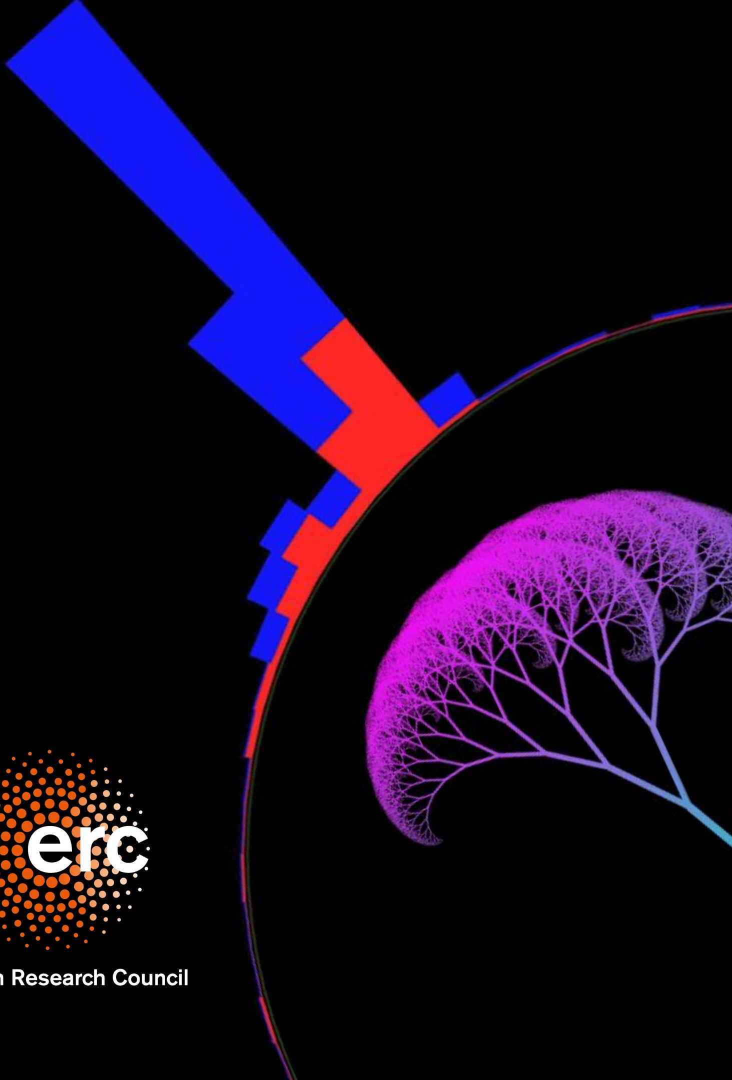
CMS results on hadronization & underlying event

Cristian Baldenegro (Sapienza Università di Roma)
on behalf of the CMS Collaboration

MPI@LHC 2023 at the University of Manchester
November 20th–25th



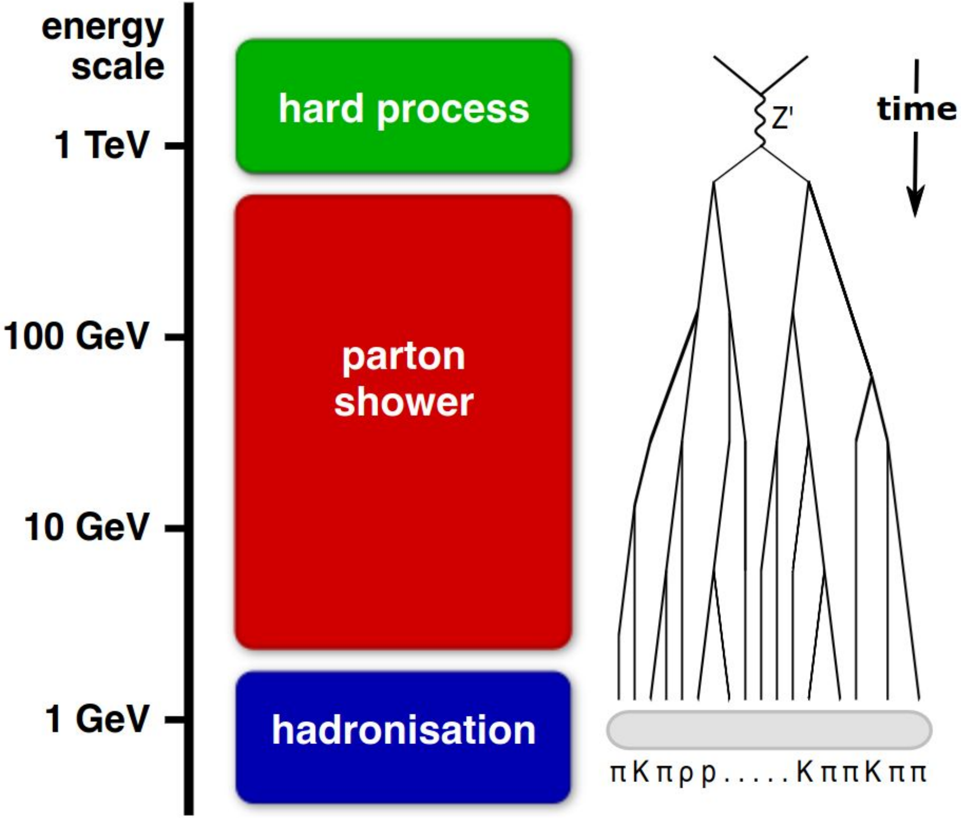
Focus on a subset of recent results:

CMS primary Lund jet plane density
(**CMS-PAS-SMP-22-007**)

Energy-energy correlations in jets
(**CMS-PAS-SMP-22-015**)

Collectivity in jets with high constituent multiplicities
(**CMS-PAS-HIN-21-013**)

Jet formation is a multiscale probe of QCD



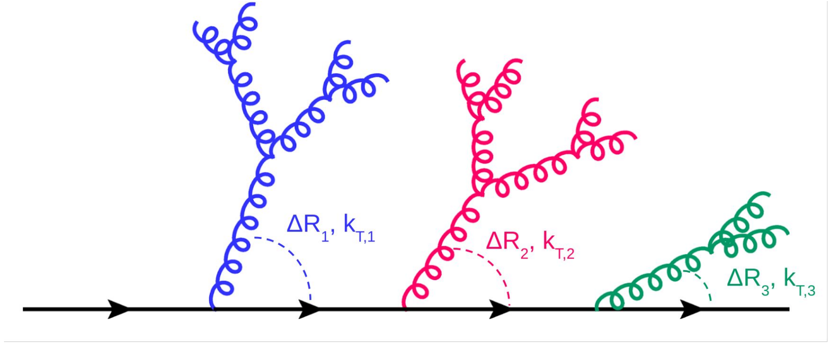
G. Salam's sketch

In principle from Q ~ 1 TeV down to Q~Λ_{QCD}

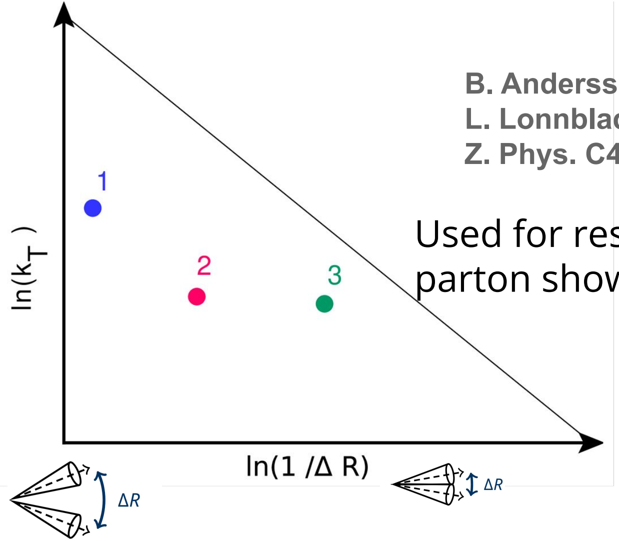
Depending on observable, jet p_T & radius *R*, one has sensitivity to parton shower, hadronization effects, underlying event, color reconnection, …

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**

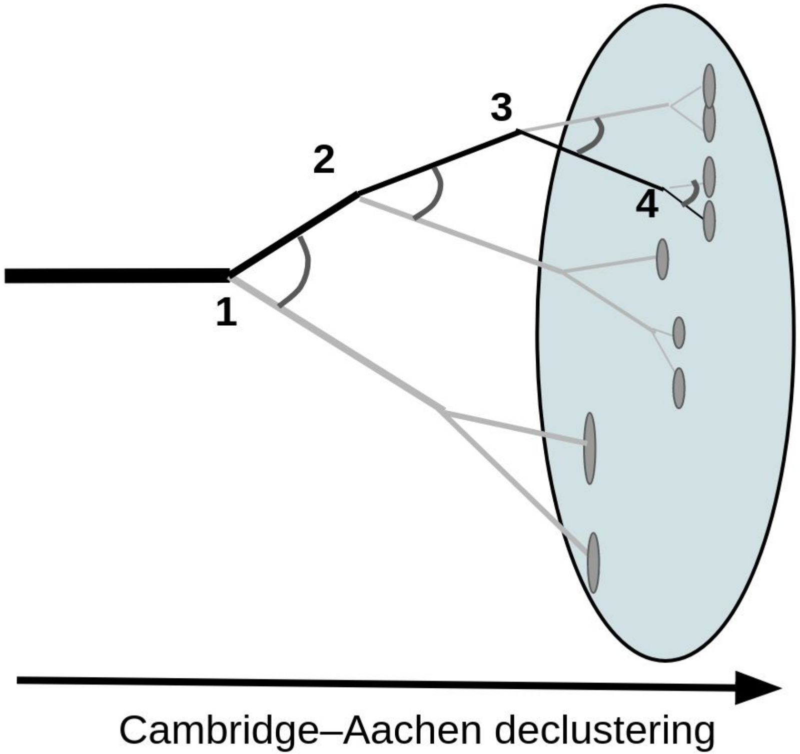
Used for resummation and parton shower development

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

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

Constructing the ***primary*** Lund jet plane (LJP)

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



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

$$\Delta R = \sqrt{(y^{\text{softer}} - y^{\text{harder}})^2 + (\phi^{\text{softer}} - \phi^{\text{harder}})^2}$$

$$k_{\mathrm{T}} = p_{\mathrm{T}}^{\text{softer}} \Delta R$$

4. Repeat until harder 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

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

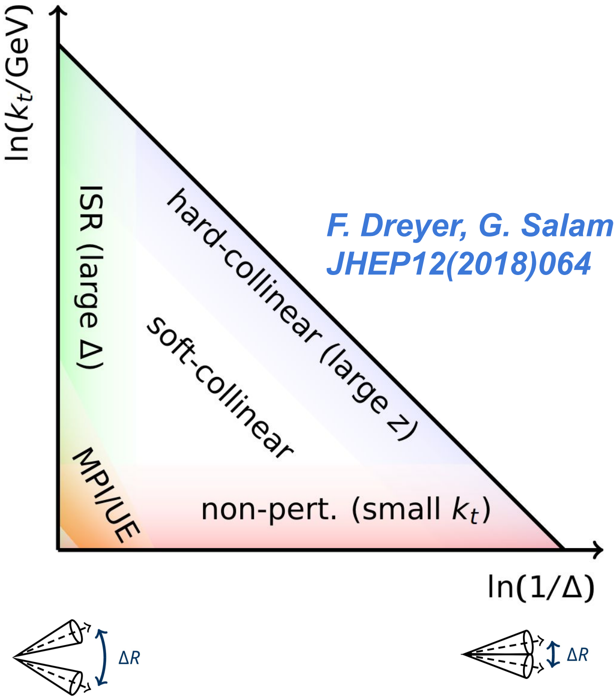
($C_R = C_A = 3$ for g→gg, $C_F = 4/3$ for q→qg)

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

- **Inclusive jet selection:**
  $p_T^{\text{jet}} > 700$ GeV, $|y^{\text{jet}}| < 1.7$,

  anti-k_T with small R = 0.4 and large R = 0.8
- Charged-particles of the jet used for Lund plane
- Distributions unfolded to particle level

Various mechanisms are separated



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

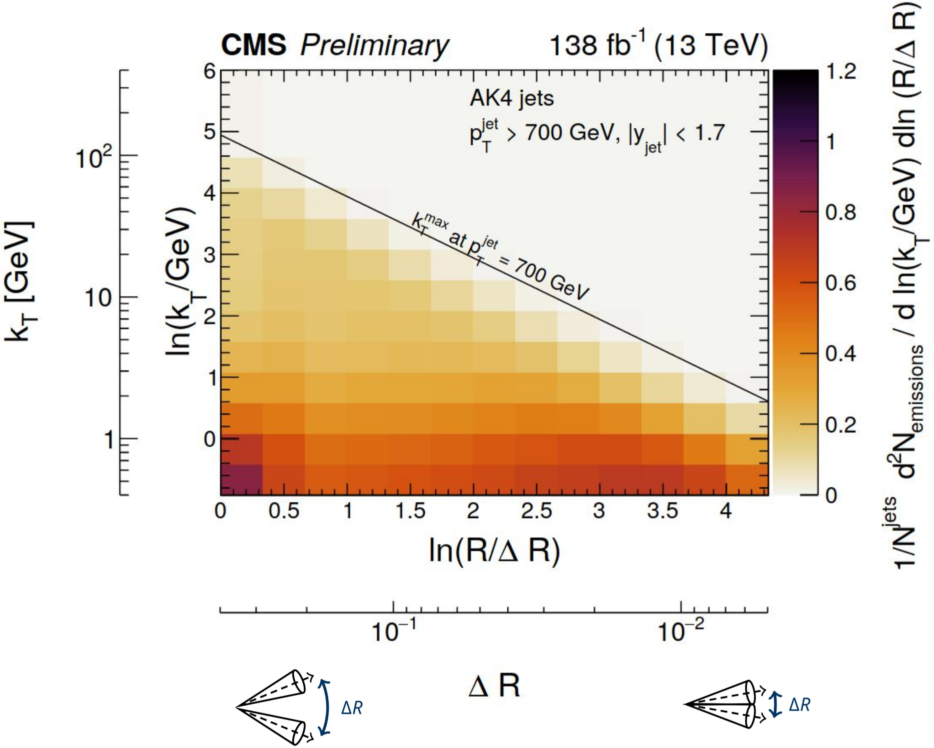
Can use Lund plane density to improve and test calculations in a “factorized” way

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

Unfolded primary Lund jet plane densities

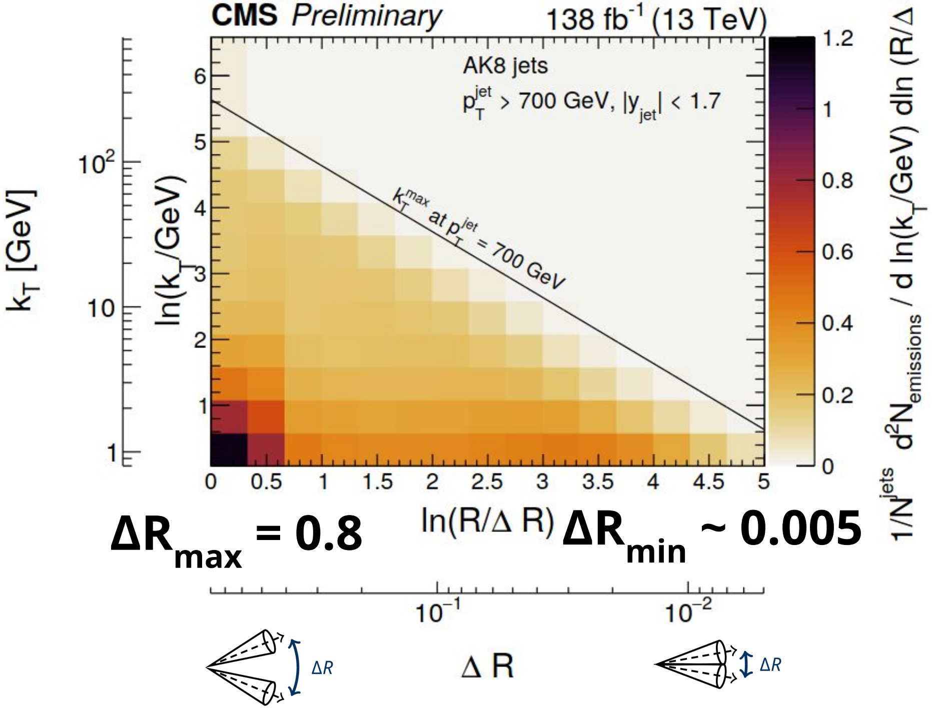
CMS-PAS-SMP-22-007

R=0.4 (standard R in Run-2)



R=0.8 (wider & harder emissions)

Up to k_T ~ 700 GeV at large angles.



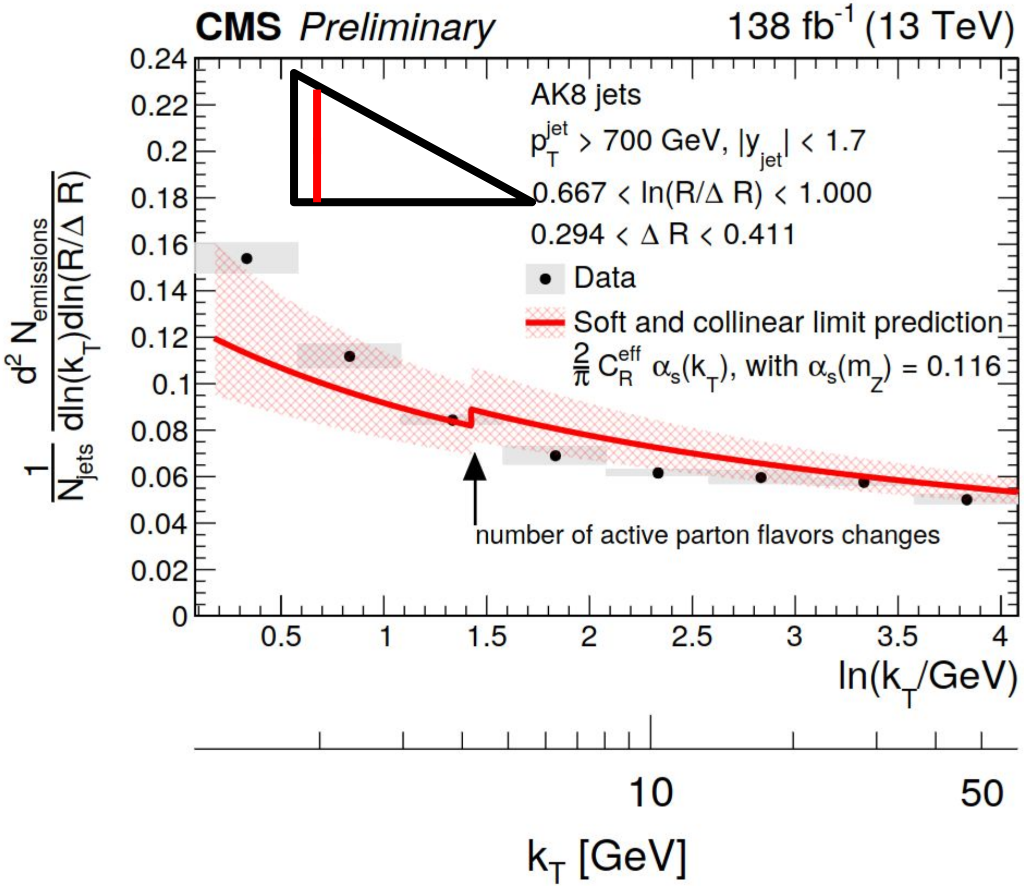
ΔR_{max} = 0.8

ΔR_{min} ~ 0.005

LJP density approximately flat for hard & collinear emissions due to running coupling $\alpha_S(k_T) \sim 1/\ln(k_T)$

Running of α_s in the jet shower!

CMS-PAS-SMP-22-007



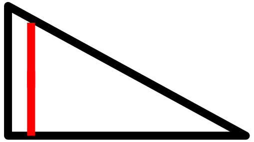
Recall LO pocket formula for Lund density:

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

Running $\alpha_s(k_T)$ from few GeV to ~60 GeV qualitatively describes the data

(Assuming q/g fractions from PYTHIA8)

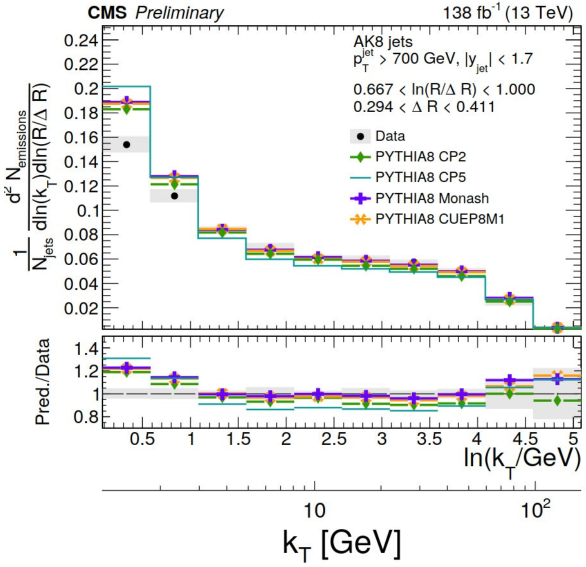
Large angle emissions



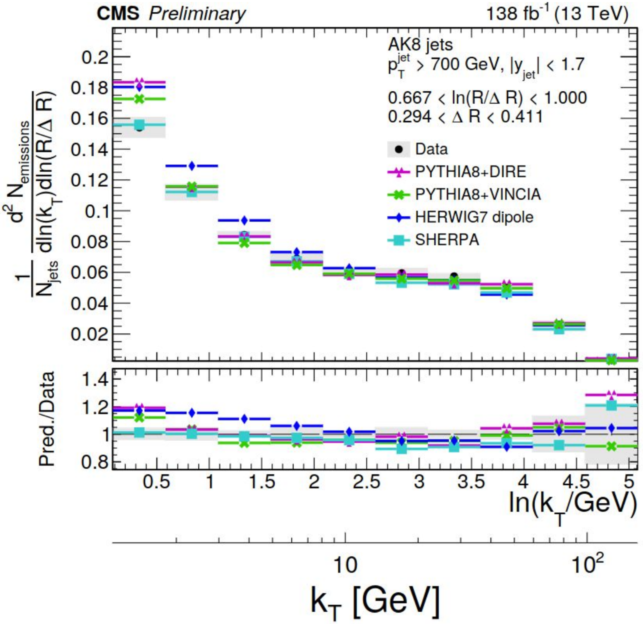
R = 0.8

Comparison to parton showers& tunes

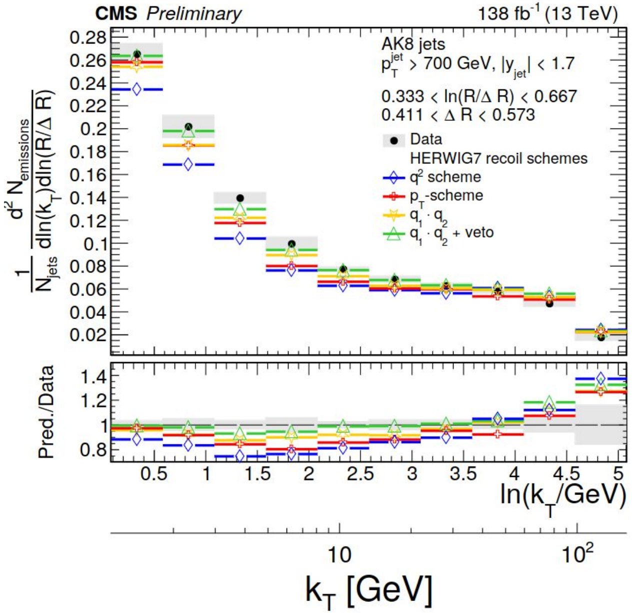
CMS-PAS-SMP-22-007



PYTHIA8 tunes
(CP2, CP5, Monash, CUEP8m1)



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

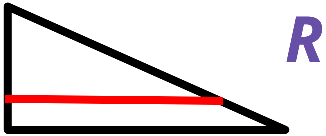


Herwig7 recoil schemes,
(angle-ordered showers)

Data/MC differences of 10–20%. Most important difference for PYTHIA8 tunes is the $\alpha_S^{FSR}(m_Z)$ value.

HERWIG7 angle-ordered describes better the data than **HERWIG7 dipole**

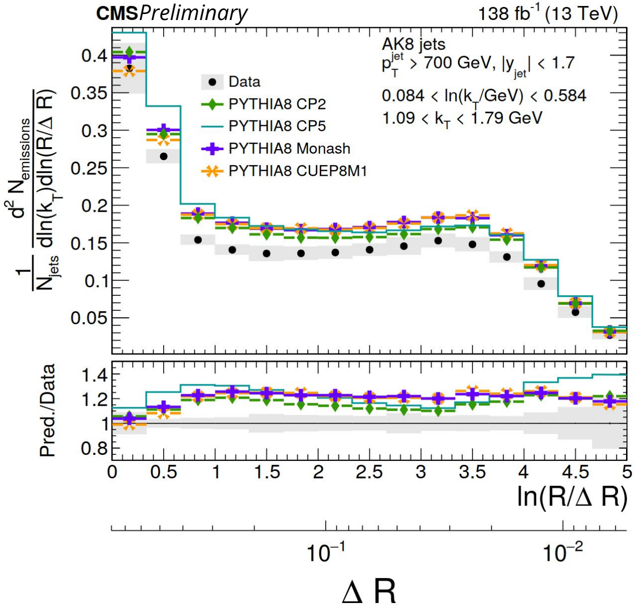
Factorization of effects can be exploited in MC tuning



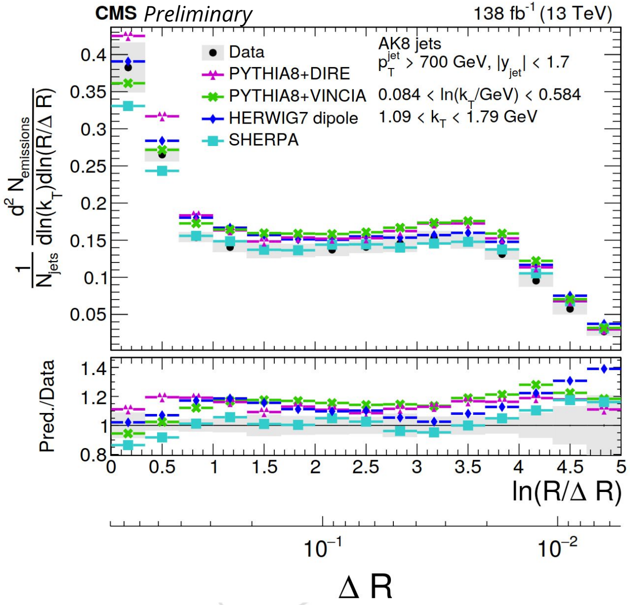
R=0.8

Low-k_T (hadronization + MPI)

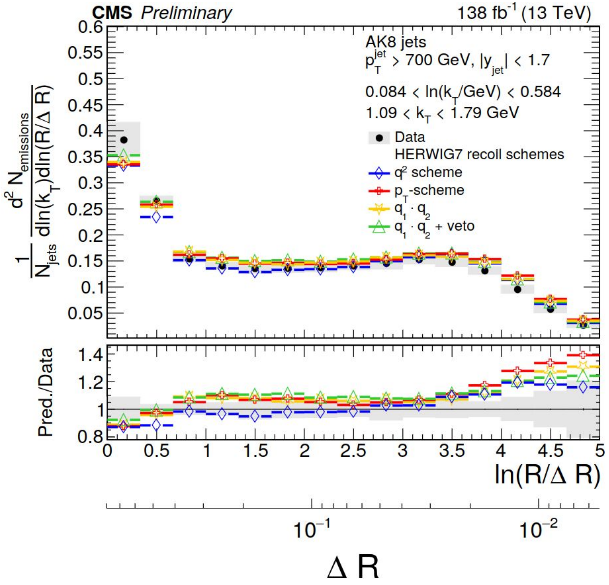
CMS-PAS-SMP-22-007



PYTHIA8 tunes
(CP2, CP5, Monash, CUEP8m1)



Dipole showers
(Vincia, Dire, Herwig7Dipole, Sherpa)



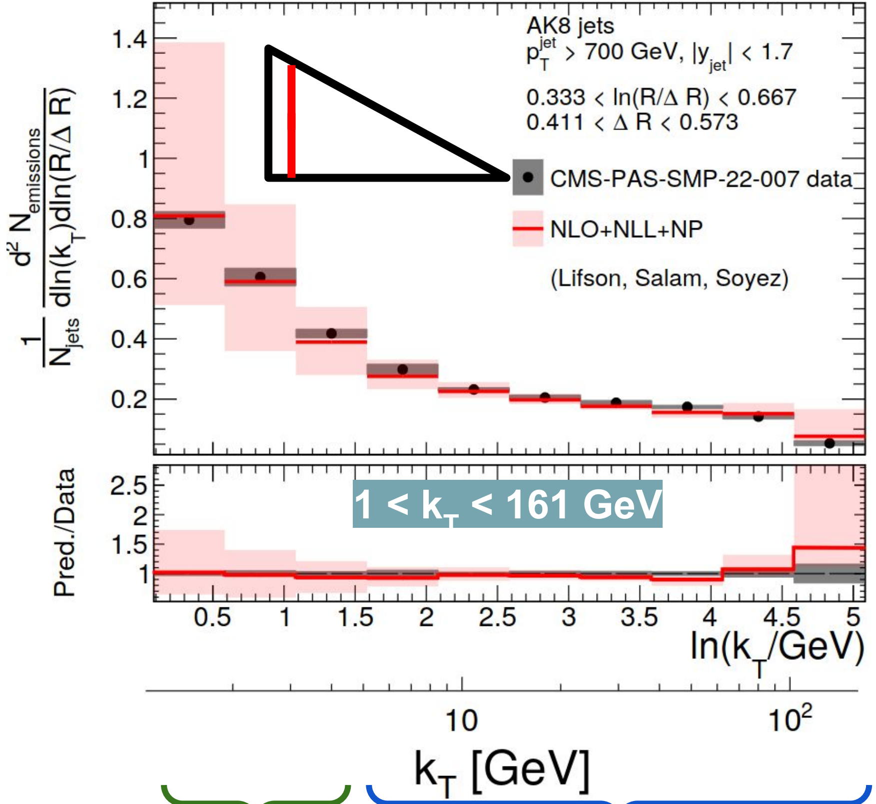
Herwig7.2 recoil schemes
(angle-ordered)

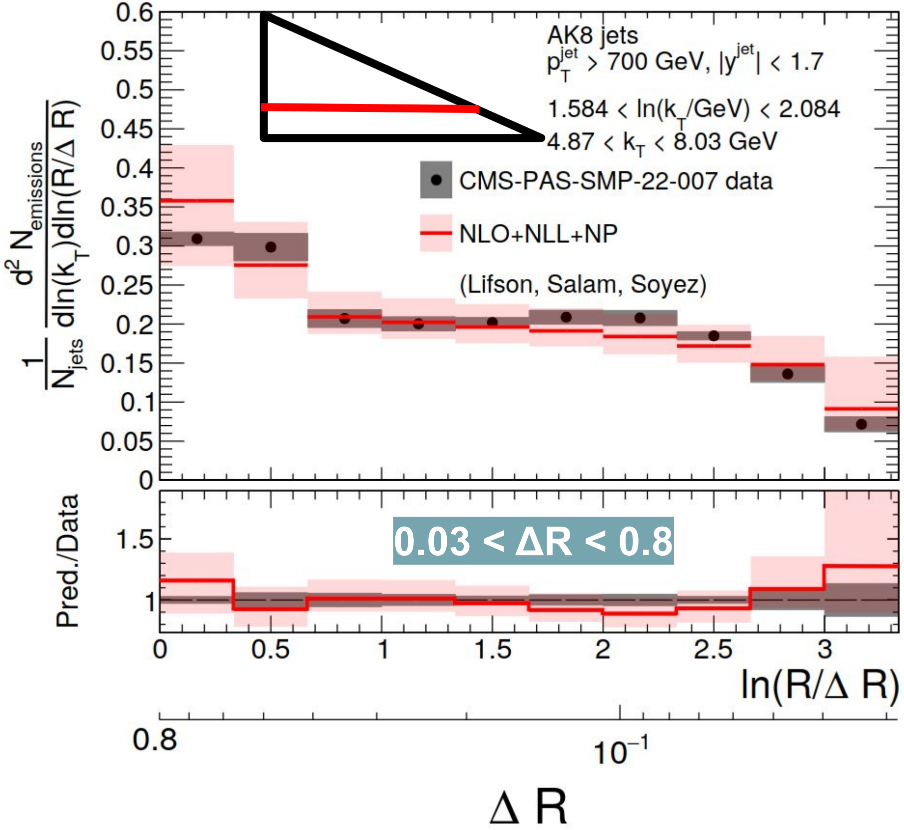
PYTHIA8 systematically overshoots LJP at low k_T by 15-20%, regardless of tune or parton shower option

HERWIG7 & **Sherpa** generally do better at low k_T. **Cluster vs string fragmentation model differences?**

pQCD analytical calculations (NLO+NLL+NP)

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





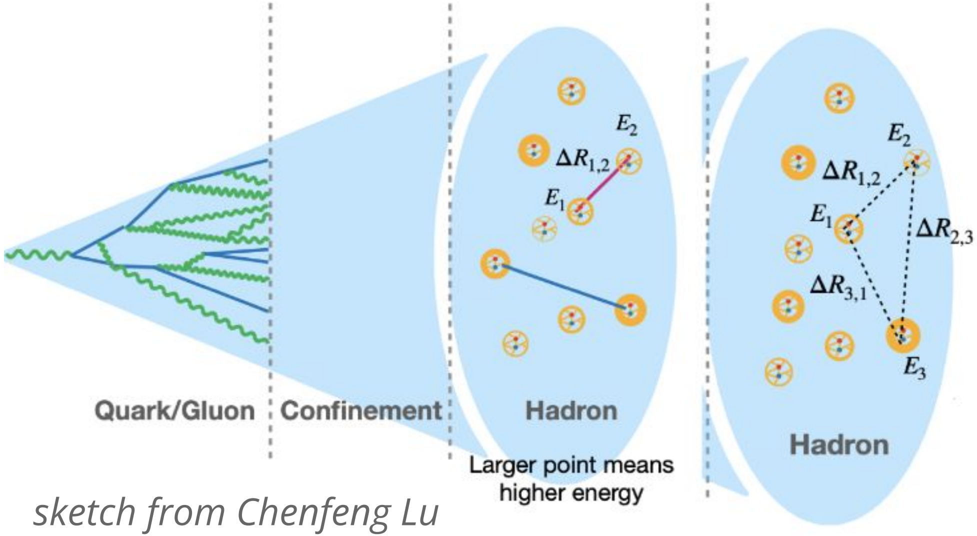
nonperturbative

**resummation
(unc. mostly perturbative)**

Nonglobal logs, clustering logs

Energy-energy correlators

Energy-weighted two-particle angular correlations



sketch from Chenfeng Lu

$$E2C = \frac{d\sigma}{dx_L} = \sum_{i,j}^{n} d\sigma \boxed{\frac{E_iE_j}{E^2}} \delta(x_L - \Delta R_{i,j})$$

$$E3C = \frac{d\sigma}{dx_L} = \sum_{i,j,k}^{n} d\sigma \boxed{\frac{E_iE_jE_k}{E^2}} \times \delta(x_L - \max(\Delta R_{i,j}, \Delta R_{i,k}, \Delta R_{j,k}))$$

Angular separation x_L == ΔR_{ij} = $\sqrt{\Delta y^2+\Delta\varphi^2}$

Energy weights: soft contributions are penalized, hard contributions are rewarded

Mapping out different stages of jet formation (small angle x_L dominated by hadronization, large x_L dominated by short distance physics)

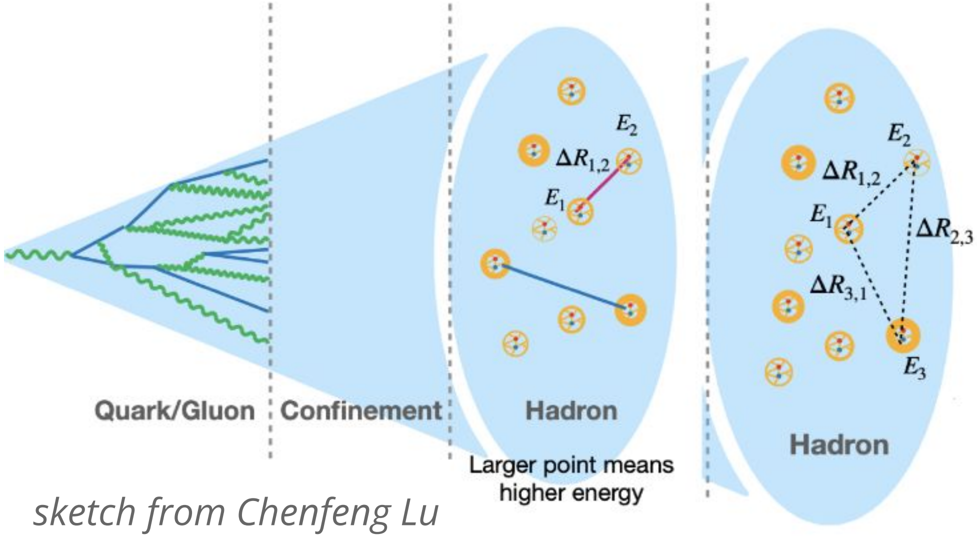
Originally proposed for e^+e^- , C.L. Basham, L. S. Brown, S. D. Ellis, S. Love PRL 41 (1978) 1585

More specific applications for jet substructure:

A. Larkoski, G. Salam, J. Thaler, JHEP 2013, 108 (2013)
Chen, Moult, Zhang, and Zhu, PRD 102, 054012 (2020)
P. Komiske, I. Moult, J. Thaler, H.X. Zhu, PRL130 (2023) 5
Lee, Meçaj, Moult, arXiv:2205.03414
Chen, Gao, Li, Xu, Zhang, Zhu, arXiv:2307.07510

Energy-energy correlators

Energy-weighted two-particle angular correlations



sketch from Chenfeng Lu

$$E2C = \frac{d\sigma}{dx_L} = \sum_{i,j}^{n} d\sigma \boxed{\frac{E_iE_j}{E^2}} \delta(x_L - \Delta R_{i,j})$$

$$E3C = \frac{d\sigma}{dx_L} = \sum_{i,j,k}^{n} d\sigma \boxed{\frac{E_iE_jE_k}{E^2}} \times \ \delta(x_L - \max(\Delta R_{i,j}, \Delta R_{i,k}, \Delta R_{j,k}))$$

Angular separation $x_L == \Delta R_{ij} = \sqrt{\Delta y^2 + \Delta\varphi^2}$

Energy weights: soft contributions are penalized, hard contributions are rewarded

Preliminary results
also by ALICE and STAR

Mapping out different stages of jet formation (small angle x_L dominated by hadronization, large x_L dominated by short distance physics)

Interesting also in heavy-ions!
Handle to expose medium resolution length:
C. Andres, F. Dominguez, R. Kunnawalkam Elayavalli, J. Holguin, C. Marquet, I. Moult PRL 130, no.26, 262301 (2023)

Energy-energy correlations with CMS

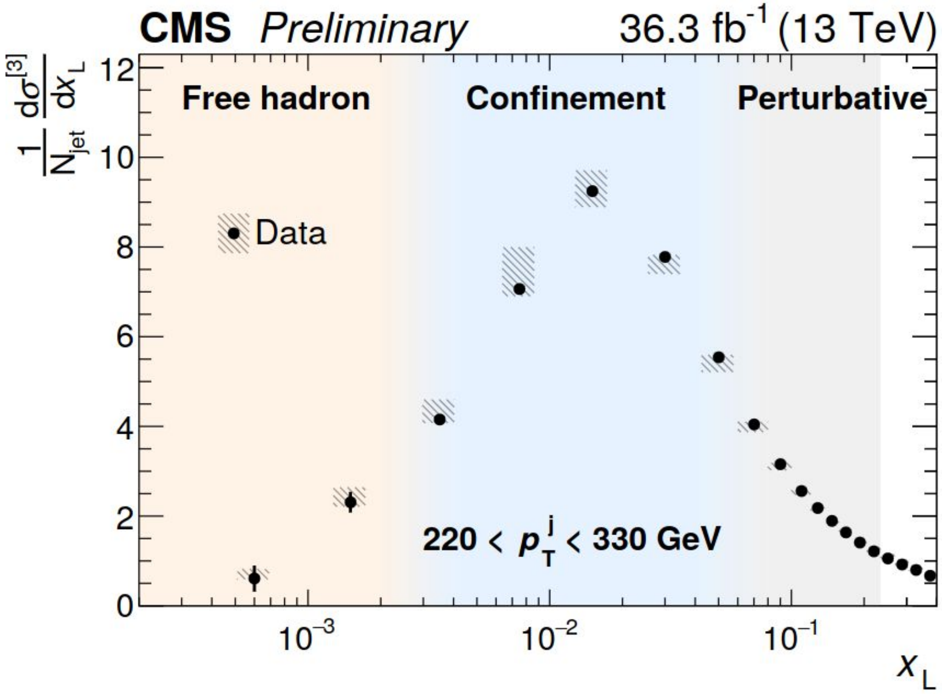
CMS-PAS-SMP-22-015

36.3 fb^{-1} of 13 TeV data,
anti-k_T R = 0.4 jets

At least two jets with $|\eta|<2.1$,
binned in jet p_T in 97~1784 GeV

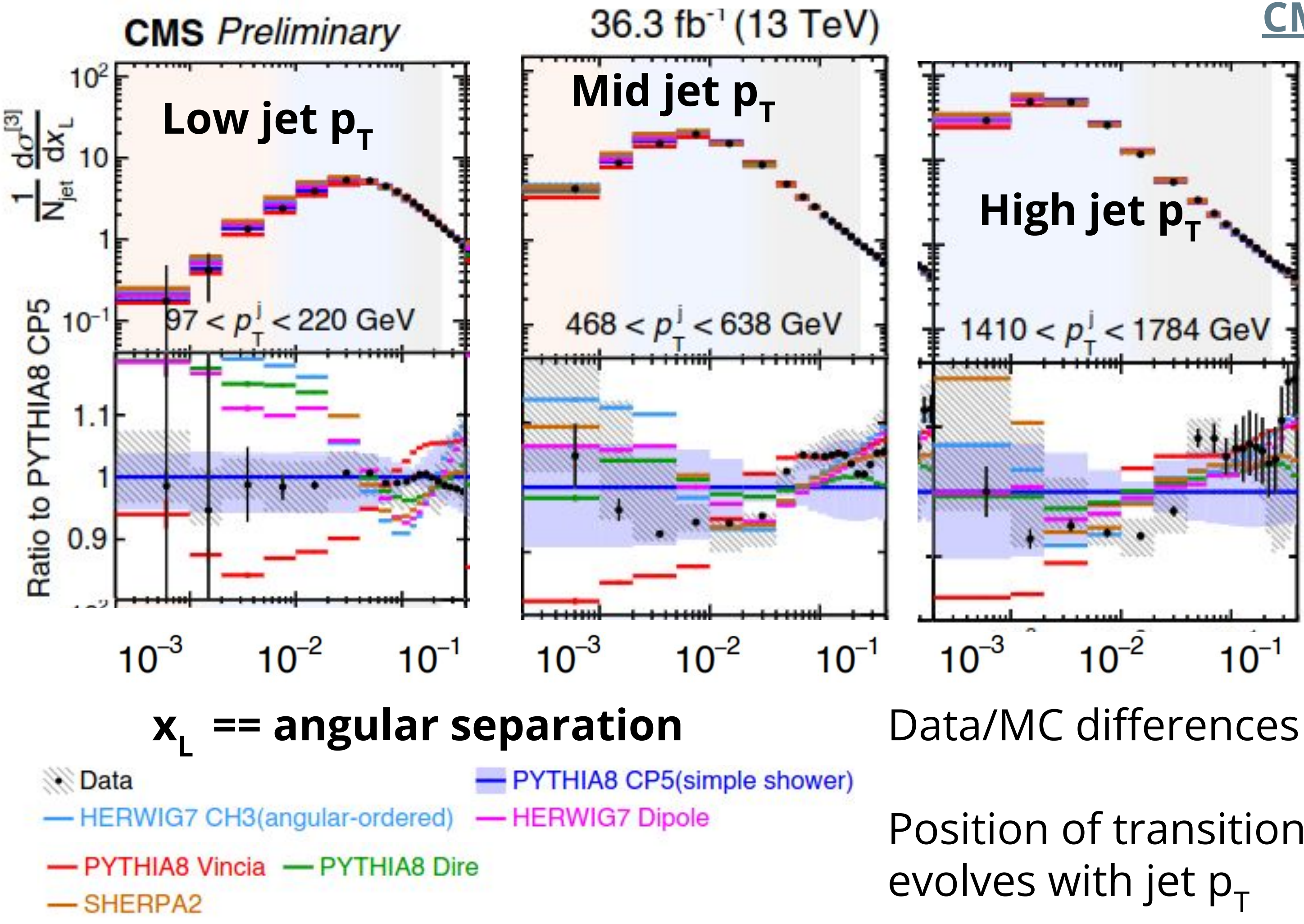
Using all particles in the jet with
p_T > 1 GeV

Distributions unfolded to stable
particle level (x_L, p_T , energy weights)



x_L == angular separation between particles

CMS-PAS-SMP-22-015



Three-point energy correlator (E3C)

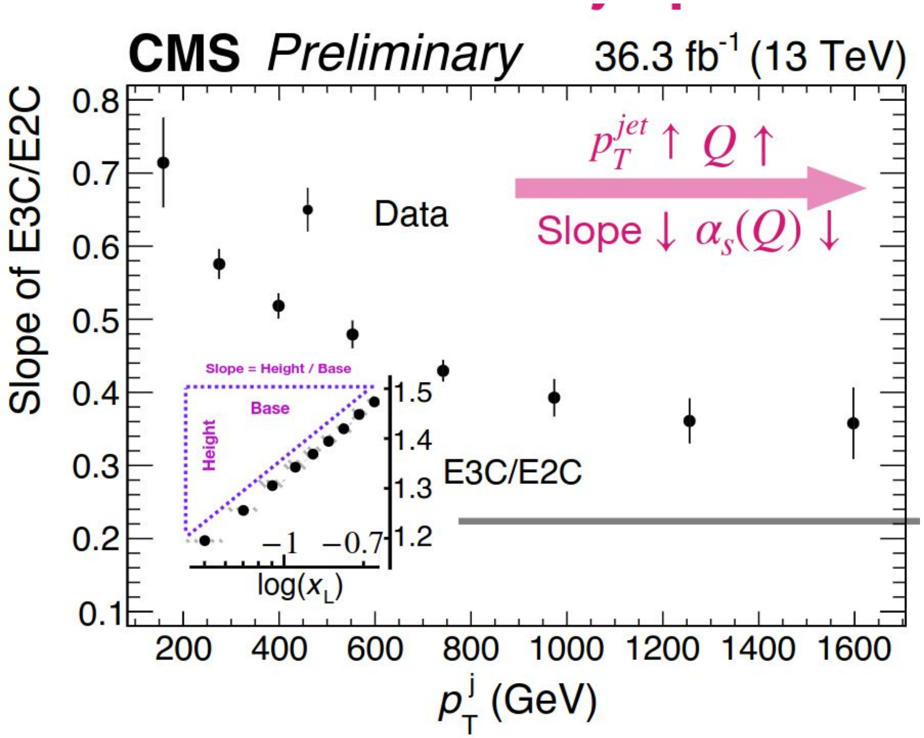
x_L == angular separation

Data/MC differences of ~10-15%

Position of transition region “bump” evolves with jet p_T

E3C/E2C sensitive to running α_s

CMS-PAS-SMP-22-015

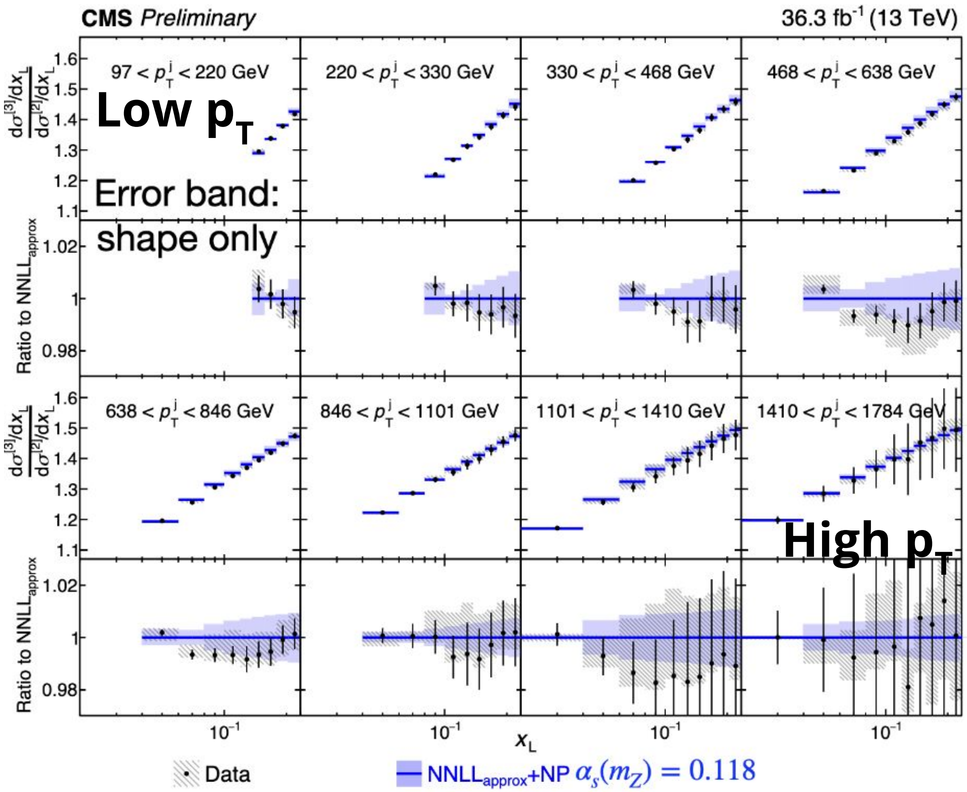


At LL, slope of E3C/E2C ratio sensitive to α_s(Q)

$$\propto \alpha_S(Q) \ln x_L + O(\alpha_s^2)$$

Quark/gluon fraction sensitivity is reduced in the E3C/E2C ratio, without losing sensitivity to α_s(Q) running

Extraction of α_s from jet substructure



Low p_T

Error band: shape only

High p_T

Fit of strong coupling (NLO+NNLL$_{approx}$ calculation)

$$\alpha_s(m_Z) = 0.1229^{+0.0040}_{-0.0050} \; (\sim 4\,\%)$$

Most precise extraction of $\alpha_s(m_Z)$ with jet substructure (dominated by theory uncertainties)

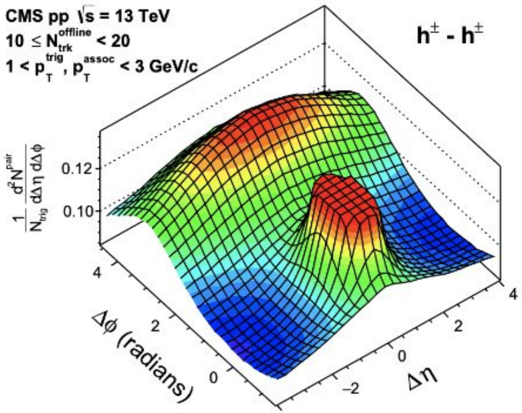
CMS-PAS-SMP-22-015

Two-particle angular correlations

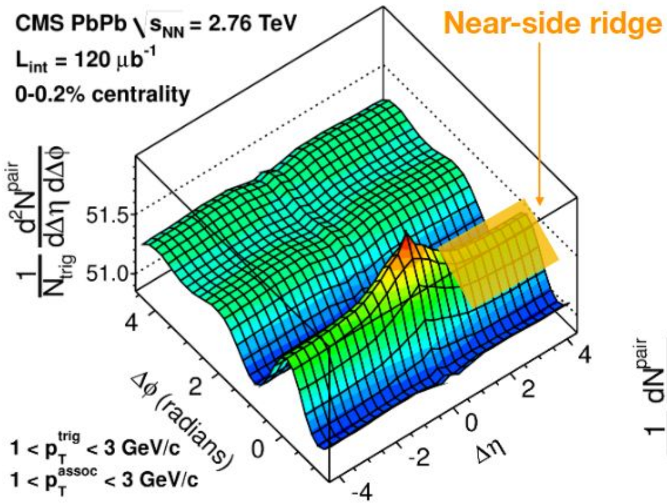
- **Near-side ridge** typical sign of collective behavior
- Fourier harmonics decomposition, nonzero $V_{2\Delta}$ associated to anisotropic expansion

$$\frac{1}{N_{\mathrm{ch}}}\frac{\mathrm{d}N^{\mathrm{pair}}}{\mathrm{d}\Delta\phi} \propto \sum_{n=1}^{\infty} \boxed{V_{\mathrm{n}\Delta}} \cos(\mathrm{n}\Delta\phi)$$

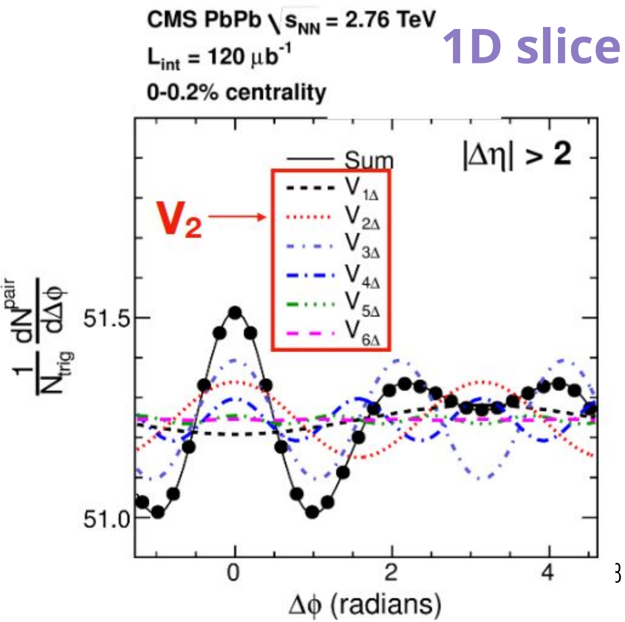
Inclusive pp



PbPb



1D slice



CMS, Phys. Lett. B 765 (2017) 193

What about smaller systems?

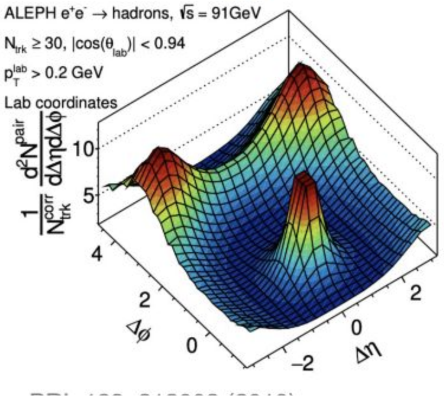
Unexpected nonzero v_2 in high-multiplicity pp and pPb by CMS (PLB 765 (2017) 193, PLB 718 (2013) 795)

→QGP formation or a general consequence of high-multiplicity QCD?

Since then, many searches pushing the boundaries towards even smaller systems (usually limited by $\mathbf{N_{ch}}$ reach)

e^+e^-

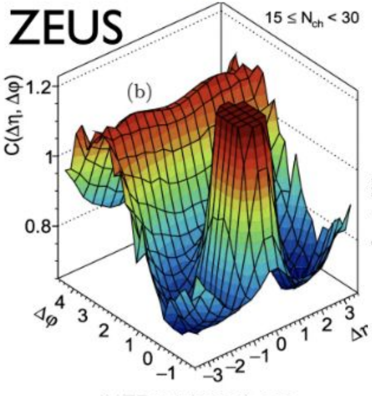
N_{ch}~30



PRL 123 212002 (2019)

ep

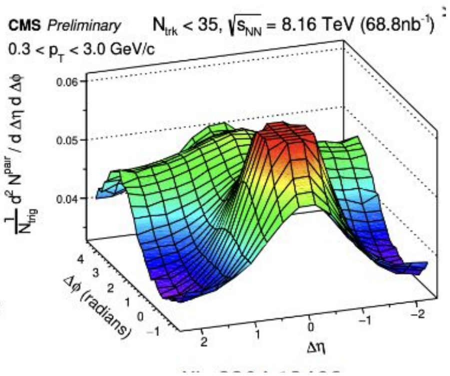
N_{ch}~30



JHEP 04 (2020) 070

γp

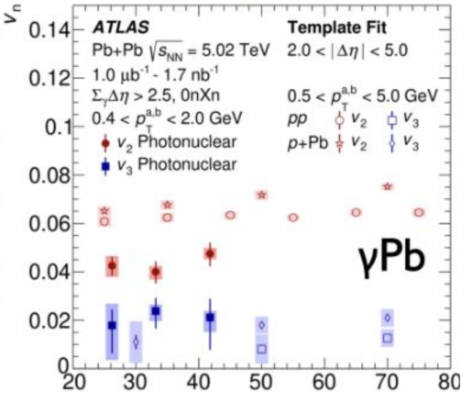
N_{ch}~20



PLB 844 (2023) 137905

γPb

N_{ch}~40



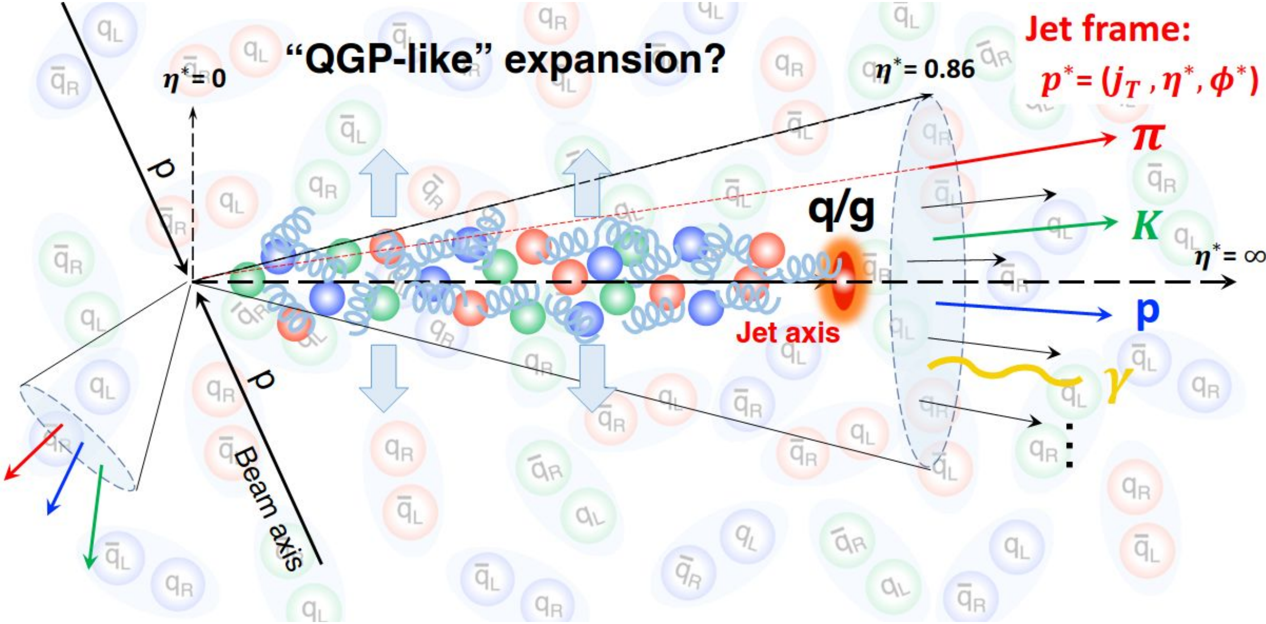
PRC, 104 014903 (2021)

Intrajet collective behavior?

Two-particle correlations in jets with high particle multiplicity

[A. Baty, P. Gardner, W. Li, Phys. Rev. C 107 (2023) 064908]

Do parton rescatterings in a ***localized*** high-density region cause any effect?



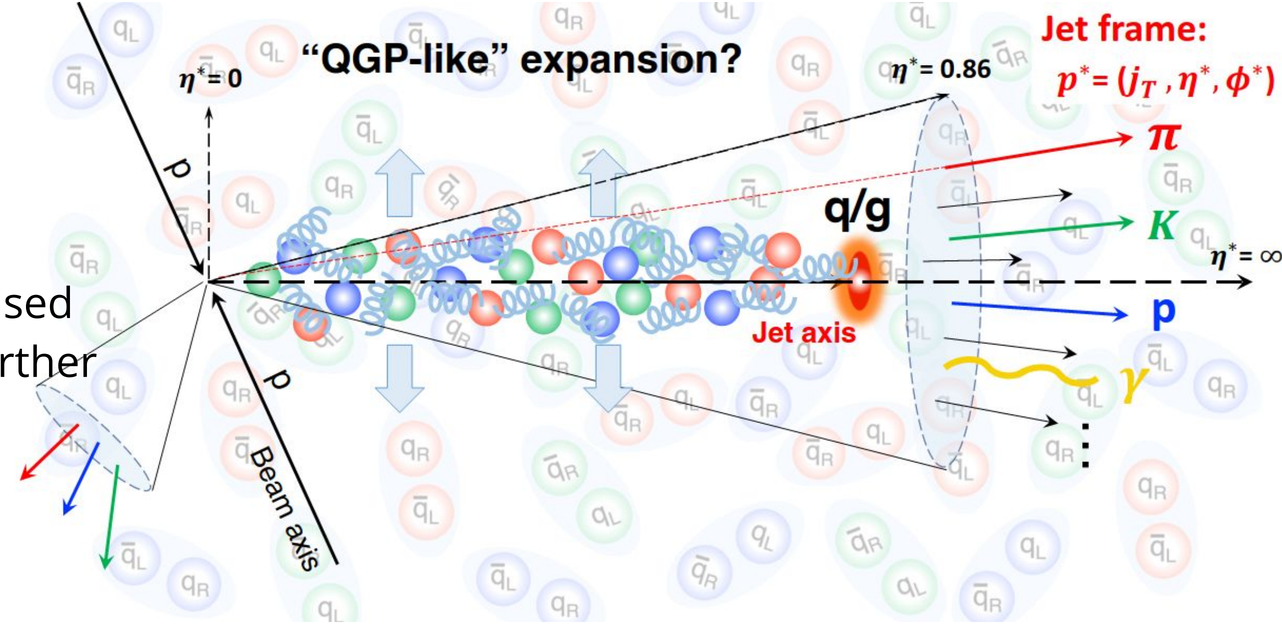
[A. Baty, P. Gardner, W. Li, Phys. Rev. C 107 (2023) 064908]

Search for intrajet collective behavior in CMS

CMS-PAS-HIN-21-013

Run-2 analysis 13 TeV analysis, PUPPI jets $p_{T,jet}$ >550 GeV, anti-k_T R = 0.8, $|\eta^{jet}| < 1.6$

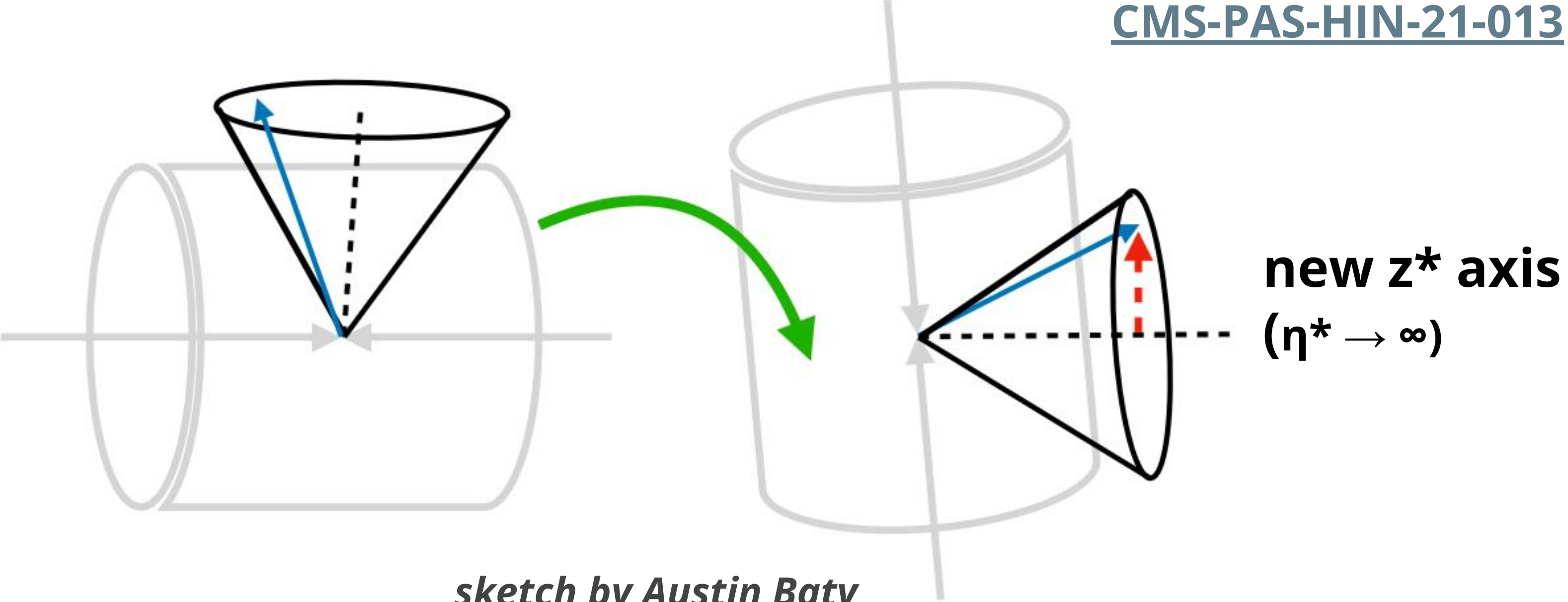
Charged-particle constituents used for two-particle correlations (further PU mitigation + low p_T reach)



A. Baty, P. Gardner, W. Li, Phys. Rev. C 107 (2023) 064908

Rotated reference frame such that z* axis is aligned with jet axis

CMS-PAS-HIN-21-013

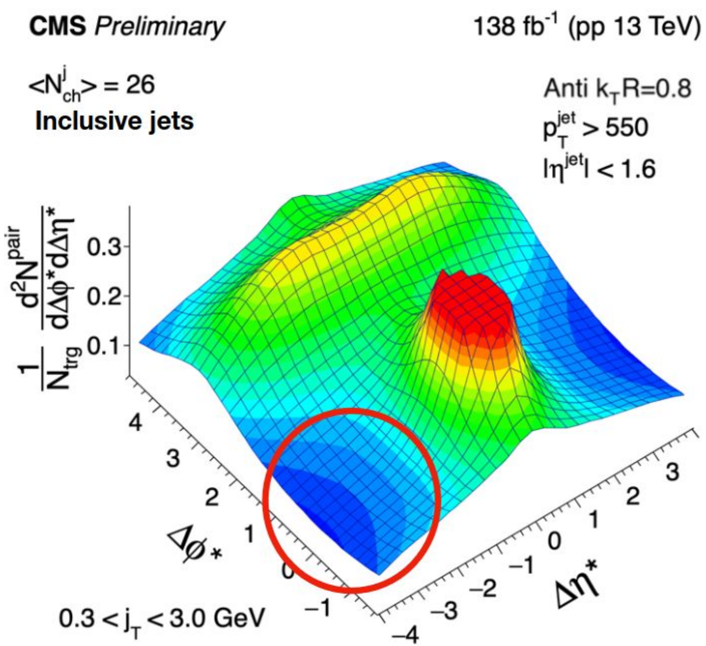


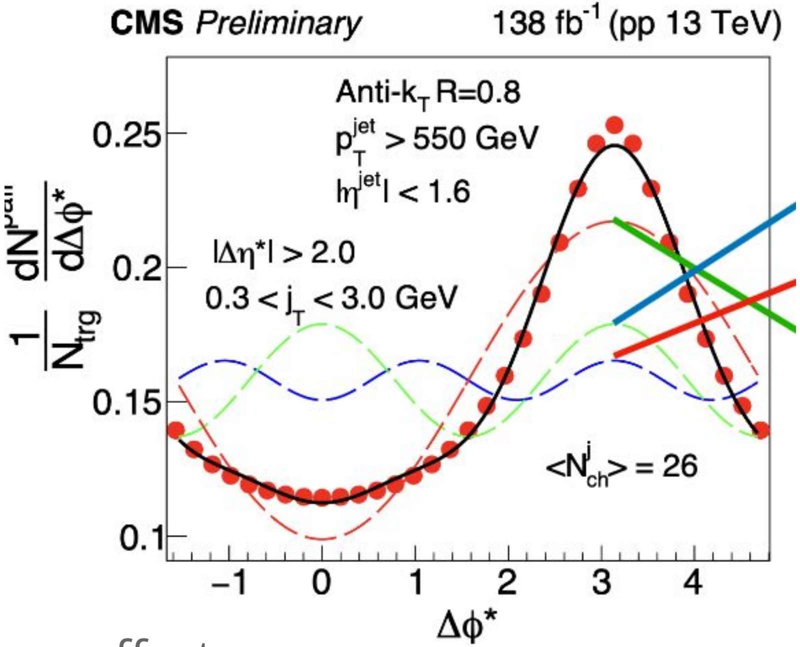
sketch by Austin Baty

Particle correlations using φ^* and η^* coordinates (restricted to $0.86 < |\eta^*| < 5$), transverse momentum relative to the jet axis j_T ($0.3 < j_T < 3$ GeV)

inclusive N_{ch} category

$$\frac{1}{N^{j}_{\mathrm{ch}}}\frac{\mathrm{d}N^{\mathrm{pair}}}{\mathrm{d}\Delta\phi^{*}} \propto \sum_{n=1}^{\infty} \boxed{V_{\mathrm{n}\Delta}} \cos(\mathrm{n}\Delta\phi^{*})$$





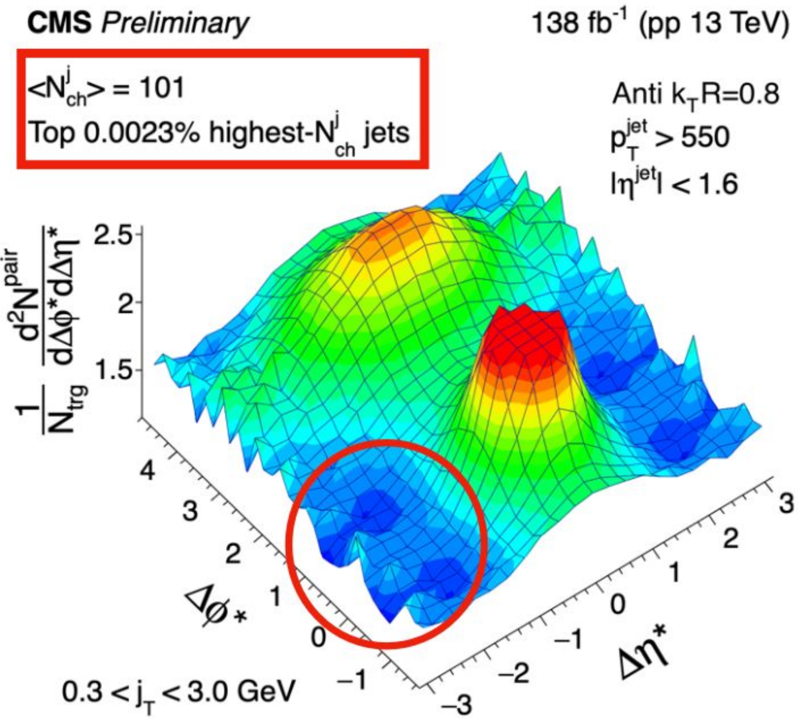
2D distributions corrected for acceptance/efficiency effects

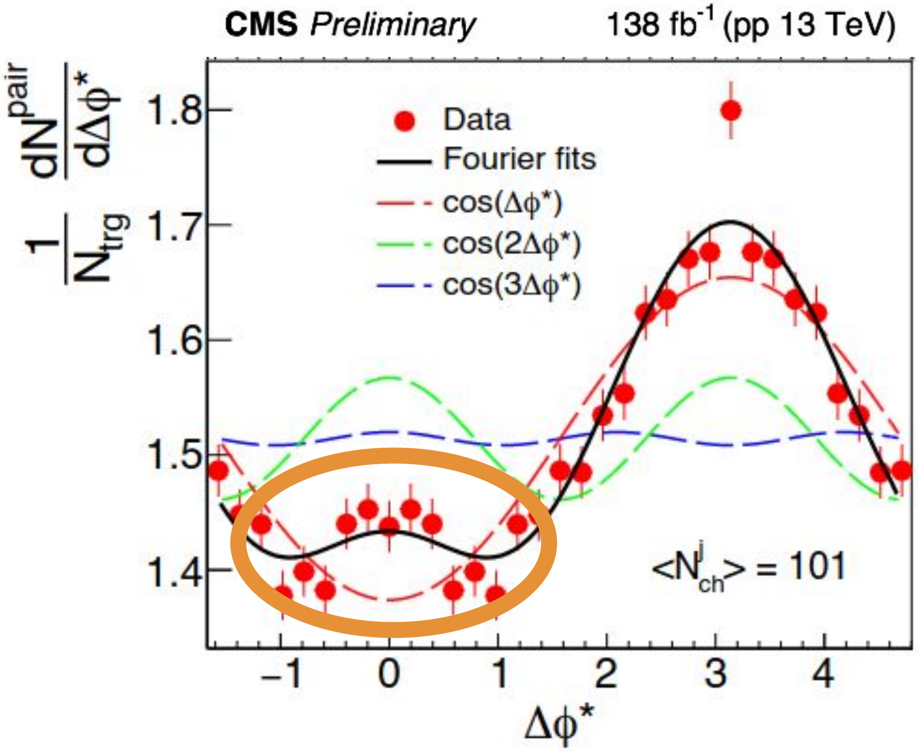
CMS-PAS-HIN-21-013

No near-side ridge at Δφ* ~ 0

high N_{ch} category

[CMS-PAS-HIN-21-013](CMS-PAS-HIN-21-013)

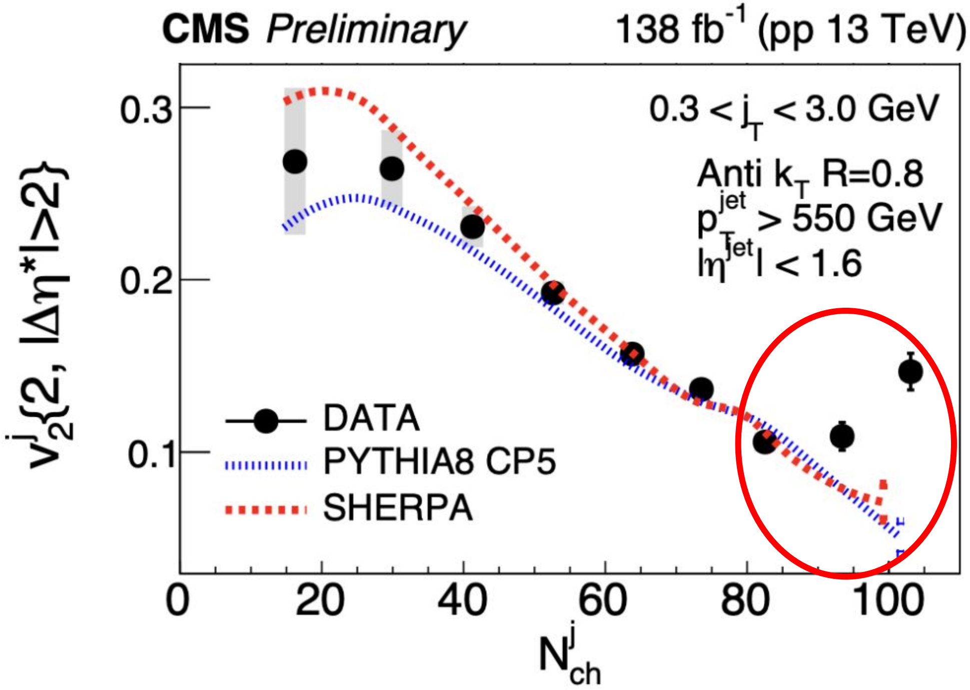




Near-side ridge-like structure at Δφ* ~ 0

single-particle $v_2 = \sqrt{V_2}$ vs N_{ch}

CMS-PAS-HIN-21-013

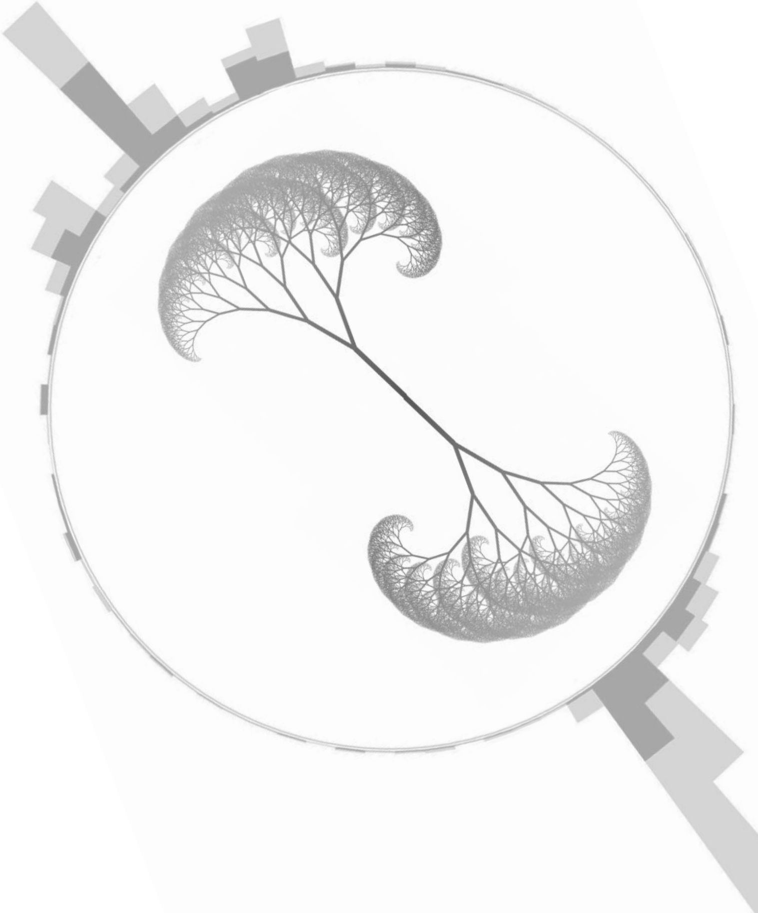


Nonzero v_2 reproduced by SHERPA2, PYTHIA8 CP5 up to N_{ch}~ 80

Increasing v_2 with large N_{ch} not expected by these predictions

Summary

- Mapping out weakly and strongly coupled regimes of the strong interaction via the primary Lund jet plane (**CMS-PAS-SMP-22-007**) and energy-energy correlations (**CMS-PAS-SMP-22-015**)
- Collective-like behavior in jets with high particle multiplicities (**CMS-PAS-HIN-21-013**).
  A great example of synergy between communities!
- **Stay tuned for respective publications & HepData records/Rivet routines!**



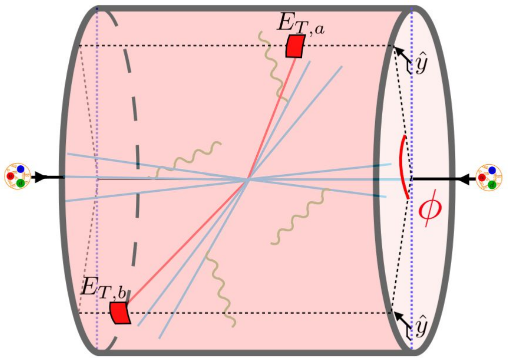
A. Larkoski, G. Salam, J. Thaler,
JHEP06(2013)108

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



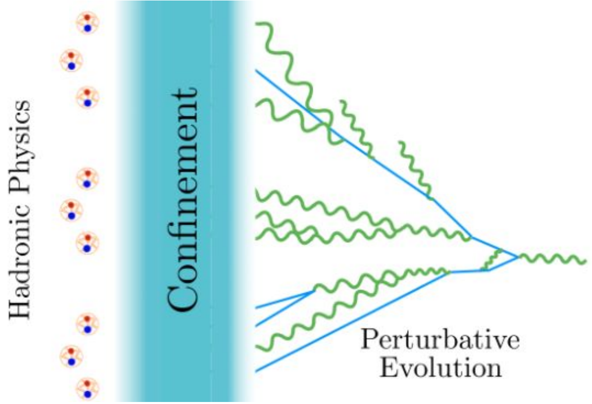
sketch from Ian moult

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

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



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

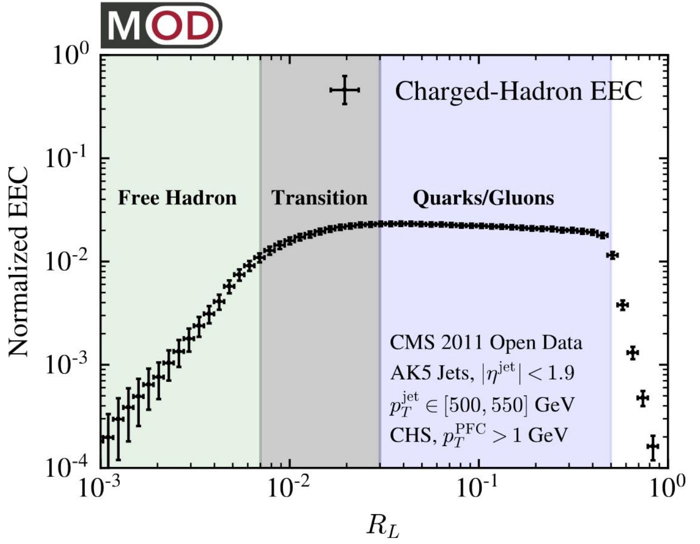
How to measure these experimentally?

1. For a given pair of jet constituents, fill a histogram with weight = $p_{T,i}\, p_{T,j} / p^2_{T,\mathrm{jet}}$ 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

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



Proof of concept using CMS OpenData

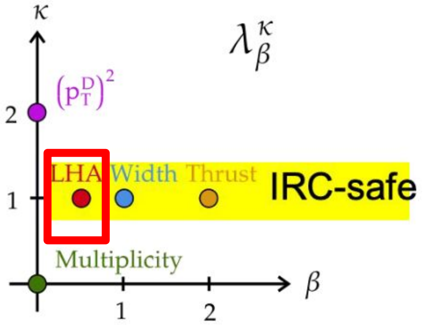
Access to scaling properties of QCD

Generalized angularities in dijet and Z+jet events

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

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

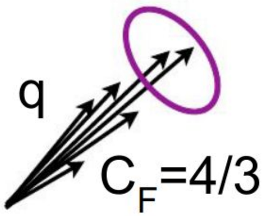
κ & β are parameters set by user



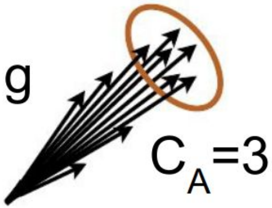
Will show a specific angularity (LHA)

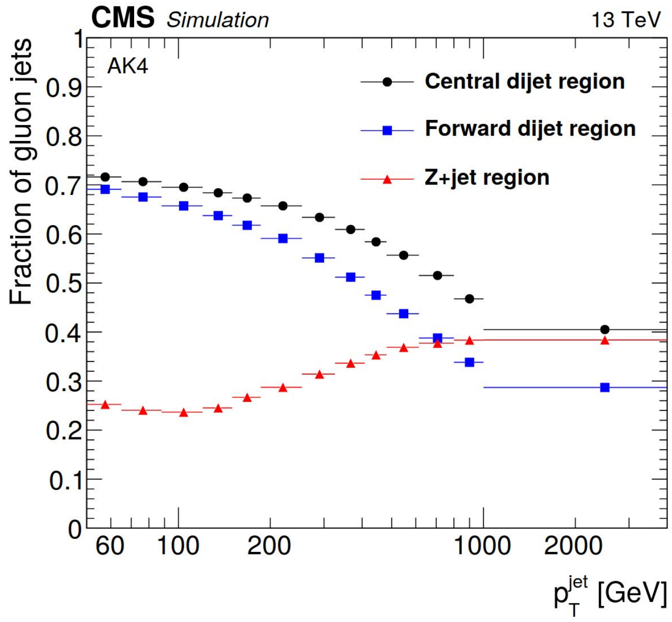
Sensitive to quark vs gluon differences (subset of them are IRC-safe)
JHEP 1707 (2017) 091

Z+jet (quark-like)



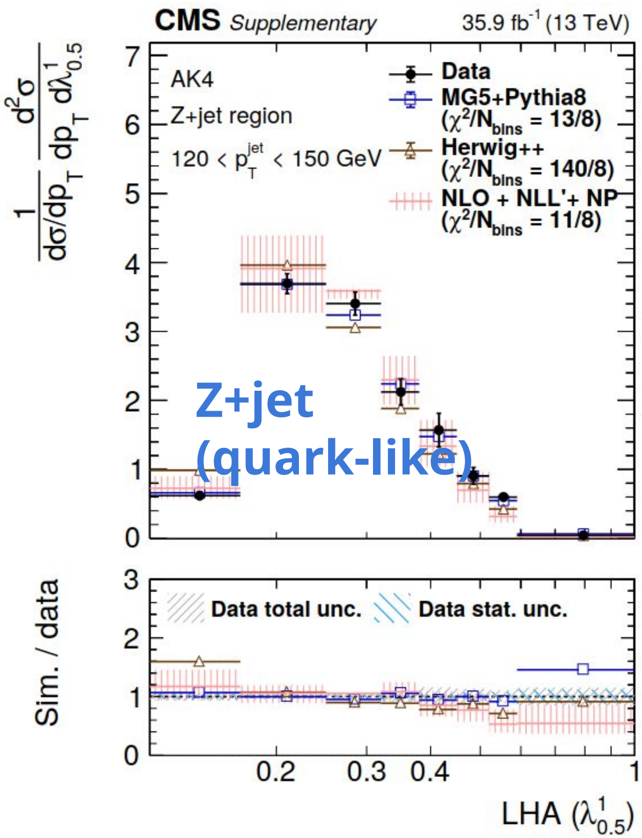
Dijet (gluon-like)



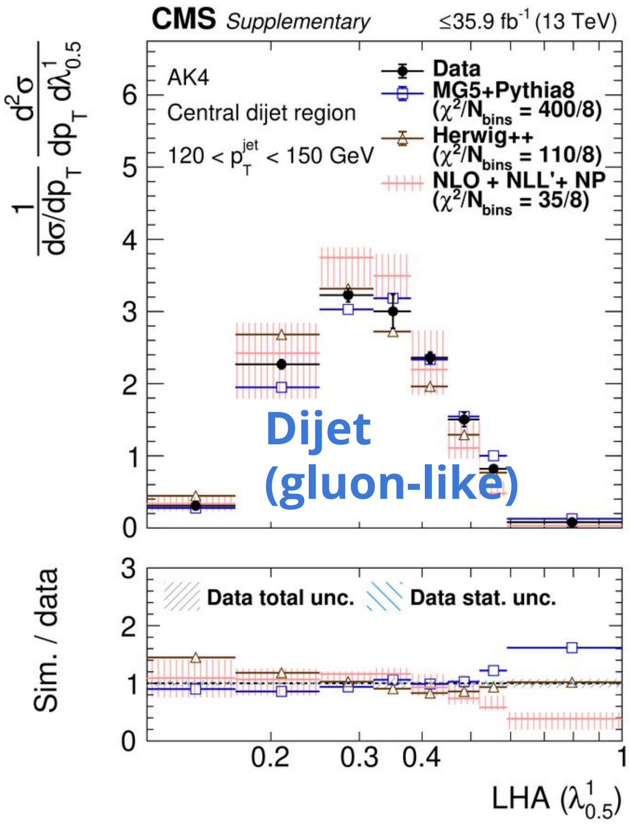


Ungroomed vs groomed with z_{cut} = 0.1, β_{SD} = 0,
R = 0.4 vs R =0.8
charged-only vs charged+neutrals

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)**

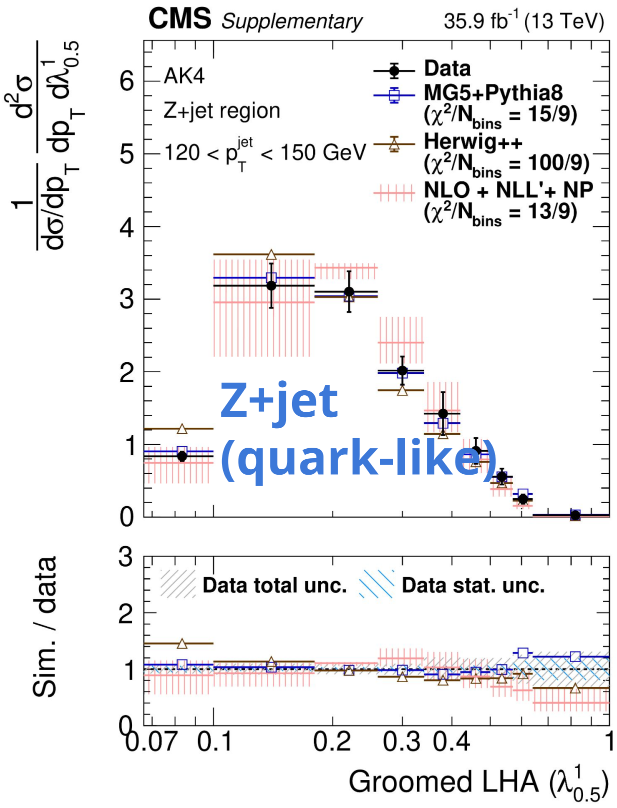
κ = 0.5, β = 1

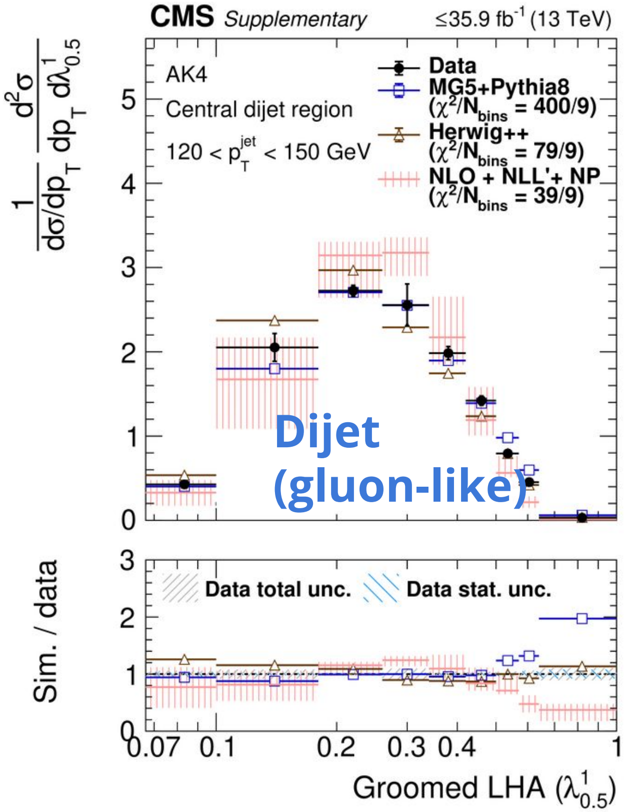
$$\lambda^{\kappa}_{\beta} = \sum_{i\in \text{jet}} z_i^{\kappa} \left(\frac{\Delta R_i}{R}\right)^{\beta} \quad z_i \equiv \frac{p_{Ti}}{\sum_{j\in \text{jet}} p_{Tj}}$$

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

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

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

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

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

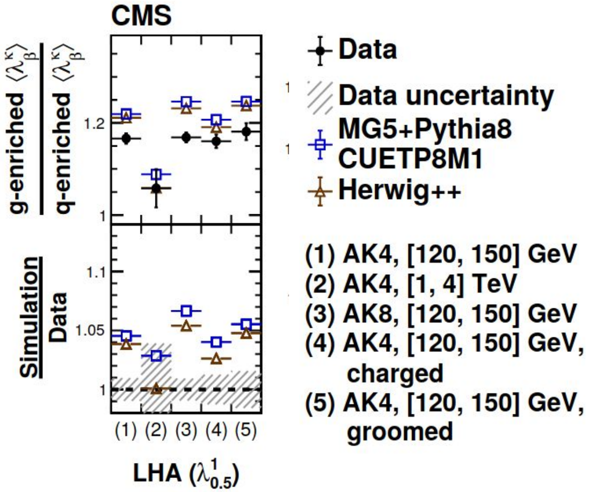
Dijet/Z+jet ratio (g-enriched/q-enriched)

CMS, [arXiv:2109.03340](https://arxiv.org/abs/2109.03340),
JHEP 01 (2022) 188

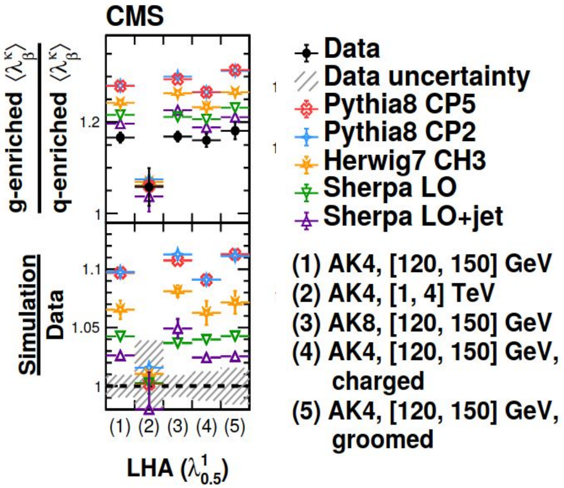
- uncertainties partially cancel in dijet/Z+jet ratio
- MC simulations overestimate g-enriched/q-enriched ratio
- g-enriched / q-enriched ratio is better modelled with “old” PYTHIA8/HERWIG7 tunes

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

“old” CMS tunes
(<~ 5% off)



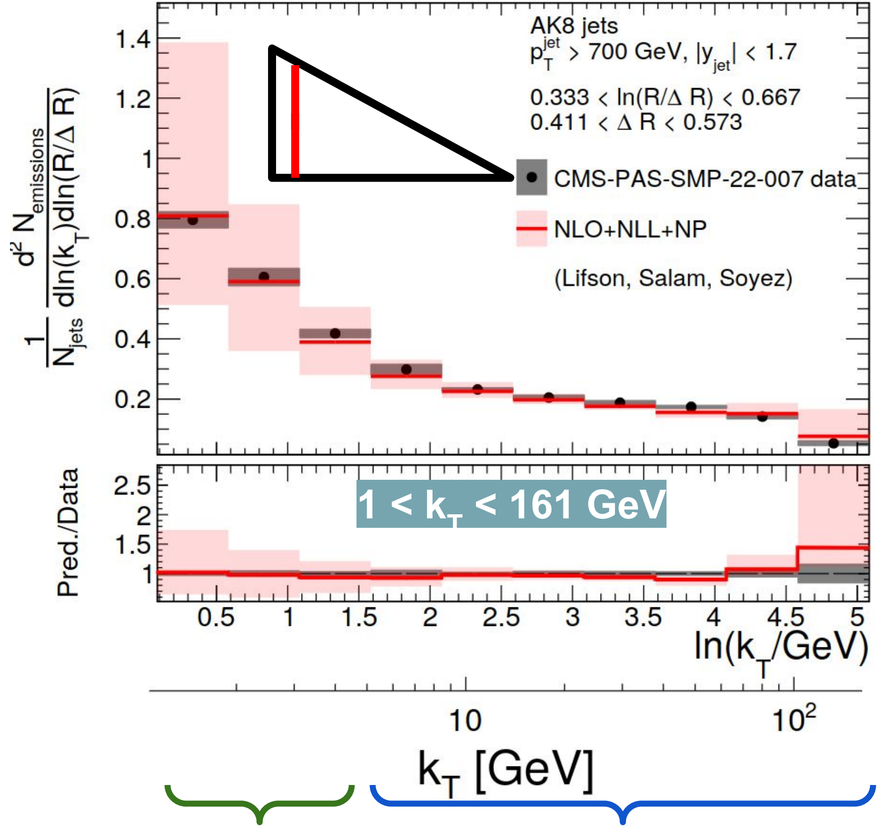
“new” CMS tunes
(up to ~10% off)



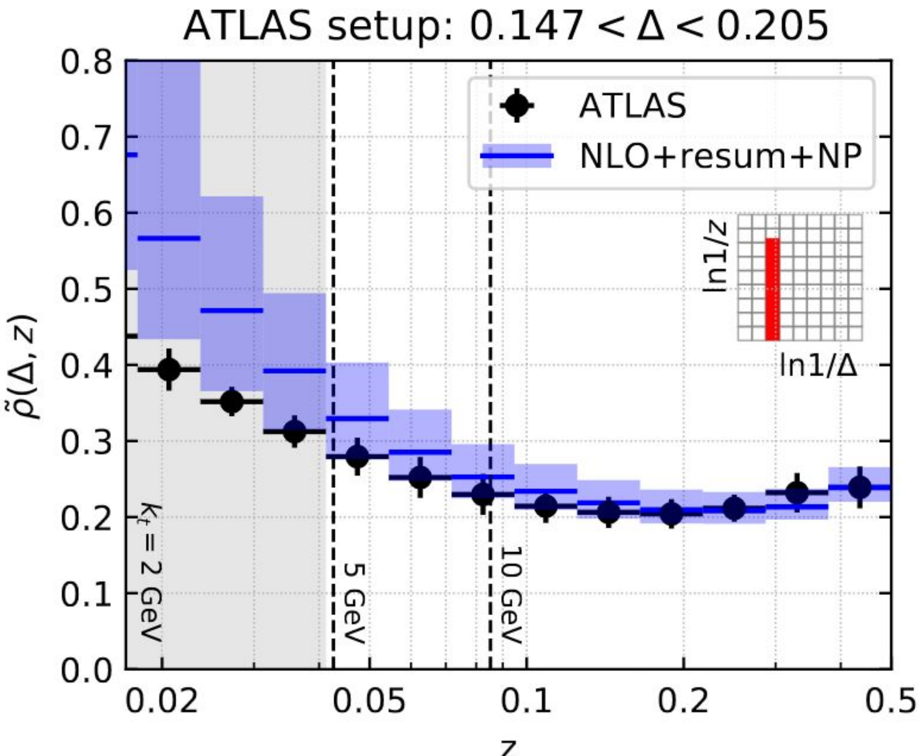
full summary plot in backup
(other angularities)

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

Calculations based on JHEP10(2020)170



nonperturbative **resummation**

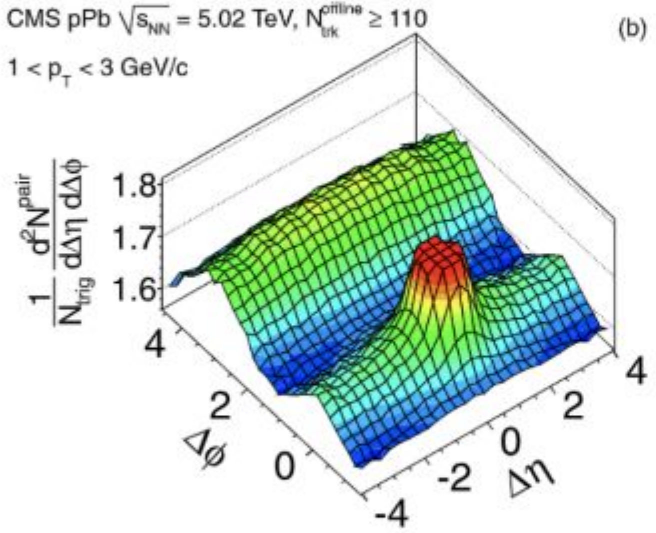


A. Lifson, G. Salam, G. Soyez JHEP10(2020)170

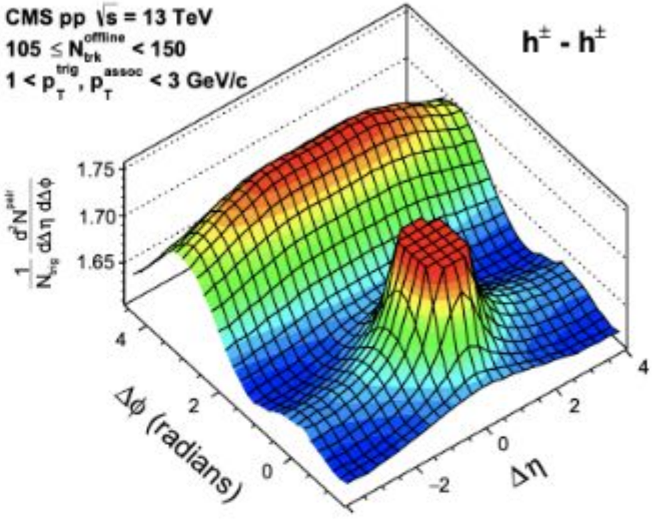
data from ATLAS Lund plane,
PRL 124, 222002 (2020)

Ridge in pPb and high-multiplicity pp

pPb

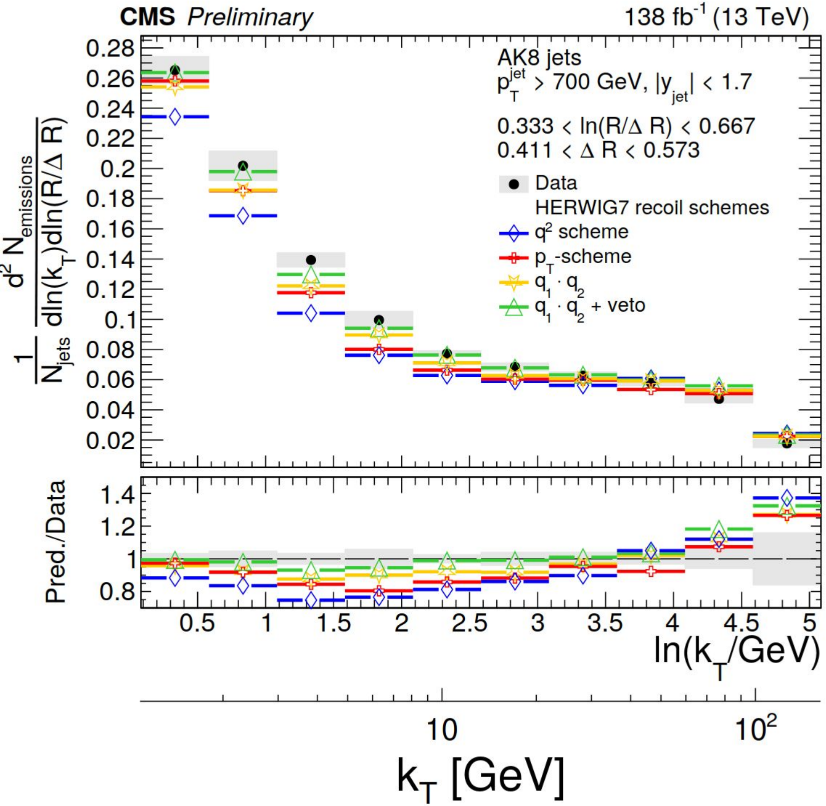


High-multiplicity pp

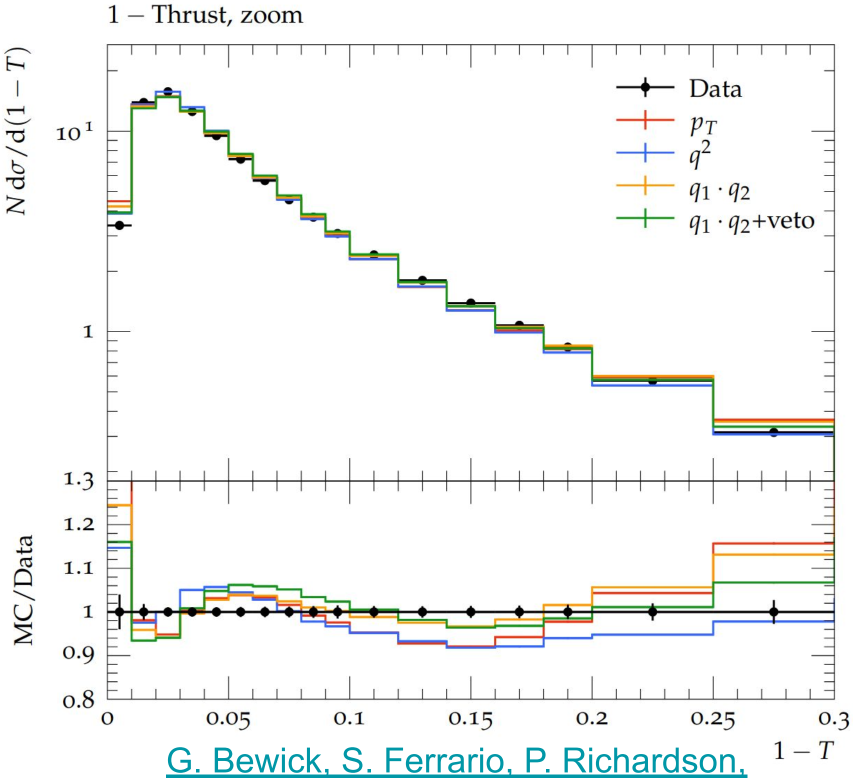


Sensitivity to recoil scheme choice, **important ingredient to reach NLL accuracy**

high-p_T quark and gluon jets



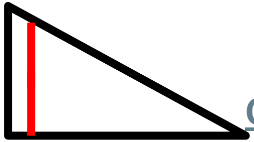
Thrust in e^+e^- at Z mass pole



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

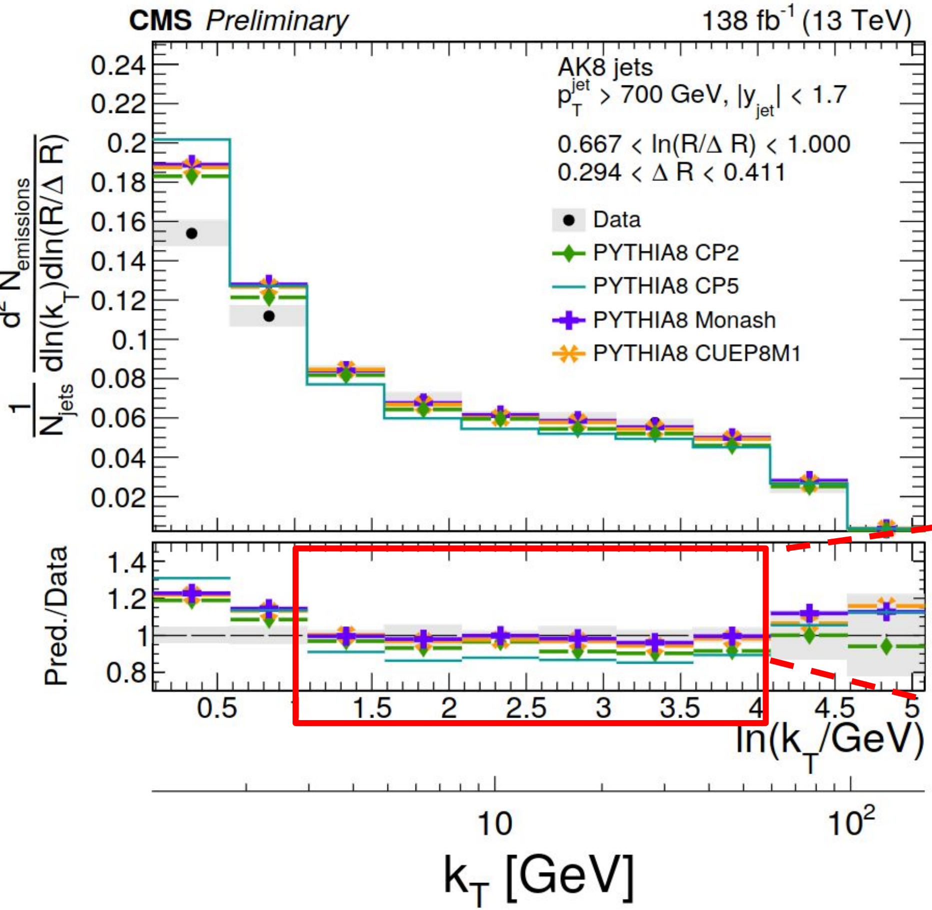
Large angle emissions



R = 0.8 **Most important difference between PY8 tunes is α_s^{FSR}**

CMS-PAS-SMP-22-007

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



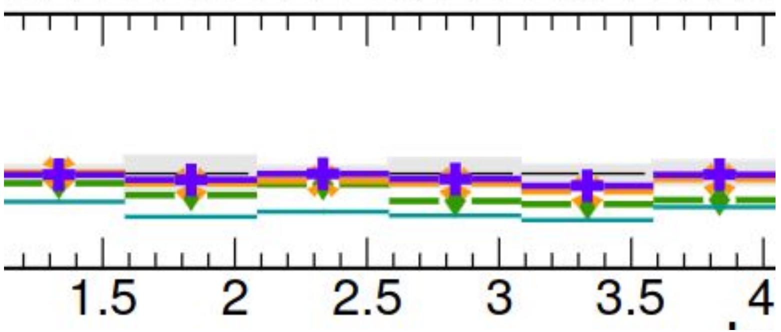
Monash*/*CUEP8M1*:** $\alpha_s^{FSR}(m_Z) = 0.1365$ ***(best description)

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

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

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

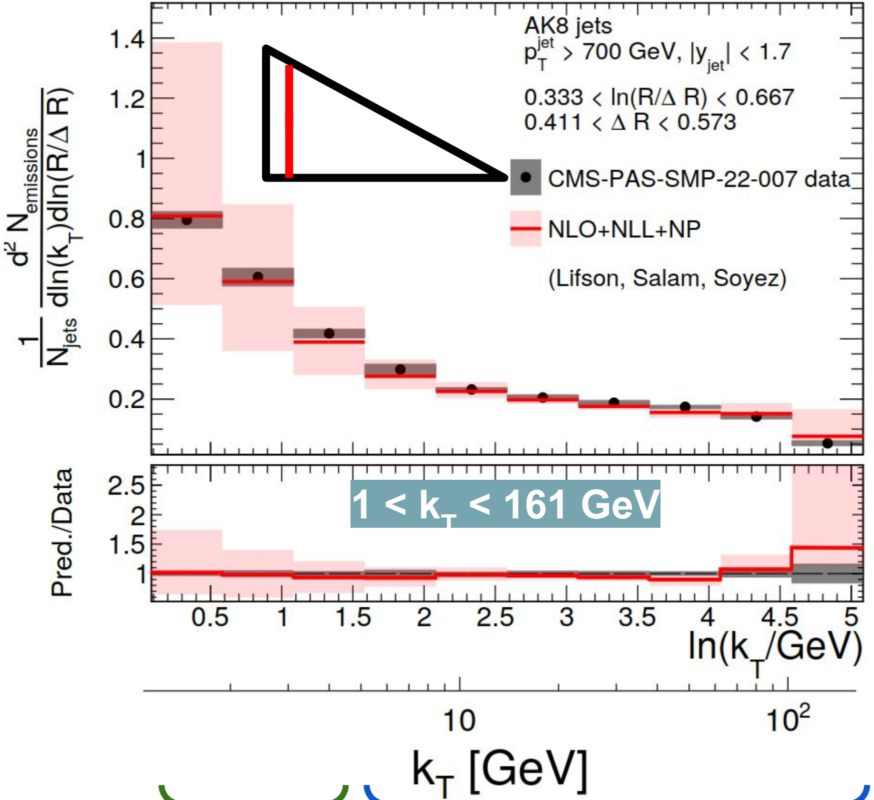
~15%



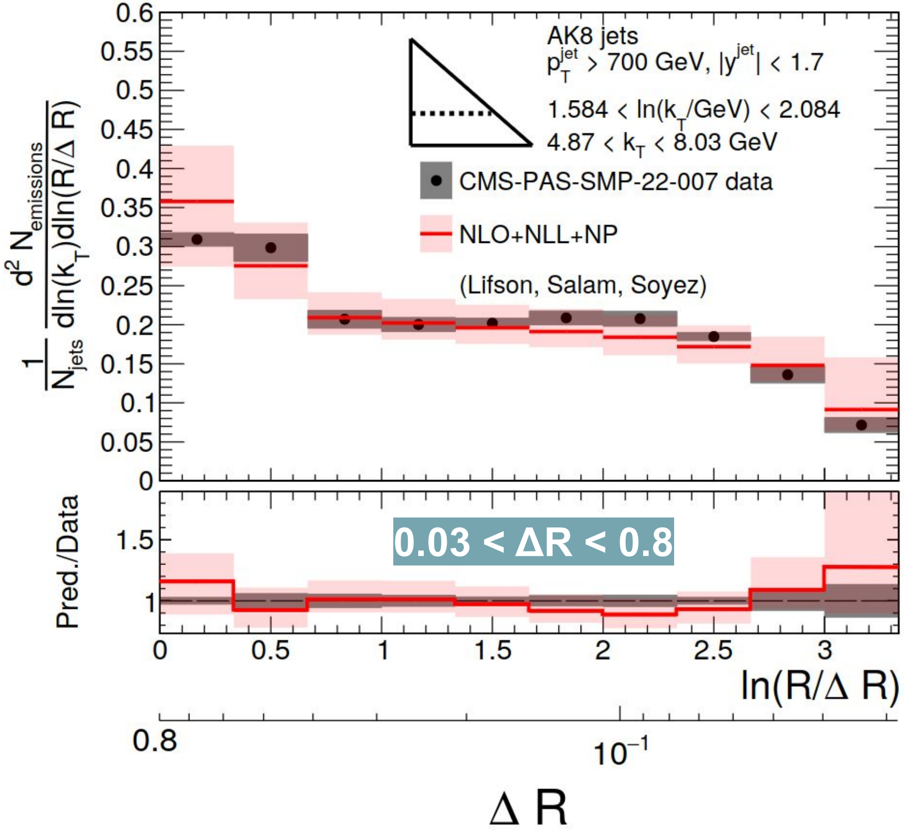
k_T between 3 – 50 GeV

pQCD analytical calculations (NLO+NLL+NP)

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



$1 < k_T < 161$ GeV



$0.03 < \Delta R < 0.8$

nonperturbative

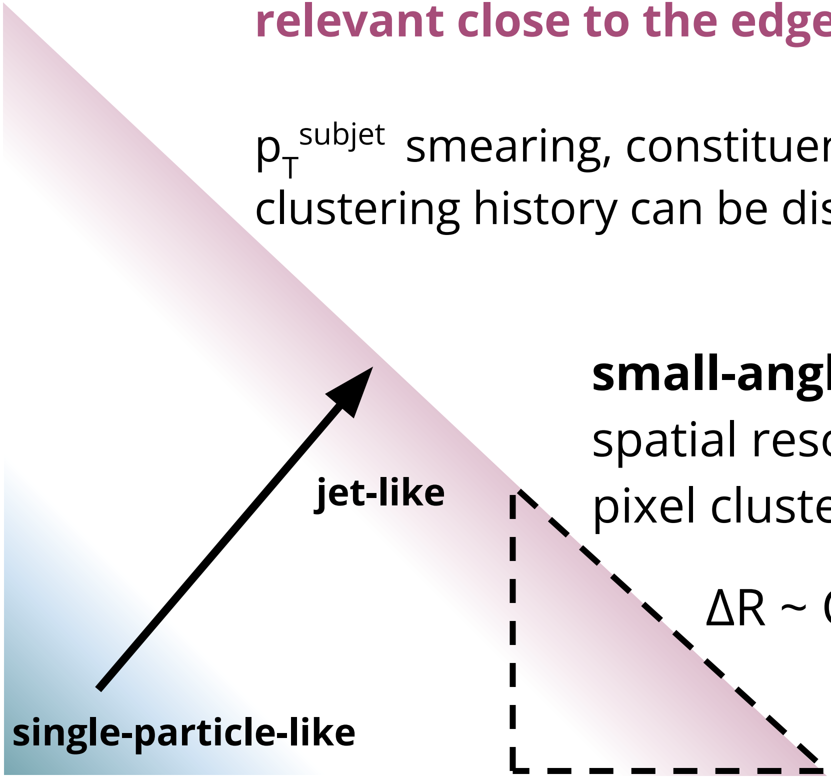
resummation
(unc. mostly perturbative)

Nonglobal logs, clustering logs

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

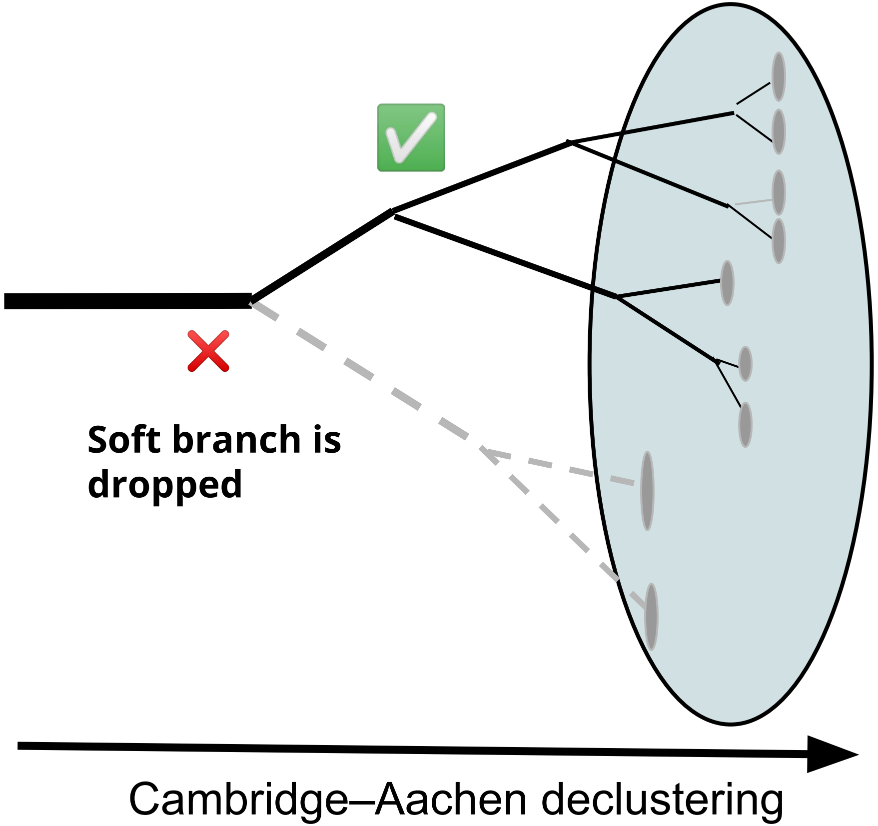


small-angles:
spatial resolution,
pixel cluster merging

$\Delta R \sim O(10^{-3} – 10^{-2})$

residual PU contributions (large ΔR, low k_T)

(Intermezzo) soft-drop grooming algorithm



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**

Corrections to particle level

Sequential set of corrections:

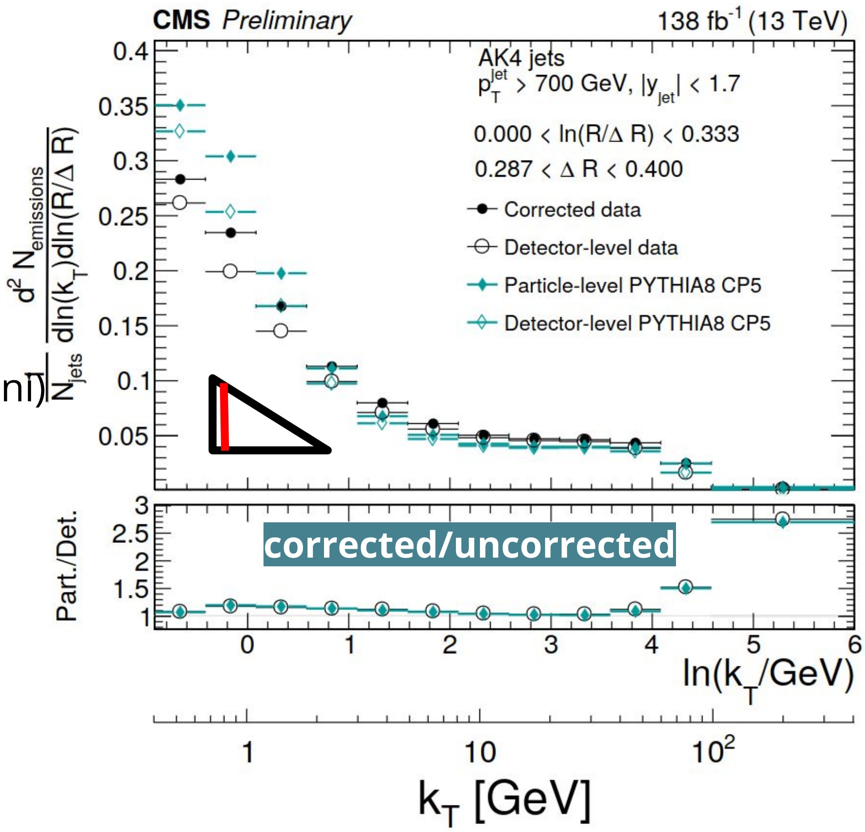
1. **Background:** bin-by-bin correction to account for det-level emissions not matched to truth-level emissions.

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

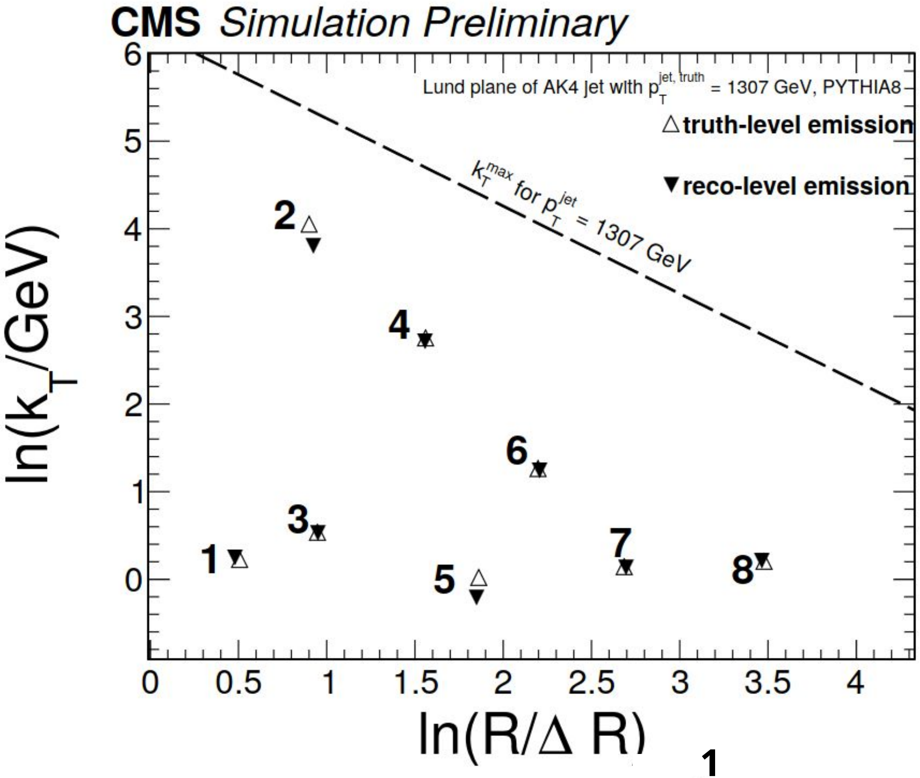
CMS-PAS-SMP-22-007

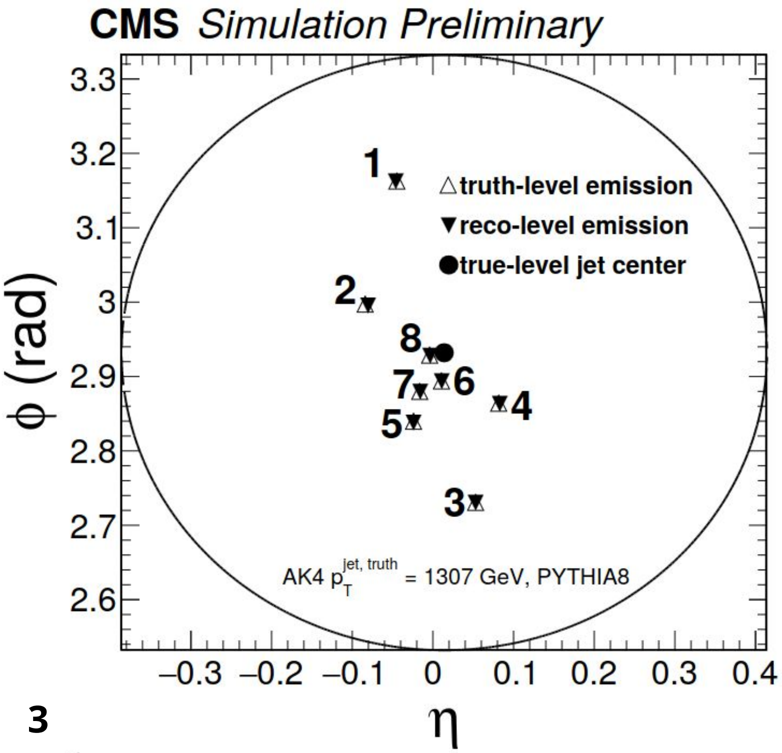


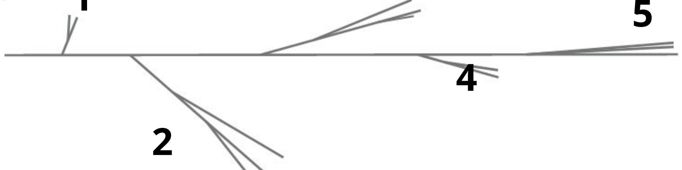
smearing becomes more important at high k_T

Matching emissions at detector level and particle level

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



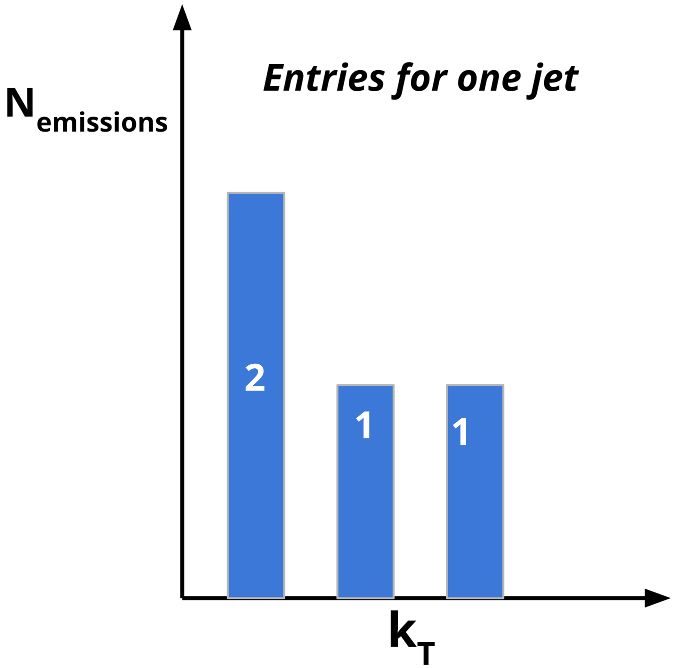


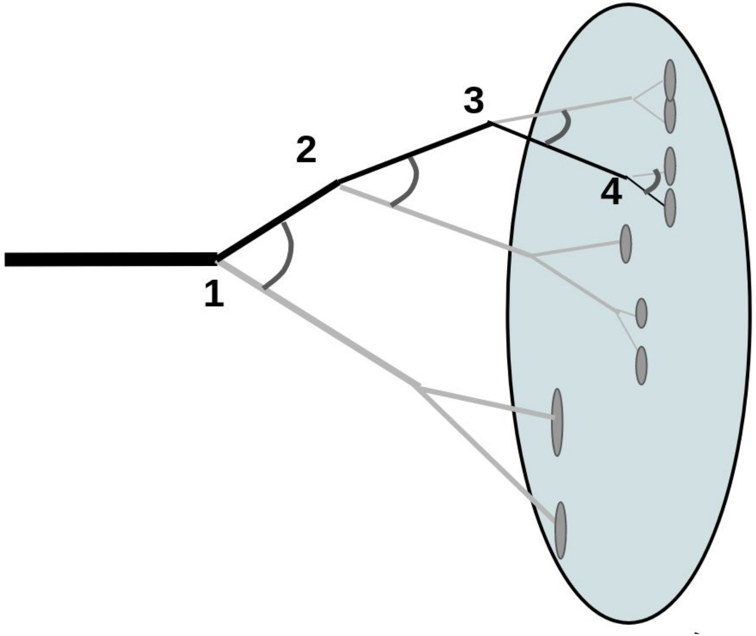


CMS-PAS-SMP-22-007

detector-level statistical correlations

LJP is a multicount observable (i.e., multiple entries per jet) → bins are statistically correlated at det level





bin-to-bin correlations of up to ~5–10%, measured covariance matrix used in unfolding

(can be important for other observables, e.g. Lund multiplicities, energy correlators, …)

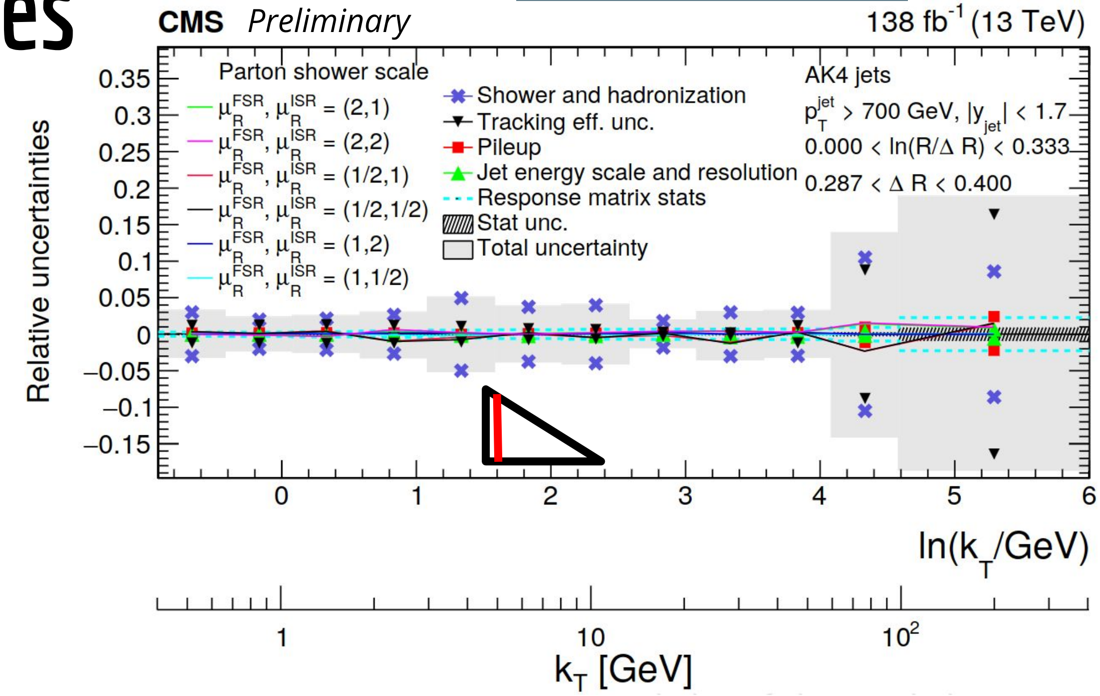
Systematic uncertainties

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 reco. efficiency model uncertainty,
1-2% in bulk, dominates at 10-20% at edge



Subleading components (<~ 1%):

Parton shower scale
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