Theory aspects of jet flavour definition

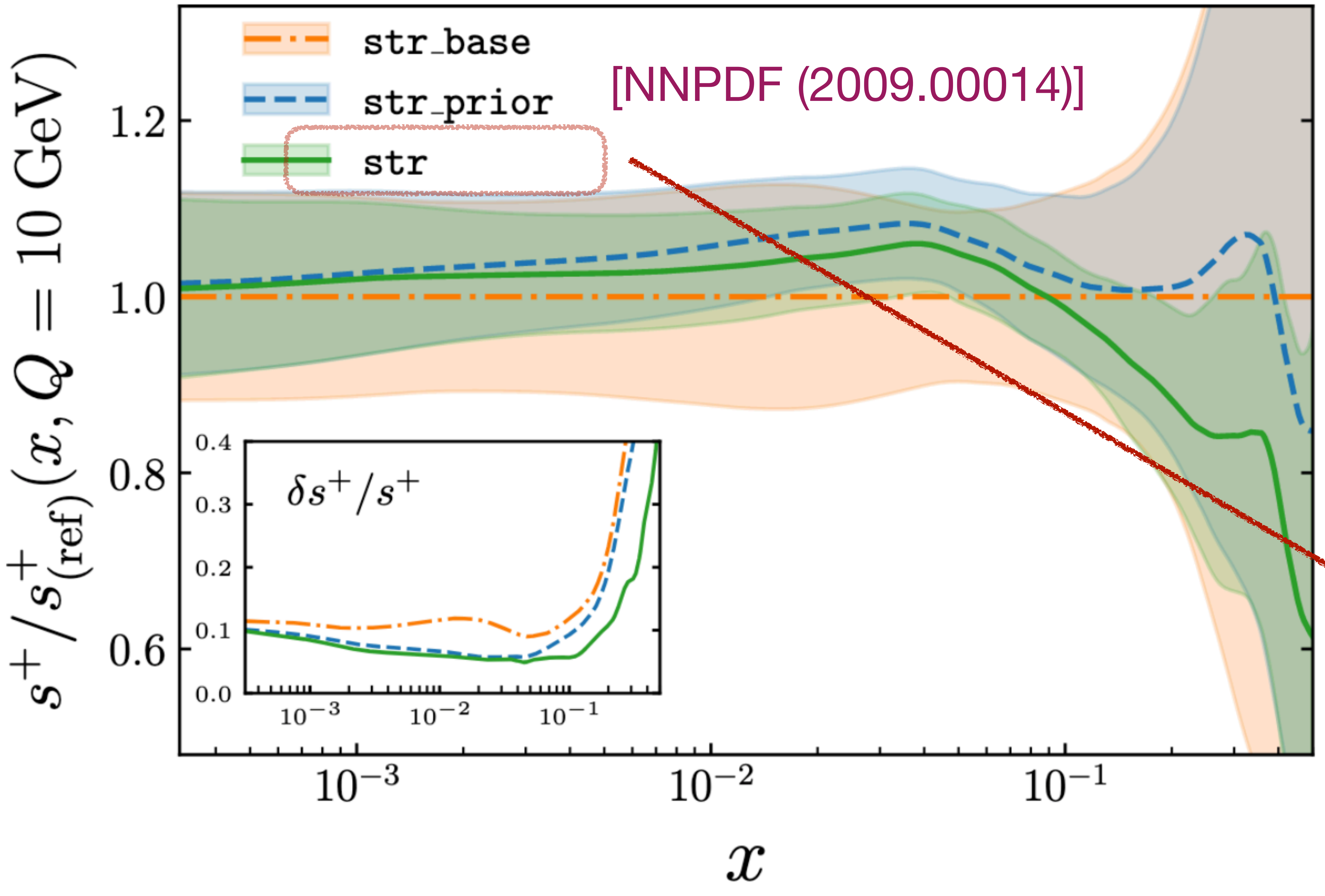
Giovanni Stagnitto

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

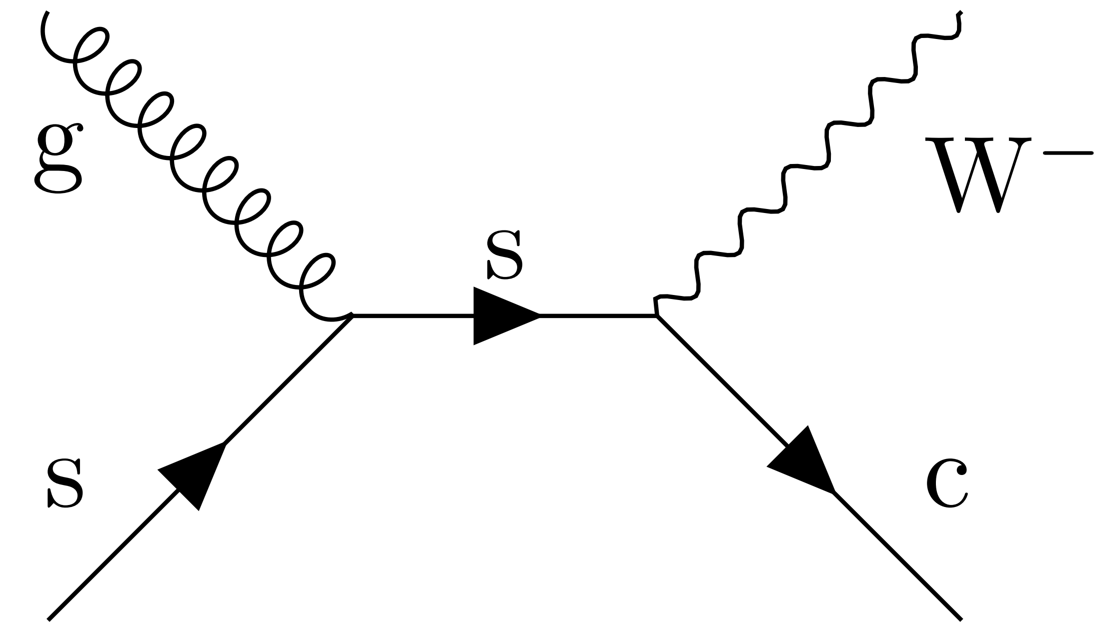
INFN

LHC EW WG General Meeting, CERN, 10-12.07.2024

Why flavoured jets? An example



W+c-jet unique probe into strange PDF



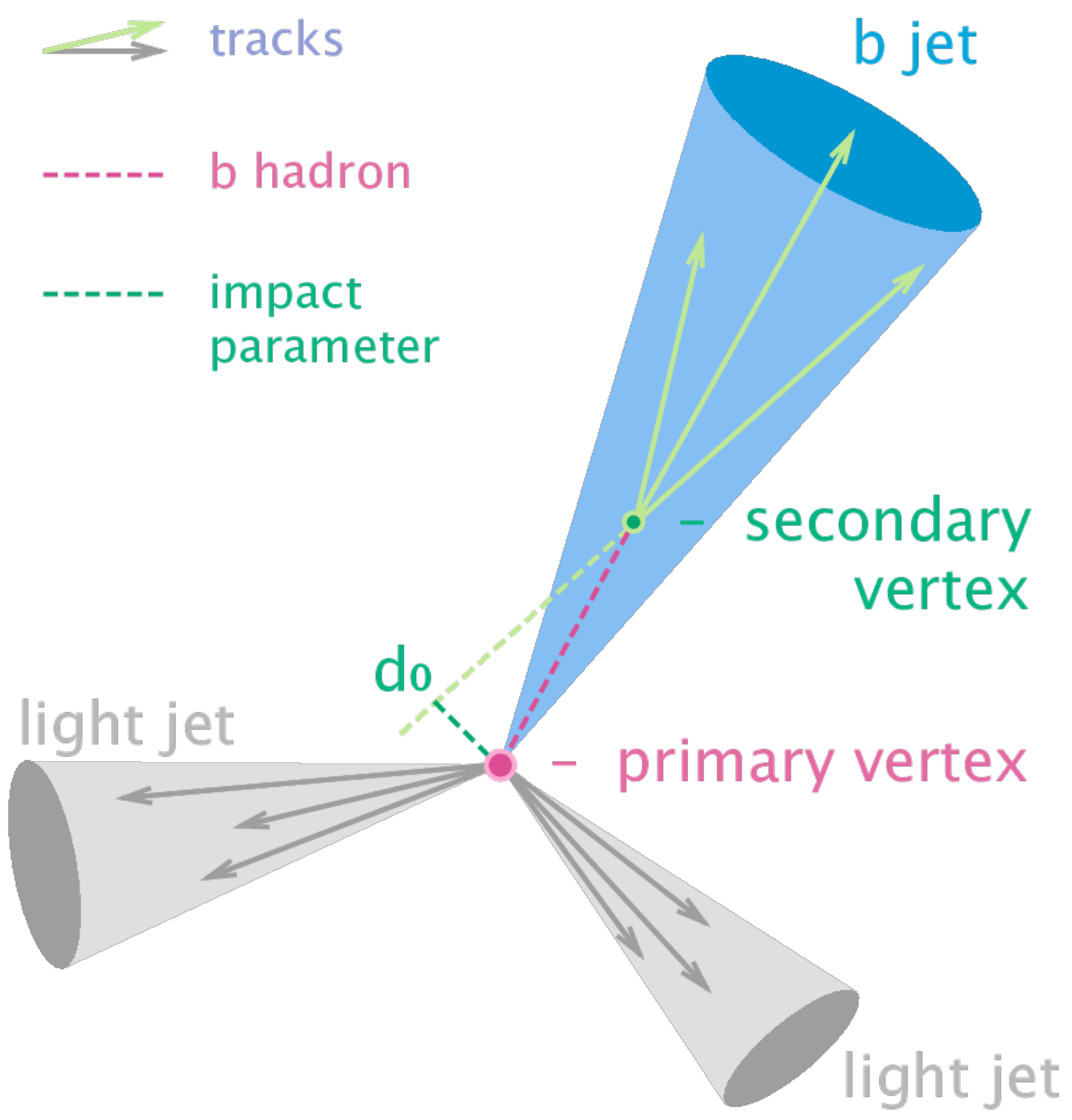
contain [ATLAS (1402.6263)] and [CMS (1310.1138)] 7 TeV data

… but **flavoured jets appear everywhere**: top, Higgs, new physics searches, … useful to pinpoint specific scattering processes and reject backgrounds

(Common) experimental definition of flavoured jet

"An (anti-k_t) jet is flavoured if it contains at least one heavy hadron within $\Delta R < R$ with $p_T > p_{T,\mathrm{cut}}$"

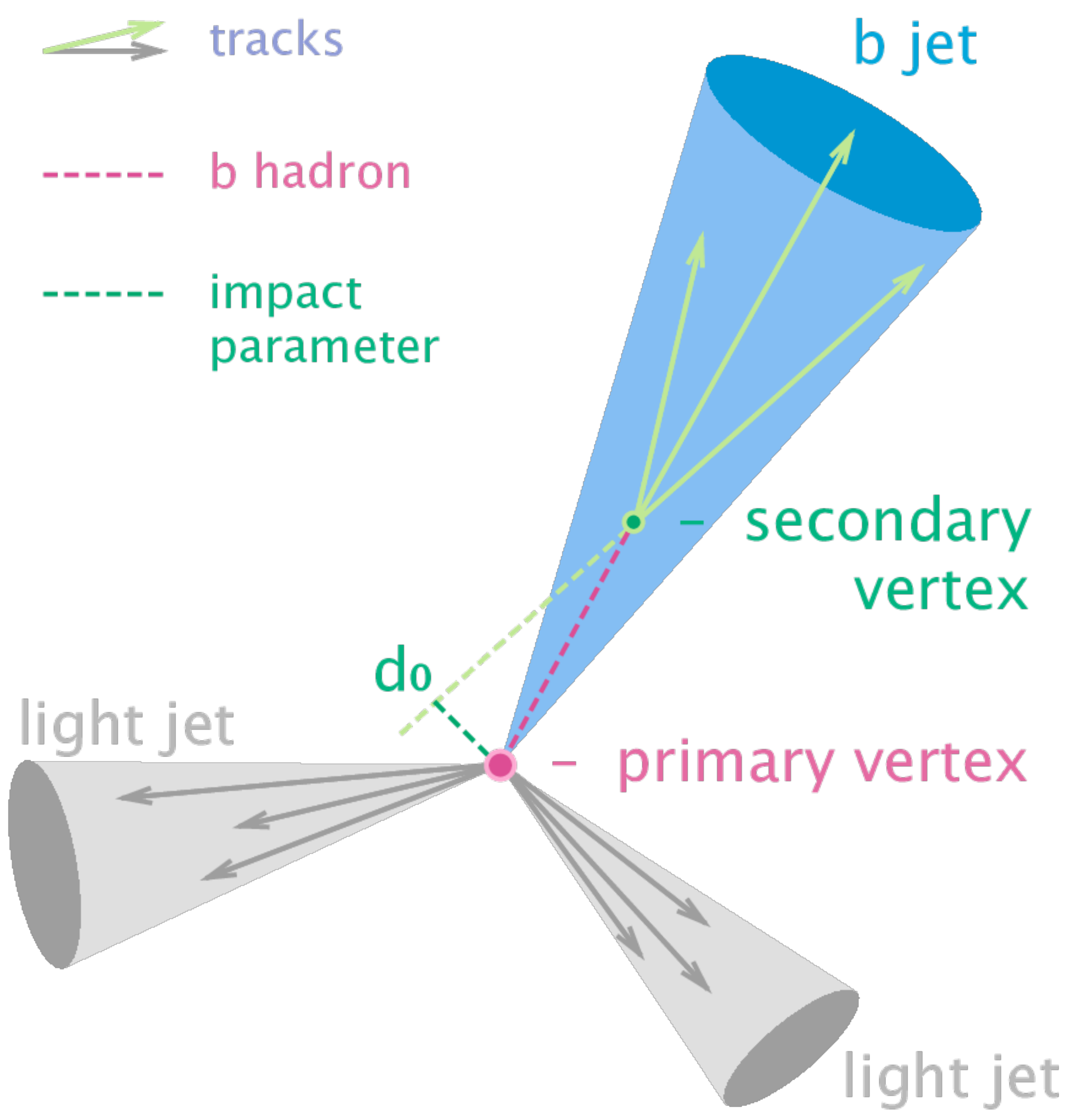
This definition is adopted as "true" label in MC samples. These samples are then used to train ML architectures ("high-level taggers"), which exploit low-level variables as inputs.



(Common) experimental definition of flavoured jet

"An (anti-k_t) jet is flavoured if it contains at least one heavy hadron within $\Delta R < R$ with $p_T > p_{T,\mathrm{cut}}$"

This definition is adopted
as "true" label in MC samples.
These samples are then used to train ML
architectures ("high-level taggers"),
which exploit low-level variables as inputs.



This definition is both **soft and collinear (IRC) unsafe**
(in massless perturbative QCD calculations)
i.e. arbitrary soft and/or collinear emissions alter the
flavour of jets i.e. infinities in calculations!

How to "fix" it?

"An (anti-k_t) jet is flavoured if it contains at least one heavy hadron within $\Delta R < R$ with $p_T > p_{T,\rm cut}$"

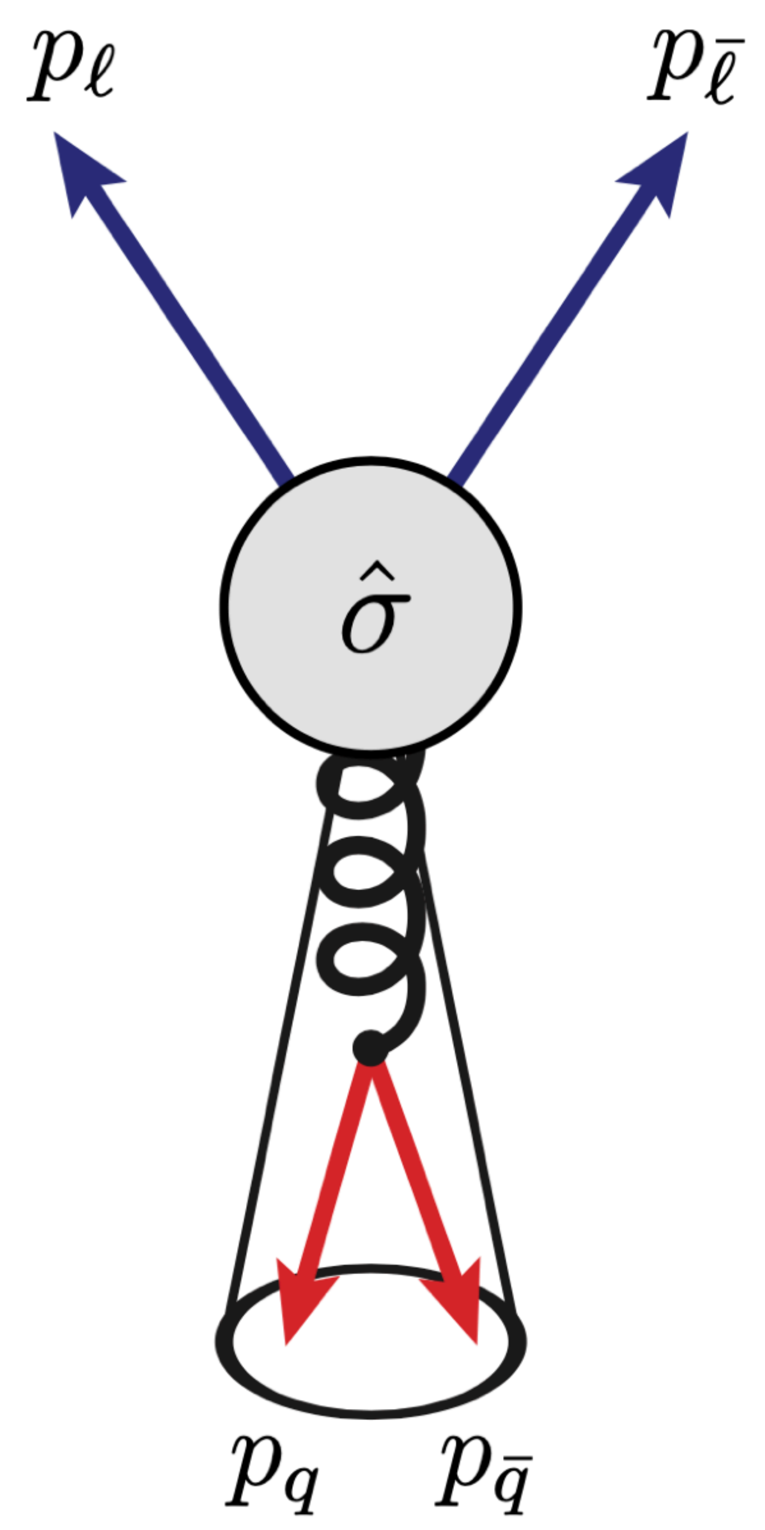
Problem:

$g \to q\bar{q}$ **is always flavoured even in the collinear limit**

Solution:

If no charge information is available, apply an *even-tag veto*

(e.g. jets with pairs of bs in it are flavourless)



How to "fix" it?

"An (anti-k_t) jet is flavoured if it contains at least one heavy hadron within $\Delta R < R$ with $p_T > p_{T,\text{cut}}$"

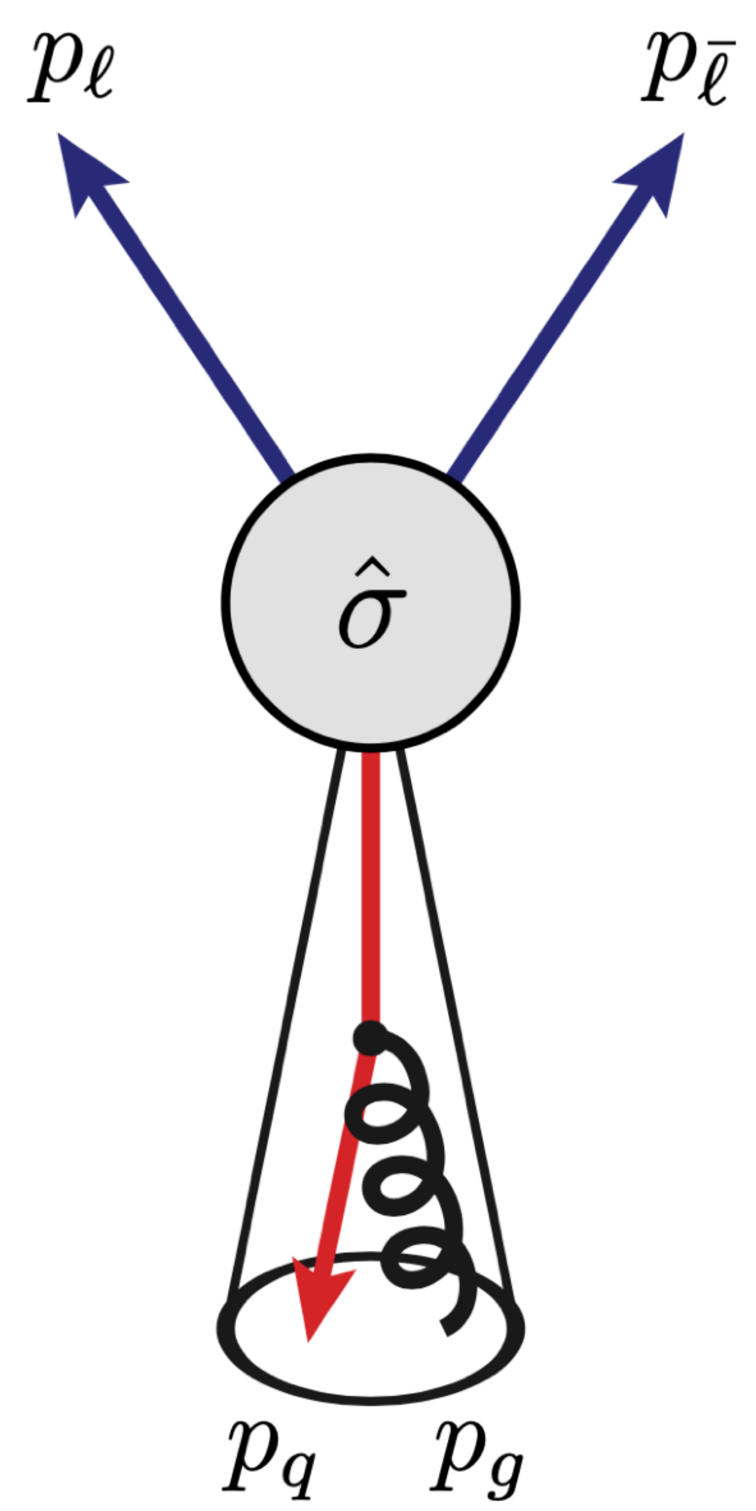
Problem:

$q \to qg$ **collinear with a hard gluon leads to a flavourless jet**

Solution:

Remove $p_{T,\text{cut}}$ requirement

(with $p_{T,\text{cut}}$, it would require non-perturbative fragmentation function)



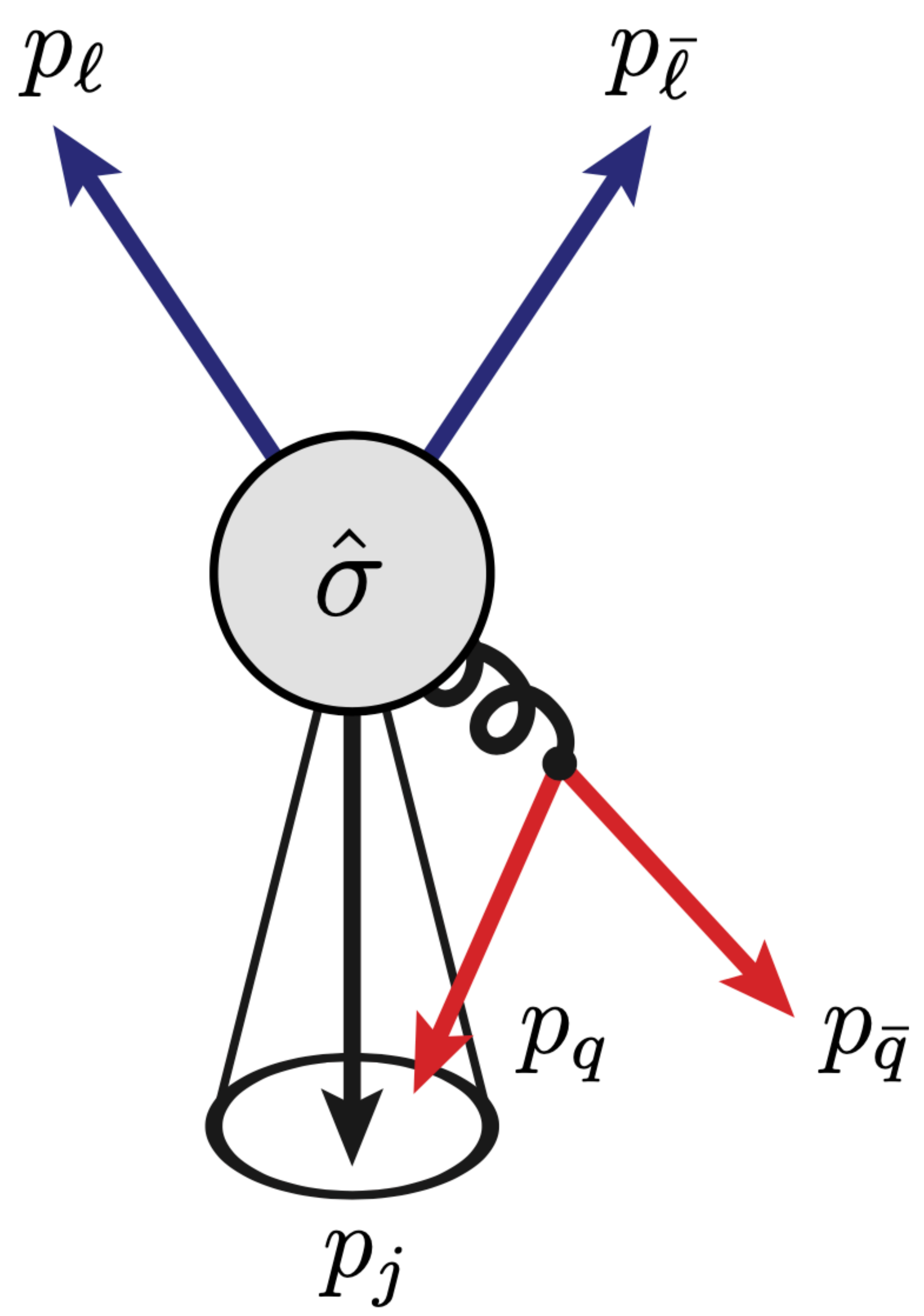
How to “fix” it?

“An (anti-k_t) jet is flavoured if it contains at least one heavy hadron within $\Delta R < R$ with $p_T > p_{T,\mathrm{cut}}$”

Problem:

Soft large-angle $g \to b\bar{b}$

polluting the flavour of other jets



Solution:

No simple solution

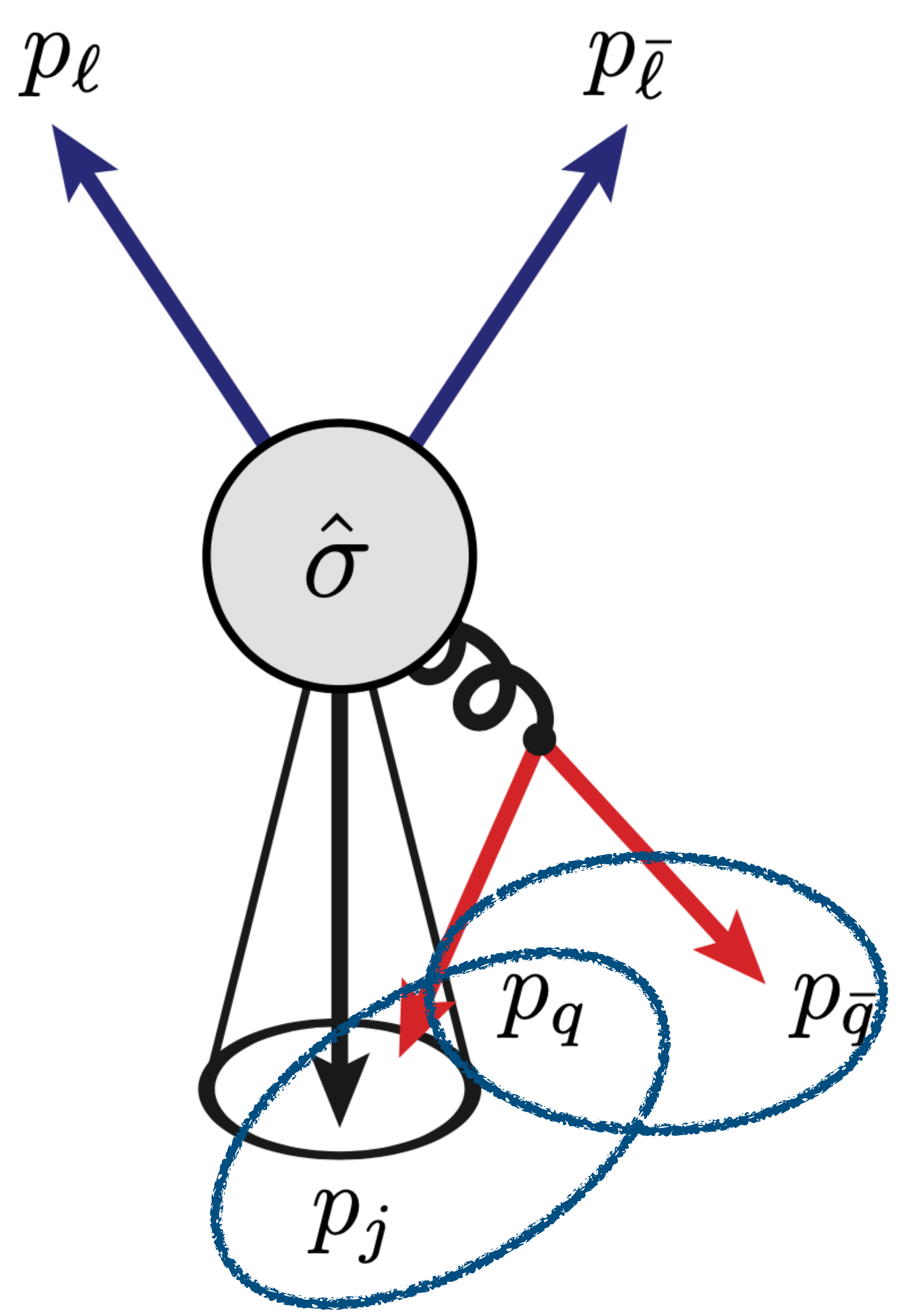
A new algorithm is needed

First IRC safe definition: the flavour-k_t algorithm (2006)

[Banfi, Salam, Zanderighi (hep-ph/0601139)]

Flavour IRC safety obtained with a flavour-aware clustering distance:

$$d_{ij}^{(F,\alpha)} = \frac{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}{R^2} \times \begin{cases} \max(k_{ti}, k_{tj})^{\alpha} \min(k_{ti}, k_{tj})^{2-\alpha}\,, & \text{softer of } i, j \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tj}^2)\,, & \text{softer of } i, j \text{ is flavourless} \end{cases}$$



d(hard flavourless quark, soft flavoured quark) = $\max(p_{T,j}, p_{T,q}) \gg 1$

d(soft flavoured quark, soft flavoured quark) = $\max(p_{T,q}, p_{T,\bar{q}}) \ll 1$

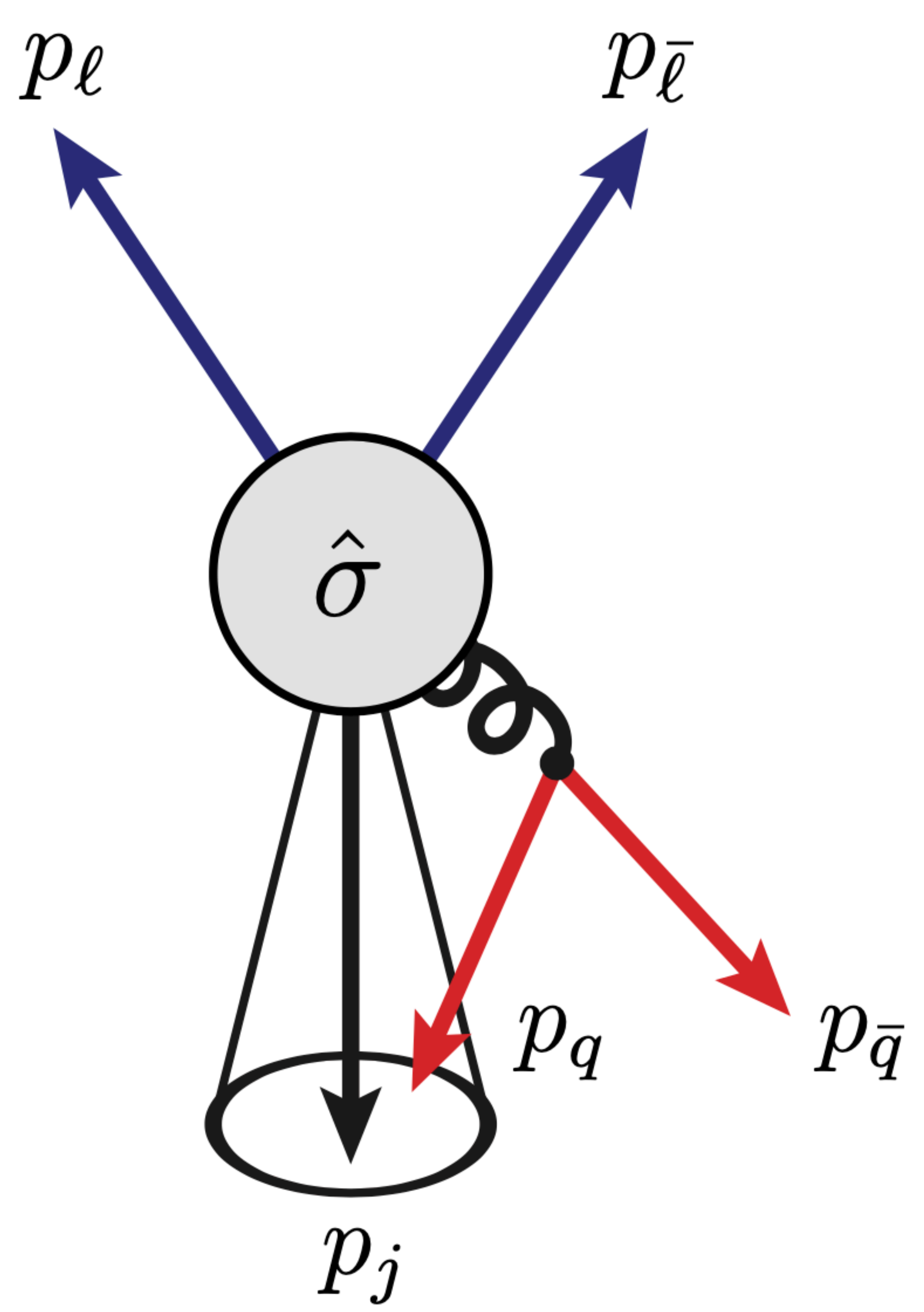
→ soft pairs of quarks recombined early in the clustering!

First IRC safe definition: the flavour-k_t algorithm (2006)

[Banfi, Salam, Zanderighi (hep-ph/0601139)]

Flavour IRC safety obtained with a flavour-aware clustering distance:

$$d_{ij}^{(F,\alpha)} = \frac{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}{R^2} \times \begin{cases} \max(k_{ti}, k_{tj})^{\alpha} \min(k_{ti}, k_{tj})^{2-\alpha}\,, & \text{softer of } i,j \text{ is flavoured,} \\ \min(k_{ti}^2, k_{tj}^2)\,, & \text{softer of } i,j \text{ is flavourless} \end{cases}$$



At the price of jets with different kinematics i.e. not anti-k_t jets

(anti-k_t was not even there in 2006!)

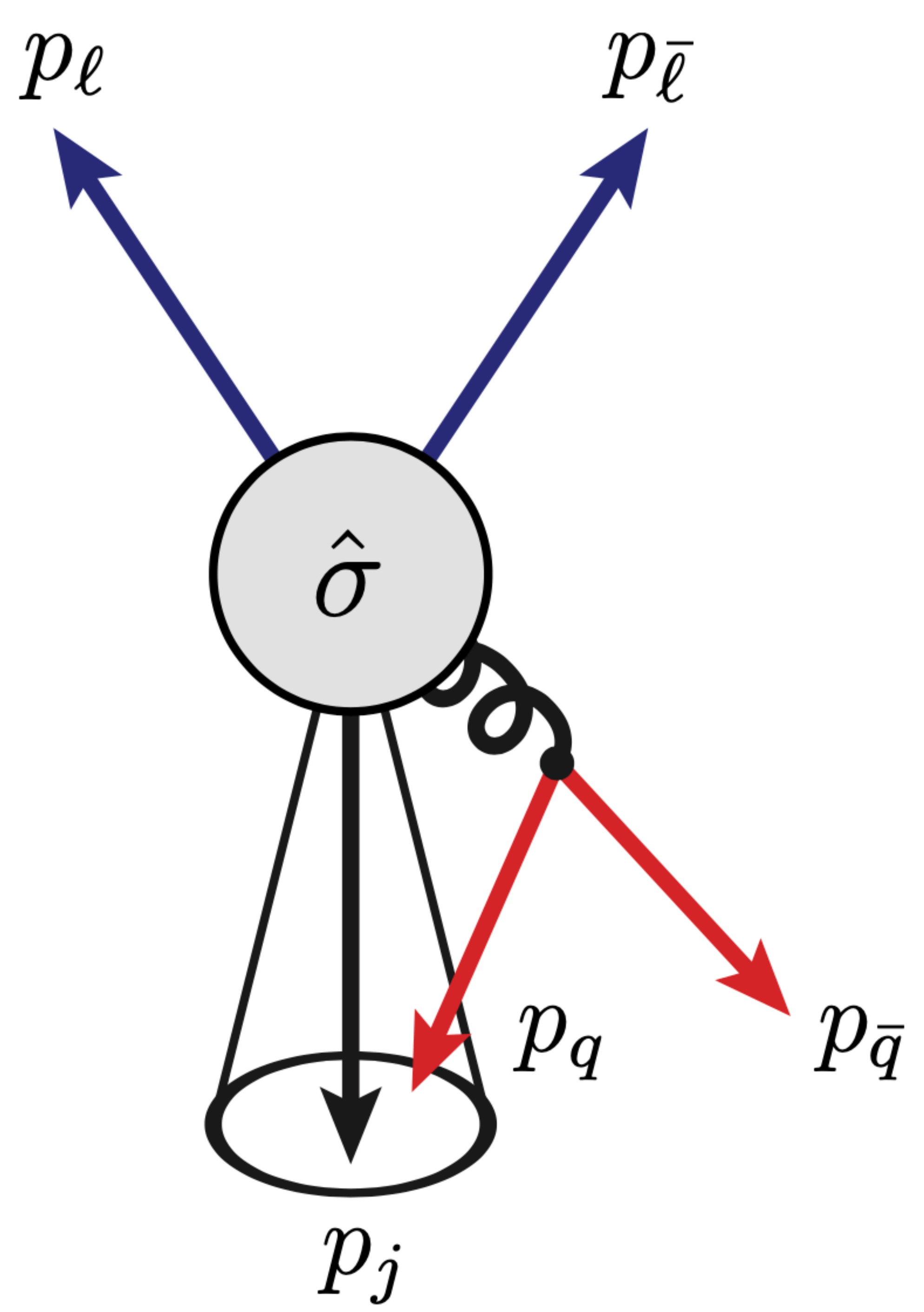
First IRC safe definition: the flavour-k_t algorithm (2006)

[Banfi, Salam, Zanderighi (hep-ph/0601139)]

Definition adopted in first NNLO QCD calculations of V + flavoured jet:

$Z + b$-jet [Gauld et al. (2005.03016)]

$W + c$-jet [Czakon et al. (2011.01011)]



Comparison with data made difficult by different jet kinematics

Availability of new NNLO calculations (2020-…)

calls for new algorithms

(Note that in V+jet the problematic configurations appear for the first time at NNLO, so up to NLO using plain anti-k_t is fine, and until recently there was then no real "need" for a true IRC safe anti-k_t-like flavoured jet algorithm)

The flavour revolution (2022-…)

4 new proposals, IRC safe to all orders (or up to high order) with exact (or close to exact) anti-k_t kinematics

[Caletti, Larkoski, Marzani, Reichelt (2205.01109)] **SDF**

[Czakon, Mitov, Poncelet (2205.11879)] **CMP**

[Gauld, Huss, Stagnitto (2208.11138)] **GHS**

[Caola, Grabarczyk, Hutt, Salam, Scyboz, Thaler (2306.07314)] **IFN**

I refer to the *Public LHCb meeting on jet flavor algorithms* (https://indico.cern.ch/event/1403404/) for talks presenting in details the different algorithms

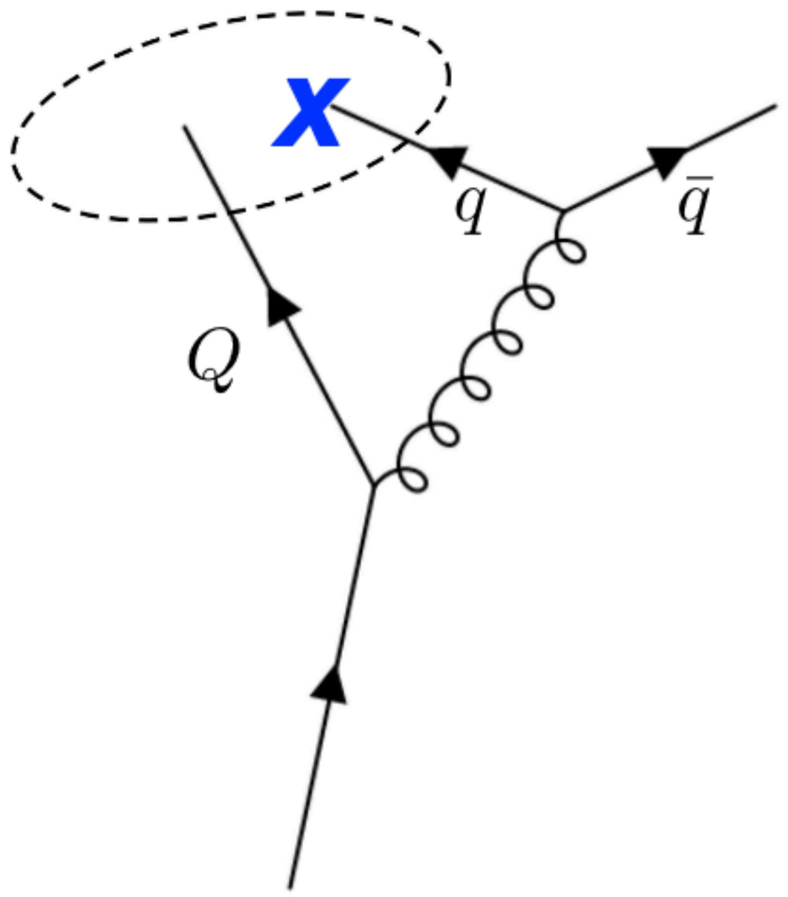
The flavour revolution (2022-…)

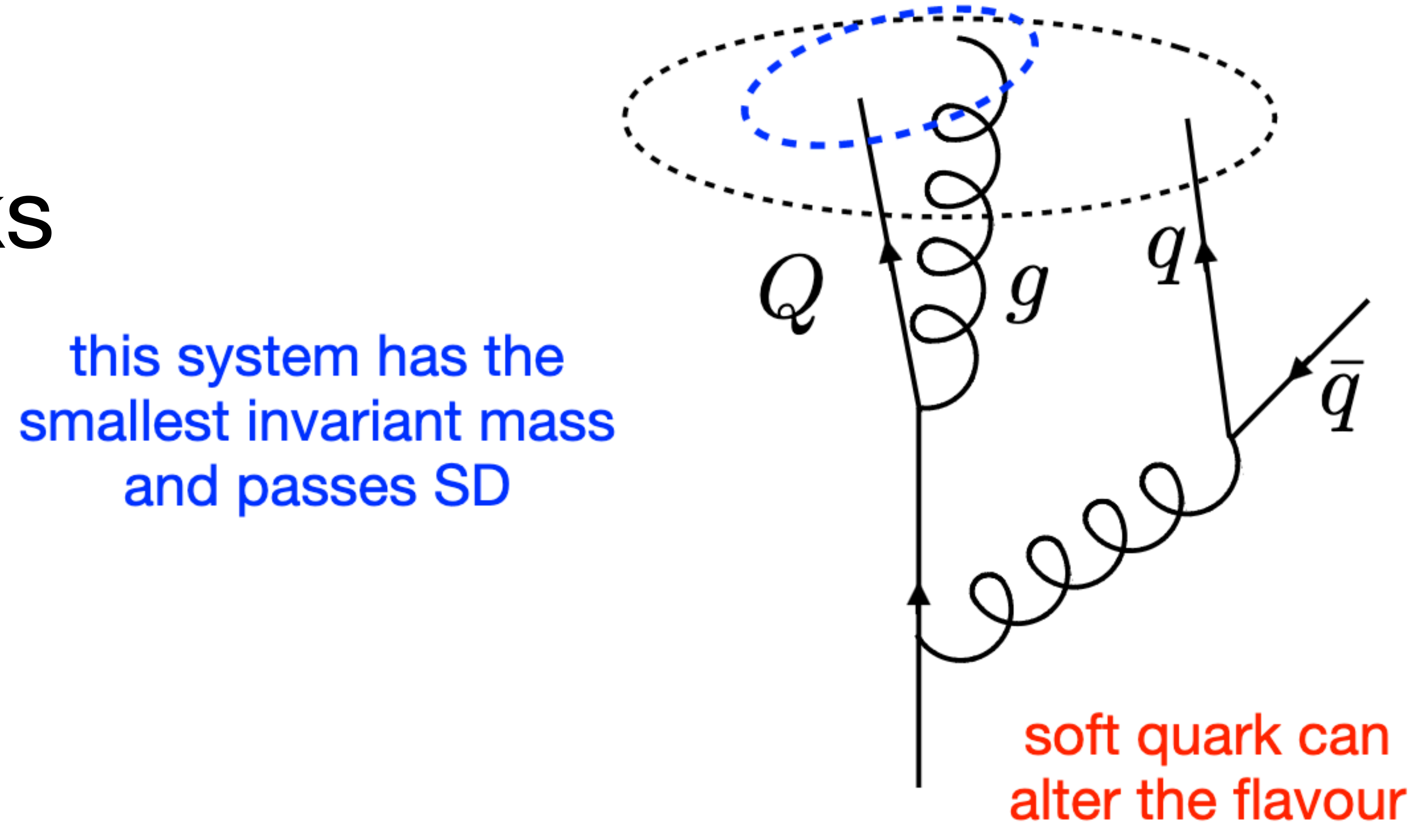
IRC safe up to N3LO

Exact anti-k_t kinematics

[Caletti, Larkoski, Marzani, Reichelt (2205.01109)] **SDF**

First recluster anti-k_t jet with JADE,
then use Soft Drop to remove soft quarks





Known to fail at N3LO

The flavour revolution (2022-…)

IRC safe to all orders

Close to exact anti-k_t kinematics

[Czakon, Mitov, Poncelet (2205.11879)] **CMP**

Modify anti-k_t distance when flavoured particles involved

$$d_{ij} = R^2 \min(k_{T,i}^{-2}, k_{T,j}^{-2}) \cdot S_{ij}^a \,, \quad d_B = k_{T,i}^{-2}$$

where $S_{ij} \neq 1$ only when i and j are of opposite flavour

$$S_{ij}^a = 1 - \theta(1-\kappa)\cos\left(\frac{\pi}{2}\kappa\right) \,, \quad \kappa = \frac{1}{a}\frac{k_{T,i}^2 + k_{T,j}^2}{2\, k_{T,\max}^2}$$

One recovers (IRC flavour unsafe) anti-k_t jets when $a \to 0$

The flavour revolution (2022-…)

IRC safe to all orders
Exact anti-k_t kinematics

[Gauld, Huss, Stagnitto (2208.11138)] **GHS**

Flavour assignment *factorised* from jet reconstruction:

1) first jets are clustered with anti-k_t (or any other algorithm)

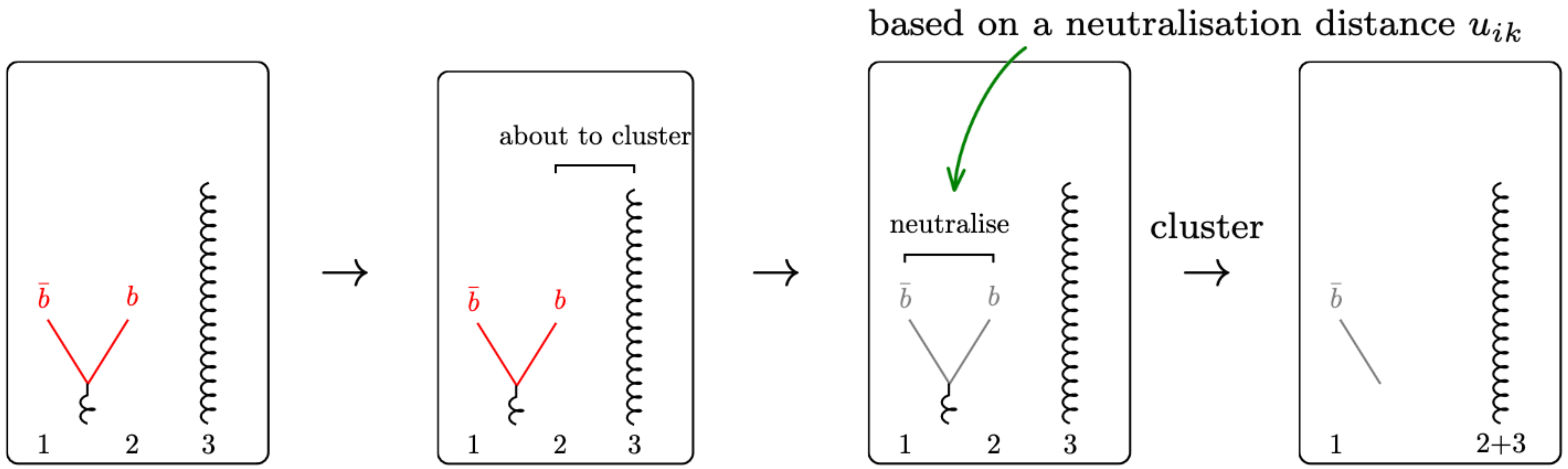
2) then flavour is assigned to jets by running a sequential algorithm with flavour-k_t-like distances between particles and jets

The flavour revolution (2022-…)

IRC safe to all orders
Exact anti-k_t kinematics

[Caola, Grabarczyk, Hutt, Salam, Scyboz, Thaler (2306.07314)] **IFN**

Interleaved Flavour Neutralisation (IFN):
cluster the particles, and "neutralise" flavour when necessary



neutralise ≡ remove the (opposite) flavours of both 1 & 2
while maintaining kinematics

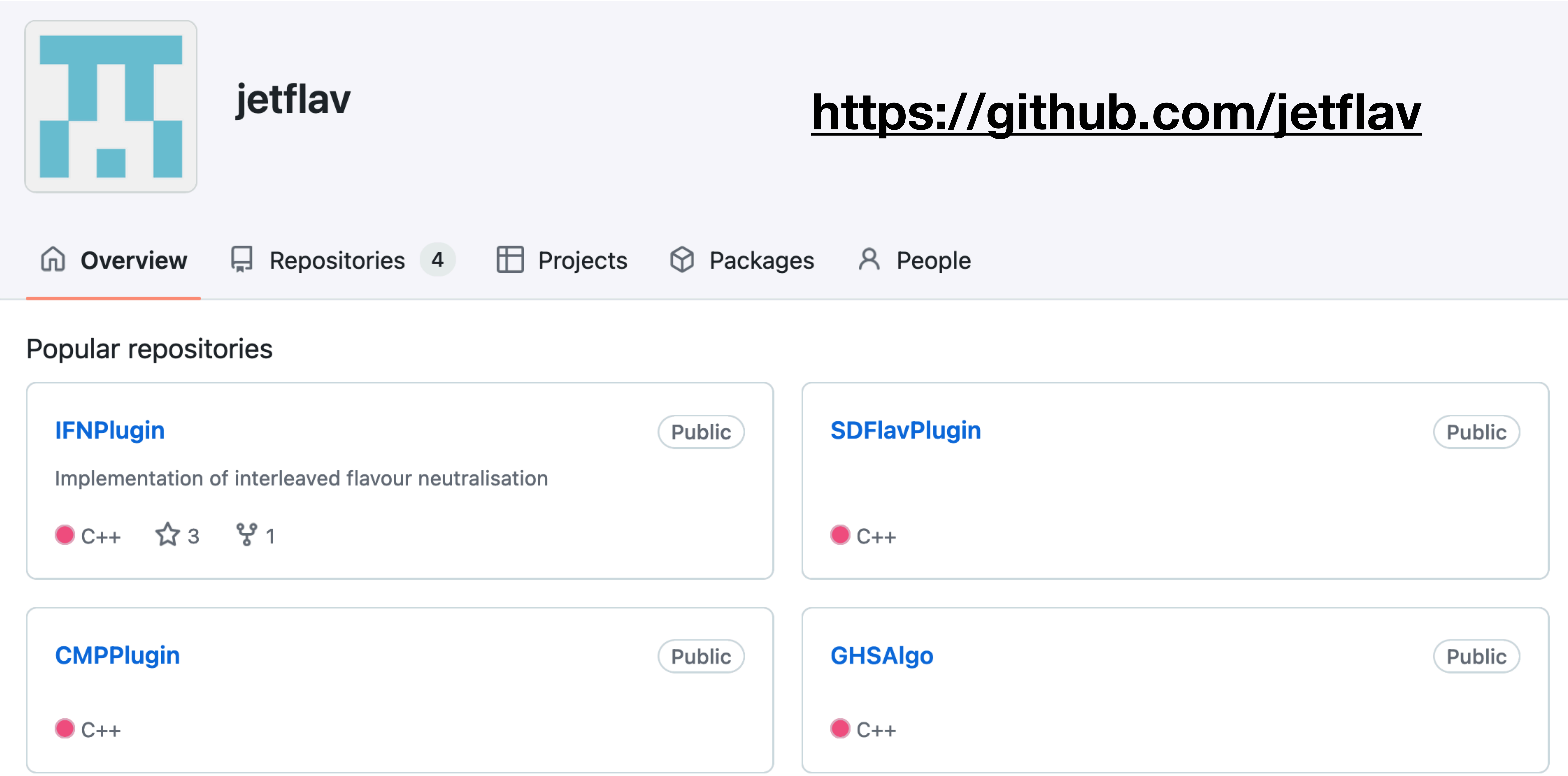
Ongoing studies started at the Les Houches PhysTeV 2023 workshop

[Standard Model WG report (2406.00708)]

Compare new generation of theory-friendly flavoured jet algorithms,
both at fixed-order and with (N)LO+PS simulations

In the following slides I will flash some selected (preliminary) results.
A detailed paper in preparation.

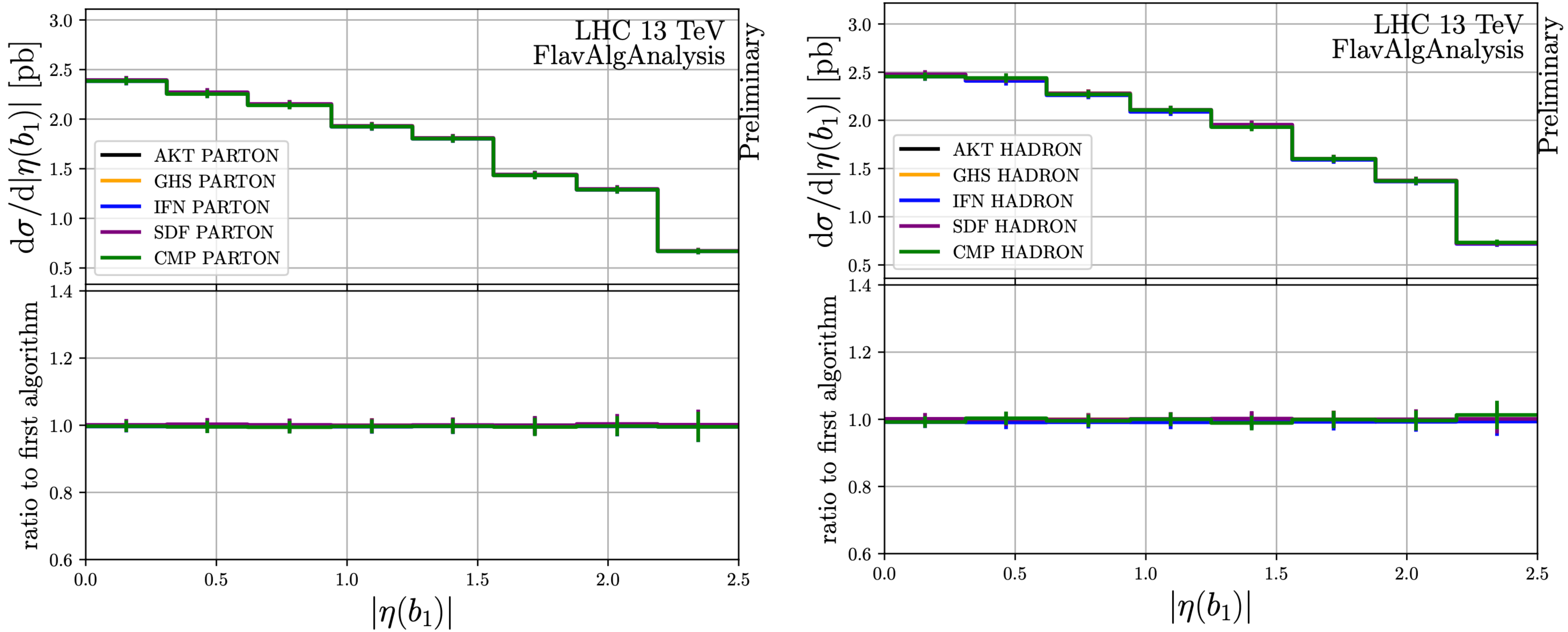
First task at LH: common framework with FastJet implementation



A proper FastJet contribution (to be part of fjcontrib) is in preparation

Z+b-jet @ NLO+PS in the central region: comparison of algorithms

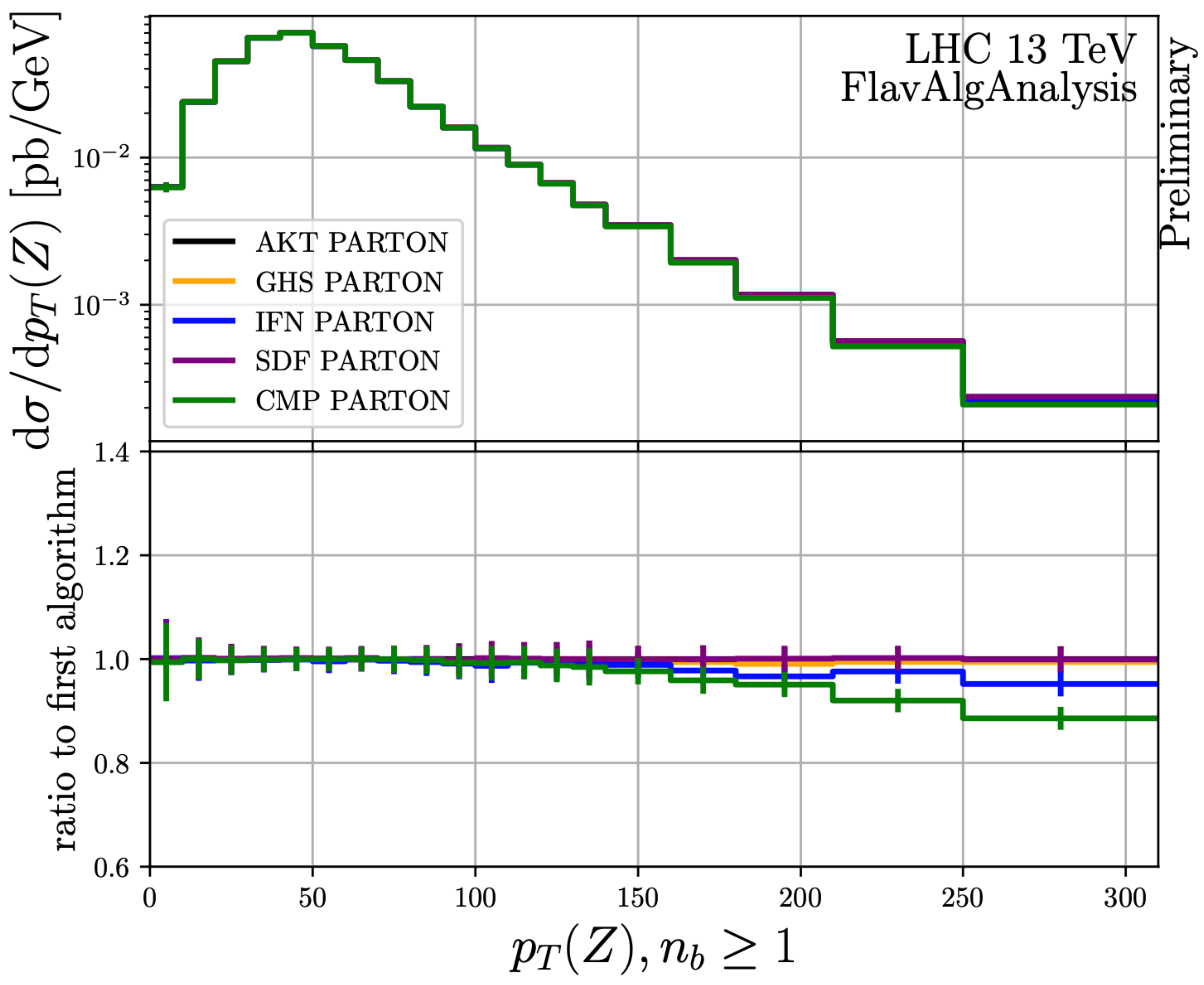
Thanks to Rene Poncelet and Daniel Reichelt!

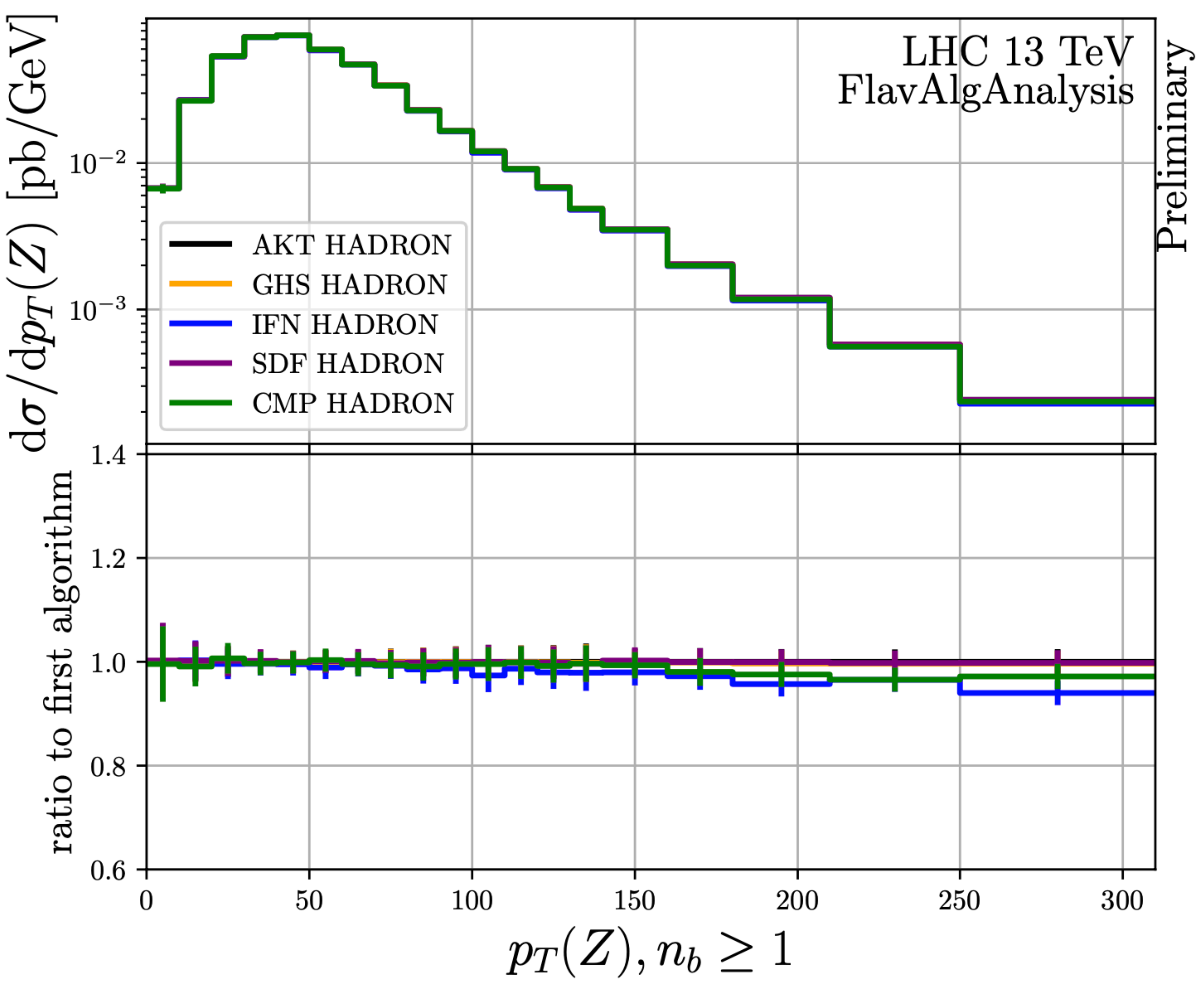


For many distributions, difference between algorithms negligible,
(once you require IRC safety, constrained behaviour in most of the phase space?)

Z+b-jet @ NLO+PS in the central region: comparison of algorithms

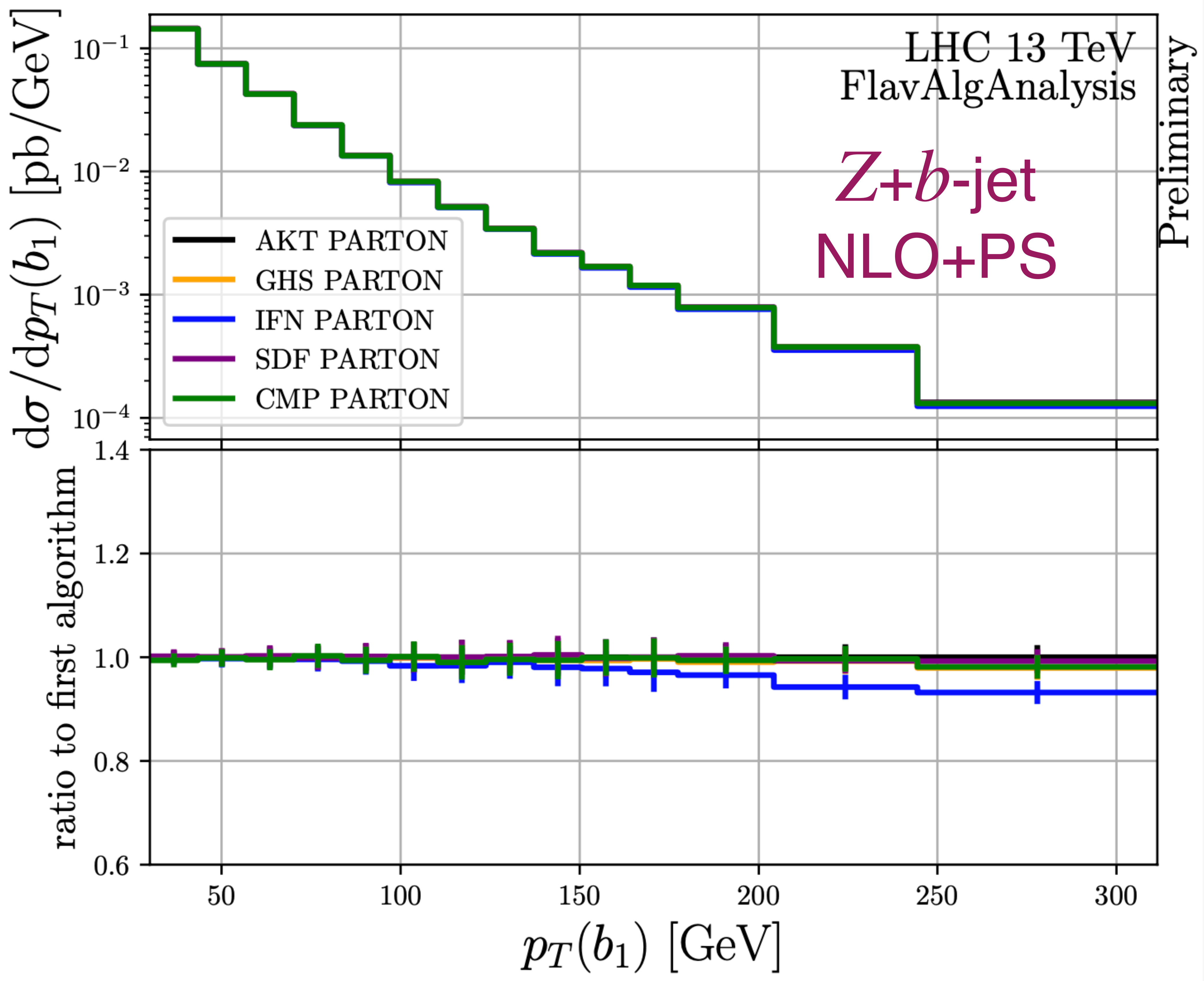
Thanks to Rene Poncelet and Daniel Reichelt!

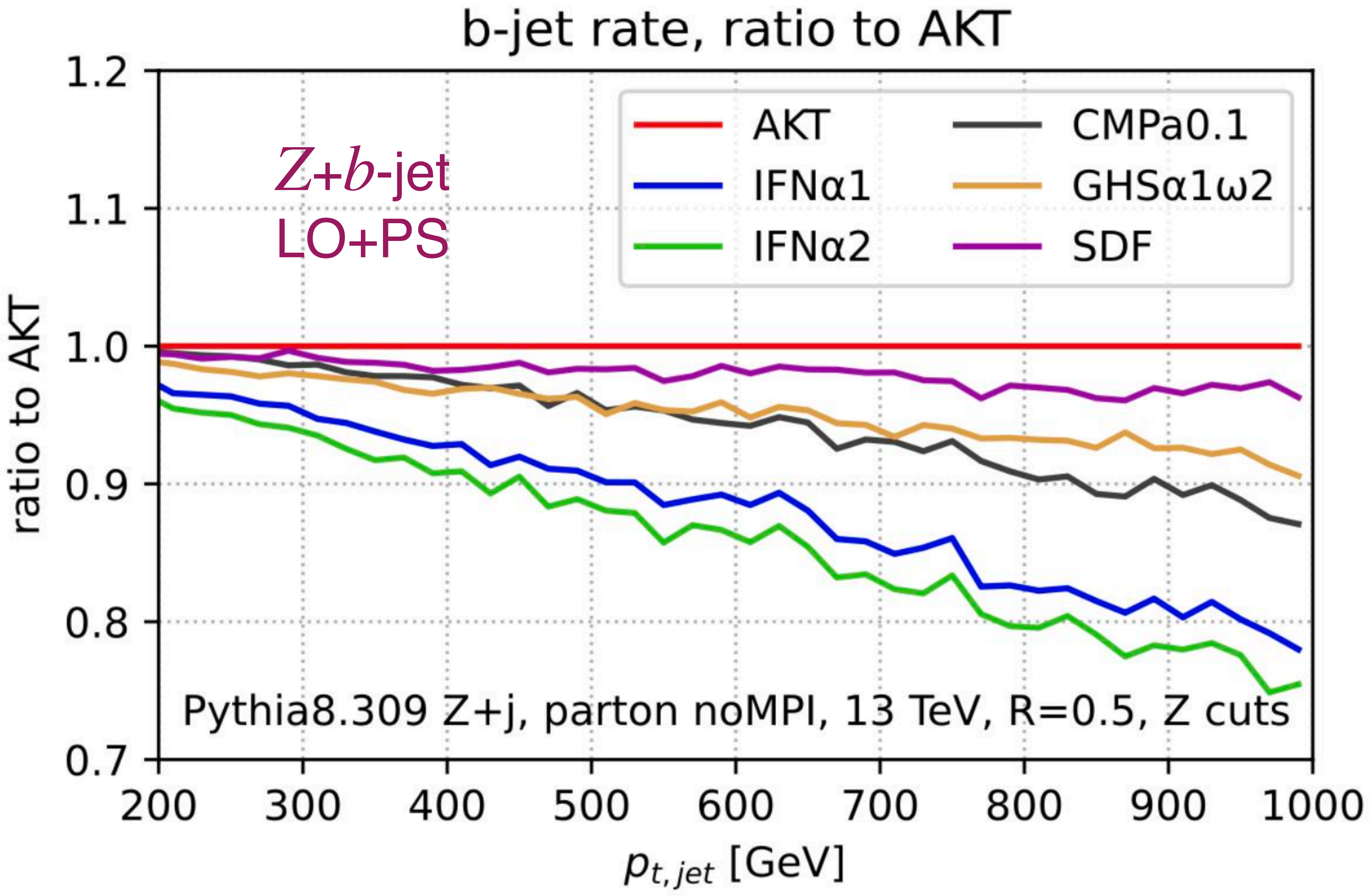




Visible difference at high-p_T: CMP and IFN outliers compared to AKT/GHS/SDF

Z+b-jet @ NLO+PS in the central region: understanding the high-p_T behaviour





As we increase the p_T, the difference between algorithms is enhanced

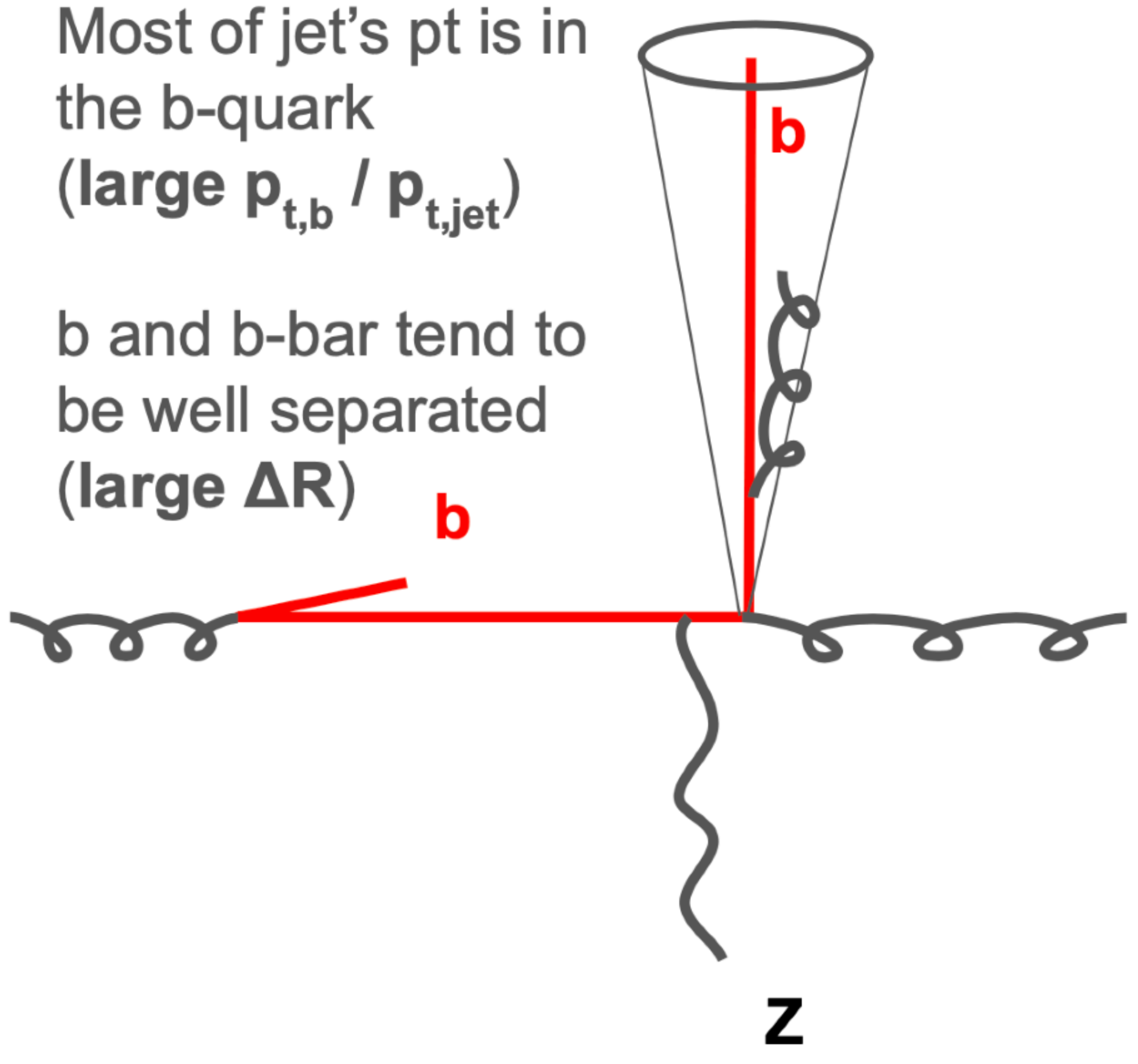
Z+b-jet in the central region: understanding the high-p_T behaviour

Thanks to Ludovic Scyboz and Gavin Salam!

Genuine b-jet

Most of jet's pt is in the b-quark (**large $p_{t,b}$ / $p_{t,jet}$**)

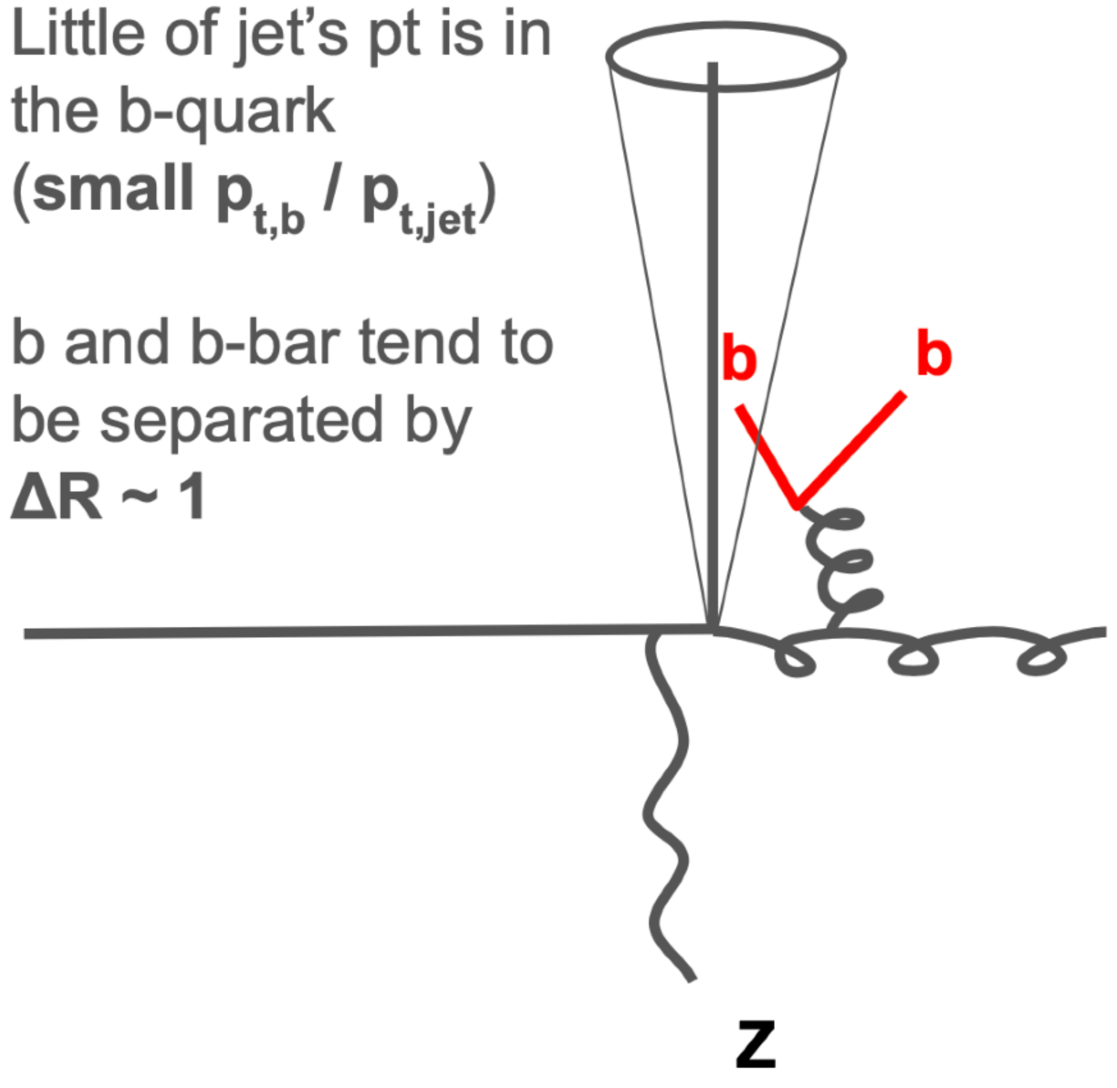
b and b-bar tend to be well separated (**large ΔR**)



Fake b-jet

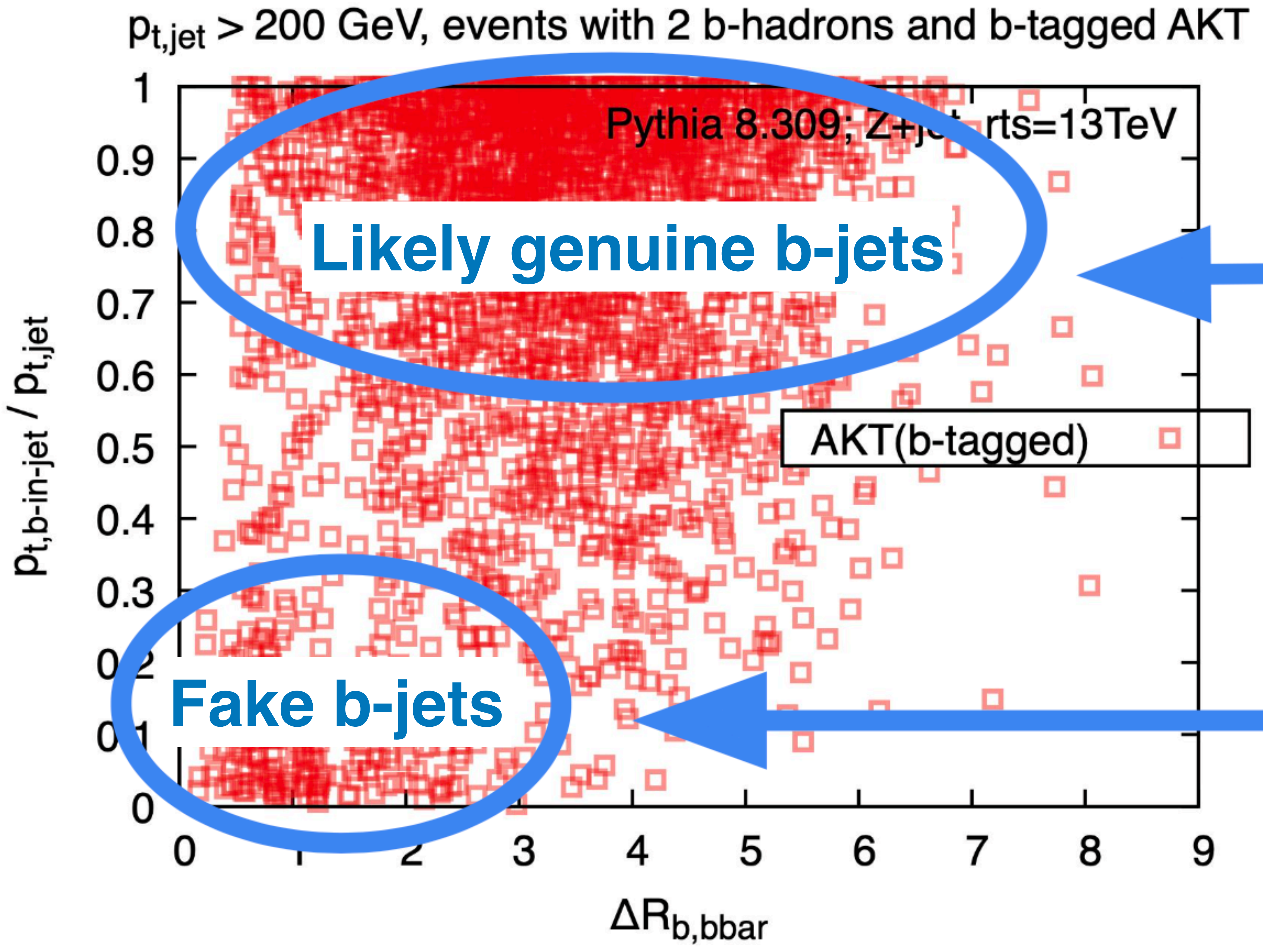
Little of jet's pt is in the b-quark (**small $p_{t,b}$ / $p_{t,jet}$**)

b and b-bar tend to be separated by **ΔR ~ 1**



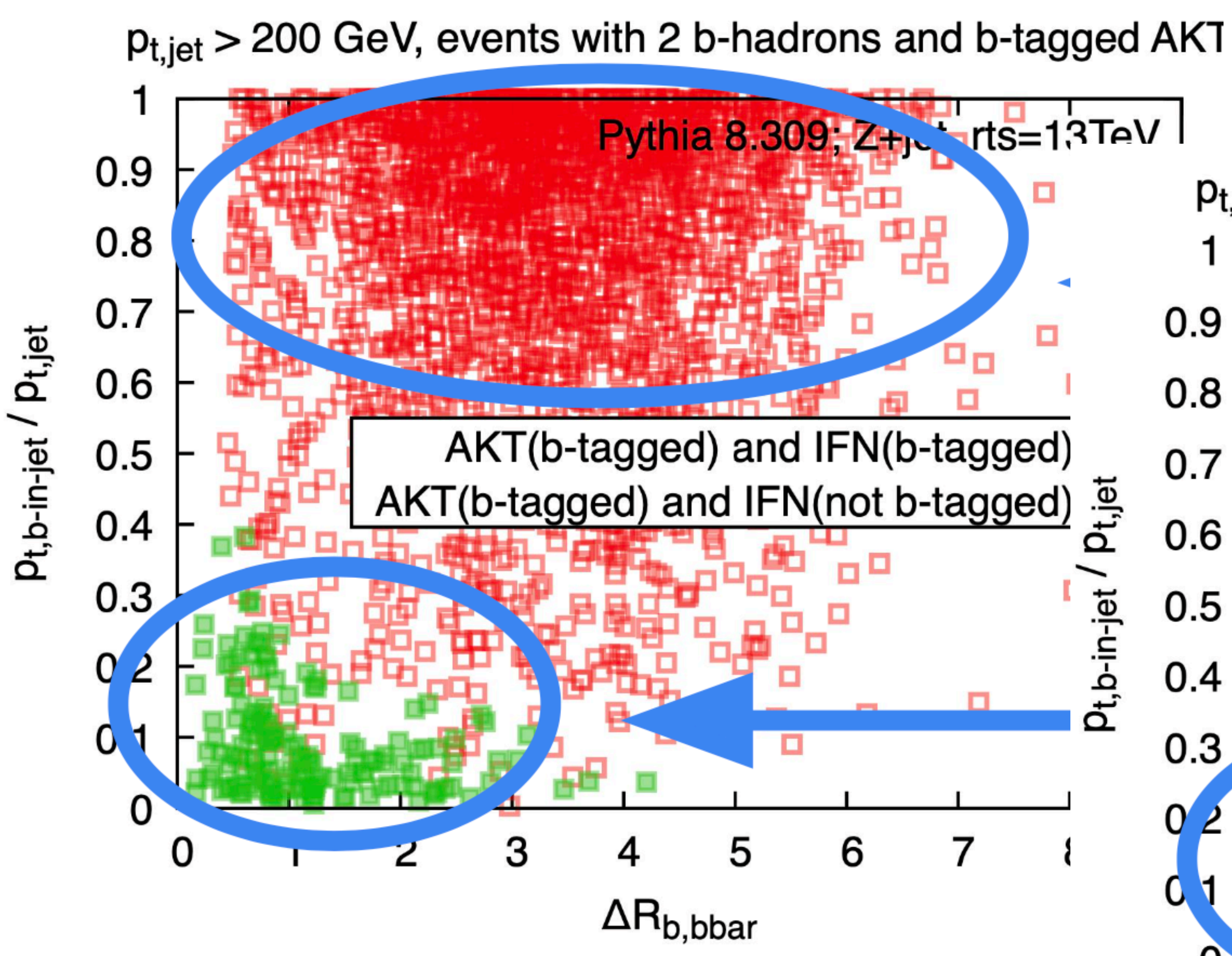
"Genuine" and "fake" b-jets can be disentangled in a ($p_{T,b}/p_{T,j}$, $\Delta R_{b,\bar{b}}$) plane

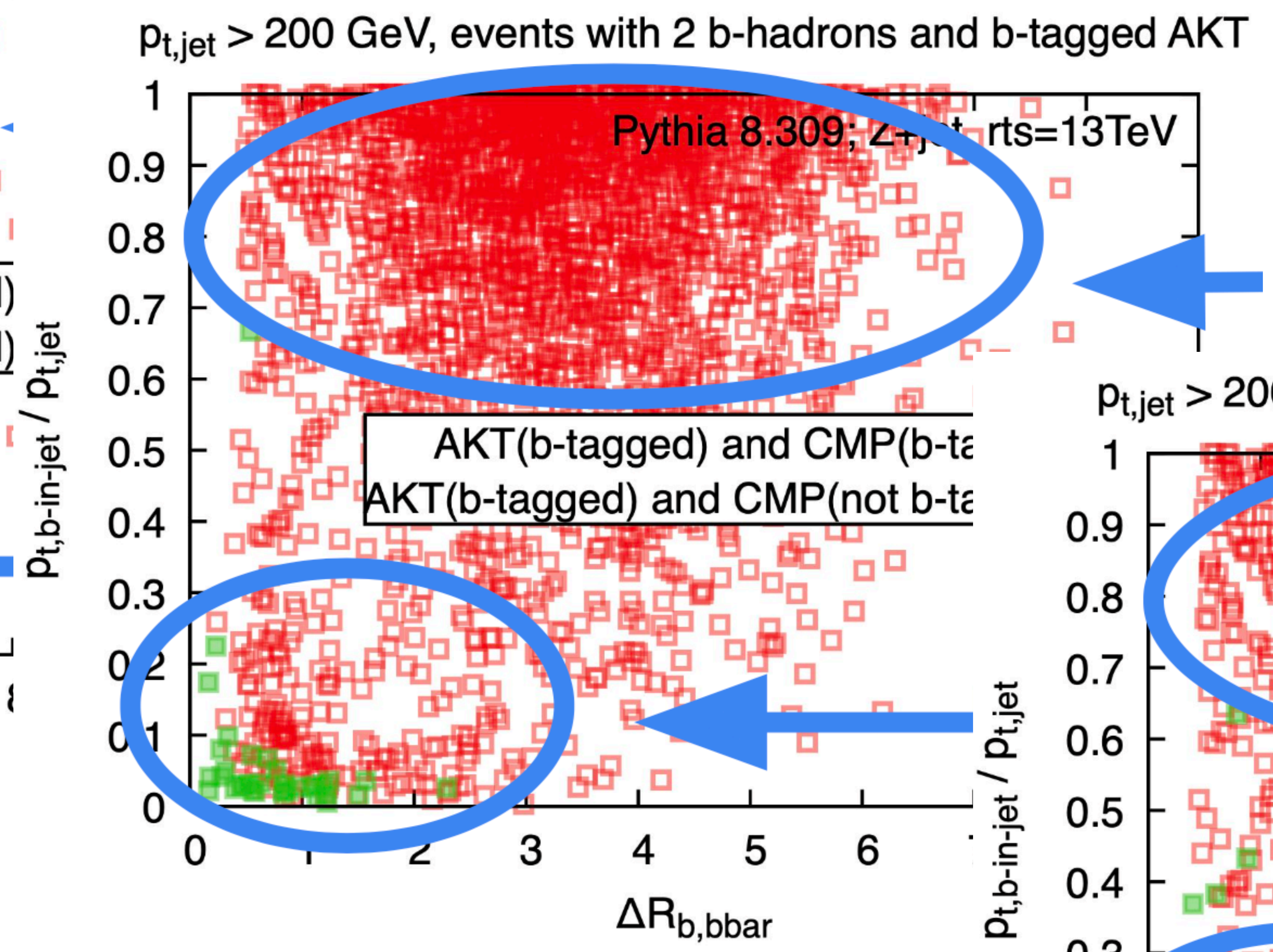
Naive AKT will tag both "genuine" and "fake" b-jets

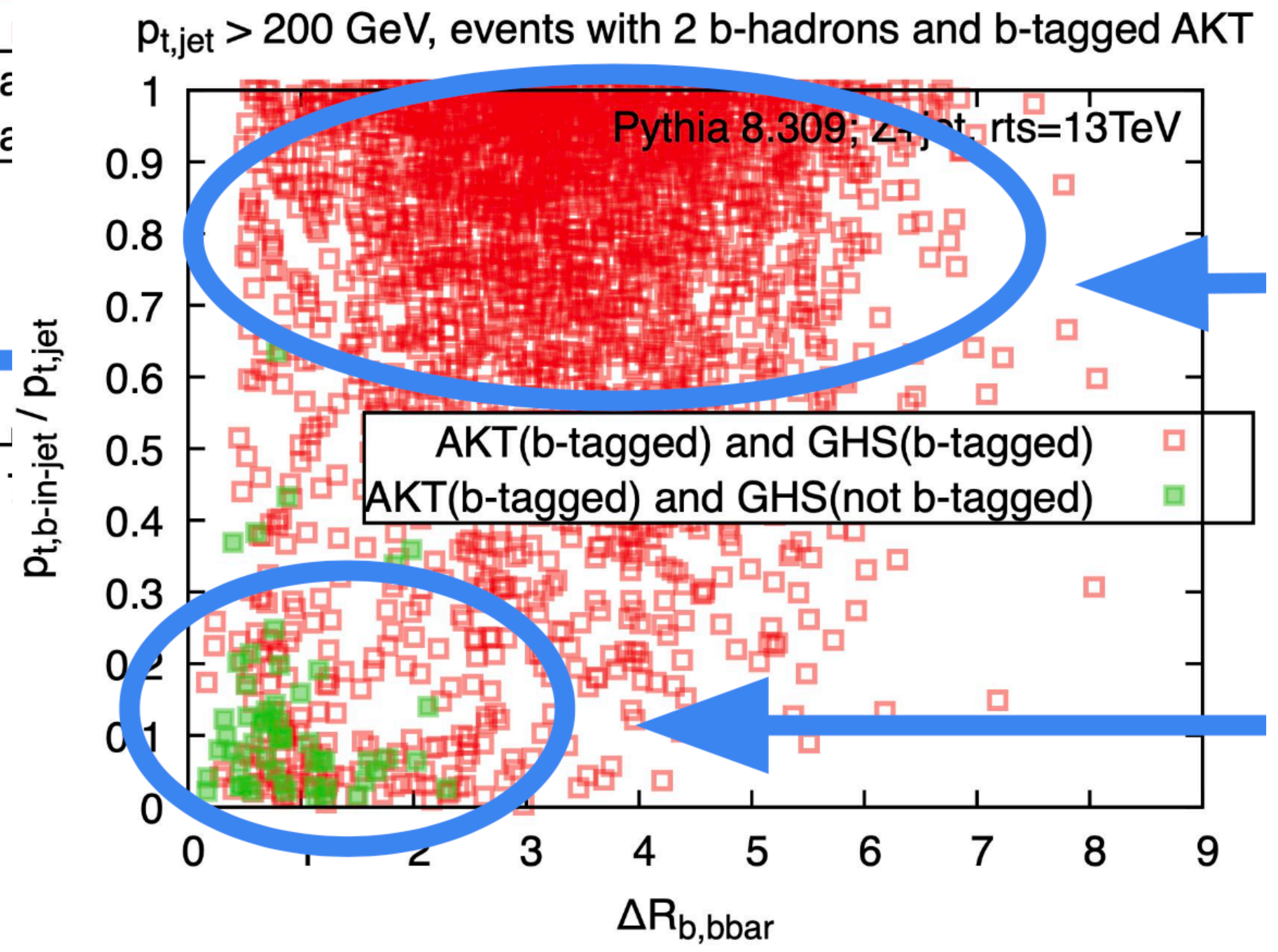


Z+b-jet in the central region: understanding the high-p_T behaviour

Thanks to Ludovic Scyboz and Gavin Salam!







Why not using calculations with massive quarks?

In principle, massive calculations do *not* require IRC safe flavour algorithms (thanks to screening effect due to quark mass m_q).

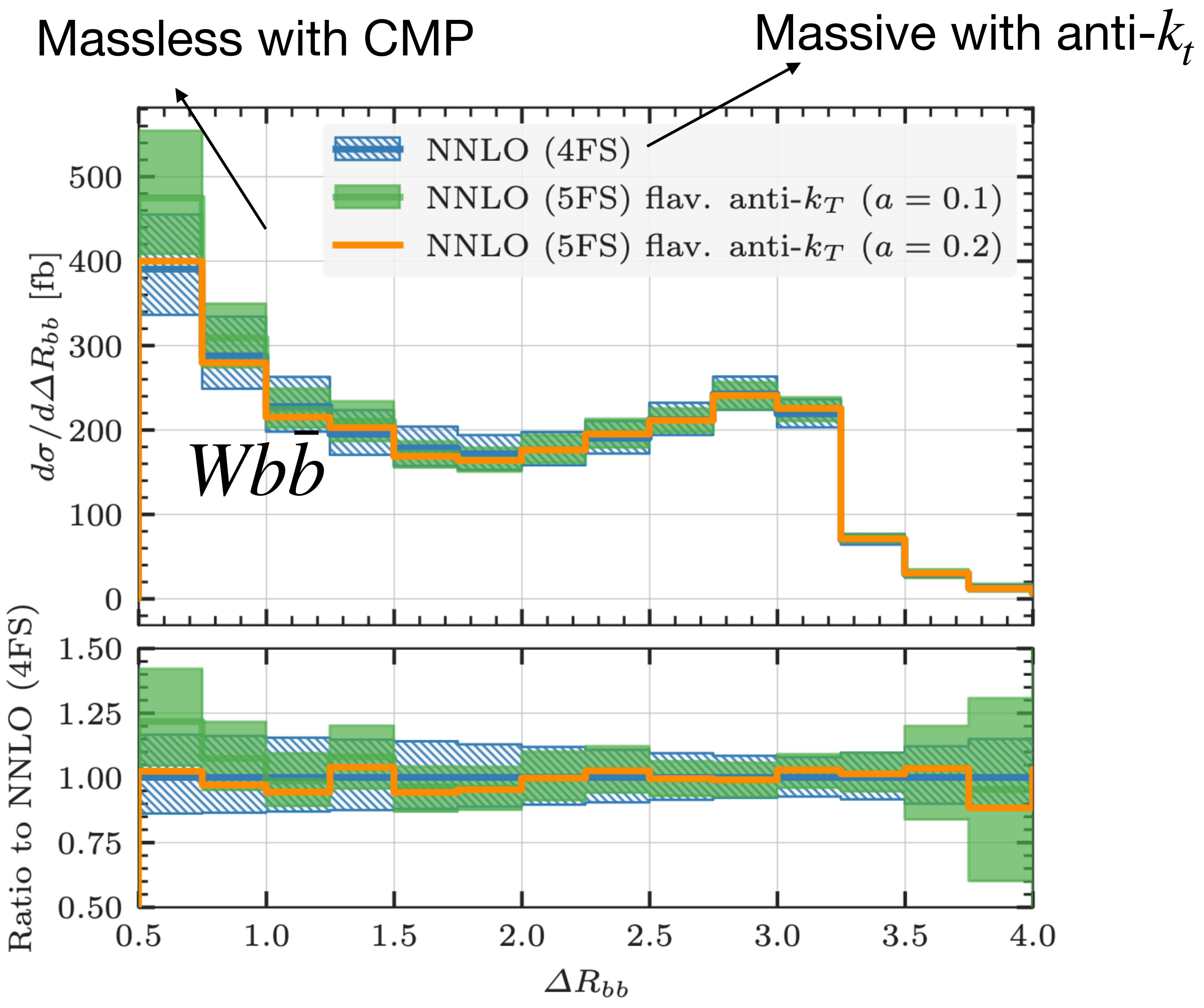
Practical answer: for many processes, massive quarks are beyond the state of the art (especially on the amplitude side)

More justified answer: presence of large logarithms $\log(Q^2/m_q^2)$ in massive calculation, spoiling the convergence of the perturbative series.

An IRC safe prescription implies a suppressed sensitivity on $\log(Q^2/m_q^2)$

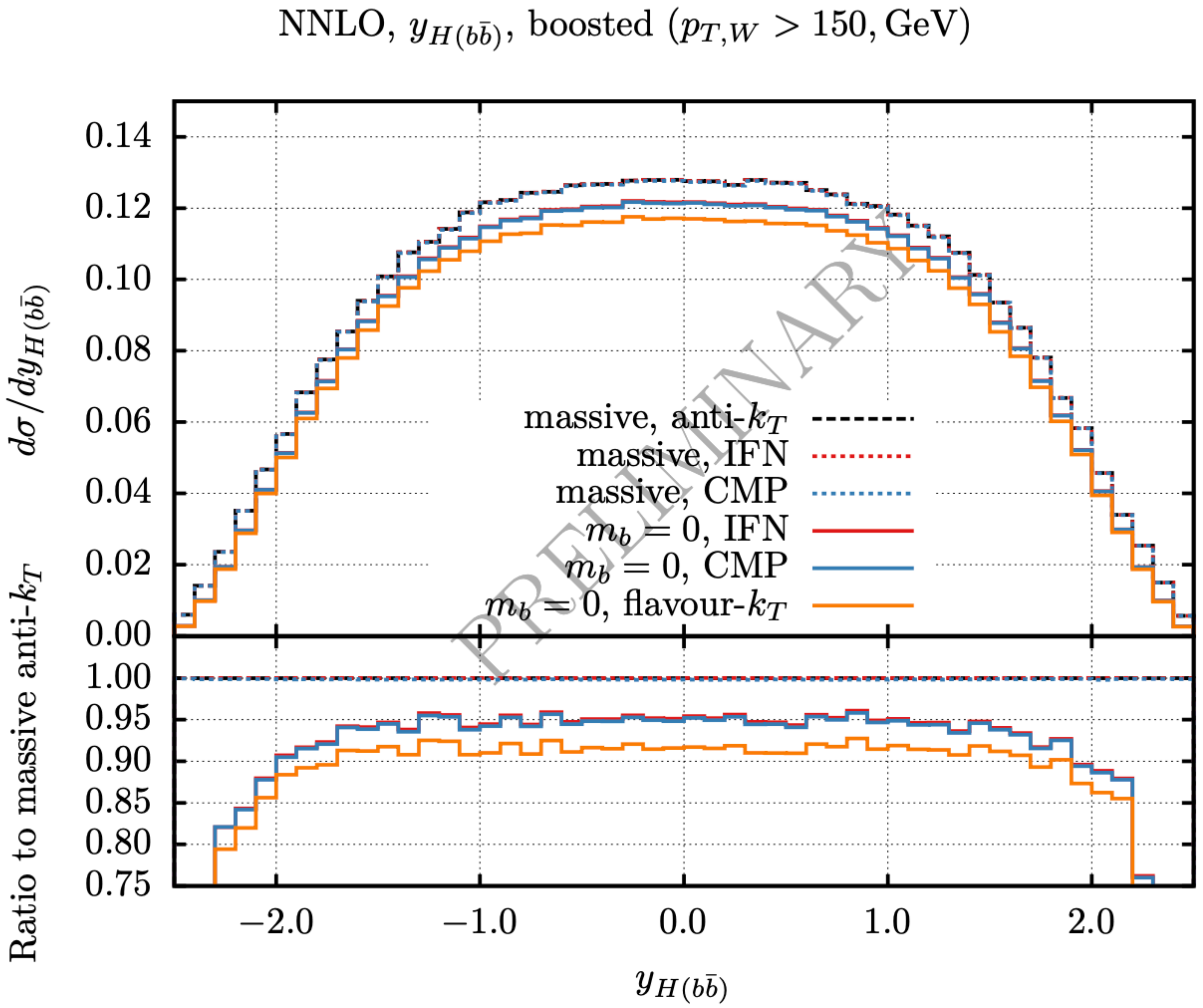
(And a massless calculation allows for a resummation of initial-state $\log(Q^2/m_q^2)$ by PDF evolution)

Compare massless and massive calculations (when available)



$Wb\bar{b}$ [Buonocore et al. (2212.04954)]

Thanks to Arnd Behring and Raoul Röntsch!



$WH(b\bar{b})$ [Behring et al. (2003.08321)]
and LH study in progress

Conclusions

Renewed interest in proper definition of flavoured jet, driven by precision requirements of the LHC and availability of new higher order calculations

Comparison between new algorithms crucial to assess size of MC corrections needed for theory-data comparison and to explore adoption of “improved” labels for flavour tagging

Many studies started in Les Houches, preliminary results available, to be finalised and collected in a dedicated paper

Not a clear conclusion yet: synergy between theory and experiments crucial at this stage!

BACKUP

Z+b-jet in the central region: detail of the analysis

MC@NLO sample of $Z(\rightarrow \mu^+\mu^-)$ + jet events with SHERPA (100M events)
LHC @ 13 TeV, PDF4LHC21, scale choice $E_T(Z)$

Generation cuts:
$p_T(\mu^\pm) > 20$ GeV; $|y(\mu^\pm)| < 2.4$; 71 GeV $< m(\mu^+\mu^-) <$ 111 GeV;
$p_T(\mu^+\mu^-) > 20$ GeV (avoid jet requirement at generation level)

Samples both at parton and stable heavy-hadron level

Rivet analysis FlavAlgAnalysis (https://github.com/DReichelt/LH23FlavAlgs)
Based on https://rivet.hepforge.org/analyses/CMS_2017_I1499471

Jet reconstructed with anti-k_t and then tagged (e.g. GHS, SDF)
or directly reconstructed and tagged with a flavoured algorithm (e.g. CMP, IFN)
$R = 0.5$, $p_T(j) > 30$ GeV, $|\eta(j)| < 2.4$

Default parameter choice for the algorithms

[Caletti, Larkoski, Marzani, Reichelt (2205.01109)] **SDF**

$\beta = 1, z_{\text{cut}} = 0.1$ (soft-drop parameters)

[Czakon, Mitov, Poncelet (2205.11879)] **CMP**

$a = 0.1$ (anti-k_t-like distance)

[Gauld, Huss, Stagnitto (2208.11138)] **GHS**

$\alpha = 1, \omega = 2$ (flavour-k_t-like and beam distance)

[Caola, Grabarczyk, Hutt, Salam, Scyboz, Thaler (2306.07314)] **IFN**

$\alpha = 1, \omega = 2$ (flavour-k_t-like distance)

In plots, **AKT** is the naive IRC-unsafe flavour tagging of anti-k_t jets
(does the jet contain a flavoured parton/hadron?)

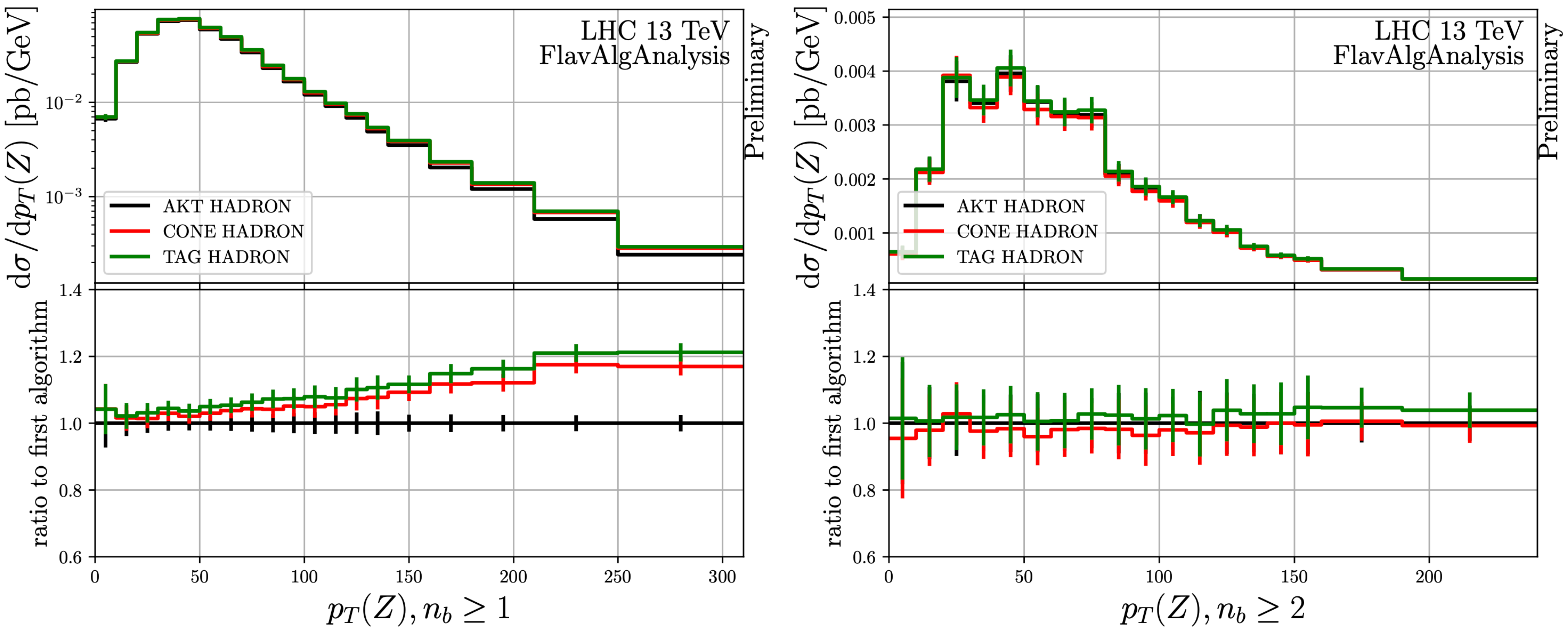
Z+b-jet in the central region: comparison of (unsafe) tagging strategies

AKT:
check if flavoured particle / hadron is inside the anti-k_t jet

CONE:
ATLAS-style tagging
(heavy hadron with $\Delta R(j, h) < 0.3$ and with $p_T > 5$ GeV)

TAG:
CMS-style tagging
(bTagged method with ghost-tagging in Rivet)

Z+b-jet in the central region: comparison of (unsafe) tagging strategies



Effect of multi-b tags important at high-p_T: by requiring presence of 2 b-jets, we have a reduced probability of double tags, mostly coming from first splitting