Jets & Substructure: experiment

Wahid Bhimji

THE UNIVERSITY of EDINBURGH

On behalf of the ATLAS and CMS collaborations

Physics at LHC and Beyond
Quy Nhon, Vietnam
10th August 2014

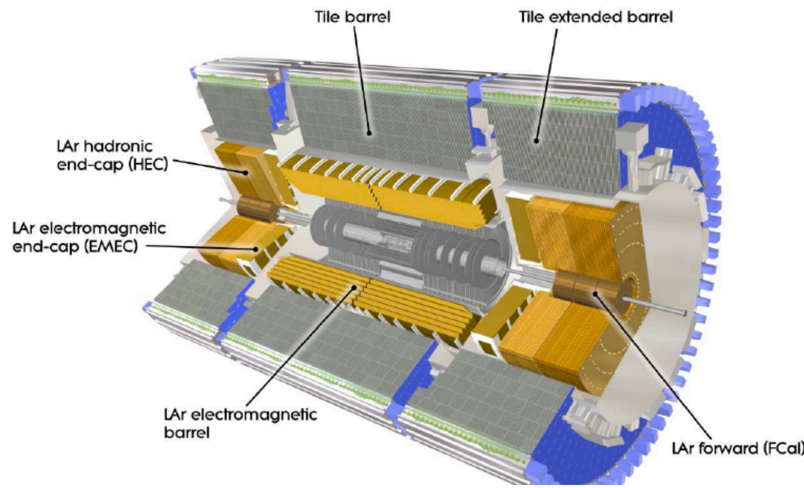
Outline

- ATLAS and CMS jet inputs
- Jets and Jet Substructure: motivation and algorithms
- Recent performance results:
  - W boson tagging
  - Top tagging
- Other recent highlights

ATLAS and CMS Jet Inputs

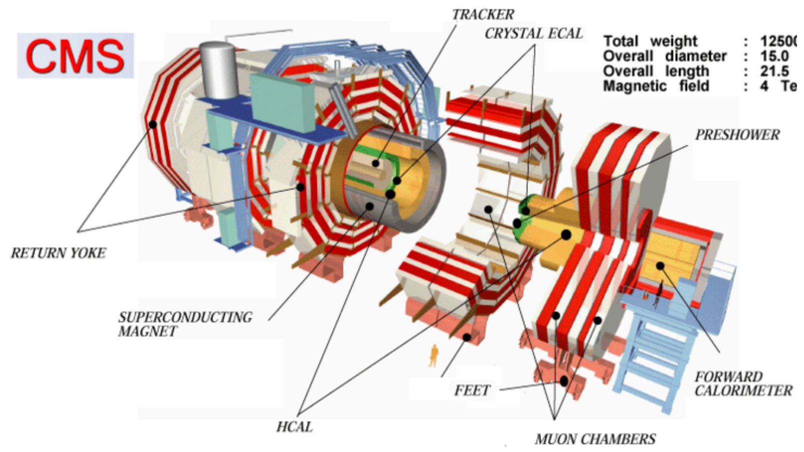
ATLAS

- LAr electromagnetic and high resolution tile hadronic calorimeter
- Inputs to Jet Reconstruction
  - 3D topological clusters: seed cell with $E_{cell} > 4\sigma$ and neighbours $E_{cell} > 2\sigma$
  - Noise suppression and calibration



CMS

- Fast and extremely high resolution ECAL with high (transverse) granularity
- Reduced material in front of ECAL and strong magnetic field
- Particle flow Reconstruction:
  - Link HCAL+ECAL clusters and tracks
  - If $E_{cal} \sim p_{track}$ fit track and calo energy

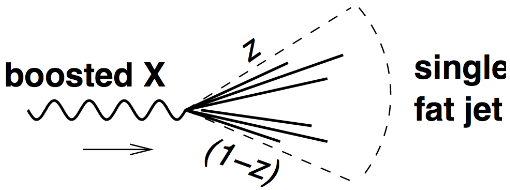


Outline

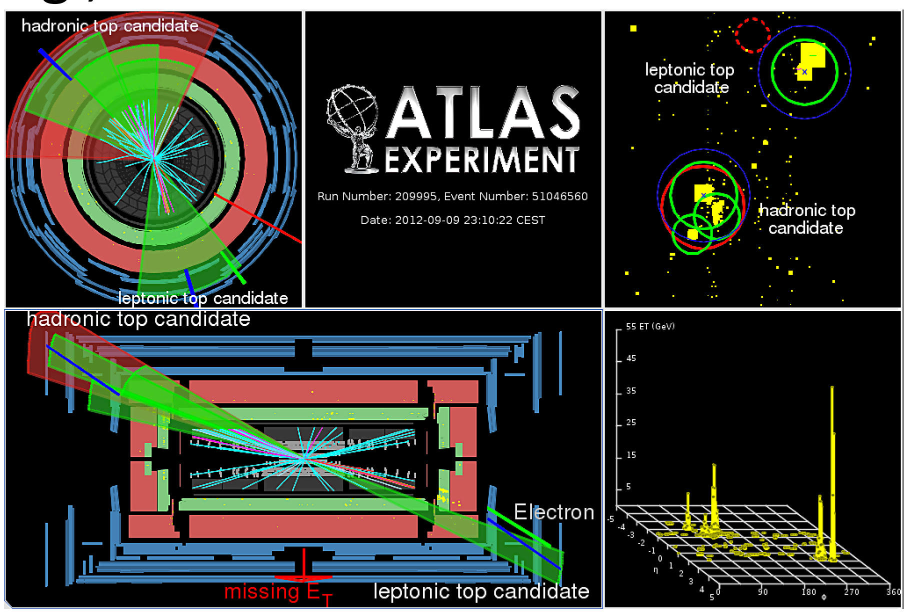
- ATLAS and CMS jet inputs
- Jets and Jet Substructure: motivation and algorithms
- Recent performance results:
  - W boson tagging
  - Top tagging
- Other recent highlights

Jet Substructure: Motivation

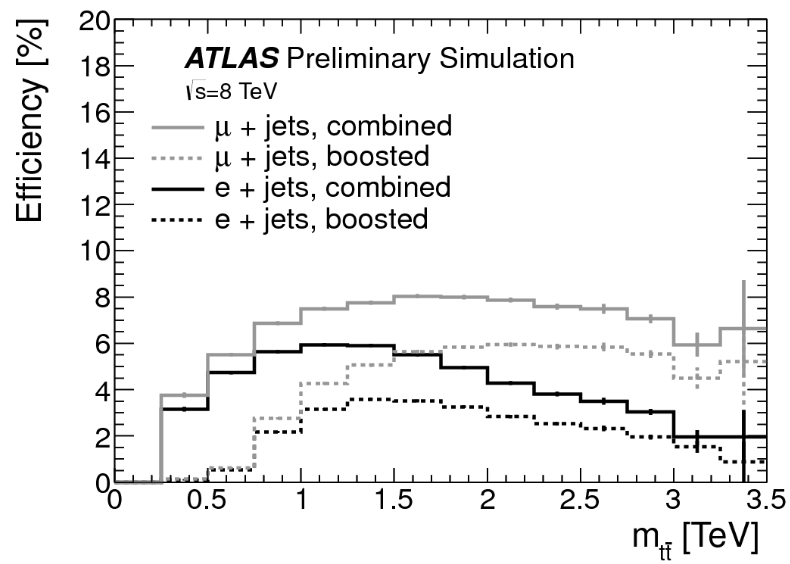
- At LHC $\sqrt{s} \gg M_{EW}$ => high p_T *boosted* objects
  - Also choose high-p_T region to reduce QCD backgrounds
  - Decay products merge into single *fat* jet
  - Need to look at *substructure* for *tagging*



E.g., Search for $t\bar{t}$ resonances



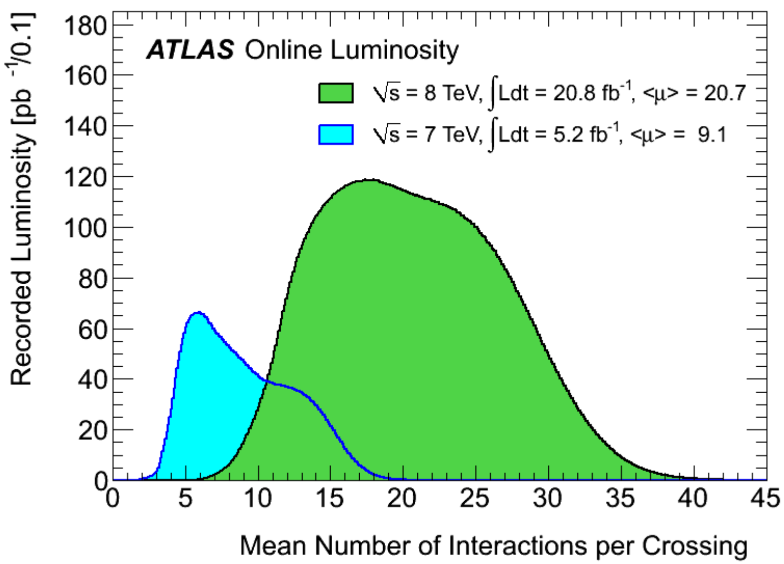
ATLAS-CONF-2013-052



Jet Substructure: Motivation

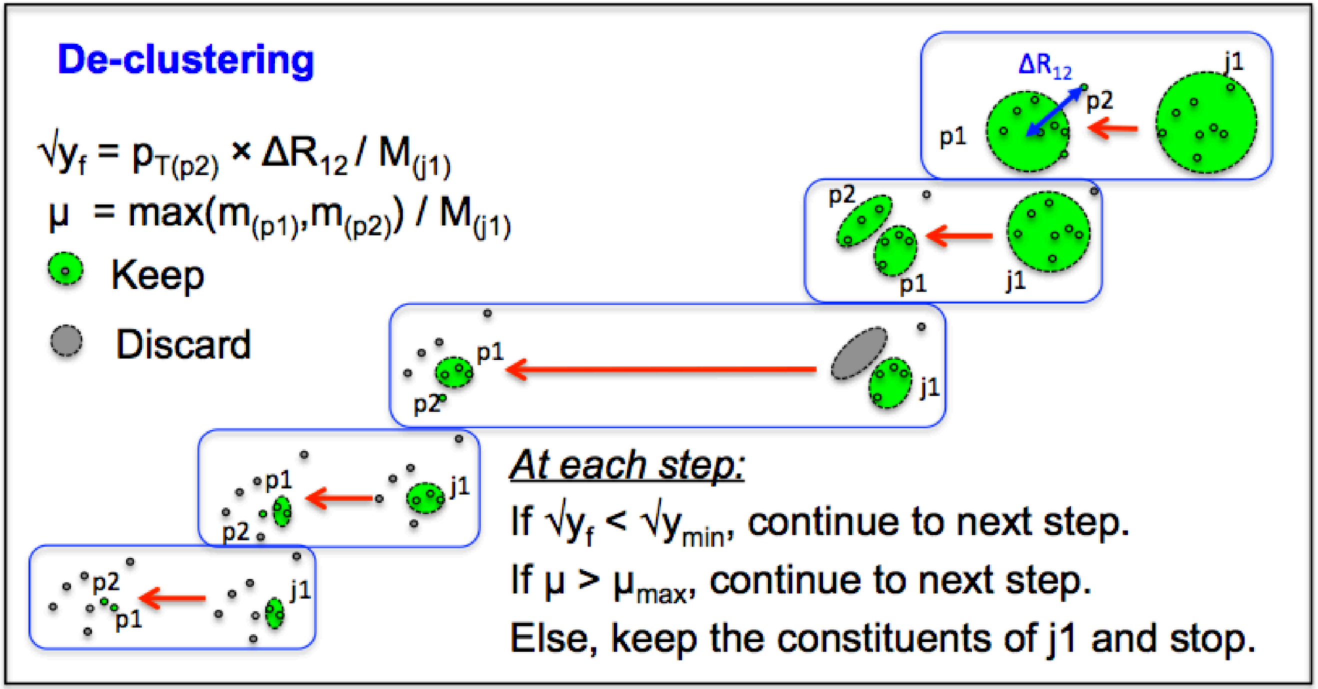
- At LHC $\sqrt{s} \gg M_{EW}$ => high p_T *boosted* objects
  - Also choose high-p_T region to reduce QCD backgrounds
  - Decay products merge into single *fat* jet
  - Need to look at *substructure* for *tagging*

- Also high-pileup conditions
  - Substructure *grooming* can remove soft contributions



Jets Substructure: Algorithms

- Identify hard constituents via *splitting, decomposition*

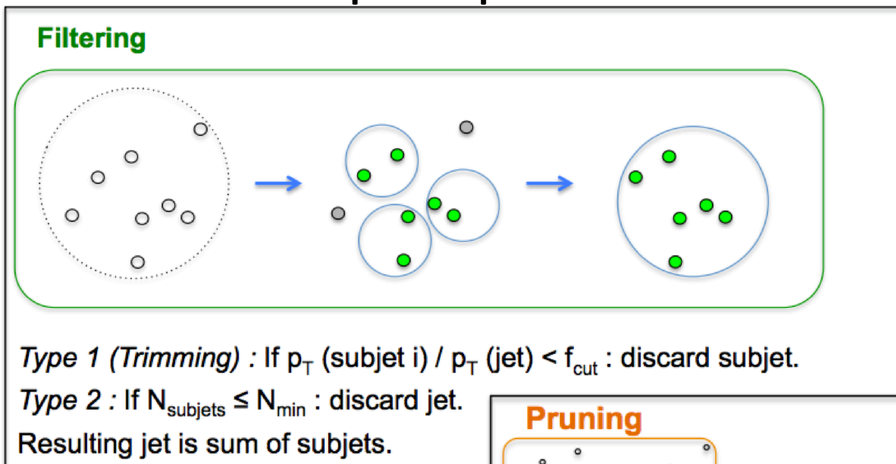


For example (explored in ATLAS results here)

- *BDRS aka Mass-drop Filtering*:
  - *Cambridge-Aachen (CA)* R=1.2 Fat jet
  - Split with $\mu_{12} < 2/3$, $\sqrt{y_f} > 0.3$
- BDRS-A:
  - CA R=1.2
  - Split $\sqrt{y_f} > 0.2$

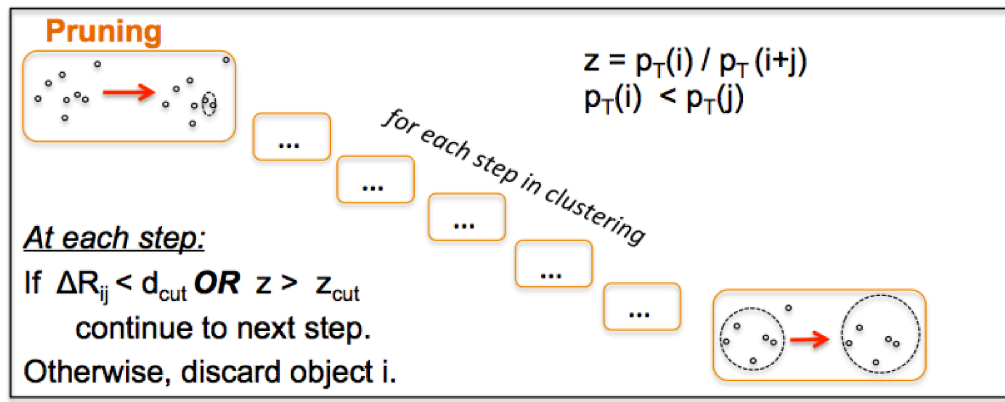
Jets Substructure: Algorithms

- Improve resolution and pileup resistance via *trimming, filtering, pruning*



For example

- *BDRS* Filtered with $R_{subjet} = \min(0.3, R_{12}/2)$
- BDRS-A Filtered with $R_{subjet} = 0.3$



E.g. (explored in CMS results here)

Pruned CA R=0.8 jets

- $z_{cut} = 0.1$; $d_{cut} = 0.5 * m^{orig}/p_T^{orig}$

Jet Substructure: *some* variables

Also make use of substructure variables, *for example*

- Mass drop μ_{12} , Momentum Fraction $\sqrt{y_f}$
- N-subjettiness: $$\tau_N = \frac{\sum_k p_{\mathrm{T},k}(\min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., R_{N,k}\})^\beta}{\sum_k p_\mathrm{T}(R_0)^\beta}$$
  - Sum over jet constituents (*k)*
  - Small if jet consistent with N subjets hypothesis
  - τ_2/τ_1 used to discriminate 2-body decays from W bosons
- Also Qjet volatility (ν_{QJets}), jet width, jet charge, planar flow, correlation functions

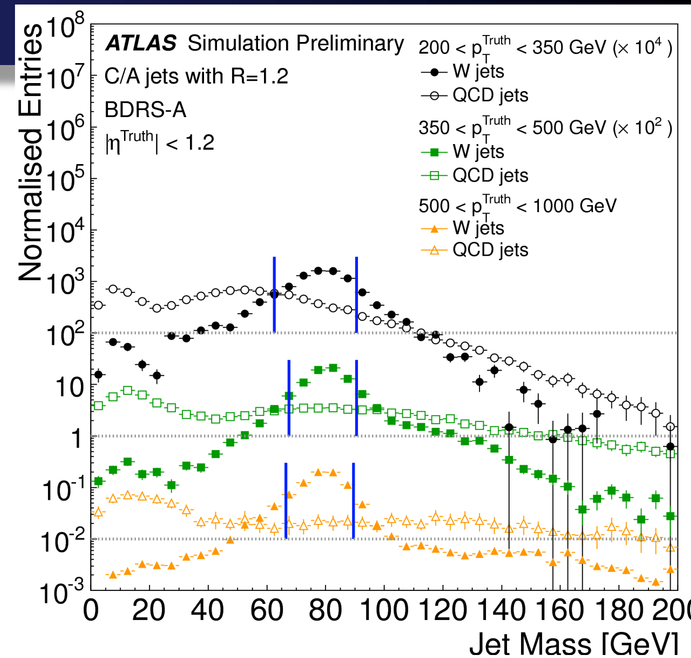
Outline

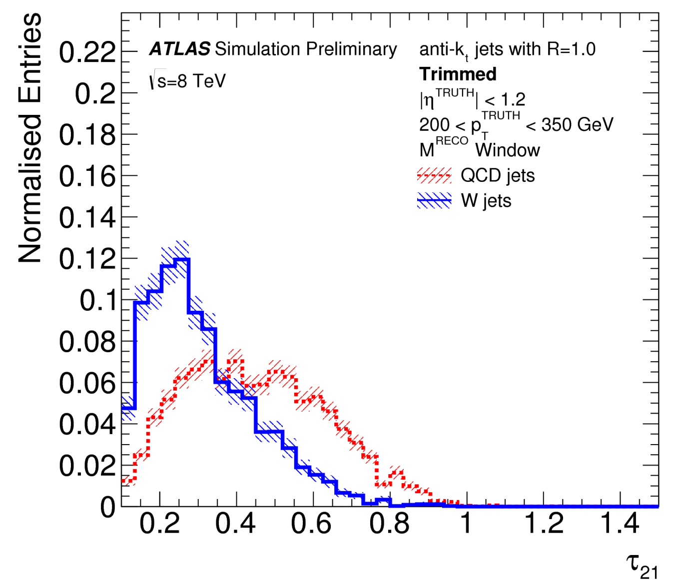
- ATLAS and CMS jet inputs
- Jets and Jet Substructure: motivation and algorithms
- Recent performance results:
    - W boson tagging
    - Top tagging
- Other recent highlights

W boson tagging ATLAS

ATL-PHYS-PUB-2014-004

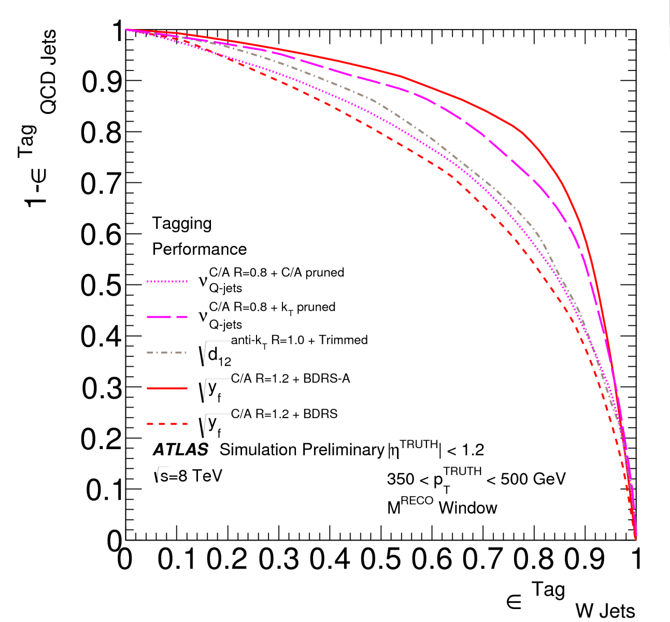
- Evaluate various algorithms in p_T ranges
  - Signal Monte-Carlo: Kaluza-Klein Graviton -> WW -> lvqq
  - Background: W+jets (Sherpa)
  - Main discriminant is jet mass (look for W peak):
    - Define window with 68% signal
- Also look at wide range of substructure variables and find optimum variable + alogorithm combinations.

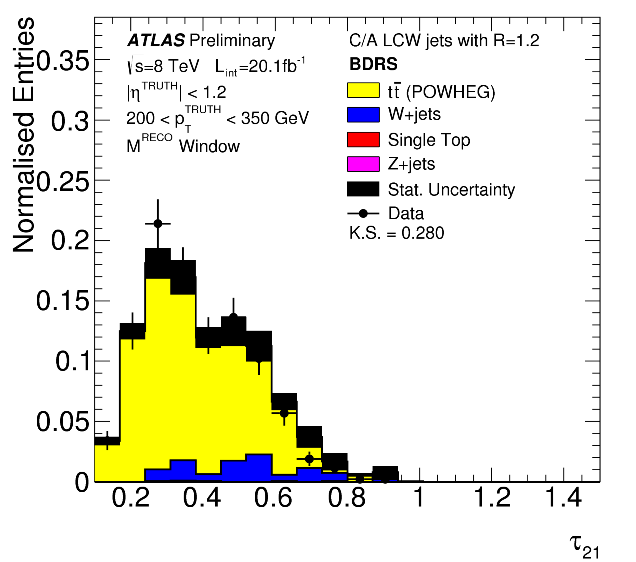




W boson tagging ATLAS

- Measured performance of each optimal combination
  - Algorithms perform similarly particularly when 'groomer + tagger' performance are taken together (i.e. within mass window)
- Data/MC comparison using W bosons from semileptonic tt-bar sample
  - High-purity (~98%) selection using HepTopTagger
  - Good agreement for relevant variables



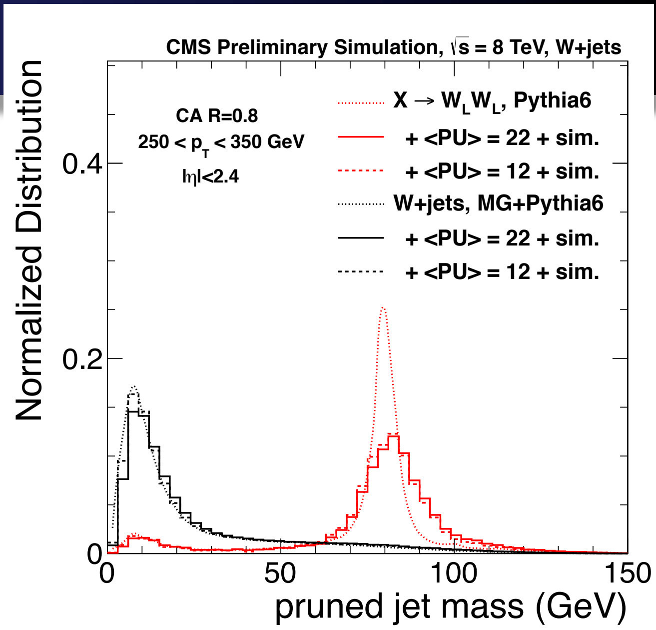


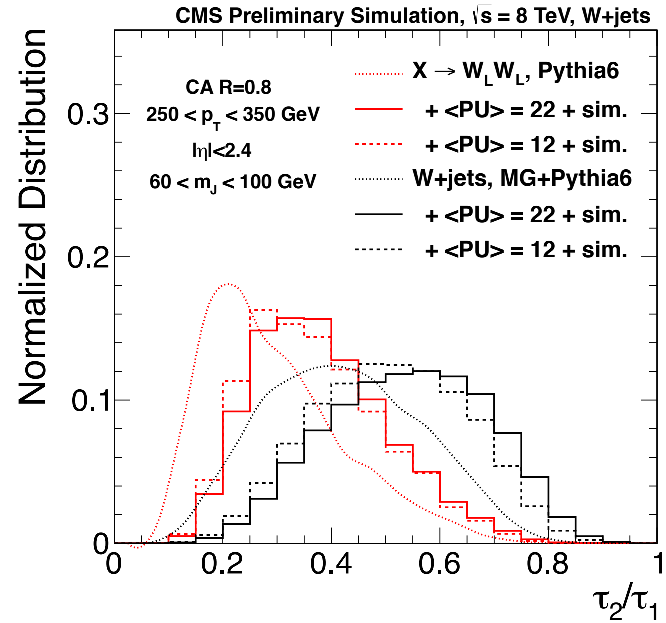
W boson tagging CMS

[CMS PAS JME-13-006]

- Use CA R=0.8 w/pruning
- Performance for Signal X->$W_L W_L$, Background: W + jets (MadGraph +Pythia6)

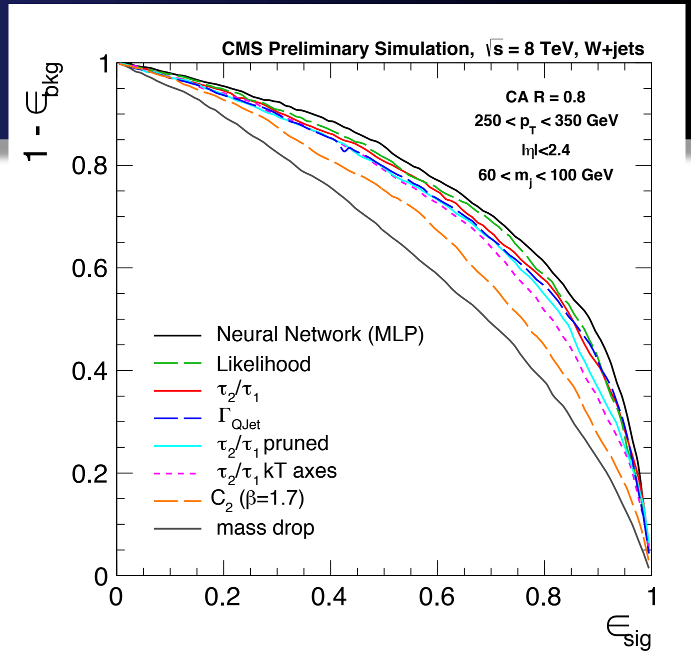
- Variety of variables:
  - And for N-subjetiness evaluate an alternative with one step optimization of exclusive k_T axes

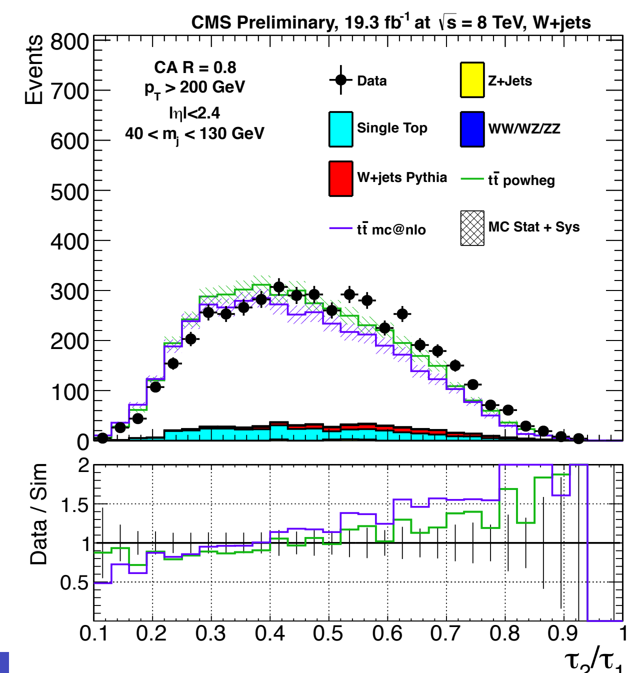




W boson tagging CMS

- Optimised τ_2/τ_1 is best performing variable
- Also look at a MVA
  - Offers little further improvement
- Data/MC Comparison in both W +jets and a semileptonic tt-bar
  - Decent agreement
- Form scale-factor for cut on τ_2/τ_1 (< 0.5) from data/MC efficiency in tt-bar sample (0.905 ± 0.08 (stat))





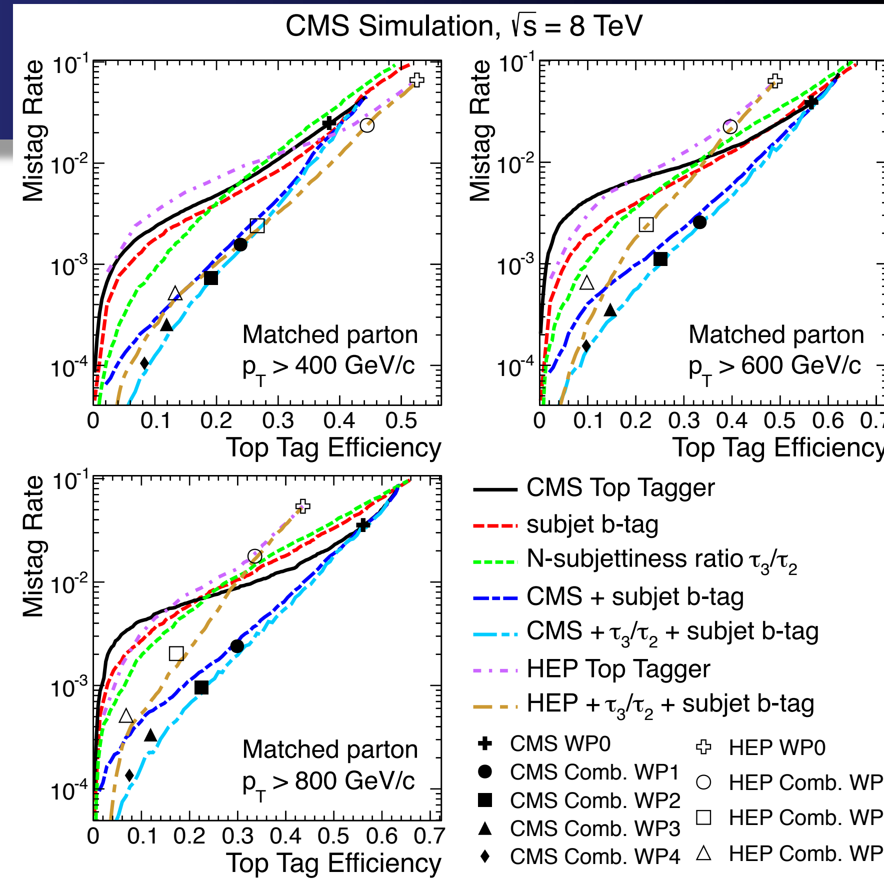
Outline

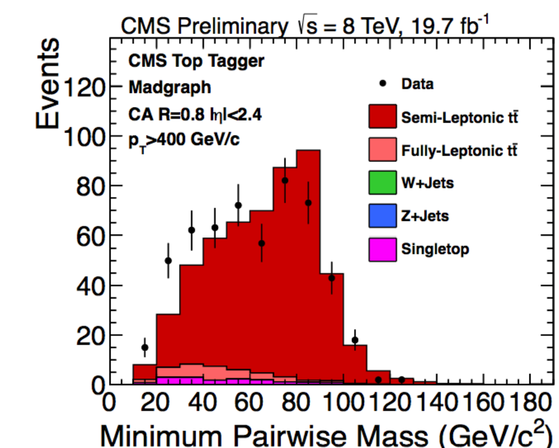
- ATLAS and CMS jet inputs
- Jets and Jet Substructure: motivation and algorithms
- Recent performance results:
    - W boson tagging
    - Top tagging
- Other recent highlights

Top Tagging CMS

CMS PAS JME-13-007

- Evaluate alternatives including *CMSTopTagger*
  - Two-stage decomposition of CA R=0.8 jets
- Add subjet b-tag and τ_3/τ_2 to form *Combined* tagger
  - Best for $p_T > 400$ GeV/c
  - Low p_T top not contained in R=0.8 jet - use *HEPTopTagger*
- η dependant scale-factors from data / MC efficiencies



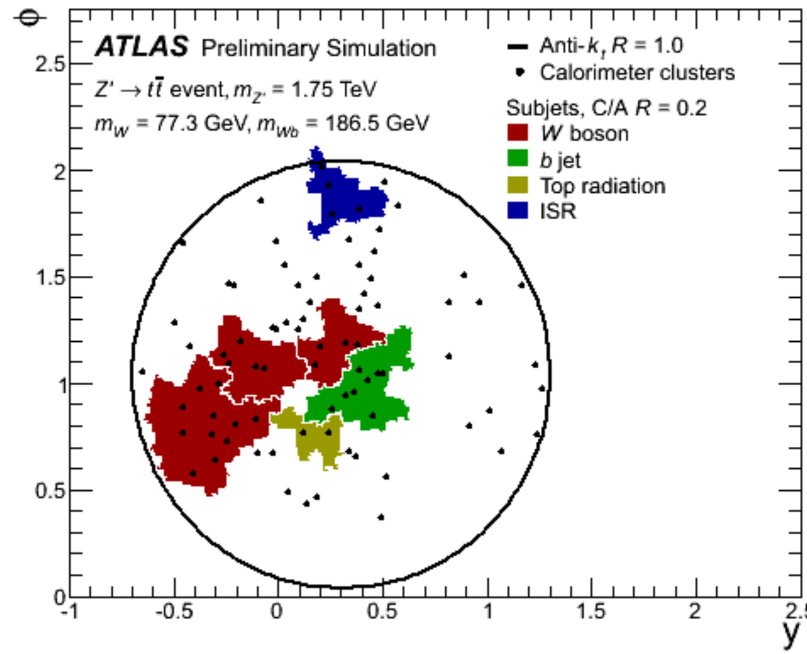


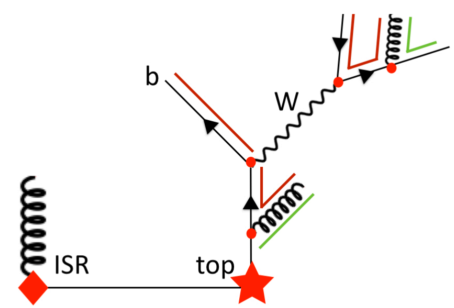
Top Tagging ATLAS

- Previous studies evaluate range of taggers e.g. ATLAS-CONF-2013-084
- Focus here on ATLAS-CONF-2014-003 on *Shower Deconstruction:*
- Input collection of CA R=0.2 subjets within Akt R=1.0 jet. Four-momenta $\{p\}_N = \{p_1,.., p_N\}$
- Different series of parton branchings that could build this gives *shower histories* $\{p,c^j\}_N$ that are assigned to categories c_i^j
- Assign splitting probabilities. Form likelihood ratio:

$$\chi_{SD}(\{p\}_N) = \frac{P(\{p\}_N|S)}{P(\{p\}_N|B)} = \frac{\sum_{\text{histories}} P(\{p, c^j\}_N|S)}{\sum_{\text{histories}} P(\{p, c^j\}_N|B)}$$

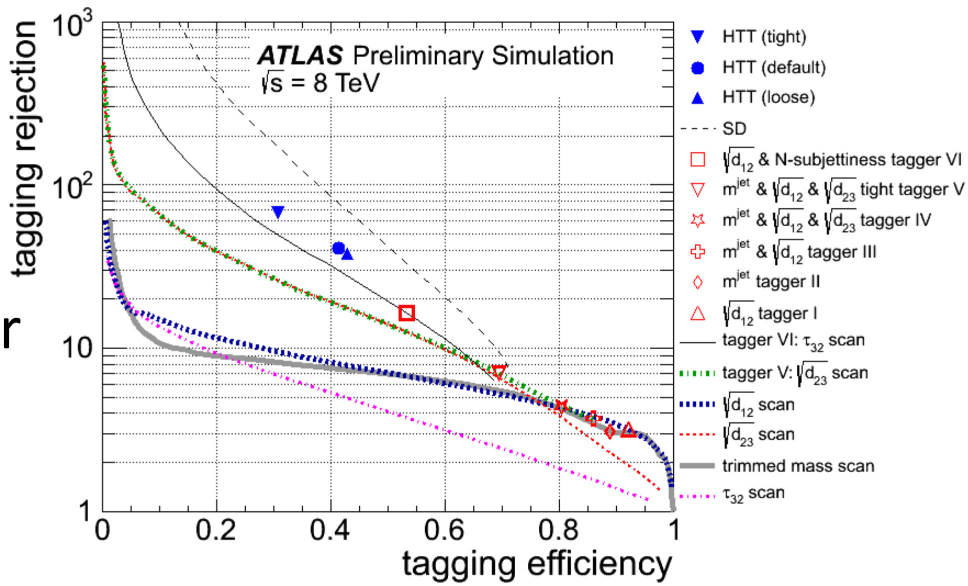
One (of >1500) (signal) shower histories:

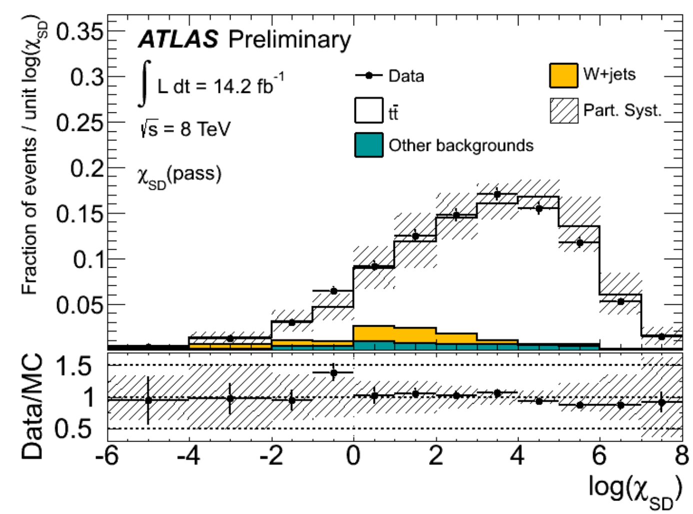




ATLAS Shower Deconstruction

- Performance measured using same samples as ATLAS-CONF-2013-084
  - Improved performance over range of efficiency
  - Not including systematics
- Data / MC comparison
  - Satisfactory agreement and stable with pileup



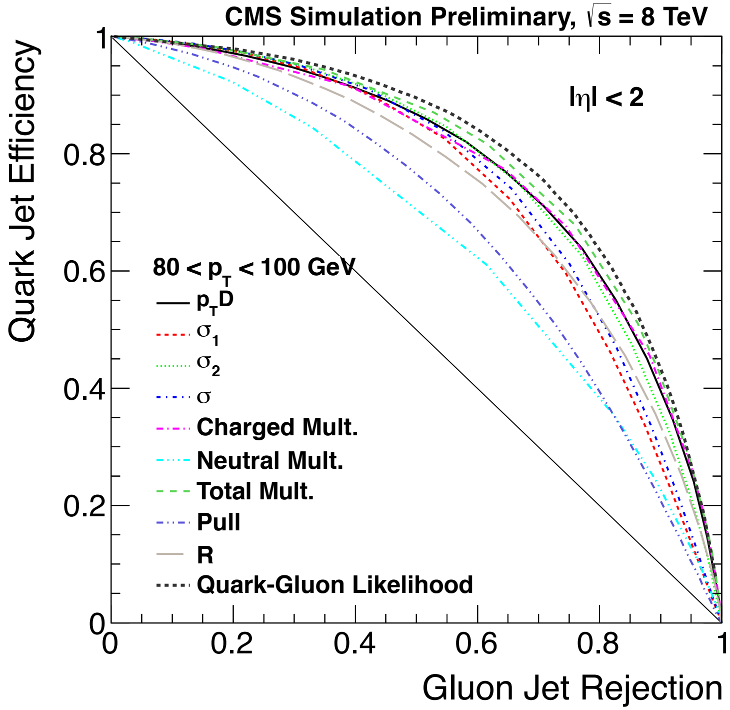


Outline

- ATLAS and CMS jet inputs
- Jets and Jet Substructure: motivation and algorithms
- Recent performance results:
    - W boson tagging
    - Top tagging
- Other recent highlights

Some other recent highlights

- Quark-gluon tagging
  CMS-JME-13-002
  ATLAS arXiv:1405.6583

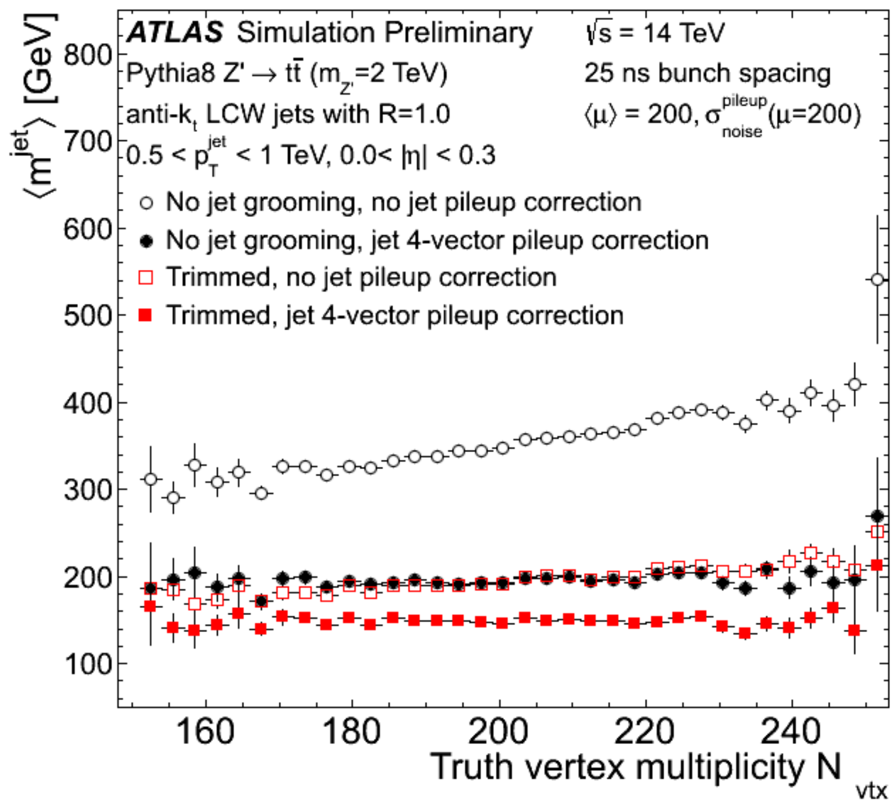


Fragmentation Function

$$p_{\mathrm{T}}D = \frac{\sqrt{\sum_i p_{\mathrm{T},i}^2}}{\sum_i p_{\mathrm{T},i}}$$

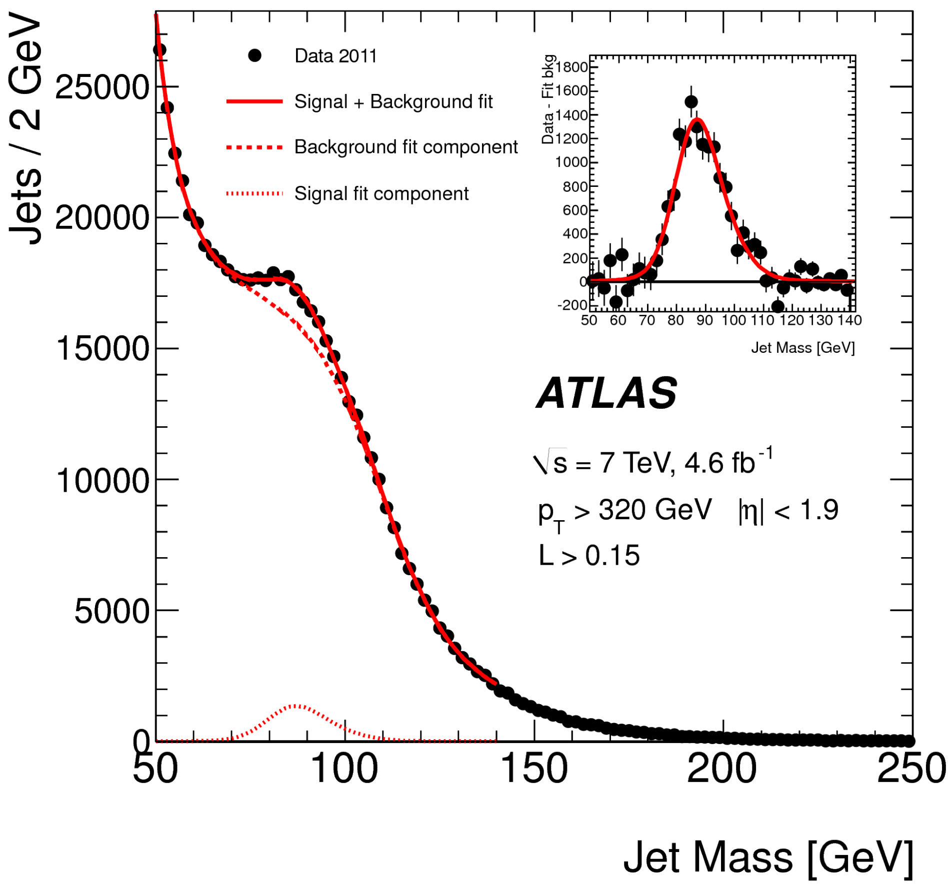
Some other recent highlights

- Quark-gluon tagging
  [CMS-JME-13-002]
  [ATLAS arXiv:1405.6583]
- b-tagging in boosted jets
  [CMS-PAS-BTV-13-001]
- Jet pull performance
  [ATLAS-CONF-2014-048]
- Pile-up jet id/subtraction
  [CMS-PAS-JME-13-005]
  [ATLAS High Mu]



Some other recent highlights

- Quark-gluon tagging
  CMS-JME-13-002
  ATLAS arXiv:1405.6583
- b-tagging in boosted jets
  CMS-PAS-BTV-13-001
- Jet pull performance
  ATLAS-CONF-2014-048
- Pile-up jet id/subtraction
  CMS-PAS-JME-13-005
  ATLAS High Mu
- Physics – e.g.Cross-section of high p_T vector bosons:
  ATLAS CERN-PH-EP-2014-123



$$\sigma_{W+Z} = 8.5 \pm 0.8 \ \ (\text{stat.}) \pm 1.5 \ \ (\text{syst.}) \text{ pb.}$$

Some other recent highlights

- Quark-gluon tagging
  CMS-JME-13-002
  ATLAS arXiv:1405.6583
- b-tagging in boosted jets
  CMS-PAS-BTV-13-001
- Jet pull performance
  ATLAS-CONF-2014-048
- Pile-up jet id/subtraction
  CMS-PAS-JME-13-005
  ATLAS High Mu
- Physics – e.g.Cross-section of high p_T vector bosons:
  ATLAS CERN-PH-EP-2014-123

And Boost2014 next week!

Conclusions

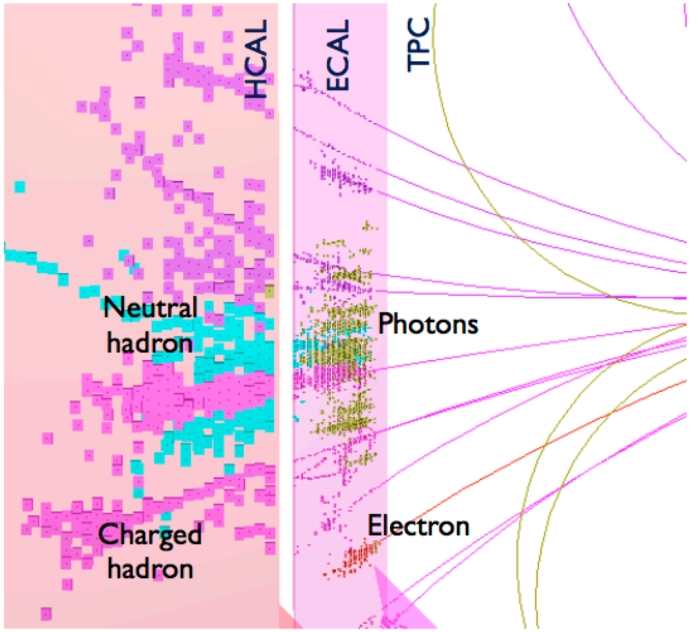
- Jet substructure offers powerful techniques, essential for analyses at LHC Run 2 and beyond
  - Tagging to unveil composition of boosted objects, grooming for resolution and pileup resistance
- Considerable recent activity in validating and optimising these techniques at the LHC
  - Building optimum taggers with scale-factors
- Now need to finalise methods for scale-factors and uncertainties – ready to use for Run 2 physics!

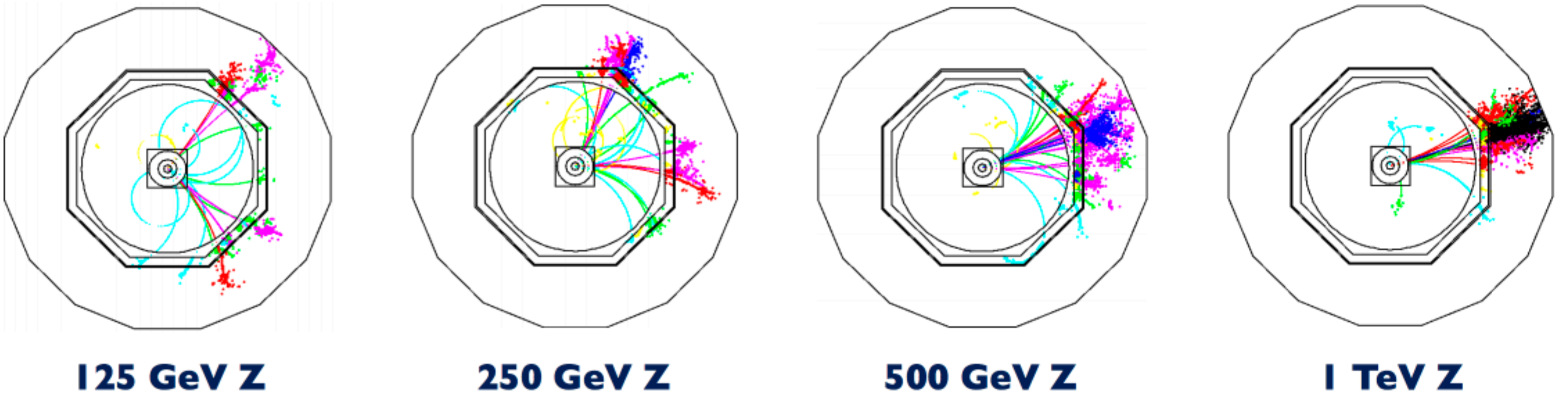
Extra Slides

And Beyond

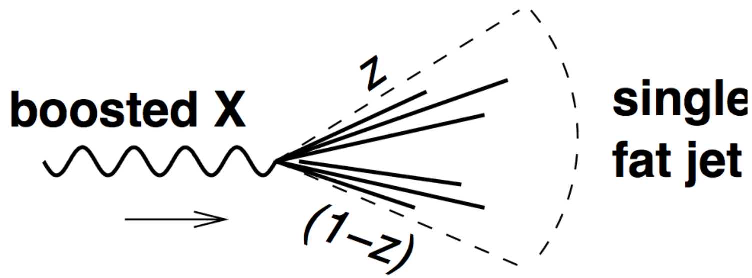
From J.S.Marshall CHEF2013

- ILC exploring fine granularity particle flow calorimetry – hardware and software
- At CLIC energies expect to see merged jets





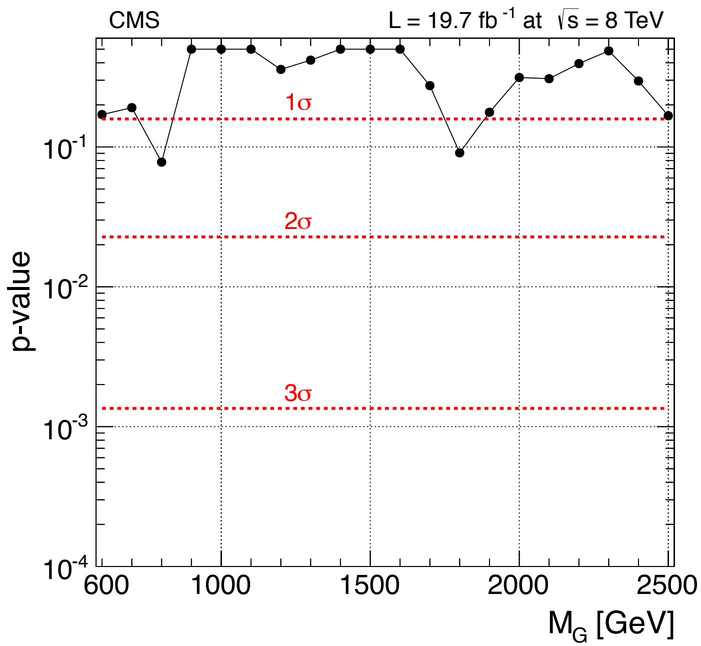
Physics Motivation- Boosted bosons



Happens for $p_t \gtrsim 2m/R$

$p_t \gtrsim 320$ GeV for $m = m_W$, $R = 0.5$

CMS
[EXO-13-009]

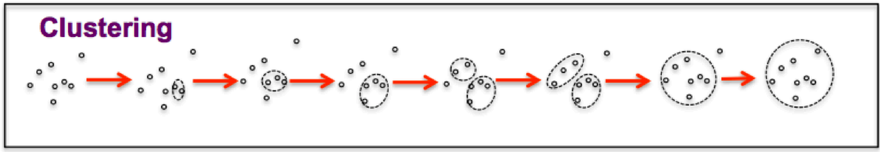


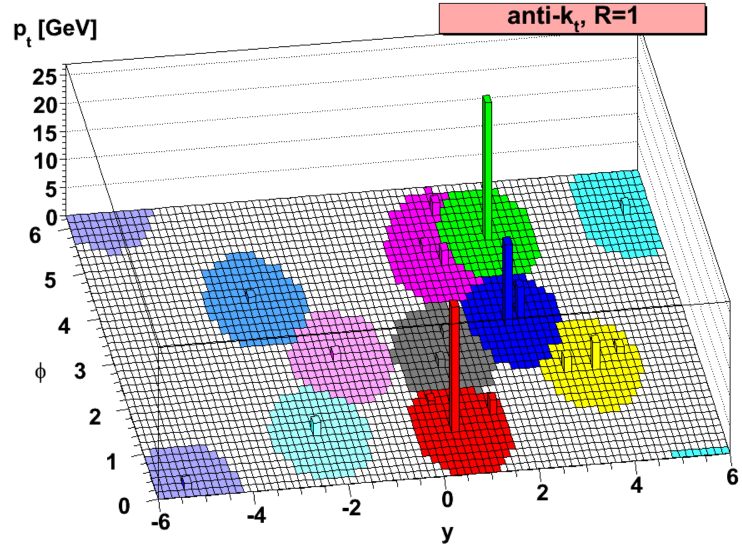
Jet Algorithms at LHC

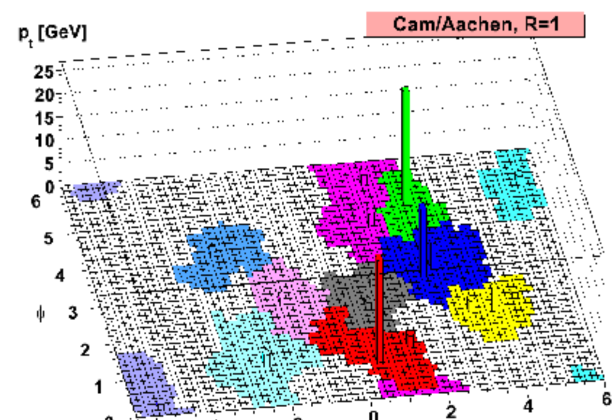
- Partons in ATLAS/CMS produce dispersed hadrons - clustered together by jet algorithm
- Standard LHC choice uses distance parameter d_{ij}:

$$d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n})\Delta R_{ij}^2 / R^2$$
$$d_{iB} = p_{Ti}^{2n},$$

  - *n=-1 Anti-k_T* used extensively at LHC, regular shaped jets, robust to pileup
  - *n=1 k_T* algorithm
  - n=0 *Cambridge-Aachen (CA)* – only angular info
- Can undo clustering to reveal hard structure







Cacciari, Salam, Soyez JHEP 0804 (2008) 063

Boson Tagging

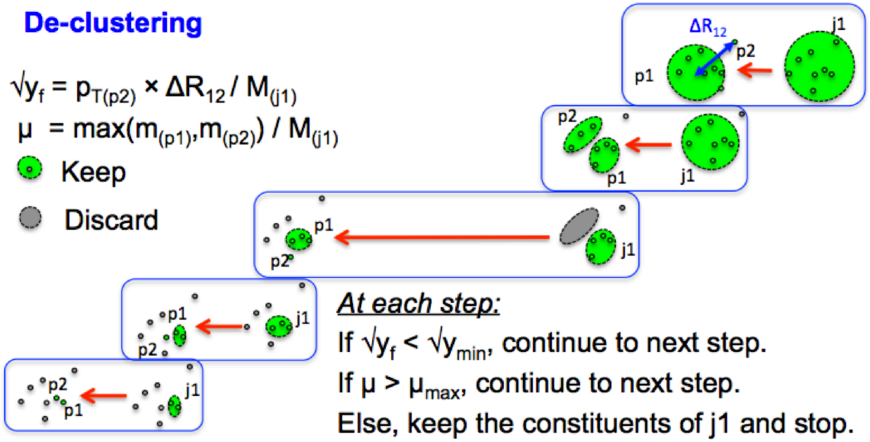
De-clustering

$\sqrt{y_f} = p_{T(p2)} \times \Delta R_{12} / M_{(j1)}$

$\mu = \max(m_{(p1)}, m_{(p2)}) / M_{(j1)}$

Keep

Discard



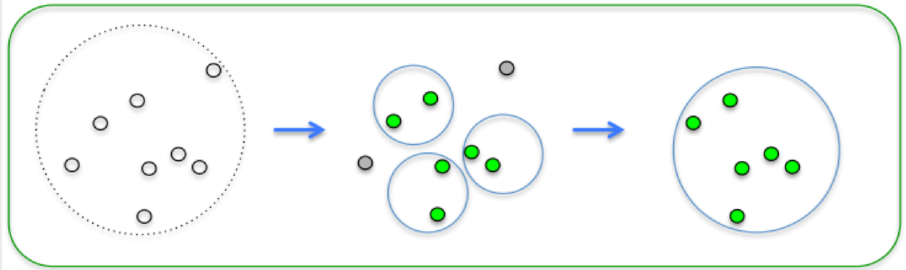
At each step:

If $\sqrt{y_f} < \sqrt{y_{min}}$, continue to next step.

If $\mu > \mu_{max}$, continue to next step.

Else, keep the constituents of j1 and stop.

Filtering



Type 1 (Trimming) : If p_T (subjet i) / p_T (jet) < f_{cut} : discard subjet.

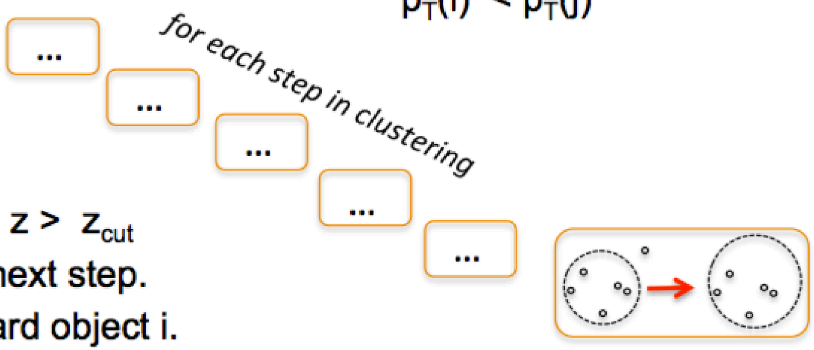
Type 2 : If $N_{subjets} \leq N_{min}$: discard jet.

Resulting jet is sum of subjets.

Pruning

$z = p_T(i) / p_T(i+j)$

$p_T(i) < p_T(j)$



At each step:

If $\Delta R_{ij} < d_{cut}$ ***OR*** $z > z_{cut}$

continue to next step.

Otherwise, discard object i.

Jet Substructure – *some* variables

- Mass drop μ_{12} , Momentum Fraction $\sqrt{y_f}$
- N-subjettiness:

$$\tau_N = \frac{\sum_k p_{T,k}(\min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., R_{N,k}\})^\beta}{\sum_k p_T (R_0)^\beta}$$

  - Small if jet has N subjets
  - τ2/τ1 used to discriminate 2-body decays from W bosons
- Qjet volatility
  - Recluster with a weight $\omega_{ij}(\alpha)$

$$\omega_{ij}(\alpha) = \exp\left\{-\alpha \frac{\Delta R^2_{ij} - \Delta R^2_{\min}}{\Delta R^2_{\min}}\right\},$$

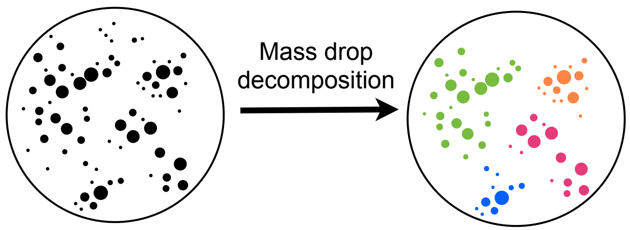
  - Measured a volatility v_{QJets}

$$v_{\mathrm{QJets}} = \frac{\sqrt{\langle m^2 \rangle - \langle m \rangle^2}}{\langle m \rangle}.$$

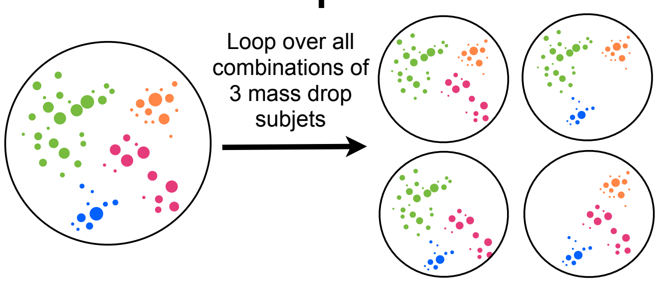
- Also jet width, jet charge, planar flow, correlation functions

HEP Top Tagger details

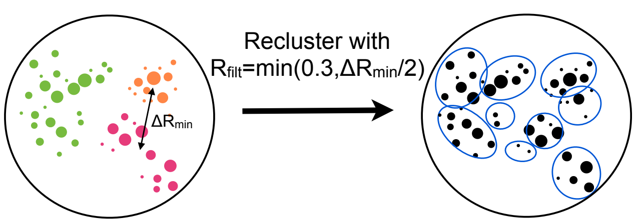
Step 1:



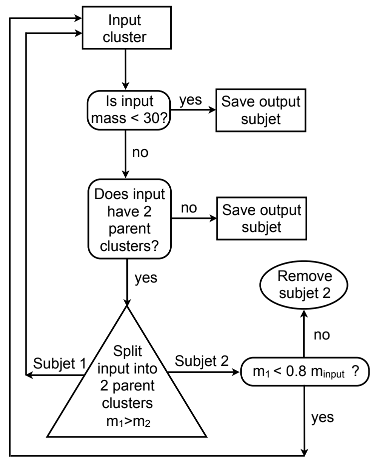
Step 2:



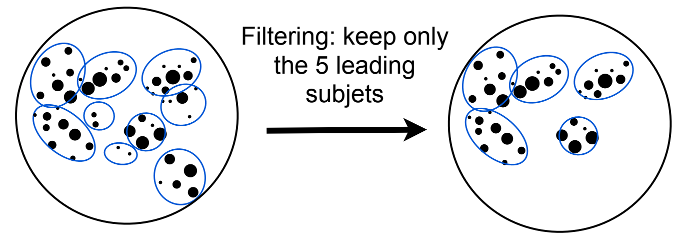
Step 3:



HEP Top Tagger
Mass drop decomposition

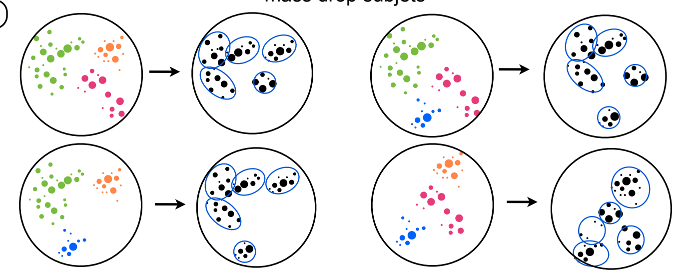


Step 4:

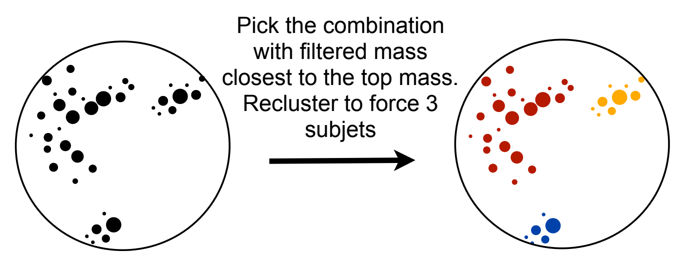


Step 5:

Repeat reclustering and filtering procedure for all combinations of 3 mass drop subjets



Step 6:



HEP Top Tagger - W mass selection

Bi-dimensional distribution based on the ratio of subjet pairwise masses

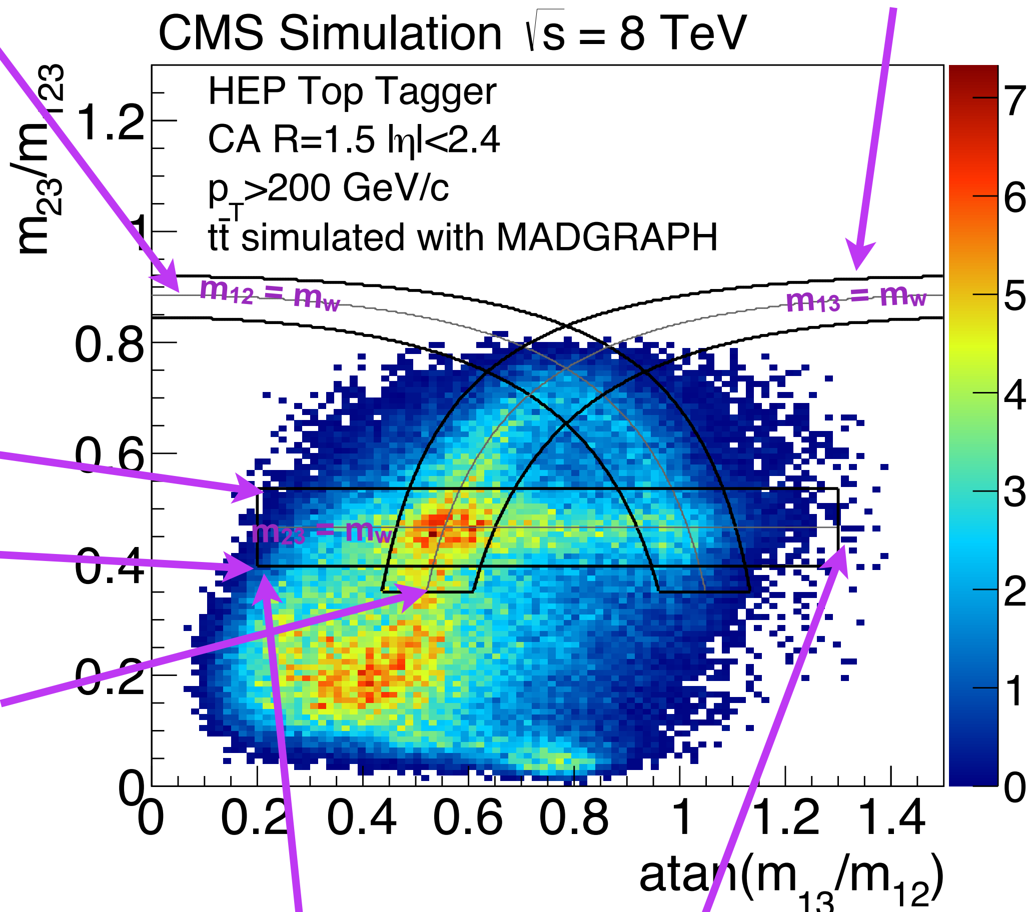
$$R_{\min}^2\left(1+\left(\frac{m_{12}}{m_{13}}\right)^2\right) < 1-\left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\max}^2\left(1+\left(\frac{m_{12}}{m_{13}}\right)^2\right) \qquad R_{\min}^2\left(1+\left(\frac{m_{13}}{m_{12}}\right)^2\right) < 1-\left(\frac{m_{23}}{m_{123}}\right)^2 < R_{\max}^2\left(1+\left(\frac{m_{13}}{m_{12}}\right)^2\right)$$

$$R_{\min} < \frac{m_{23}}{m_{123}} < R_{\max}$$

$$R_{\max} = (1+f_W) \times m_W/m_t$$

$$R_{\min} = (1-f_W) \times m_W/m_t$$

$$\frac{m_{23}}{m_{123}} > 0.35$$

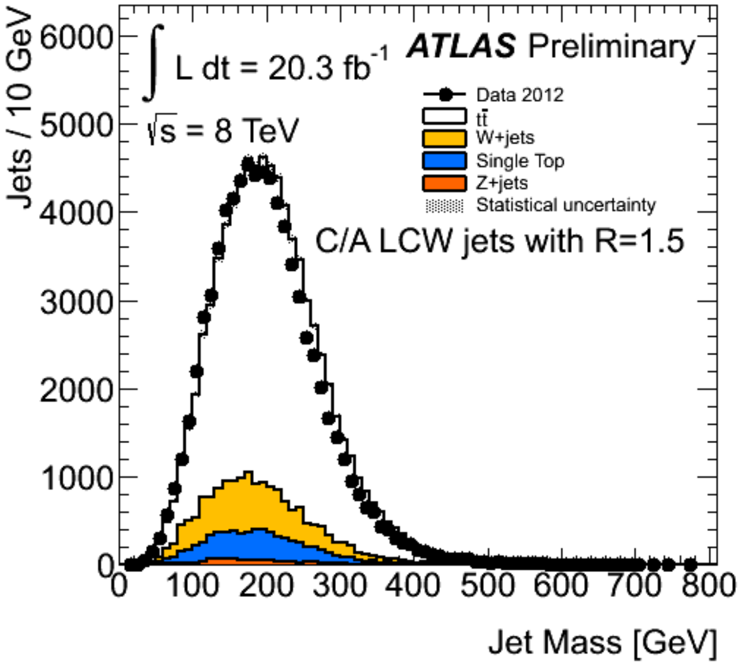


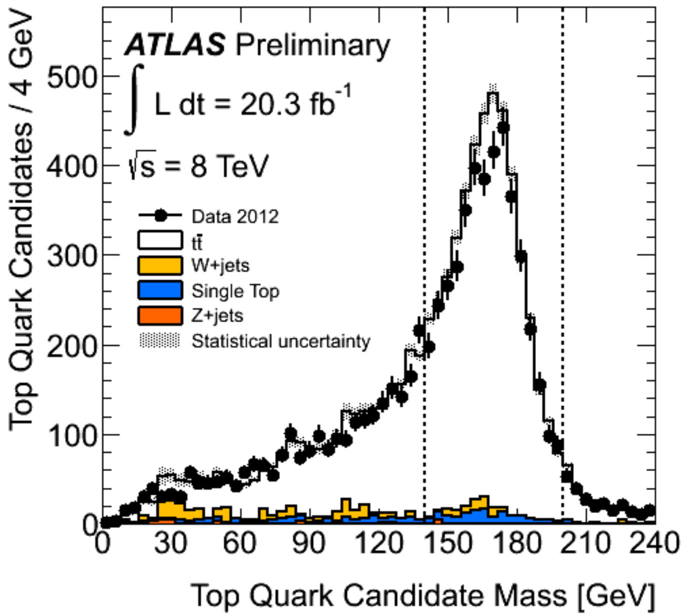
$$0.2 < \arctan\frac{m_{13}}{m_{12}} < 1.3$$

Performance: top mass

[ATLAS-CONF-2013-084]

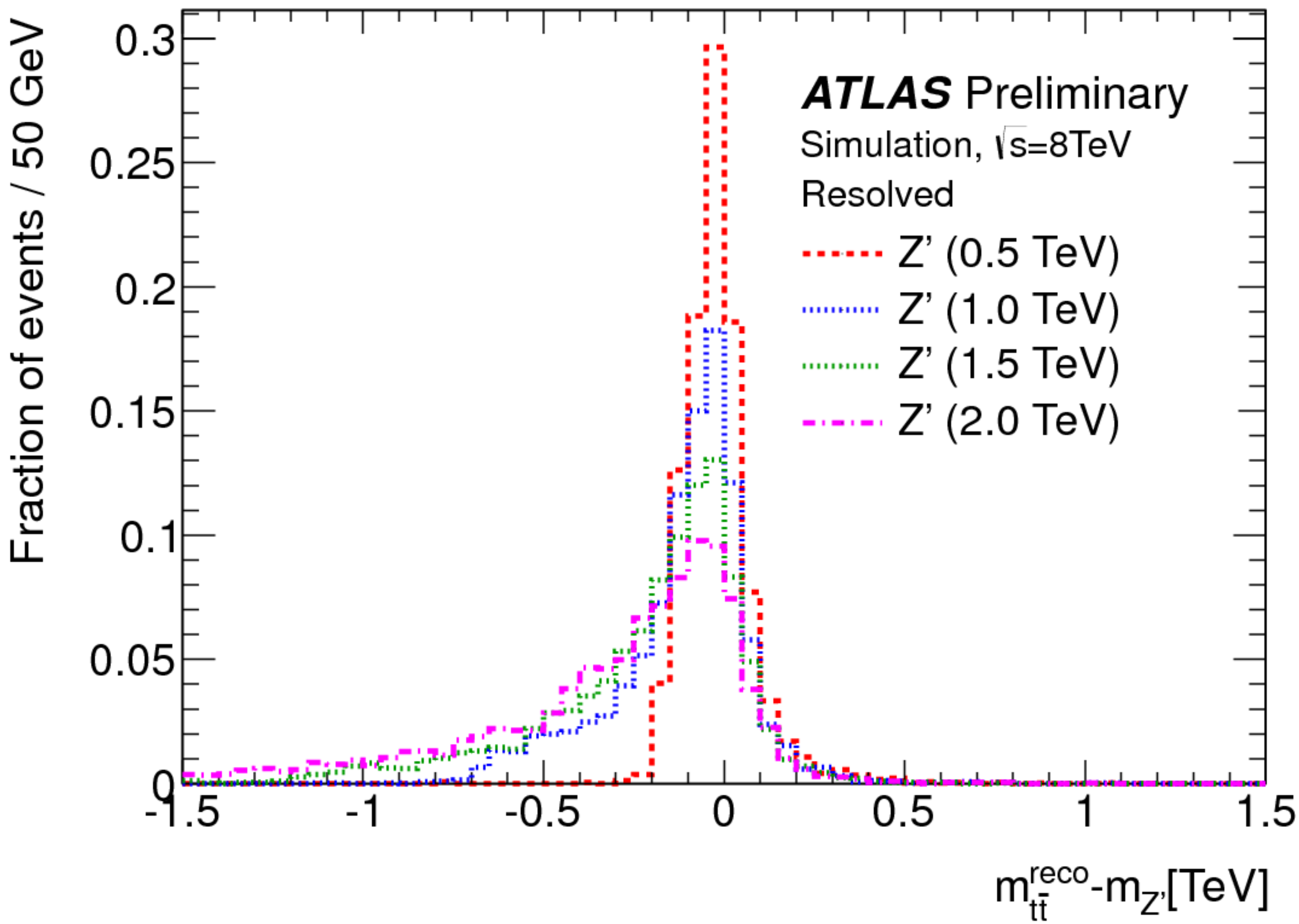
- CA R=1.5 Jets with HEPTopTagger (p_T > 200 GeV)
- Increases purity from 86% $t\bar{t}$ and single-top-quark processes before tagging to 98% in candidate mass window 140 < m_t < 200 GeV



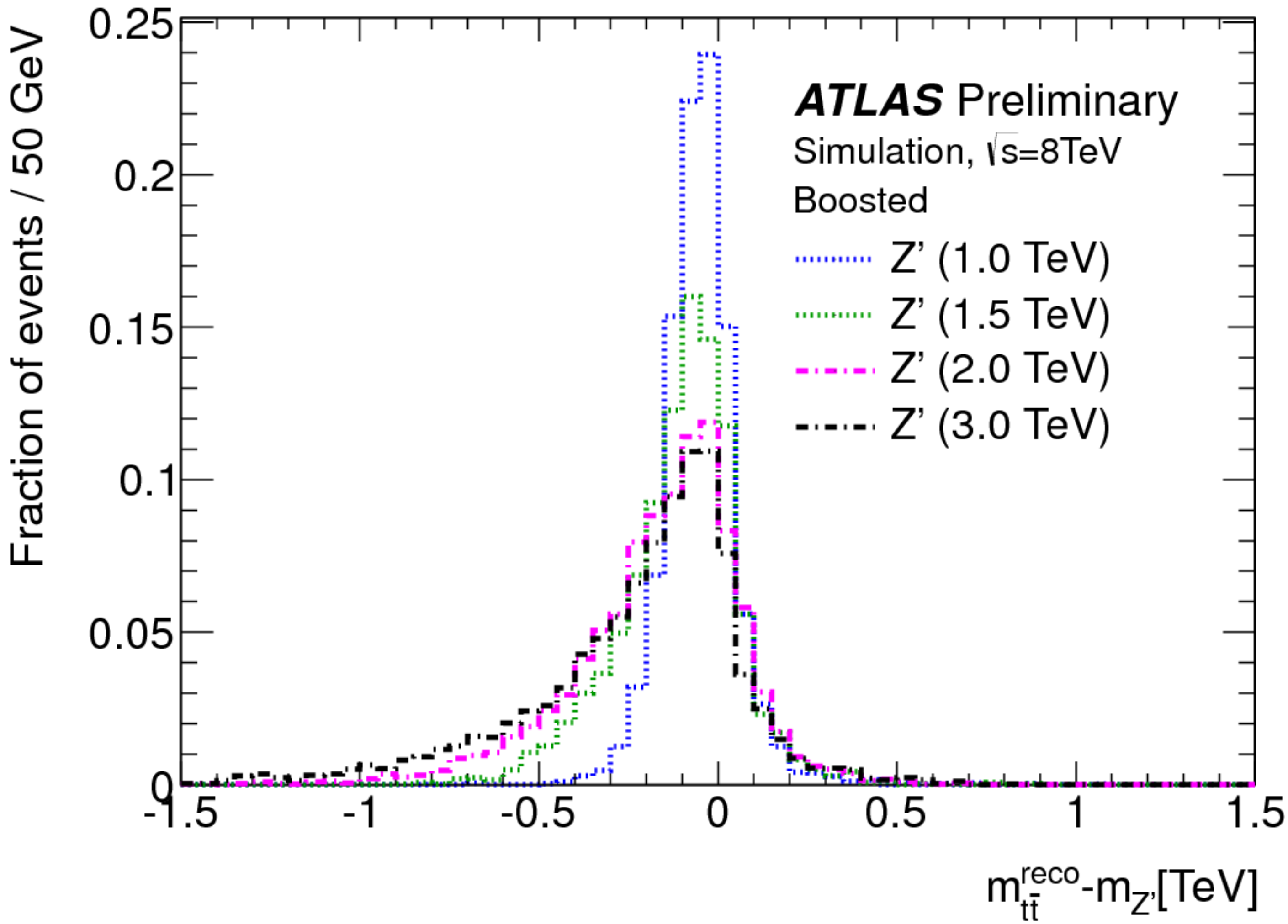


Reconstructed tt-bar mass

ATLAS-CONF-2013-052



Reconstructed tt-bar mass

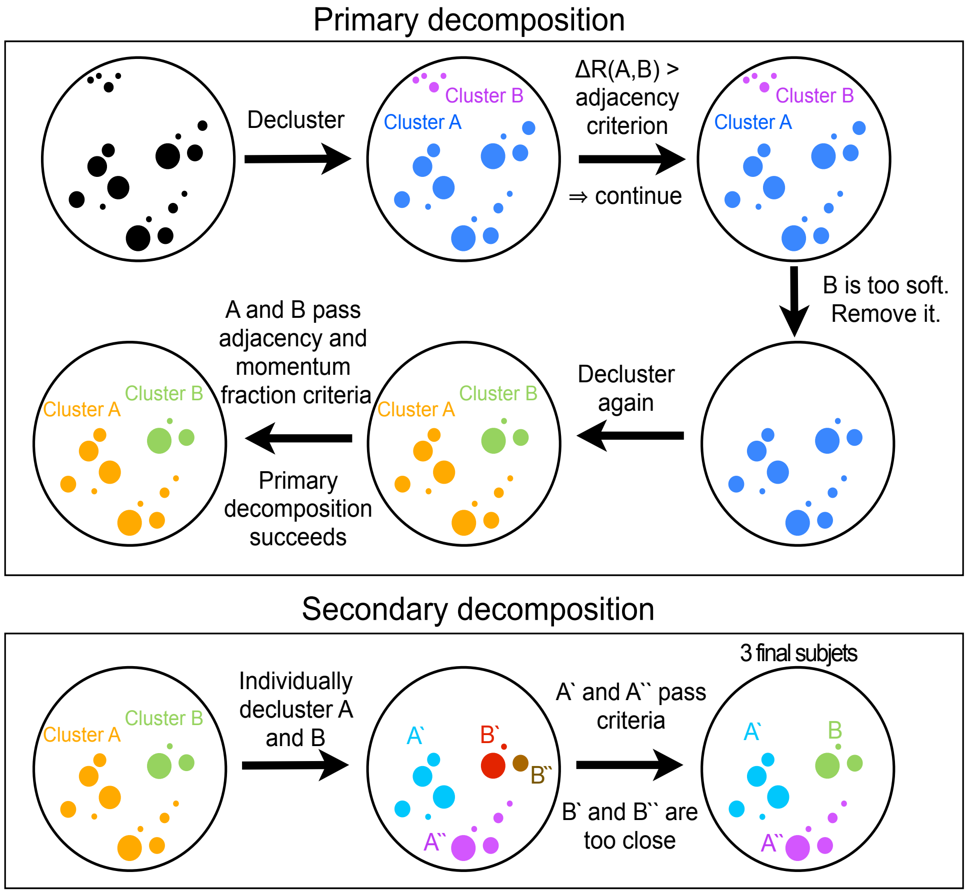


Top Tagging CMS

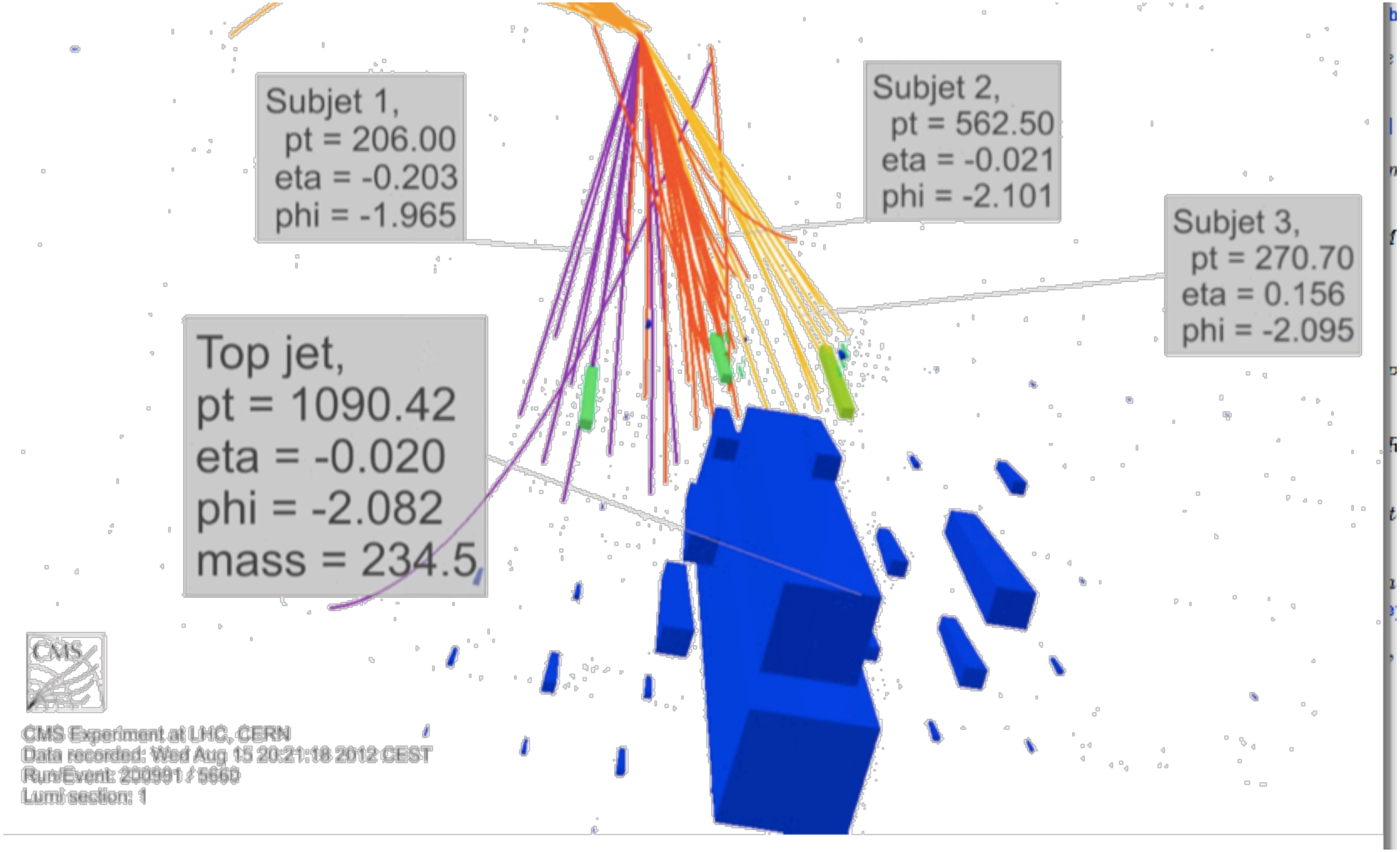
[CMS PAS JME-13-007]

Performance of multiple techniques inc. *CMS Top Tagger:*

- Based on Kaplan et. al
  Phys. Rev. Lett. **101** (2008) 142001
- Input CA R=0.8 jets
- Primary decomposition: find 2 well separated subclusters with significant p_T fraction
- If succeeds then do secondary decomposition
- Form m_{jet}, $N_{subjets}$, m_{min} (min pairwise mass)
- Top tagged if $m_{jet} \sim m_{top}$ $N_{subjets}>2, m_{min} \sim m_W$



CMS Boosted Top Event



Large-R Jet Calibration

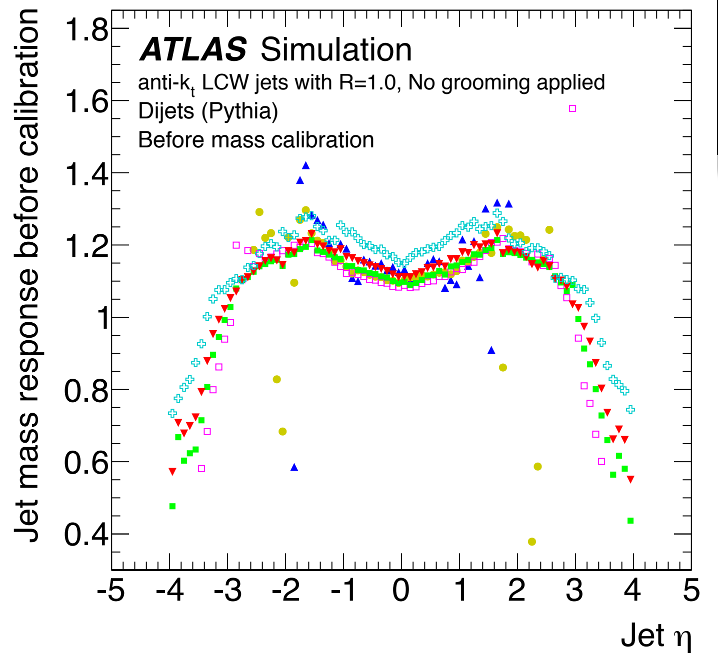
ATLAS: JHEP09 (2013) 076

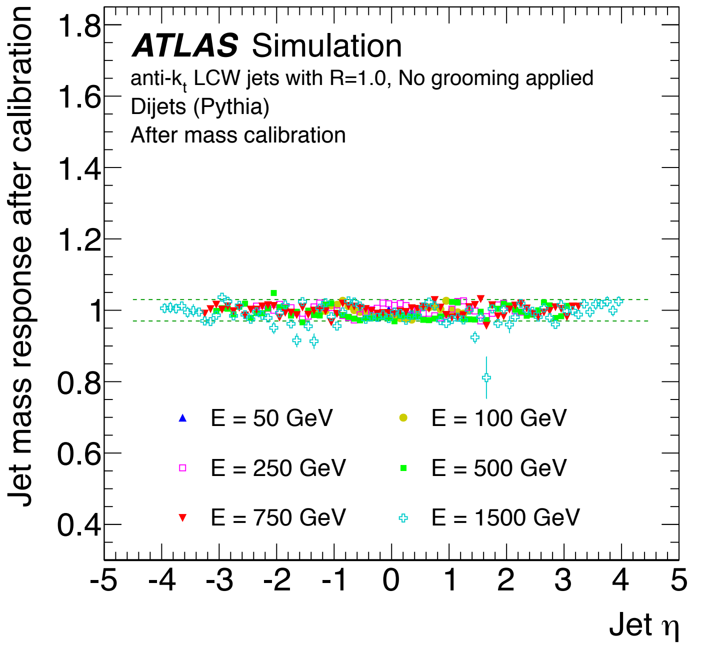
Jet energy scale calibration

- Correct calorimeter response to true jet energy as done for small-R jets on ATLAS Eur. Phys. J. C 73 (2013) 2304
- Derived from PYTHIA MC, no pileup correction

Jet mass calibration

- Mass response: mean of a Gaussian to core of reco jet mass/ true mass (m_{reco}/m_{true})



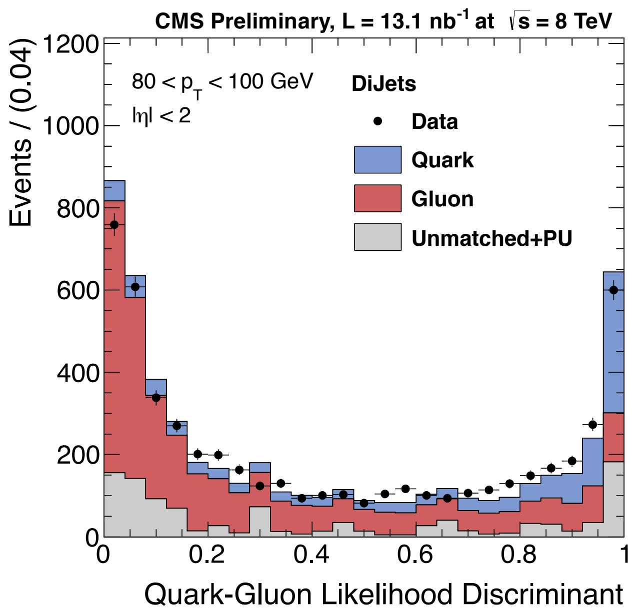


Quark-gluon tagging – more info

[CMS-JME-13-002]

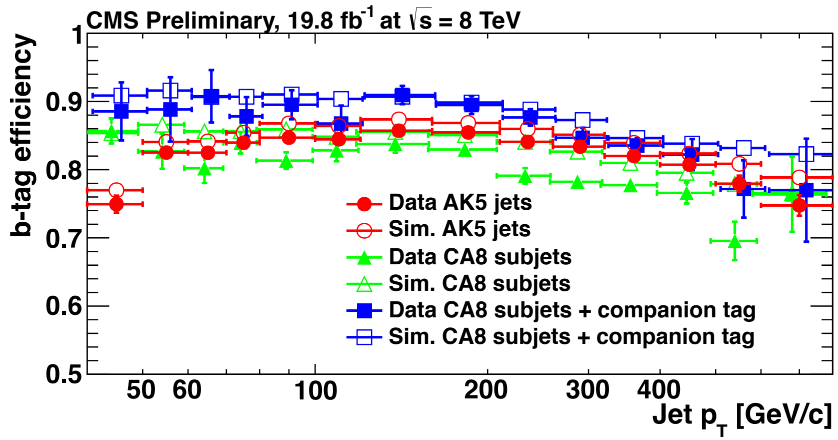
- Likelihood based on:
  - Width of minor axis of jet ellipse
  - Fragmentation
  - Total Multiplicity
- Validated with Z+jets, dijets
- Systematics from 2-parameter smearing
- Also [ATLAS arXiv:1405.6583]

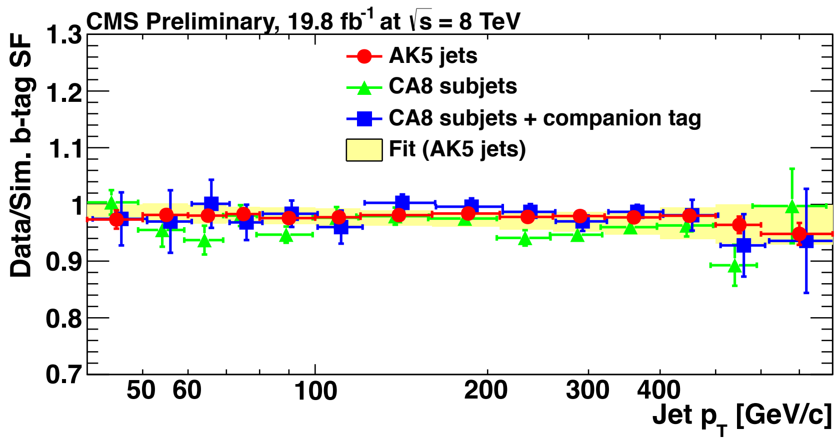
$$\sigma_2 = \left(\lambda_2 / \sum_i p_{\mathrm{T},i}^2\right)^{1/2}$$



Some other recent highlights

- Quark-gluon tagging
  [CMS-JME-13-002]
  [ATLAS arXiv:1405.6583]
- b-tagging in boosted jets
  [CMS-PAS-BTV-13-001]





Some other recent highlights

- Quark-gluon tagging
  CMS-JME-13-002
  ATLAS arXiv:1405.6583
- b-tagging in boosted jets
  CMS-PAS-BTV-13-001
- Jet pull performance
  ATLAS-CONF-2014-048