

# Secondary Lund jet plane as a gluon-enriched sample

---

**Cristian Baldenegro,<sup>a</sup> Alba Soto-Ontoso,<sup>b</sup> Gregory Soyez<sup>c</sup>**

<sup>a</sup>*Massachusetts Institute of Technology, Laboratory for Nuclear Science, Cambridge, MA, 02139*

<sup>b</sup>*Departamento de Física Teórica y del Cosmos, Universidad de Granada, Campus de Fuentenueva, E-18071 Granada, Spain*

<sup>c</sup>*Université Paris-Saclay, CNRS, CEA, Institut de physique théorique, 91191, Gif-sur-Yvette, France*

**ABSTRACT:** We propose a new strategy to obtain a high-purity sample of gluon-initiated jets at the LHC. Our approach, inspired by the Lund jet plane picture, is to perform a dijet selection where the two jets are collinear to each other and their momentum fraction share is highly asymmetric, and to measure the primary Lund plane density of emissions of the subleading jet. The subleading jet in this topology is practically equivalent to a secondary Lund jet plane. We demonstrate by means of fixed-order calculations that such a simple setup yields gluon jet fractions of around 90% for the subleading jet for both quark- and gluon-initiated jets. This observation is confirmed using hadron-level Monte Carlo generated events. We also show that the extracted gluon purities are highly resilient to the overall colour structure of the event, to the flavour of the hard-scattering process, and to the parton distribution functions. This strategy is well-suited for constraining the radiation pattern of gluon-initiated jets using a set of fiducial cuts that can readily be tested at the LHC, without relying on taggers or statistical demixing.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The secondary Lund plane density</b>	<b>4</b>
2.1	Strategies to enhance the gluon fraction	5
2.2	First-principles estimates of gluon purities	7
<b>3</b>	<b>High-gluon purities at the LHC using a dijet selection</b>	<b>11</b>
<b>4</b>	<b>Summary and Outlook</b>	<b>15</b>
<b>A</b>	<b>Analytic results for gluon purities in the grooming setup</b>	<b>17</b>
<b>B</b>	<b>First-principles estimates for the gluon purities using small-<math>R</math> jets</b>	<b>17</b>

---

## 1 Introduction

Jets, the collimated showers of particles produced in high-energy particle collisions, are multi-scale probes of the strong interaction. The description of the formation of jets involves a combined understanding of the hard scattering producing energetic quarks and gluons, the parton showering driving the jet evolution, as well as the transition from the collection of partons to the hadrons that are eventually detected experimentally. In high-energy proton-proton collisions, the underlying event activity, multi-parton interactions, and initial-state radiation also contribute to the transverse momentum,  $p_t$ , and substructure of jets.

Although jet physics can be understood analytically to a great extent, in practice Monte Carlo (MC) event simulation is often used for phenomenological and experimental applications. For jet physics specifically, MC event simulation provides the baseline in searches for new physics, is a key ingredient in the calibration and flavour tagging of such jets, and is used for precision tests of Quantum Chromodynamics (QCD) [1].

A longstanding problem in MC event generators is the correct description of the fragmentation of gluon-initiated jets. As a paradigmatic example, the leading systematic uncertainty in the jet energy scale at the LHC is the MC uncertainty on the gluon-jet response [2–4].<sup>1</sup> Along these lines, Refs. [8, 9] showed that there is a much wider spread in predictions for gluon showers than for quark fragmentation for a number of jet substructure observables. This is mostly because MC generators are tuned to describe hadronic final states produced in  $e^+e^-$  collisions as a baseline, so that quark radiation patterns are

---

<sup>1</sup>Recently, the ATLAS collaboration has published a series of paper in which they investigate the origin of the gluon-jet response uncertainty in MCs [5–7]. Their results indicate that the baryon fraction inside jets play a significant role in their detector response.

much more strongly constrained than gluon ones. The mismodeling of gluon jet showers trickles down to uncertainties in other quantities at the LHC as well as the size of the calibration factors for machine-learning-based quark-gluon discriminators, which rely on MC generated events for training [10]. Having access to a gluon-enriched sample would reduce these uncertainties.<sup>2</sup>

A high-purity gluon-initiated jet sample would also facilitate the extraction of the strong coupling  $\alpha_s$  from jet substructure observables at the LHC. This is so since at lowest order in perturbation theory jet substructure observables are sensitive to  $\alpha_s C_i$ , where  $C_i$  is the colour charge of the emitter. In other words, there is a degeneracy between  $\alpha_s$  and the quark versus gluon fraction of jets [14].<sup>3</sup> More generally, a strategy that allows for constraining final-state radiation of gluon-initiated jets in an isolated way would be beneficial for jet substructure studies and measurements at the LHC.

Crafting a gluon-dominated jet sample has been an active topic of research. In  $e^+e^-$  collisions, a sample enriched in gluon-initiated jets can be obtained in events with two  $b$ -jets recoiling from a third jet, which is used as a proxy for gluon jets. Such an event topology has been measured at LEP to constrain soft gluon jets [20–33] and has been used in recent tuning campaigns of HERWIG7 [34, 35]. In such measurements, anti- $b$  tagging on the third jet and additional topological requirements were shown to improve the gluon purities. The statistical uncertainty of this channel would be highly reduced at the future FCC- $ee$ , which will also enable constraining gluon-initiated radiation by using associated Higgs boson production [36, 37].

At the LHC, a general strategy proposed in Refs. [8, 38] consists of targeting process-based and phase-space-based selections to enrich samples in quark-like or gluon-like radiation patterns. An example of the first category would be to use inclusive dijet (gluon dominated) and boson-tagged jet (quark dominated) samples to constrain quark and gluon jet substructure simultaneously. However, the interpretation of this data in the context of MC generator tuning is convolved with other effects unrelated to final-state radiation (e.g., the choice of parton distribution function (PDF), the order of the perturbative QCD (pQCD) calculation used for the hard scattering, and others), such that it becomes difficult to isolate the effect that comes from the mismodeling of gluon-initiated jets or other mechanisms in the collision. Regarding phase-space-based selections, one could rely on an inclusive jet sample and change the relative quark/gluon fraction by exploring different jet rapidities [39–41] or centre-of-mass energies [42].

A conceptually different approach to the ones discussed above pursues quark/gluon

---

<sup>2</sup>Throughout this paper we use a definition of gluon-enrichment in the spirit of Ref. [8, 9], namely “phase-space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of gluons (as interpreted by some suitable, though fundamentally ambiguous, criterion)”. We also note that there have been extensive recent developments on flavour-aware jet clustering algorithms [11–13] attempting at providing a more practical definition of the flavour of a jet. We will comment on this when relevant in this paper.

<sup>3</sup>There are different ways of eliminating this degeneracy. On the one hand, there are certain observables where the sensitivity to this degeneracy is reduced, e.g. ratios of energy-energy correlators [15–18]. Another way is to calculate the quark and gluon fractions using perturbative QCD as was done in Ref. [19] for the SoftDrop jet mass, or to include subleading corrections which partially break the degeneracy.

discrimination on a jet-by-jet basis exploiting the simple idea that gluons radiate more than quarks. This can be done using jet substructure observables [43–46], machine learning strategies [47–49], or a combination of both [50]. Yet another strategy consists of using statistical demixing, for example with topic modeling, to separate quark and gluon distributions at the event ensemble level [39, 40, 51–55]. In the latter, the connection with the partonic degrees of freedom is not as straightforward.

In this paper, we present a strategy to create a gluon-enriched sample based on the Lund jet plane picture [56, 57]. The Lund jet plane is a tool to create a proxy for the actual parton shower using the jet clustering tree. First, it re-organises the jet into an angular ordered sequence using the Cambridge/Aachen (C/A) algorithm [58]. Then, it proceeds by unwinding the C/A clustering sequence. The first declustering, i.e., the pair of subjets at largest angles, can be viewed as a proxy for the *first* emission in the shower (in an angular-ordered picture). The *primary* declusterings are constructed by iteratively undoing the C/A clustering sequence recursing, at each step, into the hardest of the two subjets, where, in a hadron collider context, hardest means having the largest transverse momentum with respect to the beam,  $p_t$ . The set of softer subjets along this chain is referred to as the *primary emissions*. This procedure can be repeated deeper into the C/A clustering tree. For each of the primary declusterings, we can do the same iterative declustering exercise, yielding the *secondary* declusterings and emissions associated with this primary declustering. To each of the secondary declusterings we could then associate *ternary* declusterings, and so on.

One can then build observables out of the kinematics of these declusterings. The simplest one is the primary Lund plane density, i.e., the density of primary emissions as a function of their relative transverse momentum and opening angle. The primary Lund plane density has been recently measured in inclusive jets [59, 60] and boosted top quark jets [61] at the LHC and compared to first-principle calculations [62] and MC simulations. It has been shown that this observable is a powerful tool to constrain hadronisation and parton showers in a factorised way. Recently, the Lund plane picture has been used to define a multiplicity-like observable that is both infrared and collinear safe and amenable to precision calculations [63–65]. Regarding quark/gluon discrimination, Ref. [50] computed analytically a likelihood discriminator using either the primary Lund plane or the full clustering tree and compared it to machine learning tools trained with the same input.

Here, we propose to create a gluon-enriched jet sample using the secondary Lund plane and, more specifically, to constrain gluon-initiated radiation patterns via the secondary Lund plane density. Recalling that the secondary emissions are defined as emissions from each of the primary emissions, i.e., each of the softer branches in the primary declusterings, one expects from perturbative QCD that the softest branch will often be initiated by a gluon, due to the soft singularity of the splitting function. This will not always be the case, and part of this paper is devoted to exploring different definitions of the secondary Lund plane density aiming at enhancing the gluon fraction. One of such possibilities is to construct the secondary Lund plane density by using the subleading jet in a dijet pair after imposing some pQCD-inspired fiducial cuts. We show that the subleading jet is initiated by a gluon up to 90% of the times.

The secondary Lund plane density has been used in other contexts. Ref. [57] used it to improve  $W$  boson jet tagging performance at lower jet  $p_t$ . Secondary Lund jet planes have also been used by the ATLAS Collaboration in graph neural networks to separate boosted  $W$  bosons from the QCD background [66]. Another type of secondary Lund planes has been used to improve the data-to-MC agreement for multiprong jets, as presented by the CMS Collaboration [67].

With this generic picture in mind, we organise this paper in two main steps. First, in section 2, we present the algorithm to construct the secondary Lund plane density as introduced in Ref. [57] and highlight a few experimentally-feasible strategies that yield a gluon-enriched sample of jets (section 2.1). For each of these methods, we compute in Sec. 2.2 the gluon fraction using a simple tree-level based “definition” of gluon jets. We also include a brief discussion on gluon purities at all-orders based on more recent flavour-aware jet algorithms. In the second step, we show that the Lund plane density can be used as a powerful way to characterise the properties of this gluon-enriched sample. To achieve this, we perform a series of Monte Carlo studies. For this second step, specific flavour assignments are not strictly-speaking needed, only the fact that our initial sample is gluon-enriched and defined with physical cuts achievable in an experimental context. We conclude and discuss some applications of our proposed strategy in Sec. 4.

## 2 The secondary Lund plane density

For a given jet (or subjet)  $j$ , we introduce a recursive declustering procedure following the ‘hardest’ branch as follows:

- start with  $j_1 = j$ ,
- at step  $i$ , we undo the last clustering of  $j_i$  into a pair of subjets  $j_i^{(\text{hard})}$  and  $j_i^{(\text{soft})}$ , such that  $p_{ti}^{(\text{hard})} > p_{ti}^{(\text{soft})}$ . We record the *Lund coordinates*  $\mathcal{T}_i \equiv \{\Delta_i, k_{ti}, z_i, \dots\}$  as the kinematics of this declustering, i.e.,

$$\Delta_i^2 = (y_i^{(\text{soft})} - y_i^{(\text{hard})})^2 + (\phi_i^{(\text{soft})} - \phi_i^{(\text{hard})})^2, \quad (2.1a)$$

$$k_{ti} = p_{ti}^{(\text{soft})} \Delta_i, \quad z_i = \frac{p_{ti}^{(\text{soft})}}{p_{ti}^{(\text{hard})} + p_{ti}^{(\text{soft})}}, \quad (2.1b)$$

where all variables are measured with respect to the colliding beams.

- define  $j_{i+1} = j_i^{(\text{hard})}$  and iterate the procedure until no declusterings are left.

This procedure defines an ordered list of *Lund emissions* and associated *Lund coordinates*

$$\mathcal{L}^{(\text{emis})}(j) \equiv [j_1^{(\text{soft})}, \dots, j_n^{(\text{soft})}] \quad \text{and} \quad \mathcal{L}^{(\text{coords})}(j) = [\mathcal{T}_1, \dots, \mathcal{T}_n]. \quad (2.2)$$

The average density of Lund emissions per jet, also known as the Lund plane density, is then given by

$$\rho \equiv \frac{1}{N_{\text{jets}}} \frac{d\mathcal{L}^{(\text{emis})}(j)}{d \ln k_t d \ln(1/\Delta)}. \quad (2.3)$$

When applying this recursive declustering procedure to a reconstructed jet,  $j$ , in proton-proton collisions one first needs to recluster it with the C/A algorithm. Then, one can directly define the list of *primary* emissions and Lund coordinates as

$$\mathcal{L}_{\text{primary}}^{(\text{emis,coords})}(j) \equiv \mathcal{L}^{(\text{emis,coords})}(j). \quad (2.4)$$

For each primary emission  $j^{\text{soft}}$ , the *secondary* emissions and coordinates are defined as the primary declusterings of  $j^{\text{soft}}$

$$\mathcal{L}_{\text{secondary}}^{(\text{emis,coords})}(j, j^{(\text{soft})}) \equiv \mathcal{L}_{\text{primary}}^{(\text{emis,coords})}(j^{(\text{soft})}). \quad (2.5)$$

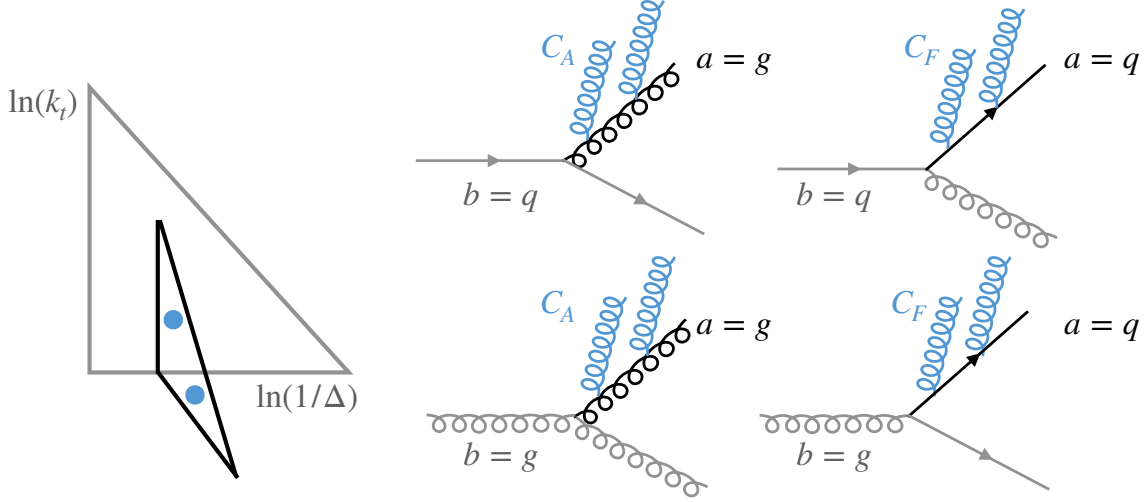
At this stage, one can either apply Eq. (2.5) to all primary emissions or to just one declustering chosen following an infrared-and-collinear (IRC) safe procedure. For the purpose of this paper, it is sufficient to focus on the latter. The key observation is that secondary emissions are mostly gluons due to the singularity structure of QCD amplitudes. To illustrate this point, let us work at first order in the strong coupling constant,  $\alpha_s$ , and in the soft-and-collinear limit, i.e., the splitting/declustering satisfies  $\Delta \ll R, k_t \ll p_{t,\text{jet}}$ , where  $R$  is the reconstructed jet cone radius. At this level of accuracy, the density of secondary emissions is given by

$$\rho_s \equiv \frac{1}{N_{\text{jets}}} \frac{d\mathcal{L}_{\text{secondary}}^{(\text{emis})}(j, j^{(\text{soft})})}{d\ln k_t d\ln(1/\Delta)} \sim \frac{2\alpha_s(k_t)C_R}{\pi} \quad (2.6)$$

where  $C_R$  is the colour factor of the first, secondary emission. The value of  $C_R$  fluctuates on a jet-by-jet basis depending on both the flavour of the jet-initiator and the kinematics of the splitting, which will determine which branch is tagged as secondary. Our goal is to maximize the configurations in which  $C_R = C_A$ . At leading order, there are four possibilities to consider, as displayed in Fig. 1. Let us first discuss the case in which the parton initiating the jet is a quark, such that the first splitting is always  $q \rightarrow qg$ . Since the  $P_{gq}$  splitting function has a soft divergence, the gluon will take, on average, a small fraction of the parent's energy and will thus be tagged as secondary, i.e.,  $C_R = C_A$ . This is spoiled by splittings with  $z > 1/2$ , since then the quark is tagged as secondary and  $C_R = C_F$ . For gluon-initiated jets, there are two possible splitting channels:  $g \rightarrow gg$  and  $g \rightarrow q\bar{q}$ . The former needs no discussion since the secondary Lund plane is always spanned by a gluon so that  $C_R = C_A$ . In turn, the latter will always contribute with a  $C_R = C_F$  factor. It is important to note that this splitting channel is naturally suppressed since it does not have a soft singularity. Thus, at leading-order, the  $C_R = C_A$  case is favored by the soft singularity of QCD.

## 2.1 Strategies to enhance the gluon fraction

So far, we have kept the discussion at the level of one emission. Further considerations are required when considering multiple emissions and when gauging the experimental feasibility of such an observable. One of them is having sufficient phase-space so as to populate the secondary Lund plane. This imposes that the reconstructed jet must have high- $p_t$  and that the emission that spawns the secondary Lund plane cannot be too soft or too collinear.



**Figure 1:** Schematic diagram showing the tree-level contributions to the secondary Lund jet plane density. The two leftmost diagrams correspond to emissions that spawn a secondary Lund plane enriched in gluon-initiated radiation. The two rightmost diagrams represent the two dominant contributions to parton flavour changes, where the secondary Lund plane is dominated by quark-initiated radiation.

The latter consideration is also important to mitigate distortions due to boundary effects, i.e., when two subjects or two jets are too close with each other. The large-angle regime is also problematic from the perspective of interpreting the results in the context of MC event generators, since they work best in the collinear limit.

With these general criteria in mind, one can consider at least three setups to enhance the gluon fraction (with selections considering Run-2 or Run-3 high-pileup conditions):

- **Grooming setup:** run anti- $k_t$  clustering [68] on the event with large jet radius  $R = 1.2$ . Select all jets with  $p_t > 1$  TeV and apply SoftDrop grooming with  $z_{\text{cut}} = 0.1$  and  $\beta = 0$  [69, 70]. For a given jet, we denote  $j_{\text{SD}}$  the primary declustering at largest angle that satisfies  $z_{\text{min}} < z_g < z_{\text{max}}$  and  $\Delta > \Delta_{\text{min}}$ . If no declustering is found we discard the jet. We then apply Eq. (2.5) with  $j_{\text{SD}} = j^{(\text{soft})}$  and obtain the secondary Lund plane density normalized to the number of jets that satisfy the above set of fiducial cuts. As we argue below, the optimal values of the free parameters are  $z_{\text{min}} = 0.1$ ,  $z_{\text{max}} = 0.2$  and  $\Delta_{\text{min}} = 0.8$ .
- **Trimming setup:** run anti- $k_t$  clustering on the event with large jet radius  $R = 1.2$ . Select all jets with  $p_t > 1$  TeV. Recluster each jet with a smaller  $R$  (e.g.,  $R = 0.4$ ) using a clustering sequence of the (generalised)  $k_t$  algorithm family. Select the two most energetic subjects,  $j_{\text{lead}}$ ,  $j_{\text{sublead}}$  and check whether they satisfy  $0.1 < z < 0.2$  and  $\Delta > 0.8$ . If so, compute the “secondary” Lund plane density using  $j = j_{\text{sublead}}$  in Eq. (2.3). If not, discard the jet.
- **Dijet selection (nominal):** run anti- $k_t$  clustering on the event with  $R = 0.4$ , as is

the standard at the LHC. Select all jet pairs in the event where the leading jet has  $p_{t,\text{lead}} > 700$  GeV and the subleading jet has  $p_{t,\text{min}} < p_{t,\text{sublead}} < p_{t,\text{max}}$ .<sup>4</sup> We also impose an angular constraint between the two jets  $\Delta_{\text{min}} < \Delta R < \Delta_{\text{max}}$ . Specifically, as we will discuss below, we use  $p_{t,\text{min}} = 150$  GeV,  $p_{t,\text{max}} = 200$  GeV,  $\Delta_{\text{min}} = 1$  and  $\Delta_{\text{max}} = 1.2$ . We build the Lund plane density using  $j \equiv j_{\text{sublead}}$  in Eq. (2.3).

The SoftDrop setup is conceptually closer to the construction of the primary/secondary Lund plane density [57], since it relies on the same C/A clustering jet tree. The trimming [71] setup mitigates distortions due to boundary effects at large angles. The third approach has the advantage that it relies on the standard anti- $k_t$  clustering sequence that is used by LHC experiments, and for which flavour taggers [72–75] and jet energy calibrations are derived. Before delving into a dedicated MC study of the gluon purities achieved with these strategies, we provide next an analytic study at fixed-order.

## 2.2 First-principles estimates of gluon purities

The secondary Lund plane density is calculable in perturbative QCD. For the purpose of this paper, we limit ourselves to a simple first-order calculation. We postpone a fully resummed prediction for  $\rho_s$  to a dedicated study.<sup>5</sup> More specifically, we focus on the double-logarithmic approximation where the Lund-plane density, given by Eq. (2.6), only depends on the colour factor of the parton emitting soft-collinear radiation in the secondary Lund plane. In this setup, we compute the fraction of events where this leading parton is a gluon at first order in  $\alpha_s$ . This is well-defined at this order of perturbation theory and serves to illustrate that our procedure yields high-purity gluon samples. At this level of accuracy, the trimming setup is equivalent to the dijet selection. We thus present results for the grooming and dijet setups both at exact  $\mathcal{O}(\alpha_s)$  using NLOJet++ [76] and using a fully-analytic approach valid in the collinear limit. This second approach is meant to study the extent to which the gluon purity depends on the flavour of the initial hard parton, the one that triggered the splitting.

Let us begin by discussing the collinear limit. In the grooming setup, the fiducial cuts presented in the previous section amount to selecting momentum fractions between some  $z_{\text{min}}$  and some  $z_{\text{max}}$ . We therefore define

$$I_{ab}^{(\text{groom})} \equiv I_{ab}^{(\text{groom})}(z_{\text{min}}, z_{\text{max}}) = \int_{z_{\text{min}}}^{z_{\text{max}}} dz P_{ab}(z), \quad (2.7)$$

where  $P_{ab}(z)$  are the Altarelli-Parisi splitting functions corresponding to the processes depicted in Fig. 1. Unless explicitly required, the  $z_{\text{min}}$  and  $z_{\text{max}}$  arguments will be omitted. If we start with a quark (or gluon), the fraction of secondary Lund planes with a gluon as a leading parton is, respectively (assuming  $z_{\text{min}} < z_{\text{max}}$ )

$$f^{(\text{groom})}(g|q) = \frac{I_{gq}^{(\text{groom})}}{I_{qq}^{(\text{groom})} + I_{gq}^{(\text{groom})}} \quad \text{and} \quad f^{(\text{groom})}(g|g) = \frac{I_{gg}^{(\text{groom})}}{I_{gg}^{(\text{groom})} + I_{gq}^{(\text{groom})}}. \quad (2.8)$$

<sup>4</sup>We have checked that a selection on the momentum imbalance,  $z$ , between the two jets also works.

<sup>5</sup>The core pieces of the resummed predictions for the primary Lund plane density [62] can be reused for the secondary Lund plane density.



All integrals in these expressions can be computed analytically and are given in Appendix A.

The case of the dijet selection is slightly more involved. Indeed, since we are imposing separate dimensionful cuts on the leading and subleading jets, the selected rates in each partonic channel depend on the underlying jet cross-section. Still working in the collinear limit, we thus introduce

$$\begin{aligned} I_{ab}^{(\text{dijet})} &= \int dp_t \frac{d\sigma_b(p_t)}{dp_t} \int_0^1 dz P_{ab}(z) \Theta((1-z)p_t > p_{t,\text{lead}}^{(\min)}) \Theta(p_{t,\text{sublead}}^{(\min)} < zp_t < p_{t,\text{sublead}}^{(\max)}), \\ &= \int dp_t \frac{d\sigma_b(p_t)}{dp_t} I_{ab}^{(\text{groom})} \left( \frac{p_{t,\text{sublead}}^{(\min)}}{p_t}, \min \left( 1 - \frac{p_{t,\text{lead}}^{(\min)}}{p_t}, \frac{p_{t,\text{sublead}}^{(\max)}}{p_t} \right) \right), \end{aligned} \quad (2.9)$$

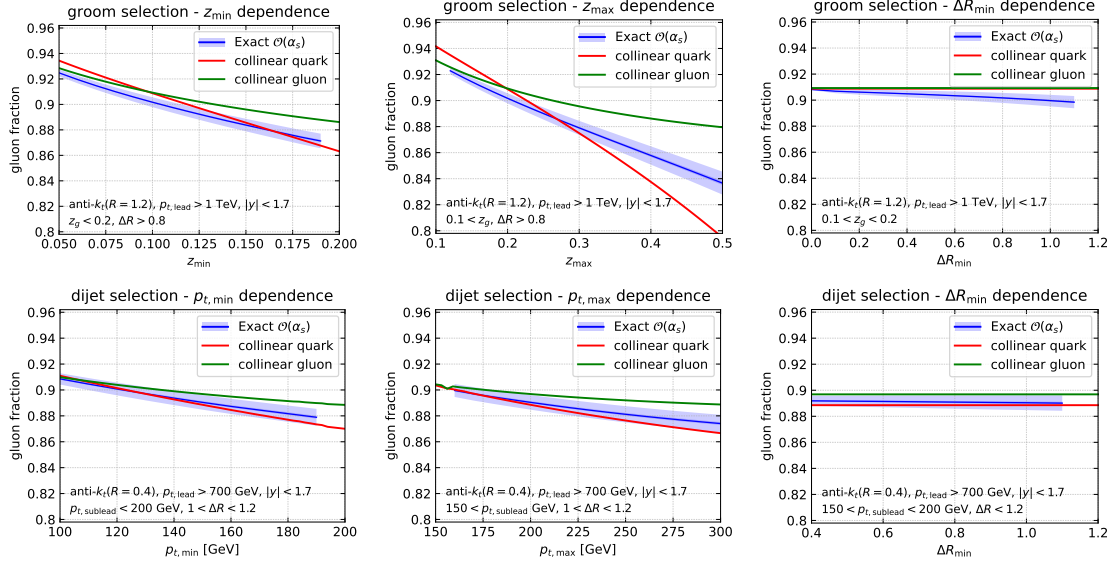
where, at leading order in  $\alpha_s$ , the flavour decomposition of the underlying Born-level process is well-defined. The fraction of secondary Lund planes with a gluon as a leading parton is

$$f^{(\text{dijet})}(g|q) = \frac{I_{gq}^{(\text{dijet})}}{I_{qq}^{(\text{dijet})} + I_{gq}^{(\text{dijet})}} \quad \text{and} \quad f^{(\text{dijet})}(g|g) = \frac{I_{gg}^{(\text{dijet})}}{I_{qq}^{(\text{dijet})} + I_{gg}^{(\text{dijet})}}. \quad (2.10)$$

respectively for a leading quark or leading gluon. It is interesting to notice that while the gluon fractions  $f^{(\text{groom})}$  are only functions of  $z_{\min}$  and  $z_{\max}$ , the corresponding fractions  $f^{(\text{dijet})}$  also depend on the choice of PDFs and of factorisation scale through the inclusive jet cross-section. In practice, we have found that this dependence is however very small, much smaller than the uncertainty associated with the actual fraction of quarks and gluons which also depends on the choice of PDFs and factorisation scale. The integrals  $I_{ab}^{(\text{dijet})}$  can only be computed analytically for a simple approximation of the inclusive jet cross-section. Since the resulting equations are not particularly helpful, the results presented below use the explicit integration with the same PDF set as the one used with `NLOJet++`.

In addition to this DGLAP-based estimate, we have also ran a MC simulation of Born-level  $pp \rightarrow jjj$  events using an extended version of `NLOJet++` that has access to flavour information. The centre-of-mass energy is set to  $\sqrt{s} = 13.6$  TeV. We choose the `PDF4LHC21_mc` PDF set and account for uncertainties using the Hessian reduction strategy [77]. The central values for the factorisation and renormalisation scales are set to the scalar sum of the transverse momentum of the final-state partons. Scale variations are those provided by `NLOJet++` and we take the envelope to define the scale uncertainty. The total perturbative uncertainty of our exact fixed-order result is obtained by adding in quadrature the scale and PDF uncertainties.

The obtained gluon fractions are shown in Fig. 2. We first discuss the groomed selection. In this case, we choose large-radius jets ( $R = 1.2$ ) at high- $p_t$  ( $p_t > 1$  TeV). We also impose a rapidity cut ( $|y| < 1.7$ ) to account for the acceptance of the CMS and ATLAS detectors. The free parameters in this setup are  $z_{\min}$  and  $z_{\max}$  together with the minimum opening angle  $\Delta_{\min}$  between the prongs. Overall, the gluon fraction ranges from 87% to 93%. We observe that it decreases when increasing either  $z_{\min}$  or  $z_{\max}$  (keeping the other one fixed). The reduction is more pronounced for a quark-initiated jet. This decrease of the gluon fraction can be traced back to the fact that the ratio  $P_{qx}(z)/P_{gx}(z)$



**Figure 2:** Analytic estimates for the gluon fraction of the subleading subjet in the grooming setup (top row) and of the subleading jet with a dijet selection (bottom row). The different columns explore the dependence of the gluon fraction on the free parameters of each setup. On average, we achieve gluon fractions of around 90%.

(with  $x = q$  or  $x = g$ ) increases for  $0 < z < 1/2$ , meaning that an increasingly large fraction of the secondary partons will be quarks when one departs from the soft limit where  $P_{qx}(z)/P_{gx}(z) = 0$ . In turn, the  $\Delta_{\min}$ -dependence is rather flat at this level of accuracy. The collinear estimates indicate that fixing  $z_{\min} = 0.1$  with  $z_{\max} = 0.2$  leads to almost identical gluon-fraction for quark or gluon initiated jets, around 91%.

In the dijet selection case, we analyse jets with standard anti- $k_t$  properties ( $R = 0.4$ ). The leading jet is again at high- $p_t$  ( $p_t > 700$  GeV) and only jets with  $|y| < 1.7$  are accepted. The free parameters for this setup are the  $p_t$ -window for the subleading jet characterized by  $p_{t,\min}$  and  $p_{t,\max}$  together with the minimum angular separation between the leading and subleading jets,  $\Delta R_{\min}$ . Once again, the collinear estimate for the gluon fraction, Eq. (2.10), correctly describes the trends of the exact fixed-order result. In this case, we find a milder dependence of the gluon fraction with respect to the different parameters. It stays around 90% in the explored parameter space. Enlarging either  $p_{t,\min}$  or  $p_{t,\max}$  results into a reduction of the gluon fraction. For just one emission,  $p_{t,\min}$  or  $p_{t,\max}$  plays the same role as the momentum imbalance cuts,  $z_{\min}$  with  $z_{\max}$ , which we discussed in the groomed setup. In other words, increasing  $p_{t,\min}$  is equivalent to increasing  $z_{\min}$ . Therefore, it is natural that the dijet and the groomed setups display a similar behaviour. Once again, we find that, at  $\mathcal{O}(\alpha_s)$ , the gluon fraction does not change when varying  $\Delta_{\min}$ .

Regarding the agreement between the analytic estimate in the collinear limit and the exact  $\mathcal{O}(\alpha_s)$  result we observe that, overall, the collinear estimate captures the trends of the exact result. However, there are some visible discrepancies specially in the groomed selection. Note that in both the groomed and dijet selection cases the jet radii together

with the other angular parameters, such as  $\Delta R$ , are not particularly small and thus the collinear approximation is expected to deviate from the exact  $\mathcal{O}(\alpha_s)$  result. In Appendix B, we demonstrate that the agreement improves when considering a more collinear setup.

Because the dijet selection approach has the possibility of branching out to other applications beyond the secondary Lund jet plane, we will mainly focus on the latter for the remainder of this paper. The conclusions from the hadron-level simulation results discussed in the next section also hold for the SoftDrop and trimming setups.

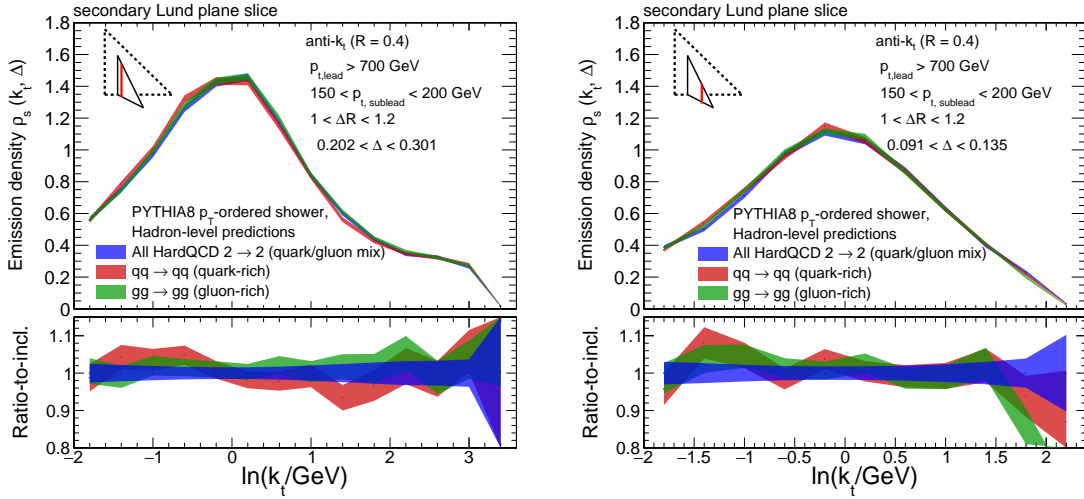
We end up this discussion by comparing our first-principle estimates for the gluon purities, based on a one-emission picture, with an all-orders, infrared and collinear safe result obtained with a recently introduced flavour-aware jet algorithm (IFN) [78]. This comparison is meant to quantify the impact of corrections beyond  $\mathcal{O}(\alpha_s)$  on the gluon purity. The results are presented in Table 1. To obtain the IFN gluon purities, we generate jet events with PYTHIA8 at parton-level either for a fixed Born-channel ( $qq \rightarrow qq$  or  $gg \rightarrow gg$ ) or considering all  $2 \rightarrow 2$  QCD scatterings. We then perform two types of event selections: (i) inclusive anti- $k_t$  jets with  $R = 0.4$  and  $p_t > 150$  GeV and (ii) the dijet selection introduced in Sec. 2.1. For each selection, we apply the IFN algorithm on a given jet and record the number of gluon-tagged jets.

Let us first discuss the results obtained for the inclusive jet selection. Trivially, the gluon at Born-level is either 0 or 100% when selecting  $qq \rightarrow qq$  and  $gg \rightarrow gg$  hard scattering process, respectively. At all orders, we observe that the gluon fraction is around 5% for the  $qq \rightarrow qq$  Born-level process. This originates from multi-jet events triggered by hard gluon emissions. Similarly, the gluon fraction is degraded at all-orders for the  $gg \rightarrow gg$  process due to relatively hard  $g \rightarrow q\bar{q}$  splittings that generate resolved quark-initiated jets.

For the dijet selection we observe a 15% reduction on the gluon purity of the subleading jet due to higher-order effects. Despite this numerical difference, the parametric trends and overall picture of using the secondary Lund plane as a gluon-enriched sample remain valid at all-orders. The gluon fraction also appears reasonably independent of the underlying hard process.

Sample	Born-level gluon fraction	IFN gluon fraction
$qq \rightarrow qq$ , inclusive jets	0%	4.5%
$gg \rightarrow gg$ , inclusive jets	100%	89.7%
$qq \rightarrow qq$ , dijet selection (subleading jet)	88.9%	75.6%
$gg \rightarrow gg$ , dijet selection (subleading jet)	89.7%	71.2%
all, dijet selection (subleading jet)	89%	75.0%

**Table 1:** Comparison between Born-level gluon fractions and those returned by the IFN algorithm for different samples. The Born-level gluon fraction for the dijet selection is obtained from the collinear estimate when fixing the Born channel and from the exact  $\mathcal{O}(\alpha_s)$  result when considering all  $2 \rightarrow 2$  QCD hard processes.



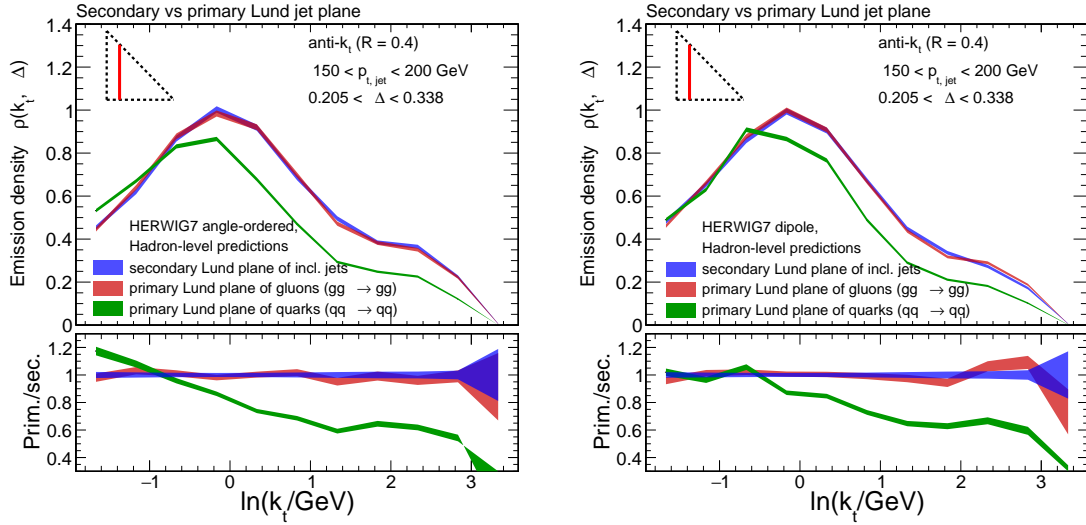
**Figure 3:** Vertical slices of the secondary Lund jet plane density at large (left) and small (right) angles. The three curves in each plot correspond to different choices for the hard  $2 \rightarrow 2$  scattering: all possible QCD channels (blue), only  $qq \rightarrow qq$  (red) and  $gg \rightarrow gg$  (green). The lower panel in each plot shows the ratio to the inclusive case. Results are shown for PYTHIA8 but similar conclusions can be drawn from HERWIG7.

### 3 High-gluon purities at the LHC using a dijet selection

To match the expected amount of high- $p_t$  jet events collected in Run 2 with the CMS or ATLAS experiments, predictions are generated for 5 million events (similar amount of events are expected for Run 3). Only jet events produced purely by the strong interaction are considered, i.e.,  $2 \rightarrow 2$  QCD scatterings. Photon bremsstrahlung off quarks is allowed in these simulation studies.<sup>6</sup> The simulation is done with three different MC setups: PYTHIA8 with its default  $p_t$ -ordered shower [79], and HERWIG7 with both the angular-ordered shower and its dipole variant [80]. In all cases, we present results at hadron-level with default settings (including the simulation of the underlying event, multiple parton interactions, colour reconnections and initial state radiation). Jets are clustered using the anti- $k_t$  algorithm and reclustered with the C/A algorithm using the FastJet package [81, 82]. From these events, about  $\mathcal{O}(10^5)$  gluon-like jets are obtained using the event selection requirements described above to select the primary emission, which corresponds to roughly  $10^6$  Lund emissions filling the secondary Lund plane. In all figures presented in this section, the bands represent the expected statistical uncertainties at the LHC with these selections.

First, we evaluate the dependence of  $\rho_s$  on the hard-scattering process itself. That is, we compute  $\rho_s$  either inclusively or fixing the  $2 \rightarrow 2$  scattering to be quark-rich ( $qq \rightarrow qq$ ) or gluon-rich ( $gg \rightarrow gg$ ). Despite being an academic exercise, it is important to demonstrate

<sup>6</sup>Occasionally, a large angle, hard photon emission from a quark will be selected as an emission to fill the secondary Lund jet plane, but these events are rejected in an experimental context with jet identification criteria and they represent a contribution at the per mille level in our MC-simulation studies.



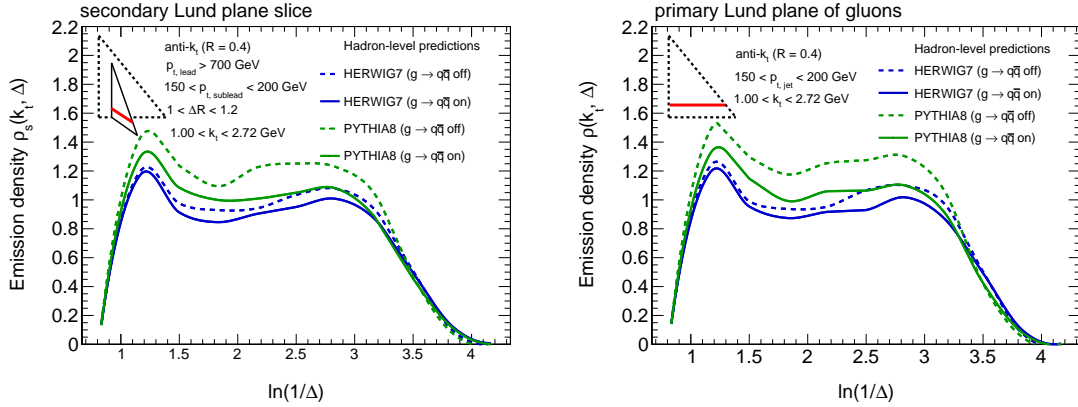
**Figure 4:** Vertical slice of a secondary Lund jet plane density derived from an inclusive jet sample (blue) compared with the primary Lund jet plane of jets produced in  $gg \rightarrow gg$  scatterings (red) and of jets produced in  $qq \rightarrow qq$  scatterings (green). Both panels are generated with HERWIG7 either using the angle-ordered shower (left) or the dipole shower (right). The bottom panels show the ratio between the primary and secondary Lund plane densities.

that the proposed methodology is independent of the flavour of the partons participating in the hard scattering. This also tests the sensitivity to the PDFs. The results are shown in Fig. 3 by choosing two angular slices of the secondary Lund plane density as obtained with PYTHIA8. First we note that  $\rho_s$  is peaked around  $k_t = 1 \text{ GeV}$  for this jet selection. Changes on the hard-scattering process translate into less than 10% differences at the level of the secondary Lund plane density. This nice feature shows up both at large and small angles. Also, we find quantitative agreement between all MC generators discussed above.

To assess if the secondary Lund jet plane yields the desired results (i.e., a sample of gluon-enriched radiation), it is useful to compare it with a reference distribution. For this purpose, we use the primary Lund jet plane of gluon-initiated jets generated from  $gg \rightarrow gg$  scatterings. The reference primary Lund jet plane is built using the two highest  $p_t$  jets in the event, with each of them required to have  $150 < p_t < 200 \text{ GeV}$  and  $|y| < 1.7$ , i.e. the same  $p_t$  and  $y$  requirements that are used for the jets used for the secondary Lund jet plane analysis.<sup>7</sup> As another reference, we also compare the results to the primary Lund jet plane density of quark-initiated jets.

We compare the secondary Lund plane density obtained from inclusive jets to the primary Lund plane density for gluon-initiated jets in Fig. 4. In this case, we show results generated with the two parton showers of HERWIG7. The curves agree within 10% deviation.

<sup>7</sup>At the simulation level, the partonic hard scale,  $\hat{p}_t$ , has to be lower for the reference sample than the  $\hat{p}_t$  used for the secondary Lund jet plane.

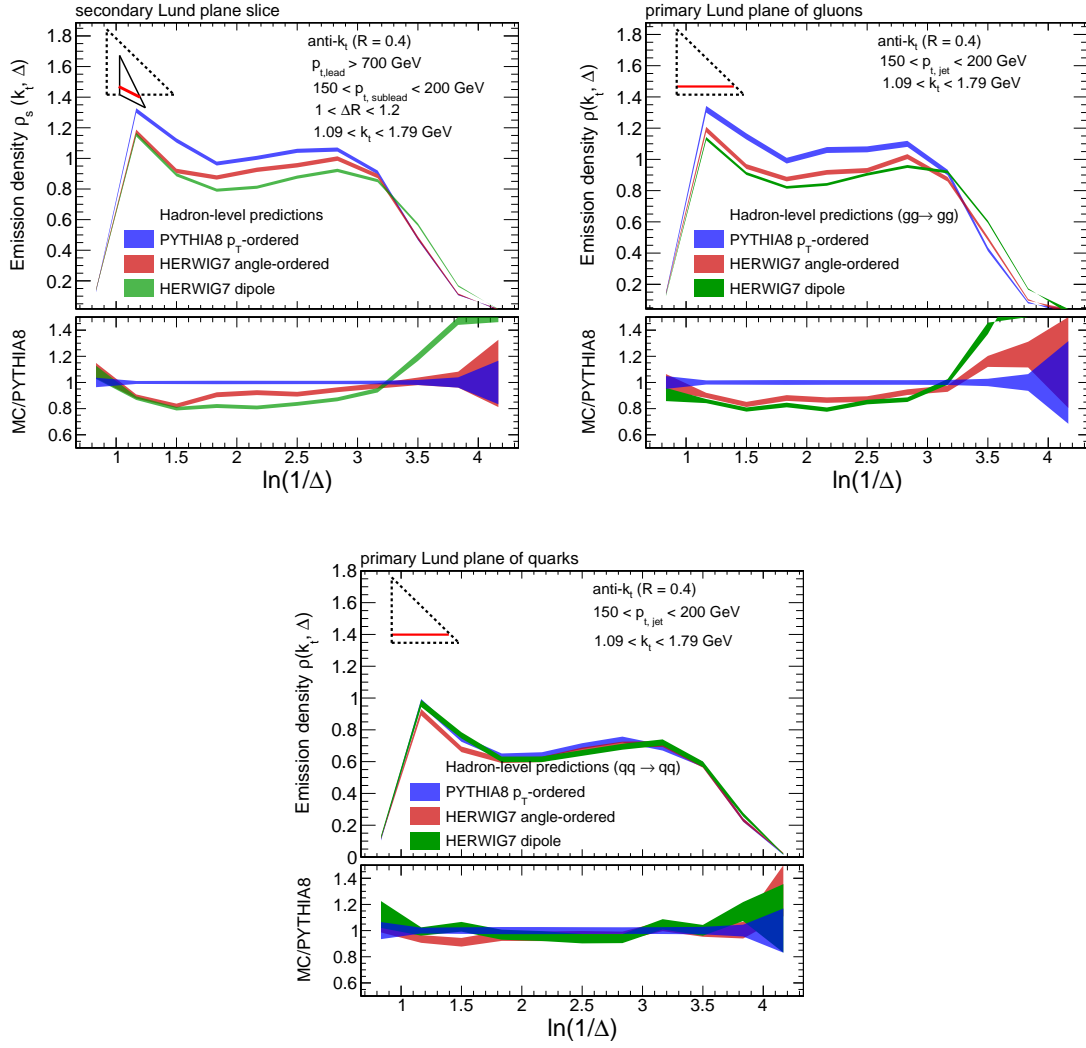


**Figure 5:** Impact of  $g \rightarrow q\bar{q}$  splittings on a horizontal slice of the secondary Lund plane density (left) and of the primary Lund plane density from  $gg \rightarrow gg$  scatterings (right). Results are shown for both PYTHIA8 and HERWIG7.

tions throughout the whole  $k_t$ -range for both parton shower models. Thus, we find that  $\rho_s$  is an excellent proxy for an artificially created sample of pure gluon-initiated jets. For reference, we also include on the right-panel the quark-initiated primary Lund jet plane density. We observe that the differences between the quark-like and gluon-like Lund jet plane densities follows Casimir scaling at higher- $k_t$  values. This further supports, with hadron-level results, that the method yields gluon-like Lund jet plane densities.

Next, we analyze the impact of gluon splittings to quark-antiquark pairs by switching them off in the simulation. Although not realistic from a physical point of view, this exercise is instructive to understand whether the proposed fiducial cuts manage to suppress this channel. We again use the primary Lund plane of gluon-initiated jets as our baseline to compare the performance of the secondary density. The results are displayed in Fig. 5 for a slice in transverse momentum. Removing  $g \rightarrow q\bar{q}$  splittings leads to an increase of emissions in both PYTHIA8 and HERWIG7 (more pronounced for the former) for the primary Lund plane of gluons. By examining the secondary Lund plane, similar modifications are found at a quantitative level. In particular, the secondary Lund plane density remains almost identical to the primary density of gluon-initiated jets with or without  $g \rightarrow q\bar{q}$  splittings. This further confirms that parton flavour changes from  $g \rightarrow q\bar{q}$  are negligible with the proposed dijet selection.

To end up this section, we explore the constraining power of the secondary Lund plane density in terms of MC generator models. We compare  $\rho_s$  to the primary Lund plane density of gluon-initiated jets in a horizontal Lund plane slice in Fig. 6. This comparison is meant to illustrate the extent to which one can discriminate parton showers and hadronisation models using only secondary Lund jet planes compared to an idealized scenario where one could unambiguously tag gluon-initiated jets. We find that the three MC setups start to differ in the deep collinear region, i.e.,  $\ln(1/\Delta) > 3$ . In this region of phase-space, the HERWIG7 dipole shower predicts a larger emission density for collinear emissions, almost



**Figure 6:** Comparison between different MC generators for a horizontal slice of the secondary Lund plane density (left) and of the primary Lund plane density for gluon-initiated jets (right). The lower plot represents the same slice, but for a primary Lund plane density for quark-initiated jets. The bottom panel in each plot shows the ratio to the PYTHIA8 results.

a factor of 2 larger with respect to HERWIG7 with an angle-ordered shower or PYTHIA8. The same trend is observed using the reference primary Lund plane for gluon jets. This shows that the secondary Lund plane densities matches the constraining power of a pure gluon-initiated jet sample. In other words, the spread among the different MC-generated predictions for secondary Lund jet planes is generally the same as the one observed for gluon-jet primary Lund jet plane densities. Repeating the same exercise, but for a slice of the primary Lund density of quark-initiated jets, we observe that the spread among the different predictions is smaller than for gluon-initiated jets. This is consistent with the



initial findings reported in Refs. [8, 9]. This further highlights the need for providing such a reference pure sample of gluon-initiated final-state radiation constraints.

## 4 Summary and Outlook

In this paper, a proposal for extending the exploration of the Lund jet tree in high energy proton-proton collisions at the LHC has been presented. Specifically, the possibility of using the average density of emissions off a primary emission, the secondary Lund jet plane, to constrain gluon-initiated radiation is investigated. Our strategy heavily relies on a fundamental property of QCD: the soft divergence of QCD splitting functions ( $q \rightarrow qg$  and  $g \rightarrow gg$ ). That is, collinear gluon emissions will typically carry a small fraction of its parent’s energy. As such, if one focuses on splittings with asymmetric momentum balance and selects the less energetic branch, the probability of it being a gluon is enhanced.

We discuss three different ways of constructing the secondary Lund jet plane density of emissions, either with well-established jet substructure techniques (SoftDrop grooming, trimming) or with a dijet (multijet) event selection. Given the simplicity of the latter, we focus on it for most of the paper. Analytically, we demonstrate by means of fixed-order calculations that an inclusive dijet selection with a pair of high- $p_t$  jets sharing asymmetric momentum balance and collinear to each other can yield subleading jet gluon fractions that are, on average, around 90% for realistic LHC fiducial cuts. This result is independent of the flavour of the jet-initiating parton to a large extent. We compute the gluon purity at all-orders by running the IFN jet flavour algorithm on a PYTHIA8 sample at parton level and confirm that the inclusive dijet selection yields high gluon purities (around 75%). Note that higher gluon purities can be reached by selecting softer subleading (sub)jets, at the expense of phase space reduction.

To further confirm the potential of such an observable, we perform hadron-level simulation studies with HERWIG7 and PYTHIA8. We select anti- $k_t$  jets with  $R = 0.4$  and pair them up imposing that the leading jet has  $p_{t,\text{lead}} > 700$  GeV and the subleading  $150 < p_{t,\text{sublead}} < 200$  GeV. On top of the  $p_t$ -selection, we also constrain their angular separation to be  $1 < \Delta R < 1.2$ . For Run-2 luminosity, this corresponds to  $\mathcal{O}(10^5)$  jets. The secondary Lund plane density is then constructed using the subleading jet.

We find that the resulting secondary Lund plane densities are resilient, within 10%, to the flavour of the partons participating in the hard-scattering process (including their PDFs). Remarkably, the secondary Lund plane density agrees with the primary Lund plane density from gluon-initiated jets again within 10%. We also study the sensitivity to flavour changes due to gluon to quark-antiquark splittings and find it to be negligible for our jet selection. Finally, the model discrimination power for secondary Lund jet plane densities is similar to the one obtained in an ideal primary Lund jet plane density from gluon-initiated jets. This supports the idea of using the secondary Lund plane density as a tool to constrain the modeling of gluon-initiated radiation.

Having access to a gluon-enriched sample using simple hadron-level phase space cuts opens up a series of interesting opportunities. At the phenomenological level, a measurement of the secondary Lund plane density can serve as a much needed input for constraining



gluon-initiated radiation in MC event generators. This could potentially reduce the scale factor uncertainties for quark-gluon likelihood discriminators. One should note that the dijet selection we propose necessarily sets a kinematic upper bound on the hardness of the softer jet. Thus, a limitation of the method is that it is not possible to obtain gluon-rich samples at the TeV scale for the expected LHC integrated luminosities without penalizing the gluon purity of the softer jet. However, the method can be used to derive a data-based anchor point for gluon-initiated radiation patterns. The hope is that tuning Monte Carlo event generators to such a reference data sample would potentially reduce the mismodeling of gluon jet radiation at higher momenta. Measurements of the hadron species composition and their energy distribution within jets in such a gluon-rich sample could serve as a guide for the reduction of uncertainties associated to the jet flavour response uncertainty for the jet energy scale. Regarding precision physics, suppressing the sensitivity to the quark/gluon jet fraction represents a major step forward towards the goal of extracting  $\alpha_s$  from jet substructure observables. To this end, there are better candidates than the secondary Lund jet plane density from the theoretical perspective. For example, observables such as the SoftDrop groomed jet mass, or the average Lund multiplicities have a simpler resummation structure. The drawback of computing observables on the secondary Lund plane is that it pushes the accuracy of the fixed-order result to one order higher. For instance, providing a theoretical prediction that could result into a competitive  $\alpha_s$  extraction would require the calculation of  $pp \rightarrow jjj$  at least at NNLO [7, 83], for which no public code currently exists. The complexity on the resummation remains the same as for observables on the primary branch. Nevertheless, we hope that the resilience of observables based on the secondary Lund plane to quark/gluon fractions and PDF uncertainties motivates the community to perform theoretical computations in this direction. In fact, the resilience of our method to the jet-initiating flavour can be readily tested experimentally using existing quark/gluon discriminators.

Another interesting spin-off of this work would be to apply this technique to heavy-ion jets. There, controlling the quark/gluon fraction represents a fantastic opportunity to understand the colour-charge dependence of energy loss. Several proposals to disentangle quark and gluon jets have been put forward in the literature [84–90]. Despite the simplicity of our approach, its applicability to a heavy-ion context faces some challenges. One of them involves mitigating the underlying event activity that has so far hampered Lund plane measurements in heavy-ion collisions. In addition, a large sample of high- $p_t$  jets is required so as to guarantee sufficient phase-space to fill up the secondary Lund plane density. Heavy-flavour tagged jets can potentially be used to overcome some of these challenges, for instance extending the dijet selection to a trijet selection with two  $b$ -tagged jets and an anti- $b$  tagged third jet.

In summary, we have presented a simple strategy to measure a high-purity sample of gluon-initiated jets that (i) can be readily applied at the LHC, (ii) has a well-defined connection to pQCD calculations, and (iii) can be easily implemented in a Rivet [91] routine, which greatly improves the ability of measurements based on this method to constrain Monte Carlo generators.

## Acknowledgments

GS and ASO thank the CERN theoretical department for the hospitality at different stages of this work. CB thanks Laboratoire Leprince-Ringuet and Sapienza Università di Roma for the support provided in the development of this work. CB thanks Jacob March for his contributions at the early stages of this work, and Leticia Cunqueiro and Matthew Nguyen for helpful discussions. This work has been supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 788223, PanScales) and by the ERC consolidator programme (grant agreement No. 1010022070, QCDHighdensityCMS). ASO is supported by the Ramón y Cajal program under grant RYC2022-037846-I. CB is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under grant contract number DE-SC0011088.

## A Analytic results for gluon purities in the grooming setup

Here we provide the analytic ingredients that enter the computation of the gluon purities in the collinear limit discussed in Sec. 2.2. We focus on the grooming setup since the results are, at  $\mathcal{O}(\alpha_s)$ , independent of the jet  $p_t$  and can thus be computed fully analytically.

The starting point is the Altarelli-Parisi splitting functions,  $P_{ab}(z)$ , that read:

$$\begin{aligned} P_{qq}(z) &= C_F \frac{1+z^2}{1-z}, & P_{gq}(z) &= C_F \frac{1+(1-z)^2}{z}, \\ P_{qg}(z) &= T_f [z^2 + (1-z)^2], & P_{gg}(z) &= 2C_A \left[ \frac{1-z}{z} + \frac{z}{1-z} + z(1-z) \right], \end{aligned} \quad (\text{A.1})$$

where the standard QCD parameters are  $C_F = 4/3$ ,  $C_A = 3$ ,  $T_R = 1/2$ ,  $T_f = T_R n_f$  and the number of flavours is set to  $n_f = 5$ . The gluon purity, Eq. (2.8), is then obtained by integrating these splitting functions between some  $z_{\min}$  and some  $z_{\max}$  (see Eq. (2.7)). These integrals, that we have denoted  $I_{ab}^{(\text{groom})}$  in the main text, are given by:

$$I_{qq}^{(\text{groom})} = C_F \left[ 2 \ln \frac{1-z_{\min}}{1-z_{\max}} - (z_{\max} - z_{\min}) \left( 1 + \frac{z_{\max} + z_{\min}}{2} \right) \right], \quad (\text{A.2})$$

$$I_{gq}^{(\text{groom})} = 2C_F \left[ \ln \frac{z_{\max}}{z_{\min}} - (z_{\max} - z_{\min}) \left( 1 - \frac{z_{\max} + z_{\min}}{4} \right) \right], \quad (\text{A.3})$$

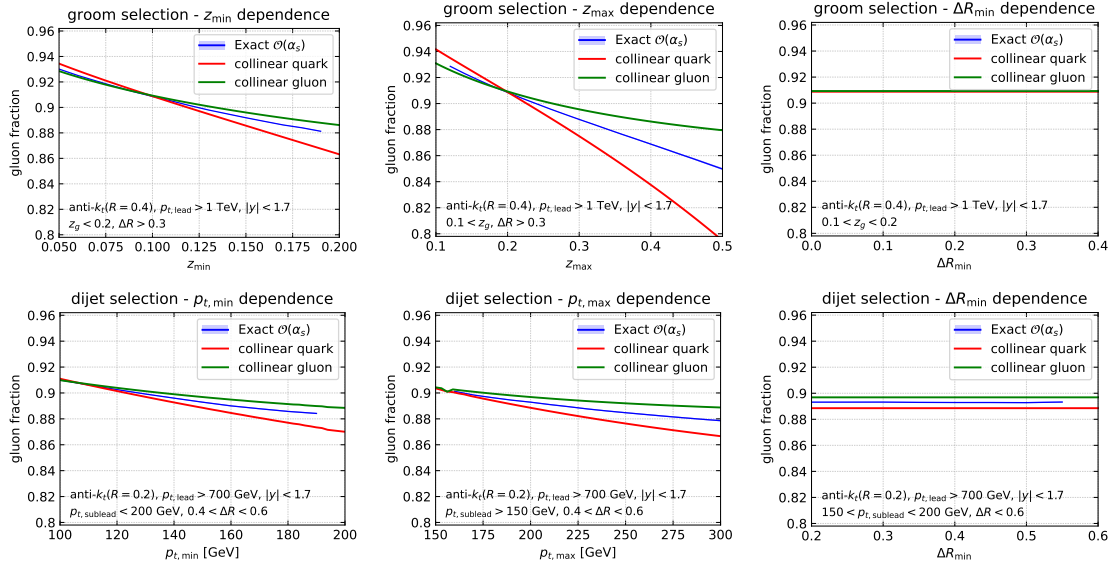
$$I_{qg}^{(\text{groom})} = \frac{2T_f}{3} [z_{\max}^3 - z_{\min}^3 + (1-z_{\min})^3 - (1-z_{\max})^3], \quad (\text{A.4})$$

$$I_{gg}^{(\text{groom})} = 2C_A \left[ \ln \frac{z_{\max}(1-z_{\min})}{z_{\min}(1-z_{\max})} - 2(z_{\max} - z_{\min}) + \frac{z_{\max}^2 - z_{\min}^2}{2} - \frac{z_{\max}^3 - z_{\min}^3}{3} \right] \quad (\text{A.5})$$

where we have assumed that  $z_{\min} < z_{\max}$ .

## B First-principles estimates for the gluon purities using small- $R$ jets

In Fig. 2 we found that the exact  $\mathcal{O}(\alpha_s)$  result for the gluon fraction did not always lie in between the pure quark and pure gluon estimates. This is related to the fact that the



**Figure 7:** Same as Fig. 2 but with a more collinear setup as discussed in the main text.

selected angular parameters, motivated by the experimental feasibility of the measurement, are rather large and thus we expect sizable corrections to the collinear approximation. Here we demonstrate this hypothesis by focusing on a more collinear setup. More concretely, in the groomed case we select anti- $k_t$  jets with  $R = 0.4$  (instead of  $R = 1.2$ ) and fix the minimum opening angle between the prongs  $\Delta_{\min} = 0.3$  (instead of  $\Delta_{\min} = 0.8$ ). In the dijet selection we select anti- $k_t$  jets with  $R = 0.2$  (instead of  $R = 0.4$ ) and the maximum angular separation between the two jets is  $\Delta_{\max} = 0.6$  (instead of  $\Delta_{\max} = 1.2$ ). For simplicity, no scale variations are considered in this case.

The results are presented in Fig. 7. We observe that the gluon fraction remains around 90% in all cases. The parametric dependences of the gluon fraction are also akin to those of Fig. 2. Further, the exact  $\mathcal{O}(\alpha_s)$  results lie partway the pure quark and pure gluon results in all cases thus confirming the validity of the collinear approximation.

## References

- [1] J. M. Campbell, et al., Event generators for high-energy physics experiments, *SciPost Phys.* 16 (5) (2024) 130. [arXiv:2203.11110](#), [doi:10.21468/SciPostPhys.16.5.130](#).
- [2] G. Aad, et al., Jet energy scale and resolution measured in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector, *Eur. Phys. J. C* 81 (8) (2021) 689. [arXiv:2007.02645](#), [doi:10.1140/epjc/s10052-021-09402-3](#).
- [3] V. Khachatryan, et al., Jet energy scale and resolution in the CMS experiment in  $pp$  collisions at 8 TeV, *JINST* 12 (2017) P02014. [arXiv:1607.03663](#), [doi:10.1088/1748-0221/12/02/P02014](#).
- [4] Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016–2018 (2020). URL <https://cds.cern.ch/record/2715872>

- [5] [Dependence of the Jet Energy Scale on the Particle Content of Hadronic Jets in the ATLAS Detector Simulation](#), Tech. rep., CERN, Geneva, all figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-021> (2022).  
URL <https://cds.cern.ch/record/2808016>
- [6] G. Aad, et al., A precise measurement of the jet energy scale derived from single-particle measurements and in situ techniques in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector (7 2024). [arXiv:2407.15627](#).
- [7] G. Aad, et al., Measurements of jet cross-section ratios in 13 TeV proton-proton collisions with ATLAS, Phys. Rev. D 110 (7) (2024) 072019. [arXiv:2405.20206](#),  
[doi:10.1103/PhysRevD.110.072019](#).
- [8] J. R. Andersen, et al., Les Houches 2015: Physics at TeV Colliders Standard Model Working Group Report, in: 9th Les Houches Workshop on Physics at TeV Colliders, 2016.  
[arXiv:1605.04692](#).
- [9] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, J. Thaler, Systematics of quark/gluon tagging, JHEP 07 (2017) 091.  
[arXiv:1704.03878](#), [doi:10.1007/JHEP07\(2017\)091](#).
- [10] G. Aad, et al., Light-quark and gluon jet discrimination in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector, Eur. Phys. J. C 74 (8) (2014) 3023. [arXiv:1405.6583](#),  
[doi:10.1140/epjc/s10052-014-3023-z](#).
- [11] S. Caletti, A. J. Larkoski, S. Marzani, D. Reichelt, Practical jet flavour through NNLO, Eur. Phys. J. C 82 (7) (2022) 632. [arXiv:2205.01109](#), [doi:10.1140/epjc/s10052-022-10568-7](#).
- [12] M. Czakon, A. Mitov, R. Poncelet, Infrared-safe flavoured anti- $k_T$  jets, JHEP 04 (2023) 138.  
[arXiv:2205.11879](#), [doi:10.1007/JHEP04\(2023\)138](#).
- [13] R. Gauld, A. Huss, G. Stagnitto, Flavor Identification of Reconstructed Hadronic Jets, Phys. Rev. Lett. 130 (16) (2023) 161901, [Erratum: Phys.Rev.Lett. 132, 159901 (2024)].  
[arXiv:2208.11138](#), [doi:10.1103/PhysRevLett.130.161901](#).
- [14] Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report.  
[arXiv:1803.07977](#).
- [15] H. Chen, I. Moutl, X. Zhang, H. X. Zhu, Rethinking jets with energy correlators: Tracks, resummation, and analytic continuation, Phys. Rev. D 102 (5) (2020) 054012.  
[arXiv:2004.11381](#), [doi:10.1103/PhysRevD.102.054012](#).
- [16] H. Chen, P. F. Monni, Z. Xu, H. X. Zhu, Scaling Violation in Power Corrections to Energy Correlators from the Light-Ray Operator Product Expansion, Phys. Rev. Lett. 133 (23) (2024) 231901. [arXiv:2406.06668](#), [doi:10.1103/PhysRevLett.133.231901](#).
- [17] A. Hayrapetyan, et al., Measurement of Energy Correlators inside Jets and Determination of the Strong Coupling  $\alpha_S(m_Z)$ , Phys. Rev. Lett. 133 (7) (2024) 071903. [arXiv:2402.13864](#),  
[doi:10.1103/PhysRevLett.133.071903](#).
- [18] K. Lee, A. Pathak, I. Stewart, Z. Sun, Nonperturbative Effects in Energy Correlators: From Characterizing Confinement Transition to Improving  $\alpha_s$  Extraction (5 2024).  
[arXiv:2405.19396](#).
- [19] H. S. Hannesdottir, A. Pathak, M. D. Schwartz, I. W. Stewart, Prospects for strong coupling

measurement at hadron colliders using soft-drop jet mass, JHEP 04 (2023) 087.  
[arXiv:2210.04901](#), [doi:10.1007/JHEP04\(2023\)087](#).

- [20] G. Alexander, et al., A Direct observation of quark - gluon jet differences at LEP, Phys. Lett. B 265 (1991) 462–474. [doi:10.1016/0370-2693\(91\)90082-2](#).
- [21] P. D. Acton, et al., A study of differences between quark and gluon jets using vertex tagging of quark jets, Z. Phys. C 58 (1993) 387. [doi:10.1007/BF01557696](#).
- [22] D. Buskulic, et al., Quark and gluon jet properties in symmetric three jet events, Phys. Lett. B 384 (1996) 353. [doi:10.1016/0370-2693\(96\)00849-0](#).
- [23] R. Akers, et al., A Model independent measurement of quark and gluon jet properties and differences, Z. Phys. C 68 (1995) 179–202. [doi:10.1007/BF01566667](#).
- [24] P. Abreu, et al., Energy dependence of the differences between the quark and gluon jet fragmentation, Z. Phys. C 70 (1996) 179. [doi:10.1007/s002880050095](#).
- [25] K. Nakabayashi, et al., Charged particle multiplicities of quark and gluon jets in  $e^+e^-$  annihilation at TRISTAN, Phys. Lett. B 413 (1997) 447–452. [doi:10.1016/S0370-2693\(97\)01191-X](#).
- [26] G. Abbiendi, et al., Experimental properties of gluon and quark jets from a point source, Eur. Phys. J. C 11 (1999) 217–238. [arXiv:hep-ex/9903027](#), [doi:10.1007/s100520050628](#).
- [27] P. Abreu, et al., Measurement of the gluon fragmentation function and a comparison of the scaling violation in gluon and quark jets, Eur. Phys. J. C 13 (2000) 573–589. [doi:10.1007/s100520050719](#).
- [28] M. S. Alam, et al., Shape studies of quark jets versus gluon jets at  $s^{*(1.2)} = 10\text{-GeV}$ , Phys. Rev. D 46 (1992) 4822–4827. [doi:10.1103/PhysRevD.46.4822](#).
- [29] M. Derrick, et al., Comparison of Charged Particle Multiplicities in Quark and Gluon Jets Produced in  $e^+e^-$  Annihilation at 29-GeV, Phys. Lett. B 165 (1985) 449–453. [doi:10.1016/0370-2693\(85\)91263-8](#).
- [30] D. Buskulic, et al., Study of the subjet structure of quark and gluon jets, Phys. Lett. B 346 (1995) 389–398. [doi:10.1016/0370-2693\(95\)00037-L](#).
- [31] G. Abbiendi, et al., Particle multiplicity of unbiased gluon jets from  $e^+e^-$  three jet events, Eur. Phys. J. C 23 (2002) 597–613. [arXiv:hep-ex/0111013](#), [doi:10.1007/s100520200926](#).
- [32] G. Abbiendi, et al., Scaling violations of quark and gluon jet fragmentation functions in  $e^+e^-$  annihilations at  $s^{*(1/2)} = 91.2\text{-GeV}$  and  $183\text{-GeV}$  to  $209\text{-GeV}$ , Eur. Phys. J. C 37 (1) (2004) 25–47. [arXiv:hep-ex/0404026](#), [doi:10.1140/epjc/s2004-01964-4](#).
- [33] G. Abbiendi, et al., Experimental studies of unbiased gluon jets from  $e^+e^-$  annihilations using the jet boost algorithm, Phys. Rev. D 69 (2004) 032002. [arXiv:hep-ex/0310048](#), [doi:10.1103/PhysRevD.69.032002](#).
- [34] D. Reichelt, P. Richardson, A. Siódmok, Improving the simulation of quark and gluon jets with HERWIG7, Eur. Phys. J. C 77 (2017) 876. [arXiv:1708.01491](#), [doi:10.1140/epjc/s10052-017-5374-8](#).
- [35] J. Mo, F. J. Tackmann, W. J. Waalewijn, A case study of quark-gluon discrimination at NNLL' in comparison to parton showers, Eur. Phys. J. C 77 (11) (2017) 770. [arXiv:1708.00867](#), [doi:10.1140/epjc/s10052-017-5365-9](#).

- [36] G. Soyez, Quark vs. Gluon Jets, in: Parton radiation and fragmentation from LHC to FCC-ee, 2017, pp. 56–60.
- [37] D. d’Enterria, A. Poldaru, G. Wojcik, Measuring the electron Yukawa coupling via resonant s-channel Higgs production at FCC-ee, *Eur. Phys. J. Plus* 137 (2) (2022) 201. [arXiv:2107.02686](#), [doi:10.1140/epjp/s13360-021-02204-2](#).
- [38] J. Gallicchio, M. D. Schwartz, Pure Samples of Quark and Gluon Jets at the LHC, *JHEP* 10 (2011) 103. [arXiv:1104.1175](#), [doi:10.1007/JHEP10\(2011\)103](#).
- [39] G. Aad, et al., Measurement of soft-drop jet observables in  $pp$  collisions with the ATLAS detector at  $\sqrt{s} = 13\text{TeV}$ , *Phys. Rev. D* 101 (2020) 052007. [arXiv:1912.09837](#), [doi:10.1103/PhysRevD.101.052007](#).
- [40] ATLAS Collaboration, Properties of jet fragmentation using charged particles measured with the ATLAS detector in pp collisions at  $\sqrt{s} = 13\text{TeV}$ , *Phys. Rev. D* 100 (2019) 052011. [arXiv:1906.09254](#), [doi:10.1103/PhysRevD.100.052011](#).
- [41] D. Pablos, A. Soto-Ontoso, Pushing forward jet substructure measurements in heavy-ion collisions, *Phys. Rev. D* 107 (9) (2023) 094003. [arXiv:2210.07901](#), [doi:10.1103/PhysRevD.107.094003](#).
- [42] P. Baroň, M. H. Seymour, A. Siódmok, Novel approach to measure quark/gluon jets at the LHC, *Eur. Phys. J. C* 84 (1) (2024) 28. [arXiv:2307.15378](#), [doi:10.1140/epjc/s10052-023-12363-4](#).
- [43] D. Krohn, M. D. Schwartz, T. Lin, W. J. Waalewijn, Jet Charge at the LHC, *Phys. Rev. Lett.* 110 (21) (2013) 212001. [arXiv:1209.2421](#), [doi:10.1103/PhysRevLett.110.212001](#).
- [44] C. Frye, A. J. Larkoski, J. Thaler, K. Zhou, Casimir Meets Poisson: Improved Quark/Gluon Discrimination with Counting Observables, *JHEP* 09 (2017) 083. [arXiv:1704.06266](#), [doi:10.1007/JHEP09\(2017\)083](#).
- [45] Z.-B. Kang, X. Liu, S. Mantry, M. C. Spraker, T. Wilson, Dynamic Jet Charge, *Phys. Rev. D* 103 (7) (2021) 074028. [arXiv:2101.04304](#), [doi:10.1103/PhysRevD.103.074028](#).
- [46] I. W. Stewart, X. Yao, Pure quark and gluon observables in collinear drop, *JHEP* 09 (2022) 120. [arXiv:2203.14980](#), [doi:10.1007/JHEP09\(2022\)120](#).
- [47] P. T. Komiske, E. M. Metodiev, M. D. Schwartz, Deep learning in color: towards automated quark/gluon jet discrimination, *JHEP* 01 (2017) 110. [arXiv:1612.01551](#), [doi:10.1007/JHEP01\(2017\)110](#).
- [48] F. A. Dreyer, H. Qu, Jet tagging in the Lund plane with graph networks, *JHEP* 03 (2021) 052. [arXiv:2012.08526](#), [doi:10.1007/JHEP03\(2021\)052](#).
- [49] M. J. Dolan, A. Ore, Disentangling quark and gluon jets with normalizing flows, *Phys. Rev. D* 107 (11) (2023) 114003. [arXiv:2211.16053](#), [doi:10.1103/PhysRevD.107.114003](#).
- [50] F. A. Dreyer, G. Soyez, A. Takacs, Quarks and gluons in the Lund plane, *JHEP* 08 (2022) 177. [arXiv:2112.09140](#), [doi:10.1007/JHEP08\(2022\)177](#).
- [51] A. J. Larkoski, J. Thaler, W. J. Waalewijn, Gaining (Mutual) Information about Quark/Gluon Discrimination, *JHEP* 11 (2014) 129. [arXiv:1408.3122](#), [doi:10.1007/JHEP11\(2014\)129](#).
- [52] E. M. Metodiev, J. Thaler, Jet Topics: Disentangling Quarks and Gluons at Colliders, *Phys.*

- Rev. Lett. 120 (24) (2018) 241602. [arXiv:1802.00008](#),  
[doi:10.1103/PhysRevLett.120.241602](#).
- [53] P. T. Komiske, E. M. Metodiev, J. Thaler, An operational definition of quark and gluon jets, JHEP 11 (2018) 059. [arXiv:1809.01140](#), [doi:10.1007/JHEP11\(2018\)059](#).
  - [54] M. LeBlanc, B. Nachman, C. Sauer, Going off topics to demix quark and gluon jets in  $\alpha_s$  extractions, JHEP 02 (2023) 150. [arXiv:2206.10642](#), [doi:10.1007/JHEP02\(2023\)150](#).
  - [55] P. T. Komiske, S. Kryhin, J. Thaler, Disentangling quarks and gluons in CMS open data, Phys. Rev. D 106 (9) (2022) 094021. [arXiv:2205.04459](#),  
[doi:10.1103/PhysRevD.106.094021](#).
  - [56] B. Andersson, G. Gustafson, L. Lönnblad, U. Pettersson, Coherence effects in deep inelastic scattering, Z. Phys. C 43 (1989) 625. [doi:10.1007/BF01550942](#).
  - [57] F. A. Dreyer, G. P. Salam, G. Soyez, The Lund jet plane, JHEP 12 (2018) 064.  
[arXiv:1807.04758](#), [doi:10.1007/JHEP12\(2018\)064](#).
  - [58] Y. L. Dokshitzer, G. D. Leder, S. Moretti, B. R. Webber, Better jet clustering algorithms, JHEP 08 (1997) 001. [arXiv:hep-ph/9707323](#), [doi:10.1088/1126-6708/1997/08/001](#).
  - [59] G. Aad, et al., Measurement of the Lund jet plane using charged particles in 13 TeV proton-proton collisions with the ATLAS detector, Phys. Rev. Lett. 124 (2020) 222002.  
[arXiv:2004.03540](#), [doi:10.1103/PhysRevLett.124.222002](#).
  - [60] A. Hayrapetyan, et al., Measurement of the primary Lund jet plane density in proton-proton collisions at  $\sqrt{s} = 13$  TeV, JHEP 05 (2024) 116. [arXiv:2312.16343](#),  
[doi:10.1007/JHEP05\(2024\)116](#).
  - [61] G. Aad, et al., Measurement of the Lund jet plane in hadronic decays of top quarks and W bosons with the ATLAS detector (7 2024). [arXiv:2407.10879](#).
  - [62] A. Lifson, G. P. Salam, G. Soyez, Calculating the primary Lund jet plane density, JHEP 10 (2020) 170. [arXiv:2007.06578](#), [doi:10.1007/JHEP10\(2020\)170](#).
  - [63] R. Medves, A. Soto-Ontoso, G. Soyez, Lund and Cambridge multiplicities for precision physics, JHEP 10 (2022) 156. [arXiv:2205.02861](#), [doi:10.1007/JHEP10\(2022\)156](#).
  - [64] R. Medves, A. Soto-Ontoso, G. Soyez, Lund multiplicity in QCD jets, JHEP 04 (2023) 104.  
[arXiv:2212.05076](#), [doi:10.1007/JHEP04\(2023\)104](#).
  - [65] G. Aad, et al., Measurements of Lund subjet multiplicities in 13 TeV proton-proton collisions with the ATLAS detector, Phys. Lett. B 859 (2024) 139090. [arXiv:2402.13052](#),  
[doi:10.1016/j.physletb.2024.139090](#).
  - [66] [Tagging boosted W bosons applying machine learning to the Lund Jet Plane](#), Tech. rep., CERN, Geneva, all figures including auxiliary figures are available at  
<https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2023-017> (2023).  
URL <https://cds.cern.ch/record/2864131>
  - [67] CMS Collaboration, [Lund Plane Reweighting for Jet Substructure Correction](#) (2023).  
URL <https://cds.cern.ch/record/2866330>
  - [68] M. Cacciari, G. P. Salam, G. Soyez, The anti- $k_t$  jet clustering algorithm, JHEP 04 (2008) 063. [arXiv:0802.1189](#), [doi:10.1088/1126-6708/2008/04/063](#).



- [69] M. Dasgupta, A. Fregoso, S. Marzani, G. P. Salam, Towards an understanding of jet substructure, JHEP 09 (2013) 029. [arXiv:1307.0007](#), [doi:10.1007/JHEP09\(2013\)029](#).
- [70] A. J. Larkoski, S. Marzani, G. Soyez, J. Thaler, Soft drop, JHEP 05 (2014) 146. [arXiv:1402.2657](#), [doi:10.1007/JHEP05\(2014\)146](#).
- [71] D. Krohn, J. Thaler, L.-T. Wang, Jet Trimming, JHEP 02 (2010) 084. [arXiv:0912.1342](#), [doi:10.1007/JHEP02\(2010\)084](#).
- [72] A. M. Sirunyan, et al., Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV, JINST 13 (05) (2018) P05011. [arXiv:1712.07158](#), [doi:10.1088/1748-0221/13/05/P05011](#).
- [73] H. Qu, L. Gouskos, ParticleNet: Jet Tagging via Particle Clouds, Phys. Rev. D 101 (5) (2020) 056019. [arXiv:1902.08570](#), [doi:10.1103/PhysRevD.101.056019](#).
- [74] G. Aad, et al., ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset, Eur. Phys. J. C 83 (7) (2023) 681. [arXiv:2211.16345](#), [doi:10.1140/epjc/s10052-023-11699-1](#).
- [75] G. Aad, et al., Transforming jet flavour tagging at ATLAS (5 2025). [arXiv:2505.19689](#).
- [76] Z. Nagy, Next-to-leading order calculation of three jet observables in hadron hadron collision, Phys. Rev. D 68 (2003) 094002. [arXiv:hep-ph/0307268](#), [doi:10.1103/PhysRevD.68.094002](#).
- [77] R. D. Ball, et al., The PDF4LHC21 combination of global PDF fits for the LHC Run III, J. Phys. G 49 (8) (2022) 080501. [arXiv:2203.05506](#), [doi:10.1088/1361-6471/ac7216](#).
- [78] F. Caola, R. Grabarczyk, M. L. Hutt, G. P. Salam, L. Scyboz, J. Thaler, Flavored jets with exact anti-kt kinematics and tests of infrared and collinear safety, Phys. Rev. D 108 (9) (2023) 094010. [arXiv:2306.07314](#), [doi:10.1103/PhysRevD.108.094010](#).
- [79] C. Bierlich, et al., A comprehensive guide to the physics and usage of PYTHIA 8.3, SciPost Phys. Codeb. 2022 (2022) 8. [arXiv:2203.11601](#), [doi:10.21468/SciPostPhysCodeb.8](#).
- [80] J. Bellm, et al., HERWIG7.2 release note, Eur. Phys. J. C 80 (2020) 452. [arXiv:1912.06509](#), [doi:10.1140/epjc/s10052-020-8011-x](#).
- [81] M. Cacciari, G. P. Salam, Dispelling the  $N^3$  myth for the  $k_t$  jet-finder, Phys. Lett. B 641 (2006) 57–61. [arXiv:hep-ph/0512210](#), [doi:10.1016/j.physletb.2006.08.037](#).
- [82] M. Cacciari, G. P. Salam, G. Soyez, FastJet user manual, Eur. Phys. J. C 72 (2012) 1896. [arXiv:1111.6097](#), [doi:10.1140/epjc/s10052-012-1896-2](#).
- [83] M. Czakon, A. Mitov, R. Poncelet, Next-to-Next-to-Leading Order Study of Three-Jet Production at the LHC, Phys. Rev. Lett. 127 (15) (2021) 152001, [Erratum: Phys.Rev.Lett. 129, 119901 (2022)]. [arXiv:2106.05331](#), [doi:10.1103/PhysRevLett.127.152001](#).
- [84] Y.-T. Chien, R. Kunnawalkam Elayavalli, Probing heavy ion collisions using quark and gluon jet substructure (3 2018). [arXiv:1803.03589](#).
- [85] H. T. Li, I. Vitev, Jet charge modification in dense QCD matter, Phys. Rev. D 101 (2020) 076020. [arXiv:1908.06979](#), [doi:10.1103/PhysRevD.101.076020](#).
- [86] J. Brewer, J. Thaler, A. P. Turner, Data-driven quark and gluon jet modification in heavy-ion collisions, Phys. Rev. C 103 (2) (2021) L021901. [arXiv:2008.08596](#), [doi:10.1103/PhysRevC.103.L021901](#).



- [87] J. Brewer, Q. Brodsky, K. Rajagopal, Disentangling jet modification in jet simulations and in Z+jet data, JHEP 02 (2022) 175. [arXiv:2110.13159](#), [doi:10.1007/JHEP02\(2022\)175](#).
- [88] Y. Ying, J. Brewer, Y. Chen, Y.-J. Lee, Data-driven extraction of the substructure of quark and gluon jets in proton-proton and heavy-ion collisions (4 2022). [arXiv:2204.00641](#).
- [89] A. M. Sirunyan, et al., Measurement of quark- and gluon-like jet fractions using jet charge in PbPb and pp collisions at 5.02 TeV, JHEP 07 (2020) 115. [arXiv:2004.00602](#), [doi:10.1007/JHEP07\(2020\)115](#).
- [90] S.-L. Zhang, E. Wang, H. Xing, B.-W. Zhang, Flavor dependence of jet quenching in heavy-ion collisions from a Bayesian analysis, Phys. Lett. B 850 (2024) 138549. [arXiv:2303.14881](#), [doi:10.1016/j.physletb.2024.138549](#).
- [91] C. Bierlich, et al., Robust Independent Validation of Experiment and Theory: Rivet version 3, SciPost Phys. 8 (2020) 026. [arXiv:1912.05451](#), [doi:10.21468/SciPostPhys.8.2.026](#).