Flavoured jet algorithms: a comparative study

An attempt at a summary

Arnd Behring

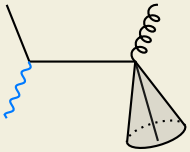
Max-Planck-Institut für Physik
Garching bei München, Germany

based on arXiv:2506.13449 [hep-ph] in collaboration with

S. Caletti, F. Giuli, R. Grabarczyk, A. Hinzmann, A. Huss, J. Huston, E. D. Lesser, S. Marzani, D. Napoletano, R. Poncelet, D. Reichelt, A. Rescia, G. P. Salam, L. Scyboz, F. Sforza, A. Siódmok, G. Stagnitto, J. Whitehead, R. Xu

18 June 2025 – PhysTeV 2025 workshop – Les Houches, France

Flavour jets

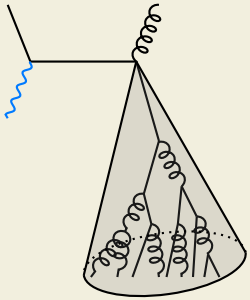


Jets are stable concepts: jets defined at

- fixed order
- parton level
- hadron level
- detector level

roughly correspond to each other

Flavour jets

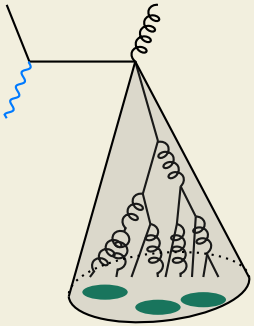


Jets are stable concepts: jets defined at

- fixed order
- parton level
- hadron level
- detector level

roughly correspond to each other

Flavour jets

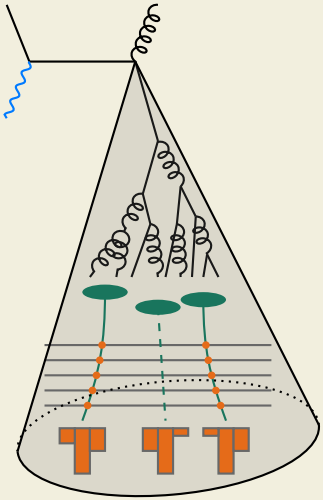


Jets are stable concepts: jets defined at

- fixed order
- parton level
- hadron level
- detector level

roughly correspond to each other

Flavour jets

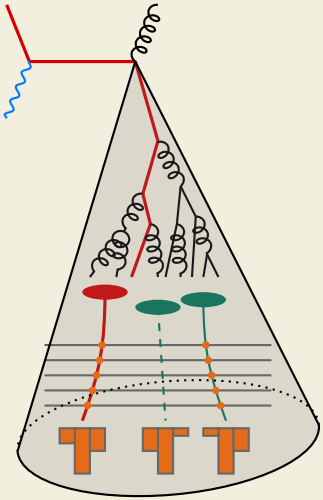


Jets are stable concepts: jets defined at

- fixed order
- parton level
- hadron level
- detector level

roughly correspond to each other

Flavour jets



Jets are stable concepts: jets defined at

- fixed order
- parton level
- hadron level
- detector level

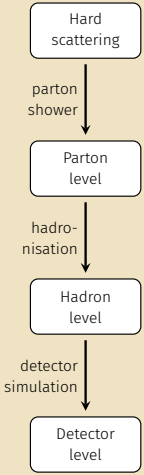
roughly correspond to each other

It is interesting to assign flavour to a jet

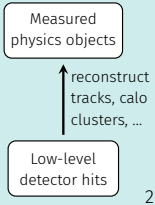
- Access to essential information about hard scattering (and potentially the initial state)
- Signature for many important signals (top, Higgs, intrinsic charm?, ...)

However: flavour assignment is a non-trivial task

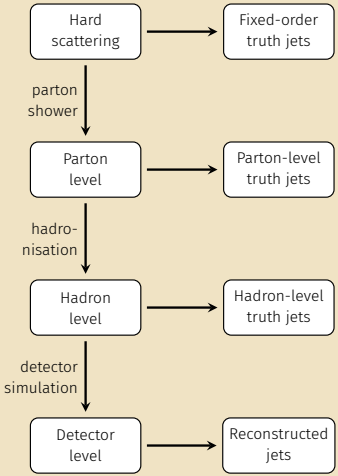
Theory / Simulation



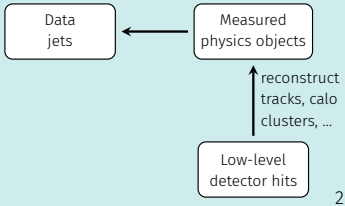
Experiment / Data

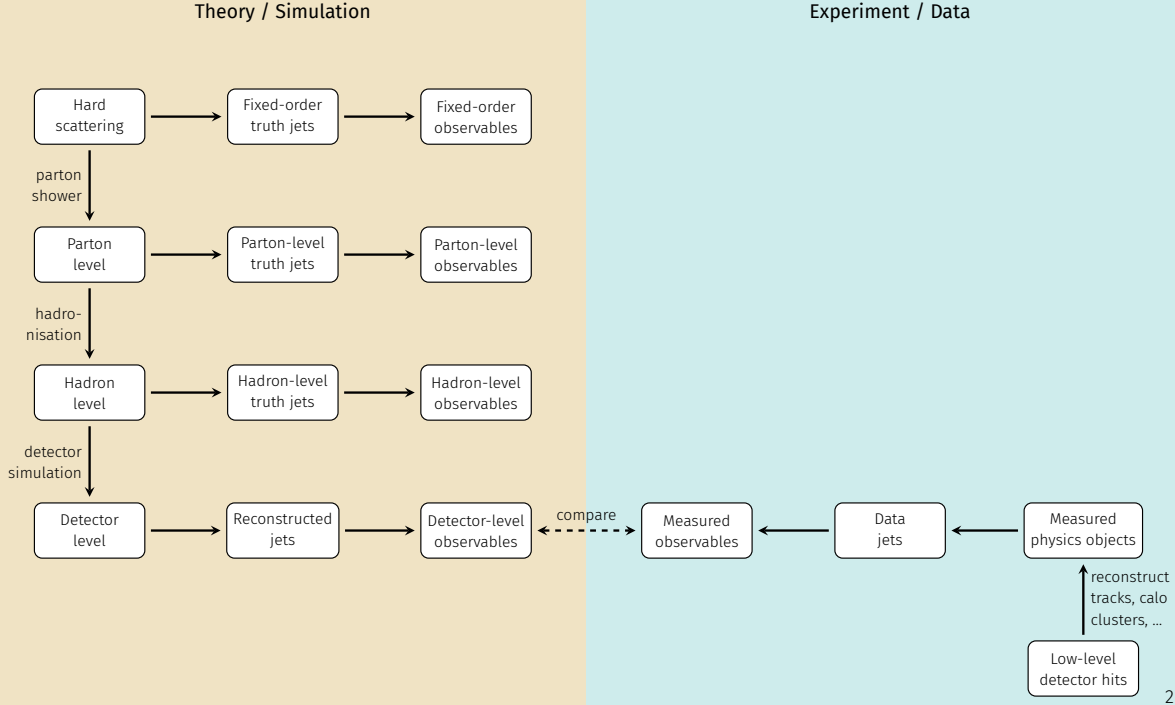


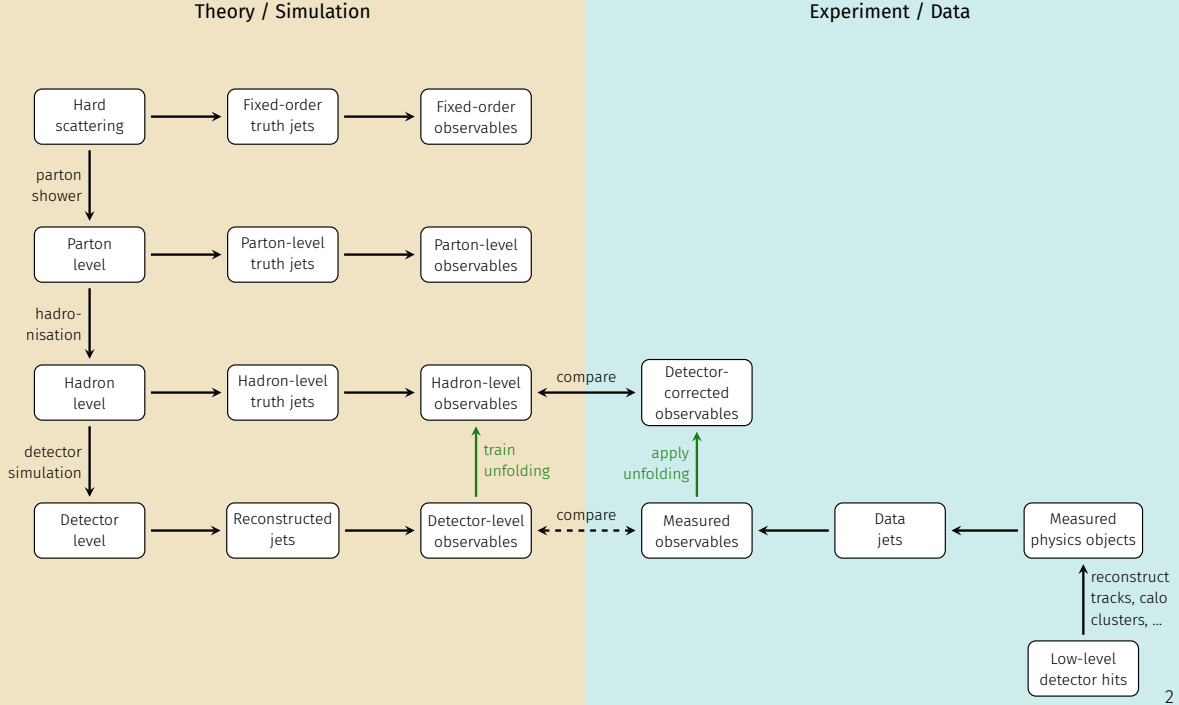
Theory / Simulation

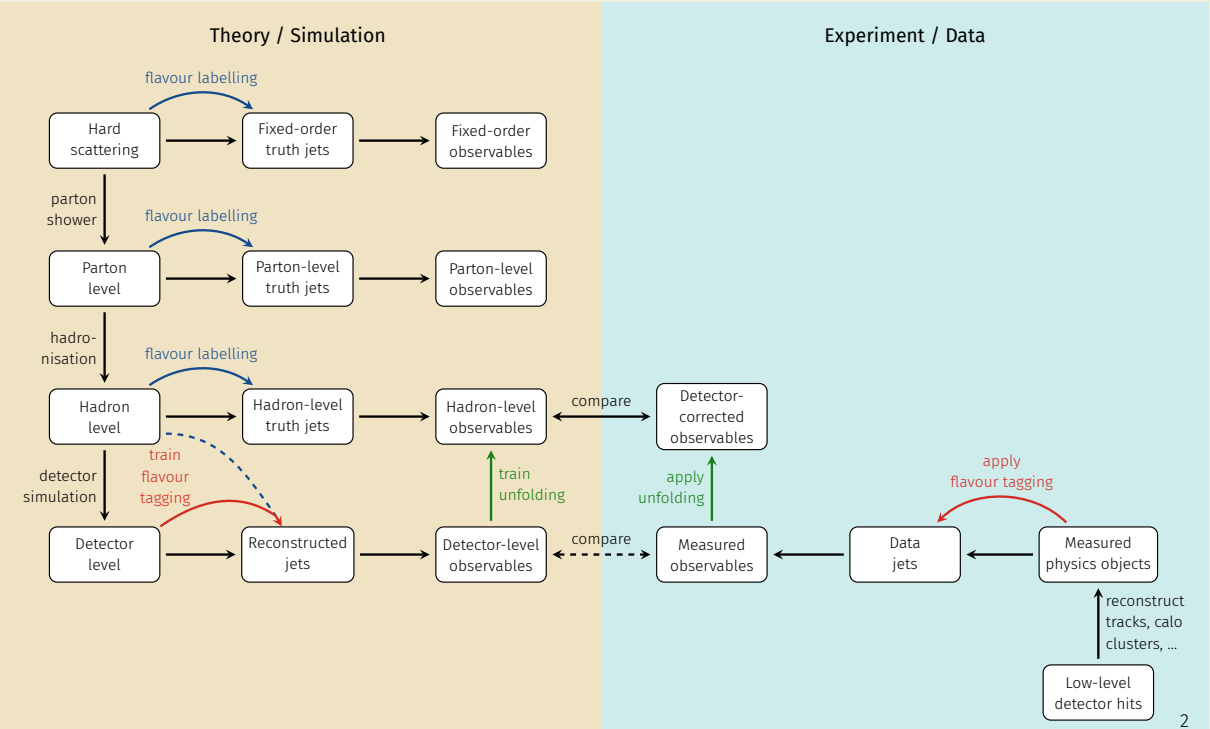


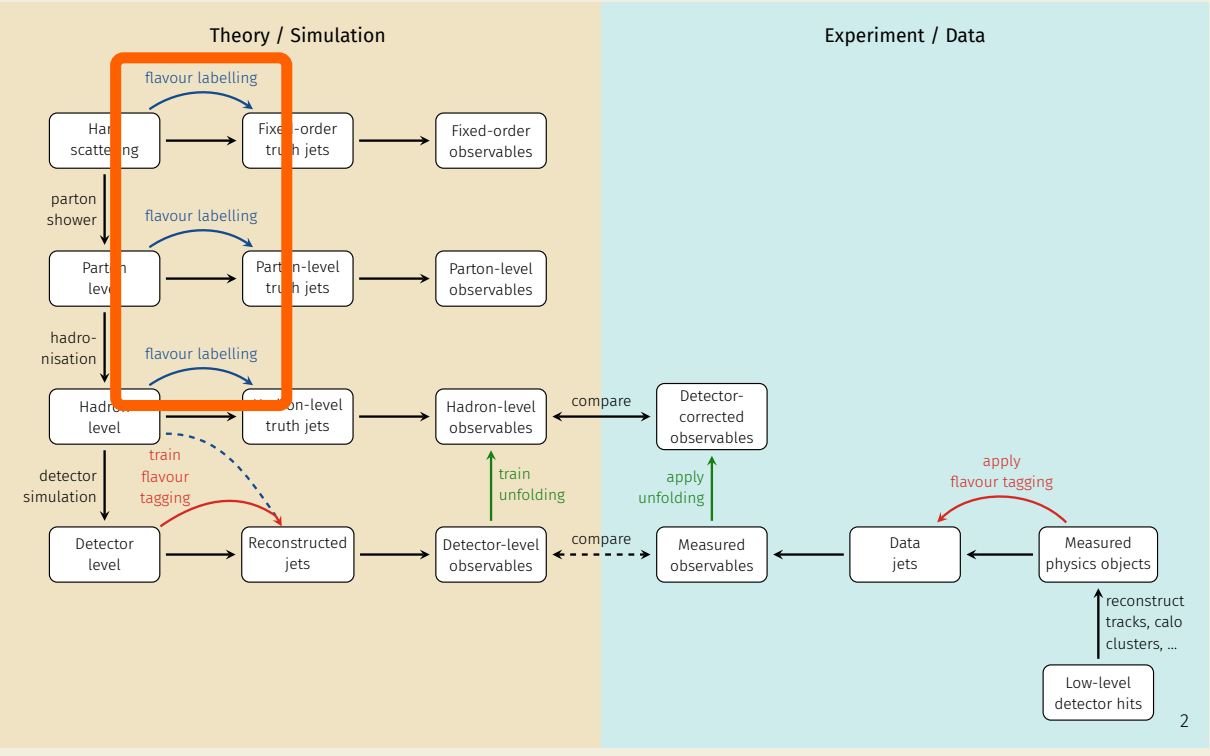
Experiment / Data



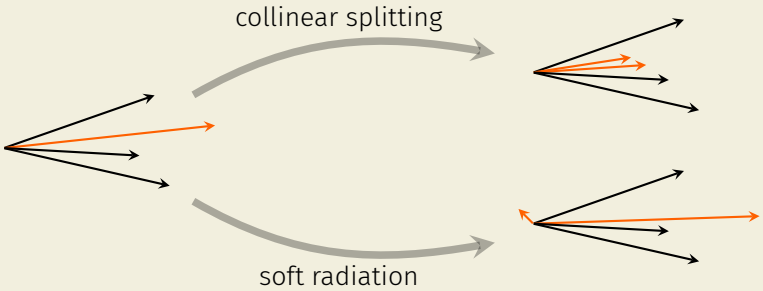

Theory / Simulation
Experiment / Data
Hard scattering
Fixed-order truth jets
Fixed-order observables
parton shower
Parton level
Parton-level truth jets
Parton-level observables
hadro-nisation
Hadron level
Hadron-level truth jets
Hadron-level observables
detector simulation
Detector level
Reconstructed jets
Detector-level observables
compare
Measured observables
Data jets
Measured physics objects
reconstruct tracks, calo clusters, ...
Low-level detector hits


Theory / Simulation
Experiment / Data
Hard scattering
Fixed-order truth jets
Fixed-order observables
parton shower
Parton level
Parton-level truth jets
Parton-level observables
hadro-nisation
Hadron level
Hadron-level truth jets
Hadron-level observables
compare
Detector-corrected observables
detector simulation
train unfolding
apply unfolding
Detector level
Reconstructed jets
Detector-level observables
compare
Measured observables
Data jets
Measured physics objects
reconstruct tracks, calo clusters, ...
Low-level detector hits


Theory / Simulation
Experiment / Data
flavour labelling
Hard scattering
Fixed-order truth jets
Fixed-order observables
parton shower
flavour labelling
Parton level
Parton-level truth jets
Parton-level observables
hadro-nisation
flavour labelling
Hadron level
Hadron-level truth jets
Hadron-level observables
compare
Detector-corrected observables
detector simulation
train flavour tagging
train unfolding
apply unfolding
apply flavour tagging
Detector level
Reconstructed jets
Detector-level observables
compare
Measured observables
Data jets
Measured physics objects
reconstruct tracks, calo clusters, ...
Low-level detector hits


Theory / Simulation
Experiment / Data
flavour labelling
Fixed-order truth jets
Fixed-order observables
parton shower
flavour labelling
Parton-level truth jets
Parton-level observables
hadro-nisation
flavour labelling
Hadron-level truth jets
Hadron-level observables
compare
Detector-corrected observables
train flavour tagging
train unfolding
apply unfolding
apply flavour tagging
detector simulation
Detector level
Reconstructed jets
Detector-level observables
compare
Measured observables
Data jets
Measured physics objects
reconstruct tracks, calo clusters, ...
Low-level detector hits

IRC safety



- Infrared and collinear safe observables must be the same when
  - a particle splits into two collinear particles
  - an additional soft particle is radiated
- Ensures that observables can be calculated in fixed-order perturbation theory (otherwise IR poles do not cancel between real and virtual corrections)
- Not only theoretical issue: IR sensitive observables pick up more dependence on long-distance / low-energy physics

IRC safety and flavour

Flavour dependent observables can easily violate IRC safety

"...we require at least one b-tagged jet with $p_{t,b} \geq 25\,\text{GeV}$ *..."*

IRC safety and flavour

Flavour dependent observables can easily violate IRC safety

"...we require at least one b-tagged jet with $p_{t,b} \geq 25\,\text{GeV}$ *..."*

Collinear $g \to b\bar{b}$ splitting

- Problematic if presence of "any" b quark determines jet flavour
- Collinear splitting influences flavour assignment
- Problem starts at NLO
- Fix: Change flavour recombination scheme

IRC safety and flavour

Flavour dependent observables can easily violate IRC safety

"...we require at least one b-tagged jet with $p_{t,b} \geq 25\,\text{GeV}$ *..."*

Collinear $g \to b\bar{b}$ splitting

- Problematic if presence of "any" b quark determines jet flavour
- Collinear splitting influences flavour assignment
- Problem starts at NLO
- Fix: Change flavour recombination scheme

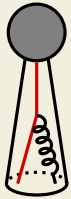
scheme \ jet contents	b	$b+\bar{b}$	$b+b$	
"any flavour"	b	b	b	simplest experimentally (but collinear unsafe for $m_b \to 0$)
net flavour	b	g	$2b$	theoretically "ideal" definition; but not robust wrt B–Bbar oscillations
flavour modulo 2	b	g	g	theoretically OK; robust wrt B–Bbar oscillations

Table: [Salam '24]

IRC safety and flavour

Flavour dependent observables can easily violate IRC safety

"...we require at least one b-tagged jet with $p_{t,b} \geq 25\,\text{GeV}$ *..."*



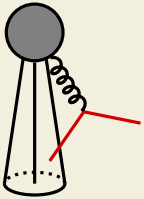
Collinear $b \to bg$ splitting

- Causes logarithmic sensitivity if flavoured hadron must pass, e.g., $p_{t,\text{cut}} > 5\,\text{GeV}$ cut
- Such a cut cannot be implemented on parton level
  $\to$ requires hadronisation/fragmentation function

IRC safety and flavour

Flavour dependent observables can easily violate IRC safety

"...we require at least one b-tagged jet with $p_{t,b} \geq 25\,\text{GeV}$ *..."*



Soft $g \to b\bar{b}$ splitting

- Causes problems when wide angle b quarks are clustered into jets
- Soft splitting influences flavour assignment
- Problem starts at NNLO
- Fix requires flavour-aware jet algorithms

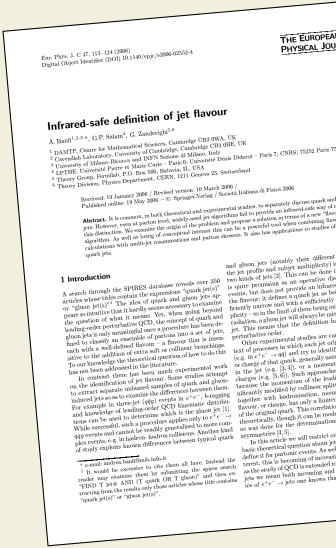
Flavoured jet algorithms

First proposal for flavoured jet algorithm:
Flavour-k_t algorithm [Banfi, Salam, Zanderighi '06]

- Based on k_t algorithm
- Modification of distance measure depending on flavour, e.g.,

$$d_{ij} = \Delta R_{ij} \begin{cases} \max(k_{t,i}^2, k_{t,j}^2) & \text{softer of } i, j \text{ is flavoured,} \\ \min(k_{t,i}^2, k_{t,j}^2) & \text{softer of } i, j \text{ is flavourless} \end{cases}$$

- Prefers clustering soft quark pairs together


Eur. Phys. J. C 47, 113–124 (2006)
Digital Object Identifier (DOI) 10.1140/epjc/s2006-02552-4

Infrared-safe definition of jet flavour

A. Banfi[1,2,3,a], G.P. Salam[4], G. Zanderighi[5,6]

[1] DAMTP, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK
[2] Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK
[3] University of Milano–Bicocca and INFN Sezione di Milano, Italy
[4] LPTHE: Université Pierre et Marie Curie – Paris 6; Université Denis Diderot – Paris 7; CNRS; 75252 Paris 75
[5] Theory Group, Fermilab, P.O. Box 500, Batavia, IL, USA
[6] Theory Division, Physics Department, CERN, 1211 Geneva 23, Switzerland

Received: 19 January 2006 / Revised version: 10 March 2006 /
Published online: 19 May 2006 – © Springer-Verlag / Società Italiana di Fisica 2006

1 Introduction

A search through the SPIRES database reveals over 350 articles whose titles contain the expressions "quark jet(s)" or "gluon jet(s)"[1]. The idea of quark and gluon jets appears so intuitive that it hardly seems necessary to examine the question of what it means. Yet, when going beyond leading-order perturbative QCD, the concept of quark and gluon jets is only meaningful once a procedure has been defined to classify an ensemble of partons into a set of jets, each with a well-defined flavour – a flavour that is insensitive to the addition of extra soft or collinear branchings. To our knowledge the theoretical question of how to do this has not been addressed in the literature.

In contrast there has been much experimental work on the identification of jet flavour. Some studies attempt to extract separate unbiased samples of quark and gluon-induced jets so as to examine the differences between them. For example in three-jet ($q\bar{q}g$) events in e^+e^-, b-tagging and knowledge of leading-order QCD kinematic distributions can be used to determine which is the gluon jet [1]. While successful, such a procedure applies only to $e^+e^- \to q\bar{q}g$ events and cannot be readily generalised to more complex events, e.g. in hadron–hadron collisions. Another kind of study exploits known differences between typical quark

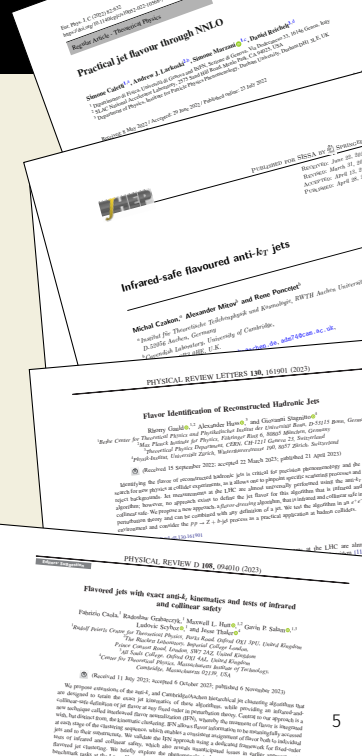
[a] e-mail: andrea.banfi@mib.infn.it
[1] It would be excessive to cite them all here. Instead the reader may examine them by submitting the spires search "FIND T jet# AND (T quark OR T gluon)" and then extracting from the results only those articles whose title contains "quark jet(s)" or "gluon jet(s)".

Flavoured jet algorithms

New generation of jet algorithms

- SDF (soft drop flavour)
  [Caletti, Larkoski, Marzani, Reichelt '22]
- CMP (flavour anti-k_T)
  [Czakon, Mitov, Poncelet '22]
- GHS (flavour dressing)
  [Gauld, Huss, Stagnitto '22]
- IFN (interleaved flavour neutralisation)
  [Caola, Grabarczyk, Hutt, Salam, Scyboz, Thaler '23]

→ Based on/closely resemble anti-k_T algorithm
→ Details: see talks by authors at ▶ LHCb public meeting
→ Available as `fjcontrib` plugins for `FastJet`



Flavoured jet algorithms

New generation of jet algorithms

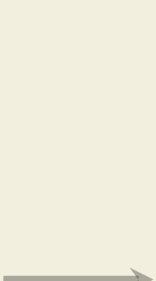
- SDF (soft drop flavour)
  [Caletti, Larkoski, Marzani, Reichelt '22]
- CMP (flavour anti-k_T)
  [Czakon, Mitov, Poncelet '22]
- GHS (flavour dressing)
  [Gauld, Huss, Stagnitto '22]
- IFN (interleaved flavour neutralisation)
  [Caola, Grabarczyk, Hutt, Salam, Scyboz, Thaler '23]

→ Based on/closely resemble anti-k_T algorithm
→ Details: see talks by authors at ▶ LHCb public meeting
→ Available as `fjcontrib` plugins for `FastJet`

Alternative, fragmentation-based approach:
WTA (winner take all flavour)
[Caletti, Larkoski, Marzani, Reichelt '22]

Les Houches 2023

ÉCOLE DE PHYSIQUE DES HOUCHES



CERN-TH-2025-113, IFJPAN-IV-2025-13, MCNET-25-14, MPP-2025-118, PUBDB-2025-01802

Flavoured jet algorithms: a comparative study

Arné Behring,[1] Simone Caletti,[2] Francesco Giuli,[3] Radosław Grabarczyk,[4,5] Andreas Hinzmann,[6] Alexander Huss,[7] Joey Huston,[8] Ezra D. Lesser,[7] Simone Marzani,[3] Davide Napoletano,[10] Rene Poncelet,[11] Daniel Reichelt,[7,12] Alberto Rescia,[9] Gavin P. Salam,[5,13] Ludovic Scyboz,[14] Federico Sforza,[9] Andrzej Siódmok,[11,15] Giovanni Stagnitto,[10] James Whitehead,[16] Ruide Xu[17]

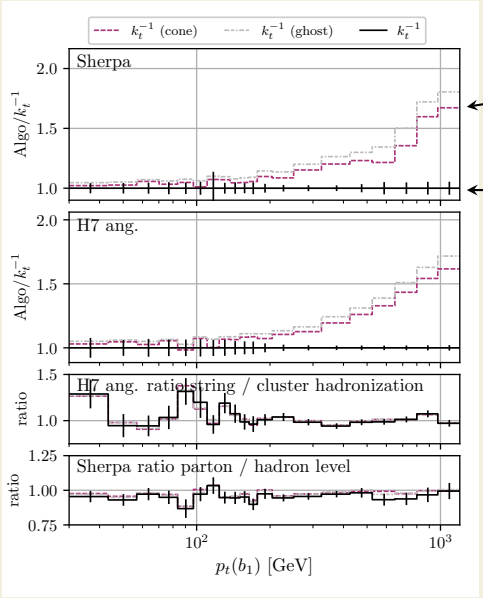
[1] *Max-Planck-Institut für Physik, Boltzmannstraße 8, 85748 Garching, Germany*
[2] *Institute for Theoretical Physics, ETH, CH-8093 Zürich, Switzerland*
[3] *Dipartimento di Fisica, Università degli Studi di Roma Tor Vergata and INFN, Sezione di Roma Tor Vergata, Via della Ricerca Scientifica 1, Roma 00133, Italy*
[4] *Department of Physics, University of Oxford, Keble Road, Oxford, OX1 3RH, UK*
[5] *Rudolf Peierls Centre for Theoretical Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK*
[6] *Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany*
[7] *CERN, 1211 Geneva 23, Switzerland*
[8] *Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA*
[9] *Dipartimento di Fisica, Università di Genova and INFN, Sezione di Genova, Via Dodecaneso 33, 16146, Italy*
[10] *Università degli Studi di Milano-Bicocca & INFN, Piazza della Scienza 3, Milano 20126, Italy*
[11] *Institute of Nuclear Physics, ul. Radzikowskiego 152, 31-342 Kraków, Poland*
[12] *Institute for Particle Physics Phenomenology, Durham University, Durham DH1 3LE, UK*
[13] *All Souls College, Oxford OX1 4AL, UK*
[14] *School of Physics and Astronomy, Monash University, Wellington Rd, Clayton VIC-3800, Australia*
[15] *Jagiellonian University, ul. prof. Stanisława Łojasiewicza 11, 30-348 Kraków, Poland*
[16] *University of Michigan, Ann Arbor, MI 48109, USA*

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

Selected findings

Impact of changing flavour recombination schemes

$pp \to Z + b$ jet in central kinematics



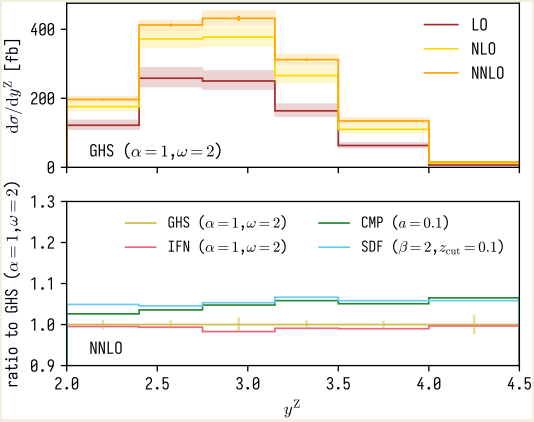
jet contents / scheme	b	$b+\bar{b}$	$b+b$	
"any flavour"	b	b	b	simplest experimentally (but collinear unsafe for $m_b \to 0$)
net flavour	b	g	$2b$	theoretically "ideal" definition; but not robust wrt B–Bbar oscillations
flavour modulo 2	b	g	g	theoretically OK; robust wrt B–Bbar oscillations

Table: [Salam '24]

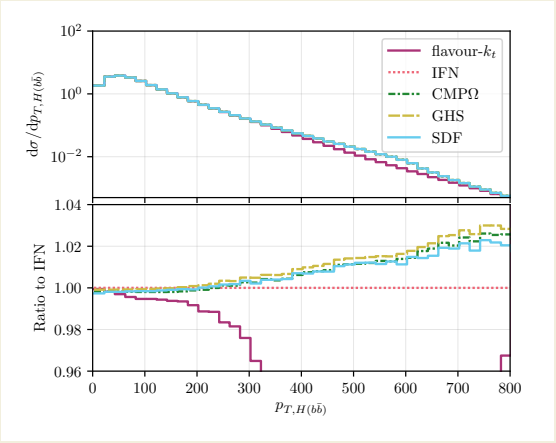
- Biggest impact from change of flavour recombination scheme
- Illustration for $pp \to Z + b$ jet
  Top two panels compare
  - anti-k_t with cone matching (any flavour)
  - anti-k_t with ghost matching (any flavour)
  - anti-k_t with mod2 flavour
- Hadronisation does not play a major role

Fixed-order observables

$pp \to Z + c$ jet in LHCb kinematics



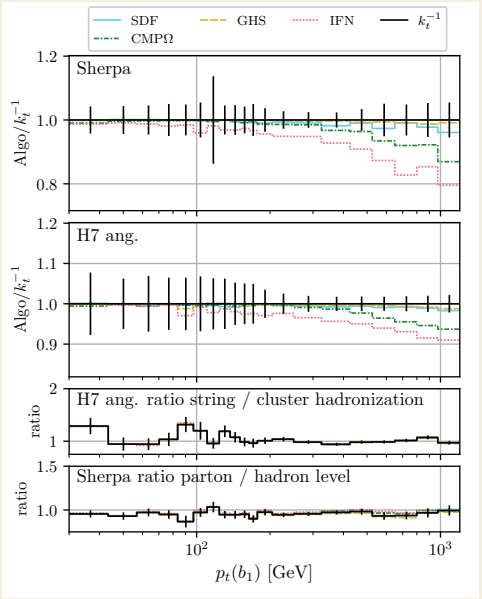
$pp \to WH(\to b\bar{b})$



- Effects of new algorithms in fixed-order calculations are fairly minor (few %)
- Consistent with low flavour multiplicity in final states

NLO+PS: Z+b/c jets

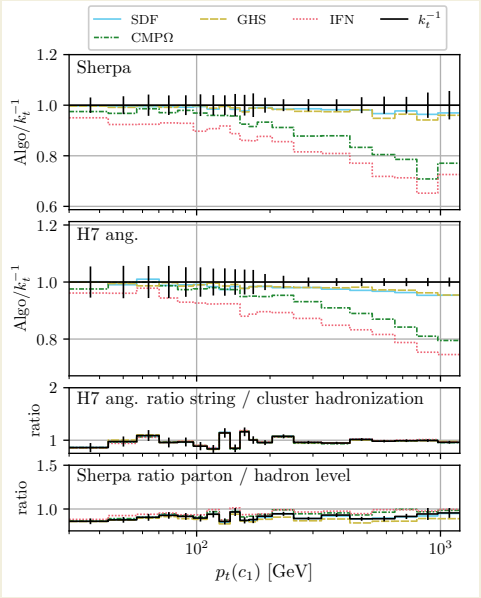
$pp \to Z + b$ jet in central kinematics



- Including parton shower exacerbates differences between algorithms
- Differences reach $\sim 20\%$ in high-p_t tails
- Going from b to c quarks further increases differences
- IFN algorithm is most “aggressive” in eliminating flavour labels

NLO+PS: Z+b/c jets

$pp \to Z + c$ jet in central kinematics

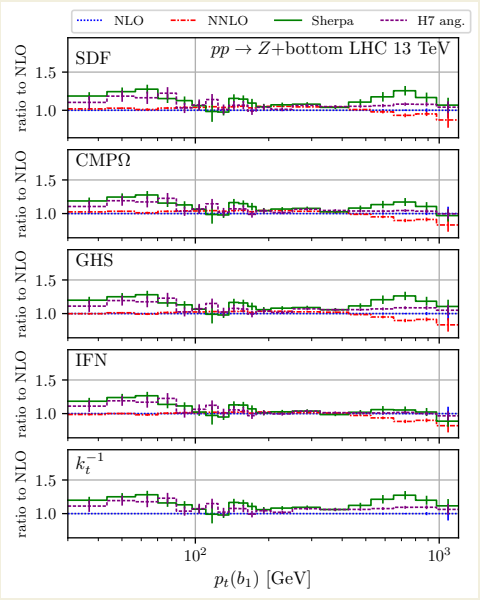


- Including parton shower exacerbates differences between algorithms
- Differences reach $\sim 20\%$ in high-p_t tails
- Going from b to c quarks further increases differences
- IFN algorithm is most “aggressive” in eliminating flavour labels

Comparing fixed-order and parton shower matched calculations

- Differences between FO and PS in threshold and tail regions
- Threshold region: related to matching and shower scale dependence (also occur in flavour-blind calculations for Z + jet)
- High-p_t tail: Lots of flavour produced by parton shower
- IFN appears to be particularly stable against such effects

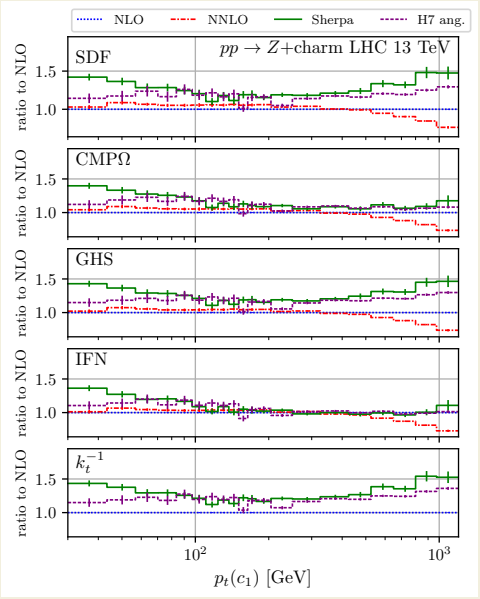
$pp \to Z + b$ jet in central kinematics



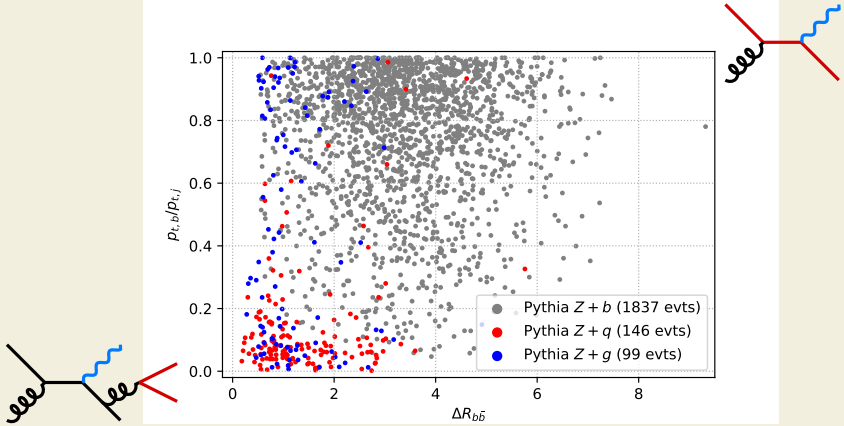
Comparing fixed-order and parton shower matched calculations

- Differences between FO and PS in threshold and tail regions
- Threshold region: related to matching and shower scale dependence (also occur in flavour-blind calculations for $Z +$ jet)
- High-p_t tail: Lots of flavour produced by parton shower
- IFN appears to be particularly stable against such effects

$pp \to Z + c$ jet in central kinematics

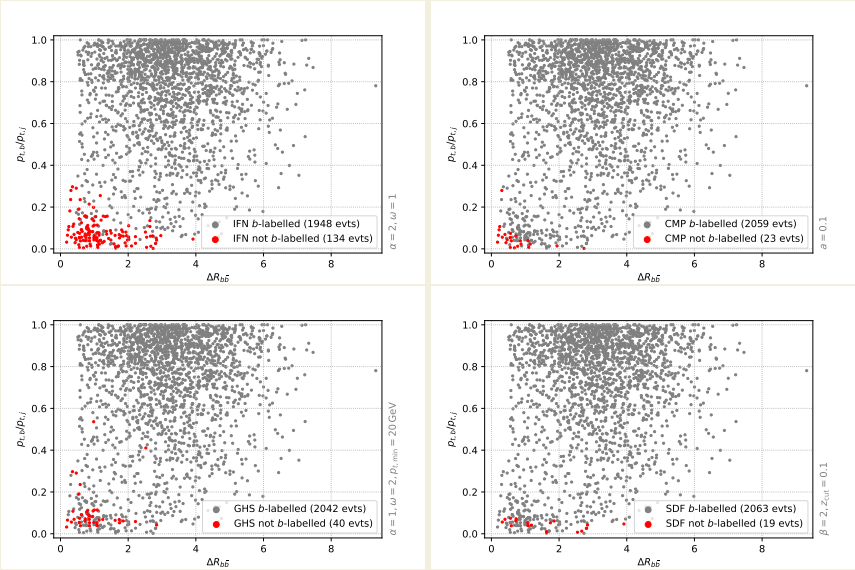


Investigate effectiveness of $g \to b\bar{b}$ rejection



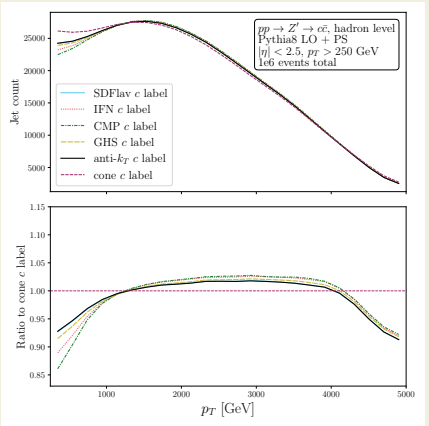
- Pythia LO+PS simulation of $pp \to Z +$ jet events
- Select events with at least one b-labelled jet (via anti-k_t with net flavour)
- Rough expectation:
  - Red dots correspond to events with $g \to b\bar{b}$ splittings
  - Grey dots correspond to events with "hard b jet"

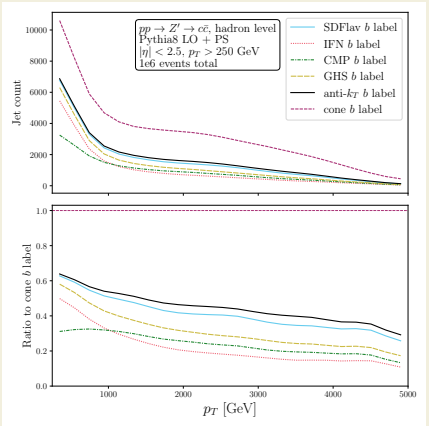
Investigate effectiveness of $g \to b\bar{b}$ rejection



Red: Events labelled by anti-k_t, but not by {IFN, CMP, GHS, SDF}

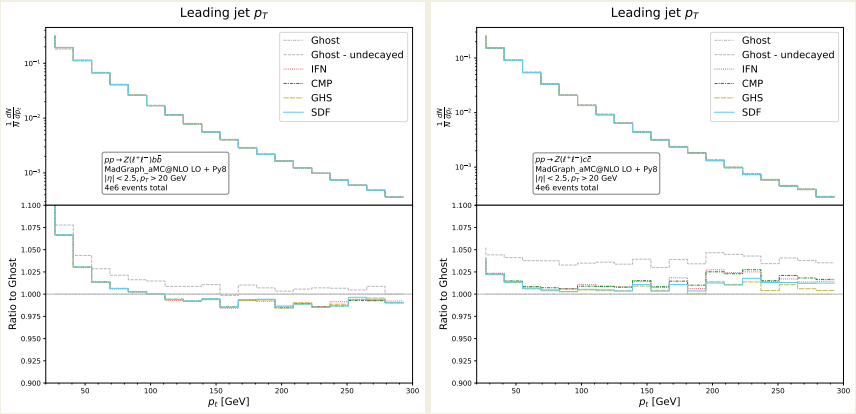
Impact on ATLAS heavy flavour tagging training samples





- Pythia LO+PS simulation of $pp \to Z' \to c\bar{c}$ events ($m_{Z'} = 4\,\text{TeV}$)
- Similar to setup used for ATLAS flavour tagger training
- $g \to c\bar{c}$ and $g \to b\bar{b}$ splittings can have sizeable effect on training sample

Including decays of heavy-flavour hadrons



- Most flavoured jet algorithms need undecayed hadrons as inputs
- Decayed vs. undecayed inputs have a percent level effect on $pp \to Z + b\bar{b}(c\bar{c})$

Exclusive labelling/tagging with Winner Take All (WTA) flavour

- Particle reconstruction abilities of LHCb and ALICE allow for different approach
- Fully reconstruct exclusive decay mode of a flavoured hadron
- Use flavour definition based on WTA axis [Caletti, Larkoski, Marzani, Reichelt '22]
  - **Recluster** into angularly-ordered tree

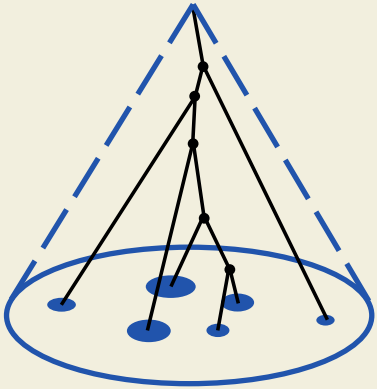


Diagram: [Ezra's LHC EW WG talk '25]

Exclusive labelling/tagging with Winner Take All (WTA) flavour

- Particle reconstruction abilities of LHCb and ALICE allow for different approach
- Fully reconstruct exclusive decay mode of a flavoured hadron
- Use flavour definition based on WTA axis [Caletti, Larkoski, Marzani, Reichelt '22]
  - **Recluster** into angularly-ordered tree
  - **Follow** the hardest branch

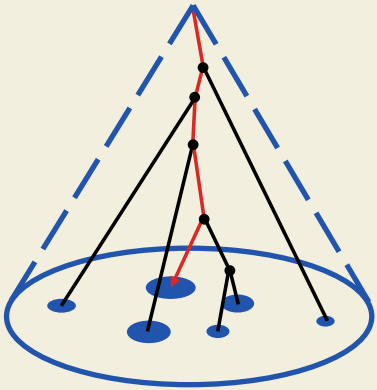


Diagram: [Ezra's LHC EW WG talk '25]

Exclusive labelling/tagging with Winner Take All (WTA) flavour

- Particle reconstruction abilities of LHCb and ALICE allow for different approach
- Fully reconstruct exclusive decay mode of a flavoured hadron
- Use flavour definition based on WTA axis [Caletti, Larkoski, Marzani, Reichelt '22]
  - **Recluster** into angularly-ordered tree
  - **Follow** the hardest branch
  - **Define** flavour by particle along the axis

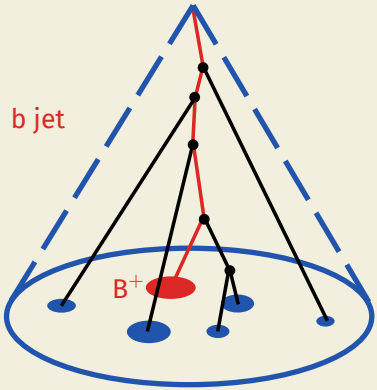
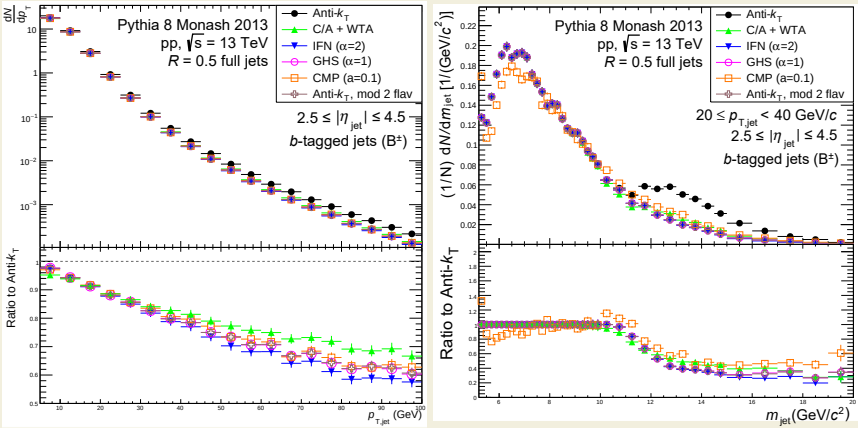


Diagram: [Ezra's LHC EW WG talk '25]

WTA flavour in an LHCb environment

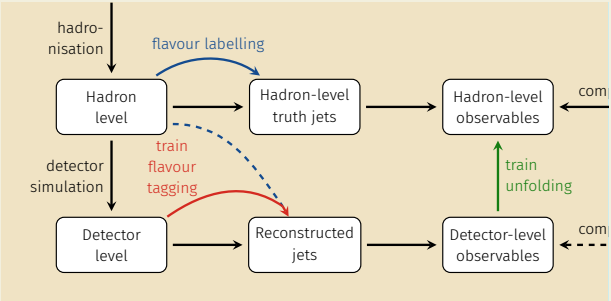


- WTA flavour behaves similar to new algorithms
- Significant contribution from $g \to b\bar{b}$ splitting
- Jet mass: second peak at $m_j \sim 2m_B$
  $\to$ suppressed by flavour recombination schemes

Conclusions

- New IRC safe flavour jet algorithms are available
- We investigate their use as flavour labelling strategies
- Implementations as `fjcontrib` plugins for `FastJet` are available
- Biggest effects come from change of flavour recombination scheme
- Algorithms largely yield comparable results
- Scenarios with large amount of flavoured particles can bring out differences

Outlook



- Investigate impact of hadronisation?
- Deeper look into jet substructure?
- Investigate interaction with unfolding and flavour tagging?
- New algorithms as label providers for tagger training?
- Interaction with $g \to f\bar{f}$ modelling?