Jet substructure and fragmentation at the LHC

Cristian Baldenegro (Laboratoire Leprince-Ringuet)
on behalf of the LHC experiments

SM@LHC 2023 at Fermilab
July 10th – July 14th

European Research Council

Jet constituents momenta are mapped onto physically transparent observables

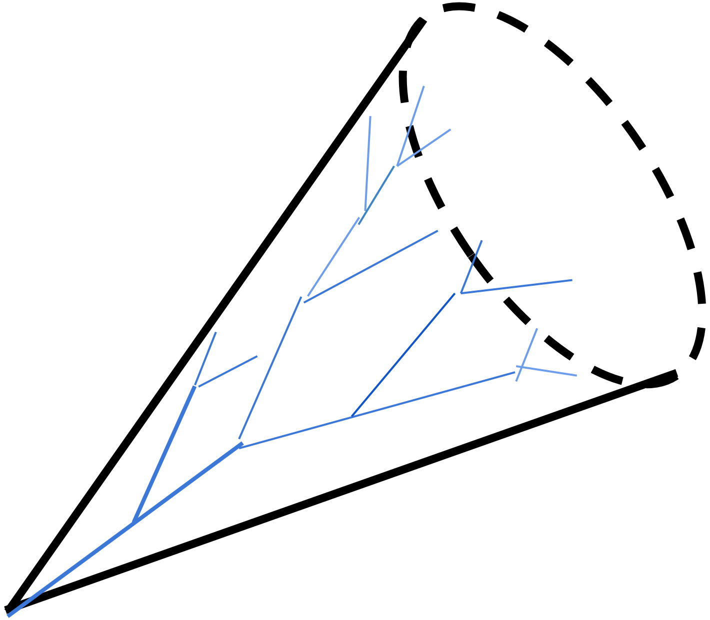
$$\{p_i\} \rightarrow \lambda$$

Fragmentation Functions | **Classic Jet Shapes** | **Groomed Observables** | **Iterative declustering**

Single hadron | *All hadrons* | *Subset of hadrons* | *Organizing hadrons*

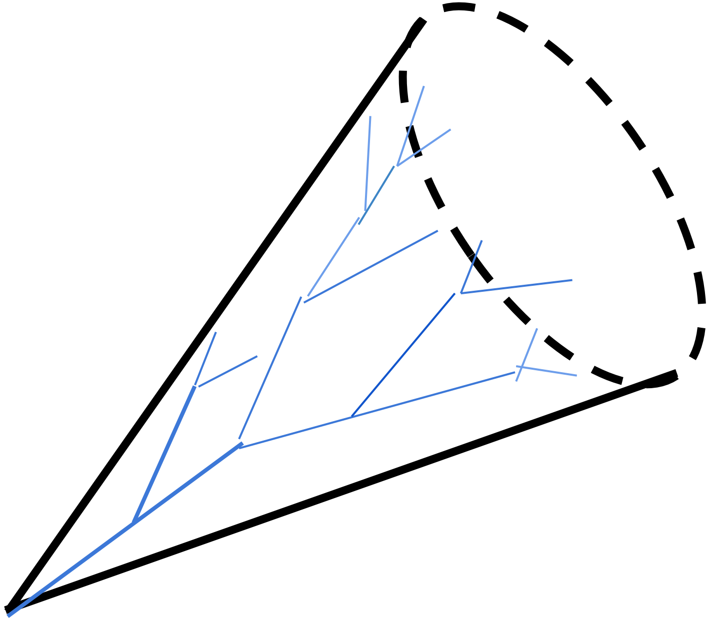
(slightly modified) sketch from Jesse Thaler

Several tools that give us access to the internal dynamics of the jet



Jet substructure can be used to:

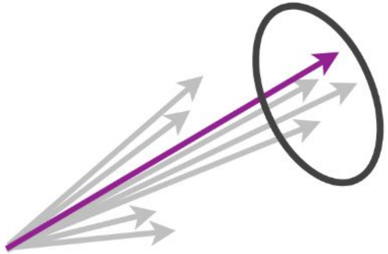
- Stress test precision “pen-and-paper” calculations (**resummation**) and MC generators (parton showers, hadronization, UE)
- Tag objects (parton flavors, boosted objects)
- mitigate pileup/UE (e.g., via grooming)
- get a space-time picture of the quark-gluon plasma



I’ll focus mostly on recent (pp) measurements:

- Stress test precision “pen-and-paper” calculations (**resummation**) and MC generators (parton showers, hadronization, UE)
- Tag objects (parton flavors, boosted objects)
- mitigate pileup/UE (e.g., via grooming)
- get a space-time picture of the quark-gluon plasma

Fragmentation Functions

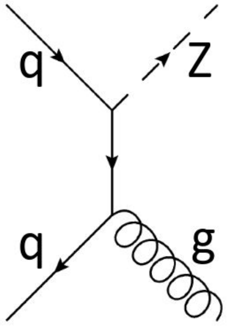


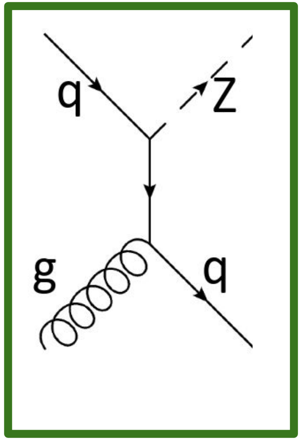
Single hadron

Fragmentation functions in Z-tagged jet events (LHCb)

LHCb designed mainly for flavor physics at the LHC
→ equipped with particle ID from 2 < p < 100 GeV, **which can be used for jet substructure**

Jets in LHCb acceptance are mostly quark jets
(~70% quark jets)





Study two different fragmentation variables

$$z \equiv \frac{\mathbf{p}_{jet} \cdot \mathbf{p}_{hadron}}{|\mathbf{p}_{jet}|^2}$$

longitudinal

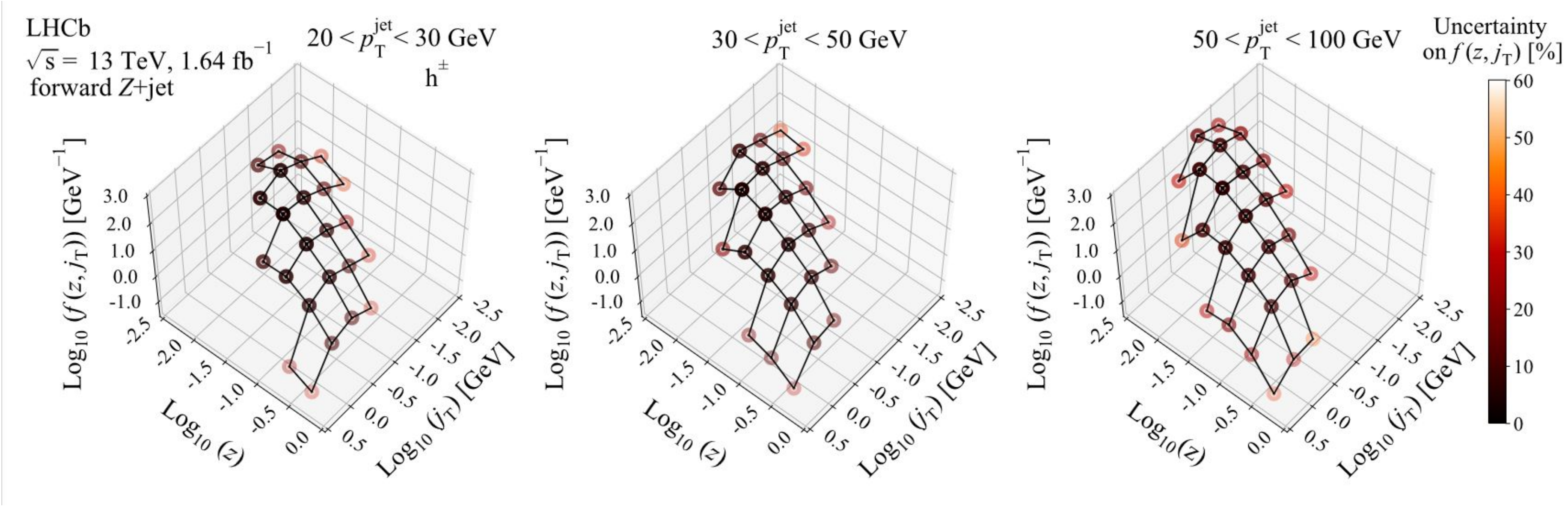
$$j_T \equiv \frac{|\mathbf{p}_{jet} \times \mathbf{p}_{hadron}|}{|\mathbf{p}_{jet}|}$$

transverse

Triple differential distributions in j_T, z & p_T^{jet} for unidentified hadrons

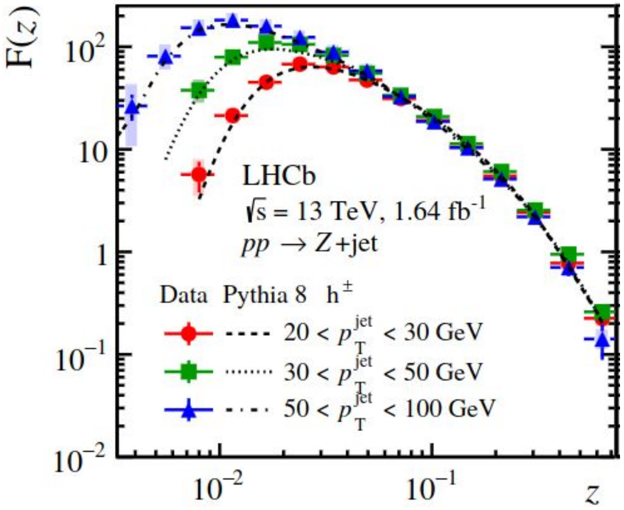
Some kinematical correlations are visible
(harder p_T^{jet} gives access to lower z, higher z gives access to larger j_T)

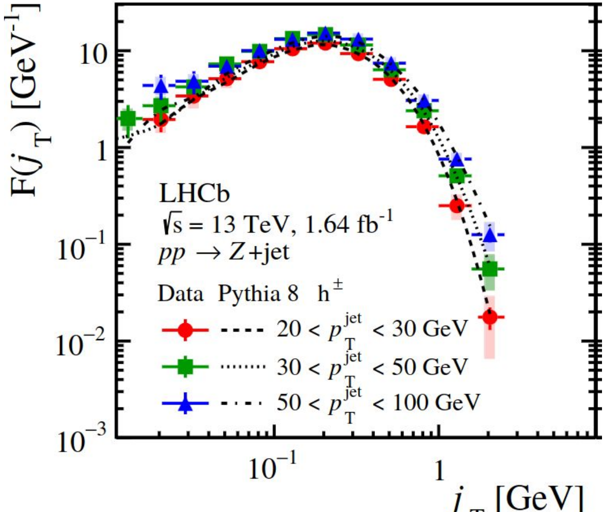
LHCb, arXiv:2208.11691



Comparison for unidentified hadrons in Z-tagged jet events

LHCb, arXiv:2208.11691





Universality between jet p_T bins at high *z*

Higher p_T^{jet} gives access to lower z and higher j_T values

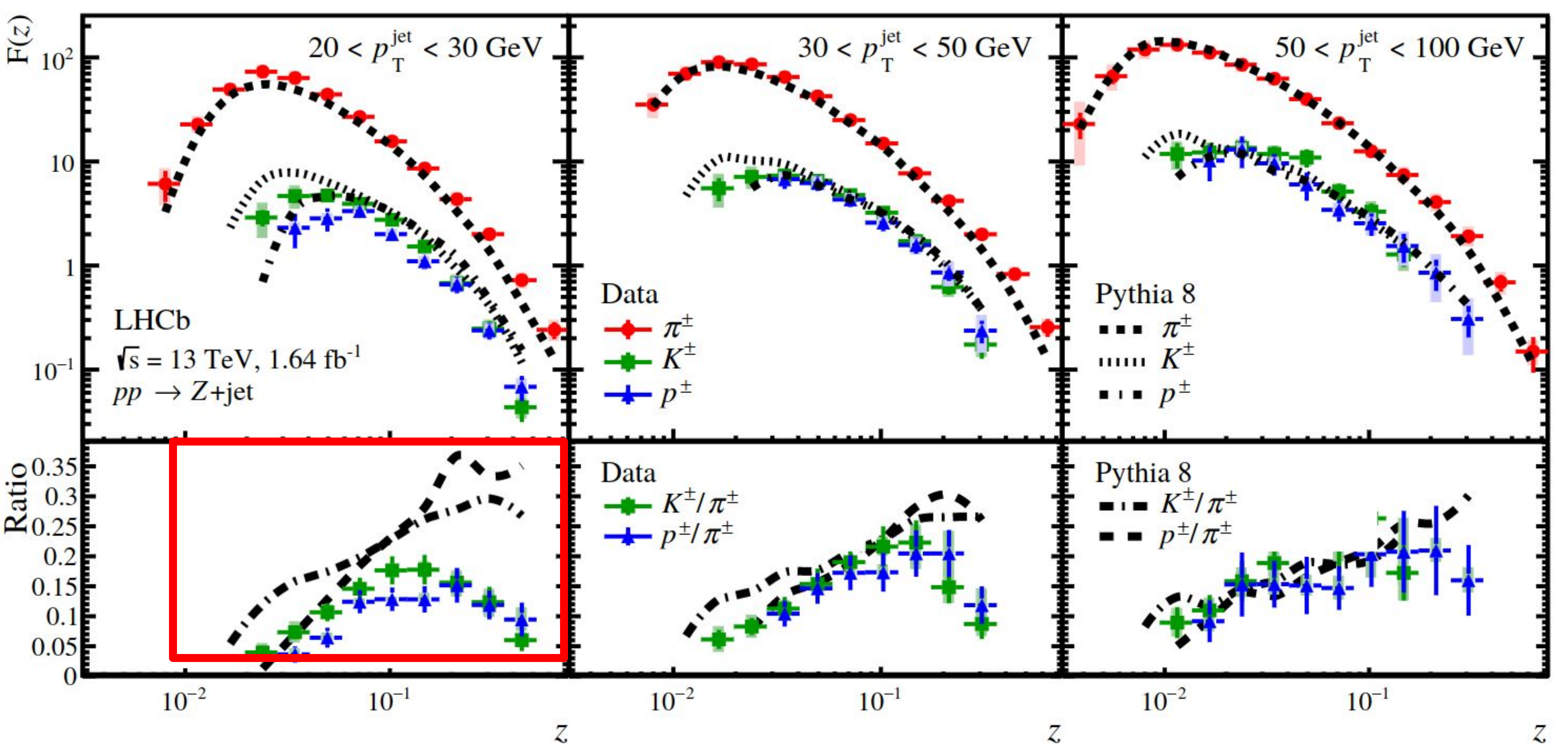
PYTHIA8 generally describes fragmentation function of **unidentified** charged hadrons

Identified charged hadrons ($K^{\pm}$, $p^{\pm}$, $\pi^{\pm}$)

PYTHIA8 predicts larger abundance of hard **$K^{\pm}$** and **$p^{\pm}$** at **low p_T^{jet}** than is seen in data

Complementary information to minimum-bias spectra analyses, can help pin down mechanisms for baryon and meson production

LHCb, arXiv:2208.11691



Fragmentation of b-quarks

- *b*-jet have a distinct fragmentation pattern compared to light-quark jets

  (harder fragmentation + dead cone effect due to mass effects)

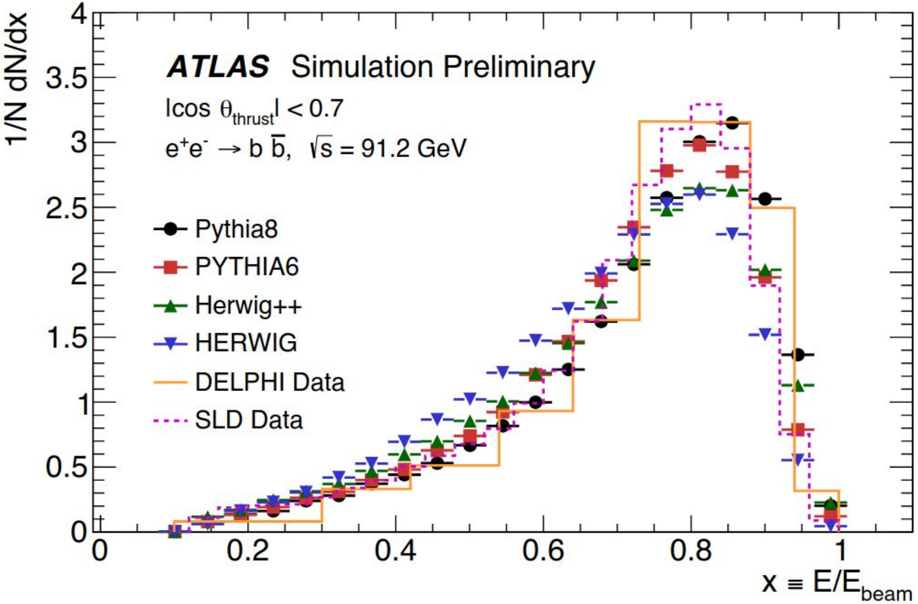
- Important for top quark and and H→ $b\bar{b}$ measurements

- *b*-fragmentation function tuned to e^+e^- data, then used at LHC energies and environment

  To what extent does this work?

ATLAS-PHYS-PUB-2014-008

particle-level e^+e^- simulation



Predicted fragmentation functions in e^+e^- collisions from simulation

Differences of O(10%) between different PS and hadronization models for LEP environment.
What can we do at the LHC?

b-quark fragmentation in top quark pair events

ATLAS, arXiv:2202.13901, PRD 106 (2022) 032008

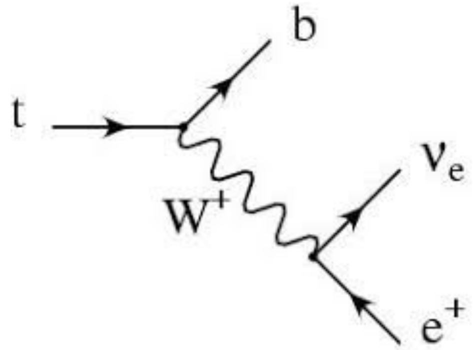
- 1μ 1e golden channel; clean events with primary b-jets

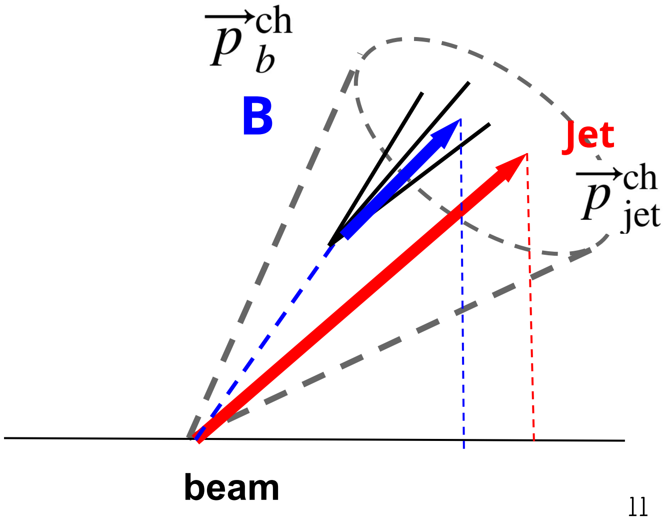
- exactly 2 *b*-jets with p_T>30 GeV, |η|< 2.1, and ΔR(jet, jet) > 0.5, clustered with anti-k_t algorithm with R = 0.4.

- **Proxy for *b*-hadron using charged-particle tracks in secondary vertex**

Reconstruct 3-momentum of “charged” b-hadron $\overrightarrow{p}^{\,\mathrm{ch}}_{b}$

Compare with charged component of jet $\overrightarrow{p}^{\,\mathrm{ch}}_{\mathrm{jet}}$



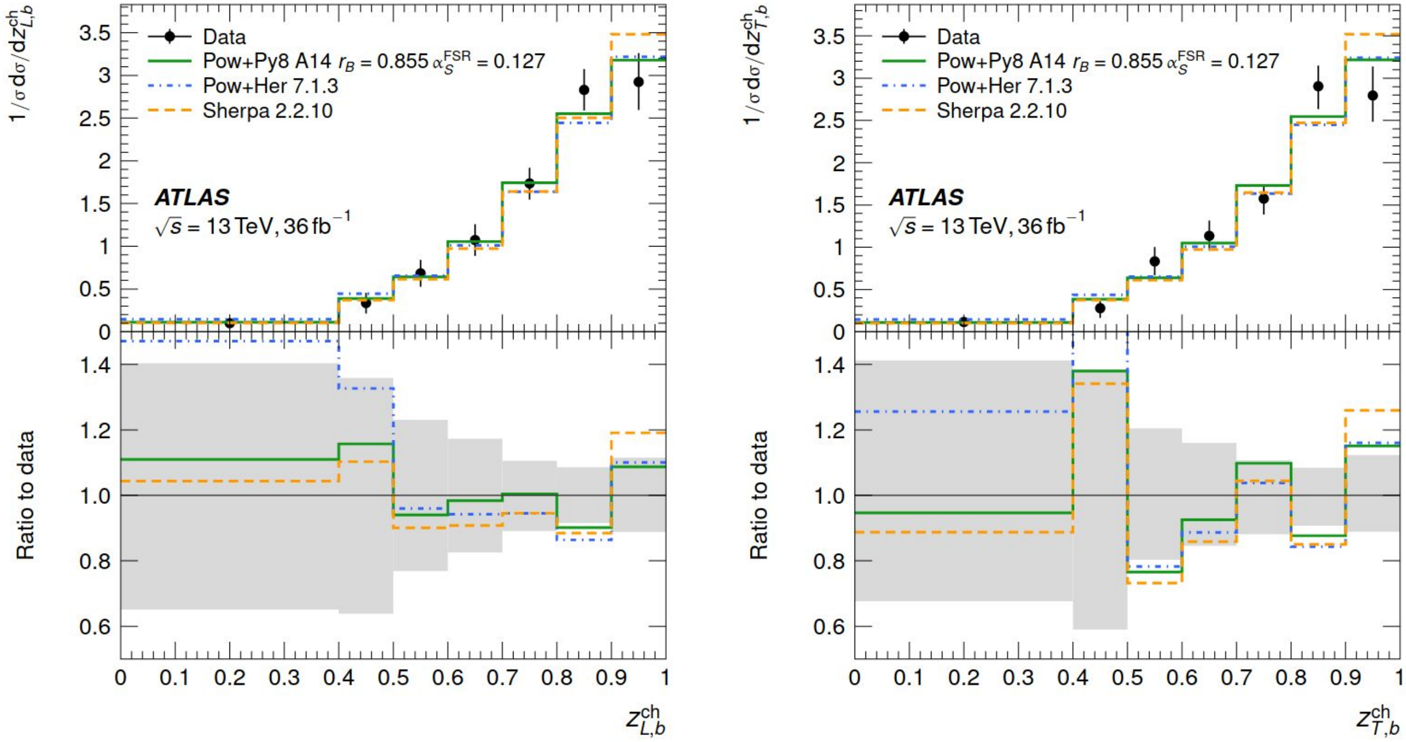


Longitudinal and transverse fragmentation variables

ATLAS, arXiv:2202.13901, PRD 106 (2022) 032008

Corrections to particle-level with fully Bayesian unfolding method

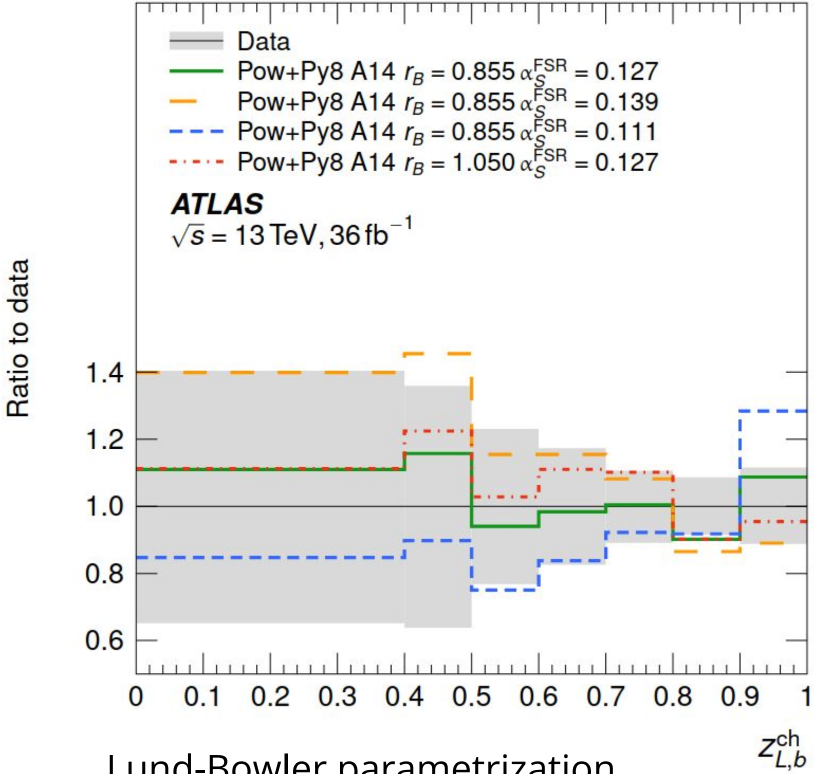
$$z^{\mathrm{ch}}_{\mathrm{L},b} = \frac{\vec{p}_b^{\ \mathrm{ch}} \cdot \vec{p}_{\mathrm{jet}}^{\ \mathrm{ch}}}{|p^{\mathrm{ch}}_{\mathrm{jet}}|^2} \qquad z^{\mathrm{ch}}_{\mathrm{T},b} = \frac{p^{\mathrm{ch}}_{\mathrm{T},b}}{p^{\mathrm{ch}}_{\mathrm{T,jet}}}$$



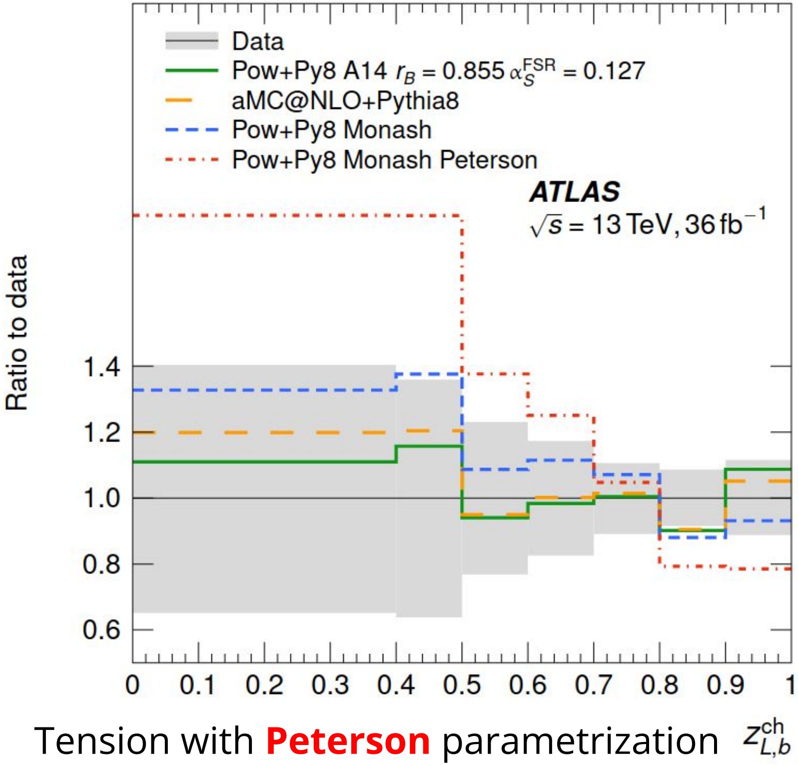
Simulations do a reasonable job in describing fragmentation variables

Fragm. function parametrization and α_s^{FSR}

ATLAS, arXiv:2202.13901
PRD 106 (2022) 032008



Lund-Bowler parametrization

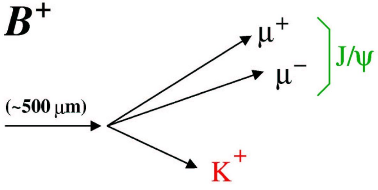


Tension with **Peterson** parametrization (when used with NLO+PS) at low $z_{L,b}^{ch}$

b-quark fragmentation via $B^\pm \rightarrow J/\psi\, K^\pm \rightarrow \mu^+\mu^- K^\pm$ decays

Fully reconstructed $B^\pm$ hadrons in b-jets using exclusive decay channel:

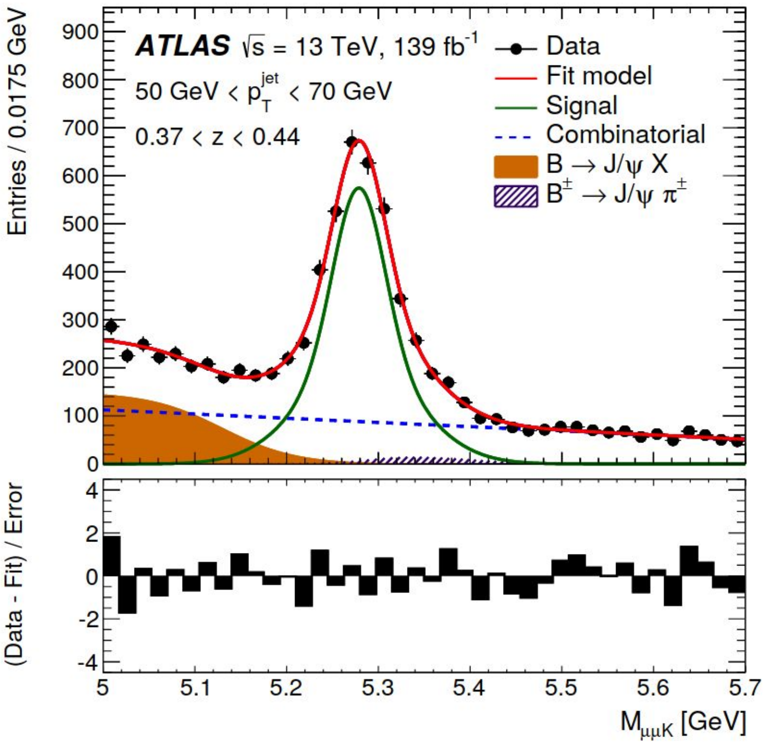
ATLAS, arXiv:2108.11650, *JHEP 12 (2021) 131*

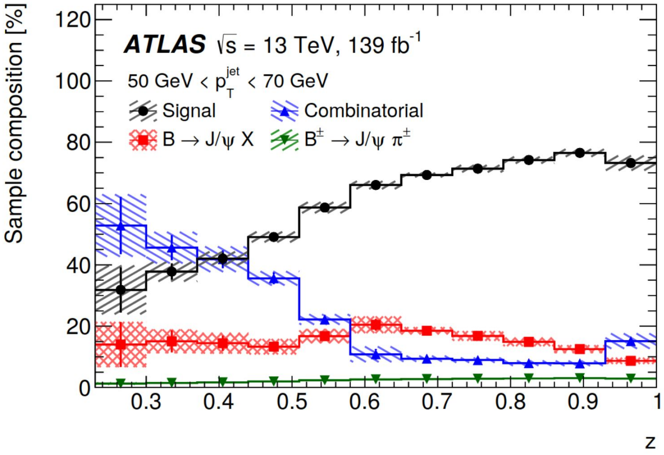


b-jets in **inclusive jet** sample (flavor excitation + flavor creation + gluon splitting) using R = 0.4 anti-k_T jets, $p_T > 35$ GeV and $|\eta| < 2.1$

Two fragmentation variables are analyzed (showing *z* in this presentation)

$$z = \frac{\vec{p}_B \cdot \vec{p}_j}{|\vec{p}_j|^2}; \quad p_{\mathrm{T}}^{\mathrm{rel}} = \frac{|\vec{p}_B \times \vec{p}_j|}{|\vec{p}_j|},$$





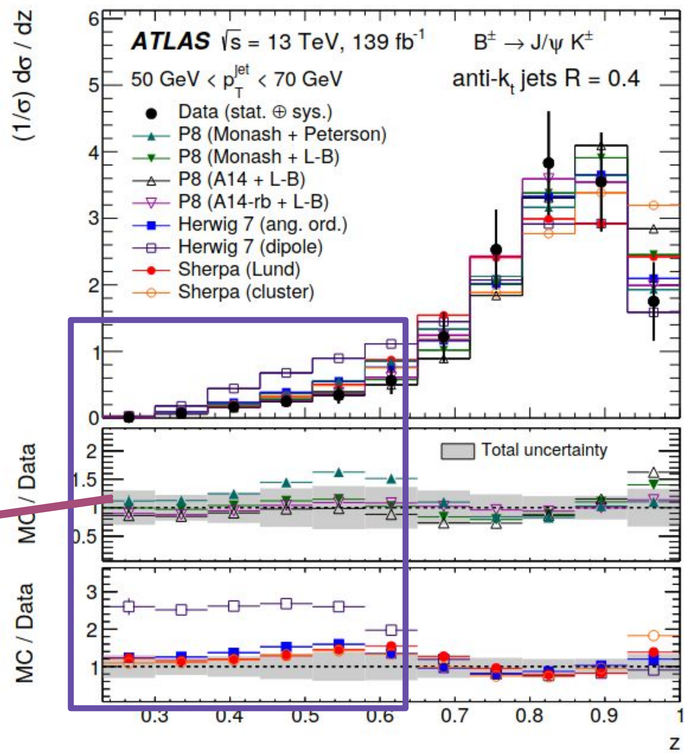
Fragmentation variable $z = \frac{\vec{p}_B \cdot \vec{p}_j}{|\vec{p}_j|^2}$

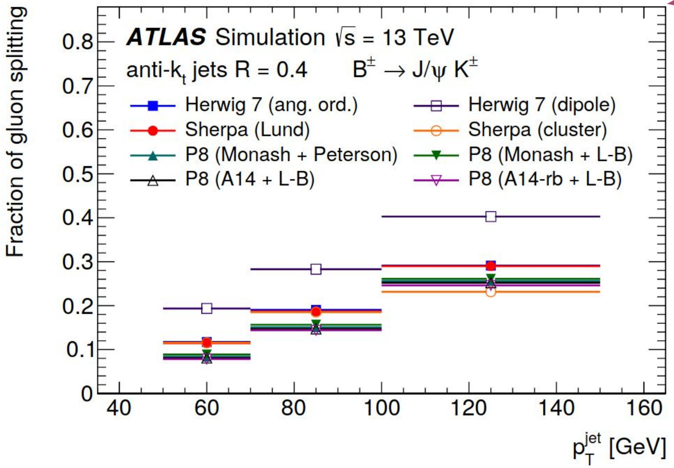
ATLAS, *arXiv:2108.11650*, *JHEP 12 (2021) 131*

Corrected to particle-level with D'Agostini unfolding

PYTHIA8 tunes and fragmentation function tested are consistent with data within the uncertainties

Differences between **H7 angle-ordered** & **H7 dipole** at low-*z* *(z < 0.5),* correlated to g→ bb modeling





Main uncertainties related to mass fits, unfolding model, jet energy calibration.

energy flow,
jet shapes

Energy-energy correlators for jet substructure

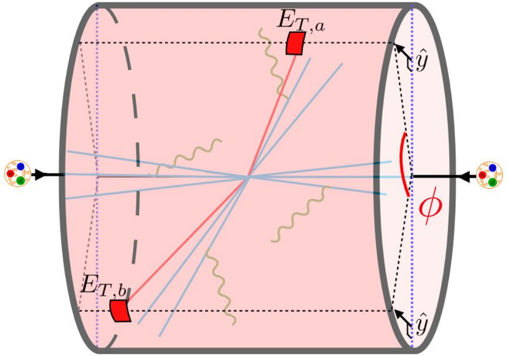
A. Larkoski, G. Salam, J. Thaler,
JHEP06(2013)108

energy-weighted cross section

$$\frac{\mathrm{d}\sigma_{\mathrm{EEC}}}{\mathrm{d}R_{\mathrm{L}}} = \sum_{i,j} \int d\sigma(R'_{\mathrm{L}}) \, \frac{p_{\mathrm{T},i}\, p_{\mathrm{T},j}}{p^2_{\mathrm{T,jet}}} \, \delta(R'_{\mathrm{L}} - R_{\mathrm{L},ij})$$

$$R_{\mathrm{L}} = \sqrt{\Delta\varphi^2_{ij} + \Delta\eta^2_{ij}}$$

Observable connected to conformal field theory approaches



sketch from Ian moult

Energy-energy correlators for jet substructure

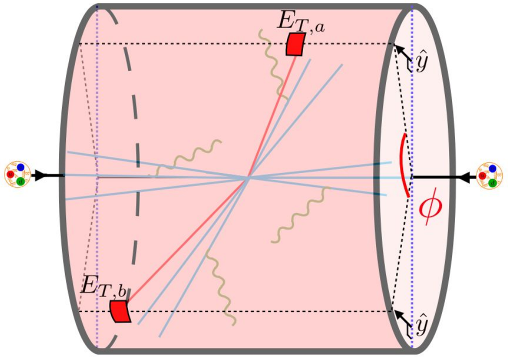
A. Larkoski, G. Salam, J. Thaler,
JHEP06(2013)108

energy-weighted cross section

$$\frac{d\sigma_{\mathrm{EEC}}}{dR_{\mathrm{L}}} = \sum_{i,j} \int d\sigma(R'_{\mathrm{L}}) \, \frac{p_{\mathrm{T},i}\, p_{\mathrm{T},j}}{p^2_{\mathrm{T,jet}}} \, \delta(R'_{\mathrm{L}} - R_{\mathrm{L},ij})$$

$$R_{\mathrm{L}} = \sqrt{\Delta\varphi_{ij}^2 + \Delta\eta_{ij}^2}$$

Observable connected to conformal field theory approaches



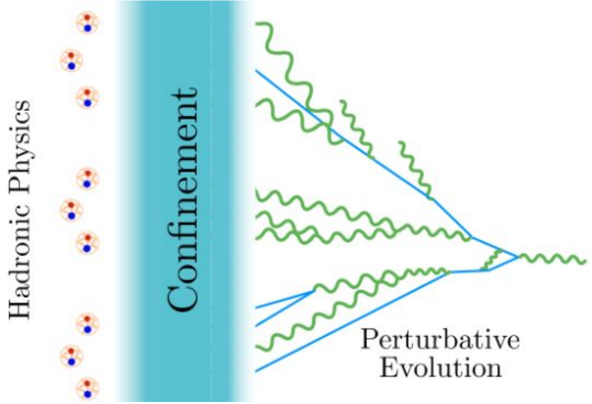
sketch from Ian moult

Soft particle pairs are “penalized” with small energy weights (typically at small R_L)

Hard radiation is “rewarded” with larger weights (typically at large R_L)

No jet grooming to suppress soft physics is required

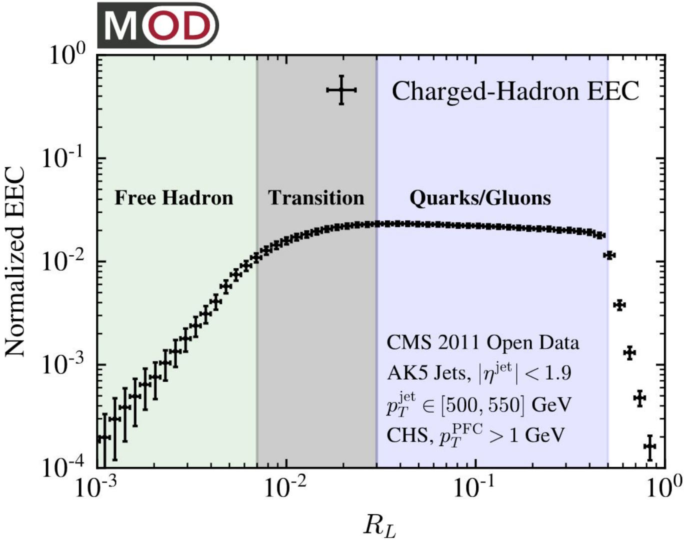
$R_L << \Lambda_{QCD}/p_{T,jet}$ $R_L >> \Lambda_{QCD}/p_{T,jet}$



$$\frac{d\sigma_{\rm EEC}}{dR_{\rm L}} = \sum_{i,j} \int d\sigma(R'_{\rm L})\, \frac{p_{{\rm T},i}\, p_{{\rm T},j}}{p^2_{\rm T,jet}}\, \delta(R'_{\rm L} - R_{{\rm L},ij})$$

P. Komiske, I. Moult, J. Thaler, H.X. Zhu, PRL 130, 051901

How to measure these experimentally?



1. For a given pair of jet constituents, fill a histogram with weight = $p_{T,i}\ p_{T,j}/p^2_{T,\text{jet}}$ at entry $R_L = \Delta R_{ij}$

2. Iterate step 1 for all possible pairs in the jet (there will be multiple histogram entries per jet)

3. Do this for all jets, and you obtain an energy-weighted two-particle correlation distribution

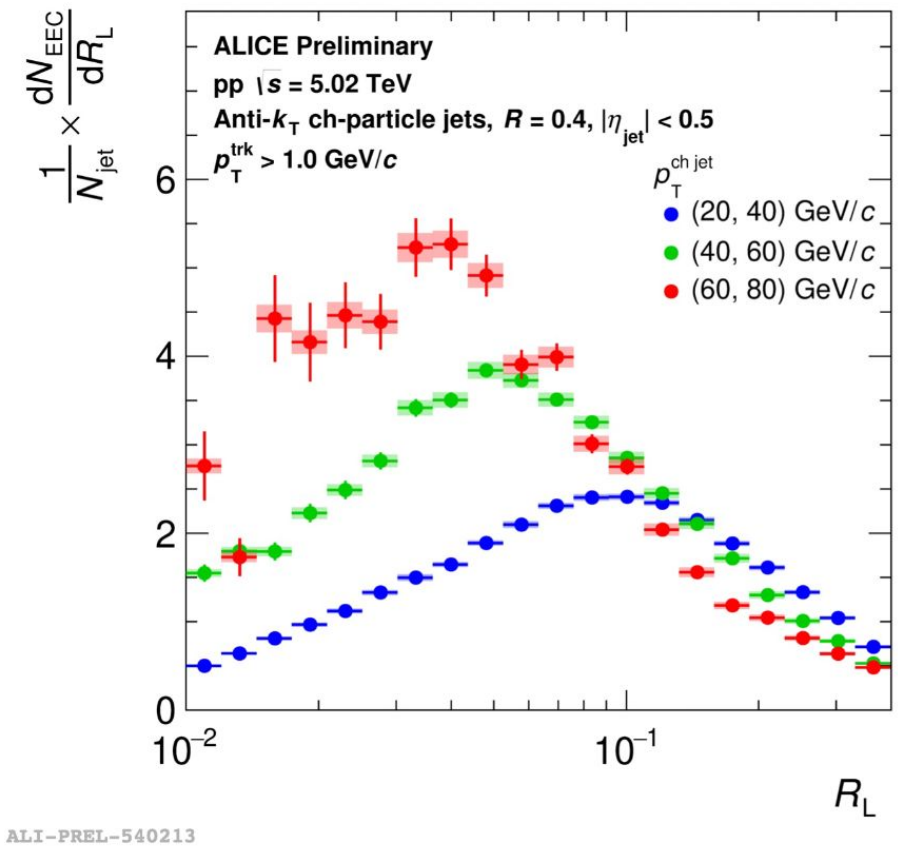
Proof of concept using CMS OpenData

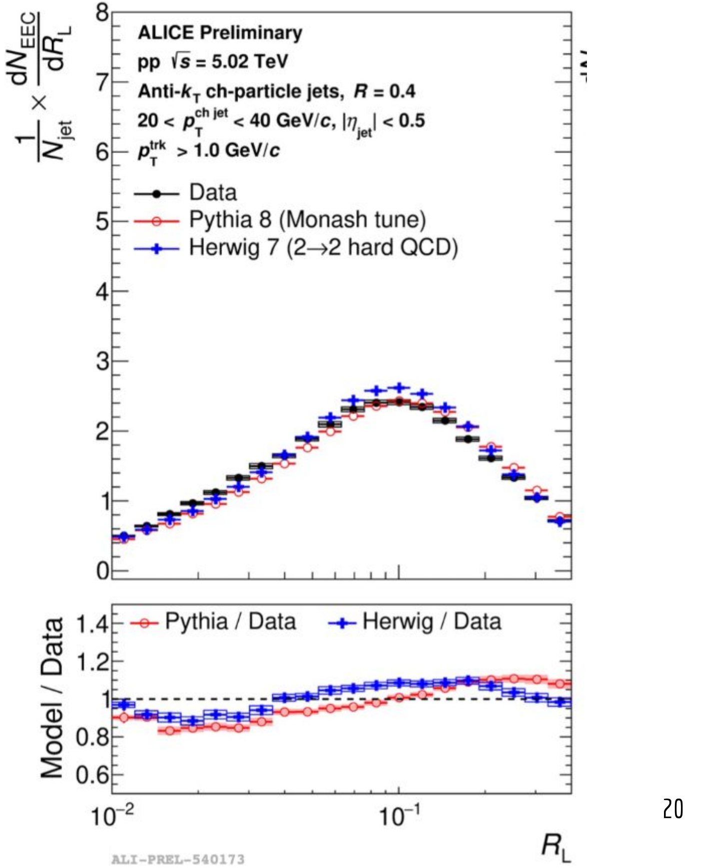
Access to scaling properties of QCD

Energy-energy correlators measurement (ALICE)

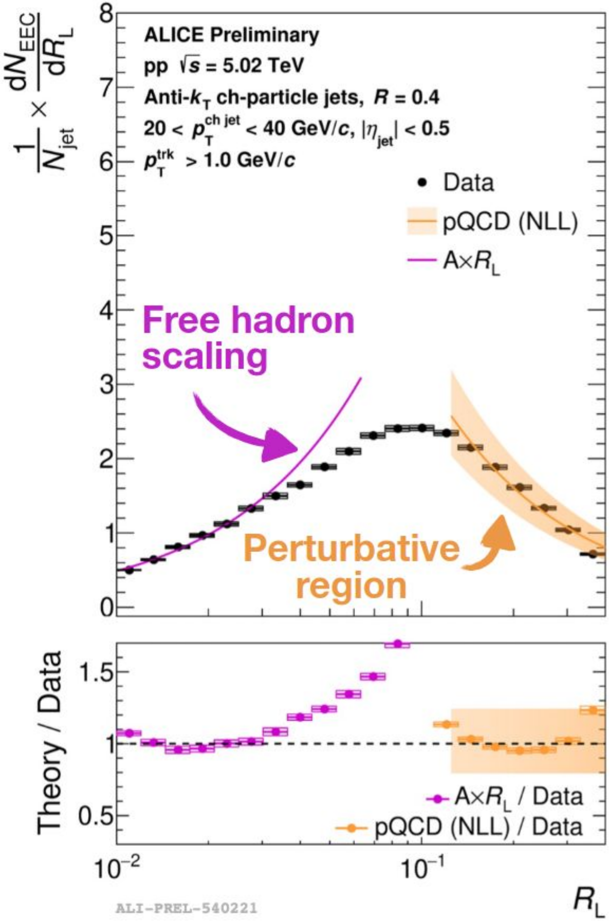
Bin-by-bin corrections to particle-level, dominated by physics model (H7 vs P8), syst. unc. at the 4%-level

PY8/**H7** undershoot data at low R_L (nonperturb.)
H7 describes data better (perturbative)





Comparison to analytical calculations



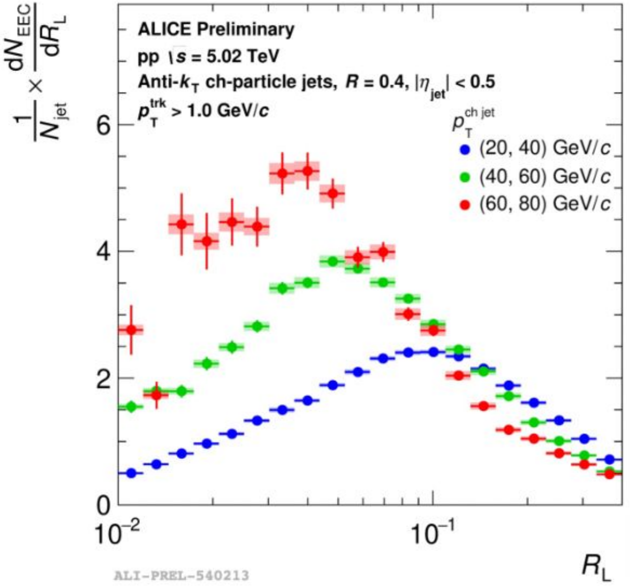
Based on K. Lee, B. Meçaj, I. Moult arXiv:2205.03414

pQCD NLL calculations correspond to neutral+charged jets & are normalized to data in **perturbative region (large R_L)**

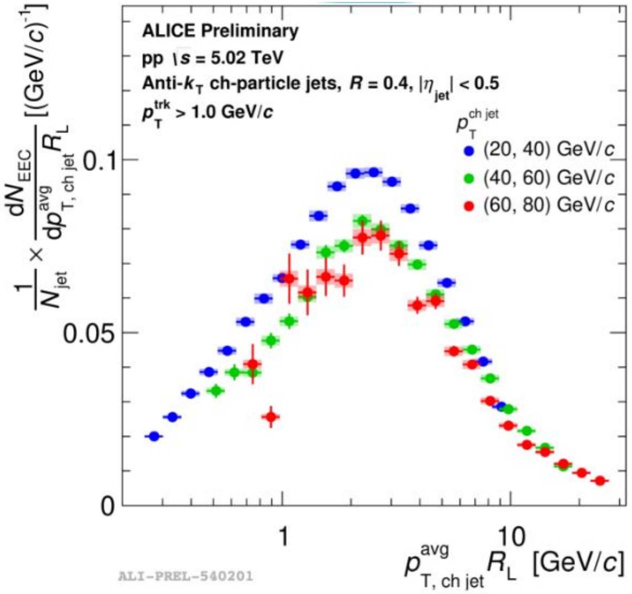
At small R_L, free hadron scaling is followed, deviation due to transition to short-distance physics

scaling properties

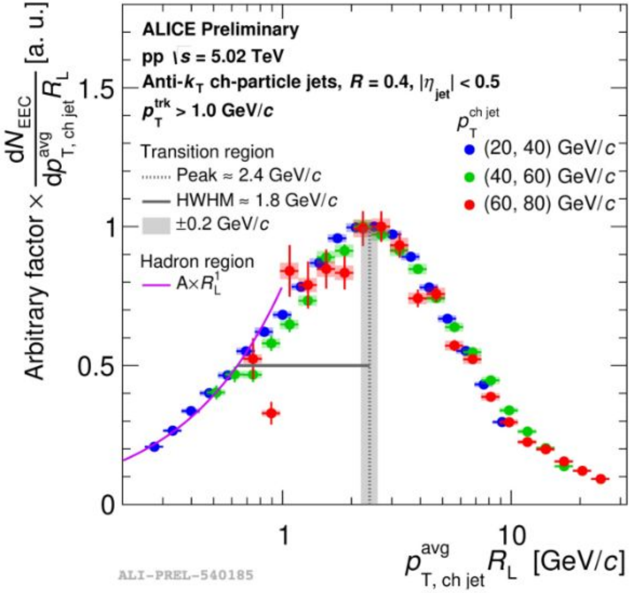
EECs as function of R_L



scale $R_L \rightarrow p^{jet}_T R_L$



Normalize to their integrals
scaling behavior observed
(within uncertainties)

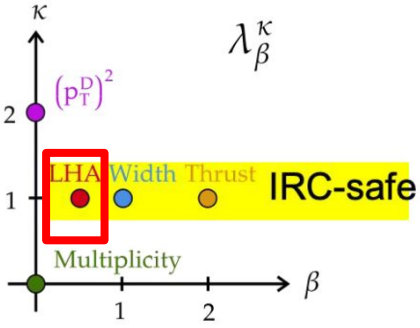


Generalized angularities in dijet and Z+jet events

CMS, *arXiv:2109.03340*,
JHEP 01 (2022) 188

$$\lambda_\beta^\kappa = \sum_{i\in \mathrm{jet}} z_i^\kappa \left(\frac{\Delta R_i}{R}\right)^\beta \quad z_i \equiv \frac{p_{Ti}}{\sum_{j\in \mathrm{jet}} p_{Tj}}$$

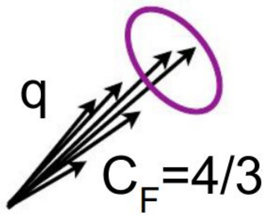
κ & β are parameters set by user



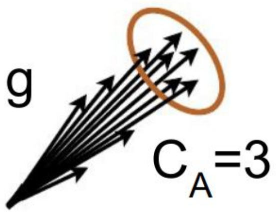
Will show a specific angularity (LHA)

Sensitive to quark vs gluon differences
(subset of them are IRC-safe)
JHEP 1707 (2017) 091

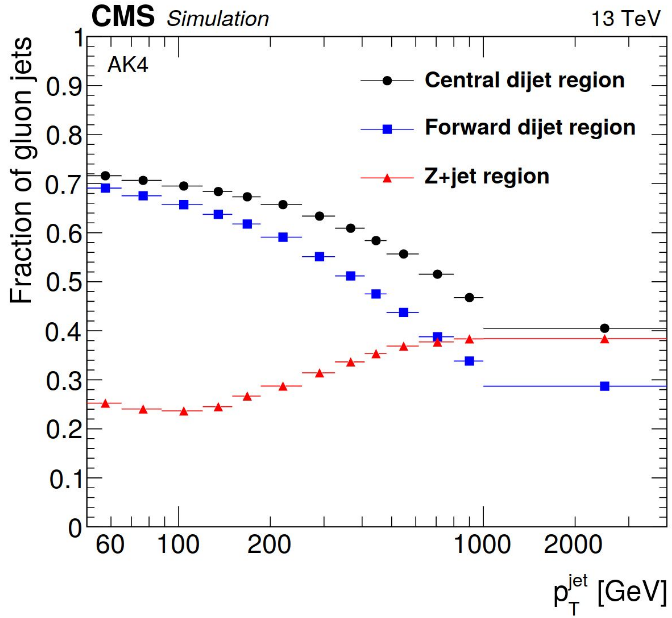
Z+jet (quark-like)



Dijet (gluon-like)

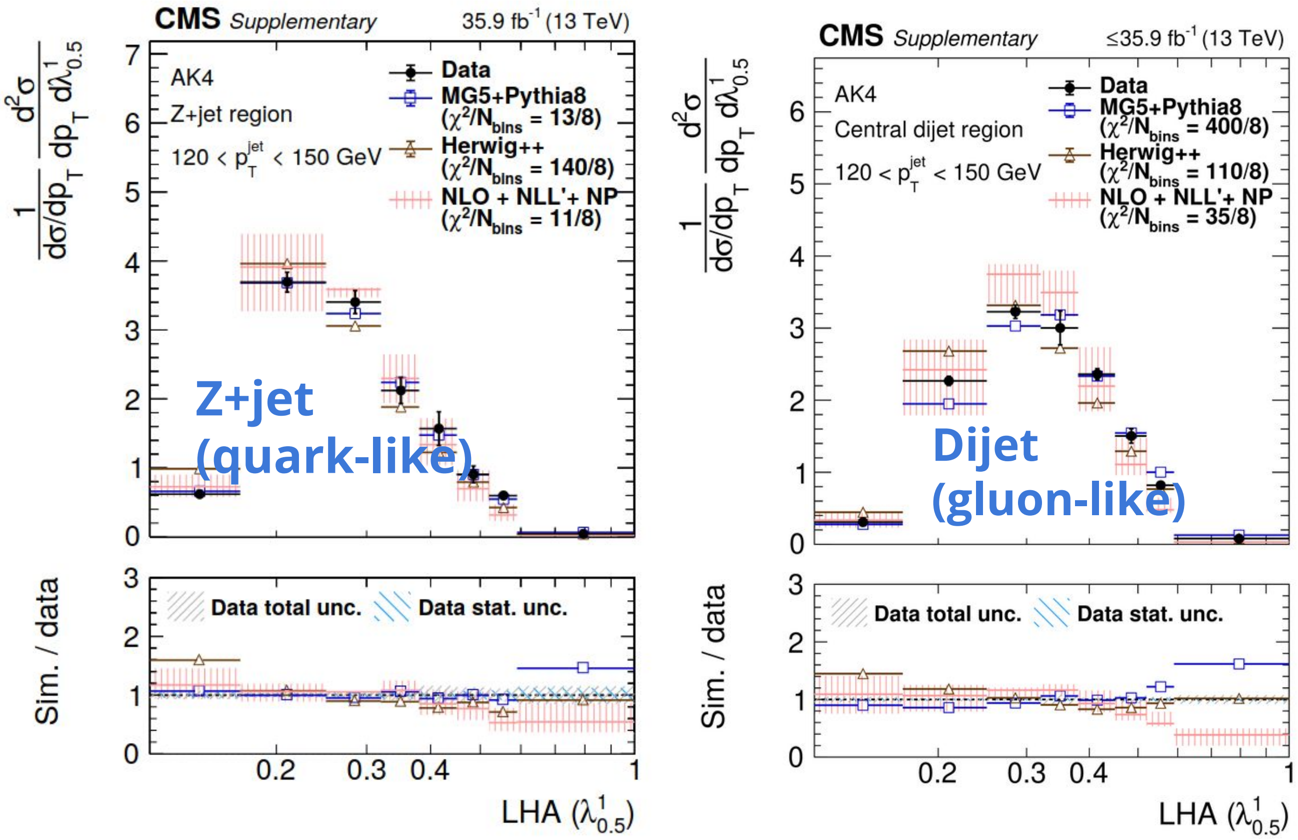


Ungroomed vs groomed with $z_{\mathrm{cut}} = 0.1$, $\beta_{\mathrm{SD}} = 0$,
R = 0.4 vs R =0.8
charged-only vs charged+neutrals



Ungroomed Les Houches Angularity ($\lambda^{1}_{0.5}$) in Z-jet and dijet events

JHEP 01 (2022) 188



Z+jet (quark-like)

Dijet (gluon-like)

Jets in dijets (gluon-like) broader than Z+jets (quark-like)

More challenging to describe **gluon-enriched jets (dijet)**

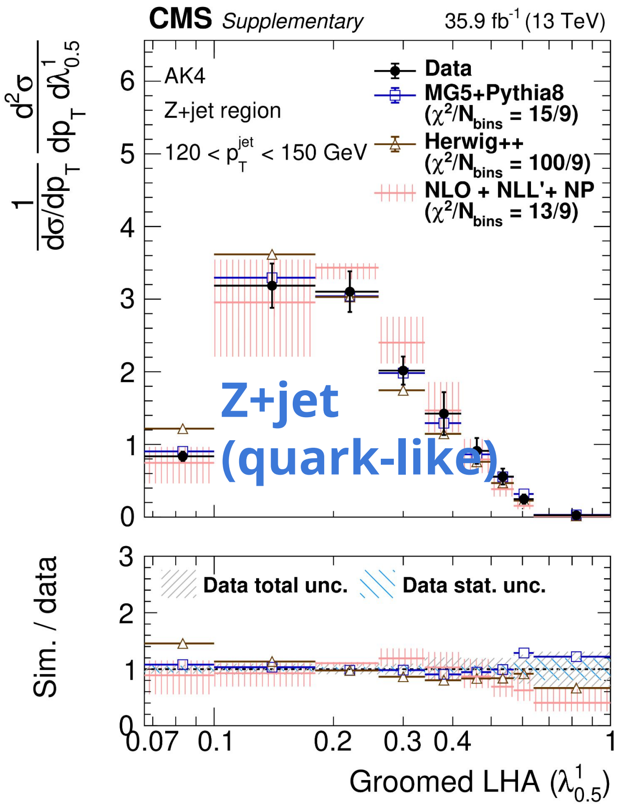
$\kappa = 0.5$, $\beta = 1$

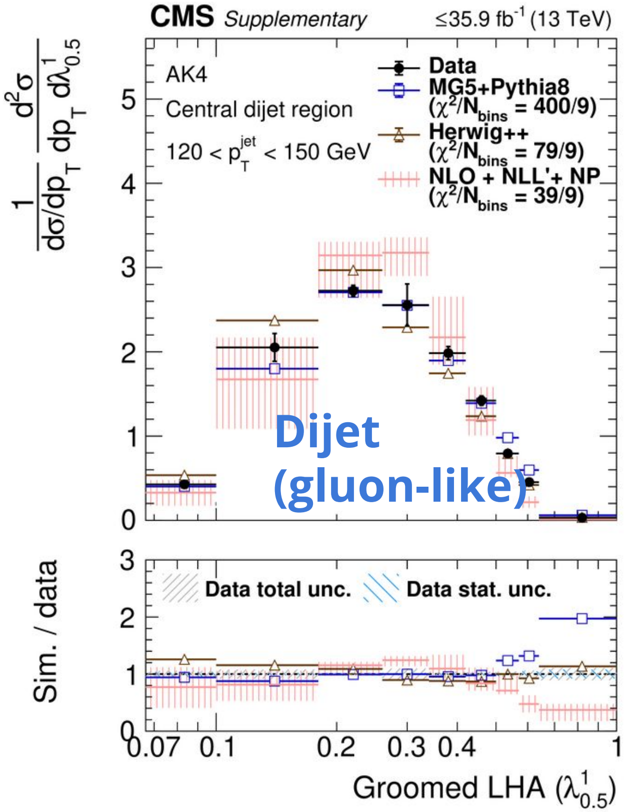
$$\lambda^{\kappa}_{\beta} = \sum_{i\in\mathrm{jet}} z_i^{\kappa}\left(\frac{\Delta R_i}{R}\right)^{\beta} \qquad z_i \equiv \frac{p_{Ti}}{\sum_{j\in\mathrm{jet}} p_{Tj}}$$

pQCD calculations D. Reichelt, S. Caletti, O. Fedkevych, S. Marzani, S. Schumann, G. Soyez, JHEP 03 (2022) 131

Groomed Les Houches Angularity in Z-jet and dijet events

JHEP 01 (2022) 188





Soft-drop grooming (z_{cut} = 0.1, β_{sd} = 0) to remove soft and wide-angle radiation

More challenging to describe **gluon-enriched jets**

Mismodeling at large LHA increase after removing soft&wide-angle radiation

$\kappa = 0.5, \beta = 1$

$$\lambda^{\kappa}_{\beta} = \sum_{i\in \text{jet}} z_i^{\kappa}\left(\frac{\Delta R_i}{R}\right)^{\beta} \quad z_i \equiv \frac{p_{Ti}}{\sum_{j\in\text{jet}} p_{Tj}}$$

pQCD calculations D. Reichelt, S. Caletti, O. Fedkevych, S. Marzani, S. Schumann, G. Soyez, JHEP 03 (2022) 131

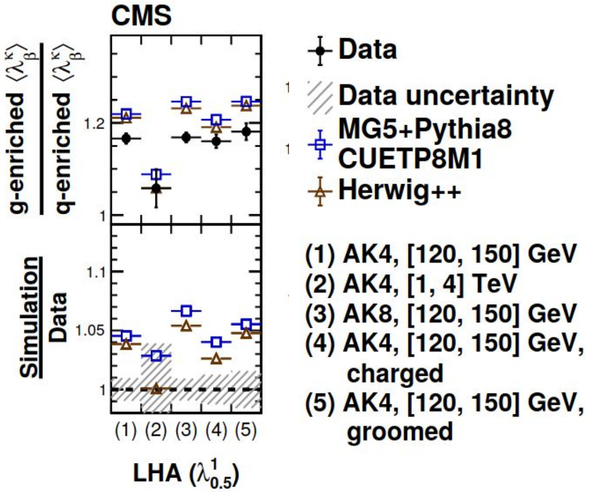
Dijet/Z+jet ratio (g-enriched/q-enriched)

CMS, *arXiv:2109.03340*, ***JHEP* 01 (2022) 188**

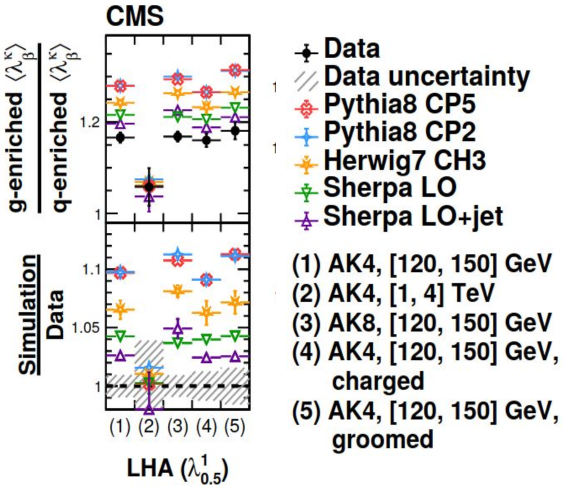
- uncertainties partially cancel in dijet/Z+jet ratio
- MC simulations overestimate g-enriched/q-enriched ratio
- g-enriched / q-enriched ratio is better modelled with “old” PYTHIA8/HERWIG7 tunes

gluon-LHA/quark-LHA > 1
(mostly due to $C_A > C_F$)

“old” CMS tunes
(<~ 5% off)

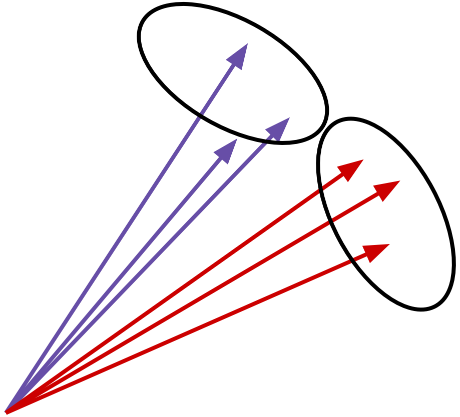


“new” CMS tunes
(up to ~10% off)



full summary plot in backup
(other angularities)

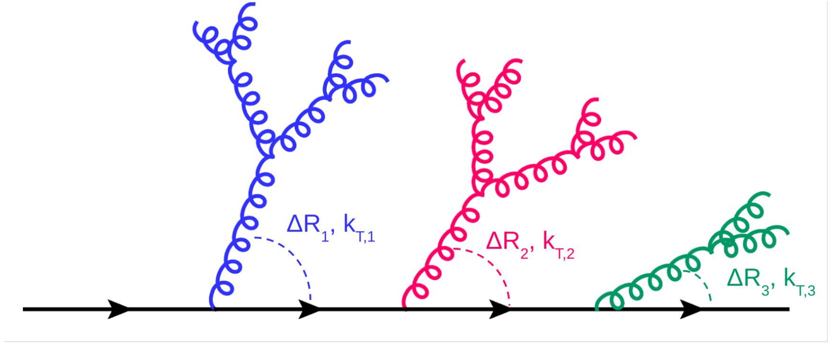
Iterative declustering



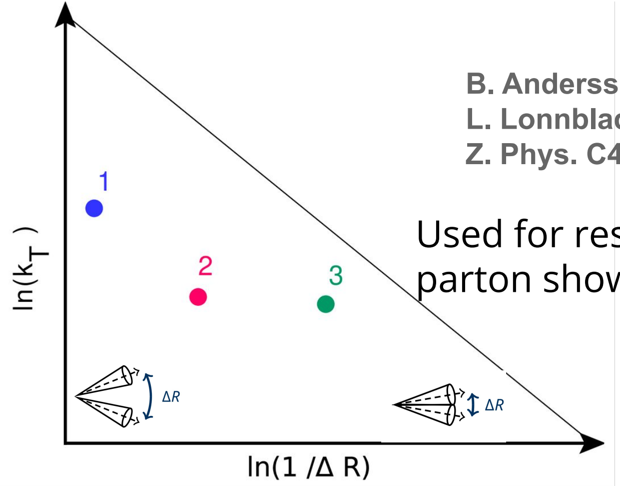
Organizing hadrons

Phase-space of QCD branchings in the Lund plane

Lund planes (or diagrams) are a 2D representation of the phase-space of 1→2 splittings:



k_T : relative transverse momentum of emission
ΔR: angular opening of emission and core



B. Andersson, G. Gustafson,
L. Lonnblad, and U. Pettersson,
Z. Phys. C43 (1989) 625

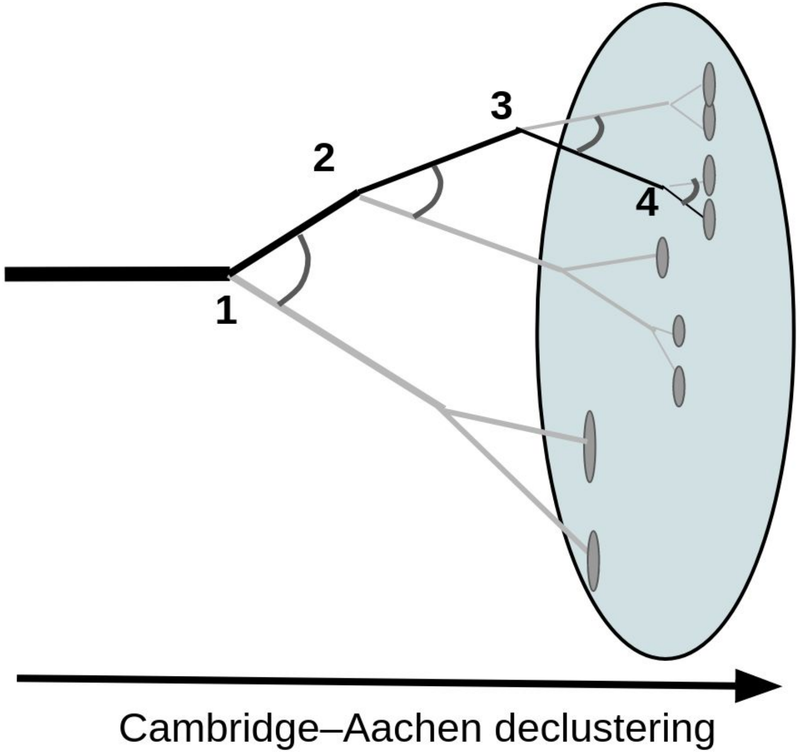
Used for resummation and parton shower development

In soft & collinear limit of QCD, emissions fill the double-logarithmic plane of k_T and ΔR uniformly

$$\mathcal{P} \propto \alpha_{\mathrm{s}} \frac{\mathrm{d}k_{\mathrm{T}}}{k_{\mathrm{T}}} \frac{\mathrm{d}\Delta R}{\Delta R} = \alpha_{\mathrm{s}} \mathrm{d}\ln(k_{\mathrm{T}}) \mathrm{d}\ln(\Delta R)$$ ← **approximate self-similarity of QCD**

Constructing the *primary* Lund jet plane

F. Dreyer, G. Salam, G. Soyez, JHEP12(2018)064



1. Jet constituents are reclustered with the CA algorithm.
2. Follow clustering tree in reverse (large → small angles), **along the hardest branch**
3. k_T and ΔR of the softer subjet (**emission**) relative to the harder subjet (**core**) is registered at each step

$$\Delta R = \sqrt{(y_{\text{soft}} - y_{\text{hard}})^2 + (\phi_{\text{soft}} - \phi_{\text{hard}})^2}$$
$$k_{\mathrm{T}} = p_{\mathrm{T}}^{\text{soft}} \Delta R$$

4. Repeat 2-3 until hard branch has a single constituent

Previously measured by ATLAS PRL 124, 222002 (2020) and ALICE ALICE-PUBLIC-2021-002

Angular ordering privileges QCD collinear divergence & mimics color coherence effects

Primary Lund jet plane density

We measure the jet-averaged density of emissions:

$$\frac{1}{N^{\text{jets}}}\frac{\mathrm{d}^2 N_{\text{emissions}}}{\mathrm{d}\ln(k_T)\mathrm{d}\ln(R/\Delta R)} \simeq \frac{2}{\pi} C_R \alpha_s(k_T)$$

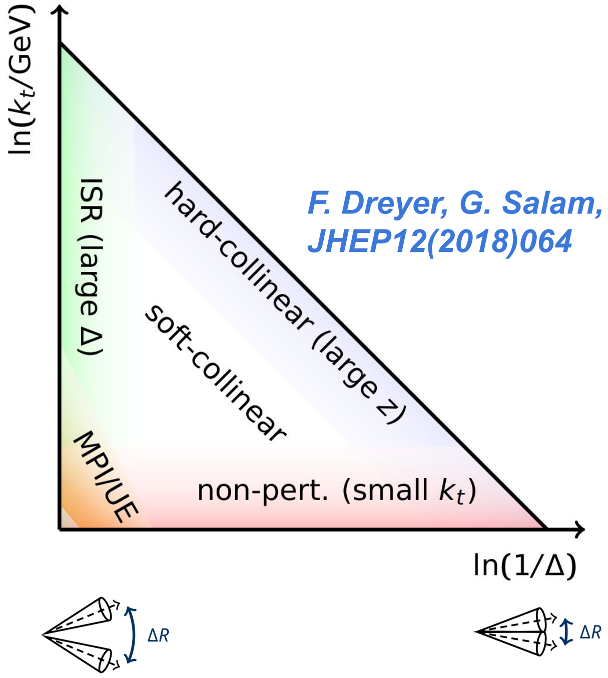
soft & collinear limit

(C_R is a color factor, C_A = 3 for g→gg, C_F = 4/3 for q→qg)

CMS Run-2 setup CMS-PAS-SMP-22-007 :

- $p_T^{\text{jet}} > 700$ GeV, $|y^{\text{jet}}| < 1.7$, anti-k_T with small R = 0.4 and large R = 0.8
- Using charged constituents for the substructure
- Unfolded with D'Agostini to particle-level (p_T^{jet} , $\ln(k_T)$, ΔR)

Various mechanisms are separated



F. Dreyer, G. Salam, G. Soyez, JHEP12(2018)064

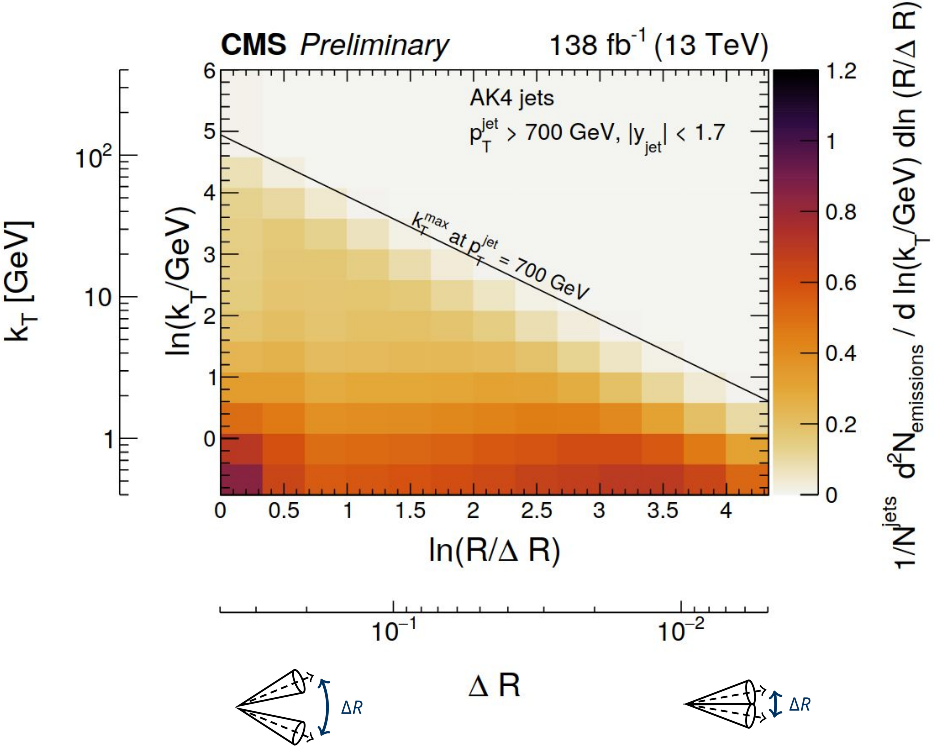
Can use Lund plane density to improve and test calculations in a "factorized" way

measured by ATLAS PRL 124, 222002 (2020) and ALICE ALICE-PUBLIC-2021-002

Unfolded primary Lund jet plane densities

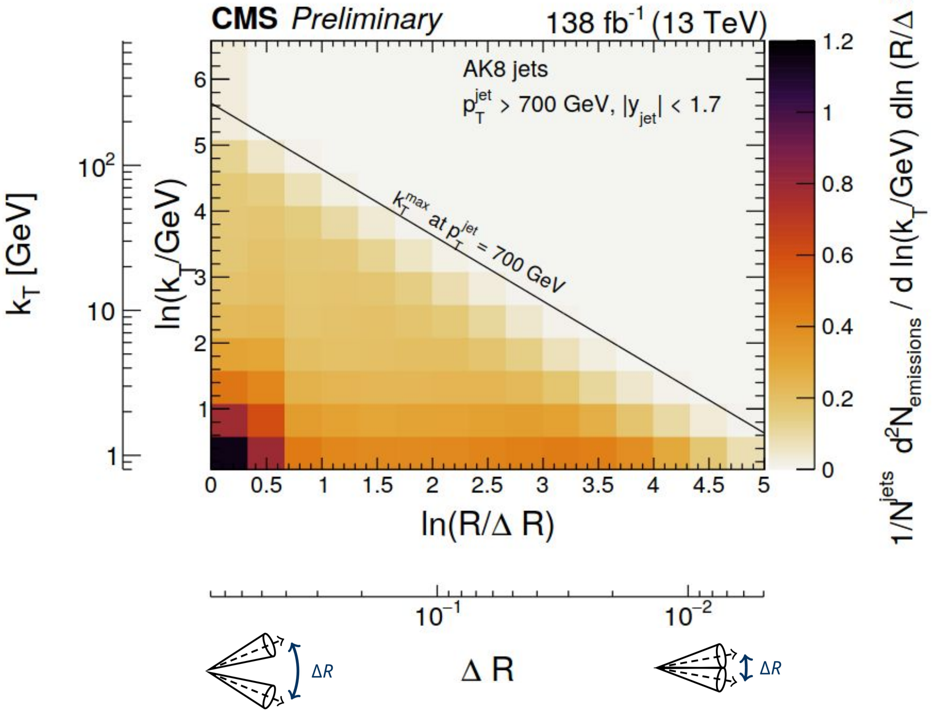
CMS-PAS-SMP-22-007

R=0.4 (standard R in Run-2)



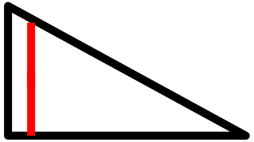
R=0.8 (wider & harder emissions)

Up to k_T ~ 700 GeV at large angles.



Densities approximately flat for hard &collinear emissions due to running of α_S

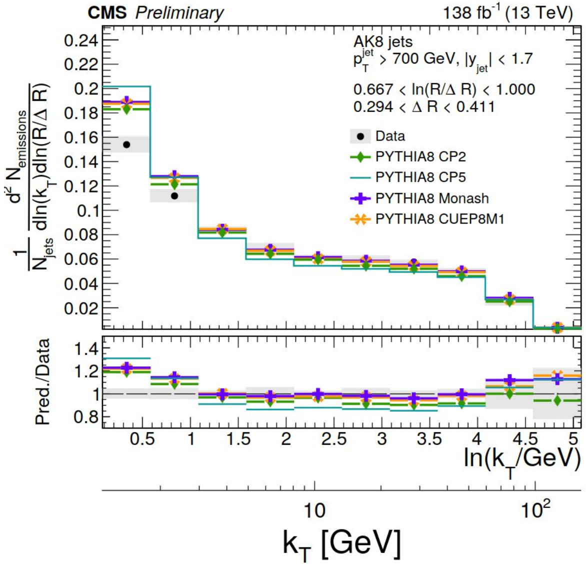
Large angle emissions



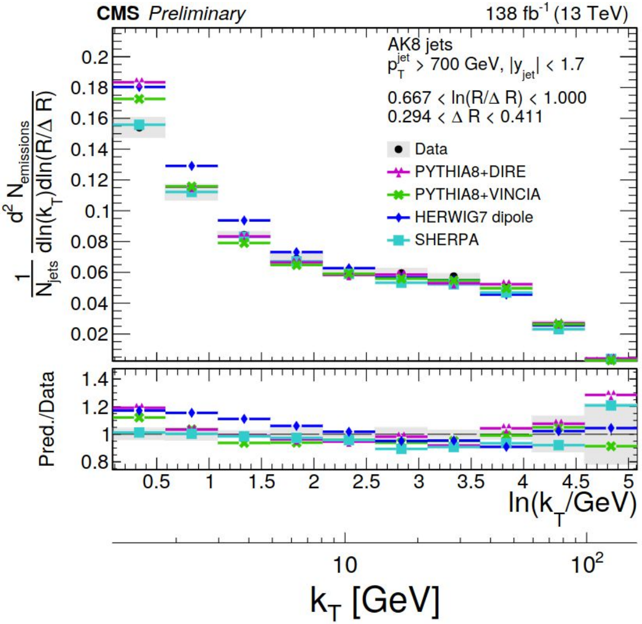
R = 0.8

Comparison to parton showers& tunes

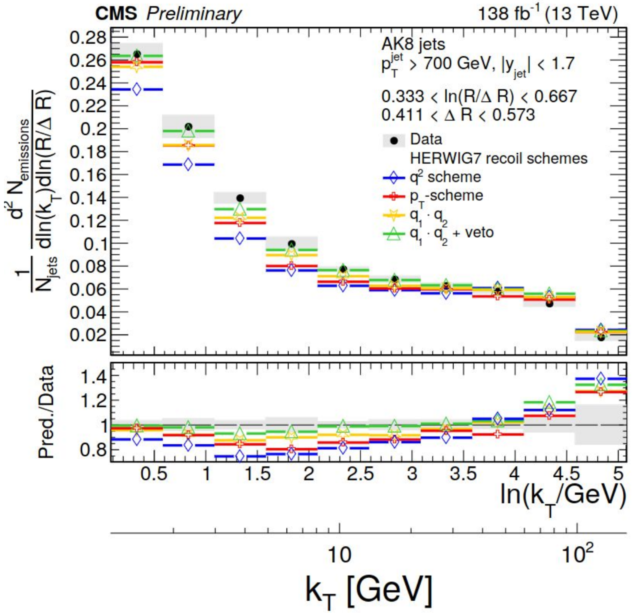
CMS-PAS-SMP-22-007



PYTHIA8 tunes
(CP2, CP5, Monash, CUEP8m1)



Dipole showers
(Vincia, Dire, Herwig7 dipole, Sherpa)



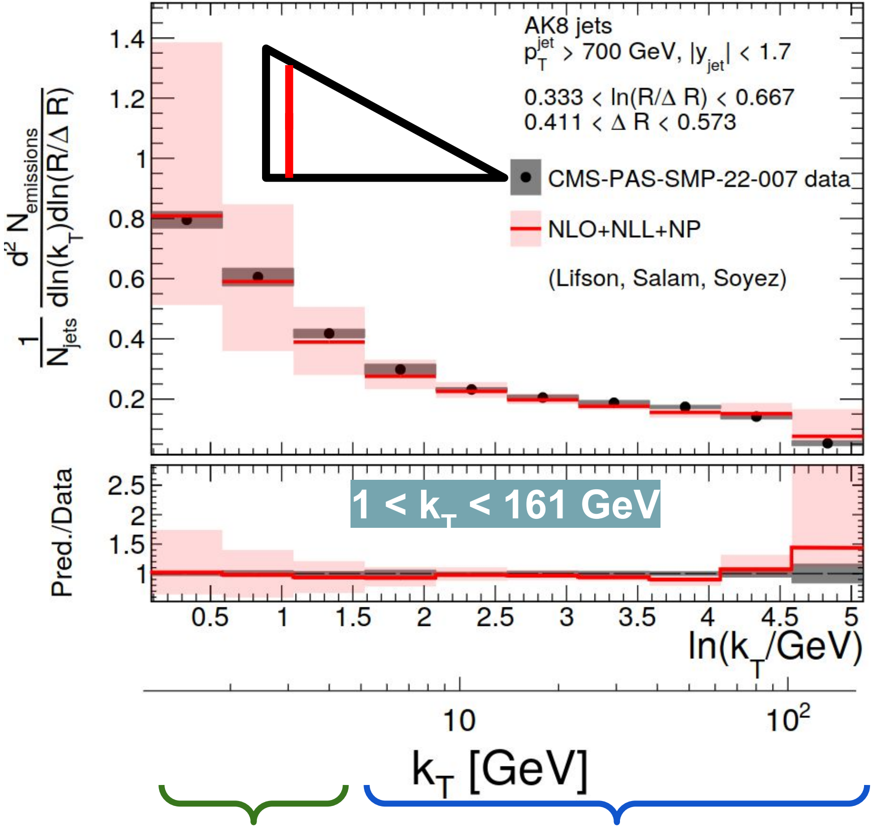
Herwig7 recoil schemes
(angle-ordered)

Differences between data & MC of the order of 10–20%. “Factorization” of effects can be used for MC tuning

Herwig7.2 angle-ordered shower performs better than **Herwig7.2 dipole shower**

Comparison to pQCD analytical calculations (NLO+NLL+NP)

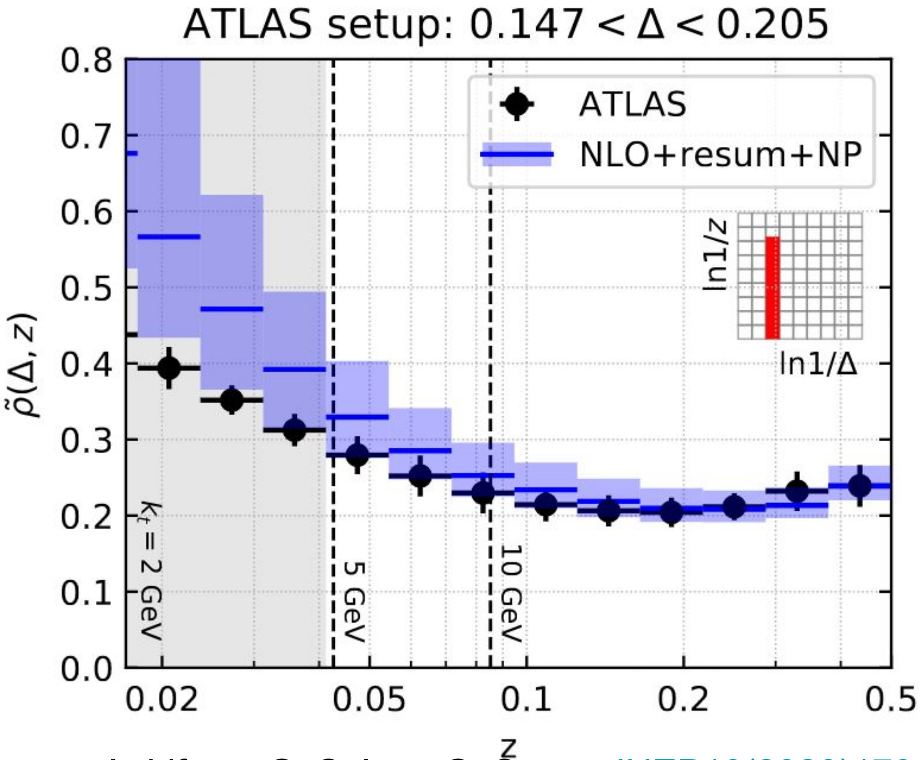
Calculations based on JHEP10(2020)170



$1 < k_T < 161$ GeV

nonperturbative

resummation



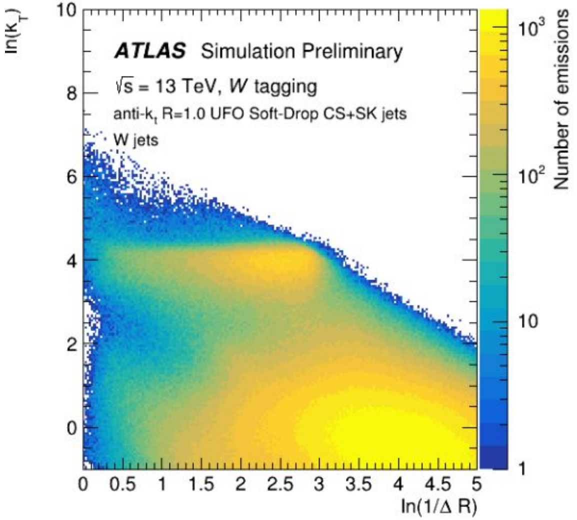
A. Lifson, G. Salam, G. Soyez JHEP10(2020)170

data from ATLAS Lund plane,
PRL 124, 222002 (2020)

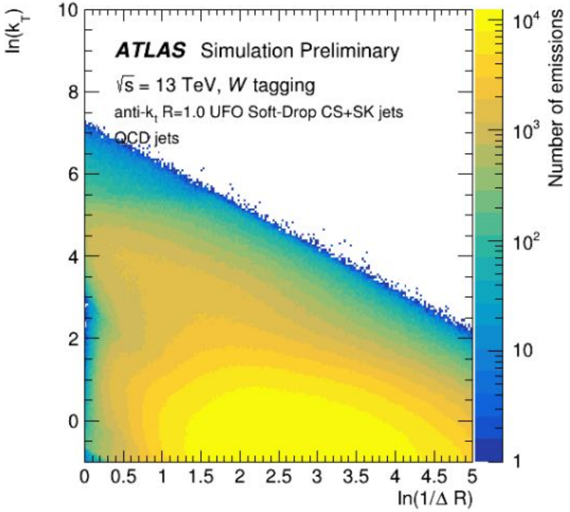
Identification of boosted W boson jets with Lund plane

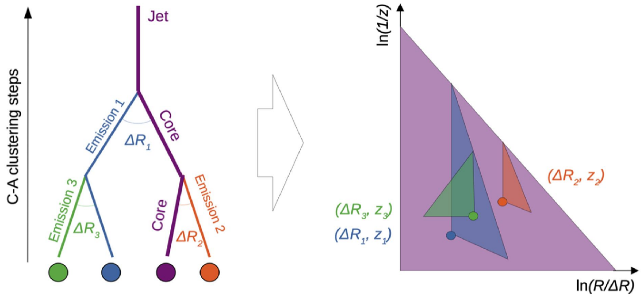
ATL-PHYS-PUB-2023-017

W boson jets (signal)



QCD jets (bkg)





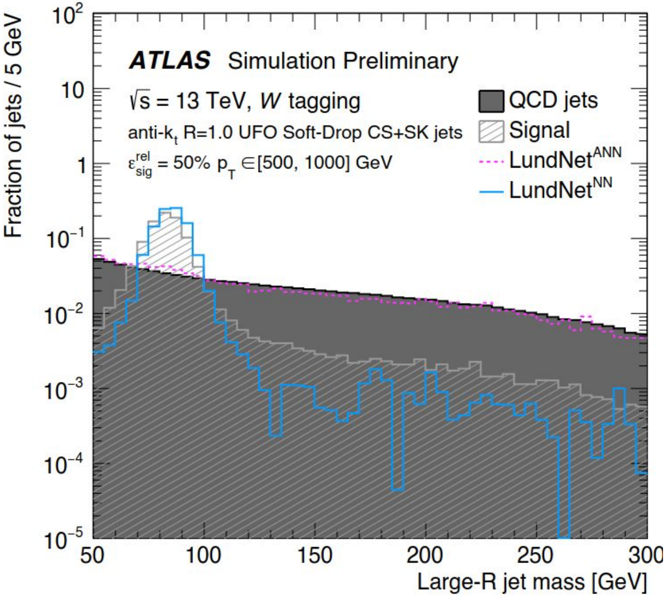
Using LundNet Dreyer, Qu arXiv:2012.08526

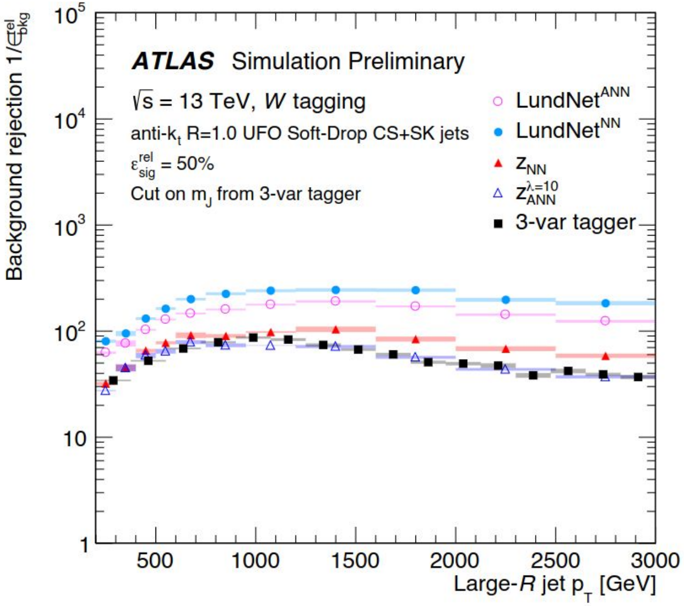
Graph neural network using CA declusterings

New results from ATLAS, more details presented by Jad Sardain at Lund plane workshop

Identification of boosted W boson jets with Lund graphs

ATL-PHYS-PUB-2023-017





Example of closure test in simulation

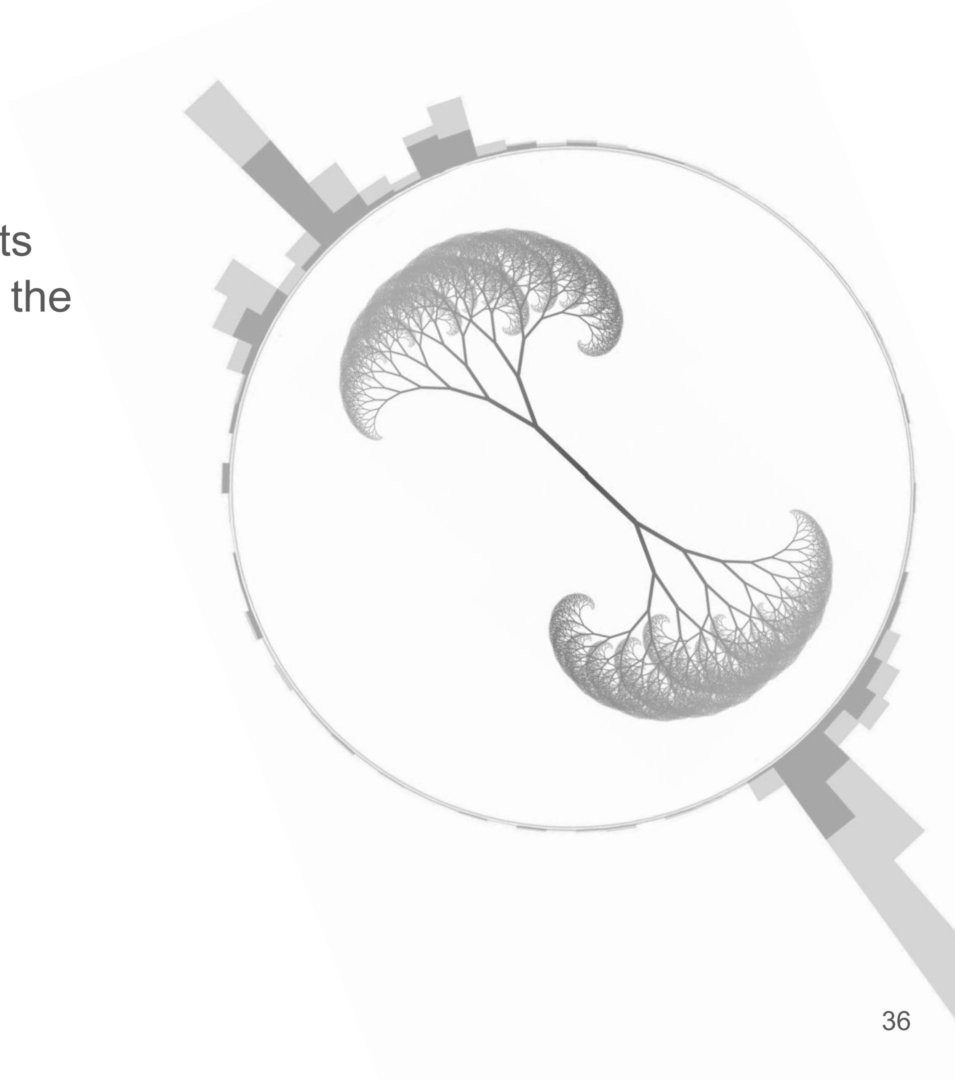
ANN is able to recover **QCD bkg**
NN is able to recover **boosted W boson signal**

LundNet^{NN} : bkg rejection by a factor of roughly ~2.5–3 wrt standard taggers (DNN, 3-variable based)

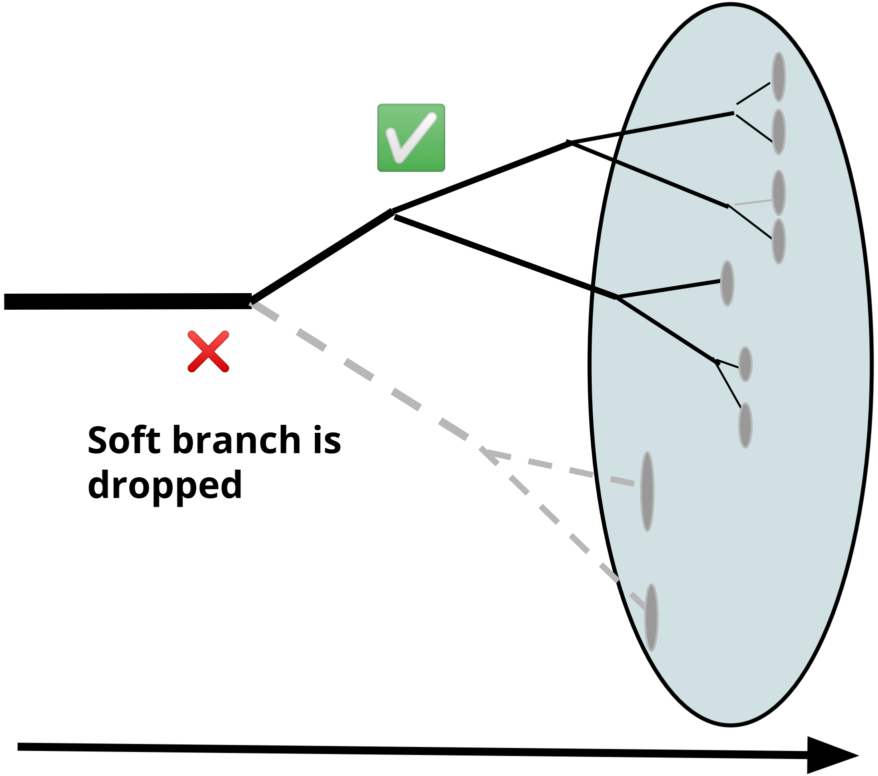
bkg rejection reduced by about 25% with **LundNet^{ANN}** (used for mass decorrelation)

Summary

- LHC precision jet substructure measurements allows us to stress test our understanding of the **internal dynamics of the jet**
- More complete list of LHC jet substructure measurements in LHC-EW WG Twiki page



(Intermezzo) soft-drop grooming algorithm



Cambridge–Aachen declustering

1. Jet is reclustered with Cambridge–Aachen (CA), which clusters particles with **angular ordering**

2. Follow the CA clustering history in reverse. Check if the branch satisfies the soft-drop condition:

   $z = p_T^{\text{softer}}/(p_T^{\text{softer}}+p_T^{\text{harder}}) > z_{\text{cut}}\,(\Delta R/R)^{\beta}$

   (a typical choice is $z_{\text{cut}} = 0.1$, $\beta = 0$)

   If the splitting fails the SD condition, the branch is removed

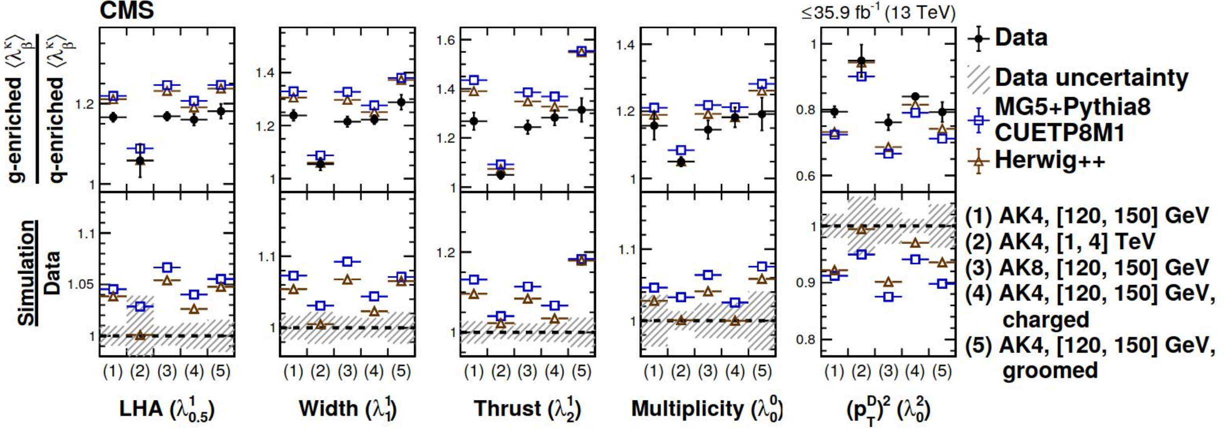
3. Repeat 2 until SD condition is satisfied, which yields a **soft-drop groomed jet**

Dijet/Z+jet ratio (gluon-like/quark-like jet ratio)

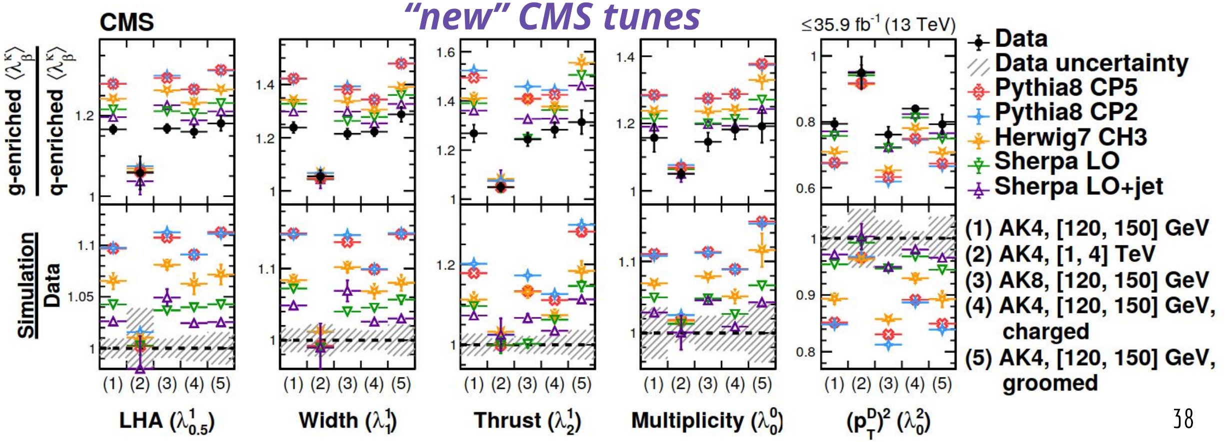
CMS, arXiv:2109.03340
JHEP 01 (2022) 188

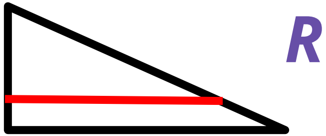
- experimental uncertainties partially cancel in dijet/Z+jet ratio
- LO+PS preds. overestimate the g-enriched/q-enriched ratio
- g-enriched / q-enriched ratio is better modelled with “old” PYTHIA8 and HERWIG7 CMS tunes.

“old” CMS tunes



“new” CMS tunes

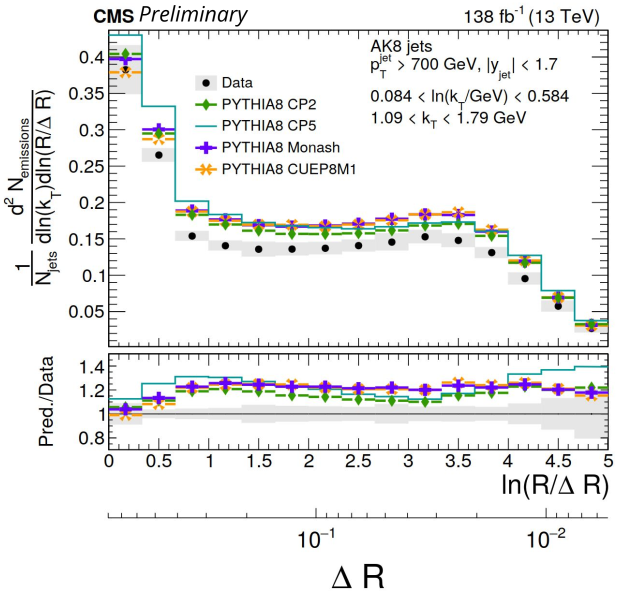




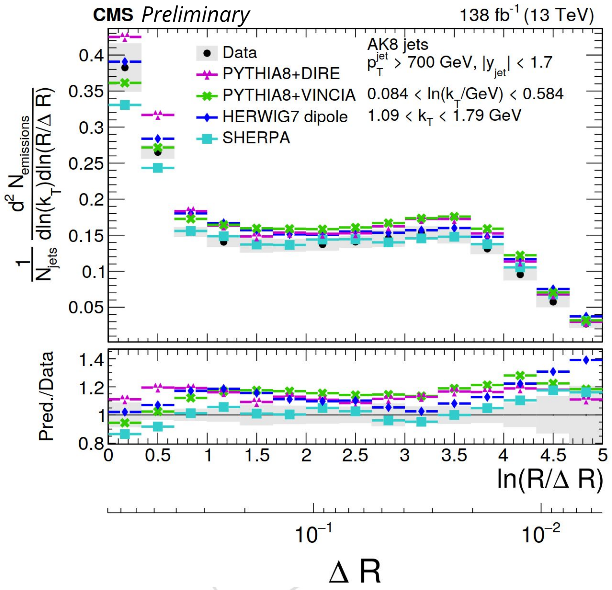
R=0.8

Low-k_T (hadronization + MPI)

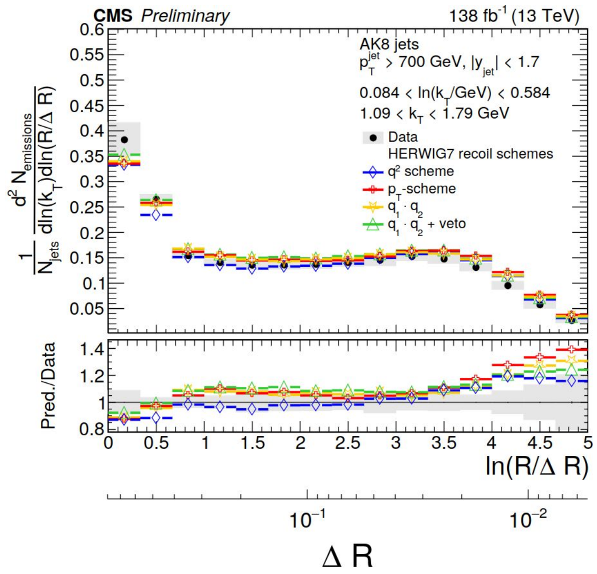
CMS-PAS-SMP-22-007



PYTHIA8 tunes
(CP2, CP5, Monash, CUEP8m1)



Dipole showers
(Vincia, Dire, Herwig7Dipole, Sherpa)



Herwig7.2 recoil schemes
(angle-ordered)

PYTHIA8 systematically overshoots the data at low k_T , **regardless of tune or shower option.**

HERWIG7 **&** Sherpa generally do better. **Cluster vs string fragmentation?**