

Boosting precision measurements using jet substructure

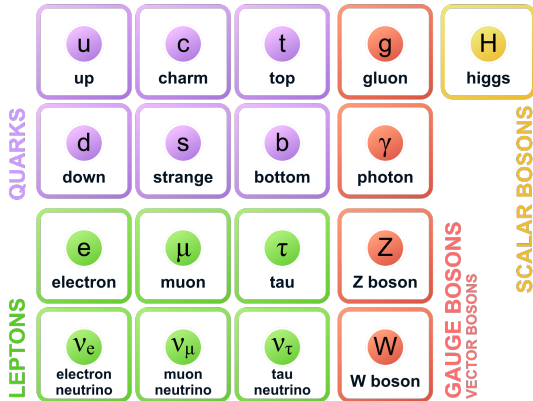


Chris Malena Delitzsch

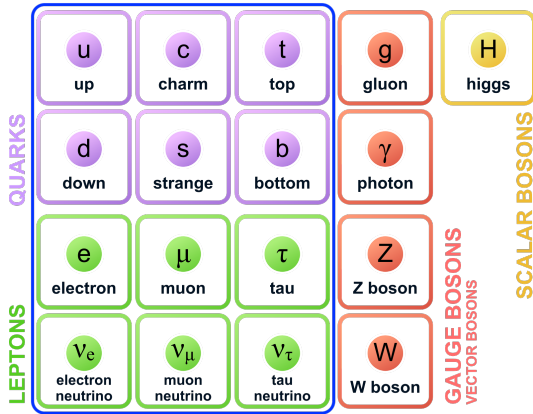
Run: 299584

Event: 563621388

The Standard Model (SM) of Particle Physics



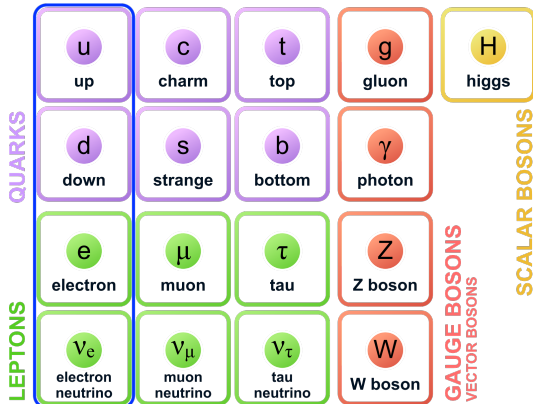
Particle content



Particle content

- *Fermions* (Quarks and leptons)
 - Spin 1/2 particles

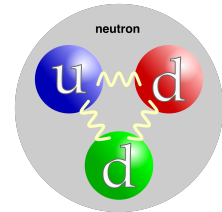
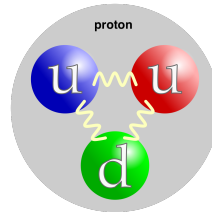
The Standard Model (SM) of Particle Physics

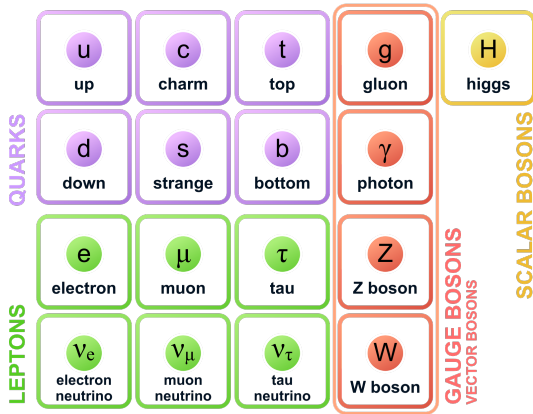


↑
building blocks of matter

Particle content

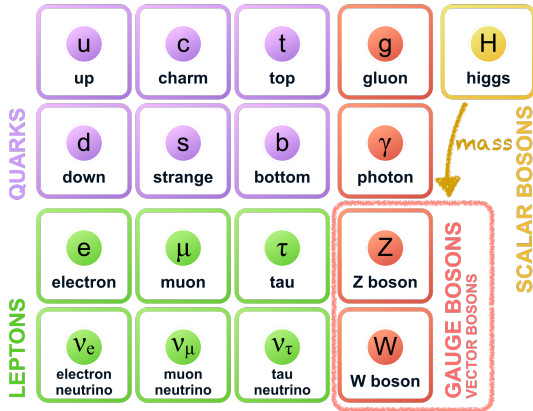
- *Fermions* (Quarks and leptons)
 - Spin 1/2 particles





Particle content

- *Fermions* (Quarks and leptons)
 - Spin 1/2 particles
- Interactions mediated via *gauge bosons*
 - Spin 1



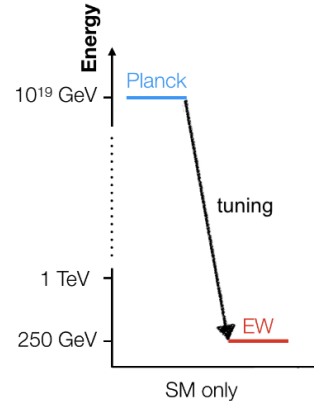
Particle content

- *Fermions* (Quarks and leptons)
 - Spin 1/2 particles
- Interactions mediated via *gauge bosons*
 - Spin 1
- Masses included via Higgs mechanism
 - Higgs boson: Spin 0

- Large differences between the Planck scale and the scale of electroweak SM sector

$$\text{Higgs Boson } (125 \text{ GeV})^2 = \text{raw mass } (10^{36} \text{ GeV}^2) - \text{quantum corrections } (10^{36} \text{ GeV}^2)$$

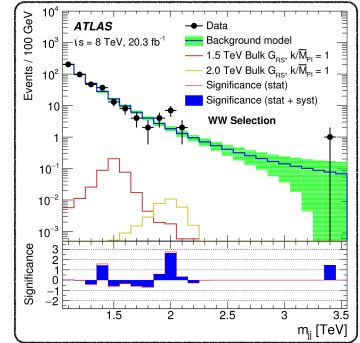
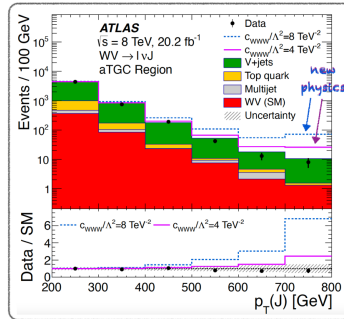
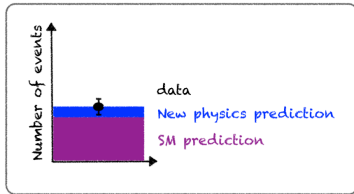
$$\begin{array}{r} 123456789012345678901234567890123456 \\ - \\ 123456789012345678901234567890107831 \\ = \\ \mathbf{15625 = (125)^2} \end{array}$$



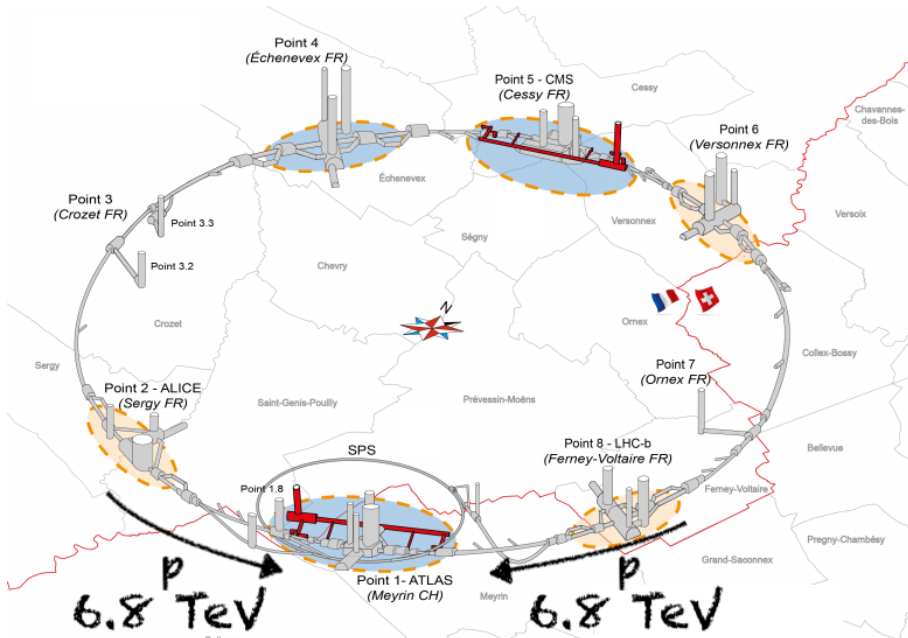
- Physics beyond the Standard Model will reduce these large corrections

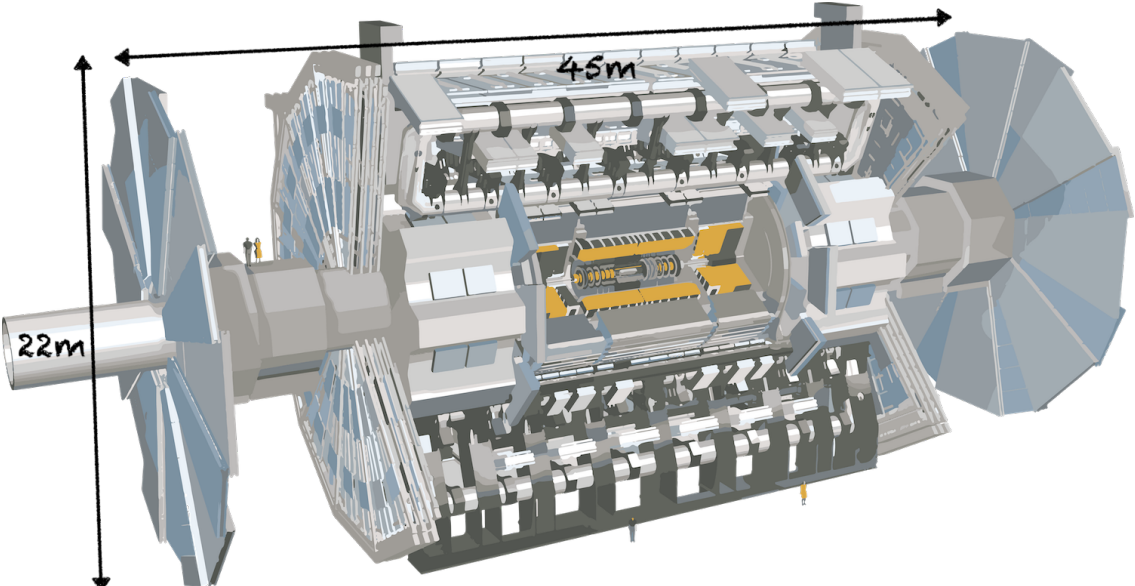
How to search for physics beyond the SM?

■ Indirect and direct searches for BSM physics



- We are specifically sensitive to BSM physics at high transverse momenta
 - Hadronic decays of interest due to their large branching ratio
 - I am particularly interested in processes involving top quarks and W/Z bosons
- We can only search for new physics if we understand the SM with utmost precision

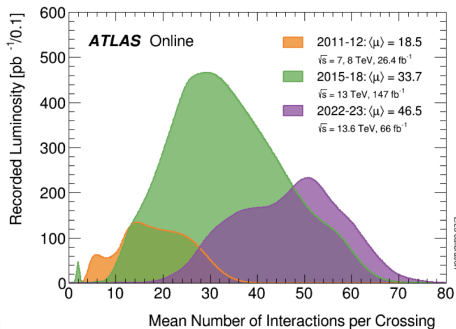
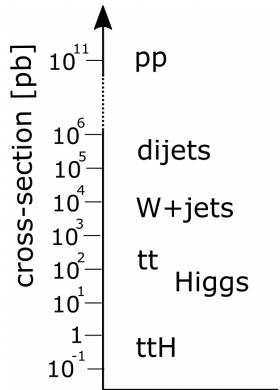




Challenge at the LHC - Pile-up

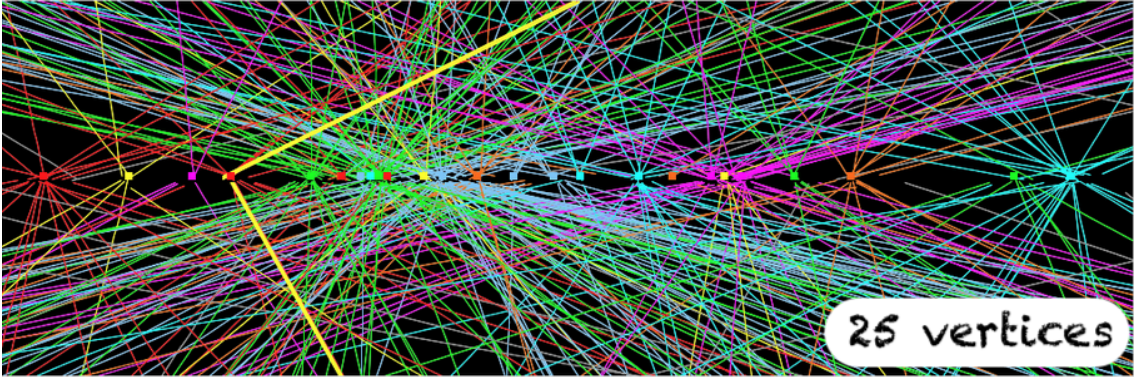
$$N = \mathcal{L} \cdot \sigma \xrightarrow{\text{theory}}$$

N → number of events of specific process
 \mathcal{L} → integrated luminosity
 σ → cross-section



LuminosityPublicResultsRun3

How does this look like in our detector?

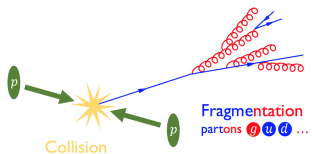


- We're only interested in one of these collisions and the associated particles

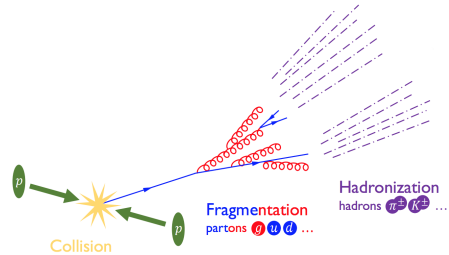
- We cannot directly measure quarks from the decay of heavy objects
 - $W \rightarrow q\bar{q}'$, $Z \rightarrow q\bar{q}$, $t \rightarrow q\bar{q}'b$



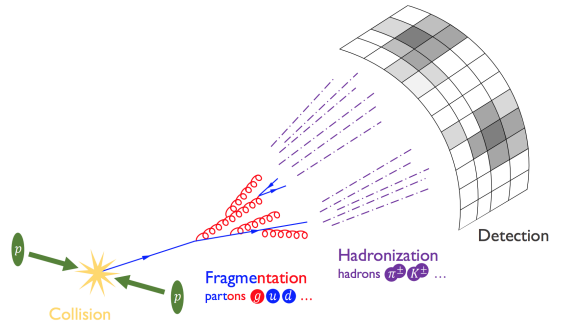
- We cannot directly measure quarks from the decay of heavy objects
 - $W \rightarrow q\bar{q}'$, $Z \rightarrow q\bar{q}$, $t \rightarrow q\bar{q}'b$



- We cannot directly measure quarks from the decay of heavy objects
 - $W \rightarrow q\bar{q}'$, $Z \rightarrow q\bar{q}$, $t \rightarrow q\bar{q}'b$



- We cannot directly measure quarks from the decay of heavy objects
 - $W \rightarrow q\bar{q}'$, $Z \rightarrow q\bar{q}$, $t \rightarrow q\bar{q}'b$

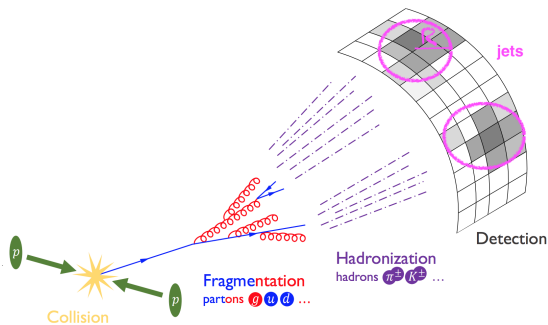


- We cannot directly measure quarks from the decay of heavy objects

- $W \rightarrow q\bar{q}'$, $Z \rightarrow q\bar{q}$, $t \rightarrow q\bar{q}'b$

- Instead define proxy for quarks - *jet*

- Collimated spray of hadrons (cone)
 - Clustering based on distance parameter

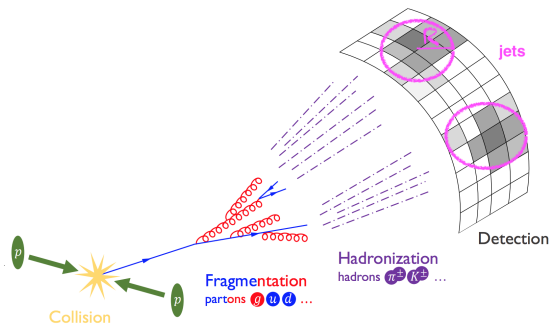


- We cannot directly measure quarks from the decay of heavy objects

- $W \rightarrow q\bar{q}', Z \rightarrow q\bar{q}, t \rightarrow q\bar{q}'b$

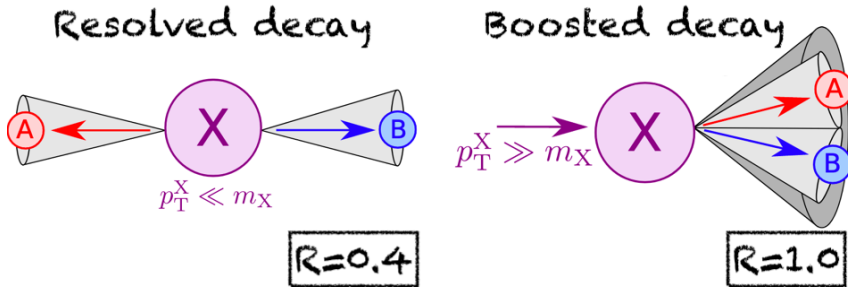
- Instead define proxy for quarks - *jet*

- Collimated spray of hadrons (cone)
 - Clustering based on distance parameter



- We would expect e.g. two (three) jets for the decay of a W/Z boson (top quark)
 - Jets have distance parameter of $R = 0.4$

- What happens at high p_T with decay products?
 - Decay products are collimated such that hadrons from quarks start overlapping
 - Reconstruct decay products instead as **single $R = 1.0$ jet**



Rule of thumb

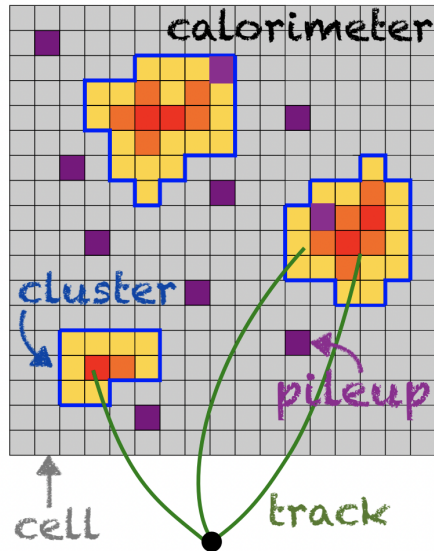
$$\Delta R \approx \frac{2 \cdot m_X}{p_T}$$

- Resolved W boson decay: two small- R jets with $m^2 = (p_1 + p_2)^2$
- Boosted W boson decay: one large- R jet with mass close to m_W

What are the inputs to jet reconstruction?

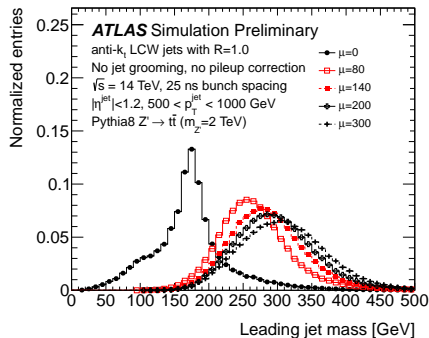
- Jets are comprised of 2/3 charged particles and 1/3 neutral particles
 - Tracker only sees charged particles, while the calorimeter sees both types
 - Energy deposits in the calorimeter are thus the key to jet reconstruction
- Deposits in EM calorimeter from e.g. $\pi^0 \rightarrow \gamma\gamma$
- *Topo-clusters* constructed to suppress pile-up
 - Group of topologically connected cells
 - Pile-up creates add. energy or clusters

[arXiv:1603.02934](https://arxiv.org/abs/1603.02934)



- The larger catchment area results in a larger pile-up susceptibility
 - Energy deposits from other simultaneous collisions pollute large- R jet
- Need to groom jet before studying its substructure

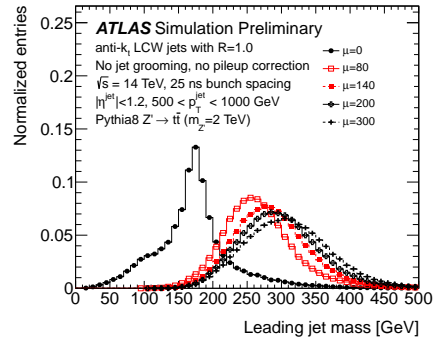
[link to figure](#)



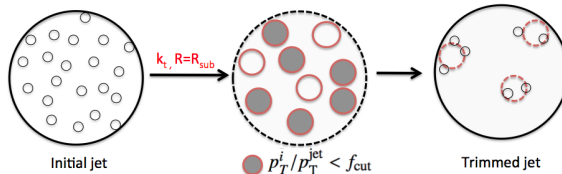
Pileup: major challenge for large- R jets

- The larger catchment area results in a larger pile-up susceptibility
 - Energy deposits from other simultaneous collisions pollute large- R jet
- Need to groom jet before studying its substructure

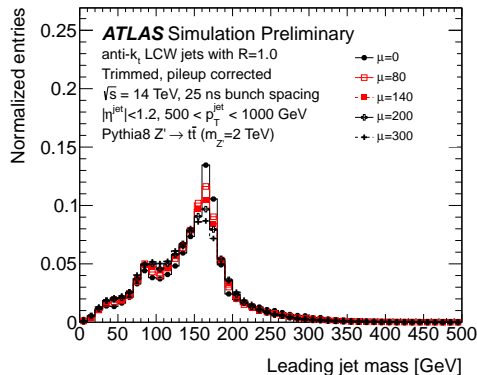
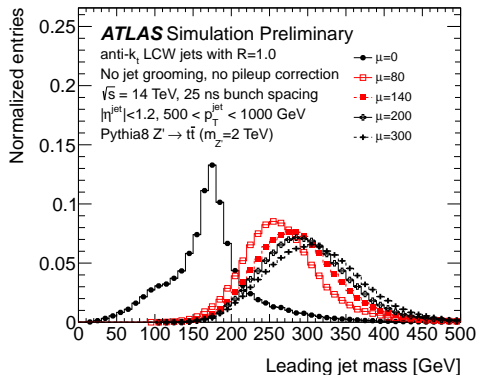
[link to figure](#)



Trimming ([arXiv:0912.1342](#))

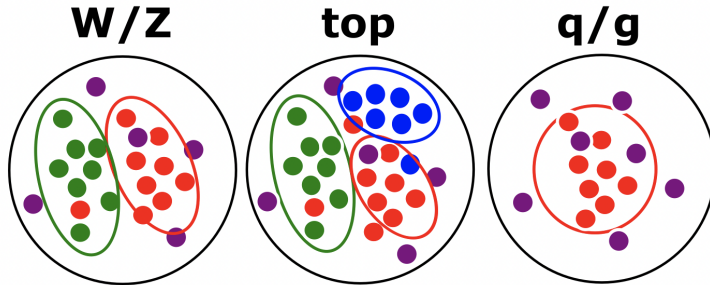


[link to figures](#)

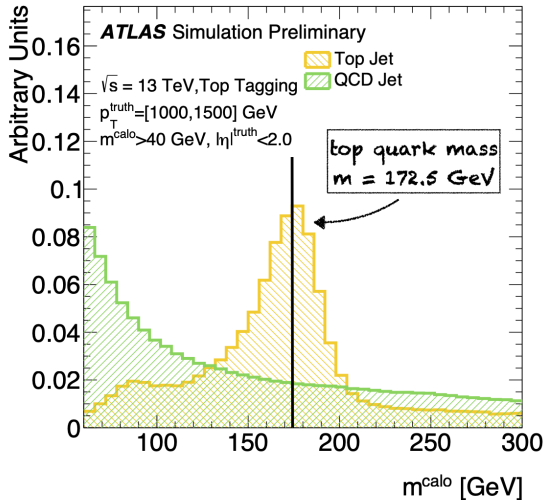


■ Pileup removal works even for pileup scenarios expected for the HL-LHC!

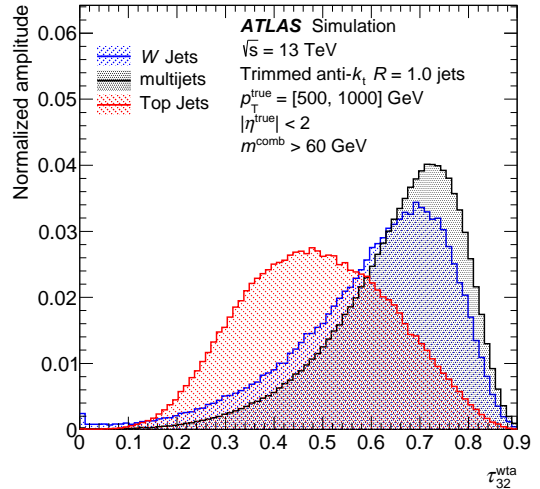
- The number of background events significantly exceeds that of signal events
- Background jets: jets initiated by one quark or gluon
- Study the jet's inner structure for signal vs. background separation: [jet substructure](#)



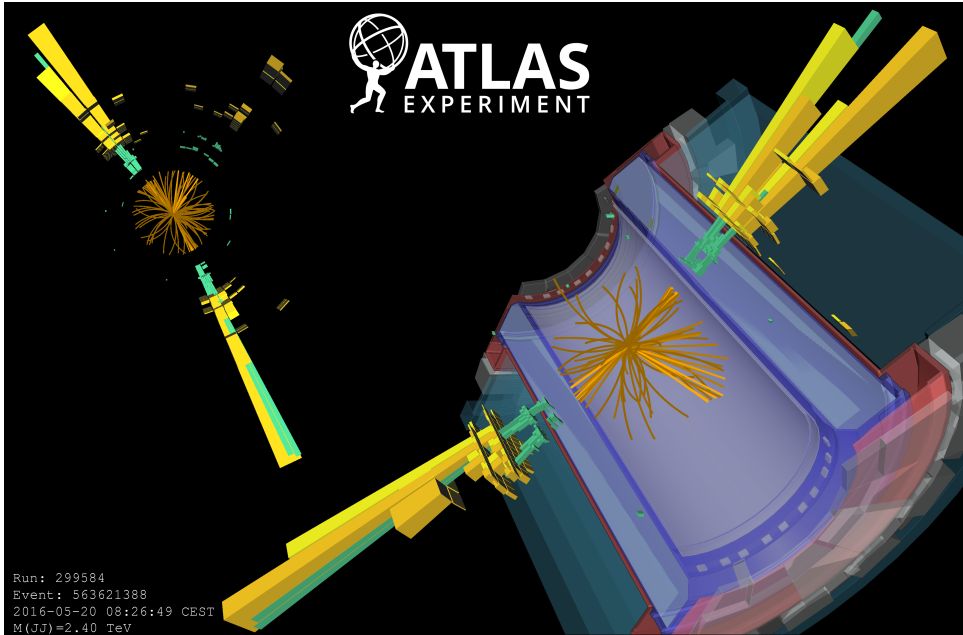
Jet mass

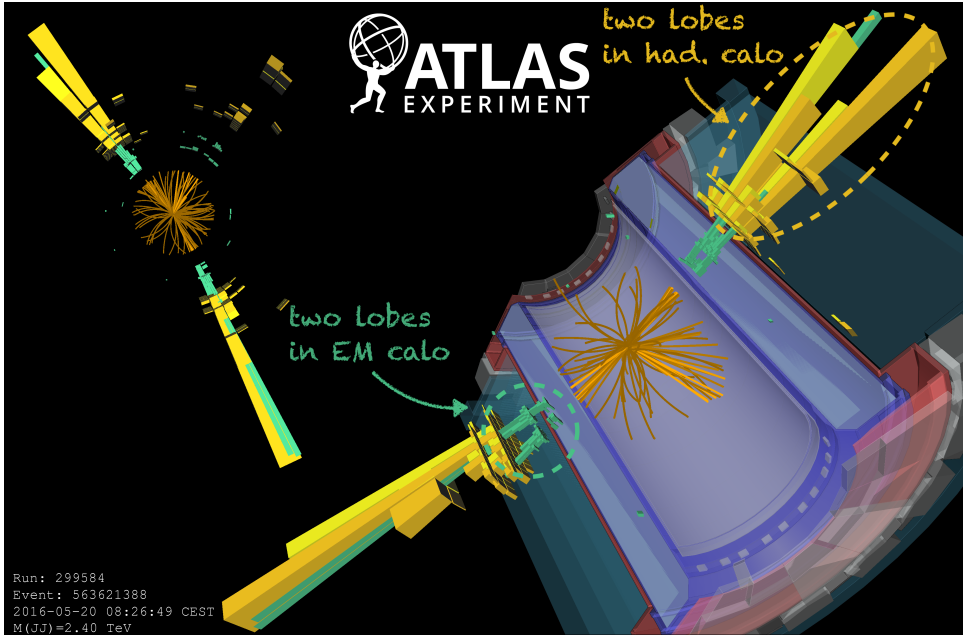


N-Subjettiness



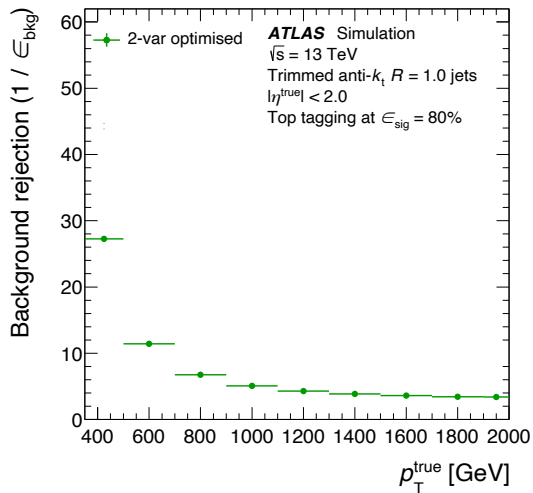
■ τ_{32} is trying to determine if the jet is composed out of 3 or 2 subjets





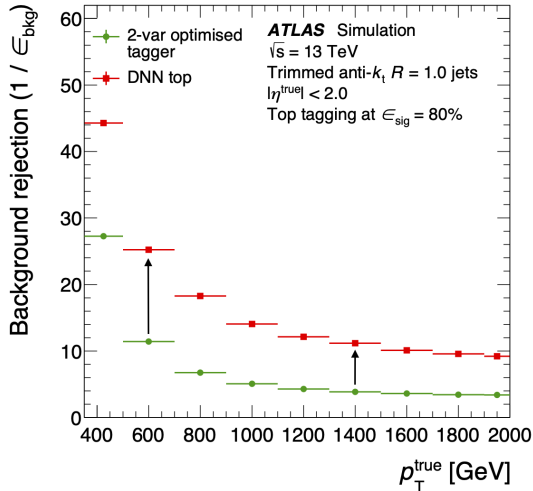
- Need powerful tools to distinguish signal from background

[arXiv:1808.07858](https://arxiv.org/abs/1808.07858)

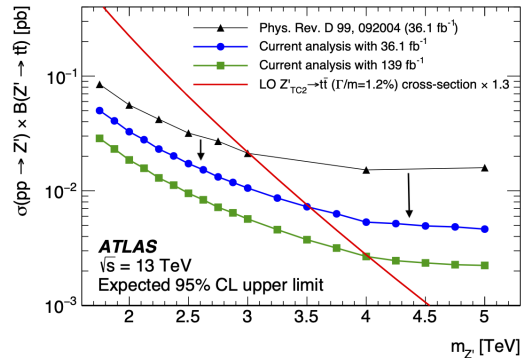


- Need powerful tools to distinguish signal from background
 - ML-based taggers (using various substructure variables), improved inputs to jet reco

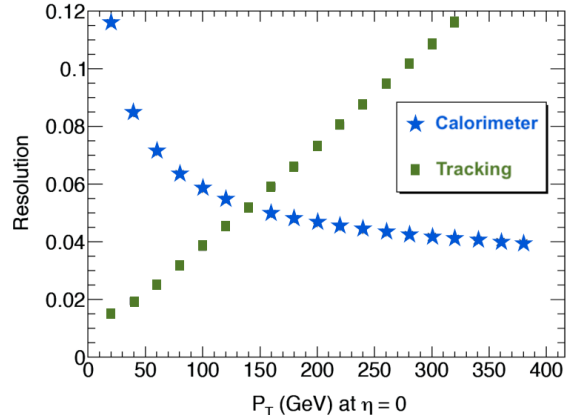
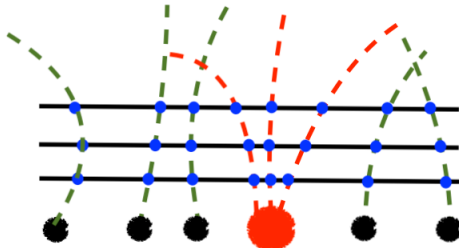
[arXiv:1808.07858](https://arxiv.org/abs/1808.07858)



- Search for new heavy particle $Z' \rightarrow t\bar{t}$
- [arXiv:2005.05138](https://arxiv.org/abs/2005.05138)

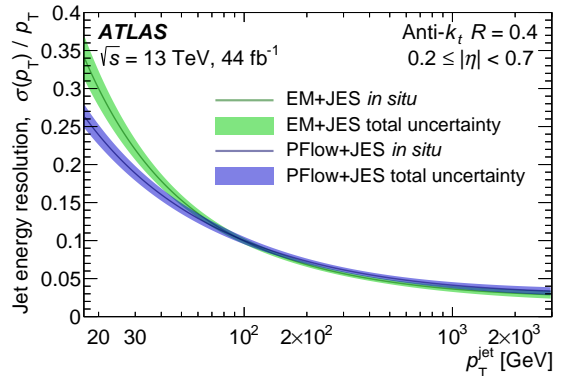
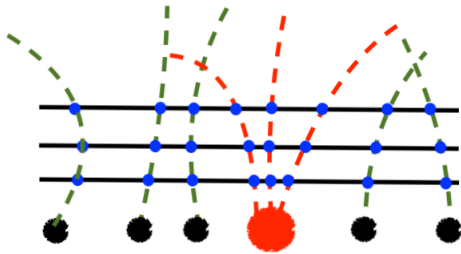


- The tracker is less susceptible to pile-up and has a better p_T resolution at low momenta
- Combine information from tracker and calorimeter to form inputs for jet reconstruction \Rightarrow Particle-Flow Algorithm ([arXiv:1703.10485](https://arxiv.org/abs/1703.10485))
- Better angular resolution of tracks



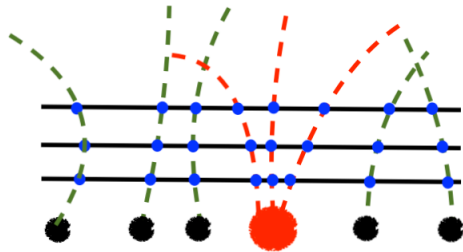
- The tracker is less susceptible to pile-up and has a better p_T resolution at low momenta
- Combine information from tracker and calorimeter to form inputs for jet reconstruction \Rightarrow Particle-Flow Algorithm ([arXiv:1703.10485](https://arxiv.org/abs/1703.10485))
- Better angular resolution of tracks

[arXiv:2007.02645](https://arxiv.org/abs/2007.02645)

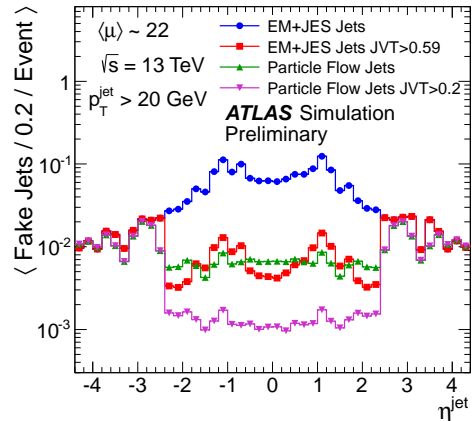


Adding tracks to the mix

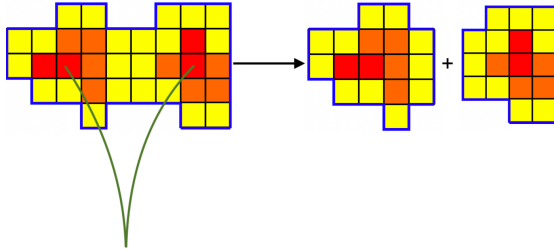
- The tracker is less susceptible to pile-up and has a better p_T resolution at low momenta
- Combine information from tracker and calorimeter to form inputs for jet reconstruction \Rightarrow Particle-Flow Algorithm ([arXiv:1703.10485](https://arxiv.org/abs/1703.10485))
- Better angular resolution of tracks



JETM-2017-006

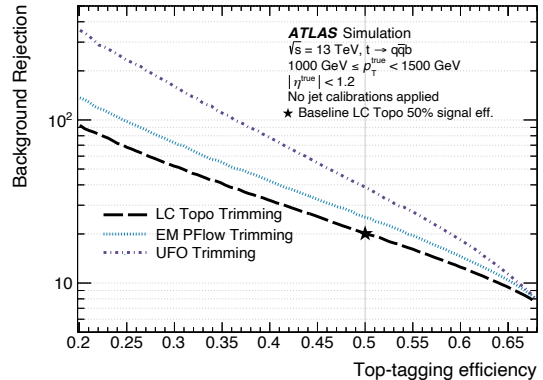


- At very high p_T , decay products could be reconstructed in only one cluster
→ loss of substructure information
- Split cluster based on tracks and replace angular position with track measurement
 - UFOs: Unified Flow Objects

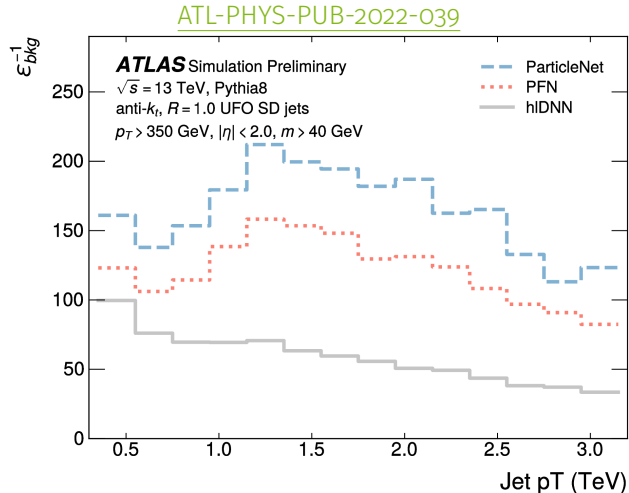


- This does not work for clusters purely from neutral hadrons
→ **development of new splitting algorithms**

[arXiv:2009.04986](https://arxiv.org/abs/2009.04986)



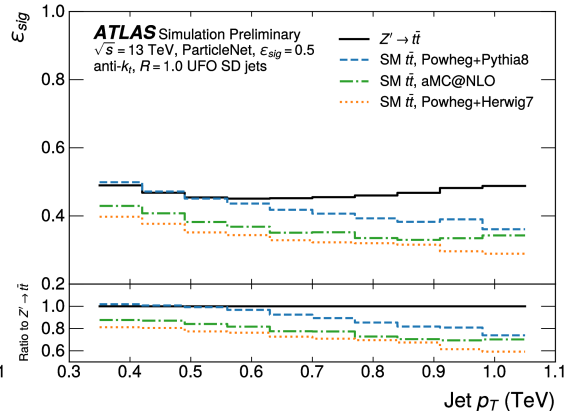
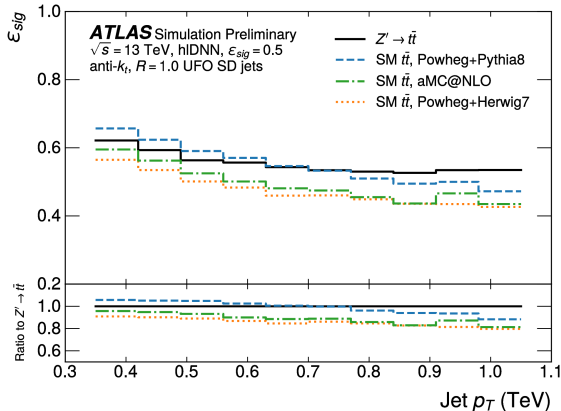
- Community is moving towards constituent-based taggers with improved performance
 - More sophisticated neural networks being developed

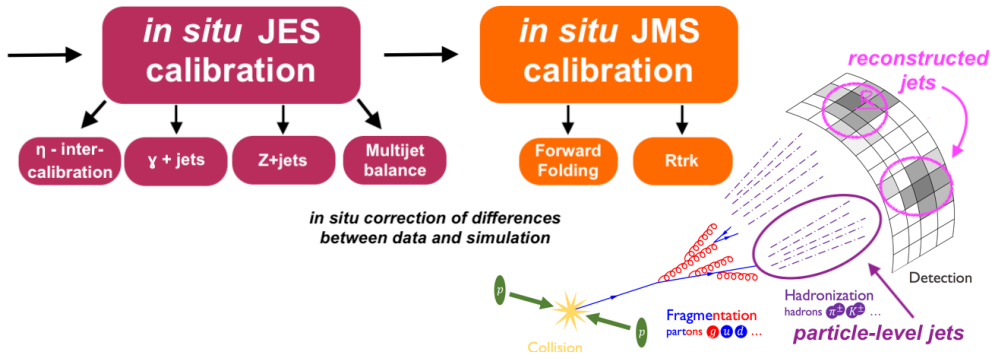


The machine learning era - the bad

- Community is moving towards constituent-based taggers with improved performance
 - More sophisticated neural networks being developed
- But new taggers show increase in modelling differences

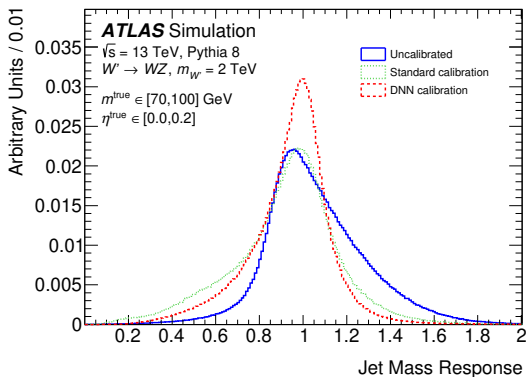
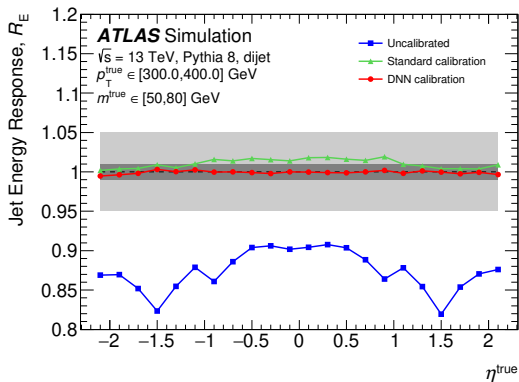
ATL-PHYS-PUB-2022-039

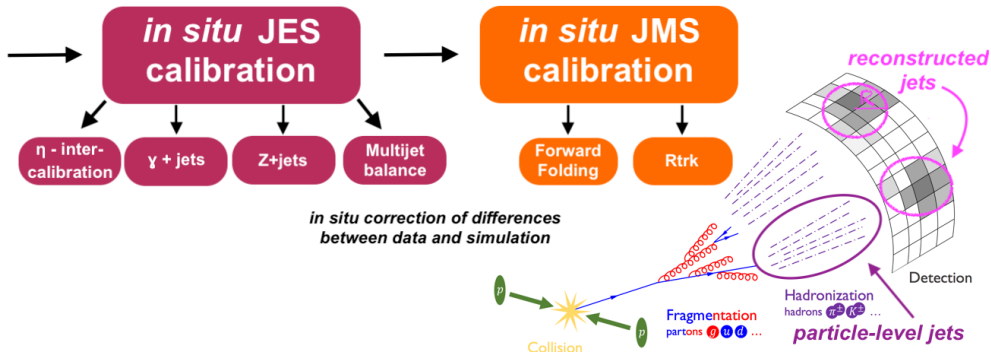


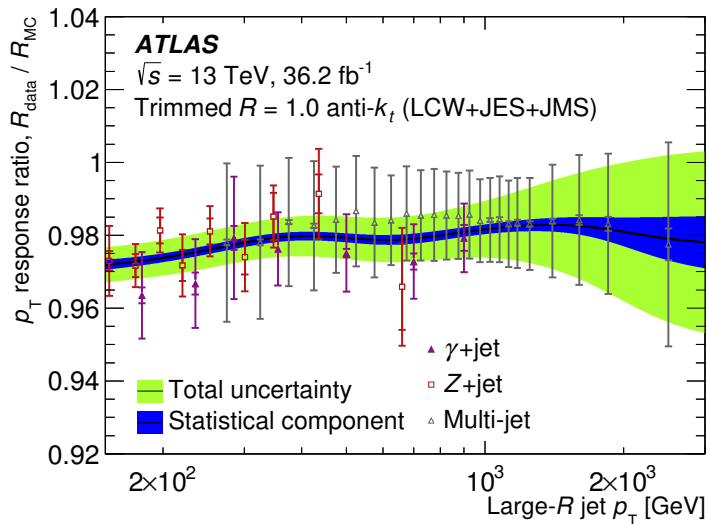


- Correct the reconstructed jet energy on average to the truth jet energy
- Energy lost due to non-compensating calorimeter, inactive material, noise thresholds, ...
- Previously, energy and mass (despite their correlation) were calibrated individually
- Improved closure with DNN taking into account shower evolution, substructure
- Accounts for differences between jet types, e.g. quark vs. gluon or q/g vs. W/Z/H/top

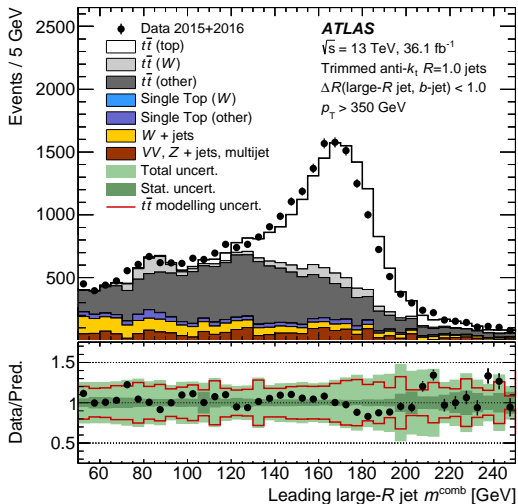
[arXiv:2311.08885](https://arxiv.org/abs/2311.08885)



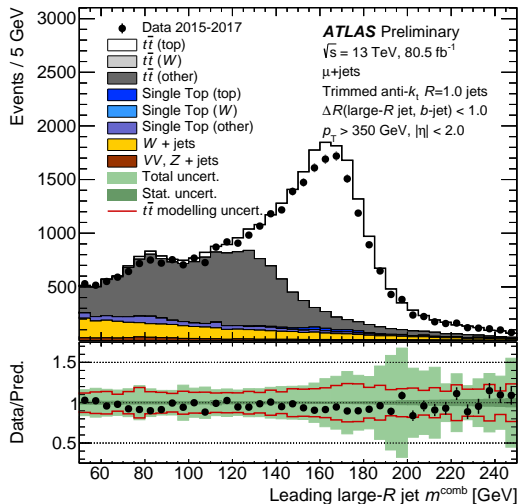




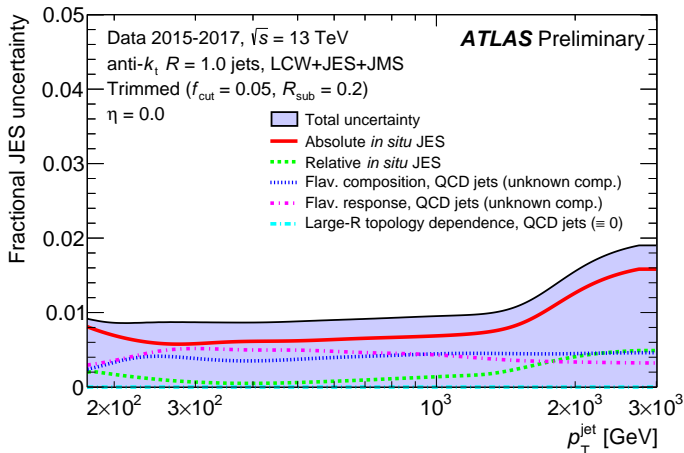
without *in situ* JES correction



with *in situ* JES correction



JETM-2019-05

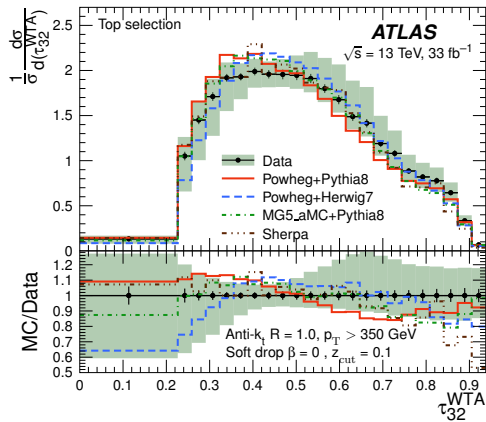


- Achieved already high precision of large- R jets
 - Compatible with small- R jets in same p_T regime
 - Reduced from an uncertainty of approx. 5% initially

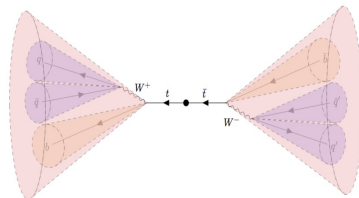
- Tools shown before are fantastic for e.g BSM searches or to select objects in measurements

[arXiv:1903.0294](https://arxiv.org/abs/1903.0294)

- But we can use them for much more
⇒ e.g. tuning of simulation
- Comparison to diff. generators to disentangle effects like parton shower vs. hadronization
- Grooming algorithms reduce sensitivity of observables to soft physics
→ less affected by non-perturbative effects



- Provide unfolded data to test analytic predictions and tune MC generators in $t\bar{t}$ events
- Measurement performed using full Run 2 data in lepton+jets and all-had channel
- [arXiv:2312.03797](https://arxiv.org/abs/2312.03797)

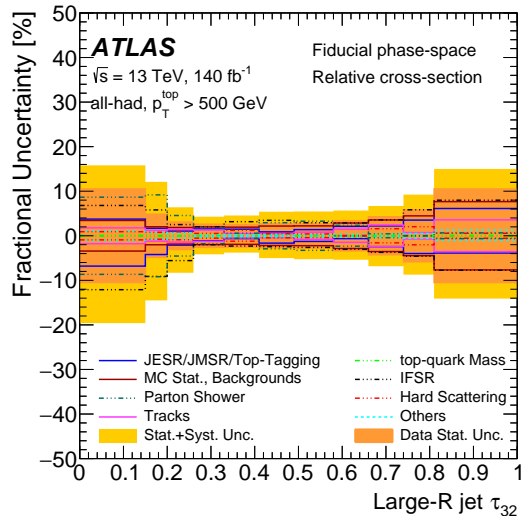
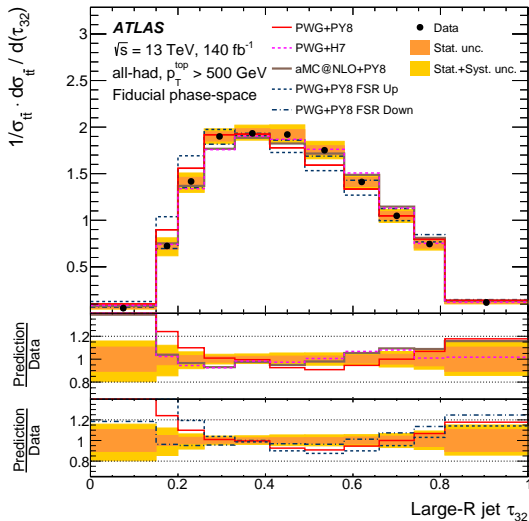


		1 st large- R jet			
		t0b0	t1b0	t0b1	t1b1
2 nd large- R jet	t1b1	J	K	L	S
	t0b1	B	D	H	N
	t1b0	E	F	G	M
	t0b0	A	C	I	O

- Require two b -tagged jets matched to large- R jet
 - Suppresses large multi-jet background
- Non-probe jet has to be top tagged
- Extended ABCD method for bkg estimation
- Only charged-particle tracks used to measure substructure

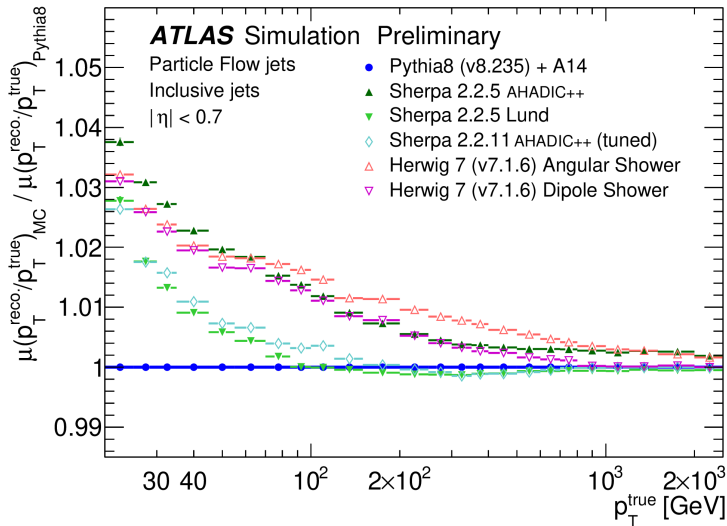
Category	Event yields ℓ +jets selection	Number of large- R jets all-hadronic selection
Data	83 069	30 524
Predictions	$97\,200 \pm 3\,700$	$36\,500 \pm 1\,400$
$t\bar{t}$ (ℓ +jets)	$90\,600 \pm 3\,400$	$1\,610 \pm 140$
$t\bar{t}$ (all-hadronic)	–	$25\,700 \pm 1\,400$
Multijet	–	$8\,100 \pm 300$
Single-top quark	$2\,200 \pm 300$	710 ± 70
NP/Misid. leptons	$1\,500 \pm 600$	
W +jets	$1\,500 \pm 700$	–
$t\bar{t}V$ ($t\bar{t}Z + t\bar{t}W + t\bar{t}H$)	920 ± 120	310 ± 40
Other	400 ± 200	–
Data/Predictions	0.85 ± 0.03	0.84 ± 0.03
(Data – Background)/Signal	0.84 ± 0.03	0.77 ± 0.05

Category	Event yields ℓ +jets selection	Number of large- R jets all-hadronic selection
Data	83 069	30 524
Predictions	97 200 \pm 3 700	36 500 \pm 1 400
$t\bar{t}$ (ℓ +jets)	90 600 \pm 3 400	1 610 \pm 140
$t\bar{t}$ (all-hadronic)	–	25 700 \pm 1 400
Multijet	–	8 100 \pm 300
Single-top quark	2 200 \pm 300	710 \pm 70
NP/Misid. leptons	1 500 \pm 600	
W +jets	1 500 \pm 700	–
$t\bar{t}V$ ($t\bar{t}Z$ + $t\bar{t}W$ + $t\bar{t}H$)	920 \pm 120	310 \pm 40
Other	400 \pm 200	–
Data/Predictions	0.85 \pm 0.03	0.84 \pm 0.03
(Data – Background)/Signal	0.84 \pm 0.03	0.77 \pm 0.05



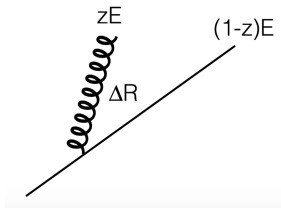
- Largest uncertainties coming from the comparison of different Monte Carlo generators
- Jet-related uncertainties can be also large in some regions of phase space
 - Some double-counting of uncertainties related to modelling → MC-to-MC scale factors

ATL-PHYS-PUB-2022-021

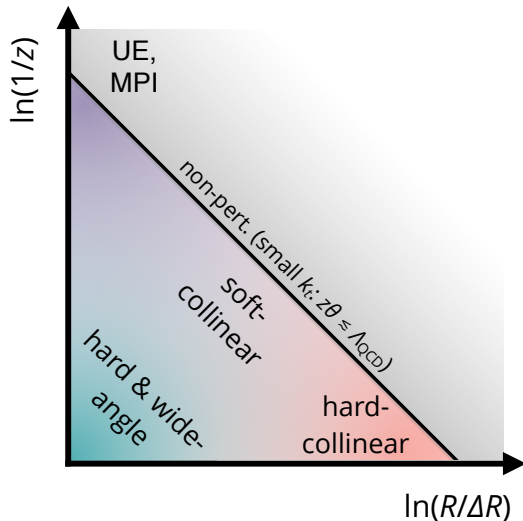


■ New AHADIC tune from Sherpa authors using LEP data (hadron fractions within jet)

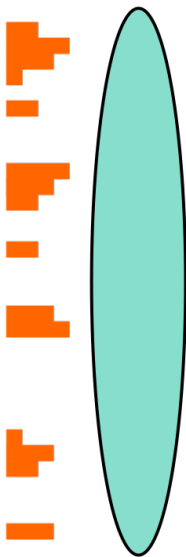
- Lund plane is a useful tool to display emissions



- Expect uniform emission pattern in $\ln(1/z)$ and $\ln(1/\theta)$
- However not directly usable because we don't observe quarks/gluons
- [arXiv:2004.03540](https://arxiv.org/abs/2004.03540)

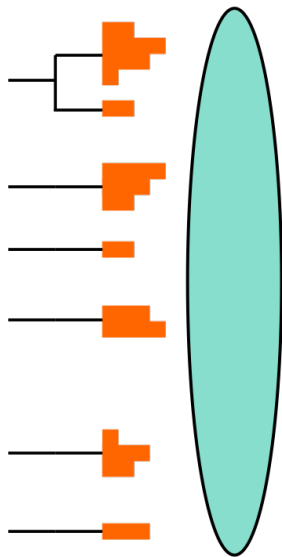


Re-clustering with the Cambridge/Aachen algorithm



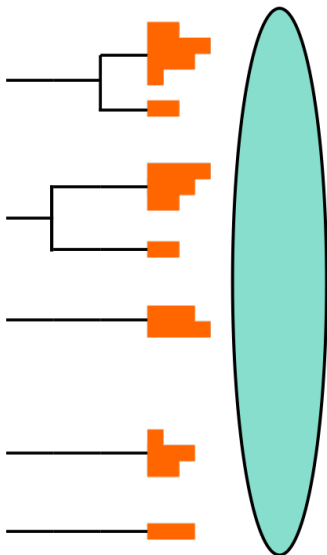
- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower

Re-clustering with the Cambridge/Aachen algorithm



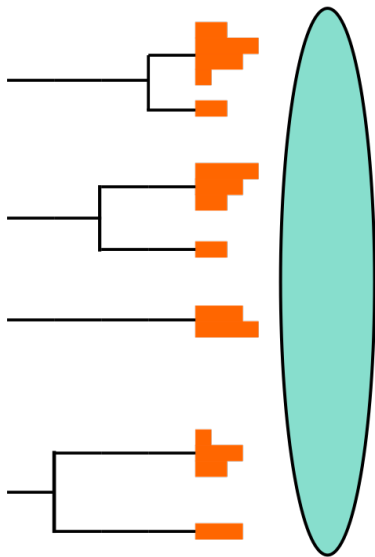
- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower

Re-clustering with the Cambridge/Aachen algorithm



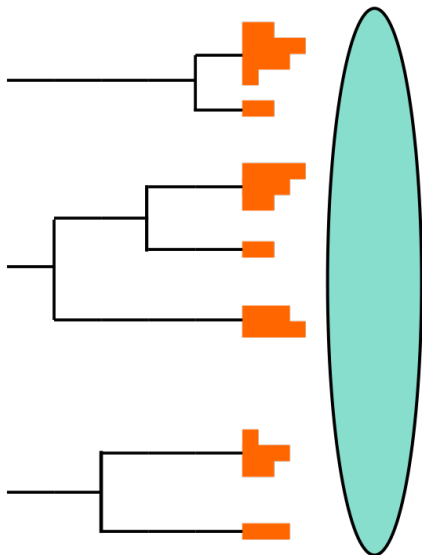
- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower

Re-clustering with the Cambridge/Aachen algorithm



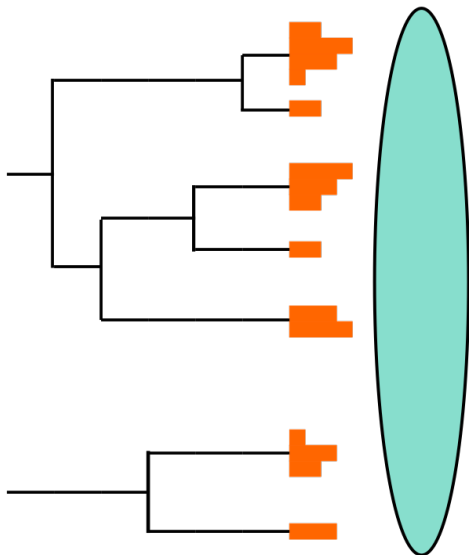
- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower

Re-clustering with the Cambridge/Aachen algorithm



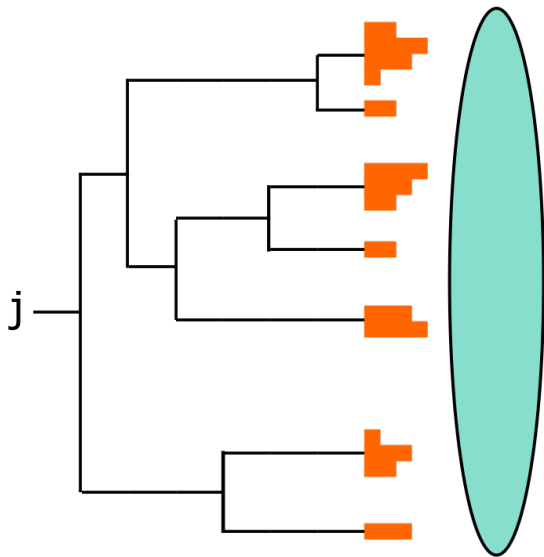
- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower

Re-clustering with the Cambridge/Aachen algorithm

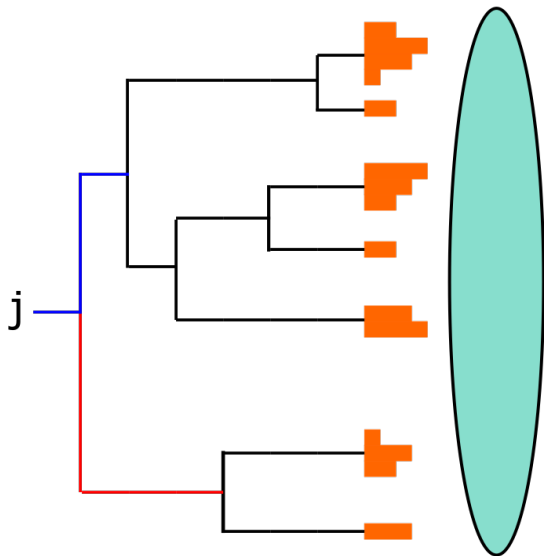


- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower

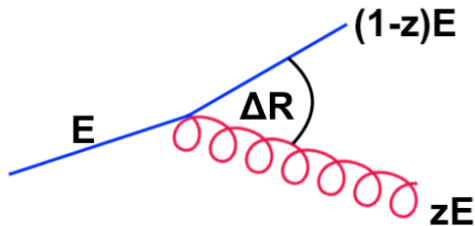
Re-clustering with the Cambridge/Aachen algorithm



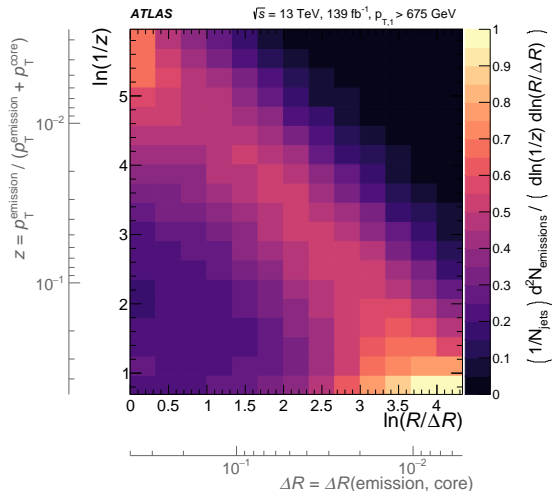
- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower



- Cluster jet constituents with C/A alg
 - Based on angular separation
- Reverse of parton shower

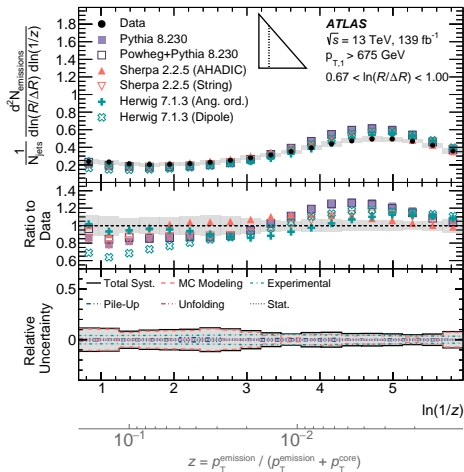
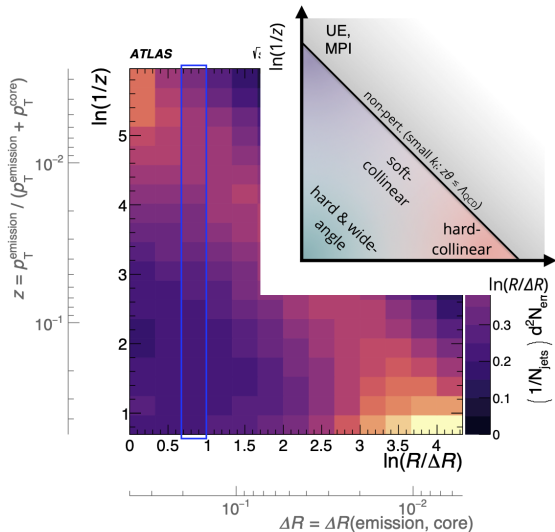


- Measured Lund Jet Plane in dijet events using $R = 0.4$ jets
- Only tracks are used here to allow for precise measurement of small splittings

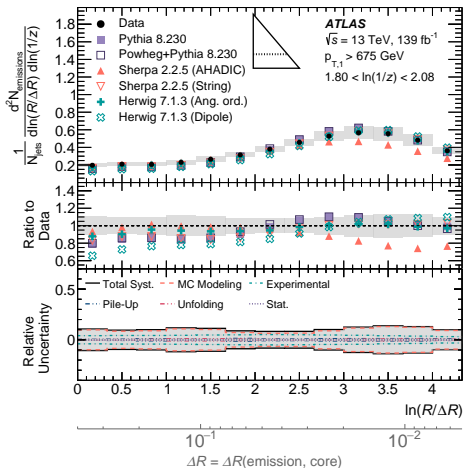
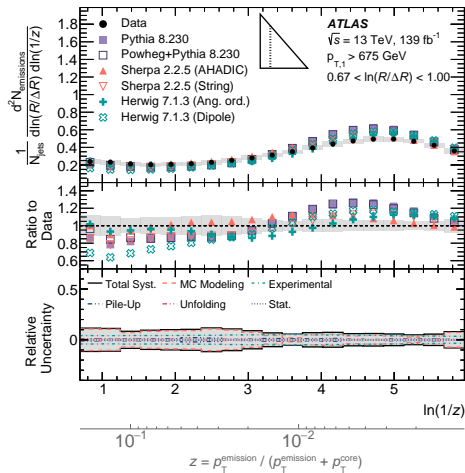


The Lund jet plane - [arXiv:2004.03540](https://arxiv.org/abs/2004.03540)

- Measured Lund Jet Plane in dijet events using $R = 0.4$ jets
- Only tracks are used here to allow for precise measurement of small splittings

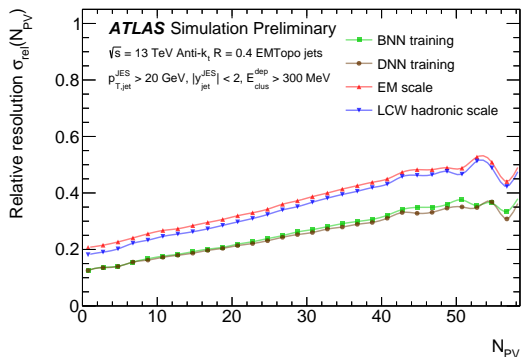


- Measured Lund Jet Plane in dijet events using $R = 0.4$ jets
- Only tracks are used here to allow for precise measurement of small splittings



- Lately, precision jet substructure measurements have been only performed with tracks
- Uncertainties associated with topoclusters are relatively large compared to tracks
 - We could be missing important discrepancies stemming from neutral particles

ATL-PHYS-PUB-2023-019



- ML-based calibrations for topoclusters
 - Response for hadronic clusters lower than for electromagnetic ones
- Efforts on-going to reduce pile-up dependence, e.g. cell-level timing cuts

■ **Jet substructure is a versatile tool to probe the SM**

- Tuning of the simulation, α_s determination, jet quenching, searches for BSM ...

■ **Large- R reconstruction has significantly improved over the past 10 years**

- We can determine the jet energy scale with 1% precision, same level as for small- R jets

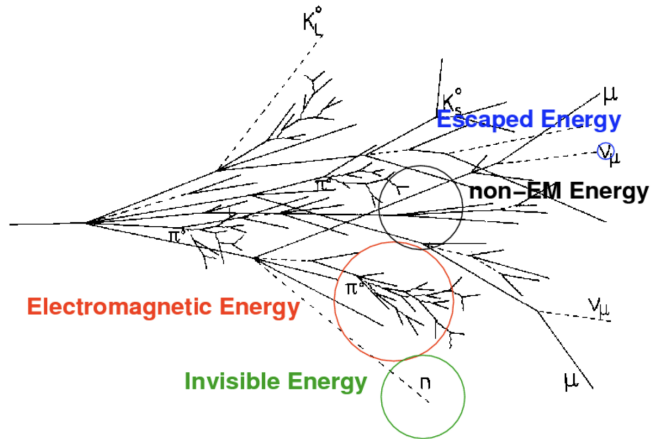
■ **There is much more that I couldn't show here today ...**

- e.g. quark vs. gluon tagging, multijet event isotropies, mass measurements, ...

■ **Interested? Join us at BOOST in Genova**

- Annual meeting on jet reconstruction, tagging, pileup mitigation, QCD calculations, ...
- Agenda: [link](#)

Backup



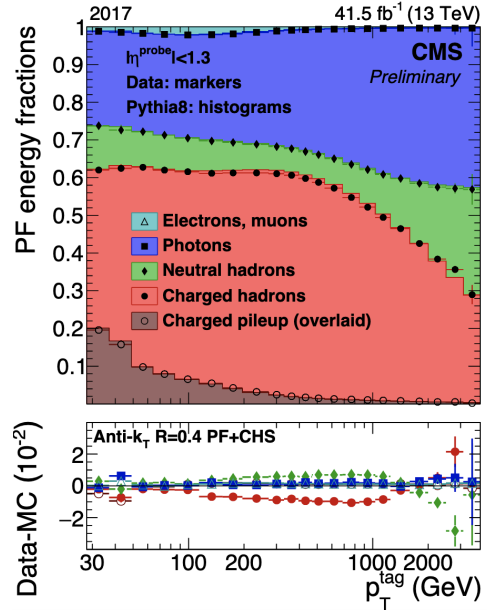
- A hadronic shower has two components: a hadronic and an electromagnetic one
- Escaped energy: e.g. muons and/or neutrinos (from hadron decays)
- Electromagnetic component: $\pi^0 \rightarrow \gamma\gamma$

- Jets are not only composed of charged & neutral hadrons but also of e, μ, γ
 - e.g. hadron decays, soft photon emissions

- $\approx 2/3$ charged and $1/3$ neutral particles

- Tracker
 - Reconstructs only charged particles

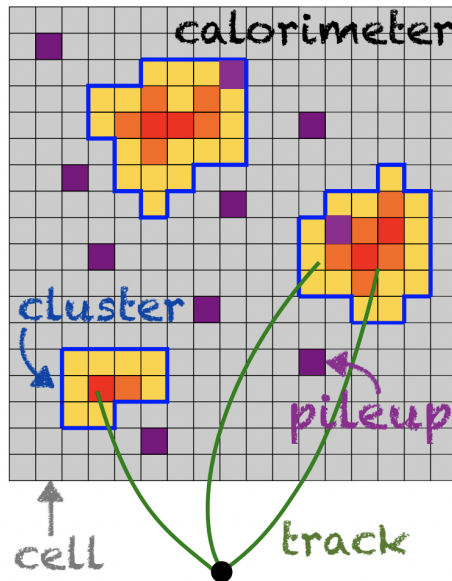
- Calorimeter:
 - Reconstructs neutral + charged particles



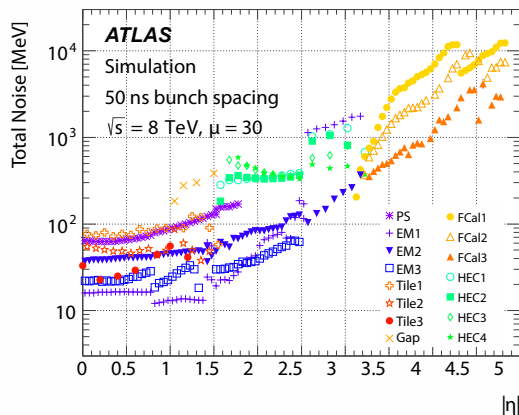
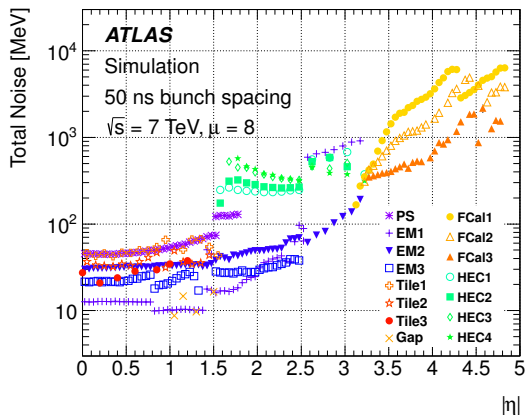
- Cells are grouped together in topoclusters based on the cell significance ς (4-2-0)
- Cell significance: ratio of signal to noise

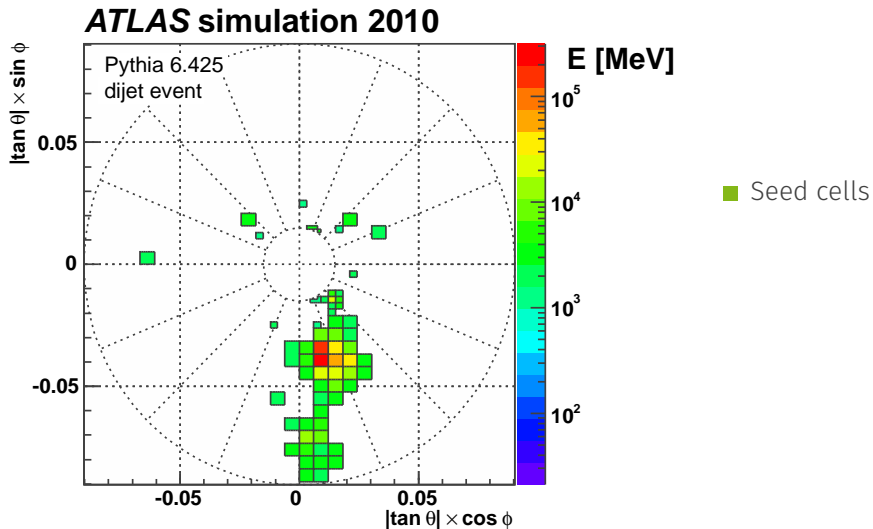
$$\varsigma_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise, cell}}^{\text{EM}}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sqrt{(\sigma_{\text{noise}}^{\text{electronic}})^2 + (\sigma_{\text{noise}}^{\text{pile-up}})^2}}$$

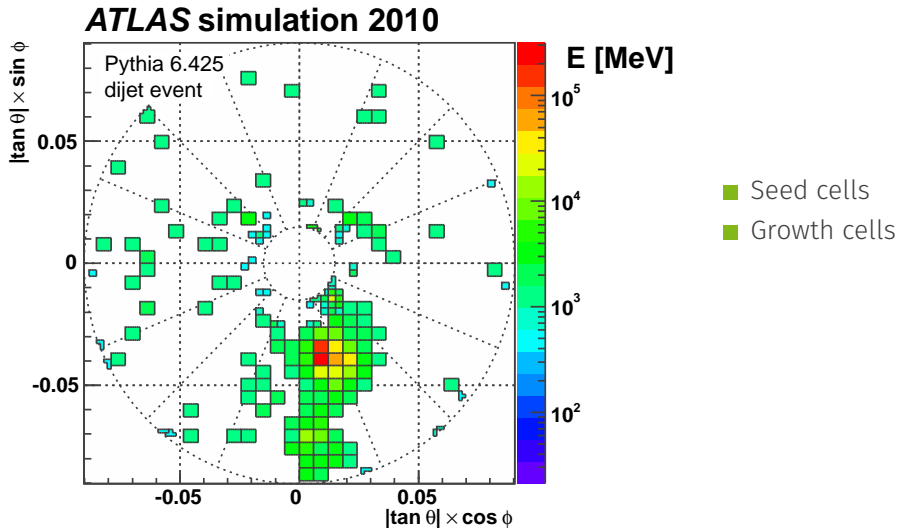
- seed cells, $\varsigma > 4$
- growth cells, $\varsigma > 2$
- boundary cells, $\varsigma > 0$
- cells from pile-up vertices

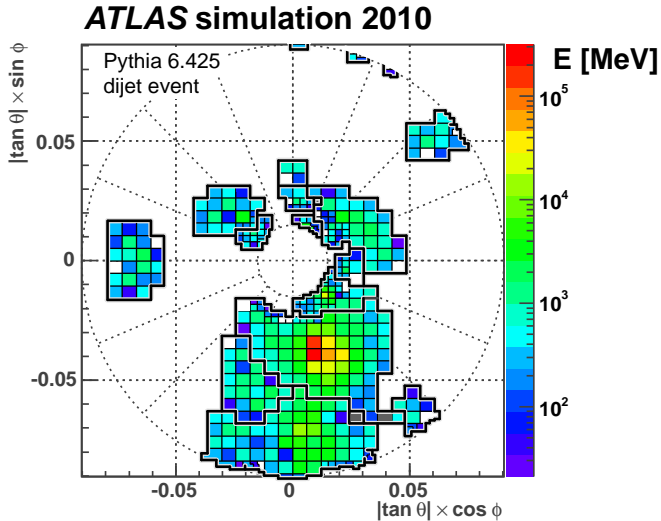


- Topo-cluster formation depends on the cell noise which is dominated by pile-up noise
- Noise from pile-up needs to be determined prior to data-taking and well tuned



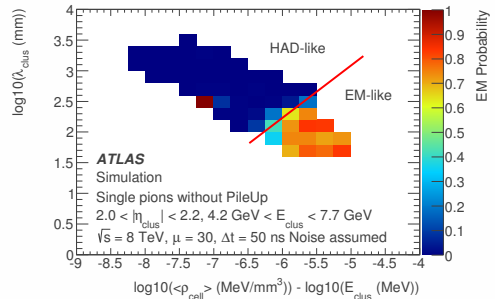




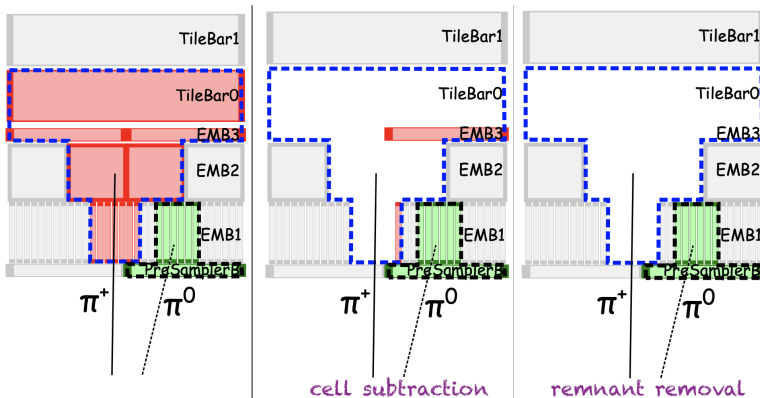


- Seed cells
- Growth cells
- Cluster formation
- Cluster splitting in case of two local maxima

- The hadronic calorimeter is a non-compensating calorimeter:
 - i.e. smaller signal for hadrons than e, γ of the same incident energy
 - caused by invisible energy: energy used to release nucleons from nuclei + $\mu + \nu_x$
- We measure the hadronic signal at the electromagnetic (EM) scale
- **Local Cell Weighted** scale: extra calibration for hadronic signals to account for:
 - Non-compensating character of the calorimeter
 - Signal losses due to inactive material
 - Energy falling in unclustered cells
- Clusters first have to be identified as either hadronic or electromagnetic
- LCW used only for large- R jets in Run-2
- Small- R jets showed larger pile-up dependence
- New efforts on-going to improve LCW using ML



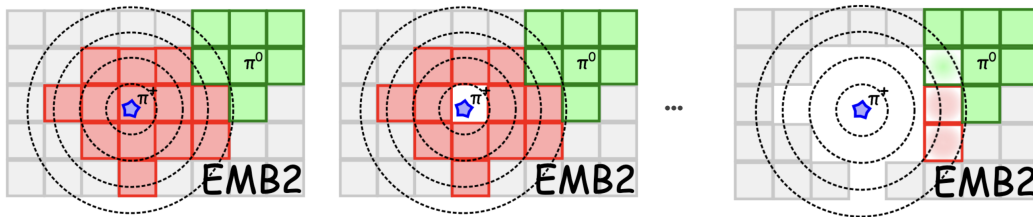
1. Match charged-particle tracks to topo-clusters in the calorimeter
2. Subtract energy deposited in the calorimeter by charged particles from matched clusters
3. Add only tracks associated with hard-scatter vertex to list of inputs to jet reconstruction
(Charged **H**adron **S**ubtraction)



- Need to know how much energy a particle with p_{trk} deposits on average in the calorimeter

$$\langle E_{\text{dep}} \rangle = p_{\text{trk}} \left\langle \frac{E_{\text{clus}}}{p_{\text{trk}}^{\text{ref}}} \right\rangle$$

- $\langle E_{\text{dep}} \rangle$ is determined in single-particle simulations without pile-up
- $\langle E_{\text{dep}} \rangle$ provided as a function of p_{trk}, η and the layer of highest energy density (LHED)
 - Shower core has a well-defined ellipsoidal shape in $\eta - \phi$
 - First perform subtraction in LHED before progressing to less regular shower periphery
- If $\langle E_{\text{dep}} \rangle > E_{\text{clus}}$: remove cluster, else cell-by-cell subtraction



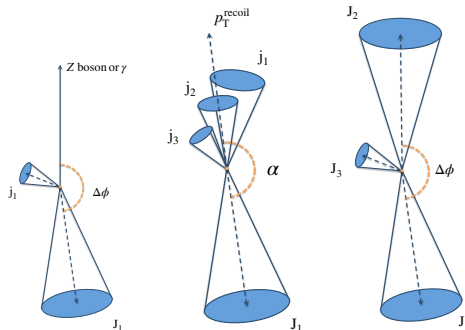
- Need to measure the jet response in **data** and **simulation**, correction factor defined as

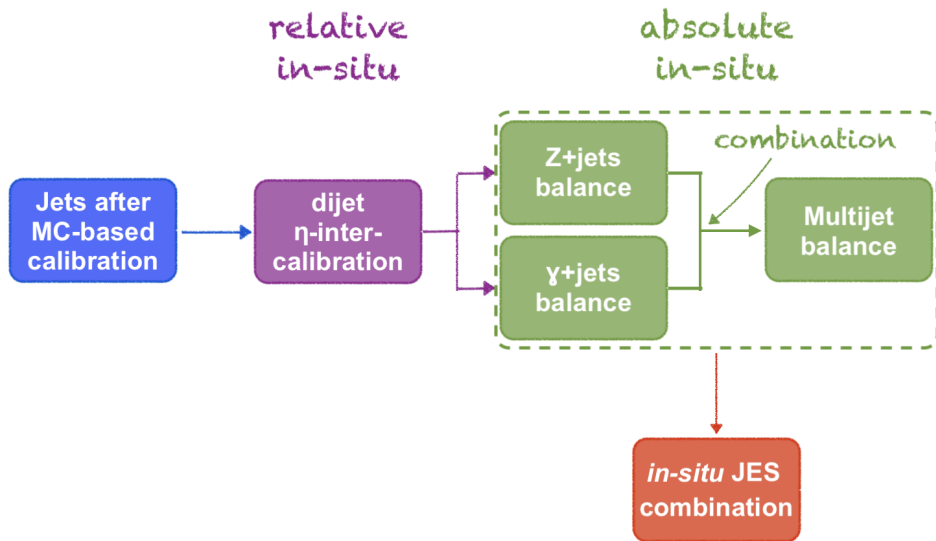
$$c = \mathcal{R}_{MC}^{in\ situ} / \mathcal{R}_{data}^{in\ situ}$$

- Response is calculated by balancing the jet p_T against a well-calibrated reference object with approx. no other hadronic activity

$$\mathcal{R}_{MC,data}^{in\ situ} = \left\langle \frac{p_T^{jet}}{p_T^{ref}} \right\rangle$$

- Distribution fitted in bins of p_T^{ref} with Gaus → extract mean
- Uncertainties of ref. object are propagated, thus use objects with high precision, e.g. $Z \rightarrow \mu\mu$, $Z \rightarrow ee$, γ or system of well-calibrated low p_T jets





- **Relative *in-situ* JES:** correct jets with $|\eta| > 0.8$ to the same energy scale as $|\eta| < 0.8$
- **Absolute *in-situ* JES:** correct jets with $|\eta| < 0.8$ to a precise reference object

- Aim: flatten the JES across the detector
- Matrix method used to increase stat.: neither of the jets needs to be within $|\eta| < 0.8$
 - Multiple reference regions are defined which are calibrated against each other
 - Calibration derived using dijet events

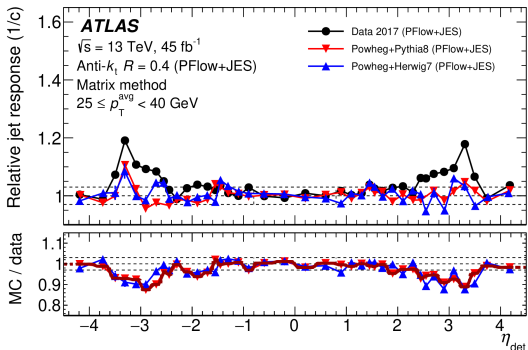
- Response difference between two regions

$$\mathcal{A} = \frac{p_T^{\text{left}} - p_T^{\text{right}}}{p_T^{\text{avg}}}$$

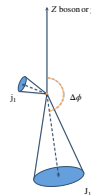
- Response ratio \mathcal{R} of the two jets defines the calib factor c for each jet

$$\mathcal{R} = \frac{c_{\text{left}}}{c_{\text{right}}} = \frac{2 + \langle \mathcal{A} \rangle}{2 + \langle \mathcal{A} \rangle} \simeq \frac{p_T^{\text{left}}}{p_T^{\text{right}}}$$

- Dominating uncertainties are Monte Carlo generator differences and the third jet veto



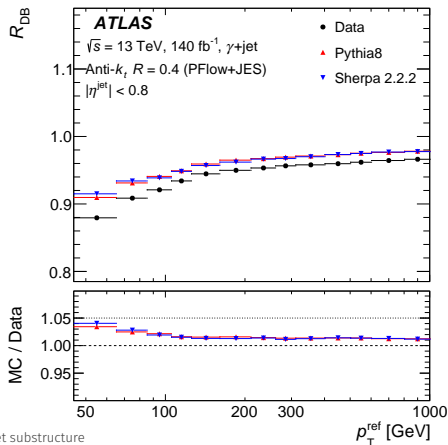
- Reference objects are either $Z(\rightarrow \mu\mu)$, $Z(\rightarrow ee)$ or γ
- Z +jets covers lowest p_T range ($p_T > 17$ GeV), γ +jets starting at ≈ 25 GeV
- Jets required to be calibrated with η -intercalibration for second jet veto



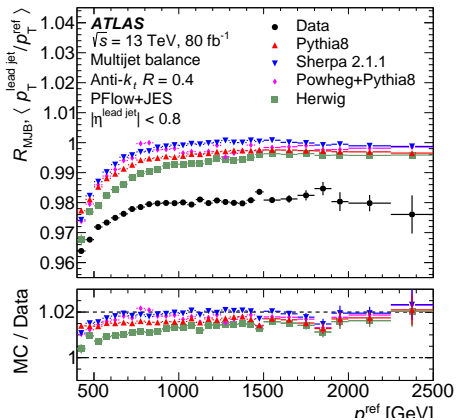
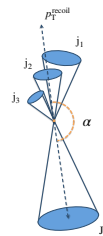
$$\mathcal{R}_{DB} = \left\langle \frac{p_T^{\text{jet}}}{p_T^{\text{ref}}} \right\rangle$$

with $p_T^{\text{ref}} = p_T^{Z/\gamma} |\cos(\Delta\phi(X, \text{jet}))|$

- p_T imbalance may be introduced by out-of-cone (OOC) effects, pile-up or ISR/FSR
→ MPF technique
- Technique also used to derive JES for b -jets, see e.g. [JTM-2022-01](#)



- Reference object: system of low- p_T small- R jets
 - Those jets are calibrated with combination of Z +jets & γ +jets calib
 - Uncertainties from low- p_T calibration are propagated through the MJB
- MJB covers roughly the range $p_T > 400$ GeV up to 2.5 TeV



$$\mathcal{R}_{\text{DB}} = \left\langle \frac{p_T^{\text{lead. jet}}}{p_T^{\text{ref}}} \right\rangle$$

- p_T^{ref} is the vectorial sum of the recoil system
- Multiple iterations to extend the reach of the technique
- Dominating uncertainties are the flavour uncertainties, propagated γ +jets uncertainties and MC generator differences