

RECEIVED: August 3, 2022

ACCEPTED: November 24, 2022

PUBLISHED: December 5, 2022

# Systematic quark/gluon identification with ratios of likelihoods

Samuel Bright-Thonney,<sup>a</sup> Ian Moult,<sup>c</sup> Benjamin Nachman<sup>d</sup> and Stefan Prestel<sup>e</sup>

<sup>a</sup>Physics Department, Cornell University,  
109 Clark Hall, Ithaca, NY 14853, U.S.A.

<sup>c</sup>Department of Physics, Yale University,  
New Haven, CT 06511, U.S.A.

<sup>d</sup>Physics Division, Lawrence Berkeley National Laboratory,  
Berkeley, CA 94720, U.S.A.

<sup>e</sup>Department of Astronomy and Theoretical Physics, Lund University,  
S-223 62 Lund, Sweden

E-mail: [skb93@cornell.edu](mailto:skb93@cornell.edu), [ian.moult@yale.edu](mailto:ian.moult@yale.edu), [bpnachman@lbl.gov](mailto:bpnachman@lbl.gov),  
[stefan.prestel@thep.lu.se](mailto:stefan.prestel@thep.lu.se)

**ABSTRACT:** Discriminating between quark- and gluon-initiated jets has long been a central focus of jet substructure, leading to the introduction of numerous observables and calculations to high perturbative accuracy. At the same time, there have been many attempts to fully exploit the jet radiation pattern using tools from statistics and machine learning. We propose a new approach that combines a deep analytic understanding of jet substructure with the optimality promised by machine learning and statistics. After specifying an approximation to the full emission phase space, we show how to construct the optimal observable for a given classification task. This procedure is demonstrated for the case of quark and gluons jets, where we show how to systematically capture sub-eikonal corrections in the splitting functions, and prove that linear combinations of weighted multiplicity is the optimal observable. In addition to providing a new and powerful framework for systematically improving jet substructure observables, we demonstrate the performance of several quark versus gluon jet tagging observables in parton-level Monte Carlo simulations, and find that they perform at or near the level of a deep neural network classifier. Combined with the rapid recent progress in the development of higher order parton showers, we believe that our approach provides a basis for systematically exploiting subleading effects in jet substructure analyses at the Large Hadron Collider (LHC) and beyond.

**KEYWORDS:** Jets and Jet Substructure, Parton Shower

ARXIV EPRINT: [2207.12411](https://arxiv.org/abs/2207.12411)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The structure of quark and gluon jets</b>	<b>3</b>
<b>3</b>	<b>Quark and gluon jet likelihoods</b>	<b>4</b>
3.1	Analysis in the eikonal limit	5
3.2	Systematic expansion beyond the eikonal limit	7
3.3	Perturbative multiplicity and parton showers	8
3.4	Discussion and extensions	9
<b>4</b>	<b>Monte Carlo studies</b>	<b>10</b>
4.1	Configurable parton showers	10
4.2	Optimality & particle flow networks	11
4.3	Results: PYTHIA & DIRE	12
4.3.1	Parton shower validation	12
4.3.2	Exponential modifications of $q \rightarrow qg$	12
4.3.3	Restoring $C_F = 4/3$	15
4.4	Further investigation: toy showers and hadronization	16
4.5	Discussion	17
<b>5</b>	<b>Conclusions</b>	<b>18</b>

---

## 1 Introduction

Jets are collimated sprays of particles produced by high energy quarks and gluons. The radiation pattern within jets — jet substructure — provides a powerful set of tools for identifying the origin of jets [1–7]. For example, jet substructure has been used to differentiate between jets originating from quarks and those originating from gluons, as well as to distinguish jets due to highly Lorentz-boosted massive particles (e.g.  $W/Z$ /Higgs bosons or top quarks) and generic quark and gluon-initiated jets.

In most cases, jet classification proceeds by constructing a small set of physically motivated observables. These observables are often required to have certain properties like infrared and collinear (IRC) safety (or looser requirements like Sudakov safety [8, 9]) so that their cross section can be calculated in perturbation theory. Furthermore, these observables are usually built to isolate different regions of phase space that are predominantly occupied by one class of jet or another. In some cases, the decision boundaries are defined using heuristic arguments, while in other applications power counting [10–16] or other theoretical tools are employed.

While this program has resulted in many interesting physics results, it is not systematically improvable. In particular, a stream of new observables have been proposed over the last decade with increasingly superior numerical performance on particular physics problems. There is no guarantee that the latest observable is optimal and so there is no natural way of knowing when further observable development is no longer necessary. In parallel to the improvement in observables, there has been significant improvement in the perturbative accuracy of parton showers [17–26], incorporating increasingly subtle features into the description of jets that can be exploited for classification tasks. Combined, this suggests that a more systematic approach to observable construction is required.

In this paper we propose such a systematic approach, that is complementary and opposite to the philosophy commonly taken in the literature. Instead of first positing an observable and then calculating it to high precision, we propose to specify a given precision and then compute the optimal observable. Given a probability density<sup>1</sup> for jets originating from particles  $i$  and  $j$ , the optimal observable for differentiating these types of jets is the likelihood ratio [27]. We specifically focus on the universal part of the cross section and consider higher order approximations to the emission phase space density. While we focus on quark versus gluon jet tagging, the core idea is applicable to a variety of jet tagging tasks. Using this approach, we will show that we can systematically incorporate sub-eikonal corrections to the splitting functions into the observable definition. The resulting observable itself can be computed to higher orders in perturbation theory, but the observable may no longer be optimal in the sense defined above.

Although our approach yields calculable and systematically improvable approximations of the likelihood ratio, its promise of optimality is strictly limited to the perturbative regime. Any nonperturbative effects — most notably hadronization in the case of jet physics — are not accounted for in our calculations, and we will show that they significantly constrain the classification power of our observables. Devising a set of classifying observables that are simultaneously interpretable, (approximately) optimal, and robust to nonperturbative physics is a significant challenge, and falls beyond the scope of this work.

While we find it unlikely that analytic classifiers can completely replace modern machine learning (ML) techniques, particularly those that exploit information beyond energy flow, we still believe that further developing an understanding of the nature of physical information in jets and how it can be exploited for classification tasks is important. First, it can help to reduce the “black box” nature of ML techniques, enabling one to gain confidence about the validity of ML techniques in jet physics. Second, it allows one to identify the physical origin of information exploited by the classification task, so that one can ensure that the relevant effects are well modelled, or to motivate further work to improve their description.

Our approach is related to a number of other proposals in the literature. Most prominently, likelihood-based methods built on cross section calculations underly the Matrix Element Method (MEM) first proposed for top quark mass measurements at the Tevatron [28]. However, the MEM focuses on the hard-scattering part of the cross section and is not read-

---

<sup>1</sup>In practice, these are computed at a given order in perturbation theory and could be negative. As long as these regions are small and isolated, zeroing them is likely sufficient.

ily extendable to other final states. In the context of jet substructure, an observable similar to the one proposed in this paper is shower deconstruction (SD) [29–31]. Like our optimal observables, SD is built as a likelihood ratio from approximations to the parton shower, however, a key difference is that we are able to translate our result into (an expansion in) standard jet observables. Other recent work in this direction can be found in [32–38].

This paper is organized as follows. In section 2 we provide a review of the structure of quark and gluon jets, emphasizing the organization of the  $1 \rightarrow 2$  splitting functions into eikonal and non-eikonal structures. In section 3 we describe our approach to constructing optimal classifiers using likelihood ratios, and consider the explicit example of quark versus gluon jet tagging at leading logarithmic (LL) and modified leading logarithmic (MLL) order. In section 4 we validate our approach using parton shower (PS) Monte Carlo (MC) simulations, and study the resulting observables. We conclude in section 5.

## 2 The structure of quark and gluon jets

We begin by analyzing the perturbative structure of a jet sourced by a hard parton radiating massless gauge bosons (the radiation of fermion pairs will be addressed later). Here, we will restrict ourselves to LL or MLL, where we can use the  $1 \rightarrow 2$  splitting functions. Extensions to higher logarithmic accuracy could be performed by analyzing the  $1 \rightarrow n$  splitting functions. We believe that this is particularly interesting in light of progress in the description of quark and gluon jets with parton shower programs.

While the structure of quark and gluon jets has been extensively discussed in the jet substructure literature, this has almost entirely been from the perspective of the soft-collinear limit, where the only distinction between quarks and gluons is their color charge. However, the  $1 \rightarrow 2$  splitting functions, which are the basis of standard parton shower programs, contain more information. Here we would like to clearly understand the physical nature of the additional information, how it can be exploited, and why it is small compared to the  $C_F$  vs.  $C_A$  color information.

Although we will ultimately be focused on the physically realized case of QCD, it is interesting to view the problem more generally. This perspective has also been inspired by arguments for the simplification of higher order splitting functions in refs. [39–41]. We therefore consider the splitting function for the radiation of a massless gauge boson

$$P_i(x) = 4C_i \frac{\alpha}{4\pi} \left( \frac{x}{1-x} + (1-x)g_i(x) \right), \quad \begin{cases} g_\phi(x) = 0 \\ g_\lambda(x) = \frac{1}{2} \\ g_V(x) = x + \frac{1}{x} \end{cases}. \quad (2.1)$$

The function  $g_i(x)$  satisfies the condition that it has a regular Taylor series expansion about  $x = 1$ . Here we see that the splitting probability depends on the parton species through two factors: the Casimir  $C_i$  which describes the color charge of the particle, and the function  $g_i(x)$ .

In the  $x \rightarrow 1$  limit, in which the emitted radiation is soft, and which is enhanced by the soft singularity, the splitting functions exhibit a universal form depending only on the

color factors  $C_i$ . Explicitly, these are

$$C_F = \frac{N^2 - 1}{2N} \rightarrow \frac{4}{3}, \quad C_A = N \rightarrow 3. \quad (2.2)$$

The result in this limit is simply the classical result for radiation in the eikonal limit<sup>2</sup> [42], which is (as it must be) independent of the spin of the emitting parton. In this limit, the information in a jet is encoded solely in the *quantity* (multiplicity) of radiation rather than the *structure* of the radiation (we will prove this rigorously in section 3). Standard studies of quark vs. gluon discrimination have focused solely on this leading eikonal term, resulting in the standard claim that multiplicity or counting observables are optimal (see e.g. ref. [43]).

One of the motivations for exploring in more detail the structure beyond the strict eikonal limit, is that there has recently been work on purely collinear jet substructure observables (the projected energy correlators, see [44–53] for more detailed discussions) that are sensitive to specific integer moments (the twist-2 spin-J anomalous dimensions) of the splitting functions. These observables are not correctly described by the leading eikonal result for the splitting function, and exhibit differences between quarks and gluons generated from the non-eikonal terms in the splitting function. One would therefore like to understand how to exploit this information for quark gluon discrimination.

The non-eikonal term,  $g_i(x)$ , depends on the nature (spin) of the emitting parton. Its contribution to the splitting function is suppressed by two powers of  $(1 - x)$ , which is guaranteed by the Low-Burnett-Kroll (LBK) theorem [54–56] describing the soft limits of gauge theories.<sup>3</sup> As expected,  $g_i(x)$  vanishes for a scalar, which has no structure beyond its color.

Expanding the function  $g_i(x)$  in powers of  $(1 - x)$ , we have

$$g_\phi = 0, \quad g_\lambda = \frac{1}{2}, \quad g_V = \frac{1}{2} + \frac{1}{4}(1 - x)^2 + \dots \quad (2.3)$$

This shows that in QCD, the difference between a quark and a gluon is in fact more suppressed than it needs to be! This makes the difficulty of quark/gluon jet discrimination abundantly clear: the structure of a jet is dominated by its classical result, with the spin information being highly suppressed. Nevertheless, this information is there, and we should be able to design observables to systematically exploit it.

In QCD, there is an additional distinction between quarks and gluons, namely that gluons can split into a  $q\bar{q}$  pair,

$$P_{g \rightarrow q\bar{q}}(z) = \frac{n_f T_R}{2C_A} [z^2 + (1 - z)^2]. \quad (2.4)$$

This splitting does not obey the LBK theorem and is down only by a single power of  $z$ .<sup>4</sup>

### 3 Quark and gluon jet likelihoods

In this section, we introduce an approach which allows us to prove the optimality of quark/gluon discriminants. In this approach, we analyze the likelihood function of jet

<sup>2</sup>This corresponds with the well-known fact that the leading double logarithmic Sudakov is classical.

<sup>3</sup>In this case, the LBK theorem is applied after taking the collinear limit.

<sup>4</sup>See refs. [57–61] for a discussion.

radiation at a given accuracy, and apply the Neyman-Pearson Lemma [27] to construct an optimal discriminant as the likelihood ratio of the two jet classes in question (in our case, quark and gluon jets). Explicitly constructing likelihood ratios has garnered increasing attention in recent years, as it provides some insight into what physical information a deep neural network might be using when learning a (presumed optimal) classification function. Our work is similar in spirit to the approaches in refs. [33, 35], and provides a complementary perspective for calculating jet likelihoods.

Our technique allows us to work in some approximation (e.g. eikonal, independent emission, etc.) and then analyze the likelihood using machine learning or information-theoretic techniques to prove results about the optimal observables. These optimality results will then apply to any observables that can be computed at a given accuracy from this likelihood. Since we will always work in some approximation for the likelihood, one can always come up with observables that can evade the proven results, if they cannot be correctly computed from this approximation. However, we view this as a virtue, as it shows how results can be systematically improved with more accurate approximations.

### 3.1 Analysis in the eikonal limit

We begin by considering the optimal quark/gluon discriminant in the eikonal limit. We define the eikonal limit as independent emissions with the universal part of the splitting functions

$$dP_{i \rightarrow ig}(z, \theta) = \frac{2\alpha_s C_i}{\pi} \frac{dz}{z} \frac{d\theta}{\theta}, \quad (3.1)$$

where  $i \in \{q/g\}$ .

Before proceeding, we would like to clarify a number of issues related to our referring to this as the eikonal, or classical eikonal limit. Another name which could be used is the LL limit. We avoid this language, since it is best defined when there is an observable of interest. While it is of course true that the above will generate the leading logarithms for Sudakov-type observables, it is straightforward to identify IRC safe observables (even those that probe only two particle correlations) for which the leading logarithm will not be produced by the above approximated splitting function.

Since we have shown in section 2 that the splitting function in this limit is identical for all partons up to the color factor, we should be able to prove that multiplicity is the optimal observable. To show this, we consider the quark/gluon likelihood ratio  $L_{q/g}$  for radiating a collection of gluons<sup>5</sup> with kinematics  $\{(z_n, \theta_n)\}_{n=1}^M$

$$\ln L_{q/g}^{\text{LL}} = \sum_{n=1}^M \ln \left( \frac{dP_{q \rightarrow qg}(z_n, \theta_n)}{dP_{g \rightarrow gg}(z_n, \theta_n)} \right) = \ln \frac{C_F}{C_A} \sum_{n=1}^M 1 = M \ln \frac{C_F}{C_A}. \quad (3.2)$$

Here  $M$  is simply the multiplicity, and so by the Neyman-Pearson Lemma<sup>6</sup> [27], the optimal observable at this level of accuracy is the multiplicity. Experimentally the multiplicity is

---

<sup>5</sup>Here we use the well known fact that there is a probabilistic interpretation for the twist two splitting functions. This, however, should not be taken for granted, and indeed such a probabilistic interpretation fails at higher twist [62].

<sup>6</sup>Monotonic functions of the likelihood ratio are also optimal since monotone functions do not change the ordering by the classifier.

indeed found to perform well as a discriminant [63, 64], and we provide justification for why this is the case.

Another perspective is that the optimality of counting follows simply from the universal nature of soft emissions in gauge theory, or in other words, the classical structure of the jet. This behavior is true for a very wide class of observables, including standard Sudakov observables, Sudakov observables that have been soft dropped [65], and groomed multiplicity [43]. We also note that the above result holds regardless of whether the coupling is taken to be running or not, since it cancels out of the ratio in eq. (3.2). We therefore state our first result:

**Theorem 1.** *For an observable whose LL result can be computed using the eikonal splitting functions of eq. (3.1) in the independent emission approximation (with or without running coupling), this LL result can not achieve better quark/gluon discrimination than multiplicity.*

In addition to showing that multiplicity is optimal, we can in fact derive its probability distribution using Poisson distributions for the multiplicity  $M$ <sup>7</sup>

$$p_i^{\text{LL}}(M) = \text{Pois} \left[ \int \frac{dz}{z} \int \frac{d\theta}{\theta} \frac{2\alpha_s C_i}{\pi} \Theta(z, \theta) \right]. \quad (3.3)$$

In the eikonal limit, emissions are uniformly distributed in the Lund plane, and we can tessellate some perturbative region of the plane using triangles (see figure 1). The probability for  $n$  emissions in a given triangle with area  $\Delta$  is then Poisson distributed according to

$$\Pr(n_i = n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad \lambda = \frac{2\alpha_s C_i \Delta}{\pi}. \quad (3.4)$$

Due to the assumption of independent emissions, we then have that the radiation counts throughout the emission plane are distributed according to

$$\Pr_i(n_1, n_2, \dots, n_N) = \prod_{j=1}^N \frac{\lambda_i^{n_j} e^{-\lambda_i}}{n_j!}. \quad (3.5)$$

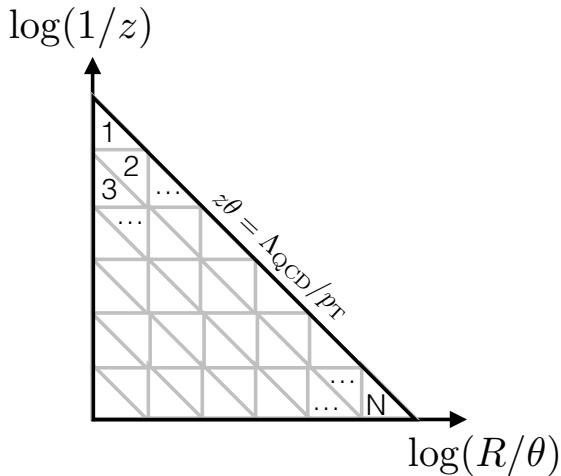
The optimal quark/gluon discriminant is then

$$\ln \frac{\Pr_q(n_1, \dots, n_N)}{\Pr_g(n_1, \dots, n_N)} = \ln \frac{C_F}{C_A} \sum_{j=1}^N n_j + \text{const.} \quad (3.6)$$

We can take the  $\Delta \rightarrow 0$  limit to find again that the optimal quark gluon discriminant is simply the multiplicity. This version of the proof makes clear that we can apply a cutoff and only consider the likelihood function in the perturbative regime.

---

<sup>7</sup>Poisson observables were emphasized in [43].



**Figure 1.** A tiled version of the Lund plane where each of the  $N$  triangles has the same area. The non-perturbative regime is defined by  $z\theta > \Lambda_{\text{QCD}}/p_T$ .

### 3.2 Systematic expansion beyond the eikonal limit

The radiation phase space from the previous section can be systematically improved. Observables computed with modified leading logarithmic accuracy should include the effects of the running coupling as well as subleading terms in the splitting function. In particular, the radiation phase space is given by

$$dP_{i \rightarrow ij}(z, \theta) = \frac{2\alpha_s(z\theta p_T) C_i}{\pi} p_{i \rightarrow ij}(z) dz \frac{d\theta}{\theta}, \quad (3.7)$$

where  $p_{ij}$  are the QCD splitting functions. In general, at this beyond-Eikonal (BE) order, there can be flavor changes between quarks and gluons, since  $\int_{1/2}^1 dz p_{q \rightarrow qg}(z) > 0$  and a finite part of the gluon splitting function contains a contribution from  $g \rightarrow q\bar{q}$ . Ignoring flavor changing,<sup>8</sup> the formalism from the previous section still holds, only now the optimal observable is more complicated than simply multiplicity. In particular:

$$\text{JL}^{\text{BE}'} \equiv \ln L_{q/g}^{\text{BE}'} = \sum_{n=1}^M \ln \frac{dP_{q \rightarrow qg}(z_n, \theta_n)}{dP_{g \rightarrow gg}(z_n, \theta_n)} = \sum_{n=1}^M \left[ \ln \frac{C_F}{C_A} + \ln \frac{p_{q \rightarrow qg}(z_n)}{p_{g \rightarrow gg}(z_n)} \right], \quad (3.8)$$

where BE' denotes the beyond Eikonal approximation for the radiation phase space, but ignoring flavor changing. By inserting the corresponding the following splitting functions<sup>9</sup>

$$p_{q \rightarrow qg}(z) = \frac{1 + (1 - z)^2}{2z}, \quad (3.9)$$

$$p_{g \rightarrow gg}(z) = \frac{1 - z}{z} + \frac{1}{2}z(1 - z), \quad (3.10)$$

the optimal observable is

$$\text{JL}^{\text{BE}'} = \sum_{n=1}^M \left[ \ln \frac{C_F}{C_A} + \ln \frac{1 + (1 - z_n)^2}{(1 - z_n)(2 + z_n^2)} \right] \approx M \ln \frac{C_F}{C_A} + \frac{1}{2} \sum_{n=1}^M z_n^3 + \mathcal{O}(z_n^4). \quad (3.11)$$

<sup>8</sup>See e.g. ref. [66] for a careful treatment of flavor changing effects.

<sup>9</sup>The  $p_{q \rightarrow qg}$  still includes the flavor change, but this fact is ignored in the BE' approximation.

Equation (3.11) shows that the multiplicity is still nearly optimal when including the non-flavor changing corrections to the radiation phase space. A full BE calculation requires a proper accounting of flavor changing, which is omitted here. When summed over final states, the quark and gluon splitting functions become

$$P_q(z) = \frac{1 + (1 - z)^2}{2z}, \quad (3.12)$$

$$P_g(z) = \frac{1 - z}{z} + \frac{1}{2}z(1 - z) + \frac{n_f T_R}{2C_A}[z^2 + (1 - z)^2]. \quad (3.13)$$

where the  $z \leftrightarrow 1 - z$  symmetry is used put the gluon singularity entirely at  $z \rightarrow 0$ . The splitting functions influence the optimal observable via the log likelihood ratio  $\ln P_q(z)/P_g(z)$ , giving

$$\text{JL}^{\text{MLL}} = \sum_{n=1}^M \left[ \ln \frac{C_F}{C_A} + \ln \frac{1 + (1 - z_n)^2}{(1 - z_n)(2 + z_n^2) + \frac{n_f T_R}{C_A} z_n(z_n^2 + (1 - z_n)^2)} \right]. \quad (3.14)$$

Expanding in powers of  $z_n$ , we have

$$\text{JL}^{\text{MLL}} = \sum_{n=1}^M \left( \ln \frac{C_F}{C_A} - \frac{n_f T_R}{2C_A} z_n + \frac{n_f T_R (4C_A + n_F T_R)}{8C_A^2} z_n^2 + \dots \right) \quad (3.15)$$

$$= \ln \frac{C_F}{C_A} n^{(\kappa=0)} - \frac{n_f T_R}{2C_A} n^{(\kappa=1)} + \frac{n_f T_R (4C_A + n_F T_R)}{8C_A^2} n^{(\kappa=2)} + \dots, \quad (3.16)$$

where

$$n^{(\kappa)} = \sum_{n=1}^M z_n^\kappa. \quad (3.17)$$

We can therefore extend to our second result:

**Theorem 2.** *For an observable whose LL or MLL result can be computed using splitting functions in the independent emission approximation, this result can not achieve better quark/gluon discrimination than a linear combination of weighted multiplicities.*

### 3.3 Perturbative multiplicity and parton showers

Parton showers are numerical tools that generate multi-parton phase-space points on which observables can be measured. The shower proceeds iteratively, producing its own initial conditions for subsequent evolution. This procedure requires physically sensible intermediate states, i.e. on-shell phase-space points during its ordered evolution. Parton showers recover leading-logarithmic results for observables that are sufficiently similar to their ordering variable, and for which a strong ordering of emissions is guaranteed. The latter aspect and the desire for physically sensible intermediate states can sometimes lead to conflicting needs, especially for observables depending/relying on states containing several partons that cannot be interpreted as “ordered”, and that are sufficiently different from the ordering variable. In such observables, kinematic recoil effects can no longer be neglected, since different schemes to ensure momentum conservation become distinguishable. Perturbative multiplicity is such an observable.

One particularly common scheme to enforce momentum conservation arises from the leading-color approximation of QCD, in which color-dipoles radiate soft gluons coherently. This dipole picture suggests to produce physical  $(n+1)$ -parton states from  $n$ -parton states by distributing the momenta of two originator partons over three new partons, thus allowing momentum-conservation and on-shell conditions throughout. The originator pair is, in most cases, determined by color connections. Quarks and antiquarks will thus contribute to the evolution of a single dipole, while gluons will participate in two dipoles.

In parton showers, quarks and gluons yield different radiation patterns (and thus perturbative multiplicities) because of differing color factors, and because of different phase-space dependence of hard-collinear contributions. Moreover, the effect of kinematic recoil is handled differently for both. This is easily seen by comparing the evolution of  $Z \rightarrow q\bar{q}$  to  $h \rightarrow g_1\bar{g}_2$ . In the former, the original color connection between  $q$  and  $\bar{q}$  is broken after the first gluon emission, while for the latter, only one of the two original connections is severed by emitting the first gluon. If momentum is re-distributed locally within color-connected dipoles (as is commonly the case in parton showers, including the DIRE shower [67] employed below), no subsequent gluon emission from the  $q$  will influence the  $\bar{q}$  momenta, while emissions from  $g_1$  can continue to influence the  $g_2$  momentum, until the second original connection is severed. This simple choice of recoil mechanism has the advantage that it makes the action of the parton shower locally invertible, and allows stopping and restarting the shower at will without consequences. These factors are crucial for matching and merging methods to improve the overall event generator fidelity. However, such a recoil scheme clearly complicates the subleading-color behavior of the parton shower, and can lead to unphysical artifacts in quark-vs-gluon discrimination.

To assess the impact of recoil on perturbative multiplicities, we introduce a simple extension of the local recoil strategy of DIRE, with the aim to obtain an identical recoil handling for quarks and gluons. For this, we introduce a “backbone” dipole for gluon evolution, which is never allowed to split. The backbone dipole is defined by the left-over original color-connection after the first gluon emission in the shower. If the original state consisted of a  $q\bar{q}$  pair, no backbone is present. The emission rate from the backbone dipole is forced to vanish. If a parton contributes both to the backbone and to another dipole evolution (i.e. is a gluon), then the rate of emissions from the parton in the latter dipole is increased to ensure that the overall  $g \rightarrow gg$  rate is recovered. This guarantees that recoil effects in  $Z \rightarrow q\bar{q}$  and  $h \rightarrow g_1\bar{g}_2$  are treated completely identically, by effectively treating the recoil of emissions from gluons connected to the backbone identical to emission from (anti)quarks.

### 3.4 Discussion and extensions

In this section, we introduced a strategy for constructing optimal discriminating observables using quark/gluon jet likelihoods at fixed perturbative accuracy. In the next section, we demonstrate this strategy for quark/gluon jet discrimination in parton-level Monte Carlo events generated at a fixed order. While this technique sheds new light on the quark/gluon jet classification problem in particular, it need not be restricted to this application. It can be applied whenever one can compute class likelihoods for an observable, and performance can be systematically improved to track with advances in theoretical calculations.

A key advantage of our approach is that it can go systematically beyond the leading eikonal limit that has been the focus of most studies of quark-versus-gluon discrimination. We envision that this could be particularly interesting for understanding how information in higher order splitting functions can be exploited for quark gluon discrimination.

However, we must note that our approach relies on having an understanding of the underlying process describing the formation of quark and gluon jets. While this is true in perturbation theory, it is not true for the hadronization process, for which there exists little analytic understanding. Therefore while we find it exciting that we are able to systematically derive optimal observables from a given set of splitting functions, we must be cautious, particularly for subleading information, that it may be washed out or dominated by hadronization effects. We will study this in section 4.4.

## 4 Monte Carlo studies

In the previous section, we introduced a systematic method to compute likelihood ratio observables (LROs) corresponding to (approximately) optimal quark/gluon jet discriminants at a given perturbative accuracy. In this section, we will study their classification performance using parton-level Monte Carlo simulations of quark and gluon jets. We alter the functional form of  $q \rightarrow qg$  emissions in a leading log (LL) shower, producing a simple class of LROs  $\ln(P_q/P_g)$ , and compare their classification performance to several benchmarks: jet multiplicity and family of deep neural networks (DNNs). When sufficiently trained, DNNs are often presumed to classify optimally, and thus provide an approximate performance ceiling for assessing LROs.

### 4.1 Configurable parton showers

We generate parton-level Monte Carlo events using PYTHIA 8.303 [68] with the DIRE parton shower [67]. DIRE features a number of settings (“kernel orders”) for the perturbative order of the shower, the simplest of which corresponds to a LL shower without  $g \rightarrow q\bar{q}$  splittings (`KernelOrder = -1`). Accounting for  $g \rightarrow q\bar{q}$  splittings significantly complicates calculating the quark/gluon jet likelihoods, so we use `KernelOrder = -1` for all studies presented here. Pure samples of quark and gluon jets are extracted from  $e^+e^- \rightarrow H \rightarrow q\bar{q}/gg$  events, respectively, generated at  $\sqrt{s} = m_H = 250$  GeV and with all initial-state radiation turned off. In each event, two sets of jets are clustered using the anti- $k_T$  algorithm with radii  $R = 0.4$  and  $R = 1.0$  [69]. To minimize the chance of including secondary jets from wide-angle emissions, we only analyze the leading jet from each event.

In addition to fixing the shower order, DIRE allows us to tweak the functional form of the splitting functions  $P_{q \rightarrow qg}$  and  $P_{g \rightarrow gg}$ . In our study, we consider splittings of the form

$$\begin{aligned} P_{g \rightarrow gg} &\propto C_A \times f_{\text{univ}}(z) \\ P_{q \rightarrow qg} &\propto C_F \times c_0 \exp(c_1 z + c_2 z^2) \times f_{\text{univ}}(z), \end{aligned} \tag{4.1}$$

where  $c_0$ ,  $c_1$ , and  $c_2$  are tunable parameters and  $f_{\text{univ}}(z)$  is the universal part of the splitting function. In this case, the log likelihood ratio takes on a particularly simple form:

$$\ln \left( \frac{P_{q \rightarrow qg}}{P_{g \rightarrow gg}} \right) = \ln \left( c_0 \frac{C_F}{C_A} \right) + c_1 z + c_2 z^2. \quad (4.2)$$

On the level of jets, this defines a likelihood ratio observable

$$\begin{aligned} \ln \mathcal{L} &= \sum_{i \in \text{jet}} \ln \left( c_0 \frac{C_F}{C_A} \right) + c_1 z_i + c_2 z_i^2 \\ &= \ln \left( c_0 \frac{C_F}{C_A} \right) n^{(0)} + c_1 n^{(1)} + c_2 n^{(2)}, \end{aligned} \quad (4.3)$$

where  $n^{(i)}$  is defined as in eq. (3.17) with  $z_i = p_{T,i}/p_{T,\text{jet}}$ . In our studies, we vary the three parameters above, as well as the color factor  $C_F$ , and track how the LRO performs as a quark/gluon jet classifier.

## 4.2 Optimality & particle flow networks

The family of LROs defined in section 4.1 are only *approximately* optimal, as the calculation makes the simplifying assumptions that (a) all emissions  $z_i$  come from the initiating hard parton, and (b) the momentum fractions  $z_{i,\text{jet}}$  relative to the jet  $p_T$  are sufficient proxies for the fractions  $z_i$  relative to the emitting partons. In this study we compare LROs to Particle Flow Networks (PFNs), which use a per-particle latent space embedding to learn arbitrary permutation-invariant jet substructure observables [70–72]. We use the constituent kinematic features  $(z_i, \eta_i, \phi_i)$  as the per-particle inputs,<sup>10</sup> where momentum fractions are defined as  $z_i = p_{T,i}/p_{T,\text{jet}}$  and  $(\eta_i, \phi_i)$  are measured relative to the jet axis.<sup>11</sup> The particle embedding map  $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^\ell$  and binary classification function  $F : \mathbb{R}^\ell \rightarrow \mathbb{R}^2$  are trained simultaneously as fully connected neural networks. Jets are represented in the latent space by summing their constituent particle representations, and the sum is classified by the function  $F$ .

In addition to the PFN we train an Energy Flow Network (EFN), which learns an explicitly infrared and collinear (IRC) safe observable of the form  $F(\sum_{i \in \text{jet}} z_i \Phi(\eta_i, \phi_i))$ . The EFN provides complementary information to the PFN, indicating whether IRC-safety restricts access to useful classification information (multiplicity, for instance, is IRC unsafe and thus could not be directly learned by an EFN). We also train two variants of a PFN — PFN[ $z$ ] and PFN[ $\eta, \phi$ ] — which use only a subset of the per-particle information, indicated by the variables in brackets. This split elucidates how much useful information comes from the angular structure of the jet versus solely the momenta. The LROs depend only on  $z_i$ , so comparing their performance against PFN[ $z$ ] will indicate if they are optimal given the restricted set of substructure information.

We implement all networks with the tools provided in the `EnergyFlow` package [71, 72], based on `TensorFlow` [73]/`Keras` [74]/`Adam` [75], and use architectures that have been shown

<sup>10</sup>We use pseudo-rapidity  $\eta$  rather than rapidity, as it is a standard observable in collider experiments.

<sup>11</sup>The jet axis is defined as  $(\hat{\eta}, \hat{\phi}) = \sum_{i \in \text{jet}} p_{T,i}(\eta_i, \phi_i)/p_{T,\text{jet}}$ .

to perform well for quark/gluon jet classification [71]. In particular, the particle embedding  $\Phi$  is trained as a fully connected network with layer widths of  $(100, 100, 128)$  (where  $\ell = 128$  is the latent space dimension). The output mapping has hidden dimensions  $(100, 100, 100)$  and an output dimension 2, with the Euclidean unit vectors  $\hat{e}_1, \hat{e}_2$  acting as quark/gluon jet truth labels. To avoid overfitting, a uniform dropout rate of 20% is applied to all nodes in  $F$ . To train each network, a sample of 1 million quark jets and 1 million gluon jets is combined, shuffled, and split into training (70%), validation (15%), and testing (15%) sets. The networks are trained and validated on the corresponding sets over three epochs,<sup>12</sup> and the receiver operating characteristic (ROC) curve and area under the curve (AUC) metrics are evaluated on the test set. We use these metrics to compare network performance with the predicted likelihood ratio observable.

### 4.3 Results: Pythia & DIRE

The central focus of our studies is varying the parameters  $c_1$  and  $c_2$  of eq. (4.3). We retain the default setting  $c_0 = 1$ , as any variation of  $c_0$  corresponds to a term proportional to multiplicity in the likelihood ratio. We instead vary  $C_F$  between its default value  $C_F = 4/3$  and  $C_F = C_A = 3$ . Setting  $C_F = 3$  removes the multiplicity term from eq. (4.3), leaving it sensitive solely to the higher order terms  $n^{(1)}$  and  $n^{(2)}$ . Using  $C_F = 4/3$  reintroduces the  $n^{(0)}$  term, making the other two sub-leading effects. Furthermore, we only consider  $(c_0, c_1) \in \{0, 1\}^2$ , since these settings are sufficient to add or remove the corresponding terms in eq. (4.3).

#### 4.3.1 Parton shower validation

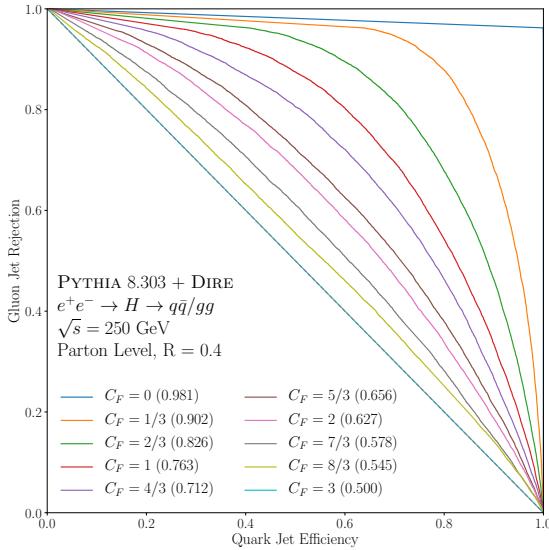
We begin by validating the expected behavior of the DIRE parton shower and PFN training procedure. Since we are working at `KernelOrder = -1`, we expect quark jets and gluon jets to look identical when  $C_F = C_A$ . In figure 2, we show ROC curves and AUC scores for PFN quark/gluon jet classifiers trained on quark jets generated with a range of  $C_F$  values between  $C_F = 0$  to  $C_F = 3$ .<sup>13</sup> As expected, the two are indistinguishable when  $C_F = 3$ . This confirms two crucial details for our later studies: first, that the  $q \rightarrow qg$  and  $g \rightarrow gg$  splitting functions can be made identical by tuning  $C_F$ , and second, that our PFN architecture and training protocol is not overfitting the limited amount of information available in the parton-level shower.

#### 4.3.2 Exponential modifications of $q \rightarrow qg$

Figure 3 shows classifier performance with an  $\exp(z^2)$  enhancement to the  $q \rightarrow qg$  splitting and  $C_F = C_A$ . The LRO in this case is simply  $n^{(2)}$ , and it performs as well as multiplicity in both the  $R = 0.4$  (left) and  $R = 1.0$  (right) cases. There is no *a priori* reason for  $n^{(2)}$  to be useful for classification, so these results demonstrate that the LRO does indeed capture useful classification information. Multiplicity, though not explicitly present in  $\ln \mathcal{L}$ , retains good performance because an enhancement of  $q \rightarrow qg$  splittings is fundamentally

<sup>12</sup>Due to the large number of events ( $\sim 1.4M$ ) in the training set, model training converged very quickly.

<sup>13</sup>The same set of gluon jets generated with  $C_A = 3$  is used in each training.



**Figure 2.** ROC curves for PFN quark/gluon jet classification over a range of quark color factors  $C_F$ . AUC scores are shown next to each legend entry. At each  $C_F$  setting,  $q \rightarrow qg$  and  $g \rightarrow gg$  splitting functions are configured to be identical up to the color factor. As expected, quark and gluon jets are indistinguishable at  $C_F = C_A = 3$  and nearly perfectly distinguishable at  $C_F = 0$ .

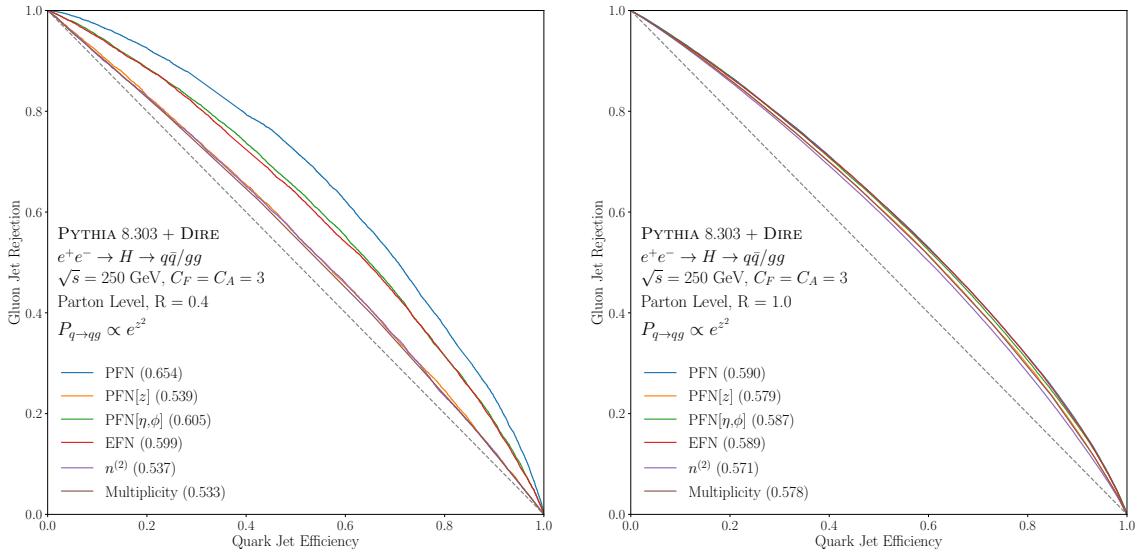
an enhancement of quark jet multiplicity. In both the  $R = 0.4$  and  $R = 1.0$  cases, multiplicity and  $n^{(2)}$  match  $\text{PFN}[z]$  — the PFN trained on constituent  $z_i$ . This indicates that both multiplicity and  $n^{(2)}$  capture virtually all of the useful information encoded in the constituent momenta.

At  $R = 0.4$ , we observe a clear performance gap between the computed observables ( $n^{(2)}$ , multiplicity) and the DNNs that use angular information (PFN,  $\text{PFN}[\eta, \phi]$ , and EFN). This largely disappears at  $R = 1.0$ , where all classifiers perform within 1-2% of one another. We speculate that the difference comes from events where the full radiation pattern is not captured within an  $R = 0.4$  jet (due to a wide-angle emission, for example). In this case, the computed observables suffer due to incomplete information, while the DNNs exploit potentially spurious angular information. Alternatively, “behind-the-scenes” behavior in PYTHIA — such as momentum conservation — may introduce angular correlations that are unaccounted for in the computed observables. The absence of a performance gap at  $R = 1.0$  slightly favors the former explanation, and for the remainder of our studies we use  $R = 1.0$ .

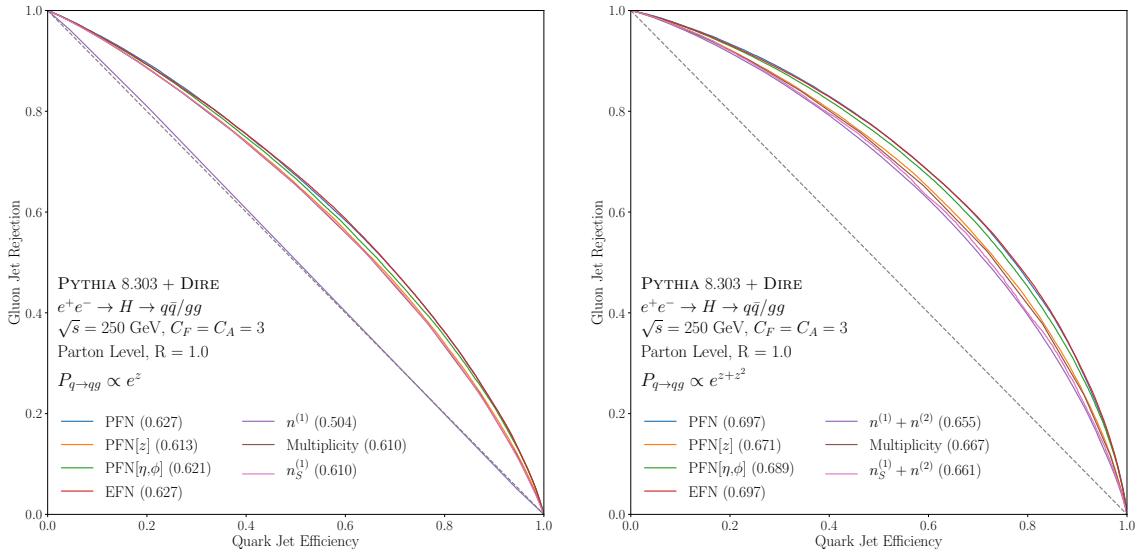
Next, we consider the  $\exp(z)$  enhancement, which corresponds to the LRO  $n^{(1)}$ . Unfortunately,  $n^{(1)}$  is a nearly the constant unity for all jets and thus useless for discrimination, as shown by the purple curve in figure 4 (left). To circumvent this, we expand the sum  $\sum_{i \in \text{jet}} z_i^\kappa$  about  $\kappa = 1$ , as in ref. [76]. To leading order, this yields:

$$\sum_{i \in \text{jet}} z_i^\kappa = 1 + \sum_{i \in \text{jet}} (1 - \kappa) z_i \ln z_i + \dots, \quad (4.4)$$

which justifies replacing  $n^{(1)}$  with the observable  $\sum_{i \in \text{jet}} z_i \ln z_i$ . Due to its similarity to Shannon entropy, we denote this observable  $n_S^{(1)}$ . Figure 4 (left) includes the ROC and



**Figure 3.** ROC curves and AUC scores (listed in the legend) for the likelihood ratio observable  $n^{(2)}$ , multiplicity, and the four benchmark DNNs with settings  $C_F = C_A = 3$  and  $P_{q \rightarrow qg} \propto \exp(z^2)$ . In the  $R = 0.4$  case (left), classifiers that use angular information appear to outperform those using only momenta. This effect mostly vanishes at  $R = 1.0$ , where all perform approximately equally. In both cases,  $n^{(2)}$  performs as well as multiplicity.



**Figure 4.** ROC curves and AUCs for the settings  $P_{q \rightarrow qg} \propto \exp(z)$  (left) and  $P_{q \rightarrow qg} \propto \exp(z + z^2)$  (right). In each case the likelihood ratio contains the trivial term  $n^{(1)}$ , and ROCs using alternative formulation  $n_S^{(1)}$  of eq. (4.4) are shown in an additional curve.

AUC for  $n_S^{(1)}$ , which clearly improves upon  $n^{(1)}$  and matches the performance of multiplicity and the DNNs, as in the  $\exp(z^2)$  case. This again indicates that the LRO captures nearly the same information as the DNNs, while having the clear advantage of interpretability, algebraic simplicity, and ease of computation.

Concluding our studies with  $C_F = C_A = 3$ , we show results from with  $P_{q \rightarrow qg} \propto \exp(z + z^2)$  in figure 4 (right). We again include a ROC curve for the corrected LRO  $n_S^{(1)} + n^{(2)}$ , which boosts performance relative to variant using  $n^{(1)}$ . In this case, however, there is less parity between the computed observables and the DNNs. The PFN, PFN $[\eta, \phi]$ , and EFN perform approximately 2% better than multiplicity and  $n_S^{(1)} + n^{(2)}$ , implying that the angular structure is helpful for classification (though this may be an artifact of the event generator). The LRO is only 1% behind PFN $[z]$ , indicating that it is still a powerful classifier using only the constituent momenta.

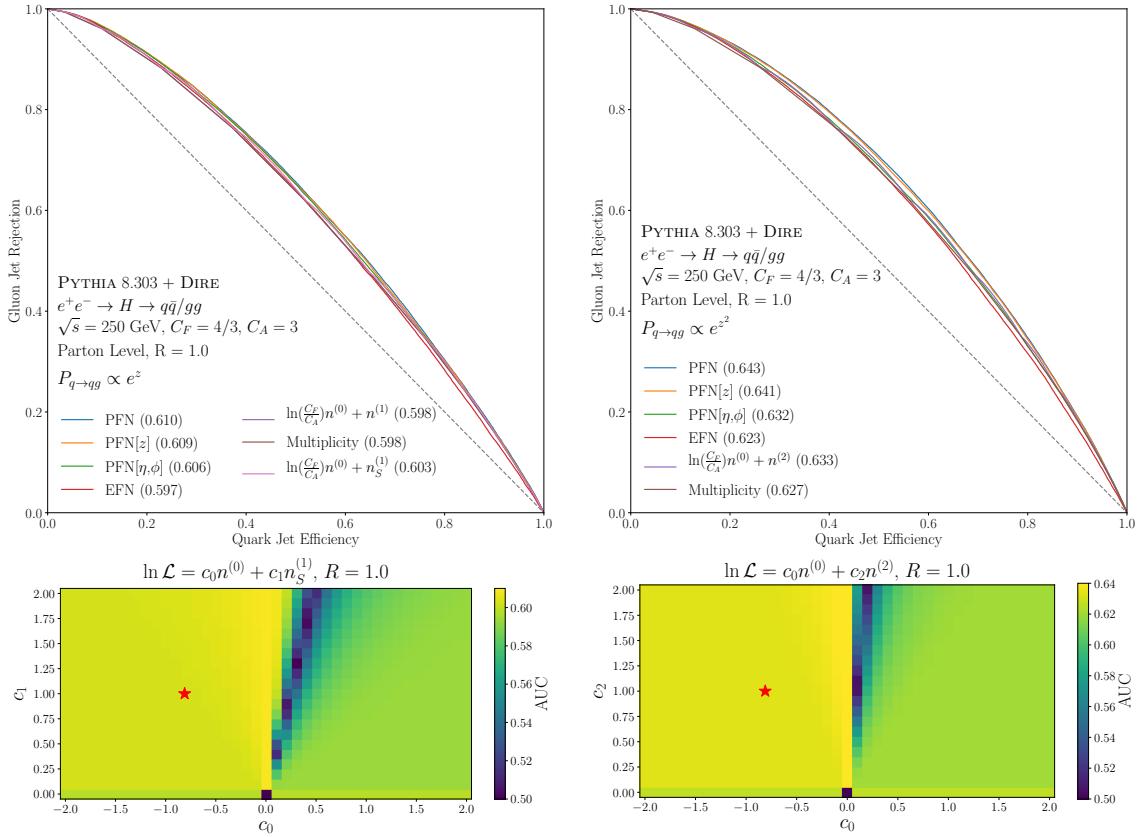
#### 4.3.3 Restoring $C_F = 4/3$

To conclude our studies in PYTHIA, we restore the physical quark color factor  $C_F = 4/3$ , which introduces an explicit multiplicity term  $\ln(C_F/C_A)n^{(0)}$  in the LRO (taking  $c_0 = 1$  as mentioned above). Figure 5 (top), shows results for  $P_{q \rightarrow qg} \propto \exp(z)$  (left) and  $P_{q \rightarrow qg} \propto \exp(z^2)$  (right). The LRO remains very close to the other benchmarks, again indicating that it performs nearly optimally. Notably, the EFN performs worse than the other classifiers, reflecting its insensitivity to jet multiplicity (the leading order effect in the LRO).

Since the LRO explicitly contains multiplicity, we are better poised to understand the relative impact of the higher order terms  $n_S^{(1)}$  and  $n^{(2)}$ . In figure 5 (bottom), we scan over the coefficients  $(c_0, c_1)$  for the observable  $c_0 n^{(0)} + c_1 n_S^{(1)}$  (left) and  $(c_0, c_2)$  for  $c_0 n^{(0)} + c_2 n^{(2)}$  (right).<sup>14</sup> We consider points  $c_0 \in [-2, 2]$  and  $c_1, c_2 \in [0, 2]$ , where negative values of  $c_{i>0}$  are omitted because the observable  $(c_0, -|c_i|)$  is equivalent to  $(-c_0, |c_i|)$  up to an overall sign. At each point, we compute an AUC score for the corresponding LRO and color the points accordingly. The red stars correspond to the default working points  $c_0 = \ln(C_F/C_A)$  and  $c_1 = c_2 = 1$ . The predominant effect appears to be the relative signs of  $c_0$  and  $c_{i>0}$ , with points in the left half (same relative sign as the predicted observable) performing about 1% better than those on the right half (opposite relative sign). This difference is small, but the performance is very consistent within each half.

The other notable feature is the behavior near the lines  $c_0 = 0$  (middle vertical line) and  $c_{i>0} = 0$  (bottom horizontal line). The points with  $c_{i>0} = 0$  — which lack a contribution from the higher order terms  $n_S^{(1)}$  or  $n^{(2)}$  — slightly underperform relative to the points in the upper left half, though the difference is only around 0.5% in each case. Interestingly, the points that contain *only* higher order terms ( $c_0 = 0$ ) perform the best overall. The differences are again quite small (0.5-1%) relative to the default point, but slightly more significant (1-1.5%) relative to the  $c_{i>0} = 0$  points. It is curious that the higher order terms alone perform so well, but this is not entirely unexpected as they are somewhat correlated with jet multiplicity. For example, large-multiplicity jets will have many particles with smaller  $z_i$ , which pushes down the average value of  $n^{(2)}$  (a similar trend also holds for  $n_S^{(1)}$ ). Lastly,

<sup>14</sup>For simplicity, have converted the term  $\ln(c_0 C_F/C_A)n^{(0)}$  into a single prefactor  $c_0 n^{(0)}$ .



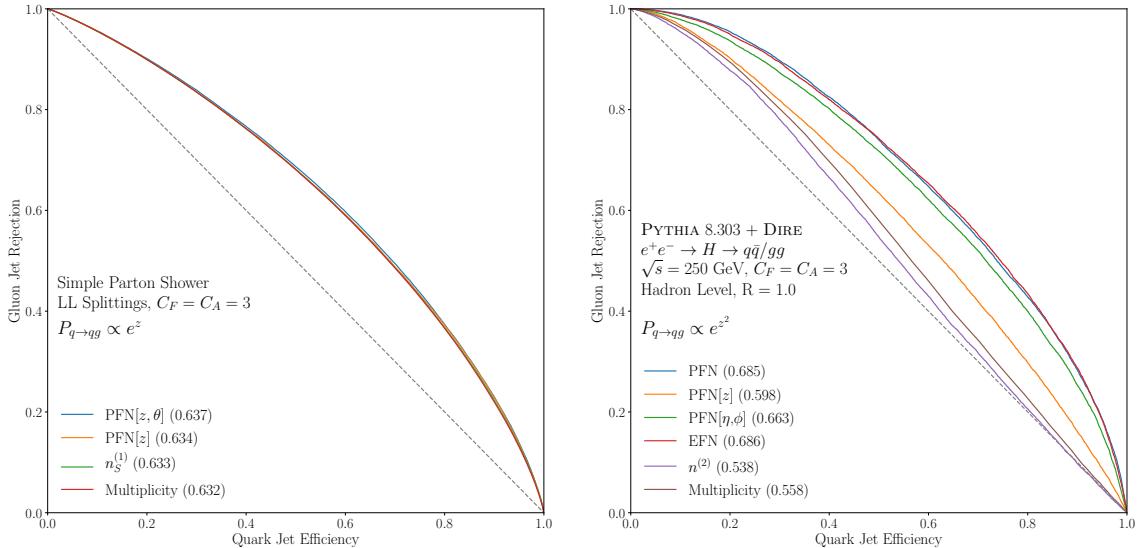
**Figure 5.** Top: ROC curves and AUC scores for  $P_{q \rightarrow gg} \propto \exp(z)$  (left) and  $P_{q \rightarrow gg} \propto \exp(z^2)$  (right), generated with the true quark color factor  $C_F = 4/3$ . Bottom: AUC scores for a scan over the coefficients  $c_0$  and  $c_1/c_2$  in the observables  $c_0 n^{(0)} + c_1 n_S^{(1)}$  (left) and  $c_0 n^{(0)} + c_1 n^{(2)}$  (right). Color indicates the AUC score, and the grid point represents the  $(c_0, c_i)$  setting. In each plot, the red star represents the default working point  $c_0 = \ln(C_F/C_A)$ ,  $c_i = 1$  defined by the computed likelihood ratio.

each plot features a line of poorly performing points emanating from the origin into the right half. These correspond to a family of linearly related observables  $(rc'_0, rc'_{i>0})$  ( $r \in \mathbb{R}$ ) for which quark and gluon jets are identically distributed. On either side of this line (i.e. the left and right halves of the grid), the distributions differ and enable better discrimination power.

As in section 4.3.2, these trends suggest that the LROs are a good approximation of the true likelihood ratio. The higher order terms clearly encode useful classification information beyond jet multiplicity, and the importance of the relative sign between  $n^{(0)}$  and  $n_S^{(1)}$  or  $n^{(2)}$  reflects the algebraic form of the LROs. Furthermore, the LROs perform within 1% of the best performing DNN classifiers, indicating that they have captured nearly all of the useful classification information in a simple, computationally tractable expression.

#### 4.4 Further investigation: toy showers and hadronization

As a final test of the LROs, we examine their performance in classifying jets generated at two opposite extremes: precise adherence to user-specified splitting functions in a simplified parton shower generator, and PYTHIA/DIRE at hadron-level. In the former case, we



**Figure 6.** ROC curves comparing predicted classifier performance to neural networks for two additional event generation paradigms: a dedicated jet generator with explicitly tunable splitting functions (left), and PYTHIA at hadron-level (right). We observe near-perfect agreement between the DNNs and the likelihood ratio observable in the former case. In the latter case, generator-level hadronization effects severely diminish the performance of the calculated observables.

expect strong agreement between LROs and DNNs. The simplified generator<sup>15</sup> computes splitting probabilities directly from hand-coded likelihood functions  $P_{q \rightarrow gg}$  and  $P_{g \rightarrow gg}$ , making the correspondence to eq. (4.1) explicit and the likelihood ratio formulated in eq. (4.3) nearly exact. In the latter case, the PYTHIA hadronization routine significantly complicates the correspondence between a parton shower and a final-state jet, and the likelihood ratio formulae we derive no longer apply. We thus expect the LRO performance to drop substantially.

Figure 6 shows the results in each scenario. In the simplified scenario (left), there is virtually no difference in performance (less than 0.5%) between the LRO and the DNNs. The parton shower routine used for these results is essentially a “jet generator”, and without any of the additional physics features of PYTHIA/DIRE it allows an even more direct probe of the LROs than the parton-level studies in the previous section. While this result is unsurprising, it is strong evidence that the LROs — where they apply nearly exactly — do in fact capture the same information as a DNN. In the case with full hadronization (right), the LRO loses nearly all of its discrimination power, performing worse than multiplicity and all of the DNNs. The performance drop is expected, but the magnitude is exceptionally large and underscores the limitations of working at parton-level when constructing the LROs.

#### 4.5 Discussion

In each result shown in sections 4.3.2 and 4.3.3, our predicted LROs track closely with jet multiplicity and PFN[z], and in most cases perform within 1 to 2% of the full PFN.

<sup>15</sup>We thank Duff Neill for sharing his parton shower code.

Assuming that  $\text{PFN}[z]$  learns the optimal classifier given the constituent momenta, this implies that the LROs are near optimal and roughly equivalent to the true likelihood ratio. The overall parity with multiplicity suggests that while LROs may not confer an explicit *advantage*, they capture the same information and can be systematically computed to one’s desired accuracy. In the  $C_F \neq C_A$  case — where multiplicity is explicitly included in the LRO — the higher order terms do appear to confer a small advantage. In the hyper-simplified case addressed in section 4.4, we see very clear evidence that the LRO — when it reflects the true “physics” in the event generator — does classify optimally. Broadly speaking, it is encouraging that the LROs perform so well relative to multiplicity and DNNs on PYTHIA jets, given that they are (a) calculated using several simplifying assumptions, and (b) insensitive to any angular substructure created by the generator. These results demonstrate the value of an approximate likelihood ratio when the true ratio is intractable, and motivate further study of physics-motivated observables that can approach DNN performance in more realistic scenarios.

## 5 Conclusions

In this paper we have presented a new approach to understanding jet substructure observables, by analyzing their likelihood function at a fixed accuracy and explicitly constructing likelihood ratios. In the case of quark/gluon jet discrimination, we have proven that multiplicity is optimal at leading order, and that going to higher perturbative order introduces sub-leading contributions in the form of weighted multiplicities. Furthermore, we have shown that for observables that can be computed using eikonal splitting functions, their LL result cannot beat multiplicity. Beyond its application in quark/gluon jet classification, our technique reveals a general strategy for computing and systematically improving approximations of the likelihood ratio for a wide range of observables.

We demonstrate our approach using Monte Carlo simulations of parton-level jets, and directly compare the analytic classifier to a collection of deep neural networks. Using jets clustered from fixed-order parton showers, we find that the analytic classifiers nearly match the performance of Particle Flow Networks. This indicates that the networks indeed “discover” the underlying structure of the shower and extract relevant classification information. In some cases, the networks retain a clear advantage which may be attributed to artifacts of the jet clustering scheme, quirks of the event generator, or constraints that are unaccounted for in our calculations (e.g. momentum conservation).<sup>16</sup> In a toy parton shower where splitting functions are hand-coded and the full shower final state is recorded, the analytic and PFN classifiers perform identically.

In summary, this work demonstrates the possibility of a first-principles physical explanation for the performance of ML classifiers. This signals a new way for theory and ML to work together, wherein ideas from statistics and deep learning (e.g. the optimality of likelihood ratios) motivate theoretical calculations that can at least partially explain the performance of a deep neural network. While it is unlikely that analytic classifiers can completely replace DNNs, especially in the noisy and high-dimensional space of collider data,

---

<sup>16</sup>A detailed analysis that explicitly considers jet clustering can be found in ref. [35].

they can illuminate some corners of the “black box”. Furthermore, they can enable an understanding of which physical features of parton shower simulations are being exploited by classifiers, to ensure that they are well described.

We look forward to continued progress in this vein, and believe that our approach is general enough to apply beyond quark/gluon discrimination. In future work, it would be interesting to modify the technique to become more robust against unconstrained behaviors of Monte Carlo generators, or apply it to other classification problems such as boosted  $W/Z/h$  boson identification.

## Acknowledgments

We are grateful to Eric Metodiev for collaboration during an early stage of this work, including the development of many components of the codebase. The authors are grateful to Gregory Soyez, and Jesse Thaler for helpful discussions and suggestions. We thank Duff Neill for sharing his parton shower code with us, and for useful discussions. The work of B.N. is supported by the DOE Office of Science under contract DE-AC02-05CH11231. Cloud computing resources were provided through a Microsoft Azure for Research award. I.M. is supported by start-up funds from Yale University.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited. SCOAP<sup>3</sup> supports the goals of the International Year of Basic Sciences for Sustainable Development.

## References

- [1] A. Abdesselam et al., *Boosted objects: a probe of beyond the Standard Model physics*, *Eur. Phys. J. C* **71** (2011) 1661 [[arXiv:1012.5412](https://arxiv.org/abs/1012.5412)] [[INSPIRE](#)].
- [2] A. Altheimer et al., *Jet substructure at the Tevatron and LHC: new results, new tools, new benchmarks*, *J. Phys. G* **39** (2012) 063001 [[arXiv:1201.0008](https://arxiv.org/abs/1201.0008)] [[INSPIRE](#)].
- [3] A. Altheimer et al., *Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23<sup>rd</sup>–27<sup>th</sup> of July 2012*, *Eur. Phys. J. C* **74** (2014) 2792 [[arXiv:1311.2708](https://arxiv.org/abs/1311.2708)] [[INSPIRE](#)].
- [4] D. Adams et al., *Towards an understanding of the correlations in jet substructure*, *Eur. Phys. J. C* **75** (2015) 409 [[arXiv:1504.00679](https://arxiv.org/abs/1504.00679)] [[INSPIRE](#)].
- [5] A.J. Larkoski, I. Moult and B. Nachman, *Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning*, *Phys. Rept.* **841** (2020) 1 [[arXiv:1709.04464](https://arxiv.org/abs/1709.04464)] [[INSPIRE](#)].
- [6] R. Kogler et al., *Jet substructure at the Large Hadron Collider: experimental review*, *Rev. Mod. Phys.* **91** (2019) 045003 [[arXiv:1803.06991](https://arxiv.org/abs/1803.06991)] [[INSPIRE](#)].
- [7] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, *Springer* (2019).
- [8] A.J. Larkoski and J. Thaler, *Unsafe but calculable: ratios of angularities in perturbative QCD*, *JHEP* **09** (2013) 137 [[arXiv:1307.1699](https://arxiv.org/abs/1307.1699)] [[INSPIRE](#)].

[9] A.J. Larkoski, S. Marzani and J. Thaler, *Sudakov safety in perturbative QCD*, *Phys. Rev. D* **91** (2015) 111501 [[arXiv:1502.01719](https://arxiv.org/abs/1502.01719)] [[INSPIRE](#)].

[10] A.J. Larkoski, I. Moult and D. Neill, *Power counting to better jet observables*, *JHEP* **12** (2014) 009 [[arXiv:1409.6298](https://arxiv.org/abs/1409.6298)] [[INSPIRE](#)].

[11] A.J. Larkoski, I. Moult and D. Neill, *Building a better boosted top tagger*, *Phys. Rev. D* **91** (2015) 034035 [[arXiv:1411.0665](https://arxiv.org/abs/1411.0665)] [[INSPIRE](#)].

[12] A.J. Larkoski, I. Moult and D. Neill, *Analytic boosted boson discrimination*, *JHEP* **05** (2016) 117 [[arXiv:1507.03018](https://arxiv.org/abs/1507.03018)] [[INSPIRE](#)].

[13] I. Moult, L. Necib and J. Thaler, *New angles on energy correlation functions*, *JHEP* **12** (2016) 153 [[arXiv:1609.07483](https://arxiv.org/abs/1609.07483)] [[INSPIRE](#)].

[14] A.J. Larkoski, I. Moult and D. Neill, *Analytic boosted boson discrimination at the Large Hadron Collider*, [arXiv:1708.06760](https://arxiv.org/abs/1708.06760) [[INSPIRE](#)].

[15] A.J. Larkoski, I. Moult and D. Neill, *Factorization and resummation for groomed multi-prong jet shapes*, *JHEP* **02** (2018) 144 [[arXiv:1710.00014](https://arxiv.org/abs/1710.00014)] [[INSPIRE](#)].

[16] P. Cal, J. Thaler and W.J. Waalewijn, *Power counting energy flow polynomials*, *JHEP* **09** (2022) 021 [[arXiv:2205.06818](https://arxiv.org/abs/2205.06818)] [[INSPIRE](#)].

[17] L. Gellersen, S. Höche and S. Prestel, *Disentangling soft and collinear effects in QCD parton showers*, *Phys. Rev. D* **105** (2022) 114012 [[arXiv:2110.05964](https://arxiv.org/abs/2110.05964)] [[INSPIRE](#)].

[18] F. Dulat, S. Höche and S. Prestel, *Leading-color fully differential two-loop soft corrections to QCD dipole showers*, *Phys. Rev. D* **98** (2018) 074013 [[arXiv:1805.03757](https://arxiv.org/abs/1805.03757)] [[INSPIRE](#)].

[19] S. Höche, F. Krauss and S. Prestel, *Implementing NLO DGLAP evolution in parton showers*, *JHEP* **10** (2017) 093 [[arXiv:1705.00982](https://arxiv.org/abs/1705.00982)] [[INSPIRE](#)].

[20] H.T. Li and P. Skands, *A framework for second-order parton showers*, *Phys. Lett. B* **771** (2017) 59 [[arXiv:1611.00013](https://arxiv.org/abs/1611.00013)] [[INSPIRE](#)].

[21] S. Höche and S. Prestel, *Triple collinear emissions in parton showers*, *Phys. Rev. D* **96** (2017) 074017 [[arXiv:1705.00742](https://arxiv.org/abs/1705.00742)] [[INSPIRE](#)].

[22] M. van Beekveld, S. Ferrario Ravasio, G.P. Salam, A. Soto-Ontoso, G. Soyez and R. Verheyen, *PanScales parton showers for hadron collisions: formulation and fixed-order studies*, *JHEP* **11** (2022) 019 [[arXiv:2205.02237](https://arxiv.org/abs/2205.02237)] [[INSPIRE](#)].

[23] K. Hamilton, A. Karlberg, G.P. Salam, L. Scyboz and R. Verheyen, *Soft spin correlations in final-state parton showers*, *JHEP* **03** (2022) 193 [[arXiv:2111.01161](https://arxiv.org/abs/2111.01161)] [[INSPIRE](#)].

[24] A. Karlberg, G.P. Salam, L. Scyboz and R. Verheyen, *Spin correlations in final-state parton showers and jet observables*, *Eur. Phys. J. C* **81** (2021) 681 [[arXiv:2103.16526](https://arxiv.org/abs/2103.16526)] [[INSPIRE](#)].

[25] K. Hamilton, R. Medves, G.P. Salam, L. Scyboz and G. Soyez, *Colour and logarithmic accuracy in final-state parton showers*, *JHEP* **03** (2021) 041 [[arXiv:2011.10054](https://arxiv.org/abs/2011.10054)] [[INSPIRE](#)].

[26] M. Dasgupta, F.A. Dreyer, K. Hamilton, P.F. Monni, G.P. Salam and G. Soyez, *Parton showers beyond leading logarithmic accuracy*, *Phys. Rev. Lett.* **125** (2020) 052002 [[arXiv:2002.11114](https://arxiv.org/abs/2002.11114)] [[INSPIRE](#)].

[27] J. Neyman and E.S. Pearson, *IX. On the problem of the most efficient tests of statistical hypotheses*, *Phil. Trans. Roy. Soc. Lond. A* **231** (1933) 289.

- [28] D0 collaboration, *A precision measurement of the mass of the top quark*, *Nature* **429** (2004) 638 [[hep-ex/0406031](#)] [[INSPIRE](#)].
- [29] D.E. Soper and M. Spannowsky, *Finding physics signals with shower deconstruction*, *Phys. Rev. D* **84** (2011) 074002 [[arXiv:1102.3480](#)] [[INSPIRE](#)].
- [30] D.E. Soper and M. Spannowsky, *Finding top quarks with shower deconstruction*, *Phys. Rev. D* **87** (2013) 054012 [[arXiv:1211.3140](#)] [[INSPIRE](#)].
- [31] D. Ferreira de Lima, P. Petrov, D. Soper and M. Spannowsky, *Quark-gluon tagging with shower deconstruction: unearthing dark matter and Higgs couplings*, *Phys. Rev. D* **95** (2017) 034001 [[arXiv:1607.06031](#)] [[INSPIRE](#)].
- [32] A.J. Larkoski and E.M. Metodiev, *A theory of quark vs. gluon discrimination*, *JHEP* **10** (2019) 014 [[arXiv:1906.01639](#)] [[INSPIRE](#)].
- [33] G. Kasieczka, S. Marzani, G. Soyez and G. Stagnitto, *Towards machine learning analytics for jet substructure*, *JHEP* **09** (2020) 195 [[arXiv:2007.04319](#)] [[INSPIRE](#)].
- [34] S. Bieringer et al., *Measuring QCD splittings with invertible networks*, *SciPost Phys.* **10** (2021) 126 [[arXiv:2012.09873](#)] [[INSPIRE](#)].
- [35] F.A. Dreyer, G. Soyez and A. Takacs, *Quarks and gluons in the Lund plane*, *JHEP* **08** (2022) 177 [[arXiv:2112.09140](#)] [[INSPIRE](#)].
- [36] Y.S. Lai, D. Neill, M. Płoskoń and F. Ringer, *Explainable machine learning of the underlying physics of high-energy particle collisions*, *Phys. Lett. B* **829** (2022) 137055 [[arXiv:2012.06582](#)] [[INSPIRE](#)].
- [37] A. Buckley, G. Callea, A.J. Larkoski and S. Marzani, *An optimal observable for color singlet identification*, *SciPost Phys.* **9** (2020) 026 [[arXiv:2006.10480](#)] [[INSPIRE](#)].
- [38] F.A. Dreyer, G.P. Salam and G. Soyez, *The Lund jet plane*, *JHEP* **12** (2018) 064 [[arXiv:1807.04758](#)] [[INSPIRE](#)].
- [39] Y.L. Dokshitzer, G. Marchesini and G.P. Salam, *Revisiting parton evolution and the large- $x$  limit*, *Phys. Lett. B* **634** (2006) 504 [[hep-ph/0511302](#)] [[INSPIRE](#)].
- [40] Y.L. Dokshitzer and G. Marchesini,  *$N = 4$  SUSY Yang-Mills: three loops made simple(r)*, *Phys. Lett. B* **646** (2007) 189 [[hep-th/0612248](#)] [[INSPIRE](#)].
- [41] M. Beccaria, Y.L. Dokshitzer and G. Marchesini, *Twist 3 of the  $sl(2)$  sector of  $N = 4$  SYM and reciprocity respecting evolution*, *Phys. Lett. B* **652** (2007) 194 [[arXiv:0705.2639](#)] [[INSPIRE](#)].
- [42] M. Levy and J. Sucher, *Eikonal approximation in quantum field theory*, *Phys. Rev.* **186** (1969) 1656 [[INSPIRE](#)].
- [43] C. Frye, A.J. Larkoski, J. Thaler and K. Zhou, *Casimir meets Poisson: improved quark/gluon discrimination with counting observables*, *JHEP* **09** (2017) 083 [[arXiv:1704.06266](#)] [[INSPIRE](#)].
- [44] L.J. Dixon, I. Moult and H.X. Zhu, *Collinear limit of the energy-energy correlator*, *Phys. Rev. D* **100** (2019) 014009 [[arXiv:1905.01310](#)] [[INSPIRE](#)].
- [45] H. Chen, M.-X. Luo, I. Moult, T.-Z. Yang, X. Zhang and H.X. Zhu, *Three point energy correlators in the collinear limit: symmetries, dualities and analytic results*, *JHEP* **08** (2020) 028 [[arXiv:1912.11050](#)] [[INSPIRE](#)].

[46] H. Chen, I. Moult, X. Zhang and H.X. Zhu, *Rethinking jets with energy correlators: tracks, resummation, and analytic continuation*, *Phys. Rev. D* **102** (2020) 054012 [[arXiv:2004.11381](https://arxiv.org/abs/2004.11381)] [[INSPIRE](#)].

[47] H. Chen, I. Moult and H.X. Zhu, *Quantum interference in jet substructure from spinning gluons*, *Phys. Rev. Lett.* **126** (2021) 112003 [[arXiv:2011.02492](https://arxiv.org/abs/2011.02492)] [[INSPIRE](#)].

[48] H. Chen, I. Moult and H.X. Zhu, *Spinning gluons from the QCD light-ray OPE*, *JHEP* **08** (2022) 233 [[arXiv:2104.00009](https://arxiv.org/abs/2104.00009)] [[INSPIRE](#)].

[49] P.T. Komiske, I. Moult, J. Thaler and H.X. Zhu, *Analyzing  $N$ -point energy correlators inside jets with CMS open data*, [arXiv:2201.07800](https://arxiv.org/abs/2201.07800) [[INSPIRE](#)].

[50] J. Holguin, I. Moult, A. Pathak and M. Procura, *A new paradigm for precision top physics: weighing the top with energy correlators*, [arXiv:2201.08393](https://arxiv.org/abs/2201.08393) [[INSPIRE](#)].

[51] H. Chen, I. Moult, J. Sandor and H.X. Zhu, *Celestial blocks and transverse spin in the three-point energy correlator*, *JHEP* **09** (2022) 199 [[arXiv:2202.04085](https://arxiv.org/abs/2202.04085)] [[INSPIRE](#)].

[52] H. Chen, I. Moult, J. Thaler and H.X. Zhu, *Non-Gaussianities in collider energy flux*, *JHEP* **07** (2022) 146 [[arXiv:2205.02857](https://arxiv.org/abs/2205.02857)] [[INSPIRE](#)].

[53] K. Lee, B. Meçaj and I. Moult, *Conformal colliders meet the LHC*, [arXiv:2205.03414](https://arxiv.org/abs/2205.03414) [[INSPIRE](#)].

[54] F.E. Low, *Bremsstrahlung of very low-energy quanta in elementary particle collisions*, *Phys. Rev.* **110** (1958) 974 [[INSPIRE](#)].

[55] T.H. Burnett and N.M. Kroll, *Extension of the low soft photon theorem*, *Phys. Rev. Lett.* **20** (1968) 86 [[INSPIRE](#)].

[56] V. Del Duca, *High-energy Bremsstrahlung theorems for soft photons*, *Nucl. Phys. B* **345** (1990) 369 [[INSPIRE](#)].

[57] I. Moult, L. Rothen, I.W. Stewart, F.J. Tackmann and H.X. Zhu, *Subleading power corrections for  $N$ -jettiness subtractions*, *Phys. Rev. D* **95** (2017) 074023 [[arXiv:1612.00450](https://arxiv.org/abs/1612.00450)] [[INSPIRE](#)].

[58] R. Boughezal, X. Liu and F. Petriello, *Power corrections in the  $N$ -jettiness subtraction scheme*, *JHEP* **03** (2017) 160 [[arXiv:1612.02911](https://arxiv.org/abs/1612.02911)] [[INSPIRE](#)].

[59] I. Moult, L. Rothen, I.W. Stewart, F.J. Tackmann and H.X. Zhu,  *$N$ -jettiness subtractions for  $gg \rightarrow H$  at subleading power*, *Phys. Rev. D* **97** (2018) 014013 [[arXiv:1710.03227](https://arxiv.org/abs/1710.03227)] [[INSPIRE](#)].

[60] R. Boughezal, A. Isgrò and F. Petriello, *Next-to-leading-logarithmic power corrections for  $N$ -jettiness subtraction in color-singlet production*, *Phys. Rev. D* **97** (2018) 076006 [[arXiv:1802.00456](https://arxiv.org/abs/1802.00456)] [[INSPIRE](#)].

[61] I. Moult, I.W. Stewart, G. Vita and H.X. Zhu, *First subleading power resummation for event shapes*, *JHEP* **08** (2018) 013 [[arXiv:1804.04665](https://arxiv.org/abs/1804.04665)] [[INSPIRE](#)].

[62] R.L. Jaffe and M. Soldate, *Twist four in electroproduction: canonical operators and coefficient functions*, *Phys. Rev. D* **26** (1982) 49 [[INSPIRE](#)].

[63] CMS collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, CMS-PAS-JME-13-002, CERN, Geneva, Switzerland (2013).

[64] ATLAS collaboration, *Discrimination of light quark and gluon jets in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, ATLAS-CONF-2016-034, CERN, Geneva, Switzerland (2016).

- [65] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *JHEP* **05** (2014) 146 [[arXiv:1402.2657](https://arxiv.org/abs/1402.2657)] [[INSPIRE](#)].
- [66] R. Medves, A. Soto-Ontoso and G. Soyez, *Lund and Cambridge multiplicities for precision physics*, *JHEP* **10** (2022) 156 [[arXiv:2205.02861](https://arxiv.org/abs/2205.02861)] [[INSPIRE](#)].
- [67] S. Höche and S. Prestel, *The midpoint between dipole and parton showers*, *Eur. Phys. J. C* **75** (2015) 461 [[arXiv:1506.05057](https://arxiv.org/abs/1506.05057)] [[INSPIRE](#)].
- [68] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](https://arxiv.org/abs/1410.3012)] [[INSPIRE](#)].
- [69] M. Cacciari, G.P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063 [[arXiv:0802.1189](https://arxiv.org/abs/0802.1189)] [[INSPIRE](#)].
- [70] M. Zaheer et al., *Deep sets*, in *Advances in neural information processing systems*, volume 30, I. Guyon et al. eds., Curran Associates Inc. (2017).
- [71] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy flow networks: deep sets for particle jets*, *JHEP* **01** (2019) 121 [[arXiv:1810.05165](https://arxiv.org/abs/1810.05165)] [[INSPIRE](#)].
- [72] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy flow polynomials: a complete linear basis for jet substructure*, *JHEP* **04** (2018) 013 [[arXiv:1712.07124](https://arxiv.org/abs/1712.07124)] [[INSPIRE](#)].
- [73] M. Abadi et al., *Tensorflow: a system for large-scale machine learning*, in *OSDI*, volume 16, (2016), p. 265.
- [74] F. Chollet, *Keras*, <https://github.com/fchollet/keras>, (2017).
- [75] D.P. Kingma and J. Ba, *Adam: a method for stochastic optimization*, 2014 [[arXiv:1412.6980](https://arxiv.org/abs/1412.6980)] [[INSPIRE](#)].
- [76] A.J. Larkoski, J. Thaler and W.J. Waalewijn, *Gaining (mutual) information about quark/gluon discrimination*, *JHEP* **11** (2014) 129 [[arXiv:1408.3122](https://arxiv.org/abs/1408.3122)] [[INSPIRE](#)].