Particle flow jets and boosted top-tagging for top physics

Beyond the first 13 TeV top measurements

Regina Moles-Valls

19-21 Abril 2016, CERN

thanks to PFlow and Boosted top-tagging people

Outline

- Particle flow (PFlow) jets in ATLAS
  - Brief introduction to the PFlow algorithm
  - Performance studies with full 8 TeV data
  - Prospects for systematic uncertainties
- Boosted top-tagging algorithms in ATLAS
  - Main top-tagging algorithms
  - Performance studies with early 13 TeV data
  - Current and future developments

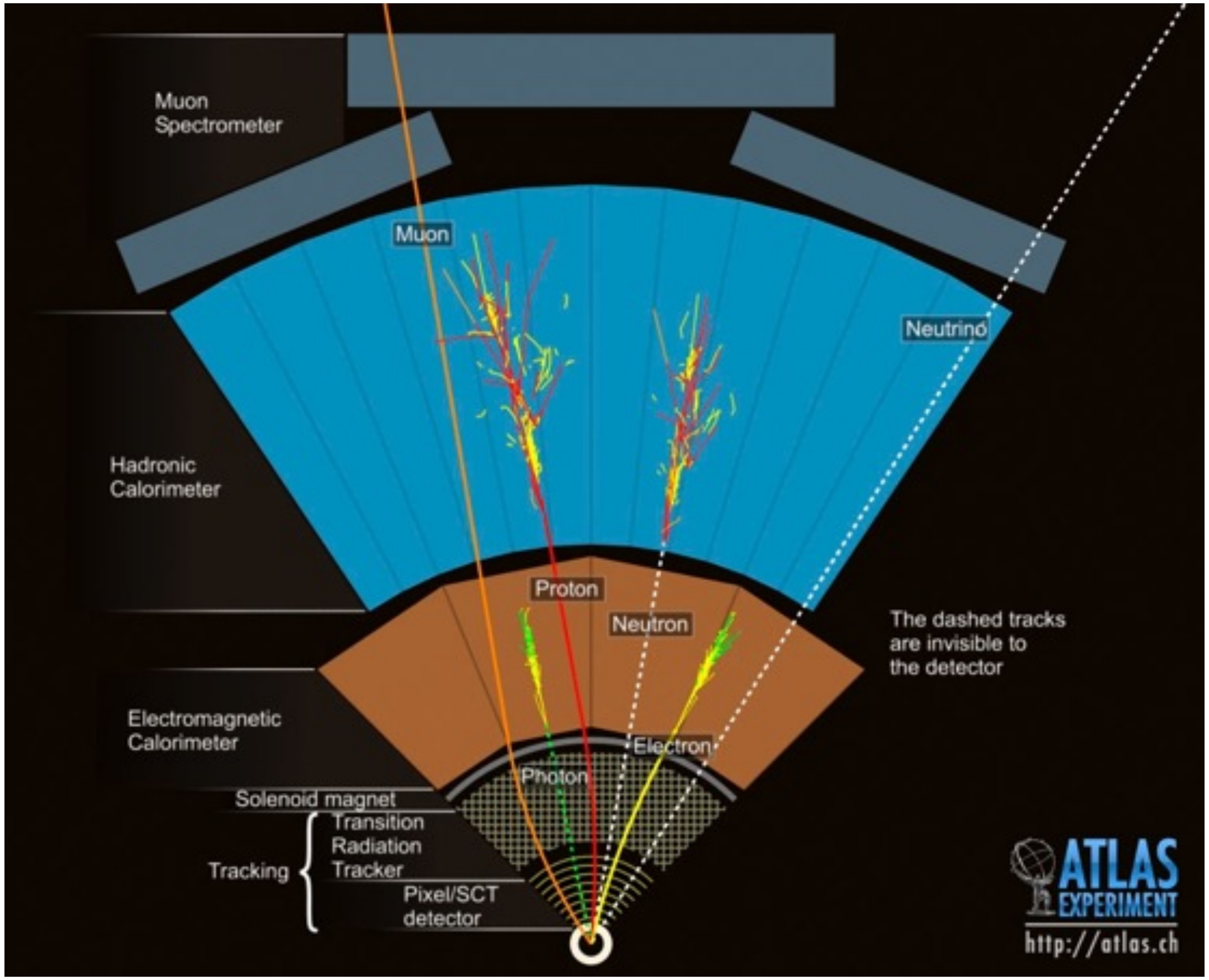
Introduction to Particle Flow

Particle Flow algorithms try to follow the path of the particles through the detector.

Main goal is to improve the energy resolution of the hadronic objects

How to do it? combining the information from different sub-detectors

➡ Emphasise the role of the tracker in jet physics.

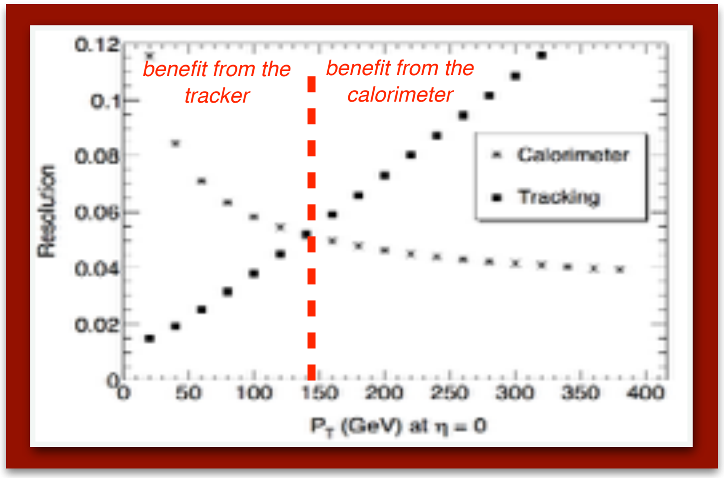


ID tracker

- Better momentum resolution at low p_T.
- Reconstruction of soft particles (0.4 GeV).
- Better angular resolution.
- Vertex information useful to mitigate the charged pileup contribution.

Calorimeter

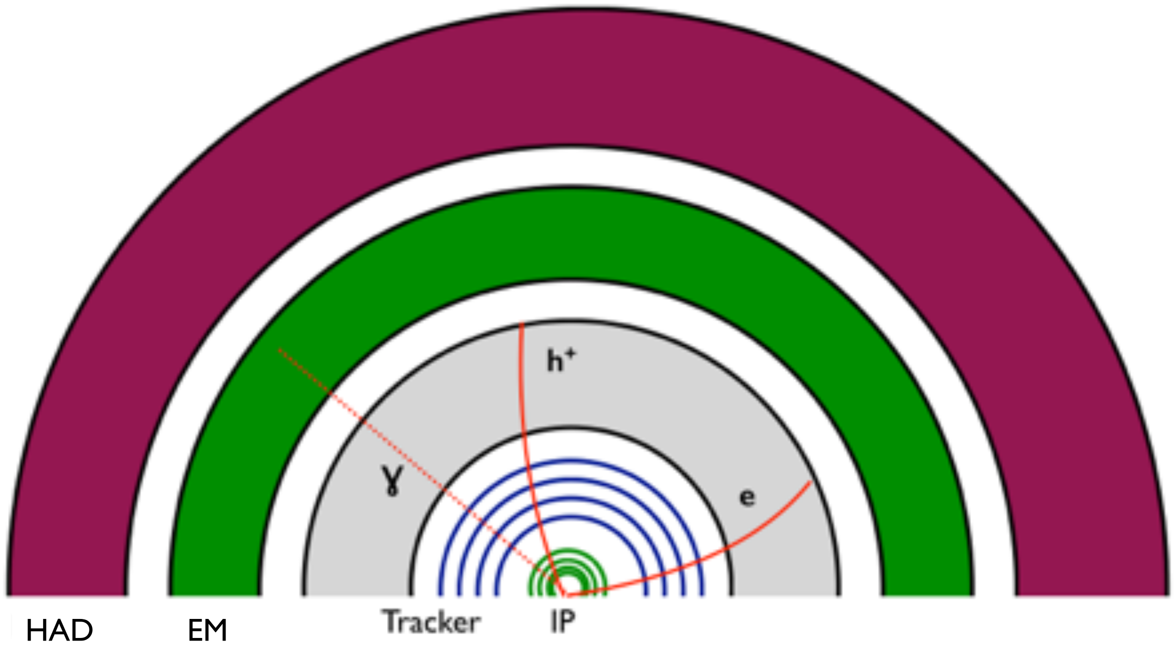
- Reconstruction of neutral particles.
- Better energy resolution at high p_T.



Introduction: Particle Flow principle

How does Particle Flow work?

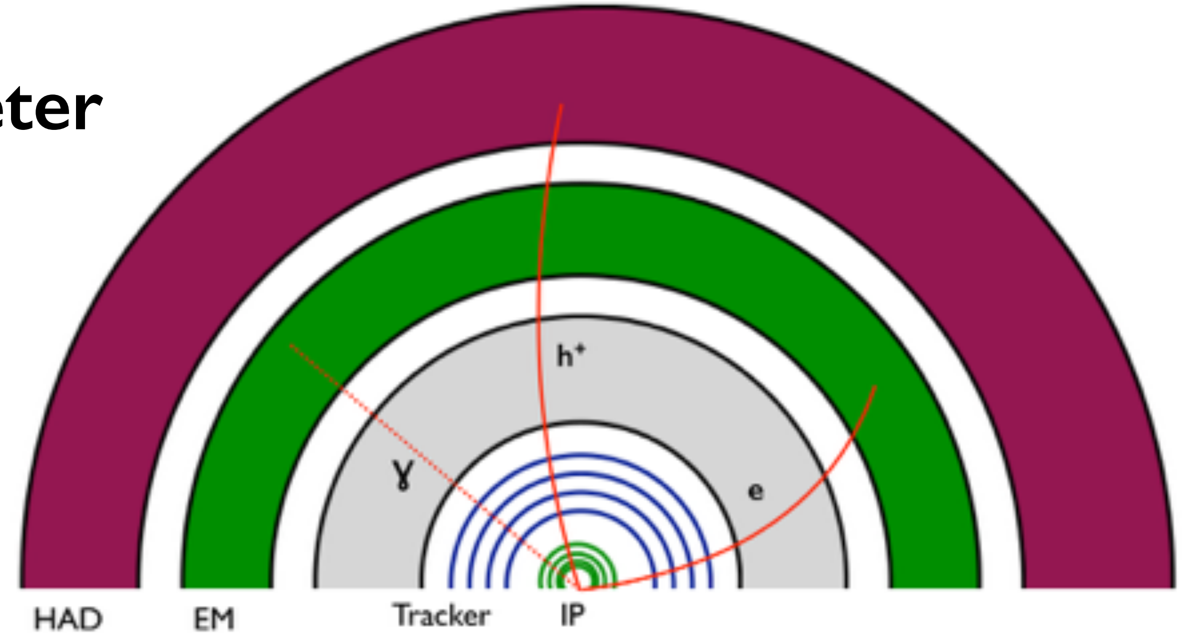
- Track reconstruction in the ID



Introduction: Particle Flow principle

How does Particle Flow work?

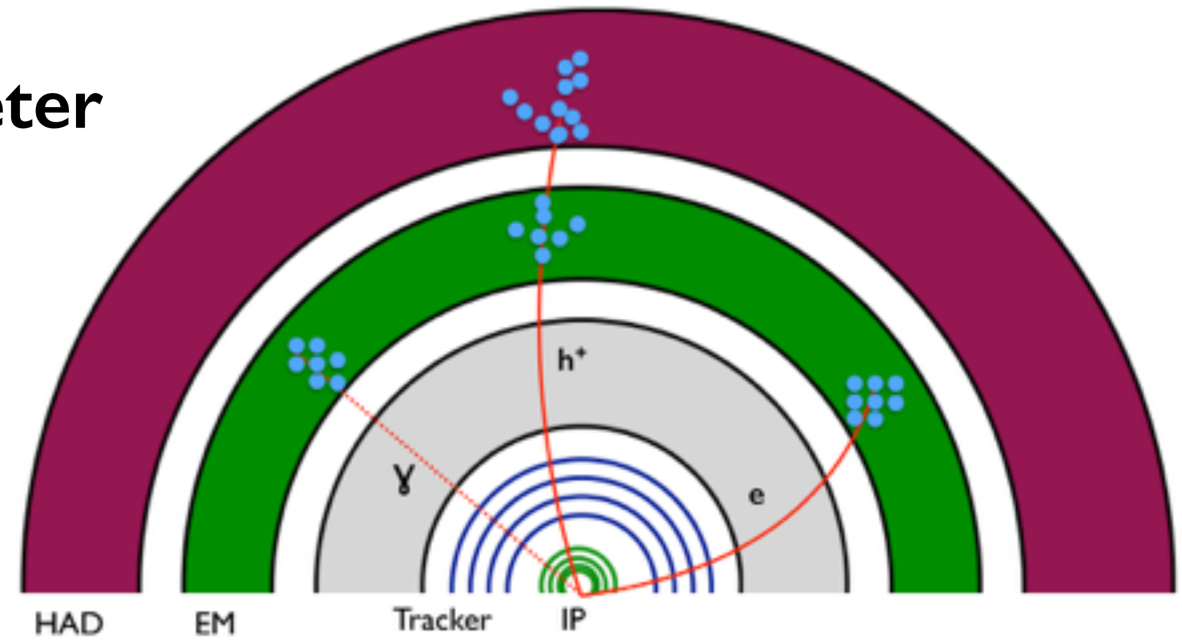
- Track reconstruction in the ID
- Extrapolate the tracks to the Calorimeter



Introduction: Particle Flow principle

How does Particle Flow work?

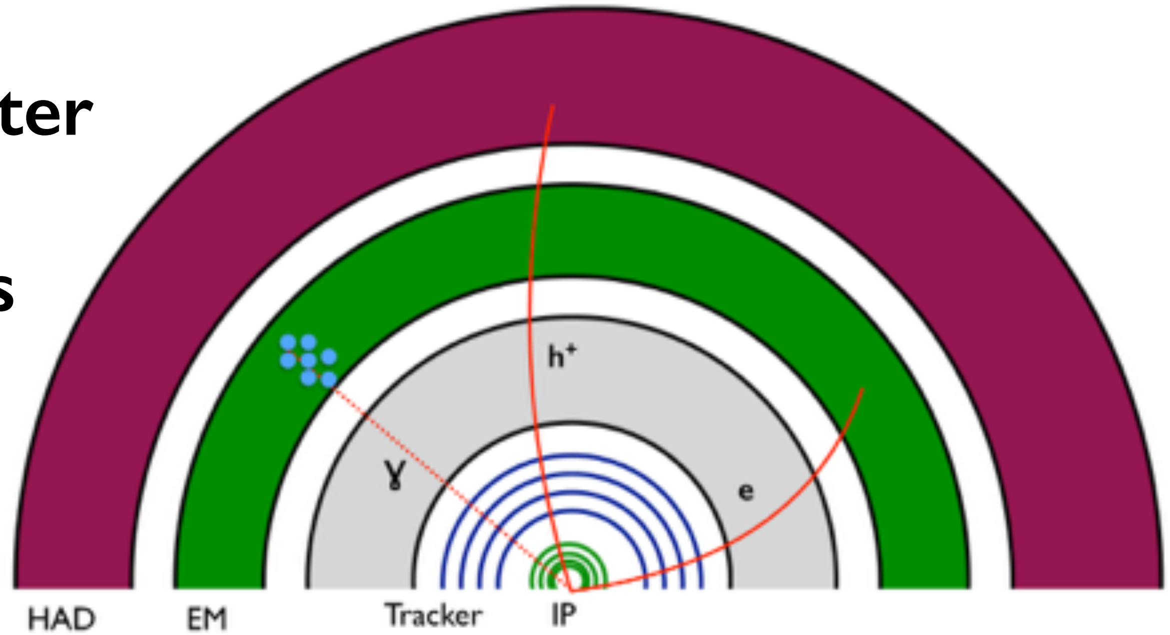
- Track reconstruction in the ID
- Extrapolate the tracks to the Calorimeter
- Match the tracks to the clusters



Introduction: Particle Flow principle

How does Particle Flow work?

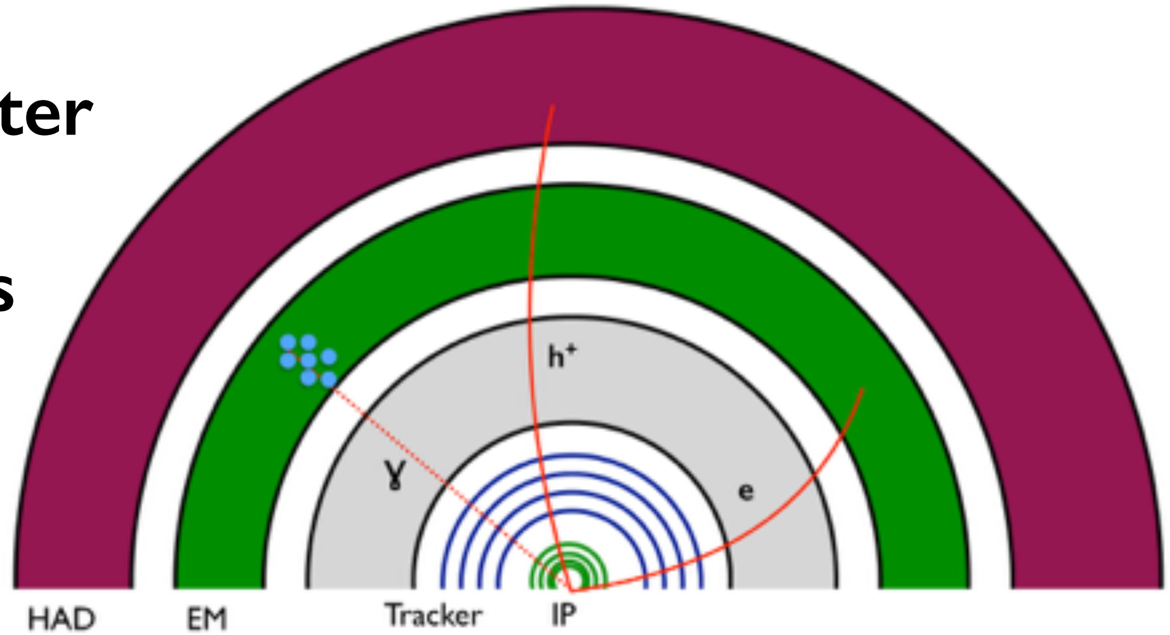
- Track reconstruction in the ID
- Extrapolate the tracks to the Calorimeter
- Match the tracks to the clusters
- Remove clusters from charged particles



Introduction: Particle Flow principle

How does Particle Flow work?

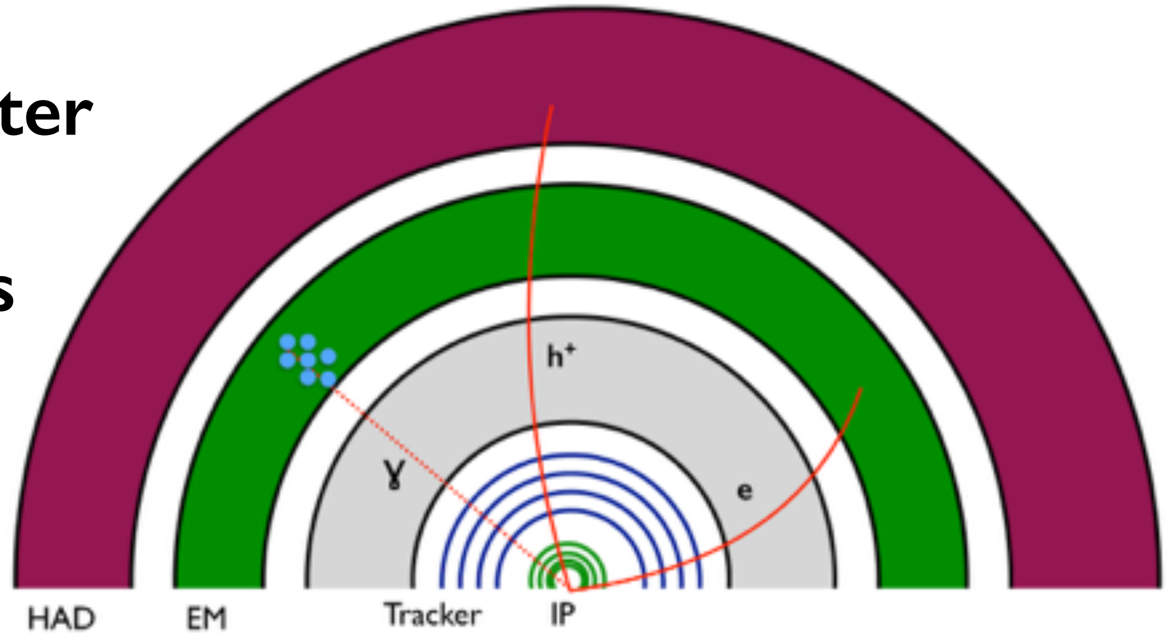
- Track reconstruction in the ID
- Extrapolate the tracks to the Calorimeter
- Match the tracks to the clusters
- Remove clusters from charged particles
- Finally are kept:
    - tracks and
    - clusters from neutral particles



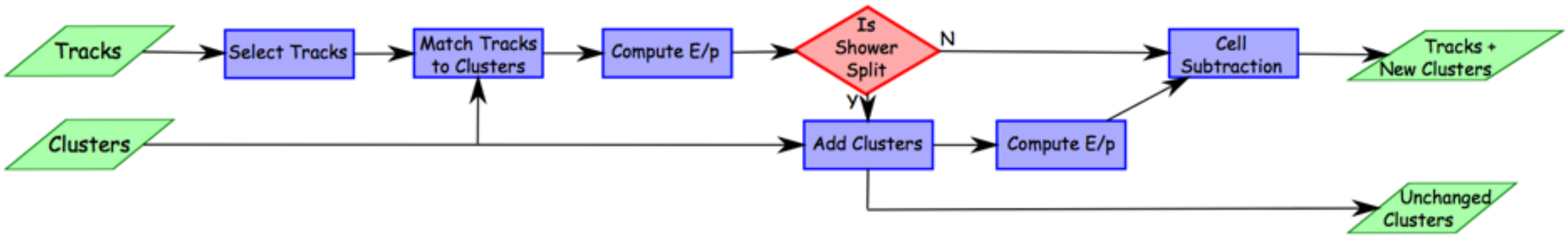
Introduction: Particle Flow principle

How does Particle Flow work?

- Track reconstruction in the ID
- Extrapolate the tracks to the Calorimeter
- Match the tracks to the clusters
- Remove clusters from charged particles
- Finally are kept:
  - tracks and
  - clusters from neutral particles



"eflowRec" algorithm in ATLAS



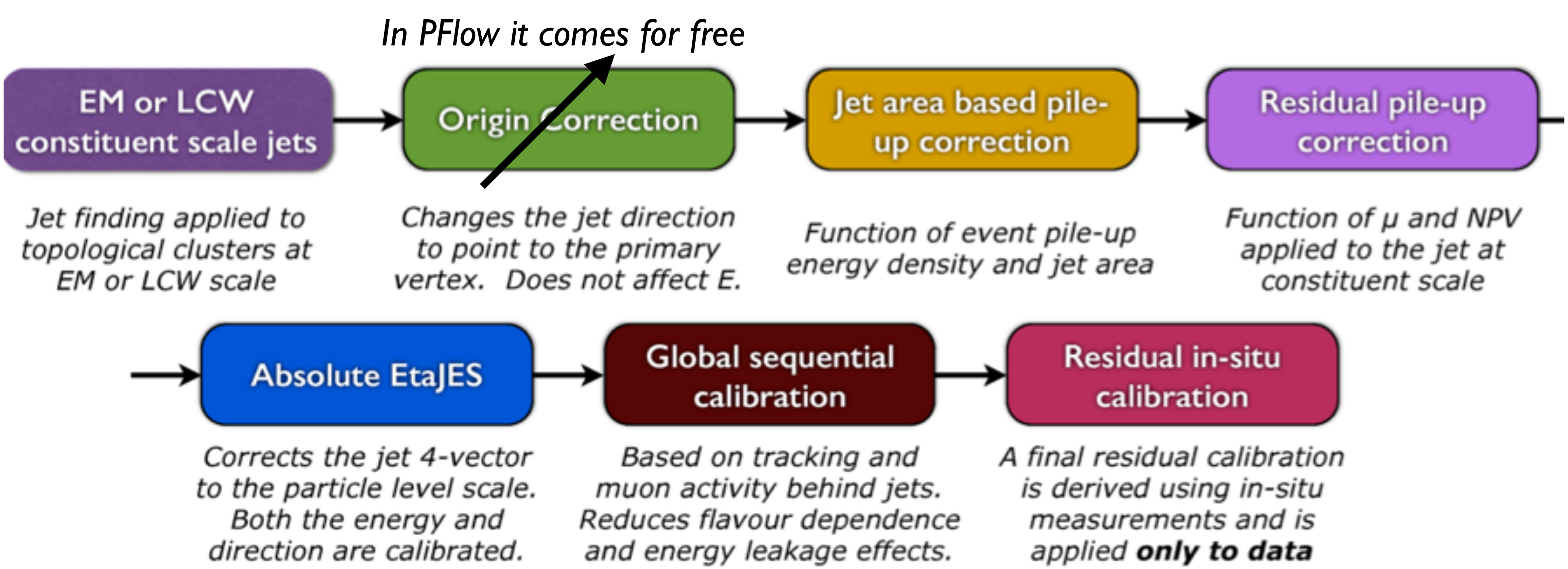
It is optimised to subtract all of a charged single particle E without removing E deposited by other particles.

Particle flow calibration

PFlow jet calibration

- PFlow jets are reconstructed using AntiKt with R=0.4:
  - From positive energy topo-clusters and the selected tracks associated to PV

- Calibration closely follows the standard procedure

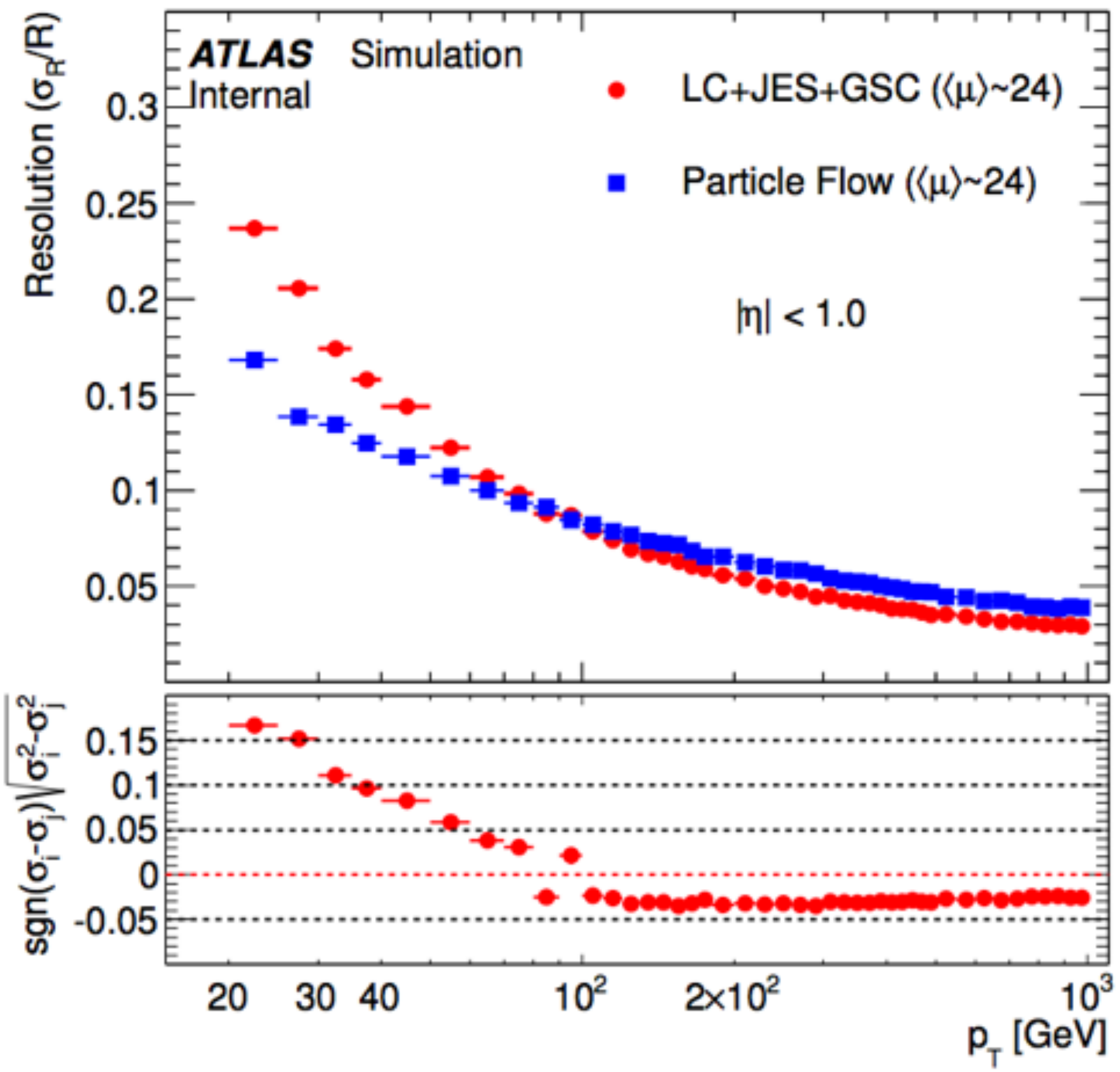


Particle flow performance studies

Results at $\sqrt{s}$ = 8 TeV

Jet P_T resolution

- The largest expected benefit from PFlow is to improve the jet resolution at low p_T.



- **<u>Better resolution p_T < 90 GeV:</u>**
  - Improved scale for low p_T hadrons.
  - Intrinsic pile-up suppression.
- Worse resolution at high p_T:
  - Charged shower subtraction (CSS) worse with dense environments.
  - New subtraction algorithm in Rel.20.7 (use the energy of calorimeter jets at high energy density environments).

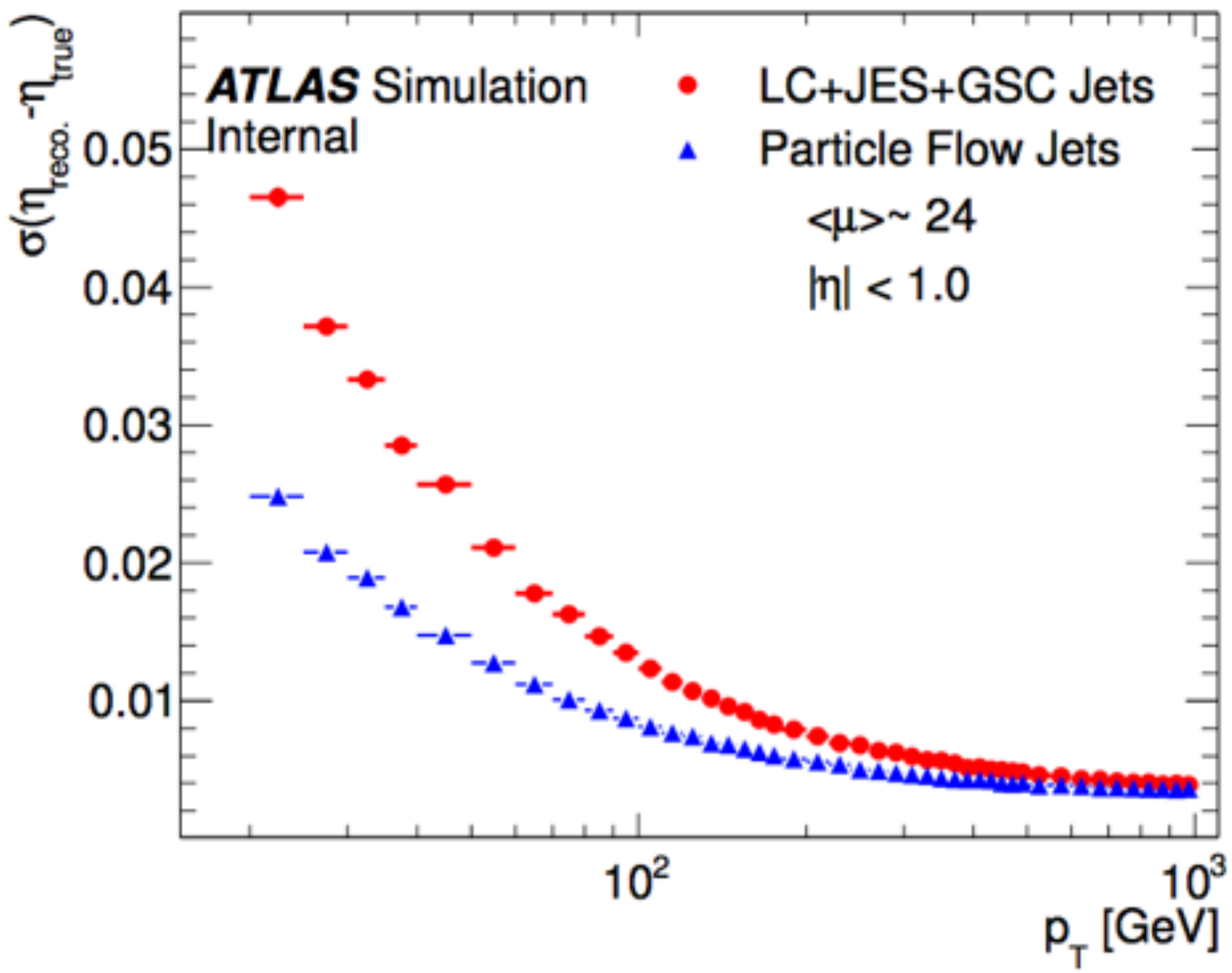
Particle flow performance studies

Results at $\sqrt{s}$ = 8 TeV

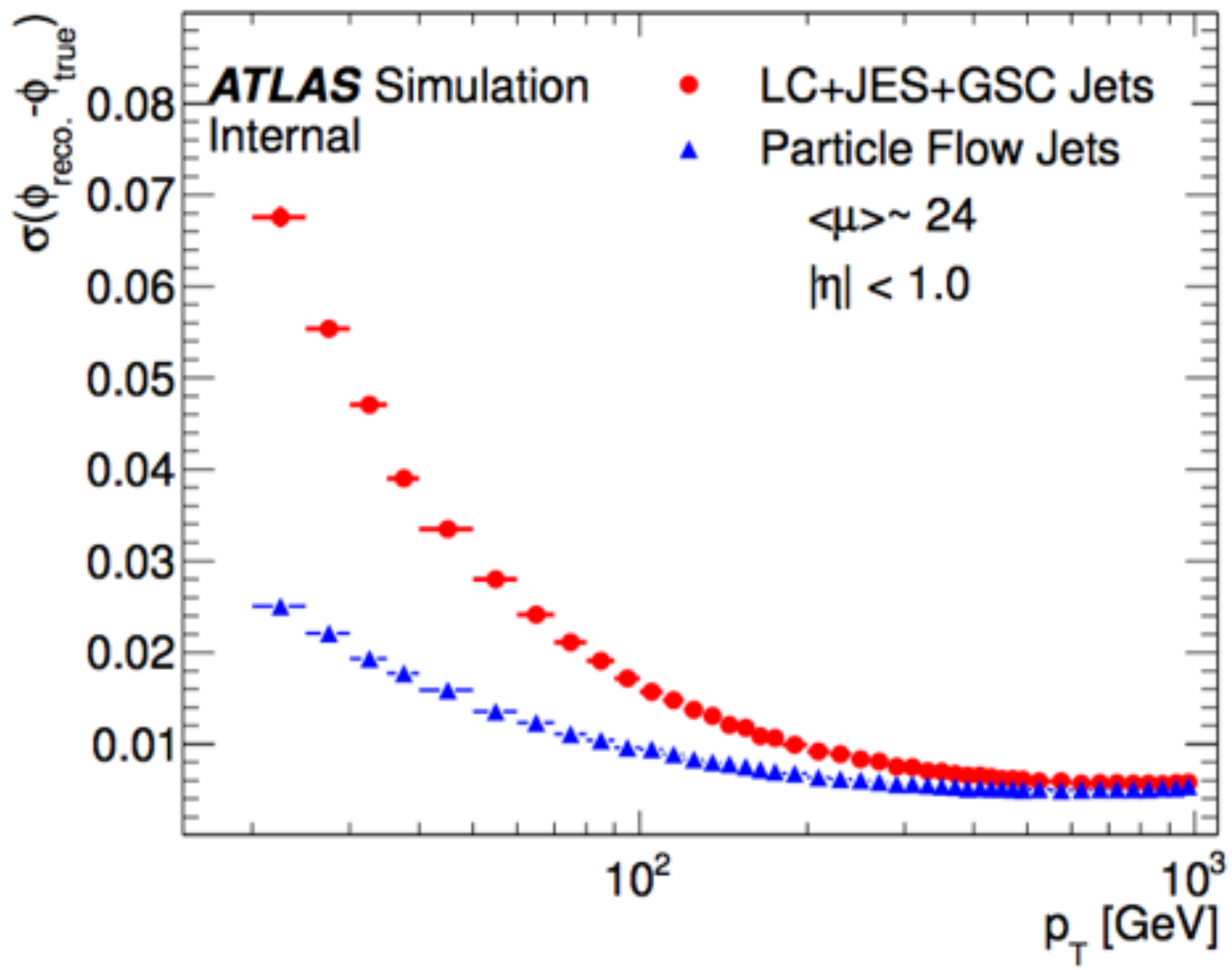
Jet angular resolution

- PFlow improves the angular resolution of the jets in both η and Φ:
  - Angular resolution of the tracker better than the calorimeter.
  - Parameters determine at perigee (no spread out due to the magnetic field).
  - Reduction of pileup also helps in the proper determination of the jet direction.

η angular resolution



Φ angular resolution

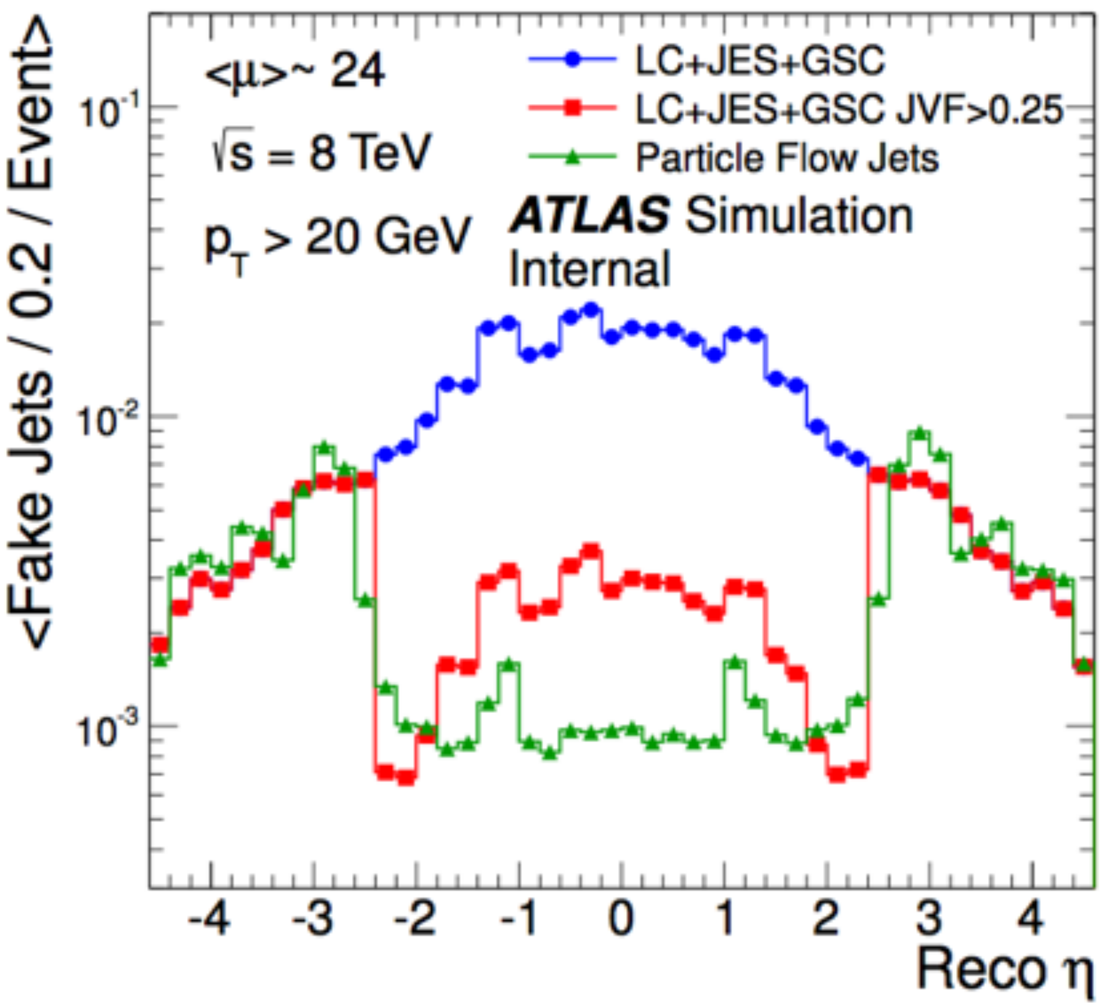


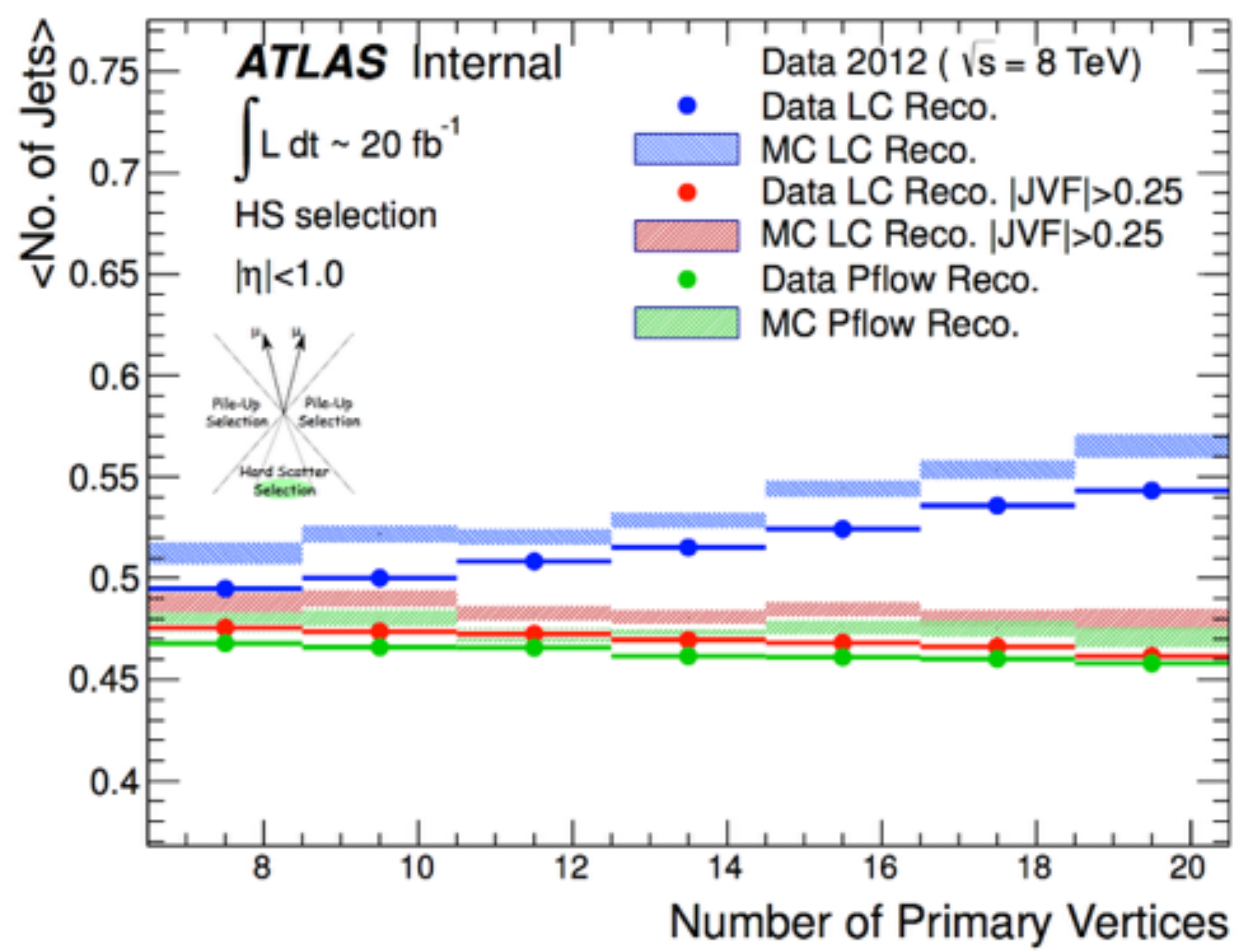
Particle flow performance studies

Results at $\sqrt{s}$ = 8 TeV

Pile-up

- Pile-up suppression is inherent to PFlow reconstruction:
  - Significantly mitigates the degradation of the jet resolution.
  - Particle flow jets originating from pile-up are much suppressed (left).
  - Average number of PFlow jets much more stable as a function of pile-up (right).

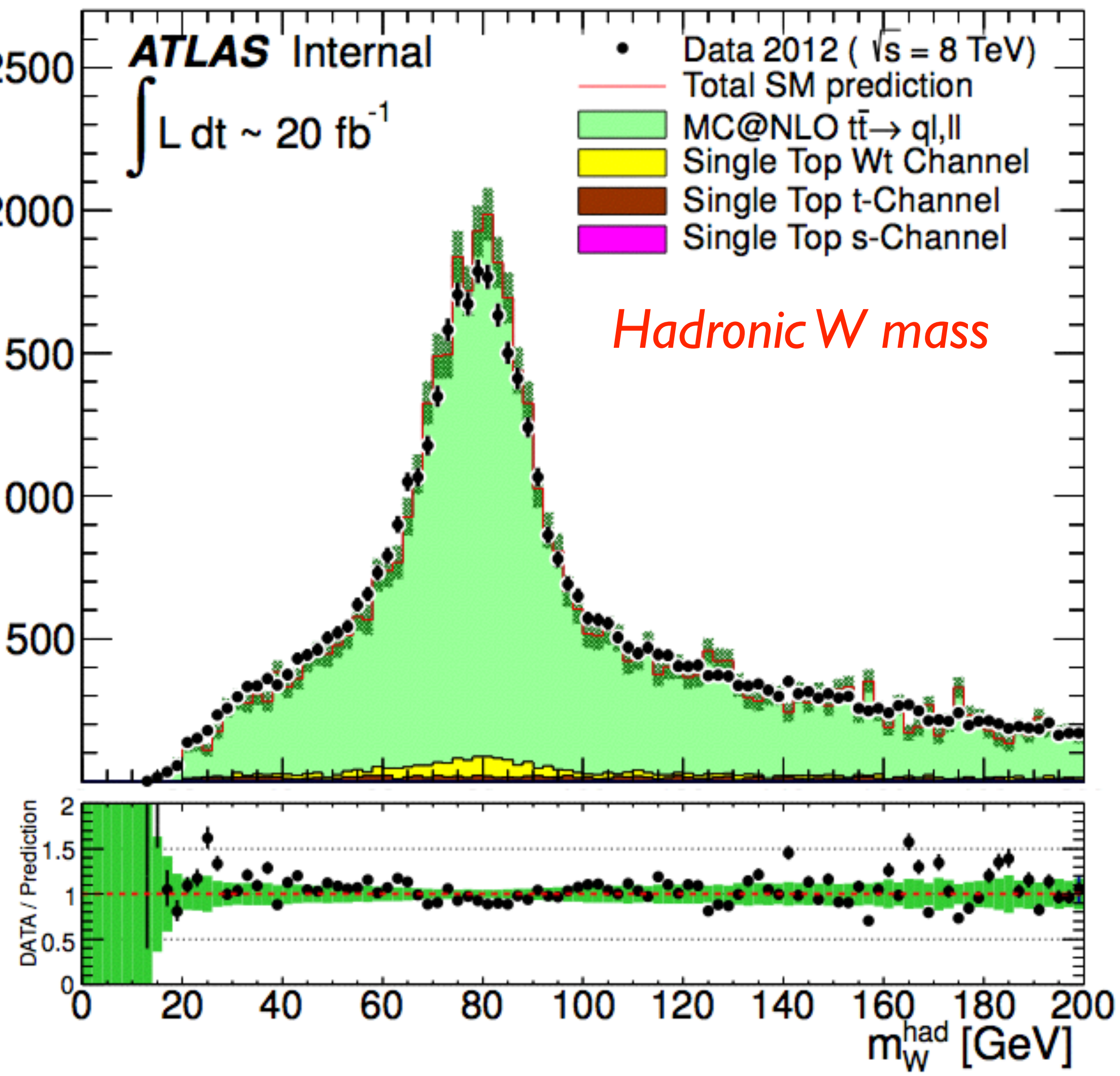


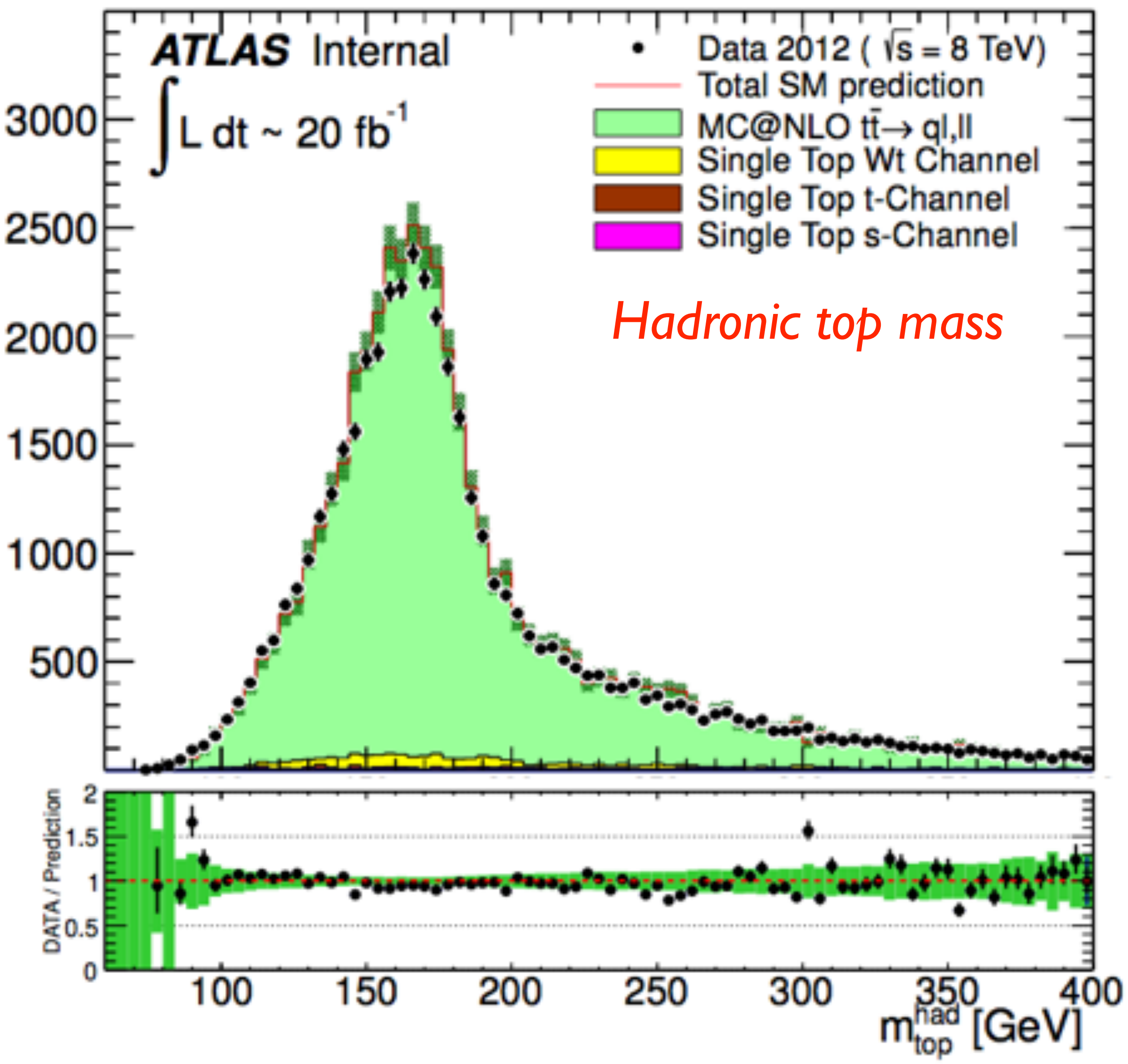


Particle flow performance studies

Results at $\sqrt{s}$ = 8 TeV

Data/MC comparison in ttbar events

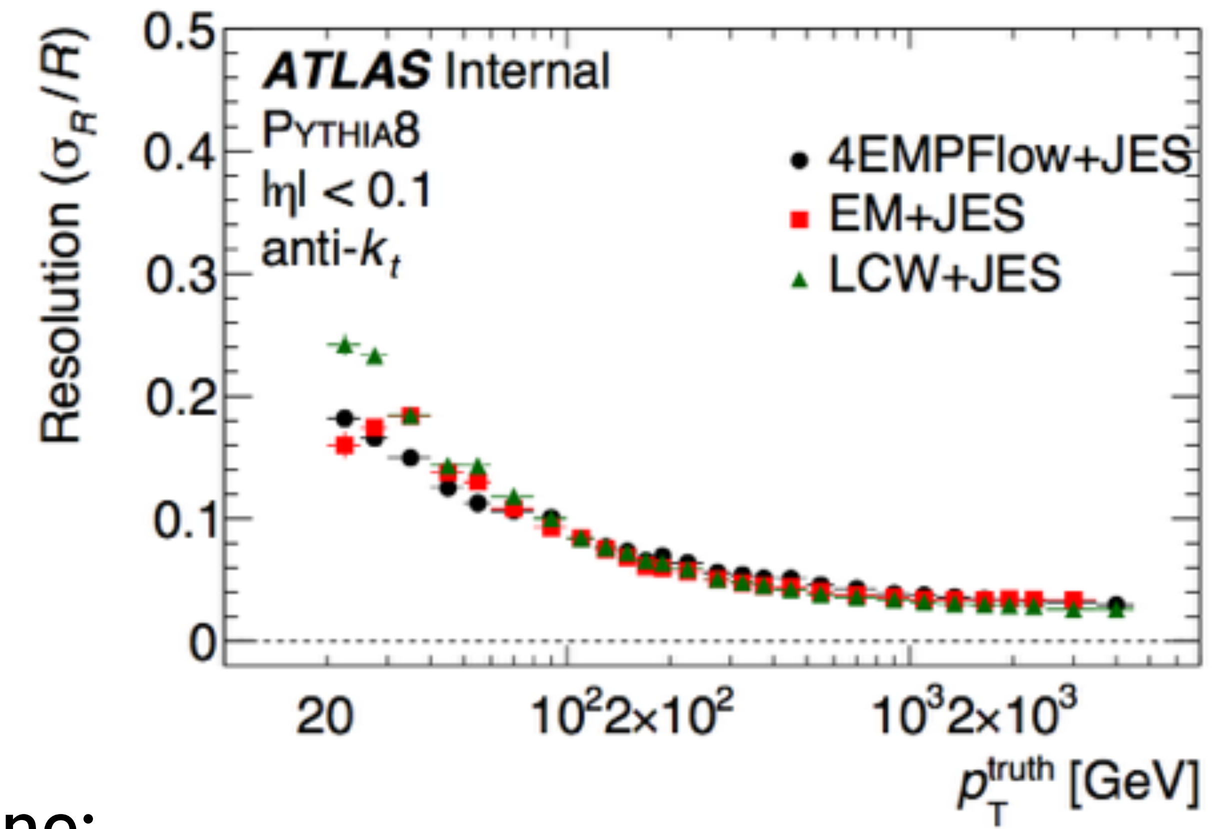




- Good agreement is seen in the distributions.
- PFlow reconstruction reduces the combinatorial bkg contribution.
- PFlow provides a good measurement of the peak:
  - $\sigma(M_W)$ decreases from 16 GeV to 14 GeV when using PFlow instead of LC jets.

Current status

- Current status and developments:
  - Full calibration sequence in 20.1 available (20.7 is being produced).
  - Optimisation of the b-tagging in progress.



Rel 20.7

- New charged shower subtraction.
- Updated lepton identification.
- Bug fix for tight tracks included.
- Good for physics!

Rel 21.

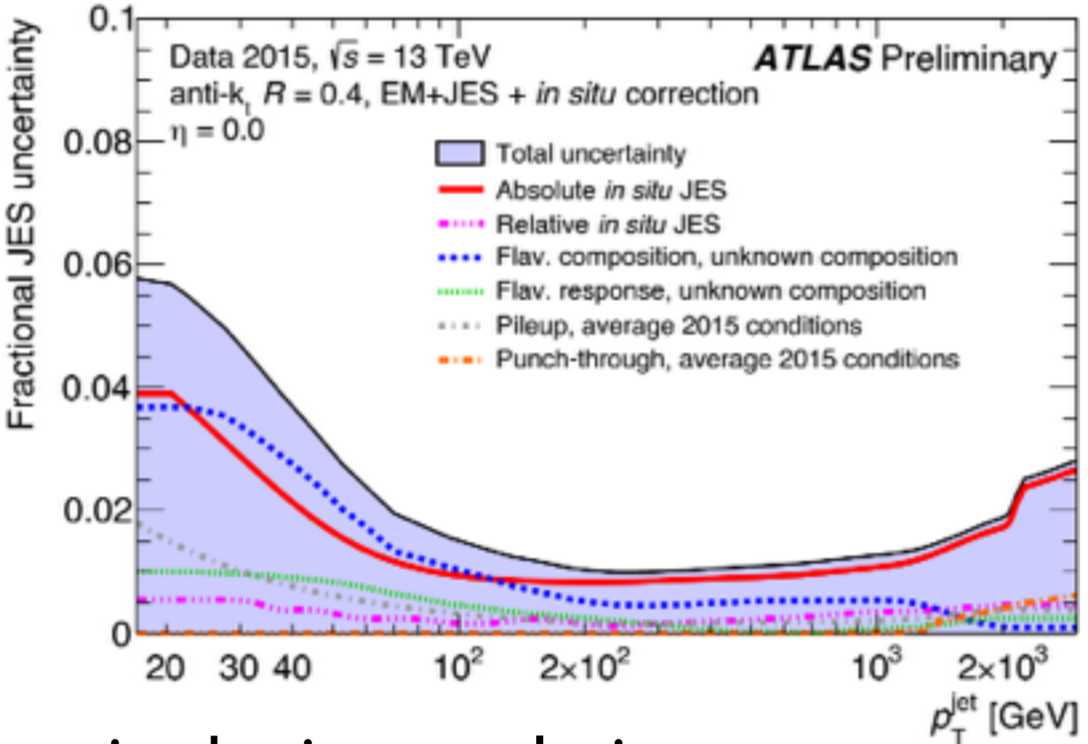
- Several improvements in the pipeline:
  - Optimisation of the charged shower subtraction
  - Vertex stored in cPFO
  - Updated cleaning moments
  - More flexible code

Prospects for systematic uncertainties 2016

The systematic evaluation at 13 TeV is ongoing but…
we can try to extrapolate the results based on our current knowledge.

JES

- High JES uncertainty from modelling
  → Improvement not totally clear
- Pile-up unc. significantly reduced
  → **Improvement expected!**



JER

- Quite significant resolution improvements in the jet resolution
  → **smaller JER uncertainty!**

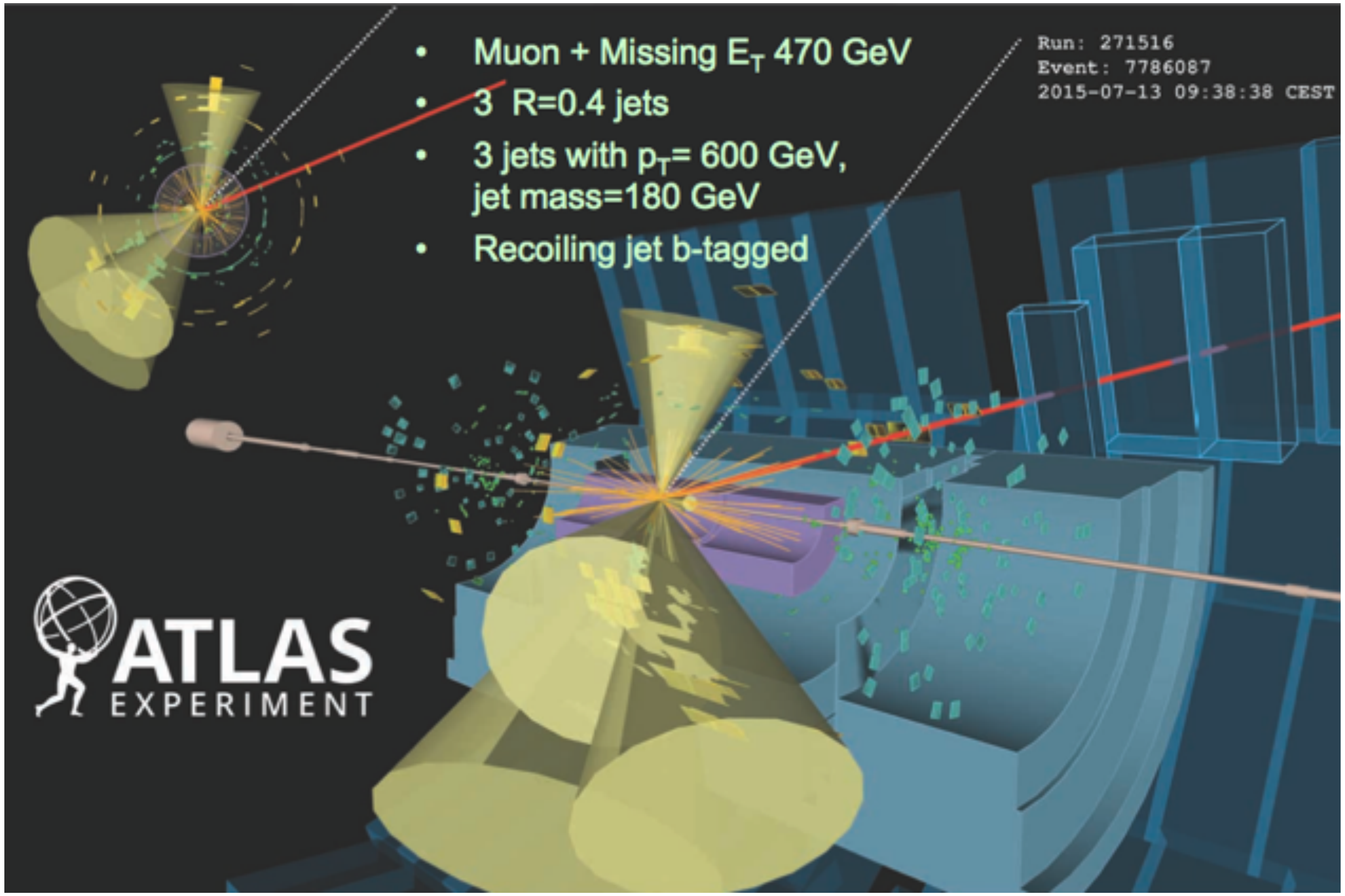
JVT

- The necessity of the JVT is reduced a lot which will have smaller systematics

Similar results with PFlow jets are expected to be ready by summer

Boosted top-tagging algorithms :

Run 2 will produce a large fraction of top-pair production events in the boosted regime!



Boosted top-tagging algorithms

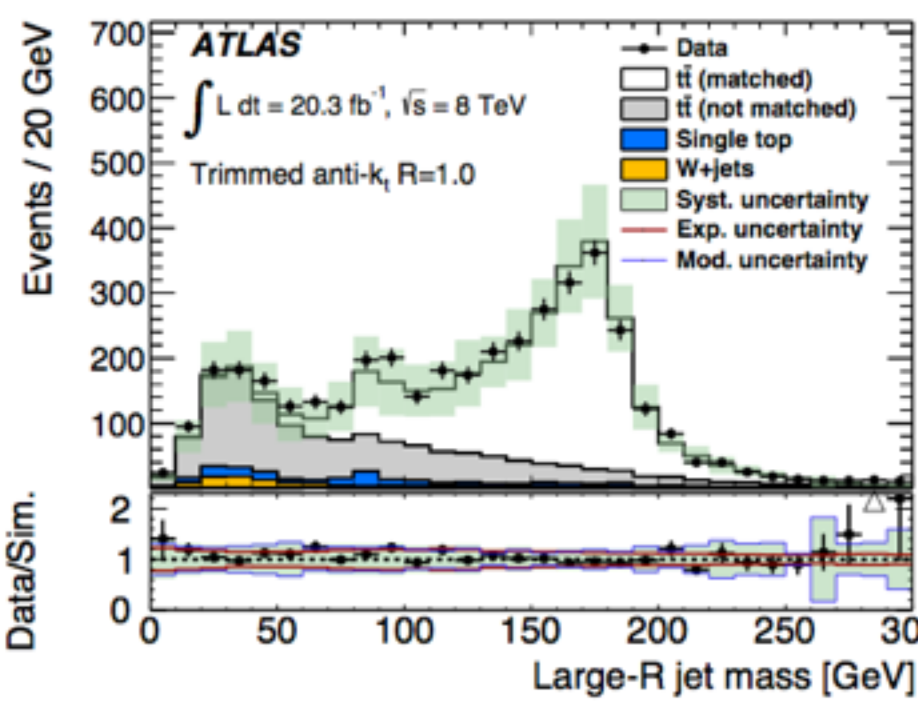
Main goal of the boosted top-tagging techniques is to properly reconstruct the hadronically decaying high p_T top quark within a fat jet

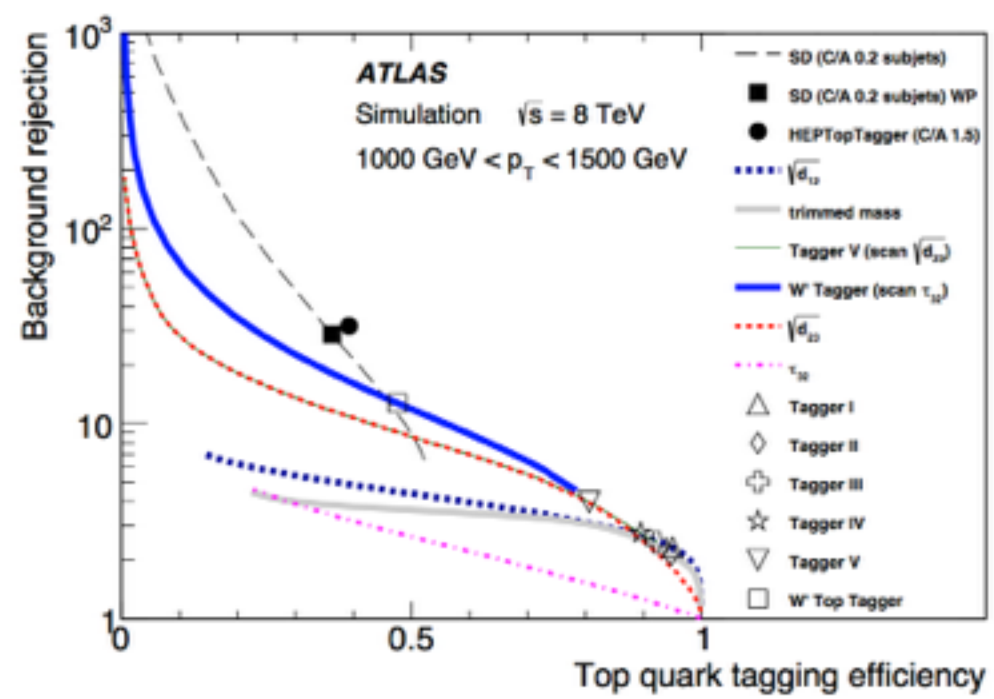
- **For very high p_T tops the decay topology is highly collimated:**
  - The individual partons can not be resolved using small-R jets
  - Boosted topologies need large-R jets

Top-tagging algorithms in Run I

- **Jet-Substructure taggers.**
- **Shower Deconstruction.**
- **HEPTopTagger.**

Tagger	Jet algorithm	Grooming	Radius parameter	p_T range	$\|\eta\|$ range
Tagger I–V W' top tagger Shower Deconstruction	anti-k_t	trimming ($R_{sub} = 0.3$, $f_{cut} = 0.05$)	$R = 1.0$	> 350 GeV	< 2
HEPTopTagger	C/A	none	$R = 1.5$	> 200 GeV	< 2





Substructure tagger performance at early 13 TeV

ATL-PHYS-PUB-2015-053

Substructure-variables tagger at early 13 TeV

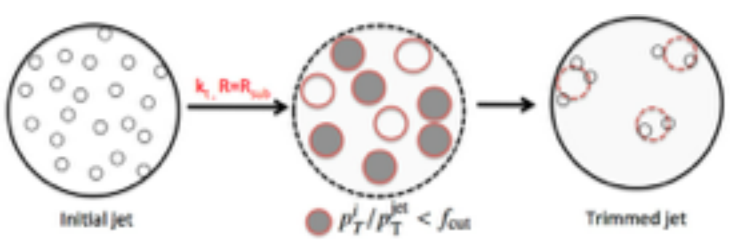
(Pre-recomendations 2016)

Suitable input of jet reconstruction

- Calibrated topo-clusters → calorimeter jets
- Stable particles → generator level jets

↓

Chose the algorithms and grooming



- Anti-k_t algorithms with R=1.0
- Jets are trimmed to mitigate the pileup:
  - Reclustered with k_t (R_{sub}=0.2)
  - Jets with $p_{T\_i}/p_{T\_jet} > 0.05$
- Jets recombined to produce the final trimmed jets
- Only jets with more than 2 constituents are kept
- MCJES + jet mass calibration

↓

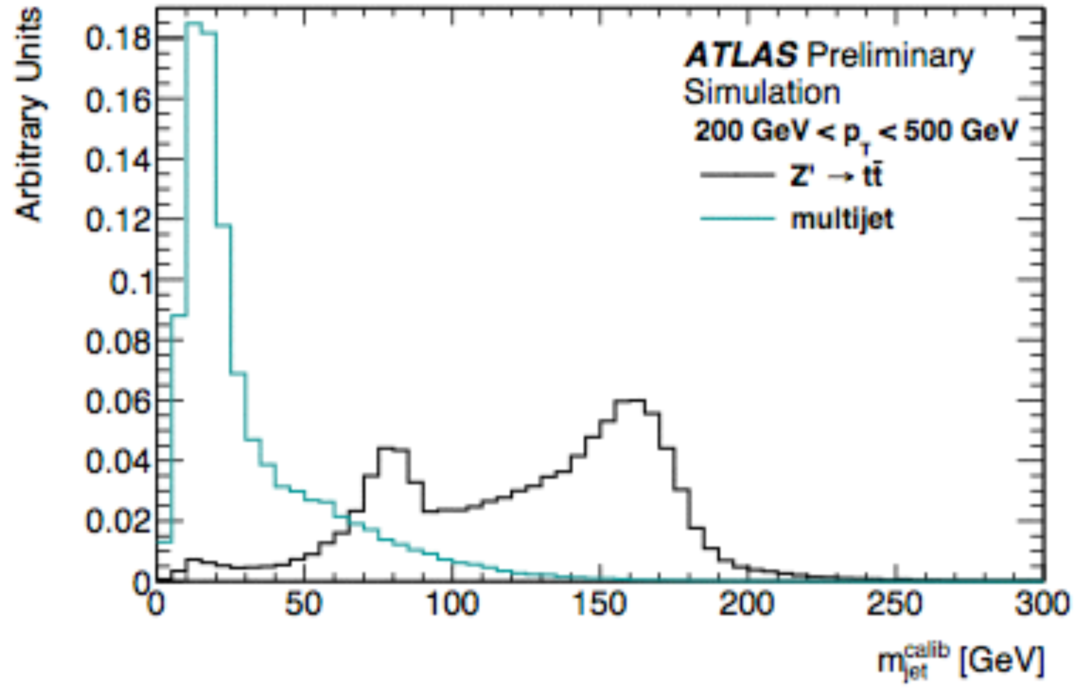
Choose the top-tagging variables

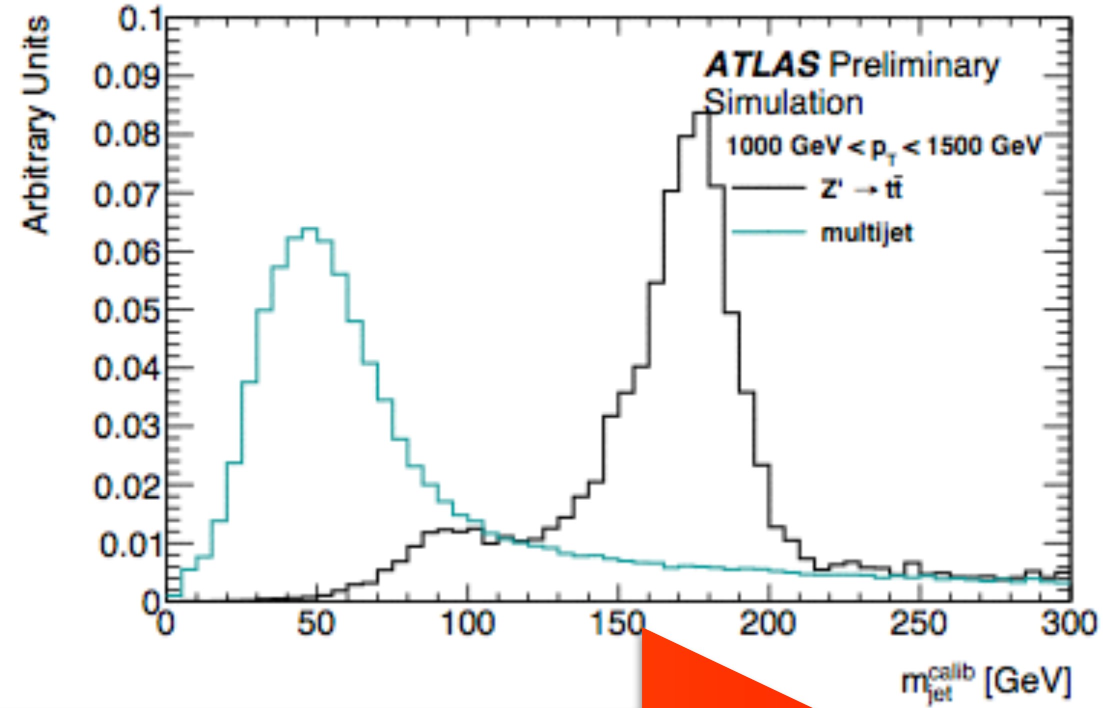
- Calibrated jet mass: $(m_{\text{jet}}^{\text{uncalib}})^2 = \left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2$, then calibrated
- N-subjetiness $\tau_{32}(\tau_3/\tau_2)$: Describe how well a jet is described as containing N subjets. Here, compatibly with 3 subjets, and with 2 subjets.

Top-tagging jet requirements

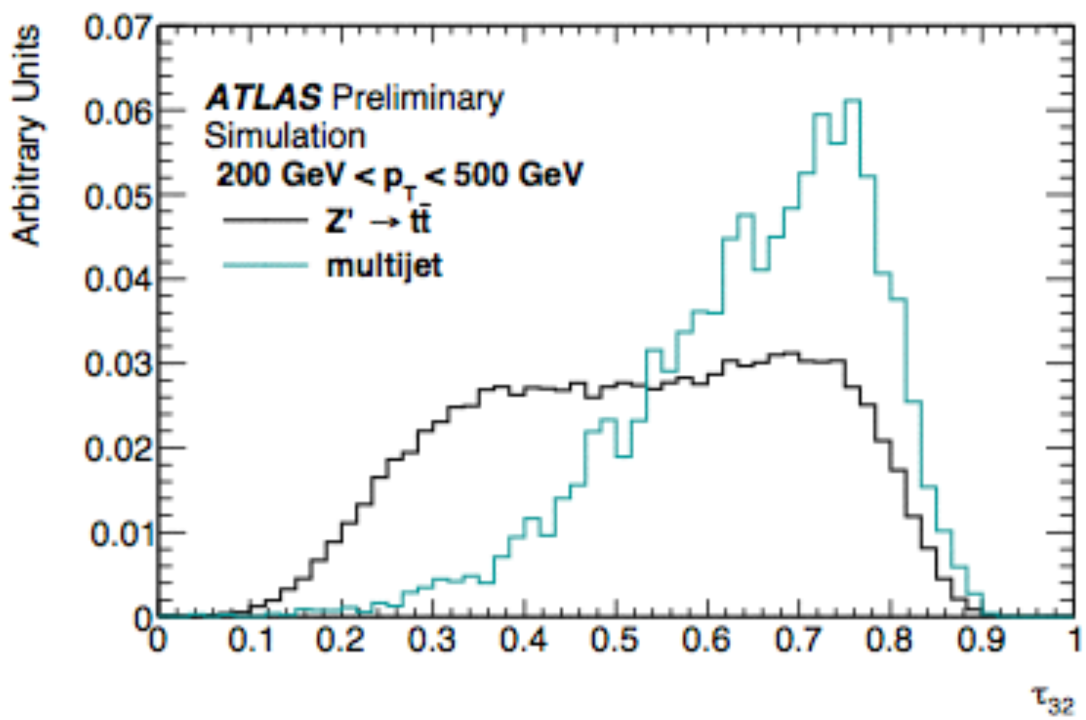
ATL-PHYS-PUB-2015-053

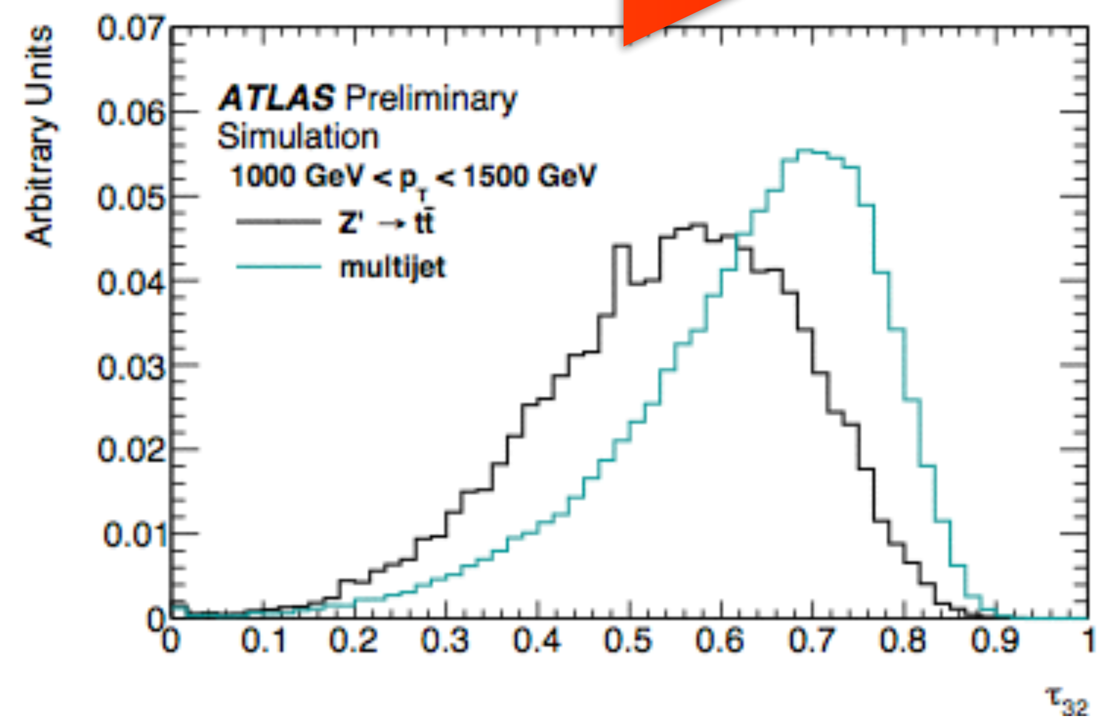
Samples:
- Signal jets: Z' → ttbar
- Background jets: QCD multi-jet





Softly boosted → Highly boosted





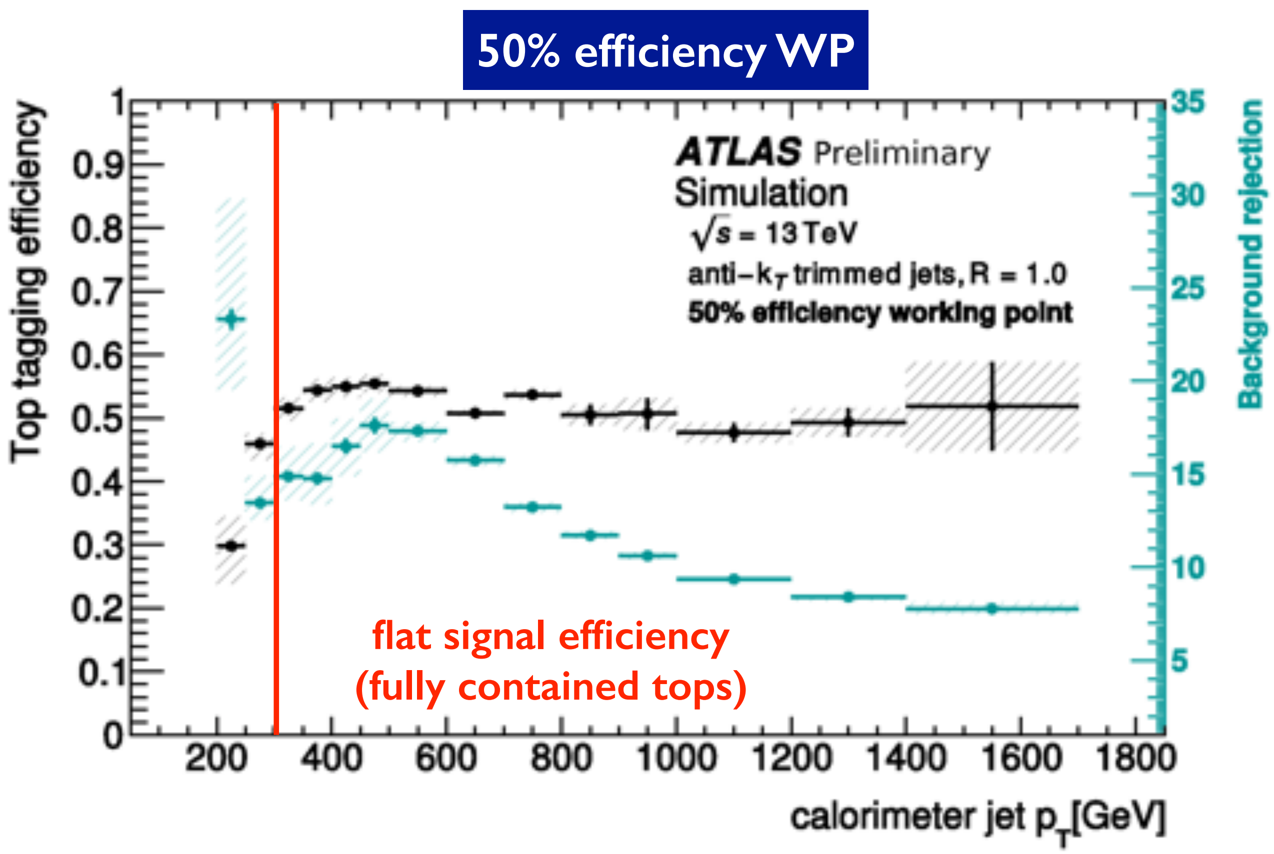
Top-tagging performance

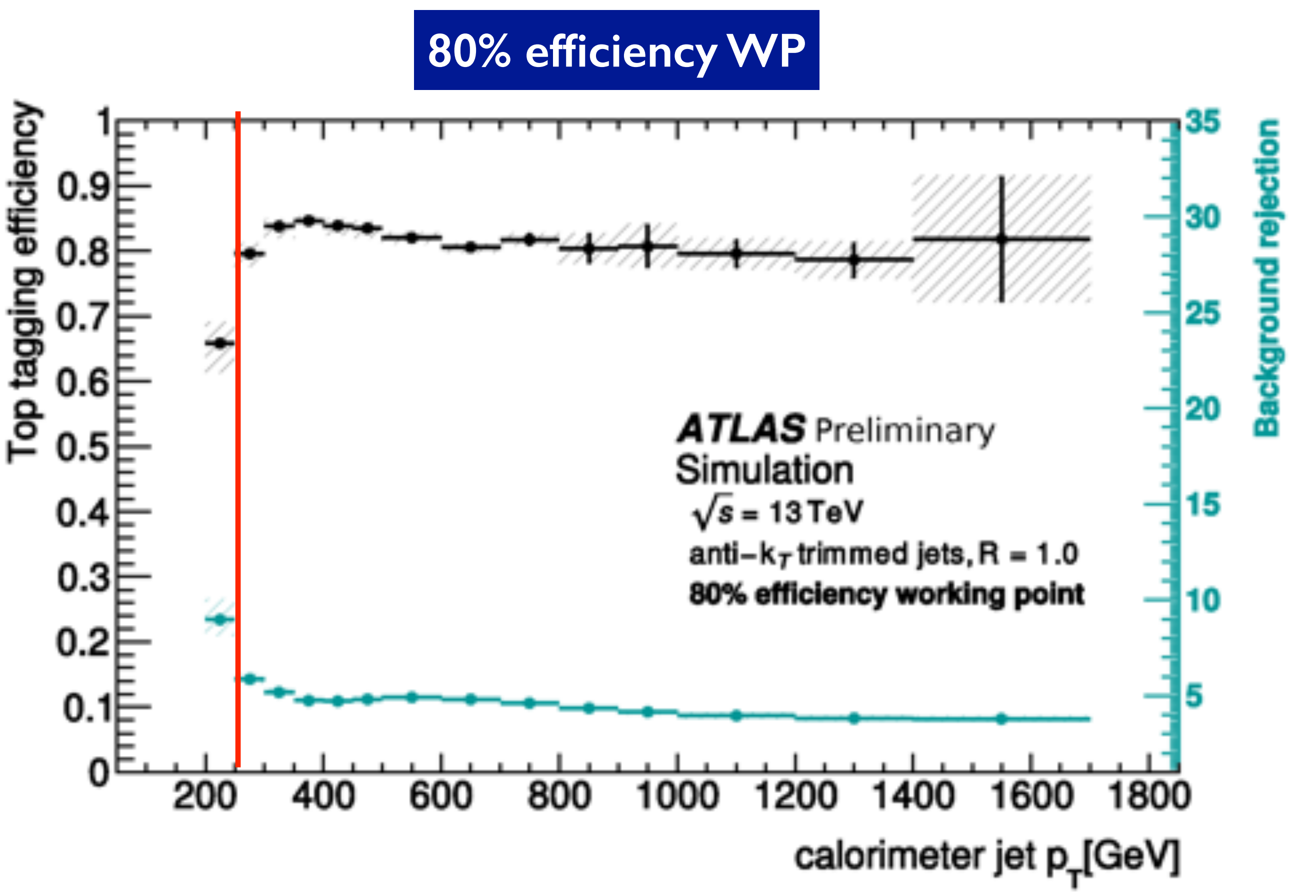
ATL-PHYS-PUB-2015-053

- **Algorithm performance**
  - Identification of hadronically decaying tops with $p_T > 200$ GeV
  - Two efficiency points are provided:

$$f^{signal} = \frac{N^{tag}_{t\bar{t}_{matched}}}{N_{t\bar{t}_{matched}}} \qquad f^{mistag} = \frac{N^{tag}_{QCD_{matched}}}{N_{QCD_{matched}}}$$

- *50% signal efficiency* → for analysis with very large reducible bkg.
- *80% signal efficiency* → for analysis dominated by and/or irreducible bkg.



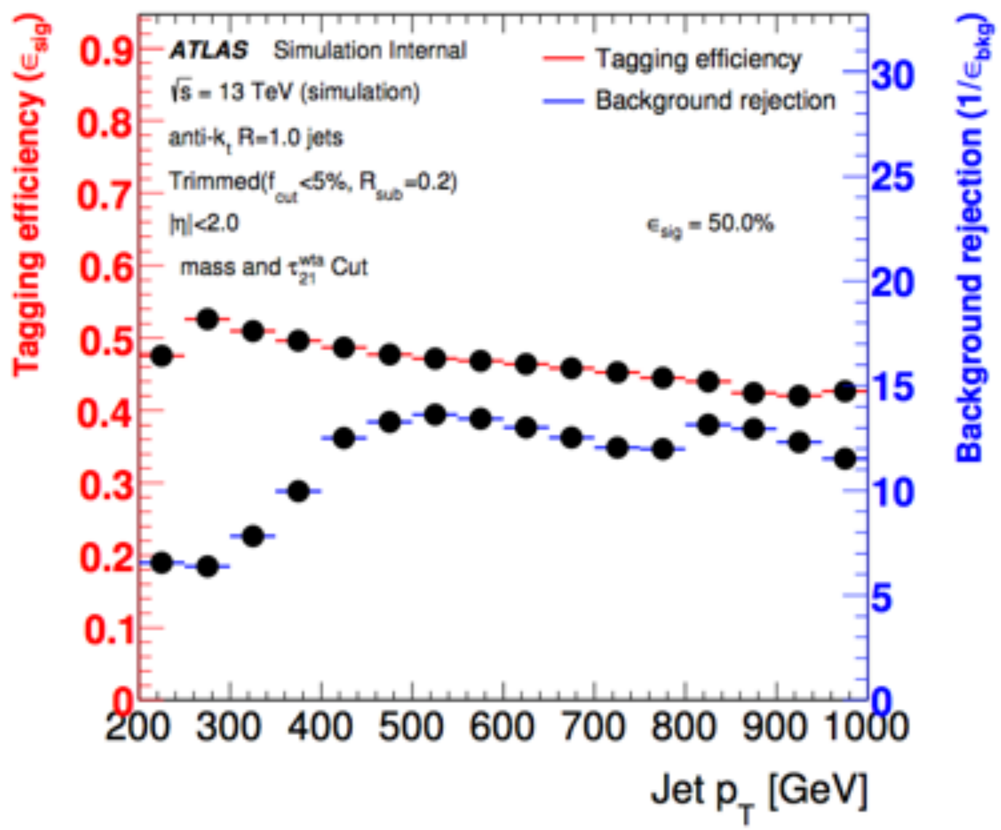


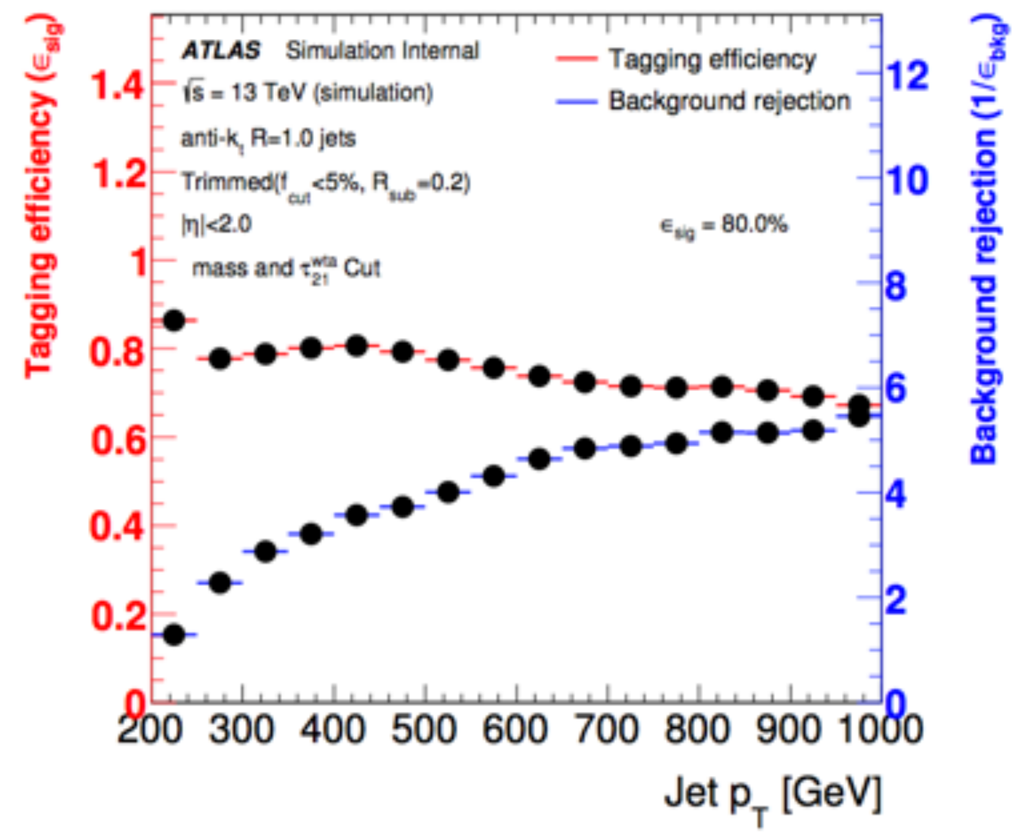
Current development:

- Top-tagging algorithms current developments:

Simple JSS tagger

- Optimisation of the cuts in mass and τ_{32} (similar performance)





- Improvements may arrive from:
  - using the track-assisted mass
  - including additional observables (the most important ones for W-tagging)

Probably useful for physics analyses in spring

Current development:

- Top-tagging algorithms current developments:

SD and HEPTopTagger

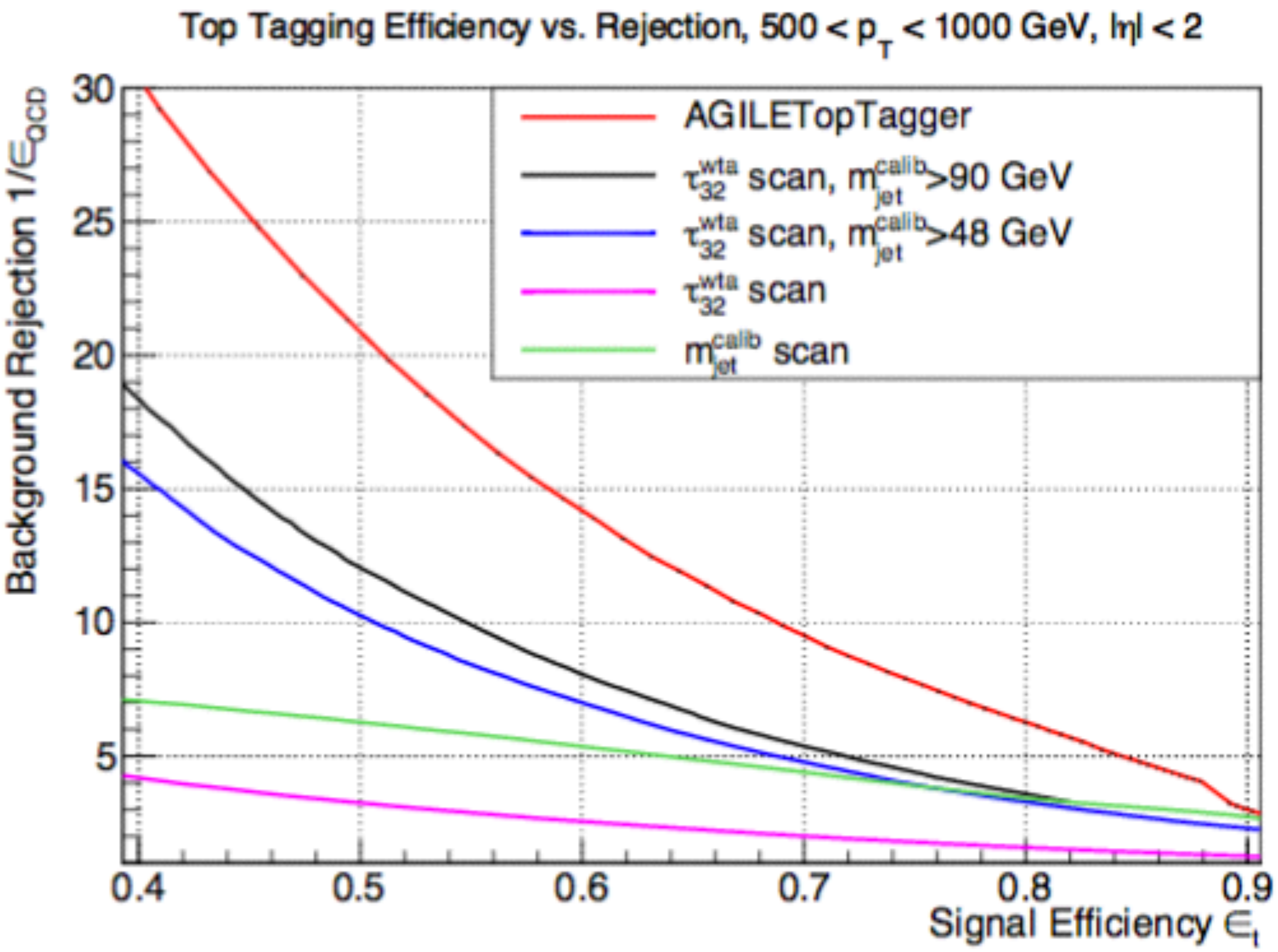
- Complicated treatment of the uncertainties from the small-R jets used as an input
- Not clear if they will be useful for physics analyses

hopefully for summer

MVA methods

- Combination of JSS moments via a BDT and AGILEPack

Results look promising but still a lot of work



feedback from top analyses to the JSS group is SUPER welcome!

Summary

PFlow jets:

- Performance studies have demonstrated advantages for the particle-flow objects:
  - PFlow jet resolution is better than EM (LC) calorimeter jets at low p_T
  - Suppression and stability of the pileup contribution
  - Better angular determination (in η and Φ)
- PFlow jets with new CSS available in mc15c
  - JetEtMiss and jet cleaning recommendation ongoing
  - Expected improvements for JES and JER systematic uncertainties (not yet quantified)

Boosted Top-Tagging techniques:

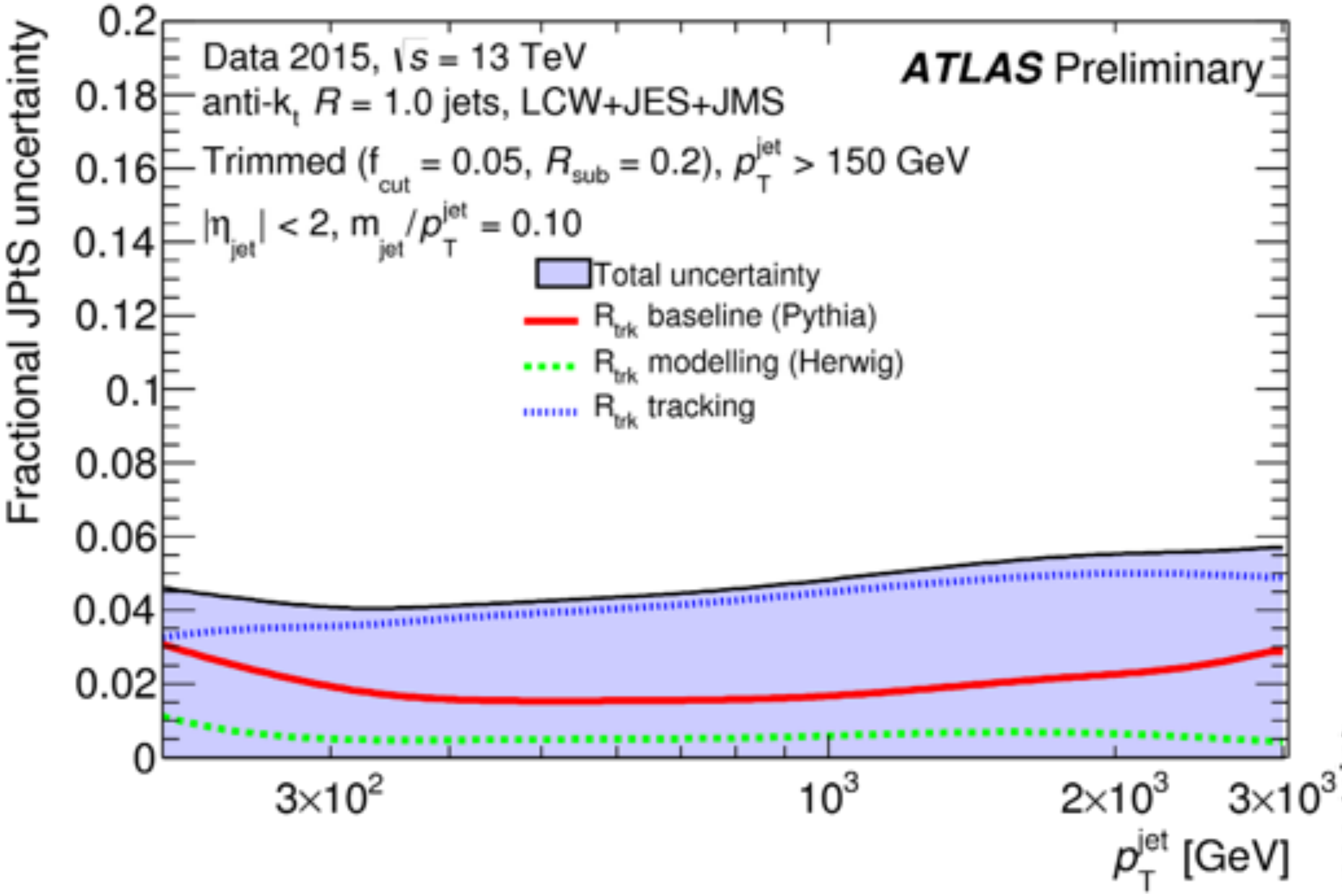
- Successfully used in heavy resonances decaying to top quarks.
- Preliminary recommendation at 13 TeV available (JSS tagger):
  - 50% and 80% W.P. efficiency available for physics analyses
  - Efficiency stable in busier topologies

It is time for starting having a look at particle-flow and using top-tagging algorithms in top physics! 😉

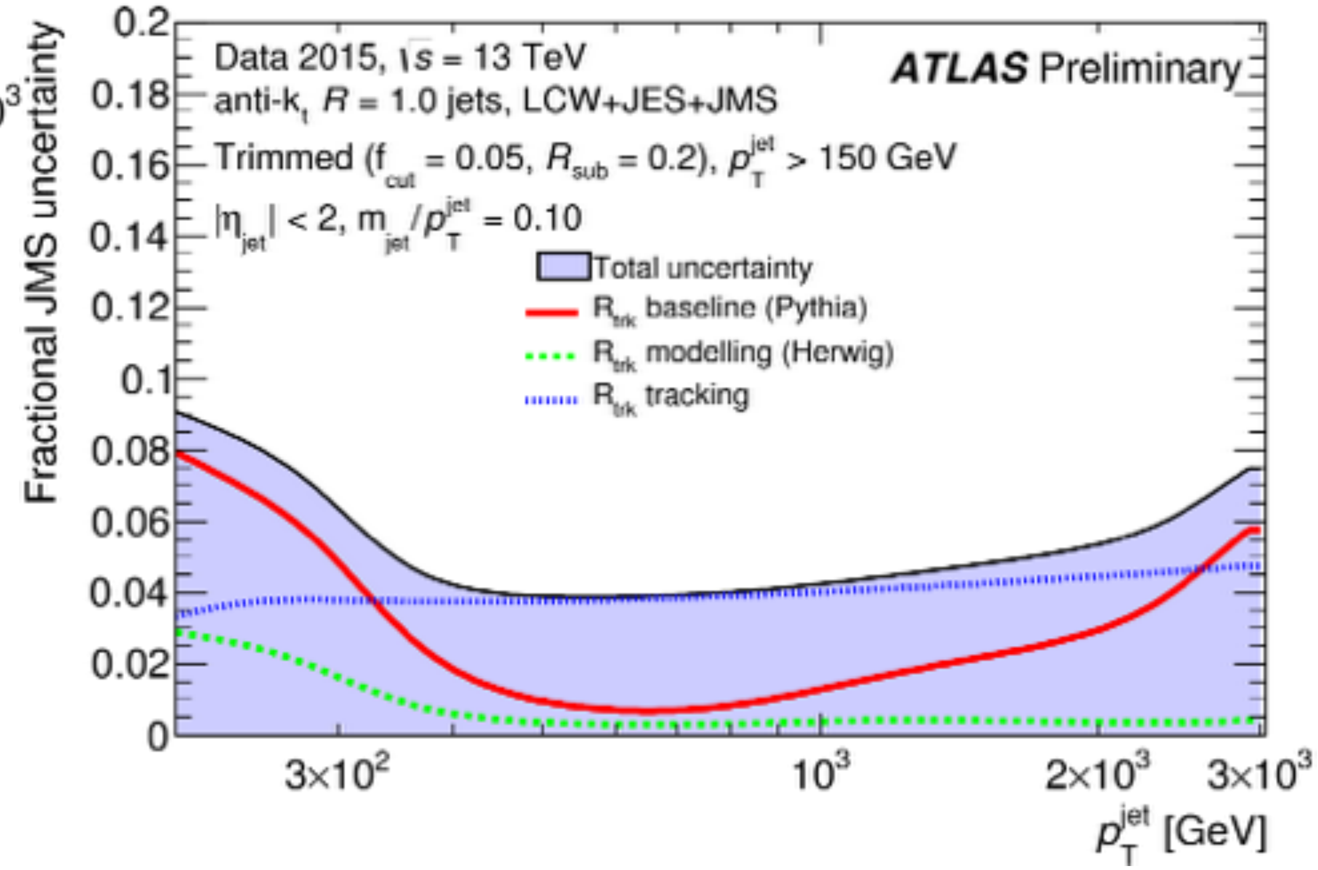
Backup

Large-R jets systematic uncertainties

- Fractional jet pT scale (JPTS) systematic unc.



- Jet mass scale systematic unc.



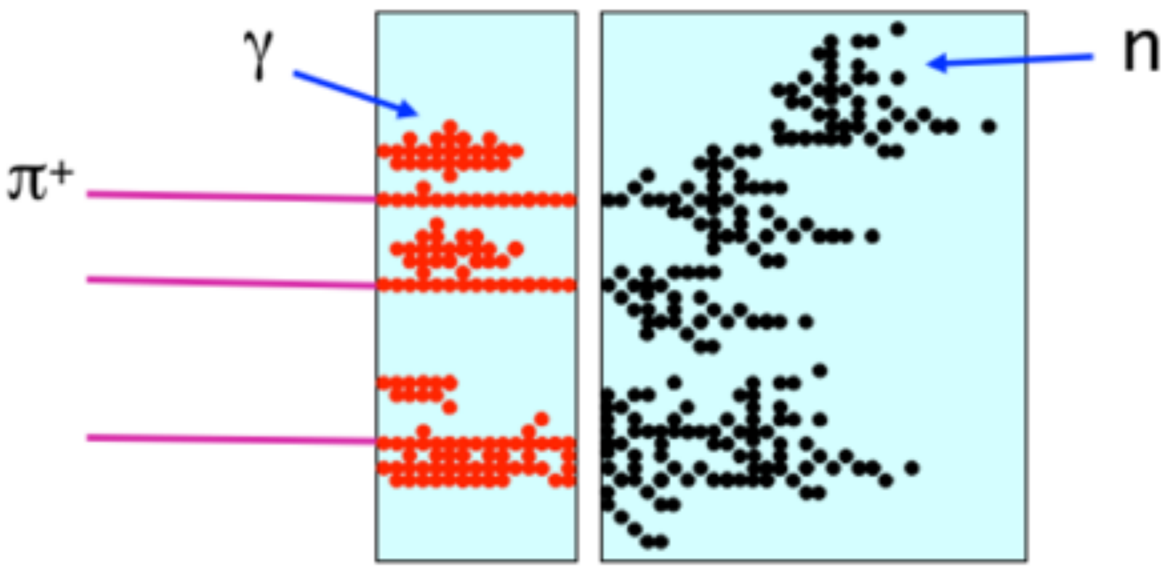
Introduction: What is Particle flow?

First, have a look at the composition of a typical jet:

- 60% of jet energy in charged hadrons
- 30% in photons (mainly from $\pi^0 \rightarrow \gamma\gamma$)
- 10% in neutral hadrons (n, K,…)

Traditional calorimetric approach:

- Measure all components of jet energy in calorimetry system
- 70 % of energy use information form the hadronic calorimeter
- Intrinsically HAD calorimeter resolution limits jet energy resolution
  - HAD Calo. ATLAS: $\frac{\sigma(E)}{E} \approx \frac{50\%}{\sqrt{E}} \oplus 3\%$



$E_{JET} = E_{ECAL} + E_{HCAL}$

Introduction: What is Particle flow?

First, have a look at the composition of a typical jet:

- 60% of jet energy in charged hadrons
- 30% in photons (mainly from $\pi^0 \rightarrow \gamma\gamma$)
- 10% in neutral hadrons (n, K,…)

ATLAS resolutions

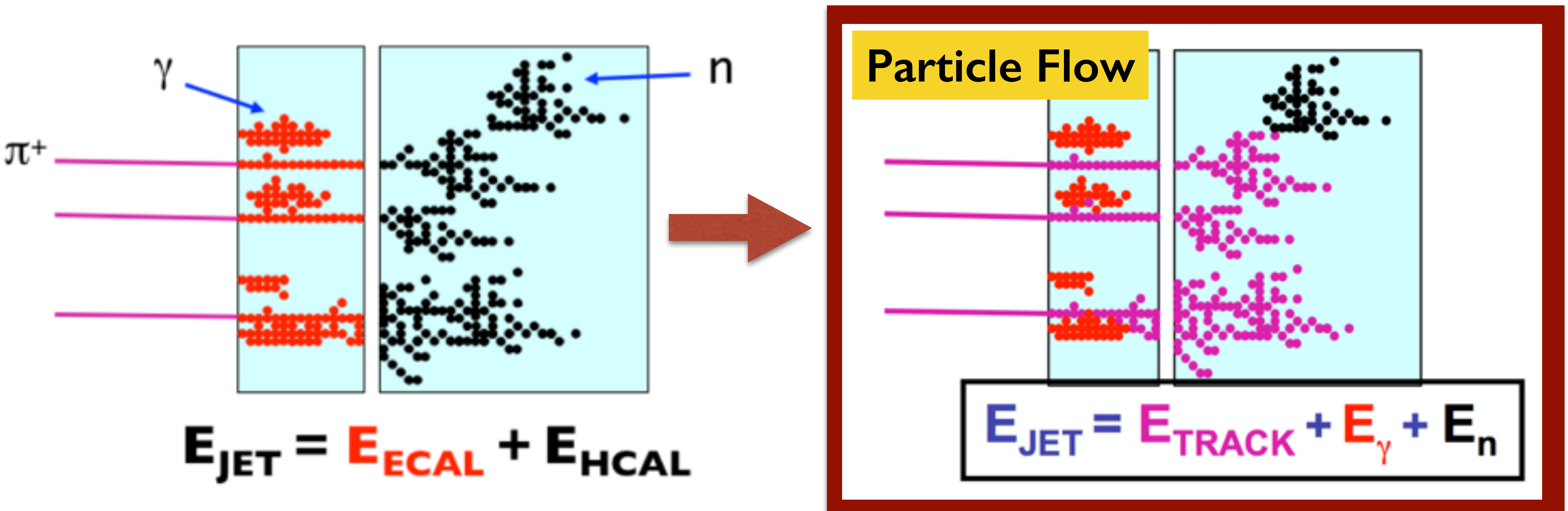
$$\frac{\sigma(p_T)}{p_T} \approx 0.036 p_T \% \oplus 1.3\%$$

$$\frac{\sigma(E)}{E} \approx \frac{50\%}{\sqrt{E}} \oplus 3\%$$

Particle Flow approach:

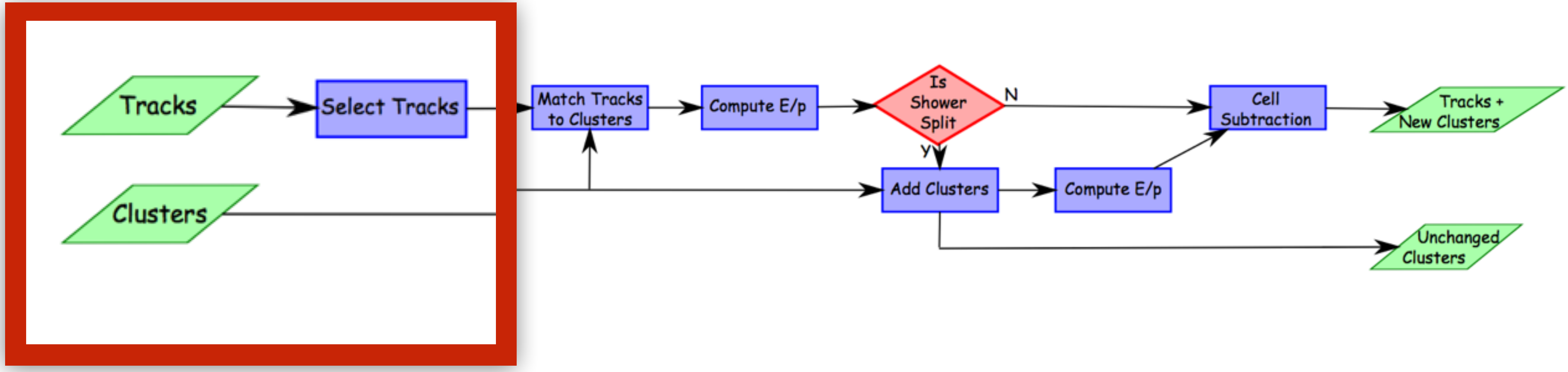
- Charged particles measured in tracker
- Photons in EM Calorimeter
- Neutral hadrons in HAD Calorimeter

Only 10 % of jet energy from HAD Calorimeter ➜ much improved resolution



Overview of the eflowRec chain:

1. Input objects selection:

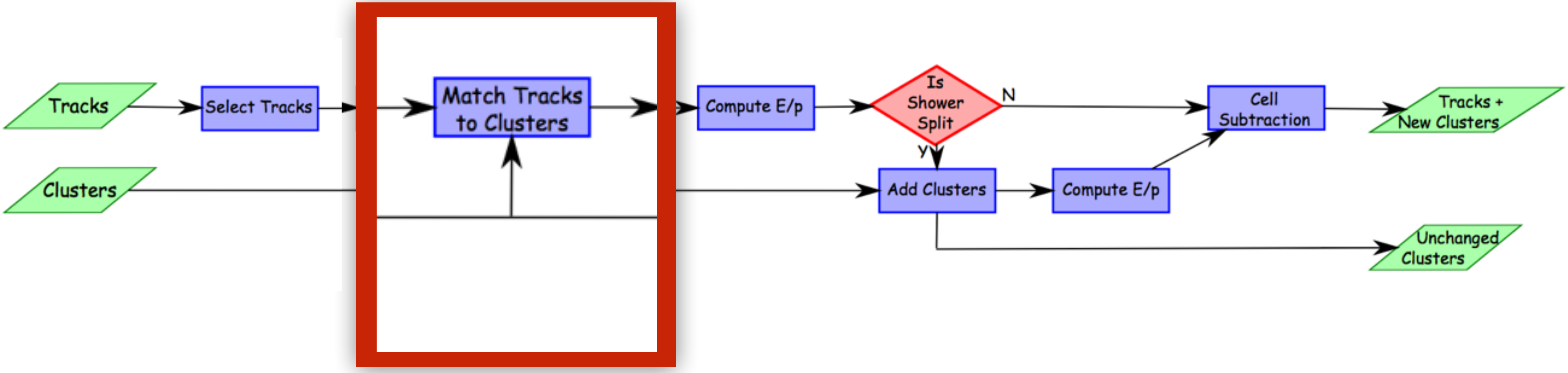


- **Tracks:** selected tracks with a reliable set of properties
  - Tight track selection (CP group recommendations)
  - $|\eta|<2.5$ & $p_T > 500$ MeV
  - $p_T < 40$ GeV

- **Topological clusters:** all considered

Overview of the eflowRec chain:

2. Track-Cluster matching:



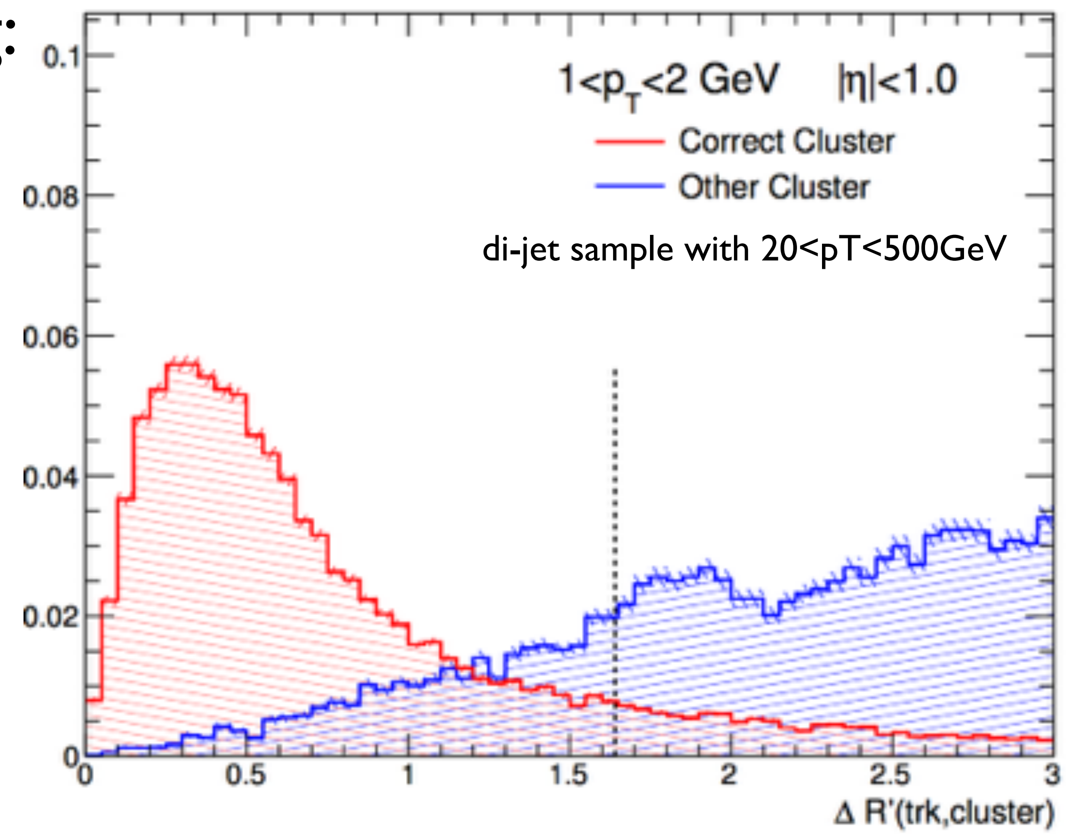
To remove the energy deposited in the calorimeter → match each track to a cluster

- Each track is extrapolated to the 2nd layer of the EM calorimeter
- The nearest topological cluster is found using:

$$\Delta R' = \sqrt{\frac{(\eta_{\mathrm{track}} - \eta_{\mathrm{clus}})^2}{\sigma_\eta^2} + \frac{(\phi_{\mathrm{track}} - \phi_{\mathrm{clus}})^2}{\sigma_\phi^2}}$$

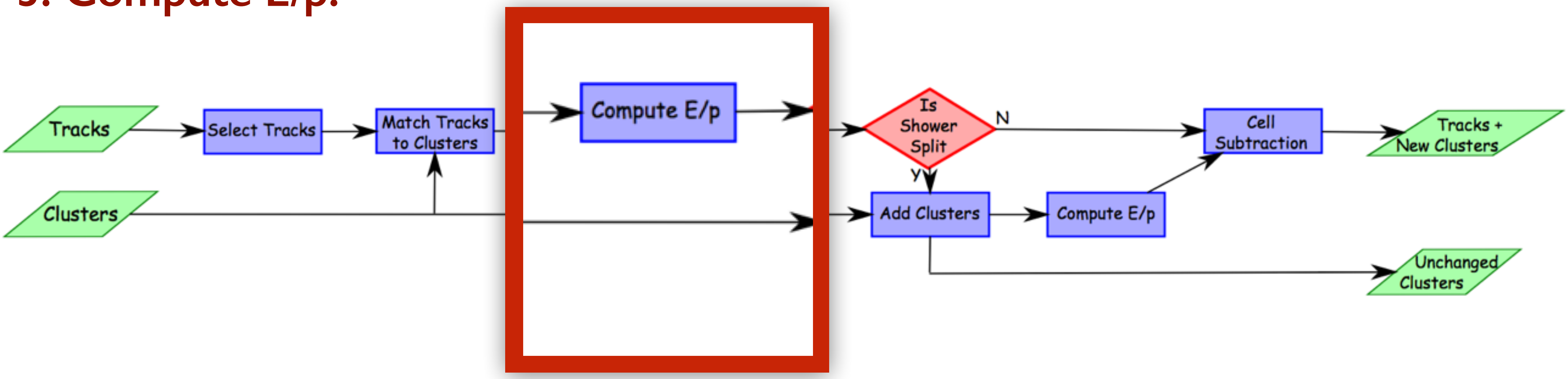
$\sigma_{\eta,\phi}$ of the cluster width

Correct cluster = cluster with E_{cl}>90% E_{true}
Other cluster = next closest cluster



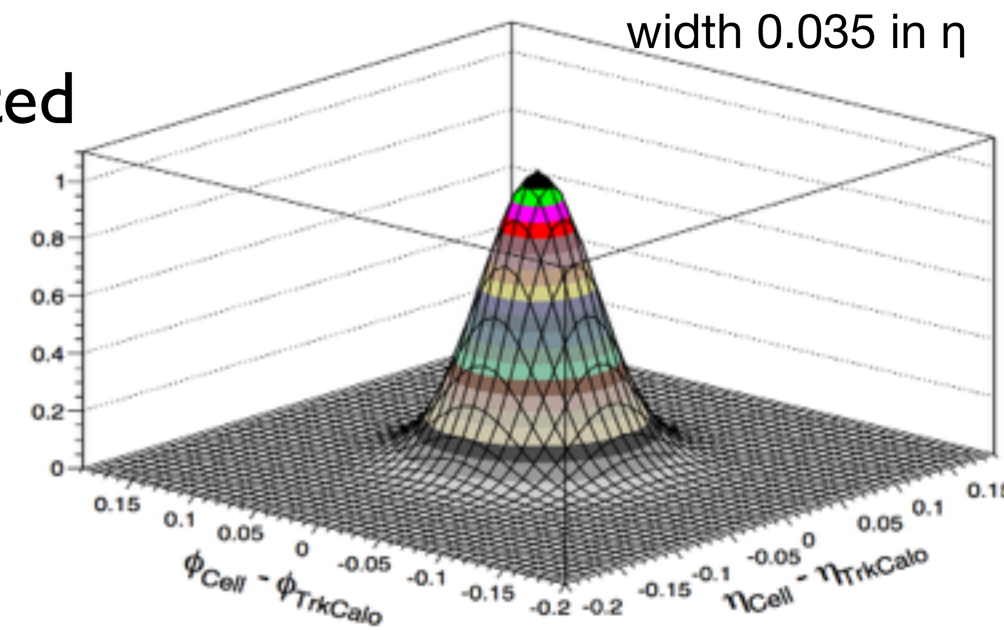
Overview of the eflowRec chain:

3. Compute E/p:



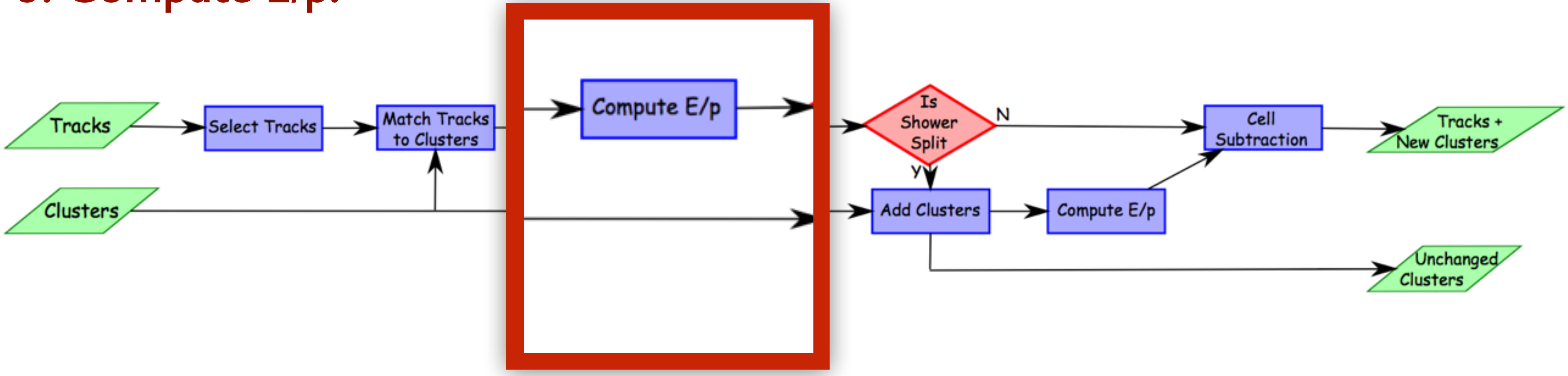
To remove the energy in the calorimeter → how much energy is expected where?

- **LHED** = layer of highest energy density
  - Motivation: find and remove the dense EM core of the shower
  - How to proceed?
    - Scale the cells around the tracks extrapolated position using a Gaussian
    - Calculate the average energy density per radiation length in each layer <ρi'>
    - Take the layer with larges <ρi'>



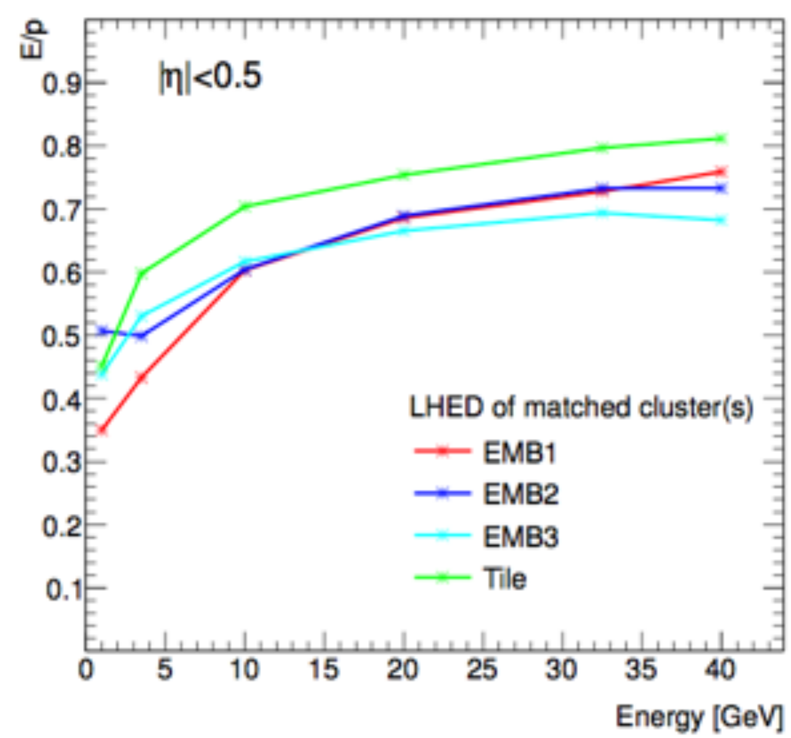
Overview of the eflowRec chain:

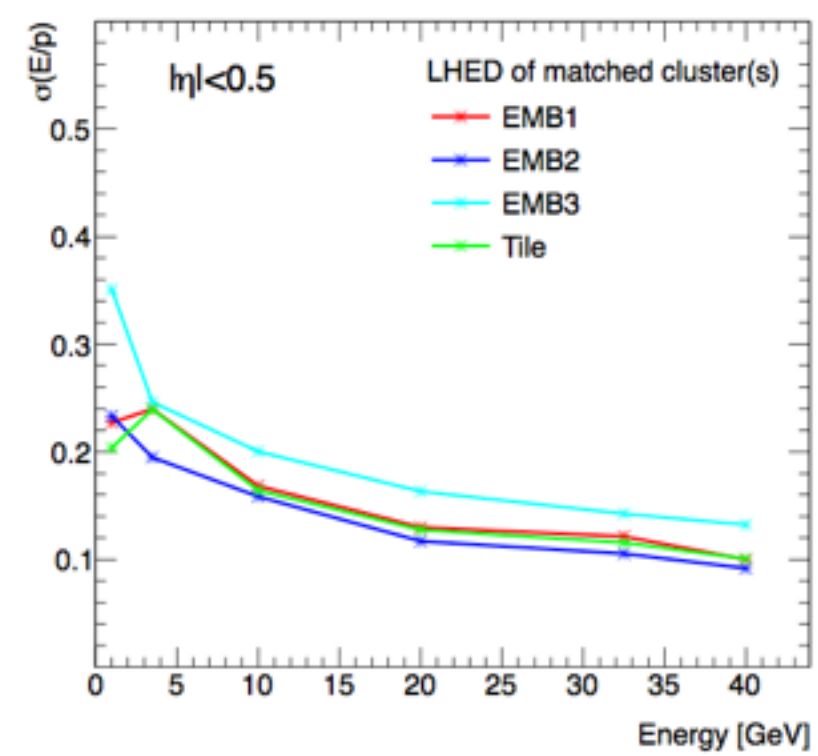
3. Compute E/p:



To remove the energy in the calorimeter → how much energy is expected where?

- E/p: ratio of the E deposited in the calorimeter divided by the p of the track.
  - Single particle samples used to determine E/p
  - E obtained as sum of the clustered E in a cone ΔR<0.2
- The E/p is parametrised in terms of η_{part}, E_{part} , LHED

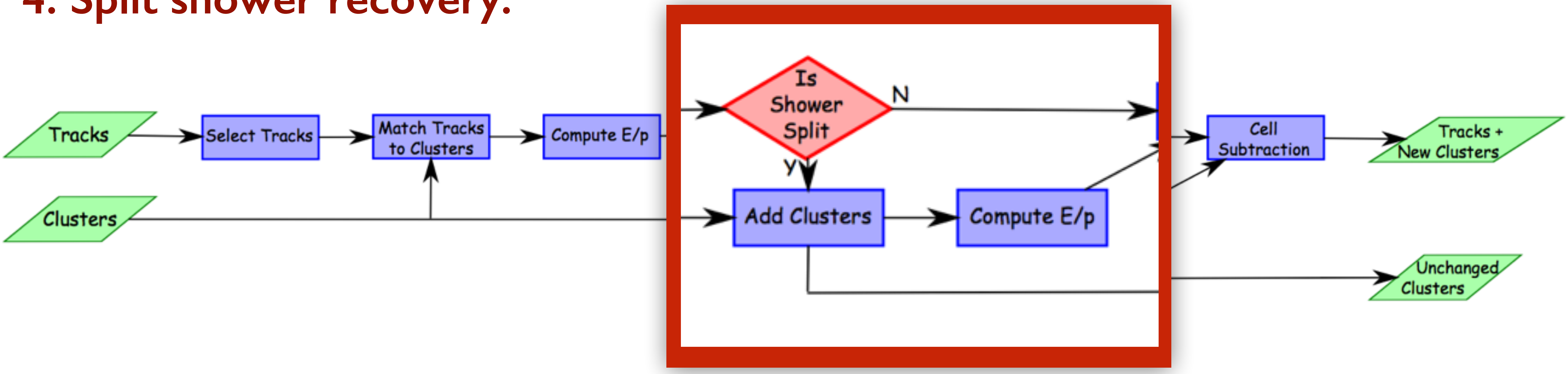




(LHED = layer of highest energy density)

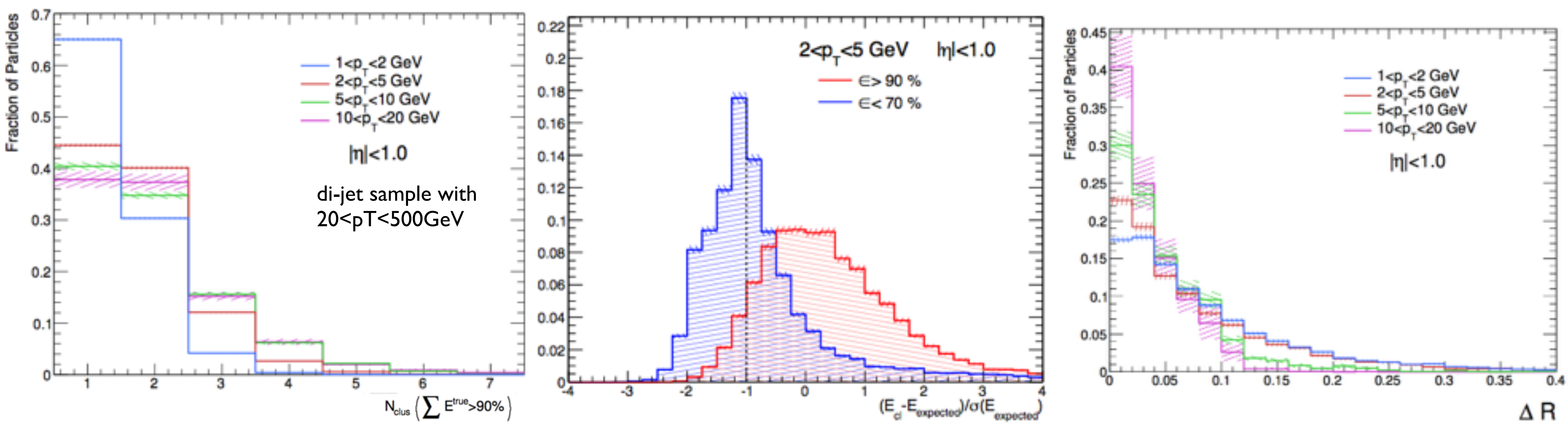
Overview of the eflowRec chain:

4. Split shower recovery:



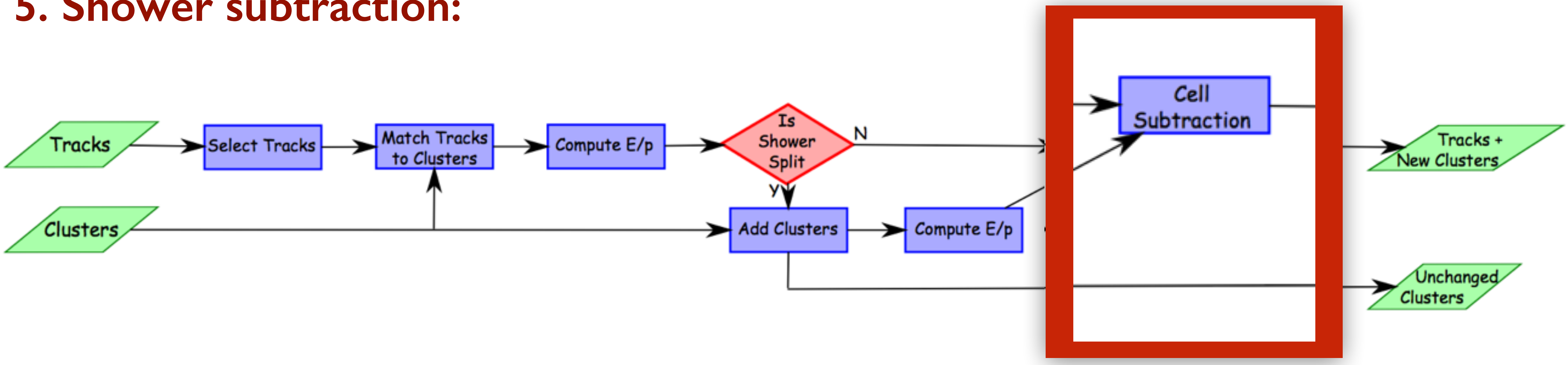
Particles do not always deposit all their energy in a single cluster.

- If E_{clus} is found to be less than E_{exp} by more than $\sigma(E_{exp})$
  - Assume energy has been split over multiple clusters (split shower recovery)
  - All clusters $\Delta R<0.2$ are considered



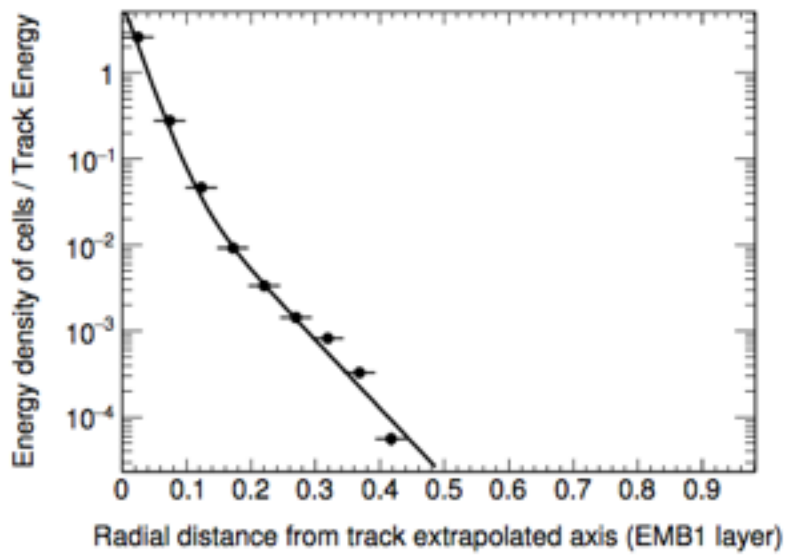
Overview of the eflowRec chain:

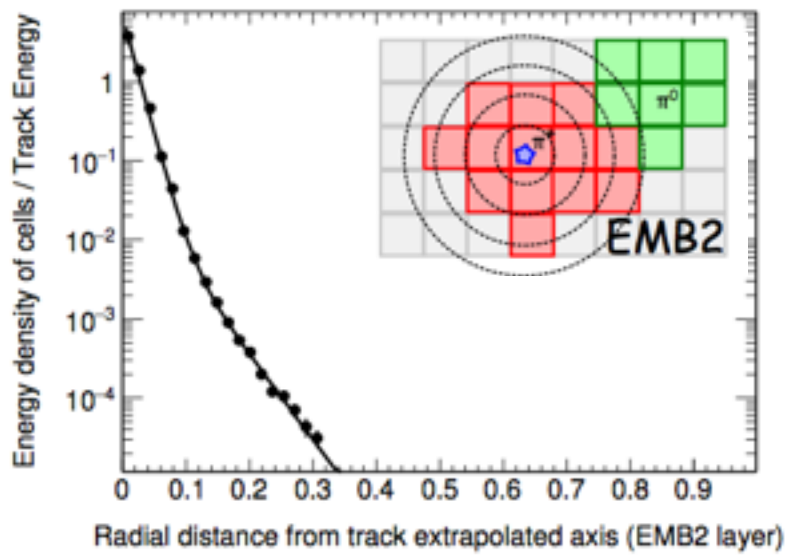
5. Shower subtraction:

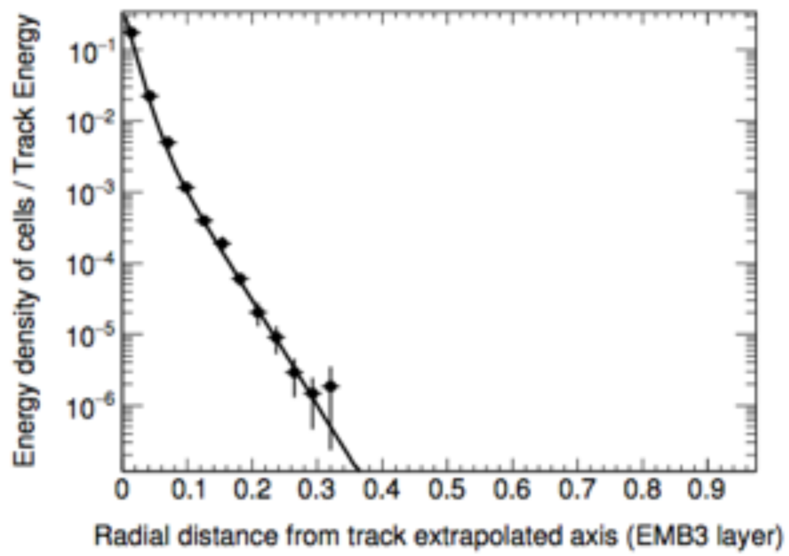


Set of clusters selected → subtraction procedure

- If E_{clus} < expected energy from E/p → completely subtracted
- If E_{clus} > expected energy from E/p → ring-by-ring subtraction
    - Parametrised shower shape in each layer (using single pion sample)
    - Using the cell size, rings in (eta, phi) around track direction defined for each layer
    - Average energy density calculated for each ring

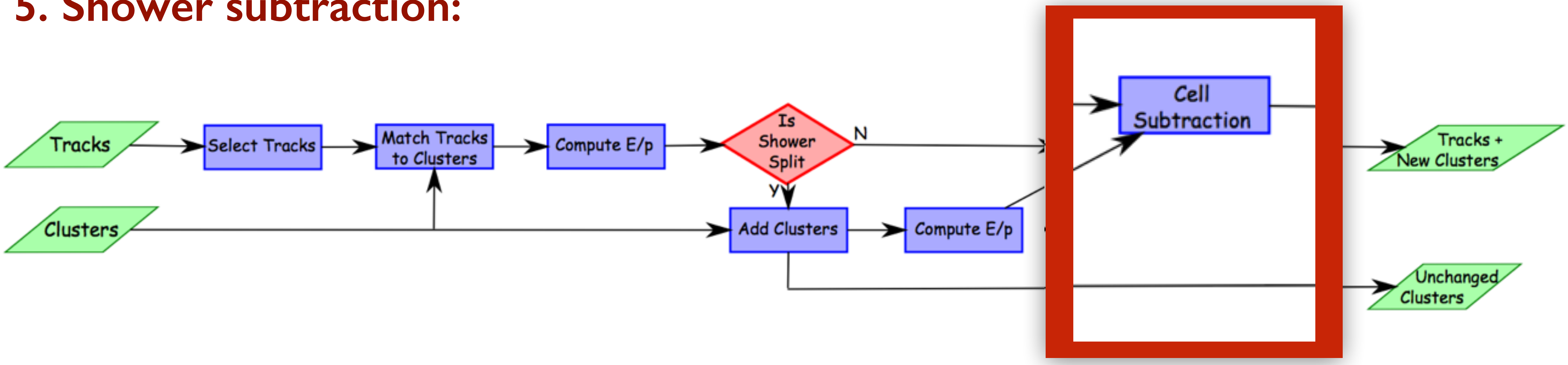






Overview of the eflowRec chain:

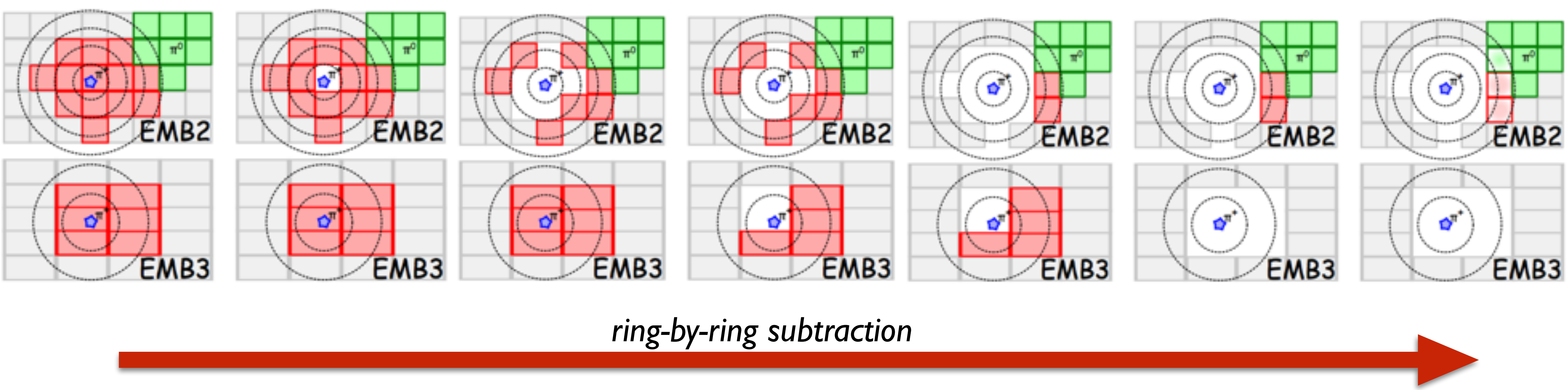
5. Shower subtraction:



Subtraction starts from the highest E density ring

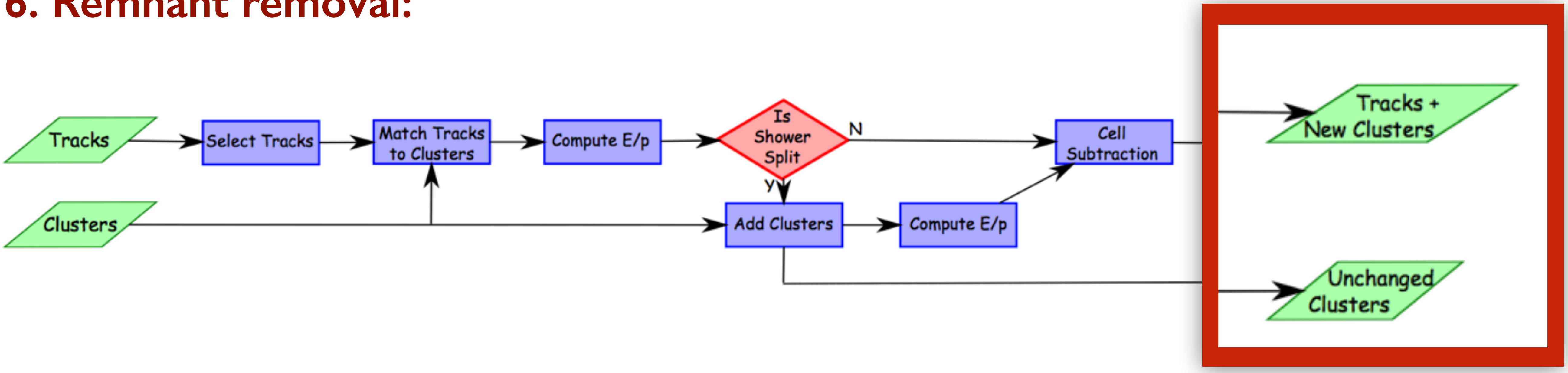
- E_{ring} < E remaining to subtract → removed
- E_{ring} > E remaining to subtract → scaled by the fraction needed

‣ The process of removing cells ring-by-ring is continued until the E_{exp} is subtracted



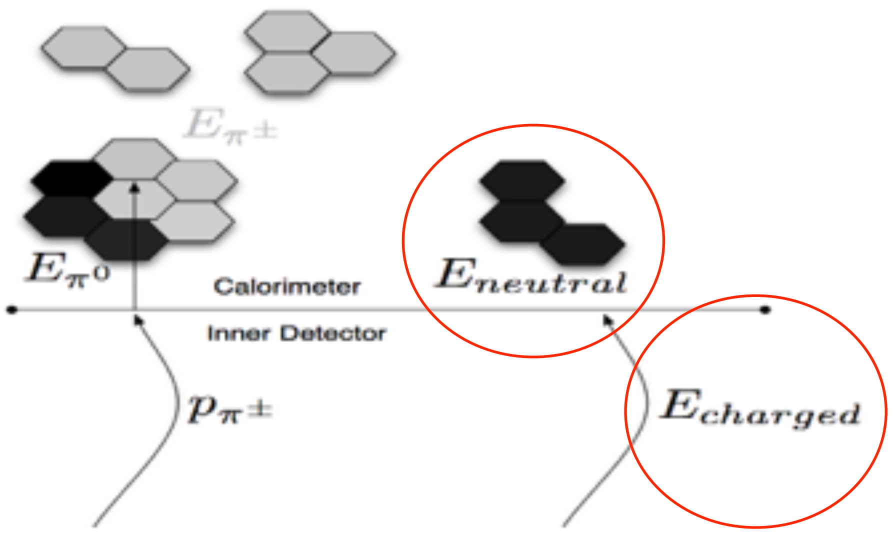
Overview of the eflowRec chain:

6. Remnant removal:

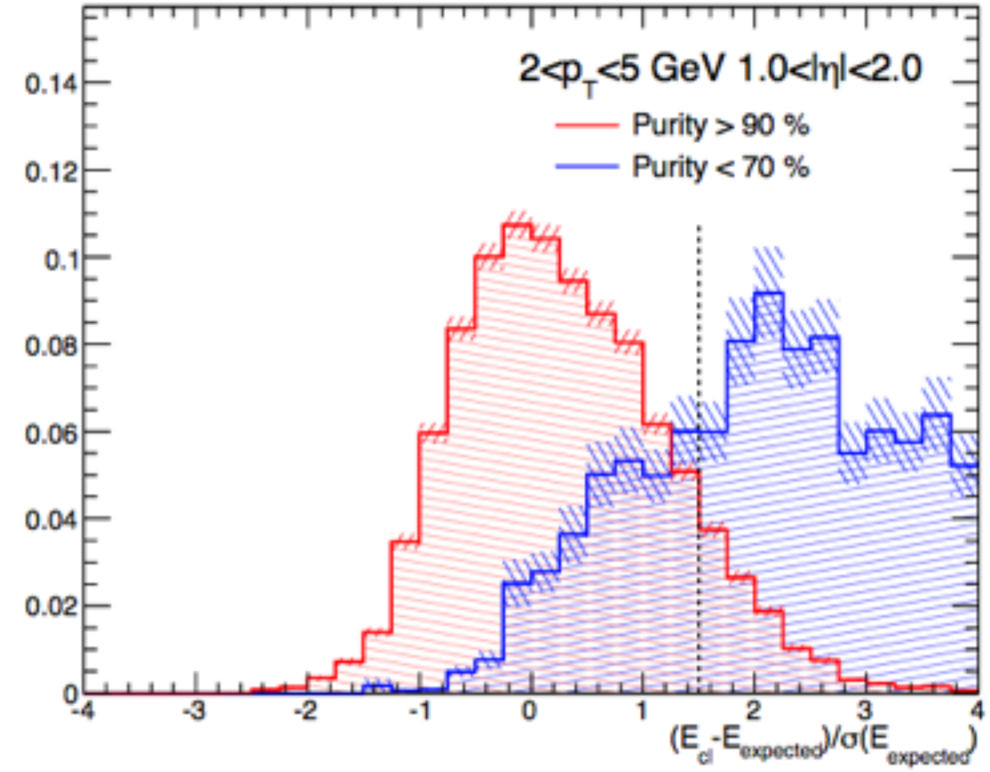


Subtraction stops when E_{exp} has been removed

- If the remaining E is consistent with σ(E/p) → purely noise → totally removed
- If the remaining E larger than σ(E/p) → other particles involved → kept



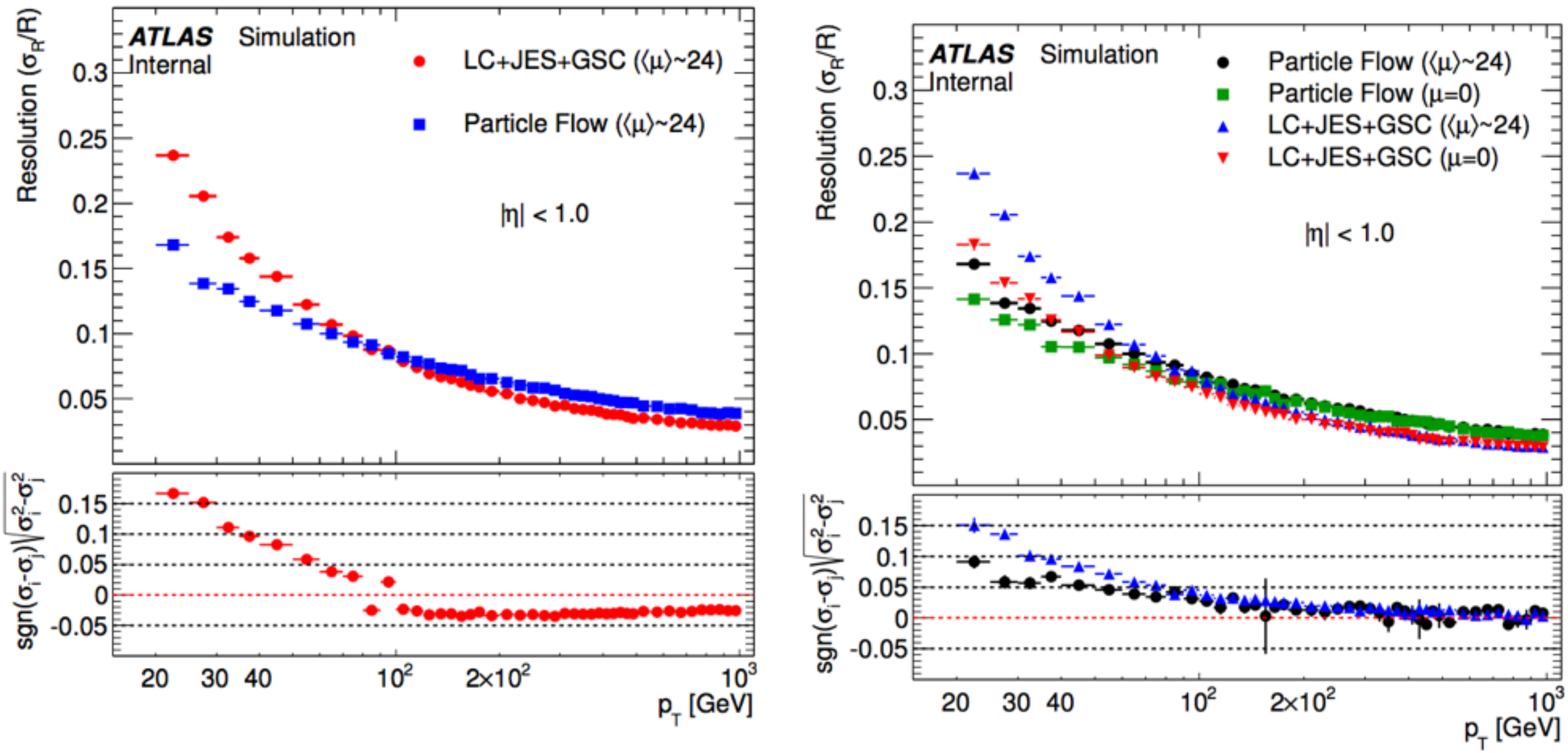
The selected tracks and the remained clusters represent the reconstructed event



Particle flow performance studies

Results at $\sqrt{s}$ = 8 TeV

Jet P_T resolution

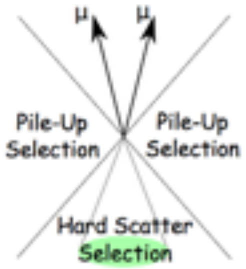


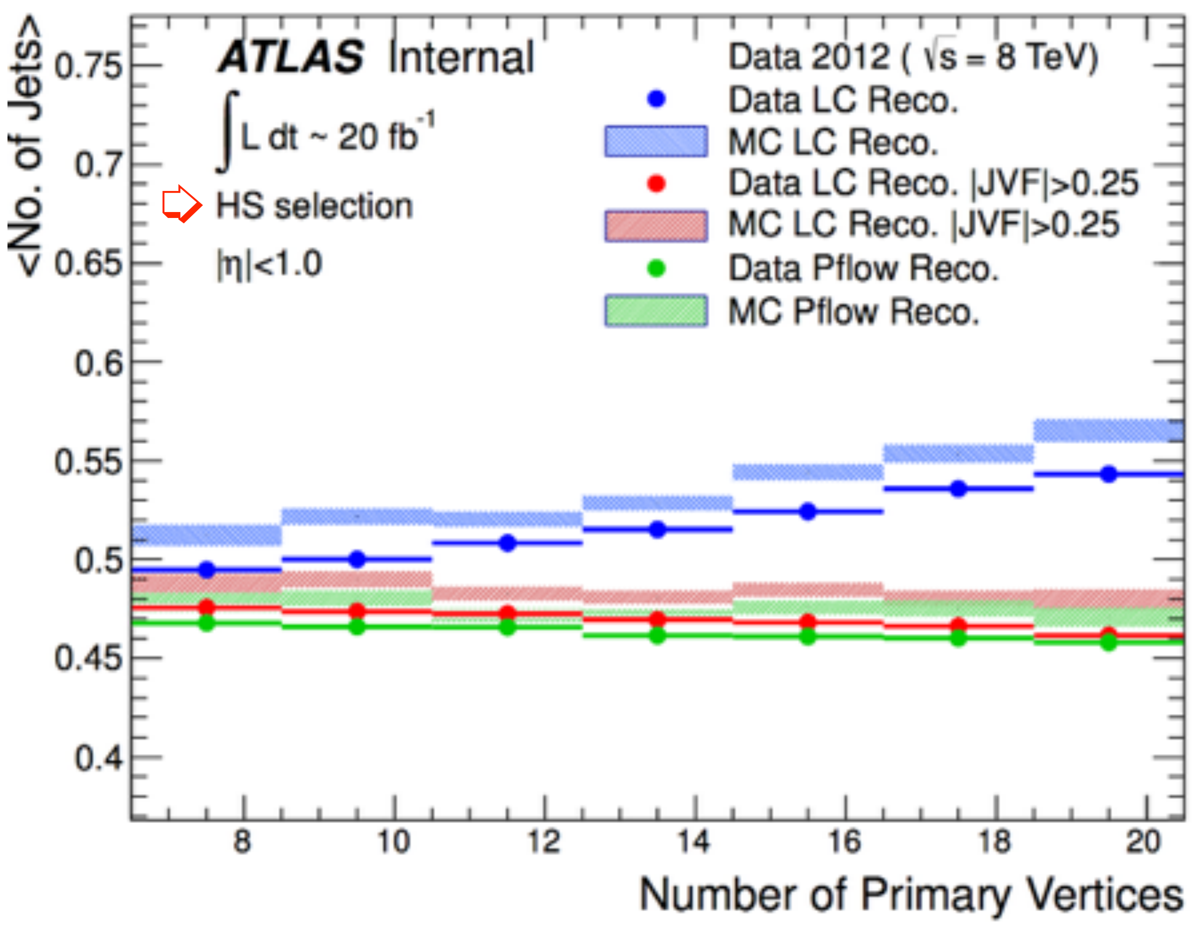
Particle flow performance studies

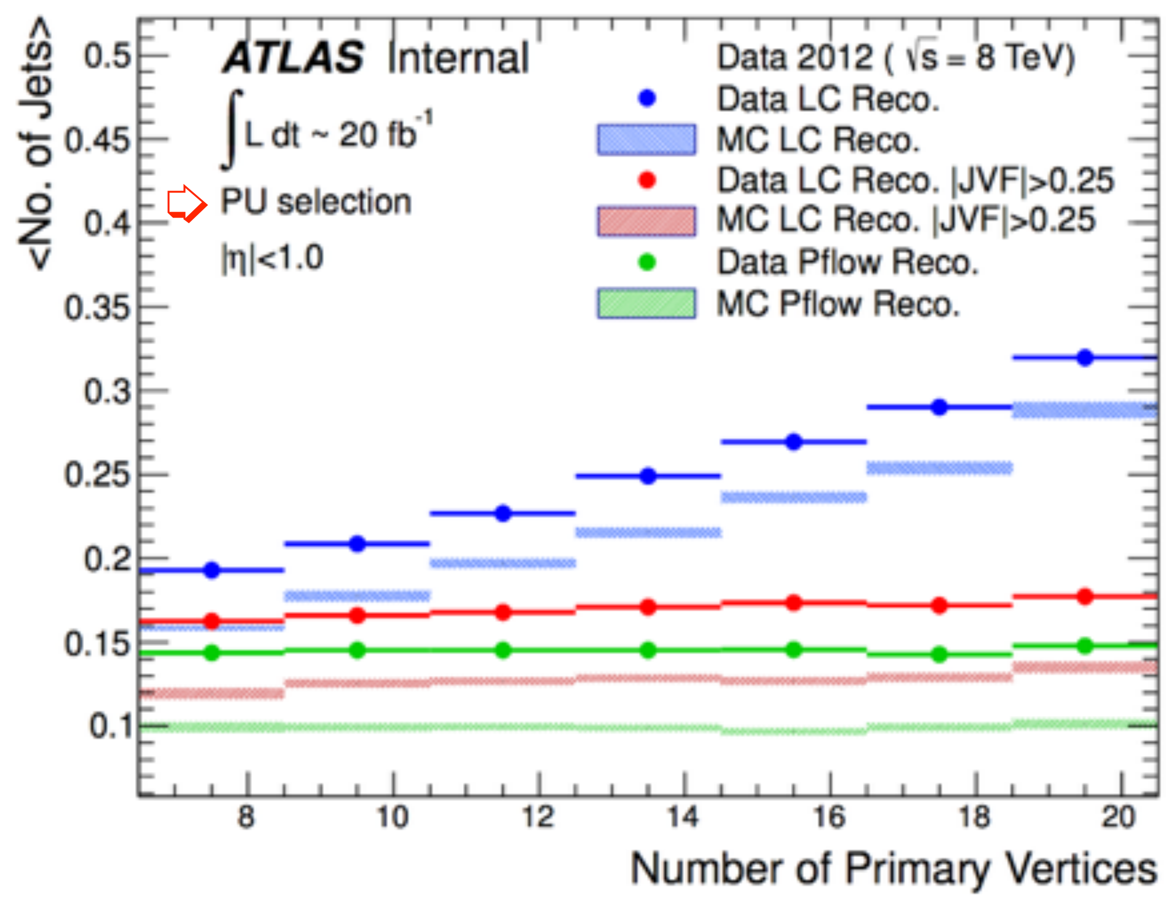
Results at $\sqrt{s}$ = 8 TeV

Pile-up

- Average number of particle flow jets originating form pile-up is much suppressed.
- Average number of particle flow jets is stable as a function of pile-up.
- Particle flow jets behaviour flatter than the LC and LC+JVF jets
  - Hard scatter (HS) selection
  - Pileup (PI) selection







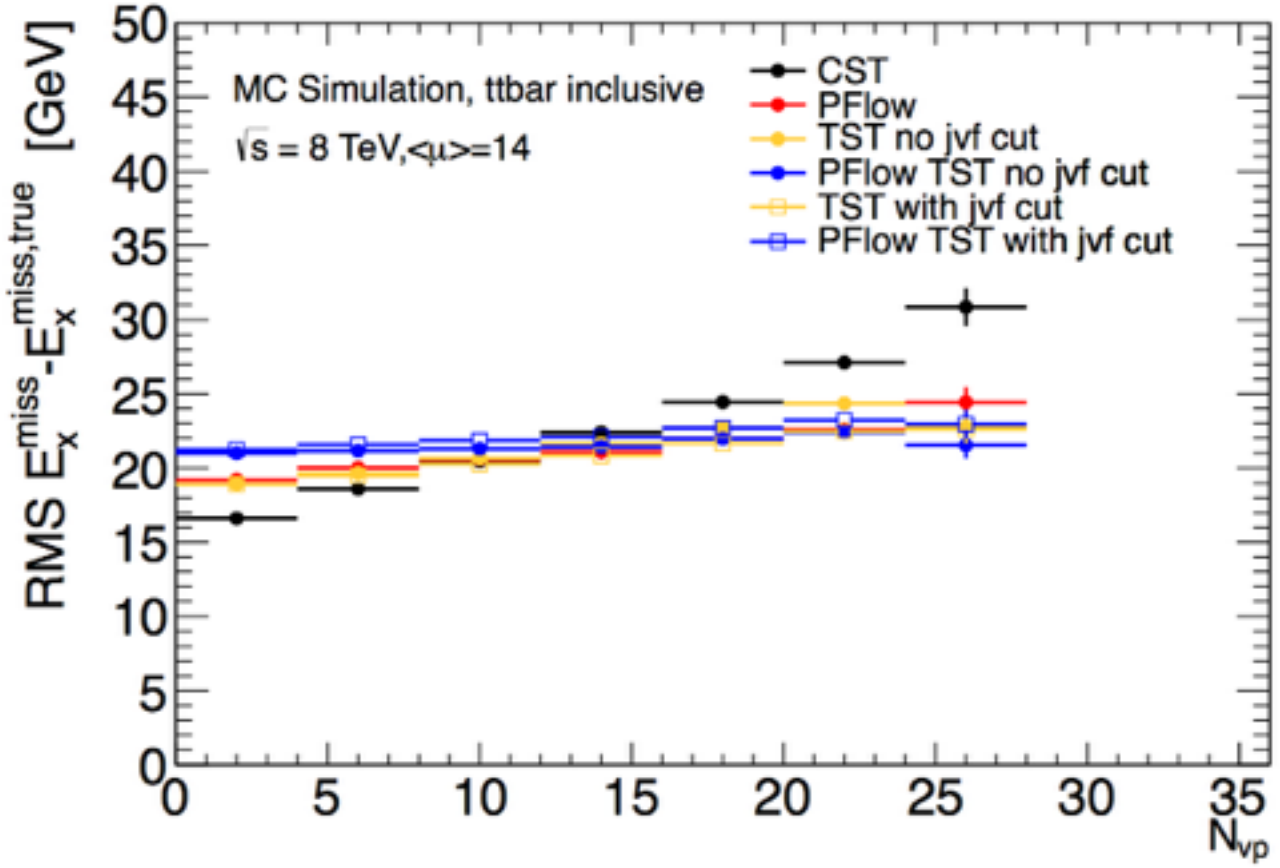
Particle flow performance studies

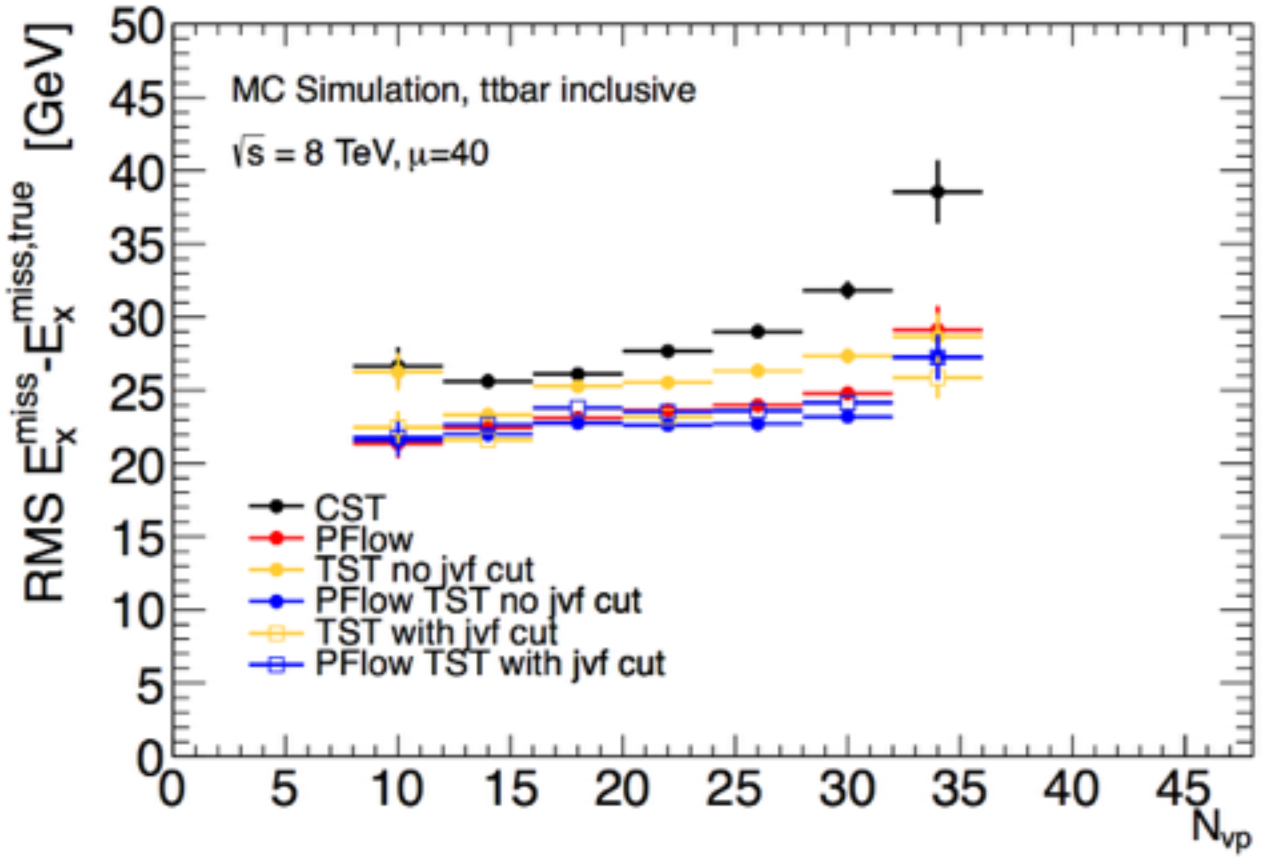
Results at $\sqrt{s}$ = 8 TeV

E_T^{miss}

$$E_{x,y}^{\text{miss}} = -\left(\Sigma E_{x,y}^{\text{particle flow jets}} + \Sigma p_{x,y}^{e} + \Sigma p_{x,y}^{\gamma} + \Sigma p_{x,y}^{\mu} + \Sigma p_{x,y}^{\text{trk, unassociated}} + \Sigma E_{x,y}^{\text{clu, unassociated}}\right)$$

- ttbar MC events with high and low pileup conditions
  - PFlow
  - CST (cluster soft term)
  - TST (track soft term)
  - PFlow TST (PFlow track soft term)





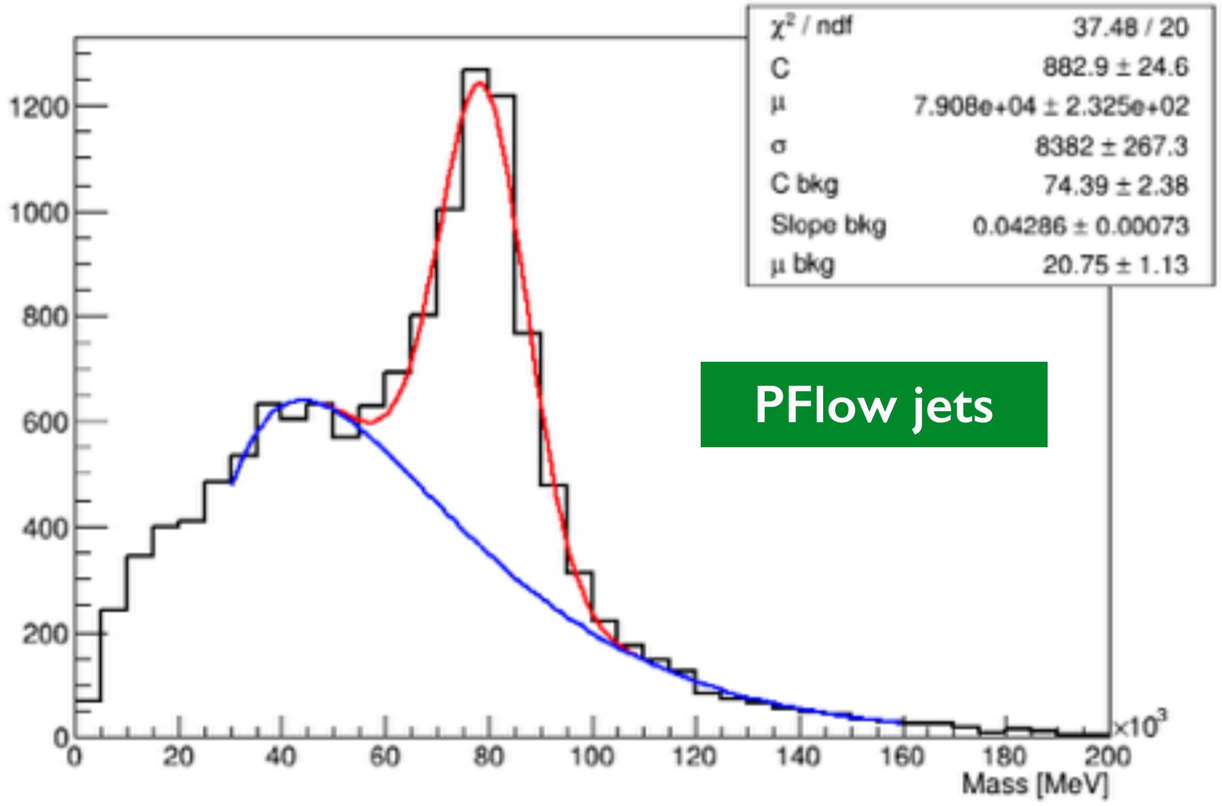
Particle flow performance studies

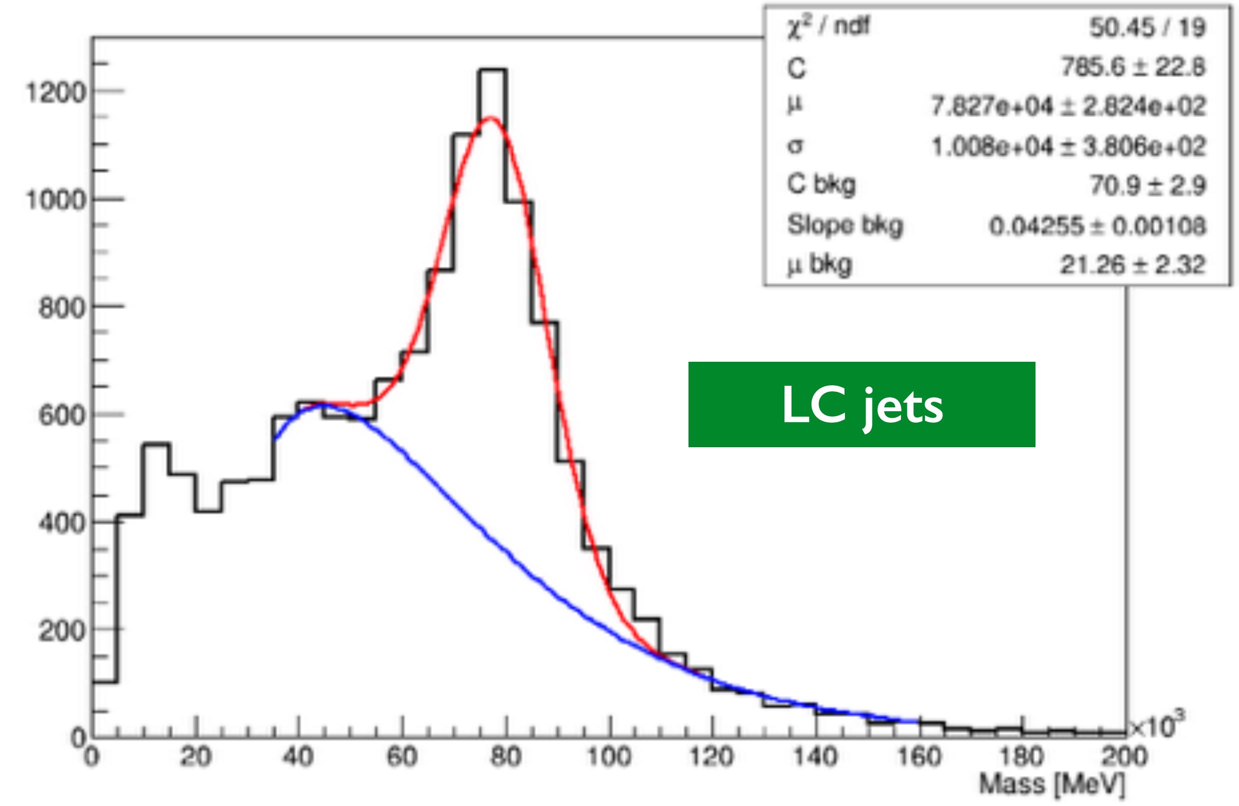
Results at $\sqrt{s}$ = 8 TeV

Large R-jets in ttbar events

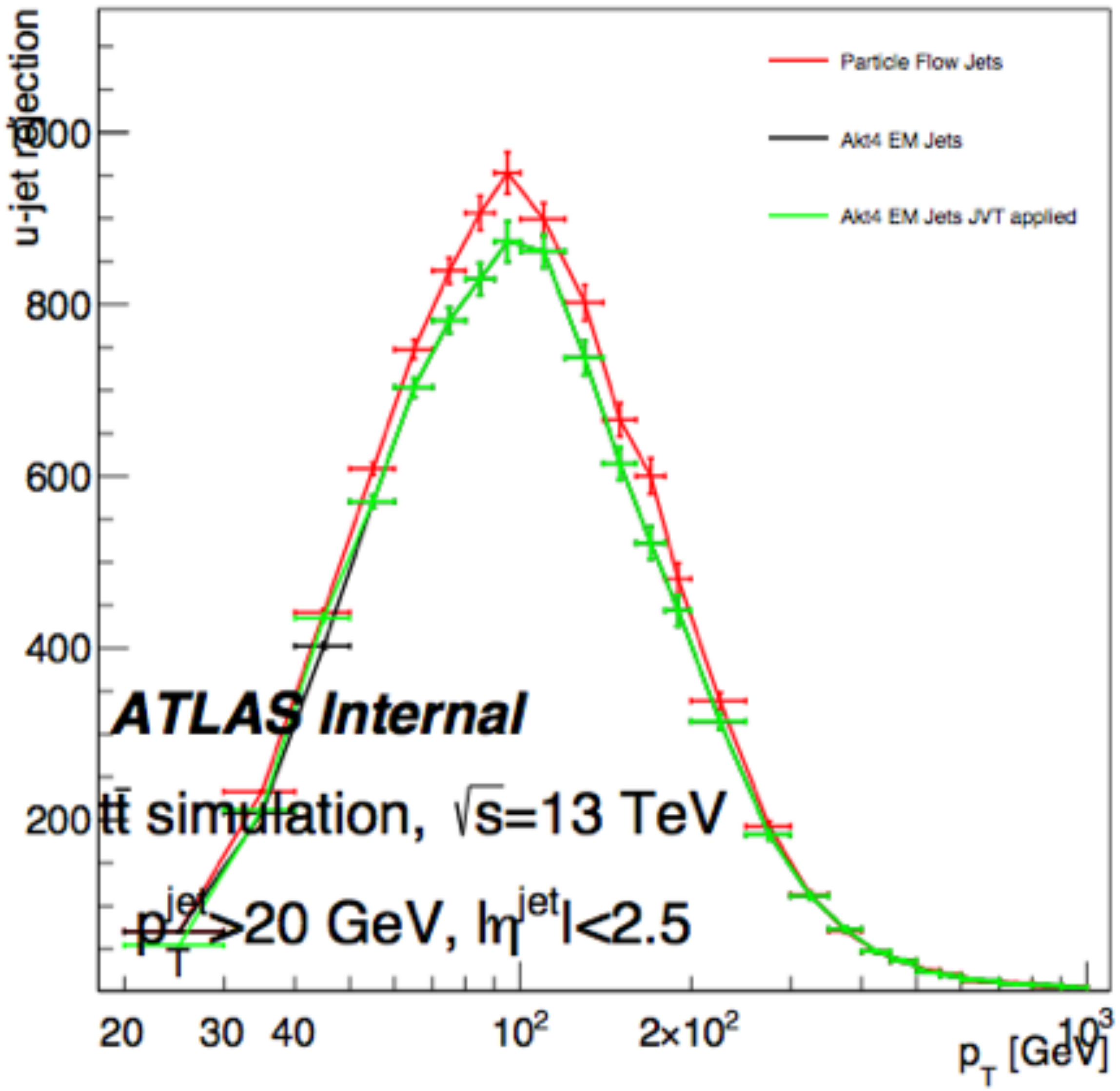
- Invariant mass reconstruction of the hadronic W boson in ttbar events
  - Signal jets' those arising from a W-boson hadronic decay
  - Improvement in the background rejection
  - Improvements in the W mass resolution

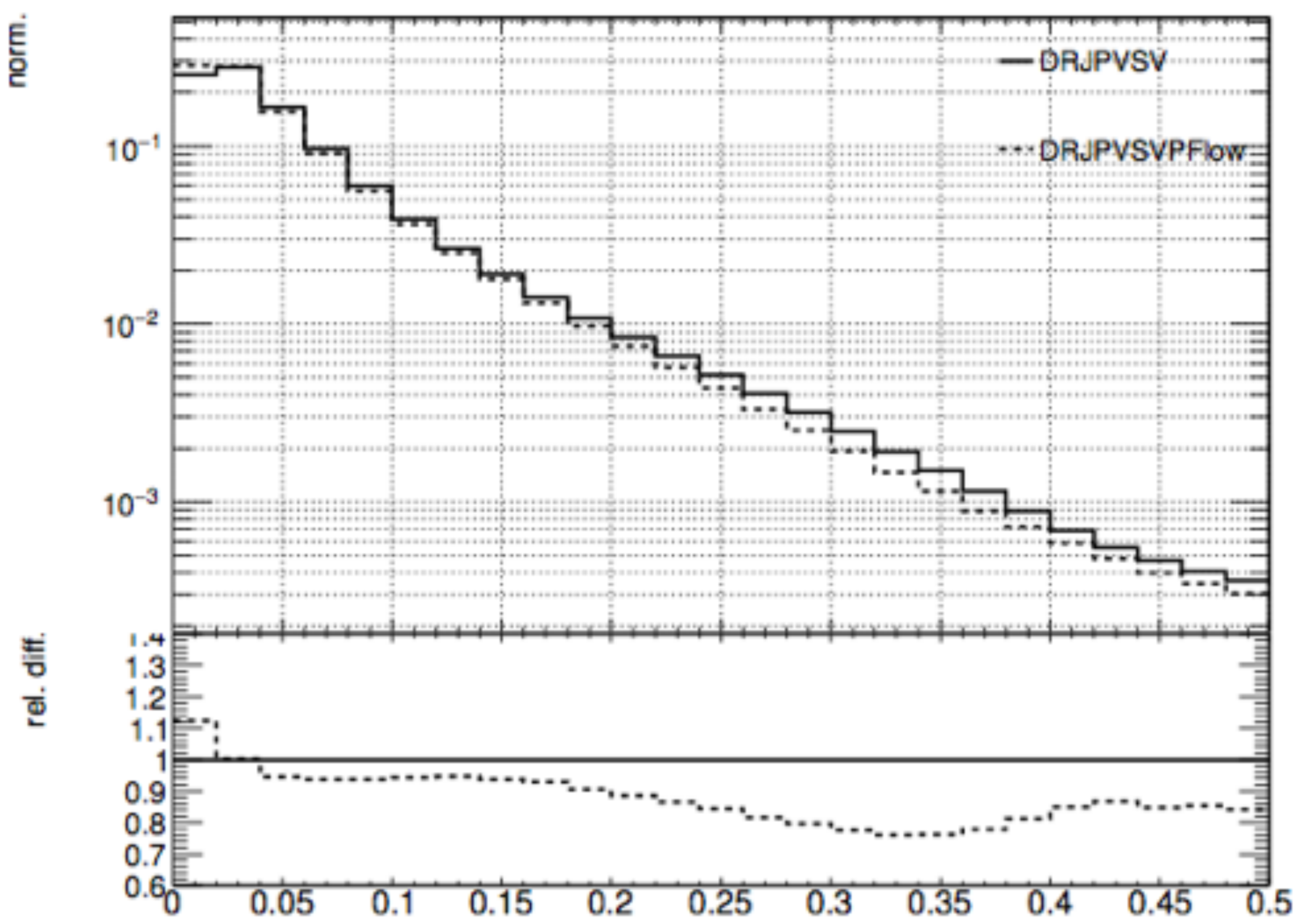
    σ = 11% (**Pflow**) σ = 13% (**LC**)





b-jets and PFlow in ttbar events

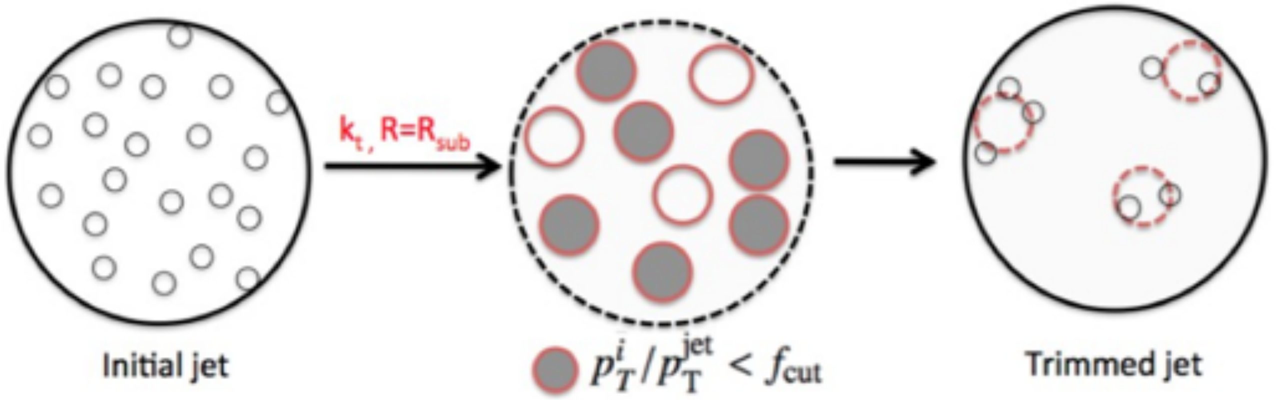




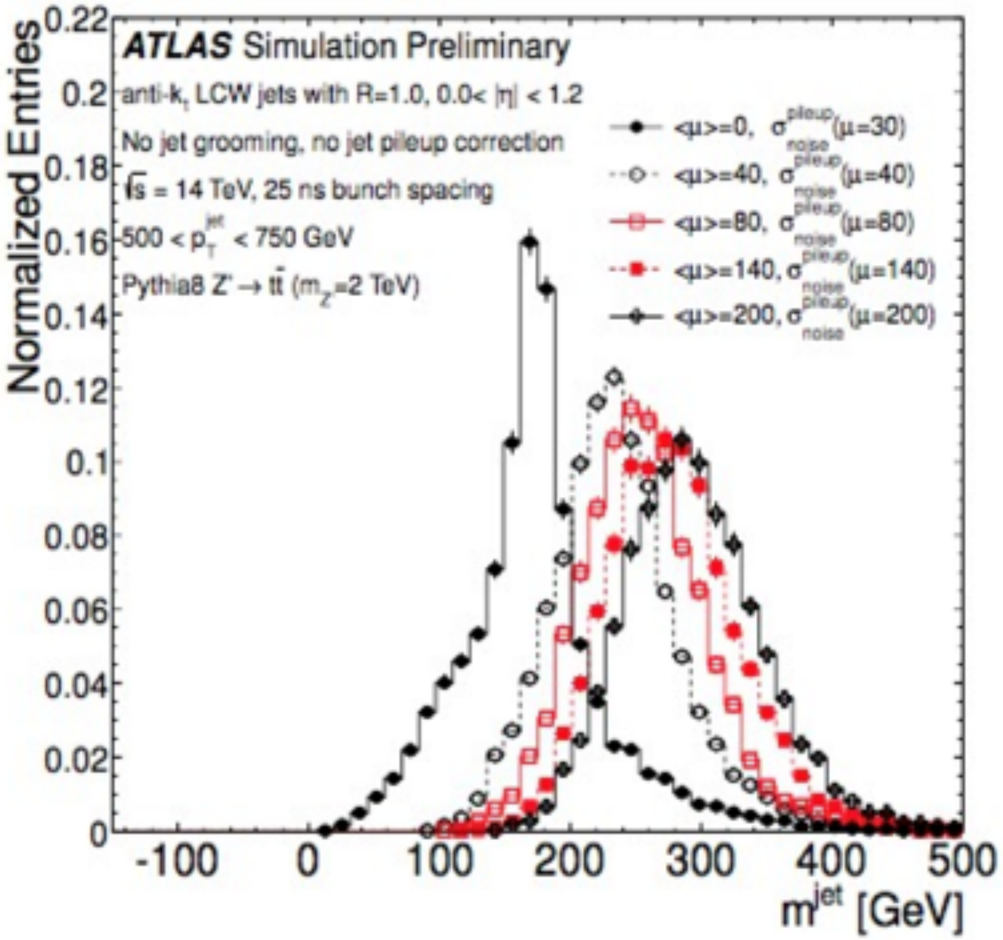
Grooming techniques: trimming

from R.Camacho

- Trimming: http://arxiv.org/abs/0912.1342
  - Reclustering of constituents of large-R jet into small-R jets of size R_{sub}
  - Remove subjet *i* if $p_T^i < f_{cut} \times p_T^{jet}$
  - Default ATLAs groomer (stable agains PU)

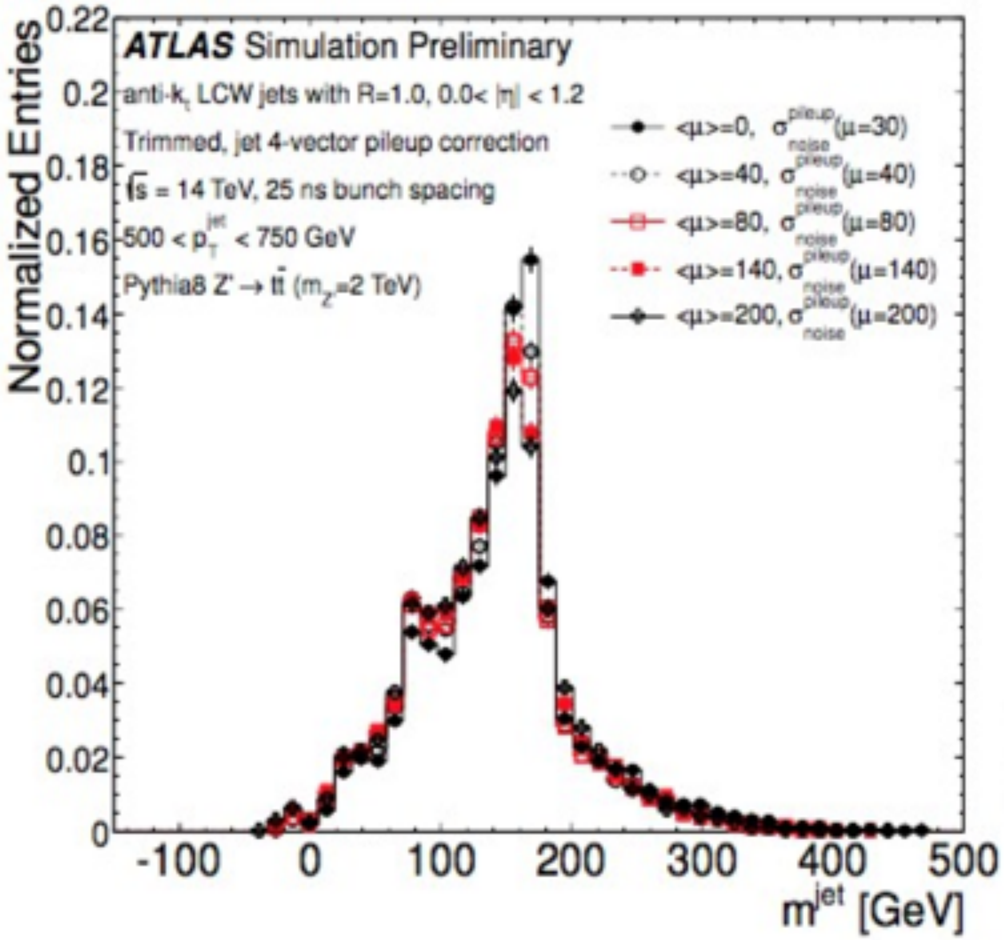


ATLAS public substructure plots



trimming →

ATLAS public substructure plots



Top-tagging jet requirements

ATL-PHYS-PUB-2015-053

- **Suitable inputs:**
  - Calibrated Topo-clusters → calorimeter jets
  - Stable particles with lifetime > 10 ps → generator level jets

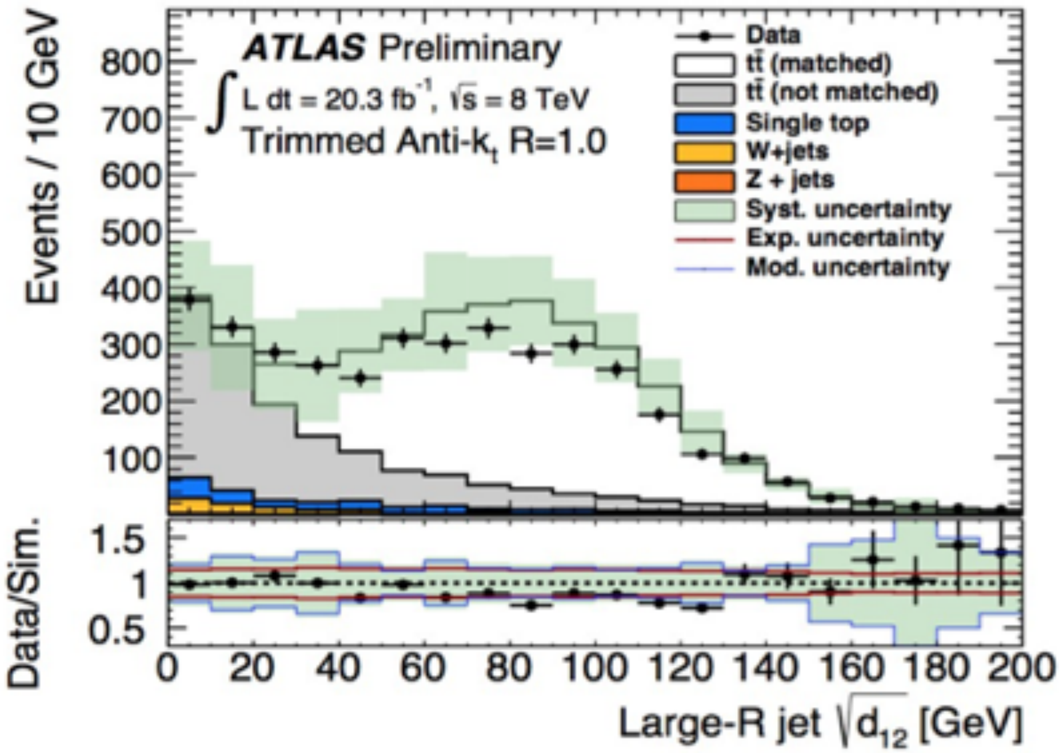
- **Choice of jet algorithm and Grooming:**
  - AntiKt algorithms with R=1.0
  - Jets are trimmed to mitigate the effect of pileup
    - The constituents are reclustered with Kt algorithm wiry R_{sub}=0.2
    - Jets with a smaller fraction than f_{cut}=0.05 of the parent jet are removed
  - Surviving jets recombined to produce the final trimmed jets
  - Only those jets with more than 2 constituents are kept

- **Calibration methodology:**
  - Calorimeter energy measured at EM scheme
  - LCW classify topological cluster in EM or HAD
  - Energy correction derived form single pions MC samples to HAD clusters
  - Extra step for large R-jets: ***mass calibration*** based on the jet mass response
  - JES calibration:
    - Calorimeter response to the true energy (Pythia MC)
    - No jet offset correction from pile-up contributions (trimming)

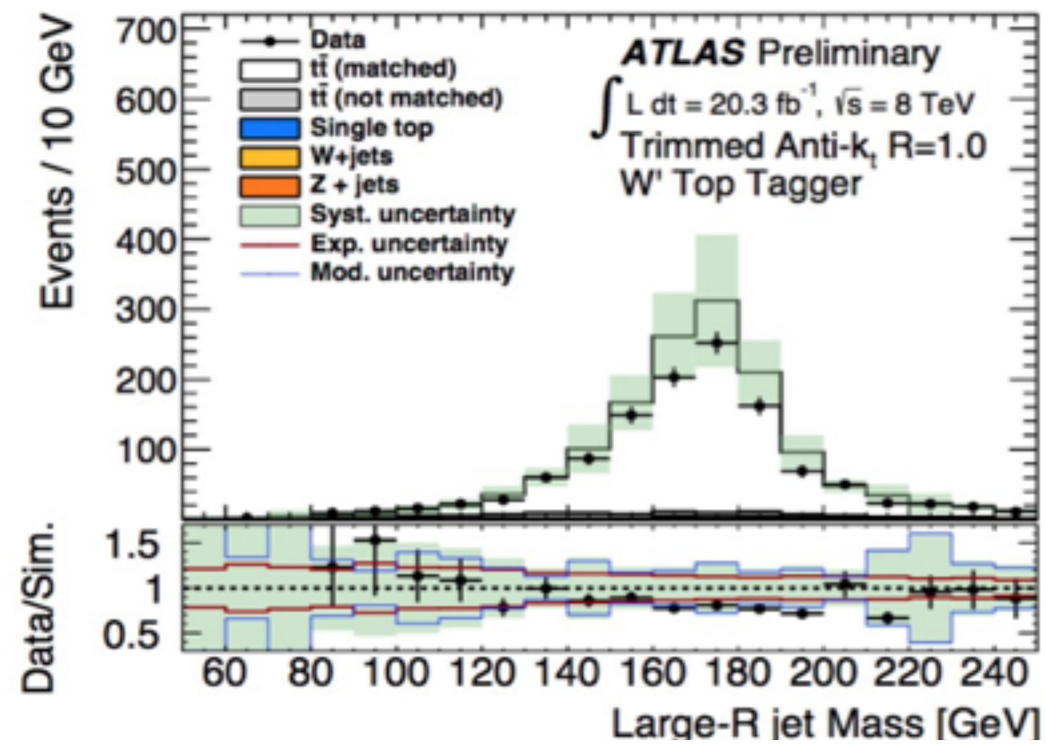
Inputs for simple substructure taggers

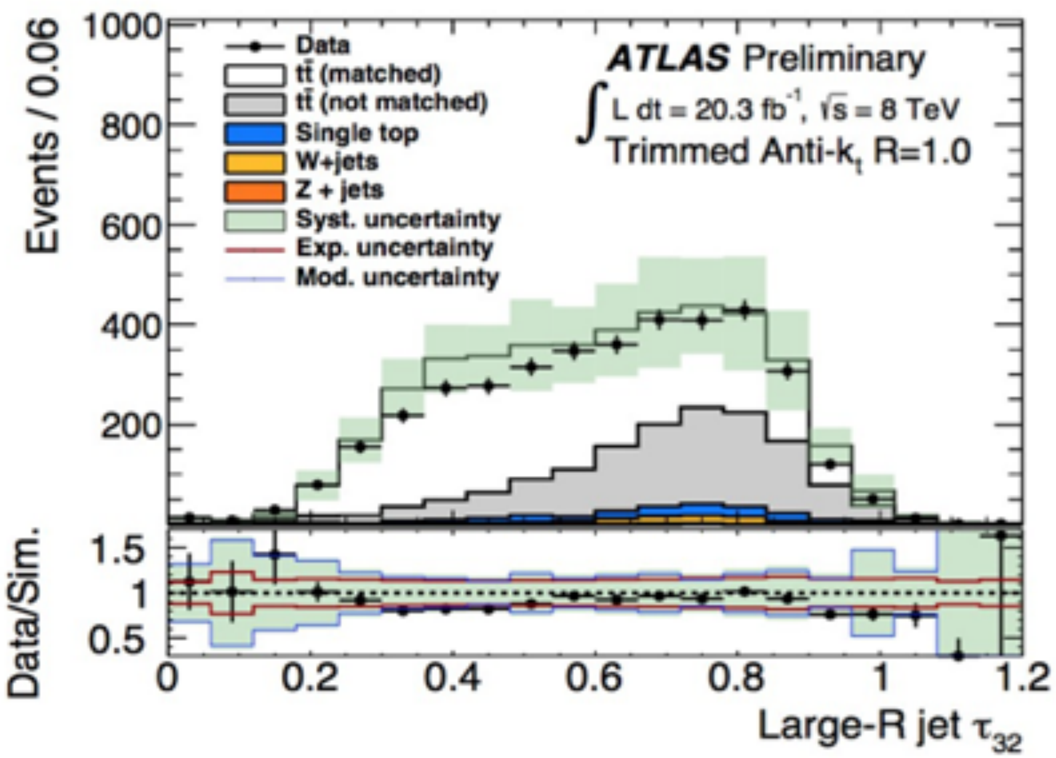
Lepton+jets tt selection (ATLAS-CONF-2015-036)

Use hadronic side to study the taggers



tagger	top tagging criterion
Substructure tagger I	$\sqrt{d_{12}} > 40$ GeV
Substructure tagger II	$m > 100$ GeV
Substructure tagger III	$m > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV
Substructure tagger IV	$m > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV and $\sqrt{d_{23}} > 10$ GeV
Substructure tagger V	$m > 100$ GeV and $\sqrt{d_{12}} > 40$ GeV and $\sqrt{d_{23}} > 20$ GeV
W' top tagger	$\sqrt{d_{12}} > 40$ GeV and $0.4 < \tau_{21} < 0.9$ and $\tau_{32} < 0.65$



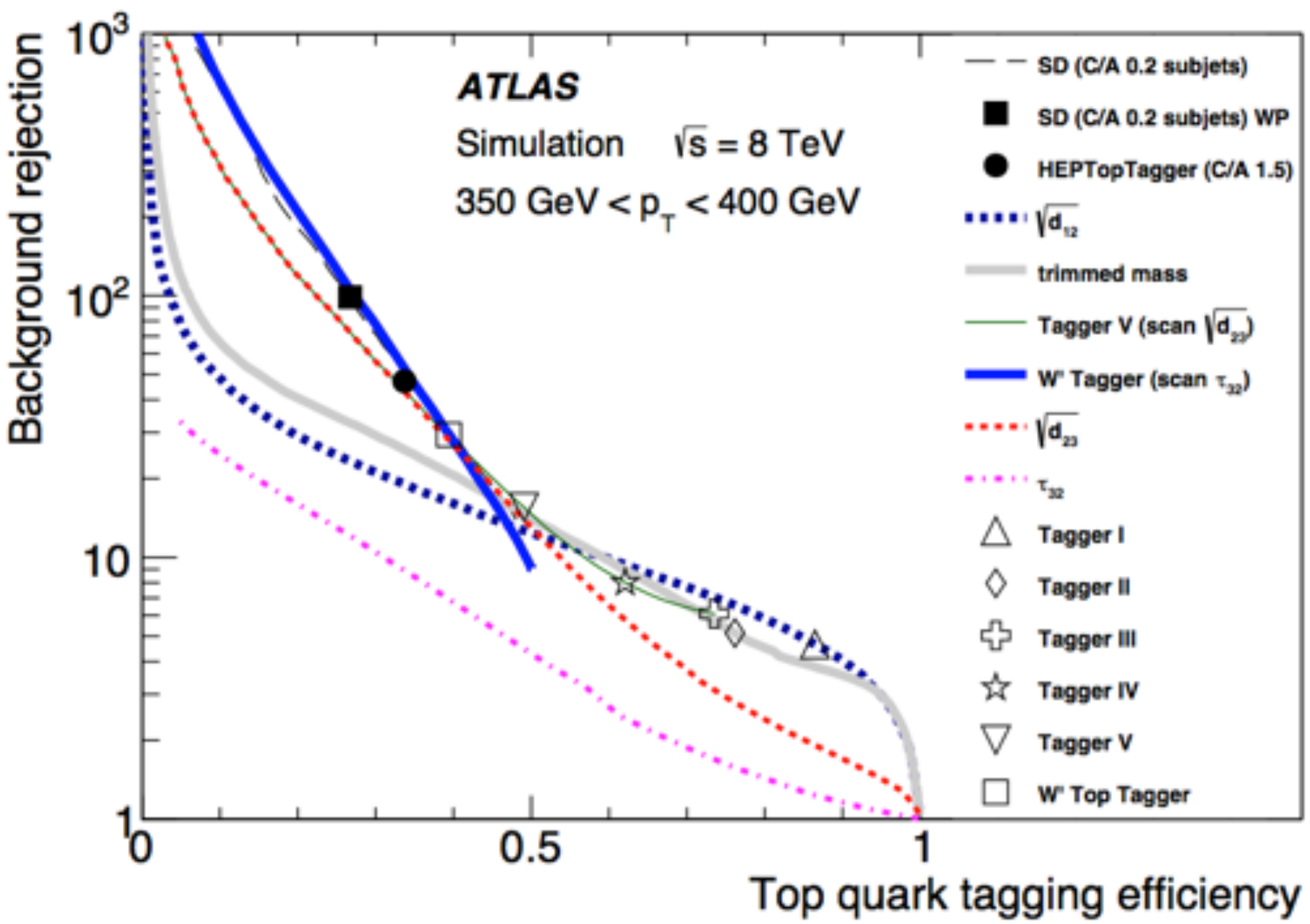


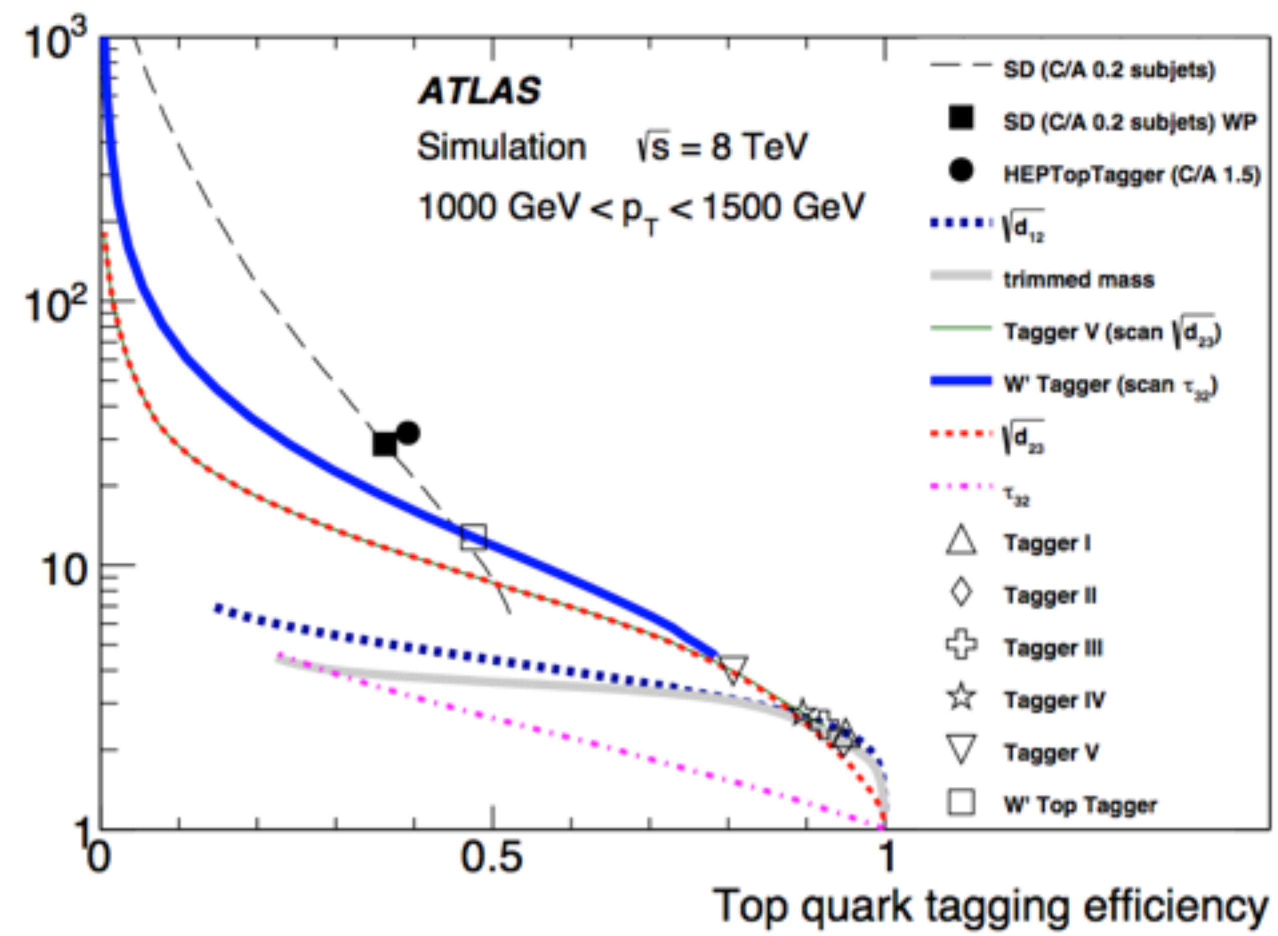
$$\tau_N = \frac{1}{d_0} \sum_k p_{Tk} \times \min(\delta R_{1k}, \delta R_{2k}, \ldots, \delta R_{Nk})$$

$$d_0 = \sum_k p_{Tk} \times R,$$

Boosted top-tagging algorithms

Top-tagging algorithms in Run I

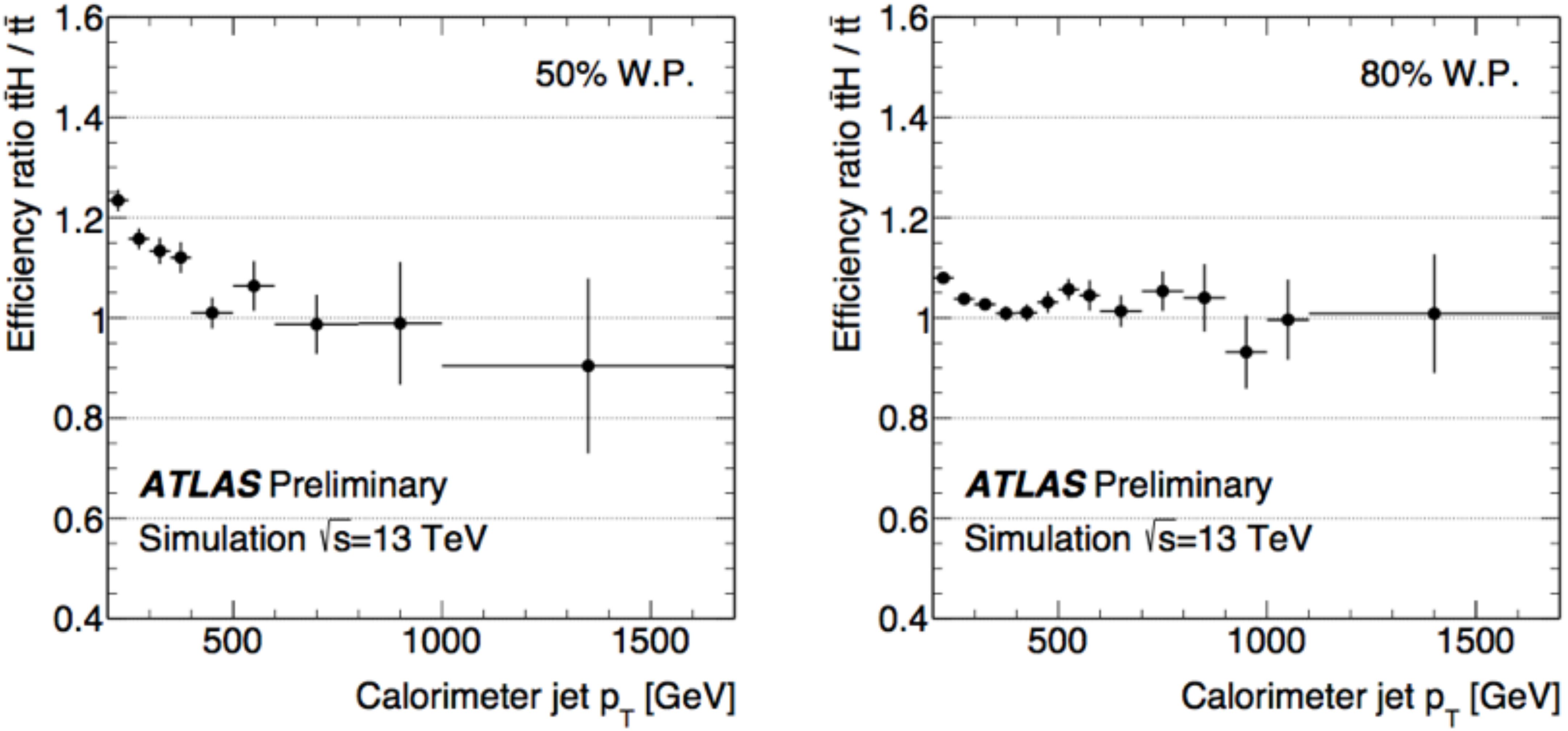




Softly boosted → Highly boosted

Top-tagging performance in ttH

- **Performance studies in ttH:**
  - The presence of H decay products can affect the large-R jets.
  - Are busier environments affecting the efficiency of the tagger?
    - responses for the m and τ_{32} variables were studied in ttbar and ttH:



The top-tagging efficiency is not diminished in busier environments