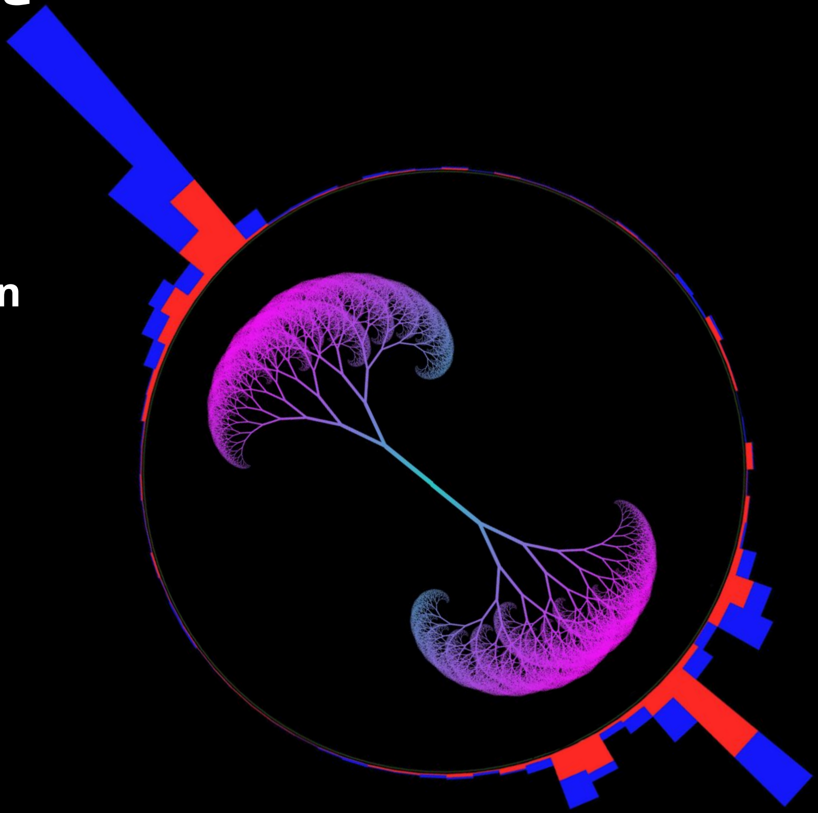
Jet substructure using the Lund tree in 13 TeV pp collisions (CMS)

Cristian Baldenegro on behalf of the CMS Collaboration
(LLR-École Polytechnique)

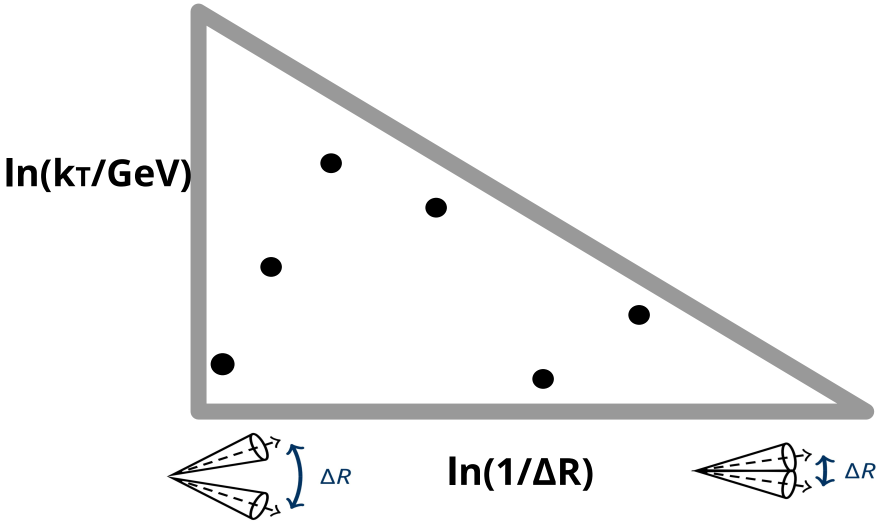
1st Lund Jet Plane Institute, July 3rd-7th 2023



European Research Council

Lund-tree related measurements

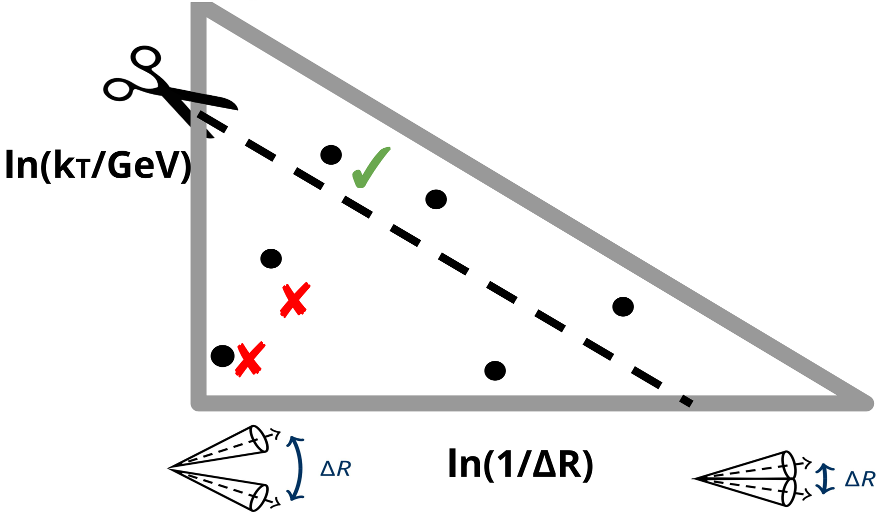
- Primary Lund jet plane density
  [CMS-PAS-SMP-22-007]



- (Un)groomed jet mass:
  JHEP 11 (2018) 113

  jet radius R_g & momentum fraction z_g:
  PRD98, 092014 (2018)

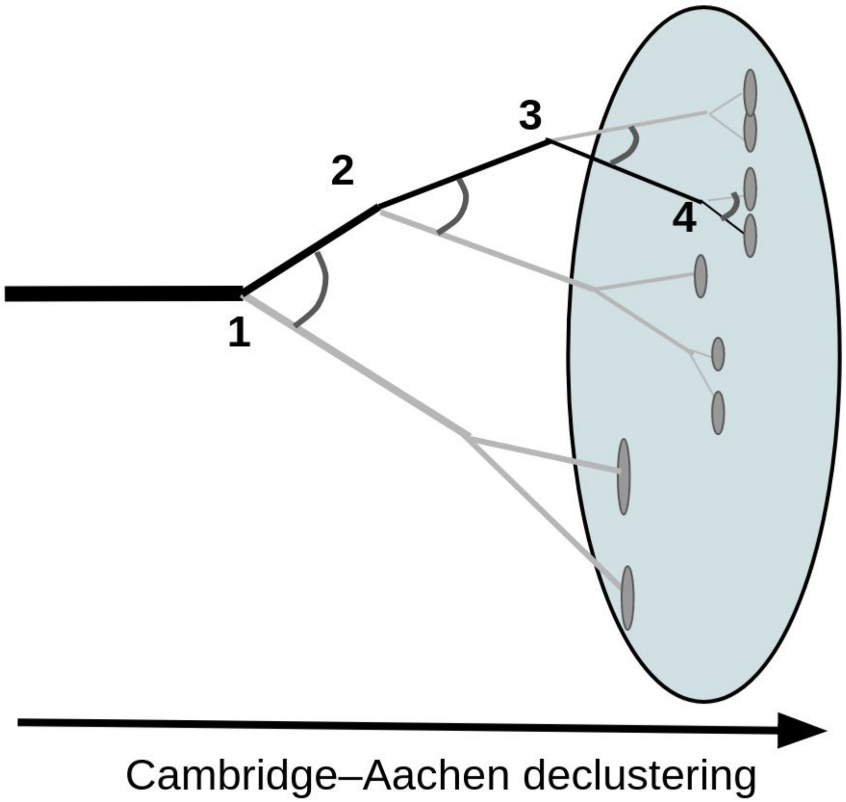
  (Un)groomed angularities:
  JHEP 01 (2022) 188



CMS measurements in heavy-ion collisions covered yesterday by Yi Chen

Constructing the *primary* Lund jet plane

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



1. All anti-k_t jet constituents are reclustered with the Cambridge–Aachen (CA) algorithm

   CA sequentially clusters pairs of particles (pseudojets) with **angular ordering** (small → large angles)

2. Then, follow clustering tree in reverse (large → small angles), **along the hardest branch (primary emissions)**

3. k_T and ΔR of the softer subjet (**emission**) relative to the harder subjet (**core**) is registered at each step

$$\Delta R = \sqrt{(y_{\rm soft} - y_{\rm hard})^2 + (\phi_{\rm soft} - \phi_{\rm hard})^2}$$

$$k_{\rm T} = p_{\rm T}\Delta R$$

4. For each jet, you get a list of paired coordinates $\{k_{T,i}, \Delta R_i\}$

5. Do this for all jets.

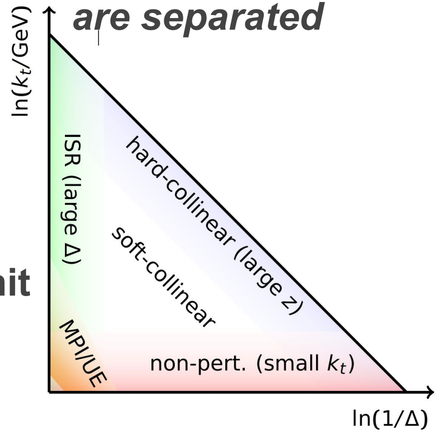
<u>Primary</u> Lund jet plane density

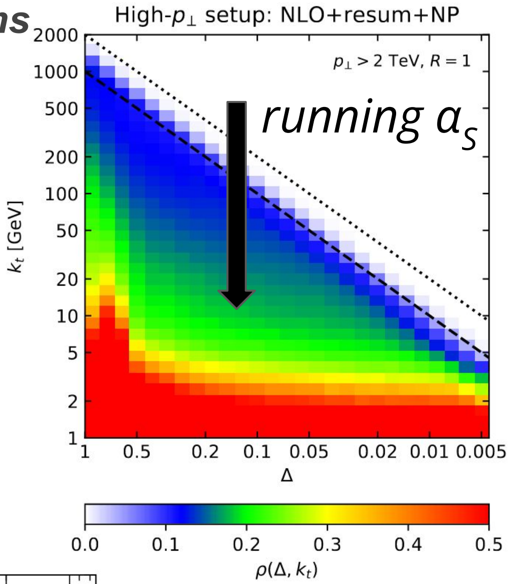
We measure the jet-averaged density of emissions:

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

soft & collinear limit

Physical mechanisms are separated



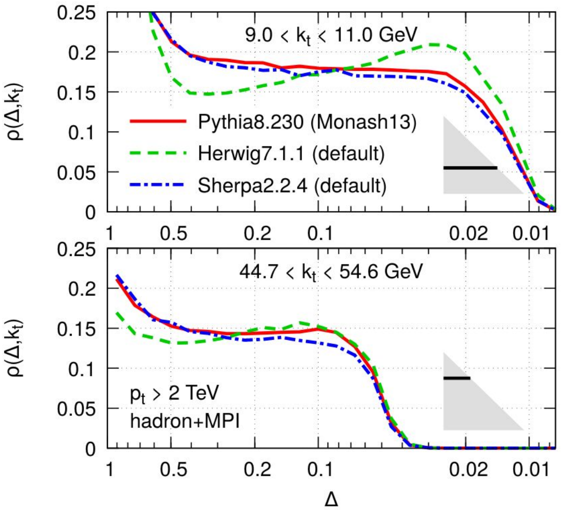


F. Dreyer, G. Salam, G. Soyez, JHEP12(2018)064

running of α_S(kT) **sculpts the Lund plane density**

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

Measurement can be used to stress test calculations in a “factorized” way



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

Density of emissions calculable in pQCD

CMS Run-2 measurement CMS-PAS-SMP-22-007

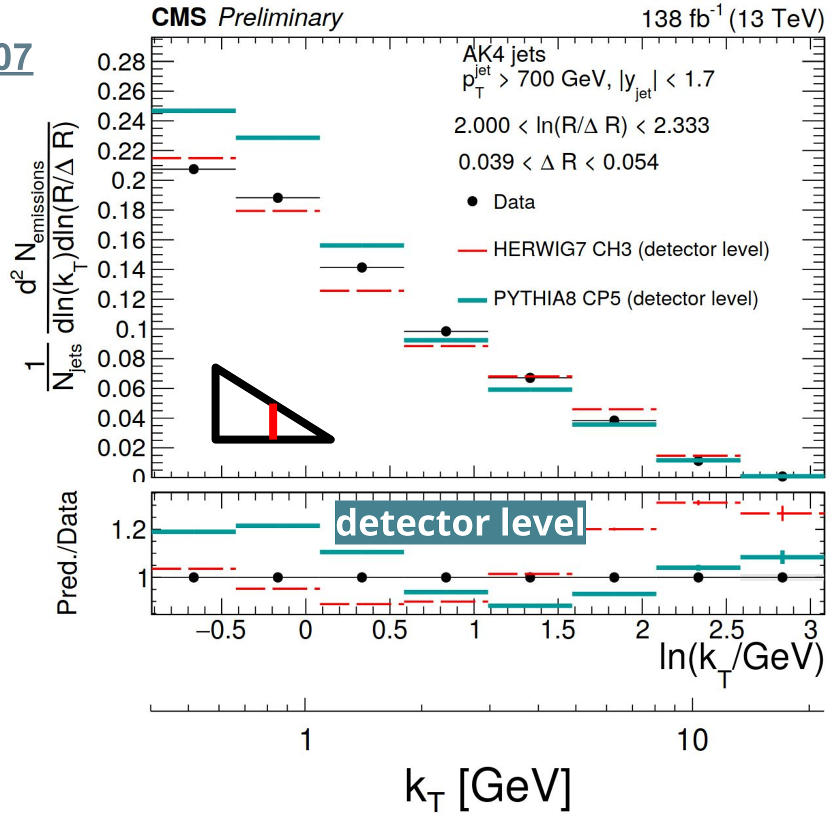
- **anti-k_T jets** with R = 0.4 and R = 0.8.

- Jets with ***p_T>700 GeV and |y| < 1.7***.

  High-p_T jets → more phase-space for hard emissions
  k_T^{max} = ½ p_T^{jet} ΔR

  Using charged-hadron subtraction
  for pileup mitigation

- Lund plane calculated by reclustering **charged PF constituents** with CA algorithm (***angular & momentum resolution + better PU control***)



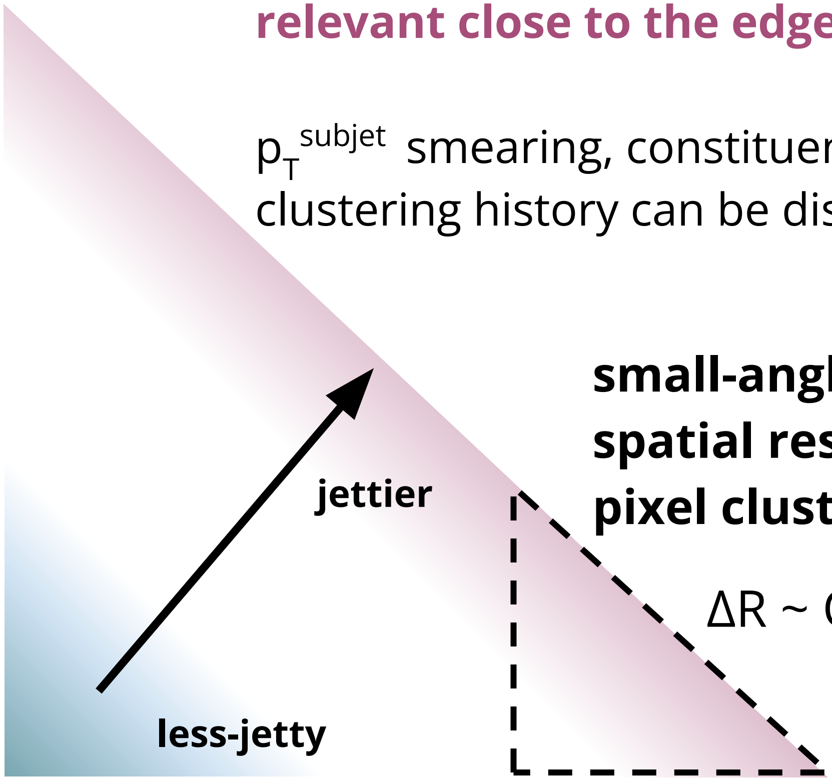
PYTHIA8 CP5 and HERWIG7 CH3 are used to derive corrections.

Generally envelop the data at det-level.

selected detector effects

relevant close to the edge ($p_T^{soft} \sim p_T^{hard}$):

p_T^{subjet} smearing, constituents lost in reconstruction, clustering history can be distorted (e.g., branch swaps)



residual PU contributions

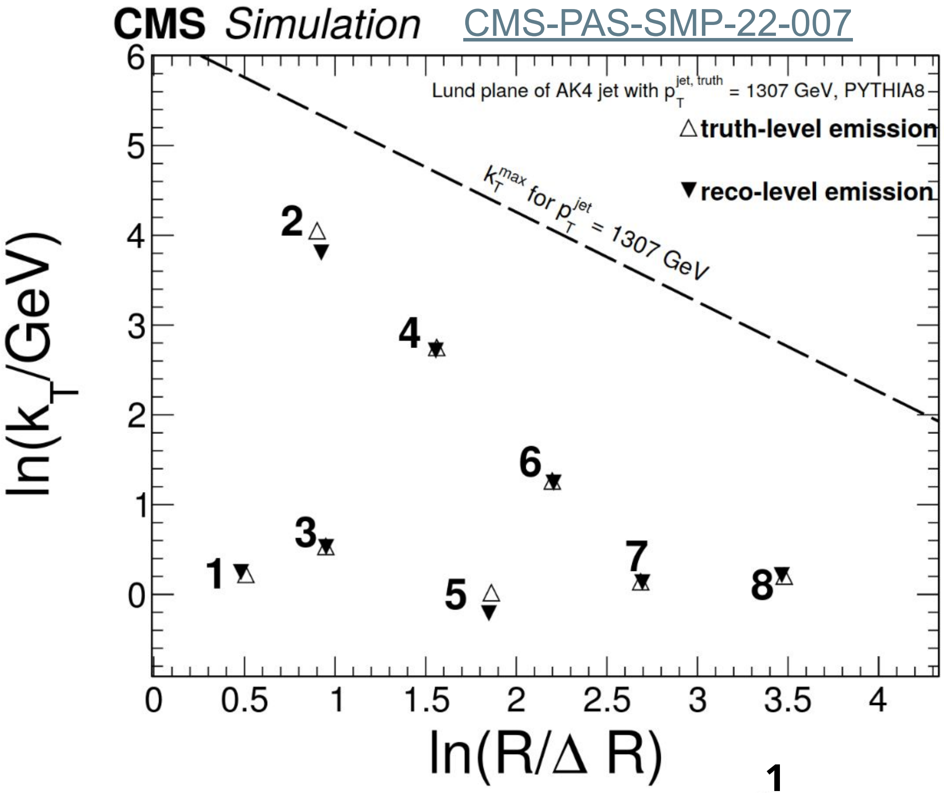
small-angles:
spatial resolution,
pixel cluster merging

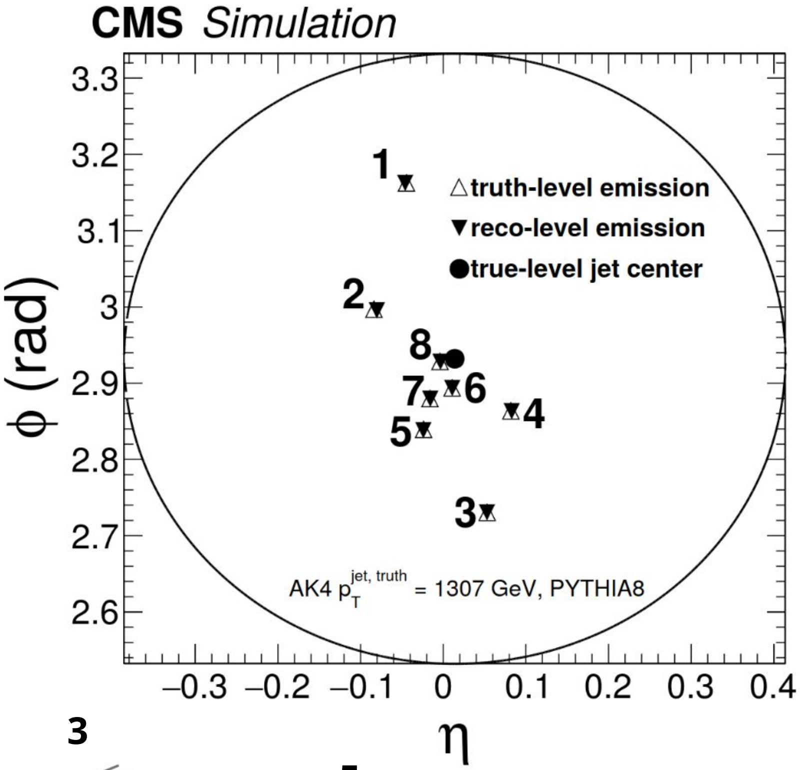
$\Delta R \sim O(10^{-3} – 10^{-2})$

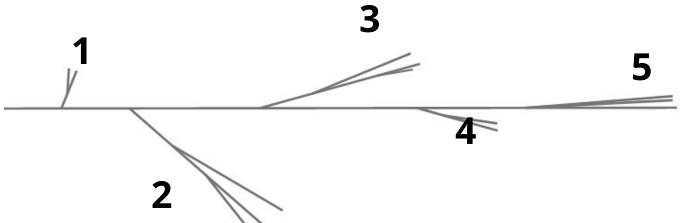
Matching emissions at detector level and particle level

Migration matrix and other MC-based corrections derived from matched part-level and det-level splittings.

CMS-PAS-SMP-22-007







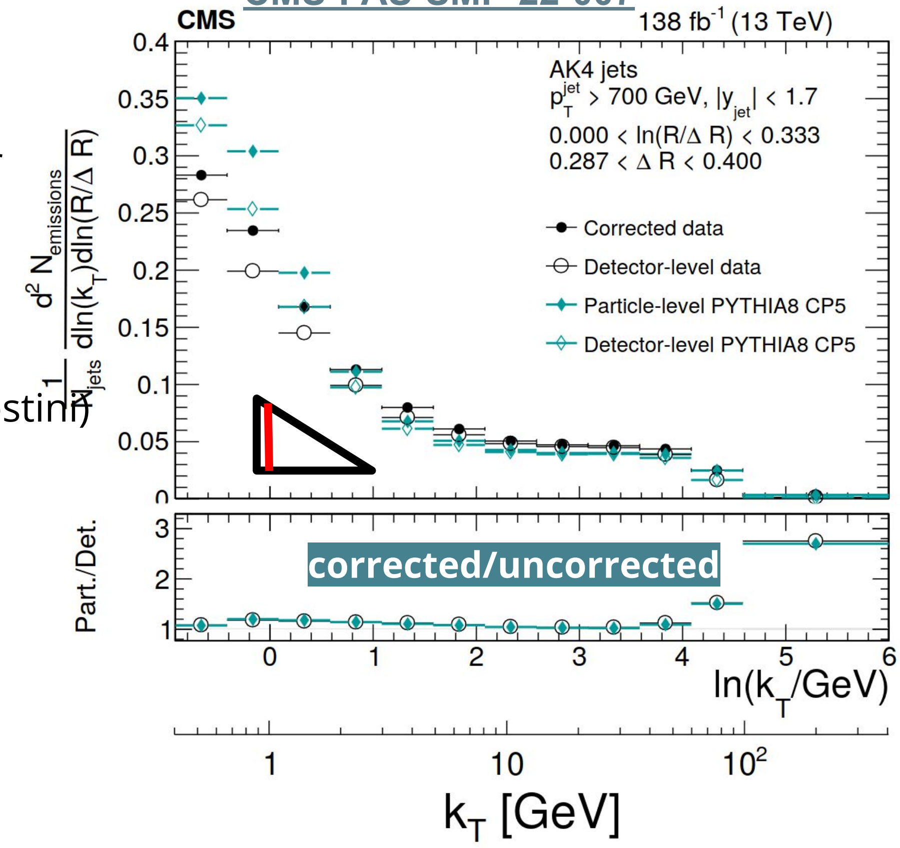
Corrections to particle level

Sequential set of corrections:

1. **Background:** bin-by-bin correction to account for det-level emissions not matched to truth-level emissions.
2. **Multidimensional regularized unfolding** (D'Agostini) of primary Lund jet plane (p_T^{jet}, k_T, ΔR).
3. **Efficiency:** bin-by-bin correction to account for hadron-level emissions without matching.

PYTHIA8 CP5 chosen as nominal to also propagate parton shower scale uncertainties

CMS-PAS-SMP-22-007



Smearing becomes more important at high k_T

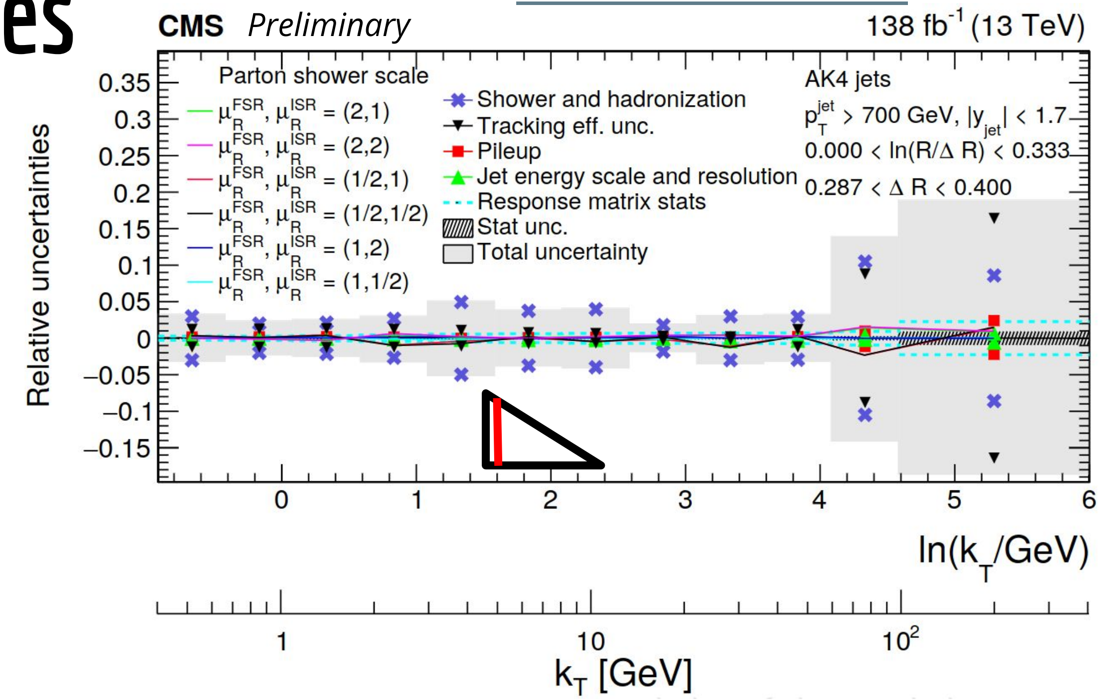
Systematic uncertainties

CMS-PAS-SMP-22-007

Shower & hadronization model uncertainty
(2–7% in the bulk, 10% at kinematical edge)

Decorrelated into prior bias × response pieces

Tracking efficiency uncertainty,
1-2% in bulk, dominates at 15-20% at edge
(covers detector modeling)



Subleading components (less than 1%):

Parton shower scale uncertainties
(six-point scale variations)
Response matrix stats
Jet energy scale **and** **resolution uncertainties**
Pileup modeling

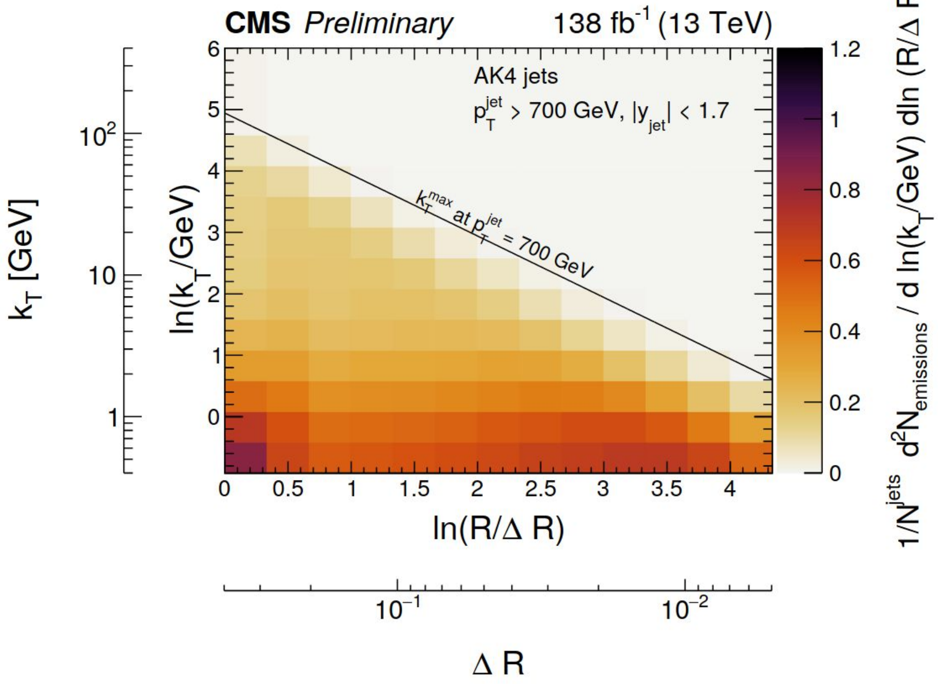
Dominated by **shower & hadronization modeling** in bulk of Lund plane & by **tracking efficiency** at high k_T

Cristian Baldenegro (LLR)

Unfolded primary Lund jet plane densities

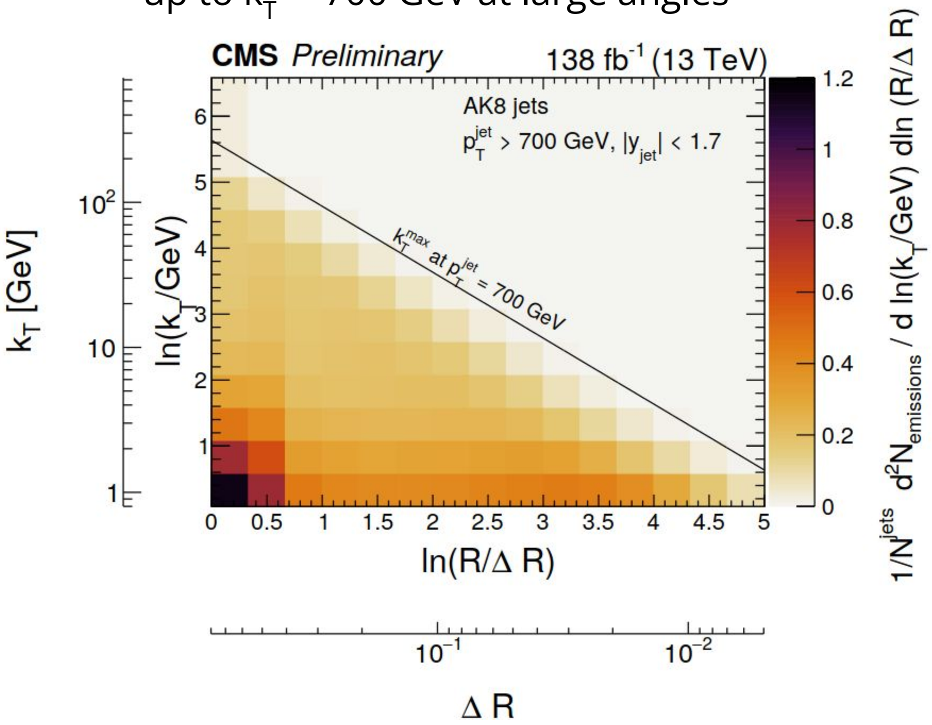
CMS-PAS-SMP-22-007

R=0.4 (standard R in Run-2)



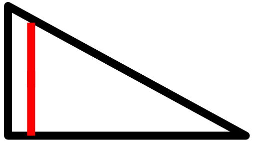
R=0.8 (wider & harder emissions)

up to k_T ~ 700 GeV at large angles



Large-R allows to mitigate clustering effects in a wider swathe of LJP

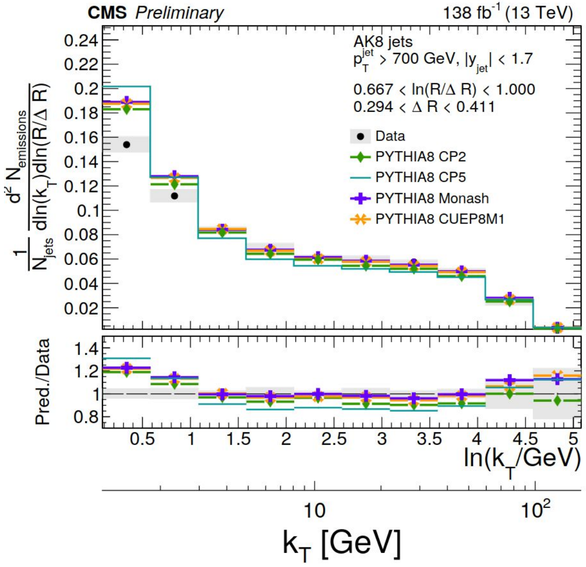
Large angle emissions



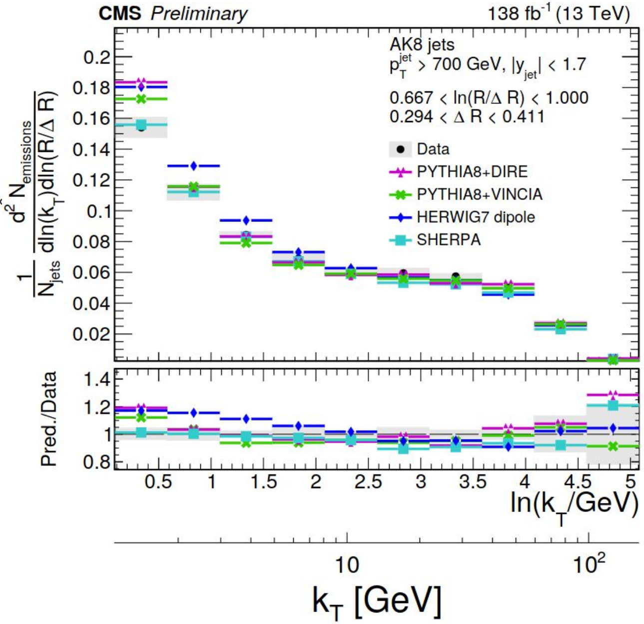
R = 0.8

Comparison to parton showers& tunes

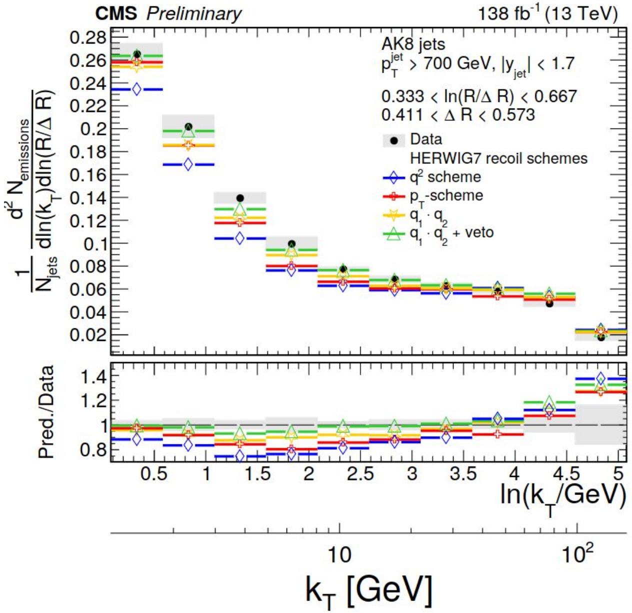
CMS-PAS-SMP-22-007



PYTHIA8 tunes
(CP2, CP5, Monash, CUEP8m1)



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

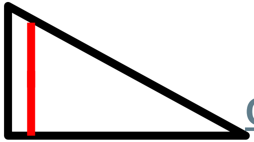


Herwig7 recoil schemes
(angle-ordered)

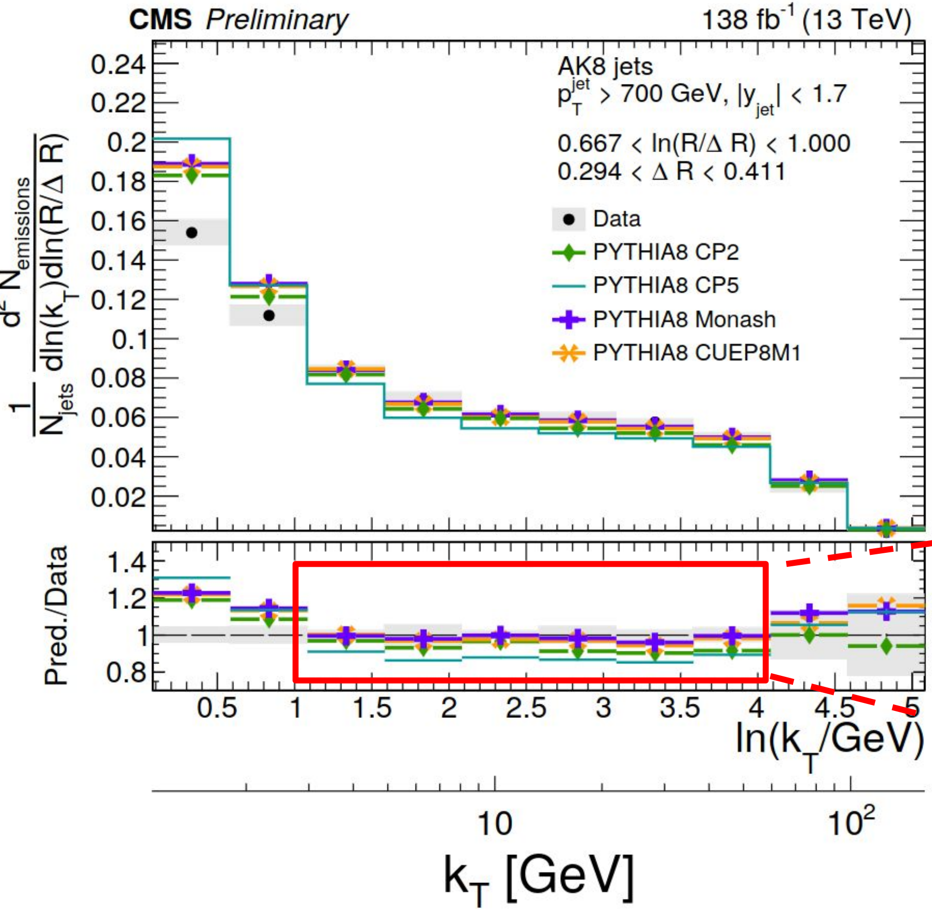
Differences between data & MC of the order of 10–20%. “Factorization” of effects can be used for MC tuning

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

Large angle emissions



CMS-PAS-SMP-22-007



Most important difference between tunes is α_s^{FSR}

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

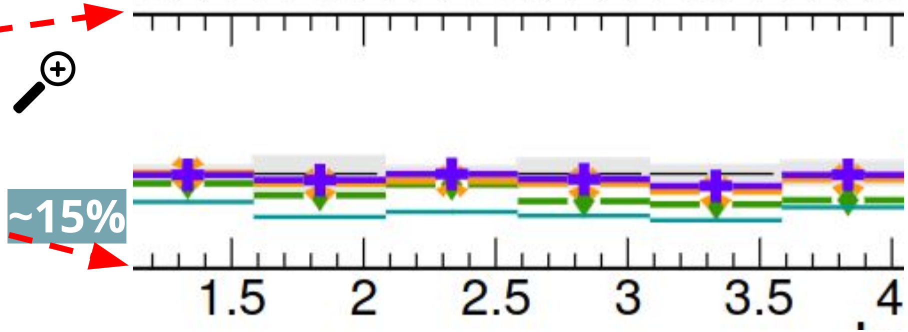
PYTHIA8 tunes:

Monash&CUEP8m1: $\alpha_s^{FSR}(m_Z) = 0.1365$

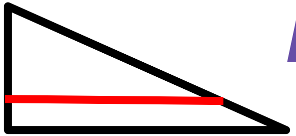
CP2: $\alpha_s^{FSR}(m_Z) = 0.130$

CP5: $\alpha_s^{FSR}(m_Z) = 0.118$

LJP data can be used to constrain $\alpha_s^{FSR}(m_Z)$ in MC tuning



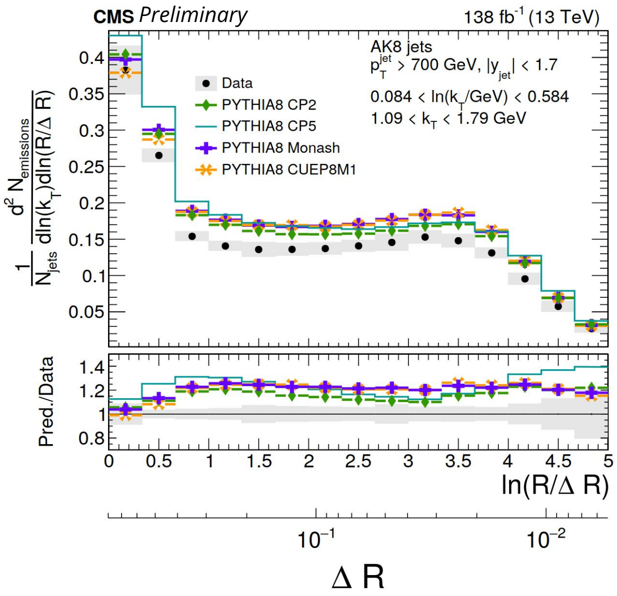
k_T between 3 – 50 GeV



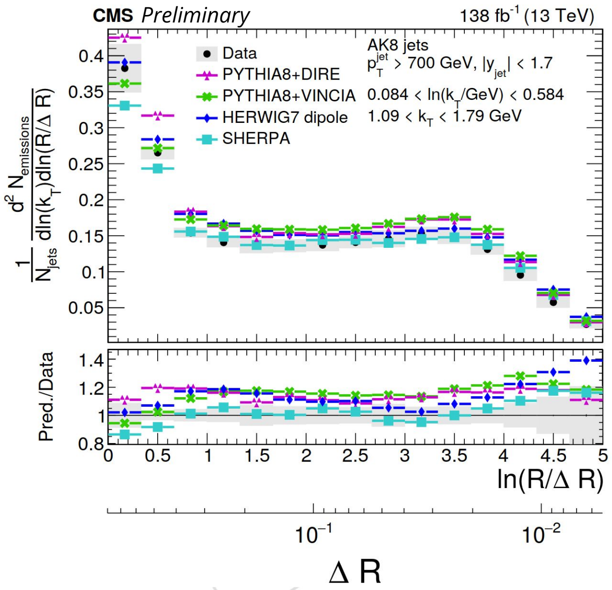
R=0.8

Low-k_T (hadronization + MPI)

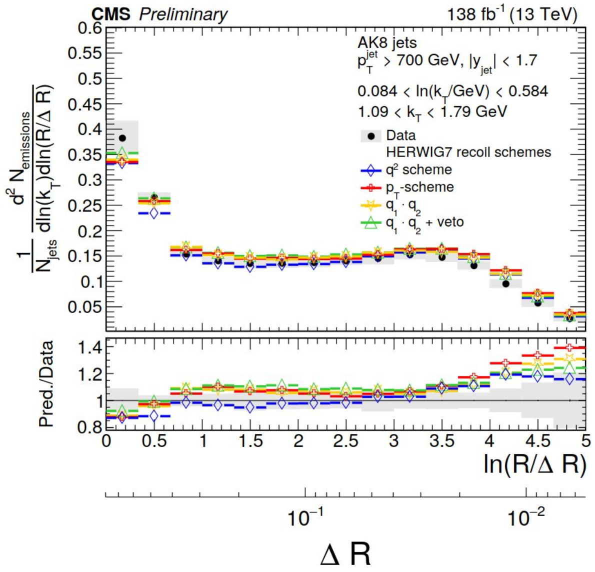
CMS-PAS-SMP-22-007



PYTHIA8 tunes
(CP2, CP5, Monash, CUEP8m1)



Dipole showers
(Vincia, Dire, Herwig7Dipole, Sherpa)



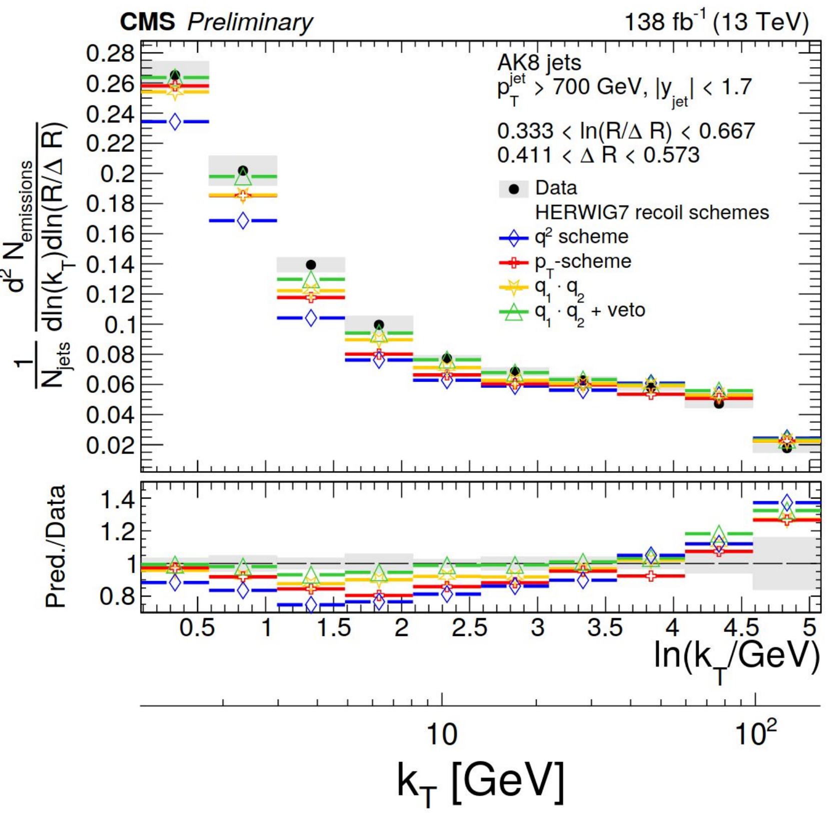
Herwig7.2 recoil schemes
(angle-ordered)

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

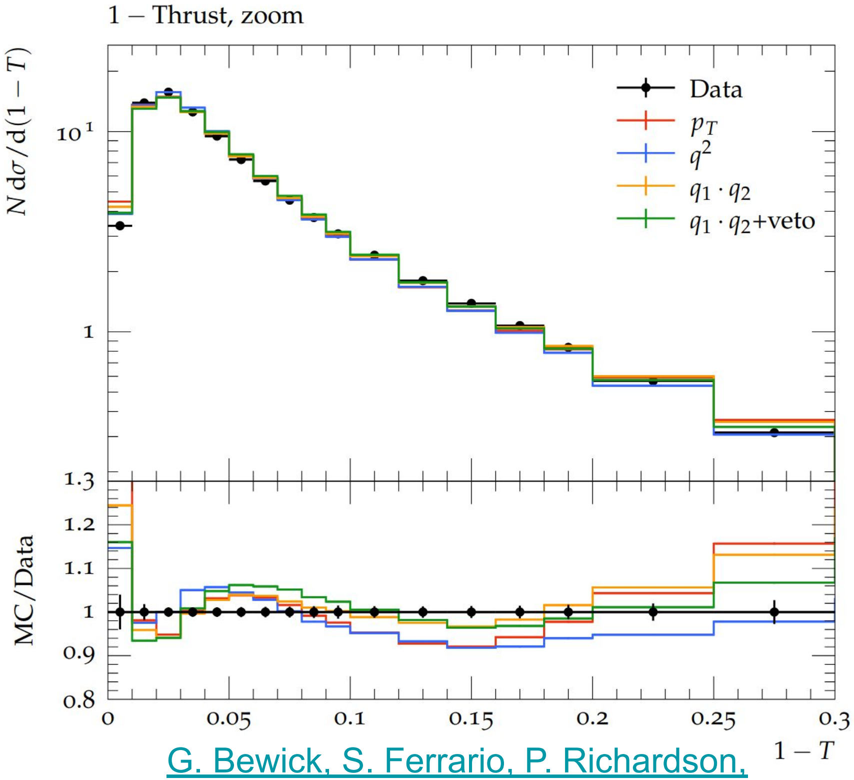
HERWIG7 **&** Sherpa generally do better. **Cluster vs string fragmentation?**

comparison to four different recoil schemes of Herwig7.2

high-p_T quark and gluon jets



Thrust in e^+e^- at Z mass pole

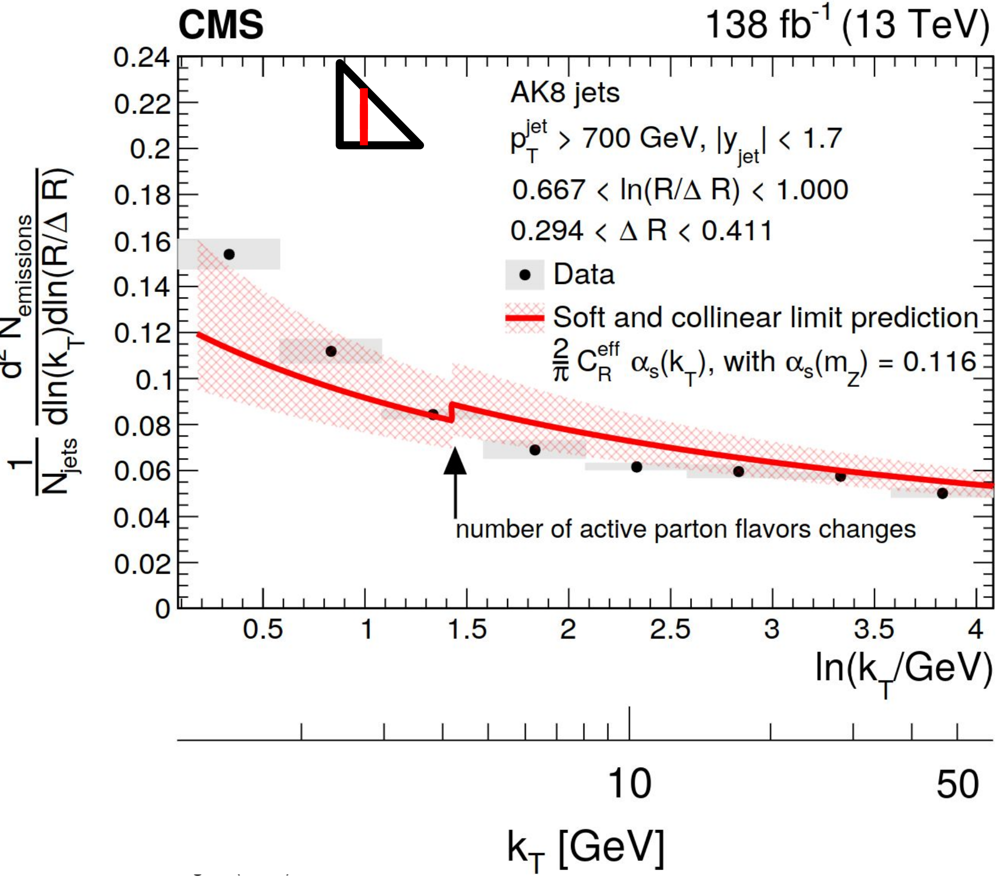


G. Bewick, S. Ferrario, P. Richardson, M. H. Seymour, arXiv:1904.11866

LJP data favors **q_1q_2+veto** scheme, consistent with trends in event shape variables at LEP

data qualitatively described by running of $\alpha_S \sim 1/\ln(k_T)$

CMS-PAS-SMP-22-007



soft & collinear limit

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

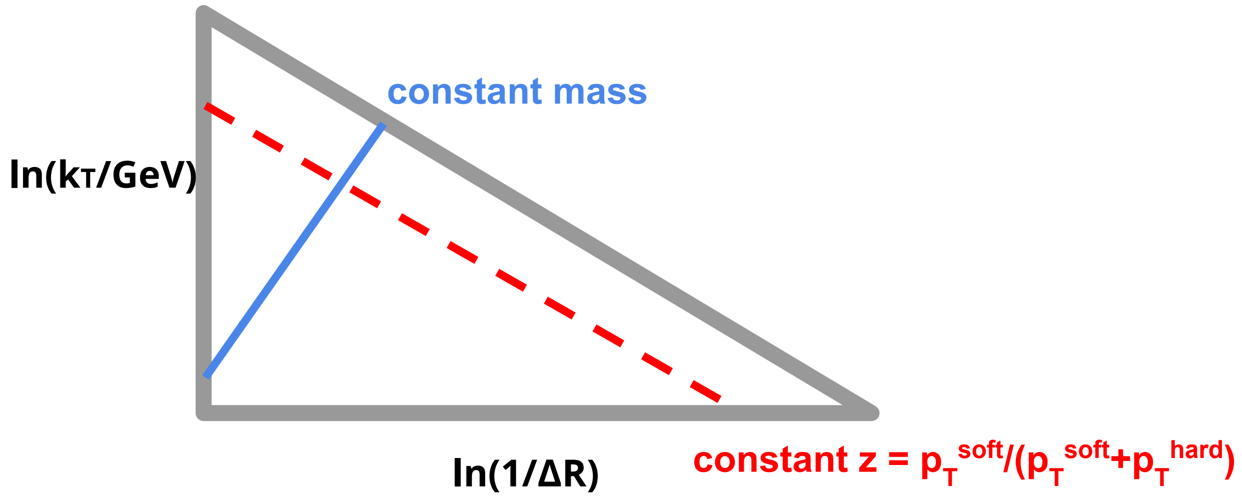
simplistic assumptions:

- $C_R^{eff} = 0.59\, C_F + 0.41\, C_A$

- one-loop β function w/ $\alpha_s\,(m_Z) = 0.116$

Can we use LHC Lund planes for an α_s extraction(s)?

Projections of the primary Lund jet plane



Select specific projections of the Lund jet plane with dedicated algorithms

(e.g., soft drop/mMDT, dynamical grooming, late k_T , …)

M. Dasgupta, A. Fregoso, S. Marzani, G. Salam, JHEP09 (2013) 029

A. Larkoski, S. Marzani, G. Soyez, J. Thaler, JHEP 1405 (2014) 146

Y. Mehtar, A. Soto, K. Tywoniuk, PRD 101, 034004 (2020)

L. Cunqueiro, D. Napoletano, A. Soto, Phys. Rev. D 107, 094008

soft drop grooming

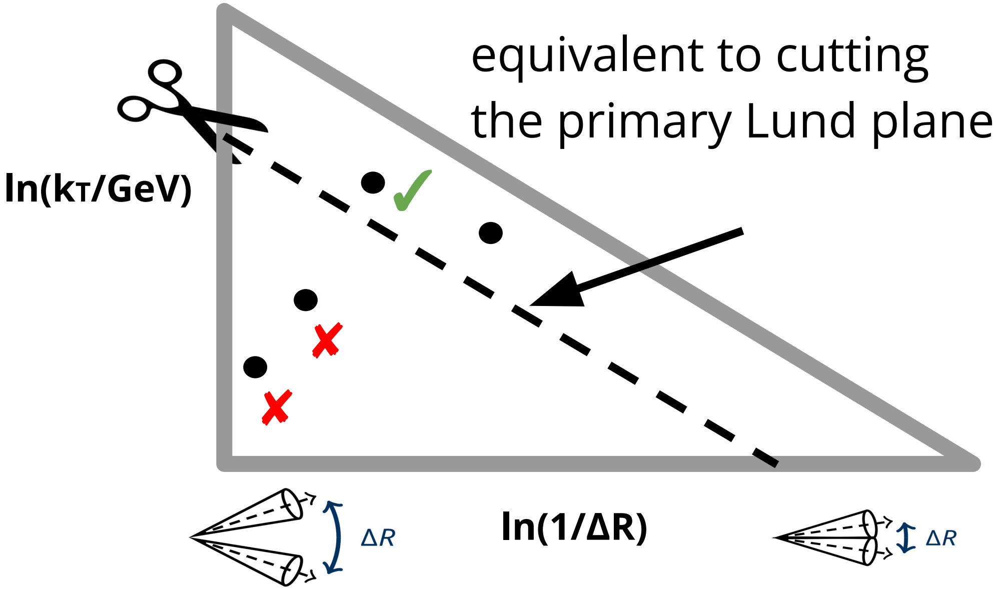
M. Dasgupta, A. Fregoso, S. Marzani, G. Salam, JHEP 1405 (2014) 146

Follow primary CA tree until you find the subjet pair that satisfies the **soft-drop** condition:

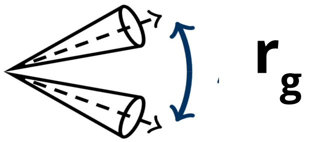
$$z_g = \frac{\min(p_T^{(1)}, p_T^{(2)})}{p_T^{(1)} + p_T^{(2)}} > z_{\rm cut} \left(\frac{\Delta R_{12}}{R} \right)^{\beta_{\rm sd}},$$

β and z_{cut} are set by the user (e.g., β= 0, z_{cut} = 0.1)

equivalent to cutting the primary Lund plane



Hard two-prong structure is exposed

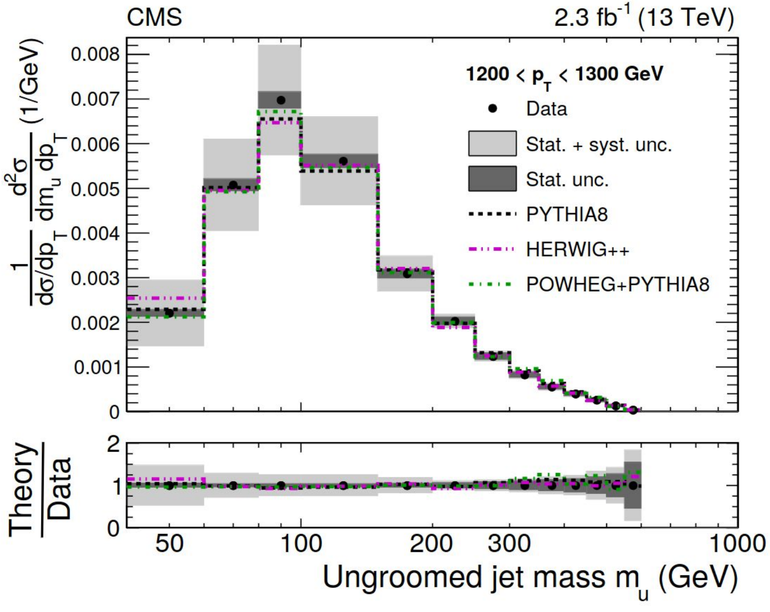


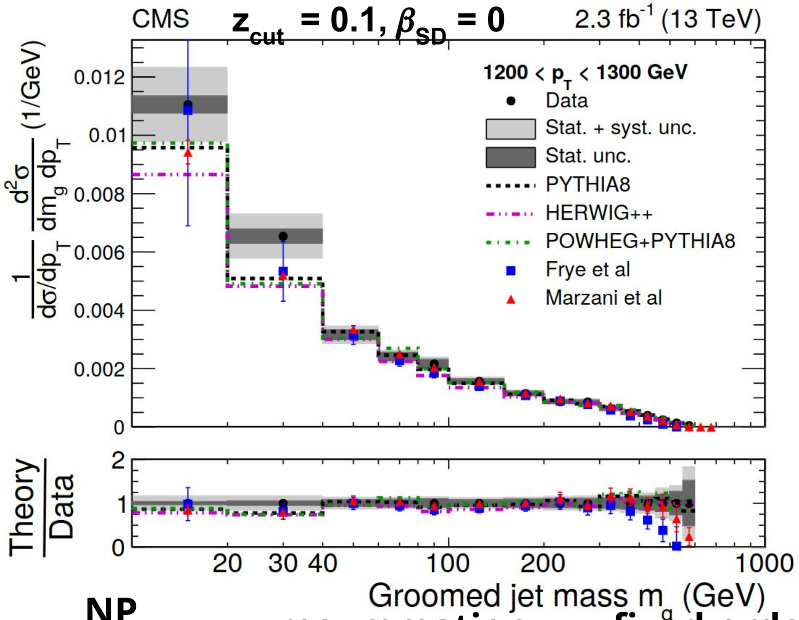
Soft-wide angle radiation removed

one-splitting observables (e.g., z_g , R_g, k_T, ...) or substructure of the groomed jet (e.g., generalized angularities, mass, ...)

ungroomed vs groomed jet mass at 13 TeV

JHEP 11 (2018) 113





NP **resummation** **fixed-order**

Initial clustering with anti-k_T R = 0.8 jets

C. Frye, A. J. Larkoski, M. D. Schwartz, K. Yan, JHEP 07 (2016) 064

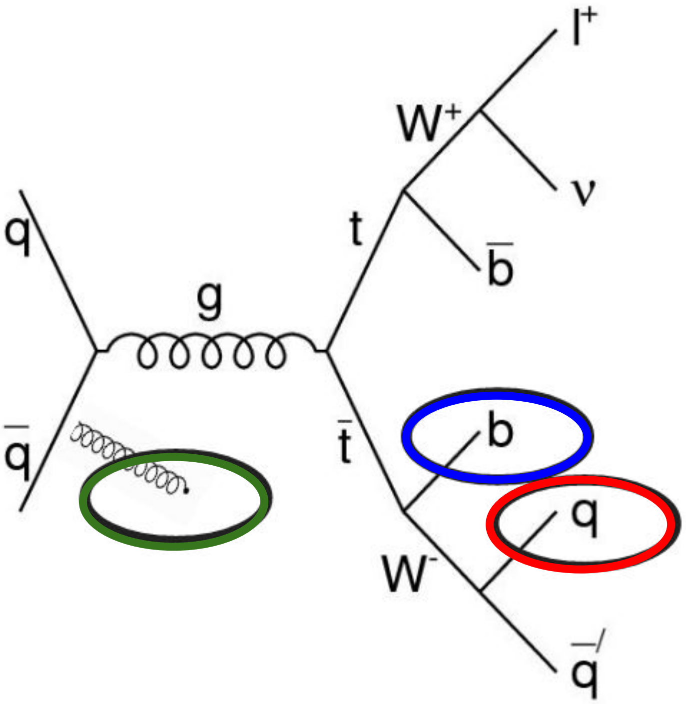
S. Marzani, L. Schunk, and G. Soyez, JHEP 07 (2017) 132

Dominant uncertainties:
* Physics model used for unfolding (PYTHIA8 vs. HERWIG++)
* jet mass scale & resolution

Also measured in pp&PbPb at 5.02 TeV, shown by Yi yesterday

Jet substructure in top quark pair + jet events

PRD98, 092014 (2018)



Can select samples enriched in:

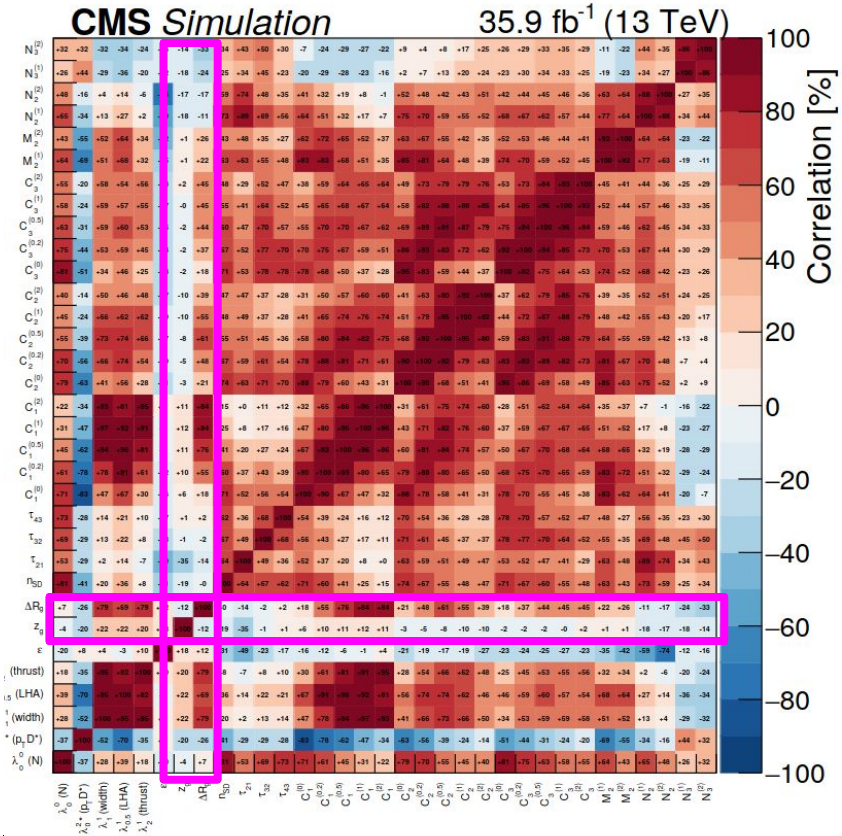
bottom quark: from top quark decay
light-quark enriched: from W boson decay
gluon-enriched: not b-tagged nor W decay

Narrow jets anti-k_T R = 0.4, using charged-only or charged+neutral particles for substructure

Jet substructure in top quark pair + jet events

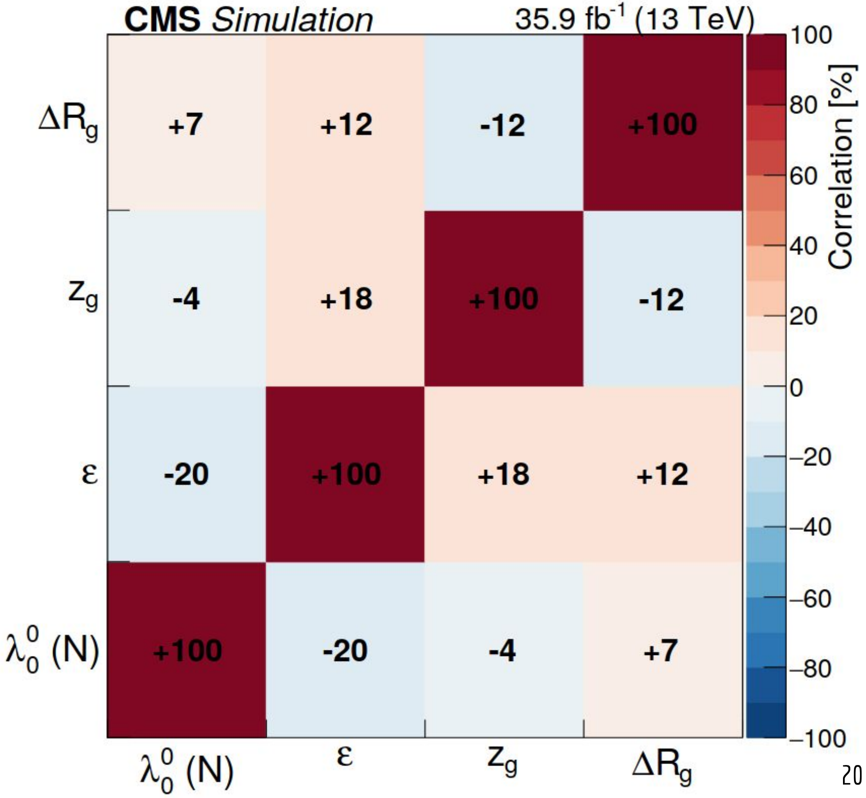
[PRD98, 092014 (2018)]

33 observables considered (EECs, angularities, Nsubjettiness, ...)

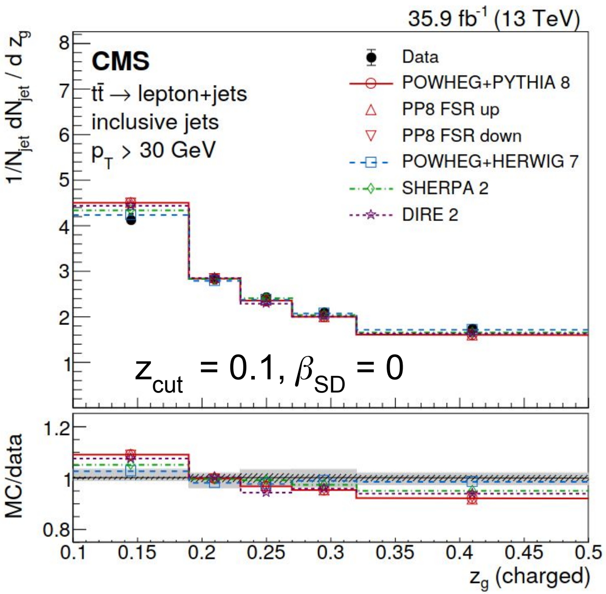


Four least correlated variables:

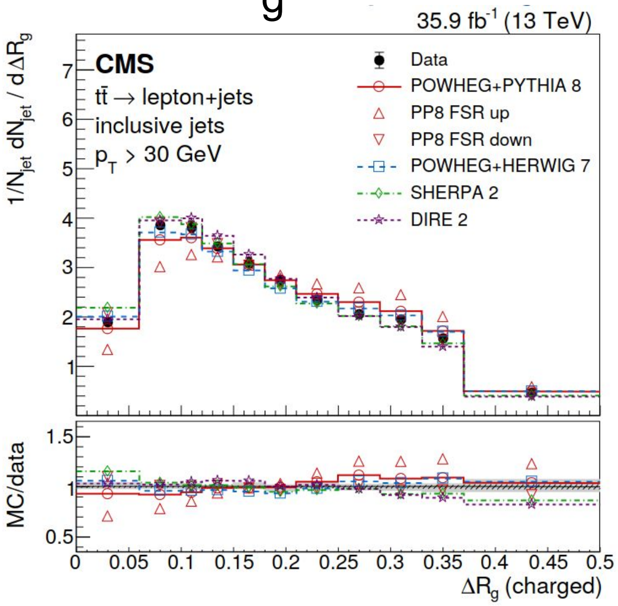
ΔR_g, z_g, multiplicity (λ_{00}), and eccentricity (ϵ)



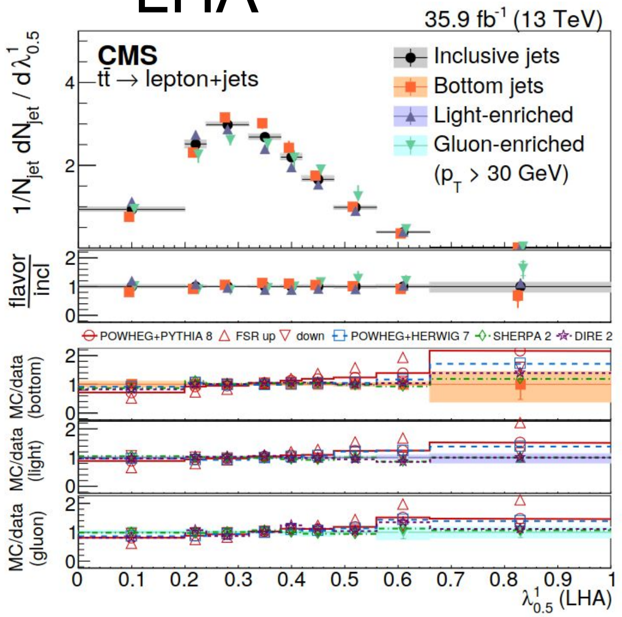
z_g



ΔR_g



LHA



- Groomed momentum fraction z_g, related to the splitting function of QCD, at LO insensitive to α_S
- Angle between groomed subjets ΔR_g: sensitivity to α_S, robust against nonperturbative effects.
- Broadening in LHA: **gluon** > **bottom** > **light-quark** jets

PRD98, 092014 (2018)

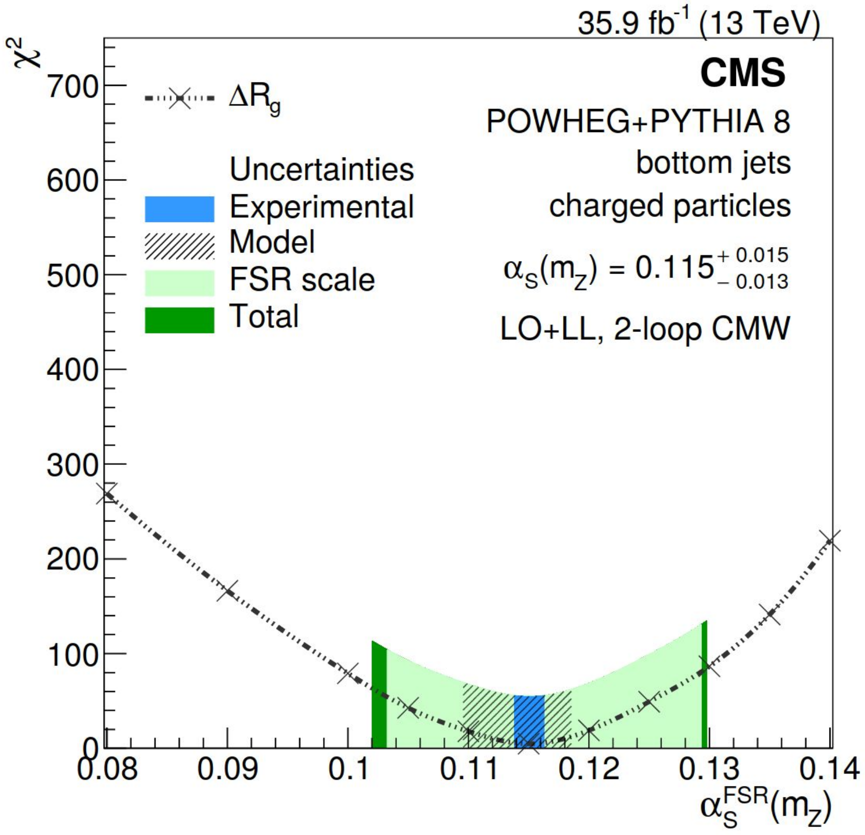
α_S extraction from R_g distribution of b-jets

- R_g of b-jets is used for the extraction of α_S (effectively LO+LL b-jet substructure)

- Result from fit to data (POWHEG+PYTHIA8)

$$\alpha_S(m_Z) = 0.115^{+0.015}_{-0.013}$$

dominated by renorm. scale uncertainties in FSR shower (~10% effect on α_s)



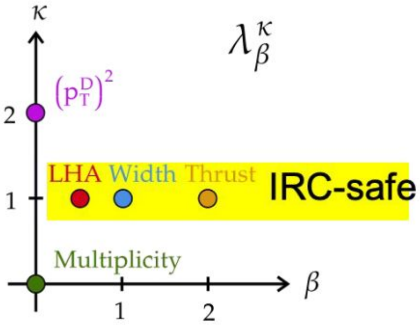
Cristian Baldenegro (LLR)

Generalized angularities in dijet and Z+jet events

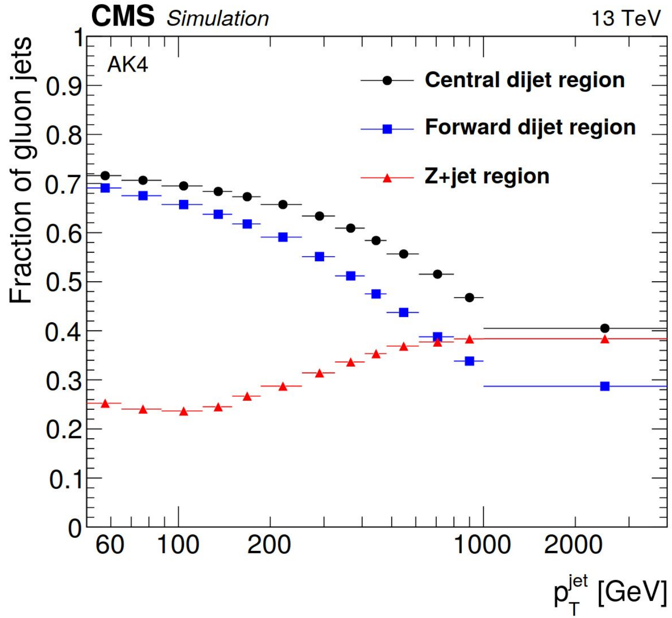
CMS, arXiv:2109.03340, *JHEP 01 (2022) 188*

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

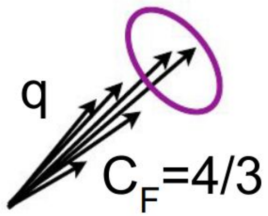
κ & β are parameters set by user



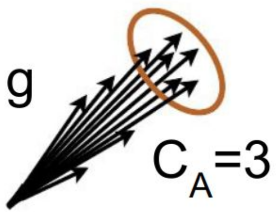
Sensitive to quark vs gluon differences
(subset of them are IRC-safe)
JHEP 1707 (2017) 091



Z+jet (quark-like)



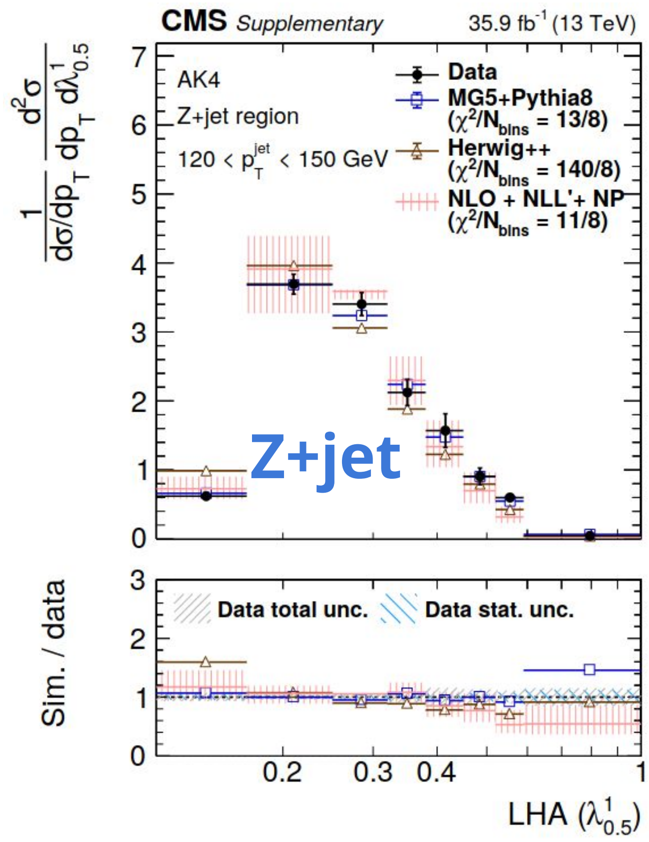
Dijet (gluon-like)

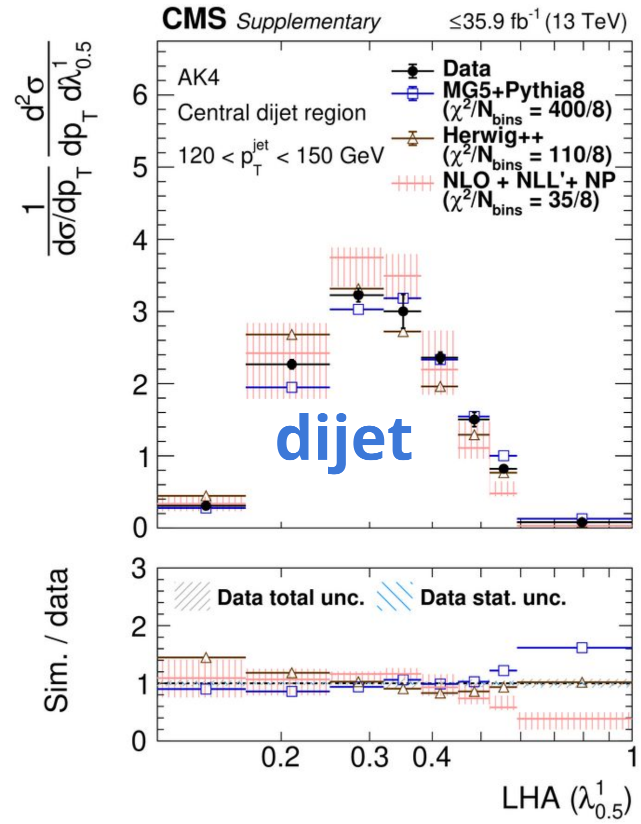


Ungroomed vs groomed with z_{cut} = 0.1, β_{SD} = 0,
R = 0.4 vs R =0.8
charged-only vs charged+neutrals

Ungroomed Les Houches Angularity in Z-jet and dijet events

JHEP 01 (2022) 188





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

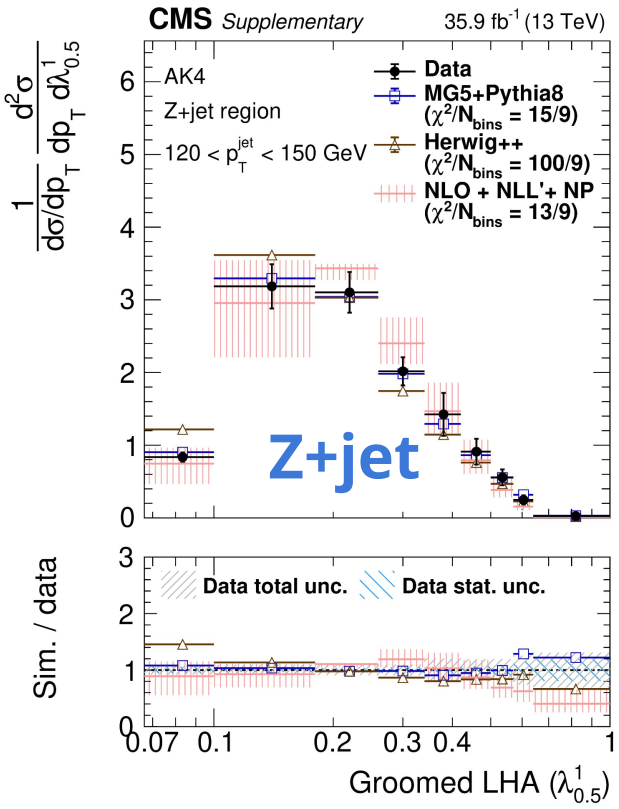
More challenging to describe **gluon-enriched jets**

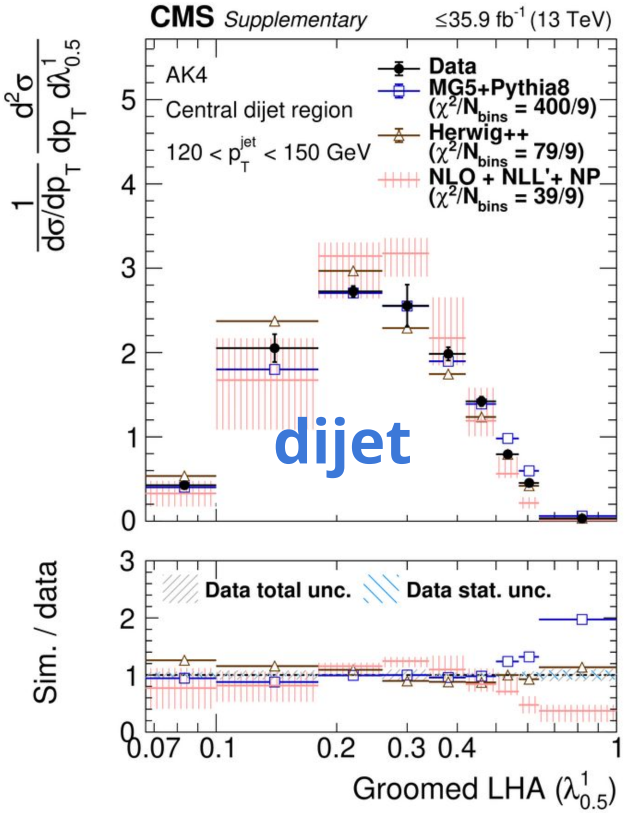
Differences at large LHA increase after removing soft&wide-angle radiation

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)

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

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

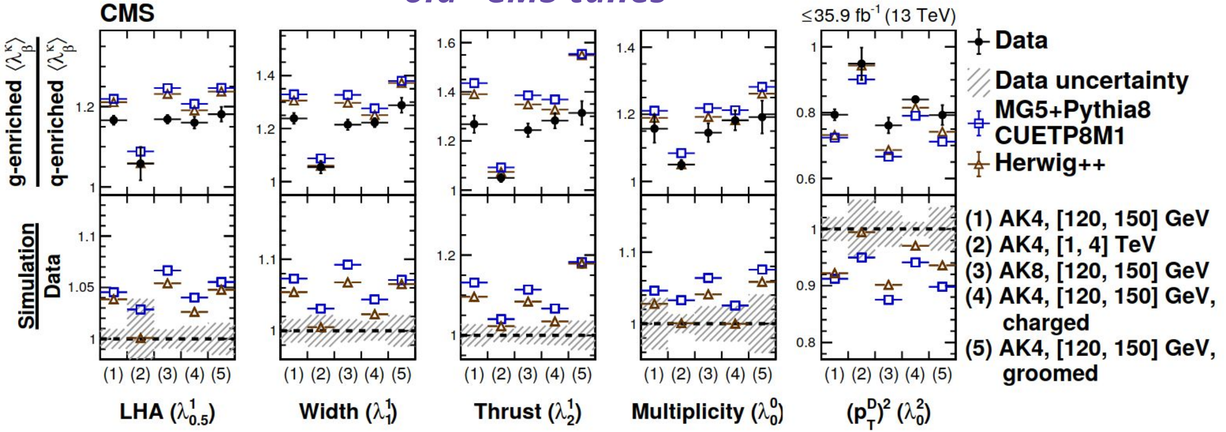
pQCD calculations D. Reichelt, S. Caletti, O. Fedkevych, S. Marzani, S. Schumann, G. Soyez, JHEP 03 (2022) 131

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

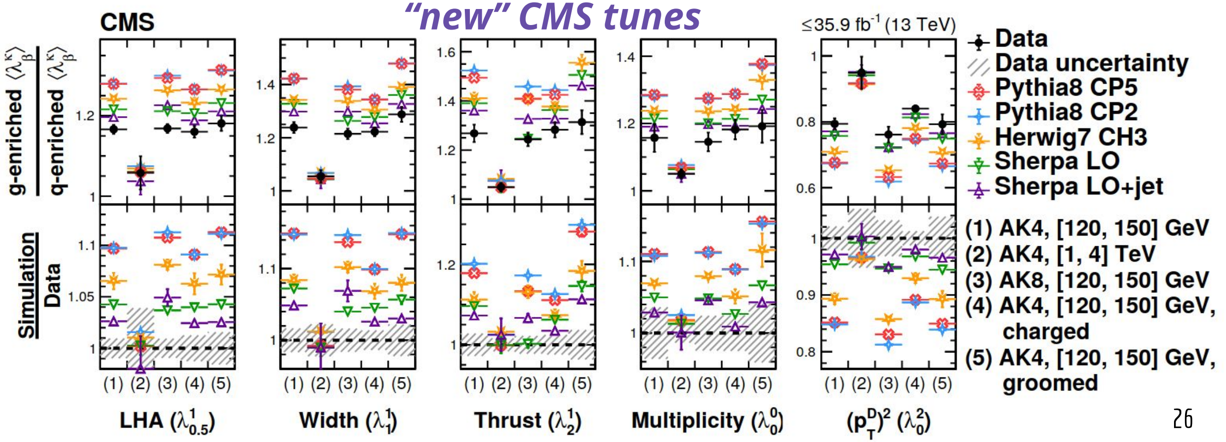
CMS, arXiv:2109.03340
JHEP 01 (2022) 188

- 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



Summary

- A number of CMS measurements based on Lund-trees of the jets.

  More related measurements in pp&PbPb presented by Yi

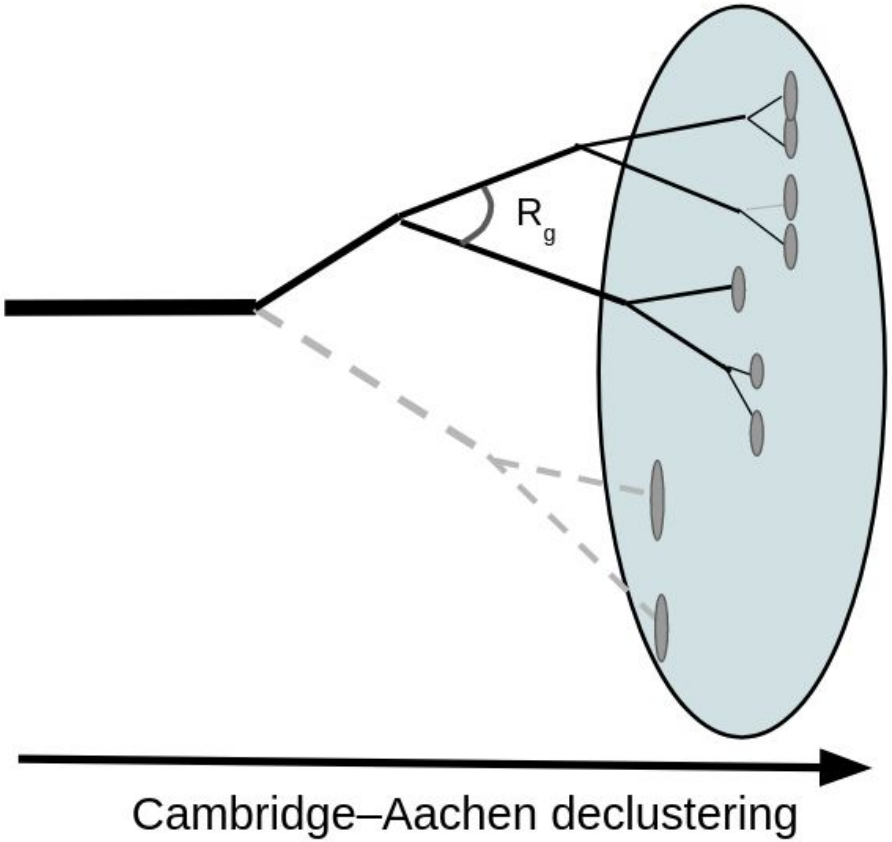
- Useful input to improve perturbative and nonperturbative ingredients

- **Measurements dominated by physics model used for unfolding corrections & detector modeling in simulation.**

groomed jet mass

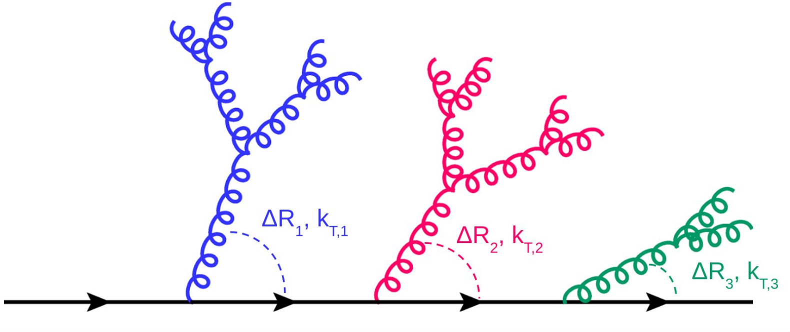
soft & wide-angle
radiation removed

allows for tests of resummation
+ fixed-order calculations

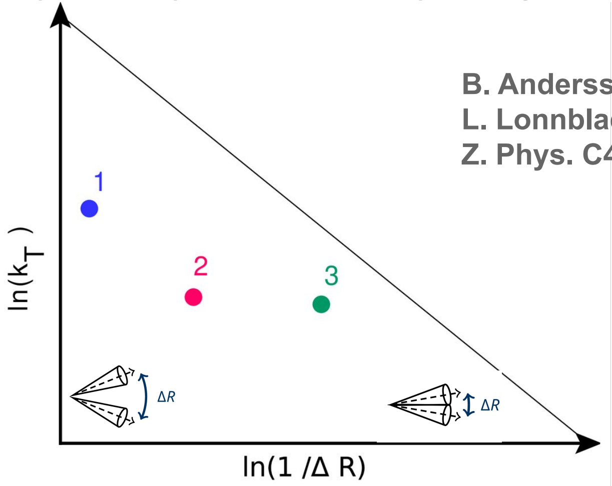


Visualizing the phase-space of QCD branchings

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

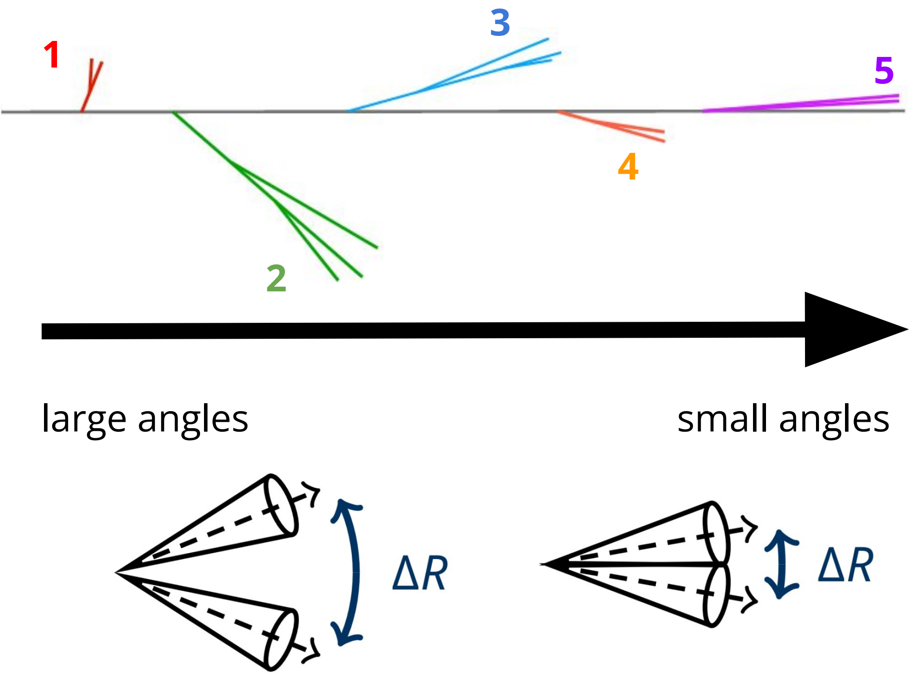
In the soft & collinear limit of QCD, emissions fill the logarithmic plane of k_T and ΔR uniformly

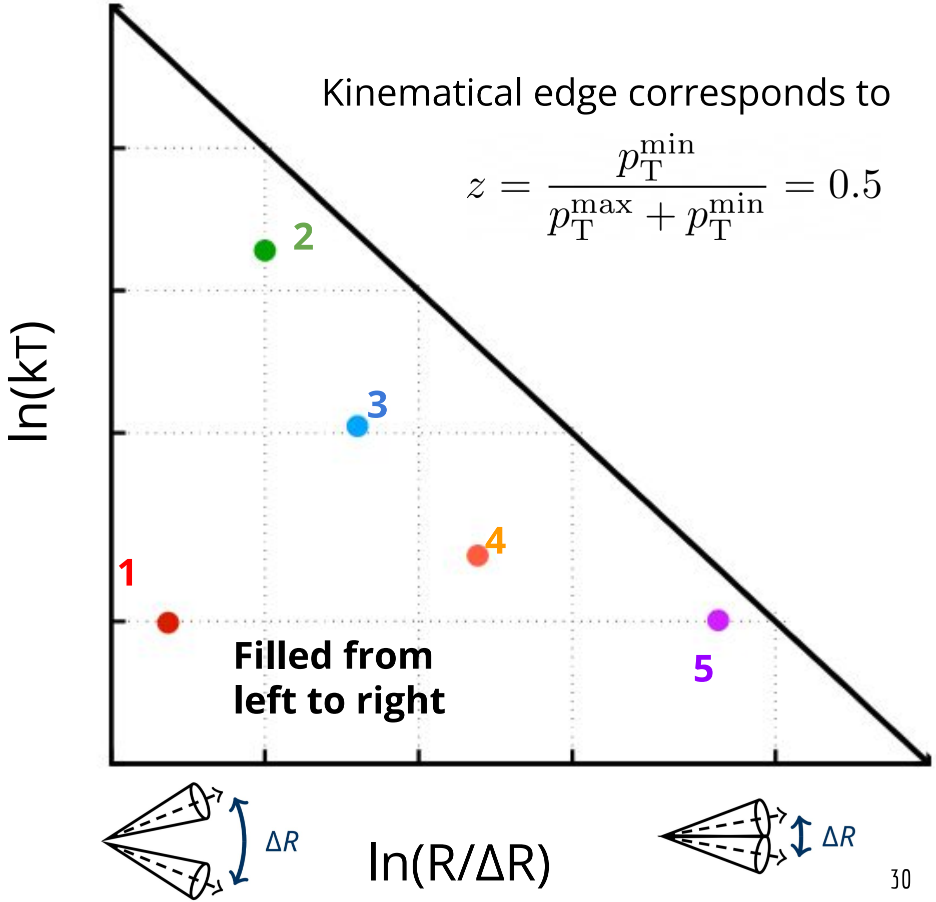
$$\mathcal{P} \propto \alpha_{\rm s} \frac{{\rm d}k_{\rm T}}{k_{\rm T}} \frac{{\rm d}\Delta R}{\Delta R} = \alpha_{\rm s} {\rm d}\ln(k_{\rm T}) {\rm d}\ln(\Delta R)$$ ← **approximate self-similarity of QCD**

Lund planes have been used for parton showers & resummations.

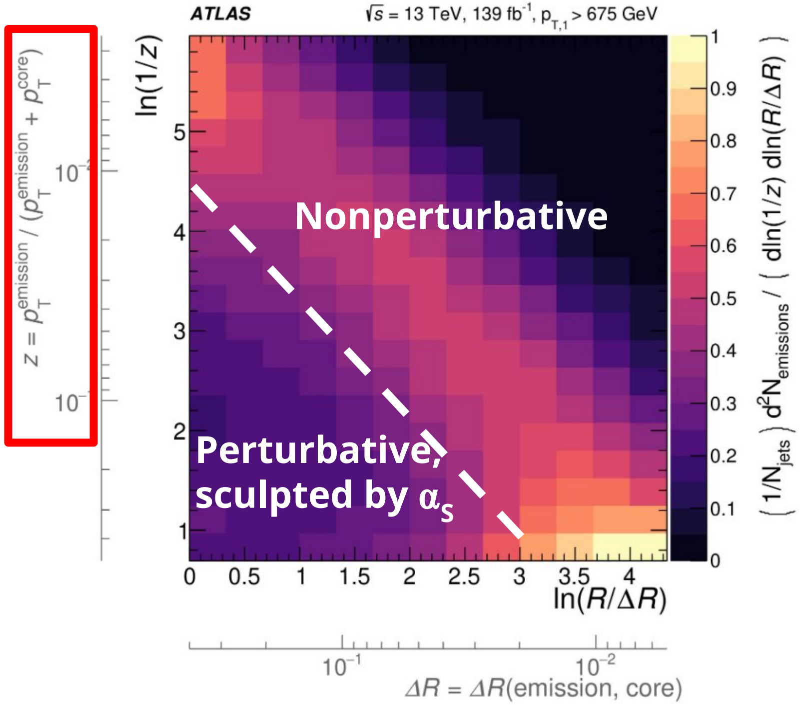
Can we visualize QCD branchings like this with data?

A given jet is represented as a number of points in the primary Lund plane



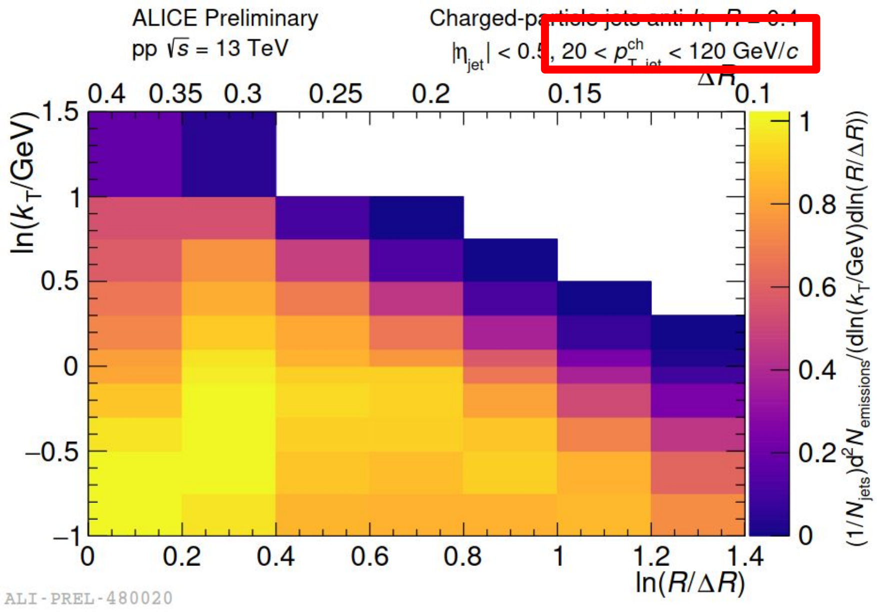


ATLAS & ALICE Lund plane measurements



Phys. Rev. Lett. 124, 222002 (2020)

momentum fraction $z = p_{T,2}/(p_{T,1}+p_{T,2})$



https://cds.cern.ch/record/2759456

ALICE used R = 0.4 jets with 20 < p_T^{jet} < 120 GeV using the $\ln(k_T)$ vs $\ln(R/\Delta R)$ representation. Sensitivity to low-k_T splittings at wide angles.

Decorrelating model dependence uncertainty

Two main sources of uncertainty in regularized unfolding:
the **response matrix** and the **prior distribution**

In D'Agostini iterative unfolding,

previous step **response matrix**

$$\lambda_j^{(t+1)} = \boxed{\lambda_j^{(t)}} \sum_{i=1}^{n} \frac{\boxed{K_{i,j}} y_i}{\sum_{k=1}^{p} \boxed{K_{i,k}} \boxed{\lambda_k^{(t)}}},$$

***t* = 0, MC ansatz for regularization**

$$\boxed{\boldsymbol{\lambda}^{(0)} = \boldsymbol{\lambda}^{\mathrm{MC}}}$$

We calculate the shower and hadronization uncertainty into two decorrelated components.