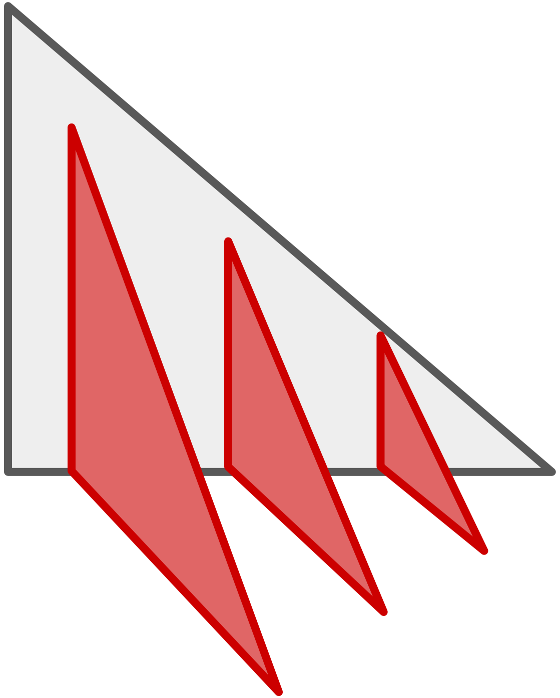
Lund jet plane @ CMS

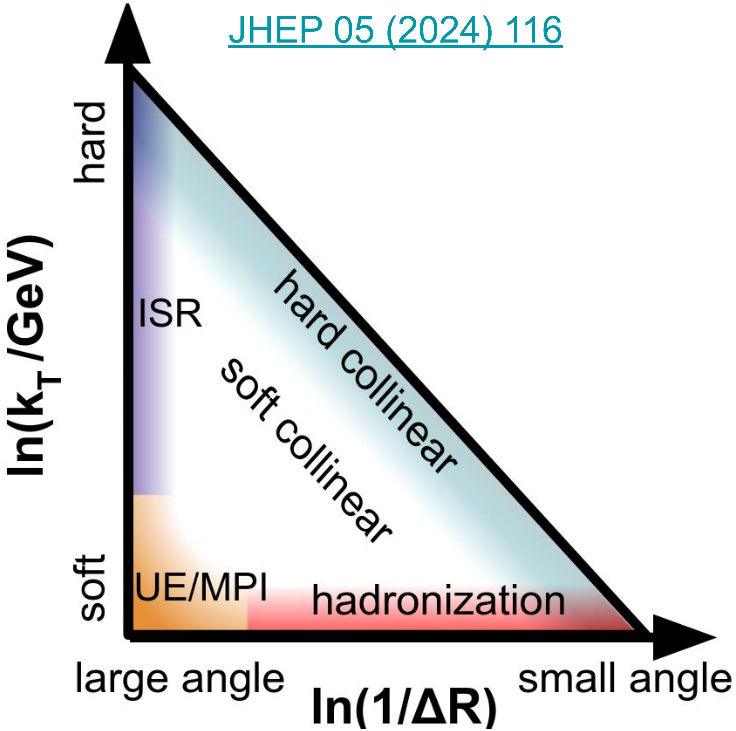
Cristian Baldenegro Barrera (MIT)

LHC-EW WG: Jets and EW bosons
Sep 18th 2024



CMS Lund plane setup

JHEP 05 (2024) 116



Anti-k_T jets with p_T > 700 GeV & |y| < 1.7 (inclusive jet selection)

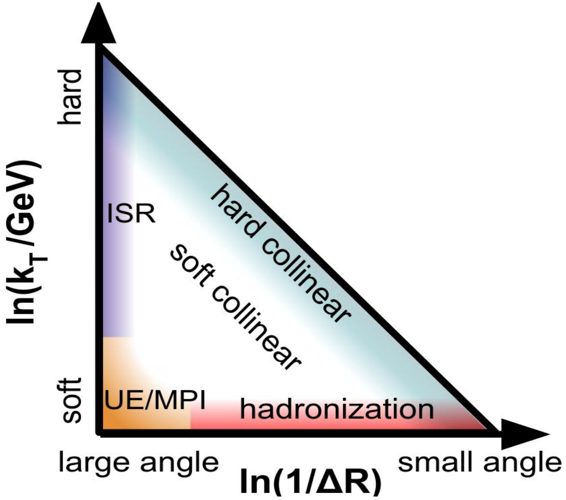
Two distance parameters: R = 0.4 and 0.8 (analysis done separately for each R)

Jet substructure using charged-particle constituents

(NB: used “charged-hadron subtraction” for pileup mitigation, PUPPI continuous weights not optimal for Cambridge–Aachen tree measurements)

ALICE/CMS Lund plane coordinates

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



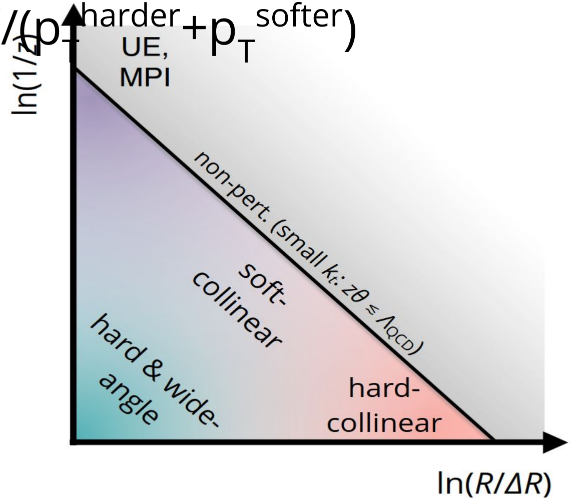
k_T : proxy for hard-scale of 1→2 branching

Weakly & strongly coupled regions separated via “horizontal” cuts

Experimentally, sensitive to tracking efficiency effects, large uncertainties at kinematical edge due to fast drop of **$\varrho(k_\mathrm{T}, \Delta R)$**

ATLAS Lund plane coordinates

$z = p_\mathrm{T}^{\mathrm{softer}}/(p_\mathrm{T}^{\mathrm{harder}}+p_\mathrm{T}^{\mathrm{softer}})$ vs ΔR



z : p_T-balance between core & emission;
less correlation with ΔR

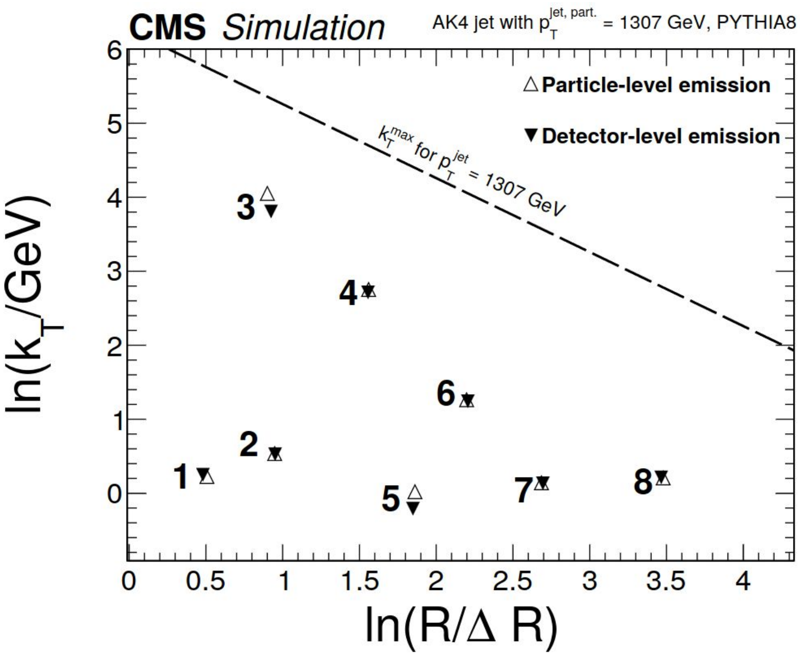
resilient against detector smearing effects,
tracking inefficiencies, charged p_T scale uncertainties, …

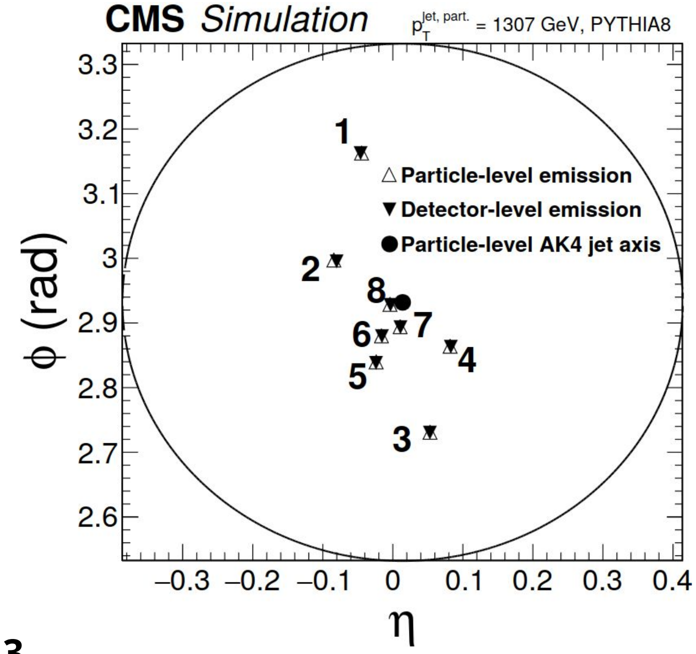
k_T scale “fuzzier” towards smaller ΔR ($k_\mathrm{T} = z\, \mathbf{p_\mathrm{T}^{mother}}\, \Delta R$)

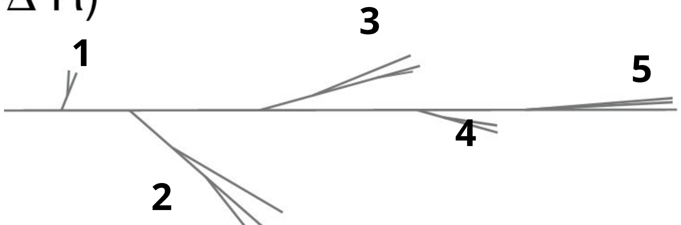
Matching emissions at detector level and particle level

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

Geometrical matching is done univocally in η vs φ (iterating through both det-level and part-level list of splittings)

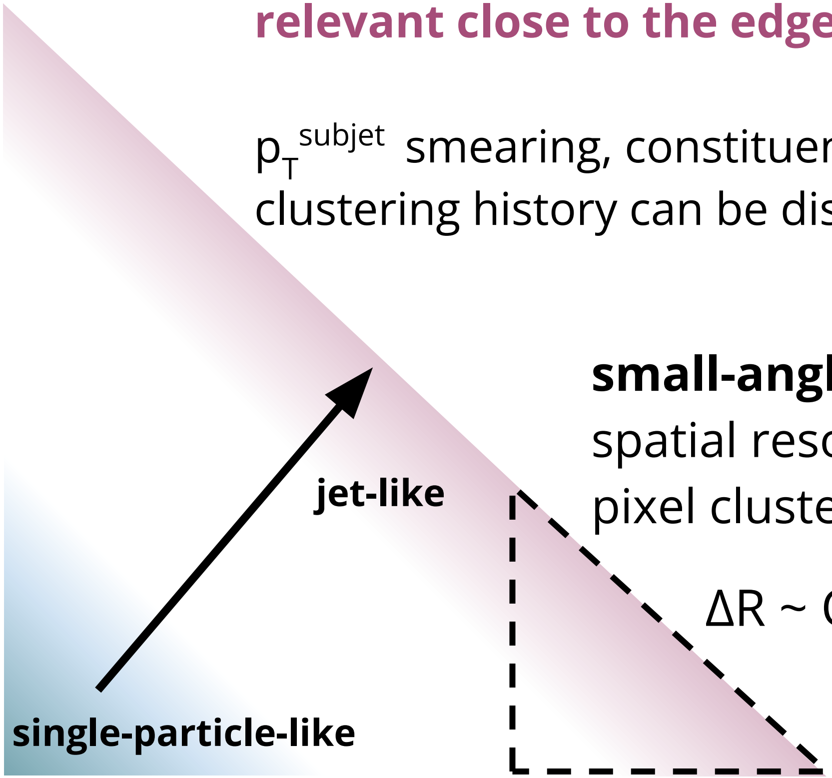






Cristian Baldenegro (MIT)

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

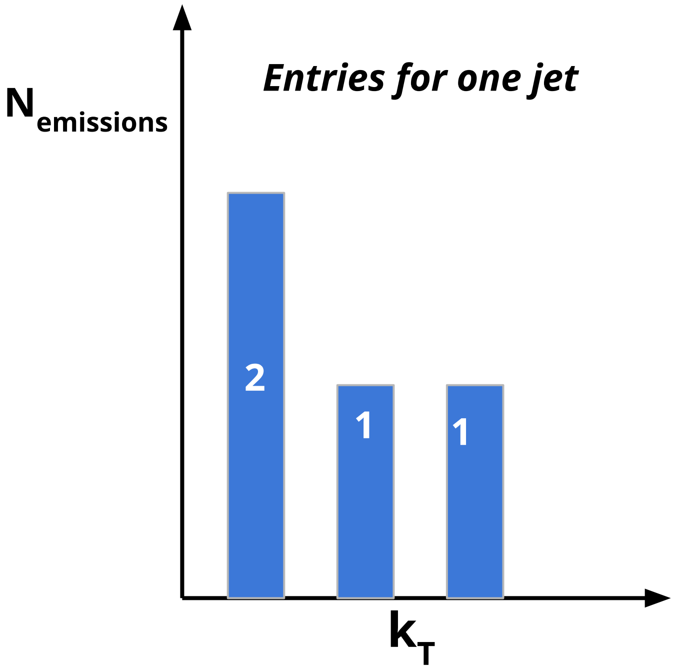
small-angles:
spatial resolution,
pixel cluster merging

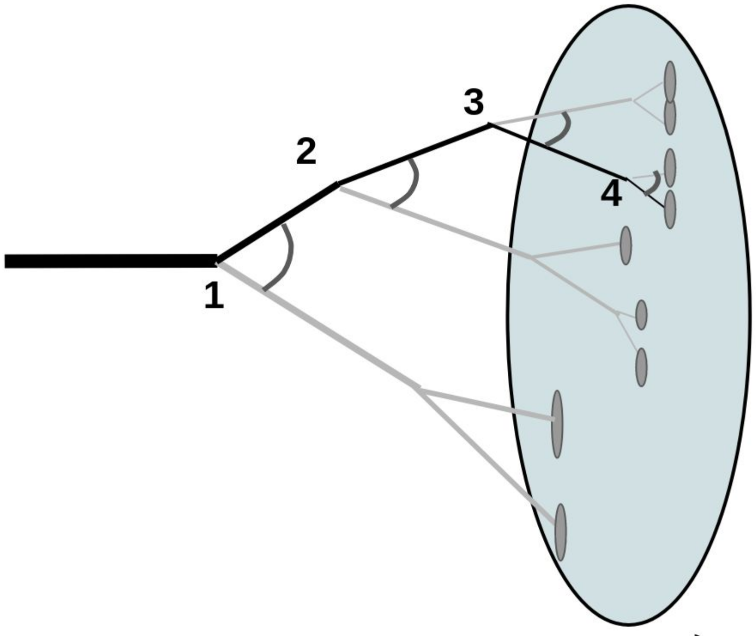
$\Delta R \sim O(10^{-3} – 10^{-2})$

residual PU contributions (large ΔR, low k_T)

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% prior to unfolding, correlations can be “long-range” due to angle-ordered CA tree

Correlations provided as input to unfolding

Systematic uncertainties

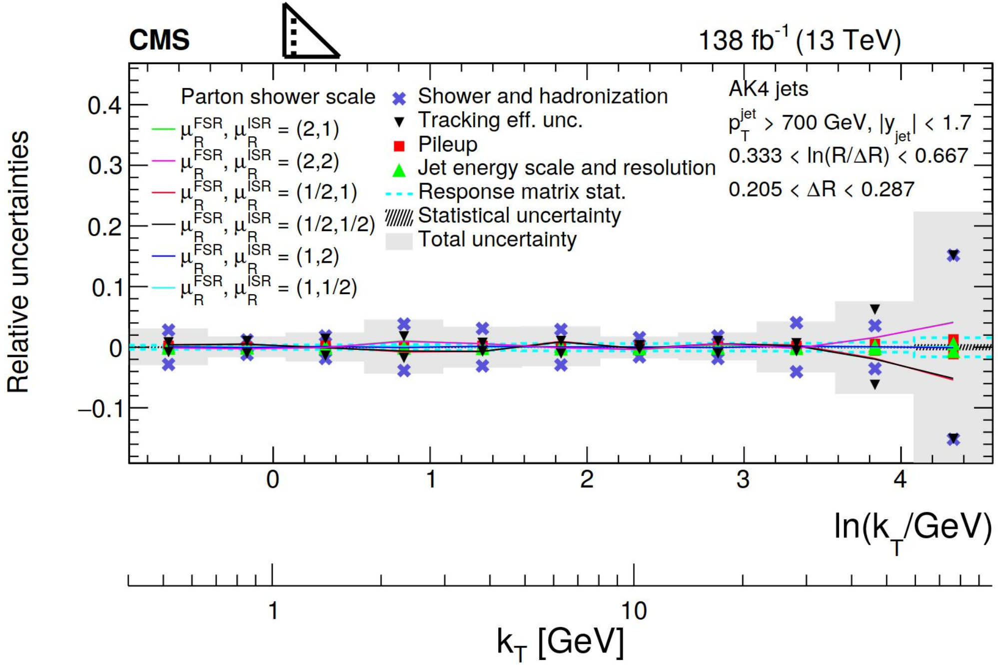
Shower & hadronization model uncertainty (HERWIG7 vs PYTHIA8)
(2–7% in the bulk, 10% at kinematical edge)

decorrelated into prior bias ⊗ response pieces

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

Dropped 3% of tracks in simulation to cover data-to-simulation differences

Procedure must be refined for future measurements (+ include ΔR dependence)



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

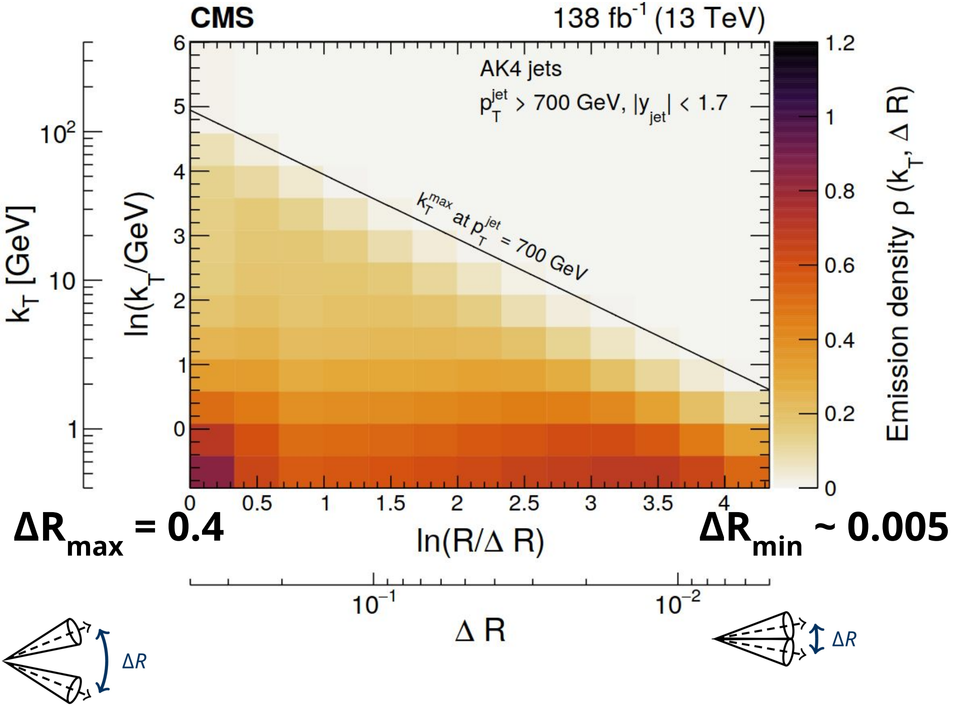
Subleading components (<~ 1%):

Parton shower scale
Response matrix stats
Jet energy scale and **resolution uncertainties**
Pileup modeling

CMS primary Lund jet plane densities

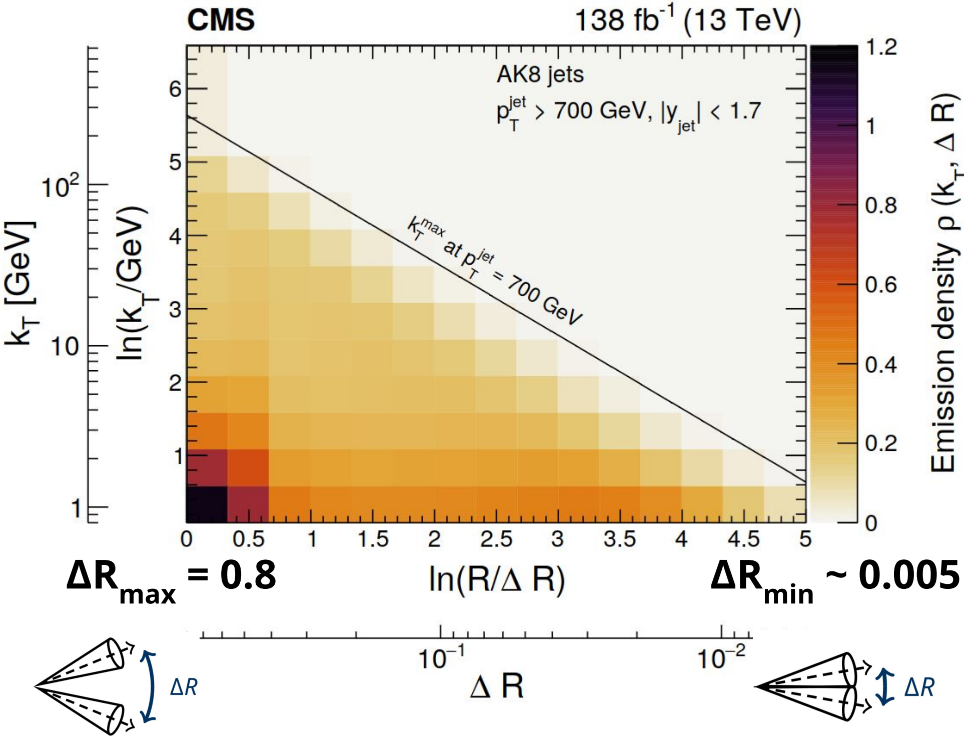
$p_T^{jet} > 700$ GeV,
charged particles for substructure

R=0.4 (standard R in Run-2)



$\Delta R_{max} = 0.4$ **$\Delta R_{min} \sim 0.005$**

R=0.8 (large-angle & harder emissions)

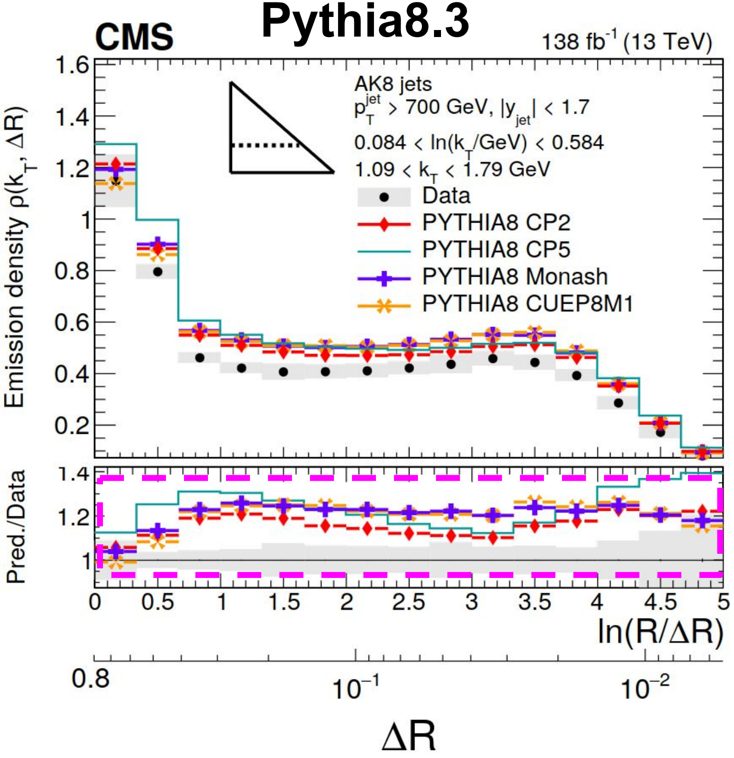


$\Delta R_{max} = 0.8$ **$\Delta R_{min} \sim 0.005$**

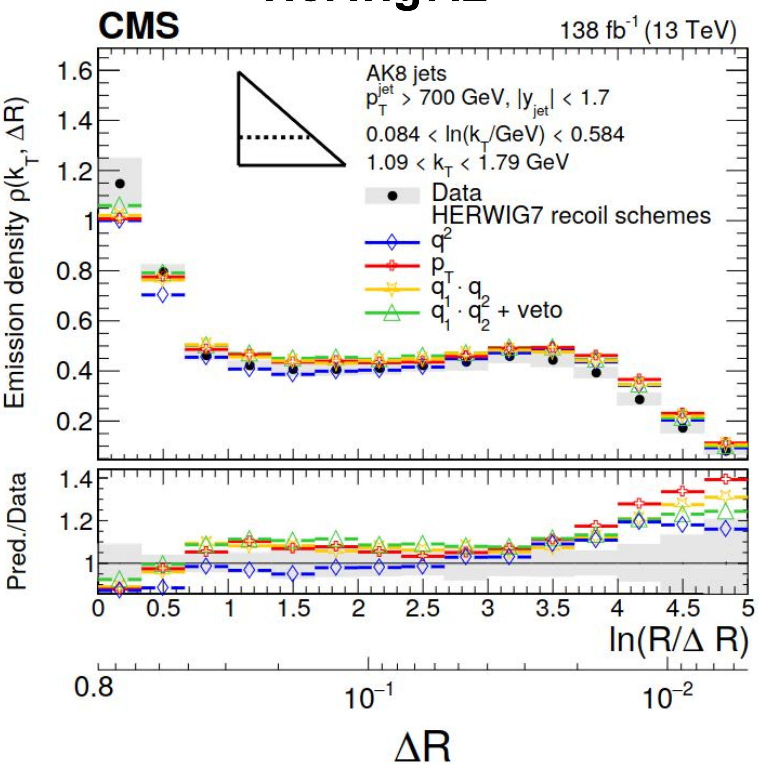
Lund vs cluster fragmentation? FSR cutoff differences?

Hadronization region (**$k_T \sim 1$ GeV**)

Pythia8.3



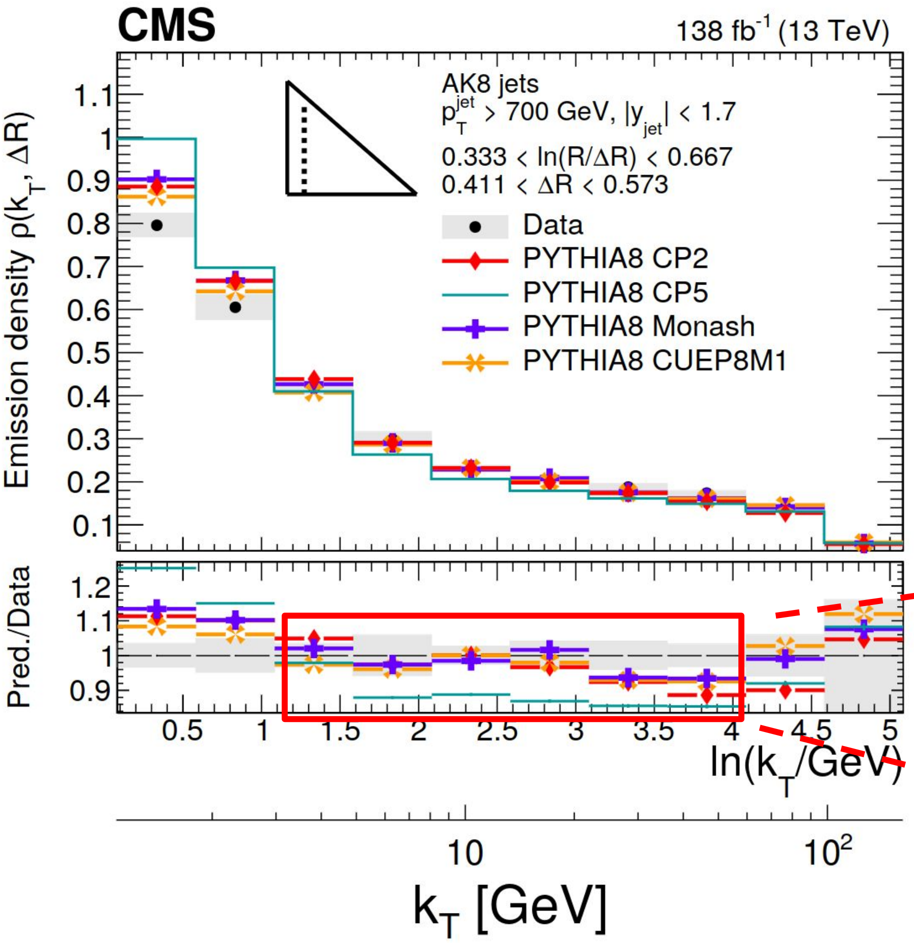
Herwig7.2



Lund string (**PYTHIA8** overshoots) data by 15-20% in hadronization region

Cluster model in better agreement… (**HERWIG7/SHERPA2**)

Reminder that $\alpha_S^{FSR}(m_Z)$ should not be treated as any other MC tuning parameter



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

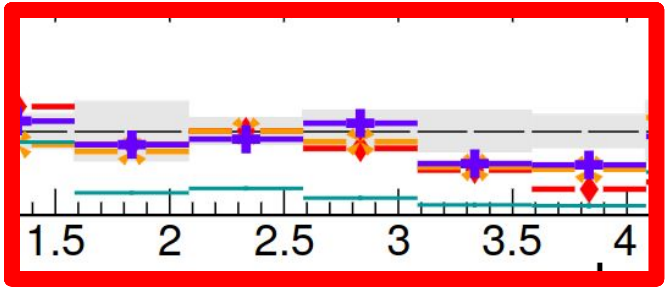
***Monash*/*CUEP8M1*:** $\alpha_S^{FSR}(m_Z) = 0.1365$

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

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

Larger value of $\alpha_S^{FSR}(m_Z)$ effectively accounts for missing NLO corrections in the soft limit (~CMW rescaling)

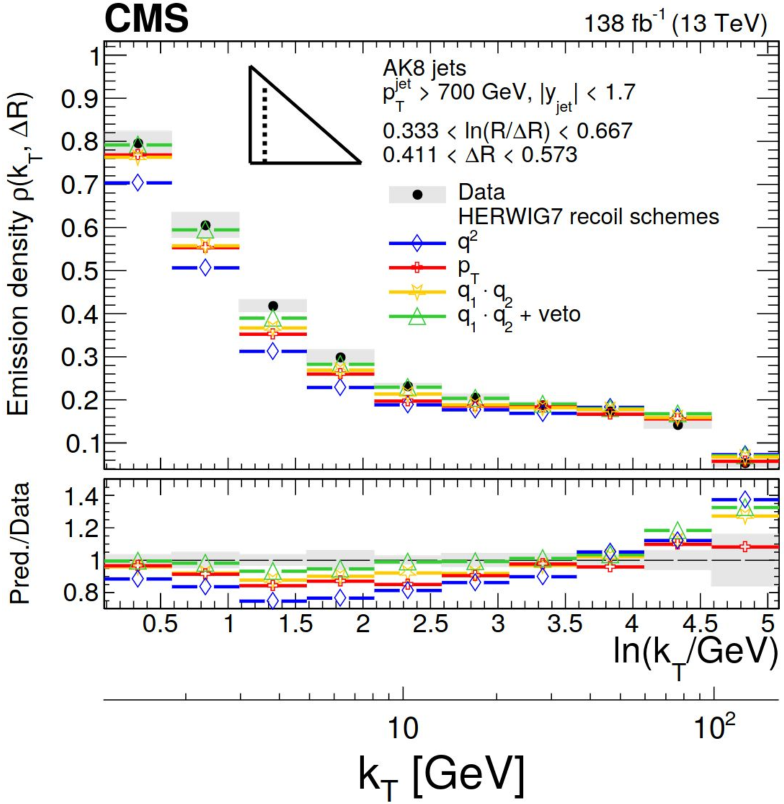
~15% off



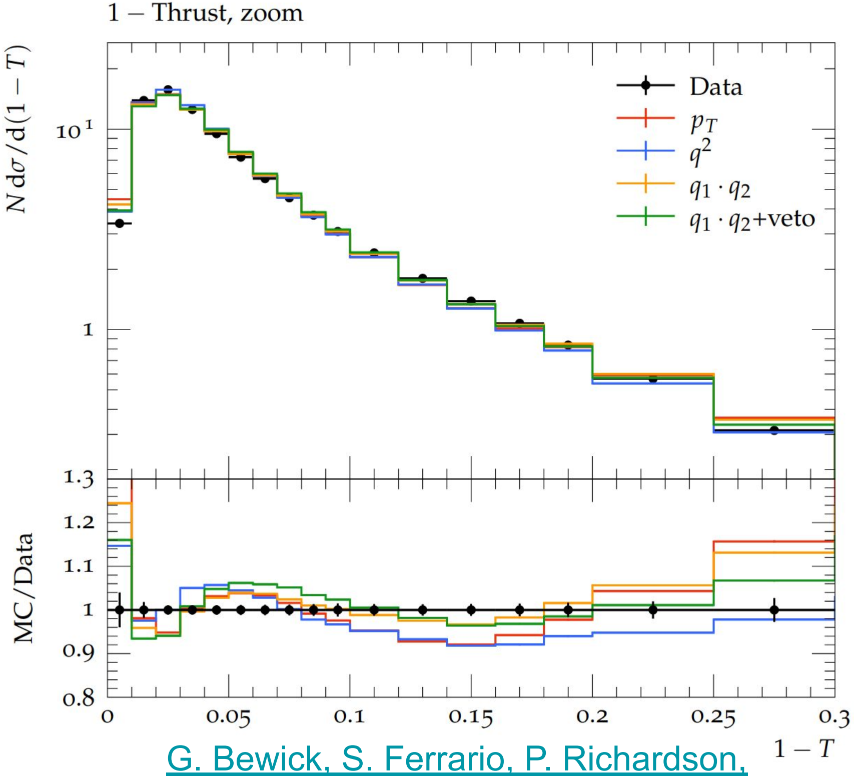
k_T between 3 – 50 GeV

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

high-p_T jets @LHC



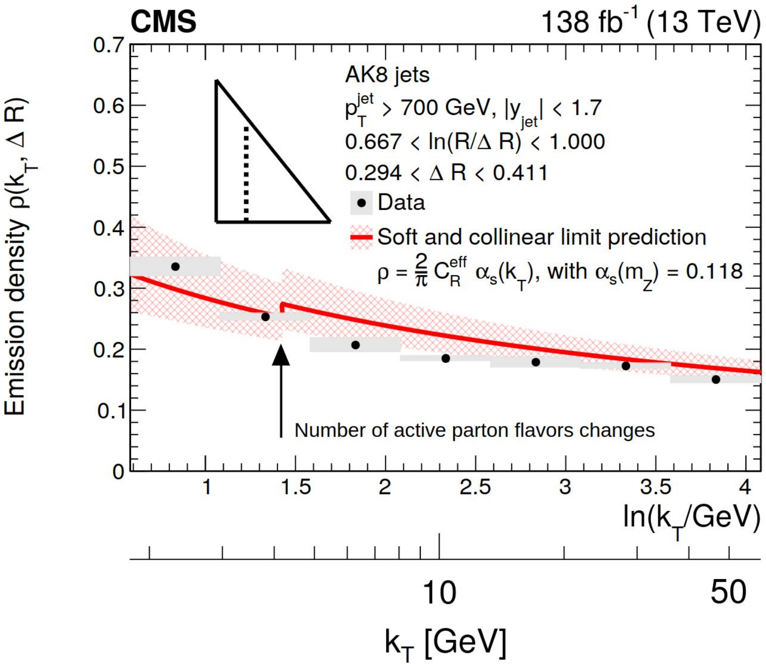
e^+e^- →hadrons at Z mass pole @ LEP



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

Running of α_s in the jet shower



Recall LO pocket formula for Lund density:

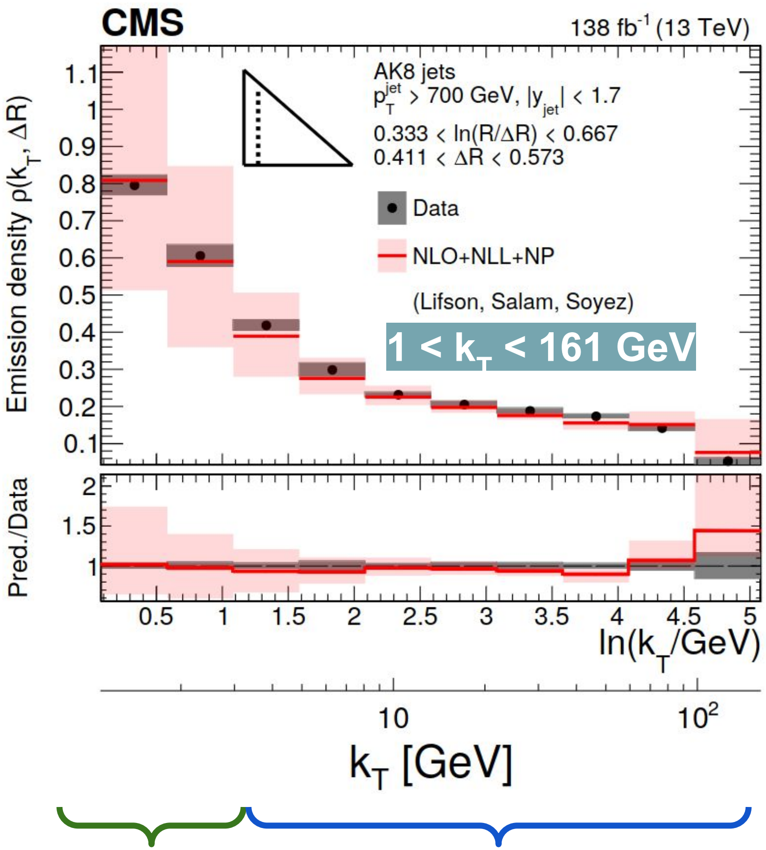
$$\frac{1}{N^{\text{jets}}}\frac{\mathrm{d}^2 N_{\text{emissions}}}{\mathrm{d}\ln(k_T)\,\mathrm{d}\ln(R/\Delta R)} \simeq \frac{2}{\pi} C_R \alpha_s(k_T)$$

Running $\alpha_s(k_T)$ from few GeV to ~60 GeV qualitatively describes the data
(Assuming q/g fractions from PYTHIA8)

Cute to see, but breaks down
at large angles ΔR, close to the edge, etc

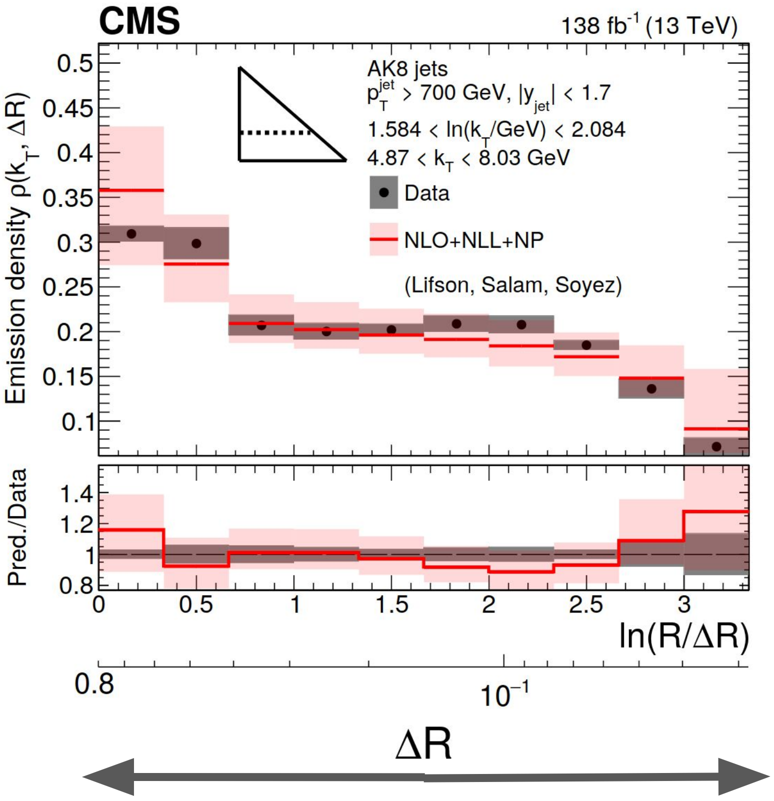
Described well by pQCD calculations (NLO+NLL+NP)

Calculations from A. Lifson, G. Salam, G. Soyez [JHEP10(2020)170]



$1 < k_T < 161$ GeV

nonperturbative **resummation**



Nonglobal logs **Parton flavor changes**

NP corrections account for $k_T \to k_T$ shift

Heavy-flavor quark jet substructure

Radiation pattern of light-quark & gluon-initiated jets governed by soft & collinear divergences of QCD

Heavy quark mass term “regularizes” QCD divergences
→Harder fragmentation, dead cone effect, …

Massive splitting function

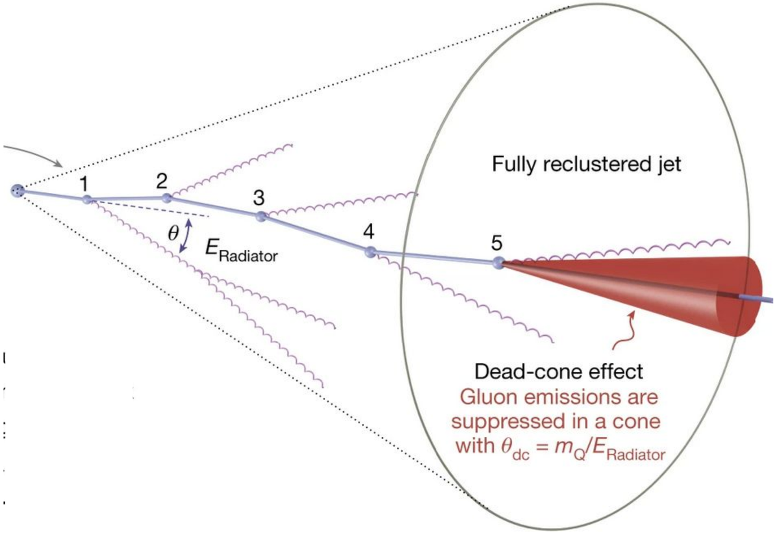
$$P_{Q\to Qg}(z) = \frac{1-z}{z} + \frac{z}{2} - 2\mu^2_{Qg}$$

$$\mu^2_{Qg} = \frac{m^2_Q}{m^2_{Qg} - m^2_Q}$$

dead cone effect

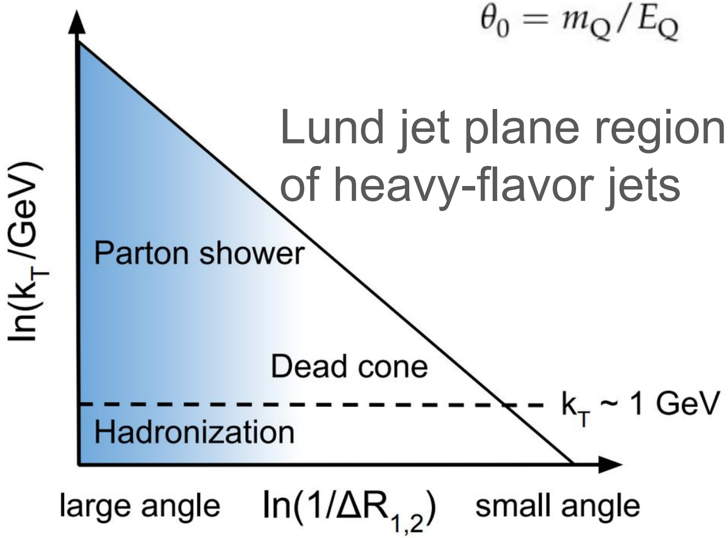
$$\mathrm{d}\mathcal{P}(\theta) \propto \frac{\mathrm{d}\theta^2}{(\theta^2 + \theta_0^2)^2}$$

$$\theta_0 = m_Q / E_Q$$



c/b quark

Lund jet plane regions of heavy-flavor jets



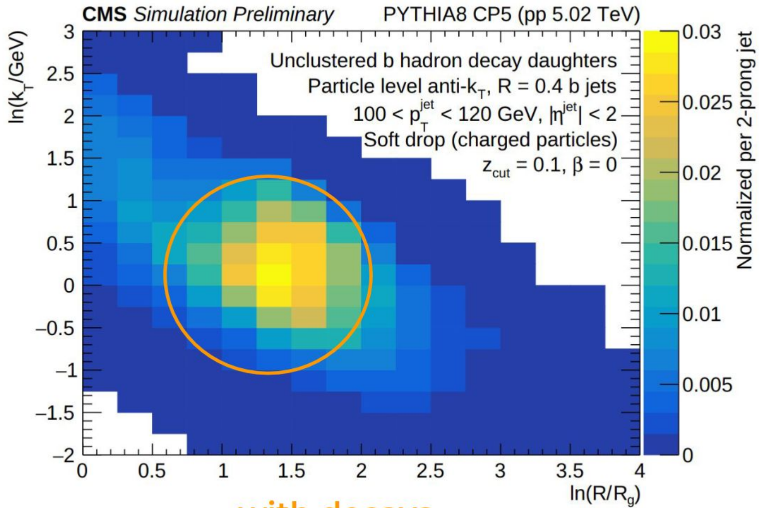
Nature **volume 605**, pages 440 (2022)

Contamination of heavy-flavor hadron decays

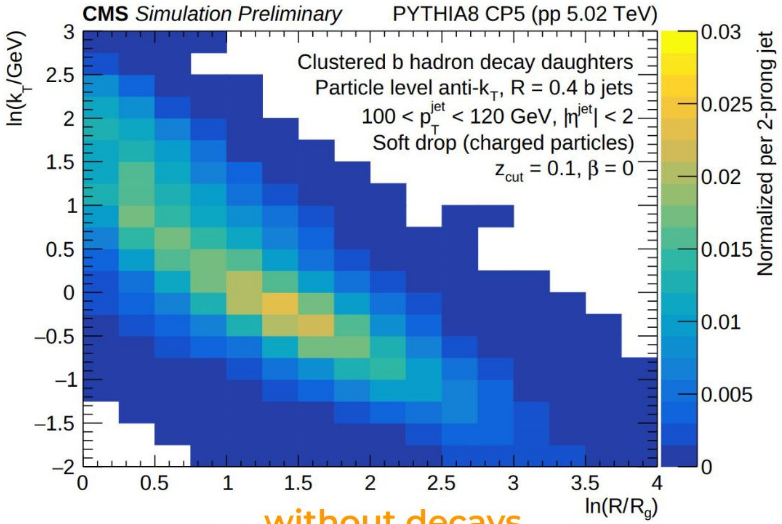
Decays **distort** the QCD radiation pattern of interest

For c-jets, one can use exclusive D meson decays (e.g., $D^0 \to K^- \pi^+$) to mitigate it

For b jets, exclusive decays (eg $B^+ \to (J/\psi \to \mu^+\mu^-)\, K^+$) are rarer, need to use other approaches (TMVA-based “clustering” of b hadron decays)



with decays



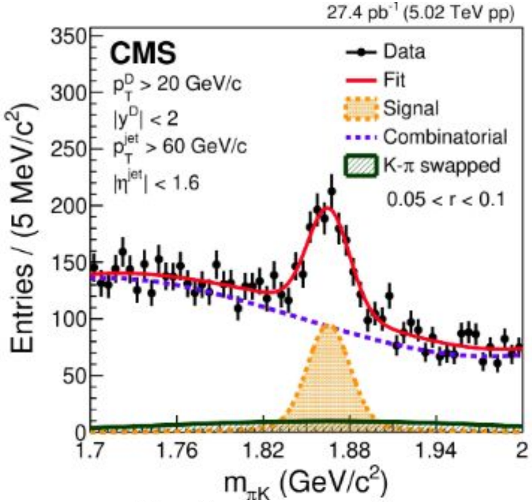
without decays

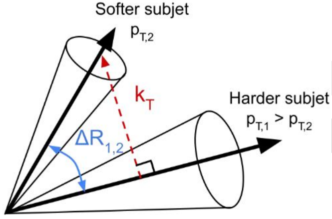
CMS-PAS-HIN-24-005

Collinear emissions suppressed for D^0-tagged jets

CMS-PAS-HIN-24-007, see also Jelena Mijuskovic's talk at BOOST'24

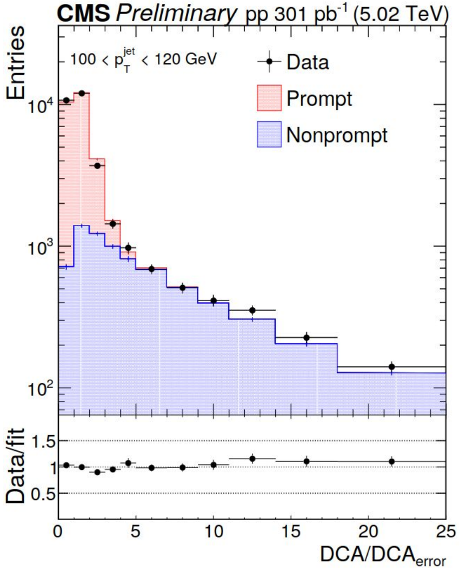
Substructure-dependent $D^0 \to K^- \pi^+$ yield extraction





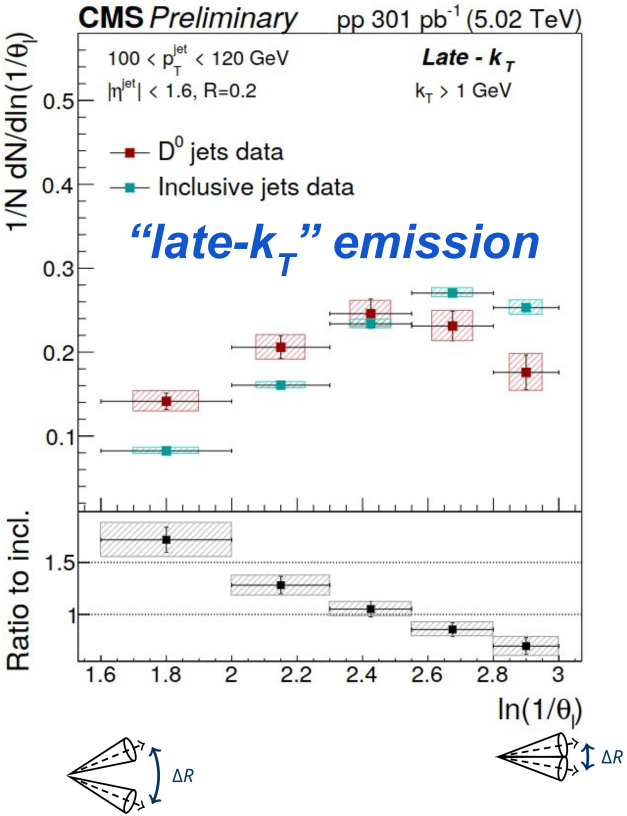
D^0 in harder subjet

Prompt fraction



Distance of closest approach significance

D-jet vs inclusive jet*



"late-k_T" emission

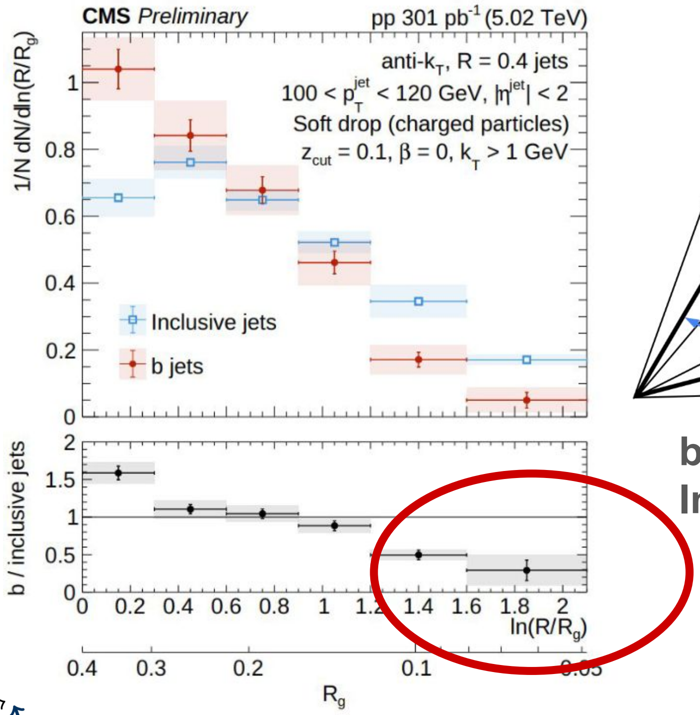
* quark+gluon jet mixture

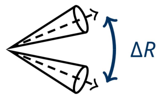
Bottom quark jet substructure

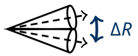
CMS-PAS-HIN-24-005

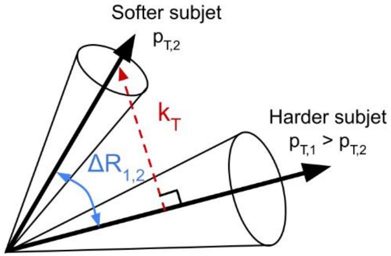
[see also Lida Kalipoliti's talk at BOOST’24]

Collinear emissions are suppressed for **b jets** relative to **inclusive jets** (quark/gluon mixture)



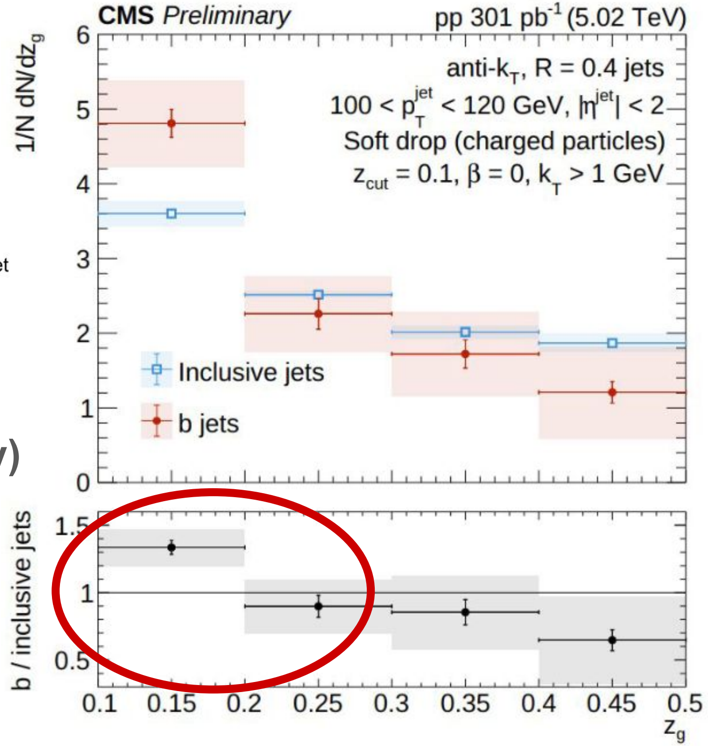






b hadron (mostly) In leading subjet

More asymmetric momentum imbalance for **b jets**



Soft-drop angle R_g

Splitting function

Closing remarks

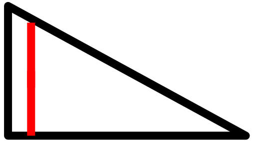
CMS Lund jet plane density,
extending to other fronts
(heavy-flavor jet substructure)

Analyses ongoing in heavy-ions
(not shown here), interest in using LJP to probe
spacetime evolution of quark-gluon plasma

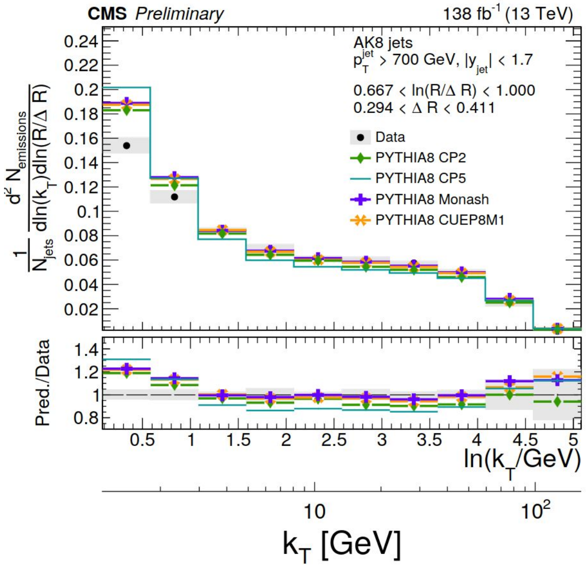
Comparison to parton showers& tunes

CMS-PAS-SMP-22-007

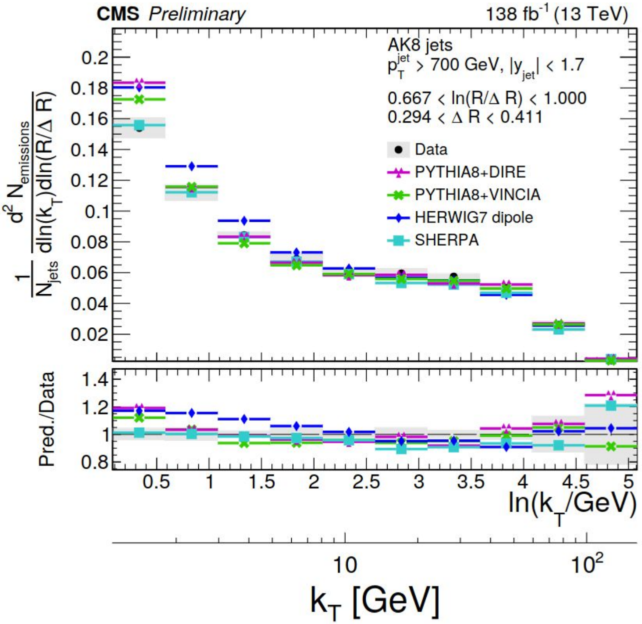
Large angle
emissions



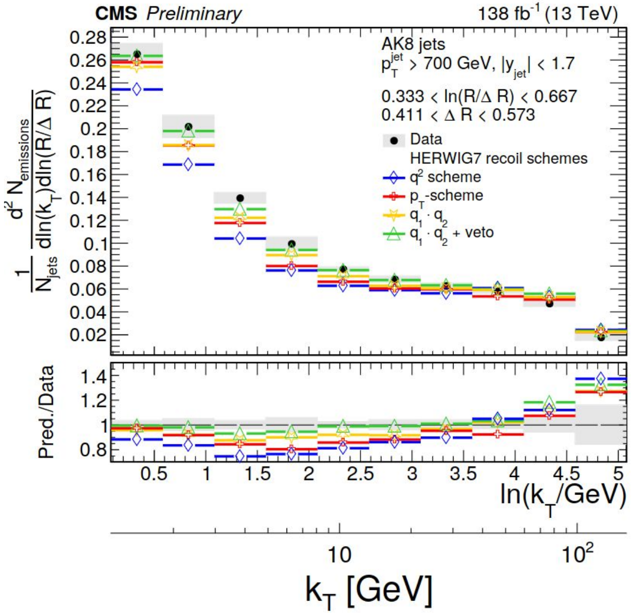
R = 0.8



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.

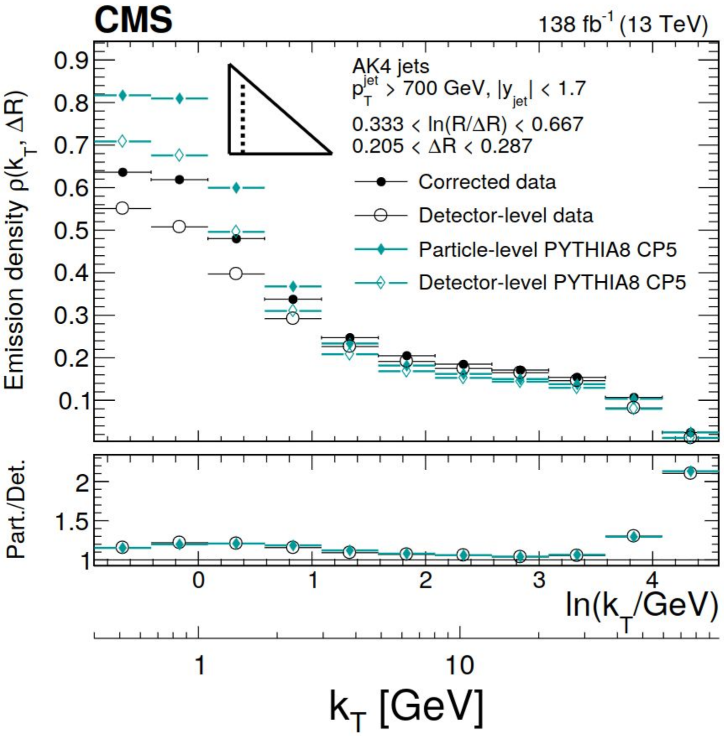
HERWIG7 angle-ordered describes better the data than **HERWIG7 dipole**

Factorization of effects can be exploited in MC tuning

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.



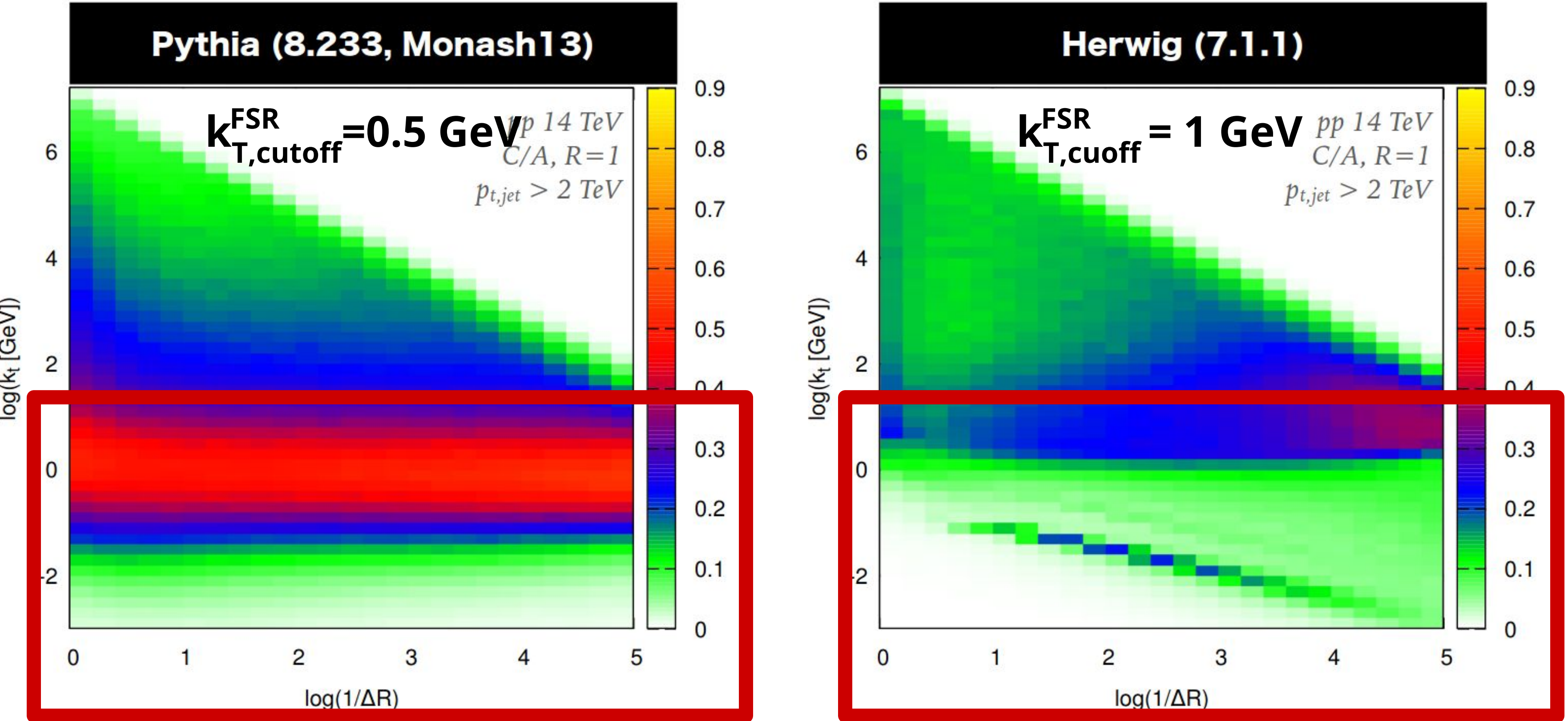
smearing becomes more important at high k_T (kinematical edge)

Cristian Baldenegro (MIT)

…Or could it be something else, e.g., the FSR k_T cutoff choice?

average pp Lund density: parton level

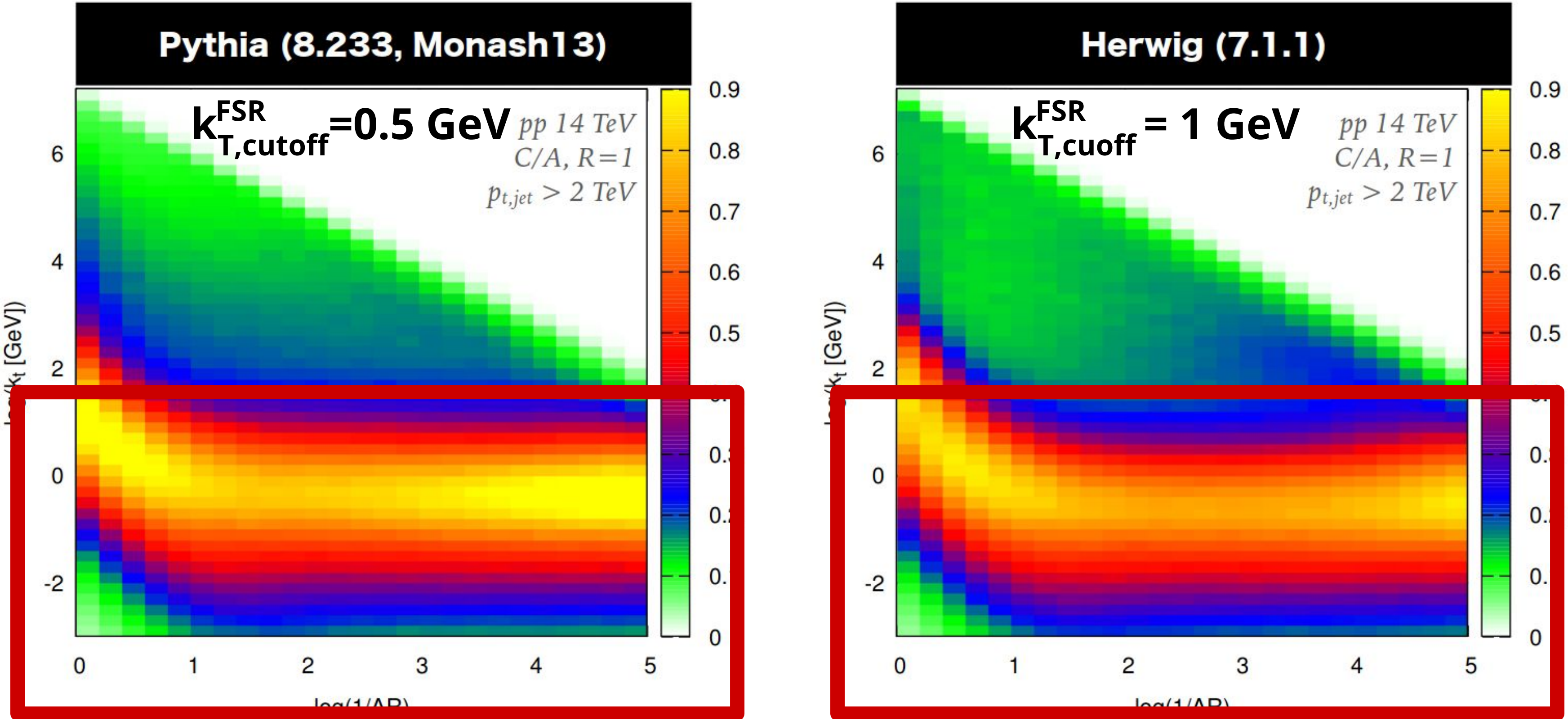
G. Salam's slide



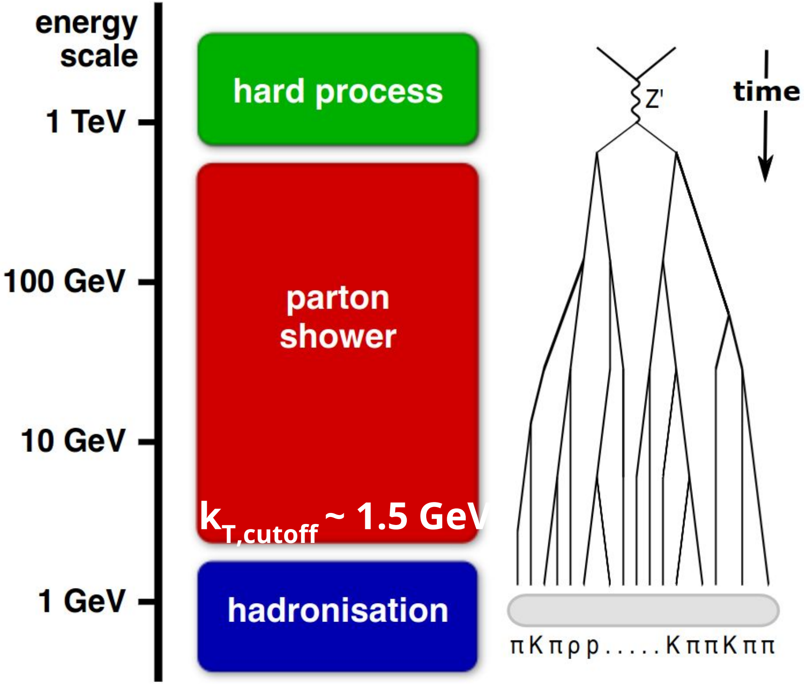
Hadron-level: FSR k_T cutoff choice shouldn't matter…

average pp Lund density: hadron level (with underlying event / MPI)

G. Salam's slide

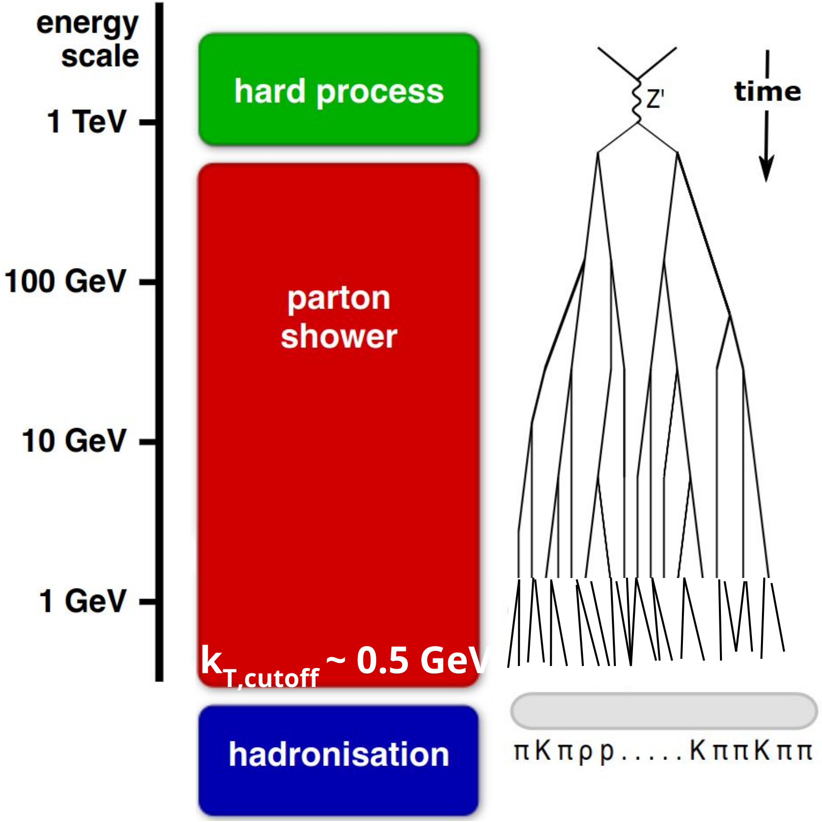


Higher shower $k_{T,\mathrm{cutoff}}$



$k_{T,\mathrm{cutoff}}$ ~ 1.5 GeV

Lower shower $k_{T,\mathrm{cutoff}}$

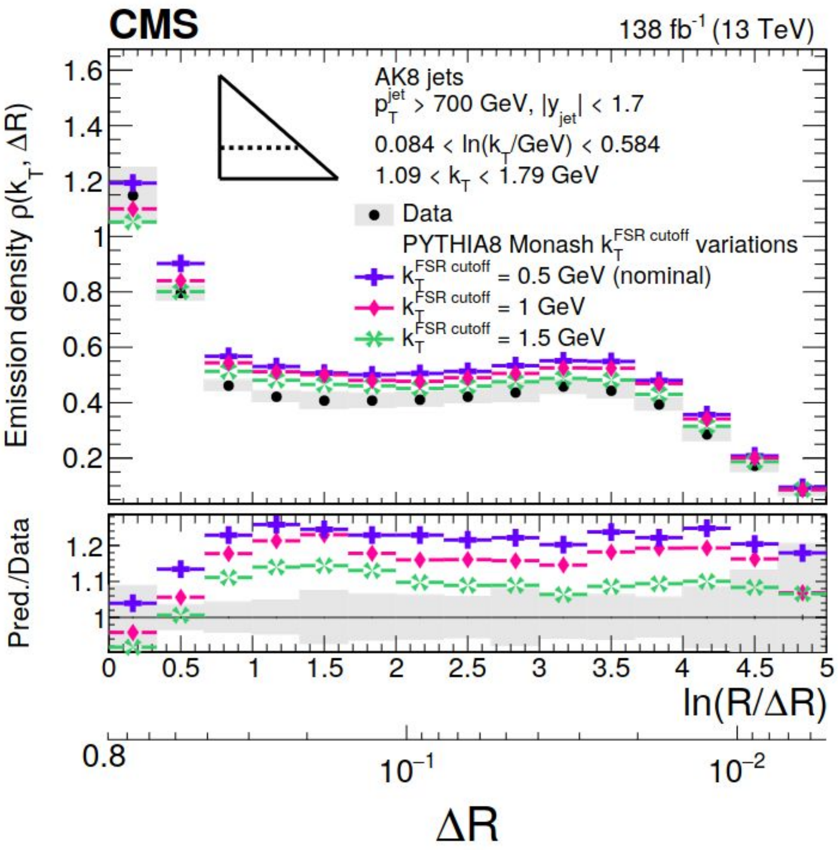


$k_{T,\mathrm{cutoff}}$ ~ 0.5 GeV

Could it lead to double counting?

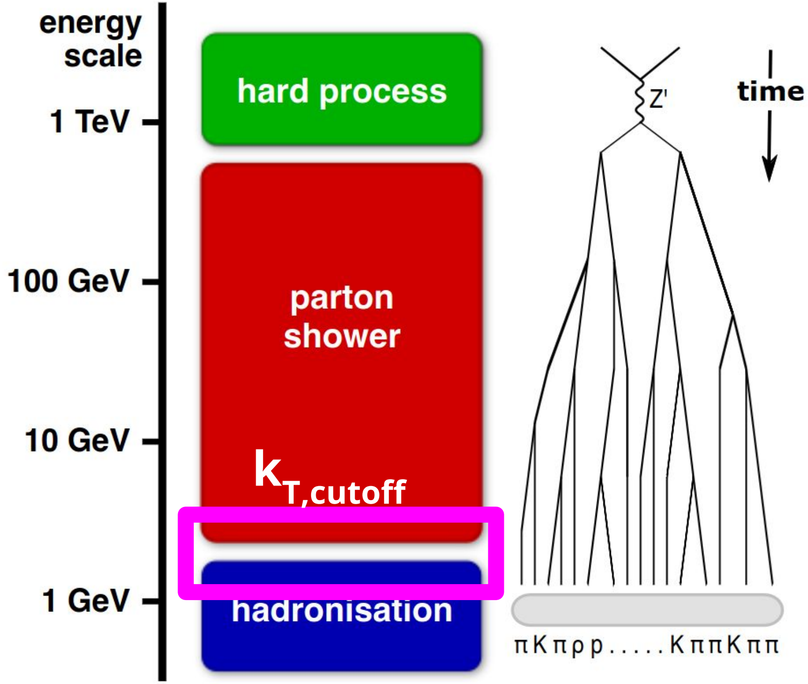
PYTHIA8 shower $k_{T,\mathrm{cutoff}}$ variations (k_T ~ 1 GeV)

arXiv:2312.16343, accepted by JHEP



Larger FSR k_T cutoff ⇔ fewer Lund emissions

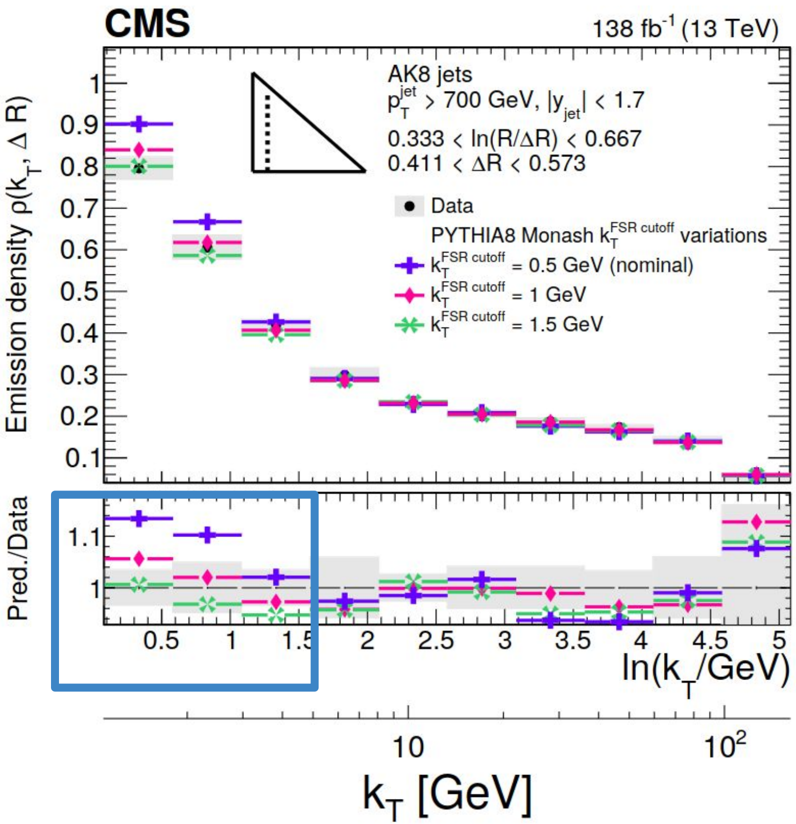
Data "prefers" higher FSR $k_{T,\mathrm{cutoff}}$ = 1.5 GeV for PYTHIA8

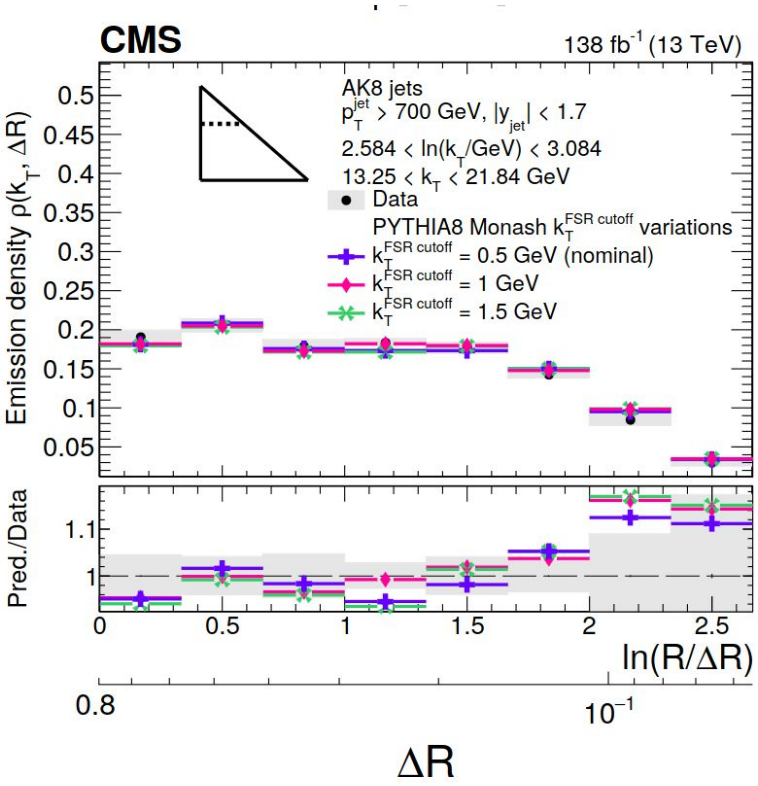


Low shower $k_{T,\mathrm{cutoff}}$ ⇒ "double counting"

Shower $k_{T,\mathrm{cutoff}}$ decouples at k_T ~ 4 GeV

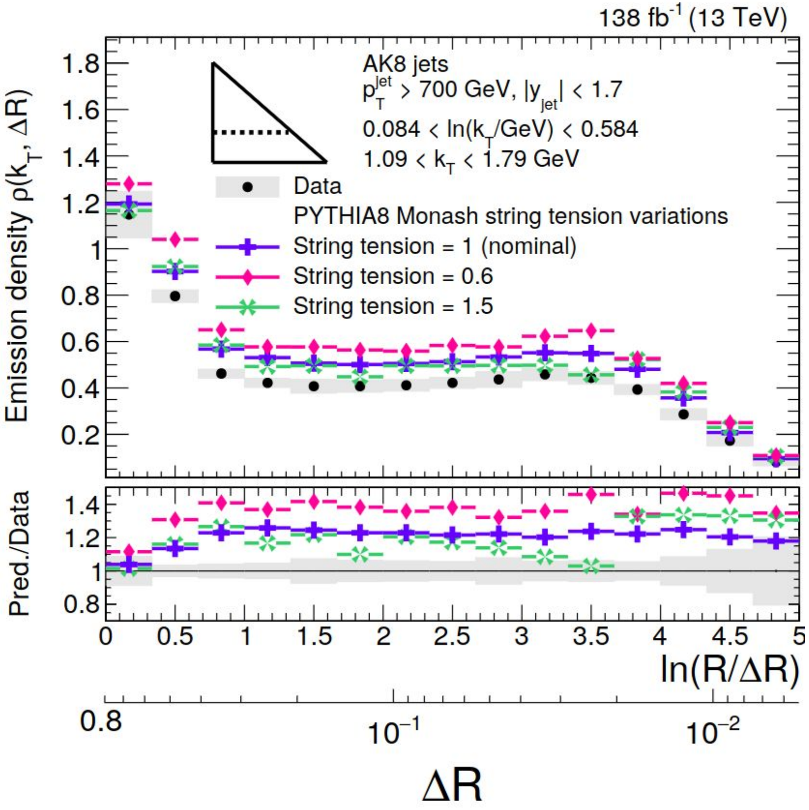
arXiv:2312.16343, accepted by JHEP





String tension sensitivity

Low $k_T \sim 1$ GeV



High $k_T > 8 – 36$ GeV