to model quarks and gluons individually, and only arise in the context of discrimination. In appendix D we show the mutual information $I(e_\alpha; e_\beta)$, which measures the degree of correlation between two angularities on separate quark and gluon samples. The four methods (LL, NLL, PYTHIA 8, and HERWIG++) show much closer agreement for $I(e_\alpha; e_\beta)$ than for $I(T; e_\alpha, e_\beta)$, suggesting that the truth overlap is more sensitive to subtle (and difficult to predict) differences between quark and gluon jets.
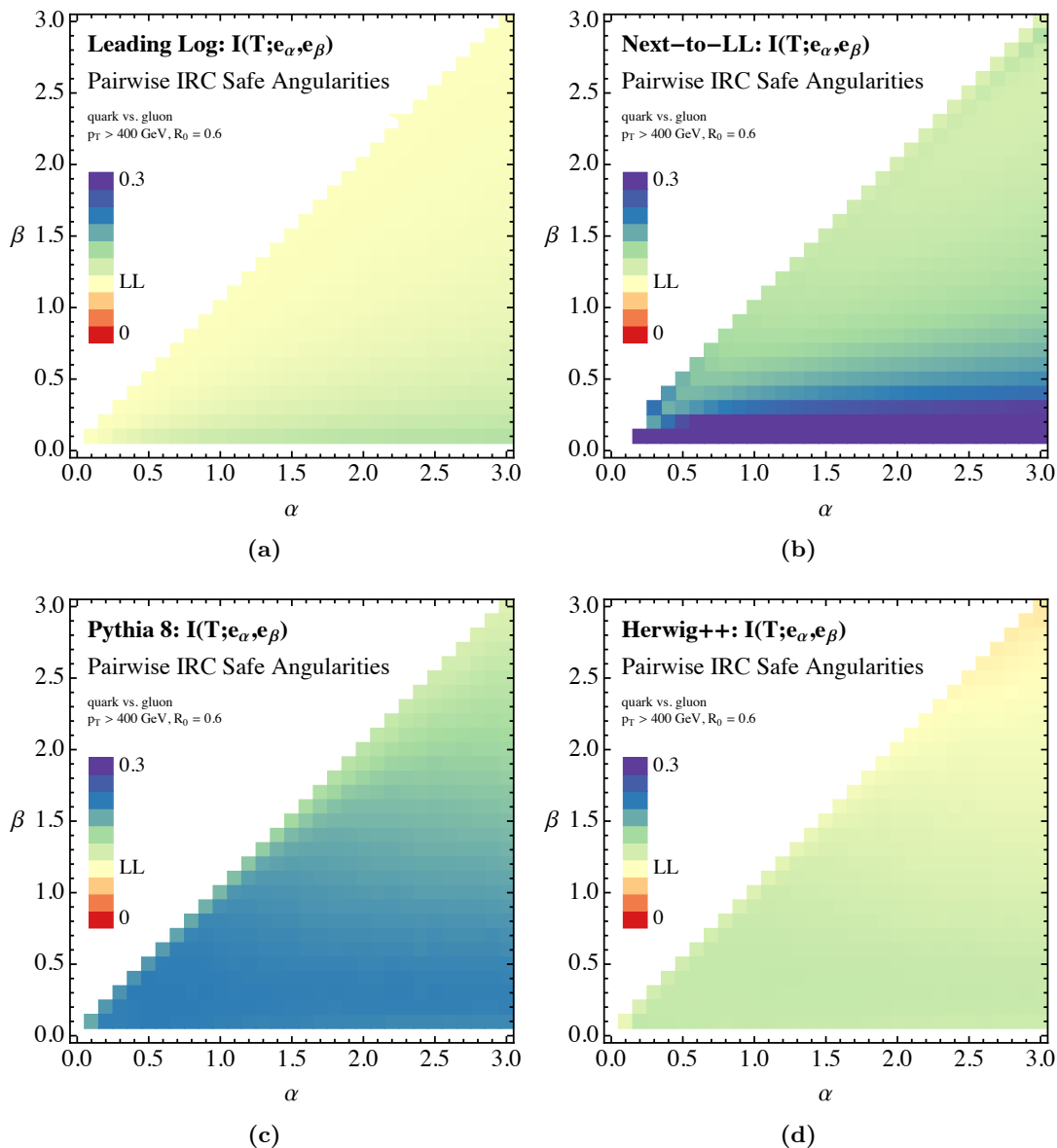
Figure 6. The quark/gluon truth overlap for pairs of IRC safe angularities (e_α, e_β). Top: the LL and NLL analytic calculations. Bottom: the PYTHIA 8 and HERWIG++ parton showers. The single observable LL baseline ($I(T; e_\beta) \simeq 0.1$) is indicated by light yellow. Note that the LL and NLL results are only trustable for $\beta \gtrsim 0.5$. Also, near the $\alpha = \beta$ diagonal, the NLL results suffer from numerical issues due to the small phase space allowed by eq. (5.4).

For completeness, in appendix D we show the truth gain

$$\Delta I(T, e_{\max} \to e_\alpha, e_\beta) \equiv I(T; e_\alpha, e_\beta) - \max\{I(T; e_\alpha), I(T; e_\beta)\}, \qquad (5.6)$$

which makes it easier to see that there is improved quark/gluon discrimination power from measuring two angularities instead of just one.[12] Roughly speaking, pairs of angularities with the smallest values of $I(e_\alpha; e_\beta)$ (i.e. least correlation) lead to the largest increase in $I(T; e_\alpha, e_\beta)$ (i.e. discrimination power), though there is considerable variability. Because the four methods have different predictions for which angularities should be combined, care should be taken when using any of these methods to estimate quark/gluon discrimination performance.

6   IRC unsafe angularities

We now turn to the more interesting case of the IRC unsafe angularities with $\kappa \neq 1$. As seen in section 4 and known in the literature, hadron multiplicity (λ_0^0) and p_T^D (λ_0^2) are effective quark/gluon discriminants. But much of the rest of the (κ, β) plane is still unexplored (apart from the angularity line at $\kappa = 1$).

One challenge to gaining an analytic understanding of the $\kappa \neq 1$ case is that λ_β^κ is collinear unsafe (unlike e_β studied above). This introduces an intrinsic sensitivity to nonperturbative physics that describes how the emitted radiation is split into hadrons, prohibiting a purely perturbative calculation. That said, using the techniques developed in refs. [6, 38–40], we can encode the nonperturbative information into a "weighted-energy function" which can be extracted from data. In fact, for $\beta > 0$, we will only need a few nonperturbative parameters (and not a whole function) to characterize the distributions.

Strictly speaking our calculations will only be valid for $\kappa \gtrsim 0.5$. The reason is that as $\kappa \to 0$ the observable also becomes infrared unsafe, further complicating calculations (as discussed in the context of hadron multiplicities in e.g. ref. [61, 62]). Also note that the $\beta = 0$ and $\beta > 0$ regimes are very different in how they treat collinear radiation, so we will consider them separately. The approximate range of validity of the calculations are shown in figure 7.

6.1   The $\beta = 0$ regime

We start with the $\beta = 0$ case with λ_0^κ. Recall from eq. (4.4) that λ_0^1 effectively refers to the observable $\sum_i z_i \ln z_i$. In order to study these observables, we need to introduce a nonperturbative object called the weighted-energy function $F_\kappa^i(x, \mu)$ that describes how the energy of a jet is distributed among its constituent hadrons. Here, i labels the flavor of the jet. This object is similar to the charge distribution [38] and the track function [39, 40] which describe other aspects of the fragmentation of quarks and gluons into hadrons.

---

[12]Of course, while mutual information is helpful to characterize the possible gains from combining observables, a multivariate analysis is still needed to realize these gains in practice. As discussed in appendix B.4, it is challenging to determine the optimal cuts analytically, even at LL. Alternatively, in the spirit of ref. [60], one could use the ratio of the quark/gluon double differential distributions as a weighting factor.
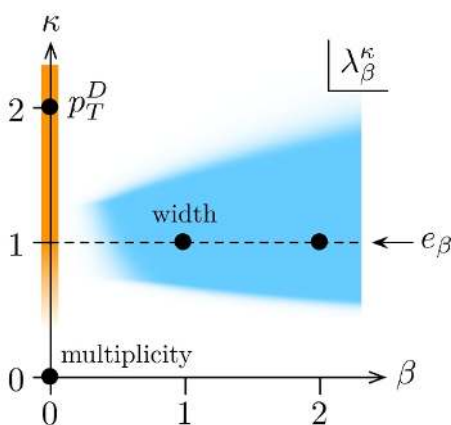
Figure 7. The regions of the space of observables λ_β^κ that we calculate are shown in orange ($\beta = 0$ and $\kappa \gtrsim 0.5$, section 6.1) and blue ($\beta\kappa \gtrsim 0.5$ and $\beta\kappa/(1-\kappa)^2 \gtrsim 6$, section 6.2). As explained in section 6.2, the funny shape of the blue region is due to a combination of perturbative and nonperturbative constraints.

The quark weighted-energy function has the following operator definition

$$F_\kappa^q(x,\mu) = \frac{1}{2N_c} \sum_H \delta\left(x - \sum_{h\in H}(z_h)^\kappa\right)$$
$$\times \operatorname{tr}\left[(\gamma^0 + \gamma^3)\left\langle 0\right|\left[(2\pi)^3\delta(k^- + \hat{p}^0 + \hat{p}^3)\delta^2(\hat{p}_\perp)\psi\right]\left|H\right\rangle\!\left\langle H|\overline{\psi}|0\right\rangle\right]. \quad (6.1)$$

Here ψ is the quark field, with momentum fixed by the δ functions involving the momentum operator $\hat{p}$, H denotes a hadronic final state, and $z_h = (p_h^0 + p_h^3)/k^-$ is the momentum fraction carried by the hadron $h \in H$. (The only dependence on k^- is through z_h.) There is a similar definition for the gluon weighted-energy function, and we have suppressed eikonal Wilson lines needed for gauge invariance. These functions are normalized such that

$$\int_0^\infty \mathrm{d}x\, F_\kappa^i(x,\mu) = 1. \quad (6.2)$$

As a point of reference, if the hadrons were weighted by their charge, then $F_\kappa^i(x)$ would be the jet charge function $D_i(x,\kappa,\mu)$ [38]. Alternatively, for $\kappa = 1$ and restricted to charged particles, this would be the track function $T_i(x,\mu)$ [39, 40].

At LO, the cross section differential in λ_0^κ for a parton of flavor i is simply

$$\frac{1}{\sigma_i}\frac{\mathrm{d}\sigma_i}{\mathrm{d}\lambda_0^\kappa} = \int \mathrm{d}x\, F_\kappa^i(x,\mu)\,\delta(\lambda_0^\kappa - x), \quad (6.3)$$

meaning that at this order, $F_\kappa^i(x,\mu)$ gives the λ_0^κ distribution directly, with $x = \lambda_0^\kappa$. The dependence on the jet p_T and jet radius R_0 enters through the scale choice $\mu = p_T R_0$. At
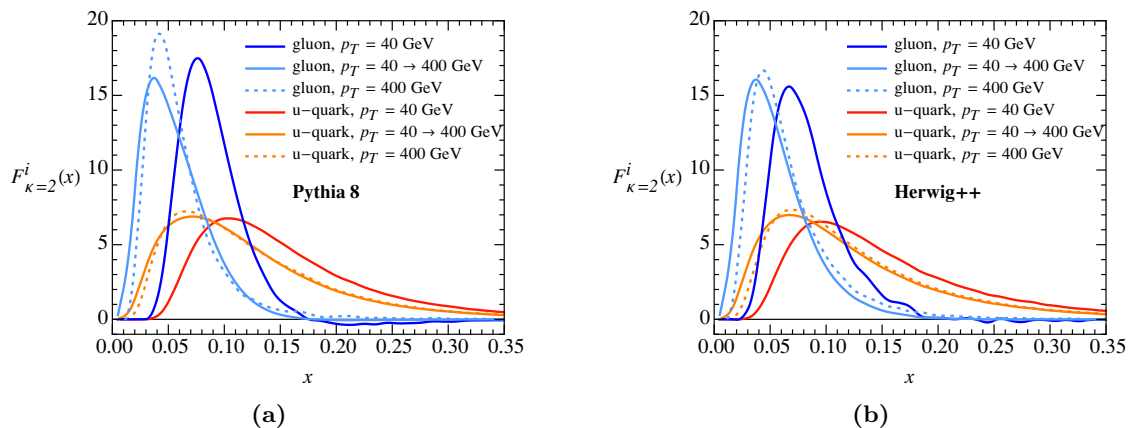
Figure 8. The weighted-energy function $F^i_{\kappa=2}(x)$ for p^D_T for u-quarks (red) and gluons (blue), extracted from PYTHIA 8 (left) and HERWIG++ (right). The darker solid curve is the parton shower results extracted at the scale $40\,\mathrm{GeV}$, the lighter solid curve is the evolution from $40\,\mathrm{GeV}$ to $400\,\mathrm{GeV}$ using eq. (6.5), and the dotted curve is the parton shower results at $400\,\mathrm{GeV}$. In all cases, we are incorporating the NLO corrections in eq. (6.4), which is why $F^i_{\kappa=2}(x)$ can be negative.

NLO, the cross section is [38]

$$\frac{1}{\sigma_i}\frac{\mathrm{d}\sigma_i}{\mathrm{d}\lambda_0^\kappa} = \frac{1}{2}\sum_{j,k}\int \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}z\,\frac{\mathcal{J}_{ij}(p_T R_0, z, \mu)}{2(2\pi)^3 J_i(p_T R_0, \mu)}$$
$$\times\, F^j_\kappa(x_1,\mu)F^k_\kappa(x_2,\mu)\,\delta\big(\lambda_0^\kappa - z^\kappa x_1 - (1-z)^\kappa x_2\big). \tag{6.4}$$

The ratio $\mathcal{J}_{ij}/J_i$ describes the perturbative splitting $i \to jk$, where j has momentum fraction z (see also refs. [63, 64]). By including these NLO corrections, the perturbative uncertainty (μ-dependence) is reduced.

The weighted-energy functions are purely nonperturbative, so in that sense, we are not really able to predict the quark/gluon discrimination power of the λ_0^κ variables. But $F^i_\kappa(x,\mu)$ does have a perturbative renormalization group evolution [38],

$$\mu\frac{\partial}{\partial\mu}F^i_\kappa(x,\mu) = \frac{1}{2}\sum_{j,k}\int \mathrm{d}z\,\mathrm{d}x_1\,\mathrm{d}x_2\,\frac{\alpha_s}{\pi}P_{i\to jk}(z)$$
$$\times\, F^j_\kappa(x_1,\mu)F^k_\kappa(x_2,\mu)\,\delta\big(x - (1-z)^\kappa x_1 - z^\kappa x_2\big). \tag{6.5}$$

Thus, one can measure $F^i_\kappa(x,\mu)$ at one scale (ideally in pure quark/gluon samples), and then evolve to a different scale. This DGLAP [65–69] evolution of F^i_κ corresponds to the emissions described by a parton shower in Monte Carlo programs. We have implemented these evolution equations for p^D_T (λ_0^2), reproducing the dependence on the jet p_T observed in PYTHIA 8 and HERWIG++, as shown in figure 8.

The weighted-energy functions are sufficient for understanding a single λ_0^κ, but if we want to study correlations between a pair of λ_0^ρ and λ_0^κ, then we would need a double weighted-energy function:
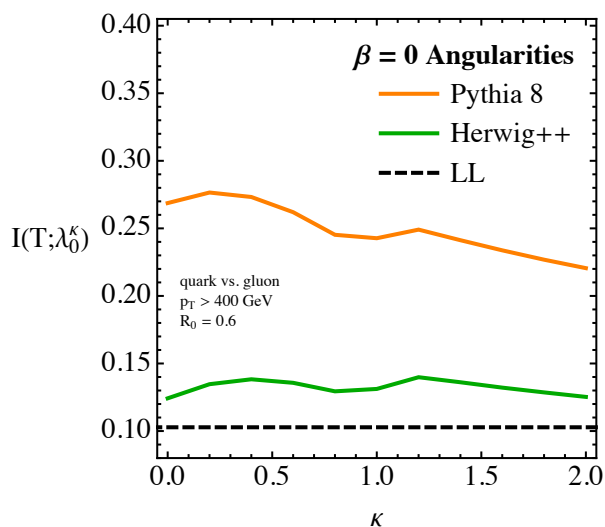
$$F^i_{\rho,\kappa}(x_1, x_2; \mu). \tag{6.6}$$

Figure 9. The quark/gluon truth overlap for an individual generalized angularity λ_0^κ as a function of the energy-weighting power κ. Here, we are comparing the PYTHIA 8 and HERWIG++ samples to the LL baseline.

This is defined analogously to eq. (6.1), albeit with the double measurement

$$\delta\Big(x_1 - \sum_{h\in H}(z_h)^\rho\Big)\,\delta\Big(x_2 - \sum_{h\in H}(z_h)^\kappa\Big). \tag{6.7}$$

This object also has a renormalization group evolution analogous to eq. (6.5). In section 6.3, we will use the fact that

$$\int \mathrm{d}x_1\, F_{\rho,\kappa}^i(x_1, x_2; \mu) = F_\kappa^i(x_2, \mu) \tag{6.8}$$

when we study correlations between the $\beta > 0$ angularities.

Since our analytic calculations are limited by our lack of knowledge of the nonperturbative function F_κ^i, we close our discussion of $\beta = 0$ by simply showing the quark/gluon truth overlap extracted from PYTHIA 8 and HERWIG++. In figure 9, we show the truth overlap of a single λ_0^κ and in figure 10 for a pair of the generalized angularities, λ_0^κ and λ_0^ρ. Both figures are quite striking in illustrating the substantial difference in discrimination power predicted by PYTHIA 8 and HERWIG++. Interestingly, the truth overlaps of PYTHIA 8 and HERWIG++ seem to be related by a simple scaling of the value, with otherwise similar structures visible over the range of observables. This might point to the source of the discrepancy in the physics descriptions between PYTHIA 8 and HERWIG++, but is beyond the scope of this paper.

6.2 The $\beta > 0$ regime

For $\beta > 0$, the generalized angularities λ_β^κ are collinear safe with respect to emissions at $\theta = 0$, but collinear unsafe with respect to wide-angle emissions. This is precisely the same situation as for the track thrust study in ref. [40], so we can adapt those methods here. In
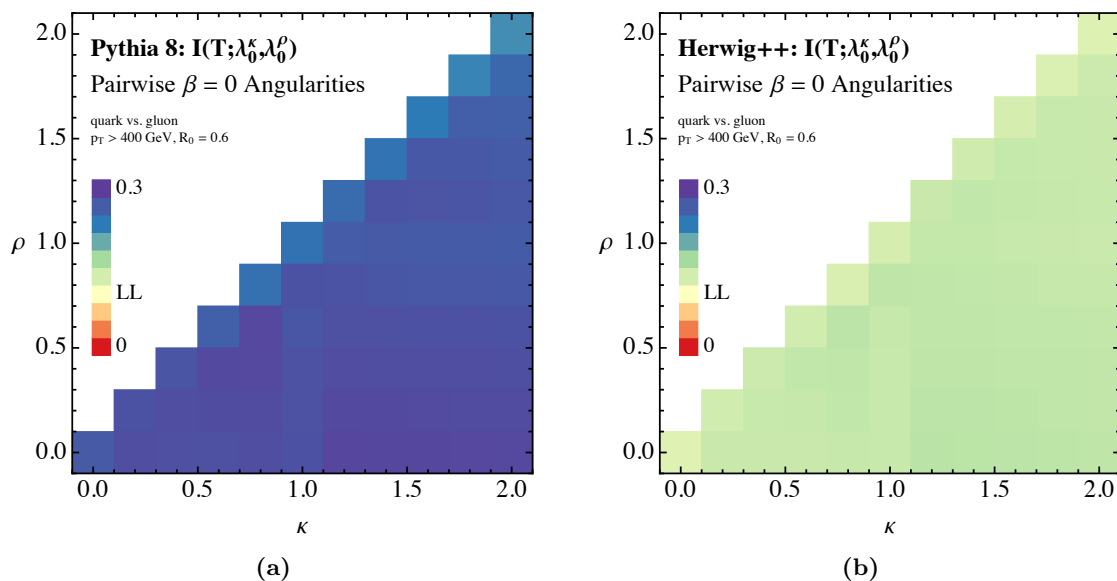
Figure 10. The quark/gluon truth overlap for pairs of $\beta = 0$ angularities $(\lambda_0^\rho, \lambda_0^\kappa)$, comparing the PYTHIA 8 and HERWIG++ parton showers.

particular, the jet can be described by a number of perturbative gluon emissions that can then be matched onto separate weighted-energy functions $F_\kappa^i(x, \mu)$. At (N)LL order, the emissions (including the weighted-energy functions) exponentiate, allowing us to predict the performance of λ_β^κ for quark/gluon discrimination.

A jet with a single parton has $\lambda_\beta^\kappa = 0$, since the reference axis will align with that parton. For two partons, the winner-take-all axis will align with the harder parton. Ignoring fixed-order corrections, we can assume that the harder parton is the initiating parton (i.e. quark for a quark jet) such that the λ_β^κ distribution is determined by the emitted soft gluon. In the LO approximation for narrow jets, we can use splitting functions

$$\frac{1}{\sigma_i} \frac{d\sigma_i}{d\lambda_\beta^\kappa} = \int_0^1 \frac{d\theta}{\theta} \int_0^1 dz \, \frac{\alpha_s}{\pi} P_{i \to ig}(z) \int_0^\infty dx \, F_\kappa^g(x, \mu) \, \delta(\lambda_\beta^\kappa - x z^\kappa \theta^\beta), \qquad (6.9)$$

where $P_{i \to ig}(z)$ is the splitting function for parton flavor i to emit a soft gluon with momentum fraction z.

We can achieve (N)LL resummation by considering the strongly-ordered limit where the λ_β^κ distribution is determined by the hardest emissions in the jet. For the strongly-ordered limit to make sense, we assume that $F_\kappa^g(x, \mu)$ is non-singular and does not have support over a hierarchically large range in x, such that the emission with, say, the largest value of $z^\kappa \theta^\beta$ also typically has the largest value of λ_β^κ. This then allows us to use the logic of CAESAR [26] to determine the λ_β^κ distribution up to NLL order (ignoring non-global effects [70]).

In the CAESAR approach, the one-gluon distribution in eq. (6.9) is interpreted as the

radiator function:

$$R_i(\lambda_\beta^\kappa) = \int_0^1 \frac{\mathrm{d}\theta}{\theta} \int_0^1 \mathrm{d}z \, \frac{\alpha_s(p_T z\theta)}{\pi} P_{i\to ig}(z) \int_0^\infty \mathrm{d}x \, F_\kappa^g(x,\mu) \, \Theta(xz^\kappa\theta^\beta - \lambda_\beta^\kappa), \qquad (6.10)$$

where α_s is now a running coupling evaluated to two-loop order in the CMW scheme [71]. This yields the cumulative distribution accurate to NLL

$$\Sigma_i(\lambda_\beta^\kappa) = \frac{e^{-\gamma_E R_i'(\lambda_\beta^\kappa)}}{\Gamma(1 + R_i'(\lambda_\beta^\kappa))} e^{-R_i(\lambda_\beta^\kappa)}, \qquad (6.11)$$

where γ_E is the Euler-Mascheroni constant, Γ is the gamma function, and $R_i' \equiv -\mathrm{d}R_i/\mathrm{d}\ln\lambda_\beta^\kappa$ is the logarithmic derivative. The cross section is obtained via

$$\frac{1}{\sigma_i} \frac{\mathrm{d}\sigma_i}{\mathrm{d}\lambda_\beta^\kappa} = \frac{\mathrm{d}}{\mathrm{d}\lambda_\beta^\kappa} \Sigma_i(\lambda_\beta^\kappa). \qquad (6.12)$$

We can already learn a lot from eq. (6.11) by considering just the LL limit. In that limit, we can drop the prefactor terms involving R', fix the coupling α_s, and take $P_{i\to ig}(z) = 2C_i/z$. In that case we find

$$R_i(\lambda_\beta^\kappa) \stackrel{\mathrm{LL}}{\simeq} \frac{\alpha_s}{\pi} \frac{C_i}{\beta\kappa} \int_{\lambda_\beta^\kappa}^\infty \mathrm{d}x \, F_\kappa^g(x,\mu) \ln^2 \frac{\lambda_\beta^\kappa}{x}. \qquad (6.13)$$

Because the bounds of integration for x depends on the value of the observable λ_β^κ, this integral cannot be simplified. However, because we assume that $F_\kappa^g(x,\mu)$ is non-singular, for small λ_β^κ we can expand the integral in powers of λ_β^κ. To leading logarithmic accuracy we then have

$$R_i(\lambda_\beta^\kappa) \stackrel{\mathrm{LL}}{\simeq} \frac{\alpha_s}{\pi} \frac{C_i}{\beta\kappa} \left(\ln^2 \lambda_\beta^\kappa - 2f_\kappa^{g,1} \ln\lambda_\beta^\kappa + f_\kappa^{g,2} \right), \qquad (6.14)$$

where other terms are suppressed by powers of λ_β^κ. The logarithmic moments are defined as

$$f_\kappa^{g,n} \equiv \int_0^\infty \mathrm{d}x \, F_\kappa^g(x,\mu) \ln^n x. \qquad (6.15)$$

In figure 11, we show values of $f_\kappa^{g,i}$ extracted from PYTHIA 8 and HERWIG++ for a range of κ values. These moments are quite similar between the two parton showers, suggesting that their extraction is robust.

Like in ref. [40], we find that logarithmic moments of the nonperturbative function appears in the exponent of the cumulative distribution. Note that the only difference between quark and gluon jets is the color factor C_i, since the same gluon-based $f_\kappa^{g,n}$ appears for both kinds of jets. Thus, this observable satisfies Casimir scaling in the LL limit, yielding the mutual information discussed in section 3. Strictly speaking, these $f_\kappa^{g,n}$ terms are only relevant at NLL order, since they multiply at most single logarithms in the observable.[13] Therefore, we will drop the $f_\kappa^{g,n}$ terms at LL order:

$$R_i^{\mathrm{LL}}(\lambda_\beta^\kappa) = \frac{\alpha_s}{\pi} \frac{C_i}{\beta\kappa} \ln^2 \lambda_\beta^\kappa. \qquad (6.16)$$

---

[13]One could imagine power counting $f_\kappa^{g,n}$ as being logarithmically enhanced instead of as $\mathcal{O}(1)$. In that case, however, one would need to keep track of every $f_\kappa^{g,n}$ starting at NLL order, so we effectively return to the $\beta = 0$ case where the full $F_\kappa^g(x,\mu)$ function is needed.
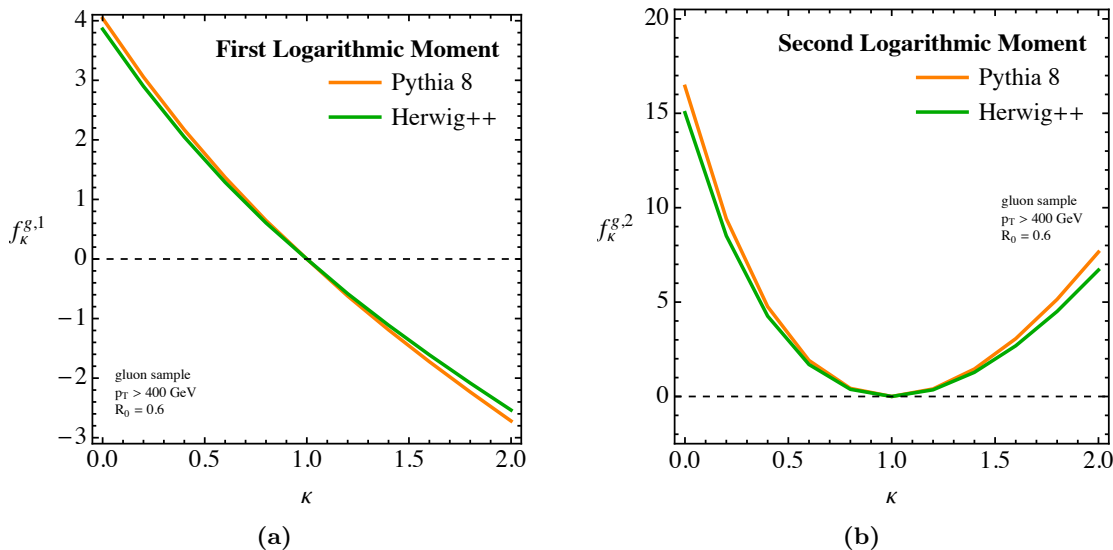
Figure 11. Extracting the logarithmic moments $f_\kappa^{g,1}$ (left) and $f_\kappa^{g,2}$ (right) defined in eq. (6.15) from the PYTHIA 8 and HERWIG++ gluon samples.

Doing the full NLL calculation using eq. (6.11) is straightforward with the help of two tricks. First, using the fact that

$$\Theta(xz^\kappa\theta^\beta - \lambda_\beta^\kappa) = \Theta\left(z\theta^{\beta/\kappa} - \left(\frac{\lambda_\beta^\kappa}{x}\right)^{1/\kappa}\right), \tag{6.17}$$

we can rewrite the radiator function in eq. (6.10) as

$$R_i(\lambda_\beta^\kappa) = \int_0^\infty \mathrm{d}x\, F_\kappa^g(x,\mu)\, \hat{R}_i\left(e_{\beta/\kappa} = \left(\frac{\lambda_\beta^\kappa}{x}\right)^{1/\kappa}\right), \tag{6.18}$$

where $\hat{R}_i$ is the radiator for the IRC safe angularity with exponent β/κ. Second, we only need to keep the first logarithmic moment $f_\kappa^{g,1}$ at NLL order, so when we do the x integral weighted by $F_\kappa^g(x,\mu)$, we can effectively replace

$$x \to \exp(f_\kappa^{g,1}) \tag{6.19}$$

in the argument of $\hat{R}_i$, up to log-suppressed terms.[14] Thus, the IRC unsafe radiator is simply

$$R_i^{\mathrm{NLL}}(\lambda_\beta^\kappa) = \hat{R}_i^{\mathrm{NLL}}\left(e_{\beta/\kappa} = \left(\frac{\lambda_\beta^\kappa}{\exp(f_\kappa^{g,1})}\right)^{1/\kappa}\right), \tag{6.20}$$

---

[14]Note that this rescaling would replace $f_\kappa^{g,2}$ with $(f_\kappa^{g,1})^2$ in eq. (6.14). Because the $f_\kappa^{g,2}$ term is formally beyond NLL accuracy, this is an allowed replacement.

where the IRC safe radiator $\hat{R}_i$ is given in appendix B.1. We find it quite remarkable that we can relate an IRC unsafe distribution to an IRC safe one in this way, and we show that these same two tricks are valid in SCET in appendix C.1.[15]

Before showing analytic results, we need to comment on the range of validity of our calculation. Because the radiator scales like $1/(\beta\kappa)$, we can only trust this perturbative expression when $\beta\kappa \gtrsim 0.5$. In addition, the validity of our approach is limited to the region where, in absolute terms, the nonperturbative parameter $f_\kappa^{g,1}$ is smaller than the typical values of $\ln \lambda_\beta^\kappa$.[16] From figure 11 we obtain the approximation $f_\kappa^{g,1} \approx 3(1-\kappa)$, and using eq. (6.16), we expect the distribution to peak at $\ln^2 \lambda_\beta^\kappa = \beta\kappa/(2\alpha_s C_i)$. This suggests that our calculation holds for

$$\frac{\beta\kappa}{(1-\kappa)^2} > c, \tag{6.21}$$

where $c = 18\alpha_s C_i$, which is $\simeq 2.7$ for quarks and $\simeq 6.0$ for gluons. We take the more restrictive value $c = 6.0$ when assessing the validity of the quark/gluon truth overlap, and this is the reason why the blue region in figure 7 is missing the upper left and lower right corners.

In figure 12, we show the truth overlap $I(T; \lambda_\beta^\kappa)$, comparing LL, NLL, PYTHIA 8, and HERWIG++. As expected, there is no difference between different choices of κ and β at LL, with differences showing up first at NLL order. The NLL calculation breaks down in the upper left region (due to large values of $f_\kappa^{g,1}$) and in the lower left region (due to $\beta\kappa$ being too small). Unfortunately, these are exactly the same regions where there is interesting behavior in the PYTHIA 8 and HERWIG++ predictions. If we naively trust the NLL results outside of their range of validity, then starting from broadening (λ_1^1) and approaching multiplicity (λ_0^0), the NLL results show an increase in discrimination power, in agreement with the parton showers. However approaching p_T^D (λ_0^2), the NLL results also show an increase in discrimination power, which is the opposite behavior from the parton showers (until one reaches the actual $\beta = 0$ line). Of course, one should be wary of this extrapolation, since our calculations are lacking important nonperturbative corrections.[17]

---

[15]It is perhaps even more remarkable that we can take logarithmic derivatives of the resulting R_i and get the right NLL expression for the multiple emissions piece. Ultimately, the only way we are able to justify this is via SCET, since the original CAESAR approach was only proven for IRC safe observables. Note that there is a Jacobian factor in R_i', so you cannot directly relate the cumulative distributions for $e_{\beta/\kappa}$ and $\lambda_\beta^\kappa/\exp(f_\kappa^{g,1})$, only the radiators. Because of the R_i' term, the discrimination power is not just a function of β/κ, and has non-trivial κ and $f_\kappa^{g,1}$ dependence at NLL.

[16]Outside of this region, the nonperturbative effects become too large to be treated using just the logarithmic moments, and we would have to include the full $F_\kappa^g(x, \mu)$ function, as also mentioned in footnote 13. In principle, this would allow us to get beyond the "nonperturbative barrier" in eq. (6.21), though we have not attempted such a calculation.

[17]Intuitively, the parton shower results make sense. Compared to gluons, quarks typically have smaller values of λ_1^1 but larger values of λ_0^2. Thus, interpolating between λ_1^1 and λ_0^2 should yield a poor discriminant variable. In the NLL approach, small β is always favored, and calculational control is lost before the $f_\kappa^{g,1}$ parameter has a chance to reverse that trend.
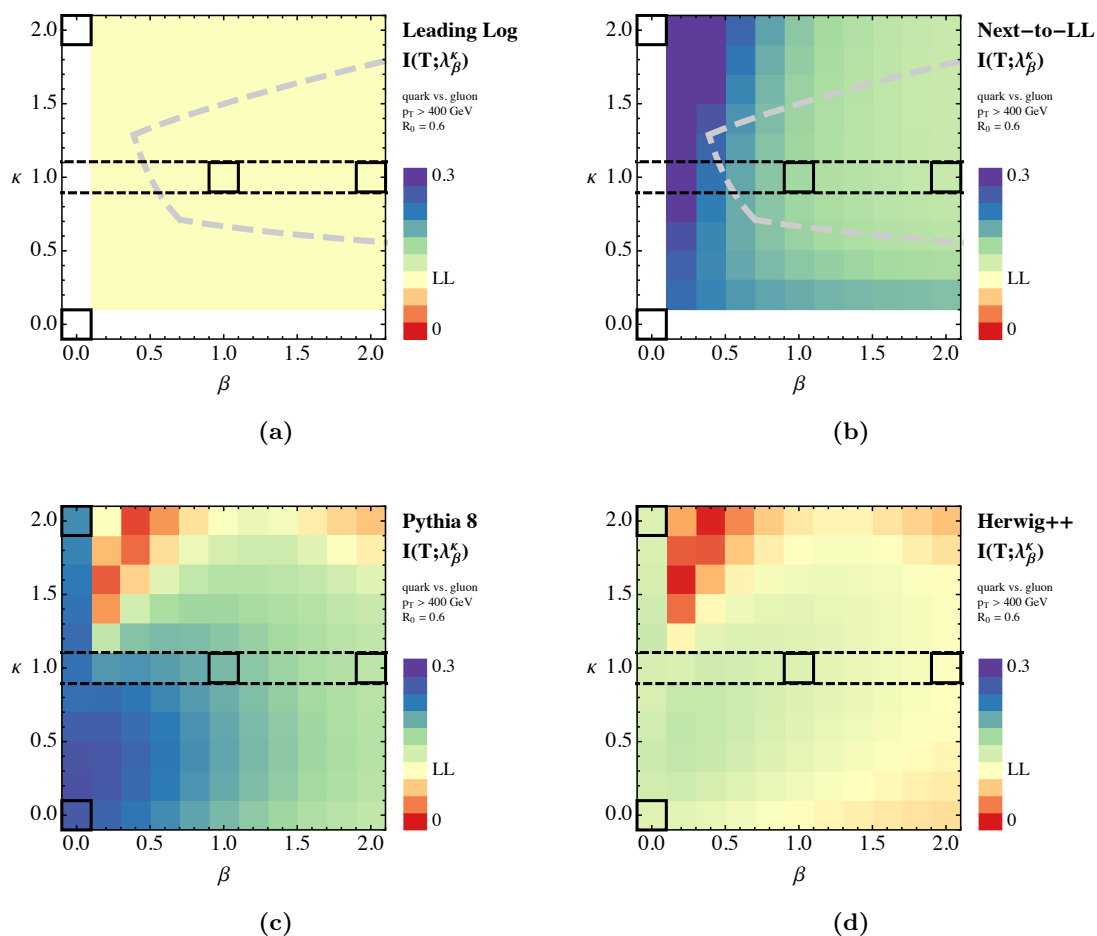
Figure 12. The quark/gluon truth overlap for an individual generalized angularity λ_β^κ. Top: the LL and NLL analytic calculations. Note that these calculations are singular at $\kappa = 0$ or $\beta = 0$. Bottom: the PYTHIA 8 and HERWIG++ parton showers (identical to figures 4a and 4b). The solid boxes correspond to dots indicated in figure 1, the dashed box corresponds to the IRC safe angularities e_β, and the grey dashed curve in the LL/NLL plots marks the range of validity of our calculations (i.e. the edge of the blue region in figure 7).

6.3 Two $\beta > 0$ angularities

Because the resummed λ_β^κ distributions for $\beta > 0$ only depend on logarithmic moments of the weighted-energy function, we have an opportunity to analytically study the correlations between two generalized angularities λ_α^ρ and λ_β^κ. We can already gain a lot of insight from a LL study, and we can use the same tricks as for eq. (6.20) to obtain an NLL result.

With the help of the double weighted-energy function in eq. (6.6), we can define a double radiator function

$$
\begin{aligned}
R_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa) = \int_0^1 \frac{\mathrm{d}\theta}{\theta} \int_0^1 \mathrm{d}z \, \frac{\alpha_s(p_T z\theta)}{\pi} P_{i\to ig}(z) \int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2 \, F_{\rho,\kappa}^g(x_1, x_2; \mu) \\
\times \left[1 - \Theta(\lambda_\alpha^\rho - x_1 z^\rho \theta^\alpha) \Theta(\lambda_\beta^\kappa - x_2 z^\kappa \theta^\beta) \right].
\end{aligned}
\tag{6.22}
$$

Again assuming that $F_{\rho,\kappa}^g$ does not have any large hierarchies (such that the observables are dominated by the hardest emissions), then we can follow the logic of ref. [53] and say that at LL accuracy

$$\frac{1}{\sigma_i}\frac{\mathrm{d}^2\sigma_i}{\mathrm{d}\lambda_\alpha^\rho\,\mathrm{d}\lambda_\beta^\kappa} = \left(\frac{\partial^2}{\partial\lambda_\alpha^\rho\,\partial\lambda_\beta^\kappa}e^{-R_i(\lambda_\alpha^\rho,\lambda_\beta^\kappa)}\right)\Theta_0(\lambda_\alpha^\rho,\lambda_\beta^\kappa), \tag{6.23}$$

where $\Theta_0(\lambda_\alpha^\rho,\lambda_\beta^\kappa)$ enforces the phase space restrictions[18]

$$\left(\frac{(\lambda_\alpha^\rho)^\kappa}{(\lambda_\beta^\kappa)^\rho}\right)^{\mathrm{sign}(\alpha/\rho-\beta/\kappa)} \leq 1, \qquad \left(\frac{(\lambda_\beta^\kappa)^\alpha}{(\lambda_\alpha^\rho)^\beta}\right)^{\mathrm{sign}(\alpha/\rho-\beta/\kappa)} \leq 1. \tag{6.24}$$

This expression does not immediately generalize to NLL accuracy, since there is no (known) factorization theorem for double differential distributions over the full phase space. Instead we will exploit the interpolation technique of ref. [54] to help find the NLL expression, as we did in section 5.2.

Calculating the double radiator in the LL limit, we find for $\alpha/\rho > \beta/\kappa$

$$R(\lambda_\alpha^\rho,\lambda_\beta^\kappa) \overset{\mathrm{LL}}{\simeq} \frac{\alpha_s}{\pi}\frac{C_i}{\alpha\kappa-\beta\rho}\int \mathrm{d}x_1\,\mathrm{d}x_2\, F_{\rho,\kappa}^g(x_1,x_2;\mu)\left(\frac{\alpha}{\beta}\ln^2\frac{\lambda_\beta^\kappa}{x_2} + \frac{\kappa}{\rho}\ln^2\frac{\lambda_\alpha^\rho}{x_1} - 2\ln\frac{\lambda_\beta^\kappa}{x_2}\ln\frac{\lambda_\alpha^\rho}{x_1}\right)$$

$$= \frac{\alpha_s}{\pi}\frac{C_i}{\alpha\kappa-\beta\rho}\Bigg[\frac{\alpha}{\beta}\left(\ln^2\lambda_\beta^\kappa - 2f_\kappa^{g,1}\ln\lambda_\beta^\kappa + f_\kappa^{g,2}\right) + \frac{\kappa}{\rho}\left(\ln^2\lambda_\alpha^\rho - 2f_\rho^{g,1}\ln\lambda_\alpha^\rho + f_\rho^{g,2}\right)$$

$$- 2\left(\ln\lambda_\beta^\kappa\ln\lambda_\alpha^\rho - f_\kappa^{g,1}\ln\lambda_\alpha^\rho - f_\rho^{g,1}\ln\lambda_\beta^\kappa + f_{\rho,\kappa}^{g,1,1}\right)\Bigg], \tag{6.25}$$

where we have used eq. (6.8) to simplify the last line (accurate to leading power in λ_α^ρ and λ_β^κ) and we have defined

$$f_{\rho,\kappa}^{g,1,1} = \int \mathrm{d}x_1\,\mathrm{d}x_2\, F_{\rho,\kappa}^g(x_1,x_2;\mu)\ln x_1\ln x_2. \tag{6.26}$$

As in the case of a single λ_β^κ, strictly speaking the nonperturbative parameters $f_{\rho,\kappa}^{g,n}$ only show up at subleading logarithmic order, and can be ignored to LL accuracy,

$$R_i^{\mathrm{LL}}(\lambda_\alpha^\rho,\lambda_\beta^\kappa) = \frac{\alpha_s}{\pi}\frac{C_i}{\alpha\kappa-\beta\rho}\left(\frac{\alpha}{\beta}\ln^2\lambda_\beta^\kappa + \frac{\kappa}{\rho}\ln^2\lambda_\alpha^\rho - 2\ln\lambda_\beta^\kappa\ln\lambda_\alpha^\rho\right). \tag{6.27}$$

Note that the exponents ρ and κ still have an effect on the discrimination power even though we are not accounting for the nonperturbative parameters at this order. The overall prefactor implies that when α and β are sufficiently different, we should only trust this distribution for

$$\alpha\kappa - \beta\rho \gtrsim 0.5. \tag{6.28}$$

---

[18]These assume that nonperturbative physics do not affect the phase space, which is fine for LL accuracy. At NLL, we use eq. (6.33) to adjust the phase space given the first logarithmic moments of the weighted-energy function.

To achieve NLL accuracy, we need to combine the interpolation technique of ref. [54] with the two tricks as we used to find the single generalized angularity distribution in eq. (6.20). Following ref. [54], we take the following ansatz for the NLL distribution:

$$\Sigma_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa) = \frac{e^{-\gamma_E \widetilde{R}_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa)}}{\Gamma(1 + \widetilde{R}_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa))} e^{-R_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa)}, \tag{6.29}$$

where R_i is the double radiator from eq. (6.22) and $\widetilde{R}_i$ is a multiple emissions term that is *not* given by any simple logarithmic derivative of R_i. We start by considering the double radiator R_i. Looking at the theta functions in eq. (6.22), we see that the integration range for the double radiator is the same as for two IRC safe angularities with

$$e_{\alpha/\rho} = \left(\frac{\lambda_\alpha^\rho}{x_1}\right)^{1/\rho}, \qquad e_{\beta/\kappa} = \left(\frac{\lambda_\beta^\kappa}{x_2}\right)^{1/\kappa}. \tag{6.30}$$

At NLL accuracy, we only need the first logarithmic moments of the weighted-energy function, so we can make the replacement

$$x_1 \to \exp(f_\rho^{g,1}), \qquad x_2 \to \exp(f_\kappa^{g,1}). \tag{6.31}$$

Thus, the double radiator is

$$R_i^{\mathrm{NLL}}(\lambda_\alpha^\rho, \lambda_\beta^\kappa) = \hat{R}_i^{\mathrm{NLL}} \left(e_{\alpha/\rho} = \left(\frac{\lambda_\alpha^\rho}{\exp(f_\rho^{g,1})}\right)^{1/\rho}, e_{\beta/\kappa} = \left(\frac{\lambda_\beta^\kappa}{\exp(f_\kappa^{g,1})}\right)^{1/\kappa} \right), \tag{6.32}$$

where $\hat{R}_i^{\mathrm{NLL}}$ is the IRC safe double cumulative distribution for angularities with exponents α/ρ and β/κ, defined in appendix B.2. Turning to the multiple emissions term $\widetilde{R}_i$, we need to find a function that interpolates between the logarithmic derivative functions $R'(\lambda_\alpha^\rho)$ and $R'(\lambda_\beta^\kappa)$ on the boundaries of phase space. Since we already found the single radiator to NLL accuracy in eq. (6.20), this interpolation is straightforward, and we give the explicit expressions in appendix B.3. Finally, to have a properly normalized distribution, we have to apply the rescaling

$$\lambda_\alpha^\rho \to \frac{\lambda_\alpha^\rho}{\exp(f_\rho^{g,1})}, \qquad \lambda_\beta^\kappa \to \frac{\lambda_\beta^\kappa}{\exp(f_\kappa^{g,1})} \tag{6.33}$$

to the phase space constraints in eq. (6.24) as well. Again, we find it remarkable that there is such a close relationship between IRC safe and IRC unsafe calculations at NLL order, and we validate this method in SCET in appendix C.2.

In figure 13, we show the truth overlap for the LL and NLL calculations, compared to results obtained from PYTHIA 8 and HERWIG++. The LL and NLL calculations do not extend to the region where α or β is zero, which are left white in the plot. We caution the reader that some of these LL and NLL results extrapolate outside the range of validity in eqs. (6.21) and (6.28). Comparing the various predictions, we observe similarities between the regions of minimal and maximal discrimination power, though the overall discrimination power is (again) larger for PYTHIA 8 and NLL than for HERWIG++ and LL. The PYTHIA
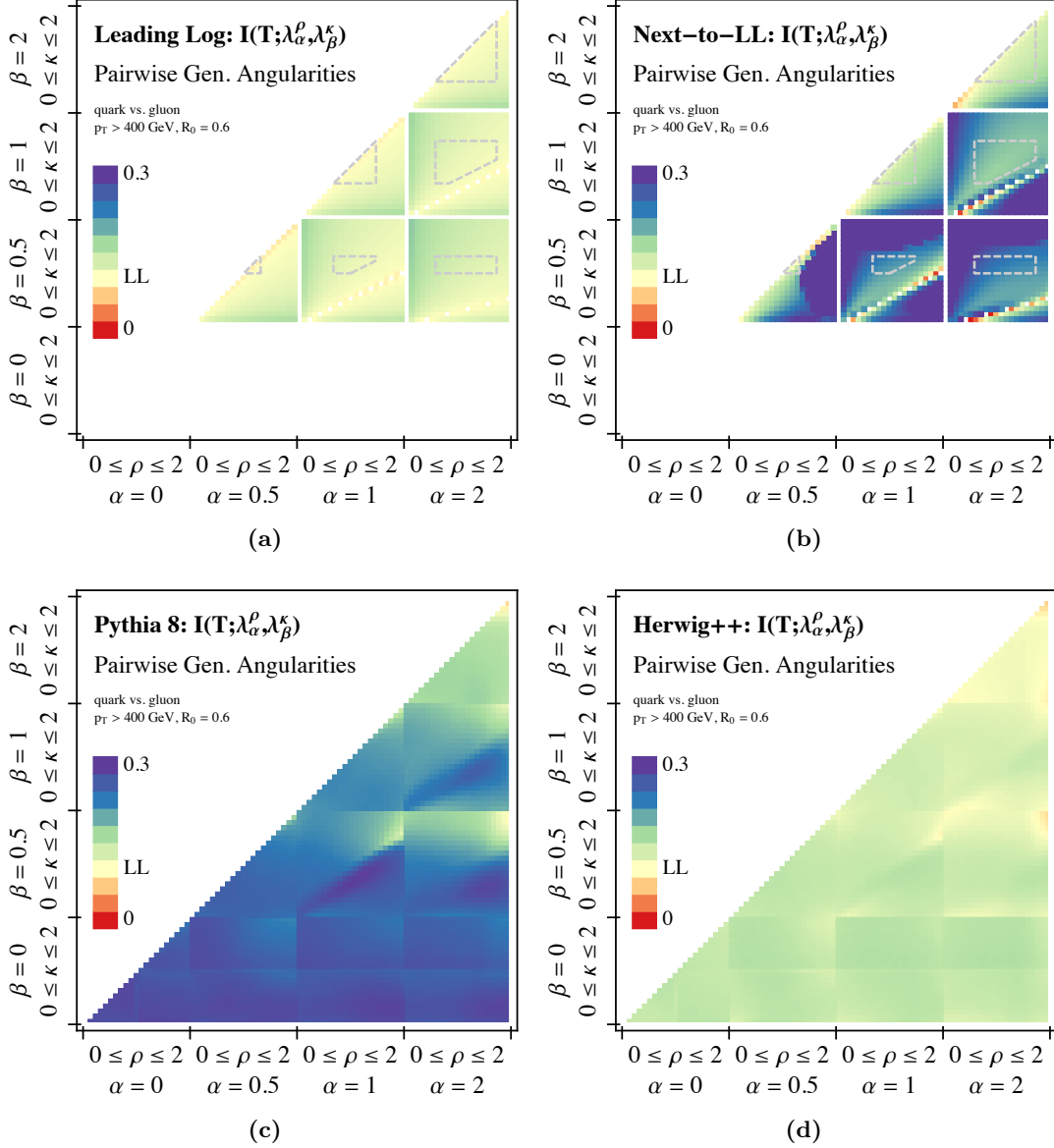
Figure 13. The quark/gluon truth overlap for pairs of generalized angularities $(\lambda_\alpha^\rho, \lambda_\beta^\kappa)$. Top: the LL and NLL analytic calculations. Bottom: the PYTHIA 8 and HERWIG++ parton showers. Here, we show four values of $\beta \in \{0, 0.5, 1, 2\}$ and for each value of β, $0 \leq \kappa \leq 2$ in steps of size 0.1. In the NLL and NLL plots, the interior of the dashed grey boxes correspond to the range of validity of our calculations.

8 and NLL results most clearly indicate that it is advantageous to pick one of the angular exponents α or β to be small. A notable difference between the predictions is that the LL and NLL calculations suggest that one should avoid the diagonal $\rho = \kappa$, whereas for PYTHIA 8, and to a lesser extent HERWIG++, the maximum discrimination power is sometimes (surprisingly) close to it (see e.g. $\alpha = 1$, $\beta = 0.5$).

7   Conclusions

Robust quark/gluon discrimination is a key goal for the jet substructure community, so to the extent possible, it is important to use first principles calculations to assess the challenges and opportunities. In this paper, we showed that mutual information is a powerful way to understand how variables are correlated, and whether or not that (lack of) correlation pertains to discrimination power. We also made progress in gaining analytic control over the tagging performance of the generalized angularities λ_β^κ. For the IRC safe angularities and the IRC unsafe angularities with $\beta > 0$, we calculated the quark/gluon truth overlap for a single angularity $I(T; \lambda_\beta^\kappa)$ and for pairs of angularities $I(T; \lambda_\alpha^\rho, \lambda_\beta^\kappa)$ to NLL order.

Ultimately, we want to extend our analysis to higher orders, but this would require a robust "truth" definition for a quark jet versus a gluon jet. While the strategy of ref. [41] is one option to define the truth flavor of a jet, we would prefer a definition for which the jet constituents are the same as for traditional flavor-less jet algorithms. Of course, quark and gluon jets do not exist in isolation, and at some point, the color correlations between the jets will be relevant for characterizing the discrimination power. The techniques introduced recently in ref. [42] should help in gaining analytic control over those color correlations. Since our NLL results are subject to large changes from scale variation, higher-order calculations will be crucial for robust uncertainty estimates.

Assuming we did have a suitable quark/gluon truth definition, then a key challenge for calculations beyond our present order is dealing with soft radiation, in particular the effect of non-global logarithms [70]. One option is to do quark/gluon tagging in concert with soft drop declustering [36] (a generalization of modified mass drop tagging [34, 72]). The soft drop procedure removes soft radiation, and therefore removes non-global contributions to the jet. The cumulative distributions for a single soft-dropped angularity were already calculated in ref. [36], where the distributions exhibited Casimir scaling at LL order. Using soft-dropped jet shapes for quark/gluon discrimination seems promising from both a theoretical and experimental point of view, and we leave a more detailed study to future work.

Based on our studies, we have two recommendations to the ATLAS and CMS experiments. The first recommendation is to make (unfolded) measurements of the recoil-free angularities distributions for a range of β values, ideally in purified quark/gluon samples.[19] The differences seen between Monte Carlo programs in figure 4 is worrisome, and while eventually calculations might be a guide to what these distributions should look like, in the short term e_β measurements can be a key reference for tuning Monte Carlo programs,

---

[19]Initial results along these lines appear in ref. [12] for the energy correlation function ratio $C_1^{(\beta)}$ [9], though only the final quark/gluon performance is shown, not the extracted quark and gluon distributions.

especially because the differences arise from effects that are formally beyond LL accuracy. The second recommendation is to measure more double differential distributions.[20] While we focused on mutual information with the truth in this paper, one would still like to understand the full correlation structure. Angularities are a good place to start, and double differential distributions of e_α and e_β would be quite valuable, especially with new calculational tools available to predict these correlations [53–55] (see appendix D for example plots along these lines).

Finally, in the spirit of refs. [6, 38–40], we have provided another example of how collinear unsafe (but soft safe) observables can be made calculable with the help of new nonperturbative objects. We introduced the weighted-energy functions $F_\kappa^i(x)$, which allowed us to understand many aspects of the $\kappa \neq 1$ regime. Because the $\beta > 0$ angularities only depend on logarithmic moments of $F_\kappa^i(x)$, they are the simplest to understand. But even the $\beta = 0$ angularities are within calculational control, since we can study the renormalization group behavior of, say, the p_T^D (λ_0^2) distribution. Of course, hadron multiplicity (λ_0^0) is not captured within our framework due to the presence of the soft singularity, but perhaps hadron multiplicities could be made analytically tractable by using soft drop declustering to remove soft radiation. We expect future studies will continue to improve (and improve our understanding of) quark/gluon discrimination.

Acknowledgments

A Properties of mutual information

A.1 Relationship to the ROC curve

The mutual information of an observable A with the truth T can be derived from the ROC curve of A. Although we phrase this discussion in terms of quark/gluon discrimination, it obviously carries over to other cases as well.

At the position (q, g) on the ROC curve, the region dq has a fraction dq of the quark jets. The fraction of gluon jets is given by the slope of the ROC curve, $dg = g'(q)\, dq$. Looking at the definition of $I(T; A)$ in eq. (2.11), we can write the integrals as

$$\int da\, p_q(a) \Rightarrow \int dq, \quad \int da\, p_g(a) \Rightarrow \int dg = \int dq\, g'(q). \tag{A.1}$$

---

[20]There are double differential results in ref. [13], but measured only in simulation and not in data.

For a sample with quark fraction f, the ratios of the probability distributions are

$$\frac{p_q(a)}{p_{\text{tot}}(a)} \Rightarrow \frac{dq}{dqf + dg(1-f)}, \qquad \frac{p_g(a)}{p_{\text{tot}}(a)} \Rightarrow \frac{dg}{dqf + dg(1-f)}. \tag{A.2}$$

This leads to

$$I(T;A) = \int dq \left(f \log_2 \frac{1}{f + (1-f)g'(q)} + (1-f)\, g'(q) \log_2 \frac{g'(q)}{f + (1-f)g'(q)} \right). \tag{A.3}$$

As a simple test of this formula, note that the ROC curve for an observable that satisfies Casimir scaling is $g(q) = q^{(C_A/C_F)}$. Plugging this into eq. (A.3), we recover $I(T;A)$ from eq. (3.8).

Inverting this relationship to obtain the ROC curve from the mutual information is not so easy, suggesting that mutual information is an easier concept to work with. Nevertheless it seems in principle possible, though we do not claim that the following simple-minded approach is optimal. Consider discretizing $g'(q)$ by treating it as a constant g'_i on the interval $q \in [i/n_{\text{bins}}, (i+1)/n_{\text{bins}}]$ with $i = 0, 1, \ldots, n_{\text{bins}} - 1$. Because an ideal ROC curve is not only monotonically increasing (i.e. $g'(q) \geq 0$) but also convex (i.e. $g''(q) \geq 0$), this means that $g'_{i+1} \geq g'_i$, and we have a chance to find a set of equations that (uniquely) determine g'_i. Then, we can integrate $g'(q)$ in the usual way to find $g(q)$. One set of equations is given by the n-th derivative of $I(T;A)$ evaluated at $f = 1$ (for $n \geq 2$)

$$\left. \frac{d^n I(T;A)}{df^n} \right|_{f=1} = -\frac{(n-2)!}{\ln 2} \int_0^1 dq\, (g'(q) - 1)^n. \tag{A.4}$$

while for the special case of $n = 1$ we can use

$$\int_0^1 dq\, (g'(q) - 1) = 0. \tag{A.5}$$

In discrete form, these become a system of polynomial equations

$$\frac{1}{n_{\text{bins}}} \sum_i (g'_i - 1) = 0,$$

$$\frac{1}{n_{\text{bins}}} \sum_i (g'_i - 1)^n = -\frac{\ln 2}{(n-2)!} \left. \frac{d^n I(T;A)}{df^n} \right|_{f=1}. \tag{A.6}$$

Because these equation are non-linear, this quickly becomes numerically unstable for $n_{\text{bins}} > 4$, but gives a proof of principle that a solution can be found. Note that the condition that the ROC curve is convex is crucial, since otherwise each permutation of g'_i would also constitute a solution. For special functional forms, there are simpler strategies to find the ROC curve. For example, if it is known that $g(q) = q^c$, then a practical way to estimate the exponent c is via

$$\left. \frac{dI(T;A)}{df} \right|_{f=0} = -\frac{1}{\ln 2} \int_0^1 dq \ln g'(q) = \frac{c - 1 - \ln c}{\ln 2}. \tag{A.7}$$

A.2 Better ROC curve implies greater truth overlap

When one observable b has a larger truth overlap than another observable w, i.e.

$$I(T; B) > I(T; W), \tag{A.8}$$

we interpreted this to mean that b is a better discriminant variable and w is worse. While this is generically true, there can be cases where the "worse" observable can have better performance at a given operating point. This occurs when the ROC curves of b and w intersect, for example when b has better background rejection at low signal efficiency but w performs better at high signal efficiency. Thus, we cannot conclude from the truth overlap alone whether b or w is better, since even from a ROC curve perspective, "better" is ill-defined.

What we can prove is that if the ROC curve for b is everywhere better than for w, then the corresponding truth overlap is strictly larger. In information theory language, we would say that b is Pareto optimal with respect to w. Consider the ROC curves $g_b(q)$ and $g_w(q)$. A better observable's ROC curve will take a smaller value (less background) at each value of signal efficiency. Thus we can define the better observable as

$$g_b(q) \equiv g_w(q) - \Delta(q), \tag{A.9}$$

where $\Delta(q) \geq 0$ for all $q \in [0, 1]$. We will now use a variational method to show that $I(T; B) > I(T; W)$.

From these two ROC curves, we can calculate the difference between their truth overlaps. Using the relationship between mutual information and the ROC curve from eq. (A.3) we have:

$$
\begin{aligned}
I(T; B) - I(T; W) &= \int dq \left(f \log_2 \frac{1}{f + (1-f)g_b'(q)} + (1-f)\, g_b'(q) \log_2 \frac{g_b'(q)}{f + (1-f)g_b'(q)} \right) \\
&\quad - \int dq \left(f \log_2 \frac{1}{f + (1-f)g_w'(q)} + (1-f)\, g_w'(q) \log_2 \frac{g_w'(q)}{f + (1-f)g_w'(q)} \right) \\
&= - \int dq\, \Delta'(q) \log_2 \frac{(1-f)g_w'(q)}{f + (1-f)g_w'(q)} + \mathcal{O}(\Delta^2).
\end{aligned}
\tag{A.10}
$$

In the last step, we expanded the integrand assuming that $\Delta(q)$ is small. Because any ROC curve g must satisfy $g(0) = 0$ and $g(1) = 1$, we can perform integration by parts on $\Delta'(q)$ without introducing any boundary terms.

$$I(T; B) - I(T; W) = \frac{1}{\ln 2} \int dq\, \Delta(q) \frac{g_w''(q)}{g_w'(q)} \frac{f(1-f)}{f + (1-f)g_w'(q)} + \mathcal{O}(\Delta^2). \tag{A.11}$$

This term is manifestly positive because for any (ideal) ROC curve $g'(q) \geq 0$ and $g''(q) \geq 0$ for all $q \in [0, 1]$, and $\Delta(q)$ is positive by assumption. Therefore,

$$I(T; B) - I(T; W) > 0 \tag{A.12}$$

up to corrections of order Δ^2. However, we can make these Δ^2 corrections arbitrarily small by applying this variational logic to a sequence of ROC curves that smoothly interpolate

between $g_b(q)$ and $g_w(q)$.[21] Thus, we have proven that an observable with an everywhere-better ROC curve also has a larger truth overlap.

A.3 Subtleties of finite statistics and binning

When dealing with a sample with a finite number of events, the Shannon entropy H is sensitive to the way in which the sample is binned. We will compute the leading effect this has on the entropy. Let N_{bins} be the number of bins and let λ_i be the average number of events in bin i. The total number of events in the sample is

$$\sum_{i=1}^{N_{\text{bins}}} \lambda_i = N_{\text{ev}}. \tag{A.13}$$

For sufficiently large N_{ev}, the fluctuations within the bins should be independent and the number of events in bin i should be Poisson distributed about λ_i. Thus, the expected entropy of the binned sample is

$$H = \sum_{i=1}^{N_{\text{bins}}} \sum_{n_i=0}^{\infty} \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \left(-\frac{n_i}{N_{\text{ev}}} \log_2 \frac{n_i}{N_{\text{ev}}} \right), \tag{A.14}$$

where n_i ranges over the possible observed events in bin i.

Assuming that $\lambda_i \gg 1$ for all bins, we can expand the entropy about the average value $\langle n_i \rangle = \lambda_i$ in each bin. The lowest order term is the expected value of the entropy in the limit of infinite statistics

$$H_\infty = - \sum_{i=1}^{N_{\text{bins}}} \frac{\lambda_i}{N_{\text{ev}}} \log_2 \frac{\lambda_i}{N_{\text{ev}}}. \tag{A.15}$$

The first order term vanishes because λ_i is the average value. The second order term is the first non-trivial effect from finite sample size. The second derivative of the entropy factor is

$$\frac{\mathrm{d}^2}{\mathrm{d}n_i^2} \left(-\frac{n_i}{N_{\text{ev}}} \log_2 \frac{n_i}{N_{\text{ev}}} \right) \bigg|_{n_i=\lambda_i} = -\frac{1}{\lambda_i N_{\text{ev}} \ln 2}. \tag{A.16}$$

Then, the entropy is

$$\begin{aligned} H &= H_\infty + \frac{1}{\ln 2} \sum_{i=1}^{N_{\text{bins}}} \sum_{n_i=0}^{\infty} \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \frac{(n_i - \lambda_i)^2}{2} \left(-\frac{1}{\lambda_i N_{\text{ev}}} \right) + \dots \\ &= H_\infty - \frac{1}{2 \ln 2} \frac{N_{\text{bins}}}{N_{\text{ev}}} + \dots, \end{aligned} \tag{A.17}$$

where we have used the fact that a Poisson distribution with average λ_i has variance λ_i. Higher terms in the expansion will depend on the distribution of λ_i, but these contributions are suppressed by inverse powers of λ_i.

---

[21]Concretely, one can always build an interpolating strategy based on randomly selecting b or w as the discriminant variable with predetermined probability.

In calculating the mutual information $I(A; B)$ on a single sample, the leading effect from finite statistics is

$$
\begin{aligned}
I(A; B) &= H(A) + H(B) - H(A, B) \\
&= H_\infty(A) + H_\infty(B) - H_\infty(A, B) - \frac{1}{2 \ln 2} \left(\frac{N_{\text{bins}}^A}{N_{\text{ev}}^A} + \frac{N_{\text{bins}}^B}{N_{\text{ev}}^B} - \frac{N_{\text{bins}}^{AB}}{N_{\text{ev}}^{AB}} \right) + \dots,
\end{aligned}
\tag{A.18}
$$

where N_{bins}^X is the number of bins used to calculate $H(X)$ and N_{ev}^X is the number of events in the corresponding sample. To avoid the leading bias term from binning when the sample sizes are the same (i.e. $N_{\text{ev}}^A = N_{\text{ev}}^B = N_{\text{ev}}^{AB}$), it is important to take N_{bins}^{AB} (approximately) equal to $N_{\text{bins}}^{AB} = N_{\text{bins}}^A + N_{\text{bins}}^B$, rather than the naive choice $N_{\text{bins}}^{AB} = N_{\text{bins}}^A N_{\text{bins}}^B$. This is the strategy we used to make figures 14 and 15. Alternatively, if one wants to use the same binning in the A and B observables such that $N_{\text{bins}}^{AB} = N_{\text{bins}}^A N_{\text{bins}}^B$, then one has to adjust the number of events accordingly. This is the strategy we used to make figures 17 and 18, where we took $N_{\text{bins}}^A = N_{\text{bins}}^B \equiv N_{\text{bins}}$, $N_{\text{ev}}^A = N_{\text{ev}}^B \equiv N_{\text{ev}}$, but $N_{\text{ev}}^{AB} = N_{\text{ev}} N_{\text{bins}}/2$.

For the calculation of the truth overlap $H(T; A)$ of an observable A, we have to deal with separate quark and gluon event samples. The truth overlap is defined in terms of the Shannon entropies as

$$
I(T; A) = H^{q+g}(A) - f H^q(A) - (1 - f) H^g(A),
\tag{A.19}
$$

where $H^{q+g}(A)$ is the entropy of the combined quark and gluon sample, $H^q(A)$ ($H^g(A)$) is the entropy of the quark (gluon) sample, and f is the fraction of quarks in the combined sample. For finite bins and sample size, the expected entropy of the combined sample is

$$
H^{q+g}(A) = -\sum_{i=1}^{N_{\text{bins}}^{q+g}} \sum_{n_i^q=0}^{\infty} \sum_{n_i^g=0}^{\infty} \frac{(\lambda_i^q)^{n_i^q} e^{-n_i^q}}{n_i^q!} \frac{(\lambda_i^g)^{n_i^g} e^{-n_i^g}}{n_i^g!} \left(\frac{n_i^q + n_i^g}{N_q + N_g} \right) \log_2 \left(\frac{n_i^q + n_i^g}{N_q + N_g} \right).
\tag{A.20}
$$

Here, λ_i^q (λ_i^g) is the average number of quark (gluon) jets in bin i, n_i^q (n_i^g) is the corresponding number of observed jets, N_{bins}^{q+g} is the total number of bins, and N_q (N_g) is the total number of quark (gluon) jets in the combined sample. Note that

$$
f = \frac{N_q}{N_q + N_g}.
\tag{A.21}
$$

Using the same technique as before, we can find the leading effect from finite statistics on the entropy by expanding n_i^q and n_i^g about λ_i^q and λ_i^g to quadratic order. We find

$$
H^{q+g}(A) = H_\infty^{q+g}(A) - \frac{1}{2 \ln 2} \frac{N_{\text{bins}}^{q+g}}{N_q + N_g} + \dots.
\tag{A.22}
$$

The resulting truth overlap $I(T; A)$ is

$$
\begin{aligned}
I(T; A) &= H_\infty^{q+g}(A) - f H_\infty^q(A) - (1 - f) H_\infty^g(A) \\
&\quad - \frac{1}{2 \ln 2} \frac{N_{\text{bins}}^{q+g}}{N_q + N_g} + \frac{f}{2 \ln 2} \frac{N_{\text{bins}}^q}{N_q'} + \frac{1 - f}{2 \ln 2} \frac{N_{\text{bins}}^g}{N_g'} + \dots,
\end{aligned}
\tag{A.23}
$$

where N_{bins}^q (N_{bins}^g) is the number of bins in the pure quark (gluon) sample and N_q' (N_g') is the number of jets in the pure quark (gluon) sample.

This finite statistics bias can be reduced by using samples and binning such that the second line of eq. (A.23) is zero. There are two useful cases to consider. When the combined sample is created by simply merging the pure samples (i.e. $N_q' = N_q$ and $N_g' = N_g$), the bias terms reduce to

$$\delta I(T; A) = \frac{1}{2\ln 2} \frac{N_{\text{bins}}^q + N_{\text{bins}}^g - N_{\text{bins}}^{q+g}}{N_q + N_g}, \tag{A.24}$$

and so the bias can be removed by binning such that $N_{\text{bins}}^q + N_{\text{bins}}^g = N_{\text{bins}}^{q+g}$. Alternatively, if we choose to use the same binning for each sample (i.e. $N_{\text{bins}}^q = N_{\text{bins}}^g = N_{\text{bins}}^{q+g} \equiv N_{\text{bins}}$), the bias terms are

$$\delta I(T; A) = \frac{1}{2\ln 2} \frac{N_{\text{bins}}}{N_q + N_g} \left(\frac{N_q}{N_q'} + \frac{N_g}{N_g'} - 1 \right). \tag{A.25}$$

This is zero, if, for example, we take the combined sample to have half the number of events of the pure samples (i.e. $N_q/N_q' = N_g/N_g' = 1/2$). This later strategy is the one we used for all of the truth overlap plots in this paper.

B   Calculational details

B.1   One IRC safe angularity at NLL

The cross section for a single recoil-free angularity e_β was derived in refs. [22, 26] (see also ref. [30]), and we summarize the results here. To NLL order, the cumulative distribution for e_β can be expressed as

$$\Sigma_i(e_\beta) = \frac{e^{-\gamma_E R_i'(e_\beta)}}{\Gamma(1 + R_i'(e_\beta))} e^{-R_i(e_\beta) - \gamma_i T_i(e_\beta)}. \tag{B.1}$$

Here, we are using a slightly different notation from ref. [26] and the body of the text (see eq. (6.11)), with

$$R_i^{\text{text}}(e_\beta) = R_i(e_\beta) + \gamma_i T_i(e_\beta). \tag{B.2}$$

The reason for separating out the T function is that it contains terms that formally start at NLL order, such that we do not need to consider the logarithmic derivative $T'(e_\beta)$ in the prefactor.

In this notation, the radiator $R(e_\beta)$ consists of the cusp pieces of the jet and soft function anomalous dimensions, while the function $T(e_\beta)$ contains the non-cusp terms (for details on these anomalous dimensions see appendix C.1). The logarithmic derivative $R'(e_\beta)$ is given by

$$R'(e_\beta) \equiv -\frac{\mathrm{d}}{\mathrm{d}\ln e_\beta} R(e_\beta). \tag{B.3}$$

To NLL accuracy, the cusp anomalous dimensions are evaluated at two-loop order and the radiator is

$$R(e_\beta) = \frac{C_i}{2\pi\alpha_s\beta_0^2} \frac{1}{\beta-1} \left[(1+\lambda)\ln(1+\lambda) - (\beta+\lambda)\ln\left(1+\frac{\lambda}{\beta}\right) \right]$$
$$+ \frac{C_i}{4\pi^2\beta_0^2} \frac{1}{\beta-1} \left[\left(\frac{\Gamma_{\text{cusp}}^1}{\Gamma_{\text{cusp}}^0} - 2\pi\frac{\beta_1}{\beta_0} \right) \left(\beta\ln\left(1+\frac{\lambda}{\beta}\right) - \ln(1+\lambda) \right) \right.$$
$$\left. + \pi\frac{\beta_1}{\beta_0} \left(\ln^2(1+\lambda) - \beta\ln^2\left(1+\frac{\lambda}{\beta}\right) \right) \right]. \tag{B.4}$$

Here, the strong coupling constant is evaluated at the hard scale

$$\alpha_s \equiv \alpha_s(p_T R), \tag{B.5}$$

using two-loop running with $n_f = 5$ from $\alpha_s(m_Z) = 0.12$. The observable is contained in $\lambda = 2\alpha_s\beta_0\ln e_\beta$, the color factor of the jet is C_i, the one- and two-loop beta functions are

$$\beta_0 = \frac{11}{12\pi}C_A - \frac{1}{6\pi}n_f, \qquad \beta_1 = \frac{17}{24\pi^2}C_A^2 - \frac{5}{24\pi^2}C_A n_f - \frac{1}{8\pi^2}C_F n_f, \tag{B.6}$$

and the ratio of the two-loop to the one-loop cusp anomalous dimensions is

$$\frac{\Gamma_{\text{cusp}}^1}{\Gamma_{\text{cusp}}^0} = \left(\frac{67}{18} - \frac{\pi^2}{6} \right) C_A - \frac{5}{9}n_f. \tag{B.7}$$

For the non-cusp terms, the function $T(e_\beta)$ is

$$T(e_\beta) = \frac{1}{\pi\beta_0}\ln\left(1+\frac{\lambda}{\beta}\right), \tag{B.8}$$

and γ_i is the non-cusp anomalous dimension to one-loop. For quarks and gluons, they are

$$\gamma_q = \frac{3}{4}C_F, \qquad \gamma_g = \frac{11}{12}C_A - \frac{1}{6}n_f. \tag{B.9}$$

For NLL accuracy, the logarithmic derivative $R'(e_\beta)$ only needs to be evaluated at one-loop:

$$R'(e_\beta)_{\text{NLL}} = \frac{C_i}{\pi\beta_0} \frac{1}{\beta-1} \left(\ln\left(1+\frac{\lambda}{\beta}\right) - \ln(1+\lambda) \right). \tag{B.10}$$

The cross section is obtained from the cumulative distribution in the standard way,

$$\frac{1}{\sigma}\frac{\mathrm{d}\sigma}{\mathrm{d}e_\beta} = \frac{\mathrm{d}}{\mathrm{d}e_\beta}\Sigma(e_\beta). \tag{B.11}$$

We will verify this single differential calculation in the language of SCET in appendix C.1.

B.2 Two IRC safe angularities at NLL

For two angularities e_α and e_β, the form of the double cumulative cross section to NLL order was conjectured in ref. [54]. Assuming $\alpha > \beta$, to NLL order it can be written as

$$\Sigma(e_\alpha, e_\beta) = \frac{e^{-\gamma_E \widetilde{R}(e_\alpha, e_\beta)}}{\Gamma(1 + \widetilde{R}(e_\alpha, e_\beta))} \, e^{-R(e_\alpha, e_\beta) - \gamma_i T(e_\alpha, e_\beta)} \,, \tag{B.12}$$

where the functions $R(e_\alpha, e_\beta)$, $T(e_\alpha, e_\beta)$, and $\widetilde{R}(e_\alpha, e_\beta)$ are given below. As in eq. (B.2), we are changing the notation from the text (see eq. (6.29)) to separate out the T piece from the radiator. Note that this has a similar form to eq. (B.1), albeit with functions that depend on two arguments. Unlike the single angularity case, $\widetilde{R}$ is not related to any logarithmic derivative of R.

The radiator $R(e_\alpha, e_\beta)$ is

$$
\begin{aligned}
R(e_\alpha, e_\beta) = {} & \frac{C_i}{2\pi\alpha_s\beta_0^2} \left[\frac{1}{\alpha - 1} \, U(2\alpha_s\beta_0 \ln e_\alpha) - \frac{\beta}{\beta - 1} \, U\left(2\alpha_s\beta_0 \frac{\ln e_\beta}{\beta}\right) \right. \\
& \left. + \frac{\alpha - \beta}{(\alpha - 1)(\beta - 1)} \, U\left(2\alpha_s\beta_0 \frac{\ln e_\alpha^{1-\beta} e_\beta^{\alpha-1}}{\alpha - \beta}\right) \right] \\
& + \frac{C_i}{4\pi^2\beta_0^2} \left[\left(\frac{\Gamma_{\text{cusp}}^1}{\Gamma_{\text{cusp}}^0} - 2\pi\frac{\beta_1}{\beta_0} \right) \left(\frac{\beta}{\beta - 1} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln e_\beta}{\beta}\right) \right. \right. \\
& - \frac{1}{\alpha - 1} \ln(1 + 2\alpha_s\beta_0 \ln e_\alpha) - \frac{\alpha - \beta}{(\alpha - 1)(\beta - 1)} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln e_\alpha^{1-\beta} e_\beta^{\alpha-1}}{\alpha - \beta}\right) \Bigg) \\
& + \pi\frac{\beta_1}{\beta_0} \left(\frac{1}{\alpha - 1} \ln^2(1 + 2\alpha_s\beta_0 \ln e_\alpha) - \frac{\beta}{\beta - 1} \ln^2\left(1 + 2\alpha_s\beta_0 \frac{\ln e_\beta}{\beta}\right) \right. \\
& \left. \left. + \frac{\alpha - \beta}{(\alpha - 1)(\beta - 1)} \ln^2\left(1 + 2\alpha_s\beta_0 \frac{\ln e_\alpha^{1-\beta} e_\beta^{\alpha-1}}{\alpha - \beta}\right) \right) \right],
\end{aligned}
\tag{B.13}
$$

where $U(z) = (1 + z)\ln(1 + z)$. The non-cusp piece $T(e_\alpha, e_\beta)$ is

$$T(e_\alpha, e_\beta) = \frac{1}{\pi\beta_0} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln e_\beta}{\beta}\right) - 2\frac{\alpha_s}{\pi} \frac{\alpha - \beta}{\alpha} \frac{e_\alpha^{-\frac{\beta}{\alpha-\beta}} e_\beta^{\frac{\alpha}{\alpha-\beta}}}{\beta + 2\alpha_s\beta_0 \ln e_\beta} \,. \tag{B.14}$$

The multiple emissions piece $\widetilde{R}(e_\alpha, e_\beta)$ is

$$
\begin{aligned}
\widetilde{R}(e_\alpha, e_\beta) = {} & \frac{C_i}{\pi\beta_0} \left[\frac{1}{\beta - 1} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln e_\beta}{\beta}\right) - \frac{1}{\alpha - 1} \ln(1 + 2\alpha_s\beta_0 \ln e_\alpha) \right. \tag{B.15} \\
& \left. - \frac{\alpha - \beta}{(\alpha - 1)(\beta - 1)} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln e_\alpha^{1-\beta} e_\beta^{\alpha-1}}{\alpha - \beta}\right) + 2\alpha_s\beta_0 \frac{\alpha - \beta}{\alpha} \frac{e_\alpha^{-\frac{\beta}{\alpha-\beta}} e_\beta^{\frac{\alpha}{\alpha-\beta}}}{\beta + 2\alpha_s\beta_0 \ln e_\beta} \right].
\end{aligned}
$$

The power suppressed terms have been chosen such that the sum of the exponents of e_α and e_β is 1. The cross section is obtained from the cumulative distribution by differentiation and imposing the phase space constraint in eq. (5.4),

$$\frac{1}{\sigma_i} \frac{\mathrm{d}^2\sigma_i}{\mathrm{d}e_\alpha \, \mathrm{d}e_\beta} = \left(\frac{\partial^2}{\partial e_\alpha \, \partial e_\beta} \Sigma(e_\alpha, e_\beta) \right) \Theta(e_\beta - e_\alpha) \, \Theta\big((e_\alpha)^\beta - (e_\beta)^\alpha\big). \tag{B.16}$$

We will verify this double differential calculation in the language of SCET in appendix C.2.

B.3 Two IRC unsafe angularities at NLL

For the IRC unsafe angularities, the double differential distribution takes the form

$$\Sigma_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa) = \frac{e^{-\gamma_E \widetilde{R}_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa)}}{\Gamma(1 + \widetilde{R}_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa))} \, e^{-R_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa) - \gamma_i T_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa)}, \tag{B.17}$$

where we are again using the notation change in eq. (B.2). Using the rescaling trick in eq. (6.32), we can determine the double radiator $R_i + \gamma_i T_i$. For the multiple emissions term $\widetilde{R}_i$, we interpolate between expressions derived at the phase space boundaries [54]. This interpolation yields

$$\widetilde{R}_i(\lambda_\alpha^\rho, \lambda_\beta^\kappa) = \frac{C_i}{\pi\beta_0}\left[\frac{1}{\beta - \kappa} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln\lambda_\beta^\kappa}{\beta}\right) - \frac{1}{\alpha - \rho} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln\lambda_\alpha^\rho}{\rho}\right) \right. \tag{B.18}$$

$$- \frac{\alpha - \beta + \kappa - \rho}{(\alpha - \rho)(\beta - \kappa)} \ln\left(1 + 2\alpha_s\beta_0 \frac{\ln(\lambda_\alpha^\rho)^{\kappa-\beta}(\lambda_\beta^\kappa)^{\alpha-\rho}}{\alpha\kappa - \beta\rho}\right)$$

$$\left. + 2\alpha_s\beta_0 \frac{(\alpha - \beta)^2}{\alpha(\alpha\kappa - \beta\rho)} \frac{(\lambda_\alpha^\rho)^{-\frac{\beta}{\alpha-\beta}}(\lambda_\beta^\kappa)^{\frac{\alpha}{\alpha-\beta}}}{\beta + 2\alpha_s\beta_0 \ln\lambda_\beta^\kappa} - 2\alpha_s\beta_0 \frac{(\rho - \kappa)^2}{\kappa(\alpha\kappa - \beta\rho)} \frac{(\lambda_\alpha^\rho)^{\frac{\kappa}{\kappa-\rho}}(\lambda_\beta^\kappa)^{-\frac{\rho}{\kappa-\rho}}}{\rho + 2\alpha_s\beta_0 \ln\lambda_\alpha^\rho}\right].$$

The final two terms, proportional to powers of λ_α^ρ and λ_β^κ, are formally power-suppressed over the entire phase space, but are necessary to satisfy the boundary conditions. When $\rho = \kappa = 1$, this reduces to the IRC safe case in eq. (B.15).

B.4 Finding the ROC curve at LL

In this paper, we focused on mutual information, but one can (in principle) use the same cross sections to determine the ROC curve for quark/gluon discrimination.

At LL accuracy, the quark/gluon probability distributions for the IRC safe angularities can be obtained from differentiating the cumulative distribution from eq. (5.3) as

$$p_i(e_\alpha, e_\beta) = \frac{1}{\sigma_i} \frac{\mathrm{d}^2\sigma_i}{\mathrm{d}e_\alpha\,\mathrm{d}e_\beta} = \frac{\partial^2}{\partial e_\alpha\,\partial e_\beta} \Sigma_i(e_\alpha, e_\beta)$$

$$= \left(\frac{2\alpha_s}{\pi} \frac{C_i}{\alpha - \beta} \frac{1}{e_\alpha e_\beta} + \frac{4\alpha_s^2}{\pi^2} \frac{C_i^2}{\beta(\alpha - \beta)^2} \frac{1}{e_\alpha e_\beta} \log\frac{e_\beta}{e_\alpha} \log\frac{(e_\alpha)^\beta}{(e_\beta)^\alpha}\right) \Sigma_i(e_\alpha, e_\beta). \tag{B.19}$$

To determine the ROC curve, one needs to find contours of constant discrimination power, which is the same as finding contours of constant quark over gluon probabilities. A displacement $(\mathrm{d}e_\alpha, \mathrm{d}e_\beta)$ along such a contour satisfies

$$\frac{p_q(e_\alpha + \mathrm{d}e_\alpha, e_\beta + \mathrm{d}e_\beta)}{p_g(e_\alpha + \mathrm{d}e_\alpha, e_\beta + \mathrm{d}e_\beta)} = \frac{p_q(e_\alpha, e_\beta)}{p_g(e_\alpha, e_\beta)}, \tag{B.20}$$

which can be rewritten in terms of a gradient $\vec{\nabla} = (\partial/\partial e_\alpha, \partial/\partial e_\beta)$ as

$$0 = (\mathrm{d}e_\alpha, \mathrm{d}e_\beta) \cdot \vec{\nabla} \ln \frac{p_q(e_\alpha, e_\beta)}{p_g(e_\alpha, e_\beta)} \tag{B.21}$$

$$= (\mathrm{d}e_\alpha, \mathrm{d}e_\beta) \cdot \vec{\nabla} \left[\ln \frac{1 + \frac{2\alpha_s}{\pi}\frac{C_q}{\beta(\alpha-\beta)}\ln\frac{e_\beta}{e_\alpha}\ln\frac{(e_\alpha)^\beta}{(e_\beta)^\alpha}}{1 + \frac{2\alpha_s}{\pi}\frac{C_g}{\beta(\alpha-\beta)}\ln\frac{e_\beta}{e_\alpha}\ln\frac{(e_\alpha)^\beta}{(e_\beta)^\alpha}} - \frac{\alpha_s}{\pi}(C_q - C_g)\left(\frac{1}{\beta}\ln^2 e_\beta + \frac{1}{\alpha - \beta}\ln^2 \frac{e_\alpha}{e_\beta}\right)\right],$$

resulting in the following differential equation for contours of constant discrimination power

$$
\begin{aligned}
\frac{\mathrm{d}e_\beta}{\mathrm{d}e_\alpha} = \frac{\beta e_\beta}{e_\alpha} \Bigg\{ & 3(\alpha-\beta)^2\beta^2 \ln e_\alpha - \beta(\alpha-\beta)^2(\alpha+2\beta)\ln e_\beta + \frac{2\alpha_s(C_q+C_g)}{\pi}\,\beta\Big[\beta(\beta-\alpha)\ln^3 e_\alpha \\
& + (\alpha-\beta)(\alpha+2\beta)\ln^2 e_\alpha \ln e_\beta + (-2\alpha^2+\alpha\beta+\beta^2)\ln e_\alpha \ln^2 e_\beta - \alpha(\beta-\alpha)\ln^3 e_\beta\Big] \\
& + \frac{4\alpha_s^2 C_q C_g}{\pi^2}\Big[\beta^2 \ln^5 e_\alpha - \beta(2\alpha+3\beta)\ln^4 e_\alpha \ln e_\beta + (\alpha^2+6\alpha\beta+3\beta^2)\ln^3 e_\alpha \ln^2 e_\beta \\
& - (3\alpha^2+6\alpha\beta+\beta^2)\ln^2 e_\alpha \ln^3 e_\beta + \alpha(3\alpha+2\beta)\ln e_\alpha \ln^4 e_\beta - \alpha^2 \ln^5 e_\beta\Big]\Bigg\}\Bigg/ \\
& \Bigg\{\beta^2(\alpha-\beta)^2(\alpha+2\beta)\ln e_\alpha - 3\alpha(\alpha-\beta)^2\beta^2 \ln e_\beta + \frac{2\alpha_s(C_q+C_g)}{\pi}\,\beta(\alpha-\beta) \\
& \times \Big[-\beta^2 \ln^3 e_\alpha + \beta(2\alpha+\beta)\ln^2 e_\alpha \ln e_\beta - \alpha(\alpha+2\beta)\ln e_\alpha \ln^2 e_\beta + \alpha^2 \ln^3 e_\beta\Big] \\
& + \frac{4\alpha_s^2 C_q C_g}{\pi^2}\Big[\beta^3 \ln^5 e_\alpha - \beta^2(3\alpha+2\beta)\ln^4 e_\alpha \ln e_\beta + \beta(3\alpha^2+6\alpha\beta+\beta^2)\ln^3 e_\alpha \ln^2 e_\beta \\
& - \alpha(\alpha^2+6\alpha\beta+3\beta^2)\ln^2 e_\alpha \ln^3 e_\beta + \alpha^2(2\alpha+3\beta)\ln e_\alpha \ln^4 e_\beta - \alpha^3 \ln^5(e_\beta)\Big]\Bigg\}.
\end{aligned}
\tag{B.22}
$$

This equation is not easy to solve, which is one of the reasons we focused on mutual information in this paper.

C Equivalent NLL results from SCET

For the IRC safe angularities, ref. [22] demonstrated that the SCET approach and CAESAR approach to resummation give the same single differential cross sections to NLL accuracy. In this appendix, we repeat the same exercise for the generalized angularities. We also show how to perform the double differential interpolation of ref. [54] in the language of SCET.

C.1 One generalized angularity in SCET

The SCET calculation for the generalized angularities λ_β^κ mirrors that of track thrust [40]. We factorize the cross section into a hard function, jet functions, and a soft function which describe physics at the corresponding scales

$$
\mu_H = p_T\,, \quad \mu_J = (\lambda_\beta^\kappa)^{1/\beta}\, p_T R_0\,, \quad \mu_S = (\lambda_\beta^\kappa)^{1/\kappa}\, p_T R_0\,.
\tag{C.1}
$$

At NLL order, the cross section is completely generated by renormalization group evolution. For simplicity we evolve the hard and jet function to the soft scale, which results in a cumulative distribution of [37, 40, 73, 74]

$$
\Sigma_i(\lambda_\beta^\kappa) = \frac{e^{K_H^i + K_J^i - \gamma_E\,\eta_J^i}}{\Gamma(1+\eta_J^i)}\left(\frac{(\lambda_\beta^\kappa)^{1/\beta}\, p_T R_0}{\mu_J}\right)^{\beta\,\eta_J^i}\left(\frac{p_T^2}{\mu_H^2}\right)^{\eta_H^i}.
\tag{C.2}
$$

The evolution kernels that enter here are

$$K_H^i(\mu_H, \mu_S) = -2C_i \, K_\Gamma(\mu_H, \mu_S) + K_{\gamma_H^i}(\mu_H, \mu_S) \,, \qquad \eta_H^i(\mu_J, \mu_S) = C_i \, \eta_\Gamma(\mu_J, \mu_S) \,, \quad \text{(C.3)}$$

$$K_J^i(\mu_J, \mu_S) = \frac{2C_i \beta}{\beta - \kappa} \, K_\Gamma(\mu_J, \mu_S) + K_{\gamma_J^i}(\mu_J, \mu_S) \,, \qquad \eta_J^i(\mu_J, \mu_S) = -\frac{2C_i}{\beta - \kappa} \, \eta_\Gamma(\mu_J, \mu_S) \,,$$

which are given in terms of

$$K_\Gamma(\mu_0, \mu) = -\frac{\Gamma_0}{4\beta_0^2} \left[\frac{4\pi}{\alpha_s(\mu_0)} \left(1 - \frac{1}{r} - \ln r \right) + \left(\frac{\Gamma_1}{\Gamma_0} - \frac{\beta_1}{\beta_0} \right)(1 - r + \ln r) + \frac{\beta_1}{2\beta_0} \ln^2 r \right] \,,$$

$$\eta_\Gamma(\mu_0, \mu) = -\frac{\Gamma_0}{2\beta_0} \ln r \,,$$

$$K_\gamma(\mu_0, \mu) = -\frac{\gamma}{2\beta_0} \ln r \,, \tag{C.4}$$

where $r = \alpha_s(\mu)/\alpha_s(\mu_0)$. The coefficients of the beta function, the cusp, and the non-cusp anomalous dimensions are

$$\beta_0 = \frac{11}{3} C_A - \frac{4}{3} T_F \, n_f \,, \qquad\qquad \beta_1 = \frac{34}{3} C_A^2 - \left(\frac{20}{3} C_A + 4C_F \right) T_F \, n_f \,, \tag{C.5}$$

$$\Gamma_0 = 4 \,, \qquad\qquad \Gamma_1 = 4 \left[\left(\frac{67}{9} - \frac{\pi^2}{3} \right) C_A - \frac{20}{9} T_F \, n_f \right] \,,$$

$$\gamma_H^q = -6C_F \,, \qquad \gamma_H^g = -2\beta_0 \,, \qquad \gamma_J^q = 6C_F + \frac{8C_F}{\beta - \kappa} f_\kappa^{g,1} \,, \qquad \gamma_J^g = 2\beta_0 + \frac{8C_A}{\beta - \kappa} f_\kappa^{g,1} \,.$$

Note that the conventions for these coefficients differ from those in section B.1, and that the nonperturbative effect described by $f_\kappa^{g,1}$ is included in the non-cusp anomalous dimension.

To test the agreement with the IRC safe result in section B.1, we use the central scale choice in eq. (C.1), such that the SCET result for e_β can be written as

$$\Sigma_i(e_\beta) = \frac{e^{-\gamma_E \, \eta_J^i}}{\Gamma(1 + \eta_J^i)} e^{K_H^i + K_J^i}. \tag{C.6}$$

Up to terms that are beyond NLL order and ignoring logarithms of the jet radius R_0, we find

$$-2C_i \, K_\Gamma(\mu_H, \mu_S) + \frac{2C_i \beta}{\beta - 1} \, K_\Gamma(\mu_J, \mu_S) \overset{\text{NLL}}{\simeq} -R(e_\beta),$$

$$K_{\gamma_H^i}(\mu_H, \mu_S) + K_{\gamma_J^i}(\mu_J, \mu_S) = K_{\gamma_H^i}(\mu_H, \mu_J) \overset{\text{NLL}}{\simeq} \frac{1}{8} \gamma_H^i T(e_\beta) = -\gamma_i T(e_\beta),$$

$$K_H^i(\mu_H, \mu_S) + K_J^i(\mu_J, \mu_S) \overset{\text{NLL}}{\simeq} -R(e_\beta) - \gamma_i T(e_\beta),$$

$$\eta_J^i(\mu_J, \mu_S) \overset{\text{NLL}}{\simeq} R'(e_\beta), \tag{C.7}$$

so the CAESAR and SCET results indeed agree at this order. One advantage of the SCET approach is that it separates the physics at different scales, allowing one to estimate the perturbative uncertainty of eq. (C.2) by (independently) varying μ_H, μ_J, and μ_S.

For the IRC unsafe case in section 6.2, we need to verify the rescaling hypothesis in eq. (6.20), which is equivalent to

$$R_i(\lambda_\beta^\kappa) \stackrel{\text{NLL}}{\simeq} \hat{R}_i \left(e_{\beta/\kappa} = \left(\frac{\lambda_\beta^\kappa}{\exp(f_\kappa^{g,1})} \right)^{1/\kappa} \right),$$

$$T_i(\lambda_\beta^\kappa) \stackrel{\text{NLL}}{\simeq} \hat{T}_i \left(e_{\beta/\kappa} = \left(\lambda_\beta^\kappa \right)^{1/\kappa} \right),$$

$$R'_i(\lambda_\beta^\kappa) \stackrel{\text{NLL}}{\simeq} \frac{1}{\kappa} \hat{R}'_i \left(e_{\beta/\kappa} = \left(\lambda_\beta^\kappa \right)^{1/\kappa} \right), \tag{C.8}$$

where X is the function in the IRC unsafe case and $\hat{X}$ is the function in the IRC safe case (with angular exponent β/κ). Note that we can drop the $f_\kappa^{g,1}$ terms from T_i and R'_i to NLL accuracy, and that there is a $1/\kappa$ Jacobian factor from the logarithmic derivative in R'. Under this rescaling, the central scales are related as

$$\ln \frac{\mu_H}{\mu_S} = \ln \frac{\hat{\mu}_H}{\hat{\mu}_S} - \frac{1}{\kappa} f_\kappa^{g,1}, \qquad \ln \frac{\mu_J}{\mu_S} = \ln \frac{\hat{\mu}_J}{\hat{\mu}_S} - \frac{\beta - \kappa}{\kappa \beta} f_\kappa^{g,1}. \tag{C.9}$$

To NLL order, this has the effect of introducing additional terms in the evolution kernels

$$-2C_i K_\Gamma(\mu_H, \mu_S) + \frac{2C_i \beta}{\beta - \kappa} K_\Gamma(\mu_J, \mu_S) \stackrel{\text{NLL}}{\simeq} -2C_i K_\Gamma(\hat{\mu}_H, \hat{\mu}_S) + \frac{2C_i(\beta/\kappa)}{(\beta/\kappa) - 1} K_\Gamma(\hat{\mu}_J, \hat{\mu}_S)$$

$$- K_{\hat{\gamma}_J^i}(\hat{\mu}_J, \hat{\mu}_S) \frac{8C_i}{\beta - \kappa} \frac{f_\kappa^{g,1}}{\hat{\gamma}_J^i},$$

$$K_{\gamma_H^i}(\mu_H, \mu_S) + K_{\gamma_J^i}(\mu_J, \mu_S) \stackrel{\text{NLL}}{\simeq} K_{\gamma_H^i}(\hat{\mu}_H, \hat{\mu}_S) + K_{\hat{\gamma}_J^i}(\hat{\mu}_J, \hat{\mu}_S) \left(1 + \frac{8C_i}{\beta - \kappa} \frac{f_\kappa^{g,1}}{\hat{\gamma}_J^i} \right),$$

$$\eta_J^i(\mu_J, \mu_S) \stackrel{\text{NLL}}{\simeq} \frac{1}{\kappa} \hat{\eta}_J^i(\hat{\mu}_J, \hat{\mu}_S). \tag{C.10}$$

Because of the $1/\kappa$ factor in the last equation, the R' condition in eq. (C.8) is immediately satisfied. The cusp and non-cusp contributions in the first and second equations of eq. (C.10) do not individually satisfy the rescaling, but their sum does. In the first equation this arises from rescaling the hard and jet cusp contributions, and in the second equation it comes from the non-cusp anomalous dimension

$$\gamma_J^i = \hat{\gamma}_J^i + \frac{8C_i}{\beta - \kappa} f_\kappa^{g,1}. \tag{C.11}$$

In the CEASAR approach, this non-cusp contribution is part of the radiator R, such that R and T each satisfy the rescaling in eq. (C.8). Because of this shuffling of contributions, we do not expect the rescaling relation to persist beyond NLL accuracy.

C.2 Double differential interpolation in SCET

We now determine the double differential cross sections with SCET, starting with two IRC safe angularities and then generalizing to the IRC unsafe case. Following ref. [54], the known results on the boundaries $e_\alpha = e_\beta$, $(e_\alpha)^\beta = (e_\beta)^\alpha$ are used to build an interpolation

on the full phase space. As discussed in appendix C.1, the cross section up to NLL order is determined by the scales μ_H, μ_J, and μ_S, so we can implement the interpolation in terms of these scales.

The interpolation for the cumulative distribution $\Sigma(e_\alpha, e_\beta)$ in the angularities e_α and e_β must satisfy the following boundary conditions:

$$\Sigma(e_\alpha, e_\beta)|_{(e_\alpha)^\beta = (e_\beta)^\alpha} = \Sigma(e_\alpha)\,, \qquad\qquad \Sigma(e_\alpha, e_\beta)|_{e_\alpha = e_\beta} = \Sigma(e_\beta)\,,$$

$$\frac{\partial}{\partial e_\alpha}\Sigma(e_\alpha, e_\beta)|_{(e_\alpha)^\beta = (e_\beta)^\alpha} = \frac{\mathrm{d}}{\mathrm{d}e_\alpha}\Sigma(e_\alpha)\,, \qquad \frac{\partial}{\partial e_\beta}\Sigma(e_\alpha, e_\beta)|_{e_\alpha = e_\beta} = \frac{\mathrm{d}}{\mathrm{d}e_\beta}\Sigma(e_\beta)\,,$$

$$\frac{\partial}{\partial e_\beta}\Sigma(e_\alpha, e_\beta)|_{(e_\alpha)^\beta = (e_\beta)^\alpha} = 0\,, \qquad\qquad \frac{\partial}{\partial e_\alpha}\Sigma(e_\alpha, e_\beta)|_{e_\alpha = e_\beta} = 0\,. \qquad (\text{C}.12)$$

We start at the boundary $(e_\alpha)^\beta = (e_\beta)^\alpha$ where we may multiply the scales by arbitrary powers of $(e_\beta)^\alpha/(e_\alpha)^\beta$. If the jet and soft scales on the one boundary maps onto the jet and soft scales on the other boundary, we find

$$\mu_{J \to J} = (e_\beta)^{1/\beta}\, p_T R_0\,, \qquad \mu_{S \to S} = e_\alpha\, p_T R_0\,. \qquad (\text{C}.13)$$

If the jet and soft scales swap from one boundary to the other boundary, we get

$$\mu_{J \to S} = \left((e_\alpha)^{1-\beta}(e_\beta)^{\alpha-1}\right)^{1/(\alpha-\beta)} p_T R_0\,, \qquad \mu_{S \to J} = \left((e_\alpha)^{\alpha-1}(e_\beta)^{\alpha(1-\beta)/\beta}\right)^{1/(\alpha-\beta)} p_T R_0\,. \tag{C.14}$$

This is of course strange from the point of view of factorization, but one should remember that the factorization theorem on the boundary does not hold in the interior. Using these interpolating scales, we can write down a candidate form for the double radiator,

$$R(e_\alpha, e_\beta) = p K_\Gamma(\mu_H, \mu_{S \to S}) + q K_\Gamma(\mu_H, \mu_{J \to S}) + r K_\Gamma(\mu_H, \mu_{S \to S}) + s K_\Gamma(\mu_H, \mu_{S \to J})$$
$$+ t K_\Gamma(\mu_{J \to J}, \mu_{S \to S}) + u K_\Gamma(\mu_{J \to S}, \mu_{S \to J}) + v K_\Gamma(\mu_{J \to J}, \mu_{J \to S})$$
$$+ w K_\Gamma(\mu_{J \to J}, \mu_{S \to J}) + x K_\Gamma(\mu_{S \to S}, \mu_{J \to S}) + y K_\Gamma(\mu_{S \to S}, \mu_{S \to J})\,. \qquad (\text{C}.15)$$

Imposing the boundary conditions in eq. (C.12) to NLL order leads to a one parameter family of solutions. The simplest one is:

$$R(e_\alpha, e_\beta) = 2 C_i \Big(K_\Gamma(\mu_H, \mu_{J \to S}) + \frac{\beta}{1-\beta} K_\Gamma(\mu_{J \to J}, \mu_{J \to S}) + \frac{1}{\alpha-1} K_\Gamma(\mu_{S \to S}, \mu_{J \to S}) \Big)\,. \tag{C.16}$$

This agrees with eq. (B.13) up to terms that are beyond NLL order.

For the non-cusp piece, we have

$$-\gamma^i\, T(e_\alpha, e_\beta) = K_{\gamma_H^i}(\mu_H, \mu_{J \to J}) - \alpha_s\, \frac{\alpha-\beta}{\alpha}\, \frac{e_\alpha^{-\beta/(\alpha-\beta)} e_\beta^{\alpha/(\alpha-\beta)}}{4\pi\beta + 2\alpha_s\beta_0 \ln e_\beta}\,. \qquad (\text{C}.17)$$

(Remember that a different convention for β_0 is used here than in appendix B.) The first term satisfies the boundary conditions on the cumulative distribution and the derivative

boundary conditions at $e_\alpha = e_\beta$. The second term is power suppressed, except at the boundary $e_\alpha = (e_\beta)^{\alpha/\beta}$, and is introduced to satisfy the derivative boundary conditions there. It is formally beyond the order we are working so other choices are possible.

The single angularity result $R'(e_\alpha) = \eta_J^i(\mu_J, \mu_S) = -2C_i/(\beta-1)\eta_\Gamma(\mu_J, \mu_S)$ suggests an ansatz similar to eq. (C.15) for $\tilde{R}$. This leads to

$$\tilde{R}(e_\alpha, e_\beta) = -2C_i\left[\frac{1}{\alpha-1}\eta_\Gamma(\mu_{J\to S}, \mu_{S\to S}) + \frac{1}{\beta-1}\eta_\Gamma(\mu_{J\to J}, \mu_{J\to S})\right.$$
$$\left. - 4\alpha_s\frac{\alpha-\beta}{\alpha}\frac{e_\alpha^{-\beta/(\alpha-\beta)}e_\beta^{\alpha/(\alpha-\beta)}}{4\pi\beta + 2\alpha_s\beta_0\ln e_\beta}\right]. \tag{C.18}$$

The first two terms satisfies the boundary condition on the cumulative distribution and the derivative boundary condition at $e_\alpha = e_\beta$. As in eq. (C.17), the additional power suppressed terms take care of the derivative boundary conditions at $e_\alpha = (e_\beta)^{\alpha/\beta}$.

The interpolation for IRC unsafe angularities is a direct generalization of eq. (C.16)

$$R(\lambda_\alpha^\rho, \lambda_\beta^\kappa) = 2C_i\left(K_\Gamma(\mu_H, \mu_{J\to S}) + \frac{\beta}{\kappa-\beta}K_\Gamma(\mu_{J\to J}, \mu_{J\to S}) + \frac{\rho}{\alpha-\rho}K_\Gamma(\mu_{S\to S}, \mu_{J\to S})\right), \tag{C.19}$$

where we assume $\alpha/\rho > \beta/\kappa$ and the scales are modified to

$$\mu_{J\to J} = (\lambda_\beta^\kappa)^{1/\beta}p_T R_0\,,$$
$$\mu_{S\to S} = (\lambda_\alpha^\rho)^{1/\rho}p_T R_0\,,$$
$$\mu_{J\to S} = \left((\lambda_\alpha^\rho)^{\kappa-\beta}(\lambda_\beta^\kappa)^{\alpha-\rho}\right)^{1/(\alpha\kappa-\beta\rho)}p_T R_0\,,$$
$$\mu_{S\to J} = \left((\lambda_\alpha^\rho)^{\kappa(\alpha-\rho)/\rho}(\lambda_\beta^\kappa)^{\alpha(\kappa-\beta)/\beta}\right)^{1/(\alpha\kappa-\beta\rho)}p_T R_0\,. \tag{C.20}$$

The interpolation of the non-cusp piece in eq. (C.17) mostly carries over. The contribution from $\gamma_H^i = -\hat\gamma_J^i$ only involves the hard and jet scales, which are the same as before. However, the nonperturbative coefficients enter through $K_{\gamma_J}(\mu_J, \mu_S)$. Although one can build an interpolation similar to eq. (C.17), we also need an interpolation between $f_\rho^{g,1}/(\alpha-\rho)$ at $(\lambda_\alpha^\rho)^\beta = (\lambda_\beta^\kappa)^\alpha$ and $f_\kappa^{g,1}/(\beta-\kappa)$ at $(\lambda_\alpha^\rho)^\kappa = (\lambda_\beta^\kappa)^\rho$. It is therefore much more convenient to use the rescaling trick in eq. (6.32). Finally, the multiple emissions contribution in eq. (C.18) generalizes to

$$\tilde{R}(\lambda_\alpha^\rho, \lambda_\beta^\kappa) = -2C_i\left[\frac{1}{\alpha-\rho}\eta_\Gamma(\mu_{J\to S}, \mu_{S\to S}) + \frac{1}{\beta-\kappa}\eta_\Gamma(\mu_{J\to J}, \mu_{J\to S})\right. \tag{C.21}$$
$$- 4\alpha_s\frac{(\alpha-\beta)^2}{\alpha(\alpha\kappa-\beta\rho)}\frac{(\lambda_\alpha^\rho)^{-\beta/(\alpha-\beta)}(\lambda_\beta^\kappa)^{\alpha/(\alpha-\beta)}}{4\pi\beta + 2\alpha_s\beta_0\ln\lambda_\beta^\kappa}$$
$$\left. - 4\alpha_s\frac{(\rho-\kappa)^2}{\kappa(\alpha\kappa-\beta\rho)}\frac{(\lambda_\alpha^\rho)^{\kappa/(\kappa-\rho)}(\lambda_\beta^\kappa)^{-\rho/(\kappa-\rho)}}{4\pi\rho + 2\alpha_s\beta_0\ln\lambda_\alpha^\rho}\right].$$

The terms on the first line satisfy the boundary condition for the cumulative distribution. The terms on the second and third line are power suppressed except at the boundaries
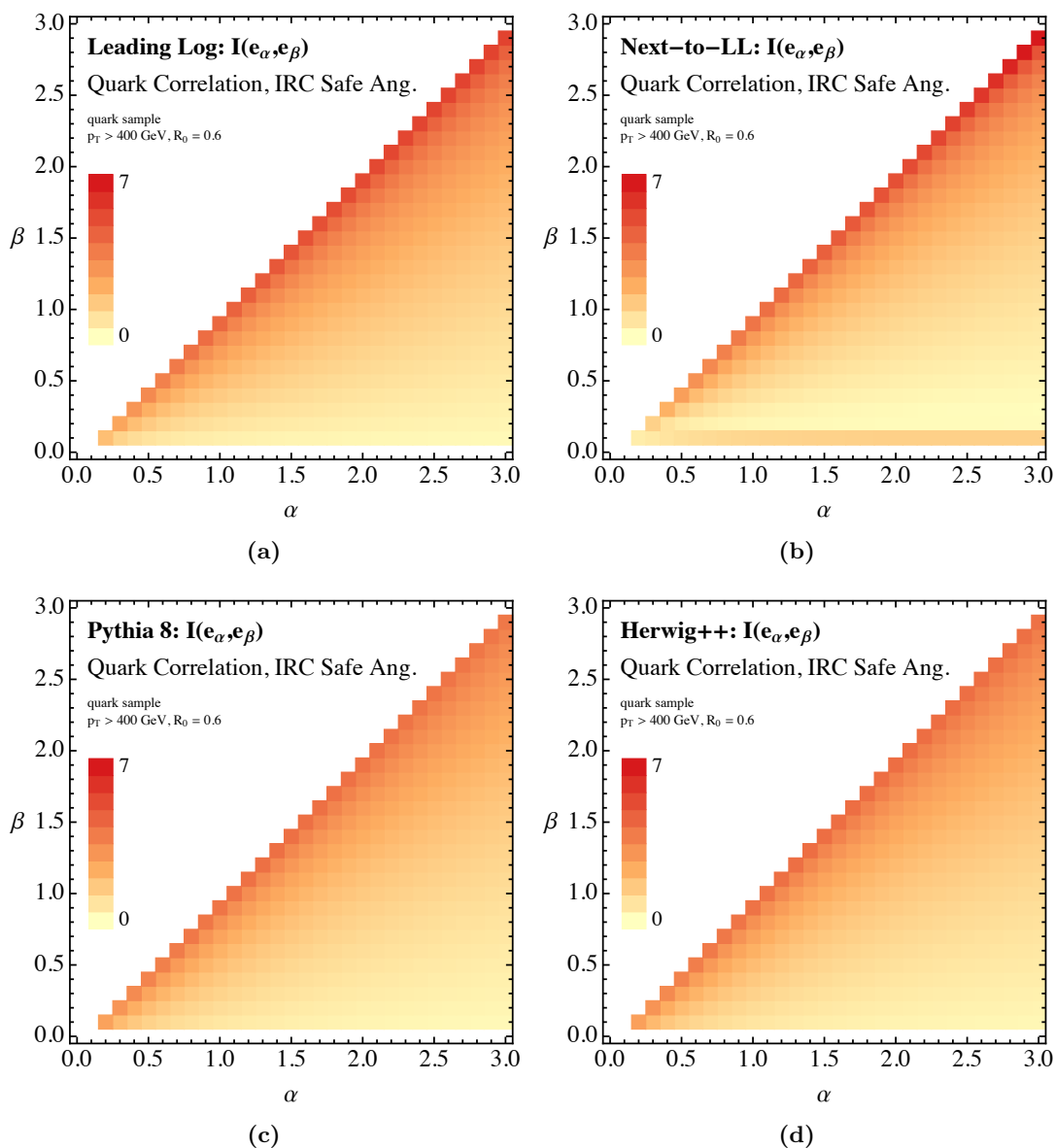
Figure 14. Correlation between two IRC safe angularities (e_α, e_β) on a pure quark jet sample. Top: the LL and NLL analytic calculations. Bottom: the PYTHIA 8 and HERWIG++ parton showers.

$(\lambda_\alpha^\rho)^\beta = (\lambda_\beta^\kappa)^\alpha$ and $(\lambda_\alpha^\rho)^\kappa = (\lambda_\beta^\kappa)^\rho$, respectively, where their inclusion enforces the derivative boundary conditions. For $\rho = \kappa$ the third line is absent, and for $\rho = \kappa = 1$ this reduces to the IRC safe case.

D   Additional plots

In this appendix, we present additional plots involving mutual information to complement the truth overlap plots in the main text. We also show some raw angularity distributions.
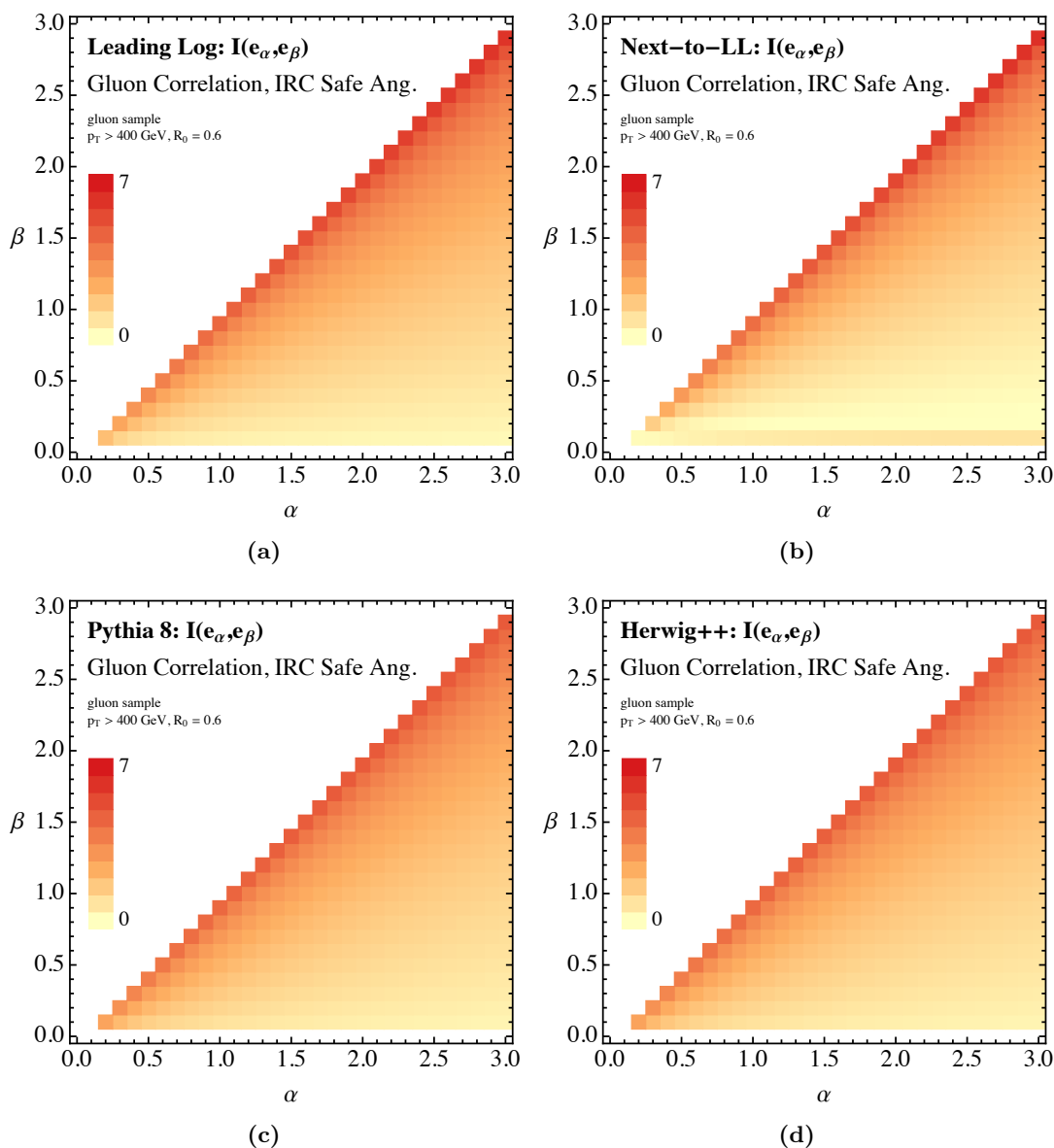
Figure 15. Same as figure 14, but on a pure gluon sample.

We first study the correlation between IRC safe angularities with different angular exponents, as measured by $I(e_\alpha; e_\beta)$. This is shown for a pure sample of quark jets in figure 14 and for a pure sample of gluon jets in figure 15. Two angularities are highly correlated when their angular exponents are close to one another, and become increasingly uncorrelated as the angular exponents move farther apart. This behavior can be understood from the definition of the angularities. For large values of angular exponent, the angularity is dominated by soft, wide-angle emissions because collinear emissions are suppressed by small angles raised to a high power. By contrast, at small values of the angular exponent, the angularity is dominated by hard collinear emissions. Thus, when two angularities have
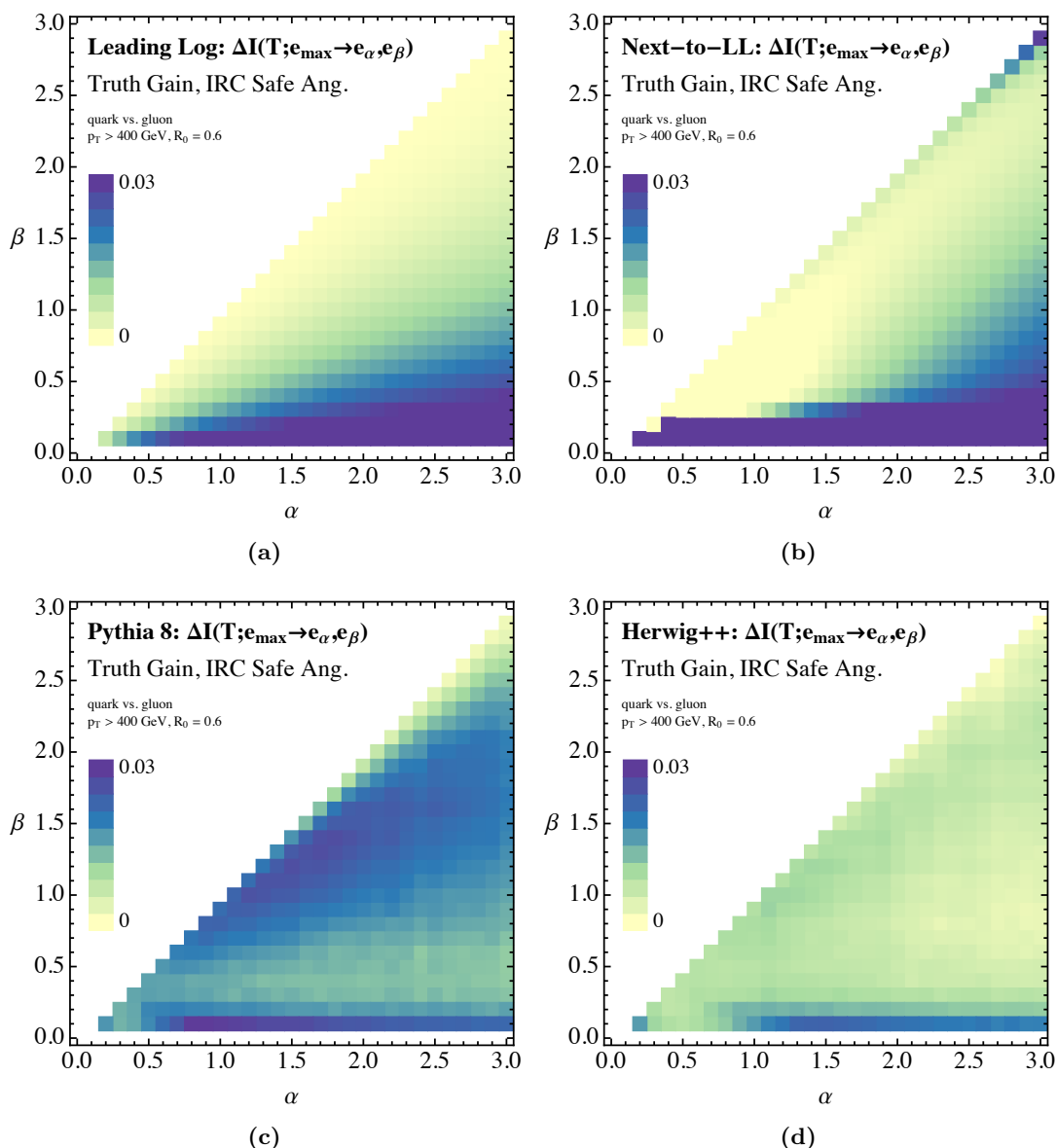
Figure 16. Improvement in the truth overlap by using two IRC safe angularities (e_α, e_β) as compared to only one, $I(T; e_\alpha, e_\beta) - \max[I(T; e_\alpha), I(T; e_\beta)]$. Top: the LL and NLL analytic calculations. Bottom: the PYTHIA 8 and HERWIG++ parton showers.

very different angular exponents, their values are dominated by different physics and so are largely uncorrelated. Unlike the case of the truth overlap in figure 5, there is broad agreement between LL, NLL, PYTHIA 8, and HERWIG++ as far as the raw correlations are concerned.

Next, we want to understand better the degree to which two angularities have more truth overlap than one angularity. In figure 16, we plot $\Delta I(T, e_{\max} \to e_\alpha, e_\beta)$ from eq. (5.6), namely the pairwise truth overlap $I(T; e_\alpha, e_\beta)$ minus the truth overlap of the stronger angularity $\max\{I(T; e_\alpha), I(T; e_\beta)\}$. The information gain is on the order of 10% ($\mathcal{O}(0.01)$

compared to a baseline truth overlap of $\mathcal{O}(0.1)$). As in figure 5, there are quite substantial differences between the various methods. It is interesting that in PYTHIA 8 one can already achieve considerable gains in performance just off the diagonal, i.e. for observables that are not very different. This may be because when α and β are close (compare to eq. (4.4)),

$$e_\alpha - e_\beta \simeq (\alpha - \beta) \sum_i z_i \theta_i^\alpha \log \theta_i + \dots, \tag{D.1}$$

and $\theta_i^\alpha \log \theta_i$ is similar to the optimal kernel found in refs. [4, 5].

Turning to the generalized angularities, in figures 17 and 18 we show the correlations on pure quark and gluon samples as measured by $I(\lambda_\alpha^\rho, \lambda_\beta^\kappa)$. As in the IRC unsafe case, there is broad agreement in the overall degree of correlation, though one has to be mindful of the restricted range of validity of the (N)LL calculations. In figure 19, we show the improvement of using two generalized angularities compared to one. The LL and NLL calculations are not shown, since those calculations are not accurate enough to assess small differences. The comparison between PYTHIA 8 and HERWIG++ is similar to the IRC safe case, with PYTHIA 8 being more optimistic about the gains possible by combining observables.

Finally, we show a few raw angularity distributions from the NLL calculation and both parton showers. In figure 20, we show single differential distributions for e_1 as an IRC safe example and $\lambda_2^{0.6}$ as an IRC unsafe example. Note that the NLL calculations lack important hadronization corrections that are modelled by the parton showers and are particularly important to correctly describe the small angularity region. The NLL result for $\lambda_2^{0.6}$ cuts off rather sharply at the low end due to our treatment of the QCD Landau pole. The PYTHIA 8 distributions are more peaked than the HERWIG++ distributions, which is part of the reason why PYTHIA 8 predicts improved discrimination power compared to HERWIG++. We then show the double differential distribution for e_1 and $\lambda_2^{0.6}$ in figure 21, showing only the half-maximum contour for readability. There is an irreducible degree of correlation between these observables due to phase space constraints (see eq. (6.24)), but one can see that, while the gluon contours are similar for PYTHIA 8 and HERWIG++, the PYTHIA 8 contour for quarks is significantly smaller than HERWIG++. This explains the enhanced discrimination power predicted by PYTHIA 8. The NLL contours are much larger in size than either PYTHIA 8 or HERWIG++ because the NLL distributions do not vanish at the phase space boundaries [54]. The analytic distribution will only vanish at the boundaries starting at NLL$'$ order, beyond the accuracy to which double differential cross sections have as-of-yet been computed.
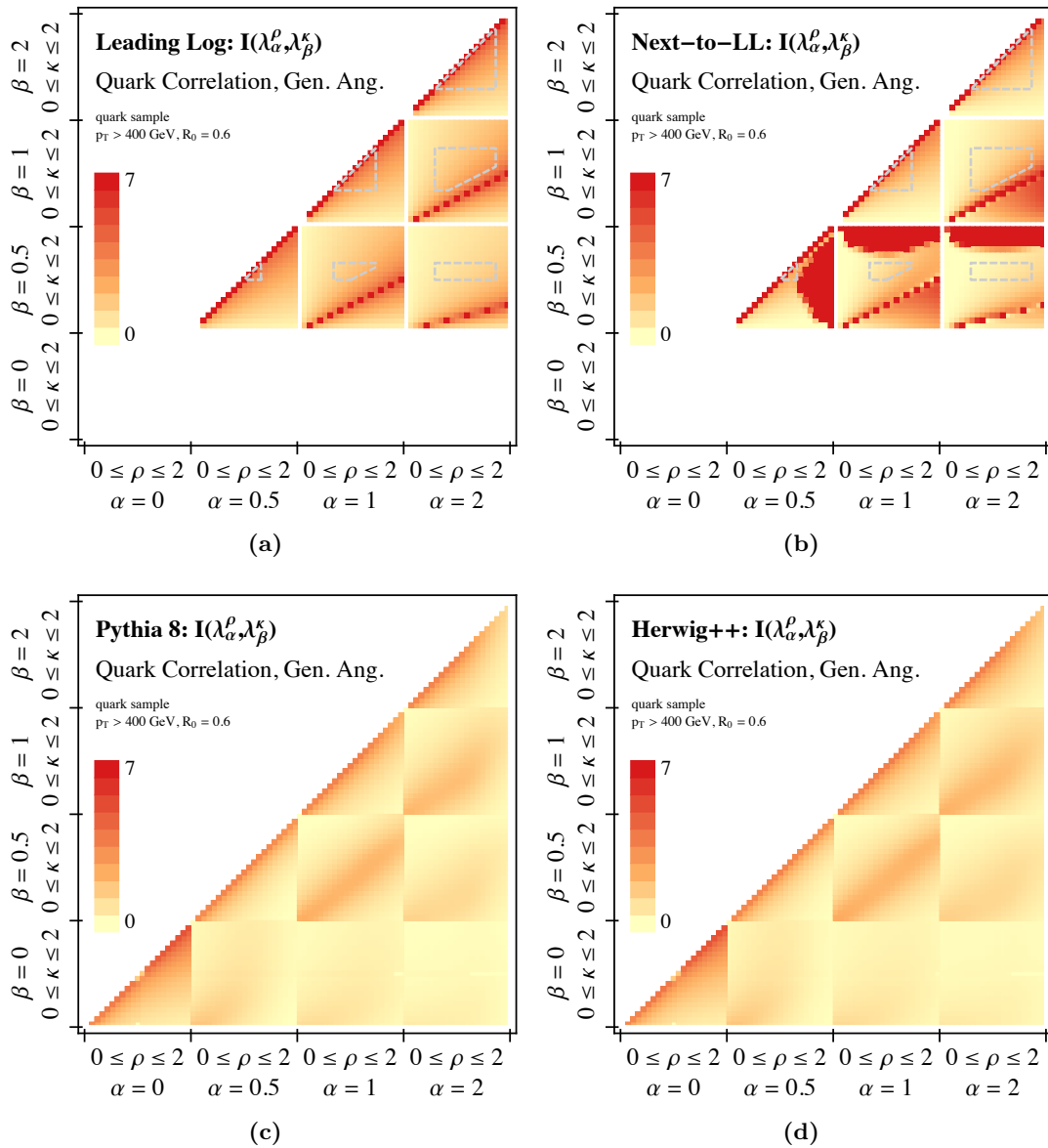
Figure 17. Correlation between two generalized angularities $(\lambda_\alpha^\rho, \lambda_\beta^\kappa)$ on a pure quark sample. Top: the LL and NLL analytic calculations. Bottom: the PYTHIA 8 and HERWIG++ parton showers. As in figure 13, $\beta \in \{0, 0.5, 1, 2\}$ and we sweep $0 \le \kappa \le 2$.
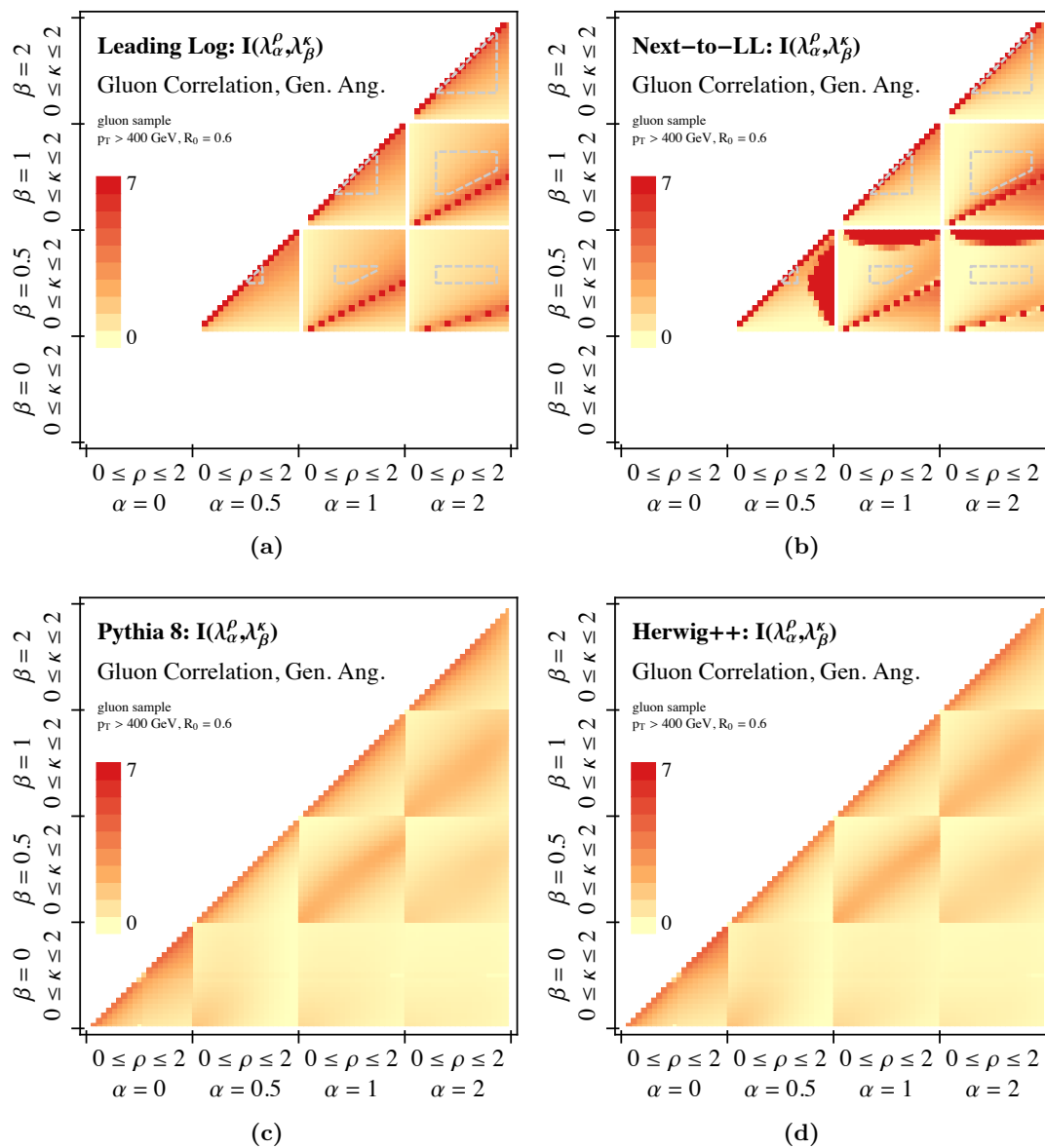
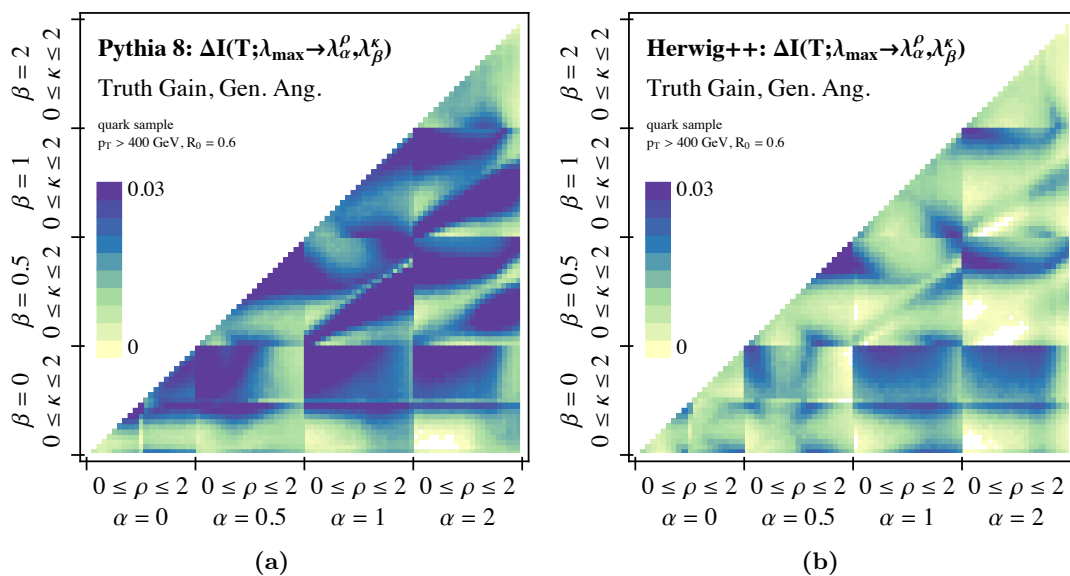**Figure 18.** Same as figure 17, but for a pure gluon sample.

Figure 19. Improvement in the truth overlap by using two generalized angularities $(\lambda_\alpha^\rho, \lambda_\beta^\kappa)$ as compared to only one, $I(T; \lambda_\alpha^\rho, \lambda_\beta^\kappa) - \max[I(T; \lambda_\alpha^\rho), I(T; \lambda_\beta^\kappa)]$. We only show the PYTHIA 8 and HERWIG++ parton showers since the LL and NLL analytic calculations are not sufficiently accurate to extract subtle differences in truth overlap.
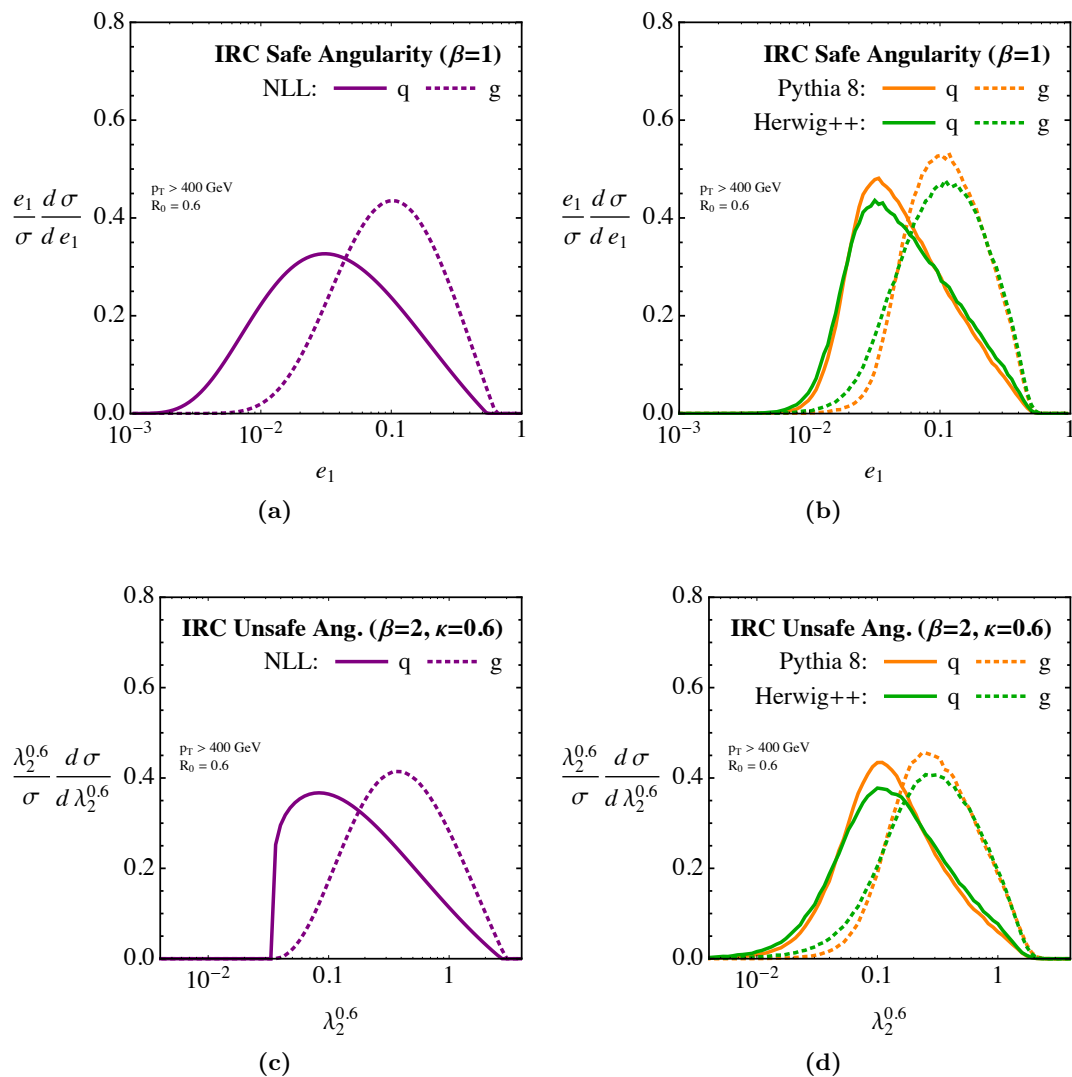
Figure 20. Raw distributions of e_1 (top) and $\lambda_2^{0.6}$ (bottom) for the NLL calculation (left) and parton showers (right). Note that the NLL distributions lack hadronization corrections that are present in the parton showers, which affects small values of the angularities.
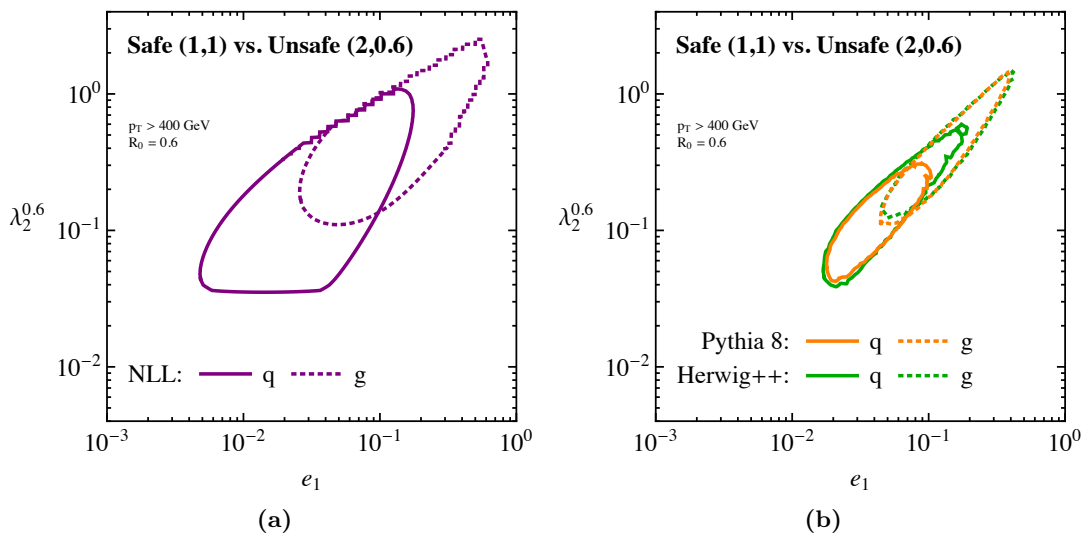
Figure 21. Double differential distributions in the e_1–$\lambda_2^{0.6}$ plane for the NLL calculation (left) and parton showers (right). The contours correspond to the half maximum of $\mathrm{d}^2\sigma/(\mathrm{d}\ln e_1\,\mathrm{d}\ln\lambda_2^{0.6})$.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] A. Abdesselam, E.B. Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth et al., *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur. Phys. J.* **C 71** (2011) 1661 [arXiv:1012.5412] [INSPIRE].

[2] A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J. Phys.* **G 39** (2012) 063001 [arXiv:1201.0008] [INSPIRE].

[3] A. Altheimer, A. Arce, L. Asquith, J. Backus Mayes, E. Bergeaas Kuutmann et al., *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd-27th of July 2012*, *Eur. Phys. J.* **C 74** (2014) 2792 [arXiv:1311.2708] [INSPIRE].

[4] J. Gallicchio and M.D. Schwartz, *Quark and Gluon Tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [arXiv:1106.3076] [INSPIRE].

[5] J. Gallicchio and M.D. Schwartz, *Quark and Gluon Jet Substructure*, *JHEP* **04** (2013) 090 [arXiv:1211.7038] [INSPIRE].

[6] D. Krohn, M.D. Schwartz, T. Lin and W.J. Waalewijn, *Jet Charge at the LHC*, *Phys. Rev. Lett.* **110** (2013) 212001 [arXiv:1209.2421] [INSPIRE].

[7] CMS collaboration, *Search for a Higgs boson in the decay channel H to ZZ(*) to q qbar ℓ⁻ l+ in pp collisions at* $\sqrt{s} = 7$ *TeV*, *JHEP* **04** (2012) 036 [arXiv:1202.1416] [INSPIRE].

[8] 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.

[9] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [INSPIRE].

[10] CMS Collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, CMS-PAS-JME-13-002.

[11] CMS Collaboration, *Pileup Jet Identification*, CMS-PAS-JME-13-005.

[12] ATLAS collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C 74** (2014) 3023 [arXiv:1405.6583] [INSPIRE].

[13] ATLAS collaboration, *Performance and Validation of Q-Jets at the ATLAS Detector in pp Collisions at $\sqrt{s}$=8 TeV in 2012*, ATLAS-CONF-2013-087.

[14] CMS Collaboration, *Identifying Hadronically Decaying Vector Bosons Merged into a Single Jet*, CMS-PAS-JME-13-006.

[15] P. Carruthers and C.C. Shih, *Mutual Information and Forward Backward Correlations in Multi — Hadron Production*, *Phys. Rev. Lett.* **62** (1989) 2073 [INSPIRE].

[16] I. Narsky and F.C. Porter, *Statistical analysis techniques in particle physics*, Wiley-VCH, Berlin, 2014.

[17] H. Casini and M. Huerta, *A Finite entanglement entropy and the c-theorem*, *Phys. Lett.* **B 600** (2004) 142 [hep-th/0405111] [INSPIRE].

[18] *Physical Review Letters* **100** (2008) 070502 [arXiv:0704.3906].

[19] P. Calabrese and J. Cardy, *Entanglement entropy and conformal field theory*, *J. Phys.* **A 42** (2009) 504005 [arXiv:0905.4013] [INSPIRE].

[20] M. Headrick, *Entanglement Renyi entropies in holographic theories*, *Phys. Rev.* **D 82** (2010) 126010 [arXiv:1006.0047] [INSPIRE].

[21] D. Bertolini, T. Chan and J. Thaler, *Jet Observables Without Jet Algorithms*, *JHEP* **04** (2014) 013 [arXiv:1310.7584] [INSPIRE].

[22] A.J. Larkoski, D. Neill and J. Thaler, *Jet Shapes with the Broadening Axis*, *JHEP* **04** (2014) 017 [arXiv:1401.2158] [INSPIRE].

[23] G. Salam, *Notes on the Winner-Take-All Axis*, unpublished.

[24] S. Catani, G. Turnock and B.R. Webber, *Jet broadening measures in e^+e^- annihilation*, *Phys. Lett.* **B 295** (1992) 269 [INSPIRE].

[25] 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] [INSPIRE].

[26] A. Banfi, G.P. Salam and G. Zanderighi, *Principles of general final-state resummation and automated implementation*, *JHEP* **03** (2005) 073 [hep-ph/0407286] [INSPIRE].

[27] A.J. Larkoski and J. Thaler, *Aspects of Jets at 100 TeV*, *Phys. Rev.* **D 90** (2014) 034010 [arXiv:1406.7011] [INSPIRE].

[28] C.F. Berger, T. Kucs and G.F. Sterman, *Event shape/energy flow correlations*, *Phys. Rev.* **D 68** (2003) 014012 [hep-ph/0303051] [INSPIRE].

[29] L.G. Almeida, S.J. Lee, G. Perez, G.F. Sterman, I. Sung et al., *Substructure of high-p_T Jets at the LHC*, *Phys. Rev.* **D 79** (2009) 074017 [arXiv:0807.0234] [INSPIRE].

[30] 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] [INSPIRE].

[31] P.E.L. Rakow and B.R. Webber, *Transverse Momentum Moments of Hadron Distributions in QCD Jets*, *Nucl. Phys.* **B 191** (1981) 63 [INSPIRE].

[32] R.K. Ellis and B. Webber, *QCD Jet Broadening in Hadron Hadron Collisions*, *Conf. Proc.* **C860623** (1986) 74.

[33] E. Farhi, *A QCD Test for Jets*, *Phys. Rev. Lett.* **39** (1977) 1587 [INSPIRE].

[34] M. Dasgupta, A. Fregoso, S. Marzani and G.P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029 [arXiv:1307.0007] [INSPIRE].

[35] M. Dasgupta, A. Fregoso, S. Marzani and A. Powling, *Jet substructure with analytical methods*, *Eur. Phys. J.* **C 73** (2013) 2623 [arXiv:1307.0013] [INSPIRE].

[36] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146 [arXiv:1402.2657] [INSPIRE].

[37] A. Hornig, C. Lee and G. Ovanesyan, *Effective Predictions of Event Shapes: Factorized, Resummed and Gapped Angularity Distributions*, *JHEP* **05** (2009) 122 [arXiv:0901.3780] [INSPIRE].

[38] W.J. Waalewijn, *Calculating the Charge of a Jet*, *Phys. Rev.* **D 86** (2012) 094030 [arXiv:1209.3019] [INSPIRE].

[39] 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] [INSPIRE].

[40] H.-M. Chang, M. Procura, J. Thaler and W.J. Waalewijn, *Calculating Track Thrust with Track Functions*, *Phys. Rev.* **D 88** (2013) 034030 [arXiv:1306.6630] [INSPIRE].

[41] A. Banfi, G.P. Salam and G. Zanderighi, *Infrared safe definition of jet flavor*, *Eur. Phys. J.* **C 47** (2006) 113 [hep-ph/0601139] [INSPIRE].

[42] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *Dissecting Soft Radiation with Factorization*, arXiv:1405.6722 [INSPIRE].

[43] J. Gallicchio and M.D. Schwartz, *Pure Samples of Quark and Gluon Jets at the LHC*, *JHEP* **10** (2011) 103 [arXiv:1104.1175] [INSPIRE].

[44] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.

[45] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[46] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[47] M. Bahr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton et al., *HERWIG++ Physics and Manual*, *Eur. Phys. J.* **C 58** (2008) 639 [arXiv:0803.0883] [INSPIRE].

[48] S. Gieseke, D. Grellscheid, K. Hamilton, A. Papaefstathiou, S. Platzer et al., *HERWIG++ 2.5 Release Note*, arXiv:1102.1672 [INSPIRE].

[49] I.W. Stewart, F.J. Tackmann and W.J. Waalewijn, *N-Jettiness: An Inclusive Event Shape to Veto Jets*, *Phys. Rev. Lett.* **105** (2010) 092002 [arXiv:1004.2489] [InSPIRE].

[50] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015 [arXiv:1011.2268] [InSPIRE].

[51] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, *JHEP* **02** (2012) 093 [arXiv:1108.2701] [InSPIRE].

[52] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C 72** (2012) 1896 [arXiv:1111.6097] [InSPIRE].

[53] A.J. Larkoski and J. Thaler, *Unsafe but Calculable: Ratios of Angularities in Perturbative QCD*, *JHEP* **09** (2013) 137 [arXiv:1307.1699] [InSPIRE].

[54] 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] [InSPIRE].

[55] M. Procura, W.J. Waalewijn and L. Zeune, *Resummation of Double-Differential Cross Sections and Fully-Unintegrated Parton Distribution Functions*, arXiv:1410.6483 [InSPIRE].

[56] C.W. Bauer, S. Fleming and M.E. Luke, *Summing Sudakov logarithms in $B \to X_s\gamma$ in effective field theory*, *Phys. Rev.* **D 63** (2000) 014006 [hep-ph/0005275] [InSPIRE].

[57] C.W. Bauer, S. Fleming, D. Pirjol and I.W. Stewart, *An Effective field theory for collinear and soft gluons: Heavy to light decays*, *Phys. Rev.* **D 63** (2001) 114020 [hep-ph/0011336] [InSPIRE].

[58] C.W. Bauer and I.W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett.* **B 516** (2001) 134 [hep-ph/0107001] [InSPIRE].

[59] C.W. Bauer, D. Pirjol and I.W. Stewart, *Soft collinear factorization in effective field theory*, *Phys. Rev.* **D 65** (2002) 054022 [hep-ph/0109045] [InSPIRE].

[60] D.E. Soper and M. Spannowsky, *Finding physics signals with shower deconstruction*, *Phys. Rev.* **D 84** (2011) 074002 [arXiv:1102.3480] [InSPIRE].

[61] A. Capella, I.M. Dremin, J.W. Gary, V.A. Nechitailo and J. Tran Thanh Van, *Evolution of average multiplicities of quark and gluon jets*, *Phys. Rev.* **D 61** (2000) 074009 [hep-ph/9910226] [InSPIRE].

[62] P. Bolzoni, B.A. Kniehl and A.V. Kotikov, *Gluon and quark jet multiplicities at $N^3LO+NNLL$*, *Phys. Rev. Lett.* **109** (2012) 242002 [arXiv:1209.5914] [InSPIRE].

[63] M. Procura and I.W. Stewart, *Quark Fragmentation within an Identified Jet*, *Phys. Rev.* **D 81** (2010) 074009 [*Erratum ibid.* **D 83** (2011) 039902] [arXiv:0911.4980] [InSPIRE].

[64] A. Jain, M. Procura and W.J. Waalewijn, *Parton Fragmentation within an Identified Jet at NNLL*, *JHEP* **05** (2011) 035 [arXiv:1101.4953] [InSPIRE].

[65] V.N. Gribov and L.N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438 [InSPIRE].

[66] H. Georgi and H.D. Politzer, *Electroproduction scaling in an asymptotically free theory of strong interactions*, *Phys. Rev.* **D 9** (1974) 416 [InSPIRE].

[67] D.J. Gross and F. Wilczek, *asymptotically free gauge theories. 2.*, *Phys. Rev.* **D 9** (1974) 980 [InSPIRE].

[68] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys.* **B 126** (1977) 298 [INSPIRE].

[69] Y.L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641 [INSPIRE].

[70] M. Dasgupta and G.P. Salam, *Resummation of nonglobal QCD observables*, *Phys. Lett.* **B 512** (2001) 323 [hep-ph/0104277] [INSPIRE].

[71] S. Catani, B.R. Webber and G. Marchesini, *QCD coherent branching and semiinclusive processes at large x*, *Nucl. Phys.* **B 349** (1991) 635 [INSPIRE].

[72] 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] [INSPIRE].

[73] T. Becher and M.D. Schwartz, *A precise determination of α_s from LEP thrust data using effective field theory*, *JHEP* **07** (2008) 034 [arXiv:0803.0342] [INSPIRE].

[74] L.G. Almeida, S.D. Ellis, C. Lee, G. Sterman, I. Sung et al., *Comparing and counting logs in direct and effective methods of QCD resummation*, *JHEP* **04** (2014) 174 [arXiv:1401.4460] [INSPIRE].