Reconstructing and calibrating jets

ULB

Andrea Malara
(Université Libre de Bruxelles)

LHC-EW WG2

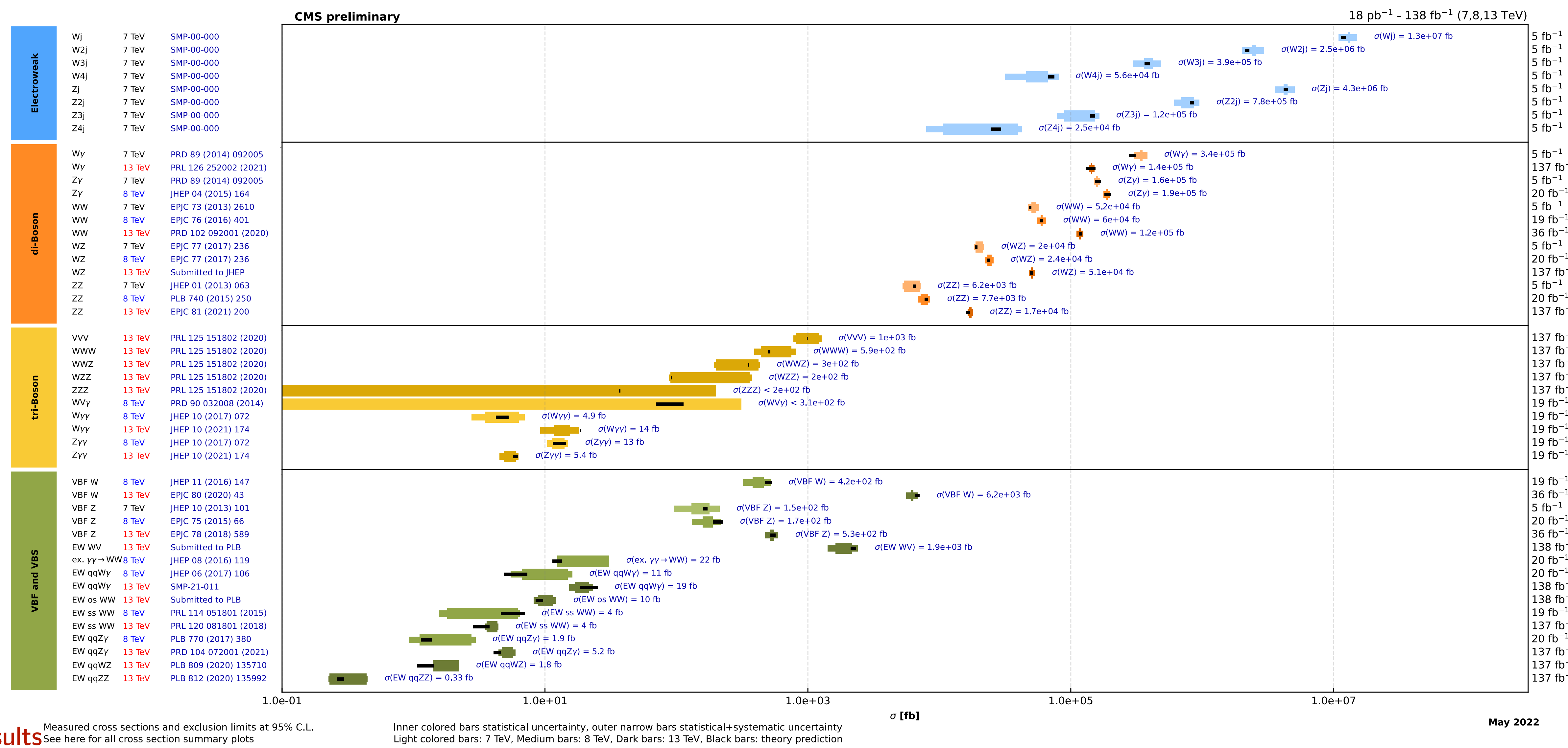
On behalf of the
CMS Collaboration

CMS
Compact Muon Solenoid

Physics programme

- Test the self-consistency of the Standard Model
  - Huge variety of processes analysed
  - Approximatively 7 orders of magnitudes of cross-sections measured
  - Majority already systematics-dominated

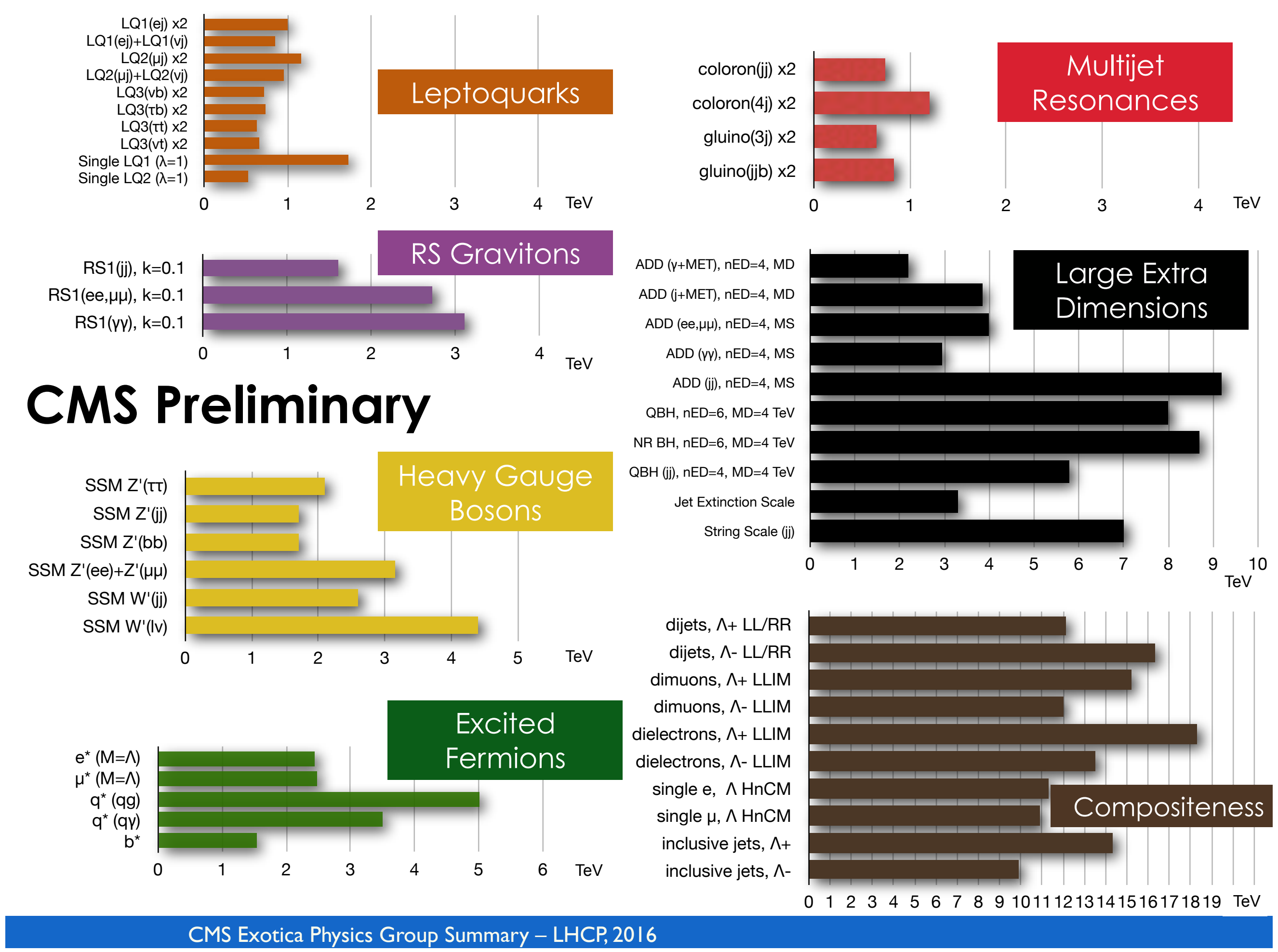
Overview of CMS cross section results

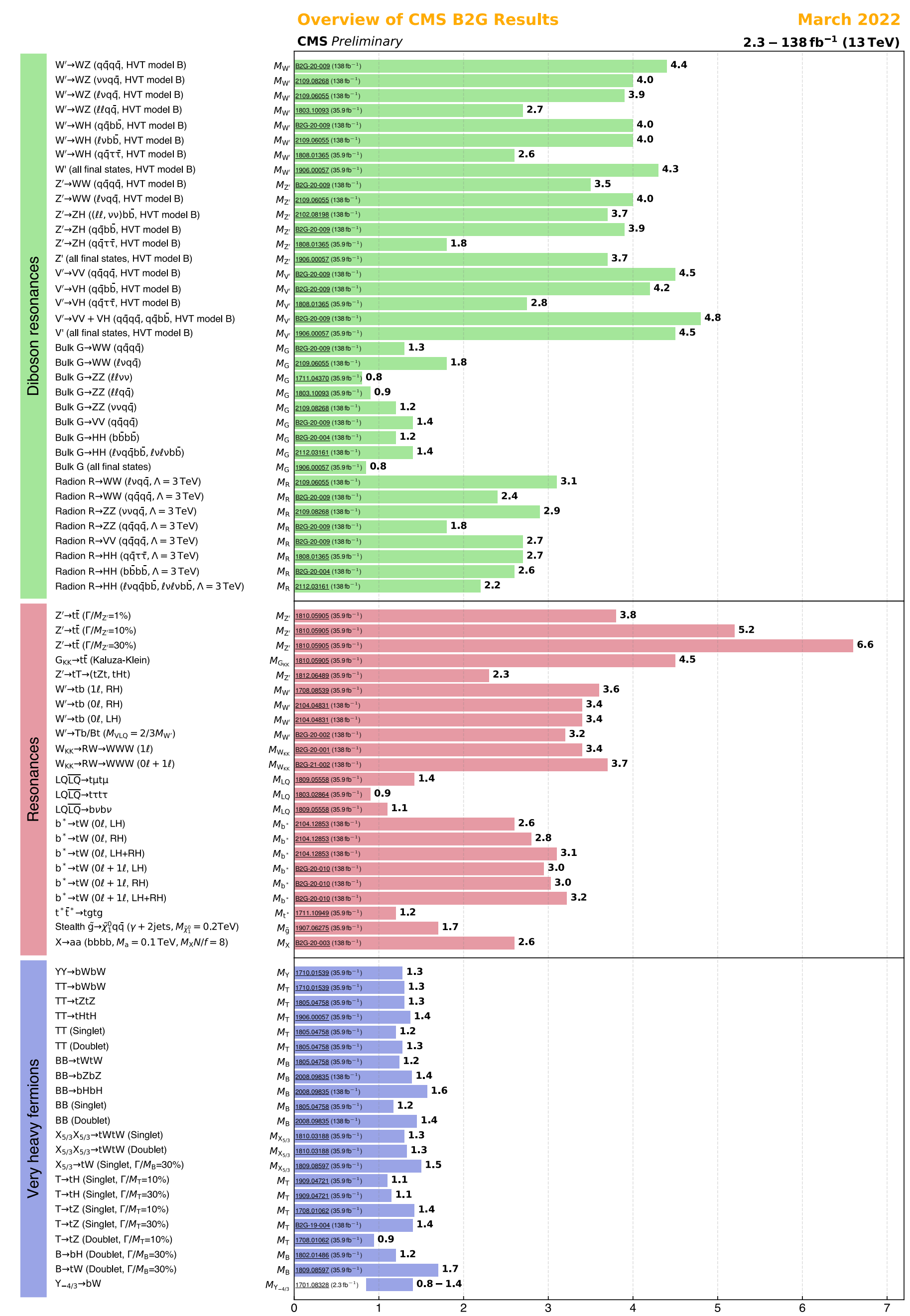


CMS results

Physics programme

- Test the self-consistency of the Standard Model
- TeV energy frontier
  - Enormous range of energy investigated (up to 10TeV)
  - Multitude of different models studied

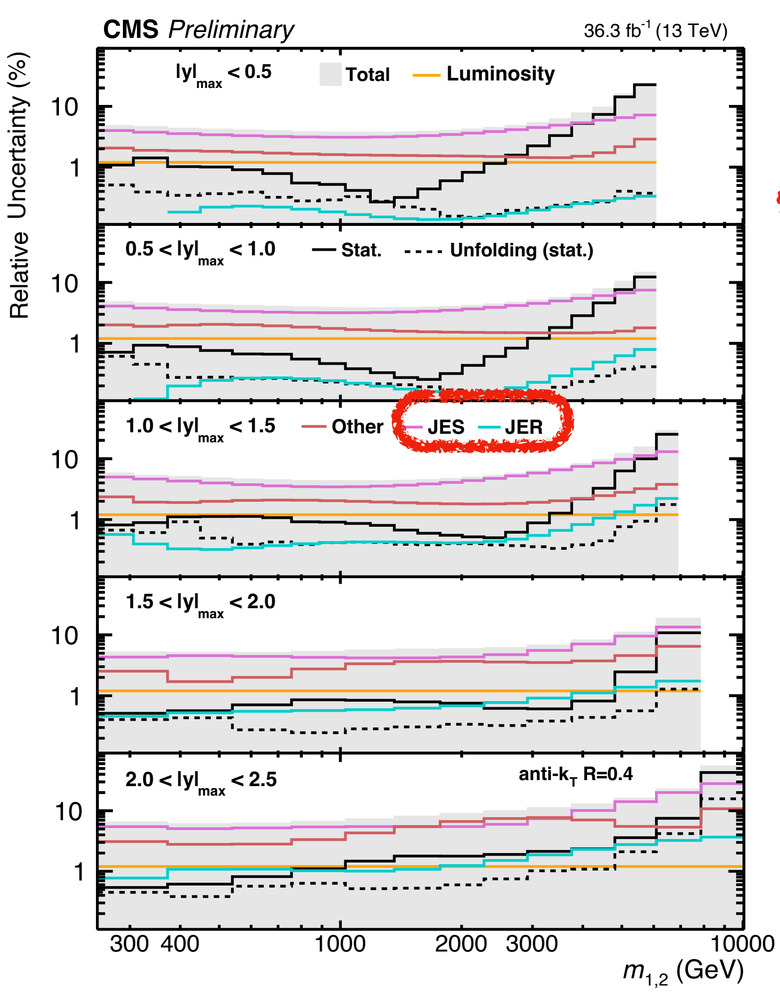




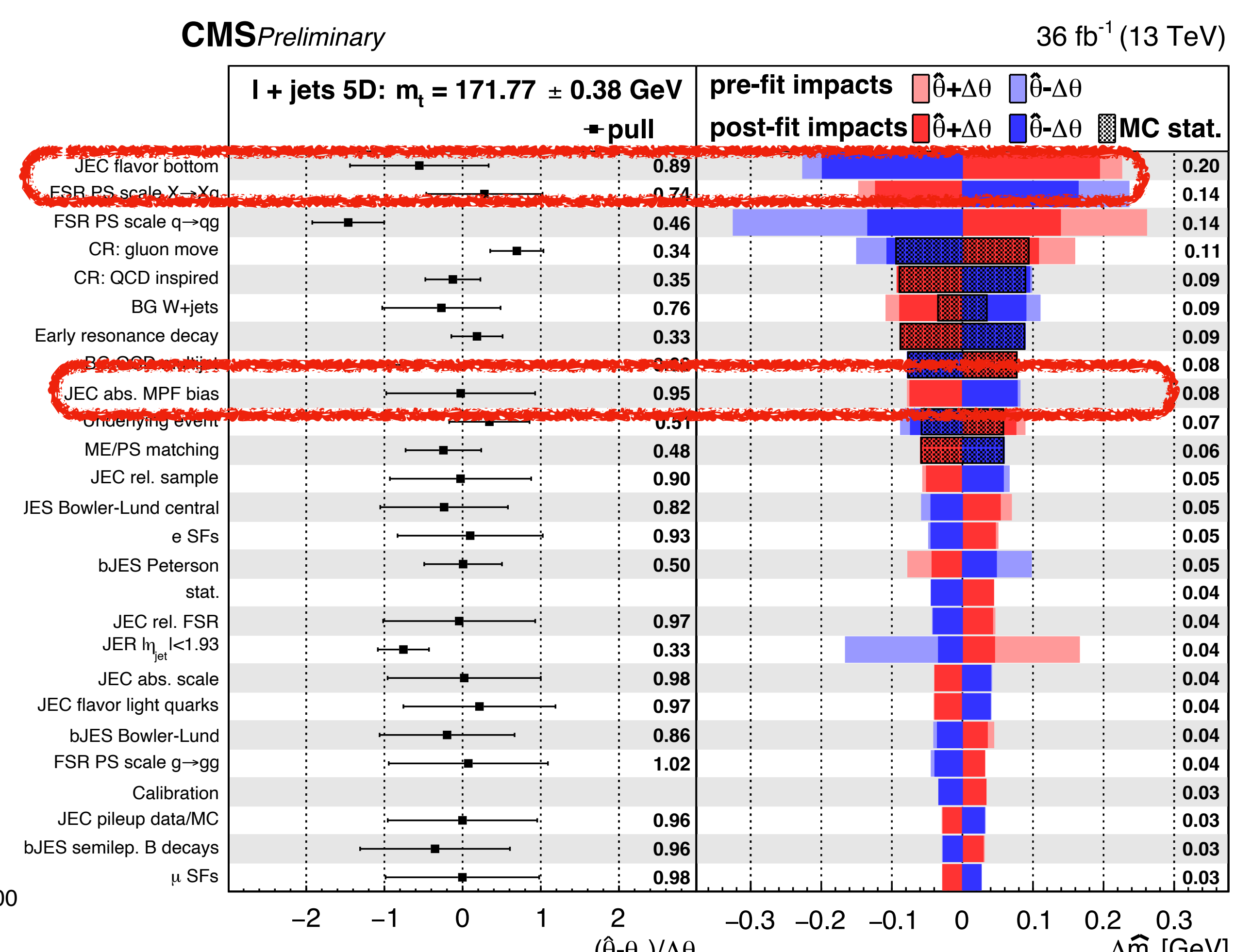
CMS results

Era of precision physics

- Jet-related uncertainties are becoming a limiting factor in many analyses
  - Jets are abundant at the LHC -> hadronic decays, associated prod. with jets...
  - Jet energy scale -> impact on: top, Higgs, multijet analyses



Dijet cross section



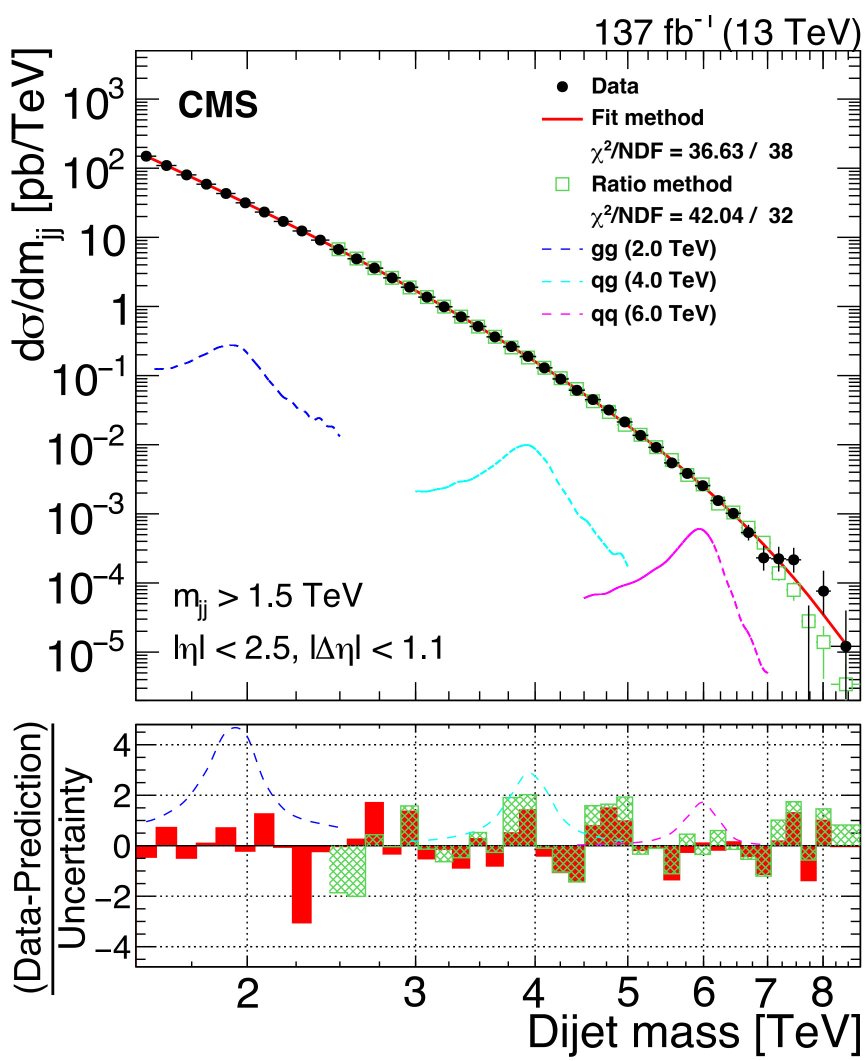
Top mass l+jet

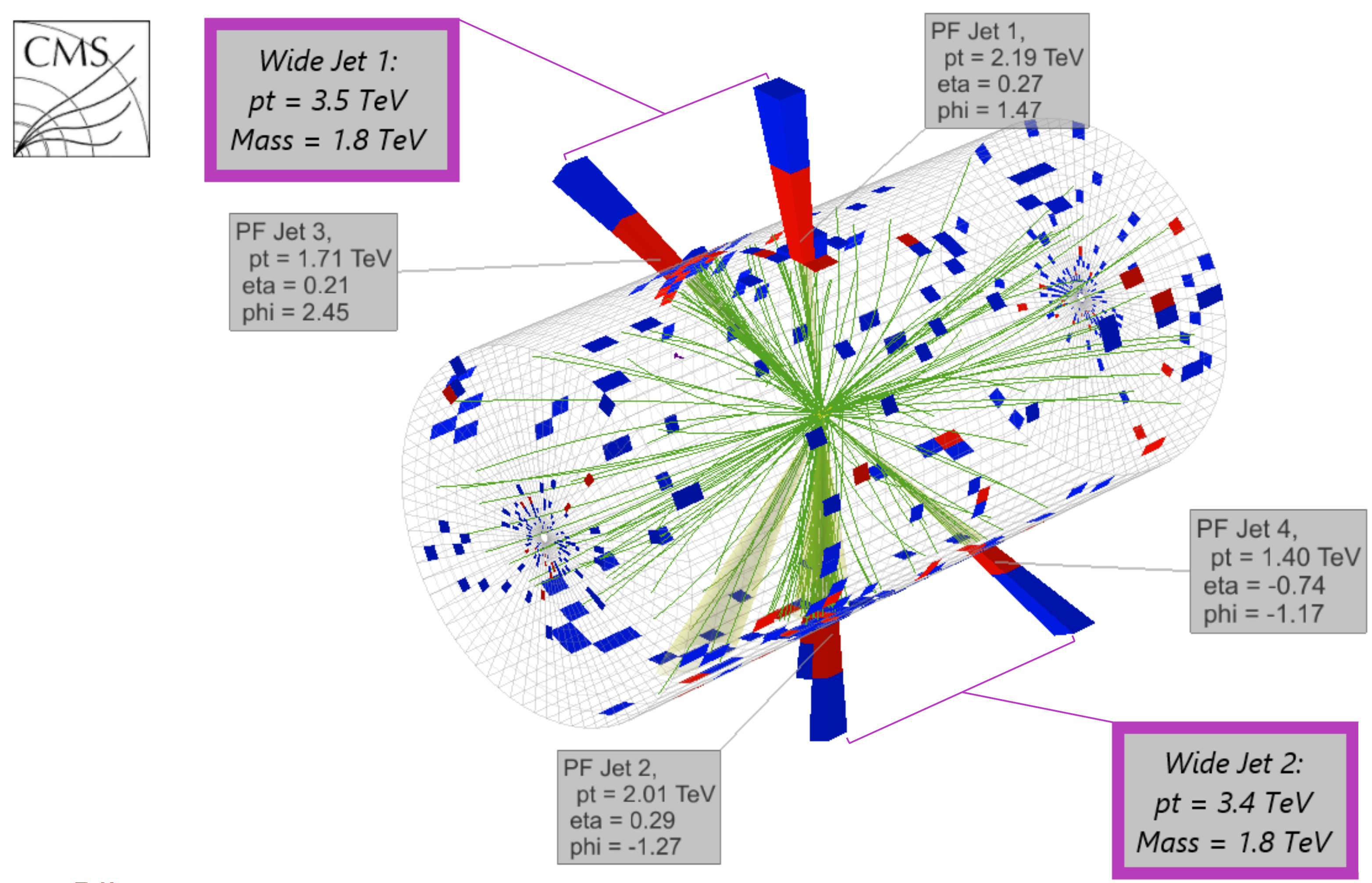
ttbar cross section dilepton

Source	Uncertainty [GeV]
Trigger	0.02
Lepton ident./isolation	0.02
Muon momentum scale	0.03
Electron momentum scale	0.10
Jet energy scale	0.57
Jet energy resolution	0.09
b tagging	0.12
Pileup	0.09
$t\bar{t}$ ME scale	0.18
tW ME scale	0.02
DY ME scale	0.06
NLO generator	0.14
PDF	0.05
$\sigma_{t\bar{t}}$	0.09
Top quark p_T	0.04
ME/PS matching	0.16
UE tune	0.03
$t\bar{t}$ ISR scale	0.16
tW ISR scale	0.02
$t\bar{t}$ FSR scale	0.07
tW FSR scale	0.02
b quark fragmentation	0.11
b hadron BF	0.07
Colour reconnection	0.17
DY background	0.24
tW background	0.13
Diboson background	0.02
W+jets background	0.04
$t\bar{t}$ background	0.02
Statistical	0.14
MC statistical	0.36
Total m_t^{MC} uncertainty	$^{+0.68}_{-0.73}$

Era of precision physics

- Jet-related uncertainties are becoming a limiting factor in many analyses
  - Jets are abundant at the LHC -> hadronic decays, associated prod. with jets...
  - Jet energy scale -> impact on: top, Higgs, multijet analyses
  - ... but also boosted searches -> merged decay products
  - Must be known very well for a wide range in energy and pseudorapidity





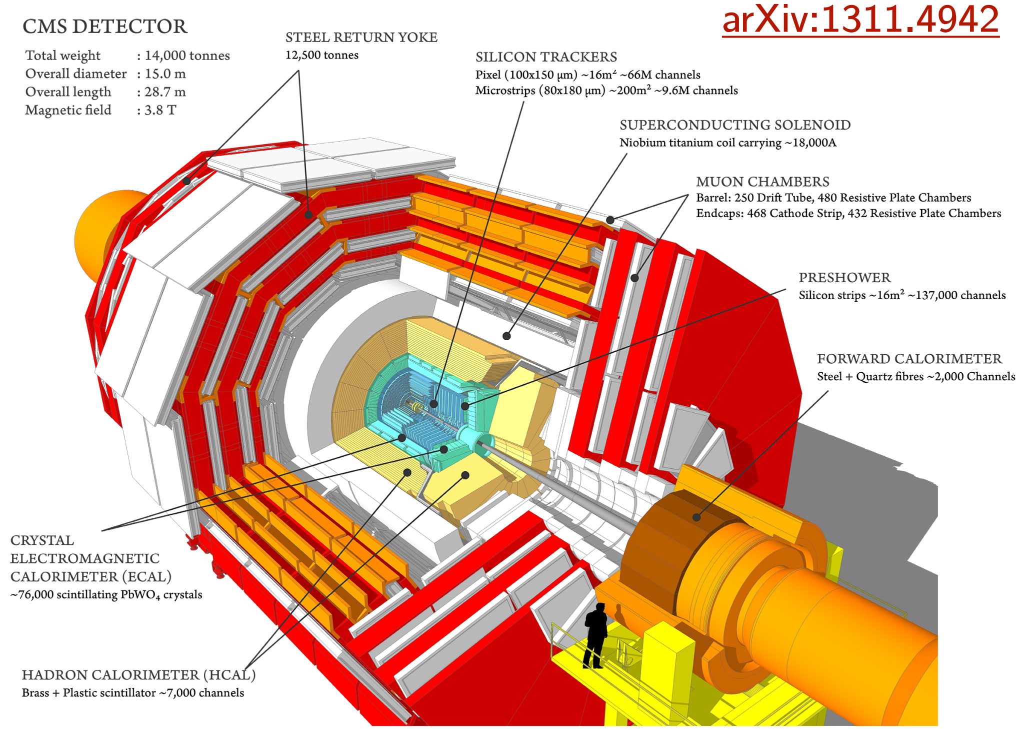
Dijet resonances

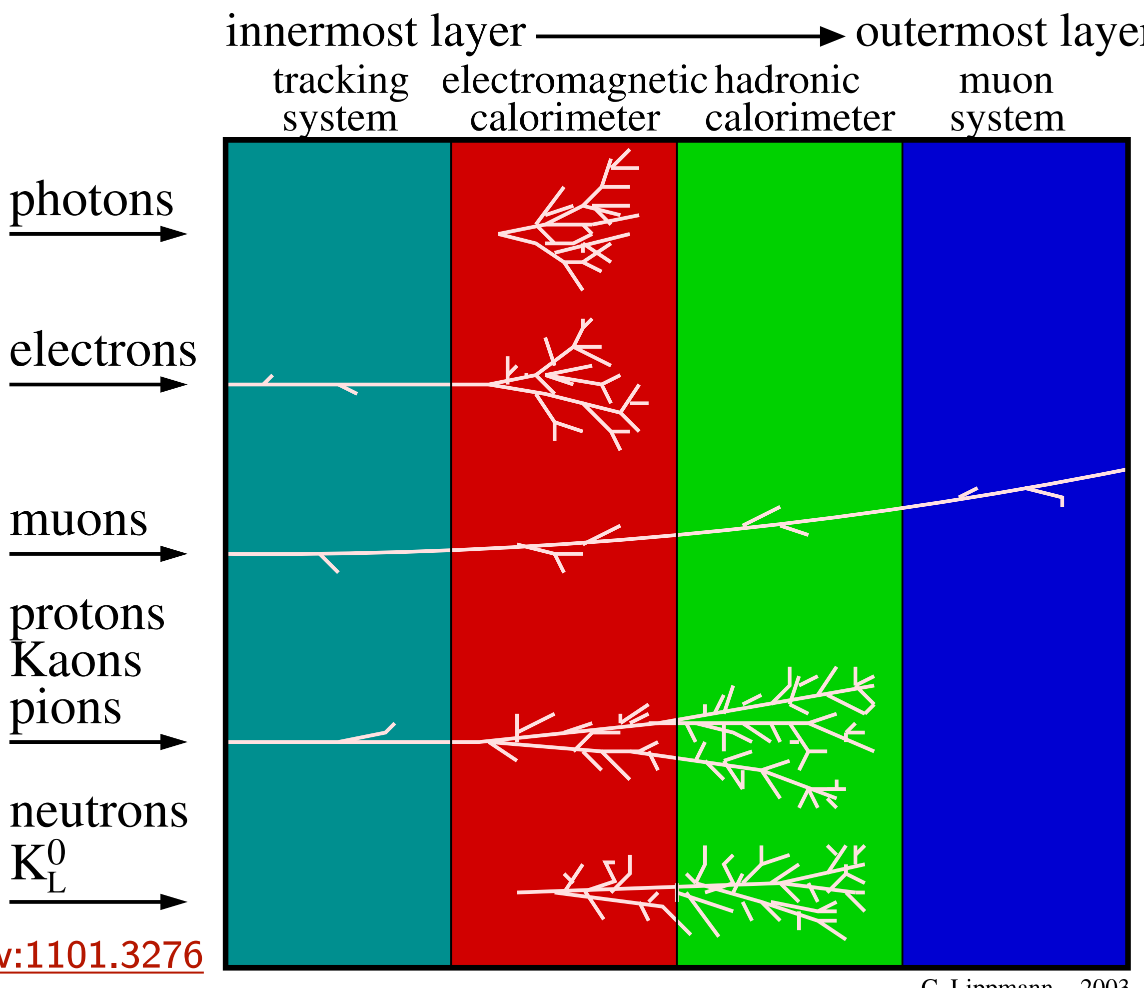
Experimental setup: CMS

- Multi-purpose detector
- Layered structure
  - Tracker
  - Electromagnetic calorimeter
  - Hadron calorimeter
  - Solenoid
  - Muon chambers

- Particle reconstruction
  - Detector signals -> physics objects
  - Based on ParticleFlow algorithm
- Operational since 2010, this talk focuses on:
  - Run2 data (2016-2018), $\sqrt{s}$ = 13 TeV
  - Run3 data (2022-ongoing), $\sqrt{s}$ = 13.6 TeV

arXiv:1311.4942

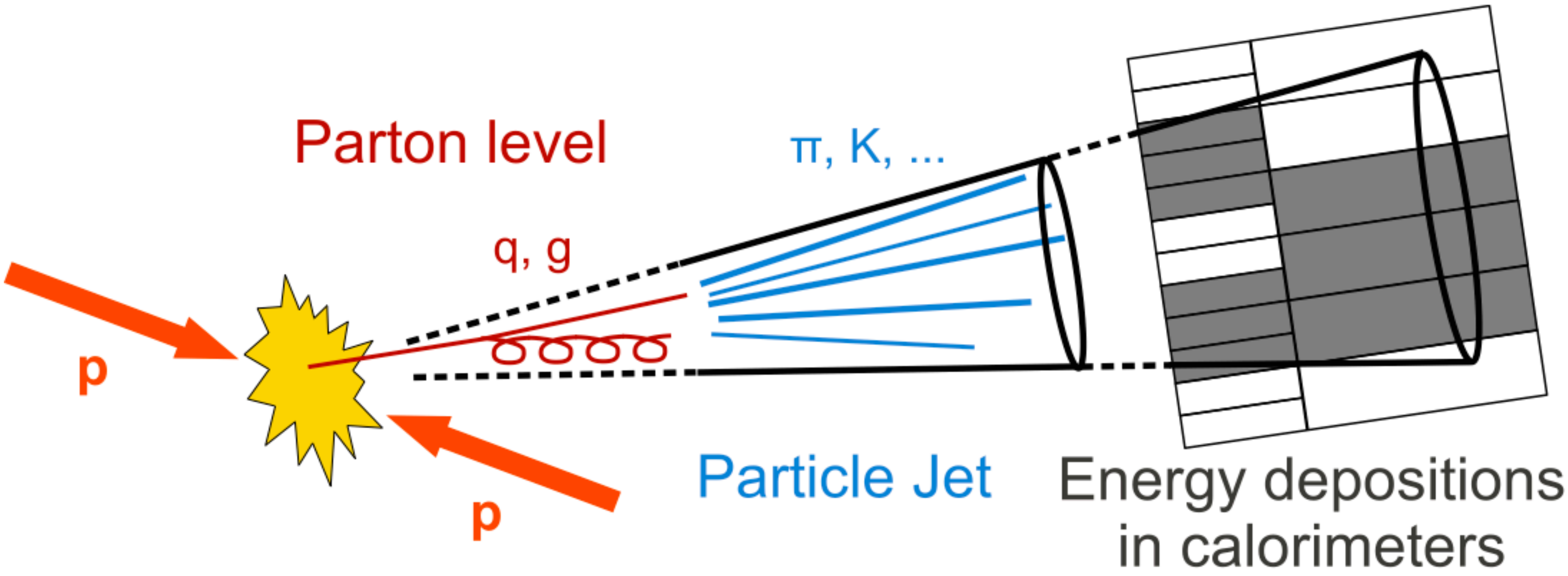




arXiv:1101.3276

Hadronic environment

▶ As a consequence of the hadronisation of quarks and gluons produced in pp collisions, a collimated shower of hadrons (jet) is produced.



[Click me](#)

From detector signals to jet calibration

Local reconstruction:
Tracks, ECAL, HCAL

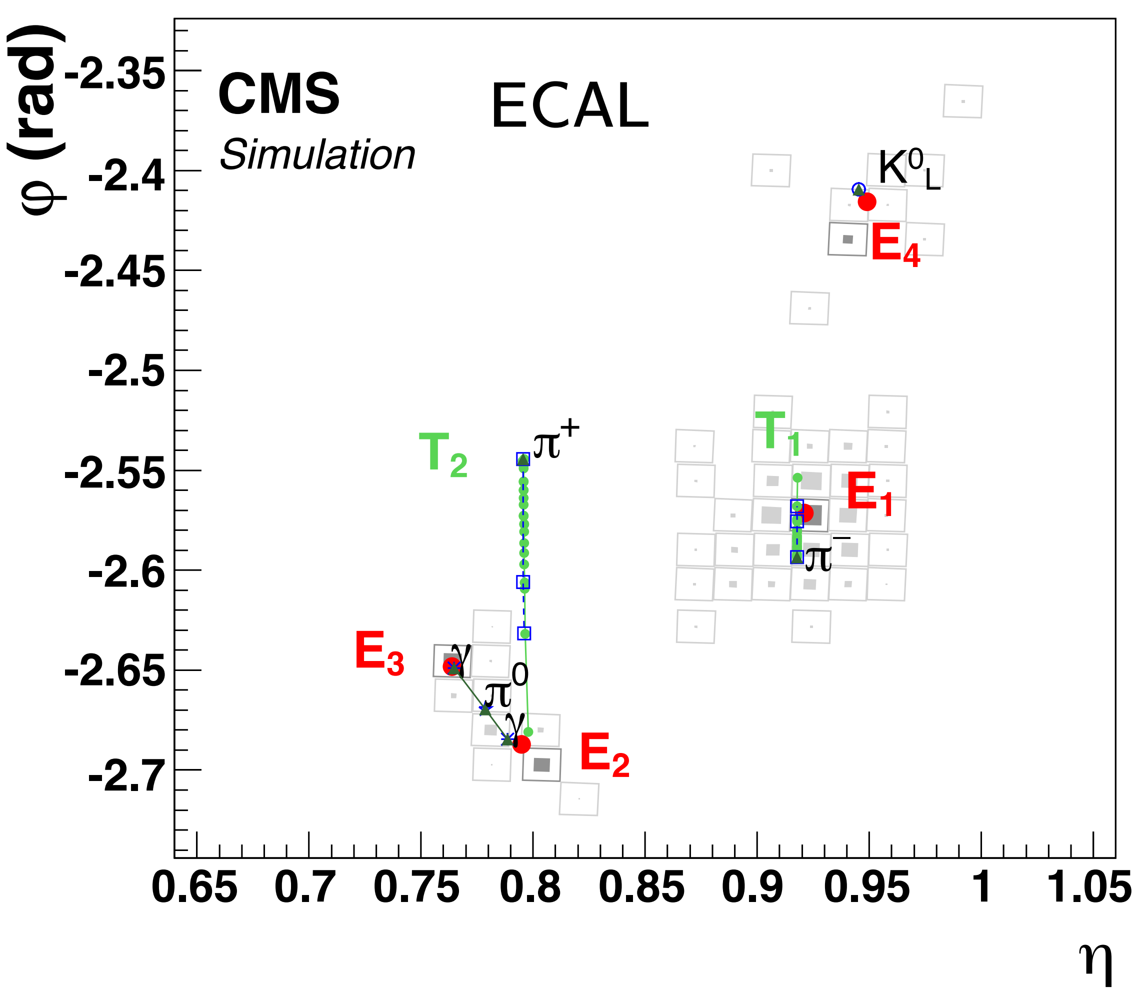
→

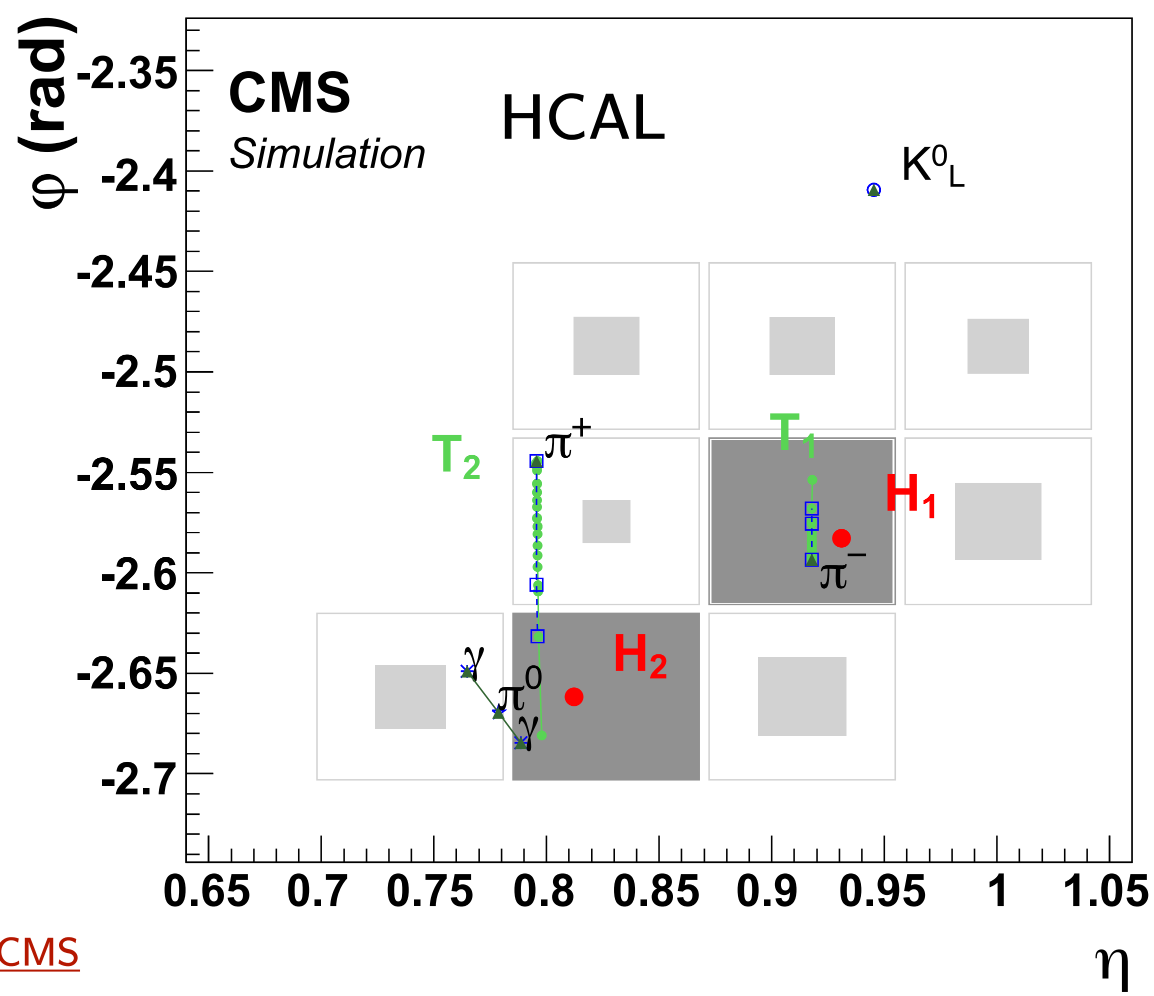
- Information from sub-detectors
- Similar method online but less detail

From detector signals to jet calibration

Local reconstruction: Tracks, ECAL, HCAL →
- Information from sub-detectors
- Similar method online but less detail

Particle flow (PF) →
- Link tracks and calorimeter signals
- Particle identification





Particle Flow in CMS

From detector signals to jet calibration

Local reconstruction: Tracks, ECAL, HCAL	→ ▶ Information from sub-detectors ▶ Similar method online but less detail
Particle flow (PF)	→ ▶ Link tracks and calorimeter signals ▶ Particle identification
Pileup mitigation	→ ▶ Charge hadron subtraction (CHS) ▶ Pileup Per Particle Identification (Puppi)

From detector signals to jet calibration

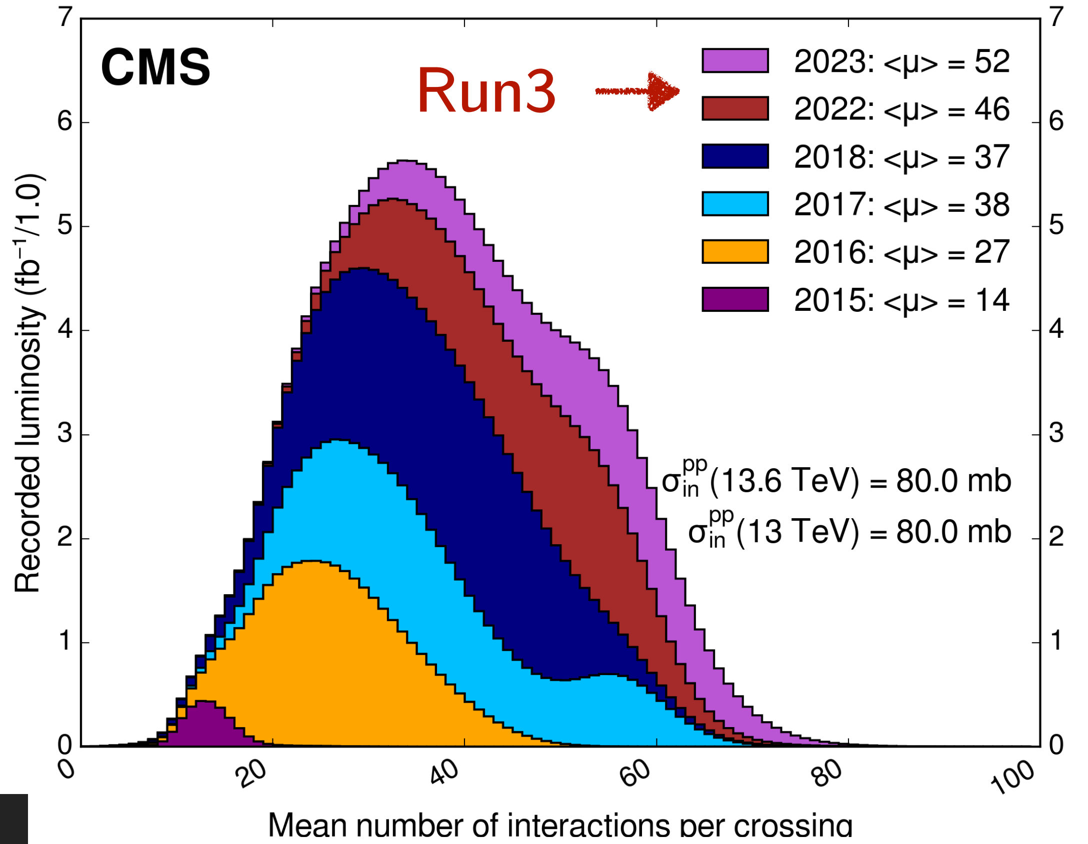
Local reconstruction: Tracks, ECAL, HCAL	▶ Information from sub-detectors ▶ Similar method online but less detail
Particle flow (PF)	▶ Link tracks and calorimeter signals ▶ Particle identification
Pileup mitigation	▶ Charge hadron subtraction (CHS) ▶ Pileup Per Particle Identification (Puppi)
Jet clustering	▶ Algorithms (AK, CA, HOTVR, Xcone) ▶ Cone radii (0.4, 0.8, 1.5, VR)

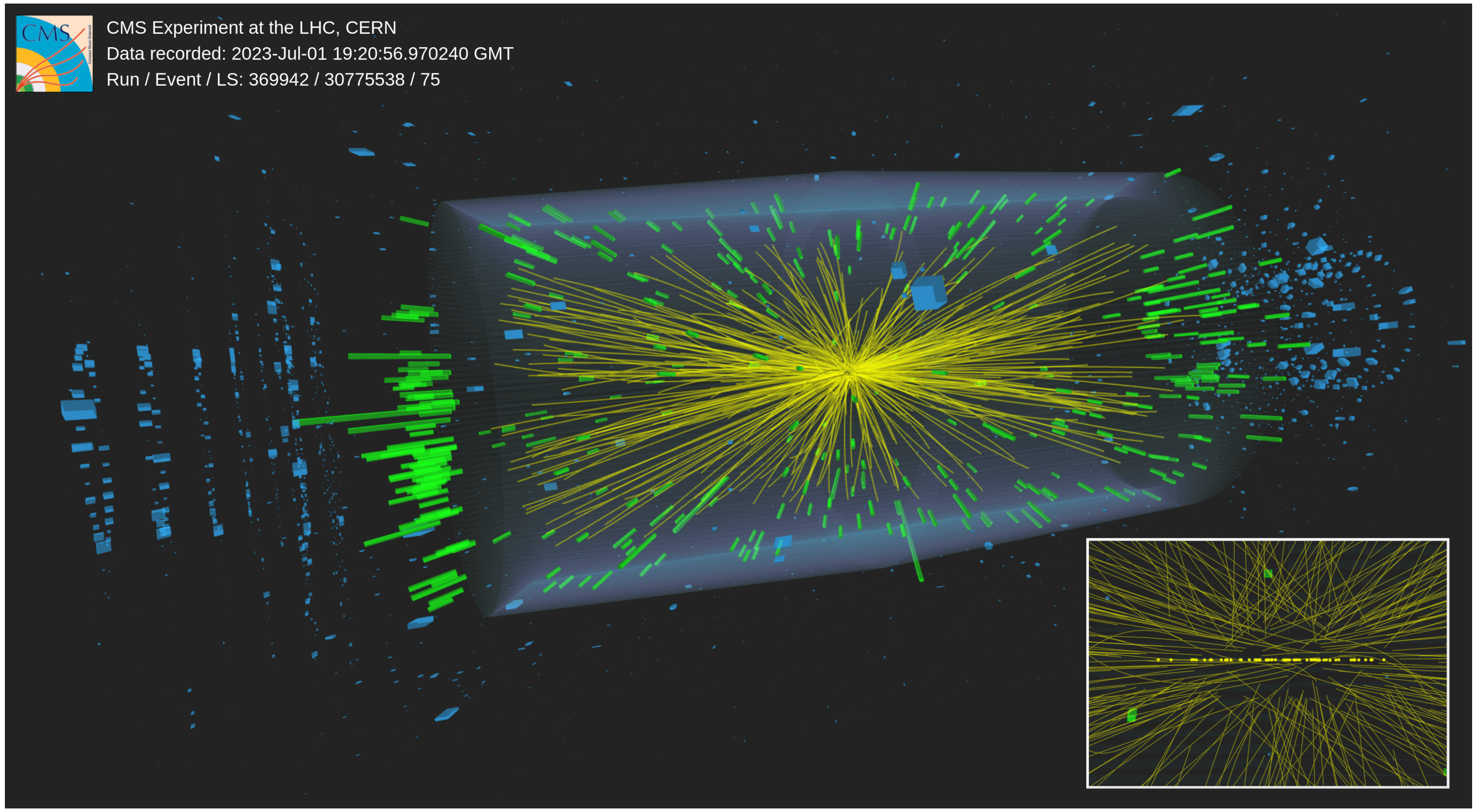
From detector signals to jet calibration

Local reconstruction: Tracks, ECAL, HCAL →	▶ Information from sub-detectors ▶ Similar method online but less detail
Particle flow (PF) →	▶ Link tracks and calorimeter signals ▶ Particle identification
Pileup mitigation →	▶ Charge hadron subtraction (CHS) ▶ Pileup Per Particle Identification (Puppi)
Jet clustering →	▶ Algorithms (AK, CA, HOTVR, Xcone) ▶ Cone radii (0.4, 0.8, 1.5, VR)
Jet calibration →	▶ Jet energy scale & resolution ▶ Jet mass scale & resolution

Pileup

- Challenging environment for LHC physics
  - Additional interactions (pileup)
  - Average of 30 interactions in Run 2
  - Average of 50 interactions in Run 3 (so far...)

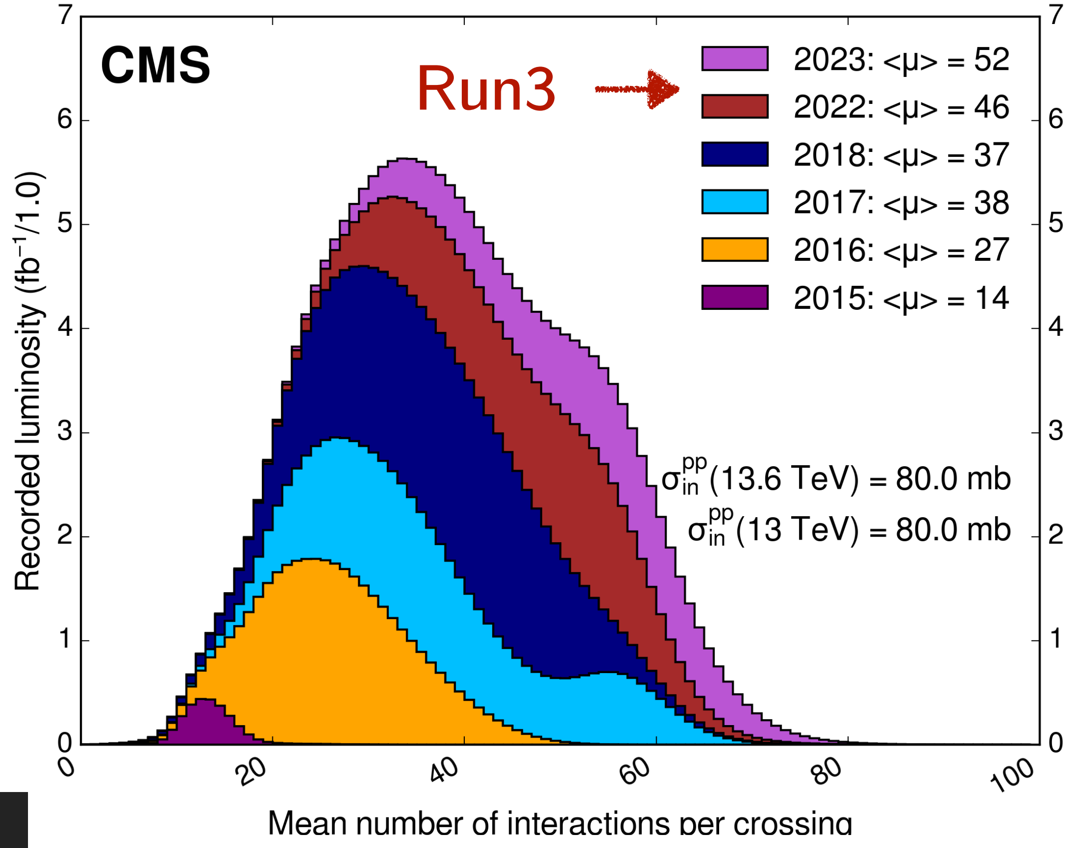


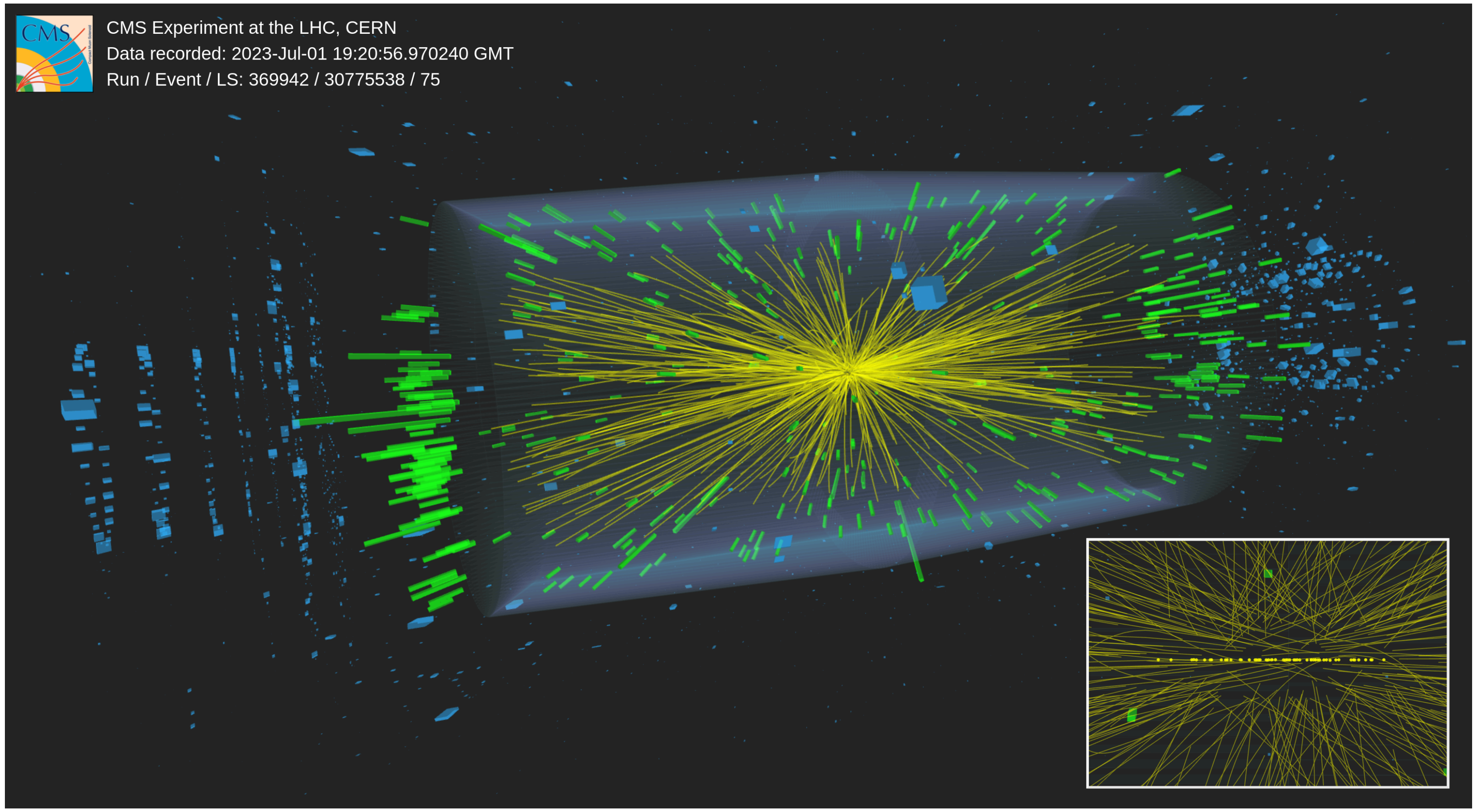


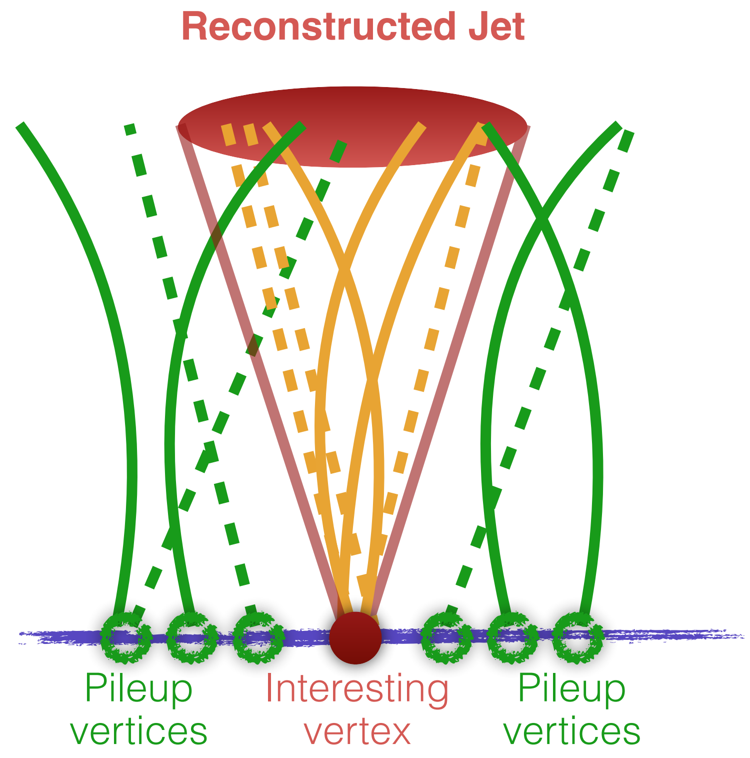
Proton-proton collision @ LHC

Pileup

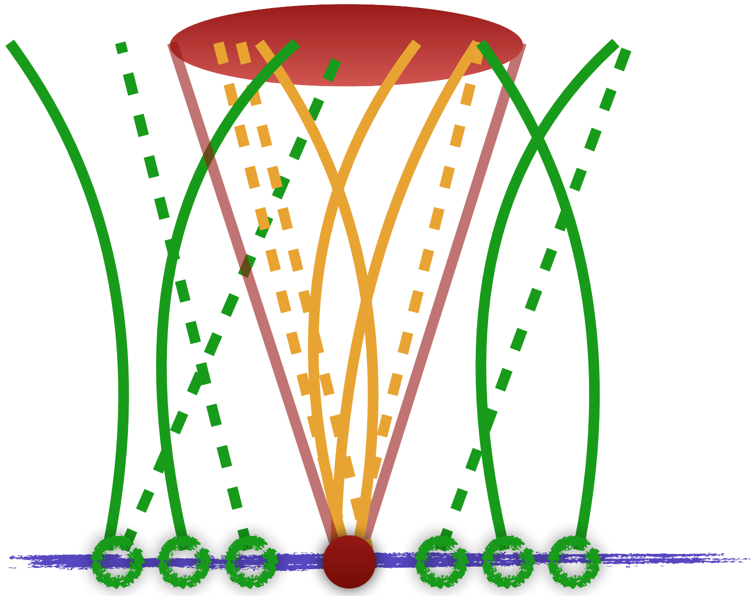
- Challenging environment for LHC physics
  - Additional interactions (pileup)
  - Average of 30 interactions in Run 2
  - Average of 50 interactions in Run 3 (so far...)
- Additional particles deteriorate measurements
- Several approaches to cope with it







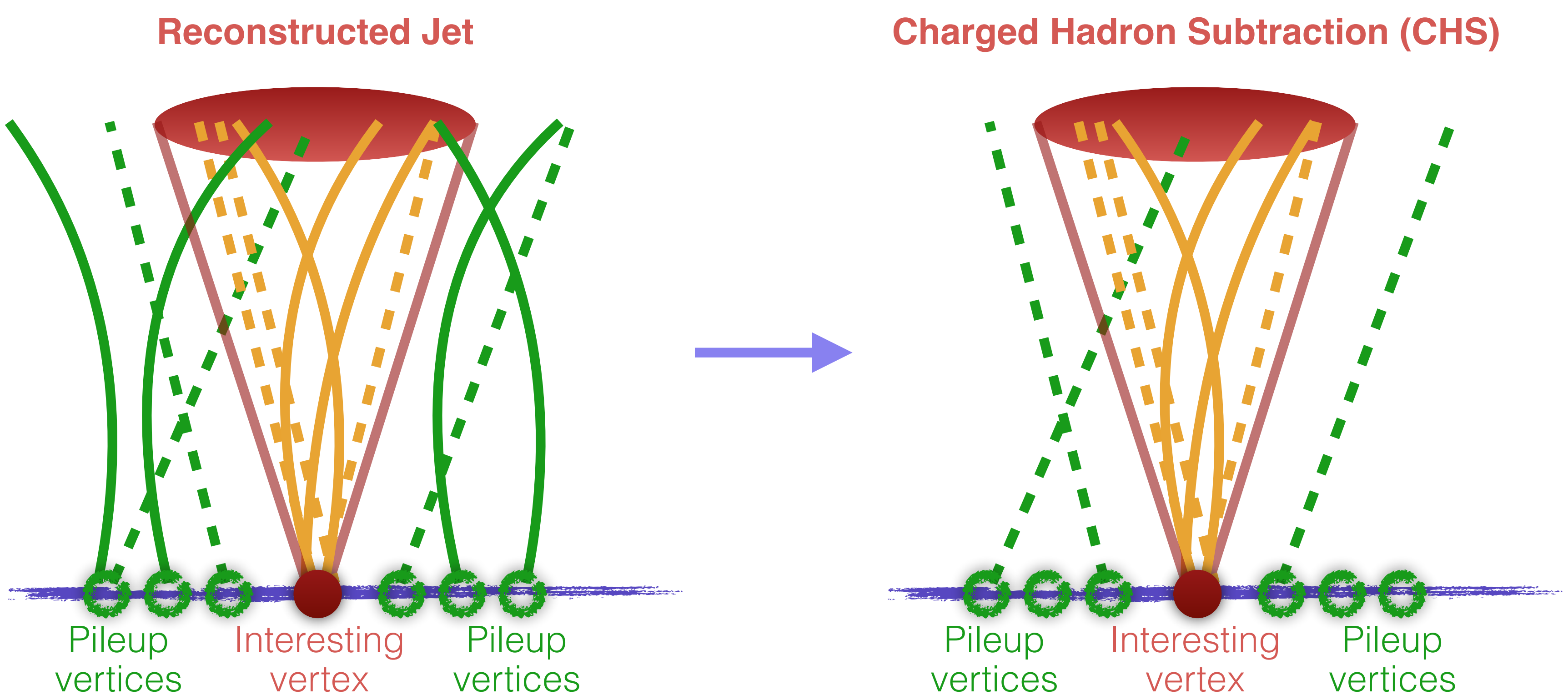
Pileup mitigation techniques



Pileup mitigation techniques

Charged Hadron Subtraction (CHS)

- Tracker information to remove charged particles associated to PU
- Neutral particles energy subtracted
- Applicable for $|\eta| < 2.4$



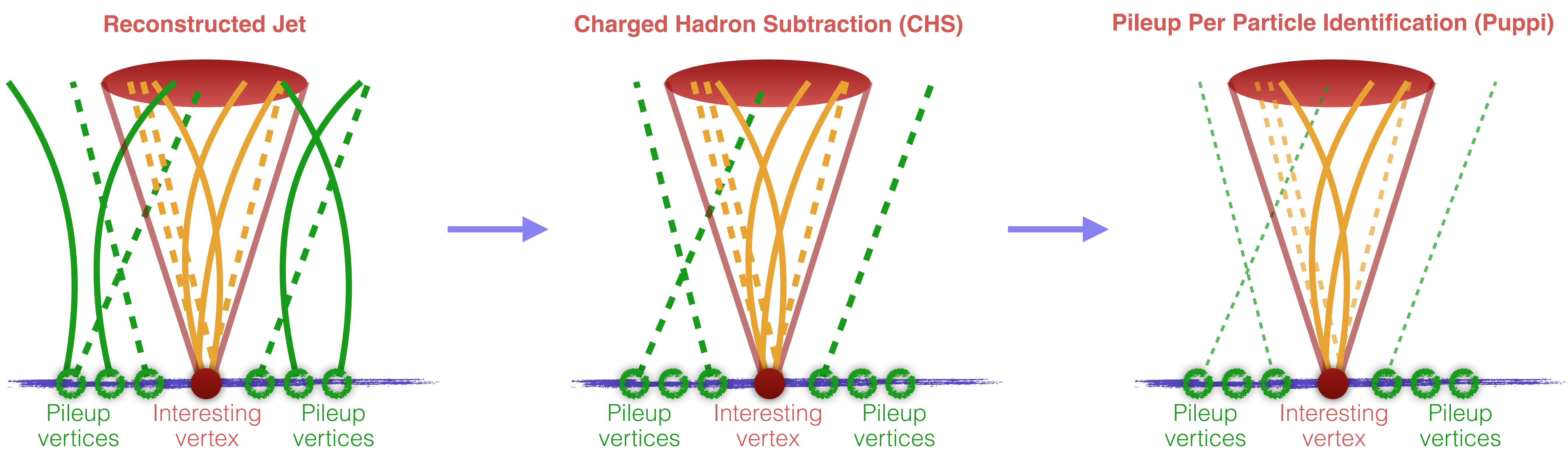
Pileup mitigation techniques

Charged Hadron Subtraction (CHS)

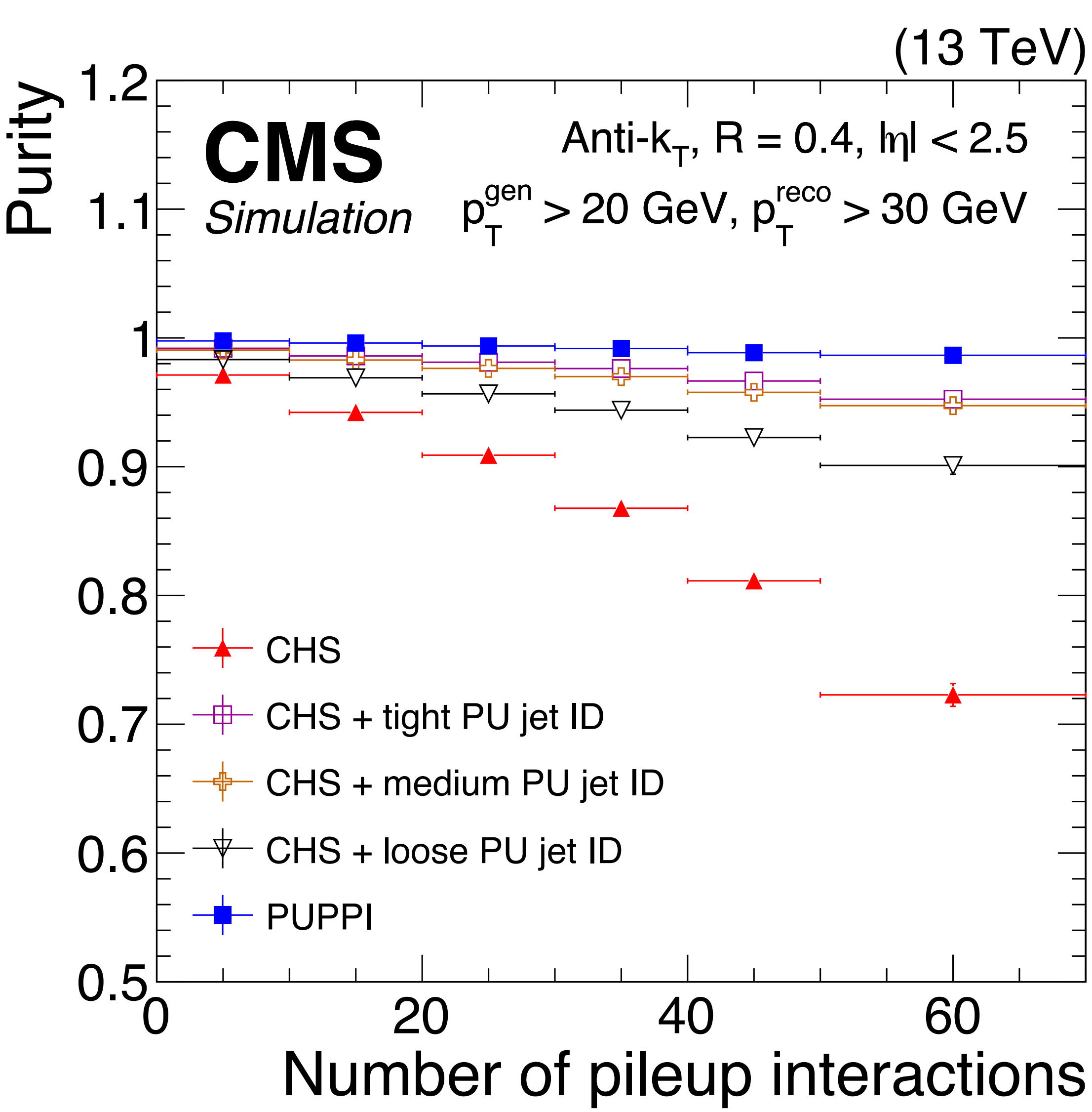
- Tracker information to remove charged particles associated to PU
- Neutral particles energy subtracted
- Applicable for $|\eta| < 2.4$

Pileup Per Particle Identification (Puppi) Puppi in CMS

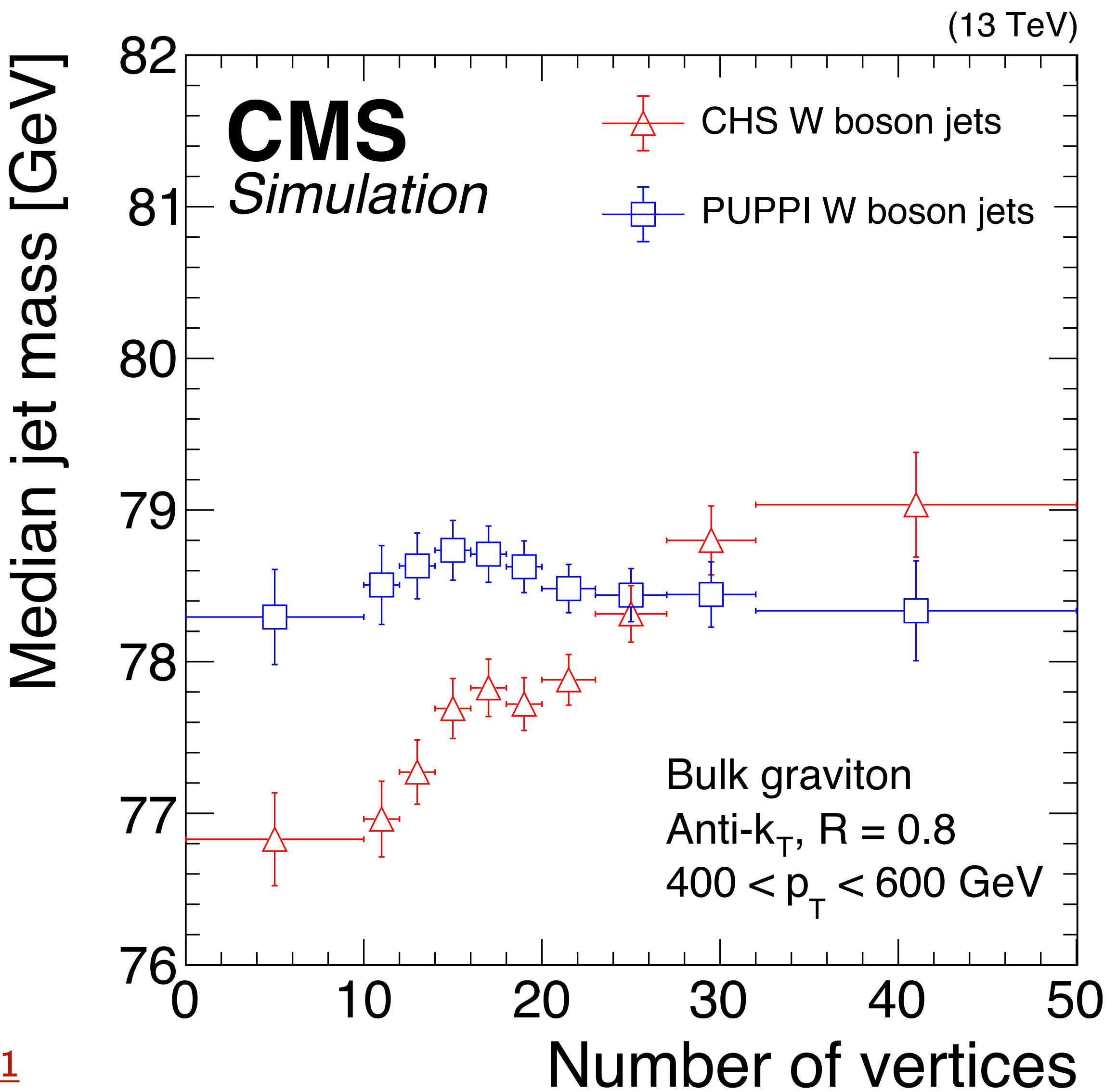
- Per-particle weight
- Scale 4-momentum before clustering
- Charged particles similar to CHS
  - Redefined track-vertex association



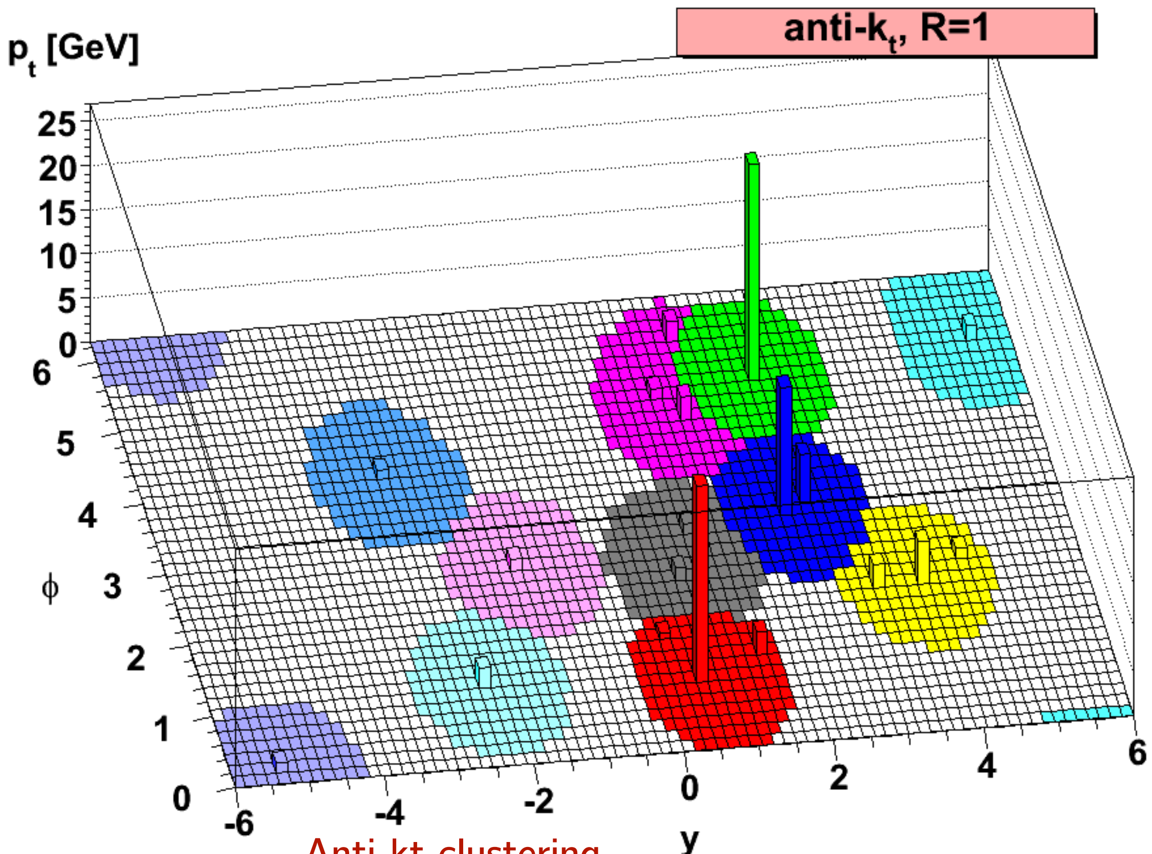
- Widely used in Run 2, default in Run 3
- Improved all jet-related variables
  - Jet efficiency and purity (matched to generator-level jets)
  - Jet substructure
  - New optimation to include hadronic tau reconstruction: CMS-DP-2024-043



CMS-DP-2021-001



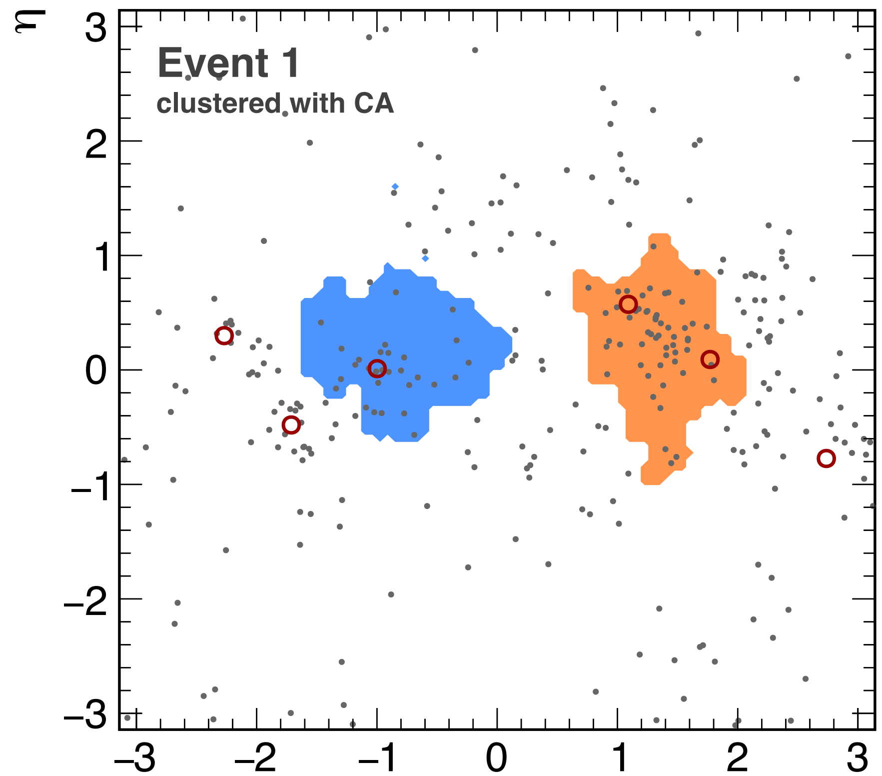
- Anti-k_{T} as default algorithm
  - small radius: R=0.4 (AK4)
  - large radius: R=0.8 (AK8)



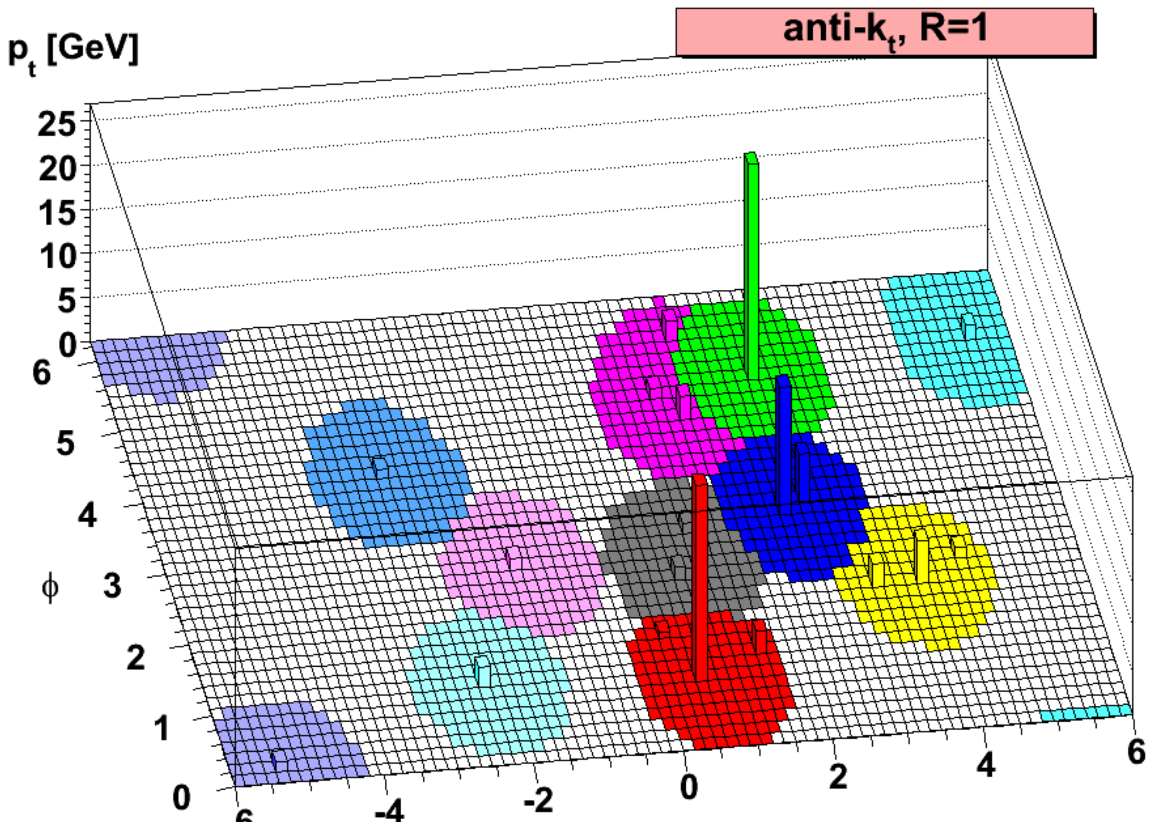
Anti-kt clustering

Jet reconstruction

- Anti-k_{T} as default algorithm
  - small radius: R=0.4 (AK4)
  - large radius: R=0.8 (AK8)
  - alternative algorithms: CA, HOTVR, XCone



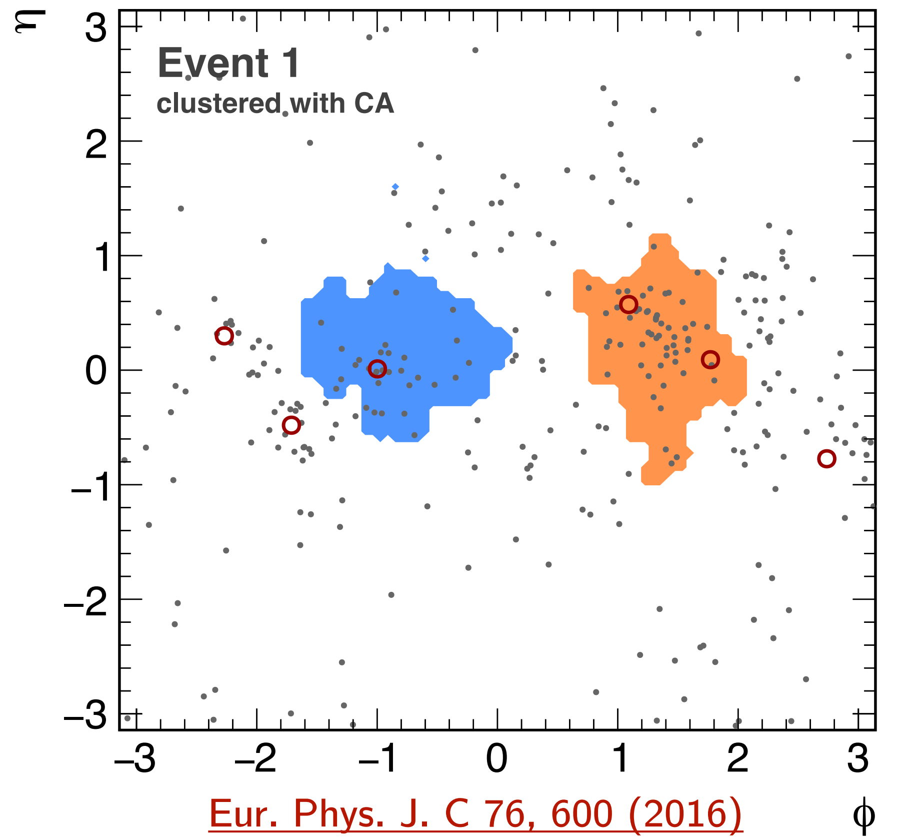
Eur. Phys. J. C 76, 600 (2016)



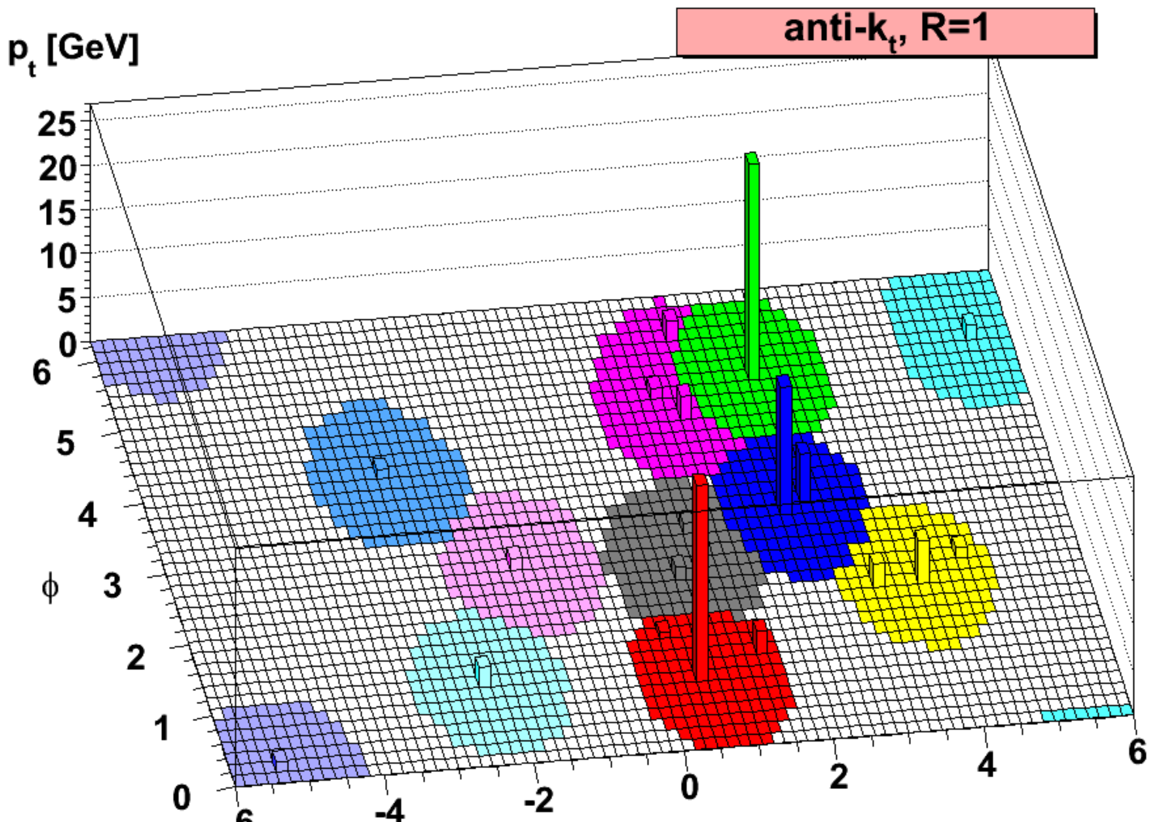
Anti-kt clustering

Jet reconstruction

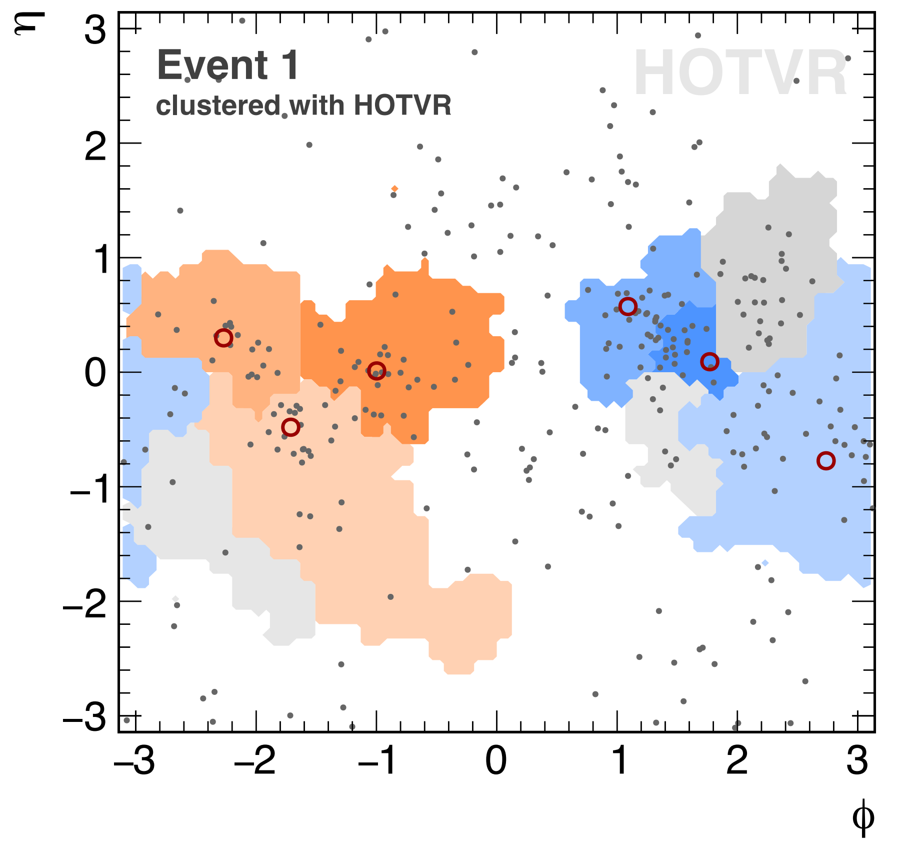
- Anti-k_T as default algorithm
  - small radius: R=0.4 (AK4)
  - large radius: R=0.8 (AK8)
  - alternative algorithms: CA, HOTVR, XCone



Eur. Phys. J. C 76, 600 (2016)

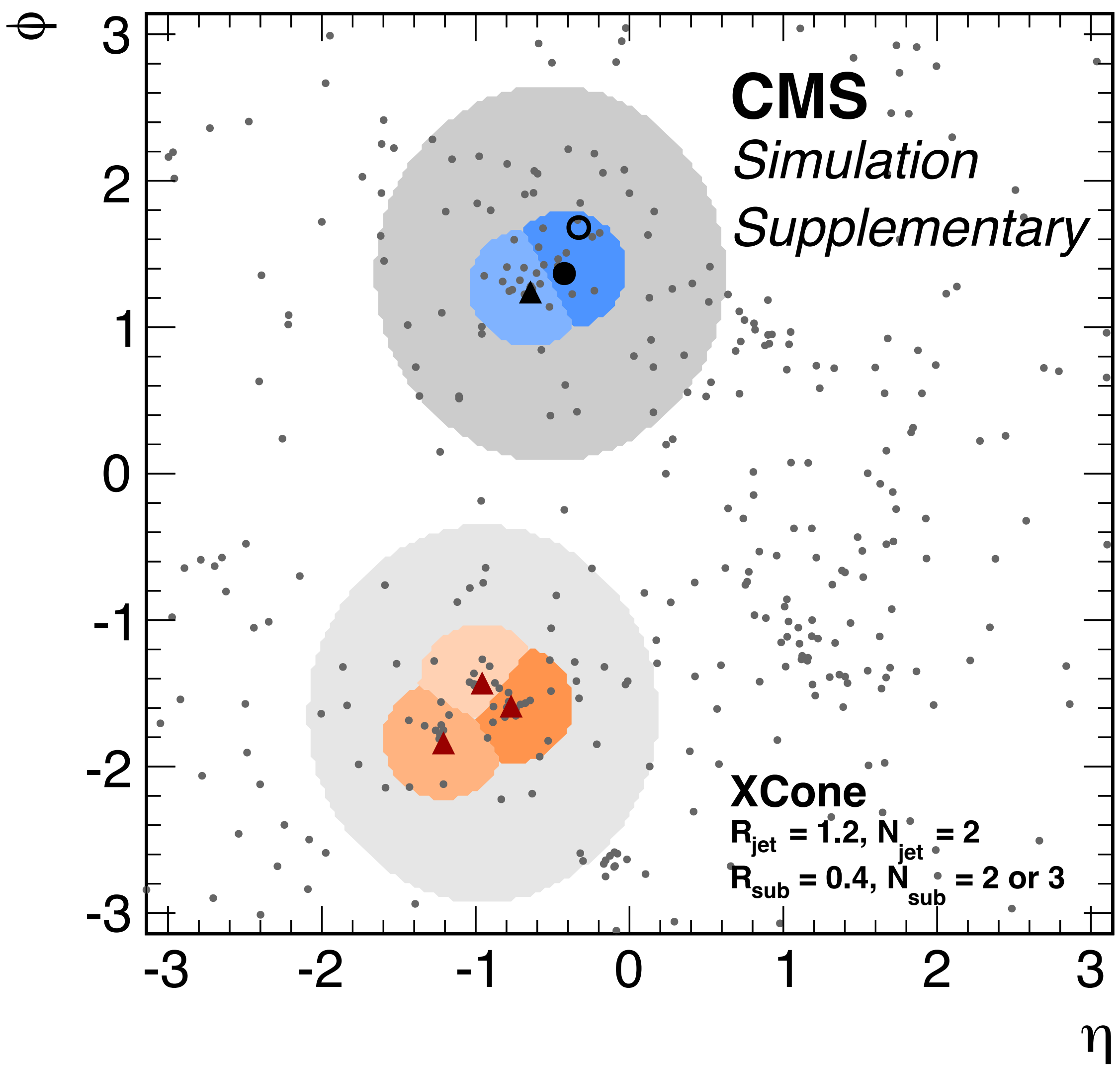


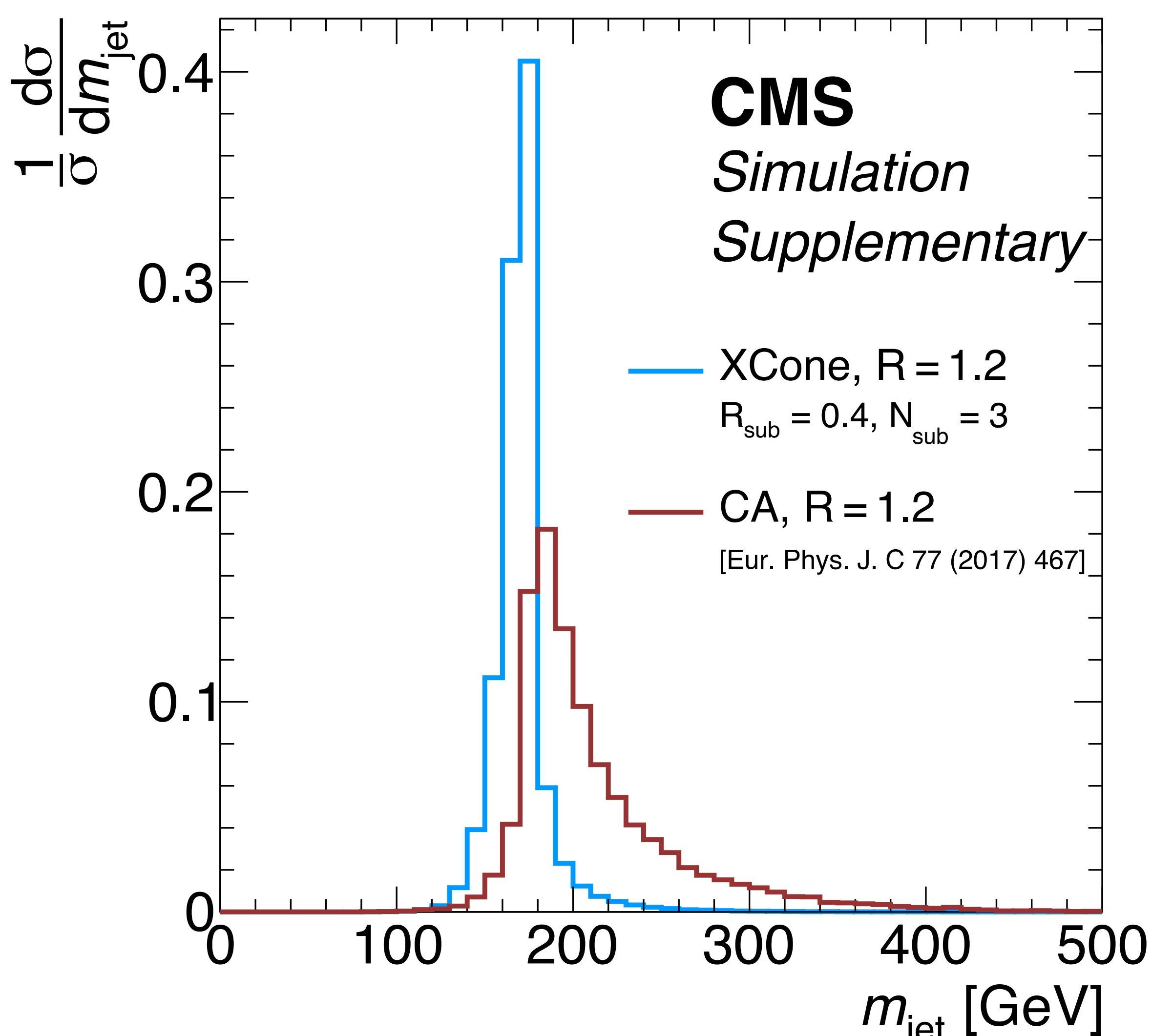
Anti-kt clustering



Jet reconstruction – XCone

- Event signature defines clustering
- Return exactly N jets
- Examples from top-mass measurement
- Large improvement for the jet mass resolution

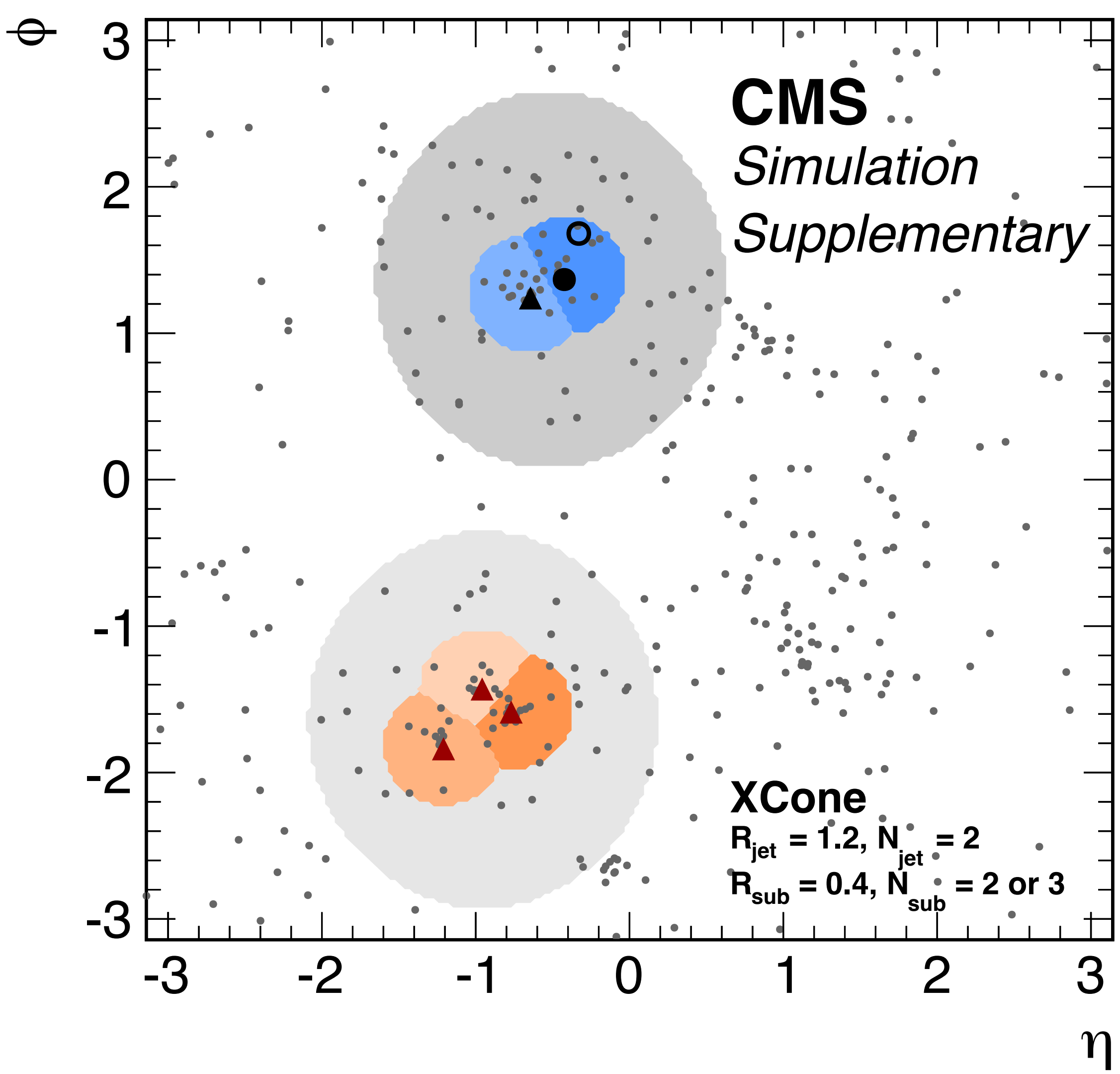




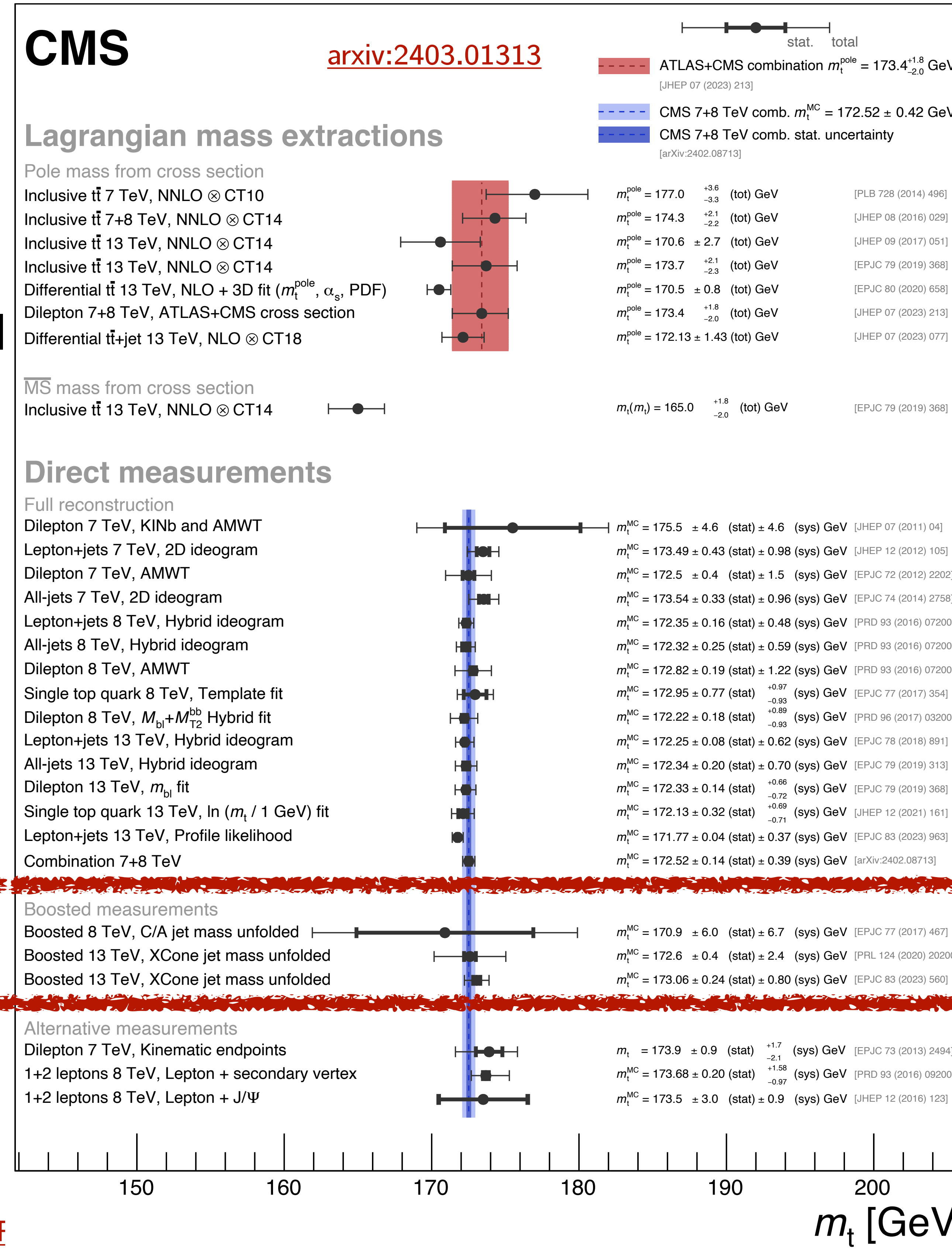
Phys. Rev. Lett. 124 (2020) 202001

Jet reconstruction – XCone

- Event signature defines clustering
- Return exactly N jets
- Examples from top-mass measurement
- Large improvement for the jet mass resol



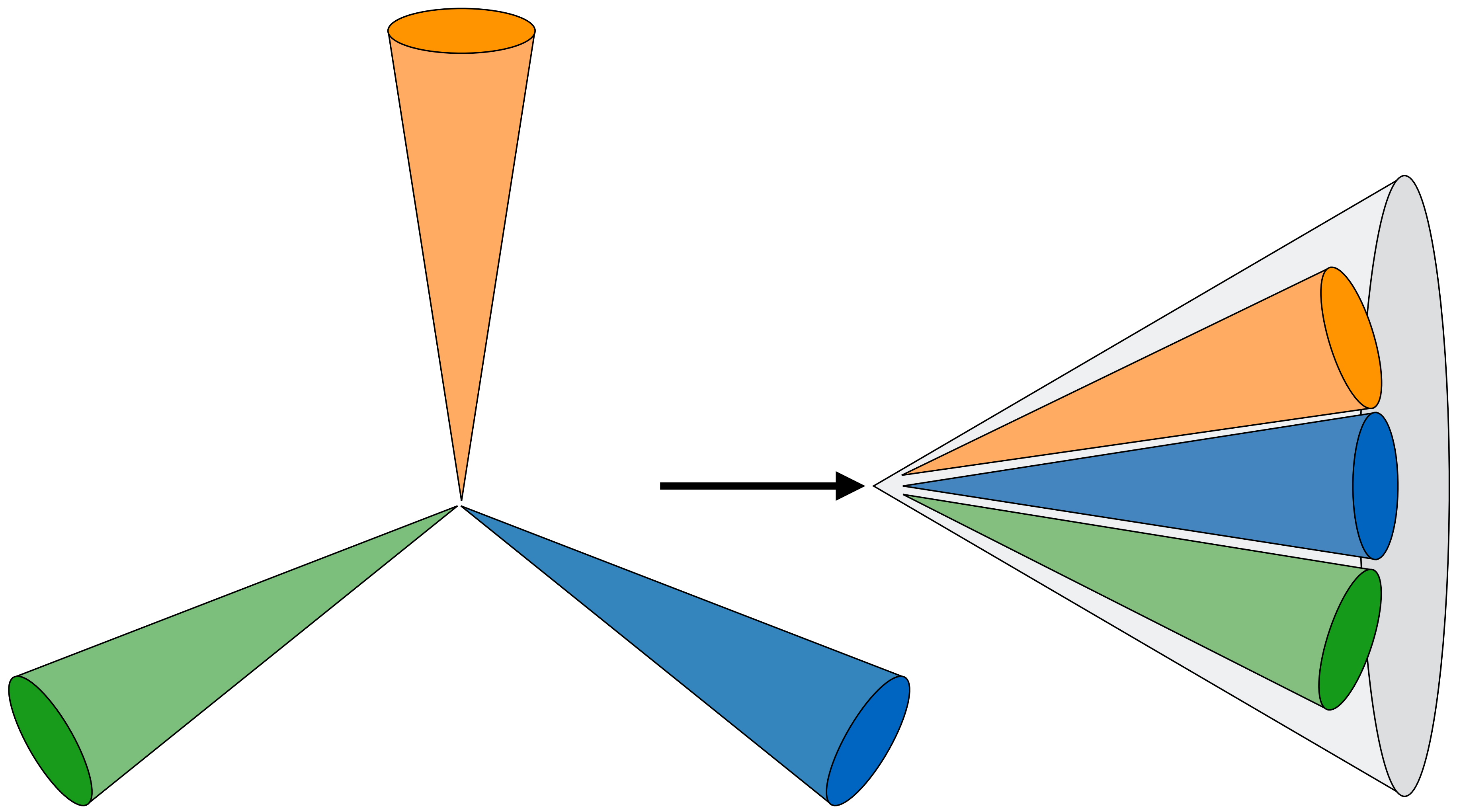
Phys. F



Jet identification ("tagging")

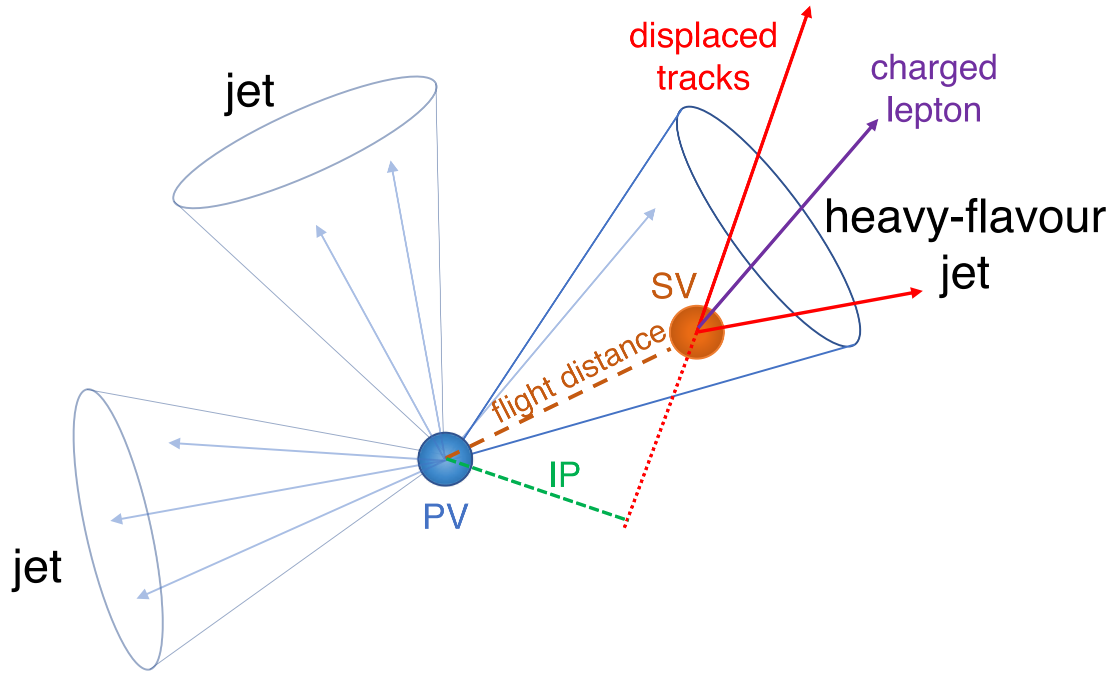
Tagging

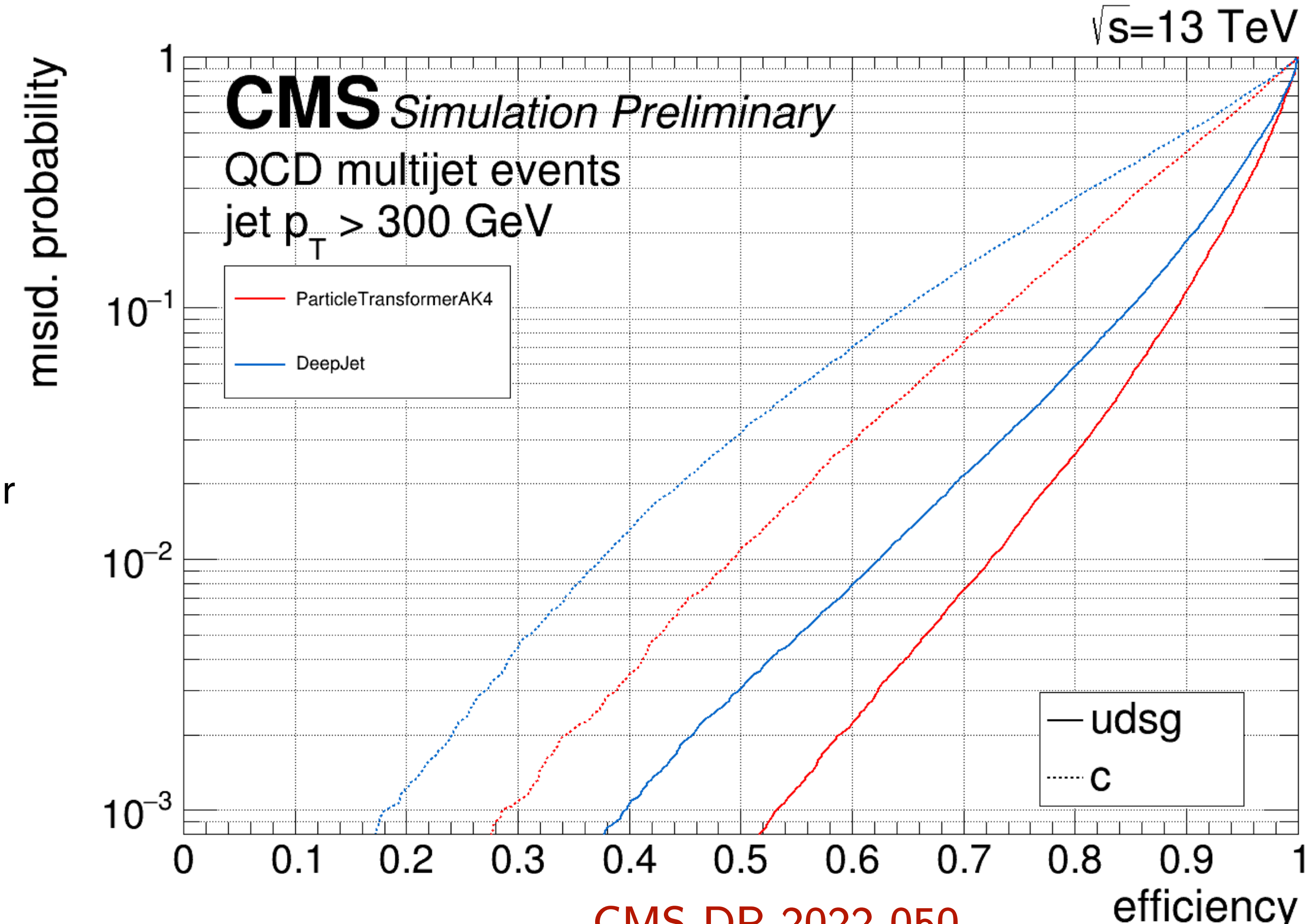
- Type of elementary particle that initiated the jet
  - Boosted topology -> Collimated decay products reconstructed as multi-prong objects



Tagging

- Type of elementary particle that initiated the jet
  - Boosted topology -> Collimated decay products reconstructed as multi-prong objects
- Jet flavor (b vs light, b vs c, ...)

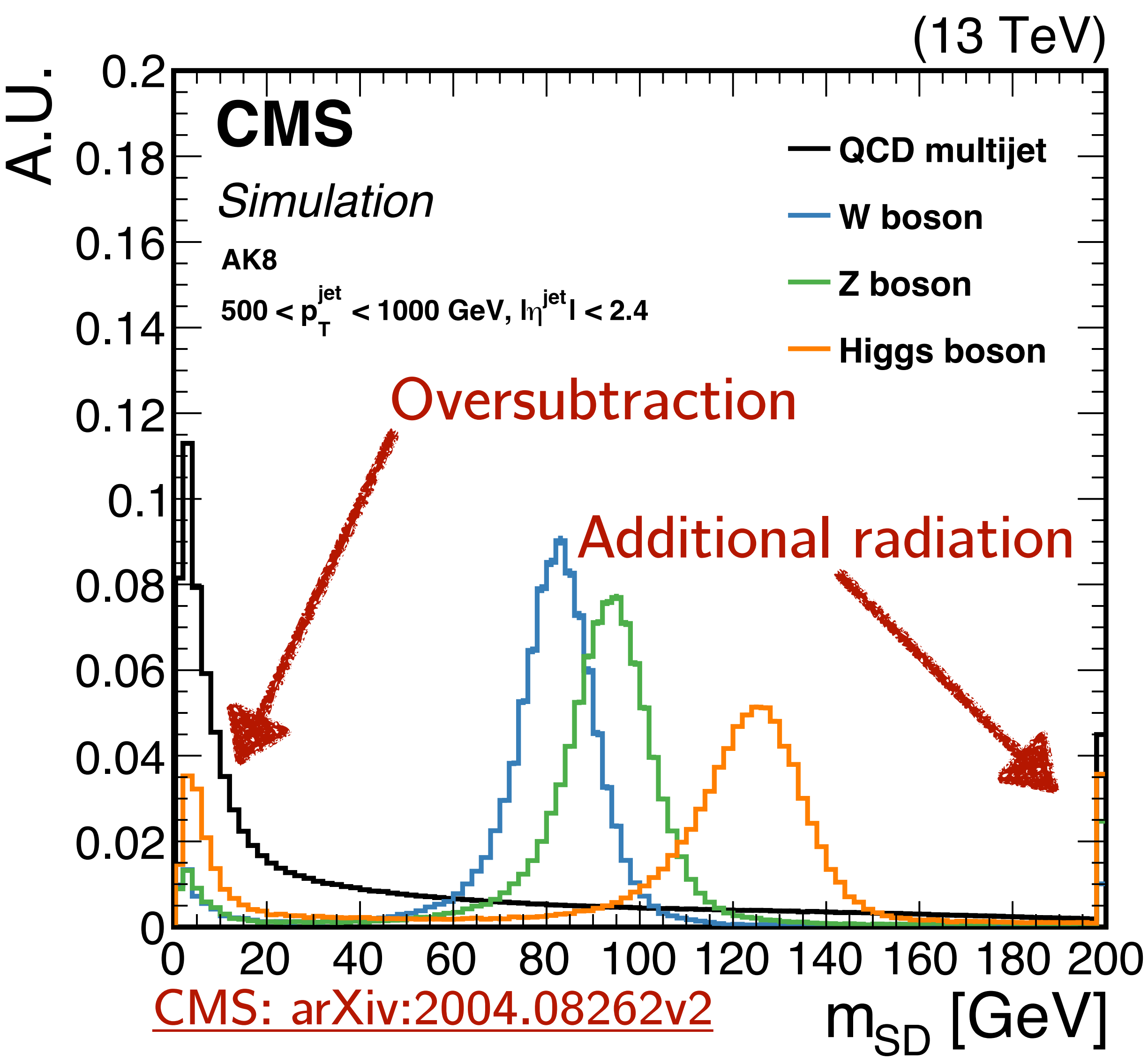




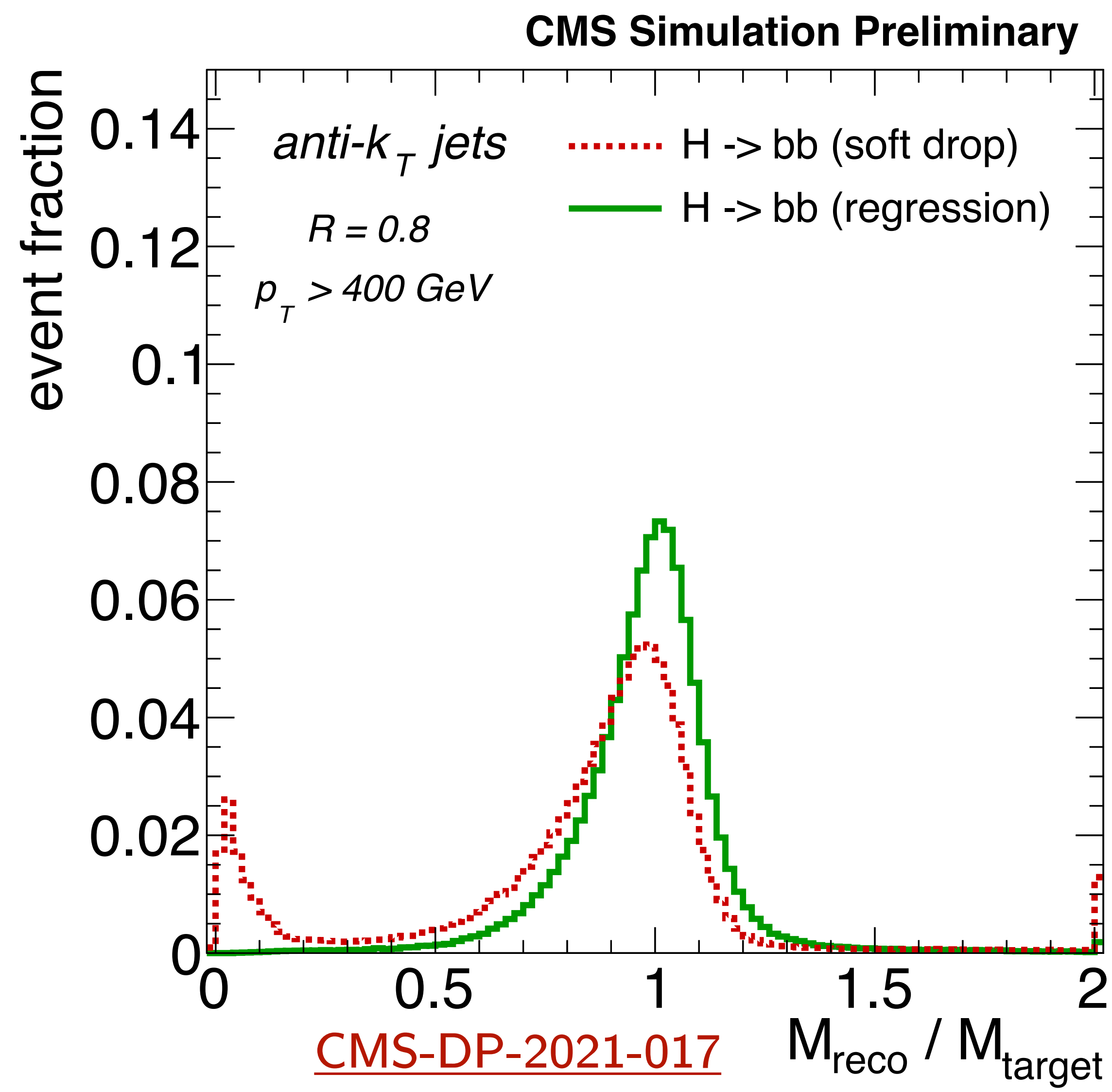
CMS-DP-2022-050

Tagging

- Type of elementary particle that initiated the jet
  - Boosted topology -> Collimated decay products reconstructed as multi-prong objects
- Jet flavor (b vs light, b vs c, ...)
- Jet mass (ML with regression)



CMS: arXiv:2004.08262v2

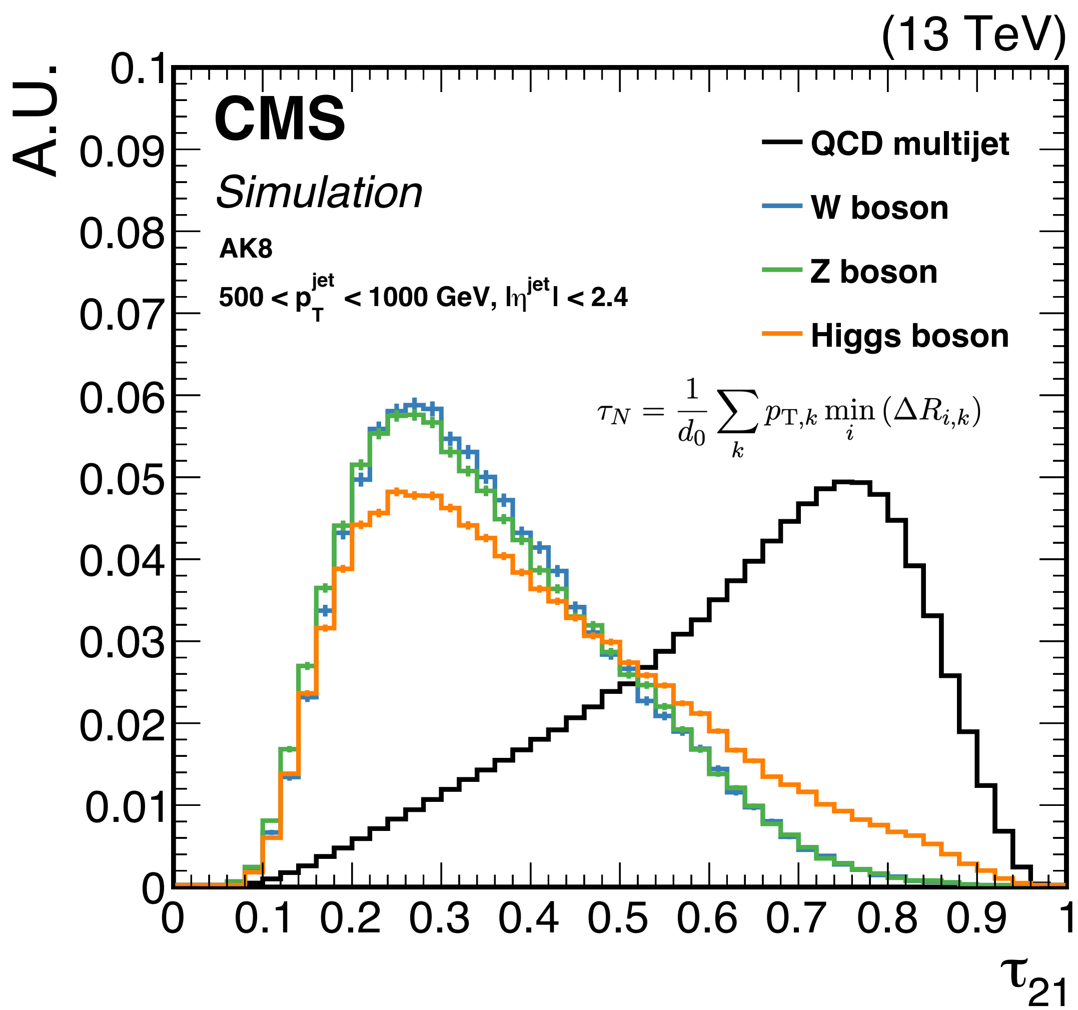


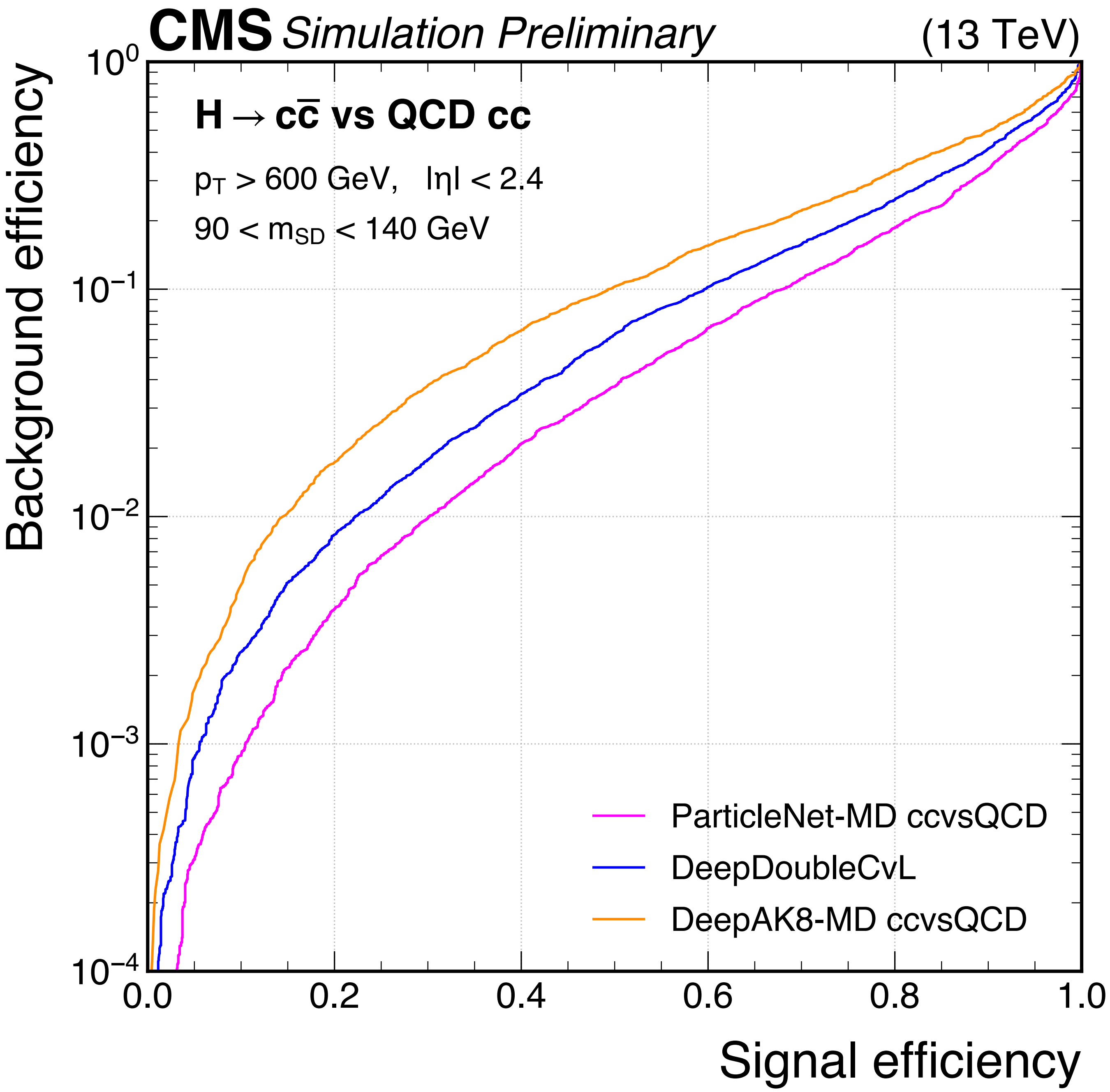
CMS-DP-2021-017

Jet identification (“tagging”)

Tagging

- Type of elementary particle that initiated the jet
  - Boosted topology -> Collimated decay products reconstructed as multi-prong objects
- Jet flavor (b vs light, b vs c, ...)
- Jet mass (ML with regression)
- Jet substructure



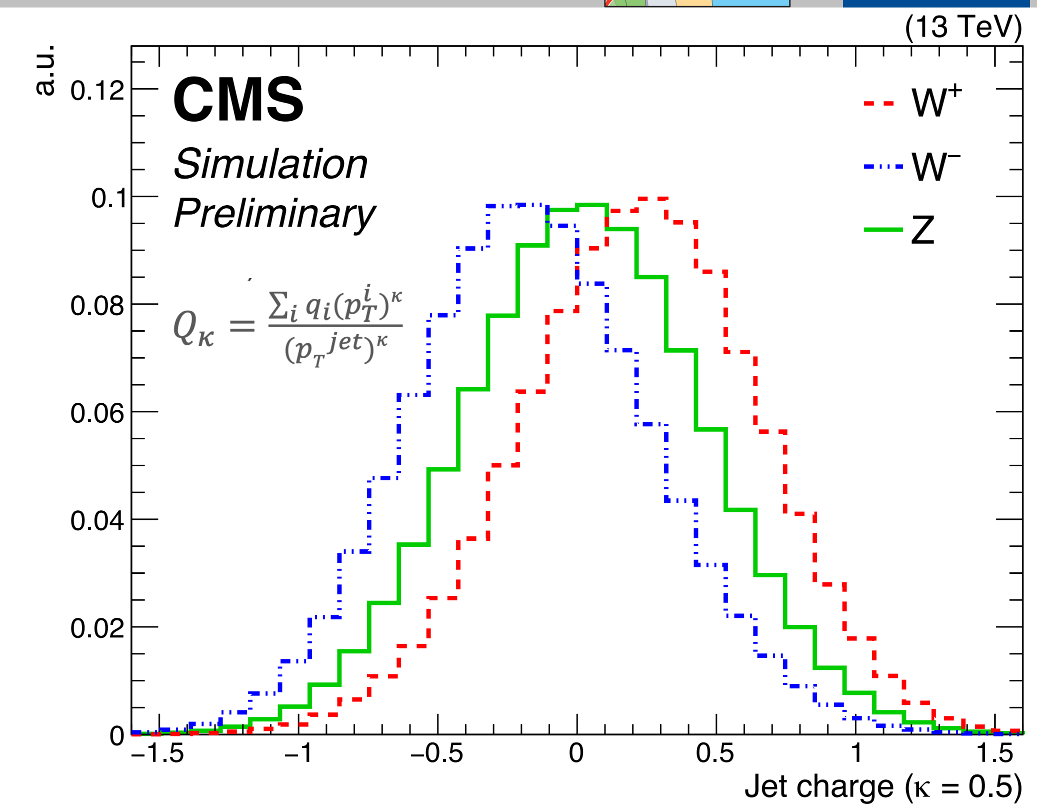


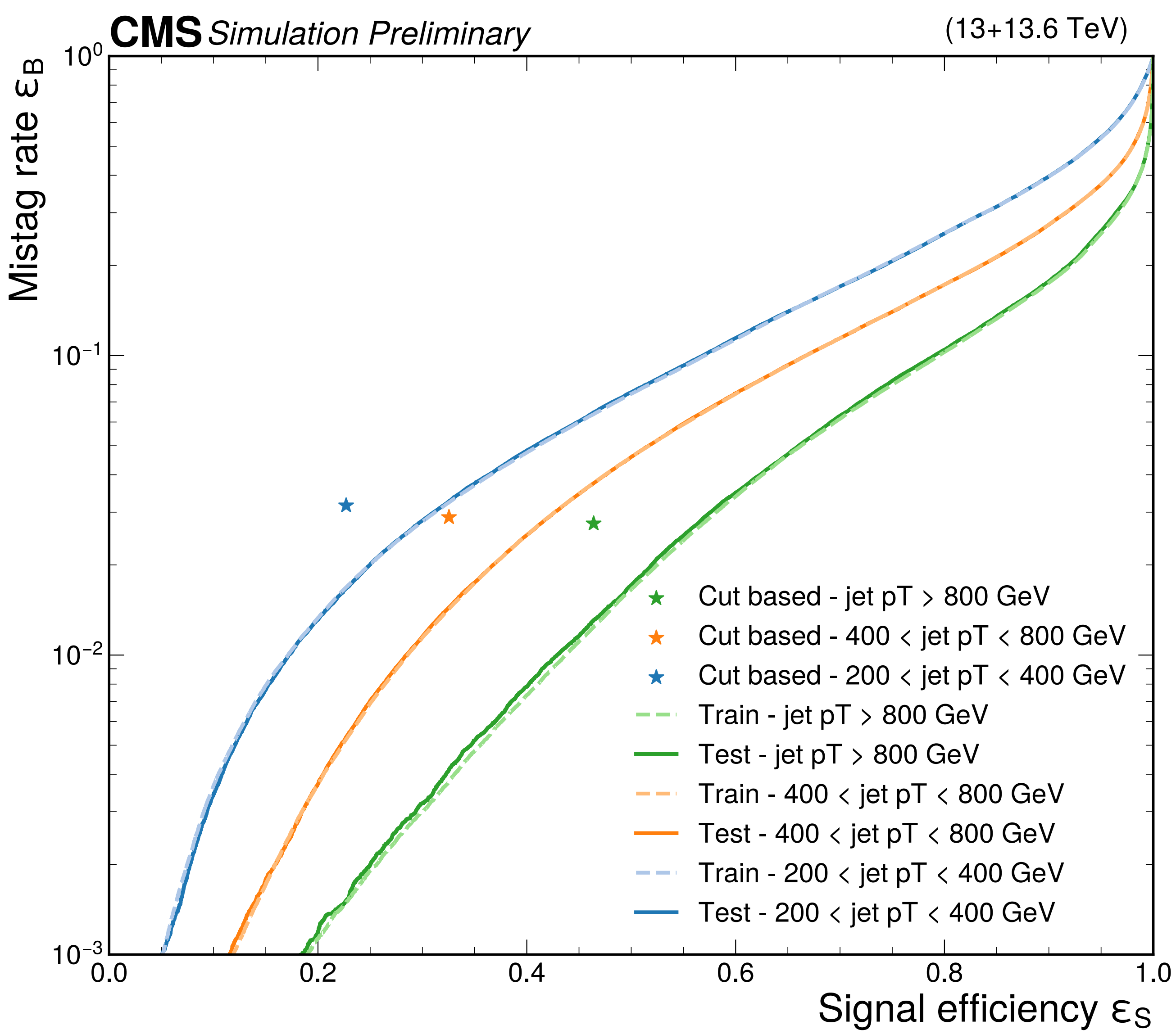
Jet identification ("tagging")

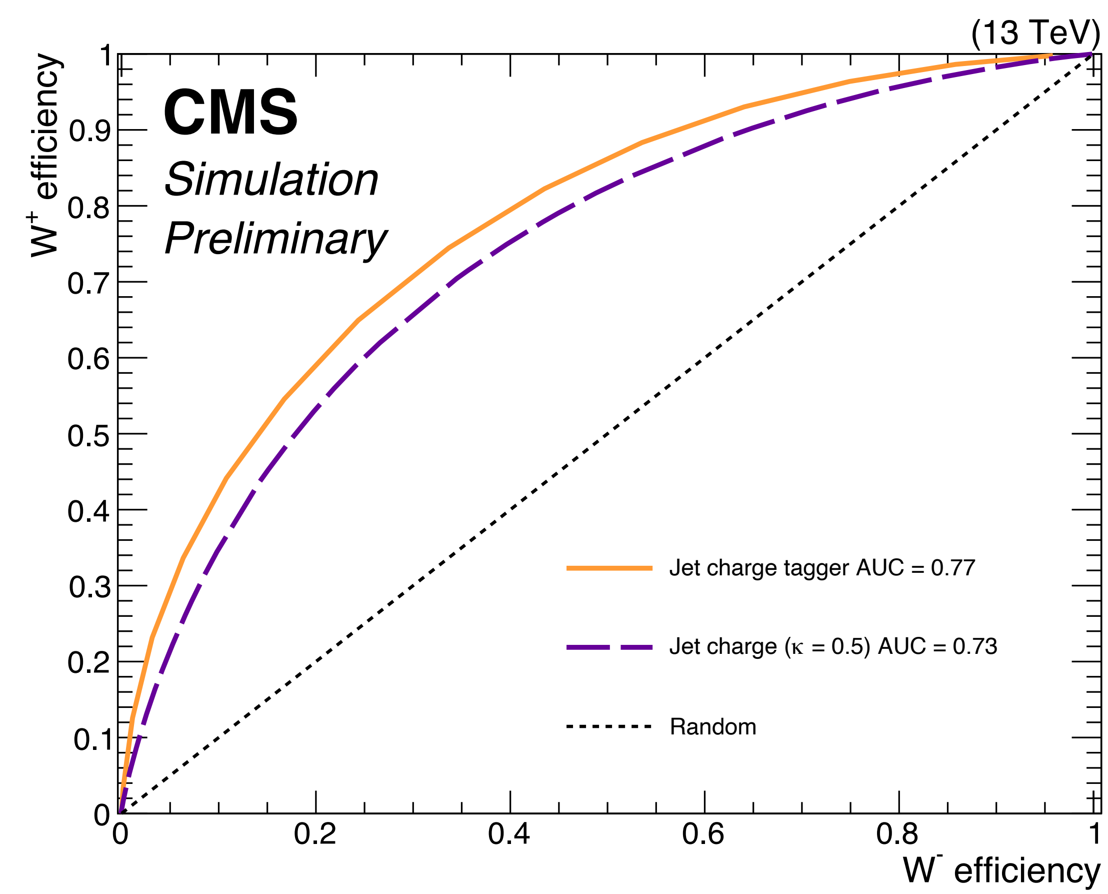
CMS-DP-2024-038
CMS-DP-2024-044

Tagging

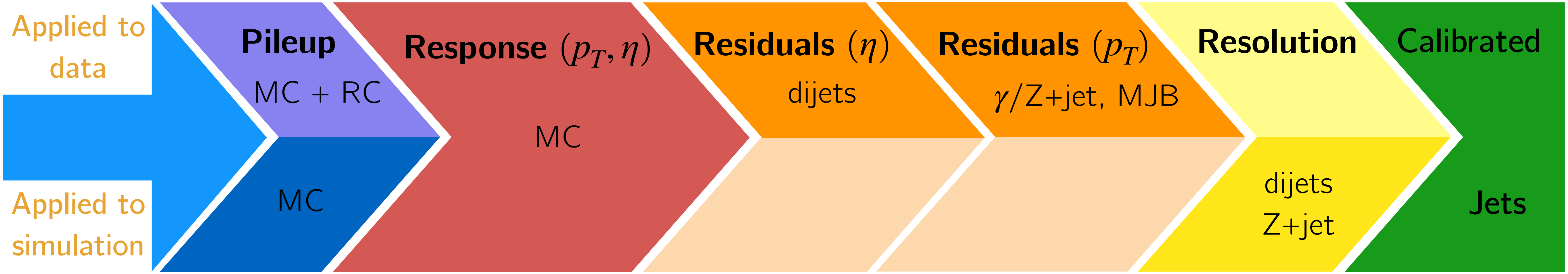
- New ML developments with:
  - HOTVR + BDT
  - Vector boson charge tagger

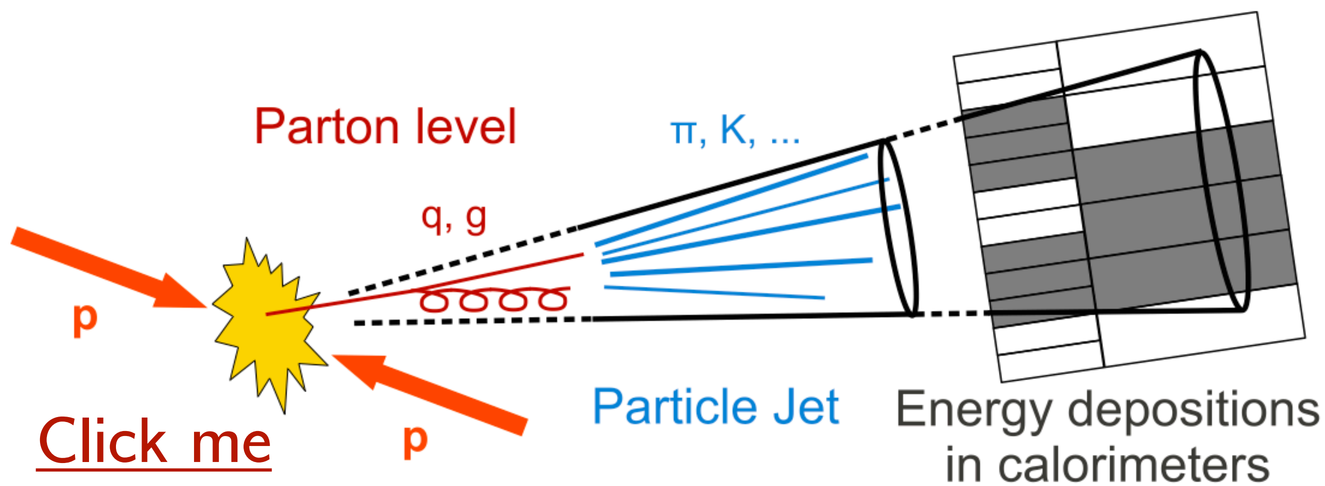




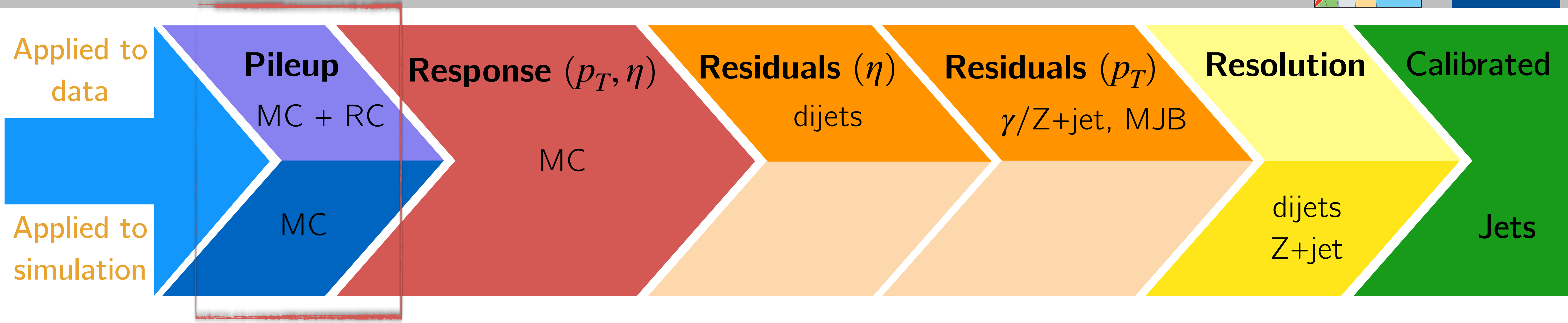


Jet calibration





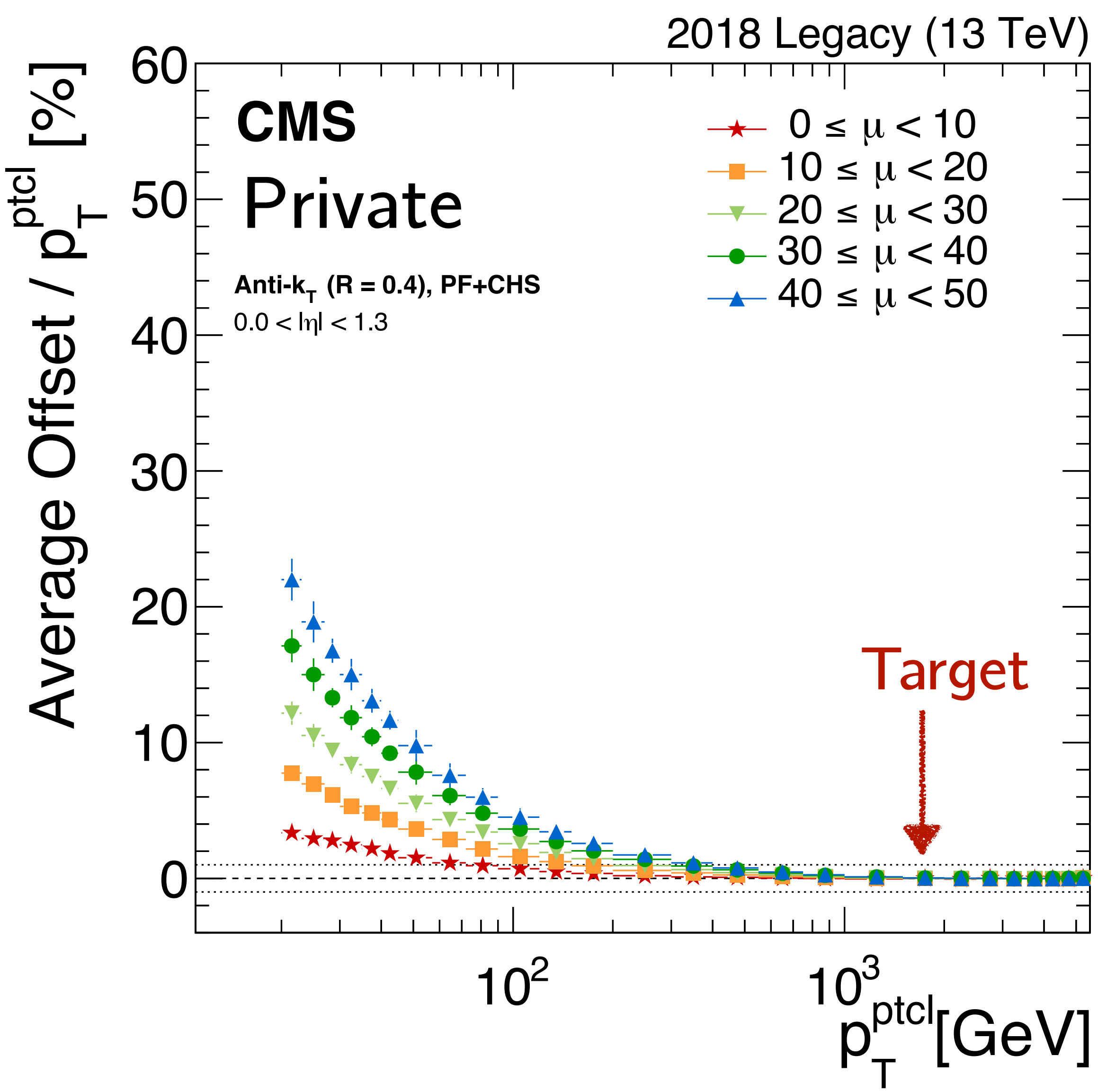
Jet calibration



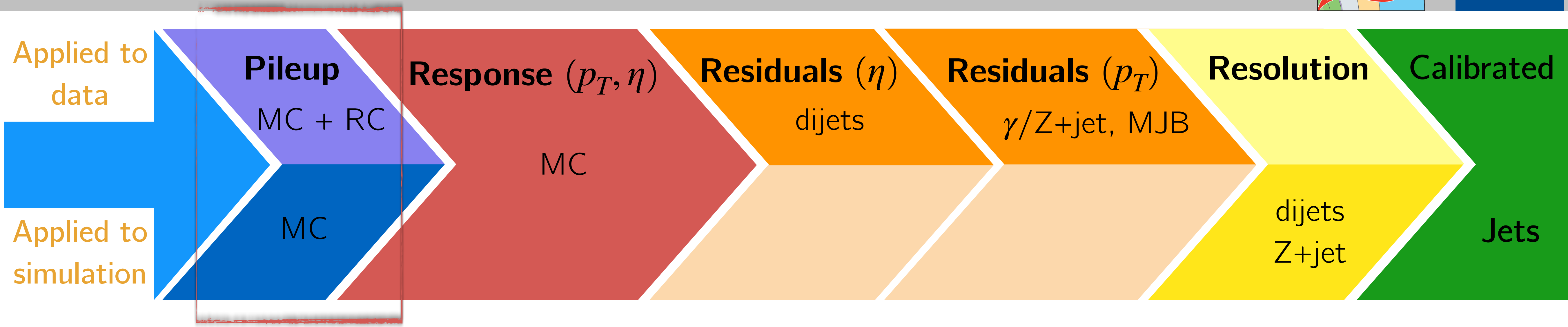
MC truth correction:

PU subtraction

- Remove average offset due to PU
- Approx. 0.5 GeV extra energy per jet
- Neutral component not removed by CHS



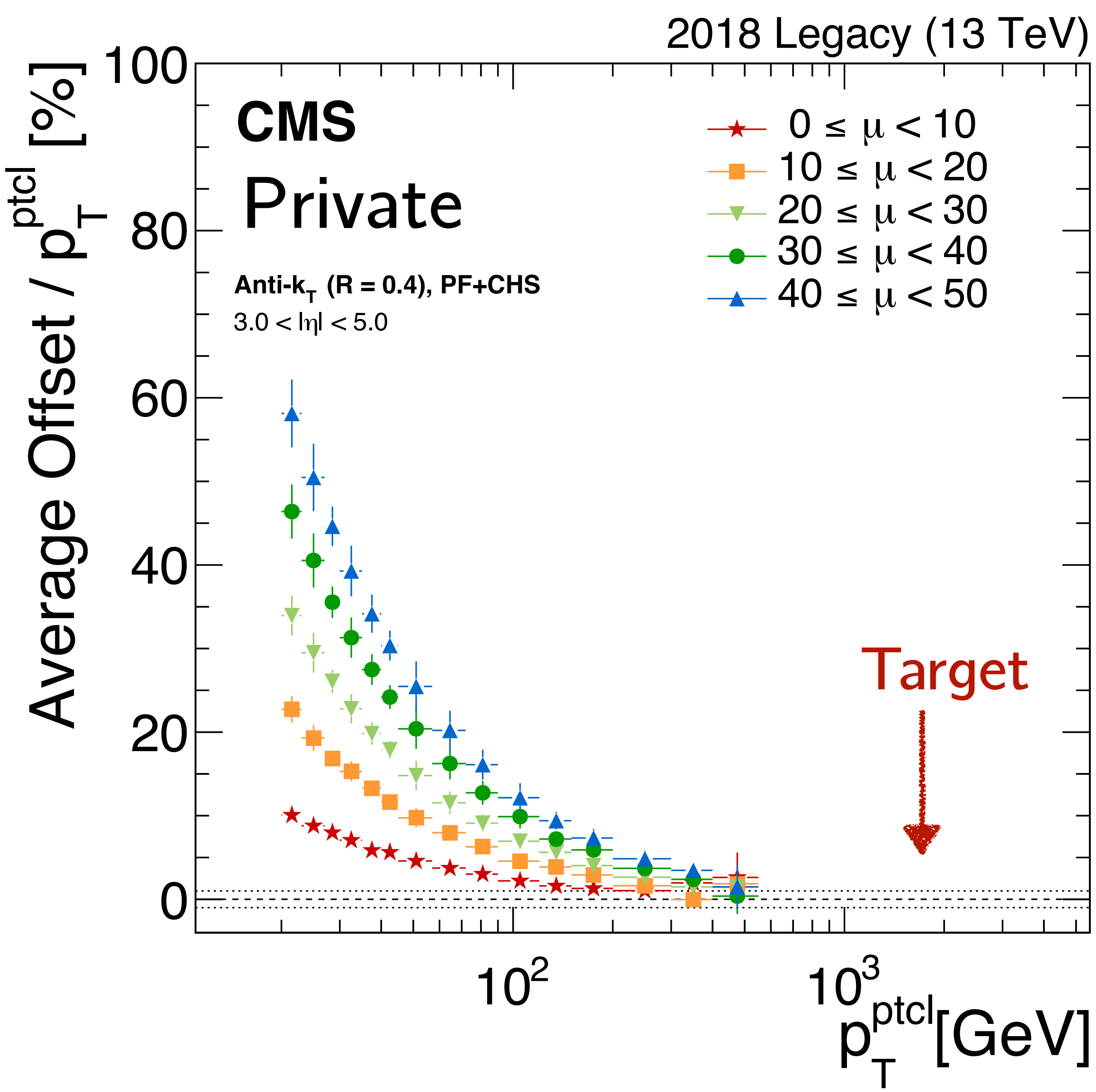
JEC DPNotes Legacy



MC truth correction:

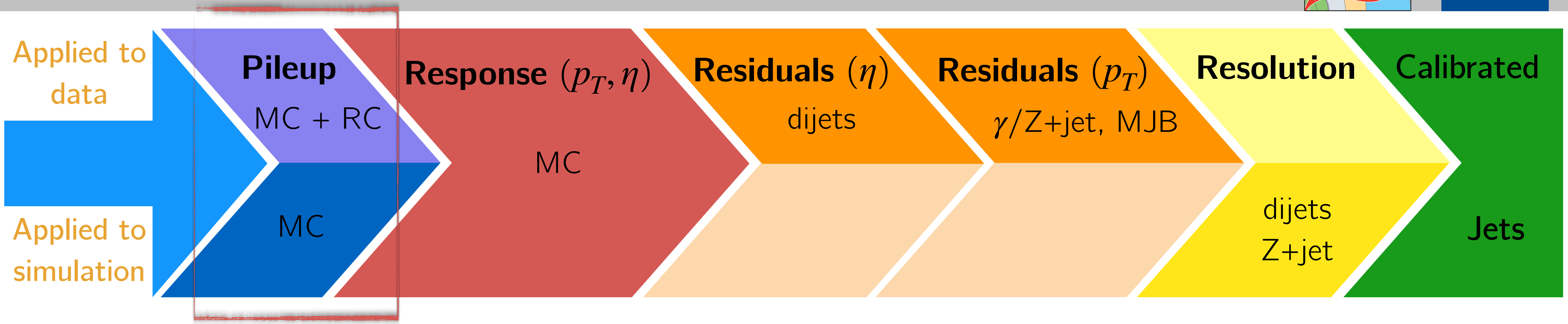
PU subtraction

- Remove average offset due to PU
- Approx. 0.5 GeV extra energy per jet
- Neutral component not removed by CHS
- Significant outside tracker acceptance



JEC DPNotes Legacy

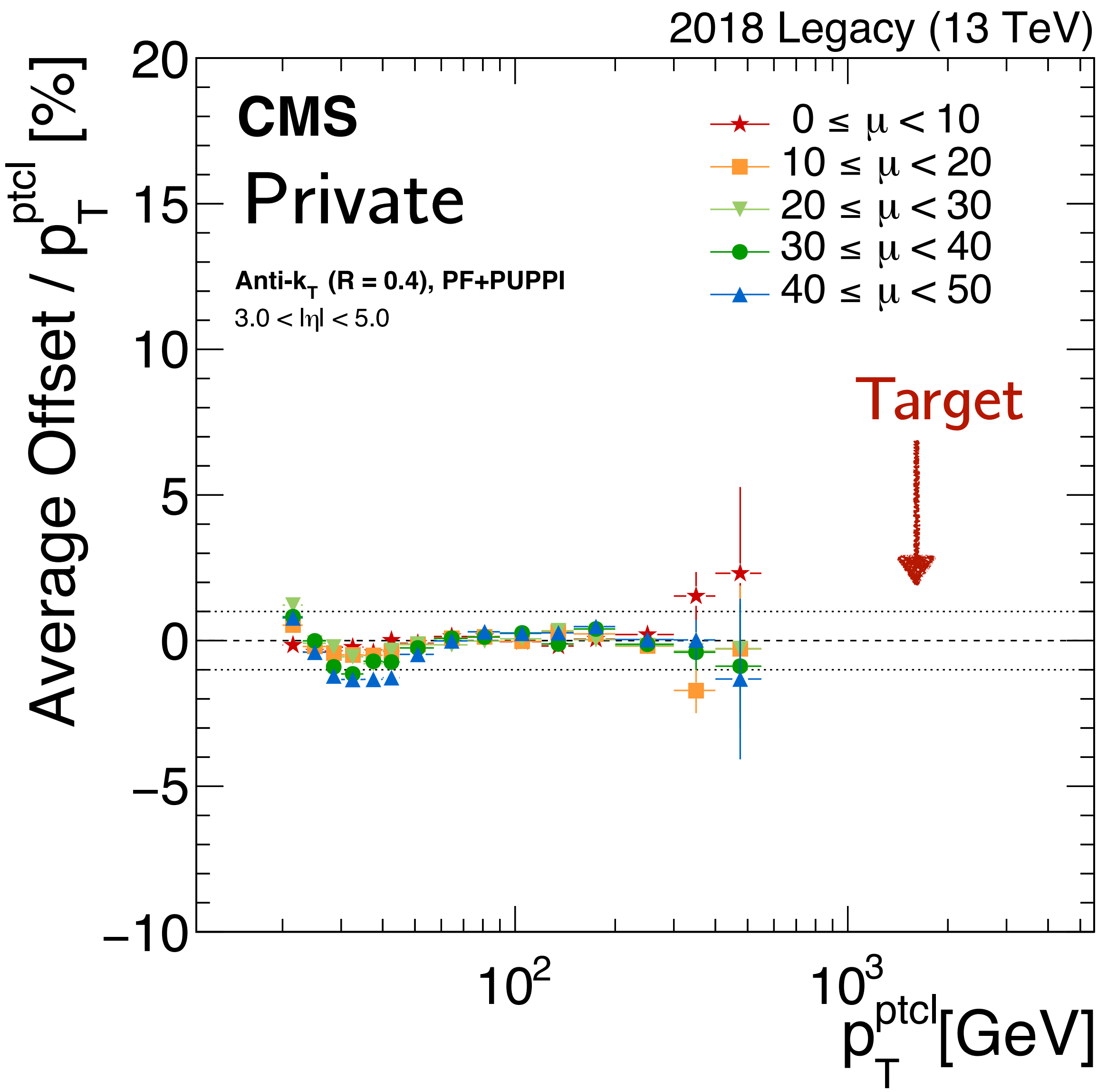
Jet calibration



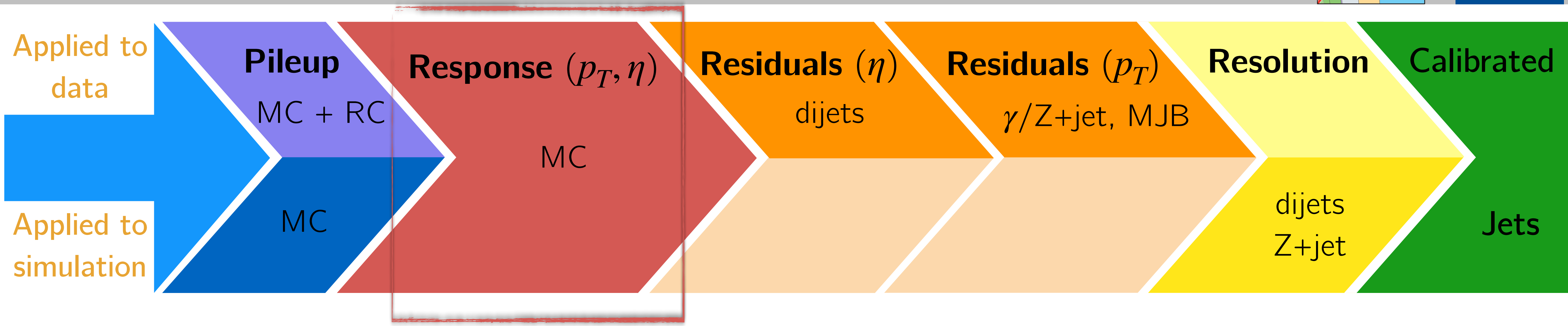
MC truth correction:

PU subtraction

- Remove average offset due to PU
- Approx. 0.5 GeV extra energy per jet
- Neutral component not removed by CHS
- Significant outside tracker acceptance
- Not needed for Puppi



JEC DPNotes Legacy

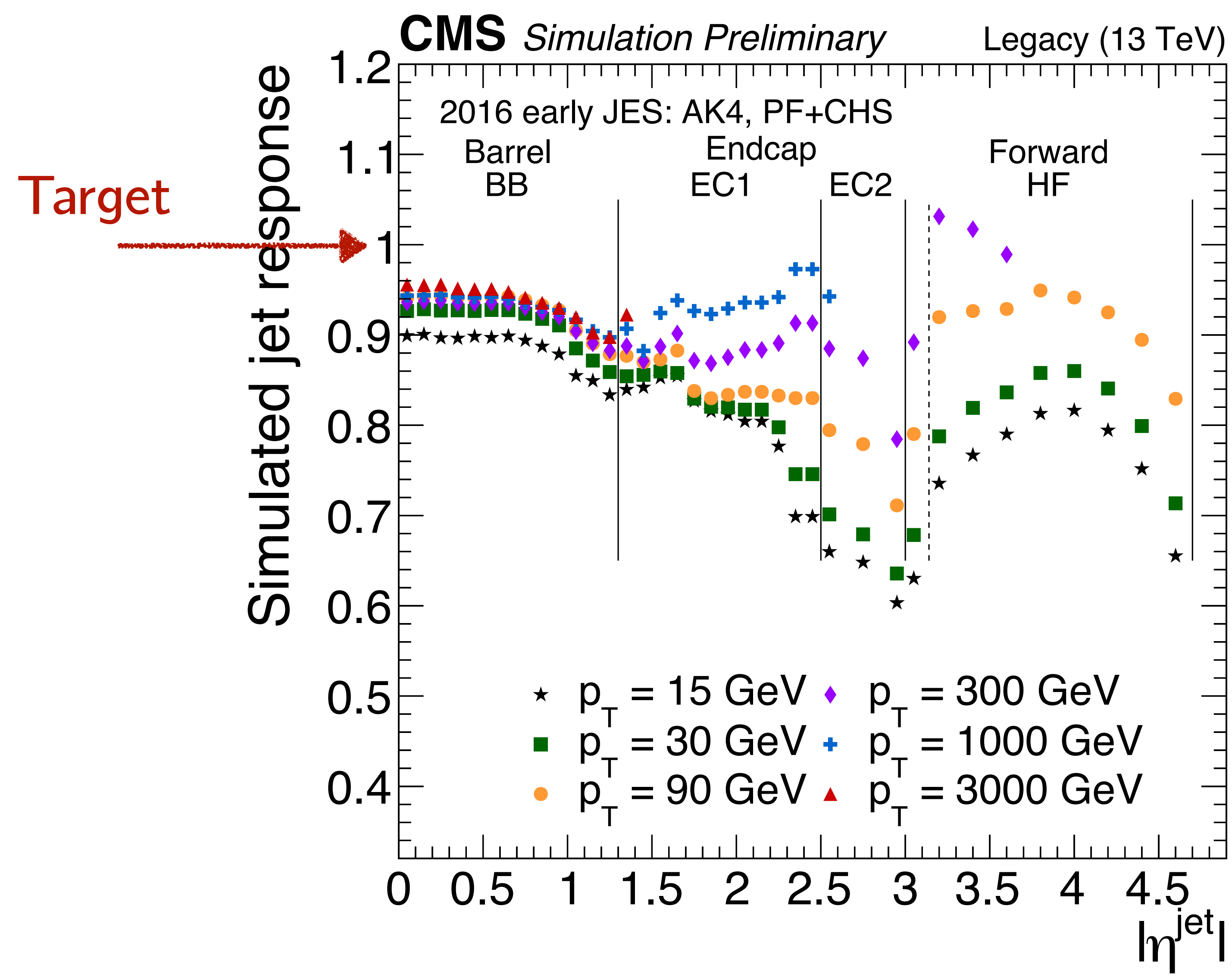


MC truth correction:

PU subtraction

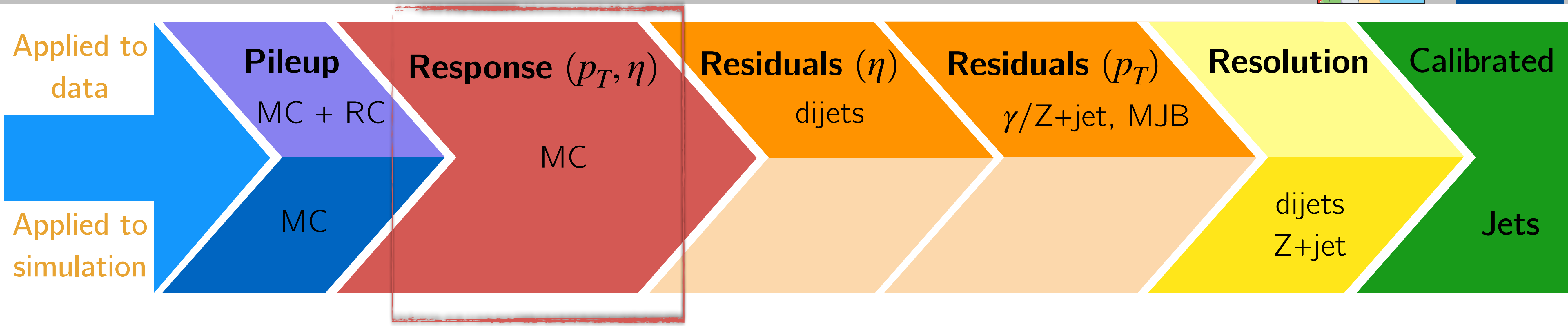
Jet response calibration

- Core of the calibration
- Simulation-based
- Accounts for detector effects
- Change in performance due to detector acceptance



JEC DPNotes Legacy

Jet calibration



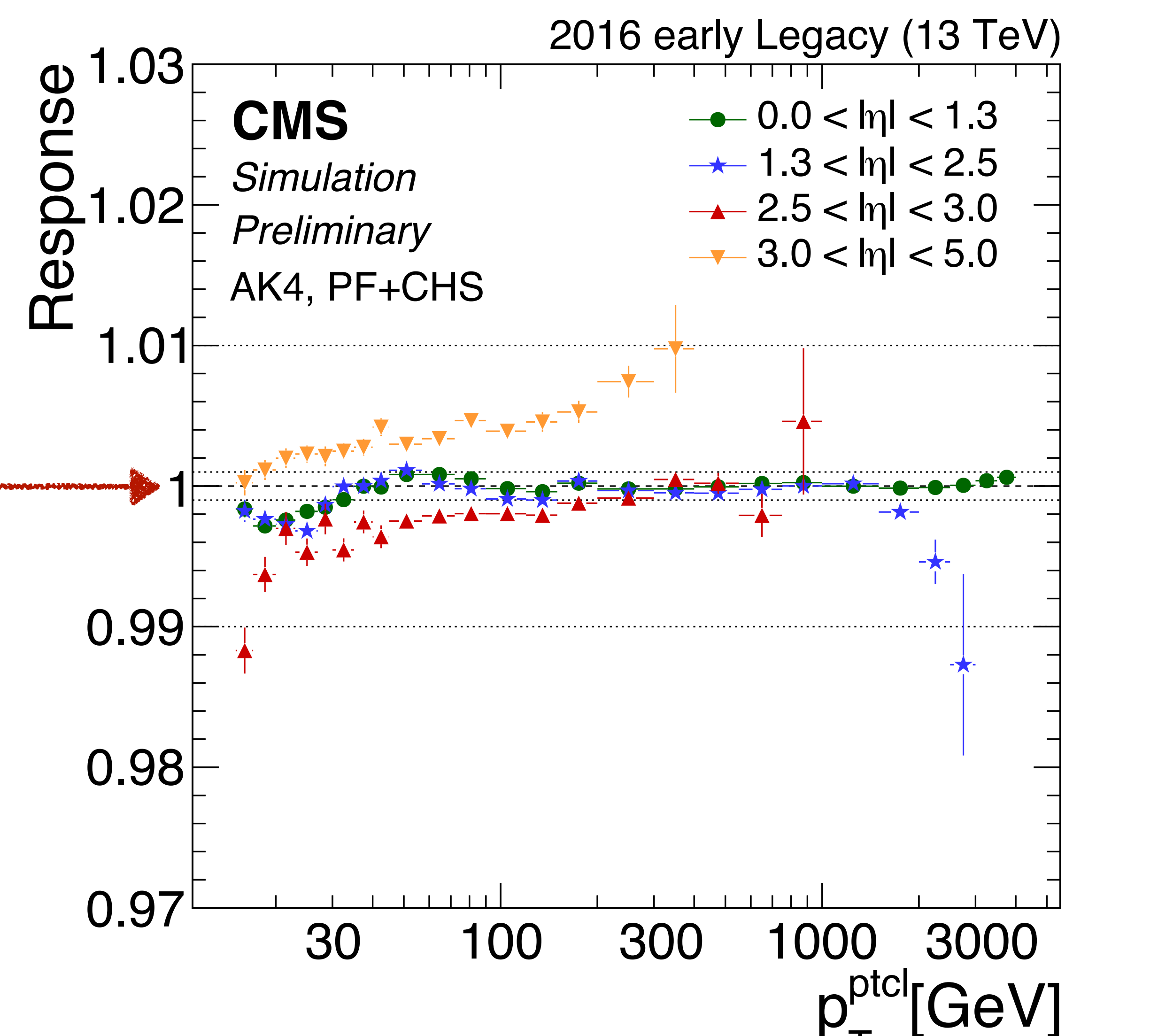
MC truth correction:

PU subtraction

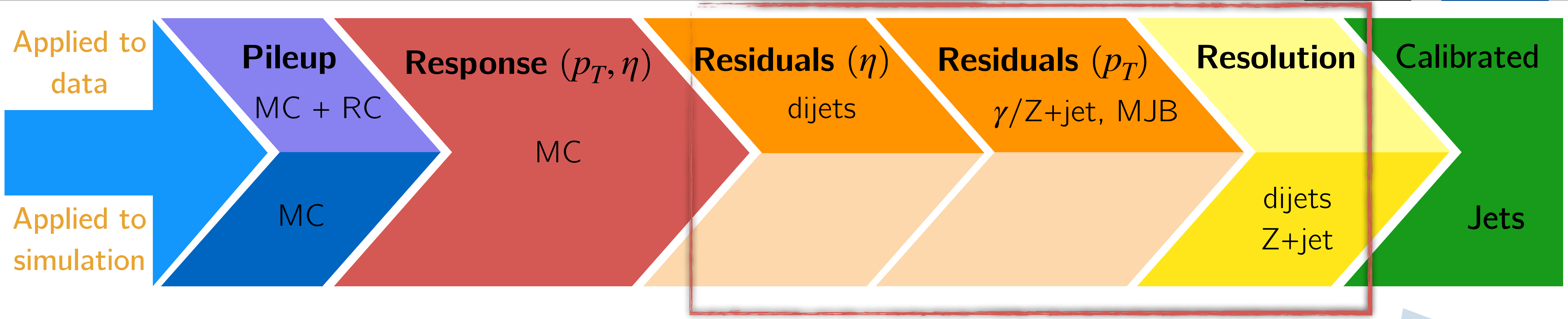
Jet response calibration

- Core of the calibration
- Simulation-based
- Accounts for detector effects
- Change in performance due to detector acceptance
- Closure better that 1% everywhere (0.1% in central region)

Target



JEC DPNotes Legacy



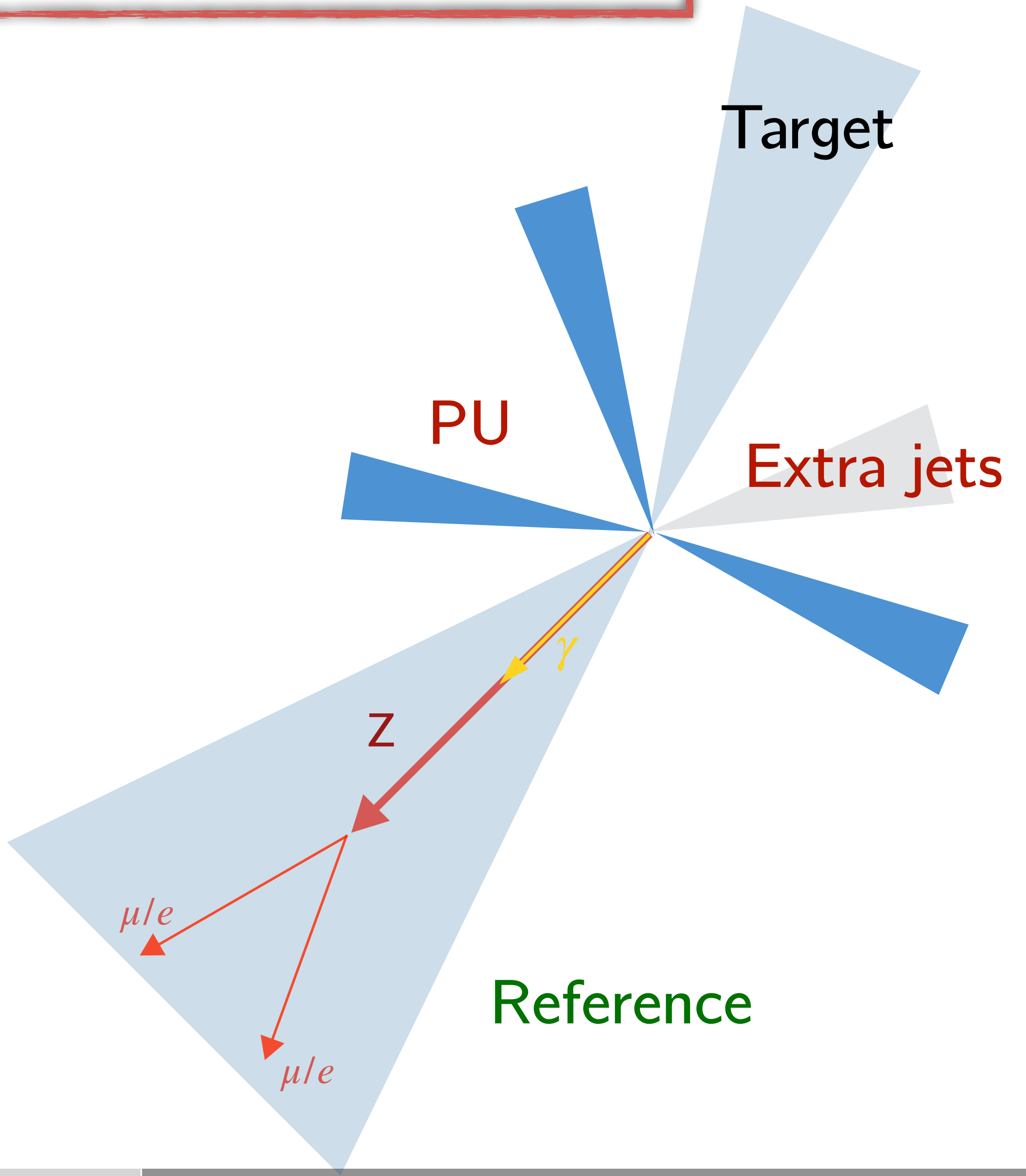
MC truth correction:

PU subtraction

Jet response calibration

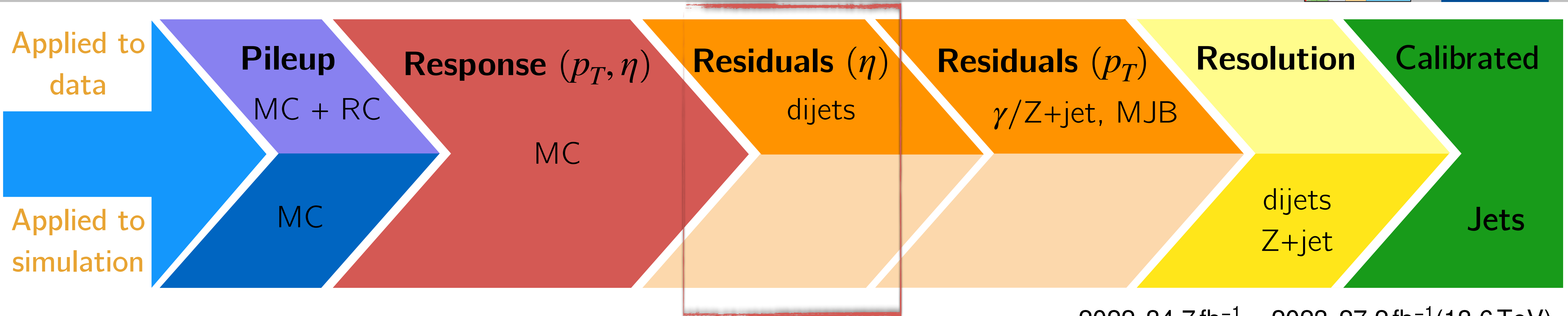
Residual corrections

- Based on precision of other reference objects
  - Electrons, photons, muons, other jets...
- Truth level unknown
  - Data/Simulation to reduce bias
- Realistic events taken into account
  - Extra jets, additional interactions,...



Jet calibration

CMS-DP-2024-039

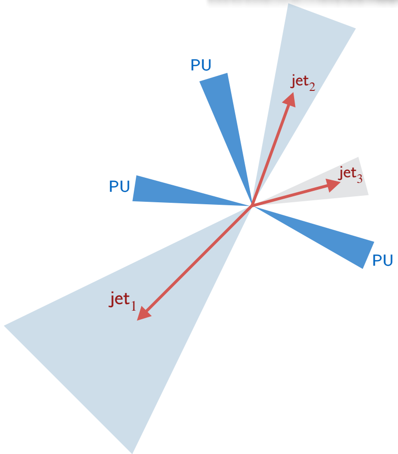


MC truth correction:

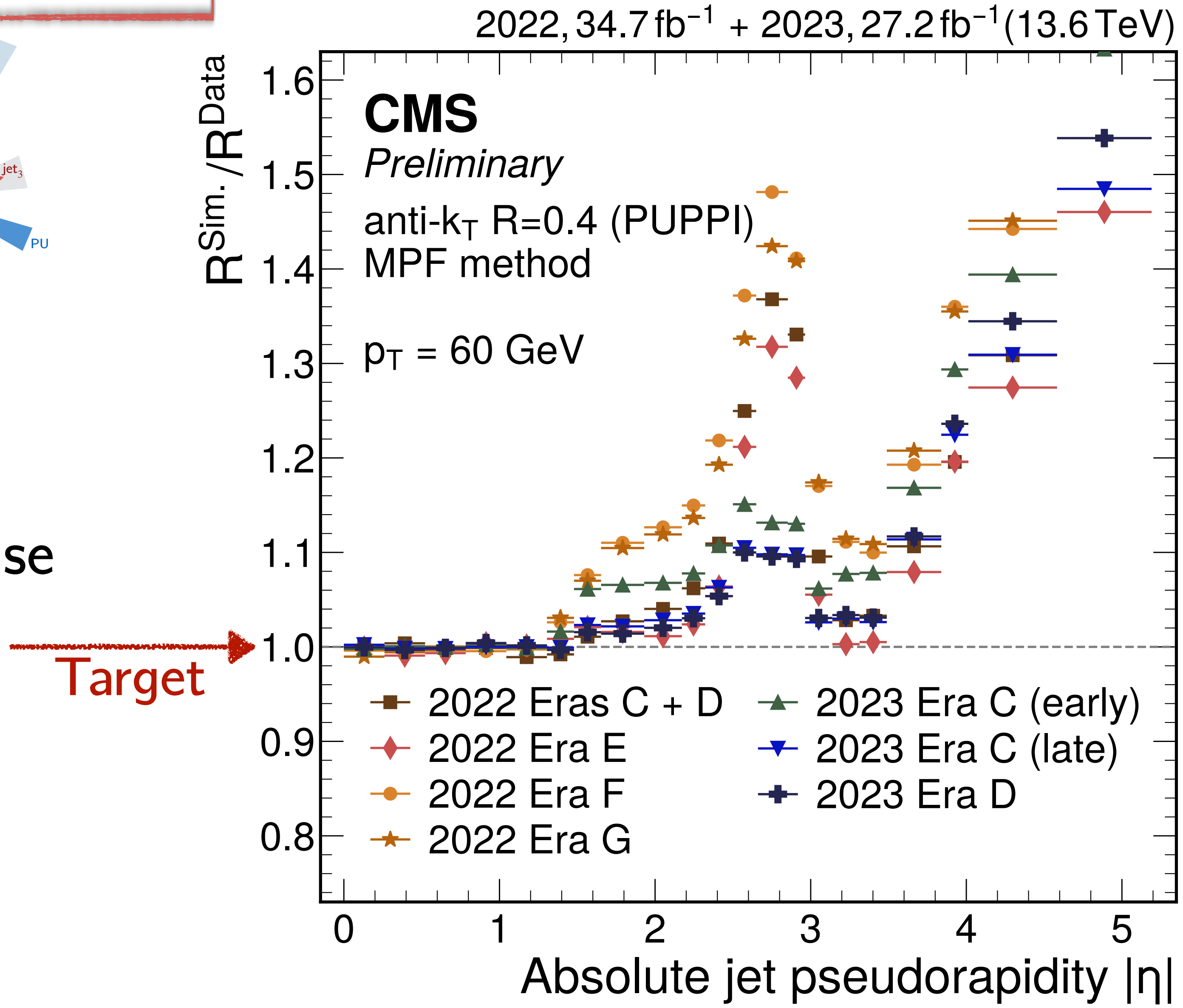
PU subtraction

Jet response calibration

Residual corrections

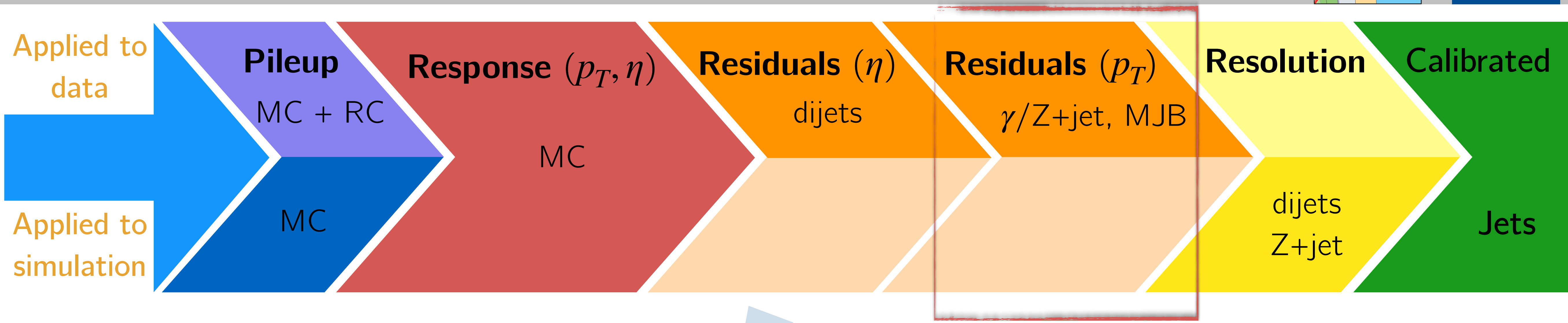


- Small residual correction of jet response applied to data based on dijet topology
- Address different response in each sub-detector (η dep.)



Jet calibration

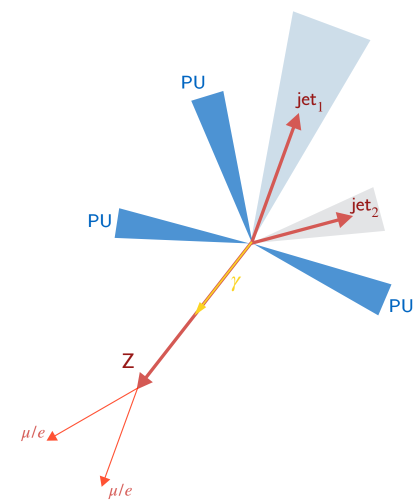
CMS-DP-2021-033



MC truth correction:

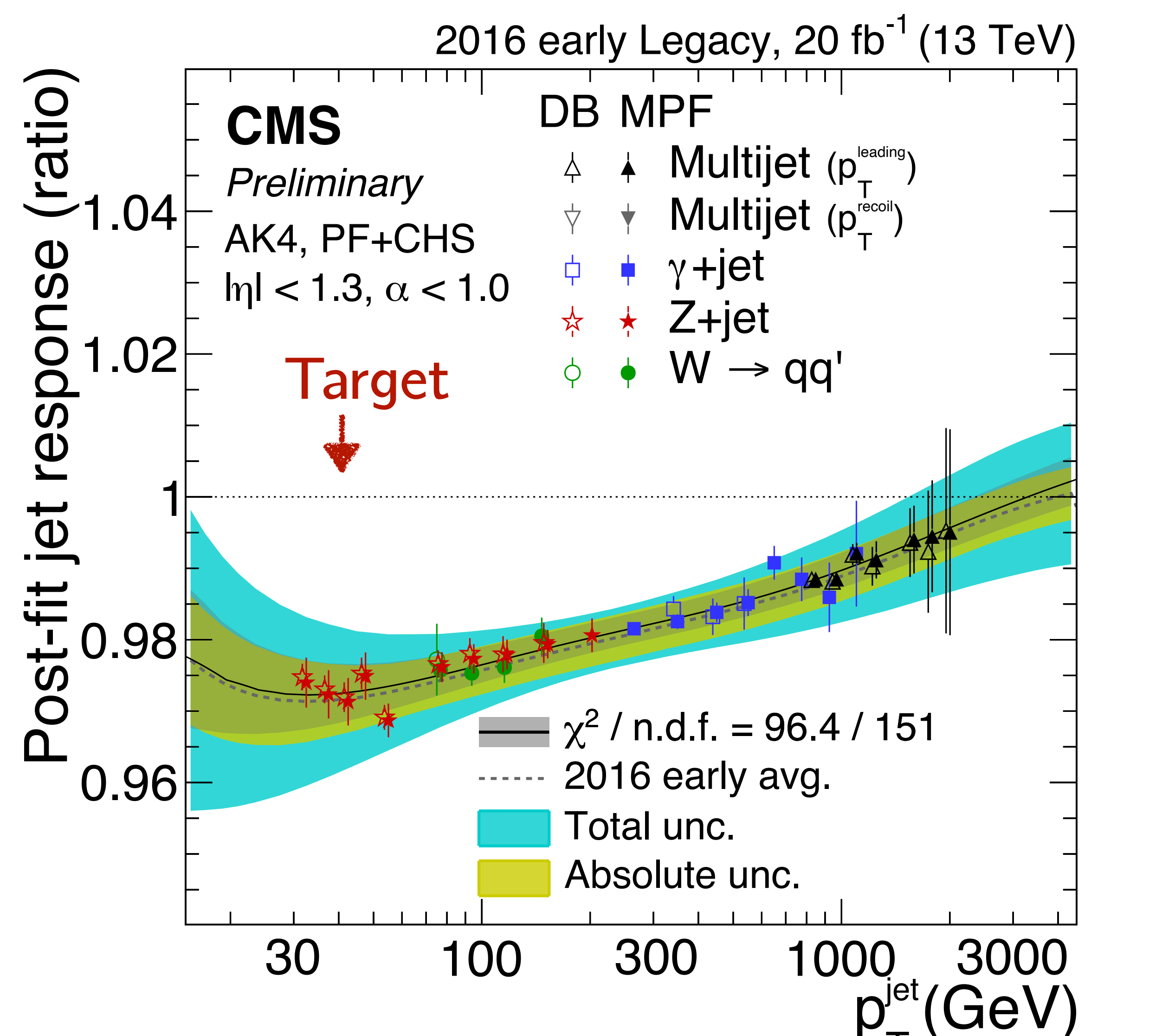
PU subtraction

Jet response calibration

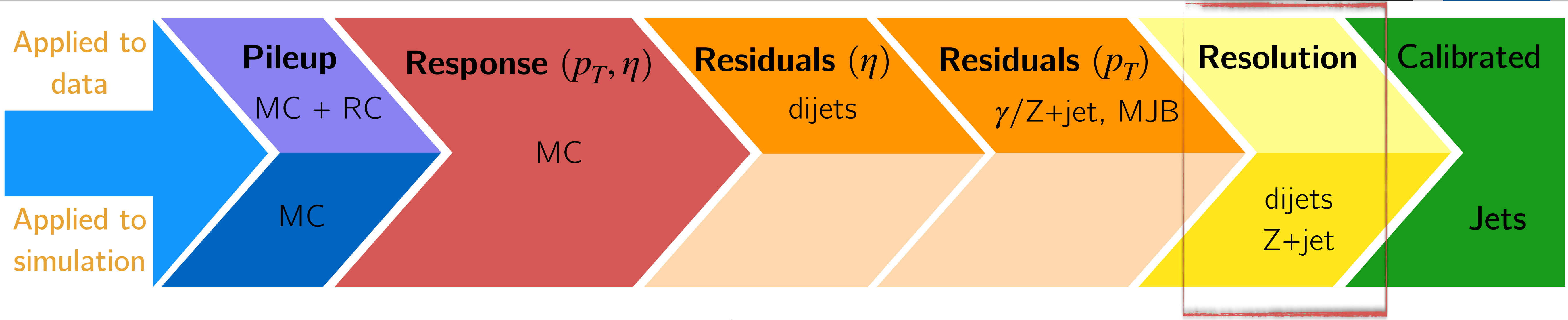


Residual corrections

- Additional p_T -dependent corrections accounting for absolute scale in barrel
- Determined relative to precisely measured reference objects (μ, e, γ, W)
- Channels combined in a global fit
  - Exploit individual precision in each phase-space



JEC DPNotes Legacy

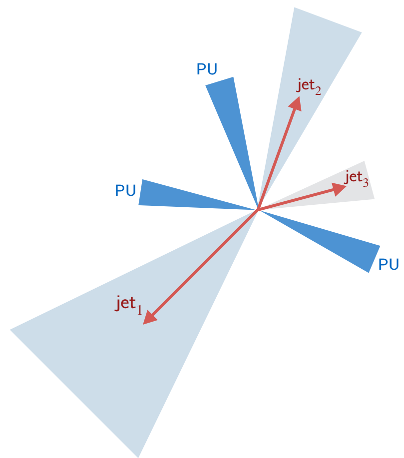


MC truth correction:

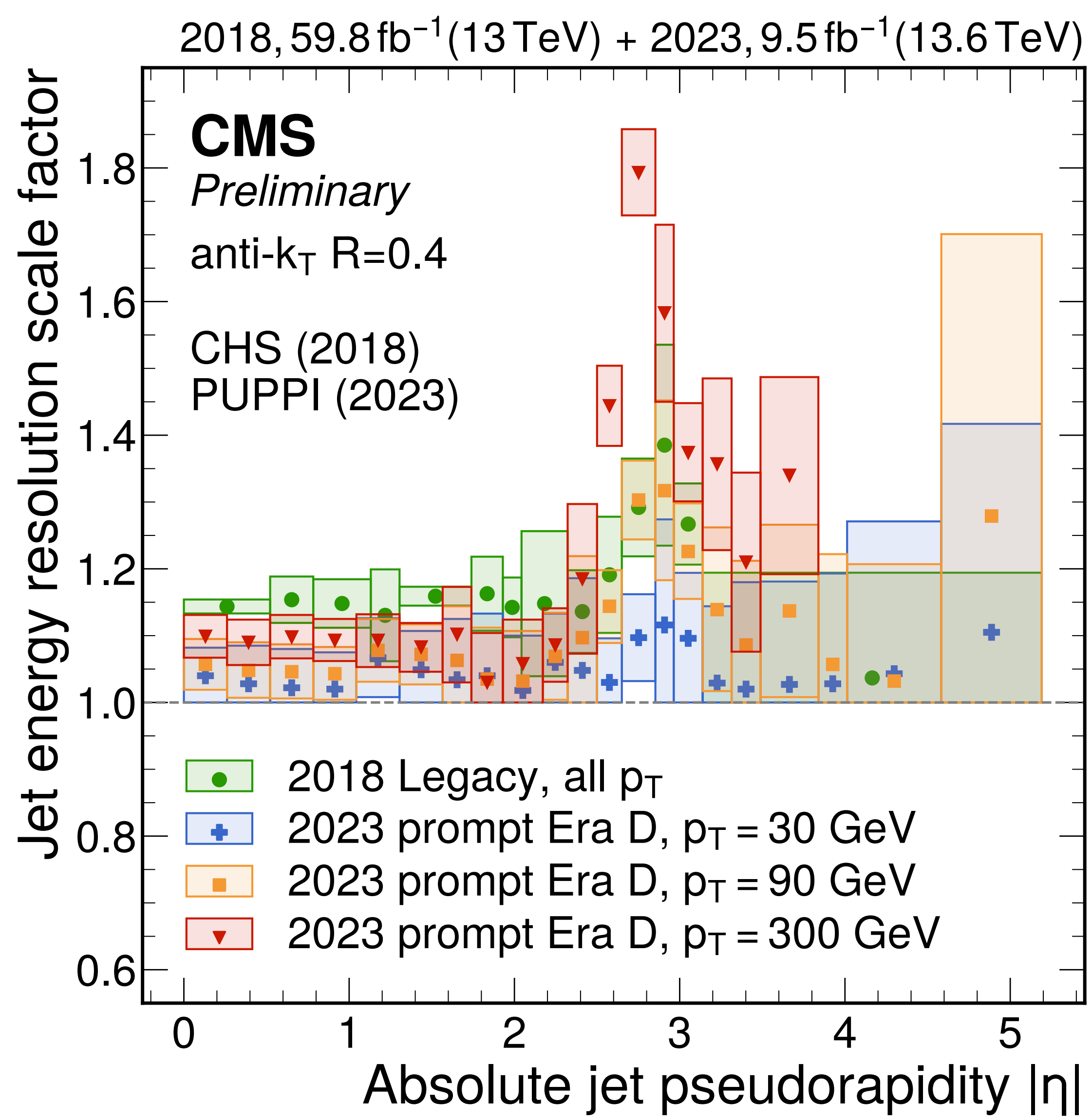
- PU subtraction
- Jet response calibration

Residual corrections

Jet energy resolution smearing

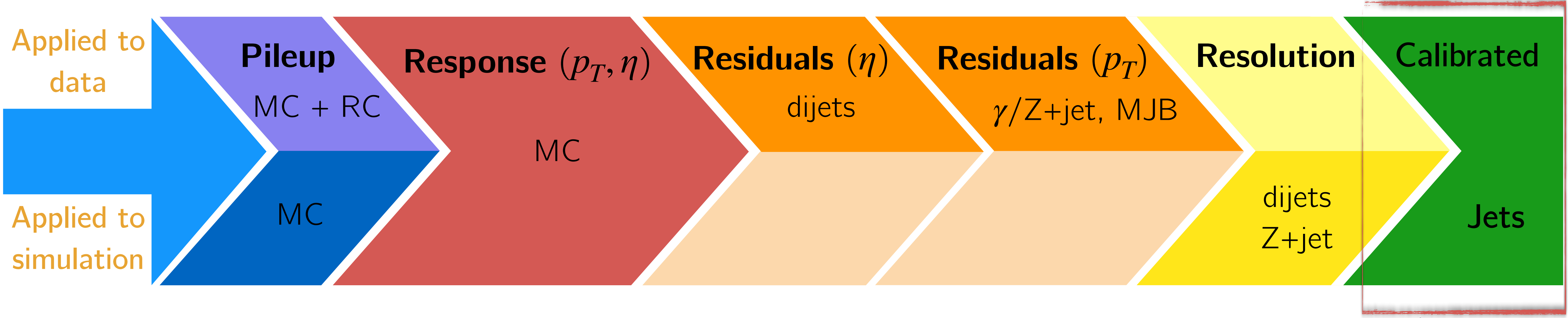


- Scale factors (SFs) applied to simulation to match resolution in data
- Direct balance in dijet events
- Sensitive to detector changes/calibration



Jet calibration

CMS-DP-2021-033



MC truth correction:

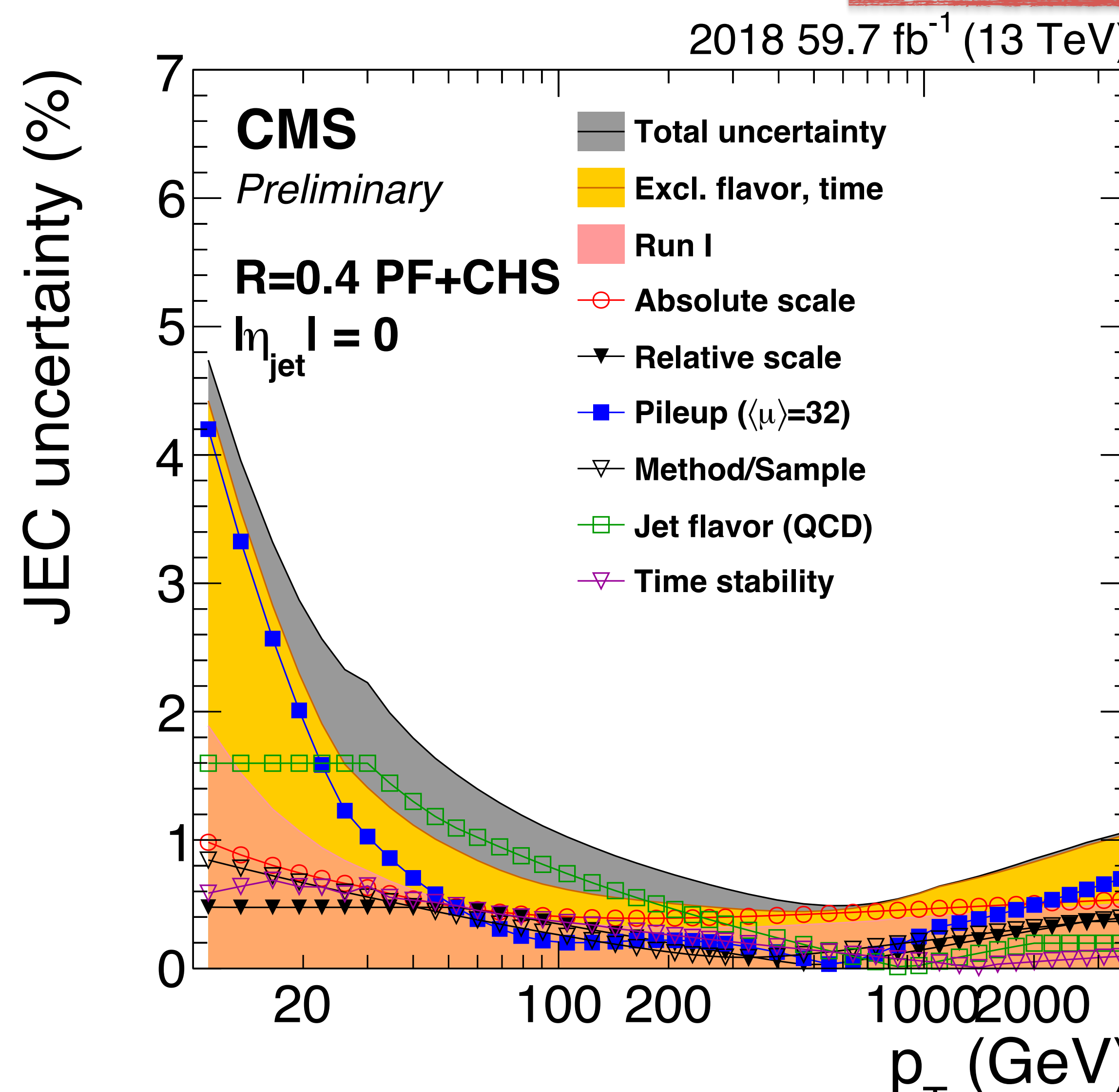
- PU subtraction
- Jet response calibration

Residual corrections

Jet energy resolution smearing

Jet energy scale uncertainties

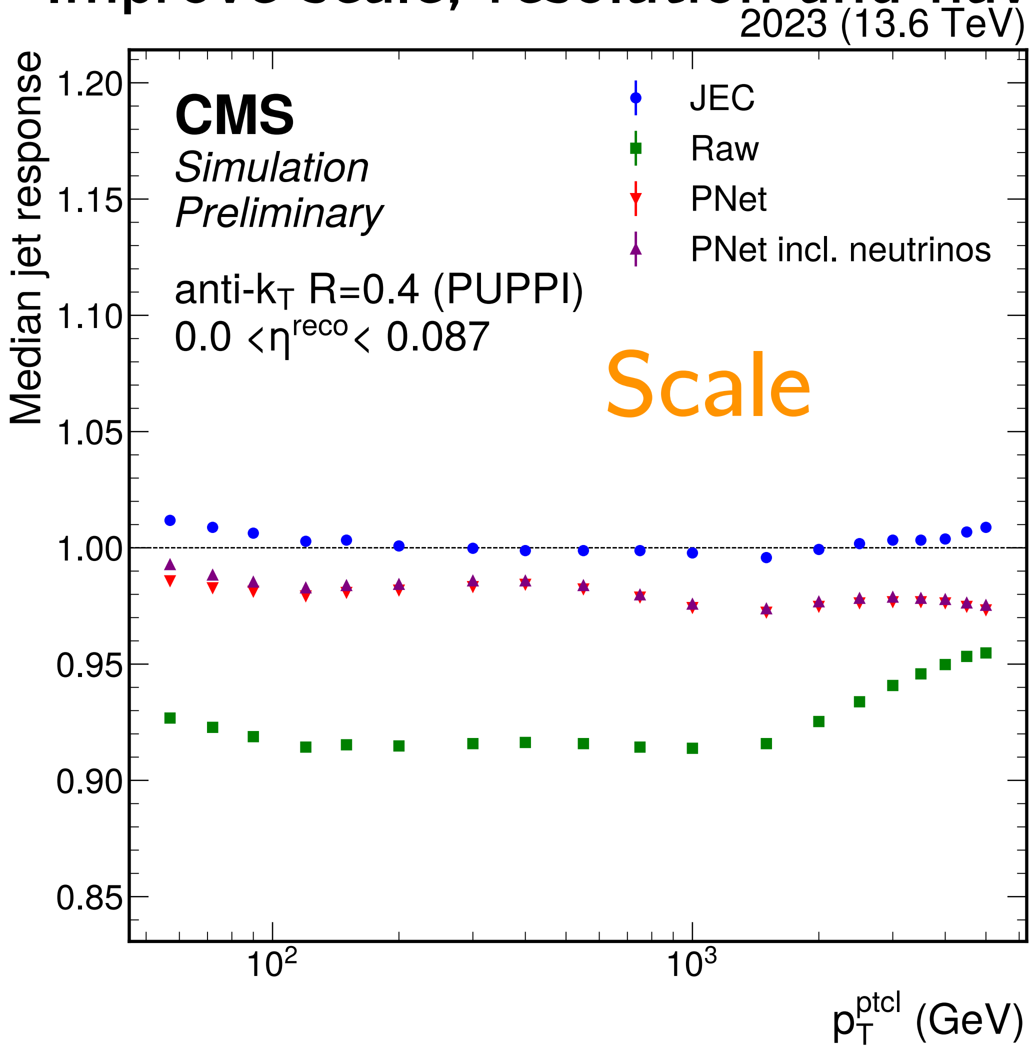
- Uncertainty ∼1% for jets pt >100 GeV
- Increasing contribution from PU
- Detector degradation:
  - Ageing, damage, ...

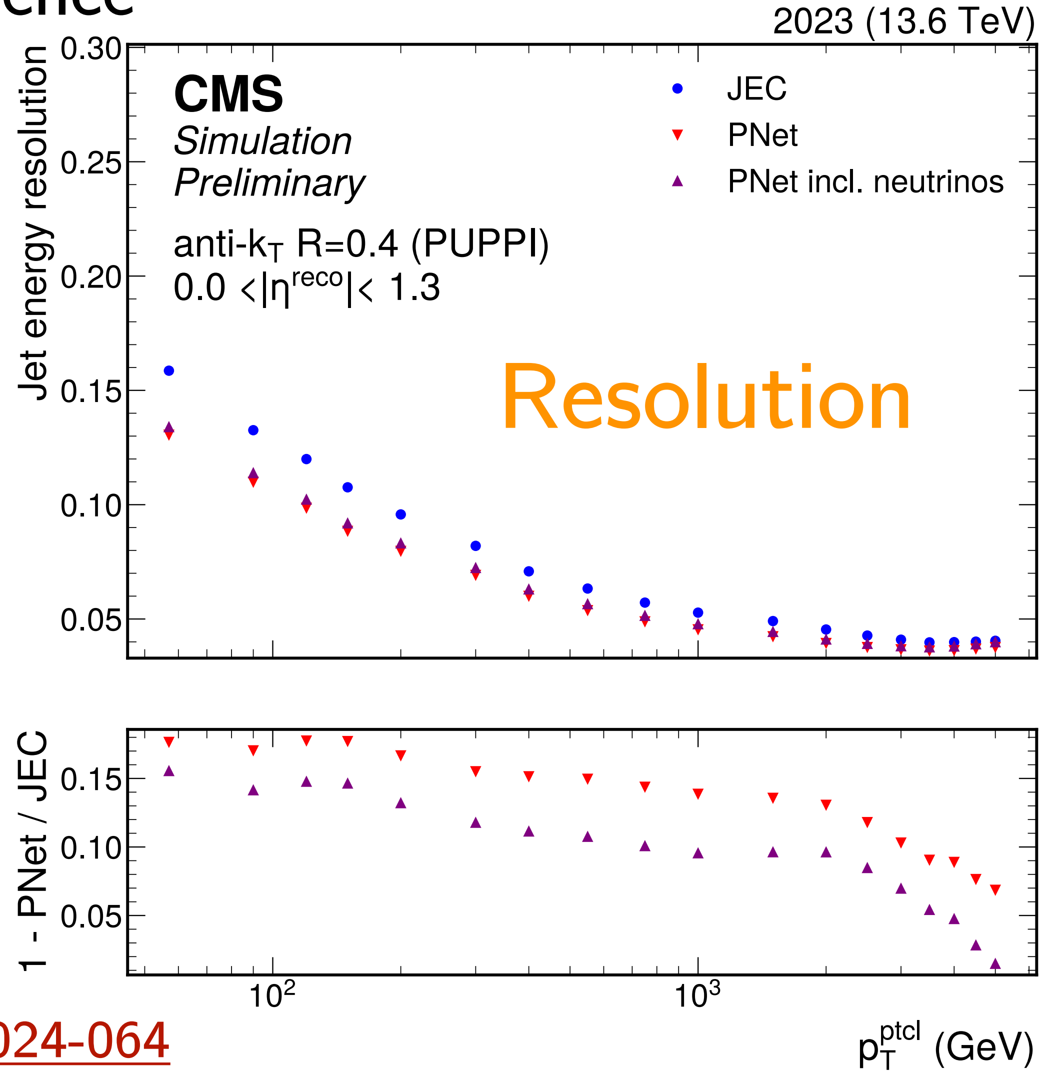


JEC DPNotes EOY

Machine learning:

- More performant wrt traditional algorithms
- Currently used for jet mass regression:
  - Direct effect on analyses's sensitivity
- Simultaneous training of tagging and regression for energy and mass:
  - Improve scale, resolution and flavour dependence

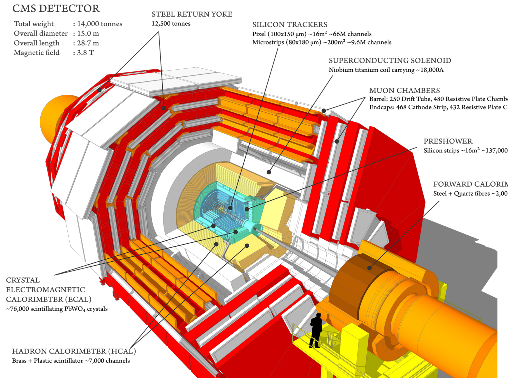


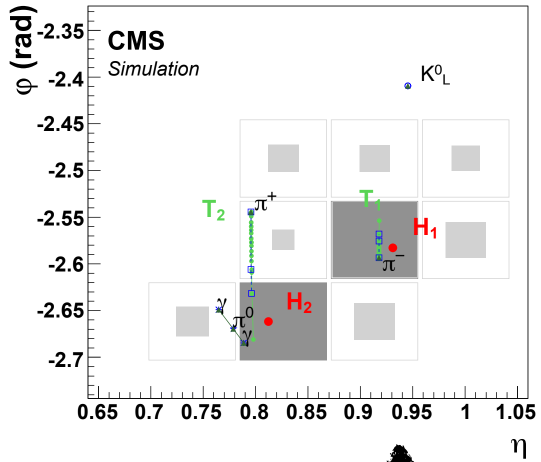


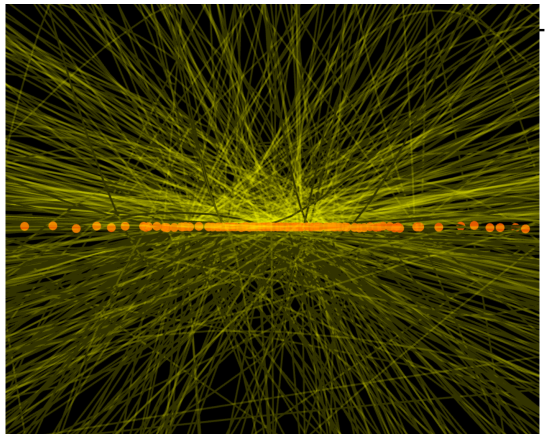
CMS-DP-2024-064

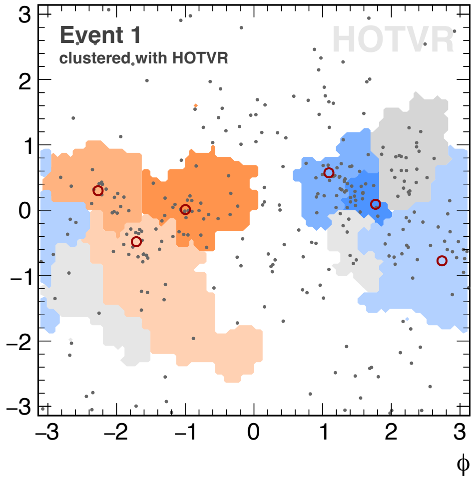
Summary and Outlook

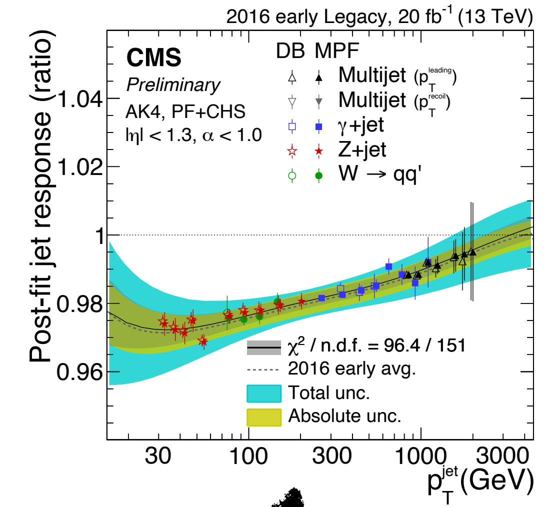
Detector	Reconstruction	Pileup	Clustering	Calibration









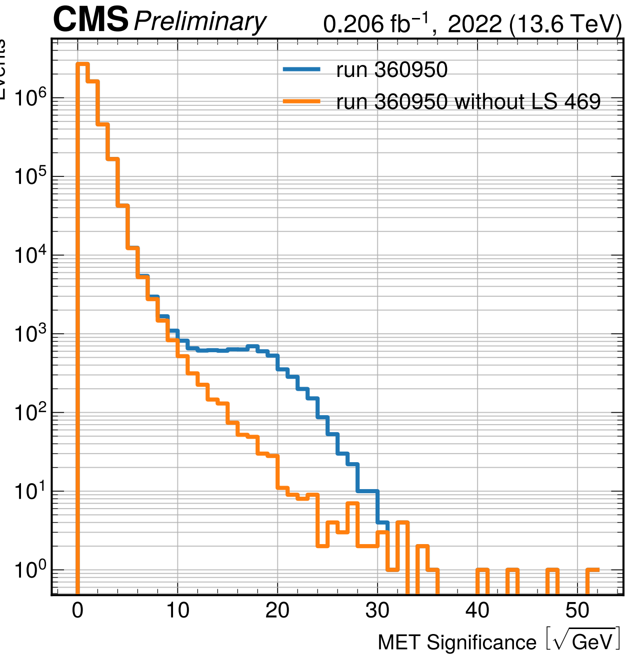


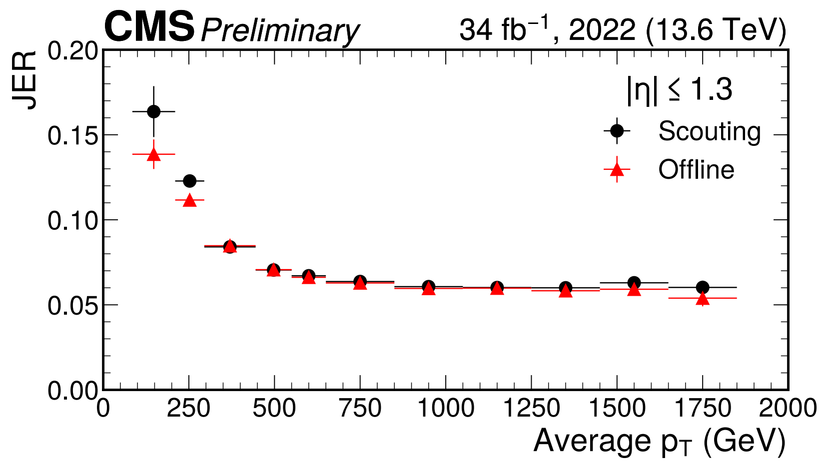
Anomaly detection

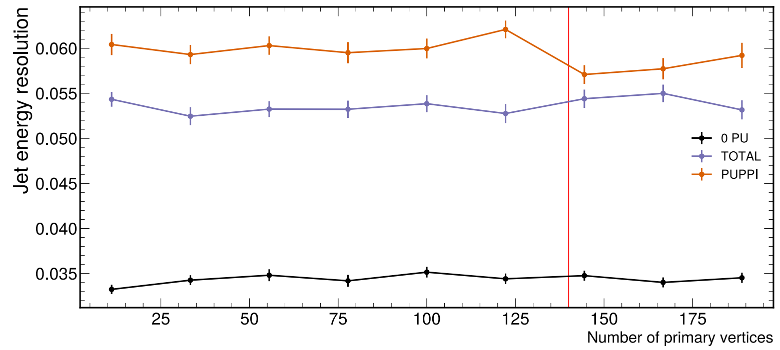
HLT scouting

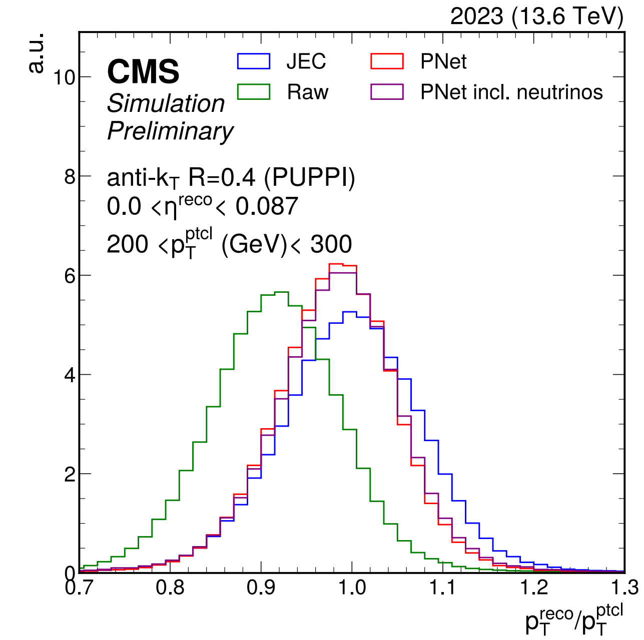
ML for PU mitigation

ML p_T regression







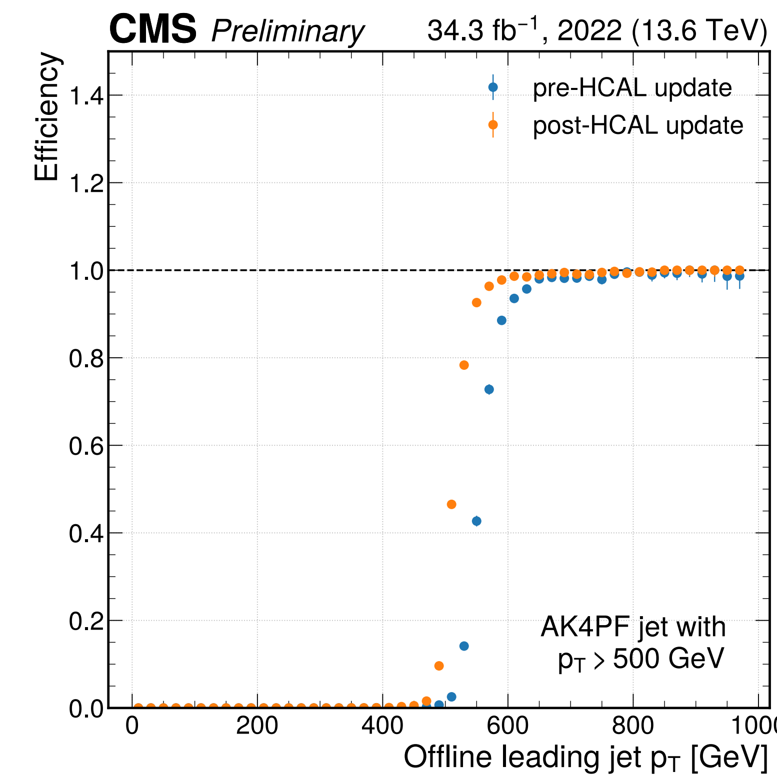


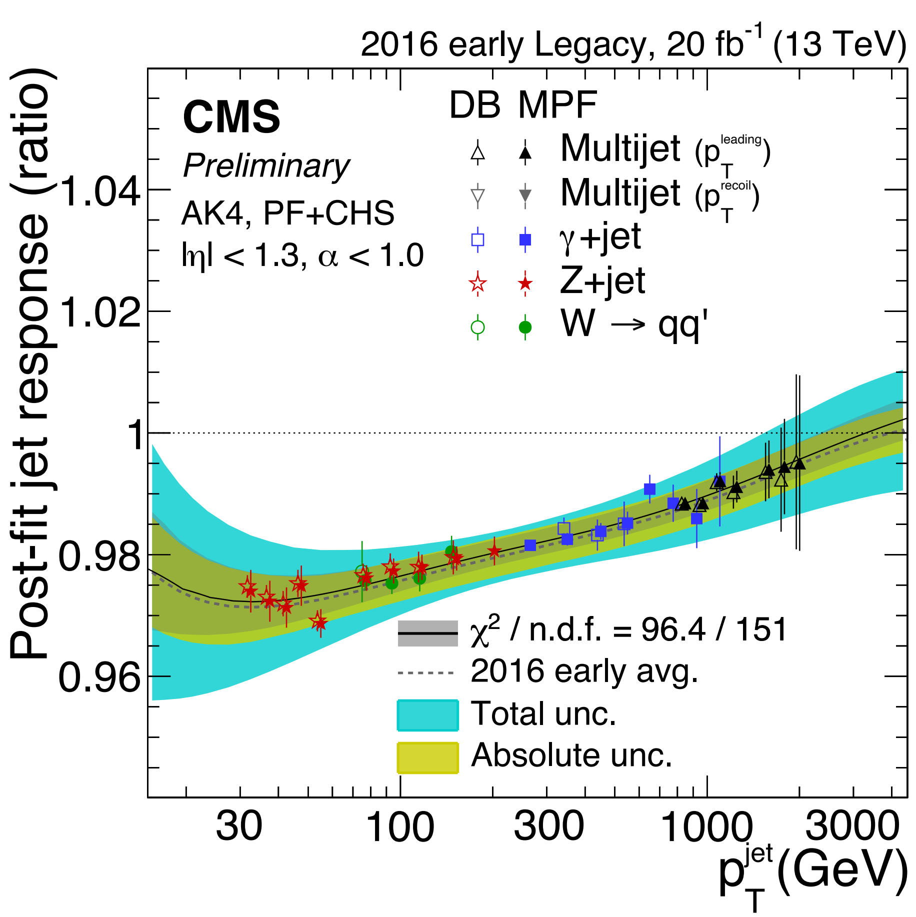
~~~ Additional Material ~~~

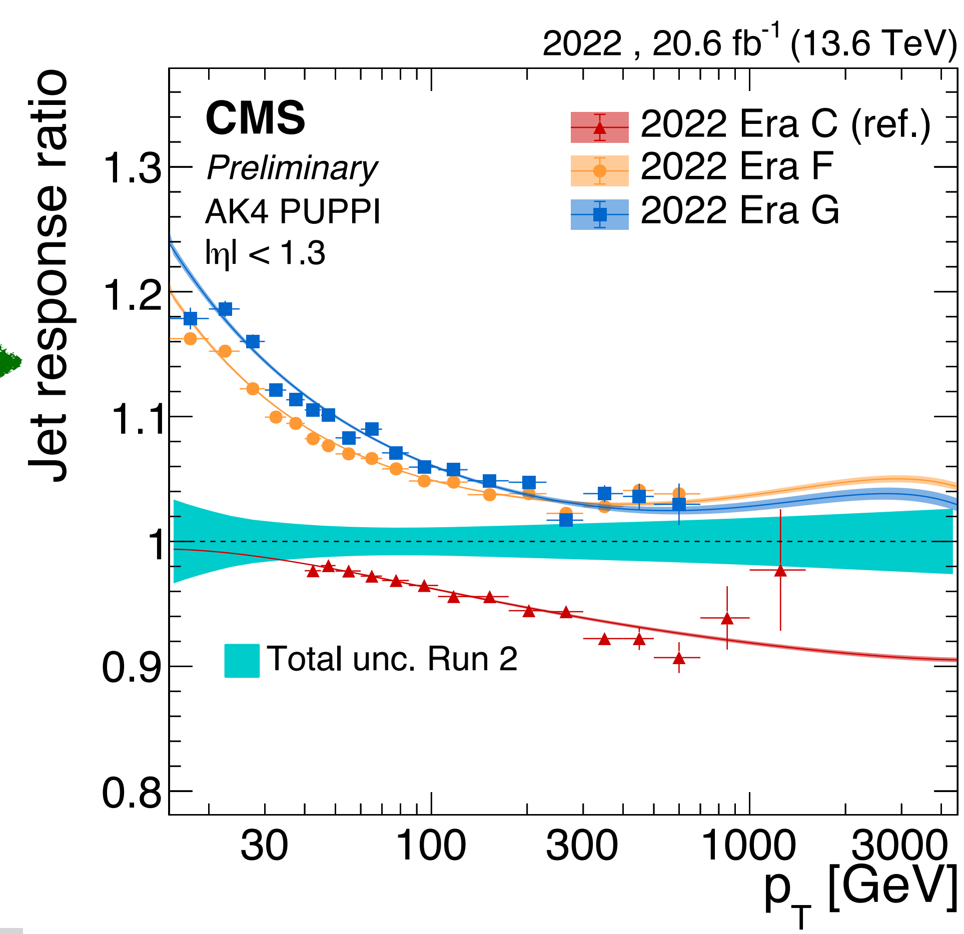
Current and future developments

Run3 calibration

- Usually small corrections for data offline...
  - ... corresponds to small effect on data online (triggers)
- Run3:
  - Substantial corrections needed ($\sim 10/20\,\%$, up to x2 in endcaps)
  - Fraction of data collected less efficiently

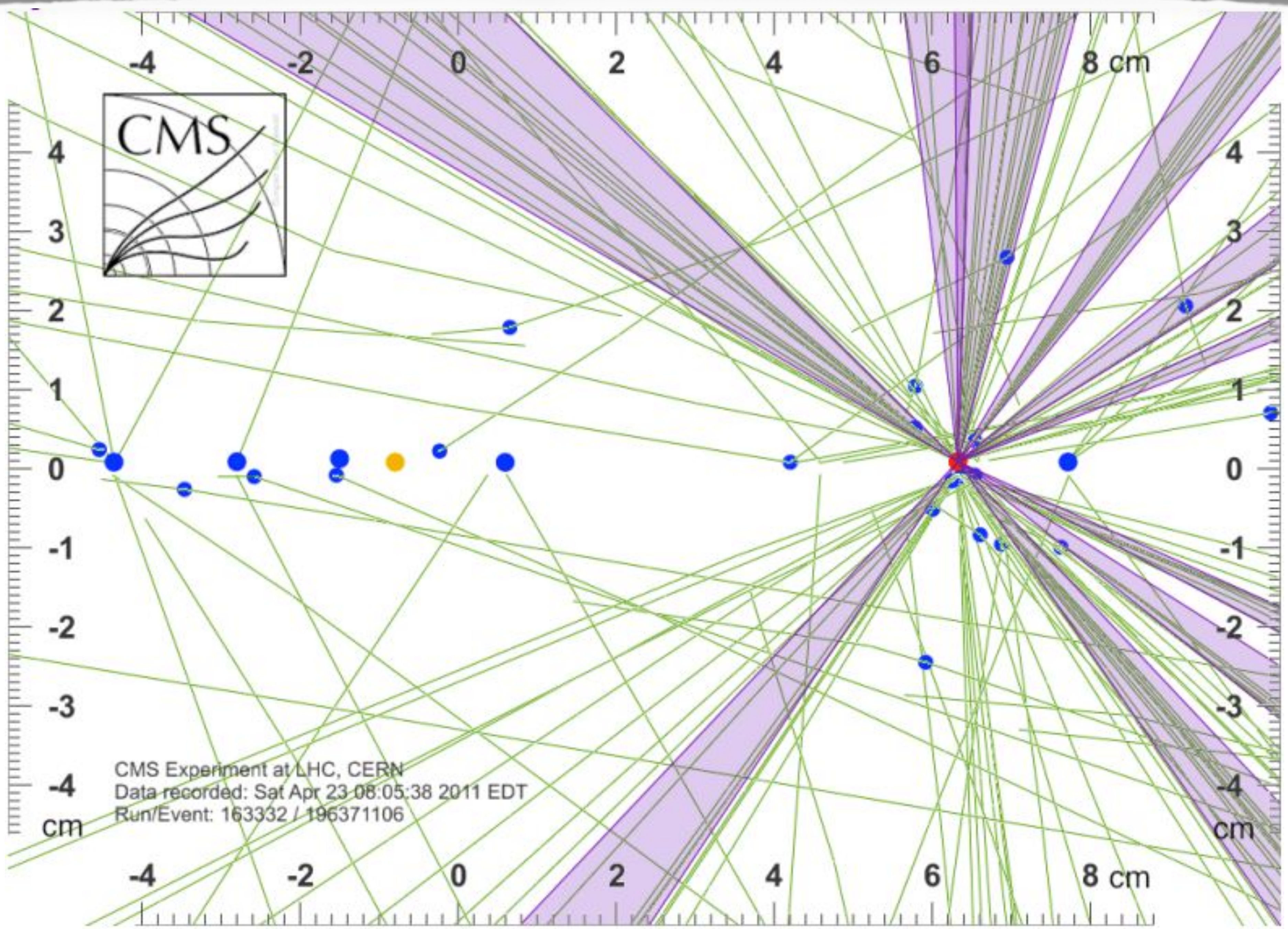






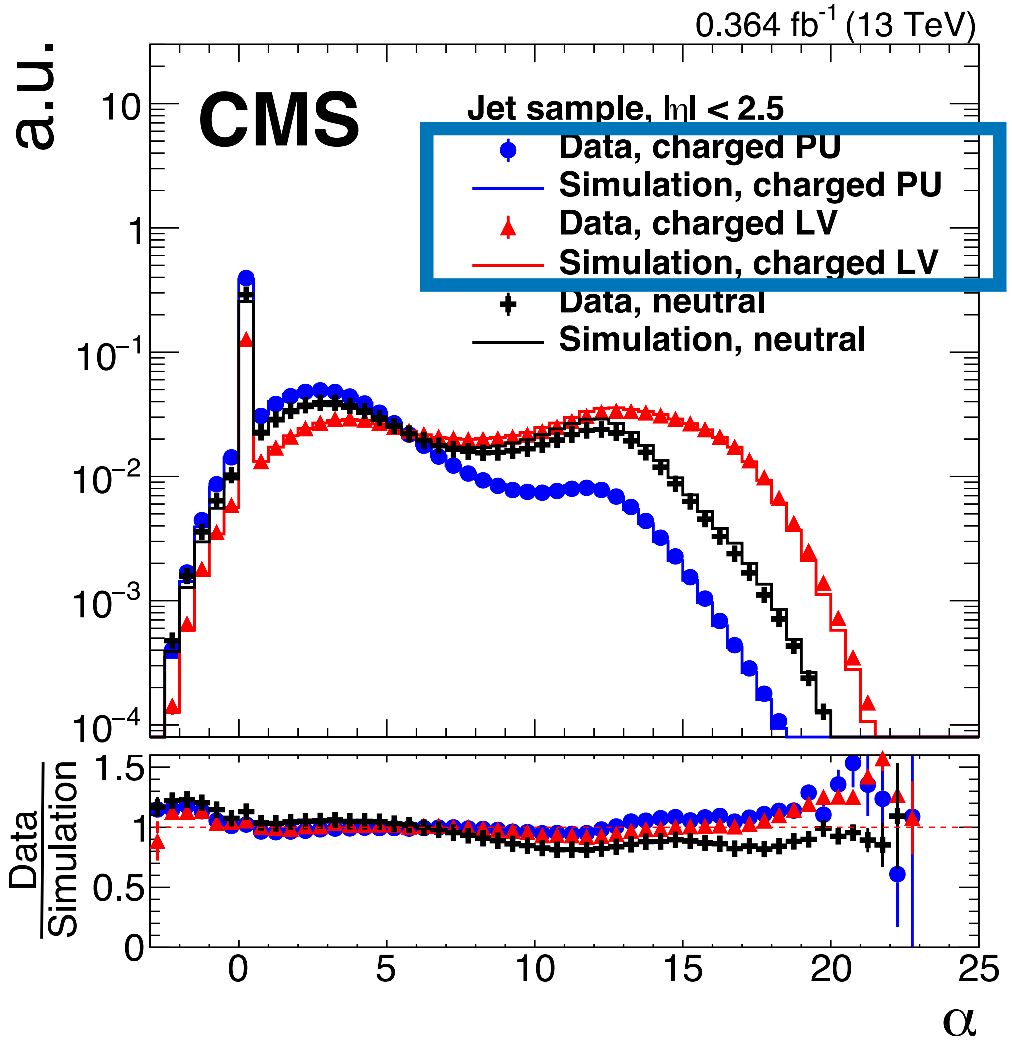
PUPPI in Detail

PUPPI in Detail

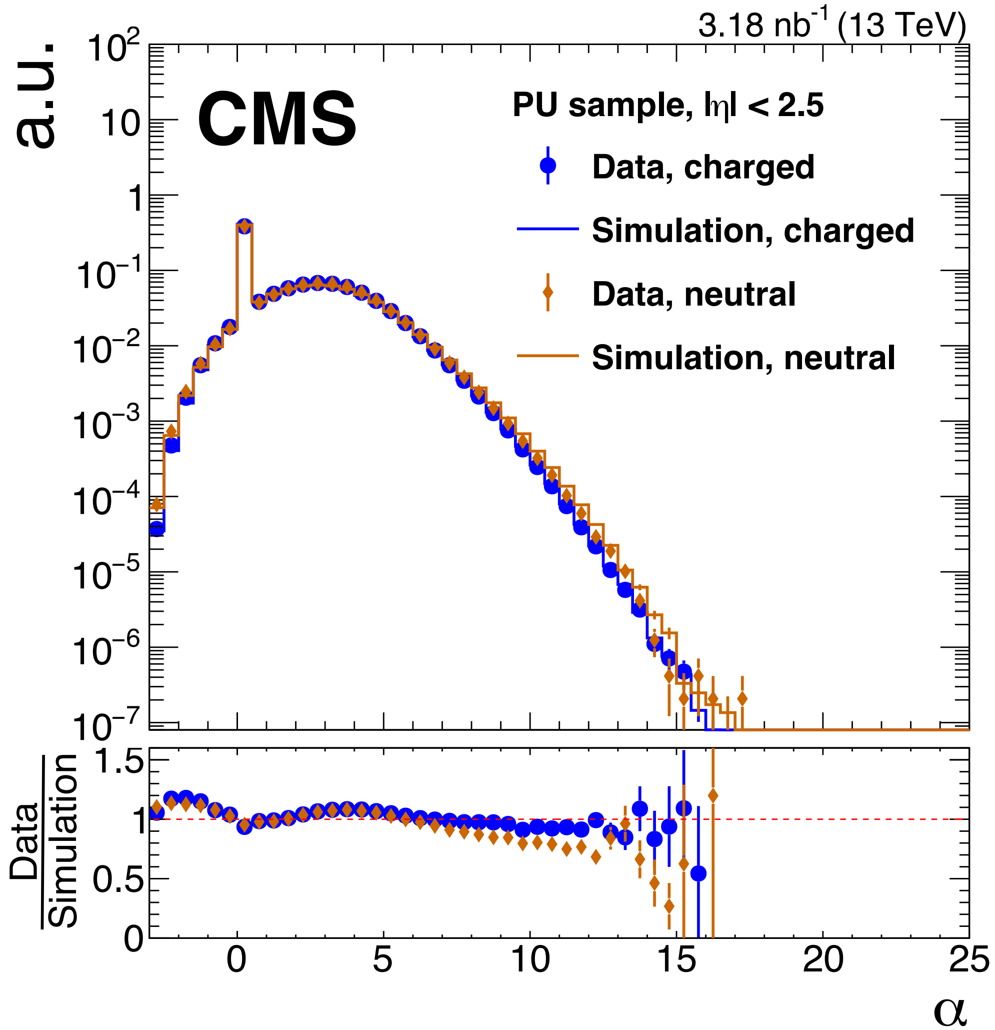


PUPPI in Detail

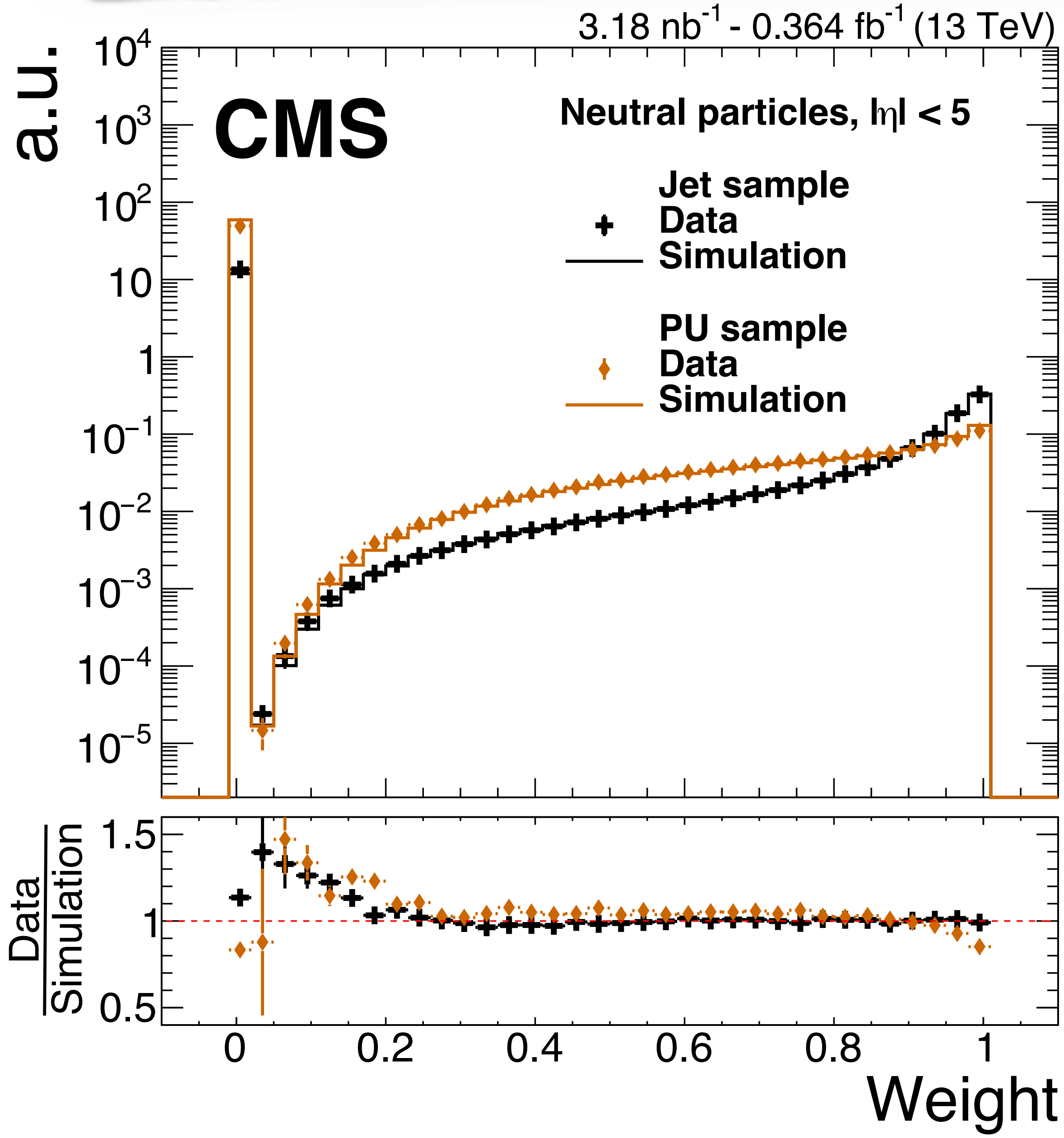
$$\alpha_i = \log \sum_{j \neq i, \Delta R_{ij} < R_0} \left(\frac{p_{\mathrm{T}j}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from leading vertex} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles} \end{cases}$$



$$\alpha_i = \log \sum_{j\neq i, \Delta R_{ij} < R_0} \left(\frac{p_{Tj}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from leading vertex} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles} \end{cases}$$



PUPPI in Detail



Pileup → Hard scattering

1. For each particle calculate

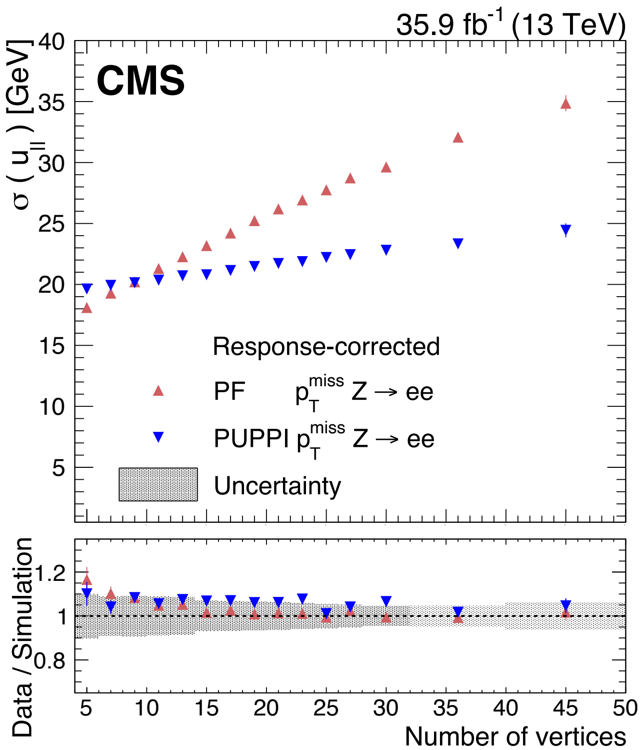
$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})\,|\alpha_i - \bar{\alpha}_{PU}|}{RMS_{PU}^2}$$

2. Assign a weight to each particle

$$w_i = F_{\chi^2, NDF=1}(\chi_i^2)$$

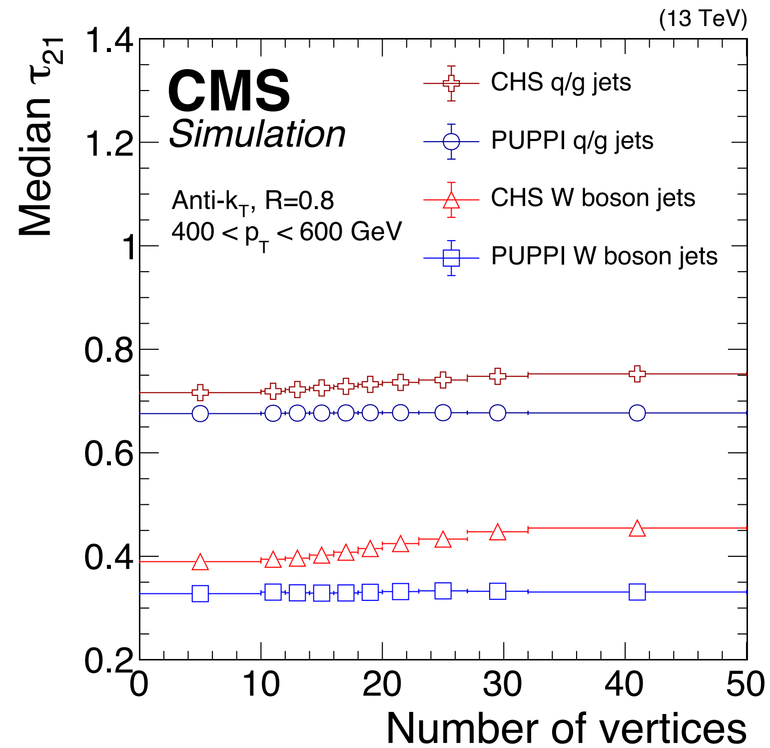
Validation of PUPPI

Performances of PUPPI jets/MET were extensively studied and compared to CHS jets/PF MET in JME-18-001



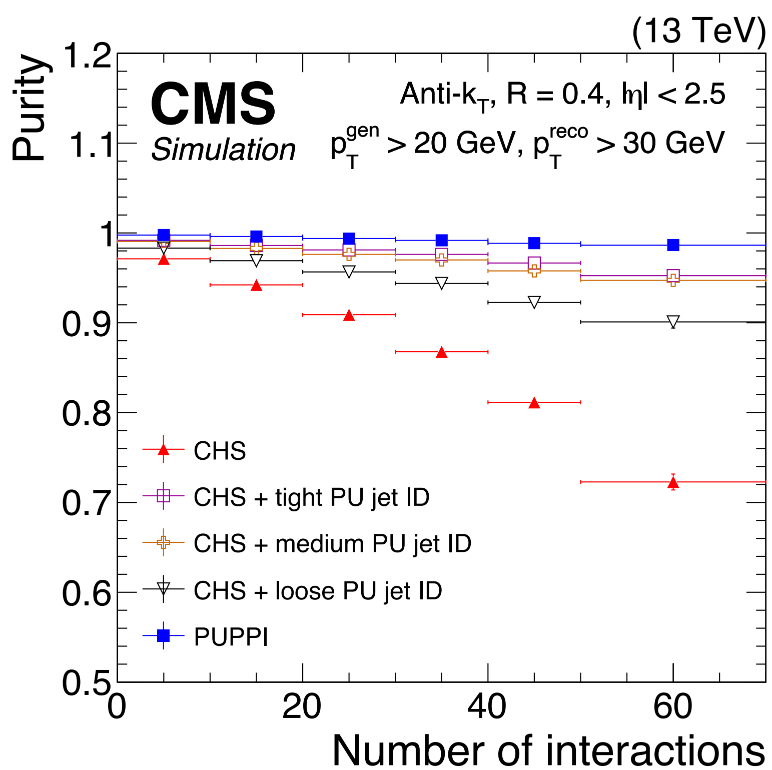
MET resolution

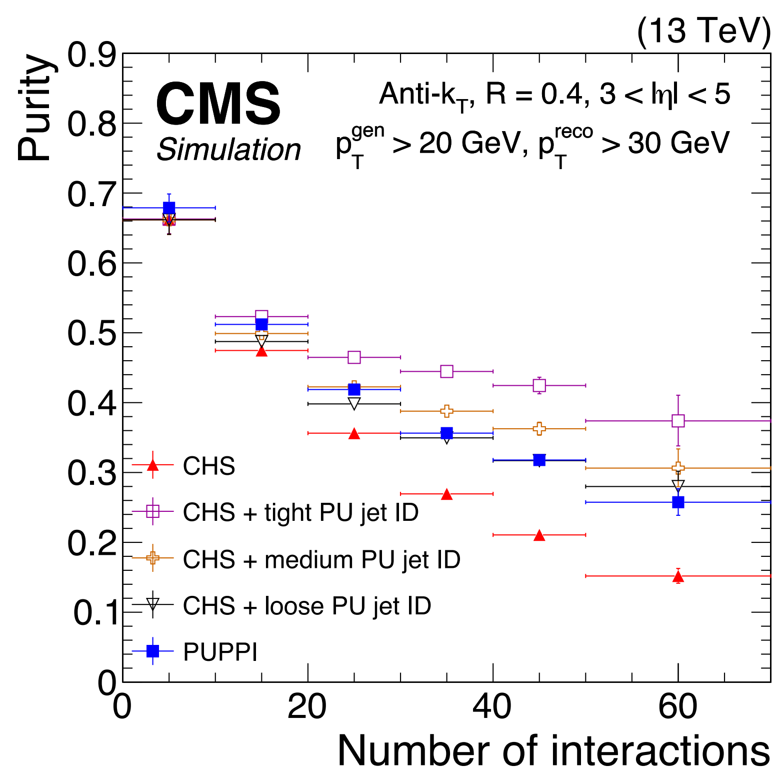
PUPPI has a better MET resolution and less fake MET.



W boson identification

PUPPI is more stable in jet mass, jet mass resolution and τ_{21}.

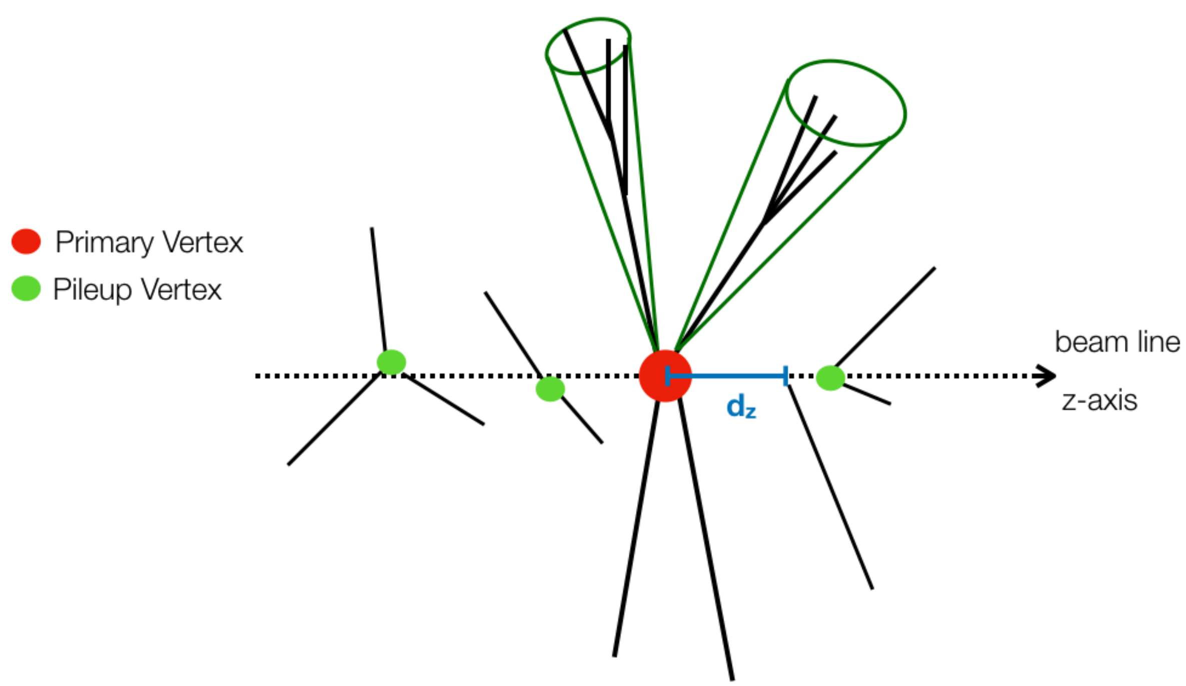




PU jets rejection

In $|\eta| < 2.5$ PUPPI rejects more PU jets than all other techniques

In $|\eta| > 2.5$ PUPPI is compatible to CHS+PUJetID



<table>
<tr><th>Charged particles</th><th>CHS</th><th>PUPPI v15</th></tr>
<tr><td>used in the LV fit</td><td>keep</td><td>keep</td></tr>
<tr><td>used in the PU vertex fit</td><td>reject</td><td>if 1st or 2nd PU verte
&& $|d_z| < 0.2$ cm keep
else reject</td></tr>
<tr><td>not used in a vertex fit</td><td>keep</td><td><table><tr><th></th><th>$p_T > 20$ GeV</th><th>$p_T < 20$ GeV</th></tr><tr><td>$|\eta| < 2.4$</td><td>keep</td><td>calculate a weight</td></tr><tr><td>$|\eta| > 2.4$</td><td>keep</td><td>if $|d_Z| < 0.3$ cm keep
else reject</td></tr></table></td></tr>
</table>

Tab. 1: Categories for charged particles in CHS and PUPPI.

A charged particle can be either: used in the fit of the LV, used in the fit of a PU vertex or not used in any fit (see plot on the top left). The categories for CHS and PUPPI for each of the cases is shown in Tab. 1.

In order to recover tracks mistakenly used in the fit of another vertex (vertex splitting or track stealing by a nearby PU vertex), charged particles belonging to one of the first two PU vertices and with $|d_Z| < 0.2$ cm are kept.

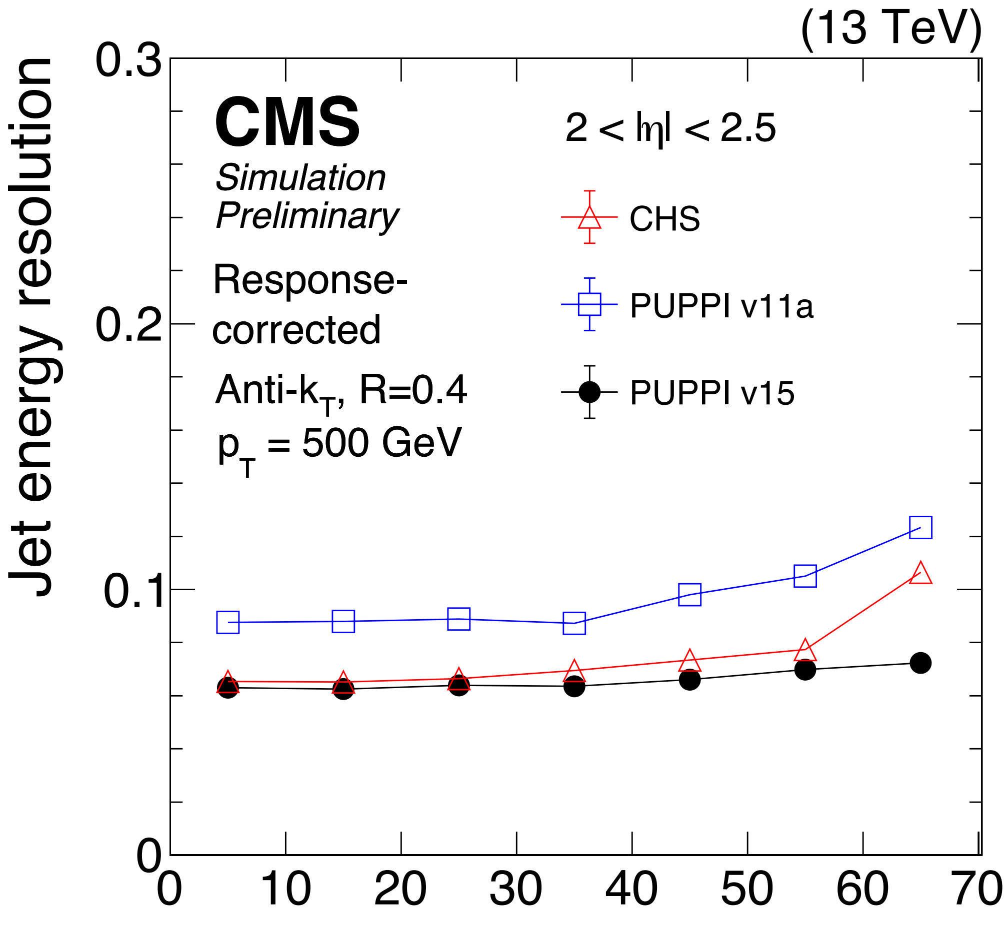
PUPPI tune v17

	$p_T > 20$ GeV	$p_T < 20$ GeV
$\|\eta\| < 2.4$	keep	calculate a weight
$\|\eta\| > 2.4$	keep	if $\|d_Z\| < 0.3$ cm keep else reject

	$p_T > 20$ GeV	$p_T < 20$ GeV	
		$p_T > 4$ GeV && FromPV $= = 2$	else
$\|\eta\| < 2.4$	keep	keep	calculate a weight
$\|\eta\| > 2.4$	keep	keep	if $\|d_Z\| < 0.3$ cm keep else calculate a weight

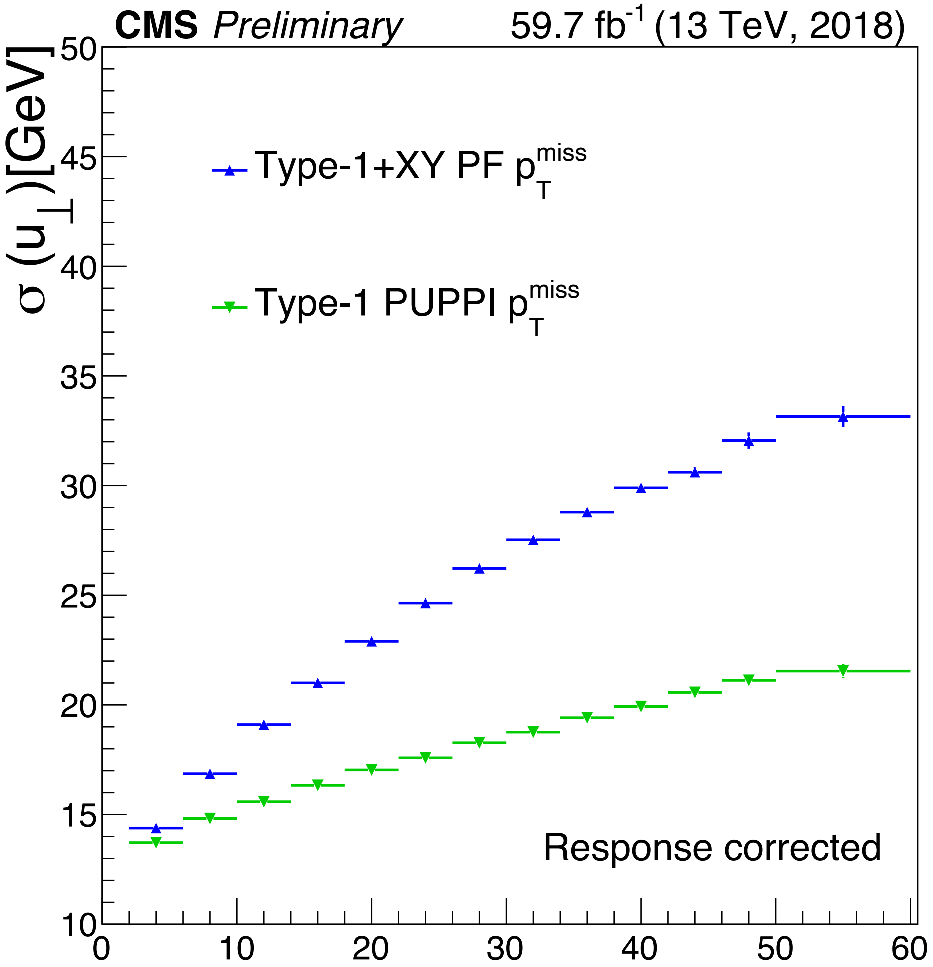
Pileup mitigation techniques – Puppi for Run 3

- Widely used in Run 2, default in Run 3
- Improved all jet-related variables
  - Jet efficiency and purity (matched to generator-level jets)
  - Jet substructure
  - Jet resolution (the lower the better)
  - Improved performance also on lepton isolation, missing transverse energy, ...



CMS-DP-2021-001

Better



CMS-DP-2020-031