Heavy-flavoured quark production in association with vector bosons at the LHC

S. Caletti, M. Muskinja

July 10, 2024

Abstract

The production of electroweak gauge bosons (V = W, Z) in association with heavy flavor (HF) quarks (b/c) serves as a crucial probe for various key physics aspects at the LHC: perturbative QCD (pQCD), HF content of the proton PDF, and non-pQCD phenomena such as the fragmentation function of HF quarks. Furthermore, V+HF modeling is a limiting factor in many searches for new physics phenomena and precision measurements. Progresses in jet flavour tagging techniques, as well as calculations and measurements with identified hadrons in the final state, are thus of paramount importance. We will review the experimental techniques used to measure the V+HF process at the LHC and discuss the theoretical challenges in modeling this process.

Contents

1 Motivation: why we measure heavy-flavour? (Miha)

Precision measurements of V+jets and V+HF processes in proton-proton collisions provide an important test of perturbative QCD (pQCD), non-perturbative QCD effects such as jet fragmentation and subsequent hadronization, and of the internal structure of the proton by constraining the proton PDFs. By identifying heavy flavor (HF) quarks in the final state produced in association with a vector boson ($V = W, Z$), we probe the HF quark content of the proton; s, c, and b PDFs. Additionally, V+HF measurements are crucial to validate and tune Monte Carlo (MC) simulation software in aspects such as the merging of hard-scatter simulations with parton shower simulations, different flavor number schemes (FNS) in the generators, different parton shower tunes, and others.

Many direct searches for New Physics processes or precision measurements of EW parameters and Higgs couplings and properties rely on V+jets and V+HF MC simulations, which are used to model the most common Standard Model backgrounds present in these measurements. In order to achieve the desired precision of these searches at the end of the LHC Run 3 or the High-Luminosity (HL) LHC, where the statistical uncertainties from the finite proton-poroton collision dataset will be greatly reduced, MC simulations of V+HF processes need to be drastically improved as well. The improvements can only be realized by carrying out dedicated V+HF precision measurements with the LHC experiments and by advancing our understanding of the pQCD and non-pQCD processes governing the production of these processes.

Recently, the ATLAS experiment performed two V+HF measurements: the Run 2 Z+HF measurement [ATL24] and the Run 2 $W+D^{\pm(*)}$ measurement [ATL23]. Similarly, the CMS experiment performed: the Run 2 Z+c [CMS21] (inclusive SV finder), Run 2 Z+HF / Z+light [CMS20] (combined secondary vertex v2), Run 2 $Z+b(b)$ (`DeepCSV`), W+c at 8 TeV [CMS22] and 13 TeV [CMS24].

Several important advancements in the theory calculations of V+HF processes along with the common experimental methods of measuring these processes at the LHC will be presented.

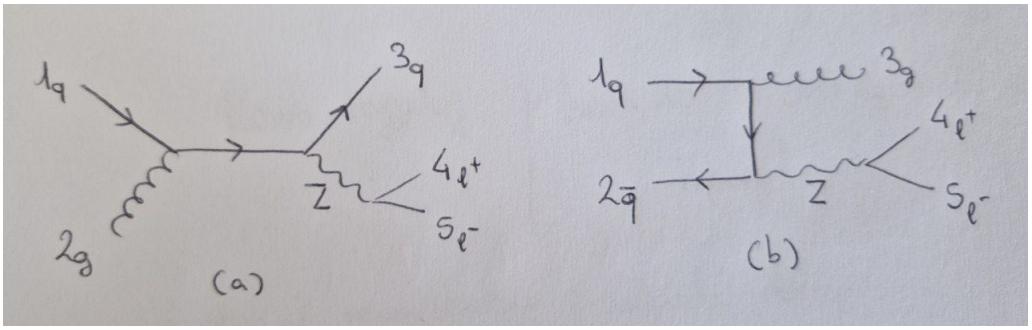

Figure 1: (Some) LO feynman diagram for Z+jet

2 $V+$ flavoured jet in the antenna framework and recent flav. algo (Simone)

The matrix elements for both the W and the Z boson processes are dependent on the flavour of the quarks involved in a non-trivial way, and this is what makes these processes particularly interesting for flavour physics. Notice that u-type and d-type quarks couple differently to the Z/W boson, thus we introduce $N_u = 2$ and $N_d = 3$, which summed are the active number of flavours, i.e. $N_f = 5$.

2.1 V+jet @ LO

In order to discuss the V+jet in the antenna formalism, we introduce the notation starting from the LO contribution. B-type matrix elements (ME) feature a single quark line. At LO, these processes, include only qg and $q\bar{q}$ initiated channels, and are denoted by:

$$d\sigma^{B,Z}_{g,(qg)} \approx d\Phi_{Z+1} B_1^{Z,0}(1_q, 2_g, 3_{\bar{q}}, 4_{l^+}, 5_{l^-}) J_1^{(1)}(3_{\bar{q}}),, \tag{1}$$

where $J_a^{(b)}$ is the jet function for a-jets out of b partons. This function must be IRC safe in order to guarantee real-virtual IRC singularities cancellations. A single Feynman diagram contributes to this channel and is represented in Fig. 1. The factor $d\Phi_{Z+1}$ is the phase space of the Z-boson plus one parton.

The other channel is given by :

$$d\sigma^{B,Z}_{g,(q\bar{q})} \approx d\Phi_{Z+1} B_1^{Z,0}(1_q, 3_g, 2_{\bar{q}}, 4_{l^+}, 5_{l^-}) J_1^{(1)}(3_g)\,. \tag{2}$$

This is the $Z + jet$ process. Notice that the $W + jet$ process is the same but $Z \to W$ and $l^{\pm} \to \nu$ (neutrinos). Specifically for the W we also have the $q\bar{Q}$ channel

$$d\sigma^{B,W}_{g,(q\bar{Q})} \approx d\Phi_{W+1} B_1^{W,0}(1_q, 3_g, 2_{\bar{Q}}, 4_{l^+}, 5_{l^-}) J_1^{(1)}(3_g)\,, \tag{3}$$

where q and $\bar{Q}$ belong to the same $SU(2)_L$ doublet. Thus, the Born contribution for the Z+jet process is given by

$$d\hat{\sigma}^{B,V} = d\hat{\sigma}^{B,V}_{q,(qg)} + d\hat{\sigma}^{B,V}_{\bar{q},(\bar{q}g)} + d\hat{\sigma}^{B,V}_{g,(q\bar{q})}\,. \tag{4}$$

2.2 V+jet @ NLO: real contribution

We describe Z+jet which is simpler in the theory part. The real level contributions listed per initial state are

- gluon-gluon initiated
- quark-gluon initiated
- quark-quark (identical flavour) initiated

- quark-quark (non-identical flavour) initiated

Thus the whole real contribution is given by

$$d\hat{\sigma}^{R,Z} = d\hat{\sigma}^{R,Z}_{qg} + d\hat{\sigma}^{R,Z}_{\bar{q}g} + d\hat{\sigma}^{R,Z}_{gg} + d\hat{\sigma}^{R,Z}_{q\bar{q}} + d\hat{\sigma}^{R,Z}_{qQ} + d\hat{\sigma}^{R,Z}_{q\bar{Q}} \tag{5}$$

We will focus instead on the qg-initiated contribution, which is represented by Fig. 4.

$$\begin{aligned} d\sigma^{R,Z}_{qg} \approx d\Phi_{Z+2}\Big\{ N\Big[B_2^{Z,0}(1_q, 2_g, i_g, j_{\bar{q}}, Z) + [2_g \leftrightarrow 3_g] \\ -\frac{1}{N^2}\tilde{B}_2^{Z,0}(1_q, \tilde{2}_g, \tilde{i}_g, j_{\bar{q}}, Z)\Big]\Big\} J_1^{(2)}(p_3, p_4) \end{aligned} \tag{6}$$

Similarly, this is true for the W+jet.
The $\tilde{}$ on top of some momenta denotes that the corresponding gluons are photon–like and the term is subleading in the colour.

2.2.1 Infrared structure of V+jet @ NLO

We focus on the qg-initiated real contributions represented in Fig. 4.
All the IRC limits are

- 2//i, i//j and i$\to$ 0 for the diagram A
- 1//i, i//2, 2//j and i$\to$ 0 for the diagram B

Notice that these divergences only come from colour connected partons. These divergences have to be subtracted in order to evaluate numerically the real contribution.

2.3 Real subtraction term

Describe briefly how fixed order results are computed.
First we have differential cross-section as convolution with PDFs

$$d\sigma = \sum_{a,b} d\hat{\sigma}_{ab} \otimes f_a \otimes f_b \tag{7}$$

The parton-level cross section can be compute in perturbation theory

$$d\hat{\sigma}_{ab} = d\hat{\sigma}^{LO}_{ab} + d\hat{\sigma}^{NLO}_{ab} + d\hat{\sigma}^{NNLO}_{ab} + \cdots \tag{8}$$

The LO contribution is given by the integral of the Born cross section over the n-body particle phase space

$$d\hat{\sigma}^{LO}_{ab} = \int_n d\hat{\sigma}^{B}_{ab} \tag{9}$$

We will discuss more in detail the NLO contribution, which is composed by the real and the virtual contributions. These contributions live in two different phase space: n+1 and n-parton respectively. We would like to evaluate numerically and we need subtraction techniques to achieve this goal. Antenna is a local subtraction and essentially is a method to explicitly derive a possible for those subtraction terms which regulates IRC divergences of the ME.

$$d\hat{\sigma}^{NLO}_{ab} = \int_{n+1} \left(d\hat{\sigma}^{R}_{ab} - d\hat{\sigma}^{S}_{ab} \right) + \int_n \left(d\hat{\sigma}^{V}_{ab} + d\hat{\sigma}^{T}_{ab} \right) \tag{10}$$

In order we have

- the real contribution, i.e. tree level cross section with an extra parton in the final state.
- real subtraction term, which must have a behaviour which mimics the real ME in the IRC limits.
- the virtual contribution, i.e. one loop corrections to the Born.

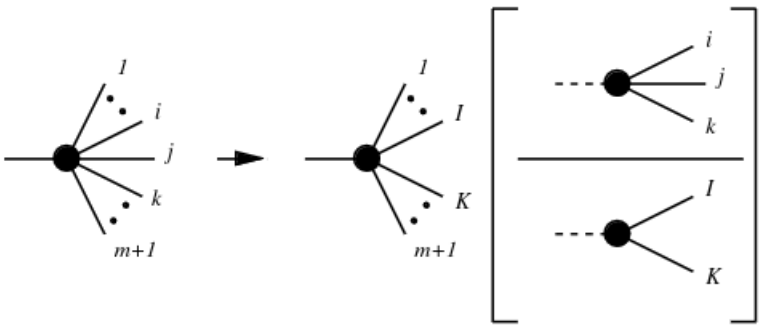

Figure 2: NLO antenna factorization in the reduced ME and the antenna.

- virtual subtraction, which can be obtained from the real one plus the mass factorization contribution.

The picture at NNLO becomes more complicated because we have RR, RV and VV corrections to the Born ME. These contributions needs to be subtracted as well by proper subtraction terms.
We discuss the structure of $d\hat{\sigma}^S$

$$d\sigma^S = \mathcal{N} \sum_{m+1} d\Phi_{m+1} \sum_j X03_{ijk} |\mathcal{M}_m(p_1, \ldots, \tilde{p}_I, \tilde{p}_K, \ldots, p_{m+1})|^2 J_m^{(m)}(p_1, \ldots, \tilde{p}_I, \tilde{p}_K, \ldots, p_{m+1}) \quad (11)$$

where we have, in order

- normalization factor
- sum over all the possible extra parton (usually for a fixed initial state configuration)
- m+1 phase space
- sum over j all the possible IRC limits for the j parton
- the antenna $X03$ which contains the IRC limits
- the reduced ME for m partons which factorize in the IRC limits
- the jet function

Momenta with the tilde, $\tilde{p}$, are after the mapping. We need to define the relation between the momenta of the m-particle ME and the $m+1$-particle ME for the real level. Eq. 11 is pictorially represented in Fig. 2.

2.4 V+jet @ NLO: virtual contribution

There is no virtual contribution for the gg-initiated channels, either for the W or for the Z boson. Similarly to before we only focus on the qg-initiated channel, for which we have

$$d\sigma_{qg}^{V,Z} = d\Phi_{Z+1} \Big\{ N \Big[B_1^{Z,1}(1_q, 2_g, j_{\bar{q}}, Z) - \frac{1}{N^2} \tilde{B}_1^{Z,1}(1_q, 2_g, j_{\bar{q}}, Z) + \frac{N_f}{N} \hat{B}_1^{Z,1}(1_q, 2_g, j_{\bar{q}}, Z) \Big] \Big\} J_1^{(1)}(p_3) \quad (12)$$

Similarly is true for the W+jet process.

2.4.1 The role of $J(\cdot)$ in NNLOJET

$J(\cdot)$ represents the jet algorithm and it is needed to distinguish between V+jet and V+2-jet configurations. But it has to do it in a IRC safe way as depicted in Fig. 3. Same will be valid for safe flavour algorithms.

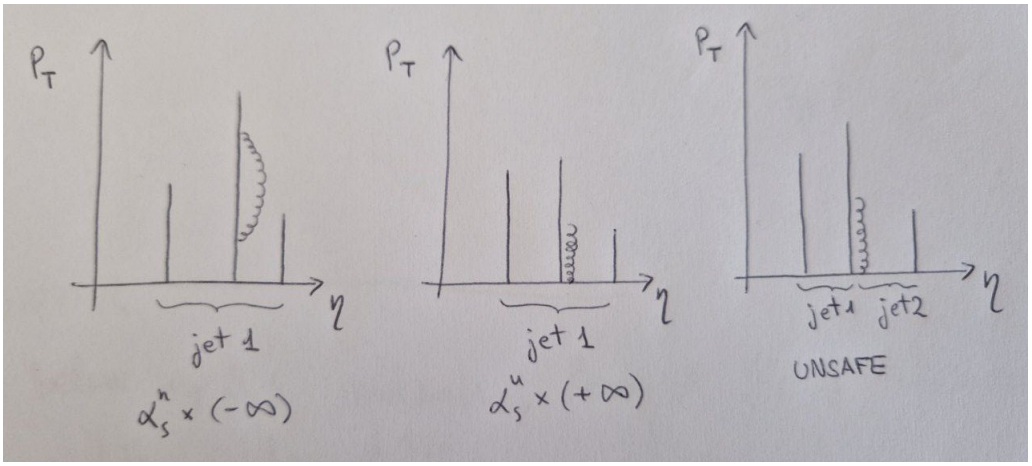

Figure 3: First two pictures represent the behaviour of an IRC safe clustering algorithm. The last one instead is unsafe if the algorithm gives us 1 jet for the virtual correction.

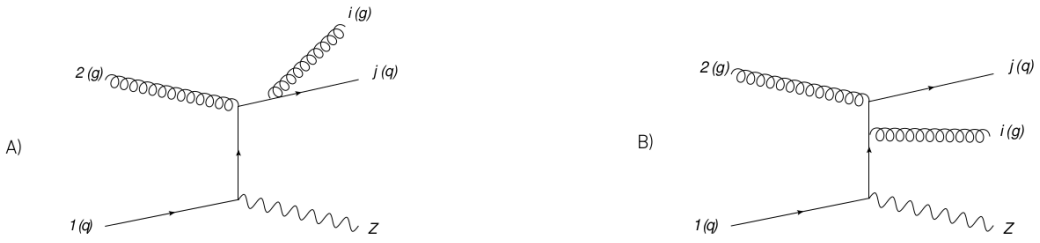

Figure 4: Real corrections to the qg-initiated Z+jet contribution.

2.5 Virtual subtraction term

The virtual subtraction term can be obtained from the real one exploiting phase space factorization, integrating over the phase space of the unresolved particle. We define the integrated antennae as

$$\mathcal{X}03_{ijk} = (8\pi^2(4\pi)^{-\epsilon}e^{\epsilon\gamma}) \int d\Phi_{X03_{ijk}} X03_{ijk} \tag{13}$$

where $d\Phi_{X_{ijk}}$ is the antenna phase space. Thus we obtain the relation between the virtual subtraction term and the integrated level real subtraction term

$$d\hat{\sigma}^T = -\int_1 d\hat{\sigma}^S + d\hat{\sigma}^{MF} \tag{14}$$

where the last one is the *mass factorization* contribution proper of the initial-final antennae. At NNLO we have to take into account separately

- single unresolved limits
- double unresolved
- single unresolved loop

and to consider the potential double counting of the IRC subtracted limits.

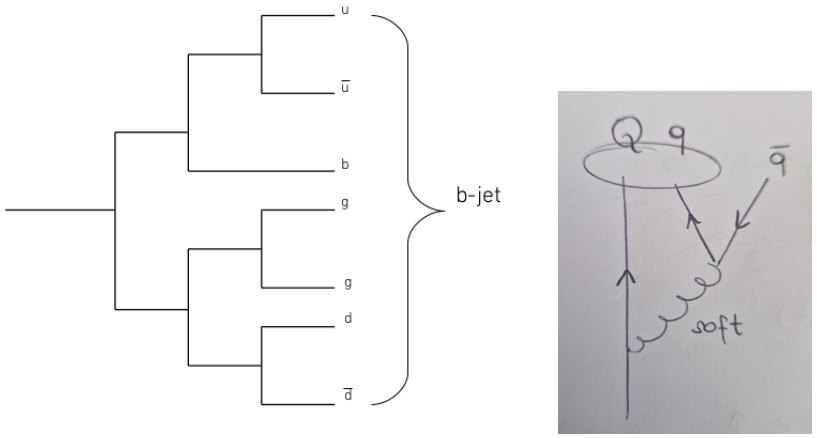

Figure 5: b-jet according to the net flavour on the left. On the right, NNLO configuration leading to soft sensitive flavour labelling with naive AKT net flavour.

2.6 Example: NLO real subtraction for V+jet qg-initiated contribution

The real subtraction term for Z+jet for the qg-initiated contribution is given by

$$\begin{aligned} d\hat{\sigma}_{qg}^{S,Z} = &+\text{qgD30II}(1,i,2) * \text{B1g0Z}([1],[2],j,Z) * \text{JET11}(j) \\ &-\text{qgA30II}(1,2,j) * \text{B1g0Z}([1],i,[2],Z) * \text{JET11}(i) \\ &+\text{gd30IF}(j,i,2) * \text{B1g0Z}(1,[2],[j,i],Z) * \text{JET11}([i,j]) \end{aligned} \tag{15}$$

where we have

- qgD30II takes care of the i//2, i//1 and i$\to$ 0 limits. D-type antennae are quark-gluon antennae with an extra gluon.
- qgA30II takes care of j//2. A-type antennae are quark-antiquark antennae with an extra gluon.
- gd30IF takes care of i//j, i//2 and i$\to$ 0. It is a sub-antenna of the D one.

The IF and II antennae differs from the FF especially at the integrated level, since in the integration we have to consider that the momenta of one (or two) parton(s) comes from the initial state and is a fraction of the momenta of the proton. A similar procedure is done also in the final state for the fragmented antennae.
We will discuss how to modify such subtraction term for V+hadron.

2.7 Flavour algorithms

To compute W+c-jet, i.e. processes with flavoured jets we modify the jet function in order to get flavour information from our parton clustering. Since we work with partons we simply define the net flavour of a collection of parton as the "sum" where particles count +1 and antiparticles count -1 if od the same flavour. Gluons count 0. An example of a b-jet is given in Fig. 5.

2.7.1 Gen-k_T naive flavour

First thing we do is just to count the net flavour of our AKT jet. This is unsafe at NNLO. The problematic configuration can be seen in Fig. 5.

To analytically see the IRC unsafety we can write the

$$d\Phi_3 |\mathcal{M}(z_q, z_{\bar{q}})|^2 \sim \frac{numerator}{\theta_{q\bar{q}}^4 (z_q + z_{\bar{q}})^2 (z_q \theta_{Qq}^2 + z_{\bar{q}} \theta_{Q\bar{q}}^2)^2} \tag{16}$$

and the measurement function for any generalized-kt algorithm is given by

$$\Theta_{\rm AKT} \sim \Theta(\min[z_Q^a, z_q^a]\theta_{Qq}^b < \min[z_Q^a, z_{\bar{q}}^a]\theta_{Q\bar{q}}^b)\Theta(\min[z_Q^a, z_q^a]\theta_{Qq}^b < \min[z_q^a, z_{\bar{q}}^a]\theta_{q\bar{q}}^b) \tag{17}$$

This means that the Qq distance is the smaller and Qq are reclustered first, as represented in the picture. The measurement function does nothing to regulate the $\theta_{q\bar{q}}$ divergence of this configuration. Without any grooming there is also a soft divergence when both the quarks becomes soft.

2.7.2 Flavour-k_T (2006)

First solution to the problem [BSZ06]. A different clustering metric, which is flavour sensitive. Since it depends on the flavour of the partons cannot be applied in experimental context. However, has been used to assign flavour labels for resummation and to match resummed results to fixed-order. This is the clustering algorithm. If the softer between i and j is flavoured, then

$$d_{ij}^{flav-kt} = \max[p_{ti}, p_{tj}]^\alpha \min[p_{ti}, p_{tj}]^{2-\alpha} \frac{\Delta R_{ij}^2}{R^2} \tag{18}$$

$$d_{iB}^{flav-kt} = \max[p_{ti}, p_{tB}(y_i)]^\alpha \min[p_{ti}, p_{tB}(y_i)]^{2-\alpha} \tag{19}$$

with

$$p_{tB}(y) = \sum_i p_{ti}\Big[\Theta(y > y_i)e^{y_i - y} + \Theta(y < y_i)\Big] \tag{20}$$

Otherwise we keep the standard measure

$$d_{ij} = \min[p_{ti}, p_{tj}]^2 \frac{\Delta R_{ij}^2}{R^2} \tag{21}$$

Consequence: for the clustering of a soft flavoured particle with a significantly harder particle, the BSZ metric is much larger than the same but with two similarly soft particles (because of the max). The result is that the soft pair is clustered first resolving the problematic configuration. Also the beam distance in the clustering algorithm is modified accordingly and is flavour sensitive. Problems: the kinematics of the jets might be very different from standard AKT jets. For example, we might have a wide angle $b\bar{b}$ pair at an angle bigger than R which is clustered first.

A "new" problematic IDS×IHC α_S^3 configuration (Fig. 6):

1. consider a hard parton 1 defining a jet, a soft wide angle qq pair IDS, and a IHC quark 4.
2. assume $p_{t2} < p_{t3}$ and $p_{t2} = z_{23}p_{t3}$
3. assume $y_1 = 0$ and $\phi_1 = 0$
4. assume $E_4 = z_{41}E_1$, then $y4 = \log \frac{2z_{41}p_{t1}}{p_{t4}}$
5. Then (when i flavoured, always here): $p_{tB}(0) \approx p_{t1}$ and $p_{tB}(y_4) \approx p_{t4} + p_{t1}e^{-y_4} \approx p_{t4}\Big(1 + \frac{1}{2z_{14}}\Big)$, where we used $p_{t4} \ll p_{t1}$
6. Without the soft quark pair there are three distances:
   - $d_{14} = p_{t1}^2 \Delta R_{14}^2 = p_{t1}^2 y_4^2$
   - $d_{1B} \approx p_{t1}^2$
   - $d_{4B} \approx p_{t4}^2\Big(1 + \frac{1}{2z_{14}}\Big)$
7. The smallest is d_{4B}, since $p_{t4} \ll p_{t1}$.
   Thus IHC particle 4 clusters first.
   Thus we have a flavoured jet made of just particle 1.

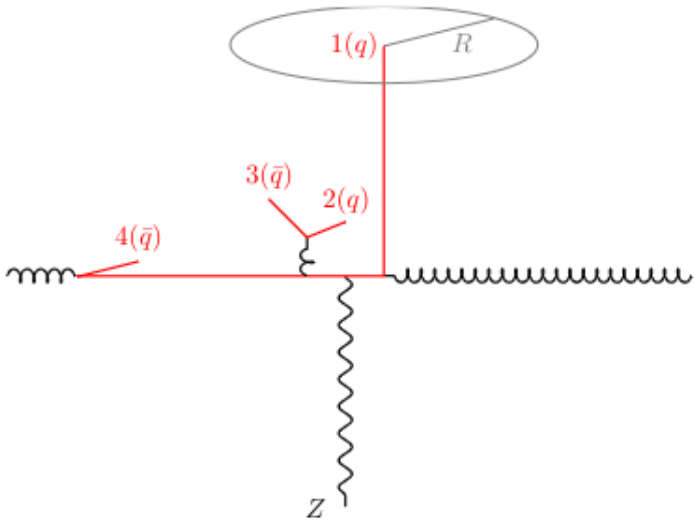

Figure 6: Example configuration to illustrate issues that arise across multiple algorithms when using a standard ΔR type angular measure in inter-particle distances.

8. Now we introduce the soft 23 pair.
   d_{2B} and d_{3B} are similar to $d_{1B} \sim p_{t1}^2$ since they are both central in rapidity.
9. New distances relevant for clustering
   - $d_{23} = p_{23}^2 \Delta R_{23}^2 \approx \frac{p_{t2}^2}{p_{t3}^2}$
   - $d_{24} = \max(p_{t4}^2, p_{t2}^2) \Delta R_{24}^2 \approx \max(p_{t4}^2, p_{t2}^2) \log^2 \frac{z_{14}^2 p_{t1}^2}{p_{t4}^2}$
   - we neglect d_{34} because we work in the $p_{t2} < p_{t3}$ limit and $d_{34} > d_{24}$
10. if d_{23} is the smallest: particle 2 and 3 annihilate, then 4 clusters with the beam and the hard jet is only 1 and is flavoured.
11. if d_{24} is the smallest: particle 2 and 4 annihilate, then 3 can (or cannot) cluster with 1 and we might have a flavourless jet.

d_{24} to be the smallest is a α_S^3 configuration that happens with a rate $\int d \log p_t / (\log p_t)^2$ which is convergent in the $p_t \to 0$ limit, but it is not a power law. This implies that this contribution might be compensated by higher order log contributions.

2.7.3 Soft Drop flavour

This algorithm resolve the problem at NNLO but it is still unsafe at N3LO [CLMR22]. However, it is relatively simple to implement from the experimental point of view, current calculation are usually NNLO apart for some exceptions, the kinematics of the jets is the same than AKT. The measure function eq. 17 is if we add Soft Drop is modified this way

$$\Theta_{\text{SDflav}} \sim \Theta_{\text{AKT}} \Theta(z_q > z_{cut}(\theta_{Qq}^2/R^2)^\beta) \Theta(z_{\bar{q}} < z_{cut}(\theta_{Q\bar{q}}^2/R^2)^\beta) \tag{22}$$

where we just added the SD condition to the AKT clustering.
Notice that the collinear $\theta_{q\bar{q}}$ is still not regularized by the measurement function, although the divergence for the gluon becoming soft (i.e. both q and $\bar{q}$ soft) has been removed.
We need to modify the reclustering step using JADE as reclustering. This way the measurement function becomes

$$\Theta_{\text{SDflav}} \sim \Theta(m_{Q\bar{q}}^2 > m_{Qq}^2) \Theta(m_{q\bar{q}}^2 > m_{Qq}^2) \Theta_{\text{SD}} \tag{23}$$

where the $\Theta_{\rm SD}$ is the two last heavisides we added for the Soft Drop.
Notice that now the $m^2_{q\bar{q}}$ becomes small both in the soft and in the collinear limit of q and $\bar{q}$, thus the reclustering heaviside regularized the $\theta_{q\bar{q}}$ divergence. However the algorithm is unsafe because of the following N3LO configuration: if there is an extra relatively hard gluon that passes the SD condition, then SD stops and does not remove the extra quark.

2.7.4 Flavour-AKT

This algorithm is safe to all order [CMP23]. However, the metric is different than AKT and it is still flavour sensitive. We can control the how much the metric is close to AKT through a parameter a and we recover AKT for $a \to 0$.
The idea is to change the clustering metric to

$$d_{ij}^{flav-AKT} = d_{ij}^{AKT} \times \begin{cases} \mathcal{S}_{ij} \\ 1 \end{cases} \tag{24}$$

where

$$d_{ij}^{AKT} = \min[p_{Ti}, p_{Tj}]^{-2} \frac{\Delta R_{ij}^2}{R^2} \tag{25}$$

$$\mathcal{S}_{ij} = 1 - \Theta(\kappa_{ij} < 1) \cos\left(\frac{\pi}{2}\kappa_{ij}\right) \quad \text{and} \quad \kappa_{ij} = \frac{1}{a}\frac{p_{Ti}^2 + p_{Tj}^2}{2p_{T,max}^2} \tag{26}$$

The metric is parametrically close to AKT and is the same in the $a \to 0$ limit.
$p_{T,max}$ is the transverse momentum of the hardest pseudojet in the event.
CMP (or flav-AKT) solves the problematic configuraton Fig. 5, i.e. when both the quark are very soft and κ is small. How it works:

- If $p_{T,max} \sim 1$
- then $\mathcal{S}_{ij} \sim \kappa^2 \sim \max[p_{Ti}^4, p_{Tj}^4]$
- then $d_{ij} \sim \max[p_{Ti}^2, p_{Tj}^2]\Delta R_{ij}^2$
- the AKT distance of a soft quark with a hard parton is $d_{ij} \sim \Delta R_{ij}$
- as a consequence, the soft $q\bar{q}$ is clustered first

However, there are still problems (IRC unsafe configurations) when $E_i, E_j \gg 1$ but $p_{Ti}, p_{Tj} \ll 1$, i.e. Fig. 7, i.e. IHC^2 configurations.
Problem explained:

1. $p_{t1}, p_{t2} \ll p_{t3}$ but $E_1, E_2 \sim E_3$
2. we assume $p_{t3} \sim \ E_3 \sim 1$ and particle 3 aligned with the x axis
3. $y_1 \sim -\log p_{t1}$ and $y_2 \sim \log p_{t2}$
4. $p_{t,max}$ is p_{t3}, so

$$\kappa_{12} = \frac{1}{2a}\frac{p_{t1}^2 + p_{t2}^2}{p_{t3}^2} \ll 1 \tag{27}$$

5. The distance between oppositely flavour particle is

$$\begin{aligned} d_{12} &= \frac{1}{\max(p_{t1}^2, p_{t2}^2)} \Delta R_{12}^2 \Big(1 - \cos(\frac{\pi}{2}\kappa_{12})\Big) \\ &\approx \frac{1}{\max(p_{t1}^2, p_{t2}^2)} \Delta R_{12}^2 \frac{1}{2}\Big(\frac{\pi}{2}\kappa_{12}\Big)^2 \\ &\approx \frac{\pi^2}{32a^2} \Delta R_{12}^2 \frac{\max(p_{t1}^2, p_{t2}^2)}{p_{t3}^4} \\ &\approx \frac{\pi^2}{32a^2} \Big(\log\frac{p_{t3}}{p_{t1}} + \log\frac{p_{t3}}{p_{t2}}\Big)^2 \frac{\max(p_{t1}^2, p_{t2}^2)}{p_{t3}^4} \end{aligned} \tag{28}$$

where ΔR_{12} is dominated by the large rapidity difference.

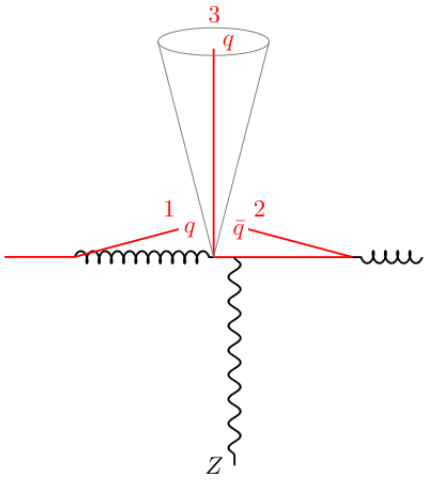

Figure 7: Example α_S^2 configuration that yields an issue for the CMP algorithm. There are two oppositely-flavoured initial-state hard-collinear splittings (b and $\bar{b}$, labelled 1 and 2), and a hard particle 3 at central rapidity.

6. The other distances are
   - $d_{iB} = \frac{1}{p_{ti}^2}$
   - $d_{i3} \approx \frac{y_i^2}{p_{t3}^2}$
7. when $p_{t1}, p_{t2} \ll p_{t3}$, then d_{12} is the smallest. The first step in the clustering is to combine 1 and 2 returning a flavourless pseudojet with
   - $p_{t,(1+2)} \approx \max(p_{t1}, p_{t2})$
   - $y_{(1+2)} \approx \frac{1}{2} \log \frac{E_1}{E_2}$
   - $m^1_{(1+2)} \approx 4E_1E_2$

   which is a weird pseudojet because it has a transverse momentum smaller than its invariant mass.
8. if $\Delta R_{(1+2),3} < R$, (1+2) and 3 cluster together (always possible for some choice of azimuth and energies), the $(1+2+3)$ cluster is kinematically very different to particle 3 alone. This exhibits a sensitivity to IRC physics, so it is unsafe.

Last variant of the algorithm

$$\bar{\mathcal{S}}_{ij} = \mathcal{S}_{ij} \frac{\Omega_{ij}^2}{\Delta R_{ij}^2} \tag{29}$$

$$\Omega_{ij}^2 = 2\Big[\frac{1}{\omega^2}(\cosh(\omega \Delta y_{ij}) - 1) - (\cos \Delta\phi_{ij} - 1)\Big] \tag{30}$$

The problem pointed out before is solved by this modification because now

$$\begin{aligned} d_{12}^{\Omega} &\sim \Big(\frac{p_{t3}^2}{p_{t1}p_{t2}}\Big)^{\omega} \frac{\max(p_{t1}^2, p_{t2}^2)}{p_{t3}^4} \\ &\sim \frac{\max(p_{t1}^2, p_{t2}^2)}{p_{t1}p_{t2}p_{t3}^2} \Big(\frac{p_{t3}^2}{p_{t1}p_{t2}}\Big)^{\omega-1} \end{aligned} \tag{31}$$

For $\omega > 1$, we ensure that $d_{12}^{\Omega} \gg d_{3B}$, thus particle 3 becomes a jet before any 1+2 clustering and we avoid kinematical issue to jet 3.

The same solution, i.e. replacing ΔR_{ij} with Ω_{ij}, can be adopted to avoid the higher order issues we discussed for the flavour-kt algorithm. Notice that Ω_{ij} is the same than ΔR_{ij} in the small Δy_{ij} and $\Delta\phi_{ij}$ limit, but it differs for large rapidity differences.
Thus the new flavour-kt measure is

$$d_{ij}^{flav-kt} = \max[p_{ti}, p_{tj}]^{\alpha} \min[p_{ti}, p_{tj}]^{2-\alpha} \Omega_{ij}^2 \tag{32}$$

$$d_{iB}^{flav-kt} = \max[p_{ti}, p_{tB_\pm}(y_i)]^{\alpha} \min[p_{ti}, p_{tB_\pm}(y_i)]^{2-\alpha} \tag{33}$$

where

$$p_{tB_\pm}(y) = \sum_i p_{ti} \Big[\Theta(\mp\Delta y_i) e^{\pm\Delta y_i} + \Theta(\pm\Delta y_i) \Big] \tag{34}$$

The original flavour-AKT algorithm (CMP) suffer also from a IHC×IHC problem, with the same configuration what we discussed for flavour-kt. Also this is solved by changing the metric using the Ω_{ij} angular distance.

2.7.5 Flavour dressing

Updated version of the algorithm does not make use of the flavour cluster, but it works directly with particles. Updated version is safe all order (tested up to α_S^6). The kinematics is exactly the same than AKT [GHS23].
We can divide the algorithm schematically in few blocks

1. flavour agnostic jets. The jet clustering step done independently of the flavour, typically AKT. We have a collection of jets $\{j_1, \ldots, j_m\}$ out of the collection of particles of the event $\{p_1, \ldots, p_n\}$.
2. Association criterion: We introduce tag_k set (initialized as empty) that will gives us the flavour information of the jet j_k. This is the criteria we use to assign the flavour of a particle p_i to a tag_k set.
3. Accumulation criterion: since many particles can populate the same tag_k set, we need a criterion to get a single flavour label from tag_k to assign to the jet j_k.

Here in the following we describe the accumulation criterion:

1. Initialise empty set tag_k for any jet j_k to accumulate all flavoured particles assigned to it.
2. Build a set $\mathcal{D}$ of all the following distance measure
   - inter-particle distance: d_{p_i,p_j} for any pair of flavoured particles or any pair of particles if associated to the same jet.
   - if particle p_j is associated to jet j_k, then add d_{p_j,j_k}.
   - At hadron colliders, add the beam distance $d_{p_j,B_\pm}$ if p_j is not associated to any jet.
3. Select the smallest distance iteratively. There are three possibilities:
   - d_{p_i,p_j} is the smallest. Merge the two particles into a new $p_{(i+j)}$ particle carrying the sum of the 4-momenta and flavour. Remove elements of $\mathcal{D}$ where p_i or p_j are involved. Add new distances for $p_{(i+j)}$.
   - d_{p_i,j_k} is the smallest. Assign the particle to the jet doing $tag_k \to tag_k \cup \{p_i\}$. Remove all the element of $\mathcal{D}$ involving p_i.
   - $d_{p_i,B_\pm}$ is the smallest. Discard particle p_i and remove elements of $\mathcal{D}$ involving p_i.
4. Assign a flavour to j_k according to the content of tag_k and the association criterion.

The particle-particle and the particle-jet distance is the same than eq. 32 for flavour-kt. The beam distance is the same than eq. 33 for flavour-kt.

Note: old version of Flav dressing has been used for Z+c-jet (still with flavour clusters), but it was ok because FHC^2 problem arise only for processes with a multiplicity at Born level higher that 4. Z+c-jet has Born multiplicity 3.
W+c-jet used the updated version of the algorithm.

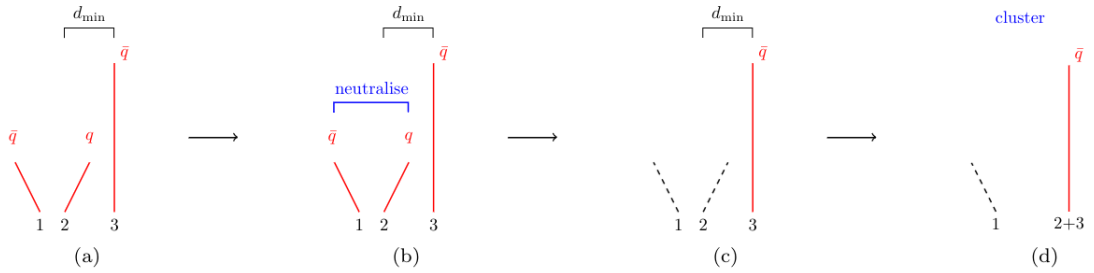

Figure 8: llustration of the flavour-neutralisation approach. The event displayed here (a) has the property that there is a soft $\bar{q}q$ pair (particles 1 and 2), and a hard $\bar{q}$ (particle 3) with $p_{t1} \sim p_{t2} \ll p_{t3}$. Additionally, we have all ΔR distances of order one, but with the constraint that $\Delta R_{23} < R$, while $\Delta R_{12} > R$, so that within the anti-kt algorithm, 2 and 3 cluster into one jet, while 1 would form a separate soft jet. In (b), just before the 2 + 3 clustering, the flavour of 1 is used to neutralise the flavour of 2, which results in the intermediate stage shown in (c), where particles 1 and 2 have lost their flavour (as represented by the black dashed lines). Finally, in (d) the (now) flavourless pseudojet 2 is clustered with 3 into a pseudojet 2+3 with the $\bar{q}$ flavour of just particle 3.

2.7.6 Interleaved Flavour Neutralization

This algorithms is safe to all order and has been numerically tested up to $\mathcal{O}(\alpha_S^6)$. The kinematics is exactly AKT [CGH+23].

- The clustering distance can be AKT.
- The neutralization distance is chosen as the flavour-kt distance eq. 32.

An explicit example of the IFN algorithm at work is given by Fig. 8.
They also introduced a numerical framework to test the algorithms up to α_S^6 by adding IRC radiation to the Born and looking to the flavour if it was affected by this. To do this is useful to classify all the possible IRC insertion you can do order by order as we discussed in the previous sections.

2.7.7 Status of the fjcontrib and comparison

During Les Houches 2023 all the groups started implementing their algorithms in the form of a fjcontrib. Still work in progress, almost done.
Some comparison between the algorithms are given in Fig. 9 and 10. Comments on Fig. 9:

- similar plots have been done for flavour dressing and IFN and removes more jets in the fake region.
- CMP going to the right is also moving far from AKT.

Comments on Fig. 10:

- on the x-axis we have the number of particle of the event, on the y-axis the time required for clustering
- The IFN and SDF perform the best
- The right plot is done with a modified rate of $g \to b\bar{b}$. For many flavoured particles SDF performs the best.

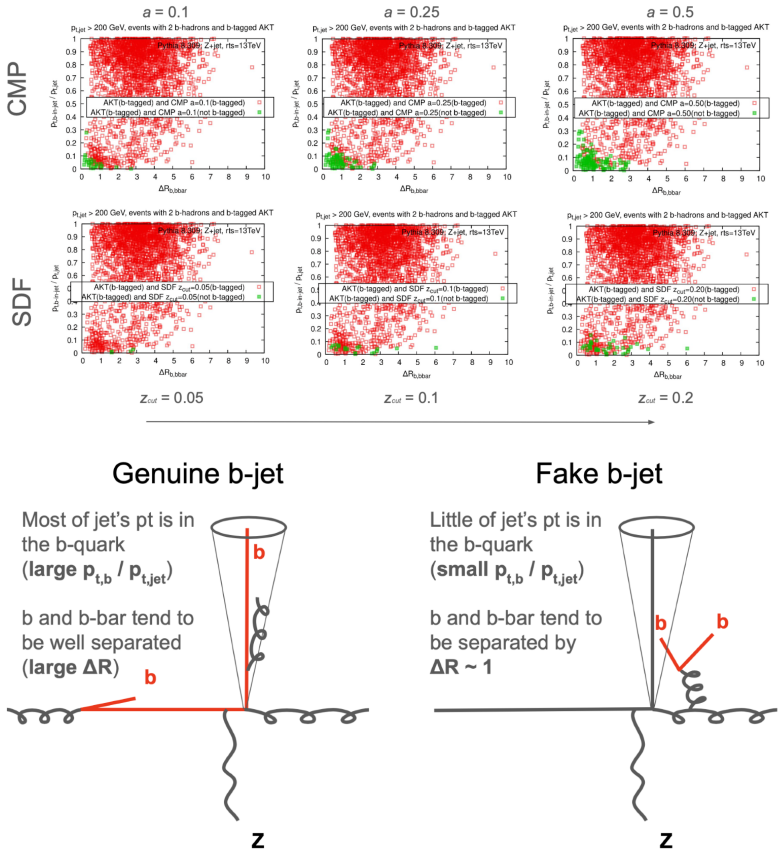

Figure 9: Comparison between CMP and SDF in the $(\Delta R_{b\bar{b}}, p_{T,b}/p_{T,jet})$ plane. Changing the parameter affects the flavour assignament.

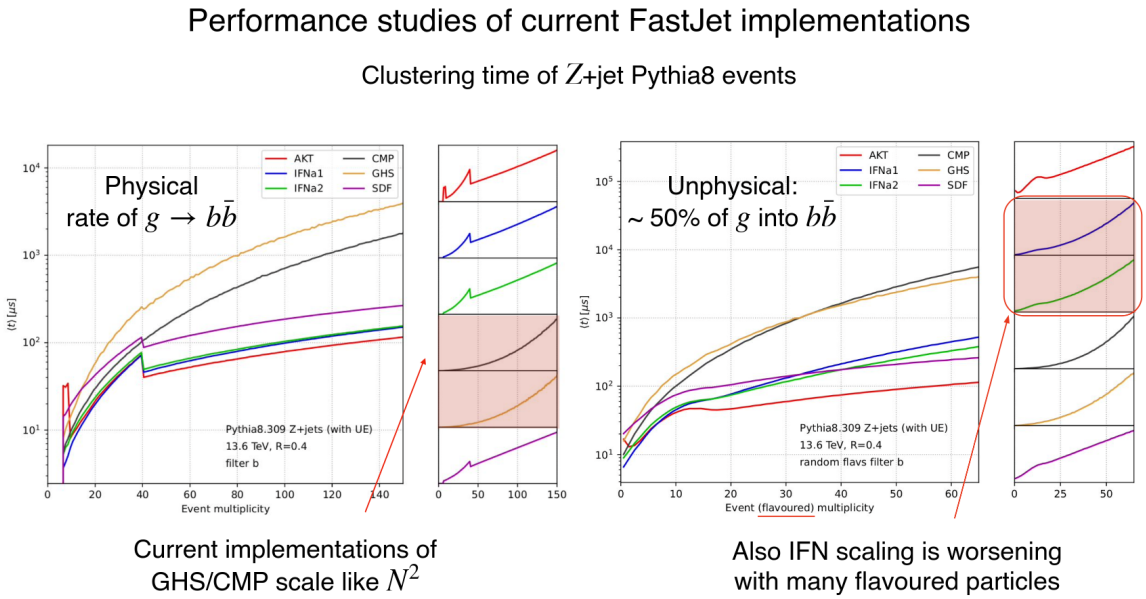

Figure 10: Comparison between different fjcontrib current implementations performances of the algorithms.

W^++c-jet	OS LO	OS NLO	SS NLO	OS NNLO	SS NNLO
$c(\bar{c})s(\bar{s})$	0.0	−0.1191(9)	0.4752(4)	−0.13(2)	0.838(1)
$c(\bar{c})c(\bar{c})$	0.0	0.2151(3)	0.2047(3)	0.3316(5)	0.3246(6)
$c(\bar{c})q(\bar{q})$	0.0	1.948(3)	1.988(4)	2.945(6)	3.038(6)
$s(\bar{s})q(\bar{q})$	0.0	−0.649(9)	0.0673(1)	−1.9(3)	0.1157(3)
$s(\bar{s})s(\bar{s})$	0.0	−0.258(1)	0.0	−0.55(5)	0.0
$q(\bar{q})q(\bar{q})$	0.0	1.431(2)	1.409(2)	2.35(2)	2.423(6)
$gq(\bar{q})$	5.8299(7)	8.257(2)	0.0	10.1(4)	0.508(4)
$gs(\bar{s})$	85.51(1)	121.04(3)	0.0	126.3(6)	−0.0430(4)
$gc(\bar{c})$	0.0	0.0	0.0	0.02(2)	−0.0293(7)
gg	0.0	−6.34(1)	0.0	−13.62(6)	0.0
Total	91.34(1)	125.51(4)	4.146(4)	125.9(7)	7.17(1)

W^-+c-jet	OS LO	OS NLO	SS NLO	OS NNLO	SS NNLO
$c(\bar{c})s(\bar{s})$	0.0	−0.1225(3)	0.4852(2)	−0.05(2)	0.842(3)
$c(\bar{c})c(\bar{c})$	0.0	0.2158(1)	0.2062(2)	0.360(2)	0.351(1)
$c(\bar{c})q(\bar{q})$	0.0	1.2392(3)	1.3132(4)	1.958(4)	2.088(4)
$s(\bar{s})q(\bar{q})$	0.0	−0.651(3)	0.03134(1)	−1.1(2)	0.0537(2)
$s(\bar{s})s(\bar{s})$	0.0	−0.2549(3)	0.0	−0.42(3)	0.0
$q(\bar{q})q(\bar{q})$	0.0	1.0314(7)	0.9838(4)	1.73(2)	1.676(6)
$gq(\bar{q})$	8.9255(6)	12.700(1)	0.0	12.7(2)	0.405(3)
$gs(\bar{s})$	86.857(4)	123.002(8)	0.0	128.9(3)	−0.0353(6)
$gc(\bar{c})$	0.0	0.0	0.0	−0.14(2)	−0.057(2)
gg	0.0	−6.355(3)	0.0	−8.31(1)	0.0
Total	95.782(5)	130.806(1)	3.020(1)	135.6(5)	5.324 (9)

Figure 11: Fiducial cross section partonic channel breakdown for W+c-jet

2.8 Results for W+c-jet @ NNLO

Here we presents NNLO results obtained in the antenna framework for W+c-jet production at the LHC [GDRGG+24]. These predictions are in agreement with the one used [CMS24] using sector decomposition (Czakon et al.). We report the channel breakdown for the total cross section in the **slides**.
Comments on Fig. 11:

- $gs(\bar{s})$ channel by far the dominant in OS events: $\sim 90\%$
- $gq(\bar{q})$ channel second largest contributions to OS events: $6 - 10\%$.
- The latter contribution is slightly larger for W^-. This explained by the presence of the d PDF in W^-, in contrast to the $\bar{d}$ PDF in the W^+.
- Third largest contribution comes from gg channel and is negative. Sometimes it compensates the gq contribution.
- The other channels contribute at the most at few-percent level.
- $c(\bar{c})c(\bar{c})$, $c(\bar{c})q(\bar{q})$ and $q(\bar{q})q(\bar{q})$ channels are all numerically very similar. In the OS-SS we enhance the (anti)strange PDF by removing channels with quarks of other flavours. Channels with a gluon PDF still remains.

3 Experimental methods: jet tagging and hadron reconstruction (Miha)

4 Identified hadrons in the antenna framework: $V+h$ (Simone)

We extend the result for W+jet to W+hadron. We have to introduce identified particles, since these final state partons have to be attached to the fragmentation function.
Antennae with fragmentation first presented in [GS22].
The fully differential cross section can be written as

$$d\sigma^H = \sum_p \int d\eta D_p^H(\eta, \mu_a^2) d\hat{\sigma}_p(\eta, \mu_a^2) \tag{35}$$

where the index p runs over all the possible partons of the process.

4.1 Real level subtraction

Now we have two roles for the partons to be considered at the same time: the identified parton p which is going to fragment into H, and the unresolved parton j. For this reason, we divide the subtraction terms as it follow

$$d\hat{\sigma}_p^S = \sum_j d\hat{\sigma}_{p,j}^S \tag{36}$$

$$d\hat{\sigma}_{p,j}^S = d\hat{\sigma}_{p,j}^{S,non-id.p} + d\hat{\sigma}_{p,j}^{S,id.p} \tag{37}$$

- where in the first term we have configuration in which p and j are not colour connected $\rightarrow$ standard NLO subtraction terms with FF kinematics
- in the latter, j is connected to p and a hard parton k and we have to introduce the following subtraction term

$$d\hat{\sigma}_{p,j}^{S,id.p} = \mathcal{N}_R d\Phi_{n+1} \frac{1}{S_{n+1}} X03(p_p, p_j, p_k) |\mathcal{M}_n(p_1, \ldots, \tilde{K}, \tilde{p}_p, \ldots, p_{n+1})|^2 J(\ldots, \tilde{K}, \tilde{k}_p, \ldots; \eta z \tilde{k}_p) \tag{38}$$

Comments:

- the X30 is the standard 3-particle tree-level antenna function for FF kinematics
- The identified parton has a fraction $1/\eta$ of the momentum of the hadron, i.e. $K_H = \eta k_p$
- $\tilde{k}_p$ is the momentum of the recoiled identified parton in the reduced matrix element (after the $n+1 \rightarrow n$ mapping), and it is a fraction z of the original momentum, i.e. $\tilde{k}_p = k_p/z$
- The overall momentum fraction entering the JET function is the product of η and z
- The identified particle cannot go unresolved

4.2 From V+jet to V+hadron: example

We recall the real level subtraction term we introduced when talking about V+jet eq. 15. We can obtain the corresponding fragmenting subtraction term from V+hadron just removing the limits where the identified particle is going unresolved, which are not allowed. Obviously, we have different cases, one for each particle that can be identified. In this particular case, we have to distinguish between identified-q and identified-g.

- $q^{id.}(j \rightarrow 3)$ i//2 i$\rightarrow$0 i//j 1//i j//2
- $g^{id.}(i \rightarrow 3)$ i//2 i$\rightarrow$0 i//j 1//i j//2

For the identified quark:

$$d\hat{\sigma}^{S,\text{Z\_hfrag}} = \text{qgD30II}(1,i,2) * \text{B1g0Z}([1],[2],3,Z) * \text{JET11}(hfrag(3)) \\ +\text{gd30IF\_q}(3,i,2) * \text{B1g0Z}(1,[2],[3,i],Z) * \text{JET11}(hfrag([3])) \tag{39}$$

For the identified gluon:

$$d\hat{\sigma}^{S,\text{Z\_hfrag}} = -\text{qgA30II}(1,2,i) * \text{B1g0Z}([1],3,[2],Z) * \text{JET11}(hfrag(3)) \\ +\text{gd303IF\_g}(i,3,2) * \text{B1g0Z}(1,[2],[3,i],Z) * \text{JET11}(hfrag([3])) \tag{40}$$

where in both cases the first line is non-id. (p non colour connected with the unresolved parton) and in the second line is id.-type (colour connected), as can be seen by the mapping within the JET function.

4.3 Virtual level subtraction

We start with IF phase space factorization and we make explicit the integration over z

$$d\Phi_{n+1} = d\Phi_n \frac{dx}{x} dz \frac{Q^2}{2\pi} d\Phi_2(k_p^{id.p}, k_j; p_1, q)\delta\Big(z - \frac{s_{1p}}{s_{1p}+s_{1j}}\Big) \tag{41}$$

Then the integrated antenna is defined as

$$\mathcal{X}_3^{0,id.p}(x,z) = \frac{1}{C(\epsilon)} \int d\Phi_2(k_p^{id.p}, k_j; p_1, q) \frac{Q^2}{2\pi} X_3^0(p_1,; k_k, k_p^{id.p})\delta\Big(z - \frac{s_{1p}}{s-1p+s_{1j}}\Big) \tag{42}$$

Then the integrated subtraction term is given by

$$\int_1 d\hat{\sigma}_{p,j}^{S,id.p} = \mathcal{N} \int \frac{dx}{x} dz d\Phi_n \frac{1}{S_n} \mathcal{X}_3^{0,id.p}(x,z) |\mathcal{M}_n^0| J(\ldots, \eta z \tilde{k}_p) \tag{43}$$

and combined with the mass factorization piece to construct the full virtual level subtraction term.

4.4 Results for W+D@NLO

We show prediction obtained within the NNLOJET framework for fragmenting hadrons with different FF in Fig. 12 from [CGDRH+24].
Comments for rapidity:

- All the prediction are OS-SS
- NLO correction about 40% and quite flat across the entire $|\eta_\ell|$ spectrum.
- CNO describes better the central region. KKKS08 describes better the forward region.
- No big differences for W^+ and W^-.

Comments for p_T:

- again NLO is 40%.
- CNO compatible in the full spectrum, while KKKS08 deviates in the high-p_T region.

5 Bonus: heavy-quark initiated jet substructure (Simone)

We started studying the Jet Substructure of Heavy-Quark (HQ) initiated jets. An example can be found in [CGM24]. Differences in massive calculations:

- threshold in the running of the coupling (horizontal lines)
- dead-cone effect (vertical dotted line)
- massive contribution to the splitting functions
- the observables might change to the massless case

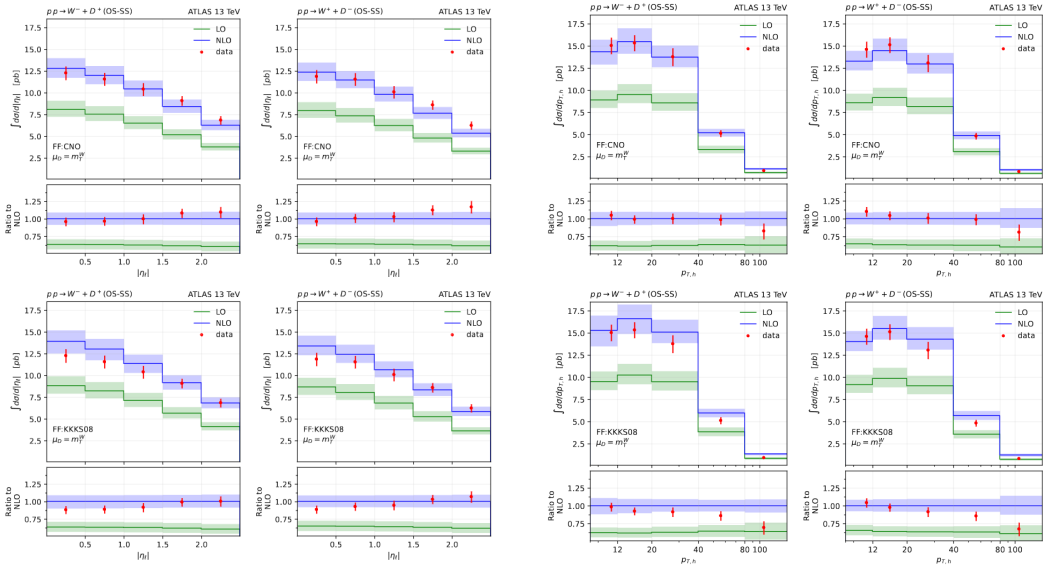

Figure 12: Comparison of LO (green) and NLO (blue) predictions with data (red). The top plots have been produced using the CNO fragmentation function, while the bottom plots using KKKS08. The left and right columns show the results for W^- and W^+, respectively. All include the OS-SS prescription

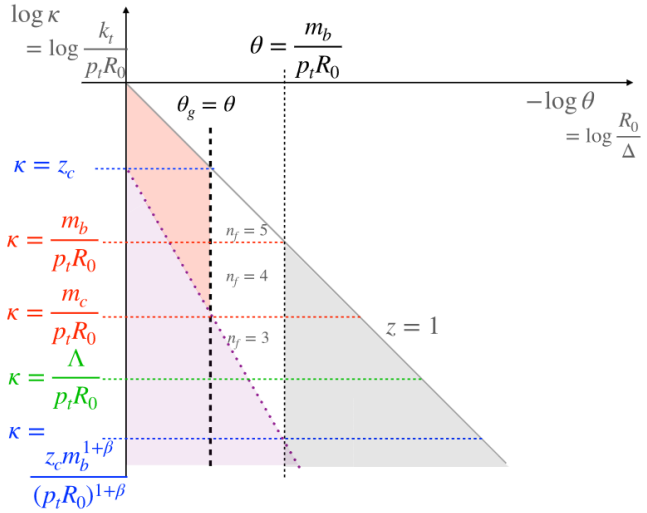

Figure 13: Lund plane for the θ_g observables with all the masses thresholds.

References

[ATL23] ATLAS Collaboration. Measurement of the production of a W boson in association with a charmed hadron in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector. *Phys. Rev. D*, 108:032012, 2023.

[ATL24] 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. 3 2024.

[BSZ06] Andrea Banfi, Gavin P. Salam, and Giulia Zanderighi. Infrared safe definition of jet flavor. *Eur. Phys. J. C*, 47:113–124, 2006.

[CGDRH+24] Simone Caletti, Aude Gehrmann-De Ridder, Alexander Huss, Adrian Rodriguez Garcia, and Giovanni Stagnitto. QCD predictions for vector boson plus hadron production at the LHC. 5 2024.

[CGH+23] Fabrizio Caola, Radoslaw Grabarczyk, Maxwell L. Hutt, Gavin P. Salam, Ludovic Scyboz, and Jesse Thaler. Flavored jets with exact anti-kt kinematics and tests of infrared and collinear safety. *Phys. Rev. D*, 108(9):094010, 2023.

[CGM24] Simone Caletti, Andrea Ghira, and Simone Marzani. On heavy-flavour jets with Soft Drop. *Eur. Phys. J. C*, 84(2):212, 2024.

[CLMR22] Simone Caletti, Andrew J. Larkoski, Simone Marzani, and Daniel Reichelt. Practical jet flavour through NNLO. *Eur. Phys. J. C*, 82(7):632, 2022.

[CMP23] Michal Czakon, Alexander Mitov, and Rene Poncelet. Infrared-safe flavoured anti-k$_T$ jets. *JHEP*, 04:138, 2023.

[CMS20] CMS Collaboration. Measurement of the associated production of a Z boson with charm or bottom quark jets in proton-proton collisions at $\sqrt{s}$=13 TeV. *Phys. Rev. D*, 102(3):032007, 2020.

[CMS21] CMS Collaboration. Measurement of differential cross sections for Z bosons produced in association with charm jets in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 04:109, 2021.

[CMS22] CMS Collaboration. Measurements of the associated production of a W boson and a charm quark in proton–proton collisions at $\sqrt{s} = 8\,\text{TeV}$. *Eur. Phys. J. C*, 82(12):1094, 2022.

[CMS24] CMS Collaboration. Measurement of the production cross section for a W boson in association with a charm quark in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 84:27, 2024.

[GDRGG+24] A. Gehrmann-De Ridder, T. Gehrmann, E. W. N. Glover, A. Huss, A. Rodriguez Garcia, and G. Stagnitto. Precise QCD predictions for W-boson production in association with a charm jet. *Eur. Phys. J. C*, 84(4):361, 2024.

[GHS23] Rhorry Gauld, Alexander Huss, and Giovanni Stagnitto. Flavor Identification of Reconstructed Hadronic Jets. *Phys. Rev. Lett.*, 130(16):161901, 2023. [Erratum: Phys.Rev.Lett. 132, 159901 (2024)].

[GS22] Thomas Gehrmann and Giovanni Stagnitto. Antenna subtraction at NNLO with identified hadrons. *JHEP*, 10:136, 2022.