




















# Flavoured jet algorithms: a comparative study

**Arnd Behring** <sup>a</sup>, **Simone Caletti** <sup>b</sup>, **Francesco Giuli** <sup>c</sup>, **Radosław Grabarczyk** <sup>d,e</sup>,  
**Andreas Hinzmann** <sup>f</sup>, **Alexander Huss**<sup>g</sup>, **Joey Huston** <sup>h</sup>, **Ezra D. Lesser** <sup>g</sup>,  
**Simone Marzani** <sup>i</sup>, **Davide Napoletano** <sup>j</sup>, **Rene Poncelet** <sup>k</sup>, **Daniel Reichelt** <sup>g,l</sup>,  
**Alberto Rescia** <sup>f,i</sup>, **Gavin P. Salam** <sup>e,m</sup>, **Ludovic Scyboz** <sup>n</sup>, **Federico Sforza** <sup>i</sup>,  
**Andrzej Siódmok** <sup>g,o</sup>, **Giovanni Stagnitto** <sup>j</sup>, **James Whitehead** <sup>o</sup> and **Ruide Xu** <sup>p</sup>

<sup>a</sup>Max-Planck-Institut für Physik, Boltzmannstraße 8, 85748 Garching, Germany

<sup>b</sup>Institute for Theoretical Physics, ETH, CH-8093 Zürich, Switzerland

<sup>c</sup>Dipartimento di Fisica, Università degli Studi di Roma Tor Vergata and INFN, Sezione di Roma Tor Vergata, Via della Ricerca Scientifica 1, Rome 00133, Italy

<sup>d</sup>Department of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, U.K.

<sup>e</sup>Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, U.K.

<sup>f</sup>Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany

<sup>g</sup>CERN, 1211 Geneva 23, Switzerland

<sup>h</sup>Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, U.S.A.

<sup>i</sup>Dipartimento di Fisica, Università di Genova and INFN, Sezione di Genova, Via Dodecaneso 33, 16146, Italy

<sup>j</sup>Università degli Studi di Milano-Bicocca & INFN, Piazza della Scienza 3, Milano 20126, Italy

<sup>k</sup>Institute of Nuclear Physics, ul. Radzikowskiego 152, 31-342 Krakow, Poland

<sup>l</sup>Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, U.K.

<sup>m</sup>All Souls College, Oxford OX1 4AL, U.K.

<sup>n</sup>School of Physics and Astronomy, Monash University, Wellington Rd, Clayton VIC-3800, Australia

<sup>o</sup>Jagiellonian University, ul. prof. Stanisława Łojasiewicza 11, 30-348 Kraków, Poland

<sup>p</sup>University of Michigan, Ann Arbor, MI 48109, U.S.A.

*E-mail:* [abehring@mpp.mpg.de](mailto:abehring@mpp.mpg.de), [scaletti@phys.ethz.ch](mailto:scaletti@phys.ethz.ch),  
[francesco.giuli@roma2.infn.it](mailto:francesco.giuli@roma2.infn.it), [adoslaw.grabarczyk@physics.ox.ac.uk](mailto:adoslaw.grabarczyk@physics.ox.ac.uk),  
[radoslaw.grabarczyk@physics.ox.ac.uk](mailto:radoslaw.grabarczyk@physics.ox.ac.uk), [andreas.hinzmann@desy.de](mailto:andreas.hinzmann@desy.de),  
[alexander.huss@cern.ch](mailto:alexander.huss@cern.ch), [huston@msu.edu](mailto:huston@msu.edu), [ezra.lessner@cern.ch](mailto:ezra.lessner@cern.ch),  
[simone.marzani@ge.infn.it](mailto:simone.marzani@ge.infn.it), [davide.napoletano@unimib.it](mailto:davide.napoletano@unimib.it),  
[rene.poncelet@ifj.edu.pl](mailto:rene.poncelet@ifj.edu.pl), [d.reichelt@cern.ch](mailto:d.reichelt@cern.ch), [alberto.rescia@ge.infn.it](mailto:alberto.rescia@ge.infn.it),  
[gavin.salam@physics.ox.ac.uk](mailto:gavin.salam@physics.ox.ac.uk), [ludovic.scyboz@monash.edu](mailto:ludovic.scyboz@monash.edu),  
[federico.sforza@ge.infn.it](mailto:federico.sforza@ge.infn.it), [andrzej.siodmok@uj.edu.pl](mailto:andrzej.siodmok@uj.edu.pl),  
[giovanni.stagnitto@unimib.it](mailto:giovanni.stagnitto@unimib.it), [james.whitehead@uj.edu.pl](mailto:james.whitehead@uj.edu.pl),  
[ruidexu@umich.edu](mailto:ruidexu@umich.edu)

**ABSTRACT:** The accurate identification of heavy-flavour jets — those which originate from bottom or charm quarks — is crucial for precision studies of the Standard Model and searches for new physics. However, assigning flavour to jets presents significant challenges, primarily due to issues with infrared and collinear (IRC) safety. This paper aims to address these challenges by evaluating recently-proposed jet algorithms designed to be IRC-safe and applicable in high-precision measurements. We compare these algorithms across benchmark heavy-flavour production processes and kinematic regimes that are relevant for LHC phenomenology. Exploiting both fixed-order calculations in QCD as well as parton shower simulations, we analyse the infrared sensitivity of these new algorithms at different stages of the event evolution and compare to flavour labelling strategies currently adopted by LHC collaborations. The results highlight that, while all algorithms lead to more robust flavour assignments compared to current techniques, they vary in performance depending on the observable and energy regime. The study lays groundwork for robust, flavour-aware jet analyses in current and future collider experiments to maximise the physics potential of experimental data by reducing discrepancies between theoretical and experimental methods.

**KEYWORDS:** Jets and Jet Substructure, Higher-Order Perturbative Calculations, Specific QCD Phenomenology

**ARXIV EPRINT:** [2506.13449](https://arxiv.org/abs/2506.13449)

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Setting the stage</b>	<b>3</b>
2.1	Approaches to jet labelling and tagging	4
2.2	Infrared and collinear safety	8
2.3	Origin of the difficulties of defining flavoured jets	9
2.4	Goals of a jet flavour definition	11
<b>3</b>	<b>A new generation of algorithms</b>	<b>12</b>
3.1	SoftDrop Flavour (SDF)	13
3.2	Flavoured anti- $k_t$ (CMP)	13
3.3	Flavour dressing (GHS)	14
3.4	Interleaved Flavour Neutralisation (IFN)	16
3.5	Winner-Take-All flavour (WTA)	17
3.6	FASTJET implementations	17
<b>4</b>	<b>Findings</b>	<b>18</b>
4.1	Impact of flavour recombination schemes	19
4.2	Fixed-order comparisons	20
4.3	Effects from parton shower and hadronisation	21
4.4	Impact of hadron reconstruction on flavoured jet definitions	25
<b>5</b>	<b>Concluding remarks</b>	<b>28</b>
<b>A</b>	<b>Flavoured jet algorithms in <math>F_{\text{ASTJET}}</math></b>	<b>30</b>
<b>B</b>	<b>NNLO fixed-order <math>pp \rightarrow WH(b\bar{b})</math></b>	<b>33</b>
B.1	Setup of the calculation	33
B.2	Results	34
<b>C</b>	<b><math>pp \rightarrow Z + b/c</math> in central kinematics: NNLO and NLO+PS</b>	<b>36</b>
C.1	Details of the simulations	36
C.2	Results	38
<b>D</b>	<b><math>pp \rightarrow Z + b</math> across kinematic regions</b>	<b>47</b>
<b>E</b>	<b><math>pp \rightarrow Z + c</math> in LHCb kinematics: NNLO and NLO+PS</b>	<b>50</b>
<b>F</b>	<b>LO+PS <math>pp 2 \rightarrow 2</math> QCD with <math>c/b</math> in LHCb kinematics</b>	<b>56</b>
<b>G</b>	<b>ATLAS <math>Z + \text{jets}</math> studies</b>	<b>62</b>
<b>H</b>	<b>Impact on jet flavour tagging in ATLAS</b>	<b>71</b>

---

## 1 Introduction

Jets, i.e. sprays of highly energetic particles, play a vital role in the precision program of particle physics [1, 2]. They are a useful concept because they are well-defined across several energy scales that characterise high-energy scattering processes, and allow for comparisons between theoretical calculations and experimental measurements. Fixed-order calculations derive jets from quarks and gluons that arise directly from the hard interaction. These partonic jets, with relatively low multiplicity, then evolve down to lower energy scales, fragmenting into more partons, a process usually referred to as the *parton shower*. Low-energy partons then eventually form particles that can be seen in the detectors. At the experimental end of the spectrum, jets are derived from tracks and/or calorimeter clusters. Jet definitions which are largely stable under the scale evolution between hard scattering and measurement in a detector capture essential features of the short-distance, high-energy scattering, while being fairly insensitive to soft or collinear emissions. Jet observables can also be defined in ways which are insensitive to details of how quarks and gluons coalesce into hadrons and how those interact with the detector. Progress on both the experimental and theoretical sides has reduced the uncertainties of measurements and improved the precision of predictions. This has led to jet observables becoming indispensable tools for high-precision analyses.

To date, jets at the LHC are most commonly defined using the anti- $k_t$  algorithm [3], which sequentially clusters particles to construct geometrically regular-shaped jets. Anti- $k_t$  jets are particularly resilient against nonperturbative effects such as multiparton interactions (MPI) that give rise to the Underlying Event (UE) and, to some extent, pile-up, which is caused by multiple proton-proton interactions per bunch crossing. While hard interactions can be described with high precision, exploiting perturbative QCD calculations, nonperturbative effects are known with limited precision and partly rely on phenomenological models. Thus, precision next-to-next-to-leading order (NNLO) calculations in QCD have been successfully compared to experimental data at large jet transverse momentum, with relatively small nonperturbative corrections needed [4–11].

Jets are initiated by energetic particles and it is natural to inspect the extent to which they maintain information about those particles through their evolution down to smaller (infrared) energy scales. Jet substructure [12] aims to answer this by characterising the radiation pattern of jets initiated by different particles. In this context, the question of meaningfully distinguishing quark-initiated jets from gluonic ones has been the subject of many dedicated studies, see, e.g., refs. [13–21], including a few that originated at past editions of the Les Houches workshop “Physics at TeV Colliders (PhysTeV)” [22–24]. Furthermore, while the flavour of the hard-scattered quark evolves through the parton shower and fragmentation processes, it carries valuable information about the short-distance physics that produced it.

At high-energy colliders, such as the LHC, it is interesting to study bottom ( $b$ ) and charm ( $c$ ) quark flavour.<sup>1</sup> This is experimentally feasible because heavy-flavoured hadrons (henceforth  $b$  and  $c$  hadrons, respectively) typically decay while passing through the detector, giving rise to recognisable experimental signatures.

---

<sup>1</sup>Jets initiated by top quarks are also widely studied but fall into a different category due to the short lifetime of top quarks; however, recent evidence of a  $t\bar{t}$  state has been observed [25]. Studies for future colliders also consider the possibility of identifying strange quark jets (see, e.g., ref. [26]).

Heavy-flavour jets play a crucial role in many physics analyses. For instance, the Higgs boson decays primarily into two  $b$  quarks ( $H \rightarrow b\bar{b}$ ) which each subsequently fragment into jets, and studying this process plays a central role in understanding the Higgs mechanism. Furthermore, any definitive statement about the presence of an intrinsic heavy-flavour component in the proton [27, 28] requires precision calculations of perturbative cross sections involving heavy quarks. Finally, top quarks predominantly decay into  $W$  bosons and bottom quarks, making the flavour of the  $b$  quark an important signal for identifying decays of top quarks.

It is therefore of great interest to be able to assign a quark flavour to a jet. An intuitive expectation could be that this should correspond to the flavour of the partons present at the beginning of the jet evolution. However, assigning such a flavour label is highly nontrivial. Experimental analyses usually assign a flavour label based on the presence of flavoured hadrons in the vicinity of a jet, i.e., information available at the late stage of evolution. This is theoretically problematic, as flavoured quarks can be introduced to a jet via radiation contributions enhanced by collinear and soft singularities. This weakens the correlation between jet flavour assigned at the hard scattering (based on partons) and the flavour assigned experimentally based on the presence of flavoured hadrons. Moreover, defining jet flavour in fixed-order calculations is complicated by the fact that solely counting the flavour of the constituent partons is not an infrared- and collinear-safe (IRC-safe) procedure, resulting in the noncancellations of infrared singularities between real and virtual corrections. This issue was recognised long ago [29, 30] and addressed thanks to the introduction of an IRC-safe version of the  $k_t$ -algorithm [31, 32], known as the flavour- $k_t$  algorithm.

With the advent of NNLO calculations for jet processes, together with the precision physics program at the LHC, the issue of assigning flavour labels to jets in an IRC-safe manner had to be reconsidered. Given the ubiquity of the anti- $k_t$  jet algorithm in modern analyses, the need for flavoured jet algorithms that maintain the kinematic features of anti- $k_t$  jets arose. Over the past few years, four proposals for novel anti- $k_t$ -like flavoured jet algorithms that are IRC-safe up to NNLO [33] or even to all orders [34–36] have appeared.<sup>2</sup> At the 2023 edition of the Les Houches PhysTeV workshop, these algorithms were compared to each other extensively [39]. The exercise reported here stems from discussions started at Les Houches. It aims to review the behaviour of these new algorithms, compare them in benchmark scenarios and improve the understanding of how they can be used in practice. The hope is that this improved understanding will lead to a better knowledge of heavy-flavour jet final states at the LHC, and will foster follow-up studies, for example at future Les Houches workshops.

## 2 Setting the stage

QCD interactions are flavour-blind, i.e., gluons couple to quarks irrespective of their flavour and, consequently, their mass. Furthermore, the highest precision in perturbative calculations is usually achieved by considering massless quarks. This enhances theoretical degeneracy with respect to flavour. Yet flavour is an indispensable physical signature of experimental final states, easily seen at colliders: hadrons containing  $b$  and  $c$  valence quarks are distinguished by

---

<sup>2</sup>Another approach that has been pursued consists of abandoning collinear safety and introducing a “flavour fragmentation function” calculable to any order [37, 38]; see section 3.5 for details.

their heavy rest masses and long lifetimes relative to strongly-decaying hadrons and vector bosons. This makes them an attractive target for experimental measurements.

Unlike for general jet observables, where parton-hadron duality provides a theoretical motivation for comparing parton-level jets with their experimental counterparts constructed from particle tracks and/or calorimeter clusters, naïve flavour assignment according to parton identity is intrinsically IRC unsafe. This has historically prevented the comparison of experimental measurements of jet flavour with precision calculations. Recently, a number of algorithms have been proposed to resolve this, providing IRC-safe methods for assigning flavour “labels” to jets clustered at parton level. These flavoured jet algorithms differ substantially from conventional experimental procedures, which despite being experimentally well-defined and measurable, are formally IRC unsafe for massless partons or develop large mass-dependent logarithms for massive quarks.

In this section we summarise current approaches to assign jet flavour, the theoretical properties which must be satisfied by any candidate algorithm for it to be calculable within perturbative QCD, and outline possible criteria by which available alternatives may be assessed.

## 2.1 Approaches to jet labelling and tagging

It is useful to first describe the typical way in which theory and experiment are used to compare jet predictions and measurements. The simplified picture of a generic jet analysis is illustrated in a flowchart given in figure 1. This chart displays how information is translated from first-principles theory and from low-level information recorded by the particle detectors through different stages of simulation and experimental analysis to quantities that can be compared directly.<sup>3</sup> In the following text, we use this simple chart as a common frame of reference to define certain terminology which we use throughout this article.

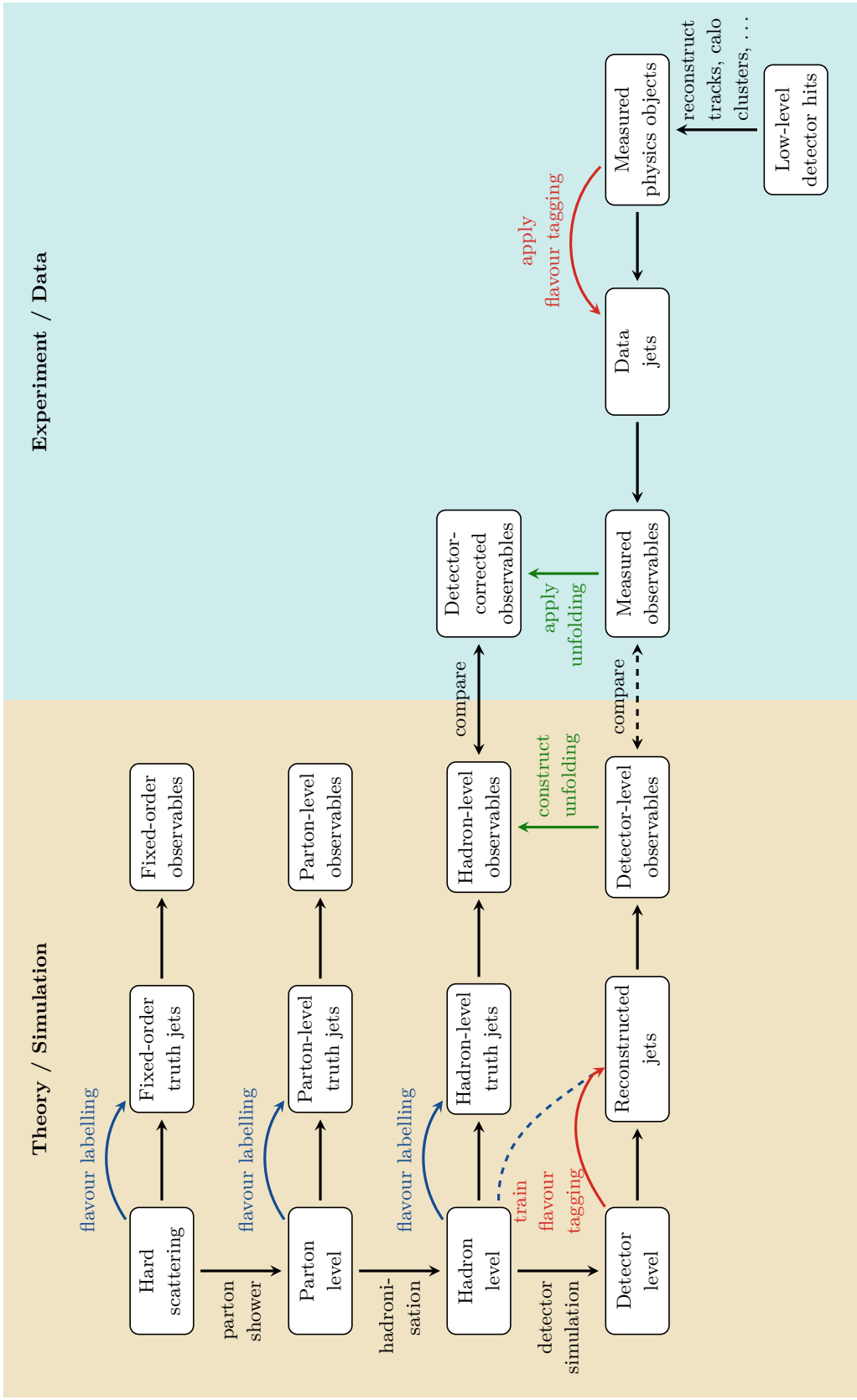
The left-hand side of figure 1 sketches how theory predictions are obtained, starting from the description of events at the level of hard scattering. Events are successively translated to parton level by applying a parton shower, then to the hadron level through hadronisation<sup>4</sup> and finally to the detector level by feeding the hadronic final state through a simulation of an experimental detector’s response. In principle, we can analyse the events at any of these stages, construct jets by applying a jet algorithm to the events and calculate observables such as fiducial cross sections or distributions from these objects. Jets constructed from simulated events at parton or hadron levels are called *truth jets*. From them we can calculate *fixed-order*, *parton-level*, or *hadron-level observables*, respectively. Jets obtained from simulated detector-level events are called *reconstructed jets*, and we calculate *detector-level observables* from them.

On the right-hand side of figure 1, low-level detector objects (tracker hits, energy deposits in calorimeter cells, etc.) are used to reconstruct higher-level physics objects (tracks, topological cell clusters, displaced vertices, ParticleFlow objects, etc.). We call

---

<sup>3</sup>The distinction between theory and experiment on the left- and right-hand sides of the diagram should not imply who is producing these predictions: experimental collaborations frequently use simulation tools to produce predictions, generate corrections, etc. Instead, the idea is to illustrate how information derived from either theory or from measurements is propagated between different levels.

<sup>4</sup>At hadron level one can further distinguish between hadrons produced from hadronising the partonic final state and their decay products.



**Figure 1.** Simplified overview of a generic jet-based analysis, illustrating both the main steps for calculating theoretical predictions and for analysing experimental data. The coloured arrows show the propagation of information about flavour. See the main text for details.

jets constructed from these physics objects *data jets* and observables calculated from them *measured observables*. There is therefore a level of parity between reconstructed jets and data jets, as well as detector-level observables and measured observables, which are at the same level but reconstructed using either simulation or experimental data, respectively.

The predictions for detector-level observables can be compared to measured observables (dashed black arrow in the flowchart), and this is done both in the context of measurements of Standard Model processes and in the context of searches for new physics. But since these quantities are detector specific, *unfolding* is often used to translate detector-specific measurements to hadron level (green arrows in the flowchart), making them more universal. Unfolding requires constructing a response matrix by comparing predictions for hadron-level observables to detector-level observables calculated on the same simulated events.<sup>5</sup> This unfolding procedure is then applied to the measured observables to obtain *detector-corrected observables*. These unfolded results can finally be compared to predictions for hadron-level observables, as well as to results from other experiments.

Given the focus of this study, it is useful to describe how flavour information is propagated both theoretically and experimentally. Here and in the following, we distinguish the procedure of *flavour tagging* (red arrows in the flowchart) from *flavour labelling* (blue arrows). We define flavour tagging as what is used to assign flavour to *reconstructed* or *data* jets using just detector-level information as input. On the other hand, by flavour labelling we refer to the procedure which assigns flavour based on truth-level flavour information available from simulation. Usually, this is applied to *truth* jets, in which case this is sometimes called “truth-level flavour tagging” in the literature. (Sometimes flavour labelling is also applied to reconstructed jets in simulation for the purpose of training flavour tagging algorithms.) While this distinction in terminology is not used universally throughout the literature, we find it useful to keep the concepts separate in this paper.

For flavour labelling, a number of different strategies are used by experimentalists and theorists.

- *Naive/constituent labelling* (also called anti- $k_t$  labelling in this paper): jets are first reconstructed without making reference to flavour. Flavour labels are then assigned based on the flavour of the jet constituents. This method requires the flavoured objects (quarks at parton level or undecayed heavy flavour hadrons at hadron level) to be in the list of inputs to the jet algorithm. It has been used in particular in theory calculations, but it is known to be IRC unsafe at higher orders in perturbation theory, as we will discuss below.
- *Exclusive labelling* [44–47]: this can be thought of as an extension of naive labelling to situations where an exclusive decay mode of a particular flavoured hadron can be fully reconstructed. The flavour of this hadron can be used to assign flavour to a jet. Before jet reconstruction, the decay products of the flavoured hadron are removed from the list of inputs and replaced by the reconstructed hadron. In this case, flavour labelling and flavour tagging can be done using essentially identical procedures (up to standard

---

<sup>5</sup>Recently, methods to unfold events as a whole rather than differential distributions of observables are being studied, e.g., refs. [40–43].

detector corrections to account for particles which were missed or poorly reconstructed). However, experimentally this full-hadron reconstruction is typically only performed in a single decay channel for a single hadron, meaning that the hadronic selection is exclusive and necessarily misses most hadrons of the same flavour.

- *Ghost labelling* [48, 49]: in general, “ghost association” is a method for studying jet clustering in which approximately-zero-momentum “ghost” particles are introduced to an event, to see which jet they are clustered into by the jet algorithm. “Ghost labelling” applies this to flavour labelling by adding prompt, undecayed  $b$  or  $c$  hadrons from the truth record of the event to the input to the jet algorithm as “ghosts”. These ghosts, with momentum scaled by a very small factor, e.g.,  $10^{-20}$ , label the jet they are clustered into with the flavour they are carrying. When ghost labelling is applied at hadron level with undecayed hadrons using the anti- $k_t$  algorithm, the method is equivalent to naive labelling.
- *Cone labelling* [50] (or geometrical matching): after jet reconstruction, the truth record of the generator is searched for the undecayed  $b$  or  $c$  hadrons, potentially above a certain threshold in transverse momentum, and jets in the event are labelled as  $b$  or  $c$  jets if a previously unmatched flavoured hadron is found within a maximum  $\Delta R$  of the jet axis.
- *Flavoured jet algorithms*: dedicated jet algorithms that come with a prescription to assign flavour to the resulting jets in an IRC-safe way. This includes the flavour- $k_t$  algorithm [29, 30] as well as the newly developed ones that we will discuss in more detail below.

Exclusive, ghost and cone labelling are frequently used in experimental analyses,<sup>6</sup> while naive labelling and flavoured jet algorithms have so far been predominantly applied in theoretical calculations.

The flavour of reconstructed or data jets is determined using *flavour tagging* algorithms that rely, depending on the algorithm, on information about physics objects such as tracks or calorimeter clusters or also higher-level derived objects like fully-reconstructed heavy-flavour hadrons (when available), or secondary vertices to infer the flavour. Many early measurements relied mostly on the impact parameter of tracks from decays of flavoured hadrons or the reconstruction of secondary vertices [51–54]. Modern flavour tagging algorithms rely more and more on complex machine learning approaches in order to combine low-level information into powerful discriminants [55–57].

These algorithms must be validated or trained on simulated datasets where reconstructed jets have been labelled in some way. Assigning flavour to reconstructed jets for training or validation (dashed blue arrow) can be accomplished in several ways. One possibility is to use a flavour labelling procedure on hadron-level truth jets and associate these jets to reconstructed jets at detector level on an event-by-event basis using geometric matching between the truth and reconstructed jets. The reconstructed jets then inherit the flavour label from the truth jet. Another possibility is to include undecayed hadrons from the event

---

<sup>6</sup>Given that a  $b$  hadron can decay to  $c$  hadrons, a flavour definition priority must be assigned, with the usual convention that a  $b$ -labelled jet is not further labelled as a  $c$  jet.

record at hadron level as ghosts in the clustering of the event at detector level. Finally, more recent approaches also include matching flavoured hadrons from hadron level to tracks after detector simulation and assigning these tracks to reconstructed jets via ghost association. All of these approaches have in common that flavour labelling procedures influence flavour tagging algorithms via their training.

The tension between experimental and theoretical approaches to jet flavour labelling results in additional systematic uncertainties that hamper comparisons between high-precision calculations and experimental data. In addition, it is well-known that flavour labelling based solely on the presence of flavoured quarks in the partonic jet is not IRC safe [29] for massless flavoured quarks.

## 2.2 Infrared and collinear safety

When calculating the probability of a particular outcome of a measurement, quantum mechanics compels us to treat all degenerate states as equal and to sum over their contributions. Since we cannot distinguish states in which a massless particle splits into two collinear particles or where a massless particle radiates another massless particle with vanishing energy, we have to sum over those indistinguishable contributions even though they have different multiplicities. This is the basic tenet of the Kinoshita-Lee-Nauenberg (KLN) theorem [58, 59], which states that infrared singularities from soft or collinear massless particles that arise in real emission and virtual contributions cancel in “sufficiently inclusive” observables. If the cancellation holds true to all orders in perturbation theory for a given observable, it is called *infrared and collinear safe*.

A necessary (but not sufficient) condition for IRC safety is that observables must be consistent between final states with different multiplicities in singular kinematic regions.<sup>7</sup> More concretely, this means, for example, that an observable must be the same for a configuration with a final state gluon as for a configuration where this gluon has been replaced by a collinear quark-anti-quark pair. If an observable does not obey these consistency conditions for all soft or collinear limits of massless particles the observable is *IRC unsafe*. In particular, this implies that an unsafe observable cannot be computed in fixed-order perturbation theory.

When constructing an observable, one way to ensure IRC safety is to pay attention to whether it agrees across different multiplicities in all soft and collinear limits. Another way is to reduce the list of kinematic limits in which the matrix elements become singular. For example, this can be achieved by considering massive quarks. This regulates collinear singularities involving this quark (giving rise to the well-known dead cone effect [61, 62]). Therefore, if an observable does not agree on configurations in these limits, it is still technically IRC safe. However, the observable does pick up a logarithmic dependence on the ratio of the mass of the quark to a hard scale of the process. If we were to send the quark mass to zero, the logarithm would diverge, recovering the noncancellation of infrared singularities in the massless case. While the observables are IRC safe in this case, they become strongly sensitive to infrared scales, such as the quark mass, and they would diverge if we were to send the quark mass to zero. We call observables that exhibit this behaviour *infrared sensitive*.<sup>8</sup>

Besides introducing a regulator in the underlying theory, we can also make observables IRC safe by suitably excluding singular regions of phase space. Usually, the observables then

<sup>7</sup>For a detailed discussion of this point see, e.g., ref. [60].

<sup>8</sup>Note that some authors would classify infrared-sensitive observables as IRC unsafe. However, we think that making this distinction is useful in the context of this study.

also pick up a logarithmic dependence on the parameter controlling this exclusion. Thus, they become infrared sensitive and depend on low-energy details of the theory. Good examples would be the jet radius, a photon-isolation cut or the cut on transverse momentum in weak boson production. Observables sensitive to these parameters can be computed in fixed-order perturbation theory. However, theoretical predictions are often supplemented with all-order resummation of infrared effects and, potentially, require the inclusion of nonperturbative contributions.

Despite the theoretically motivated definitions of IRC safety and infrared sensitivity, these classifications are also relevant for experimental analyses. While IRC-unsafe observables can be still measured in experiments, the availability of theory predictions largely depends on the observables being IRC safe. Moreover, IRC-safe and infrared-insensitive observables tend to give a better handle on the underlying short-distance physics, whereas infrared-sensitive observables, as the name suggests, are influenced more by long-distance, low-energy details of the theory.

### 2.3 Origin of the difficulties of defining flavoured jets

This brings us to the origin of the challenges in defining flavoured jets in theoretical and experimental situations. Crucially, the object that carries flavour information and, therefore, distinguishes a flavoured from a nonflavoured jet is inherently of infrared origin. For example, at parton level, the mass of the quark is (much) smaller than the energy scales of the labelled jet and thus represents a potential infrared parameter.

Depending on the specifics of the jet-flavour definition, we can face two different situations. If the mass is needed to regulate infrared singularities, perturbative predictions will exhibit logarithmic corrections that need to be resummed. However, the resummation of these contributions is challenging because not all logarithms are of collinear origin and thus not accessible with DGLAP or LL/NLL parton shower methods. Further, such a definition would make strictly fixed-order perturbative computations, which typically achieve the highest precision, more complicated and less reliable due to the aforementioned logarithms. Another way to ensure the predictivity of the theory is to define flavoured jets such that the quark mass is not required as a regulator. This second avenue has the additional advantage of allowing a definition of jet flavour in massless QCD computations that is IRC safe and naturally removes the infrared sensitivity to the mass parameter, i.e., all associated logarithms.

Including flavour assignments in the observable definition can cause the observable to become IRC unsafe. This happens if the flavour assignment can be changed by particle splittings with soft or collinear kinematics. For concreteness, we pick  $b$  flavour as the feature we want to track. As examples, we discuss three cases:

- (i) collinear  $g \rightarrow b\bar{b}$  splitting,
- (ii) collinear  $b \rightarrow bg$  splitting,
- (iii) soft  $g \rightarrow b\bar{b}$  emissions,

which illustrate the problem and point to different solutions that need to be combined.

Scheme	Consider a set of particles flavoured if ...
any flavour	$\sum_i  f_i  > 0$
net flavour	$\sum_i f_i \neq 0$
mod2 flavour	$\sum_i  f_i  \equiv 1 \pmod{2}$

**Table 1.** Summary of the different flavour recombination schemes. The sum goes over all particles  $i$  under consideration whose flavour should be combined. Flavoured particles count as  $f_i = +1$ , flavoured anti-particles count as  $f_i = -1$ . Unflavoured particles count as  $f_i = 0$ .

If we assign flavour to jets based on the presence of flavoured jet constituents (partons or particles), the flavour assignment depends on whether a gluon in the jet splits collinearly into a  $b\bar{b}$  quark pair. Given their collinearity, any sensible jet algorithm will cluster those quarks into the same jet, but whether or not this splitting happens will determine the presence of flavour in the jet. Since a collinear  $g \rightarrow b\bar{b}$  splitting introduces a singularity for massless  $b$  quarks, this makes the observable IRC unsafe. For massive  $b$  quarks, the observable acquires a logarithmic dependence on the quark mass, i.e., it becomes infrared sensitive. To cure this problem, we can change the flavour assignment from being based on the presence of any flavoured jet constituent (*any flavour* scheme) to being based on the net flavour number (*net flavour* scheme): if a gluon splits into a  $b\bar{b}$  pair, the net flavour number does not change. Alternatively, we can also count the number of flavoured jet constituents ( $b$  and  $\bar{b}$  quarks each count positively) and ask for an odd number of flavoured jet constituents (*mod2 flavour* scheme). This might be closer to what is possible in experiments since measuring the charge of the flavoured particles can be challenging and  $B-\bar{B}$  oscillations further blur the picture. Also the “mod2 flavour” number does not change under  $g \rightarrow b\bar{b}$  splitting. Both with the “net flavour” and “mod2 flavour” schemes, a jet with two  $b$ -flavoured jet constituents would have to be regarded as unflavoured. This cures the IRC unsafety or infrared sensitivity that arises from collinear  $g \rightarrow b\bar{b}$  splittings. Since the discussions below will frequently refer to the different flavour recombination schemes, we summarise them in table 1.

The second configuration concerns collinear-gluon emissions from flavoured quarks (i.e.,  $b \rightarrow bg$ ). A hard collinear emission from a flavoured quark might result in a quark that would fail transverse momentum constraints at particle or detector level. In a partonic calculation, if we try to implement this cut directly at parton-level, we will encounter a collinear singularity (in the case of massless quarks) or a mass logarithm for massive ones. In either case, the proper way to account for this is to introduce a fragmentation function that can resum this contribution through DGLAP evolution. This is a well-established formalism and we shall not discuss this further in this work.

Finally, another problem arises for soft  $g \rightarrow b\bar{b}$  splittings. If the splitting happens at wide angles, one of the  $b$  quarks can end up in a hard jet (or both  $b$  quarks can end up in separate jets), and thereby change the flavour assignment in those jets — even if the net or mod2 flavour scheme is applied. Since the gluon is soft, this induces IRC unsafety for massless quarks or infrared sensitivity for massive quarks. To remedy the issue, we have to change the flavour assignment from just being derived from the flavour of the jet constituents, to a jet algorithm that takes flavour into account.

The first such algorithm was the flavour- $k_t$  algorithm [29], where the metric that underlies the clustering was changed in order to prefer early recombinations of soft flavoured quark pairs. Being based on the  $k_t$  algorithm, the resulting jets do not have the same nice properties as jets based on the now ubiquitous anti- $k_t$  algorithm. To solve these issues, different anti- $k_t$ -like jet algorithms have been proposed [33–37]. Each of them introduces a new “flavoured” jet definition with different properties. The details of these algorithms are briefly reviewed in section 3. Most of the novel flavoured jet algorithms need complete and detailed flavour and kinematical information about the events, and this is not achievable given the efficiencies and resolutions of realistic detectors. Therefore, they cannot be used for flavour tagging directly. Instead they are useful for flavour labelling, because flavour labelling is applied at the truth level where flavour information is available from the simulation tools for all final state particles. The study presented here employs the algorithms as flavour labelling strategies. We will briefly comment about possible implementations in experimental analyses in the Conclusions, section 5.

## 2.4 Goals of a jet flavour definition

Given the fundamental differences in the way jet flavour is defined in theoretical and experimental situations, it is obvious that a sensible definition of jet flavour that captures most of the theoretical requirements but is also implementable in experiment will rely on finding common ground. This is very similar to the agreement of using the anti- $k_t$  algorithm for (nonflavoured) jets. Adopting the anti- $k_t$  algorithm, allows us to compare calculations and measurements at various stages of the event evolution and is one of the cornerstones of LHC phenomenology. We would argue that a good definition of flavoured jets would allow similarly consistent comparisons between theory and experiment.

There are several perspectives under which we can evaluate the new flavoured jet algorithms. One open question is how similar these algorithms are to existing flavour labelling procedures that are currently used in experimental analyses and theoretical studies. This is helpful for understanding the size of the effect caused by switching from the current approaches to using one of the new algorithms. The similarity has been a design consideration for recent flavoured jet algorithms. In particular, they all strive to yield jets with kinematics that are as close as possible (if not identical) to nonflavoured anti- $k_t$  jets. This is useful because the flavour tagging algorithms applied to detector-level objects mostly work with reconstructed or data jets that are constructed using the nonflavoured anti- $k_t$  algorithm. Therefore, one could expect that a greater similarity between truth- and reconstructed-level jet flavour could reduce the size of uncertainties introduced by the unfolding procedure. It is tempting to use similarity to current flavour labelling procedures as the primary evaluation criterion. However, as we argue above, current flavour labelling strategies suffer from sensitivity to infrared effects. Using similarity as the sole evaluation criterion runs the risk of perpetuating these problems.

Another interesting aspect is the stability of predictions across different stages of the event description: how much do predictions change when comparing predictions for fixed-order, parton-level and hadron-level observables? Keeping corrections small when going from one level to another is advantageous because it lets us more easily compare predictions from fixed-order calculations, where usually the highest precision is available, to experimental

measurements. Stability under “event evolution” is also tightly linked to infrared (in)sensitivity. Depending on the goals of a given analysis infrared insensitivity may be a major consideration. For example, if we want to study PDFs and fit them with perturbative input, we want to aim for an infrared-insensitive jet flavour definition, i.e., a definition which truthfully represents the short-distance process. If we are interested in the event evolution, for instance studying the dead-cone effect of a heavy quark/hadron, we might want to adopt a more infrared-sensitive definition. This choice also comes with an expectation on the modelling precision and accuracy. An insensitive definition should allow for precise computations within perturbation theory and should be stable under hadronisation and soft-physics modelling. Definitions that are more sensitive to fragmentation, hadronisation and soft-physics modelling allow us to validate or constrain them but of course also will suffer from the associated uncertainties. Over the course of the study, the stability of the predictions turned out to be a useful complementary evaluation criterion.

Further criteria include the computational efficiency of the implementations, their ease of use in existing experimental analysis chains, and the applicability of the algorithms to jet substructure analyses. While these criteria play a role in some parts of the study, they were not the main focus and we leave their evaluation to future work.

As discussed above, the introduction of new flavoured jet algorithms lays the foundations for infrared insensitive flavour labelling. We note that IRC safety of these algorithms has been investigated through high perturbative orders in ref. [36]. However, the aspect of infrared sensitivity and, in particular, the overall dependence of observables on the event evolution is less well studied. In addition, there are no detailed phenomenological comparisons between these algorithms. To fill this gap, the goal of this study is two-fold:

1. compare the phenomenology of the different jet flavour definitions and study their infrared sensitivity, and
2. investigate the impact of these modified jet definitions in experimental analyses at truth level.

### 3 A new generation of algorithms

In this section, we review the main features of the four new flavoured jet algorithms that have been recently introduced. We also describe their implementation in FASTJET [63]. All algorithms are sequential clustering algorithms that are based on the anti- $k_t$  algorithm, which performs two-to-one recombination steps of final state partons or particles. In the following, we use the term “pseudojet” to denote any object that results from the recombination ( $i + j$ ), of particles, or previous pseudojets ( $i$  and  $j$ ), in the clustering sequence. The clustering sequence of the anti- $k_t$  algorithm is determined by a distance measure between the pseudojets ( $d_{ij}$ ) as well as the beam ( $d_{iB}$ ):

$$d_{ij} = \min(k_{t,i}^{-2}, k_{t,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = k_{t,i}^{-2}. \quad (3.1)$$

The algorithm recombines the pair  $i, j$  if their distance is the smallest among all pair- and beam-distances or removes pseudojet  $i$  from the list if its beam distance is the smallest and declares it a jet. The sequence ends if all pseudojets are removed.

### 3.1 SoftDrop Flavour (SDF)

The two principles that guided the construction of the SoftDrop flavour algorithm [33] can be summarised as follows. First, one would like an IRC-safe jet flavour labelling procedure that does not alter the kinematics of the original jet, likely defined with the anti- $k_t$  algorithm. Second, one would like to design a procedure as simple as possible, that makes use of tools and algorithms already used both in phenomenology and by the experimental collaborations.

The idea of using a grooming procedure such as SoftDrop [64] originates from the realisation that the configuration that naive flavour counting IRC unsafe at NNLO is analogous to the one responsible for nonglobal logarithms [65] in jet shape observables. In fact, similar to nonglobal logarithms, it is possible to eliminate the infrared ambiguities of jet flavour grooming. There are, however, two restrictions or modifications to the SoftDrop algorithm that one must implement to ensure IRC safety of jet flavour through NNLO. First, the angular exponent  $\beta$  in the grooming constraint must be strictly positive. Secondly, the JADE clustering algorithm [66, 67] must be used as a reclustering algorithm to properly order and groom soft emissions that can render the jet flavour ambiguous.

Thus, the SoftDrop flavour algorithm proceeds as follows:

1. Recluster the jet with the JADE algorithm which has a metric  $d_{ij}$  corresponding to the invariant mass of  $i$  and  $j$ .
2. At each stage of the clustering, require that particles  $i$  and  $j$  pass the SoftDrop grooming requirement, where:

$$\frac{\min(p_{ti}, p_{tj})}{p_{ti} + p_{tj}} > z_c \left( \frac{\Delta_{ij}}{R_0} \right)^\beta, \tag{3.2}$$

$$\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2},$$

with the initial jet radius  $R_0$ , the energy cut  $z_c$ , and the angular exponent  $\beta > 0$ .

3. If the stage in the clustering passes the grooming requirement, terminate and return the flavour of the jet using a flavour recombination scheme on the jet's constituents that are left. If the grooming requirement fails, then remove the softer of the two branches, and continue to the next stage of the JADE clustering along the harder branch.

The above procedure ensures IRC safety at NNLO, but it is known to fail beyond it.

### 3.2 Flavoured anti- $k_t$ (CMP)

A proposal for a flavoured anti- $k_t$  algorithm has been made in ref. [34], for the purpose of the comparisons in this article dubbed CMP. The guiding principle for achieving IRC safety is similar to the flavour- $k_t$  algorithm, which introduces a modified distance measure used for jet clustering that captures soft and flavoured quark pairs. Starting from the standard anti- $k_t$  distance measure [3], a suppression factor  $\mathcal{S}_{ij}$  is introduced for flavoured quark or pseudojet pairs while all other distances are kept:

$$d_{ij}^{(\text{flavoured})} = d_{ij}^{(\text{standard})} \times \begin{cases} \mathcal{S}_{ij}, & \text{if both } i \text{ and } j \text{ have nonzero flavour of opposite} \\ & \text{sign and magnitude,} \\ 1, & \text{otherwise.} \end{cases} \tag{3.3}$$

where  $\mathcal{S}_{ij}$  is supposed to vanish if the quark pair is soft compared to a hard process scale. While the flavour of a pseudojet might be decided by different flavour recombination schemes, i.e., net or mod2 flavour, the recommendation is to use net flavour. This modification has two implications: 1) it makes the algorithm identical to the anti- $k_t$  algorithm in the absence of flavoured pairs, and 2) it leads to modifications in the kinematics of the clustered jets compared to nonflavoured anti- $k_t$  jets. If  $\mathcal{S}_{ij}$  is properly chosen, it will only modify the distances of pseudojets that are soft with respect to the relevant scales of the process, and therefore, the modification of the clustered jet kinematics is expected to be small. This distinguishes this approach from the other algorithms, which use nonflavoured anti- $k_t$  jets and modify the flavour labels of the constituents.

The original proposal considered the following definition for the soft term:

$$\mathcal{S}_{ij} = 1 - \theta(1 - \kappa_{ij}) \cos\left(\frac{\pi}{2}\kappa_{ij}\right) \quad \text{with} \quad \kappa_{ij} \equiv \frac{1}{a} \frac{k_{t,i}^2 + k_{t,j}^2}{2k_{t,\max}^2}. \quad (3.4)$$

Here  $k_{t,\max}$  is some (possibly dynamic) hard scale relevant to the process. To facilitate variations and the study of the sensitivity to the choice made, the parameter  $a$  is introduced, which is expected to be of  $\mathcal{O}(10^{-1})$ . The usage of  $k_{t,i}$ 's as a measure of the hardness of the quarks is motivated by boost invariance along the collision axis but leads to an unwanted suppression in the case of double collinear limits to the two incoming hadrons (i.e., each quark becomes collinear to a different initial-state parton). This has been pointed out in ref. [36] together with a possible solution to this issue. Originally, the proposed solution involved the replacement (see discussion in section 3.4)

$$\mathcal{S}_{ij} \rightarrow \bar{\mathcal{S}}_{ij} = \mathcal{S}_{ij} \frac{\Omega_{ij}^2}{\Delta R^2}, \quad \text{where} \quad \Omega_{ij}^2 = 2 \left[ \frac{1}{\omega^2} (\cosh(\omega \Delta y_{ij}) - 1) - (\cos \Delta \phi_{ij} - 1) \right]. \quad (3.5)$$

However, it turns out that this replacement leads to undesirable deviations between the plain anti- $k_t$  algorithm and the CMP algorithm in regions of phase space where modifications are neither necessary nor expected. Instead, we found that the replacement<sup>9</sup>

$$\kappa_{ij} \rightarrow \bar{\kappa}_{ij} = \kappa_{ij} \sqrt{\frac{\Omega_{ij}^2}{\Delta R^2}} \quad (3.6)$$

shows better behaviour. We describe the observations and reasoning in more detail in section B. We denote the variant of the CMP algorithm with the replacement in eq. (3.6) as CMP $\Omega$  and employ this variant throughout this study. Unless stated explicitly, the label CMP also refers to this variant in the rest of the article.

To minimise the impact on the jet kinematics through the modified clustering, it is advised to consider only the flavour of interest, i.e.,  $b$  quarks if  $b$  jets are to be studied, as “flavoured” and not to use the modification for several partonic flavours at once.

### 3.3 Flavour dressing (GHS)

The flavour dressing algorithm [35] aims at a flavour assignment that is IRC safe in perturbation theory and can be combined with any definition of a jet. For the purpose of this

---

<sup>9</sup>Such a modification was considered as an alternative to eq. (3.5) to restore IRC safety in the code underlying the numerical tests of ref. [36], as well.

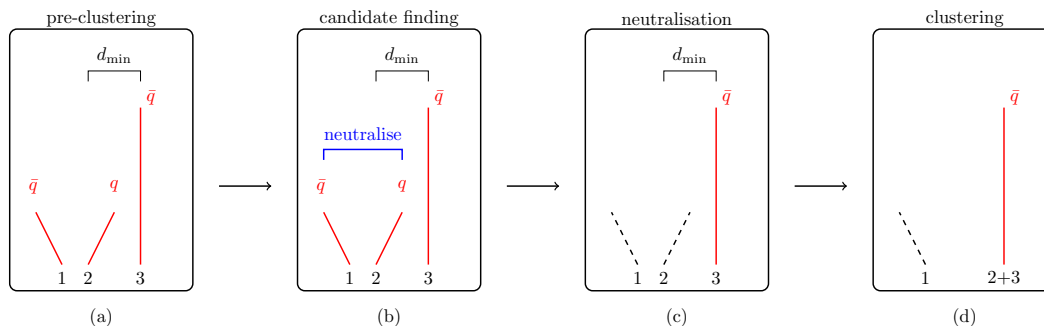
study it will be dubbed GHS. In this algorithm, flavour assignment is *factorised* from jet reconstruction: in a first step, flavour-agnostic jets are obtained with an IRC-safe jet algorithm (such as anti- $k_t$ ); in a second step, jets are *dressed* with flavour quantum numbers according to flavour information about the event (e.g., flavoured particles in parton-level predictions or stable heavy-flavour hadrons in hadron-level predictions). The set of flavour-agnostic jets is considered to be composed of truth or reconstructed jets that have passed a fiducial selection criterion, such as a minimum transverse momentum requirement,  $p_{t,\text{cut}}$ . The flavour dressing algorithm then proceeds as follows:

1. given the set of all particles in the event (that can be identified with the same input that enters the flavour-agnostic jet algorithm), populate a set of distances:
  - if the particle  $p_i$  is associated to jet  $j_k$ , add the distance measure  $d_{p_i j_k}$ . By *association*, in parton-level predictions, we mean that a particle is a constituent of a jet (other options are possible, see ref. [35]). In a hadron collider environment, the beam distances  $d_{p_i B_\pm}$  should be added if  $p_i$  is not associated to any jet;
  - for each unordered pair of particles  $p_i$  and  $p_j$ , add the distance measure  $d_{p_i p_j}$  if either *both particles are flavoured* or *at least one particle is unflavoured and  $p_i$  and  $p_j$  are associated with the same jet* (the latter requirement guarantees collinear flavour safety when recombining particles close to the kinematic boundary of a jet);
2. select the pairing with the smallest distance measure:
  - (a)  $d_{p_i p_j}$  is the smallest: the two particles merge into a new particle  $k_{ij}$  carrying the sum of the four-momenta and flavour. All entries that involve  $p_i$  or  $p_j$  are removed and new distances for  $k_{ij}$  are added.
  - (b)  $d_{p_i j_k}$  is the smallest: assign the particle  $p_i$  to the jet  $j_k$  and remove all entries that involve  $p_i$ .
  - (c)  $d_{p_i B_\pm}$  is the smallest: discard particle  $p_i$  and remove all entries that involve  $p_i$ .
3. iterate until there are no more entries in the set of distances;
4. the flavour assignment for jet  $j_k$  is determined according to the accumulated flavours: one can either perform a net summation of flavours or mod2 summation.

The distances adopted are inspired by the flavour- $k_t$  algorithm [29], with the generalisations of ref. [36]. For the distance  $d_{ab}$  between any two final-state objects (particles or jets), we adopt eqs. (3.8) and (3.9). For the distance between a particle  $p_i$  and beam  $B_\pm$  in direction of positive/negative rapidity, we adopt:

$$\begin{aligned}
 d_{p_i B_\pm} &= \max(p_{ti}^\alpha, p_{tB_\pm}^\alpha(y_i)) \min(p_{ti}^{2-\alpha}, p_{tB_\pm}^{2-\alpha}(y_i)), \\
 p_{tB_\pm}(y) &= \sum_{k=1}^m p_{t,j_k} \left[ \Theta(\pm \Delta y_{j_k}) + \Theta(\mp \Delta y_{j_k}) e^{\pm \Delta y_{j_k}} \right], \tag{3.7}
 \end{aligned}$$

with  $\Delta y_{j_k} = y_{j_k} - y$  and  $\Theta(0) = \frac{1}{2}$ . Hence, the flavour dressing algorithm depends on two parameters,  $\alpha$  and  $\omega$ , the same parameters entering the IFN algorithm (see eqs. (3.8) and (3.9)), with default choice  $(\alpha, \omega) = (1, 2)$ .



**Figure 2.** Illustration of the IFN algorithm (see main text for description).

### 3.4 Interleaved Flavour Neutralisation (IFN)

Interleaved Flavour Neutralisation (IFN) [36] takes a hybrid approach with respect to the flavoured jet algorithms mentioned above. It retains the set of distances from a given underlying generalised- $k_t$  algorithm (such as anti- $k_t$ , or Cambridge/Aachen), but allows the possibility of “neutralising” flavour at each stage of the kinematic clustering.

The IFN is triggered during the clustering of pseudojets  $i$  and  $j$  whenever the softer pseudojet,  $p_{ti} < p_{tj}$ , carries flavour. Before performing the clustering step, the algorithm runs a global search for potential neutralisation candidates among a list  $L$  of all current, flavoured pseudojets,  $k \in L$ . The decision to neutralise flavour within  $i$  from flavour in  $k$  is based on a separate set of distances,  $u_{ik}$ , which are taken to be

$$u_{ik} = \max(p_{ti}, p_{tk})^\alpha \min(p_{ti}, p_{tk})^{2-\alpha} \Omega_{ik}^2, \tag{3.8}$$

$$\Omega_{ik}^2 = 2 \left[ \frac{1}{\omega^2} (\cosh(\omega \Delta y_{ik}) - 1) - (\cos \Delta \phi_{ik} - 1) \right], \tag{3.9}$$

with  $\Delta y_{ik} = y_i - y_k$  and  $\Delta \phi_{ik} = \phi_i - \phi_k$ , and parameters  $0 < \alpha \leq 2$  and  $\omega$ .

Figure 2 shows a schematic representation of the IFN algorithm for the typical IRC-unsafe configuration at NNLO, with a soft gluon splitting to a quark-antiquark pair (numbered 1 and 2), and a hard jet (here a  $\bar{q}$ , numbered 3). (a) The anti- $k_t$  algorithm is about to cluster the soft quark 2 with the hard jet 3, which would result in a change of flavour in the hard jet. (b) Instead, with the IFN, a neutralisation candidate search is triggered (as the softer recombining pseudojet is flavoured). The algorithm finds a neutralisation candidate, jet 1 for the soft quark 2. (c) In this particular case, the algorithm decides to neutralise the flavours of 1 ( $\bar{q}$ ) and 2 ( $q$ ). From this point onwards, pseudojets 1 and 2 have had their flavour neutralised, and are to be considered flavourless (as indicated by the black dashed lines). (d) As there is no more flavour to neutralise globally in the event, the clustering step is now allowed to proceed (with net flavour summation). Pseudojet 2, which is now flavourless, recombines with pseudojet 3 and the final flavour of the jet 2 + 3 is  $\bar{q}$ .

The crucial point is that the kinematics of the underlying anti- $k_t$  jets are left unchanged (the same pseudojets are combined at each clustering step). The (flavour) IRC unsafety of the typical NNLO configuration is lifted, because soft quark-antiquark pairs coming from a gluon splitting are neutralised before their flavour can contaminate harder jets.

The form of the  $u_{ik}$  distances is critical. In particular the angular component,  $\Omega_{ik}^2$ , is designed to eliminate IRC divergences stemming from the interplay of initial-state radiation with soft flavour at large angle (similar configurations plague other flavoured algorithms, and IRC unsafety issues are resolved by the same  $\Omega_{ik}^2$  choice, instead of the standard  $\Delta R_{ik}^2$  angular factor). In the IFN, IRC safety requires a recursive definition of the neutralisation search, as well as

$$\omega > 2 - \alpha \tag{3.10}$$

for the choice of the algorithm’s parameters. We usually choose  $(\alpha, \omega) = (1, 2)$  or  $(2, 1)$ .

### 3.5 Winner-Take-All flavour (WTA)

An alternative approach to defining jet flavour, first presented in ref. [37], uses the Winner-Take-All (WTA) axis [68, 69] to define jet flavour in the infrared limit. The procedure is as follows:

1. reconstruct the jet using any desired jet reconstruction algorithm, for example standard anti- $k_t$  using  $E$ -scheme recombination and any desired resolution (radius) parameter;
2. recluster the jet using any IRC-safe jet clustering algorithm and recursively follow the hardest splitting until reaching the outermost branch in the reclustered jet tree (WTA axis);<sup>10</sup>
3. define the flavour of the jet by the sum of flavours which are collinear along the resulting WTA axis.

Since the WTA axis is completely insensitive to soft radiation, defining jet flavour using the WTA algorithm is also infrared safe. However, the WTA flavour definition is not collinear safe. This collinear unsafety is handled by introducing a “flavour fragmentation function,” which is calculable to any perturbative order [38].

Due to its collinear unsafety and its fundamentally different approach from the four IRC-safe algorithms described above, this algorithm is not included in the majority of studies reported in this article. However, the algorithm also presents experimental benefits in the case when a subset of heavy-flavoured hadrons can be fully reconstructed experimentally, since the flavour definition does not require all flavoured particles in the jet to be known, only whether or not a single flavoured particle is collinear with the WTA axis. Therefore, some experimentally-motivated studies are performed and their findings are presented below in section 4, as well as in the corresponding section F.

### 3.6 FASTJET implementations

The four IRC-safe flavoured jet algorithms presented above are available in the form of FASTJET [63] plugins (SDFPlugin, CMPPlugin, GHSAlgo, and IFNPlugin respectively), and can be found as part of the FASTJET contrib repository at

---

<sup>10</sup>This follows the original formulation of the algorithm [37] whereas more recent developments that focus on high-accuracy resummation [38, 70] explicitly adopt the  $k_t$  algorithm for the reclustering step.

<https://fastjet.hepforge.org/contrib/>.

To compile the plugins one needs `fastjet-config` in the environment path. (Due to its comparatively simple implementation using existing FASTJET tools, namely the absence of a new flavour reclustering scheme, no discrete plugin for WTA flavour is provided.) The flavour machinery relies on FASTJET’s `PseudoJet::UserInfo` extra user information holder, which we initialise with a `FlavInfo` class instance (or the more generic class `FlavHistory`, for cases where one is interested in the flavour information at different stages of the clustering), which is part of the `IFNPlugin` installation. The `IFNPlugin` requires at least FASTJET v. 3.4.1. The `SDFPlugin`, `CMPPPlugin` and `GHSAlgo` packages all depend on the `IFNPlugin` to handle flavour information, through the already mentioned flavour infrastructure. Furthermore, the `SDFPlugin` also requires the `RecursiveTools` [64, 71, 72] library to be installed since it exploits its `SoftDrop` and `JADE` clustering implementations.

An example file (`example.cc`) is provided in each algorithm’s directory, which demonstrates how to set up the `JetDefinition` and run the jet clustering on a given event (an input vector of `PseudoJet` objects). The main steps are summarised for each algorithm in section A. The example files show how to set up the flavour information of the input particles, and access the flavour of the resulting jets. An input particle’s extra `UserInfo` can be initialised, e.g., by its PDG id,

```
PseudoJet particle(px, py, pz, E);
particle.set_user_info(new FlavHistory(pdg_id));
event.push_back(particle);
```

and the flavour of the final recombined jet can be accessed by

```
vector<PseudoJet> jets = jet_def(event);
for (const auto &j : jets) {
    string flav = FlavHistory::current_flavour_of(j).description();
    cout << "Jet flavour: " << flav << endl;
}
```

to print each jet’s flavour in human-readable format, or to label flavour in a jet according to the FASTJET setting (“net” or “mod2”, set through the `FlavRecombiner` class):

```
bool is_blabelled = FlavHistory::current_flavour_of(j)[5] != 0;
```

## 4 Findings

The main objective of this study is to compare the algorithms presented in various LHC-based scenarios and to provide a comprehensive discussion of the phenomenological differences. To this end, individual studies for different final states or phase-space regions have been performed. The details of the setup and results for each individual study are collected in dedicated appendices. In this section, we condense the most relevant findings to provide the reader with a compact overview. We refer interested readers to the extensive discussions in sections B–G. We will see that, despite the different kinematic configurations, selection procedures, and details in each analysis, a coherent picture emerges.

All algorithms we compare introduce new parameters. We identify the following baseline setup for the new algorithms:

- SDF: the SoftDrop algorithm described in section 3.1 with parameters  $z_c = 0.1$  and  $\beta = 1$ , or  $\beta = 2$  (where the value of  $\beta$  is indicated explicitly in the various studies).
- CMP $\Omega$ : the flavour anti- $k_t$  algorithm described in section 3.2 with parameters  $a = 0.1$ ,  $\omega = 2$  and the choice of  $k_{t,\max}$  as the hardest transverse momentum of a pseudojet at each clustering step.
- GHS: the flavour dressing algorithm described in section 3.3 with  $\alpha = 1$ ,  $\omega = 2$  and  $p_{t,\text{cut}} = 15 \text{ GeV}$ .<sup>11</sup>
- IFN: the interleaved flavour neutralisation algorithm from section 3.4 with  $\alpha = 2$  and  $\omega = 1$ .

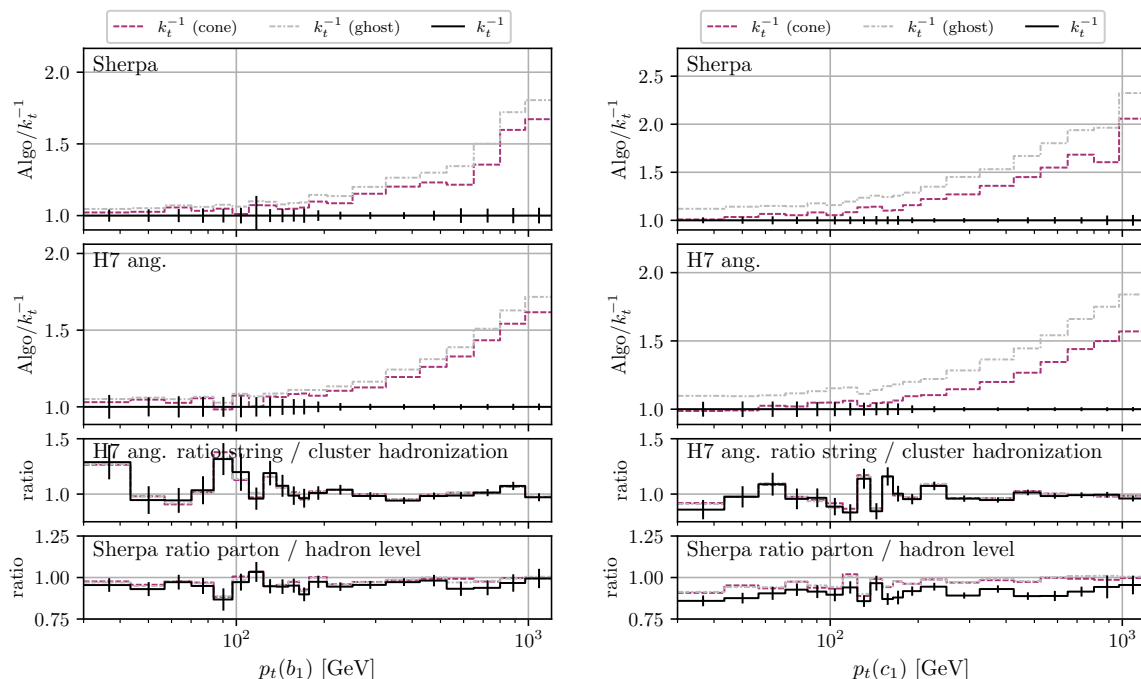
For all algorithms, we choose either the “mod2” or “net-flavour” recombination scheme. In addition to flavoured jet algorithms, we show results for the plain anti- $k_t$  algorithm equipped with the “mod2-flavour” recombination scheme in parton-shower computations. This algorithm is infrared sensitive due to the missing treatment of soft  $b\bar{b}$  pairs, but we use it to gauge the impact of the corresponding modifications in the flavoured jet algorithms.

#### 4.1 Impact of flavour recombination schemes

Before discussing the new flavoured jet algorithms, we point out the impact of the flavour recombination scheme. As discussed above, experimental analyses often correspond to the “any-flavour” scheme, where one or more flavoured particles with the correct kinematical properties are required to assign the jet flavour. To ensure IRC safety from NLO QCD onwards, “mod2-flavour” or “net-flavour” must be employed as discussed in section 2. The phenomenological impact of these schemes is profound because they remove the flavour that arises from the splitting of gluons ( $g \rightarrow b\bar{b}$  or  $g \rightarrow c\bar{c}$ ) within a jet. To assess the size of this effect, we compare the anti- $k_t$  algorithm using the “mod2-flavour” scheme to the anti- $k_t$  algorithm with cone and ghost labelling methods used in some experimental analyses for central  $Z + \text{jet}$  production (see section C for details). In the simulation, the decay of all  $b$  and  $c$  hadrons is deactivated, and the anti- $k_t$  algorithm with the labelling of the “any-flavour” scheme would be the same as ghost labelling. In figure 3 we present the transverse momentum spectrum for  $b$ - and  $c$ -labelled jets. Focusing on the second and third panels from the top, we see that the three methods agree at low  $p_t$ , but deviate increasingly at high  $p_t$ . The effect at a  $p_t$  of about 300 GeV is  $\sim 20\%$  and reaches up to 50% at 1 TeV. The interpretation of this observation is that the cone and ghost labelling strategies are picking up additional  $b/c$ -labelled jets from  $g \rightarrow b\bar{b}$  splittings and are therefore infrared sensitive. We want to stress that this observation is well known (see, for example, refs. [73, 74]), and phenomenological analyses typically take these into account by Monte Carlo and unfolding corrections. These corrections are however inherently infrared sensitive since they are computed with algorithms that are not infrared flavour safe.

---

<sup>11</sup>Natural units with  $\hbar = c = 1$  are used throughout.



**Figure 3.** NLO+PS predictions from SHERPA and HERWIG for  $pp \rightarrow Z + b$  (left) and  $pp \rightarrow Z + c$  (right) in central kinematics. Different experimentally inspired jet labelling algorithms are compared to the anti- $k_t$  algorithms with mod2 recombination for the transverse momentum of the leading  $b/c$  jet. Additionally, differences between the cluster and string hadronisation model for HERWIG, as well as parton and hadron level for SHERPA, are shown.

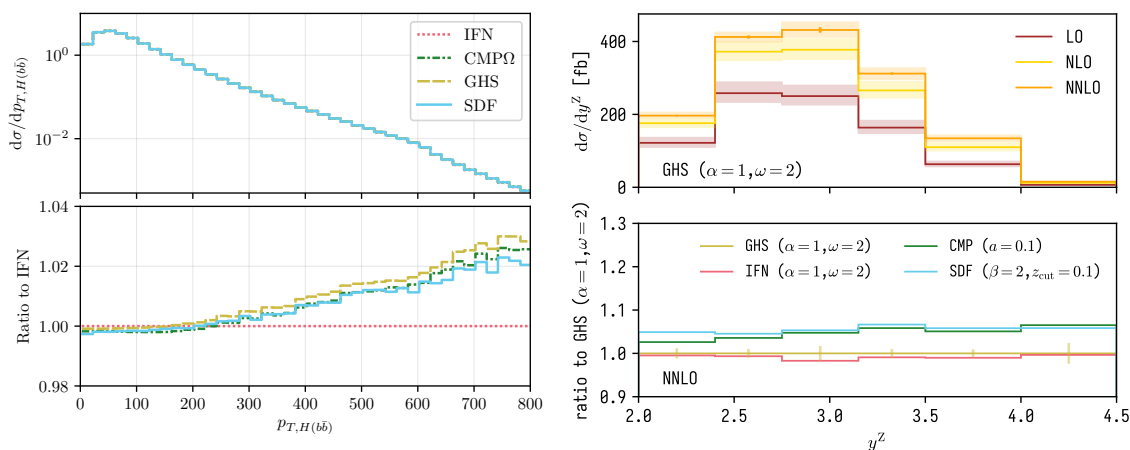
Differences between “net-flavour” and “mod2-flavour” arise due to different flavour assignments in the case of jets with  $b\bar{b}$  or  $\bar{b}b$  content. Phenomenologically, these effects are comparatively small, see for example the discussion in section C, because such configurations do not correspond to infrared-enhanced regions of the cross section.

## 4.2 Fixed-order comparisons

The algorithms can be compared at different stages of the event evolution: the hard process (i.e. in fixed-order computations), the parton level, or the particle level (i.e. in matched parton shower predictions with and without hadronisation). We will first discuss the scenarios accurately describing the highest energy scales, i.e., fixed-order predictions, and move then to the more complete simulations of the event evolution to lower scales. This will allow us to judge the stability of the flavour assignment across energy scales.

Fixed-order computations have been performed through NNLO QCD for the final states  $H(\rightarrow b\bar{b})W$  and  $Z + b/c$ -jet. Details of  $HW$  studies are collected in section B. The  $Z + b/c$ -jet final state has instead been studied in a central phase space selection, inspired by ATLAS and CMS analyses, and at forward rapidities, as appropriate for LHCb. Details on the phase-space selection cuts can be found in section C and section E.

We start with the  $HW$  final state and show on the left-hand side of figure 4 the transverse momentum spectrum of the Higgs boson reconstructed from two identified  $b$  jets. The upper



**Figure 4.** Comparison of the flavoured jet algorithms at fixed-order. The  $p_t$  spectrum of the Higgs boson, reconstructed from two  $b$ -labelled jets in  $pp \rightarrow WH$  production is shown on the left. The right shows the  $Z$  rapidity spectrum in a LHCb  $Z + c$  selection.

panel shows the absolute spectra, while the lower panel shows the ratio to the IFN algorithm. Generally, the predictions agree with each other across the phase space, with only a slight shape difference between IFN and all the other flavoured algorithms that reaches 2% at 800 GeV. Noticeably, similar observations hold for a rather different process, with different selection cuts, namely the  $Z + c$ -jet final state in the LHCb kinematic regions. Here, we show on the right-hand side of figure 4 the rapidity spectrum of the  $Z$  boson. We find qualitative agreement between the different algorithms, with some shape differences in the forward region between GHS/IFN and SDF/CMP. The same holds for central phase space selections, see section C.

From the fixed-order studies, we can conclude that the differences produced by the algorithm on NNLO QCD calculations are largely negligible, and the predictions are consistent. The low sensitivity to the jet algorithms in these fixed-order scenarios originates from the low flavour multiplicity. Up to NNLO QCD, there are at most two additional ( $b$  or  $c$ ) flavoured partons in the final state on which the algorithms act. The corresponding channels are also colour-suppressed compared to the same process with  $g \rightarrow gg$  splitting, and therefore have small numerical contributions to the cross sections. Although not explicitly shown here, at NLO QCD, all algorithms almost identically reproduce the plain anti- $k_t$  algorithm (equipped with “mod2-flavour” recombination for IRC-safety).

### 4.3 Effects from parton shower and hadronisation

The phenomenological differences become more apparent when looking at full Monte Carlo predictions that include the parton shower and nonperturbative effects. To guide this discussion, we consider studies performed for vector boson plus jet final states, in the central phase-space region. We also study the impact of labelling schemes in the development of flavour tagging algorithms.

In figure 5, we summarise the findings of the  $Z + \text{jet}$  studies in the central phase space. We focus on the transverse momentum spectrum of  $b$ -labelled jets (left-hand side) and  $c$ -

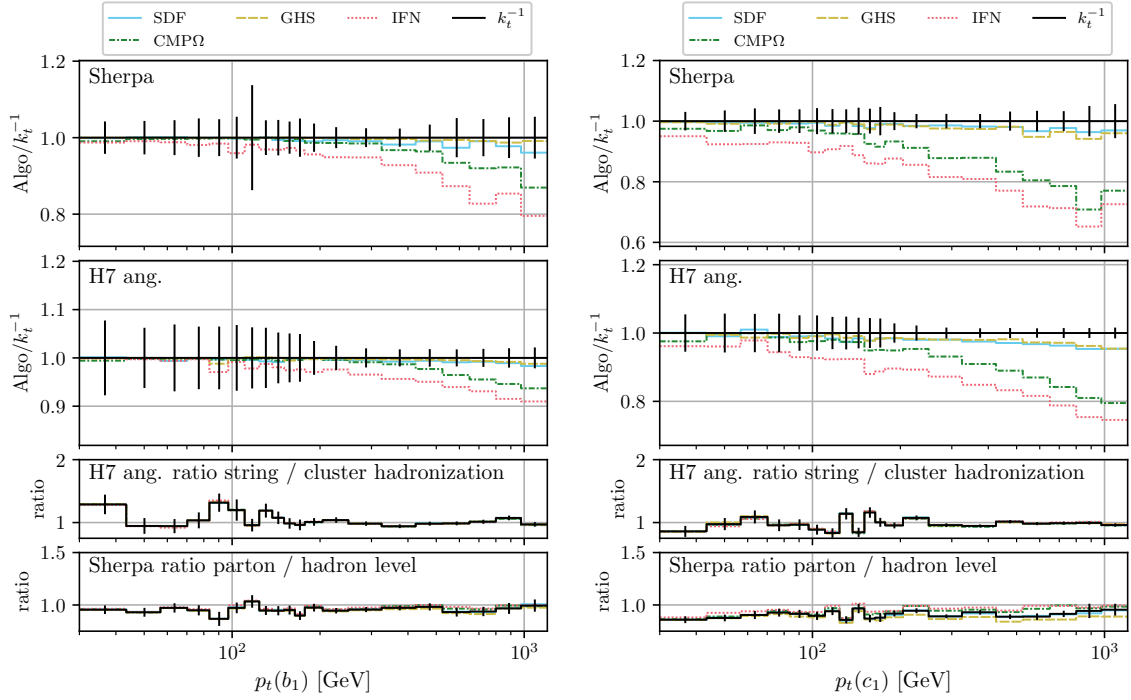
labelled jets (right-hand side), but similar conclusions essentially hold for all energy-dependent observables. The statistical uncertainties are in some cases large compared to the observed differences between the algorithms. Still, the observed differences are meaningful since the analysis has been performed on the same sample, which leads to a strong statistical correlation.

The first two panels from the top in figure 5 compare different jet algorithms with respect to the anti- $k_t$  algorithm (with “mod2-flavour” recombination) at the particle level for SHERPA and HERWIG, respectively. The differences between the flavoured jet algorithms are smaller than what we observed for the flavour-recombination-scheme dependence, but also become more pronounced for high transverse momentum. The SDF and GHS algorithms reproduce the plain anti- $k_t$  behaviour throughout the spectrum, while the CMP and IFN algorithms give reduced rates at high transverse momentum. The effect is more pronounced for the  $c$ -label case, i.e., the case of smaller mass, for which we also expect a larger flavour multiplicity (i.e., number of flavoured quarks/hadrons) in the final state, see section C.

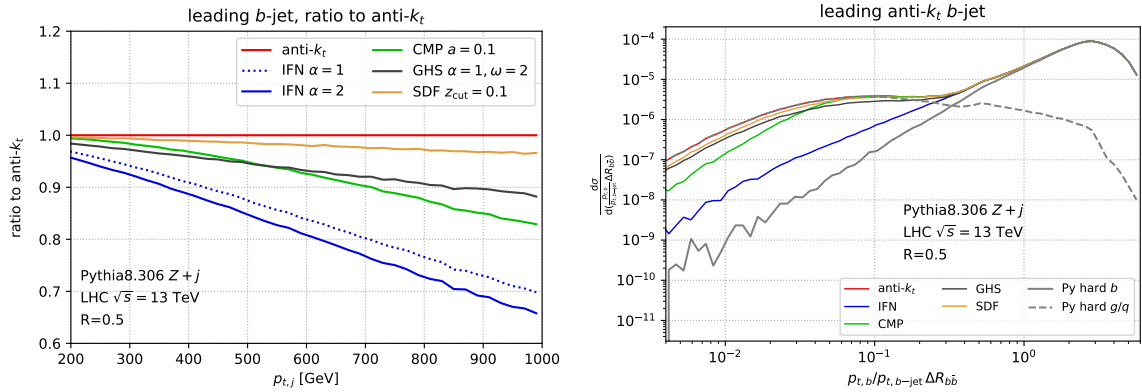
In the lowest two panels in figure 5, we show the ratio of two different hadronisation models implemented in HERWIG and the ratio between the hadron and parton level for SHERPA, for the various algorithms. The difference between the hadronisation models is very minor and within the statistical errors. The differences between parton and hadron levels for the  $b$ -label case are also rather small, as are the differences between the responses of the jet algorithms. However, for the  $c$ -label case, the hadronisation effects are more significant, in particular for a small transverse momentum. The algorithms’ responses are comparable. Still, differences arise again in the tail of the distribution, where we can observe that the CMP, IFN, and experimental labelling techniques for anti- $k_t$  jets lead to smaller hadronisation corrections than the rest.

Before we interpret these results, it is instructive to view the problem from the perspective that the most robust algorithm is the one that most reliably reproduces the correct flavour of the hard-scattering process after shower evolution. This has been studied in detail for  $Z + \text{jet}$  events in section D. First, this study confirms the qualitative behaviour of the algorithms as a function of  $p_t$  as seen on the left-hand side in figure 6, where the same quantity is shown for a LO PYTHIA setup (see section D for details).

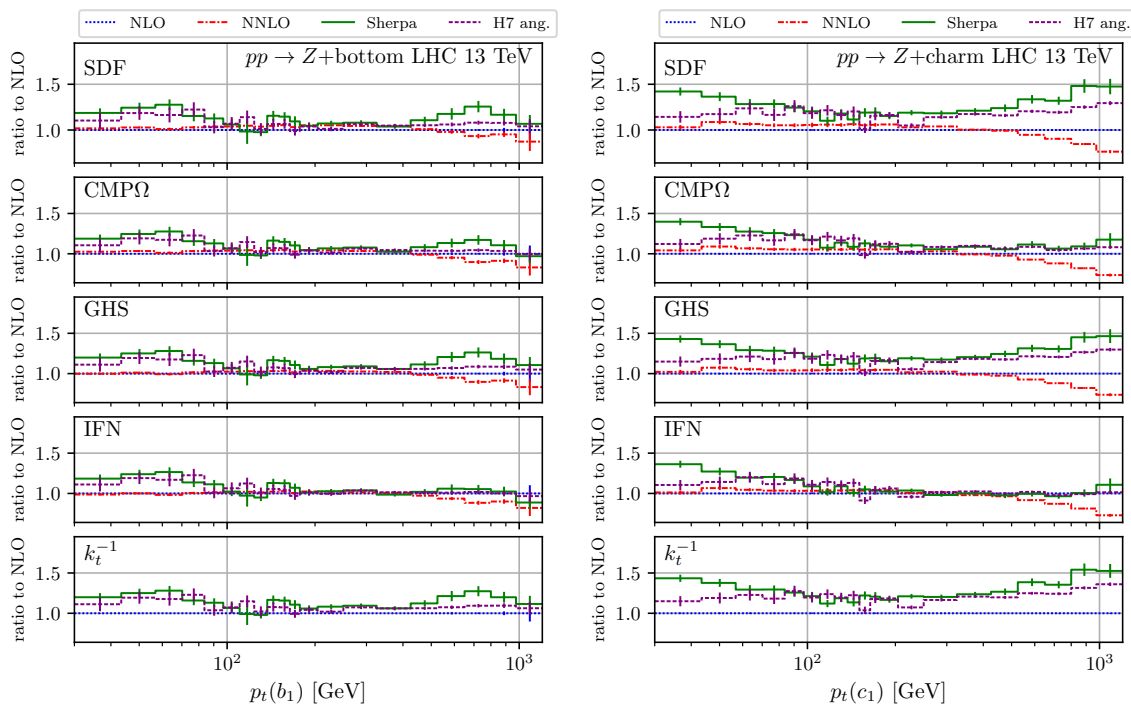
For a selection of events with at least two  $b$  quarks in the final state, the right-hand side of figure 6 shows the quantity  $p_{t,b}/p_{t,b\text{-jet}}\Delta R_{b\bar{b}}$ , where  $p_{t,b\text{-jet}}$  is the transverse momentum of the  $b$  jet,  $p_{t,b}$  denotes the transverse momentum of the hardest  $b$  quark in the jet and  $\Delta R_{b\bar{b}}$  in the angular distance between the two  $b$  quarks in the event. This quantity separates the hard- $b$ -initiated events (solid grey line) from the  $g/q$ -initiated events (dashed grey line) in a clean way; see section D for more details. The  $b$ -labelled jets originating from light flavours are likely to have obtained their label from a soft  $b$  pair (small  $p_{t,b}/p_{t,b\text{-jet}}$ ) while  $g$ -initiated jets are changed into  $b$ -labelled jets through collinear  $g \rightarrow b\bar{b}$  splittings (small  $\Delta R_{b\bar{b}}$ ). We can find qualitatively different behaviours by looking at the  $b$ -labelled rates for this observable provided by the various algorithms. The cleanest labelling, i.e., the labelling that is least contaminated by  $q/g$ -initiated events, is provided by IFN. The other algorithms pick up a significant fraction of the  $q/g$ -initiated events. These events populate the differences observed in the  $p_t$  spectrum, i.e., the differences arise from mislabelled jets.



**Figure 5.** NLO+PS predictions from SHERPA and HERWIG for  $pp \rightarrow Z + b$  (left) and  $pp \rightarrow Z + c$  (right) in central kinematics. Different jet algorithms (CMPΩ — green, SDF — light blue, GHS — yellow, IFN — red) are compared for the transverse momentum of the leading  $b/c$  jet. Additionally, differences between the cluster and string hadronisation model for HERWIG, as well as parton and hadron level for SHERPA, are shown. The vertical bars indicate statistical uncertainties.



**Figure 6.** LO+PS  $Z + j$ . Left: the ratio, to net-flavour anti- $k_t$  (in red), of IFN ( $\alpha = 1$  and  $\alpha = 2$ , blue), CMP ( $a = 0.1$ , green), GHS ( $\alpha = 1$ ,  $\omega = 2$ , black) and SDF ( $\beta = 2$ ,  $z_{\text{cut}} = 0.1$ , yellow), for the transverse momentum of the leading  $b$ -labelled jet  $p_{t,b\text{-jet}}$ . Right: distribution of  $\frac{p_{t,b}}{p_{t,b\text{-jet}} \Delta R_{b\bar{b}}}$ , in events with two  $b$  quarks at parton level, compared to classification of events according to the PYTHIA hard matrix element (hard  $b$ , solid grey, vs. hard  $g$  or light  $q$ , dashed grey).

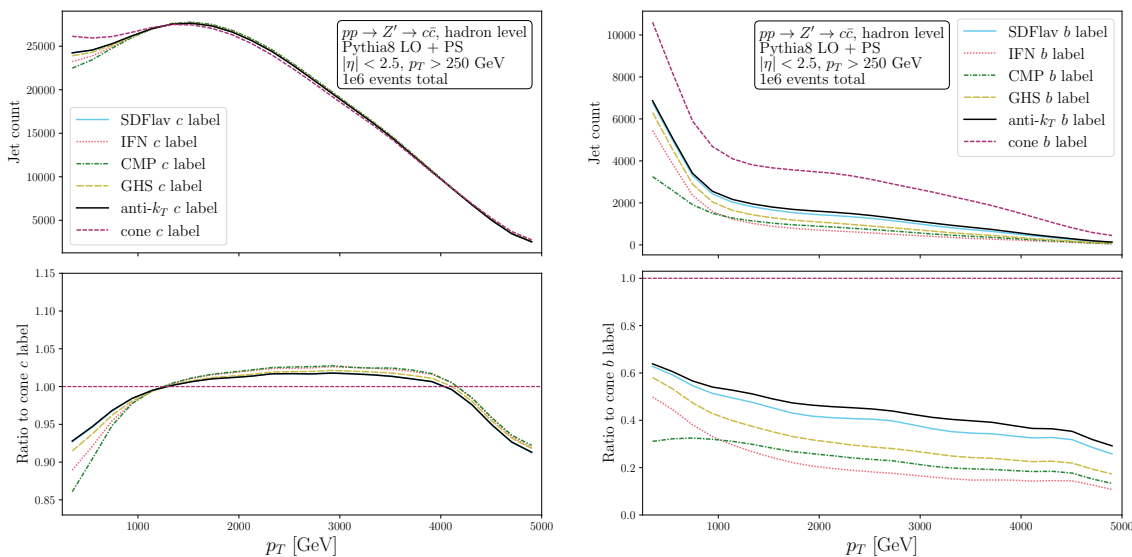


**Figure 7.** Comparison between fixed-order and NLO+PS predictions for  $pp \rightarrow Z + b/c$  in central kinematics. Different jet algorithms are compared for the transverse momentum of the leading  $b/c$  jet in the bulk and tail of the phase space.

This conclusion can be confirmed by comparing the showered and hadronised results with the fixed-order predictions for  $Z + \text{jet}$  from section C. For this purpose, we plot the ratio of the various predictions to the fixed-order NLO QCD results<sup>12</sup> in figure 7. Several observations can be made. First, we see a visible parton-shower effect in the threshold region, which can be traced back to the matching and shower scale dependence as the same differences occur for nonflavoured jets (compare to the discussion in section C). Focusing on the  $c$ -label case, we find different behaviour of the various algorithms when looking into the tail of the distribution. SDF and GHS start to have more significant corrections from the shower at higher  $p_t$ , comparable to the effects observed for nonflavoured anti- $k_t$  jets. The IFN algorithm receives minor corrections that approach unity at a very high  $p_t$ , while the CMP algorithm is somewhat between these extremes. A similar behaviour, though smaller in scale, can be observed for the  $b$ -label case (the SHERPA predictions’ statistical uncertainty is significant here, but for HERWIG 7, this observation is statistically significant).

The comparison of fixed-order with showered cross sections in the high- $p_t$  limit indicates the infrared sensitivity of the observable definition provided by a given algorithm. The smaller the corrections, the less sensitive the algorithm is to infrared physics. We observe that IFN delivers the most robust definition of infrared-insensitive jet flavour from that measure, followed by CMP, GHS and SDF.

<sup>12</sup>The motivation is that all NLO QCD results, regardless of the algorithm, are very similar for the observables considered here (see, for example, figure 14 in section C).



**Figure 8.** The  $p_t$  spectra of  $c$  and  $b$  jets found in  $pp \rightarrow Z' \rightarrow c\bar{c}$  events, simulated at LO and showered with PYTHIA 8. All labelling schemes and algorithms shown use the “net-flavour” recombination scheme, except for the cone label, which uses the “any-flavour” scheme.

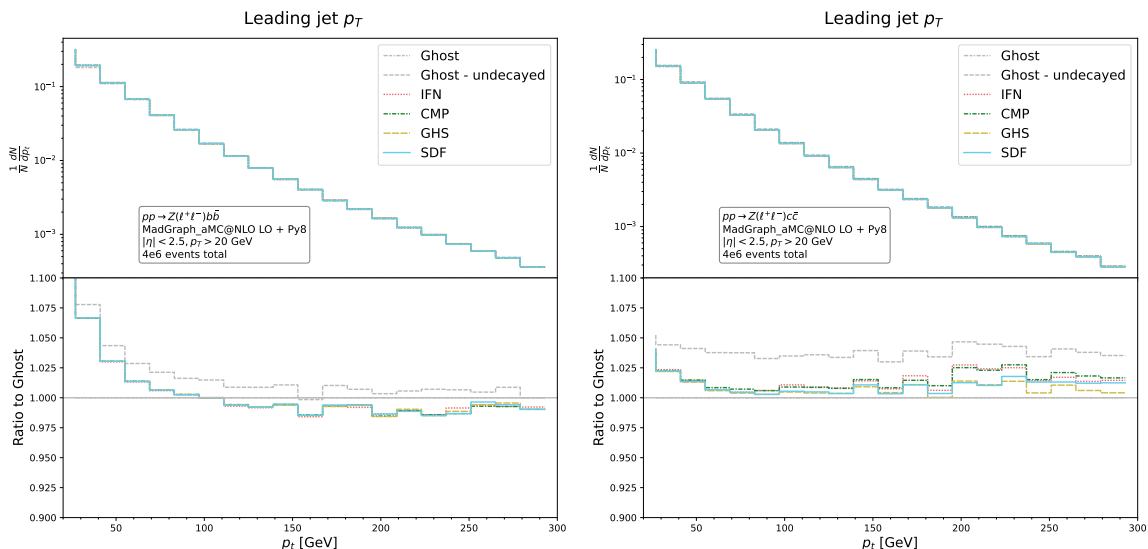
Finally, it is also interesting to investigate the impact of the different flavour labelling schemes in the development of multivariate flavour tagging algorithms. In particular, section H reports flavour labelling studies on simulated  $pp \rightarrow Z' \rightarrow c\bar{c}$  events, generated with PYTHIA 8 LO+PS with  $m_{Z'} = 4$  TeV, because such a MC sample is used by the ATLAS Collaboration for the training of multivariate flavour tagging classifiers at the jet- $p_t$  scale of 1 TeV. The sample is used, together with other MC simulations for the other flavours and phase space regions, in order to create a training dataset with well-separated  $c$ -jet signatures in the detector. However, figure 8 shows that such an assumption is not robust: a jet produced with  $p_t$  far from the bulk of the events is likely to be polluted by gluon splitting into quark flavours different from the  $c$  quarks present in the hard-scattering matrix element. The effect is even more striking considering the right part of figure 8, where the experimental flavour labelling algorithms find a much larger contamination of  $b$  jets in a sample that has no  $b$  quarks from the hard scattering with respect to the IRC-safe algorithms. The pollution of the training sample of the multivariate flavour tagging classifiers is considered as one of the possible reason for a sizeable systematic uncertainty due to model dependence in the efficiency calibration of the flavour tagging algorithm [55, 56, 75]. In this respect, the use of IRC-safe algorithms in the definition of new training samples would surely improve this bias.

#### 4.4 Impact of hadron reconstruction on flavoured jet definitions

Studies presented so far have focused on the impact of an IRC-safe flavour definition on inclusive jet observables. However, we realized that additional effects arise when considering flavour in comparisons with experimental measurements handling the treatment of heavy-flavour hadrons in different ways. We consider two examples: the inclusion of the decay products of  $b$  and  $c$  hadrons in inclusive jet observables, as well as an example of an analysis that tackles jet flavour from a different perspective, namely using the WTA flavour algorithm.

**Including heavy flavour decays.** For what concerns the first, a comparison of the different flavour labelling algorithms has been performed using  $pp \rightarrow Z + bb$  and  $pp \rightarrow Z + cc$  events and selected within a fiducial phase with a di-lepton mass cut and jets of  $p_t > 20$  GeV, similar to “standard” measurements of a  $Z$  boson produced in association with flavoured jets in the ATLAS or CMS experiments [74, 76, 77]. Detailed results are reported in section G. The comparison, visible in figure 9, shows a sizeable discrepancy in the low jet  $p_t$  spectrum when heavy flavour jets are selected in the analysis phase space using the experimental labelling scheme — namely ghost labelling — versus the new labelling algorithms — namely SDF, CMP, GHS, and IFN. The simple, yet initially unexpected, reason behind this comes from the need to define experimental flavour labelling schemes using final state particles, i.e., including the hadrons originating from heavy flavour decay chains, while the new labelling algorithms need to rely on undecayed heavy-flavour hadrons in order to be correctly evaluated. This introduces a difference in the kinematics of the clustered jet, stronger for  $b$  jets than for  $c$  jets, according to the larger decay pattern multiplicity of the first. The same effect was already reported in a recent ATLAS  $Z +$  heavy-flavour measurement [74], signalling the need of a second unfolding step to pass from phase space definition using final state particles to a phase space defined using undecayed heavy hadrons. In order to investigate smaller differences between the various labelling schemes, the analysis has been repeated switching off the heavy-flavour decay in the events, as visible in figure 9 with the label “ghost-undecayed”. Percent-level differences in the selection of  $b$  jets appear over the full jet  $p_t$  range; while for  $c$  jets the effect starts at the jet  $p_t$  threshold, of 20 GeV, and it grows to about 5% at higher  $p_t$ . This is consistent with an increasing  $g \rightarrow q\bar{q}$  contribution at higher jet  $p_t$ , with two heavy-flavour quarks (produced by the hard ME or the parton shower) winding up in the same jet: in such cases all the IRC-safe labelling algorithm — with tiny differences between all of them — are classifying the jets as flavourless, while the current experimental algorithms assign a flavour to the jets.

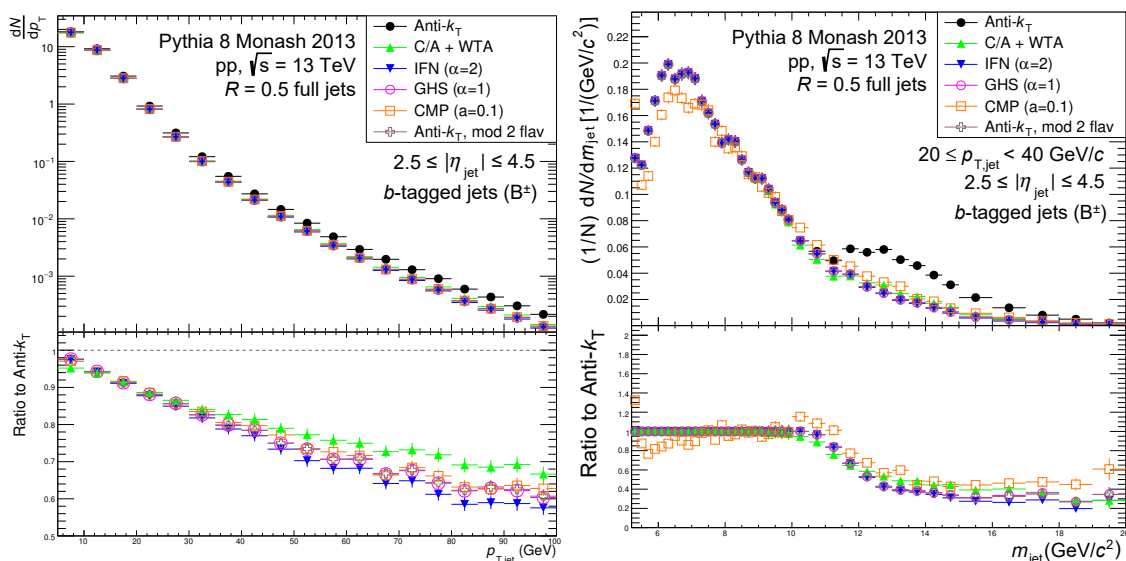
**Exclusive  $b$  tagging with WTA.** As expected, more significant differences between the currently available experimental labelling algorithms and the IRC safe ones arise in special region of the phase space where soft or collinear emission is more abundant. This prompted further investigation using simulated samples used in experiments and where these effects may be enhanced. Other LO+PS studies were carried out in approximate LHCb kinematics, as documented in section F. The LHCb experiment has a forward detector geometry which is currently unique at the LHC, covering roughly the pseudorapidity range  $2 < \eta < 5$  [78]. Because of its access to forward kinematics, flavoured hadrons are produced with a larger longitudinal boost than at mid-rapidity, providing a longer distance between the interaction point of colliding hadrons and the subsequent decay vertex of heavy-flavour hadrons. This makes it easier to separate and identify boosted decays of single hadrons in LHCb kinematics. These benefits, coupled with the ability to identify charged hadron identities across a vast range of momenta [79], enables LHCb to exclusively reconstruct specific heavy-flavour hadrons at detector level before jet reconstruction. In this case, the WTA flavour algorithm (see section 3.5) offers an advantage over other algorithms, as knowledge of all flavoured particles in the event is not required; it is only important whether or not the reconstructed flavoured hadron is collinear with the WTA axis.



**Figure 9.** The  $p_t$  distribution of leading  $b$ -jets (left) and leading  $c$ -jet (right) identified with the ghost and flavoured jet algorithms labels using final state particles. Ratios are compared to the distribution for jets obtained with ghost labelling.

Although WTA flavour tagging (as discussed above) follows a fundamentally different approach to the flavour tagging problem, similar behaviour is observed between WTA flavour and other new flavoured jet algorithms. These results confirm the dominance of flavour recombination effects over different algorithms and show that while WTA flavour tagging does not use any modified flavour recombination scheme, it similarly tends to reject gluon splitting at a rate comparable to other algorithms, in the case of  $b$ -jets.

Figure 10 shows the  $b$ -tagged jet cross section for jets containing an explicitly reconstructed  $B^\pm$ -meson for the different tagging algorithms as compared to an “any flavour” anti- $k_t$  baseline. The tagging fraction is observed to fall as a function of  $p_{t,\text{jet}}$  for all algorithms, which is expected due to the growing gluon splitting contribution at higher  $p_{t,\text{jet}}$ . The same figure also shows the  $b$ -jet invariant mass  $m_{\text{jet}}$  distribution for the same sample of jets. A secondary peak is observed in the tail of the distribution at the kinematic limit of  $m_{\text{jet}} = 2 \cdot m_{B^\pm}$ , where a significant proportion of jets with  $b$  flavour originating from gluon splitting populate the spectrum. All new algorithms are observed to suppress this secondary peak with little-to-no modification of the primary mass peak, with the exception of CMP, which has small modifications at low  $m_{\text{jet}}$  due to the slightly modified kinematics of the clustering redefinition.



**Figure 10.** LO+PS predictions for an exclusive sample of  $b$  jets containing a fully-reconstructed  $B^\pm$ -meson, comparing flavour defined using “any-flavour” anti- $k_t$  (black) with the new algorithms and flavour recombination schemes. *Left:*  $b$ -jet cross section. *Right:*  $b$ -jet invariant mass. For more details, see section F.

## 5 Concluding remarks

In this paper, we have performed a first comprehensive phenomenological study that focuses on the challenges and opportunities presented by new definitions of jet flavour in the context of high-precision collider physics. After a short introduction, we have reviewed the concepts of flavour labelling and flavour tagging in the context of a typical workflow of an experimental analysis in particle physics, highlighting tensions with an IRC-safe definition of jet flavour from a theory viewpoint.

The new generation of algorithms — SoftDrop Flavour (SDF), Flavoured anti- $k_t$  (CMP), Flavour Dressing (GHS), and Interleaved Flavour Neutralisation (IFN)—represent a significant advancement in jet flavour labelling, offering IRC-safe definitions applicable at fixed order and throughout the simulation chain. Their key role is to define jet flavour labels that are resilient against the contamination from collinear splittings and soft emissions, particularly  $g \rightarrow b\bar{b}$ , which traditionally hampers our ability to compare high-precision calculations to flavour-tagged measurements. In this context, we have stressed the importance of adopting a flavour recombination scheme such as “net-flavour” or “mod2-flavour”, rather than “any-flavour”, which is the current standard.

We have performed several studies for LHC phenomenology, which are detailed in the appendices, exploiting theory predictions computed with different tools and levels of precision. In particular, we have considered fixed-order predictions at NNLO, Monte Carlo parton showers, both at hadron and parton level, as well as their matching to NLO matrix elements. In the bulk of the phase space, all new algorithms give compatible results and most differences with respect to current experimental strategies can be traced back to the use of different flavour recombination schemes. However, as we move to regions of phase space where the

hierarchy of scales is more pronounced, e.g., smaller quark masses or higher transverse momenta, soft and collinear emissions are abundant, and the algorithms start to differ from each other. Among these, IFN stands out for its robustness against soft and collinear effects, delivering distributions that change little through the simulation chain.

In this study, we have focused on the role played by the novel algorithms on flavour labelling, which is performed on Monte Carlo pseudodata (truth level), for which full flavour information is available. However, the new flavour algorithms may impact other components of experimental workflows. We can identify three main aspects:

- definition of appropriate samples for background estimate and/or signal yield evaluation in new physics searches or measurements;
- accurate estimate of detector efficiencies and selection phase space used to unfold data events to a fiducial phase space of a measurement with flavoured final states;
- optimisation and training of flavour tagging algorithms to be used in flavour identification in collision data events.

Indeed, using new flavour definitions as truth-level labels could improve the fidelity of unfolding procedures, providing a more stable baseline for correcting detector effects and modelling uncertainties. Since unfolding is typically performed separately for each analysis, this could be the easiest first step toward adoption. A critical open question is whether a mismatch between the labelling schemes used in unfolding and flavour tagging, which is performed using actual experimental inputs, where flavour knowledge is limited, significantly degrades analysis performance. While definitive answers require more study, the large effects observed when switching from any-flavour to mod2 or net-flavour recombination schemes — despite potential efficiency losses — strongly motivate further experimental investigations.

Despite their formal appeal, these algorithms face practical limitations due to detector resolution and reconstruction capabilities. For example, their ability to identify the correct initiating flavour may be hindered in the presence of  $B$ - and  $D$ -hadron decays. However, we can envision the use of IRC-safe flavour labels to train flavour taggers. Because these are typically based on rather complex neural networks, making a switch in labelling schemes to re-train them is likely more laborious and deserves more investigation. Still, these algorithms offer a path toward defining more accurate training targets and mitigating model-induced bias due to poorly modelled infrared physics. Notably, training mismatches between PYTHIA and HERWIG are already observed and may originate from such sensitivity [55–57]. Furthermore, with improved particle identification capabilities in LHCb and ALICE, and the advent of new vertex reconstruction algorithms like the GN2 tagger in ATLAS [80] or CMS’s ML-based  $b$ -hadron momentum estimators [81], some of these algorithms could soon be implemented experimentally for select final states.

Our findings suggest the best approach to a successful implementation may be gradual: using these algorithms first at truth level, to improve Monte Carlo-based corrections, and progressively moving toward direct application to reconstructed data as experimental techniques mature. Several open directions remain. Firstly, deeper studies are needed on the modelling of gluon splitting into  $b\bar{b}$  pairs, possibly through double  $b$ -tagged jets. Secondly, validations

of the algorithms' performance under different hadronisation and parton shower models (e.g., string vs. cluster models in HERWIG and SHERPA) must continue. These variations are typically small, but still non-negligible in some regimes, especially for charm.

## Acknowledgments

First and foremost, we are grateful to the organisers of Les Houches 2023 PhysTeV workshop and of the subsequent IPPP ‘Flavoured Jets at the LHC’ workshop. A.B. and A.S. thank the CERN Theoretical Physics Department for hospitality while part of this research was being carried out. The work of S.M. and of F.S. is supported is supported by the Italian Ministry of University and Research (MUR) grant PRIN 2022SNA23K funded by the European Union — Next Generation EU, Mission 4, Component 2, CUP D53D23002880006 and by ICSC Spoke 2 under grant BOODINI, CUP I57G21000110007. L.S. is supported by an ARC Discovery Early Career Researcher Award (project number DE230100867). The work of A.S. was supported by the OpenMAPP project via National Science Centre, Poland under CHIST-ERA programme (grant No. NCN 2022/04/Y/ST2/00186) and by the Priority Research Area Digiworld under the program ‘Excellence Initiative — Research University’ at the Jagiellonian University in Kraków. J.W. gratefully acknowledges the support of the Polish National Science Centre grant 2019/34/E/ST2/00457, and the Polish high-performance computing infrastructure PLGrid (HPC Centre: ACK Cyfronet AGH) for providing computing facilities and support within computational grants PLG/2024/017081 and PLG/2025/018389. A.Hinzmann acknowledges support from DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, and support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2121 “Quantum Universe” – 390833306. G.P.S. is supported by a Royal Society Research Professorship (RP\R1\23100) and by the Science and Technology Facilities Council (STFC, grant ST/X000761/1). D.R. is supported by the European Union under the HORIZON program in Marie Skłodowska-Curie project No. 101153541. D.R. was further supported by the STFC under grant agreement ST/P006744/1 during the initial stages of this work. R.P. acknowledges the support from the Cambridge Service for Data Driven Discovery (CSD3), part of which is operated by the University of Cambridge Research Computing on behalf of the STFC DiRAC HPC Facility ([www.dirac.ac.uk](http://www.dirac.ac.uk)). The DiRAC component of CSD3 was supported by STFC grants ST/P002307/1, ST/R002452/1 and ST/R00689X/1. R.P. acknowledges further support from the Leverhulme Trust and the Isaac Newton Trust in the initial stages of the project. S.C. acknowledges the support of the Swiss National Supercomputing Centre (CSCS) under project ID ETH5f. R.G. is supported by an STFC studentship (project ST/X508664/1), the Clarendon Scholarship, and the Wolfson Harrison UK Research Council Physics Scholarship. R.X. acknowledges support from the Homer A. and Donna J. Neal Endowment Fund of the University of Michigan.

## A Flavoured jet algorithms in *FASTJET*

The following code snippets allow the user to construct the resulting jets (starting from a collection of input particles `vector<PseudoJet> event`, set up with flavour information as explained in section 3.6). Note that one can consider all possible flavour at once, or treat

flavours separately, one at a time (e.g.,  $b$  and  $c$ ).<sup>13</sup> To consider a single flavour, one can reset all flavours but the one of interest; for example, when setting up flavour from the input particles, to consider *only*  $b$  flavour:

```
int filter_flav = 5;
auto flav_info = new FlavInfo(pdg_id);
flav_info->reset_all_but_flav(filter_flav);
```

In all cases below we set the jet radius to  $R = 0.4$ , and we assume that the code has access to the `fastjet` and `fastjet::contrib` namespaces. The algorithms make use (directly or indirectly) of the `FlavRecombiner` class, which implements the flavour recombination scheme (either “mod2” or “net”) during the clustering of the jets. Each algorithm returns a list of jets whose flavour can be accessed, as described at the end of section 3.6, with the `FlavHistory` class. The return value inherits the choice of the recombination scheme. The structure of the code, however, allows for a different criterion at the stage of the label assignment. In particular, it is possible to use “net” flavour recombination during clustering and assign the label based on “mod2” flavour:<sup>14</sup>

```
bool is_blabelled_mod2 = abs(FlavHistory::current_flavour_of(j)[5])
```

The following examples demonstrate the basic usage for each algorithm.

#### IFN:

```
// jet radius parameter
double R = 0.4;
// create a base jetdef (anti-kt or C/A)
JetDefinition base_jet_def(antikt_algorithm, R);
// flavour recombiner (default is net flavour)
FlavRecombiner flav_recombiner;
base_jet_def.set_recombiner(&flav_recombiner);
double alpha = 2.0;
// IFN plugin (sets the parameter omega to 3-alpha by default)
JetDefinition jet_def(new IFNPlugin(base_jet_def, alpha));
// get the IFN jets
vector<PseudoJet> IFN_jets = jet_def(event);
```

#### CMP:

```
// jet radius parameter
double R = 0.4;
// CMP parameters
```

<sup>13</sup>For instance in the case of the IFN algorithm this choice can modify the neutralisation sequence. Also for CMP the choice of which flavours to consider has an impact on the clustering sequence and the general recommendation is to only consider one flavour at a time.

<sup>14</sup>If “mod2” is used during clustering, “net” and “mod2” yield the same result during assigning.

```

double a = 0.1;
CorrectionType CMP_corr = CorrectionType::SqrtCosyCosPhiArgument_a2;
ClusteringType CMP_clust = ClusteringType::DynamicKtMax;
// CMP plugin
JetDefinition jet_def(new CMPPlugin(R, a, CMP_corr, CMP_clust));
// flavour recombiner (default is net flavour)
FlavRecombiner flav_recombiner;
jet_def.set_recombiner(&flav_recombiner);
// get the CMP jets
vector<PseudoJet> CMP_jets = jet_def(event);

```

**GHS:**

```

// jet radius parameter
double R = 0.4;
// anti-kt jet definition
JetDefinition base_jet_def(antikt_algorithm, R);
// flavour recombiner (default is net flavour)
FlavRecombiner flav_recombiner;
base_jet_def.set_recombiner(&flav_recombiner);
// run the jet clustering with the base jet definition
vector<PseudoJet> base_jets = base_jet_def(event);

// run the flavour dressing (requires some minimal pt cut)
double ptcut = 20.0;
double alpha = 1.0;
double omega = 2.0;
// get the GHS jets
vector<PseudoJet> GHS_jets = run_GHS(base_jets, ptcut, alpha,
    omega, flav_recombiner);

```

**SDF:**

```

// jet radius parameter
double R = 0.4;
// anti-kt jet definition
JetDefinition base_jet_def(antikt_algorithm, R);
// flavour recombiner (default is net flavour)
FlavRecombiner flav_recombiner;
base_jet_def.set_recombiner(&flav_recombiner);

// SoftDrop parameters
const double beta = 2.0;
const double zcut = 0.1;
// initialise the SoftDrop flavour algorithm

```

```

SDFlavourCalc sdFlavCalc(beta, zcut, R);

// run the jet clustering with the base jet definition
vector<PseudoJet> SDF_jets = base_jet_def(event);
// run the SDFlav algorithm on the base jets
sdFlavCalc(SDF_jets);

```

## B NNLO fixed-order $pp \rightarrow WH(b\bar{b})$

**Author:** Arnd Behring.

We adapt the calculation from refs. [82, 83] for the process

$$pp \rightarrow W^+(\rightarrow e^+\nu_e)H(\rightarrow b\bar{b}) \quad (\text{B.1})$$

to study differences between the flavoured jet algorithms. The calculation is performed in the nested-soft collinear subtraction scheme [84–87], which is closely related to the sector-improved residue subtraction scheme [88–91]. We treat the decays of the Higgs and  $W$ -boson in the narrow-width approximation. The decay  $W^+ \rightarrow e^+\nu_e$  is treated at LO, but both the production subprocess  $pp \rightarrow HW^+$  and the decay subprocess  $H \rightarrow b\bar{b}$  are calculated through NNLO in QCD. In this study we treat the  $b$  quark as massless.

### B.1 Setup of the calculation

The calculation closely follows the setup from ref. [83], but we collect a few details below for the convenience of the reader.

**Input parameters.** For the input parameters of the calculation we use the values

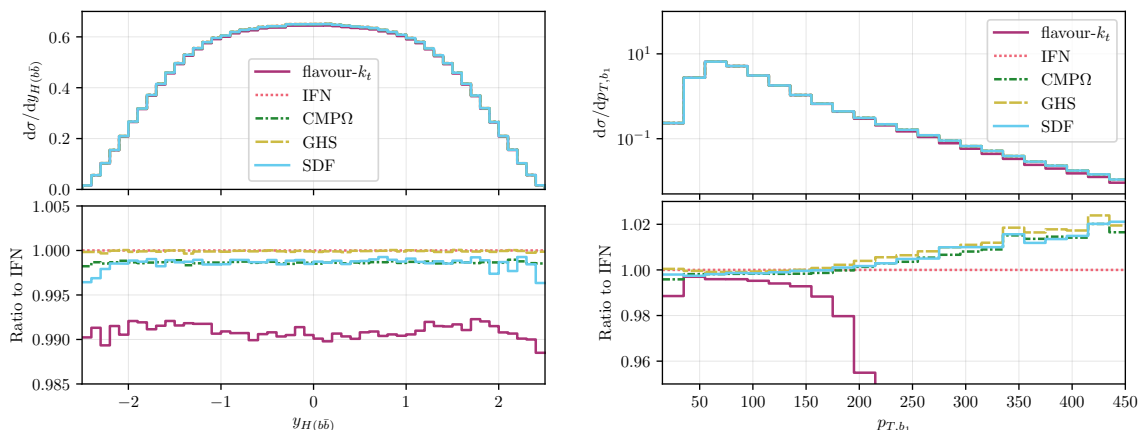
$$m_H = 125 \text{ GeV}, \quad m_W = 80.399 \text{ GeV}, \quad \Gamma_W = 2.1054 \text{ GeV}, \quad (\text{B.2})$$

$$m_t^{\text{pole}} = 173.2 \text{ GeV}, \quad m_b^{\text{pole}} = 4.78 \text{ GeV}, \quad G_F = 1.16639 \times 10^{-5} \text{ GeV}^{-2}, \quad (\text{B.3})$$

$$\sin^2 \theta_W = 0.2226459. \quad (\text{B.4})$$

The value of the  $b$ -quark mass  $m_b^{\text{pole}}$  is only used to calculate the  $b$ -quark Yukawa coupling. For details about this and about the treatment of the narrow width approximation we refer to ref. [83]. We use PDFs in the  $n_f = 5$  flavour scheme (NNPDF31\_nnlo\_as\_0118 [92]), as provided through the LHAPDF library [93]. The running of the strong coupling  $\alpha_s$  is treated in the  $\overline{\text{MS}}$  scheme with  $n_f = 5$  active flavours.

**Jet algorithms.** We use the four new IRC-safe flavoured jet algorithms (IFN, CMP $\Omega$ , GHS, and SDF) via their implementations in FASTJET [63]. For reference, we also include the flavour- $k_t$  algorithm. In all algorithms we use the net flavour recombination scheme and a jet radius of  $R = 0.4$ . The remaining parameters that are specific to each jet algorithm are chosen as listed in the baseline setup in section 4, except for the minimal jet transverse momentum  $p_{t,\text{cut}}$  in the GHS algorithm, where we use  $p_{t,\text{cut}} = 25 \text{ GeV}$  to match the fiducial cuts, and we use  $\beta = 2$  in the SDF algorithm.



**Figure 11.** Fixed-order NNLO QCD predictions for the rapidity of the reconstructed Higgs boson in  $pp \rightarrow WH(bb)$  (left) and for the transverse momentum of the leading  $b$  jet for different jet algorithms. The distributions for different jet algorithms are calculated on the same events. The lower panes show the ratio of the distributions for the respective algorithms to that for IFN.

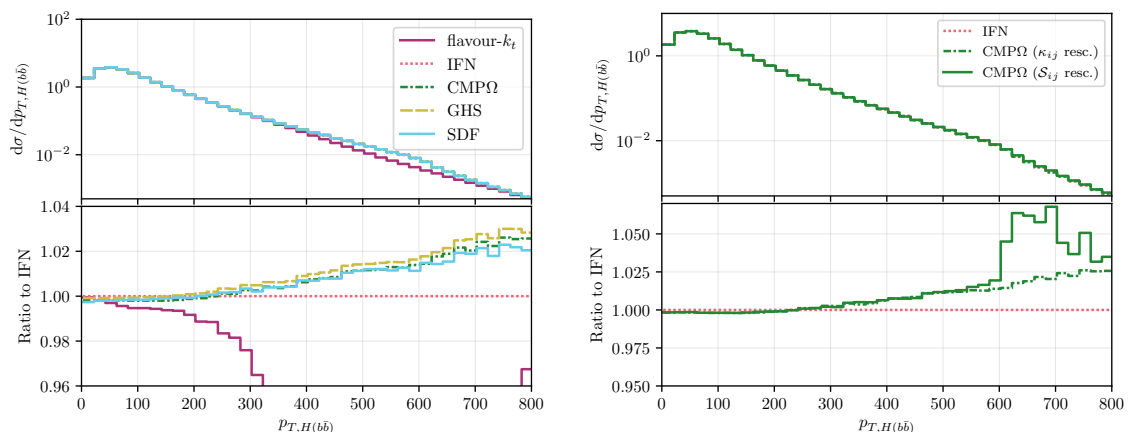
**Fiducial cuts.** We consider the following cuts on jets. In order to be counted as a  $b$  jet, the jet in question has to have a transverse momentum of at least  $p_t > 25$  GeV and a rapidity of at most  $|\eta| < 2.5$ . Similarly, flavourless jets also have to have at least  $p_t > 25$  GeV of transverse momentum. We require at least two  $b$ -labelled jets. The positron from the decay of the  $W$  boson has to fulfil  $p_t > 15$  GeV and  $|\eta| < 2.5$ .

For certain observables, we reconstruct a candidate for the Higgs boson by selecting among all  $b$ -labelled jets the pair whose invariant mass is closest to the invariant mass of the Higgs boson. The sum of momenta of these two  $b$  jets is then treated as the reconstructed momentum of the Higgs boson, from which further observables (transverse momentum, rapidity, invariant mass, etc.) can be derived. We will label those observables with a subscript  $H(bb)$ .

## B.2 Results

For many of the observables one can define for this process, the differences between IFN,  $\text{CMP}\Omega$ , GHS, and SDF, are very small, at or below the percent level. Examples for this behaviour are the rapidity distribution of the reconstructed Higgs momentum,  $y_{H(bb)}$  (figure 11, left), or the transverse momentum of the leading  $b$  jet,  $p_{t,b_1}$  (figure 11, right). Larger differences are observed between the new flavoured jet algorithms and the flavour- $k_t$  algorithm, which is to be expected given that it is based on the  $k_t$  instead of the anti- $k_t$  algorithm.

Somewhat more pronounced effects can be observed in the tail of the transverse momentum distribution of the reconstructed Higgs,  $p_{t,H(bb)}$ , see figure 12 (left). There is a stark difference between the flavour- $k_t$  algorithm and the novel flavoured jet algorithms, starting from around  $p_{t,H(bb)} \sim 200$  GeV. A similar effect was already observed in ref. [83] when comparing the massless calculation using the flavour- $k_t$  algorithm to the massive calculation using the plain anti- $k_t$  algorithm. There, it was interpreted to be due to the fact that the flavour- $k_t$  algorithm start to cluster the  $b$  quarks coming from the Higgs boson decay into a single flavourless jet for much less boosted configurations compared to the anti- $k_t$  algorithm. Since we require at least



**Figure 12.** Predictions for the transverse momentum distribution of the reconstructed Higgs boson momentum,  $p_{t,H(bb)}$ , at NNLO in QCD. The distributions for the different jet algorithms were calculated using the same events. The lower pane shows the ratio of the distributions for the respective algorithms to that for IFN. Left: comparison of different flavour jet algorithms with each other; right: comparison of the different modifications of the  $\text{CMP}\Omega$  algorithm discussed in eqs. (3.5) and (3.6), see main text for details.

two  $b$  jets, many of those events will not pass this cut. Interestingly, also the distributions for the anti- $k_t$ -based algorithms seem to change slope around  $p_{t,H(bb)} \sim 600$  GeV and towards higher transverse momenta the curves start to approach the curve of the flavour- $k_t$  algorithm. Around this value of  $p_{t,H(bb)}$ , also the anti- $k_t$ -based algorithms start to predominantly cluster the  $b$  quarks from the Higgs boson decay together into a flavourless jet, as was shown analytically in ref. [36].

During the study it was observed that the  $\text{CMP}\Omega$  algorithm with the rescaling of  $\mathcal{S}_{ij}$  shown in eq. (3.5) shows a somewhat surprising behaviour above  $p_{t,H(bb)} \gtrsim 600$  GeV (shown as the solid green line in figure 12 (right)). Compared to the other algorithms, the  $\text{CMP}\Omega$  algorithm yields about 2% to 4% larger results, with a very sudden onset. It turns out that this effect is already present at LO and the position of the onset scales linearly in  $R$ : when the jet radius is chosen twice as large, the jump in the distribution moves to about 300 GeV. This indicates that the issue is connected to configurations where the  $b$  quarks from the Higgs decay are boosted in such a way that they are on the verge of being clustered into a single jet. Detailed investigations show that the effect is in fact due to the rescaling of  $\mathcal{S}_{ij}$  by  $\frac{\Omega_{ij}^2}{\Delta R^2}$ . This modifies the metric compared to the anti- $k_t$  algorithm also away from singular configurations. In particular, it can modify whether boosted Higgs decay events are clustered into a single jet. Since the three other flavoured jet algorithms obey exact anti- $k_t$  kinematics, this modification shows up in the ratio plot. If we instead rescale  $\kappa_{ij}$  by  $\frac{\Omega_{ij}^2}{\Delta R^2}$ , as described in eq. (3.6) and shown as the dash-dotted green line in figure 12 (right), the modification of the metric is confined to fairly soft configurations, due to the step function in  $\mathcal{S}_{ij}$ . This ensures that close to the edges of the jet the metric agrees with the anti- $k_t$  metric. Switching to this alternative fix removes the sudden jump in the  $p_{t,H(bb)}$  distribution for the  $\text{CMP}\Omega$  algorithm.

## C $pp \rightarrow Z + b/c$ in central kinematics: NNLO and NLO+PS

**Authors:** Davide Napoletano, Rene Poncelet, Daniel Reichelt, Andrzej Siódmok, and James Whitehead.

In this appendix, we report on our study of  $Z$ +jet events, at fixed order up to NNLO QCD and with matched NLO-plus-parton-shower calculations from event generators HERWIG and SHERPA, with and without hadronisation.

### C.1 Details of the simulations

We consider the process

$$pp \rightarrow \mu^+ \mu^- + \text{jet} + X . \quad (\text{C.1})$$

Loosely inspired by the selection criteria used in ref. [94], we identify events corresponding to central nearly-on-shell  $Z$  bosons through the requirements:

$$p_{\text{T}}^{\mu} > 20 \text{ GeV} , \quad |y^{\mu}| < 2.4 , \quad (\text{C.2})$$

$$p_{\text{T}}^{\mu\mu} > 20 \text{ GeV} , \quad m_{\mu\mu} \in (77, 111) \text{ GeV} . \quad (\text{C.3})$$

Each accepted event contains exactly one  $\mu^+ \mu^-$  pair fulfilling these conditions.

We use FASTJET [63] to implement the various jet algorithms as described in section A. We include all flavoured jet algorithms listed in section 4 and the standard anti- $k_t$  with a jet radius of  $R = 0.5$ , applied to all visible particles except the identified muon pair. For the flavoured jet algorithms, we use the net-flavour combination scheme during clustering and initialise only the flavour in question, e.g., either the  $b$  or  $c$  flavours, see section A. The labelling of the clustered jets is performed with the mod2-flavour scheme in the case of flavoured jet algorithms, while we employ the mod2, ghost and cone labelling strategies for the anti- $k_t$  algorithm. The labelling is performed separately for  $b$ - and  $c$ -labelled jets, i.e., we do not consider multiple labels at a time. We identify as flavour-labelled jets those passing the selection cuts

$$p_t^{j_{b,c}} > 30 \text{ GeV} , \quad |y^{j_{b,c}}| < 2.4 . \quad (\text{C.4})$$

For all predictions, we use the PDF4LHC21 PDF set [95] within the LHAPDF library [93], and as factorisation and renormalisation scales, we adopt the transverse mass of the muon pair,

$$\mu_{\text{F}}^2 = \mu_{\text{R}}^2 = M_{\text{T},W}^2 \equiv p_{\text{T},\mu\mu}^2 + M_{\mu\mu}^2 . \quad (\text{C.5})$$

For the hard process, we are working in QCD with  $n_f = n_l = 5$  flavours, and the electromagnetic coupling is obtained in the  $G_{\mu}$  scheme using the Fermi constant

$$\alpha = \frac{\sqrt{2}}{\pi} G_{\mu} m_W^2 \left( 1 - \frac{m_W^2}{m_Z^2} \right) \quad \text{with} \quad G_{\mu} = 1.16638 \times 10^{-5} \text{ GeV}^{-2} . \quad (\text{C.6})$$

The values of the masses and widths used for the numerical simulations read:

$$\begin{aligned} m_Z &= 91.1535 \text{ GeV}, & \Gamma_Z &= 2.4943 \text{ GeV}, \\ m_W &= 80.352 \text{ GeV}, & \Gamma_W &= 2.084 \text{ GeV}. \end{aligned} \quad (\text{C.7})$$

The fixed-order LO, NLO and NNLO computations have been performed with STRIPPER, a C++ implementation of the four-dimensional formulation of the sector-improved residue subtraction scheme [88–91] which has been used for various processes involving flavoured jets [34, 96–99]. Conventional 7-point scale variations are used to estimate uncertainties from missing higher-order corrections.

For the NLO-plus-parton shower predictions we generate large event samples for the process  $pp \rightarrow Z/\gamma^*[\rightarrow \mu^+\mu^-] + \text{jet}$  at NLO accuracy, using the SHERPA [100] and HERWIG 7 (H7) [101] event generators (versions 3.0.0 and 7.3.0 [102], respectively). We use the MC@NLO [103] method for matching the perturbative hard process with the parton shower, as implemented in SHERPA [104] and HERWIG MATCHBOX [105] respectively. In anticipation of the analysis cuts (given in eq. (C.2)) we apply generator cuts:

$$p_T^\mu > 20 \text{ GeV}, \quad |y^\mu| < 2.4, \quad (\text{C.8})$$

$$p_T^{\mu\mu} > 20 \text{ GeV}, \quad m_{\mu\mu} \in (77, 111) \text{ GeV}. \quad (\text{C.9})$$

Hadron decays are disabled, such that stable  $B$  hadrons can be used as unambiguous tags of flavour. A single sample is generated for each generator set-up, which is used to study both  $b$ - and  $c$ -tagged jets. The renormalisation scale  $\mu_R$  is also used as the dipole-shower starting scale.

Within SHERPA, matrix elements at tree level are provided by the internal generators AMEGIC++ [106] and COMIX [107]. Virtual corrections for this process are likewise included internally. The events are showered with SHERPA’s default CSSHOWER [108] based on Catani-Seymour dipoles and hadronised using the cluster hadronisation model [109] as implemented in SHERPA [110, 111] with the parameter tune from refs. [21, 112]. The underlying event is simulated in SHERPA following the Sjöstrand-Zijl model [113].

Within HERWIG, events are showered using either the HERWIG dipole shower [114] or the HERWIG angular-ordered shower [115]. Within the HERWIG dipole shower, by default, heavy quarks are treated as massless within the shower, as in the perturbative hard process; their physical masses are restored prior to hadronisation. In the angular-ordered shower, heavy quarks are treated as massless within the perturbative hard process only, and as massive within the shower. For hadronisation, we employ the HERWIG cluster model [109, 116] and, for comparison, the string model [117] from PYTHIA 8.308 [118] with colour-reconnection model [119] interfaced to HERWIG via the THEP8I C++ package [120]. The Lund string model, together with the angular-ordered parton-shower model, have been tuned to LEP and LHC data sets as described in ref. [121].

In summary, we consider mainly the following event generator setups:

- SHERPA:  $p_t$ -ordered dipole shower with massive  $b$  and  $c$  quarks with a cluster hadronisation model,
- H7 ang.: angular-ordered shower with massive  $b$  and  $c$  quarks with a cluster hadronisation model and with the string hadronisation model,
- H7 dip.:  $p_t$ -ordered dipole shower with massless  $b$  and  $c$  quarks with a cluster hadronisation model.

The fiducial phase-space selection described above, and routines filling the histograms displayed elsewhere in this appendix, have been implemented in a RIVET [122] analysis which was used for the SHERPA and HERWIG results presented below.<sup>15</sup> The code is based on the Rivet analysis CMS\_2017\_I1499471 [94]. In the SHERPA sample, we use the code developed in ref. [123] for the extraction of nonperturbative corrections to simultaneously access the parton and hadron level of a given event.

## C.2 Results

We compare the different algorithms at each of the three possible levels at which a jet or clustering algorithm can be applied:

- (i) to the hard-process final state, at fixed-order LO, NLO and NNLO;
- (ii) to the partonic final-state obtained from a parton shower (matched to NLO with MC@NLO), using the SHERPA and HERWIG dipole showers, and the HERWIG angular-ordered shower;
- (iii) to the hadronic final-state obtained from (ii) with a hadronisation model, using the SHERPA and HERWIG cluster models and the PYTHIA string model.

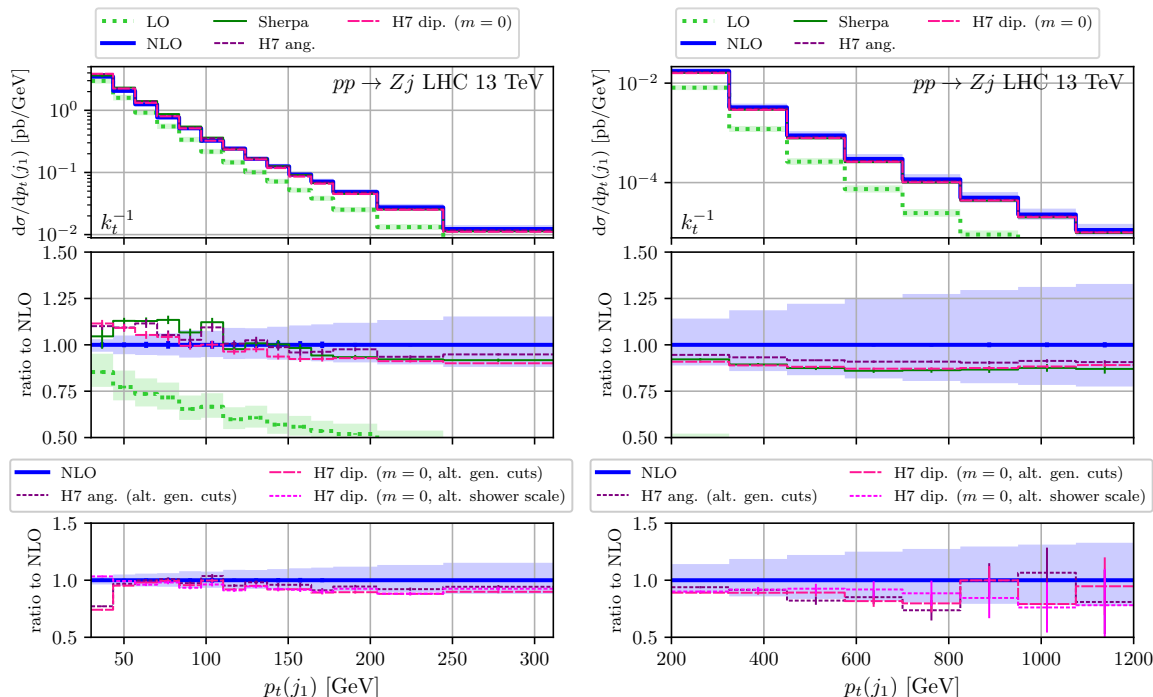
The set-ups described in section C.1 have been chosen to maximise the consistency, and hence comparability, of the predictions between different combinations of codes and algorithms. We studied various jet observables without flavour labelling to validate the consistency. A comparison of the leading jet transverse momentum  $p_t(j_1)$  can be found in figure 13, the left-hand side shows the low- $p_t$  and the right-hand side the high- $p_t$  regime. The top panel displays the absolute results, and the middle panel shows the ratio with respect to fixed-order NLO QCD. We find reasonable agreement between the different event generators and good agreement with fixed-order NLO QCD within the theory uncertainties. The largest differences appear near the phase space threshold at low  $p_t$ , a region which is, however, sensitive to the details of the matching and shower scale. We demonstrate this sensitivity in the lowest panel, which contains HERWIG predictions with altered shower starting scale or generation cuts.

### C.2.1 Sensitivity to perturbative order

Figure 14 shows the perturbative predictions through LO (light green), NLO (blue) and NNLO QCD (red) for the SDF algorithm in the transverse momentum spectrum of the leading  $b/c$  jet ( $p_t(b_1)/p_t(c_1)$ ) in the upper two panels in absolute and relative to NLO QCD. The perturbative corrections are very similar for all the other algorithms and show the known stabilisation of the perturbatively predicted cross section at NNLO QCD and a significant reduction of the scale dependence. The  $|\eta(b_1)|$  ( $|\eta(c_1)|$ ) distributions in figure 15 exhibit similar behaviour.

The different algorithms are compared directly in the lower two panels in figures 14 and 15 at NLO and NNLO QCD, respectively. The NNLO QCD predictions for the four algorithms are comparable within a few percent, and the differences are similar to the uncertainties of missing higher-order corrections. Focusing on the  $p_t(b_1)/p_t(c_1)$  distributions, there are smaller shape differences at low and high  $p_t(b_1)/p_t(c_1)$  of up to 4%. The IFN, GHS, and CMPΩ algorithms seem to approach each other at higher  $p_t(b_1)$  ( $p_t(c_1)$ ).

<sup>15</sup>The analysis code is available at <https://github.com/DReichelt/LH23FlavAlgs>.



**Figure 13.** The inclusive leading jet transverse momentum spectrum at low (left) and high  $p_t$  (right) in central  $pp \rightarrow Z + j$  kinematics. The comparison shows SHERPA, H7 dipole, H7 angular-ordered and FO QCD.

### C.2.2 Sensitivity to parton showering

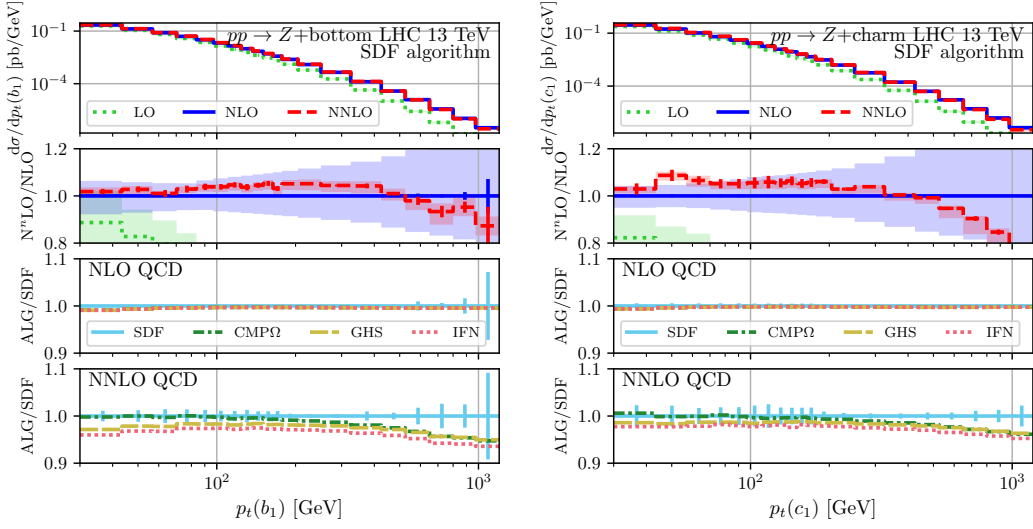
We investigate various jet distribution observables. As outlined earlier, this analysis is conducted using both  $b$ - and  $c$ -flavoured objects. Our primary observation is that the qualitative effects are similar in both cases, though they tend to be more pronounced for charm flavour due to the smaller mass of the  $c$  quark. It turned out that the sensitivity to the different flavour labelling algorithms depends significantly on the mass treatment within the parton shower.

We start by considering the transverse momentum of the leading tagged jet,  $p_t(b_1)$  and  $p_t(c_1)$  for the two flavour selections, in figure 16.<sup>16</sup> The left column of plots shows results for  $b$ -labelled jets, while the right column corresponds to the  $c$ -flavour selection. The top panel displays the absolute results for the different event generators at the hadron level, using the applied standard anti- $k_t$  algorithm with mod2-flavour labelling. Below, we present a comparison of the different algorithms for each event generator in three panels.

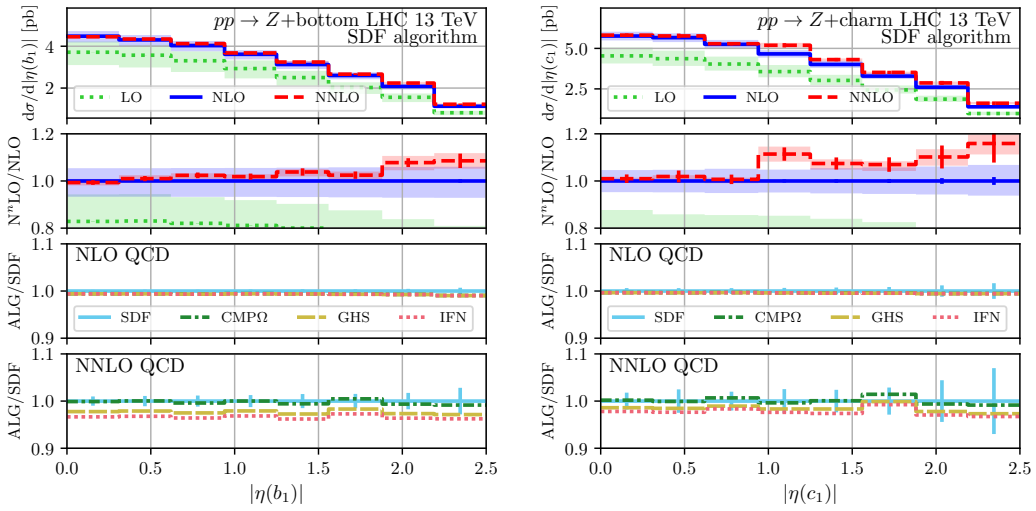
The predictions from SHERPA and HERWIG angular-ordered agree very well with each other across the spectrum. Strikingly, the HERWIG dipole prediction is much higher than the other two, which can be traced back to an enhanced production of flavoured particles if the  $b/c$  quarks are treated massless in the parton shower. We discuss this effect in more detail below.

We now turn our attention to the comparison of the jet algorithms. First, we can observe, by looking at the predictions with the standard anti- $k_t$  algorithm, that there is a growing

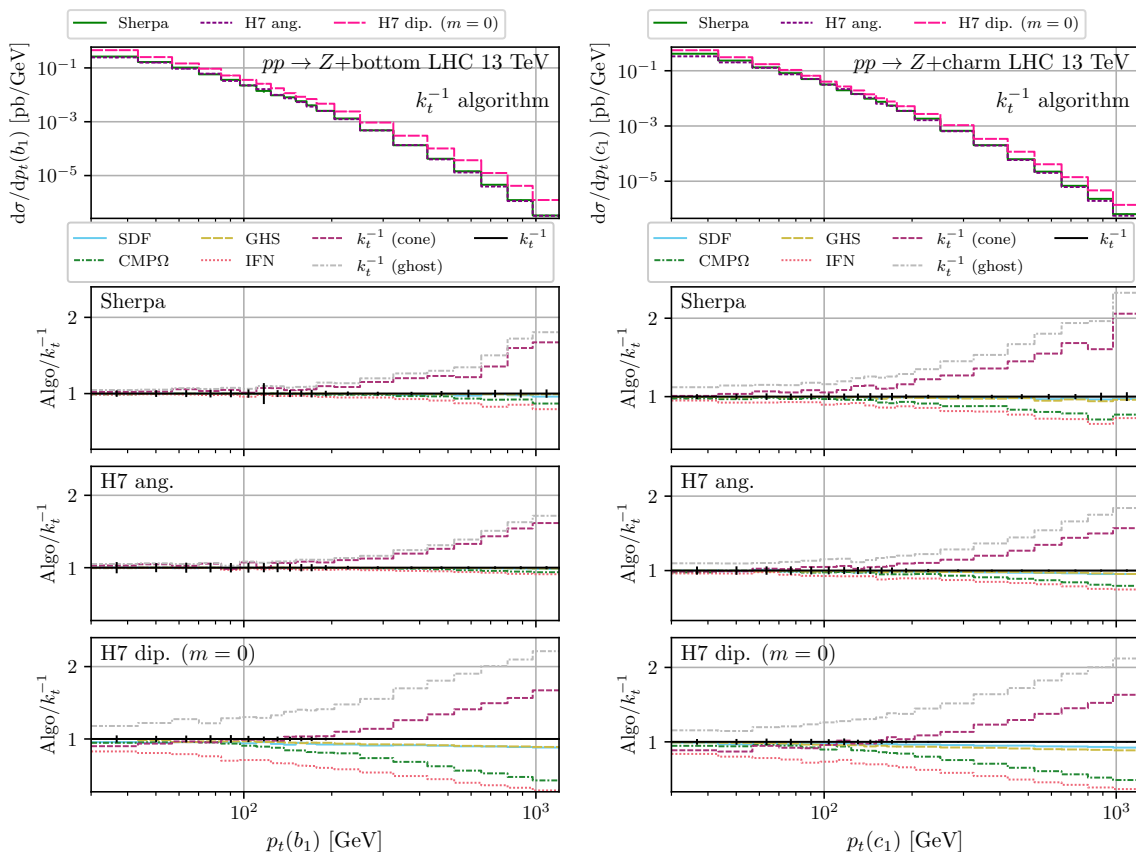
<sup>16</sup>Many more observables have been computed and can be found in the PDF files attached to this submission.



**Figure 14.** Perturbative predictions for the transverse momentum spectrum of the leading flavoured jet in central  $pp \rightarrow Z + j$  kinematics. The left column shows bottom-labelled jets and the right column charm-labelled jets. The upper two panels show the differential cross section obtained with the SDF algorithm through different perturbative orders (LO — light green, NLO — blue, NNLO — red), absolute and relative to NLO QCD, respectively. The envelope of the 7-point scale uncertainties is shown as coloured bands. Different jet algorithms (CMPΩ — green, SDF — light blue, GHS — yellow, IFN — light red) are compared at NLO and NNLO QCD in the lower two panels. The Monte Carlo errors indicated by the vertical lines are almost 100% correlated among the different algorithms and are only shown for the central prediction.

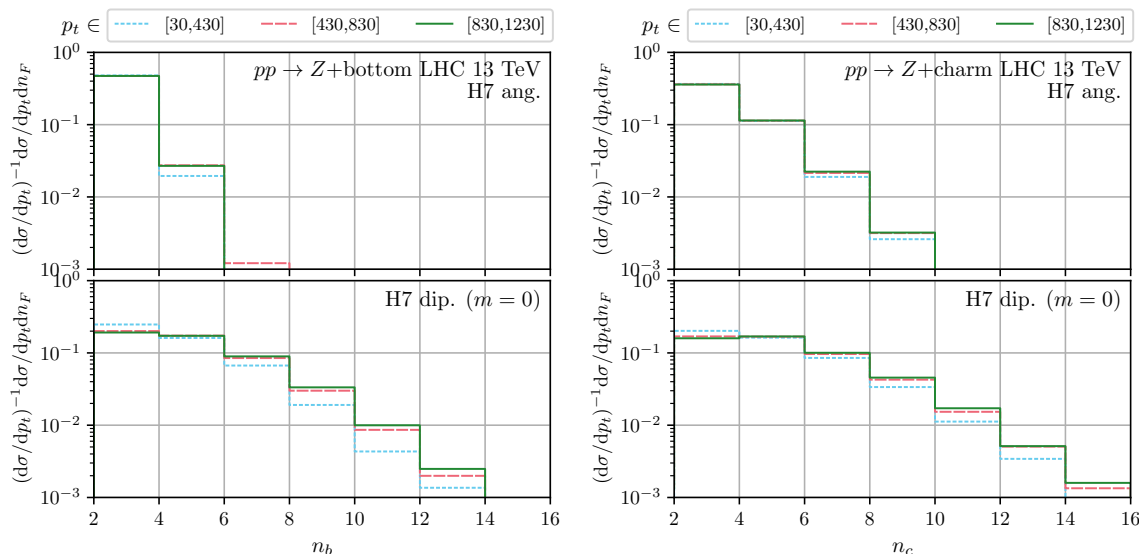


**Figure 15.** Same as figure 14 but for the rapidity distribution of the leading flavoured jet.



**Figure 16.** NLO+PS predictions from SHERPA and HERWIG for the transverse momentum of the leading flavoured jet in central  $pp \rightarrow Z + j$  kinematics. The left column shows bottom-labelled jets and the right column charm-labelled jets. The top panel shows the differential cross section obtained with the standard anti- $k_t$  algorithms with mod2-flavour combination. Different jet algorithms (CMPΩ — green, SDF — light blue, GHS — yellow, IFN — light red), as well as anti- $k_t$  with different labelling: mod2 (black), ghost (grey) and cone (wine-red)) are compared for each generator in the lower three panels. The Monte Carlo errors indicated by the vertical lines are almost 100% correlated among the different algorithms and are only shown for the central prediction.

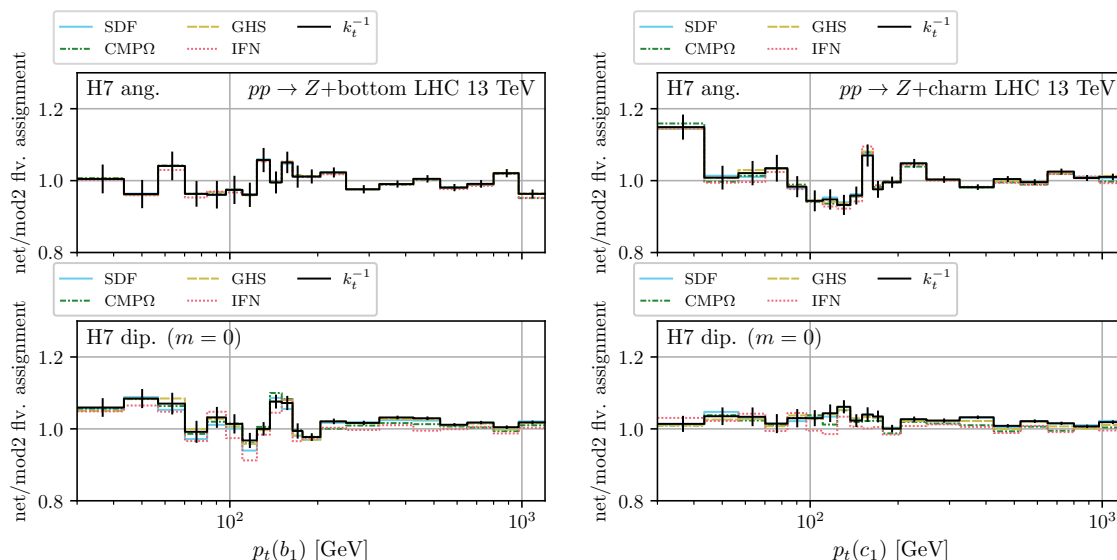
sensitivity to the different labelling strategies, i.e., mod2, cone and ghost. We want to note that since the decays of hadrons are deactivated in the samples used for this study, the ghost labelling and any-flavour scheme are the same because the ghosts will always be exactly collinear to the hadrons, i.e., zero clustering distance, and thus clustered immediately. Therefore, we can conclude that the difference between the ghost and mod2 flavour originates from the removal of  $g \rightarrow q\bar{q}$  splittings of the latter. Furthermore, we can observe that differences also appear at high  $p_t$  when comparing the actual IRC-safe algorithms with anti- $k_t$ . Comparing the  $b$  and  $c$  results, we see that in the  $c$ -label case the sensitivity is larger, reaching up to 10–20% at 1 TeV. We find consistent results between the angular-ordered HERWIG and SHERPA event samples and a significantly enhanced sensitivity for the dipole shower. This enhancement is again due to the increased flavour production within this sample, which naturally increases the sensitivity to the algorithms.



**Figure 17.** The normalised differential cross section as a function of the flavour multiplicity (the number of hadrons with the corresponding flavour) predicted by a massive (HERWIG angular-ordered — top) and massless (HERWIG dipole — bottom) parton shower, in three transverse momentum bins of the leading flavoured jet. The  $b$  case is shown on the left and the  $c$  case on the right.

The striking feature that the massless dipole HERWIG sample predicts a much larger cross section and is much more sensitive to the jet algorithms can be traced back to the amount of flavoured hadrons produced. In figure 17 we show the fraction of the differential cross section as a function of the flavour multiplicity, i.e., the number of flavoured hadrons. We show results for HERWIG angular (with massive quarks in the shower) and for the HERWIG dipole shower (massless quarks in the shower) for the  $b$  and  $c$  case separately. We find a significantly enhanced hadron multiplicity in the case of the massless treatment within the shower. Additionally, we observe that the multiplicities for the massive quarks in the shower are larger for the charm case than for the bottom case, as one would expect naively. In contrast, the difference predicted by the shower with massless treatment is significantly smaller. Consequently, the differences between massless and massive treatments are larger for bottom than for charm.

As discussed in section A, the different flavoured jet algorithms can have a different flavour assignment on the clustered jets during the clustering sequence itself. The impact of such a choice is shown in figure 18 where we directly compare the mod2 and net-flavour assignment for the massive HERWIG angular and massless HERWIG dipole sample. We want to stress that in both cases, the clustering has been performed with the net-flavour recombination scheme. Overall, the ratio is close to 1, indicating only minor differences in the predicted flavoured jet rates. In most cases, the statistical uncertainties are of comparable size. In particular, in the high- $p_t$  region, we observe a negligible difference between the two flavour assignments. A systematic difference can be seen for the massless HERWIG dipole sample at low  $p_t$  of about 5%, which is, however, independent of the jet algorithm used. The most relevant potential source of differences can come from jets containing two particles of flavour



**Figure 18.** Comparison between the net and mod2 flavour label assignment on jets which are recombined with the net-flavour recombination scheme in case of the massive HERWIG angular shower (top row) and the massless HERWIG dipole shower (bottom row). The left column shows the bottom case, and the right column shows the charm case. The ratio is shown for a collection of algorithms (CMPΩ — green, SDF — light blue, GHS — yellow, IFN — light red, anti- $k_t$  — black).

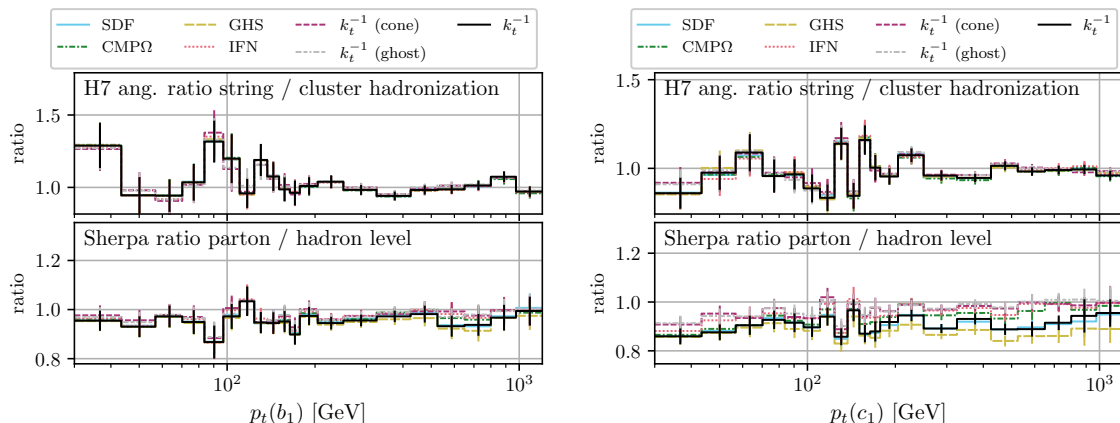
$f$ , which in the case of mod2 will be labelled as unflavoured and as flavoured with net-flavour. These cases are not infrared enhanced at NNLO since an  $ff$  flavour cannot arise from a gluon splitting. As such, we do not expect significant differences and, if at all, of geometric nature (i.e., proportional to the area of the jet) due to accidental wide-angle radiation of a flavour into another flavoured jet.

### C.2.3 Sensitivity to hadronisation

To investigate the sensitivity of the jet algorithms to hadronisation, we conducted two additional comparisons. First, we investigated the impact of the hadronisation model in HERWIG on the results, which we show differentially in the leading flavoured jet  $p_t$  in the top row in figure 19. We compare the cluster and string models, and we observe only minor effects, which are generally within the statistical uncertainties. The second comparison is between the parton and hadron level within the SHERPA sample, which we show in the lower row in figure 19. For the charm case, we observe some minor effects, differentiating the algorithms in particular at high  $p_t$ .

### C.2.4 Comparison of fixed-order and NLO+PS

Looking at the high- $p_t$  end of the  $p_t(b_1)/p_t(c_1)$  spectrum (see figure 14 and figure 16), it appears that the NNLO fixed-order prediction and the NLO+PS predictions behave differently. Qualitatively, all algorithms appear similar in the case of fixed-order prediction, while they diverge (in the sense that they differ more and more) for NLO+PS. The scale of the differences is also very different. We see at most a few percent for the fixed-order calculation; for the



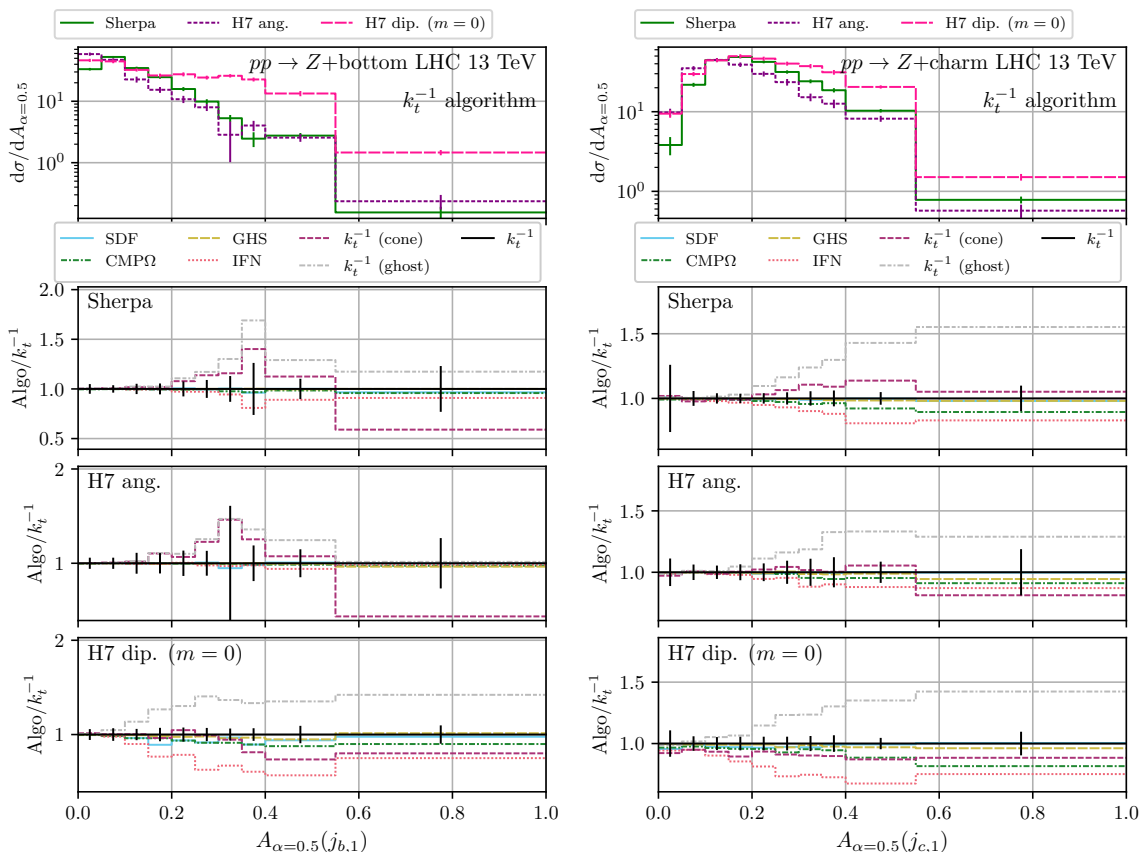
**Figure 19.** The top row shows a comparison between the string and cluster hadronisation in HERWIG for the transverse momentum spectrum of the leading  $b$ -labelled jet on the left and  $c$ -labelled jet on the right. The lower row shows the ratio between parton level and hadron level in the SHERPA sample.

NLO+PS, we see differences of  $\mathcal{O}(10\%)$ . When directly comparing the NLO+PS and fixed-order predictions for a fixed jet algorithm, which we show in figure 7, we can observe that there is generally a fair agreement in the sense that the NLO+PS is (almost) within the NLO QCD uncertainty. However, we observe shape differences in the high- $p_t$  region for the SDF, GHS and CMP algorithms, while the IFN results are the same using NLO FO and NLO+PS. The shape differences are consistent with the spread we see when comparing the algorithms at NLO+PS and FO. The effect is more clearly visible for the  $c$ -jet case than for the  $b$ -jet case, again due to the larger sensitivity in the  $c$  case.

In other words, the agreement between NLO+PS and FO for SDF, GHS, and CMP depends on the mass parameter of the flavoured quark; the smaller the mass, the worse the agreement. The reason for this observation is likely to be found in the multiplicity of flavoured quarks, which is higher for smaller masses, leading to more significant differences between the algorithms. IFN shows a much smaller dependence and therefore provides a more robust prediction.

### C.2.5 Jet substructure observables

We also take a first look at the jet substructure of the tagged jets. There are several subtleties to consider when discussing the jet substructure of jets tagged with IRC-safe jet algorithms. We here only take a simple look at jet substructure observables defined on all kinematical constituents of the anti- $k_t$  jets that are labelled as  $b$  flavoured by the respective algorithm, i.e., we ignore any information about particles outside in the jet that the flavour algorithm identified to belong to the jet in addition to the raw anti- $k_t$  constituents (and vice versa, we keep constituents even if they did not contribute to the jet flavour label). The only exception is the CMP algorithm that slightly modifies the jet kinematics and hence defines jets that can not necessarily be associated with the raw anti- $k_t$  jets. We concentrate on the angularity ( $\alpha = 0.5$ , Les Houches angularity (LHA)) in figure 20 and the jet mass normalised to the transverse momentum of the jet in figure 21.

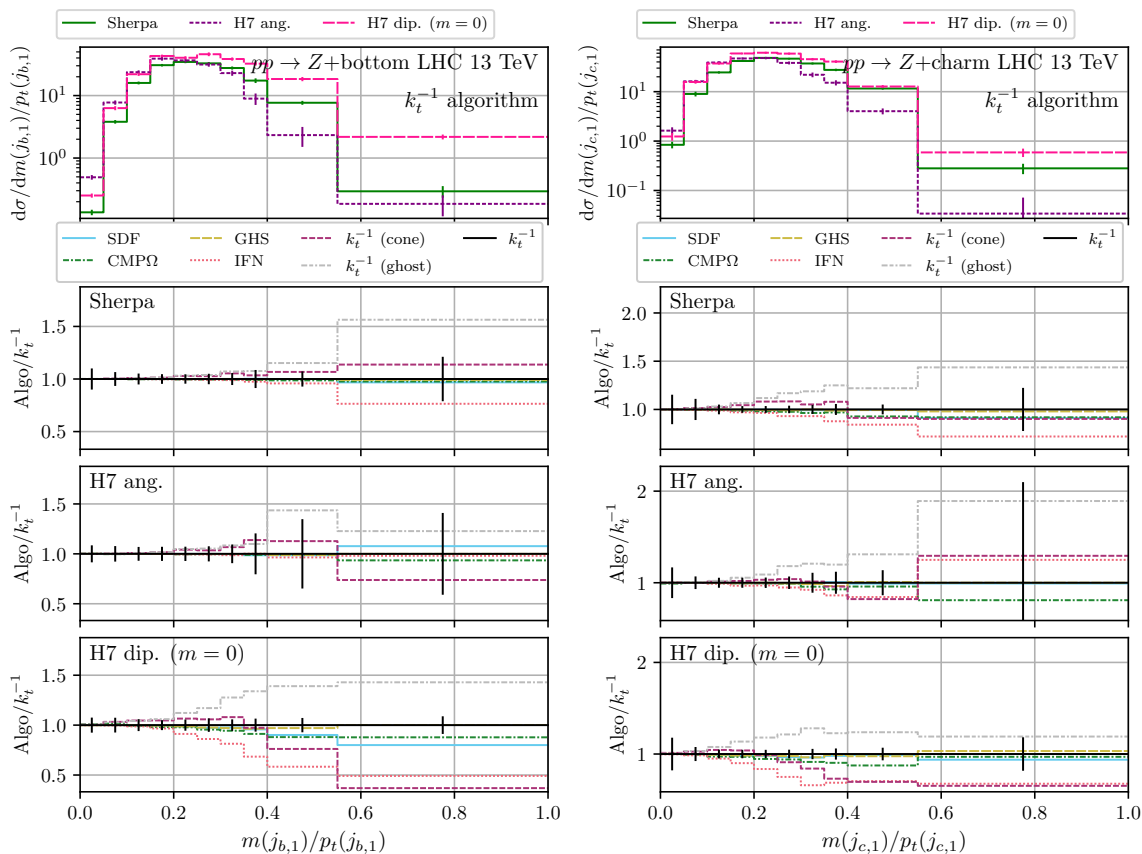


**Figure 20.** Same structure as figure 16 but for the Les Houches angularity of the leading flavoured jet,  $A_{\alpha=0.5}(j_{b,c,1})$ .

We illustrate the general features of the distributions in the respective top panels of the figures, for the case of anti- $k_t$  jets tagged with the mod2 prescription. SHERPA and HERWIG’s angular-ordered shower (using massive splitting functions) agree reasonably well, while the distribution is significantly shifted to larger values of jet mass and angularity, indication the presence of additional or harder radiation inside the jet. The differences between the showers are generally more pronounced for the LHA compared to the jet mass.

The different algorithms for each simulation are compared in the three lower panels. Comparing  $b$  (left plot) and  $c$  jets (right), we first note that, as expected, the behaviour is the same when considering the HERWIG dipole shower that treats both as equally massless quarks. For the other showers the differences between the algorithms are significantly enhanced for the  $c$ -jet case, while they represent a minor correction for  $b$  jets, consistent with our earlier observations.

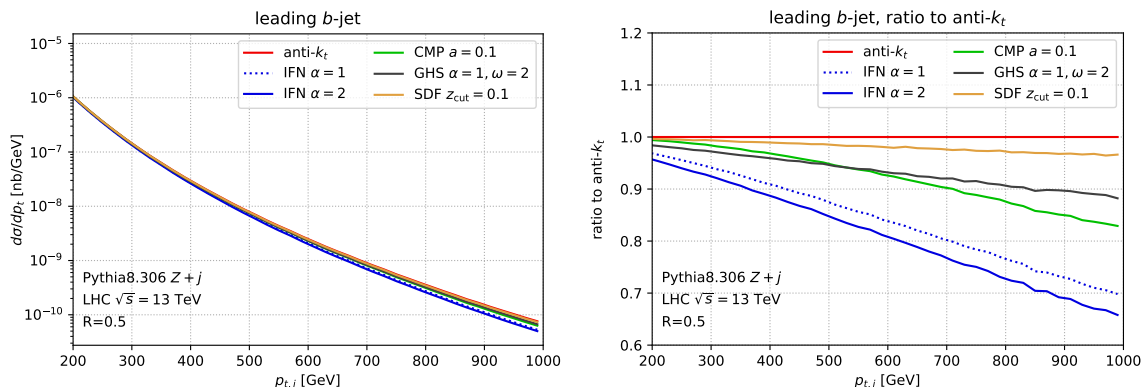
Focusing on the larger visible differences in  $c$  jets, we observe consistent qualitative results between the different showers. SDF, GHS, and CMP behave roughly similarly to anti- $k_t$  (and hence each other). This should not be surprising, since they behave similarly when considering the  $p_t$  spectrum, indicating that they label roughly the same jets as flavoured, and we simply evaluate the substructure observables of those jets. We can hence mainly learn something about the characteristics of jets that are labelled in some algorithm but not the other. for example, the excess  $b$ -jet cross section predicted by the ghost labelling



**Figure 21.** Same structure as figure 16 but for the normalised jet mass of the leading flavoured jet,  $m(j_{b/c,1}) / p_t(j_{b/c,1})$ .

technique appears to be associated with a large jet mass and angularity. The reduced cross section in the IFN algorithm relative to anti- $k_t$  on the other hand appears to be largely based on removing jets with large mass and angularities. This reinforces the idea that differences are mainly caused by configurations with a lot of radiation (and hence additional flavoured particles) in the jets. The cone-labelled jets tend to have a similar spectrum to anti- $k_t$  and the algorithms behaving similarly, tentatively shifted towards the soft end of the distribution.

Having taken a first look at the jet substructure of flavour-labelled jets with the new generation of algorithms, we leave a detailed study, exploring the effects observed above and going beyond this simple procedure, to future studies.



**Figure 22.** Left: the distribution of the transverse momentum of the leading  $b$ -labelled jet  $p_{t,b}$  for net-flavour anti- $k_t$  (red), IFN ( $\alpha = 1$  and  $\alpha = 2$ , blue), CMP ( $a = 0.1$ , green), GHS ( $\alpha = 1$ ,  $\omega = 2$ , black) and SDF ( $\beta = 2$ ,  $z_{\text{cut}} = 0.1$ , yellow). Right: the ratio to net-flavour anti- $k_t$ .

## D $pp \rightarrow Z + b$ across kinematic regions

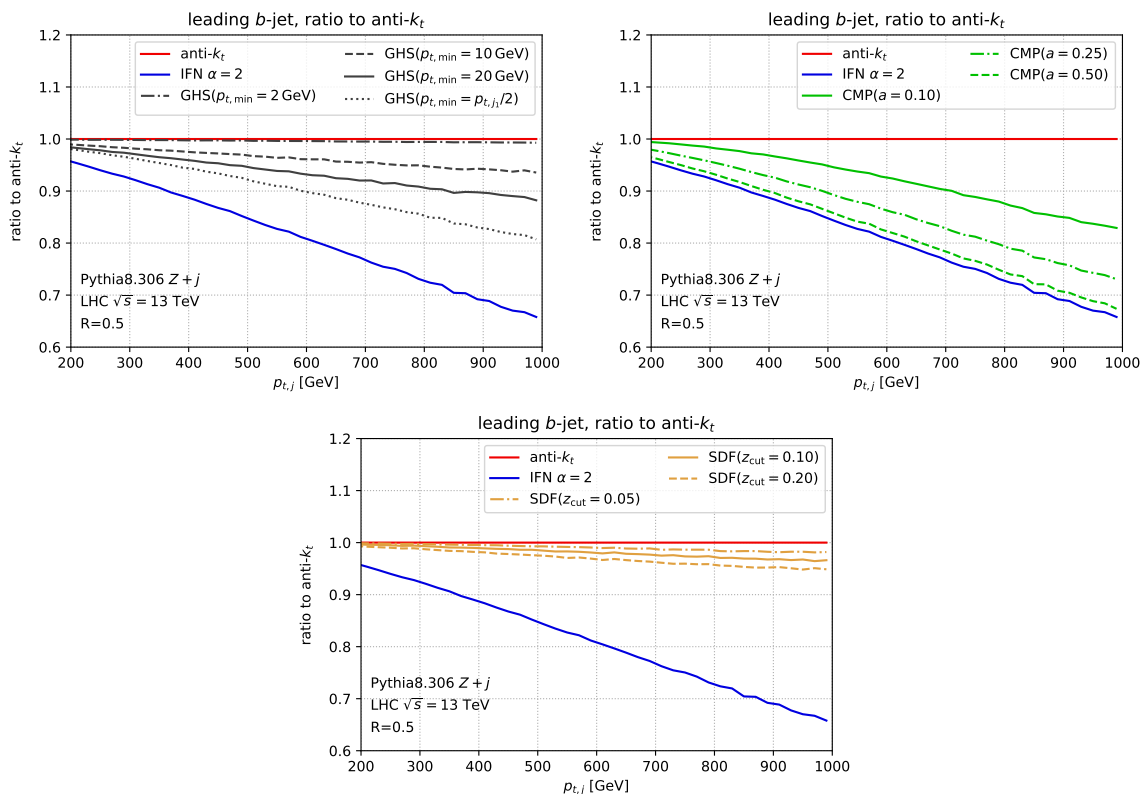
**Authors:** Gavin P. Salam, and Ludovic Scyboz.

Here we investigate further the results of the last section, in particular deviations of the various flavoured jet algorithms from anti- $k_t$  spectra (see, e.g., the transverse momentum of the  $b$ -labelled jet in figure 16). For this purpose we generate a sample of  $Z + j$  events at parton level (i.e., without hadronisation) and follow a roughly similar selection to section C in our analysis. The events are produced by PYTHIA 8.306 at leading order, at a centre-of-mass energy of  $\sqrt{s} = 13$  TeV, with no flavour constraints on the Born  $Z + j$  process. We require, as before, a  $Z$  boson to be reconstructed in an invariant mass range  $m_{\mu^+\mu^-} \in [71 \text{ GeV}, 111 \text{ GeV}]$ , where the muons have to satisfy the set of cuts ( $p_{t,\mu} > 20 \text{ GeV}$ ,  $|\eta_\mu| < 2.4$ ), and we cluster jets with a radius  $R = 0.5$ .

In figure 22 we consider the distribution of the transverse momentum of the leading identified  $b$  jet, focusing on a similarly high range of  $p_{t,b}$  (from 200 GeV to 1 TeV), as in figure 16. As can be seen in particular in the ratio to net-flavour anti- $k_t$  (figure 22, right), while the various flavoured jet algorithms differ from the anti- $k_t$  distribution by only a couple of percent at  $p_t \sim 300$  GeV (and by  $\sim 8\%$  for IFN  $\alpha = 2$ ), all algorithms start to develop larger differences at higher values of  $p_t$ .

Furthermore, this observation quantitatively depends on the actual set of parameters used by the flavoured algorithms. To illustrate this, we show the same ratio to the anti- $k_t$  result in figure 23, for variations of the IR-regularising parameters: the minimum transverse momentum  $p_{t,\text{min}}$  for the input jets in GHS (top left), the parameter  $a$  for CMP (top right), and the SoftDrop cut  $z_{\text{cut}}$  for SDF (bottom). The default values used in figure 22 are pictured as full lines, and the curves corresponding to variations of the parameters are shown as dashed/dotted lines. In particular the CMP and GHS algorithms tend to the anti- $k_t$  distribution in a limit where their defining parameters go to zero (in the exact limit, those algorithms would become IRC unsafe).

The fact that the flavoured jet algorithms differ in assigning flavour labels is not surprising. As an aid to understand whether the  $b$ -label assignment is sensible, however, we investigate

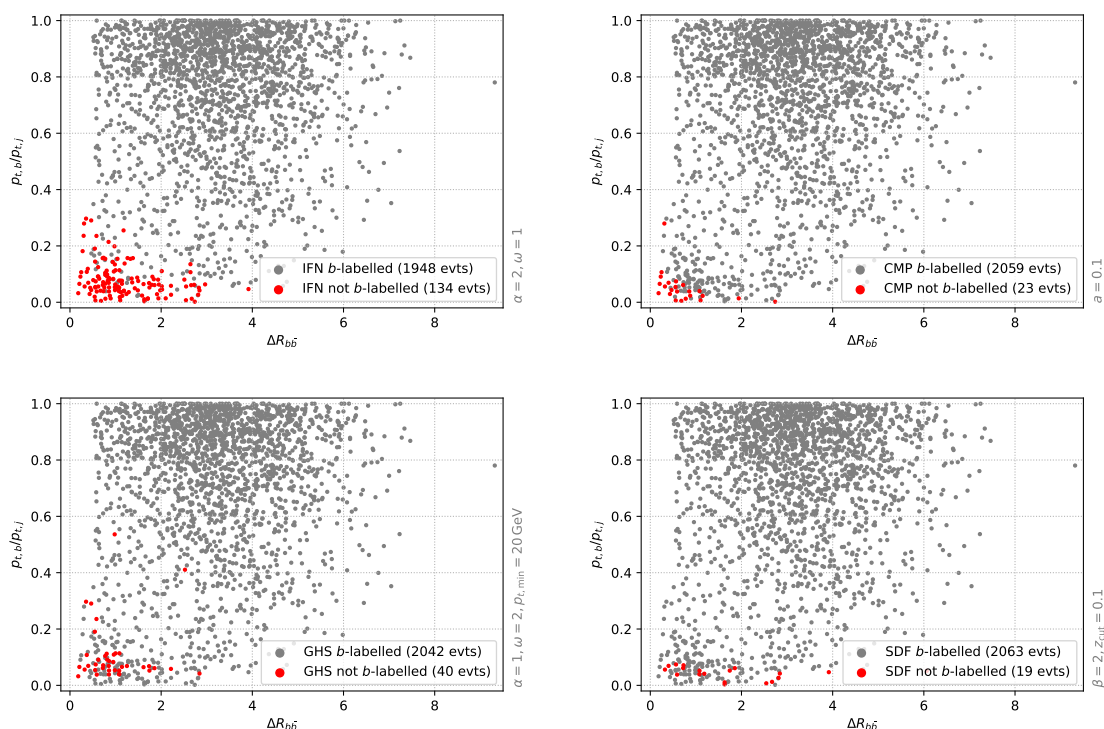


**Figure 23.** Same as figure 22, right, for variations of the flavoured algorithms’ parameters, as explained in the text. Top left: GHS, top right: CMP, bottom: SDF.

a set of individual  $Z + j$  events, again generated with PYTHIA, with no flavour constraints in the generation. We select events which have at least two  $b$  quarks at parton level, and where the leading anti- $k_t$  jet satisfies  $p_{t,j} > 200$  GeV and is classified as a  $b$  jet by the net-flavour anti- $k_t$  algorithm.

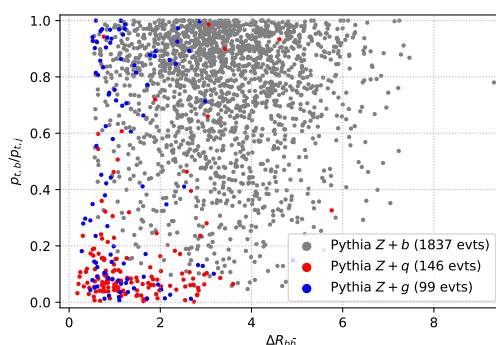
In figure 24 we depict such events as dots in a scatter plot, where the  $x$  coordinate is the angular distance between the two  $b$  quarks in the event,  $\Delta R_{b\bar{b}}$ , and the  $y$  coordinate is the transverse momentum fraction,  $p_{t,b}/p_{t,b\text{-jet}}$ , carried by the  $b$  quark in the  $b$  jet as identified with anti- $k_t$  (if there are two  $b$  (anti-)quarks,  $p_{t,b}$  is the larger of the two  $p_t$ ’s). Roughly, one would expect a majority of events in the bottom-left region of the plot to have originated from soft large-angle gluon splitting  $g \rightarrow b\bar{b}$ , where one  $b$  entered the anti- $k_t$  jet, while the other did not. Regions of relatively hard  $b$ -jet constituents and angular separations of order one or larger,  $\Delta R_{b\bar{b}} \gtrsim 1$ , will typically be populated by genuine  $b$  jets produced in the hard collision (a large  $\Delta R_{b\bar{b}}$  generally indicates that the other  $b$  quark that is outside the jet is collinear to the beam, generated by an initial-state  $g \rightarrow b\bar{b}$  splitting). We colour events in grey if the leading jet is  $b$  labelled by *both* net-flavour anti- $k_t$  and the flavoured algorithm; in contrast, if the jet is  $b$  labelled by anti- $k_t$  but not by the flavoured algorithm (because, e.g., the latter neutralised the  $b$  flavours), we show these points in red.

As evidenced by the set of plots in figure 24, the flavoured jet algorithms remove  $b$  flavour to varying degrees, with IFN removing the most. For comparison, in figure 25 we show the



**Figure 24.**  $Z + j$  events with identified  $b$  jets (grey: labelled as  $b$  by both net-flavour anti- $k_t$  and the algorithm under consideration, red: labelled as  $b$  by anti- $k_t$ , but not by the flavoured algorithm), as a function of  $\Delta R_{b\bar{b}}$ ,  $p_{t,b}/p_{t,b\text{-jet}}$ . Top left: IFN, top right: CMP, bottom left: GHS, bottom right: SDF.

same jets ( $b$  labelled by net-flavour anti- $k_t$ ), with a colouring that corresponds to the (truth level) flavour generated in the hard process by PYTHIA at leading order:  $Z + b$  events (in grey) are the most common source of  $b$  jets, and one may think of them as producing “genuine”  $b$  jets;  $Z + q$  events (in red) would typically not lead to hard  $b$  jets, but can occasionally do so, most frequently when an additional soft gluon from the parton shower splits into a  $b$  and a  $\bar{b}$ , one of which enters the leading jet, while the other goes outside the jet (physically, we would consider this to be a fake  $b$  jet); for  $Z + g$  events (in blue), the situation is similar, but additionally the hard gluon can itself split to  $b\bar{b}$ , which is the source of those blue scatter points that are at large  $p_{t,b}$  with a moderate  $\Delta R_{b\bar{b}}$  (physically, we would consider these to be genuine  $b$  jets). For the same selection, we also showed the distribution of  $\frac{p_{t,b}}{p_{t,b\text{-jet}}}\Delta R_{b\bar{b}}$  in figure 6 (right), which highlights the separation of events with a hard  $b$  (at high values of the variable) from those with a hard  $g$  or a light  $q$  (at low values), similarly to the scatter plots of figures 24 and 25. At the lower end, where standard anti- $k_t$  labels many hard  $g/q$ -initiated jets as  $b$  jets, the jet flavour algorithms differ in how they assign  $b$  flavour, with IFN being the closest to the truth labelling from the hard matrix element in PYTHIA. One should keep in mind that in PYTHIA the map from the Born flavour to the final event depends on the details of the parton shower. Nevertheless a comparison between figures 24 and 25 gives an idea of the degree of similarity between the flavour classification performed by the jet algorithms in figure 24, and a jet flavour as determined by a LO matrix-element-based likelihood.



**Figure 25.** Same as figure 24, with the colour of the MC data points determined by the truth-level flavour of the hard process generated by PYTHIA:  $Z + b$  (in grey, mostly genuine  $b$ -jets),  $Z + q$  (in red, mostly light-quark jets with a soft  $b$  contaminating them),  $Z + g$  (in blue, with some soft- $b$  contamination giving “fake”  $b$  jets and some hard  $g \rightarrow b\bar{b}$  splitting giving genuine  $b$  jets).

## E $pp \rightarrow Z + c$ in LHCb kinematics: NNLO and NLO+PS

**Authors: Alexander Huss and Giovanni Stagnitto.**

In this appendix, we consider the associated production of a  $Z$  boson with a charmed jet in the forward region, in particular focusing on the LHCb measurement [27] at  $\sqrt{s} = 13$  TeV. Such a measurement is valuable for the determination of intrinsic charm component in the proton [124–126]. Precise theory predictions for  $Z + c$ -jet production have been provided in ref. [127], both at the fixed-order level (up to NNLO) and with NLO predictions matched to a parton shower. The purpose of the current study is to extend the analysis of ref. [127] to the new generation of flavoured jet algorithms.

The fixed-order theory predictions up to NNLO are obtained with the public NNLOJET code [128], with the additional flavour tracking procedure at parton level developed for the studies in refs. [127, 129–131]. For the purpose of the present analysis, we have interfaced NNLOJET to FASTJET to access the implementation of the flavoured algorithms. Both the flavour tracking capabilities in NNLOJET and the interface to FASTJET will be included in a future version of NNLOJET.

We further consider NLO predictions matched with a parton shower (NLO+PS), obtained within the MADGRAPH5\_AMC@NLO (v. 2.7.3) [132] framework interfaced to PYTHIA (v. 8.243) [133] (default  $p_T$ -ordered parton shower) or HERWIG (v. 7.2.2) [101, 134, 135] (default angular-ordered parton shower). Both the NLO+PS predictions shown here are at the hadron level, with stable heavy hadrons. It was noticed in ref. [127] that Multiple Particle Interaction (MPI) effects are very sizeable in the forward region and for charmed observables. Hence, we will show the impact of MPI effects on NLO+PS predictions, by turning on and off the MPI flag in the PYTHIA sample (we have checked that the impact of MPI is similar when using HERWIG).

For the sake of completeness we report here the numerical setup and the fiducial cuts of ref. [127]. We use the PDF4LHC21 Monte Carlo PDF set [95]. We work in the electroweak  $G_\mu$  scheme, using a complex mass scheme for the unstable internal particles, with the following values for the input parameters:  $M_Z^{\text{os}} = 91.1876$  GeV,  $\Gamma_Z^{\text{os}} = 2.4952$  GeV,  $M_W^{\text{os}} = 80.379$  GeV,  $\Gamma_W^{\text{os}} = 2.085$  GeV, and  $G_\mu = 1.1663787 \times 10^{-5}$  GeV $^{-2}$ .

We retain events featuring jets with  $20 \text{ GeV} < p_{t,j} < 100 \text{ GeV}$  and  $2.2 < \eta_j < 4.2$  (note the upper cut on the transverse momentum of jets). Charged leptons are required to have  $p_{t,\ell} > 20 \text{ GeV}$  and  $2.0 < y_\ell < 4.5$ , with an invariant mass  $M_{\ell\bar{\ell}} \in [60, 120] \text{ GeV}$ . A minimum angular distance between leptons and jets is further imposed,  $\Delta R(j, \ell) > 0.5$ .

To define (flavoured) jets, we adopt several IRC-safe algorithms with variation of parameters:

- GHS algorithm, applied to anti- $k_t$  jets with  $R = 0.5$ , with parameters  $(\alpha = 1, \omega = 2)$ ;
- IFN algorithm, with anti- $k_t$  distance and  $R = 0.5$ , with parameters  $(\alpha = 1, \omega = 2)$  or  $(\alpha = 2, \omega = 1)$ ;
- CMP algorithm, with  $R = 0.5$ , with parameter  $a = 0.1$  or  $a = 0.25$ ;
- SDF algorithm, applied to anti- $k_t$  jets with  $R = 0.5$ , with parameters  $(\beta = 1, z_{\text{cut}} = 0.1)$ ,  $(\beta = 2, z_{\text{cut}} = 0.1)$  or  $(\beta = 2, z_{\text{cut}} = 0.05)$ .

In the case of NLO+PS predictions, we further consider two IRC unsafe variants. We first reconstruct jets with the flavour-agnostic anti- $k_t$  algorithm with  $R = 0.5$  and in the end we assign flavour in two different ways:

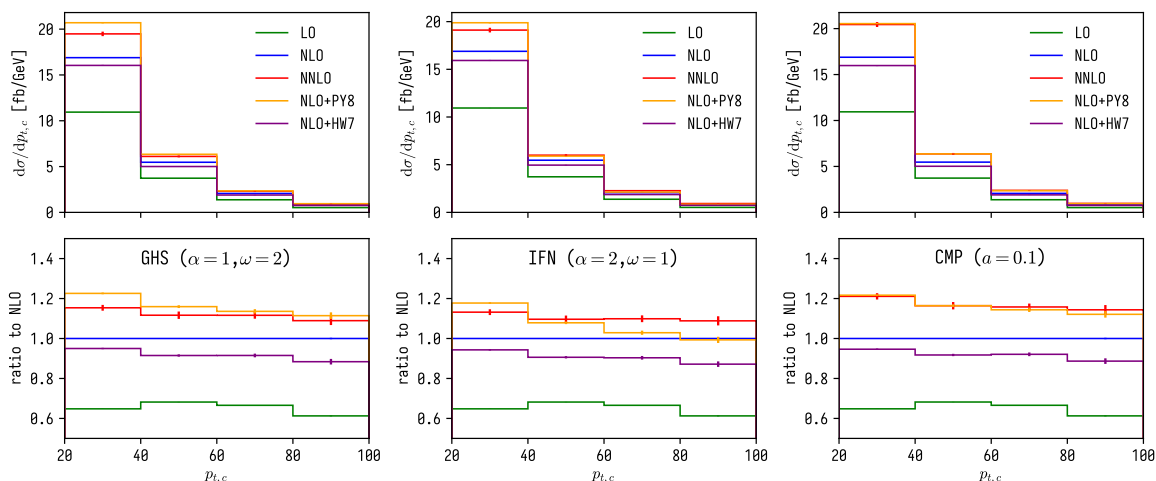
- “anti- $k_t$  naive”: we assign flavour to jets according to their flavoured constituents (here we adopt the mod2 recombination).
- “cone labelling” (matching the definition of charm jets used in ref. [27]): we check for the presence of at least one flavoured hadron satisfying both  $p_{t,c \text{ hadron}} > 5 \text{ GeV}$  and  $\Delta R(j, c \text{ hadron}) < 0.5$ .

In addition, events are only retained if the flavour-labelled  $c$  jet is the jet carrying the largest transverse momentum of reconstructed jets passing the selection cuts.

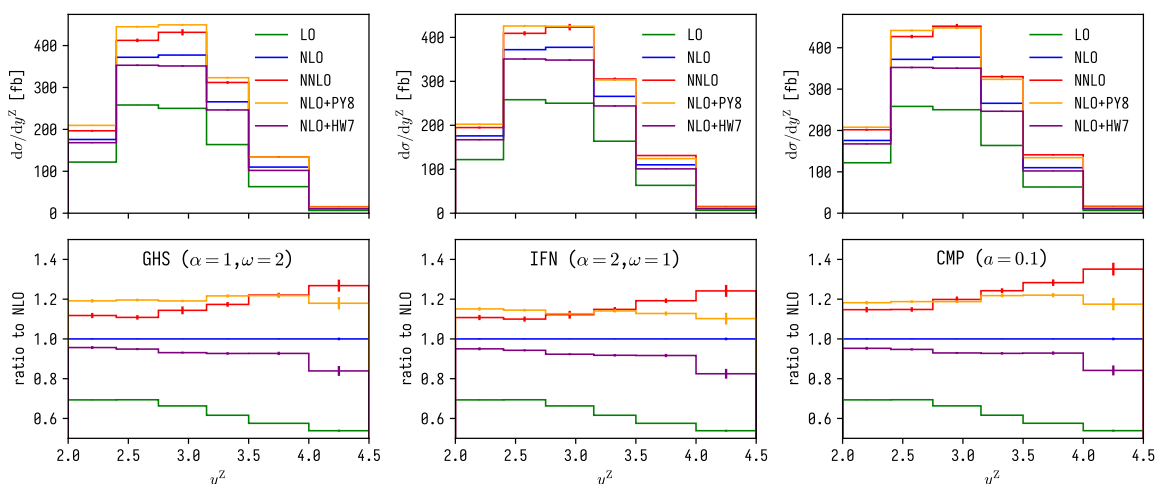
For each set of plots, we will consider two observables: the transverse momentum of the leading- $p_t$  charm jet,  $p_{t,c}$ , and the rapidity of the Z-boson,  $y^Z$ . We will further denote as “NLO+PY8” the NLO+PS results with PYTHIA, and “NLO+HW7” the NLO+PS results with HERWIG.

In figures 26 and 27 we study the relative agreement between fixed-order predictions and NLO+PS predictions for each flavoured jet algorithm. The algorithms or variations of parameters not shown in figures 26 and 27 are very similar to the ones considered. For both observables, we see that NLO+PY8 predictions are very similar to the fixed-order NNLO results, both consistently above NLO by 20%. NLO+HW7 is instead similar or slightly smaller than the NLO result. This may be due to the fact that  $p_t$ -ordered parton showers are better able to populate the forward region starting from hard events in the central region compared to angular-ordered parton showers. The same behaviour seems to hold for all of the jet algorithms.

In figures 28 and 29 we compare flavoured jet algorithms, both IRC-safe and IRC-unsafe variants, at the NLO+PS level. The “cone labelling” stands as an outlier, both for NLO+PY8 and NLO+HW7: compared to “anti- $k_t$  naive” (used as reference in the upper plots of figures 28 and 29), “cone labelling” is systematically higher than predictions with other algorithms,

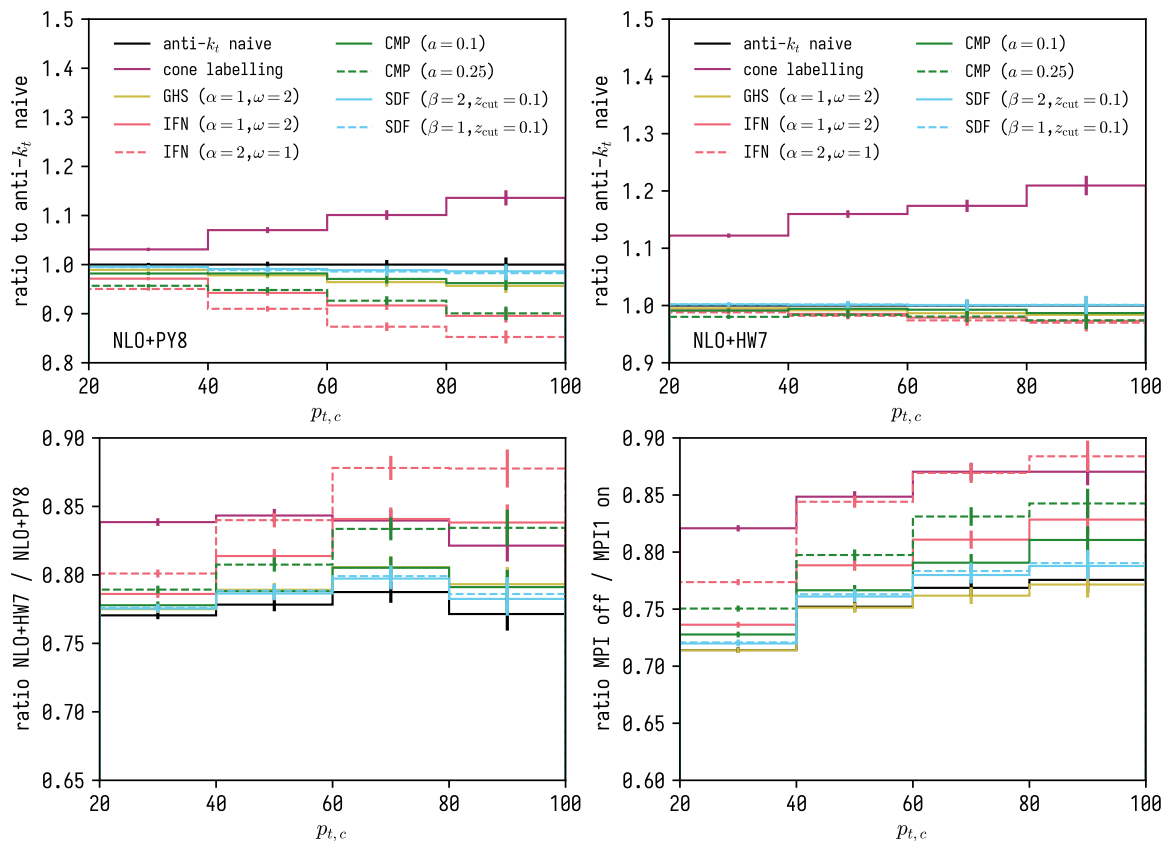


**Figure 26.** Comparison of predictions at fixed-order and NLO+PS for the transverse momentum of the leading- $p_t$  charm jet,  $p_{t,c}$ .



**Figure 27.** Comparison of predictions at fixed-order and NLO+PS for the rapidity of the Z-boson,  $y^Z$ .

ranging from 5% to 10% above “anti- $k_t$  naive” in NLO+PY8 and up to 20% above “anti- $k_t$  naive” in NLO+HW7. This is likely due to the difference between the IRC-unsafe requirement of having at least one flavoured particle inside a jet to declare it as flavoured (as done in “cone labelling”) and the IRC-safe mod2 prescription for the sum of flavours within a jet (as done in all of the other options). Focusing on the NLO+PY8 sample, we see that SDF and GHS are similar to “anti- $k_t$  naive” prescription, whereas IFN is systematically lower. This is line with the other studies of this paper: IFN seems to be more aggressive in removing  $c\bar{c}$  pairs from the event. The behaviour in the NLO+HW7 sample is similar, though with reduced differences between flavoured algorithms (at most 3%). If we compare the predictions for NLO+PY8 and NLO+HW7 for a fixed algorithm, we observe a ratio close to 0.8, in agreement with the behaviour of the NLO+PS curves in figures 28 and 29. Finally, we study the effect of MPI in the NLO+PY8 sample, which turns out to be sizeable, from 10% to 30% according to the algorithm and the observable considered. If we compare the different

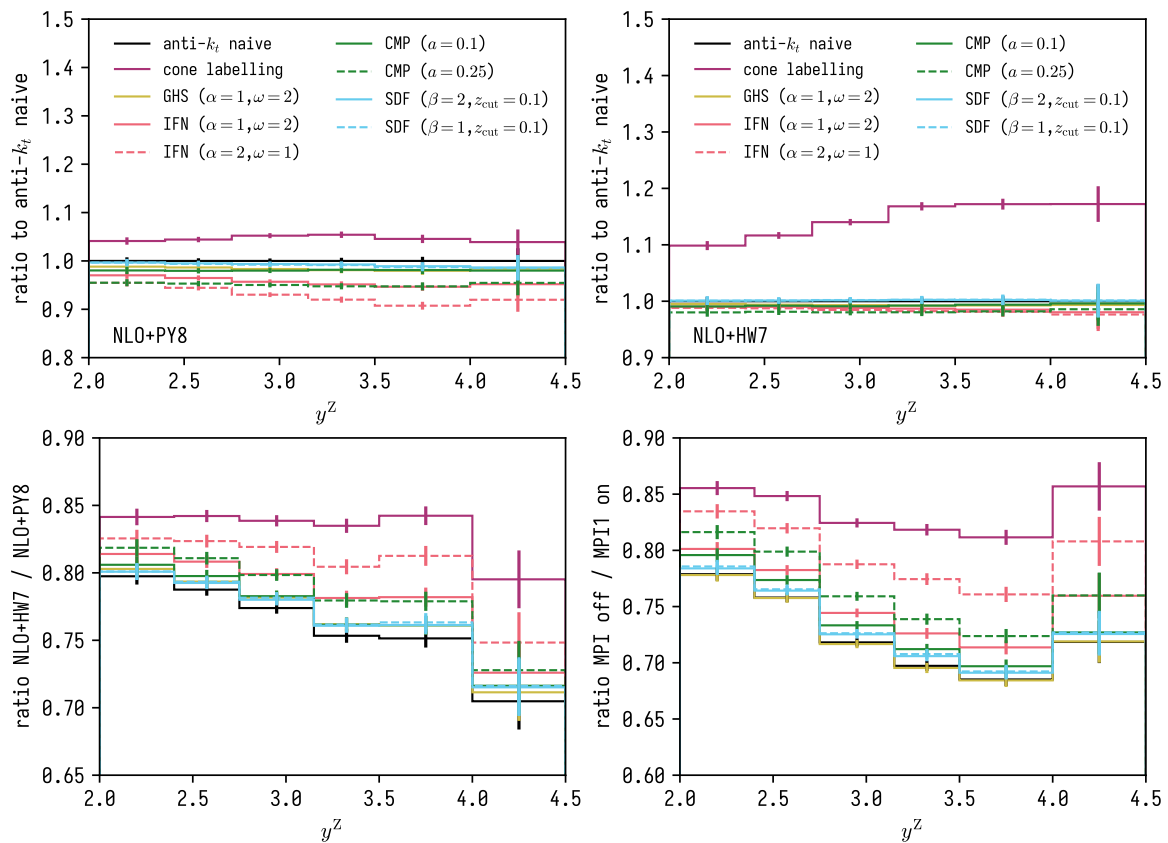


**Figure 28.** NLO+PS predictions with PYTHIA (upper right) and HERWIG (upper left) for the transverse momentum of the leading- $p_t$  charm jet,  $p_{t,c}$ . For each algorithm, we further compare PYTHIA and HERWIG (lower left) and we study the size of MPI effects within PYTHIA (lower right).

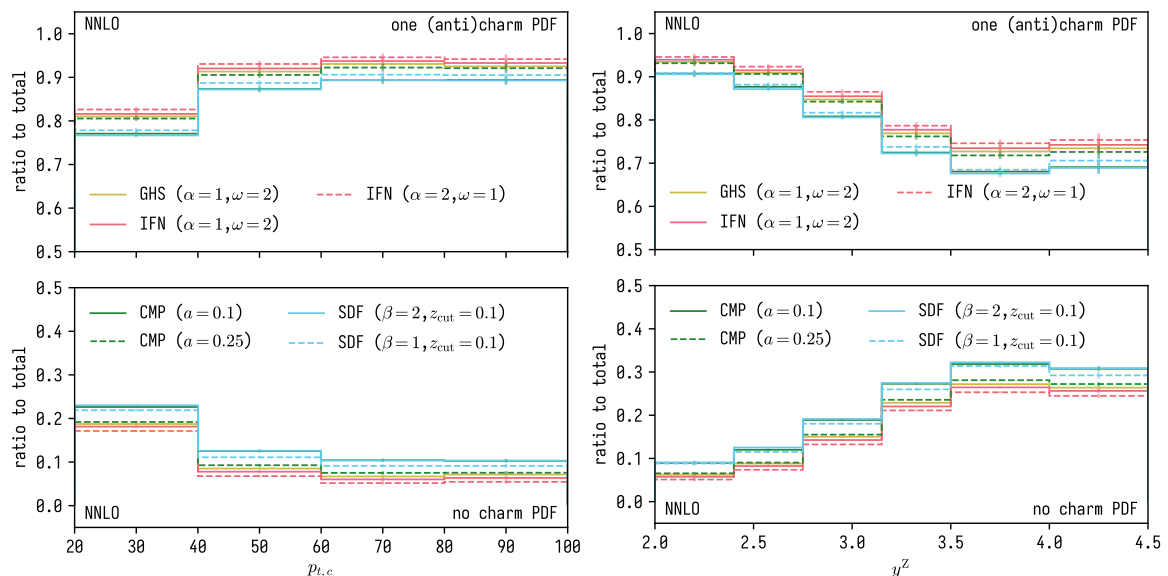
algorithms, we see that the algorithms more aggressive in removing flavour from the event (for which the NLO+PY8 predictions were smaller) are also more effective in reducing the MPI contribution, as already noted in ref. [36]. Again, the “cone labelling” is an outlier: in this case it seems to be the algorithm most resilient against the effects of MPI. In this case, the minimum  $p_{t,c}^{\text{hadron}}$  requirement, imposed only in “cone labelling” and absent in the other algorithms because of IRC safety, is effective in suppressing the MPI contribution.

In figure 30 we study the size of individual contributions associated to different partonic luminosities for each of the algorithms. Our aim is to determine which algorithm is the best in preserving the dominance of Born-like events (with a single charm PDF) at higher orders. We then consider contributions with one  $c$  or  $\bar{c}$  PDF, or contributions with no  $c$  or  $\bar{c}$  PDF (contributions with two  $c$  or  $\bar{c}$  PDFs are negligible). We see that the differences between algorithms at NNLO are at the order of a few percent, with IFN and GHS providing similar results, followed by CMP with  $a = 0.25$ . The set of SDF variations (and CMP with  $a = 0.1$ ) seems to increase the contributions coming from channels with no charm PDF.

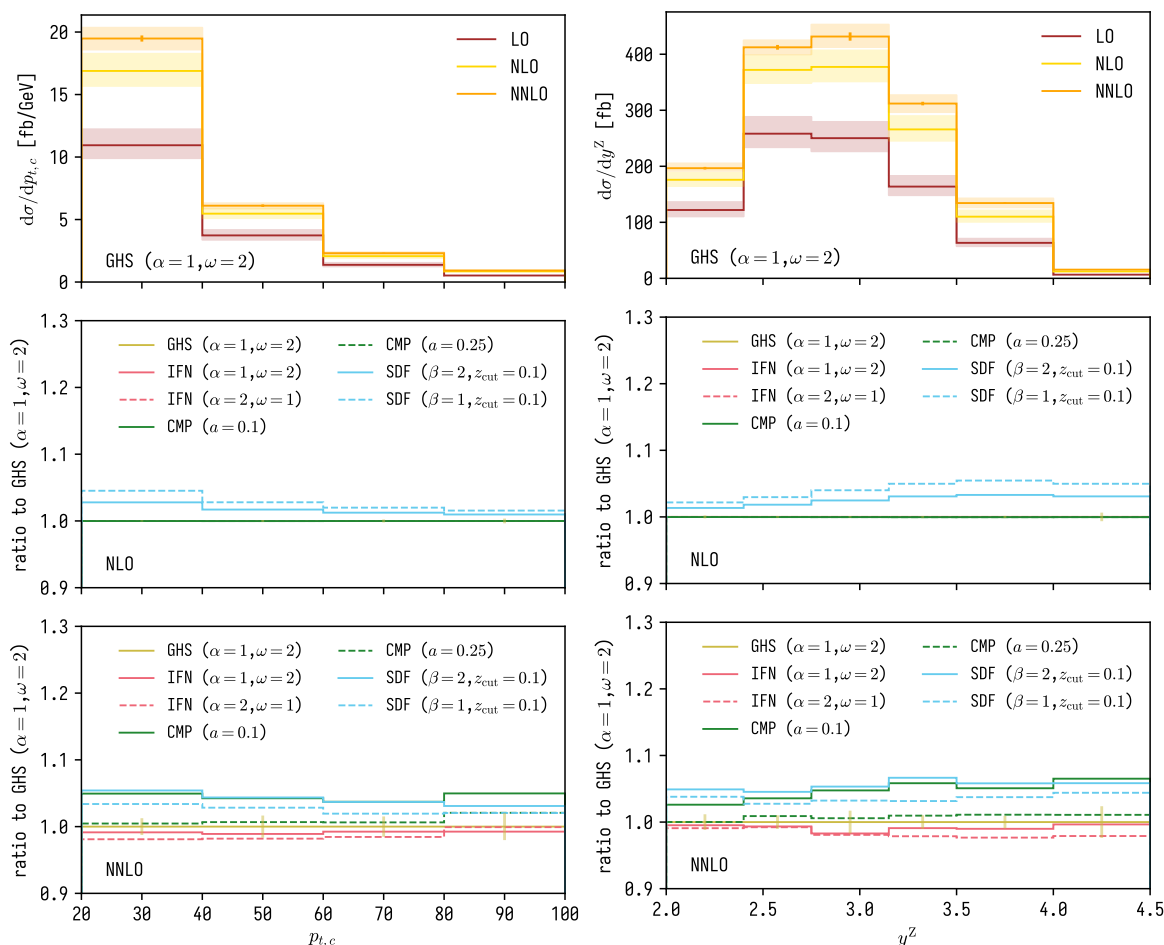
Finally, in figure 31, we compare the flavoured jet algorithms at the fixed-order level. For the two observables and for all of the algorithms, the NNLO corrections are sizeable, outside of the NLO uncertainty band. A similar behaviour was observed in ref. [127]. At NLO, we note that SDF differs from the other algorithms: while all the IRC-safe algorithms



**Figure 29.** NLO+PS predictions with PYTHIA (upper right) and HERWIG (upper left) for the rapidity of the Z-boson,  $y^Z$ . For each algorithm, we further compare PYTHIA and HERWIG (lower left) and we study the size of MPI effects within PYTHIA (lower right).



**Figure 30.** Channel breakdown at NNLO for the transverse momentum of the leading- $p_t$  charm jet,  $p_{t,c}$  (left) and for the rapidity of the Z-boson,  $y^Z$  (right).



**Figure 31.** Comparison of fixed-order predictions for the transverse momentum of the leading- $p_t$  charm jet,  $p_{t,c}$ , and for the rapidity of the Z-boson,  $y^Z$ .

provide identical results at this order (by construction), there are events that are labelled differently with SDF. These are configurations with a  $c\bar{c}$  pair from a gluon splitting, with one of the two quarks removed by the Soft Drop criterion, resulting in a flavoured jet. The impact of this effect depends on the specific value for the SDF parameters, that in turn enters the Soft Drop criterion and drives the percentage of discarded events. At NNLO, we note at most differences of the order of 5% between algorithms: GHS provides results similar to IFN for both choices of parameters, whereas the value of the cross section with SDF and CMP with  $a = 0.1$  is a bit larger.

## F LO+PS $pp \rightarrow 2 \rightarrow 2$ QCD with $c/b$ in LHCb kinematics

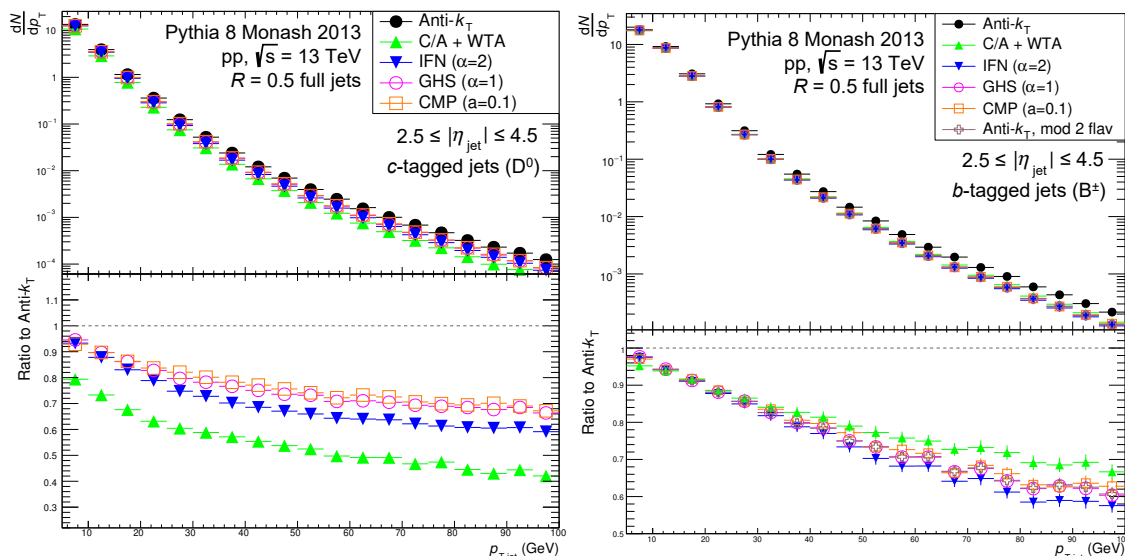
**Authors: Ezra D. Lesser and Ruide Xu.**

Many recent experimental studies of heavy-flavour-tagged (HF) jet substructure utilise jets which contain a particular heavy-flavour hadron explicitly reconstructed from its individual decay daughters [136–138]. These ‘exclusive’ HF jet measurements cannot make any discrimination on the perturbative process which initiated the jet formation, and in principle can come from any leading-order process, including from a gluon or light-quark-initiated jet, where the HF is created from a gluon splitting ( $g \rightarrow q\bar{q}$ ). Performing a study of HF jet tagging in a simulated MC sample of events in which all leading-order processes are allowed to contribute therefore offers valuable insight into how these recent HF jet flavour tagging algorithms might perform when used in similar experimental conditions.

In order to study this effect, a MC simulation was generated using PYTHIA 8.244 [133] configured with the default Monash 2013 tune and `HardQCD:all = on`, which allows all hard QCD  $2 \rightarrow 2$  processes to contribute. Approximate LHCb kinematics are chosen for this study. Since the production of HF particles is rare relative to gluons and light flavour in this configuration, optimisations were implemented to reduce the running time. First, the simulation was stopped after the parton shower, and the quark of interest (either  $c$  or  $b$ , depending on the flavour of the final-state hadron of interest) was required to be among the list of particles. The quark is also required to be within one unit of pseudorapidity from the final hadron-level LHCb kinematic requirements, allowing for some kinematic smearing between parton and hadron levels. If a suitable HF quark was not found in the event after the parton shower, the next event is generated. If a suitable HF quark is found, the event is allowed to hadronise, and the event is used for jet reconstruction.

Simulations were performed using  $c$ -tagged jets via the  $D^0$  hadronic channel and  $b$ -tagged jets via the  $B^\pm$  hadronic channel. All  $c$  and  $b$  hadrons are prevented from decaying, except for the tag hadrons, which are decayed through the  $K^-\pi^+$  (+ c.c.) and  $J/\psi (\rightarrow \mu^+\mu^-) K^\pm$  channels, respectively, and then promptly reconstructed before jet reconstruction. Following previous LHCb jet analyses [27, 139–142], anti- $k_t$  jets are reconstructed using a moderate resolution parameter ( $R = 0.5$ ) with low transverse momenta (between  $5 < p_{t,\text{jet}} < 100$  GeV). Results are compared to the WTA, IFN, CMP, and GHS flavour tagging algorithms; SDF was not used for these studies, due to difficulties building the plug-in libraries for use in the Python 3 + cppy simulation framework [143].

The differential  $p_{t,\text{jet}}$  cross section for  $c$ - and  $b$ -tagged jets produced using this framework is shown in figure 32. The tagging fraction of the new flavour algorithms relative to traditional anti- $k_t$  tagging is observed to drop rapidly as a function of  $p_{t,\text{jet}}$ , especially for  $c$ -tagged jets, where gluonic production is larger (since  $m_c < m_b$ ). Jets reclustered with the Cambridge/Aachen algorithm [144, 145] and tagged using the Winner-Take-All axis (C/A + WTA) [37] are observed to drop most rapidly as a function of  $p_{t,\text{jet}}$ , followed by IFN, GHS, and CMP. At these forward rapidities the tagging fraction is reduced from nearly 100% at  $p_{t,\text{jet}} = 5$  GeV to about 40-70% ( $c$  tagged) and 60-70% ( $b$  tagged) at  $p_{t,\text{jet}} = 100$  GeV, depending on the choice of algorithm.

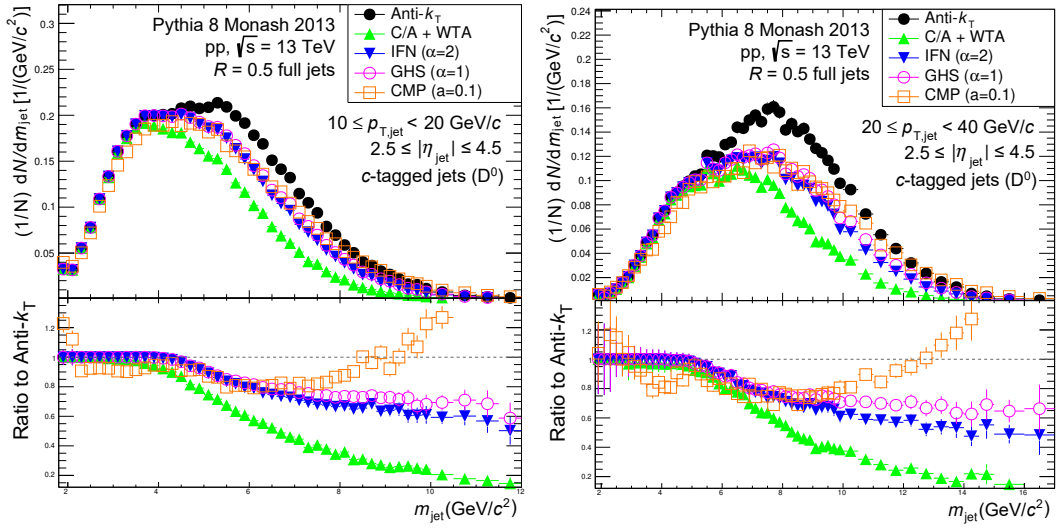


**Figure 32.** Differential cross section for  $c$ -tagged (left) and  $b$ -tagged (right) jets from  $5 < p_{t,\text{jet}} < 100$  GeV using several flavour tagging algorithms. Simulations are produced in PYTHIA 8 using all possible LO QCD  $2 \rightarrow 2$  processes. The tagging rates are found to be similar for  $b$  and  $c$  jets, except for C/A + WTA which exhibits a higher tagging rate for  $b$  jets.

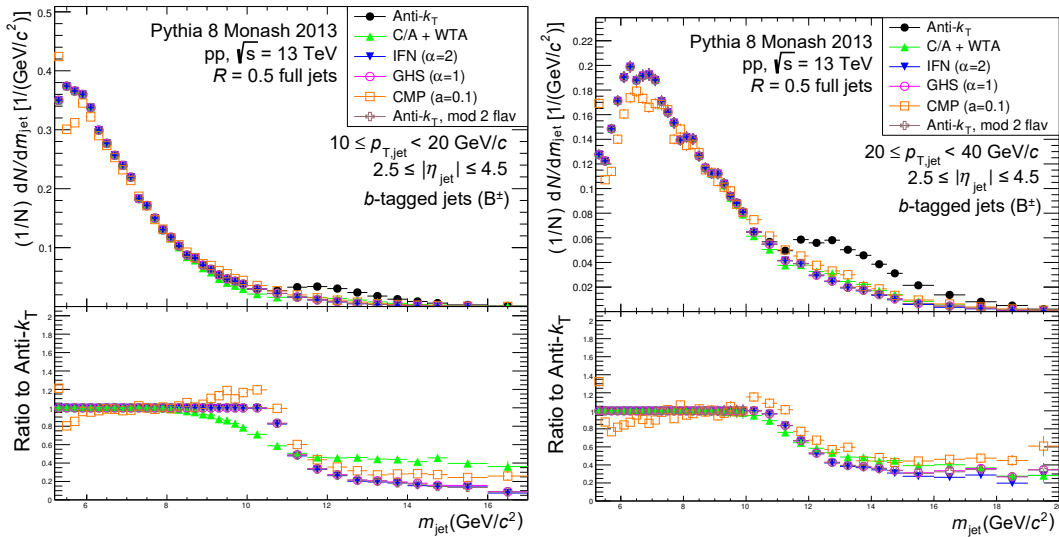
In order to study the effect of this tagging fraction on the jet substructure, this sample of HF-tagged jets was used to calculate the jet invariant mass,  $m_{\text{jet}} = \sqrt{E_{\text{jet}}^2 + p_{\text{jet}}^2}$ , where  $E_{\text{jet}}$  is the jet energy and  $p_{\text{jet}}$  is its total momentum. The  $m_{\text{jet}}$  distributions at low  $p_{t,\text{jet}}$ , where the tagging fractions are largest, are shown in figure 33 for  $c$ -tagged jets and figure 34 for  $b$ -tagged jets. In both cases, a two-peaked structure is observed, with the second  $m_{\text{jet}}$  peak observed around  $m_{\text{jet}} \sim 2 \cdot m_{c/b}$ , corresponding to jets containing two HF hadrons; these jets can be thought of as originating from a gluon splitting process. This second peak of HF jets associated with gluon splitting are suppressed by the new flavour tagging algorithms in comparison to anti- $k_t$  tagging, resulting in a substantial dip in the tagging fraction in the tail of the  $m_{\text{jet}}$  spectrum. The primary mass peak, where  $m_{c/b} \lesssim m_{\text{jet}} \lesssim 2 \cdot m_{c/b}$  and the presence of two HF hadrons is kinematically forbidden, is relatively unaffected by the choice of flavour tagging algorithm.

Although the overall tagging fractions depicted in figure 32 are relatively large at low  $p_{t,\text{jet}}$ , the tagging fractions for the new algorithms in the  $m_{\text{jet}}$  tail are much lower. This is due to the fact that  $m_{\text{jet}}$  is more strongly biased by the jet production process than the  $p_{t,\text{jet}}$  cross section. Since the relative number of counts in the tails of the  $m_{\text{jet}}$  distributions is small relative to the primary mass peak, a significant loss of counts in that region only accounts for an overall small change in the total tagging fraction. This effect is aggravated by going to higher  $p_{t,\text{jet}}$ : the second mass peak grows, as the phase space for gluon splitting increases, which corresponds to the overall greater reduction in the tagging fraction seen in figure 32.

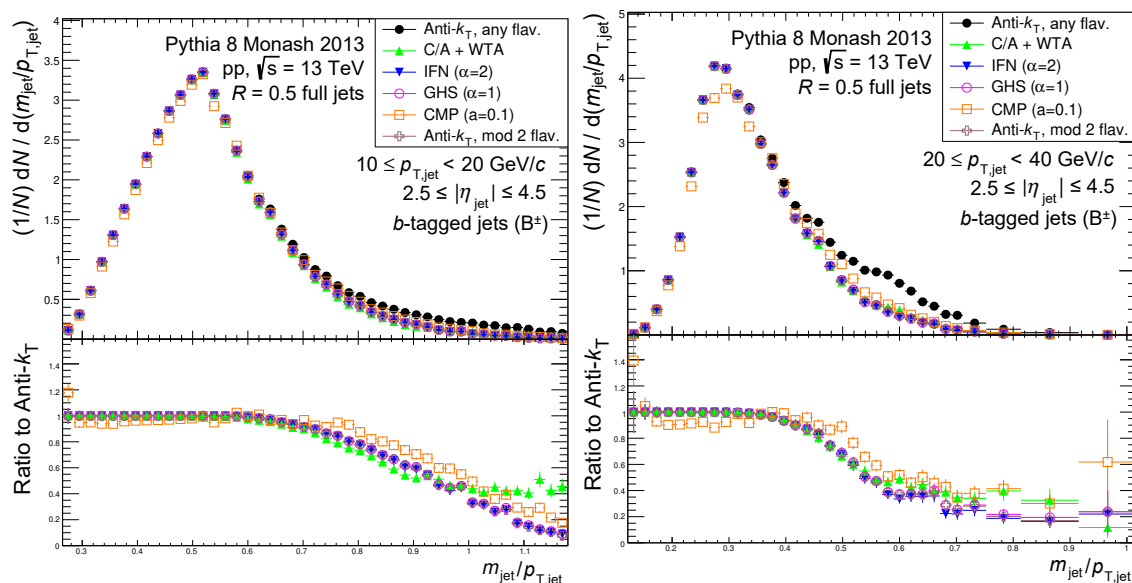
It is worthwhile to study whether these differences are particular to a bias in the jet invariant mass or whether the shape differences between algorithms can be reduced by a careful redefinition of the observable. For example, an imaginary substructure observable



**Figure 33.** Distributions of the  $c$ -tagged jet invariant mass for jets containing a  $D^0$  hadron at low  $p_{t,jet}$ . Simulations are produced in PYTHIA 8 using all possible LO QCD  $2 \rightarrow 2$  processes. A two-peaked structure is observed, with the second  $m_{jet}$  peak observed around  $m_{jet} \sim 2 \cdot m_c \sim 5$  GeV, corresponding to jets containing two  $c$  hadrons. These jets are suppressed by the new flavour tagging algorithms (with the exception of CMP).



**Figure 34.** Distributions of the  $b$ -tagged jet invariant mass for jets containing a  $B^\pm$  hadron at low  $p_{t,jet}$ . Simulations are produced in PYTHIA 8 using all possible LO QCD  $2 \rightarrow 2$  processes. A two-peaked structure is observed, with the second  $m_{jet}$  peak observed around  $m_{jet} \sim 2 \cdot m_b \sim 10$  GeV, corresponding to jets containing two  $b$  hadrons. These jets are suppressed by the new flavour tagging algorithms.



**Figure 35.** Distributions of the  $b$ -tagged  $m_{\text{jet}}/p_{t,\text{jet}}$  for jets containing a  $B^\pm$  hadron at low  $p_{t,\text{jet}}$ . Simulations are produced in PYTHIA 8 using all possible LO QCD  $2 \rightarrow 2$  processes.

which is completely insensitive to whether or not an HF-tagged jet contains 1 or 2 HF hadrons will have an equal fraction of entries discarded across the entire spectrum by the elimination of gluon splitting HF jets. This would mean that, after renormalising the distributions to the number of flavour-tagged jets, there would be no difference in the shapes of the spectrum. Observables which are insensitive to these tagging effects could be useful for future experimental measurements, where reconstruction of 2 or more HF hadrons suffers from low combinatorial efficiencies, which can only be corrected by MC simulations.

One such observable of possible interest is the jet invariant mass relative to the jet transverse momentum,  $m_{\text{jet}}/p_{t,\text{jet}}$ . While there is some correlation between  $m_{\text{jet}}$  and  $p_{t,\text{jet}}$ , inducing a degree of smearing in the higher, di-HF  $m_{\text{jet}}$  peak such that entries appear across the spectrum could improve the overall shape resiliency. Example distributions of the  $b$ -tagged  $m_{\text{jet}}/p_{t,\text{jet}}$  spectrum are shown in figure 35, generated using the same sample of jets. The  $p_{t,\text{jet}}$  scaling marginally affects the shape of the distributions, but the tagging fractions are largely unaffected; however, the shape of the CMP tagging algorithm is observed to most closely approximate anti- $k_t$ , with excesses and dips in the unscaled  $m_{\text{jet}}$  distribution being smeared out by the  $p_{t,\text{jet}}$  scaling.

As observed in central rapidity, the predominant difference between standard “any flavour” anti- $k_t$  and the new algorithms is the modified flavour recombination scheme, followed by the choice of algorithm. This is explicitly confirmed in the  $b$ -jet case, where comparisons to anti- $k_t$  with “mod2 flavour” are included. The main text of this article establishes this difference as originating primarily from the exclusion of gluon splitting. However, this has not been explicitly verified for WTA flavour, which exhibits a similar tagging fraction for  $b$  jets but appears more selective for  $c$  jets.

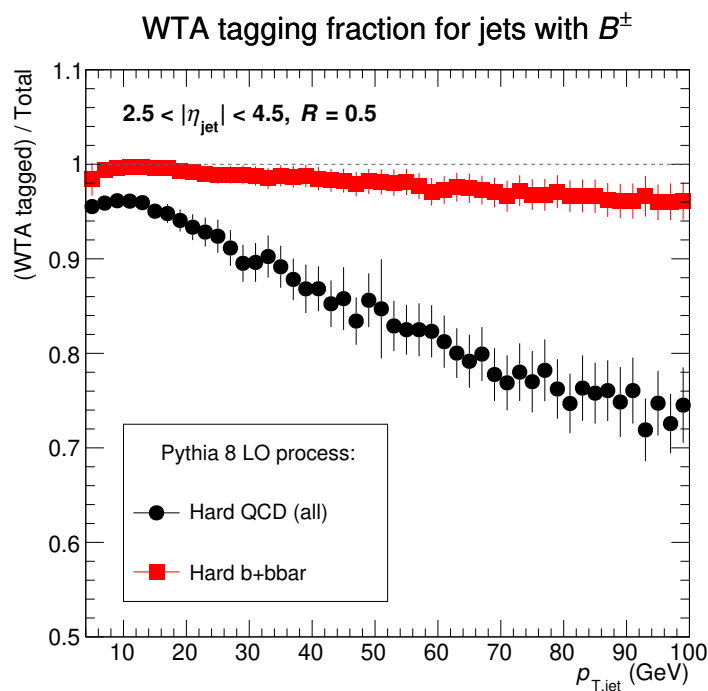
The origin of jet suppression for WTA flavour is studied by producing another simulation using the same parameters as above, but allowing all heavy-flavour hadrons to decay except

the one reconstructed experimentally (i.e., the  $B^\pm$  hadron). This more closely matches the ALICE and LHCb experimental methods, where heavy-flavour candidates are typically only reconstructed in a single decay channel. Any events containing more than one of the same heavy-flavour hadron are rejected due to limited experimental efficiencies. The identities of the LO “parent” partons, labelled 5 and 6 in the PYTHIA scheme, are saved. Jets are then reconstructed at hadron level, and the jet containing the  $B^\pm$  hadron is evaluated using the WTA flavour tagging algorithm.

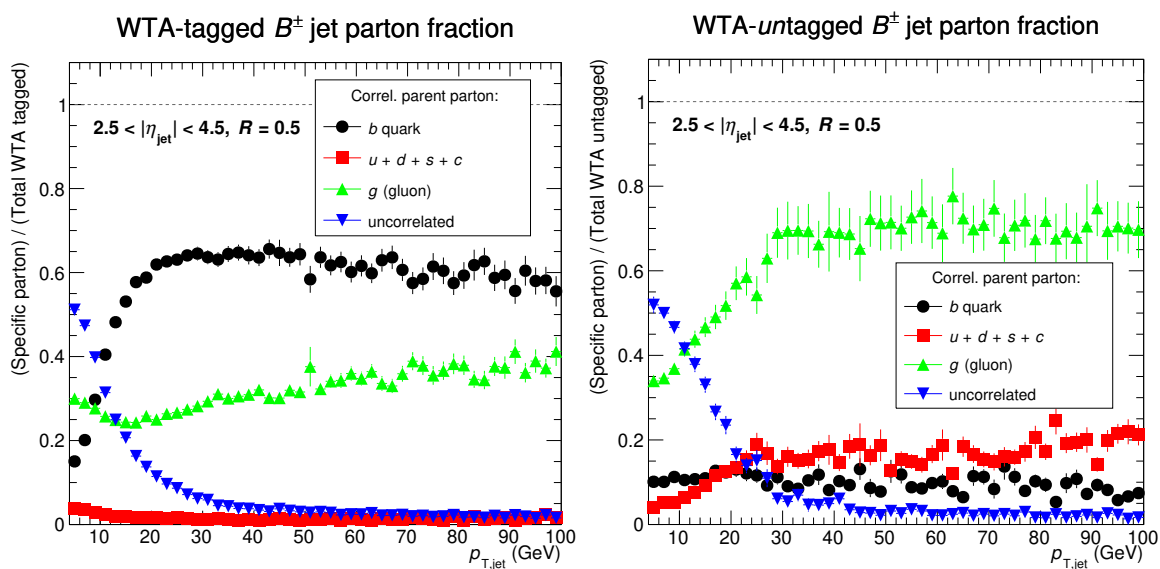
The fraction of jets tagged with WTA flavour as a function of  $p_{t,\text{jet}}$  is shown in figure 36 for events generated with all hard  $2 \rightarrow 2$  QCD processes as well as events generated purely with a hard  $b\bar{b}$  pair generated at LO. At  $p_{t,\text{jet}} = 100$  GeV, the tagging fraction is observed to be roughly 95% for the latter  $b\bar{b}$  sample while only 70-75% for the former inclusive sample, showing that the inclusion of  $b$  jets generated via gluon splitting naturally decreases the fraction of jets positively tagged as  $b$  flavoured using WTA.

To understand this further, jets are matched to either parent parton by checking whether the standard jet axis (using  $E$ -scheme recombination) is within  $\Delta R = 2$  of the momentum vector of either parent parton. If there is a match to one or both partons, the entry is stored as correlated to that parton’s PID in a histogram. Otherwise, if the jet is not correlated with either parent parton within the geometrical limit, then it is saved as “uncorrelated”. The resulting correlation fractions are shown in figure 37. For jets tagged with WTA flavour, the most likely parent parton is observed to be a  $b$  quark ( $\sim 60\%$  of the  $B^\pm$ -tagged jet sample), though a substantial fraction ( $\sim 30\text{-}40\%$ ) of jets also correlate to parent gluons, with this gluon-initiated fraction slowly increasing as a function of  $p_{t,\text{jet}}$ .

For jets that are rejected by WTA flavour tagging, the fraction of jets originating from these partons is substantially different. Most of the jets rejected by WTA flavour tagging ( $\sim 70\%$ ) are correlated to parent gluons, whereas only a small fraction ( $\sim 10\%$ ) are correlated with a hard  $b$  quark. This means that, as with the other flavour tagging algorithms, WTA tagging primarily rejects gluon splitting and therefore increases the purity of HF quark-initiated jets in the LO+PS picture.



**Figure 36.** WTA tagging fraction for  $B^\pm$  jets in events using two different LO production mechanisms: all hard  $2 \rightarrow 2$  QCD channels (black), or hard  $b\bar{b}$  only (red).



**Figure 37.** Correlations between LO parent partons and WTA tagged (left) or rejected (right)  $B^\pm$  jets, as evaluated in PYTHIA 8. A jet’s axis is required to be within  $\Delta R = 2$  of either parent parton’s four-momentum to be correlated; otherwise, it is labelled as ‘uncorrelated’.

## G ATLAS $Z$ + jets studies

**Authors:** Francesco Giuli, Radosław Grabarczyk, Alberto Rescia, and Federico Sforza.

Here we investigate the effect of the new flavoured jet algorithms in an experimental context. We aim to understand how these algorithms can affect current particle-level definitions of jet flavour, if at all, and seek to gain insight on how flavour labelling can be improved in the LHC collaborations.

Our study focuses specifically on comparing the algorithms to ATLAS and CMS labelling. The former uses cone labelling while the latter utilises ghost labelling to define  $b$  jets, i.e., the requirement is that an ultra-soft  $b$  hadron is clustered together with the other jet constituents.

As a caveat, it is important to note that this study is *not an apples-to-apples* comparison. Such a study is impossible, as the input particles for the new algorithms and the experimental particle-level jet definitions differ in a significant way: while the algorithms require undecayed  $b$  hadrons to meaningfully ascribe flavour to a jet, the experimental jets are defined with an eye towards realism. Indeed, in these cases jets are clustered using the visible decay products of the hadron. This difference inevitably leads to *ab initio* kinematic discrepancies.

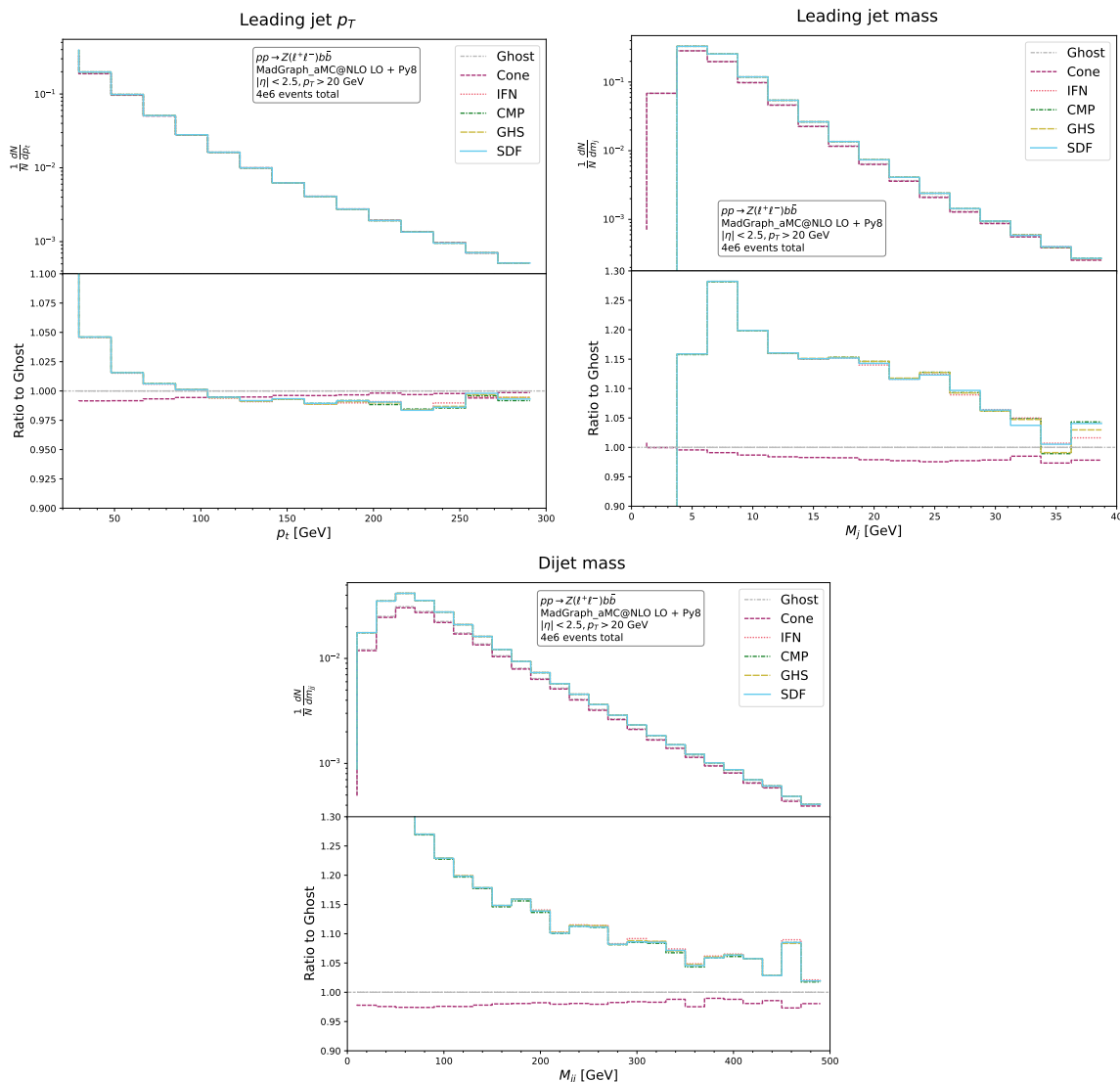
The study focuses on the  $Z + b$ -jet or  $Z + c$ -jet final states, where the  $Z$  boson decays leptonically to either electrons or muons. Leptons are required to have a  $p_t > 27$  GeV due to trigger thresholds and a pseudorapidity  $|\eta| < 2.5$ . The dilepton pair must have a mass compatible with that of the  $Z$  boson, within the range 76 GeV to 106 GeV.

Jets are clustered according to the ATLAS prescription for `AntiKt4TruthJets`, with final state particles within  $|\eta| < 4.5$  and with a radius  $R = 0.4$ . The particles used as input differ as above. Jets found to have a  $p_t > 20$  GeV and  $|\eta| < 2.5$  are labelled by the appropriate means, and final states containing at least one  $b$ -jet are selected. When more  $b$  jets are found, the two  $b$  jets with the highest  $p_t$  are selected.

Samples are simulated with `MADGRAPH5_AMC@NLO` (v3.5.8) [132] interfaced with `PYTHIA` (v8.313) [133] for the parton shower. Matrix elements are generated at LO, requiring final states containing a  $Z$  boson along with a  $b\bar{b}$  pair, or a  $c\bar{c}$  pair.

As can be seen in figure 38, the various algorithms form two distinct populations. The first is composed of cone and ghost labelled jets, which, although they differ slightly in the way they ascribe flavour labels to jets, use jets clustered including the decayed  $b$  hadrons as mentioned previously. Another population includes jets clustered with the flavour algorithms. As can be seen in the ratio plots, the  $p_t$  and mass distributions obtained in this population nearly overlap, although in some regions, such as at high  $p_t$ , the differences between the flavour algorithms become apparent.

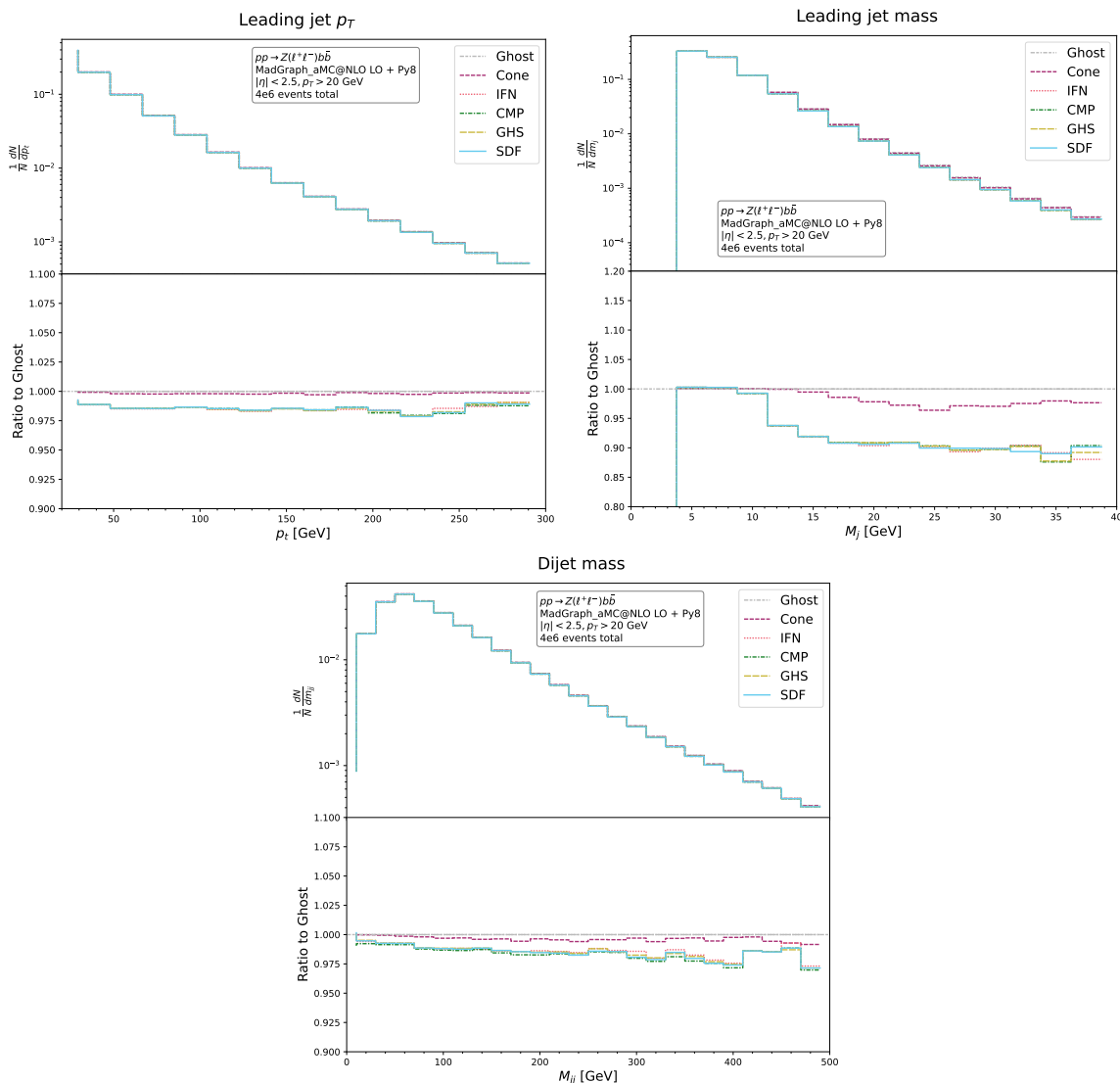
Surprisingly, figure 38 shows that the flavour algorithms ascribe flavour to more jets than the experimental algorithms. Naively, this is unexpected as the flavour algorithms should be *cancelling* flavour pairs originating from gluon splitting, which the experimental algorithms would be unable to identify. This difference can be attributed to the difference in constituents. This can be especially seen in the leading jet mass plot in figure 38, as it shows that the ghost and cone algorithms find a significant proportion of jets with a mass below that of a



**Figure 38.** The  $p_t$  distribution of leading  $b$  jets identified with the cone, ghost, and flavour algorithms (top left), the leading jet mass distribution (top right), and the dijet mass distribution of the two highest- $p_t$   $b$  jets identified in the same manner (bottom). Ratios are compared to the distribution for jets clustered with ghost labelling.

$b$  hadron. When considering undecayed hadrons and ascribing flavour with all algorithms, results match more closely, as can be seen in figure 39.

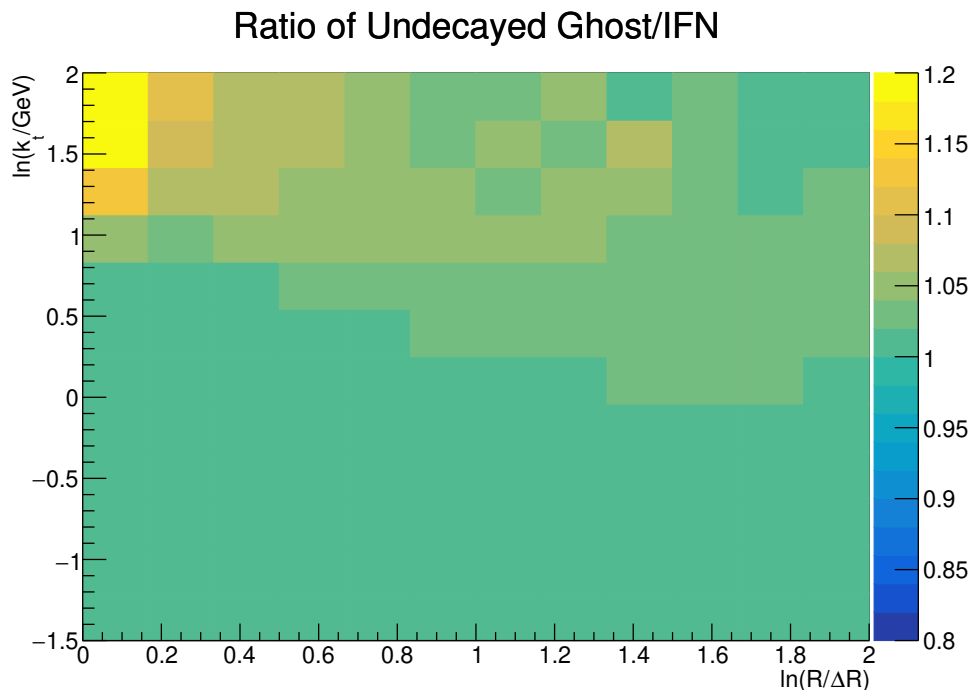
We have also investigated some aspects of jet substructure beyond just the jet mass. In particular, we have looked at any differences in the Lund Jet Plane [146] which may arise due to the flavour labelling strategy. For jet substructure measurements, it is often best to exploit the higher resolution of tracking detectors compared to calorimeters. For this reason, measurements focus on the charged component of the jet to elucidate more information. However, in this study we choose to study the substructure of the jet as a whole, thus imposing no cut on either the  $p_t$  of the constituents or on their charge.



**Figure 39.** The  $p_t$  distribution of leading  $b$  jets identified with the cone, ghost, and flavour algorithms with undecayed hadrons (top left), the leading  $b$ -jet mass (top right), and the dijet mass distribution of the two highest- $p_t$   $b$  jets identified in the same manner (bottom). Ratios are compared to the distribution for jets clustered with ghost labelling.

Another important point to mention is the fact that some substructure effects, such as the dead cone effect [61] (see e.g., refs. [147–149] for recent developments on heavy-flavour jet substructure), rely on the identification of the heavy-flavour hadron and accounting for its four-momentum. For jets obtained with the algorithms, this is straightforward as the hadron is already included in the jets’ constituents. On the other hand, experimentally it is necessary to either identify the heavy hadron, as the ALICE collaboration has done [150], or reconstruct it as the CMS collaboration has done [81]. In this study, we have opted to only consider undecayed heavy hadrons.

To compare the Lund Plane reconstructed using one of the flavour algorithms and one of the standard experimental labelling algorithms, in figure 40 we show the ratio of the



**Figure 40.** The ratio of the Lund Jet Plane for leading  $b$  jets identified with ghost labelling containing undecayed  $b$  hadrons and those identified with IFN.

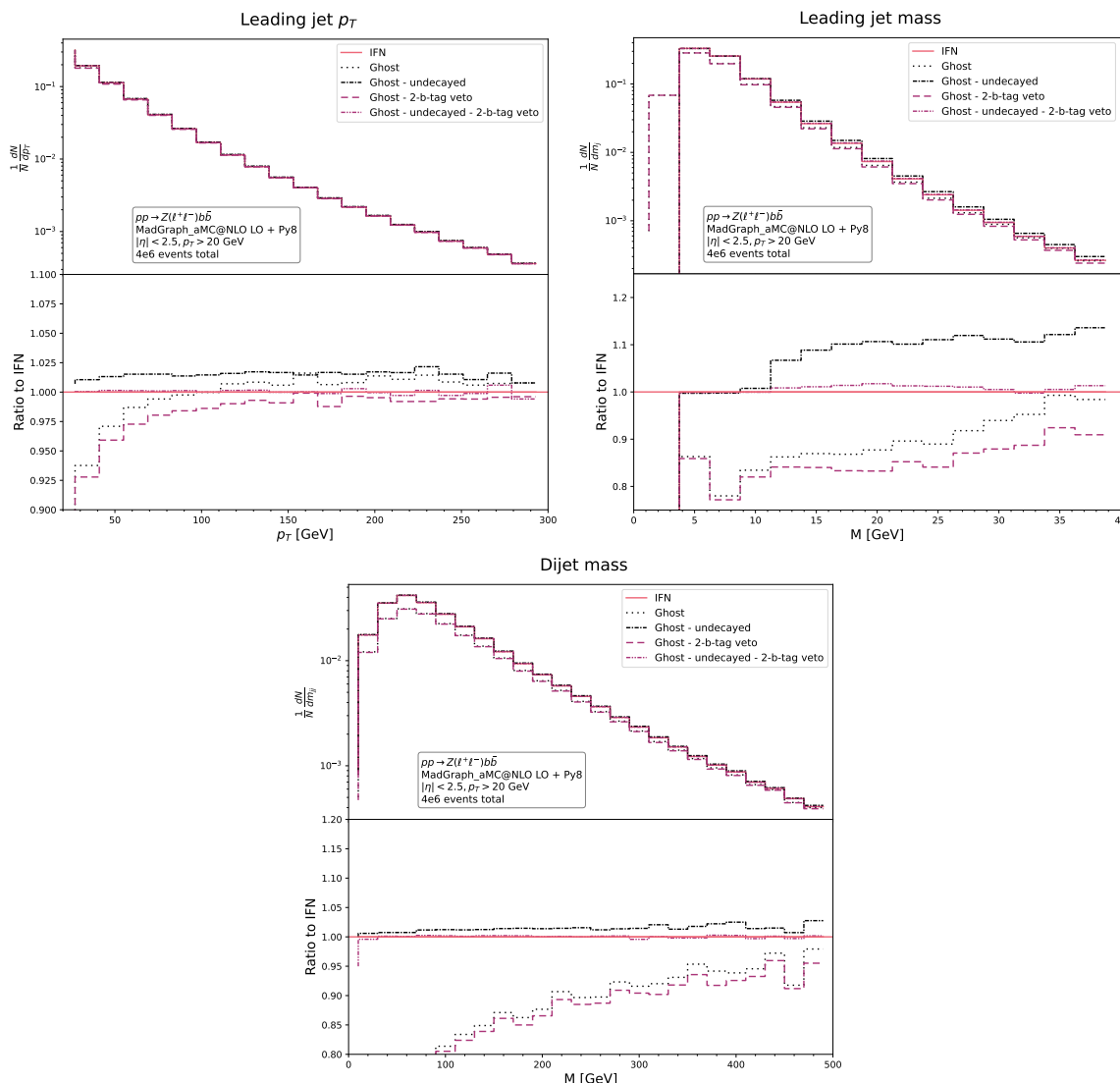
two-dimensional Lund Plane for the leading  $b$  jets found with ghost and IFN labelling. It is clear that in the central part of the Lund Plane the results are very similar, while a mild excess is visible in the high- $k_t$  and small angle regions of the Lund Plane. We understand the excess at high- $k_t$  and wide angle to be due to contributions from initial state radiation.

The motivation for this work is to ensure that jets labelled as heavy-flavour jets are indeed such, and not misidentified due to higher-order effects. Experimentally, it is of course extremely difficult, if not infeasible to disentangle the true origin of a  $b$  hadron, though for labelling purposes, a simple way to correct the experimental labelling procedure currently implemented is by imposing a 2- $b$ -tag veto. By doing so, we require jets to contain (in the ghost case) or be matched to (in the cone case) *only one*  $b$  hadron. This explicitly avoids some problematic configurations, such as soft gluons splitting within jets.

When accounting for this, we can try to see if the distributions obtained with experimental flavour labelling approach those for the flavour algorithms.

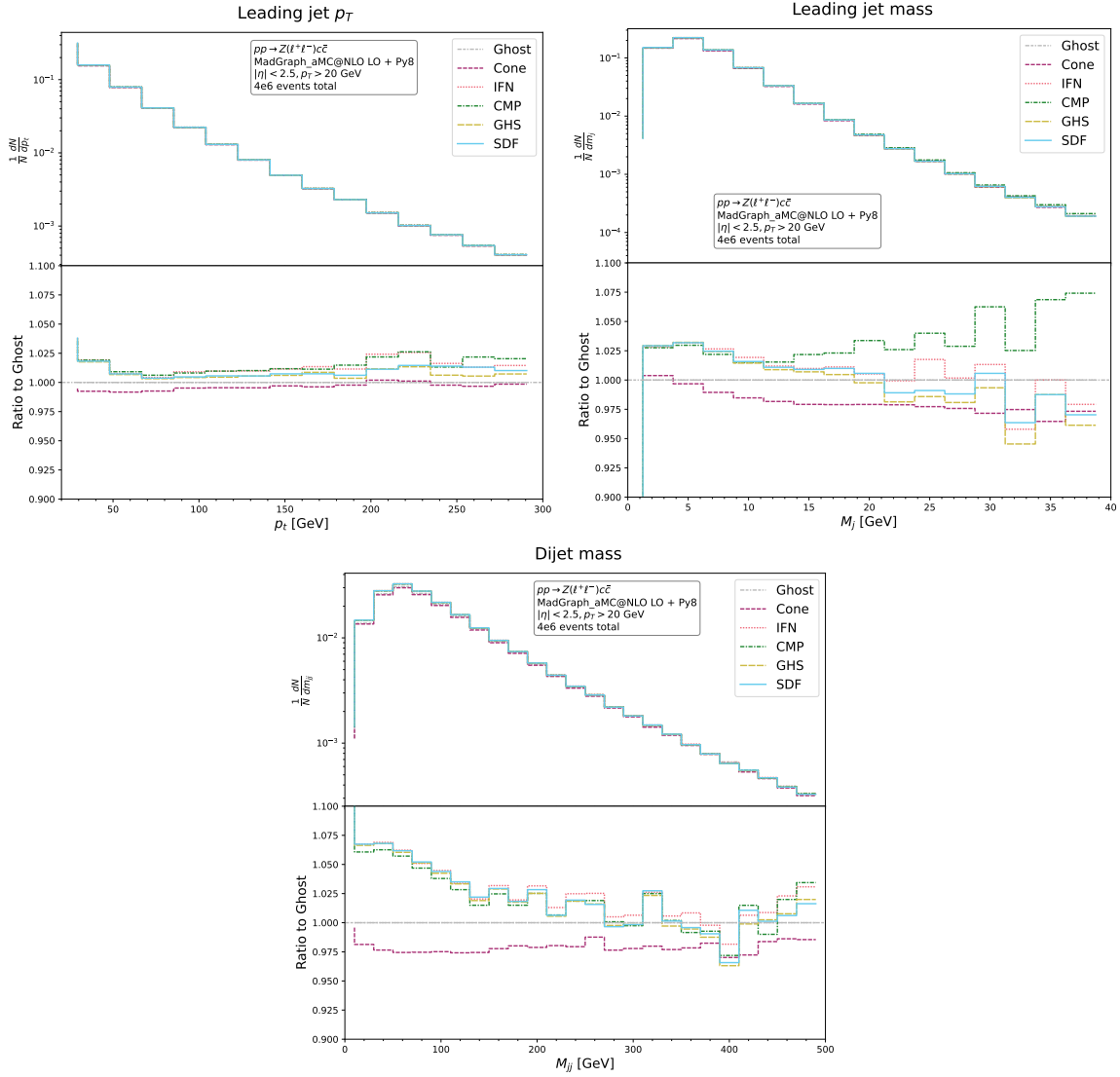
As can be seen in figure 41, applying a 2- $b$ -tag veto has a small effect on the result of the corresponding distribution for all observables considered for the LO samples considered, of the order of a few percent, both in the case of decayed and undecayed heavy hadrons. This effect, however, is negligible with respect to the kinematic differences between jets with different constituent definitions, as previously stated.

The study was repeated for  $Z + c\bar{c}$  events, as shown in figures 42–45. Some effects, such as the kinematic differences in the jets due to the presence of the heavy-hadron decay products, are not as strong as in the  $b$ -jet case due to the lower particle multiplicity of the decays compared to those for  $b$  hadrons, as can be seen in figure 42. On the other hand,

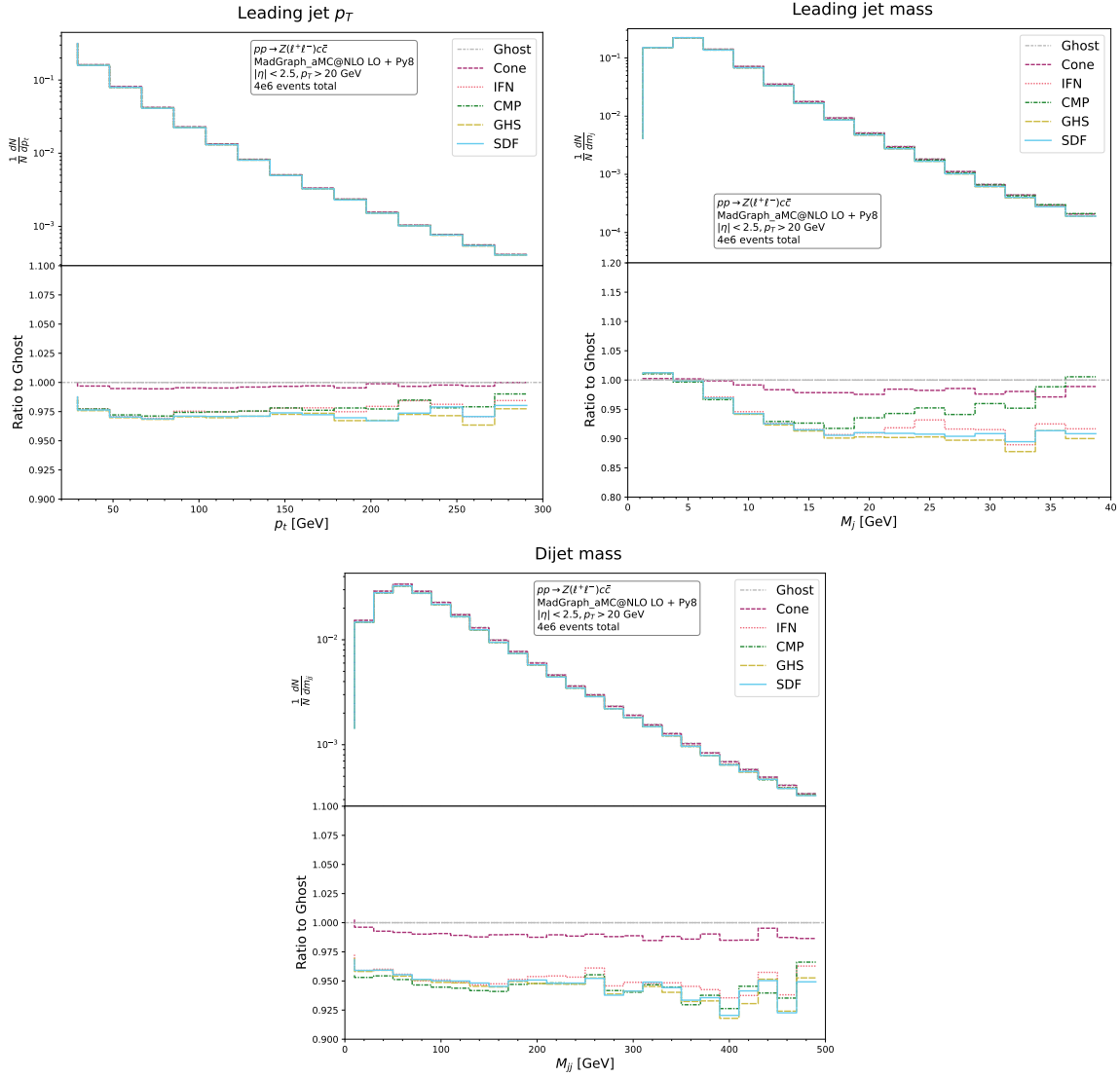


**Figure 41.** The  $p_t$  distribution of the leading  $b$  jet (top left), the leading jet mass (top right), and the dijet mass (bottom), comparing  $b$  jets clustered with IFN (continuous line) to those whose flavour was identified with the ghost prescriptions. The dotted and the dashed lines correspond to the standard ghost definitions (decayed hadrons) without or with 2- $b$ -tag veto, respectively, while the dotted-dashed lines refer to the distributions with undecayed hadrons.

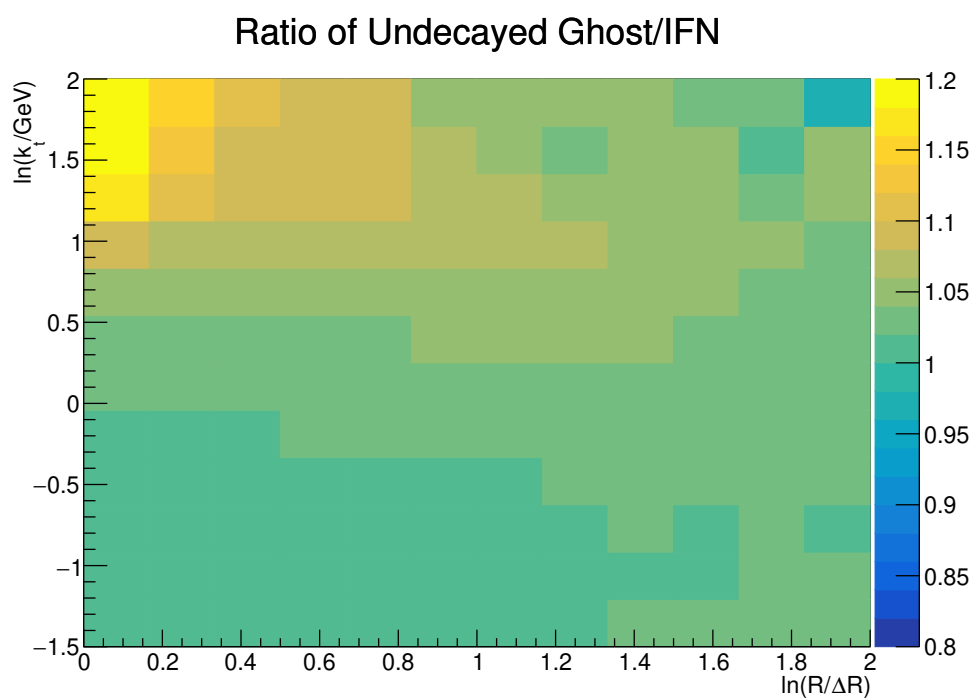
differences between the algorithms are more prominent due to the higher likelihood of  $g \rightarrow c\bar{c}$  splittings. Overall, however, the same trends and conclusions hold.



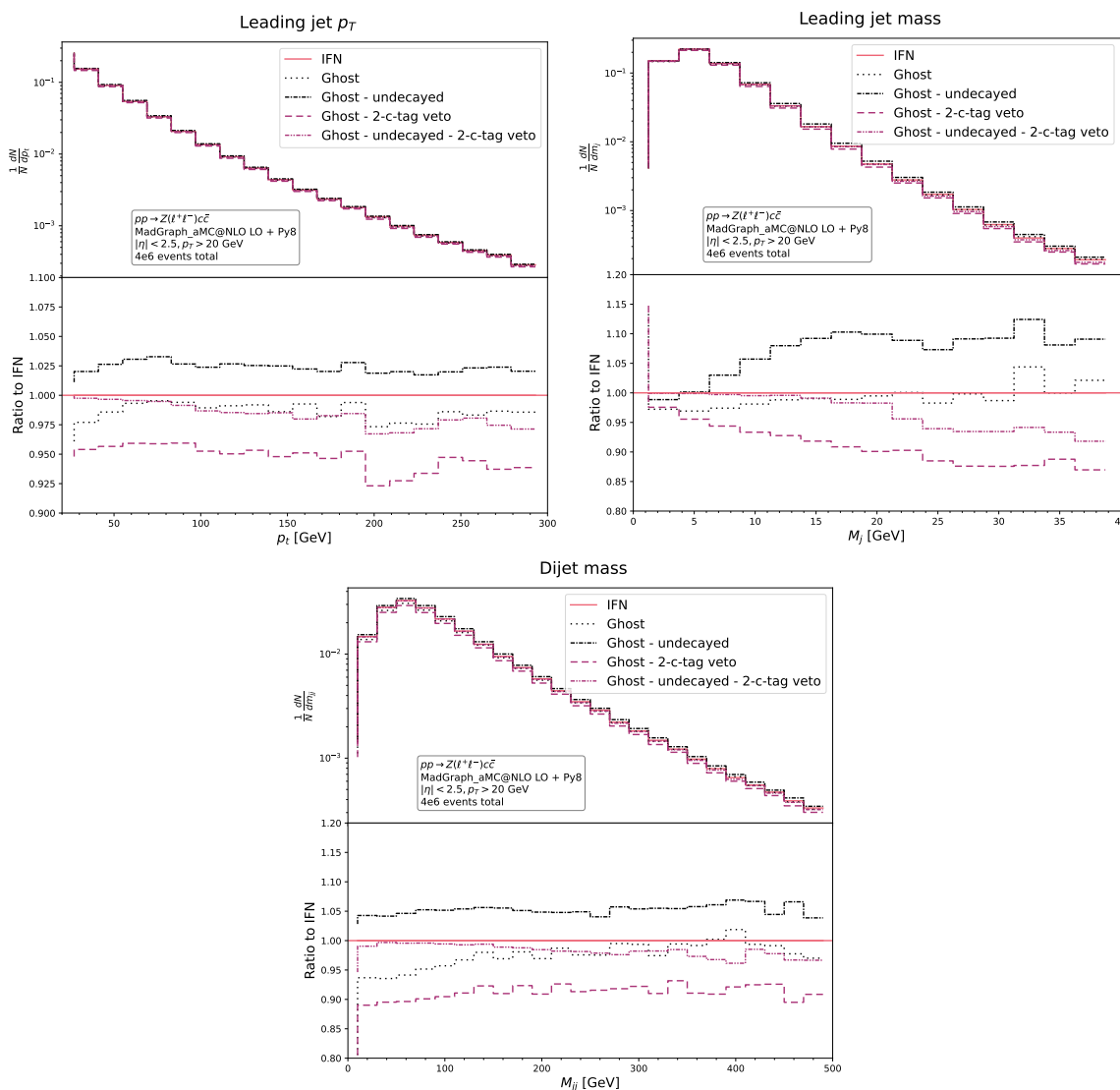
**Figure 42.** The  $p_T$  distribution of leading  $c$  jets identified with the cone, ghost, and flavour algorithms (top left), the leading jet mass (top right), and the dijet mass distribution of the two highest- $p_T$   $c$  jets identified in the same manner (bottom). Ratios are compared to the distribution for jets clustered with ghost labelling.



**Figure 43.** The  $p_t$  distribution of leading  $c$  jets identified with the cone, ghost, and flavour algorithms with undecayed hadrons (top left), the leading jet mass distribution (top right), and the dijet mass distribution (bottom) of the two highest- $p_t$   $c$  jets identified in the same manner. Ratios are compared to the distribution for jets clustered with ghost labelling.



**Figure 44.** The ratio of the Lund Jet Plane for leading  $c$  jets identified with ghost labelling containing undecayed  $c$  hadrons and those identified with IFN.



**Figure 45.** The  $p_t$  distribution of the leading  $c$  jet (top left), the leading jet mass distribution (top right), and the dijet mass (bottom) comparing  $c$  jets clustered with IFN (continuous line) to those whose flavour was identified with the ghost prescriptions. The dotted and the dashed lines correspond to the standard ghost definitions (decayed hadrons) without or with 2- $c$ -tag veto, respectively, while the dotted-dashed lines refer to the distributions with undecayed hadrons.

## H Impact on jet flavour tagging in ATLAS

**Authors: Francesco Giuli, Radosław Grabarczyk, Alberto Rescia, and Federico Sforza.**

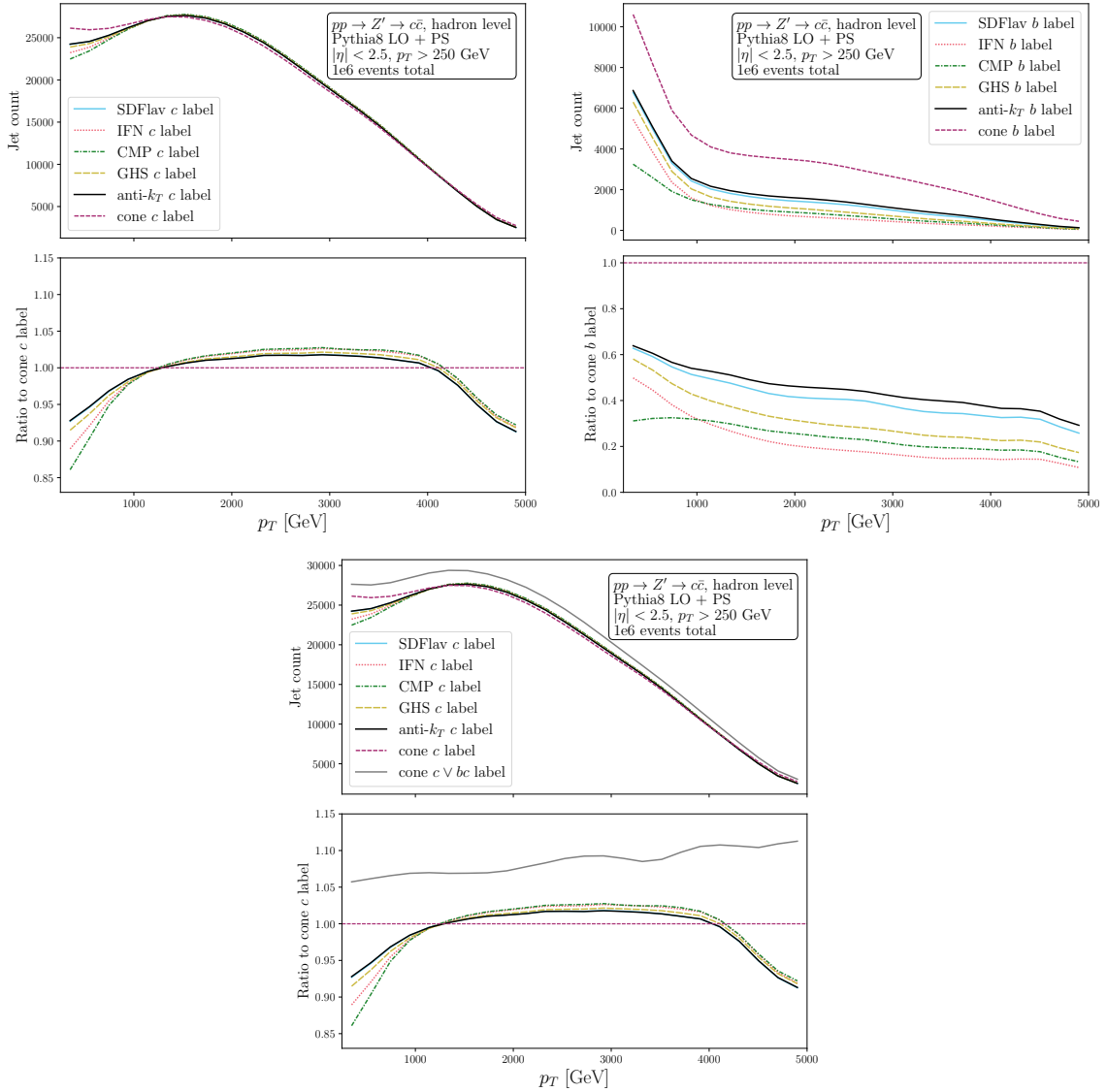
Jet flavour tagging in ATLAS is done by training neural network classifiers on Monte Carlo simulated data. In the case of high- $p_t$  jets, for which the differences between different jet flavour algorithms are enhanced, the training sample is obtained by simulating a heavy resonance  $Z'$ . This resonance is set to decay into  $b\bar{b}$  1/3rd of the time,  $c\bar{c}$  1/3rd of the time, and light flavours with equal branching ratios otherwise. The width of  $Z'$  is tuned to ensure a large spread of different transverse momenta of the jets. The event generation is done entirely with PYTHIA. Following simulation of the ATLAS detector response and reconstruction, all jets that satisfy  $|\eta| < 2.5$  and  $p_t > 250$  GeV enter the training, which means that they can often originate from the parton shower instead of the hard event.

To investigate the effect of using jet flavour algorithms on these types of events, we simulate the same  $Z'$  resonance with PYTHIA 8, with the modification that it decays into  $c\bar{c}$  pairs 100% of the time, which means that any  $b$  jets found in this sample originate from the parton shower. After hadronisation and decay of unstable hadrons, we identify all decay products of  $b$  and  $c$  hadrons, replace them with their parent hadrons and remove them. Next, we cluster the event with different jet flavour algorithms using the mod2 recombination scheme. The results are shown in figure 46. In the case of  $c$  jets, we find that the differences between cone labelling and other algorithms are largest at the lower and higher ends of the  $p_t$  spectrum, where we expect more gluon splitting into heavy flavour generated by the parton shower.

Interestingly, in the middle of the  $p_t$  spectrum, IRC safe jet flavour algorithms yield more  $c$ -labelled jets than anti- $k_t$  or cone labelling. This is due to a disagreement between labelling schemes regarding jets having *both*  $b$  and  $c$  labels, most often arising in this sample from a hard  $c$  quark and a soft, large-angle  $b\bar{b}$  pair. The most “aggressive” IRC-safe labelling schemes will tend to only assign a  $c$  label in this case. Other algorithms may label it as  $bc$ , which, due to  $b$  label always taking priority, results in a final  $b$  label of the jet. To show that this effect is significant, we provide the result of using a “cone  $c \vee bc$ ” labelling scheme, where the  $c$  label takes priority instead.

In the case of  $b$  labelling, due to the nature of our sample, we are guaranteed to see effects of gluon splittings. On the lower end of the  $p_t$  spectrum, similarly to the  $c$ -jet case, CMP is the most aggressive algorithm. On the higher end, IFN becomes the most aggressive one. The effect of collinear gluon splittings into  $b\bar{b}$ , signified by the difference between the cone-labelled  $b$  jets and anti- $k_t$ -labelled ones slowly increases with  $p_t$ , but is likely to be significant for all transverse momenta in the range considered.

We conclude that collinear gluon splittings into heavy flavour can have a sizeable effect on the ATLAS flavour tagging training sample throughout the whole  $p_t$  spectrum, while differences between jet flavour algorithms may become relevant especially at transverse momenta just above 250 GeV, where a significant amount of jets that enter the training do not originate from the hard event.



**Figure 46.** The  $p_t$  spectra of  $c$  jets (left) and  $b$  jets (right) found in  $pp \rightarrow Z' \rightarrow c\bar{c}$  events, simulated at LO and showered with PYTHIA 8. All plots use labelling schemes with “net-flavour” recombination scheme except cone, for which a difference is also shown if  $c$ -jet definition is prioritised over the presence of a  $b$  hadron or not.

**Data Availability Statement.** This article has no associated data or the data will not be deposited.

**Code Availability Statement.** This article has associated code in a code repository.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY4.0](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.

## References

- [1] G.P. Salam, *Towards jetography*, *Eur. Phys. J. C* **67** (2010) 637 [[arXiv:0906.1833](https://arxiv.org/abs/0906.1833)] [[INSPIRE](#)].
- [2] S. Marzani, *Jets at colliders*, in *Instrumentation and techniques in high energy physics*, D. Lincoln ed., World Scientific, Singapore (2024), p. 179–211 [[DOI:10.1142/9789819801107\\_0006](https://doi.org/10.1142/9789819801107_0006)] [[INSPIRE](#)].
- [3] 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](#)].
- [4] ATLAS collaboration, *Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *JHEP* **05** (2018) 195 [[arXiv:1711.02692](https://arxiv.org/abs/1711.02692)] [[INSPIRE](#)].
- [5] ATLAS collaboration, *Cross-section measurements for the production of a Z boson in association with high-transverse-momentum jets in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *JHEP* **06** (2023) 080 [[arXiv:2205.02597](https://arxiv.org/abs/2205.02597)] [[INSPIRE](#)].
- [6] CMS collaboration, *Measurement of the double-differential inclusive jet cross section in proton-proton collisions at  $\sqrt{s} = 5.02$  TeV*, *JHEP* **01** (2025) 011 [[arXiv:2401.11355](https://arxiv.org/abs/2401.11355)] [[INSPIRE](#)].
- [7] CMS collaboration, *Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *JHEP* **02** (2022) 142 [*Addendum ibid.* **12** (2022) 035] [[arXiv:2111.10431](https://arxiv.org/abs/2111.10431)] [[INSPIRE](#)].
- [8] R. Abdul Khalek et al., *Phenomenology of NNLO jet production at the LHC and its impact on parton distributions*, *Eur. Phys. J. C* **80** (2020) 797 [[arXiv:2005.11327](https://arxiv.org/abs/2005.11327)] [[INSPIRE](#)].
- [9] ATLAS collaboration, *Measurements of jet cross-section ratios in 13 TeV proton-proton collisions with ATLAS*, *Phys. Rev. D* **110** (2024) 072019 [[arXiv:2405.20206](https://arxiv.org/abs/2405.20206)] [[INSPIRE](#)].
- [10] CMS collaboration, *Measurement of multijet azimuthal correlations and determination of the strong coupling in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **84** (2024) 842 [[arXiv:2404.16082](https://arxiv.org/abs/2404.16082)] [[INSPIRE](#)].
- [11] ATLAS collaboration, *Determination of the strong coupling constant from transverse energy-energy correlations in multijet events at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *JHEP* **07** (2023) 085 [[arXiv:2301.09351](https://arxiv.org/abs/2301.09351)] [[INSPIRE](#)].
- [12] S. Marzani, G. Soyez and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology*, Springer (2019) [[DOI:10.1007/978-3-030-15709-8](https://doi.org/10.1007/978-3-030-15709-8)] [[INSPIRE](#)].
- [13] J. Gallicchio and M.D. Schwartz, *Quark and gluon tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [[arXiv:1106.3076](https://arxiv.org/abs/1106.3076)] [[INSPIRE](#)].

- [14] 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](#)].
- [15] T. Cheng, *Recursive neural networks in quark/gluon tagging*, *Comput. Softw. Big Sci.* **2** (2018) 3 [[arXiv:1711.02633](#)] [[INSPIRE](#)].
- [16] S. Caletti, O. Fedkevych, S. Marzani and D. Reichelt, *Tagging the initial-state gluon*, *Eur. Phys. J. C* **81** (2021) 844 [[arXiv:2108.10024](#)] [[INSPIRE](#)].
- [17] E.M. Metodiev and J. Thaler, *Jet topics: disentangling quarks and gluons at colliders*, *Phys. Rev. Lett.* **120** (2018) 241602 [[arXiv:1802.00008](#)] [[INSPIRE](#)].
- [18] F.A. Dreyer, G. Soyez and A. Takacs, *Quarks and gluons in the Lund plane*, *JHEP* **08** (2022) 177 [[arXiv:2112.09140](#)] [[INSPIRE](#)].
- [19] A. Butter, B.M. Dillon, T. Plehn and L. Vogel, *Performance versus resilience in modern quark-gluon tagging*, *SciPost Phys. Core* **6** (2023) 085 [[arXiv:2212.10493](#)] [[INSPIRE](#)].
- [20] M.J. Dolan, J. Gargalionis and A. Ore, *Quark-versus-gluon tagging in CMS open data with CWoLa and TopicFlow*, *JHEP* **08** (2025) 024 [[arXiv:2312.03434](#)] [[INSPIRE](#)].
- [21] M. Knobbe, F. Krauss, D. Reichelt and S. Schumann, *Measuring hadronic Higgs boson branching ratios at future lepton colliders*, *Eur. Phys. J. C* **84** (2024) 83 [[arXiv:2306.03682](#)] [[INSPIRE](#)].
- [22] J.R. Andersen et al., *Les Houches 2017: physics at TeV colliders Standard Model working group report*, [arXiv:1803.07977](#) [[INSPIRE](#)].
- [23] P. Gras et al., *Systematics of quark/gluon tagging*, *JHEP* **07** (2017) 091 [[arXiv:1704.03878](#)] [[INSPIRE](#)].
- [24] S. Amoroso et al., *Les Houches 2019. Physics at TeV colliders: Standard Model working group report*, in the proceedings of the *11th Les Houches workshop on physics at TeV colliders: PhysTeV Les Houches*, (2020) [[arXiv:2003.01700](#)] [[INSPIRE](#)].
- [25] CMS collaboration, *Observation of a pseudoscalar excess at the top quark pair production threshold*, *Rept. Prog. Phys.* **88** (2025) 087801 [[arXiv:2503.22382](#)] [[INSPIRE](#)].
- [26] FCC collaboration, *Future Circular Collider feasibility study report: volume 1, physics, experiments, detectors*, [arXiv:2505.00272](#) [[DOI:10.17181/CERN.9DKX.TDH9](#)] [[INSPIRE](#)].
- [27] LHCb collaboration, *Study of Z bosons produced in association with charm in the forward region*, *Phys. Rev. Lett.* **128** (2022) 082001 [[arXiv:2109.08084](#)] [[INSPIRE](#)].
- [28] NNPDF collaboration, *Evidence for intrinsic charm quarks in the proton*, *Nature* **608** (2022) 483 [[arXiv:2208.08372](#)] [[INSPIRE](#)].
- [29] A. Banfi, G.P. Salam and G. Zanderighi, *Infrared safe definition of jet flavor*, *Eur. Phys. J. C* **47** (2006) 113 [[hep-ph/0601139](#)] [[INSPIRE](#)].
- [30] A. Banfi, G.P. Salam and G. Zanderighi, *Accurate QCD predictions for heavy-quark jets at the Tevatron and LHC*, *JHEP* **07** (2007) 026 [[arXiv:0704.2999](#)] [[INSPIRE](#)].
- [31] S. Catani, Y.L. Dokshitzer, M.H. Seymour and B.R. Webber, *Longitudinally invariant  $K_t$  clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187 [[INSPIRE](#)].
- [32] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160 [[hep-ph/9305266](#)] [[INSPIRE](#)].

- [33] S. Caletti, A.J. Larkoski, S. Marzani and D. Reichelt, *Practical jet flavour through NNLO*, *Eur. Phys. J. C* **82** (2022) 632 [[arXiv:2205.01109](#)] [[INSPIRE](#)].
- [34] M. Czakon, A. Mitov and R. Poncelet, *Infrared-safe flavoured anti- $k_T$  jets*, *JHEP* **04** (2023) 138 [[arXiv:2205.11879](#)] [[INSPIRE](#)].
- [35] R. Gauld, A. Huss and G. Stagnitto, *Flavor identification of reconstructed hadronic jets*, *Phys. Rev. Lett.* **130** (2023) 161901 [*Erratum ibid.* **132** (2024) 159901] [[arXiv:2208.11138](#)] [[INSPIRE](#)].
- [36] F. Caola et al., *Flavored jets with exact anti- $k_T$  kinematics and tests of infrared and collinear safety*, *Phys. Rev. D* **108** (2023) 094010 [[arXiv:2306.07314](#)] [[INSPIRE](#)].
- [37] S. Caletti, A.J. Larkoski, S. Marzani and D. Reichelt, *A fragmentation approach to jet flavor*, *JHEP* **10** (2022) 158 [[arXiv:2205.01117](#)] [[INSPIRE](#)].
- [38] A.J. Larkoski and D. Neill, *Flavor fragmentation function factorization*, *JHEP* **01** (2024) 119 [[arXiv:2310.01486](#)] [[INSPIRE](#)].
- [39] J. Andersen et al., *Les Houches 2023. Physics at TeV colliders: Standard Model working group report*, in the proceedings of the *Physics of the TeV scale and beyond the Standard Model: intensifying the quest for new physics*, (2024) [[arXiv:2406.00708](#)] [[INSPIRE](#)].
- [40] A. Andreassen et al., *OmniFold: a method to simultaneously unfold all observables*, *Phys. Rev. Lett.* **124** (2020) 182001 [[arXiv:1911.09107](#)] [[INSPIRE](#)].
- [41] M. Bellagente et al., *Invertible networks or partons to detector and back again*, *SciPost Phys.* **9** (2020) 074 [[arXiv:2006.06685](#)] [[INSPIRE](#)].
- [42] A. Butter et al., *Generative unfolding with distribution mapping*, *SciPost Phys.* **18** (2025) 200 [[arXiv:2411.02495](#)] [[INSPIRE](#)].
- [43] N. Huetsch et al., *The landscape of unfolding with machine learning*, *SciPost Phys.* **18** (2025) 070 [[arXiv:2404.18807](#)] [[INSPIRE](#)].
- [44] UA1 collaboration, *A study of the  $D^*$  content of jets at the CERN  $p\bar{p}$  collider*, *Phys. Lett. B* **244** (1990) 566 [[INSPIRE](#)].
- [45] CDF collaboration, *A measurement of  $D^*$  production in jets from  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*, *Phys. Rev. Lett.* **64** (1990) 348 [[INSPIRE](#)].
- [46] LHCb collaboration, *Opposite-side flavour tagging of  $B$  mesons at the LHCb experiment*, *Eur. Phys. J. C* **72** (2012) 2022 [[arXiv:1202.4979](#)] [[INSPIRE](#)].
- [47] ALICE collaboration, *The ALICE experiment: a journey through QCD*, *Eur. Phys. J. C* **84** (2024) 813 [[arXiv:2211.04384](#)] [[INSPIRE](#)].
- [48] M. Cacciari and G.P. Salam, *Dispelling the  $N^3$  myth for the  $k_t$  jet-finder*, *Phys. Lett. B* **641** (2006) 57 [[hep-ph/0512210](#)] [[INSPIRE](#)].
- [49] M. Cacciari, G.P. Salam and G. Soyez, *The catchment area of jets*, *JHEP* **04** (2008) 005 [[arXiv:0802.1188](#)] [[INSPIRE](#)].
- [50] D0 collaboration,  *$b$ -jet identification in the D0 experiment*, *Nucl. Instrum. Meth. A* **620** (2010) 490 [[arXiv:1002.4224](#)] [[INSPIRE](#)].
- [51] CDF collaboration, *Heavy flavor properties of jets produced in  $p\bar{p}$  interactions at  $\sqrt{s} = 1.8$  TeV*, *Phys. Rev. D* **69** (2004) 072004 [[hep-ex/0311051](#)] [[INSPIRE](#)].
- [52] CMS collaboration, *Identification of  $b$ -quark jets with the CMS experiment*, *2013 JINST* **8** P04013 [[arXiv:1211.4462](#)] [[INSPIRE](#)].

- [53] ATLAS collaboration, *Performance of b-jet identification in the ATLAS experiment*, 2016 *JINST* **11** P04008 [[arXiv:1512.01094](#)] [[INSPIRE](#)].
- [54] ALICE collaboration, *Measurement of inclusive charged-particle b-jet production in pp and p-Pb collisions at  $\sqrt{s_{NN}} = 5.02$  TeV*, *JHEP* **01** (2022) 178 [[arXiv:2110.06104](#)] [[INSPIRE](#)].
- [55] ATLAS collaboration, *ATLAS flavour-tagging algorithms for the LHC run 2 pp collision dataset*, *Eur. Phys. J. C* **83** (2023) 681 [[arXiv:2211.16345](#)] [[INSPIRE](#)].
- [56] CMS collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, 2018 *JINST* **13** P05011 [[arXiv:1712.07158](#)] [[INSPIRE](#)].
- [57] LHCb collaboration, *New algorithms for identifying the flavour of  $B^0$  mesons using pions and protons*, *Eur. Phys. J. C* **77** (2017) 238 [[arXiv:1610.06019](#)] [[INSPIRE](#)].
- [58] T. Kinoshita, *Mass singularities of Feynman amplitudes*, *J. Math. Phys.* **3** (1962) 650 [[INSPIRE](#)].
- [59] T.D. Lee and M. Nauenberg, *Degenerate systems and mass singularities*, *Phys. Rev.* **133** (1964) B1549 [[INSPIRE](#)].
- [60] A. Banfi, G.P. Salam and G. Zanderighi, *Principles of general final-state resummation and automated implementation*, *JHEP* **03** (2005) 073 [[hep-ph/0407286](#)] [[INSPIRE](#)].
- [61] Y.L. Dokshitzer, V.A. Khoze and S.I. Troian, *On specific QCD properties of heavy quark fragmentation ('dead cone')*, *J. Phys. G* **17** (1991) 1602 [[INSPIRE](#)].
- [62] Y.L. Dokshitzer, V.A. Khoze and S.I. Troian, *Specific features of heavy quark production. LPHD approach to heavy particle spectra*, *Phys. Rev. D* **53** (1996) 89 [[hep-ph/9506425](#)] [[INSPIRE](#)].
- [63] M. Cacciari, G.P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896 [[arXiv:1111.6097](#)] [[INSPIRE](#)].
- [64] A.J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft drop*, *JHEP* **05** (2014) 146 [[arXiv:1402.2657](#)] [[INSPIRE](#)].
- [65] M. Dasgupta and G.P. Salam, *Resummation of nonglobal QCD observables*, *Phys. Lett. B* **512** (2001) 323 [[hep-ph/0104277](#)] [[INSPIRE](#)].
- [66] JADE collaboration, *Experimental studies on multi-jet production in  $e^+e^-$  annihilation at PETRA energies*, *Z. Phys. C* **33** (1986) 23 [[INSPIRE](#)].
- [67] JADE collaboration, *Experimental investigation of the energy dependence of the strong coupling strength*, *Phys. Lett. B* **213** (1988) 235 [[INSPIRE](#)].
- [68] D. Bertolini, T. Chan and J. Thaler, *Jet observables without jet algorithms*, *JHEP* **04** (2014) 013 [[arXiv:1310.7584](#)] [[INSPIRE](#)].
- [69] A.J. Larkoski, D. Neill and J. Thaler, *Jet shapes with the broadening axis*, *JHEP* **04** (2014) 017 [[arXiv:1401.2158](#)] [[INSPIRE](#)].
- [70] A.J. Larkoski, *Flavor factorization at two-loops*, *Eur. Phys. J. C* **84** (2024) 1131 [[arXiv:2402.06735](#)] [[INSPIRE](#)].
- [71] M. Dasgupta, A. Fregoso, S. Marzani and G.P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029 [[arXiv:1307.0007](#)] [[INSPIRE](#)].
- [72] 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](#)].
- [73] CMS collaboration, *Measurement of the production cross section for a W boson in association with a charm quark in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **84** (2024) 27 [[arXiv:2308.02285](#)] [[INSPIRE](#)].

- [74] ATLAS collaboration, *Measurements of the production cross-section for a Z boson in association with b- or c-jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Eur. Phys. J. C* **84** (2024) 984 [[arXiv:2403.15093](#)] [[INSPIRE](#)].
- [75] ATLAS collaboration, *ATLAS b-jet identification performance and efficiency measurement with  $t\bar{t}$  events in pp collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79** (2019) 970 [[arXiv:1907.05120](#)] [[INSPIRE](#)].
- [76] ATLAS collaboration, *Measurements of the production cross-section for a Z boson in association with b-jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *JHEP* **07** (2020) 044 [[arXiv:2003.11960](#)] [[INSPIRE](#)].
- [77] CMS collaboration, *Measurement of the production cross section for Z+b jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **105** (2022) 092014 [[arXiv:2112.09659](#)] [[INSPIRE](#)].
- [78] LHCb collaboration, *The LHCb upgrade I*, 2024 *JINST* **19** P05065 [[arXiv:2305.10515](#)] [[INSPIRE](#)].
- [79] LHCb RICH GROUP collaboration, *Performance of the LHCb RICH detector at the LHC*, *Eur. Phys. J. C* **73** (2013) 2431 [[arXiv:1211.6759](#)] [[INSPIRE](#)].
- [80] ATLAS collaboration, *Transforming jet flavour tagging at ATLAS*, [arXiv:2505.19689](#) [[INSPIRE](#)].
- [81] CMS collaboration, *Jet fragmentation function and groomed substructure of bottom quark jets in proton-proton collisions at 5.02 TeV*, CMS-PAS-HIN-24-005, CERN, Geneva, Switzerland (2024) [[INSPIRE](#)].
- [82] F. Caola, G. Luisoni, K. Melnikov and R. Röntsch, *NNLO QCD corrections to associated WH production and  $H \rightarrow b\bar{b}$  decay*, *Phys. Rev. D* **97** (2018) 074022 [[arXiv:1712.06954](#)] [[INSPIRE](#)].
- [83] A. Behring et al., *Bottom quark mass effects in associated WH production with the  $H \rightarrow b\bar{b}$  decay through NNLO QCD*, *Phys. Rev. D* **101** (2020) 114012 [[arXiv:2003.08321](#)] [[INSPIRE](#)].
- [84] F. Caola, K. Melnikov and R. Röntsch, *Nested soft-collinear subtractions in NNLO QCD computations*, *Eur. Phys. J. C* **77** (2017) 248 [[arXiv:1702.01352](#)] [[INSPIRE](#)].
- [85] F. Caola, K. Melnikov and R. Röntsch, *Analytic results for color-singlet production at NNLO QCD with the nested soft-collinear subtraction scheme*, *Eur. Phys. J. C* **79** (2019) 386 [[arXiv:1902.02081](#)] [[INSPIRE](#)].
- [86] F. Caola, K. Melnikov and R. Röntsch, *Analytic results for decays of color singlets to gg and  $q\bar{q}$  final states at NNLO QCD with the nested soft-collinear subtraction scheme*, *Eur. Phys. J. C* **79** (2019) 1013 [[arXiv:1907.05398](#)] [[INSPIRE](#)].
- [87] K. Asteriadis, F. Caola, K. Melnikov and R. Röntsch, *Analytic results for deep-inelastic scattering at NNLO QCD with the nested soft-collinear subtraction scheme*, *Eur. Phys. J. C* **80** (2020) 8 [[arXiv:1910.13761](#)] [[INSPIRE](#)].
- [88] M. Czakon, *A novel subtraction scheme for double-real radiation at NNLO*, *Phys. Lett. B* **693** (2010) 259 [[arXiv:1005.0274](#)] [[INSPIRE](#)].
- [89] M. Czakon, *Double-real radiation in hadronic top quark pair production as a proof of a certain concept*, *Nucl. Phys. B* **849** (2011) 250 [[arXiv:1101.0642](#)] [[INSPIRE](#)].
- [90] M. Czakon and D. Heymes, *Four-dimensional formulation of the sector-improved residue subtraction scheme*, *Nucl. Phys. B* **890** (2014) 152 [[arXiv:1408.2500](#)] [[INSPIRE](#)].
- [91] M. Czakon, A. van Hameren, A. Mitov and R. Poncelet, *Single-jet inclusive rates with exact color at  $O(\alpha_s^4)$* , *JHEP* **10** (2019) 262 [[arXiv:1907.12911](#)] [[INSPIRE](#)].

- [92] NNPDF collaboration, *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663 [[arXiv:1706.00428](#)] [[INSPIRE](#)].
- [93] A. Buckley et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132 [[arXiv:1412.7420](#)] [[INSPIRE](#)].
- [94] CMS collaboration, *Measurements of the associated production of a Z boson and b jets in pp collisions at  $\sqrt{s} = 8$  TeV*, *Eur. Phys. J. C* **77** (2017) 751 [[arXiv:1611.06507](#)] [[INSPIRE](#)].
- [95] PDF4LHC WORKING GROUP collaboration, *The PDF4LHC21 combination of global PDF fits for the LHC run III*, *J. Phys. G* **49** (2022) 080501 [[arXiv:2203.05506](#)] [[INSPIRE](#)].
- [96] M. Czakon, A. Mitov, M. Pellen and R. Poncelet, *NNLO QCD predictions for  $W+c$ -jet production at the LHC*, *JHEP* **06** (2021) 100 [[arXiv:2011.01011](#)] [[INSPIRE](#)].
- [97] H.B. Hartanto, R. Poncelet, A. Popescu and S. Zoia, *Next-to-next-to-leading order QCD corrections to  $Wb\bar{b}$  production at the LHC*, *Phys. Rev. D* **106** (2022) 074016 [[arXiv:2205.01687](#)] [[INSPIRE](#)].
- [98] H.B. Hartanto, R. Poncelet, A. Popescu and S. Zoia, *Flavour anti- $k_T$  algorithm applied to  $Wb\bar{b}$  production at the LHC*, [arXiv:2209.03280](#) [[INSPIRE](#)].
- [99] M. Czakon, A. Mitov, M. Pellen and R. Poncelet, *A detailed investigation of  $W+c$ -jet at the LHC*, *JHEP* **02** (2023) 241 [[arXiv:2212.00467](#)] [[INSPIRE](#)].
- [100] SHERPA collaboration, *Event generation with Sherpa 3*, *JHEP* **12** (2024) 156 [[arXiv:2410.22148](#)] [[INSPIRE](#)].
- [101] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *Eur. Phys. J. C* **76** (2016) 196 [[arXiv:1512.01178](#)] [[INSPIRE](#)].
- [102] G. Bewick et al., *Herwig 7.3 release note*, *Eur. Phys. J. C* **84** (2024) 1053 [[arXiv:2312.05175](#)] [[INSPIRE](#)].
- [103] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029 [[hep-ph/0204244](#)] [[INSPIRE](#)].
- [104] S. Hoeche, F. Krauss, M. Schonherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049 [[arXiv:1111.1220](#)] [[INSPIRE](#)].
- [105] S. Plätzer and S. Gieseke, *Dipole showers and automated NLO matching in Herwig++*, *Eur. Phys. J. C* **72** (2012) 2187 [[arXiv:1109.6256](#)] [[INSPIRE](#)].
- [106] F. Krauss, R. Kuhn and G. Soff, *AMEGIC++ 1.0: a matrix element generator in C++*, *JHEP* **02** (2002) 044 [[hep-ph/0109036](#)] [[INSPIRE](#)].
- [107] T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039 [[arXiv:0808.3674](#)] [[INSPIRE](#)].
- [108] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038 [[arXiv:0709.1027](#)] [[INSPIRE](#)].
- [109] B.R. Webber, *A QCD model for jet fragmentation including soft gluon interference*, *Nucl. Phys. B* **238** (1984) 492 [[INSPIRE](#)].
- [110] J.-C. Winter, F. Krauss and G. Soff, *A modified cluster hadronization model*, *Eur. Phys. J. C* **36** (2004) 381 [[hep-ph/0311085](#)] [[INSPIRE](#)].
- [111] G.S. Chahal and F. Krauss, *Cluster hadronisation in Sherpa*, *SciPost Phys.* **13** (2022) 019 [[arXiv:2203.11385](#)] [[INSPIRE](#)].

- [112] M. Knobbe, D. Reichelt and S. Schumann, *(N)NLO+NLL' accurate predictions for plain and groomed 1-jettiness in neutral current DIS*, *JHEP* **09** (2023) 194 [[arXiv:2306.17736](#)] [[INSPIRE](#)].
- [113] T. Sjostrand and M. van Zijl, *A multiple interaction model for the event structure in hadron collisions*, *Phys. Rev. D* **36** (1987) 2019 [[INSPIRE](#)].
- [114] S. Plätzer and S. Gieseke, *Coherent parton showers with local recoils*, *JHEP* **01** (2011) 024 [[arXiv:0909.5593](#)] [[INSPIRE](#)].
- [115] S. Gieseke, P. Stephens and B. Webber, *New formalism for QCD parton showers*, *JHEP* **12** (2003) 045 [[hep-ph/0310083](#)] [[INSPIRE](#)].
- [116] S. Gieseke, P. Kirchga er and S. Plätzer, *Baryon production from cluster hadronisation*, *Eur. Phys. J. C* **78** (2018) 99 [[arXiv:1710.10906](#)] [[INSPIRE](#)].
- [117] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, *Parton fragmentation and string dynamics*, *Phys. Rept.* **97** (1983) 31 [[INSPIRE](#)].
- [118] 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](#)] [[INSPIRE](#)].
- [119] J.R. Christiansen and P.Z. Skands, *String formation beyond leading colour*, *JHEP* **08** (2015) 003 [[arXiv:1505.01681](#)] [[INSPIRE](#)].
- [120] L. L nnblad et al., *TheP8I: an interface between the Pythia8 and the ThePEG toolkit*, <https://gitlab.cern.ch/TheP8I/TheP8I>.
- [121] D. Divisova, M. Myska, P. Sarmah and A. Si dmok, *Herwig 7 with the Lund string model: tuning and comparative hadronization studies*, in preparation.
- [122] C. Bierlich et al., *Robust independent validation of experiment and theory: Rivet version 3*, *SciPost Phys.* **8** (2020) 026 [[arXiv:1912.05451](#)] [[INSPIRE](#)].
- [123] D. Reichelt et al., *Phenomenology of jet angularities at the LHC*, *JHEP* **03** (2022) 131 [[arXiv:2112.09545](#)] [[INSPIRE](#)].
- [124] S.J. Brodsky, P. Hoyer, C. Peterson and N. Sakai, *The intrinsic charm of the proton*, *Phys. Lett. B* **93** (1980) 451 [[INSPIRE](#)].
- [125] NNPDF collaboration, *Evidence for intrinsic charm quarks in the proton*, *Nature* **608** (2022) 483 [[arXiv:2208.08372](#)] [[INSPIRE](#)].
- [126] M. Guzzi et al., *The persistent nonperturbative charm enigma*, *Phys. Lett. B* **843** (2023) 137975 [[arXiv:2211.01387](#)] [[INSPIRE](#)].
- [127] R. Gauld et al., *NNLO QCD predictions for Z-boson production in association with a charm jet within the LHCb fiducial region*, *Eur. Phys. J. C* **83** (2023) 336 [[arXiv:2302.12844](#)] [[INSPIRE](#)].
- [128] NNLOJET collaboration, *NNLOJET: a parton-level event generator for jet cross sections at NNLO QCD accuracy*, [arXiv:2503.22804](#) [[INSPIRE](#)].
- [129] R. Gauld et al., *Associated production of a Higgs boson decaying into bottom quarks and a weak vector boson decaying leptonically at NNLO in QCD*, *JHEP* **10** (2019) 002 [[arXiv:1907.05836](#)] [[INSPIRE](#)].
- [130] R. Gauld et al., *Predictions for Z-boson production in association with a b-jet at  $\mathcal{O}(\alpha_s^3)$* , *Phys. Rev. Lett.* **125** (2020) 222002 [[arXiv:2005.03016](#)] [[INSPIRE](#)].
- [131] A. Gehrmann-De Ridder et al., *Precise QCD predictions for W-boson production in association with a charm jet*, *Eur. Phys. J. C* **84** (2024) 361 [[arXiv:2311.14991](#)] [[INSPIRE](#)].

- [132] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [[arXiv:1405.0301](#)] [[INSPIRE](#)].
- [133] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [[arXiv:1410.3012](#)] [[INSPIRE](#)].
- [134] M. Bahr et al., *Herwig++ physics and manual*, *Eur. Phys. J. C* **58** (2008) 639 [[arXiv:0803.0883](#)] [[INSPIRE](#)].
- [135] J. Bellm et al., *Herwig 7.2 release note*, *Eur. Phys. J. C* **80** (2020) 452 [[arXiv:1912.06509](#)] [[INSPIRE](#)].
- [136] ALICE collaboration, *The ALICE experiment: a journey through QCD*, *Eur. Phys. J. C* **84** (2024) 813 [[arXiv:2211.04384](#)] [[INSPIRE](#)].
- [137] LHCb collaboration, *First measurement of b-jet mass with and without grooming*, [arXiv:2505.11955](#) [[INSPIRE](#)].
- [138] LHCb collaboration, *Measurement of the Lund plane for light- and beauty-quark jets*, [arXiv:2505.23530](#) [[INSPIRE](#)].
- [139] LHCb collaboration, *Study of  $J/\psi$  production in jets*, *Phys. Rev. Lett.* **118** (2017) 192001 [[arXiv:1701.05116](#)] [[INSPIRE](#)].
- [140] LHCb collaboration, *Measurement of charged hadron production in Z-tagged jets in proton-proton collisions at  $\sqrt{s} = 8$  TeV*, *Phys. Rev. Lett.* **123** (2019) 232001 [[arXiv:1904.08878](#)] [[INSPIRE](#)].
- [141] LHCb collaboration, *Multidifferential study of identified charged hadron distributions in Z-tagged jets in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, *Phys. Rev. D* **108** (2023) L031103 [[arXiv:2208.11691](#)] [[INSPIRE](#)].
- [142] LHCb collaboration, *Measurements of  $\psi(2S)$  and  $\chi_{c1}(3872)$  production within fully reconstructed jets*, *Eur. Phys. J. C* **85** (2025) 562 [[arXiv:2410.18018](#)] [[INSPIRE](#)].
- [143] M. Płoskoń and E.D. Lesser, *heppy Github repository*, <https://github.com/matplo/heppy>.
- [144] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [[hep-ph/9707323](#)] [[INSPIRE](#)].
- [145] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in the proceedings of the *Workshop on Monte Carlo generators for HERA physics (plenary starting meeting)*, (1998) [[hep-ph/9907280](#)] [[INSPIRE](#)].
- [146] F.A. Dreyer, G.P. Salam and G. Soyez, *The Lund jet plane*, *JHEP* **12** (2018) 064 [[arXiv:1807.04758](#)] [[INSPIRE](#)].
- [147] A. Ghira, S. Marzani and G. Ridolfi, *A consistent resummation of mass and soft logarithms in processes with heavy flavours*, *JHEP* **11** (2023) 120 [[arXiv:2309.06139](#)] [[INSPIRE](#)].
- [148] S. Caletti, A. Ghira and S. Marzani, *On heavy-flavour jets with soft drop*, *Eur. Phys. J. C* **84** (2024) 212 [[arXiv:2312.11623](#)] [[INSPIRE](#)].
- [149] P.K. Dhani et al., *Heavy flavour jet substructure*, *JHEP* **02** (2025) 046 [[arXiv:2410.05415](#)] [[INSPIRE](#)].
- [150] ALICE collaboration, *Direct observation of the dead-cone effect in quantum chromodynamics*, *Nature* **605** (2022) 440 [*Erratum ibid.* **607** (2022) E22] [[arXiv:2106.05713](#)] [[INSPIRE](#)].