Karlsruhe Institute of Technology

Jet-substructure tools and boosted hadronic boson identification in CMS

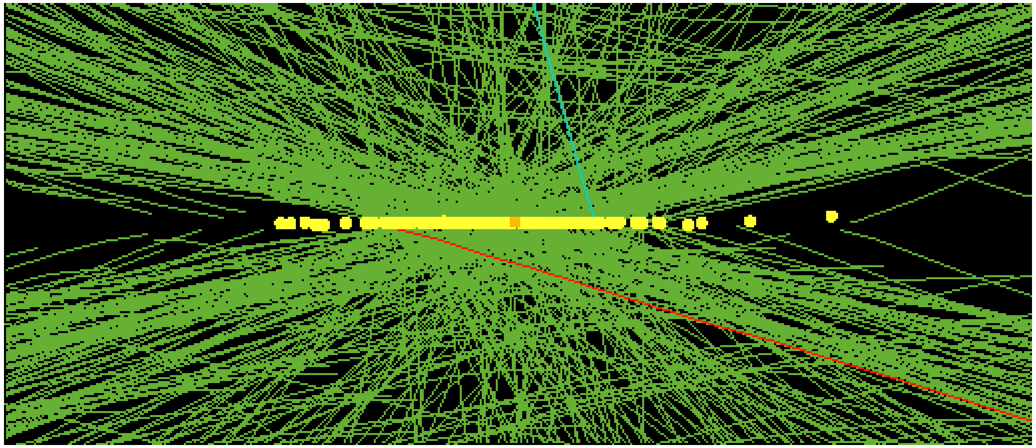
Ivan Shvetsov, Matthias Mozer, Thomas Müller, Jeannine Wagner-Kuhr

8th Annual Meeting of Helmholtz Aliance "Physics at the Terascale"
DESY, Hamburg, 2nd December 2014

Institut für Experimentelle Kernphysik, Karlsruhe Institute of Technology

Introduction

- One of the challenges of the coming LHC Run would be increase of instantaneous luminosity which will result in a *large number of pileup interactions.*
- In such pileup environment reconstruction of the jet properties will become more demanding
- Contamination from pileup degrades the ability to reconstruct jet observables
  → Study several pileup mitigation tools used in CMS



- Since LHC will work at doubled energy in Run II reconstruction of boosted hadronic vector boson will of greater importance.
- Several V-tagging algorithms were studied.

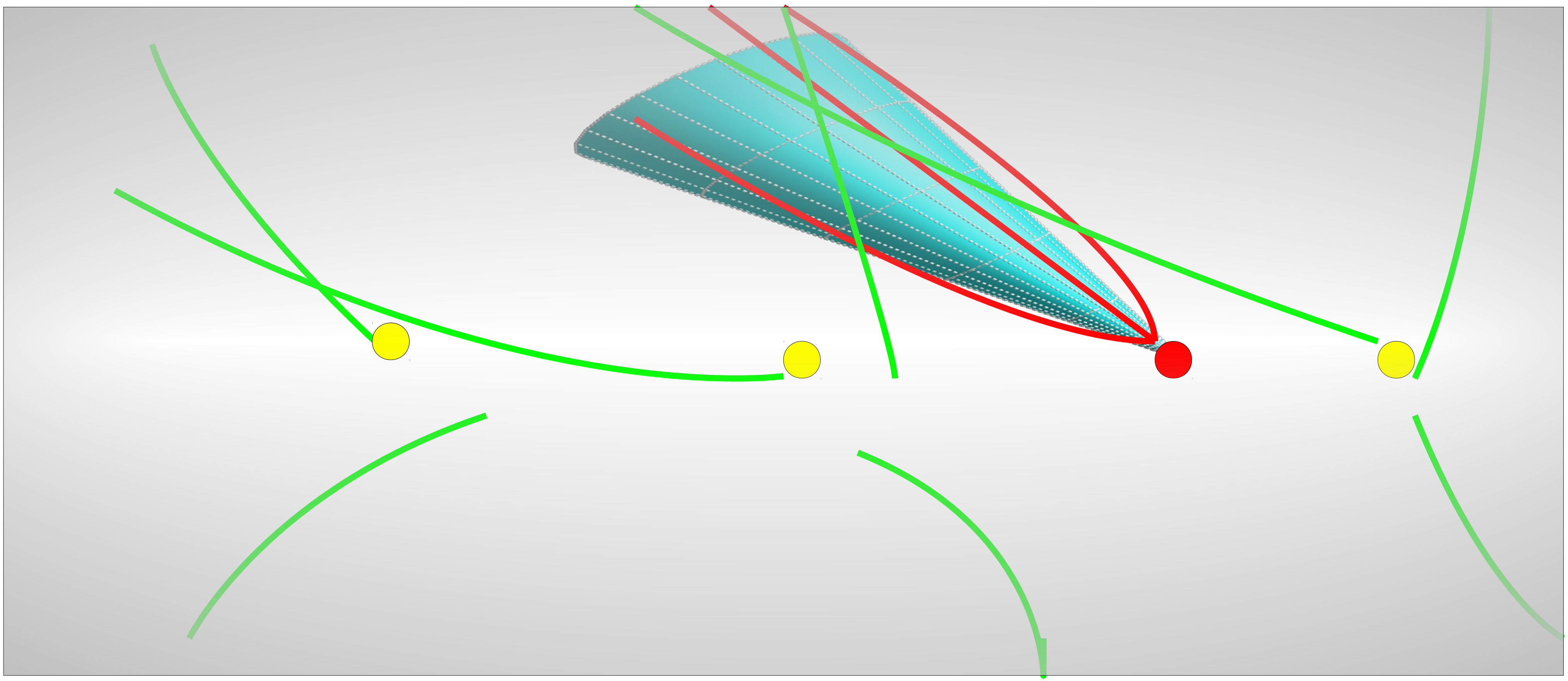
Pileup mitigation tools

- Increase of pileup →
  pileup mitigation tools:
  - Charge Hadron Subtraction **Run I**
  - PUPPI **NEW**

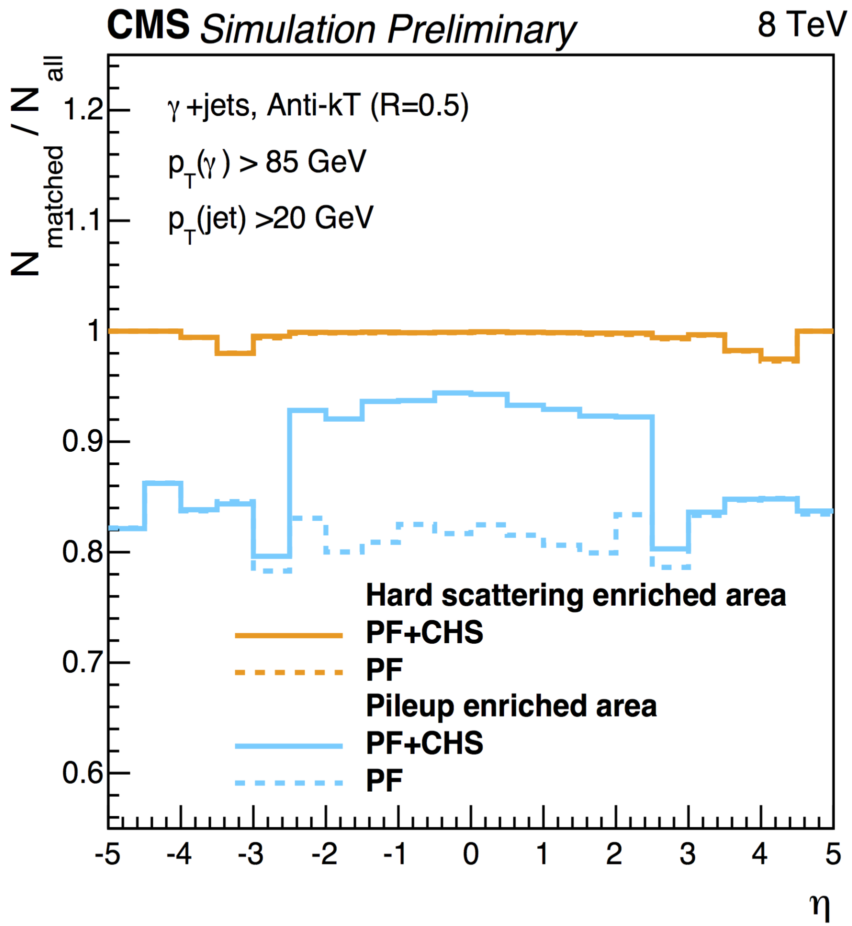
- **Jet grooming algorithms**:
  - pruning **Run I**
  - trimming **ATLAS**
  -softdrop **NEW**

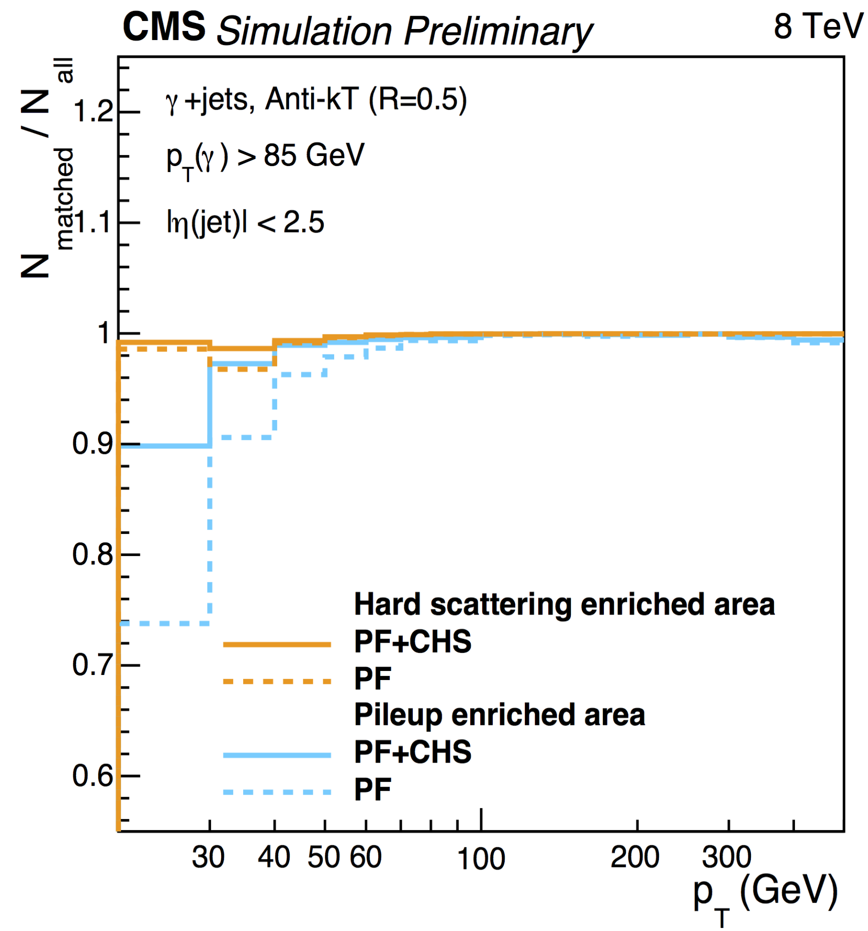
Charged Hadron Subtraction

- Leading vertex (LV) : highest sum $\mathbf{\Sigma\, |\, p_T^{track}\, |^2}$
- Subleading vertices are classified as pileup vertices
- All tracks that are associated to the pile vertices are removed
- Remaining PF candidates are allowed to cluster



CHS: matching efficiency





CHS reduces the rate of unmatched pileup jets from 20% to 5% in tracker region.

Unmatched jets have high portion of energy from pileup and are refered as pileup jets.

PUPPI

- **PUPPI = Pile Up Per Particle Identification**

PUPPI overview

for a particle *i* with nearby particles *j*

[1] define a local metric, α, that differs between pileup (PU) and leading vertex (LV)

[2] using tracking information (e.g. charged particles) "sample" the event, define unique distributions of α for PU and LV

[3] for the neutrals, ask "how PU-like is α for this particle?", compute a weight for how un-PU-like (or LV-like) it is

[4] reweight the four-vector of the particle by this weight, then proceed to cluster the event as usual

Let's assume something similar to **Particle Flow inputs** (*not a requirement*):
neutral hadrons
charged hadrons from LV
charged hadrons from PU

PUPPI: metrics

- Various metrics and settings are possible
- Metric used in the analysis is defined as :

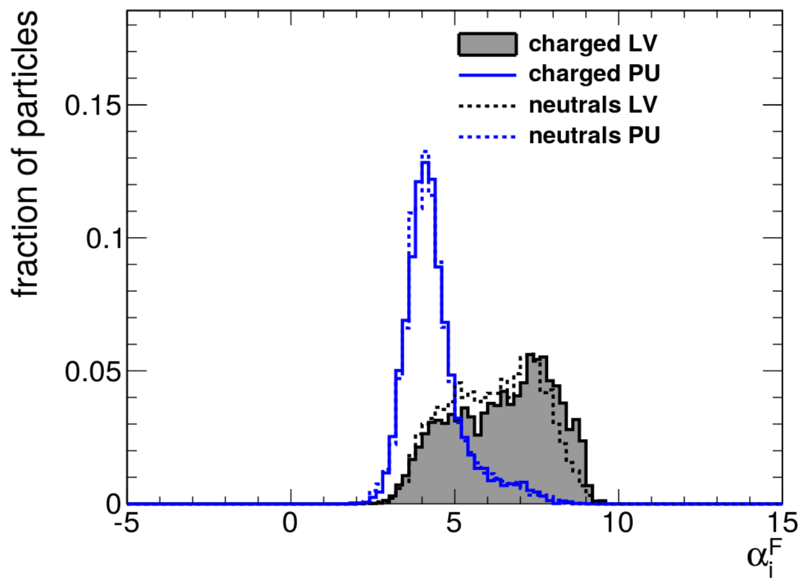
  Inside tracker region ($|\eta| < 2.5$), R_0=0.3

$$\alpha_i = \log \sum_{\substack{j \in Ch,PV \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij})$$

- In forward region ($|\eta| > 2.5$), R_0=0.3

$$\alpha_i = \log \sum_{j \neq i} \frac{p_{T,j}}{\Delta R_{ij}} \Theta(R_0 - \Delta R_{ij})$$

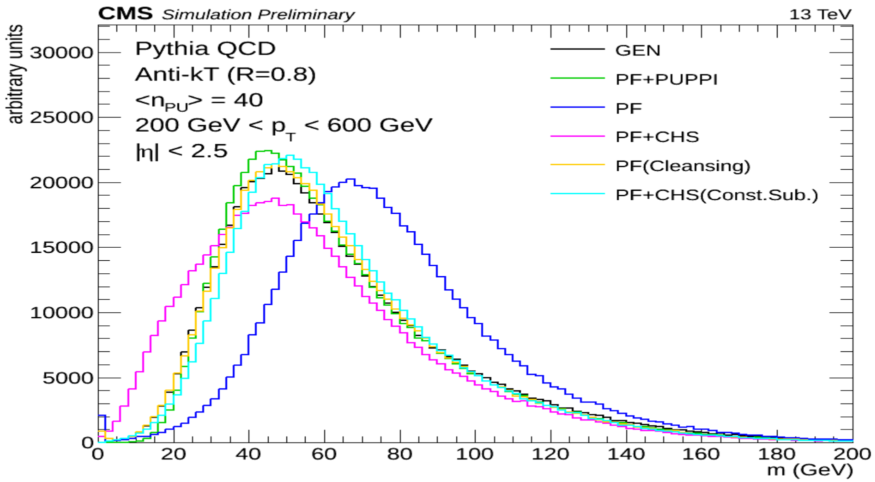
$$\alpha_i = \log \sum_{j \neq i} p_{T,j} \Theta(R_0 - \Delta R_{ij}).$$

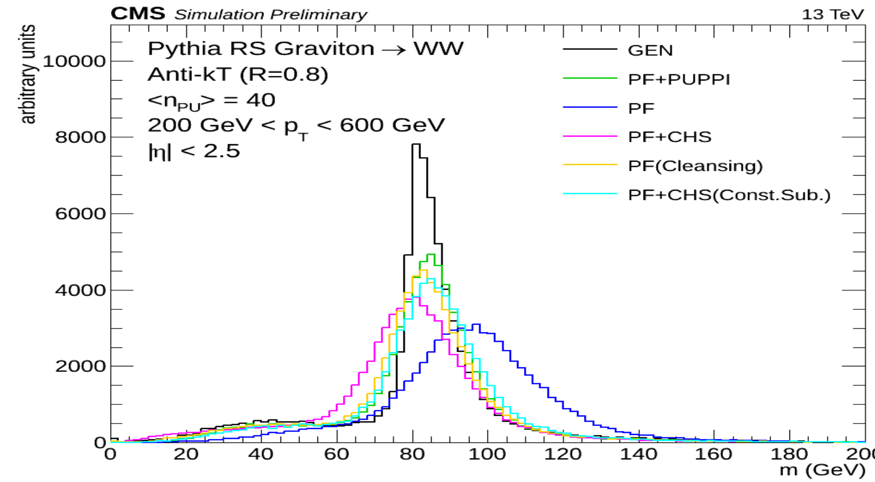


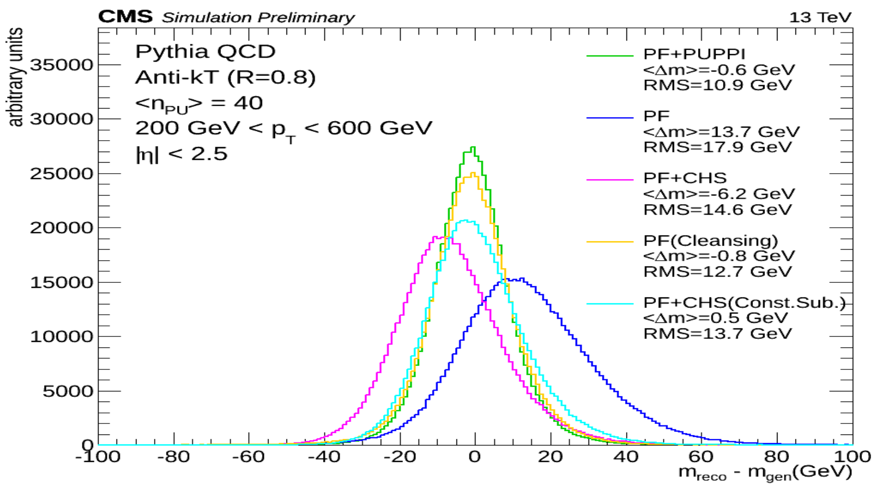
$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{RMS_{PU}^2}$$

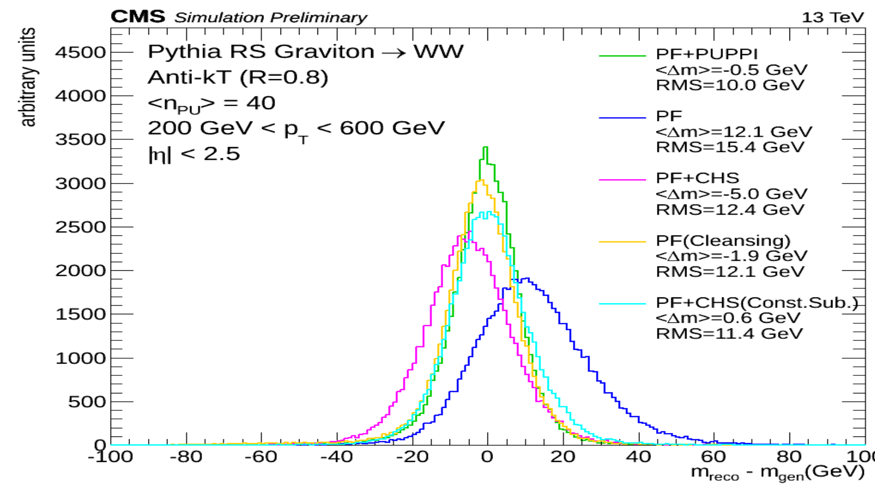
$$w_i = 1 - p_i$$

Performance









PUPPI demonstrates
the best performance

Grooming algorithms

- Grooming is intended to remove soft and wide-angle radiation from the jet
- Typically is used to reduce the overall jet mass of QCD (quark- and gluon-jets) while retaining the larger jet mass for jets originating from heavy particles (W, Z, H) bosons
- Help to reduce the pileup dependence of jet mass and shape observables
- Algorithms considered: **pruning, trimming and softdrop**
- Study is done with large R anti-kT jets (R=0.8)

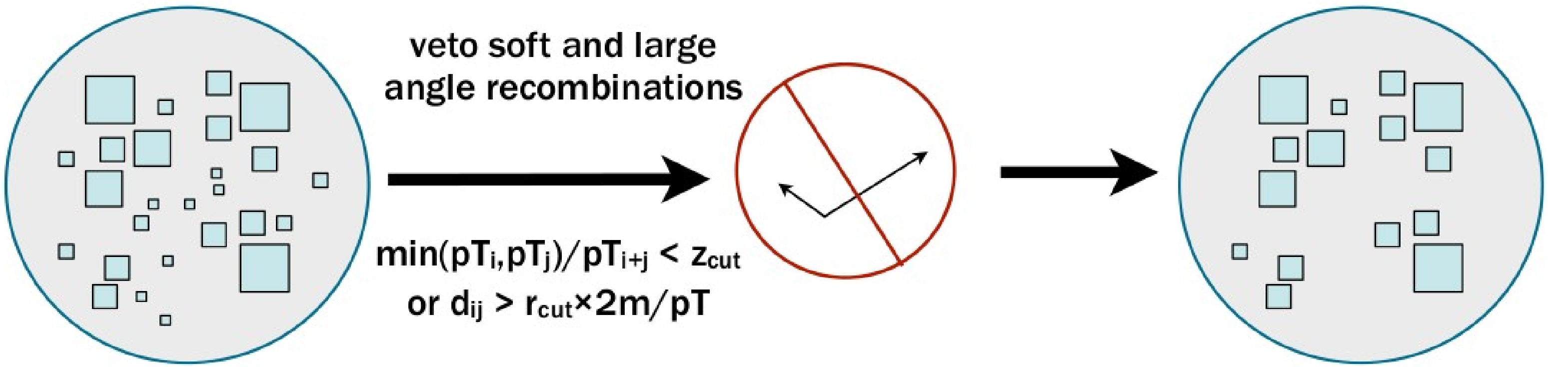
Pruning

- Recluster + veto soft and large-angle recombinations:

  - at each step the softer of two particles i and j to be merged is removed when the following conditions are met:

$$z_{ij} = \frac{\min(p_{Ti}, p_{Tj})}{p_{Ti} + p_{Tj}} < z_{\mathrm{cut}}$$
$$\Delta R_{ij} = \frac{2 \times r_{\mathrm{cut}} \times m_J}{p_T} > D_{\mathrm{cut}}$$

- Parameters of the algorithm: **z_{cut}, r_{cut}**



Trimming

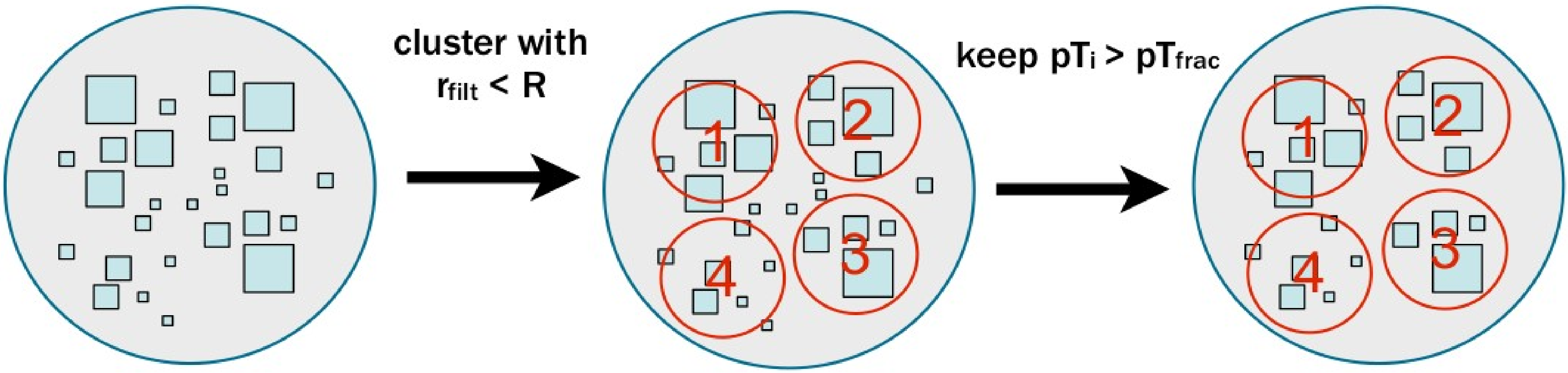
- Keep subjets over a dynamic p_T threshold:

  - recluster constituents with anti-kT into subjets with $R=R_{sub}$

  - keep if:

$$p_{T\,subjet} > f_{cut} \times p_{T\,jet}$$

- Parameters: $\mathbf{f_{cut}}$, $\mathbf{R_{sub}}$



Softdrop

- Decluster the jet recursively removing soft and wide angle radiation from the jet

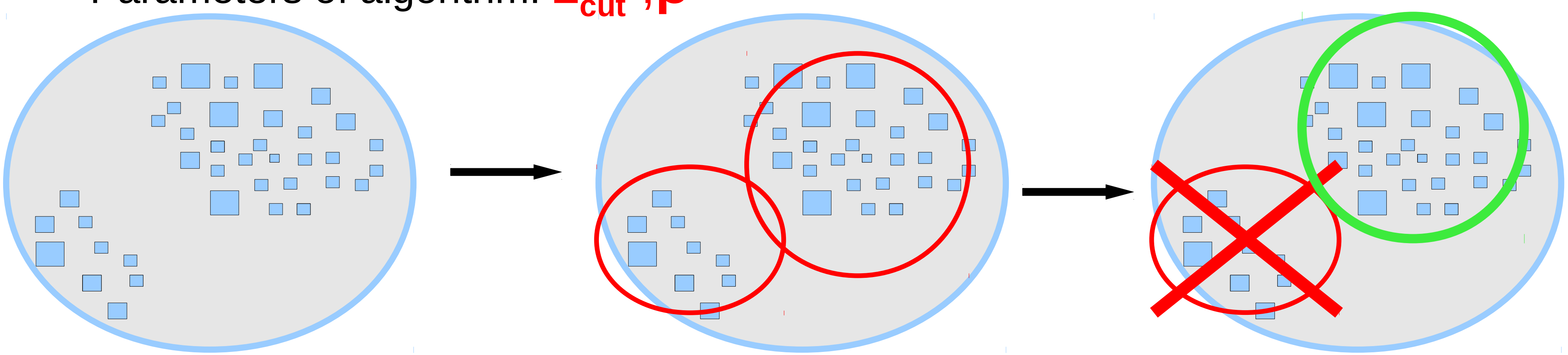
  - Decluster with CA

  - at each step for subjet j1 and j2, check the condition:

$$\frac{\min(p_{Tj1}, p_{Tj2})}{p_{Tj1} + p_{Tj2}} > z_{\mathrm{cut}} \times \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

  If this is met, keep the jet otherwise drop the softer constituent and reiterate declustering

- Parameters of algorithm: **z_{cut} ,β**



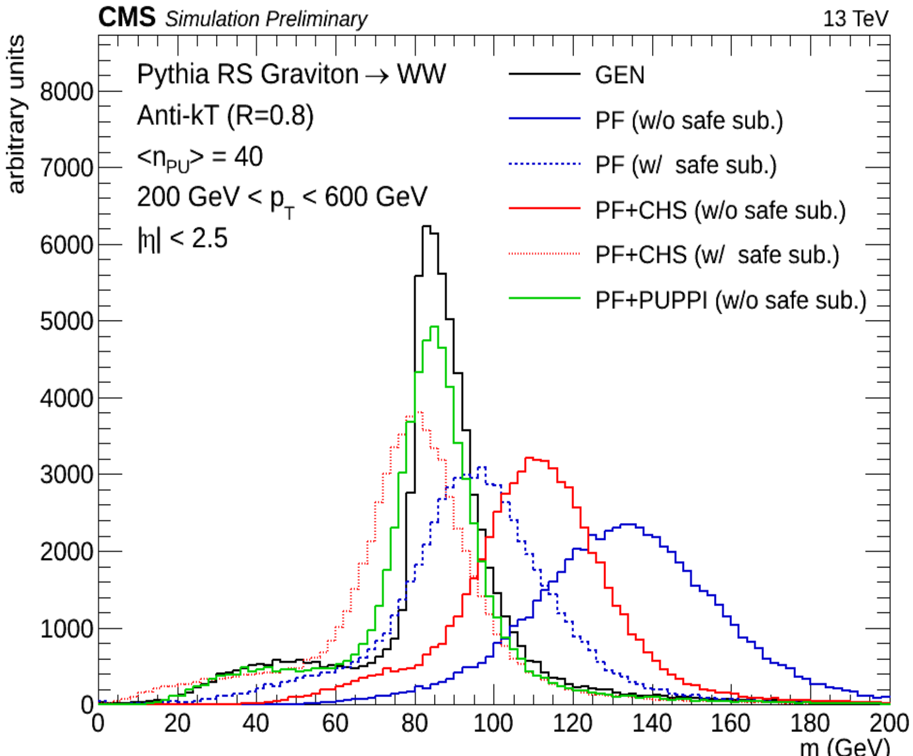
Safe subtraction

- Safe area correction is applied:

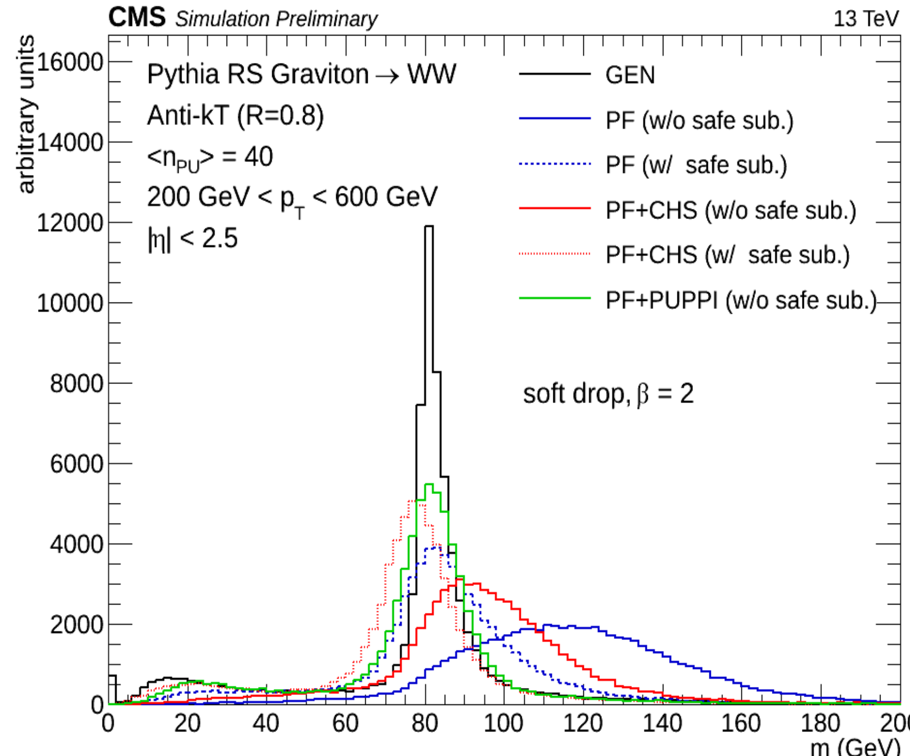
$$p^{\mu}_{sub} = p^{\mu} - \rho A^{\mu} - \rho_m A^{\mu}_m$$

- Term ρ_m is taken into account because of the low p_T pileup jets have non-negligible mass compared to their p_T :
- Basically that leads to the fact that low p_T pileup jets don't get negative masses.

Safe subtraction



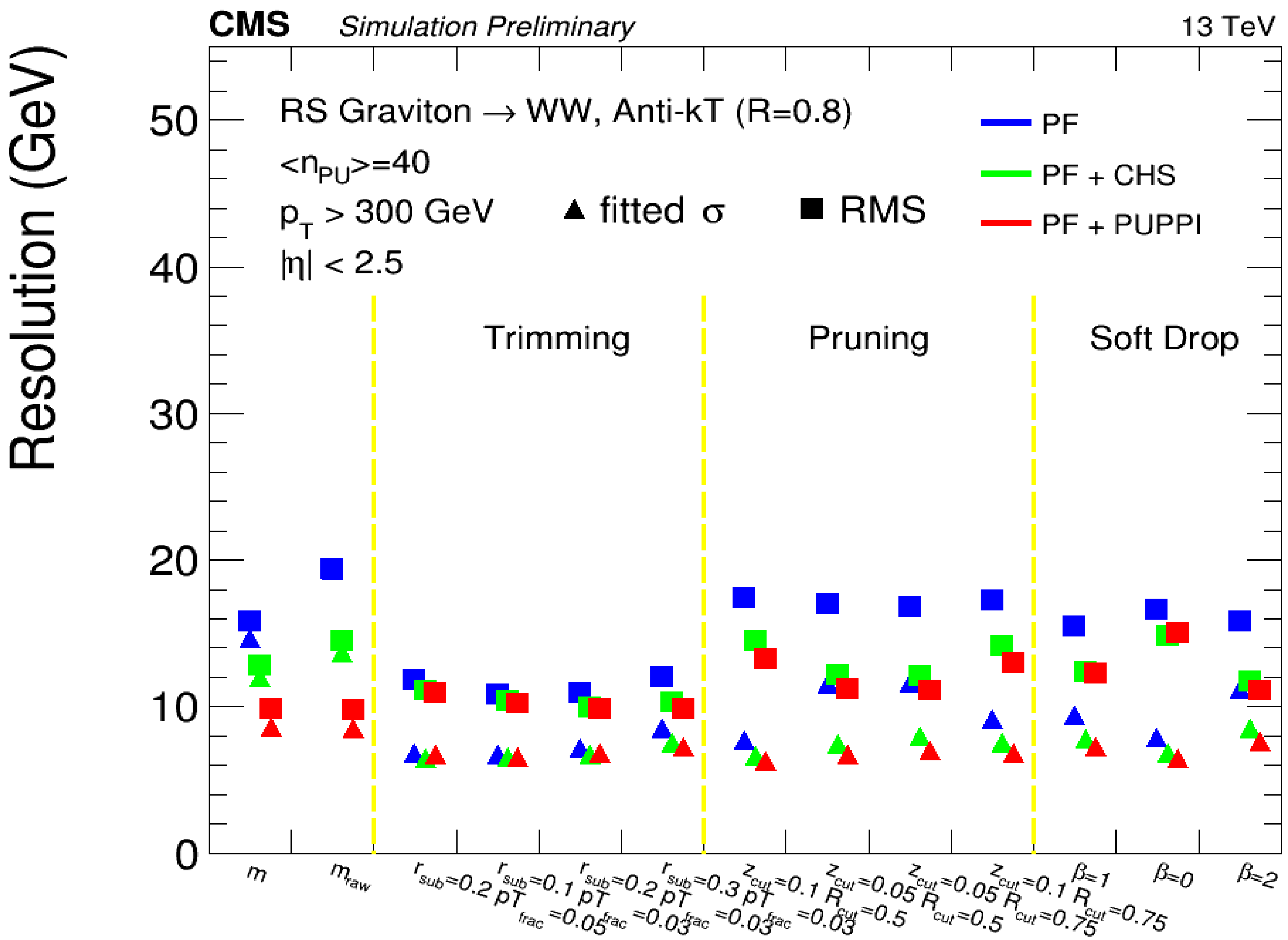
Ungroomed



Groomed

Safe subtraction moves peak on the mass distribution closer to 80 GeV and improves resolution

Mass response resolution: summary

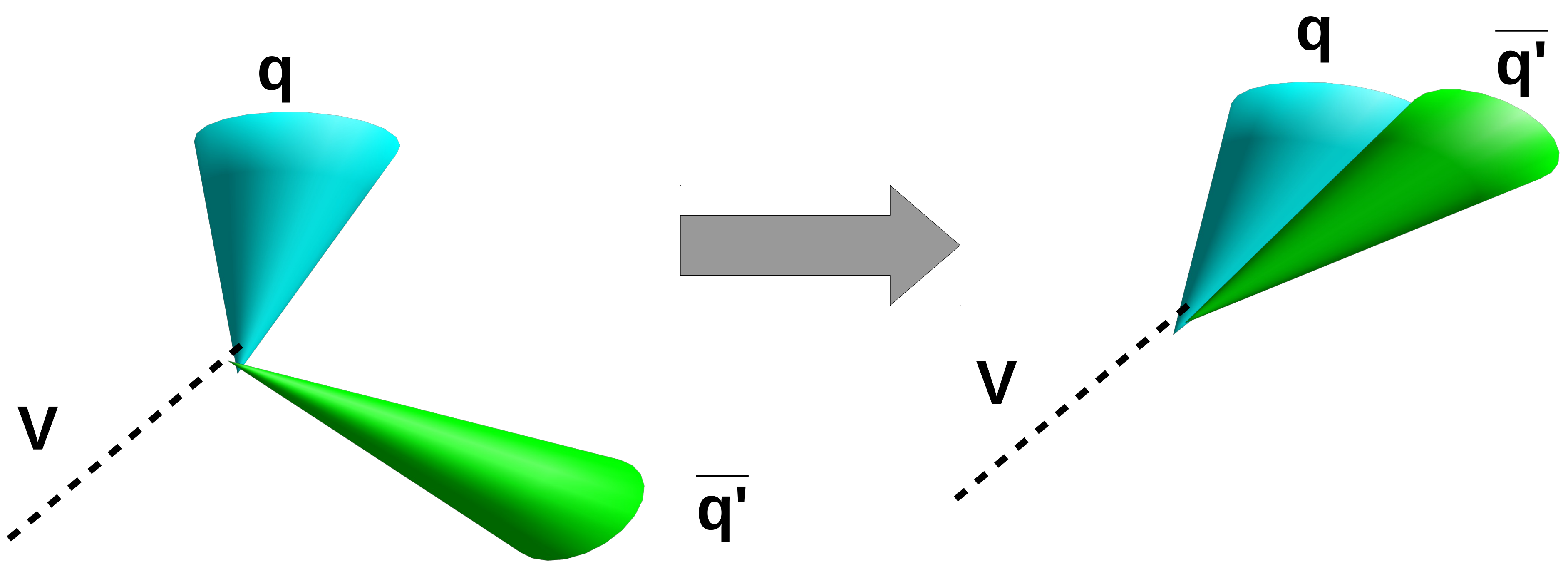


Overall PUPPI demonstrates the best performance for resolution

Boosted regime

- Higher energies → jets start to merge
- Cannot find to jets anymore
- Have to look into the **substructure** of the jet to find 2 subjets =>
- **V-tagging algorithms**

$$\Delta R_{qq} \sim \frac{M_V}{p_{T,V}}$$

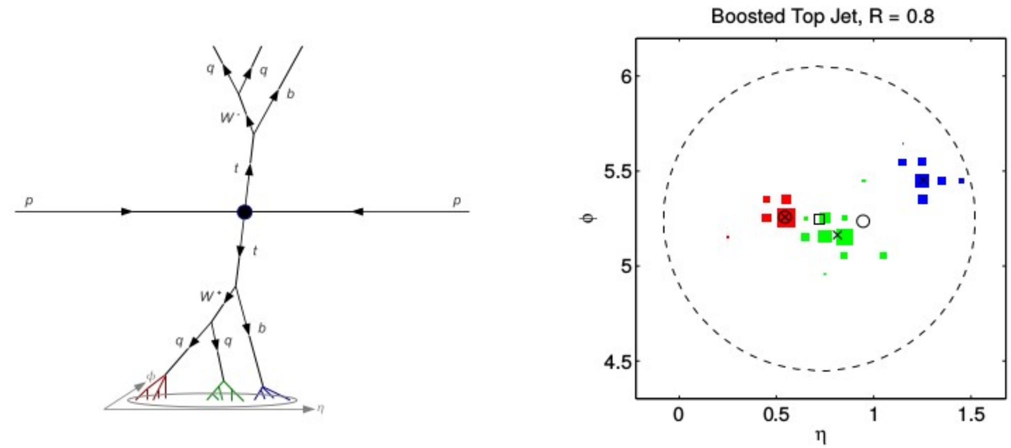


N-subjettiness

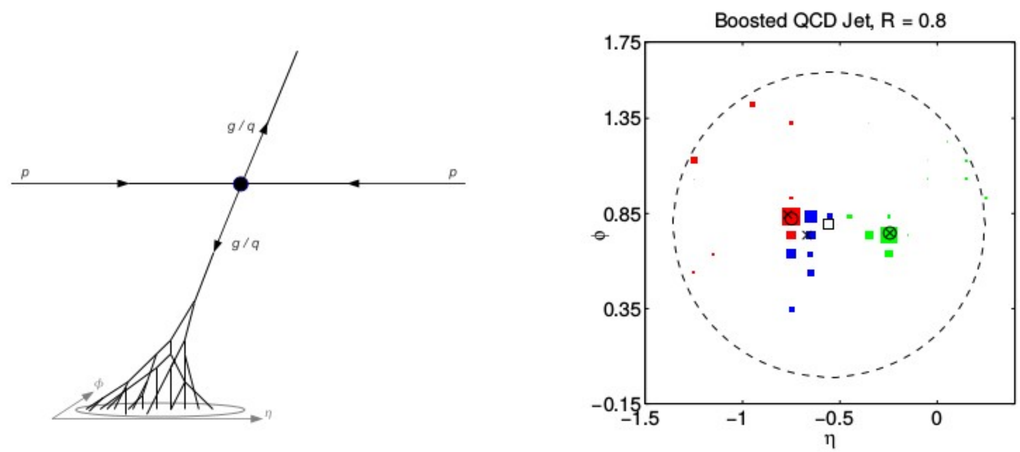
- Quantifies to what degree jet can be regarded as a jet composed of N jets

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} min\{\Delta R_{1,k}, ..., \Delta R_{N,k}\}$$

$$d_0 = \sum_k p_{T,k} R_0$$

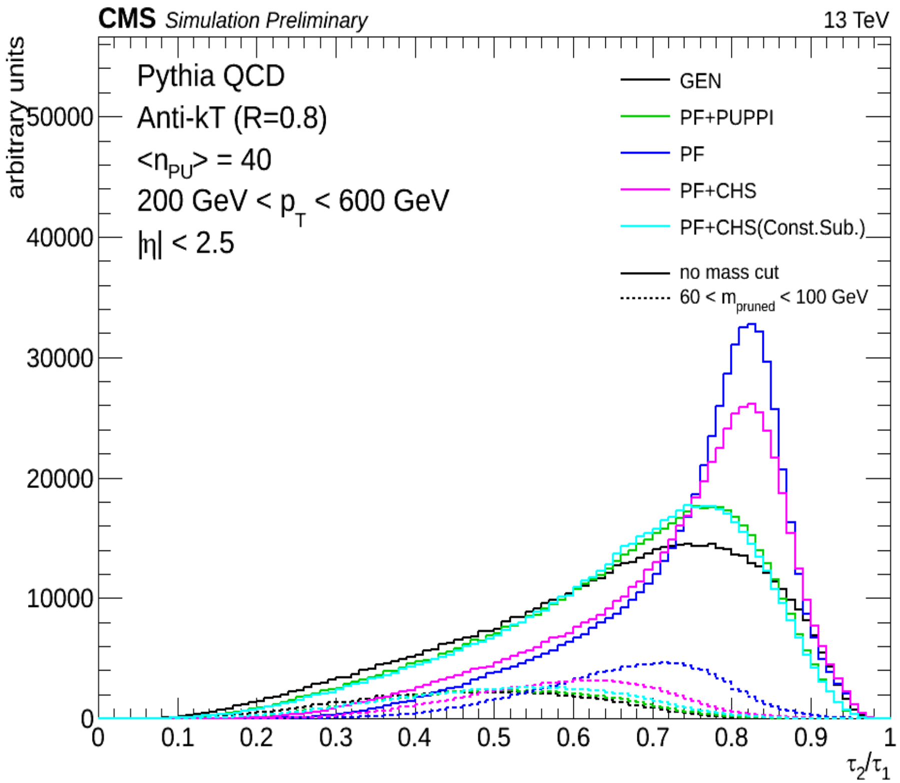


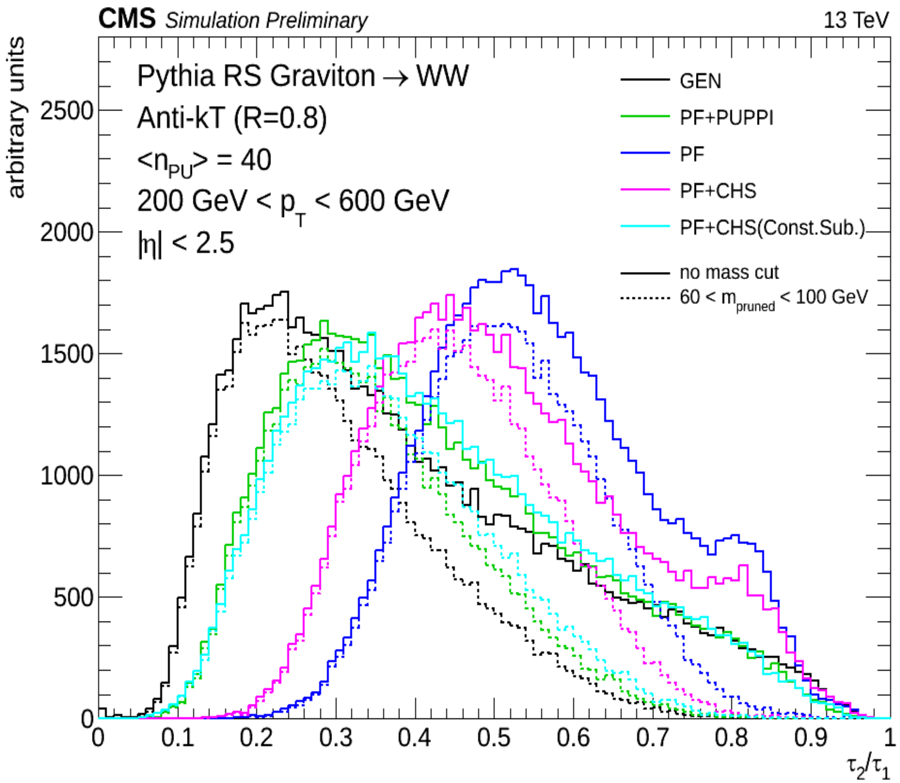
(a) (b)



τ_{21} performance

$$\tau_{21} = \tau_2/\tau_1$$

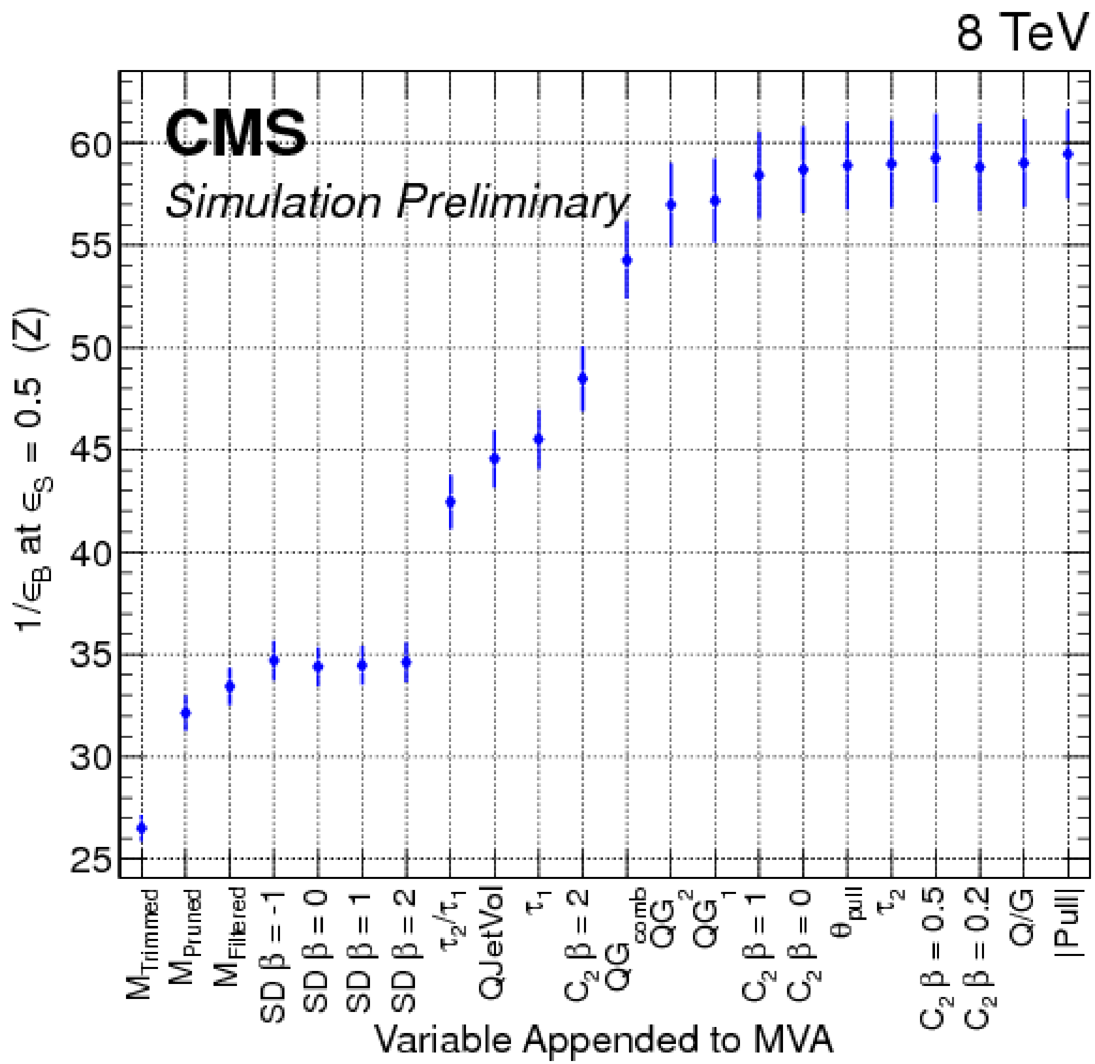




PUPPI demonstrates the best performance

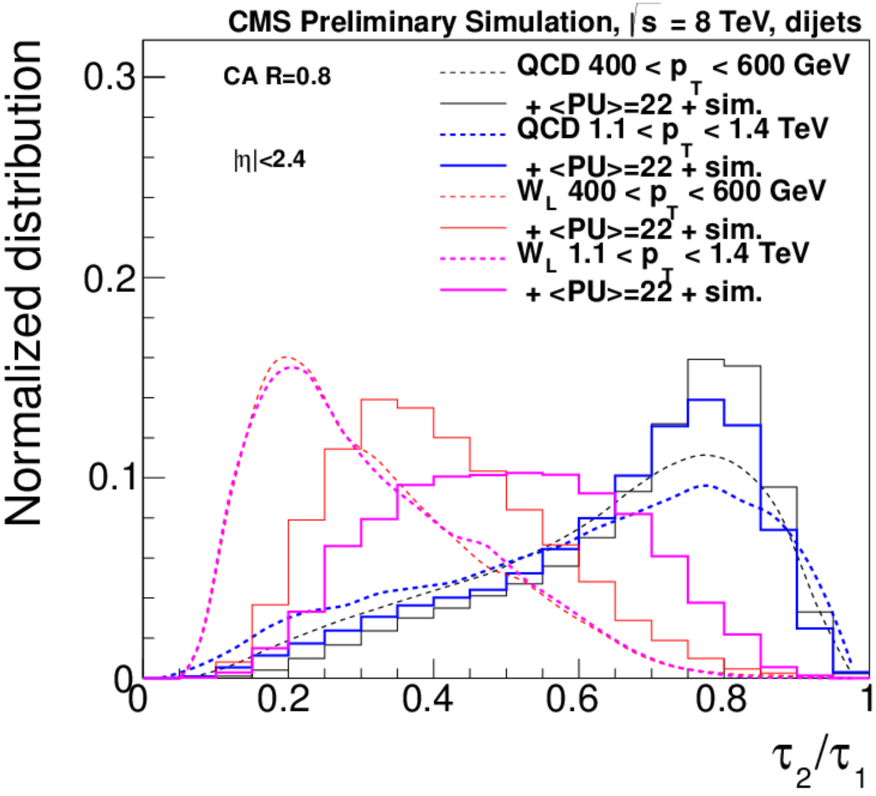
Performance studies

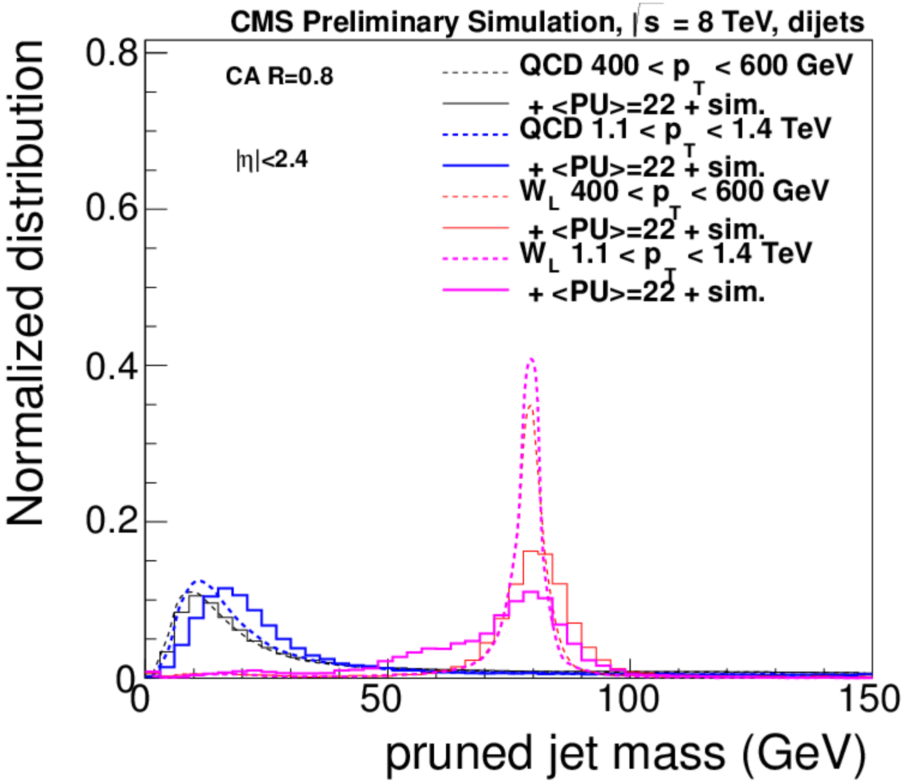
Signal: ttbar
Background: Z+jets



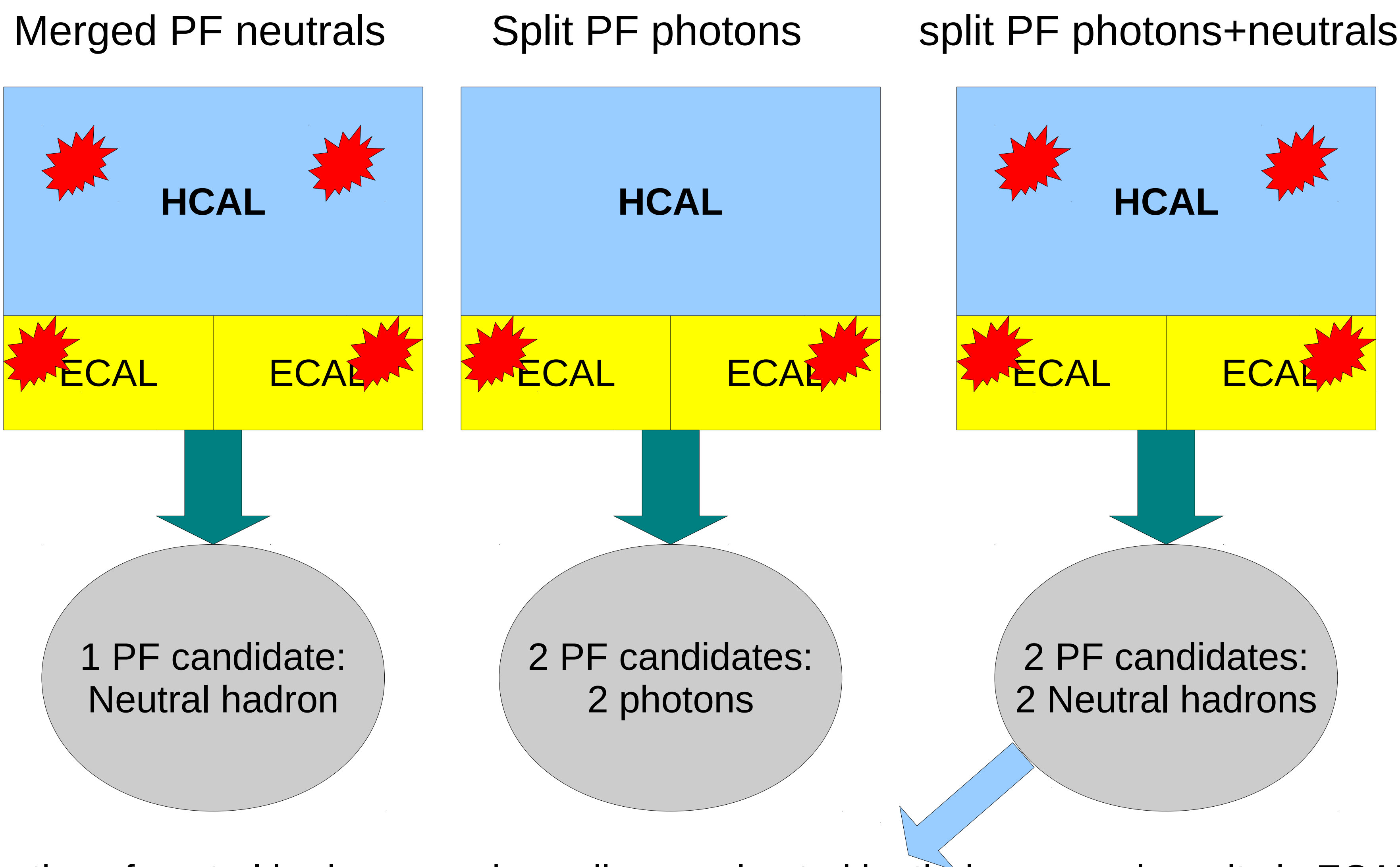
Performance at high transverse momenta

Resolution degrades dramatically with p_T: JME-13-006



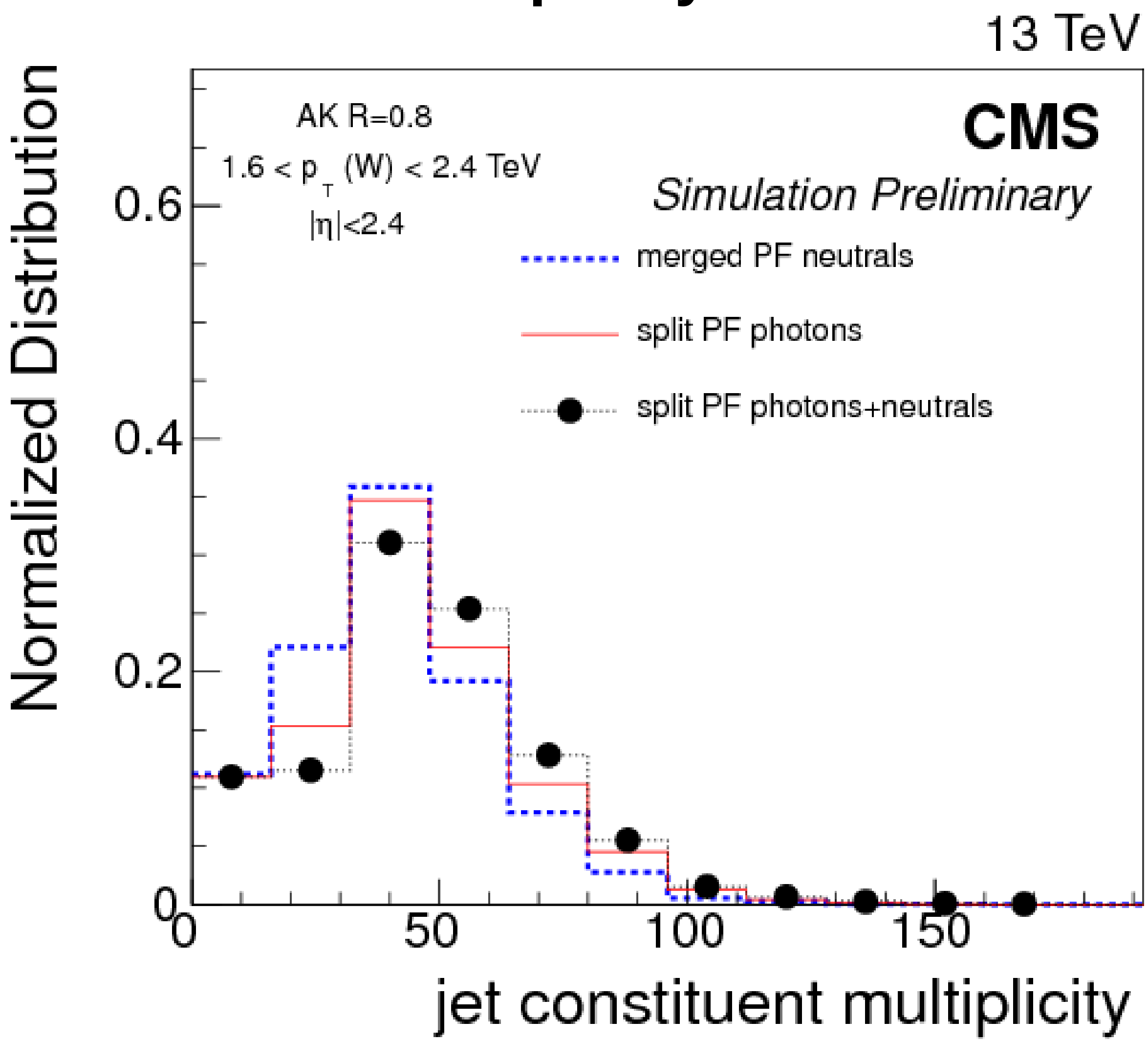


Particle Flow modification



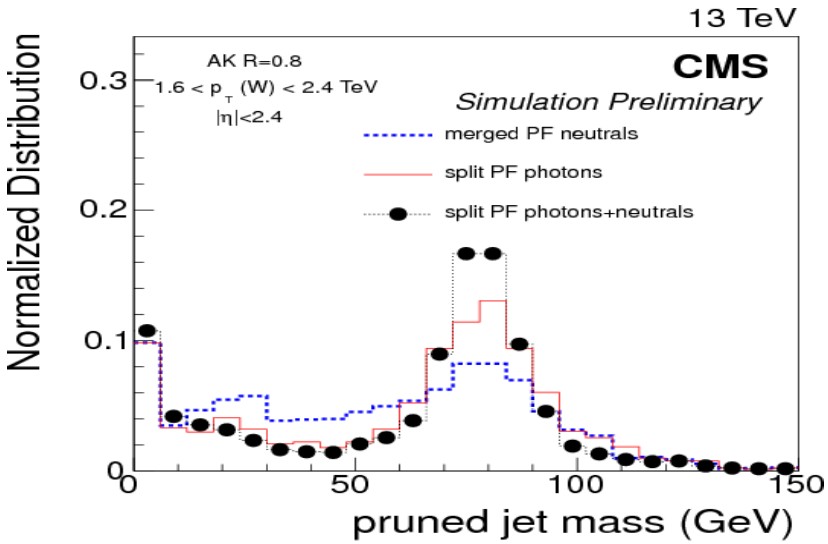
Direction of neutral hadrons can be well approximated by their energy deposits in ECAL (ECAL granularity is 5 times finer than HCAL in η and φ).

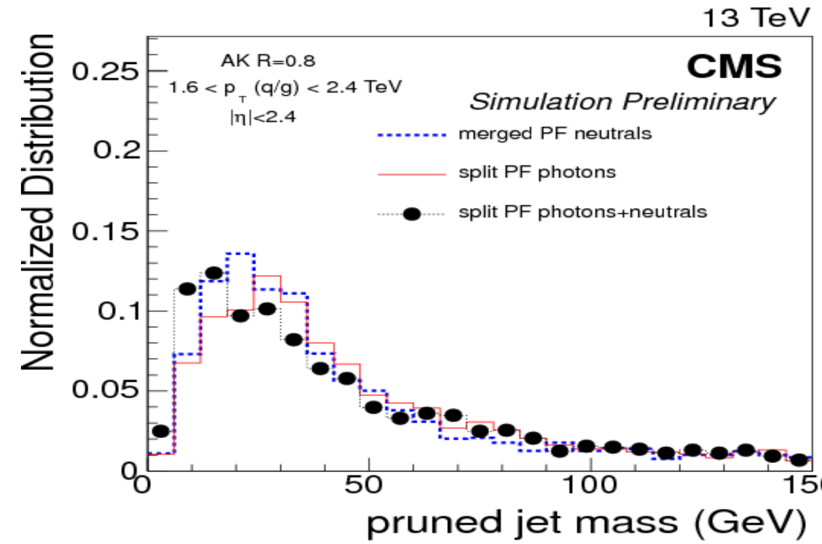
Jet constituent multiplicity

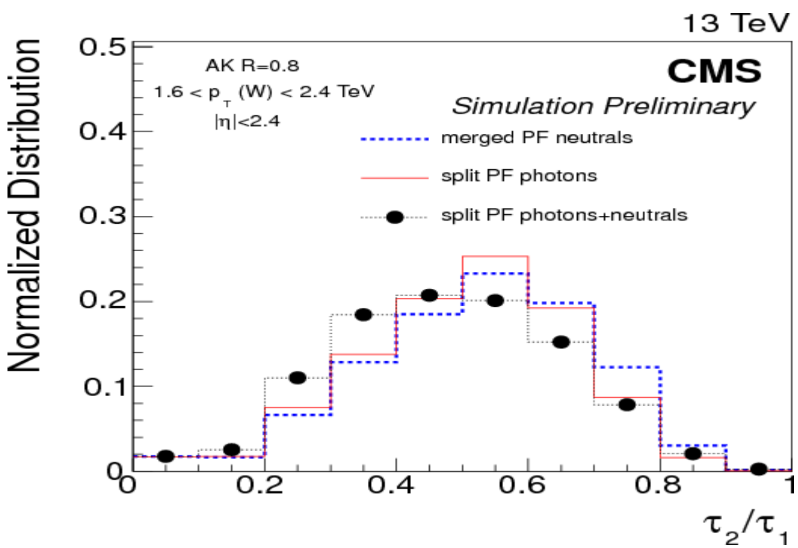


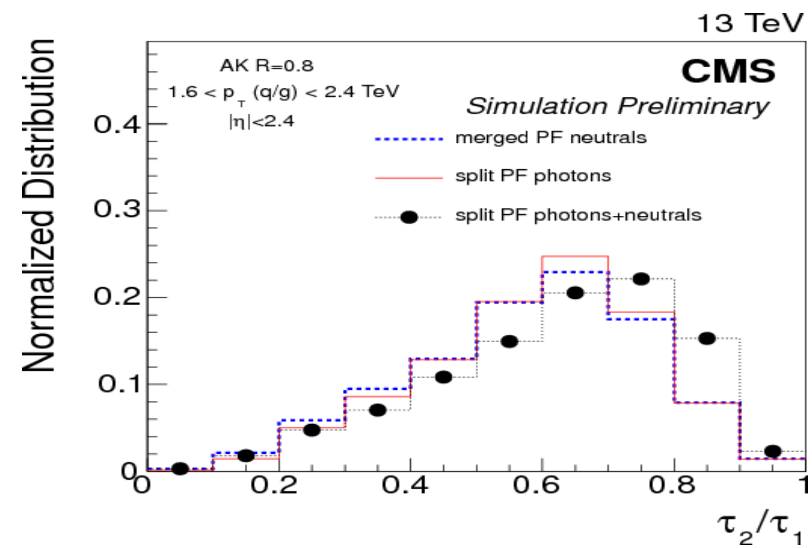
This increase in number of particles has a large impact on the performance of W-tagging

Improvement of W-tagging performance









Significant improvement can be seen!

Conclusions

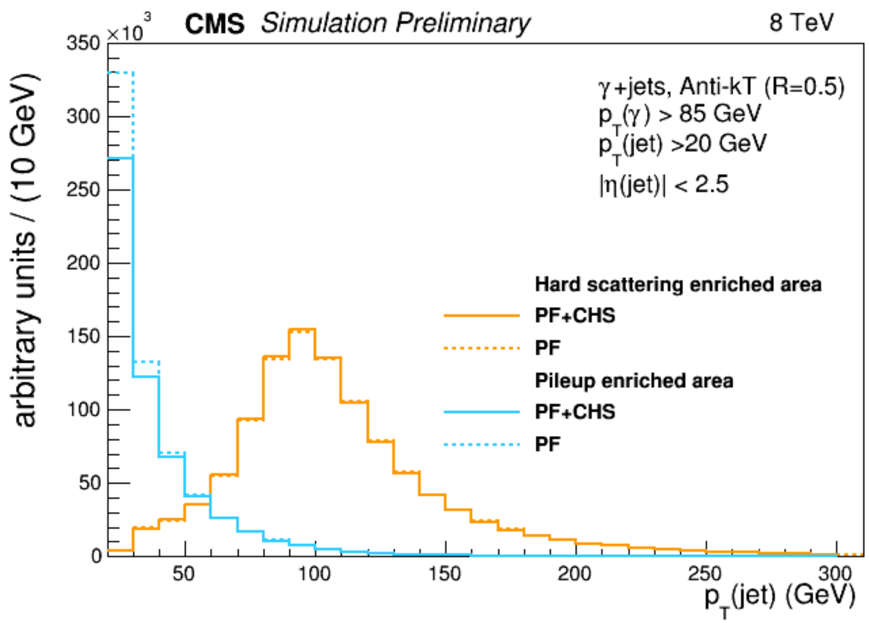
- Performance of various pileup mitigation tools on jets was studied with focus on preparation for Run II (PU ~ 40).
- PUPPI performs better than CHS.
- Grooming algorithms help to improve reconstruction of heavy particle decays.
- Combination of pileup mitigation tools and grooming algorithms gives the best result.
- PUPPI is effective in reducing the pileup jet rate and shows very good performances in terms of mass resolution, shape reconstruction and pileup stability
- Effectiveness of V-tagging was expanded up to ~2 TeV of boson p_T.
- Analysises are documented in:

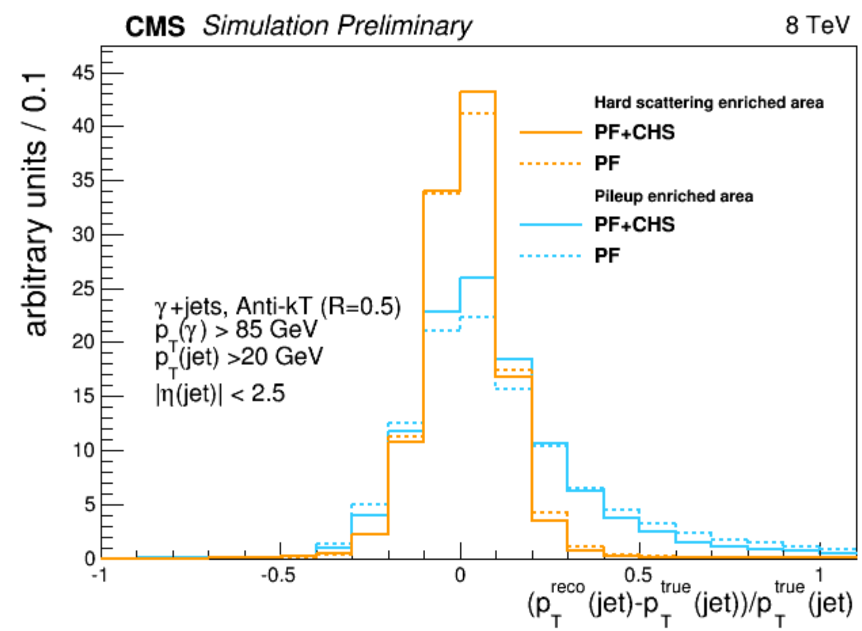
  CMS-PAS-JME-14-001 "Pileup Removal Algorithms"

  CMS-PAS-JME-14-002 "V Tagging Observables and Correlations"

Backup slides

Charged Hadron Subtraction (CHS): MC based study





Jet p_T response is peaking more sharply when CHS is applied

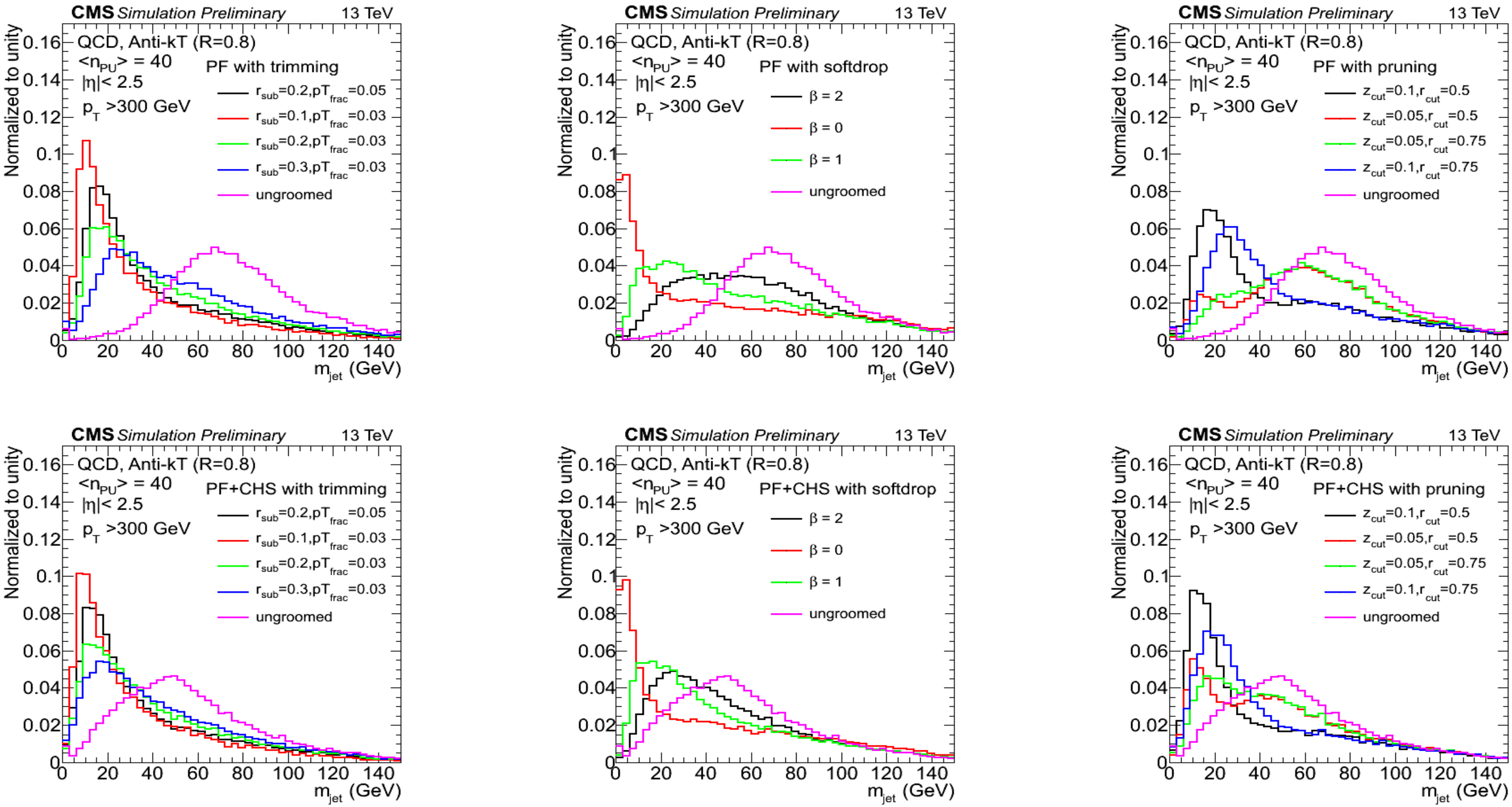
Parameters considered

- List of parameters :

grooming algorithm	parameters
Pruning	$z_{\rm cut} = 0.1,\ r_{\rm cut} = 0.5$
	$z_{\rm cut} = 0.05,\ r_{\rm cut} = 0.5$
	$z_{\rm cut} = 0.1,\ r_{\rm cut} = 0.75$
	$z_{\rm cut} = 0.05,\ r_{\rm cut} = 0.75$
Trimming	$r_{\rm sub} = 0.2,\ _{\rm frac} = 0.05$
	$r_{\rm sub} = 0.2,\ _{\rm frac} = 0.03$
	$r_{\rm sub} = 0.1,\ _{\rm frac} = 0.03$
	$r_{\rm sub} = 0.3,\ _{\rm frac} = 0.03$
Soft drop	$z_{\rm cut} = 0.1, \beta = \text{-}1$
	$z_{\rm cut} = 0.1, \beta = 0$
	$z_{\rm cut} = 0.1, \beta = 1$
	$z_{\rm cut} = 0.1, \beta = 2$

- For **PF and PF+CHS** inputs safe subtraction is applied.

Mass distributions: QCD jets for PF and PF+CHS



Mass resolution

- Fit with Gaussian is done in the range **mean ± 1*RMS**
- Truncated RMS is calculated in the range **μ ± 3σ**