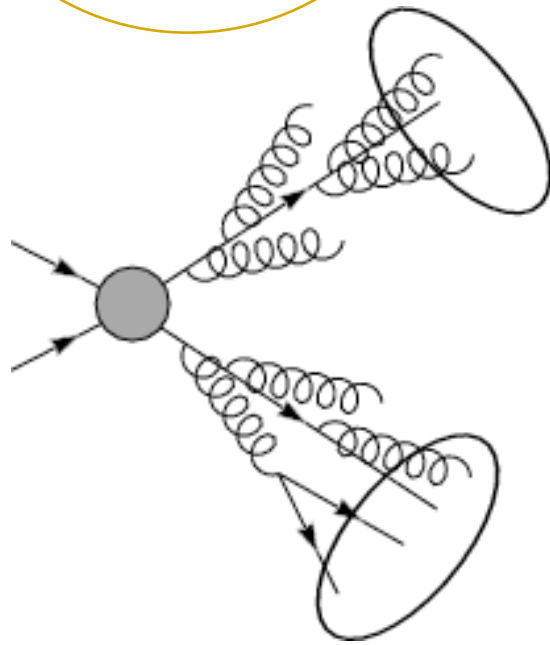




Hands-on Advanced Tutorial Session on Jet Substructure



A pedagogical introduction

The Jet Substructure HATS team
(Jake Anderson, Jim Dolen, Kalanand Mishra,
Ilya Osipenkov, Nhan Tran)

Twiki: [https://twiki.cern.ch/twiki/bin/view/CMS/
SWGuideHATSJetSubstructure](https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHATSJetSubstructure)

Indico: [https://indico.cern.ch/
conferenceDisplay.py?confid=251174](https://indico.cern.ch/conferenceDisplay.py?confid=251174)

Outline



- Jet clustering: algorithms and calibration
- Jet substructure: introduction and motivation
 - grooming: filtering, pruning, trimming, mass drop
- Jet substructure distributions for garden-variety jets
 - comparisons of data & simulation, experimental effects
- Jet substructure observables and signal reconstruction
 - illustrate the techniques with W and top tagging
- Summary



Some bibliography

Phenomenology

- ▶ Les Houches 2007 proceedings, [arXiv:0803.0678](#)
(Introduced modern lingo and conventions)
- ▶ Gavin Salam, 'Towards Jetography', [arXiv:0906.1833](#)
(Exhaustive description of jet algorithms, “manifesto” of Fastjet)
- ▶ Proceedings of BOOST 2010 and 2011:
[arXiv:1012.5412](#) and [arXiv:1201.0008](#)
- ▶ Gavin Salam’s lectures at CERN Academic Training (March 2011):
<http://indico.cern.ch/event/115078>
- ▶ Fastjet user manual, MC, G.P. Salam and G. Soyez, [arXiv:1111.6097](#).
Also available from fastjet.fr

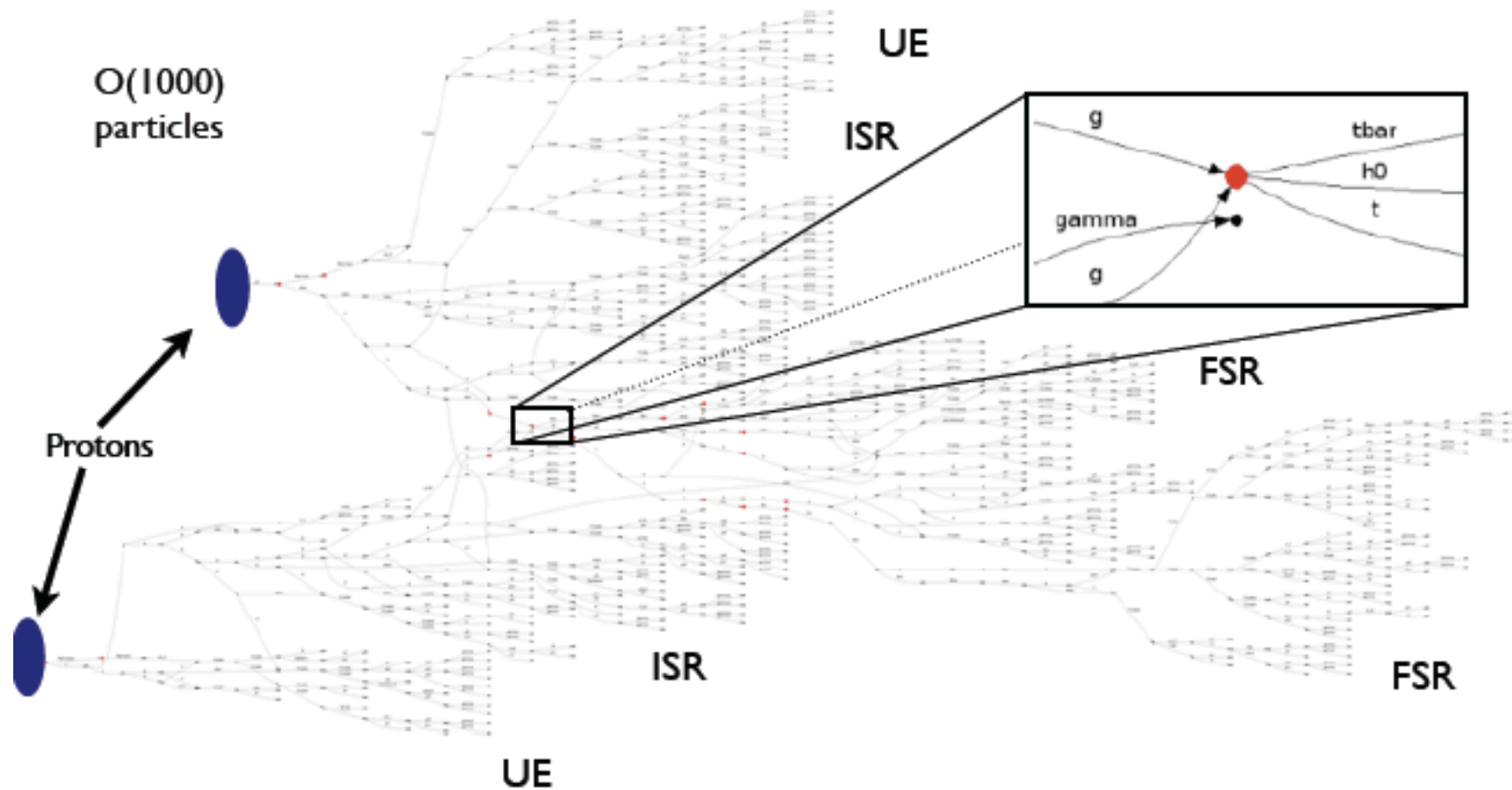
Experimental results

- ▶ **CMS collaboration**, “Jet mass in dijet and W/Z+jet events”, [arXiv:1303.4811](#)
- ▶ **ATLAS collaboration**, “Jet mass & substructure in incl. jets”, [arXiv:1203.4606](#)

Reality of event complexity at hadron colliders



A $t\bar{t}$ event, generated by Herwig++



Tedious for theorists and experimentalists

Taming the reality: a reductionist approach

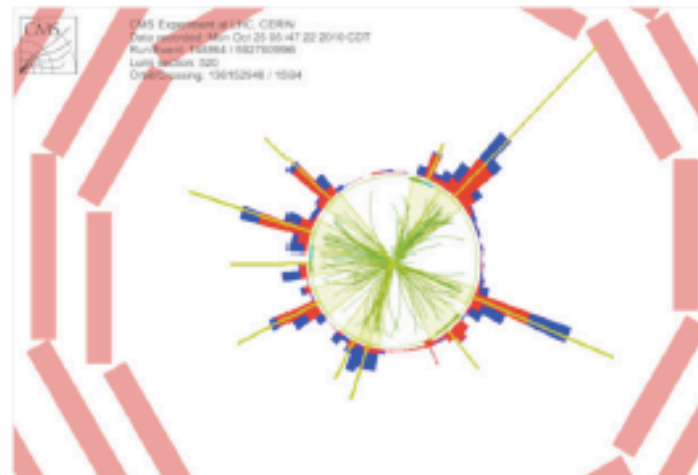


Multileg + PS



QCD predictions

??

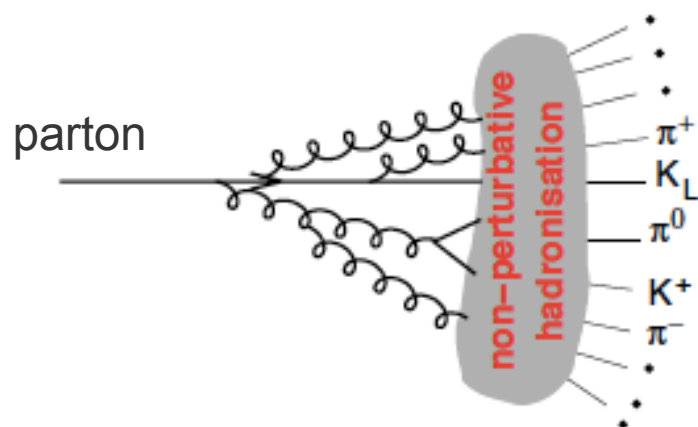


Real data

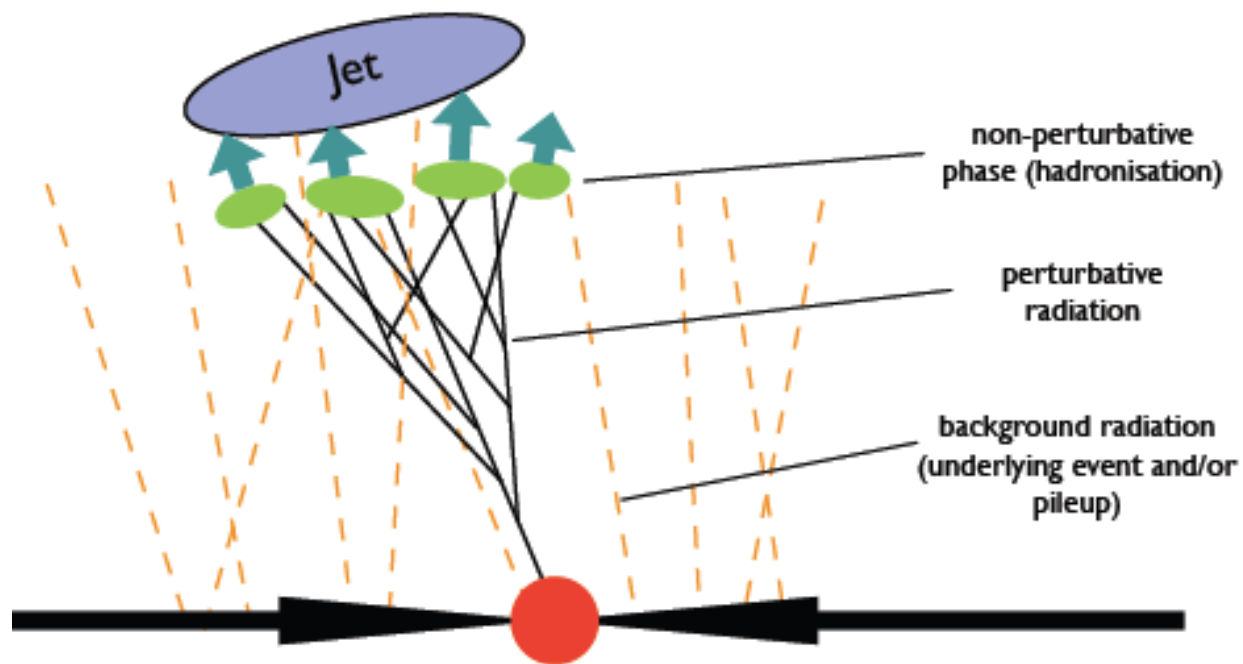
Jets

One purpose of a 'jet clustering' algorithm is to **reduce the complexity** of the final state, simplifying many hadrons to **simpler objects** that one can hope to **calculate**

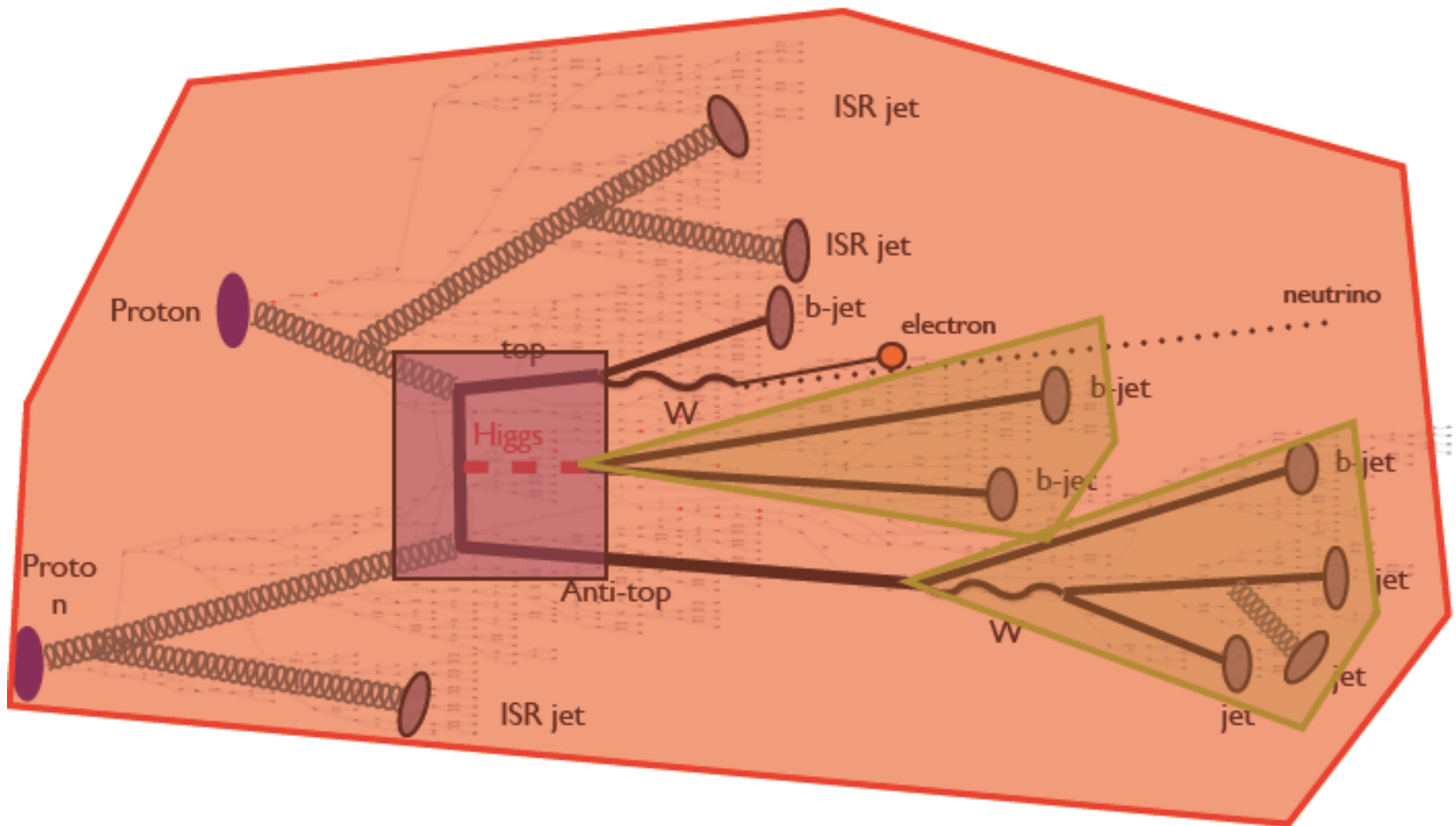
Zooming in to single objects



High energy partons unavoidably lead to collimated bunch of hadrons



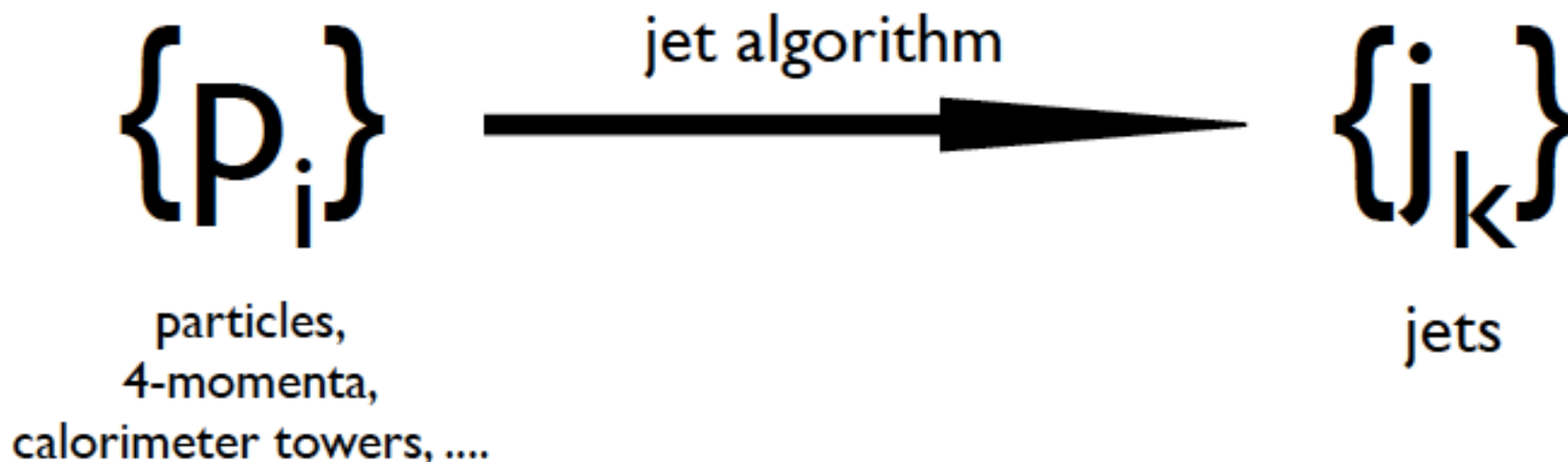
Our reductionist view of the same event



Jet algorithm



A **jet algorithm** maps the momenta of the final state particles into the momenta of a certain number of jets:

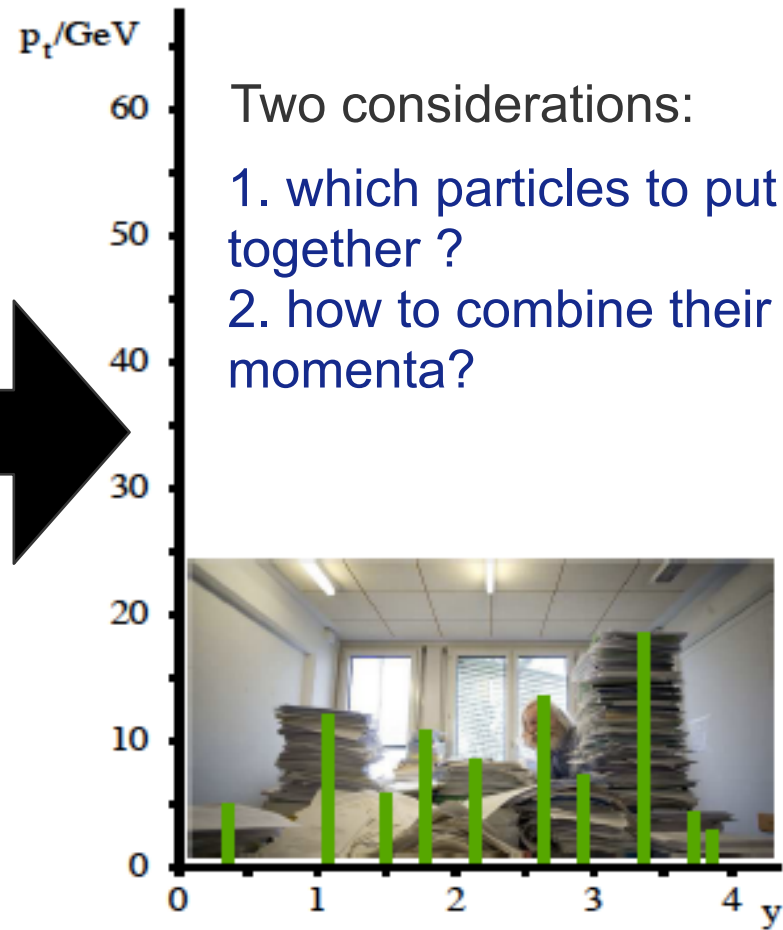
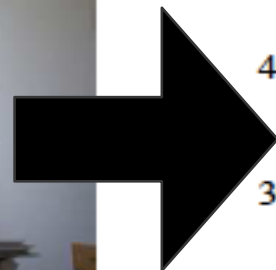


Most algorithms contain a resolution parameter, **R**, which controls the extension of the jet

A fun exercise



Tell me how many jets you see here? **0, 1, 2, ..., 10, or more?**



So, what is a jet then? Formal definition



in modern lingo, i.e., post circa 2007

A jet algorithm
+
its parameters (e.g. R)
+
a recombination scheme
=
a **Jet Definition**

← Most commonly
used: direct 4-vector
sums (E-scheme)

“Jet [definitions] are legal contracts between theorists and experimentalists”

-- MJ Tannenbaum



Two main classes of jet algorithms



Sequential recombination algorithms

Bottom-up approach: combine particles starting from **closest ones**

How? Choose a **distance measure**, iterate recombination until few objects left, call them jets

Examples: k_t , Cambridge-Aachen, anti- k_t , ...

- ▶ These algorithms are the most widely used nowadays, will discuss more about them in the following slides.

Cone algorithms

Top-down approach: find coarse regions of energy flow.

How? Find **stable cones** (i.e. their axis coincides with sum of momenta of particles in it)

Examples: JetClu, MidPoint, SIScone, ...

- ▶ These are mostly historic, **will NOT discuss them further**.



Sequential recombination algorithms

These are hierarchical clustering algorithms

Typically they work by calculating a **'distance'** between particles, and then recombine them pairwise according to a given order, until some condition is met (e.g. no particles are left, or the distance crosses a given threshold)

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta y^2 + \Delta \phi^2}{R^2} \quad d_{iB} = k_{ti}^{2p}$$

If $d_{ij} < d_{iB}$ then merge them together

p = 1 k_t algorithm

S. Catani, Y. Dokshitzer, M. Seymour and B. Webber, Nucl. Phys. B406 (1993) 187
S.D. Ellis and D.E. Soper, Phys. Rev. D48 (1993) 3160

p = 0 Cambridge/Aachen algorithm

Y. Dokshitzer, G. Leder, S. Moretti and B. Webber, JHEP 08 (1997) 001
M. Wobisch and T. Wengler, hep-ph/9907280

p = -1 anti- k_t algorithm

MC, G. Salam and G. Soyez, arXiv:0802.1189

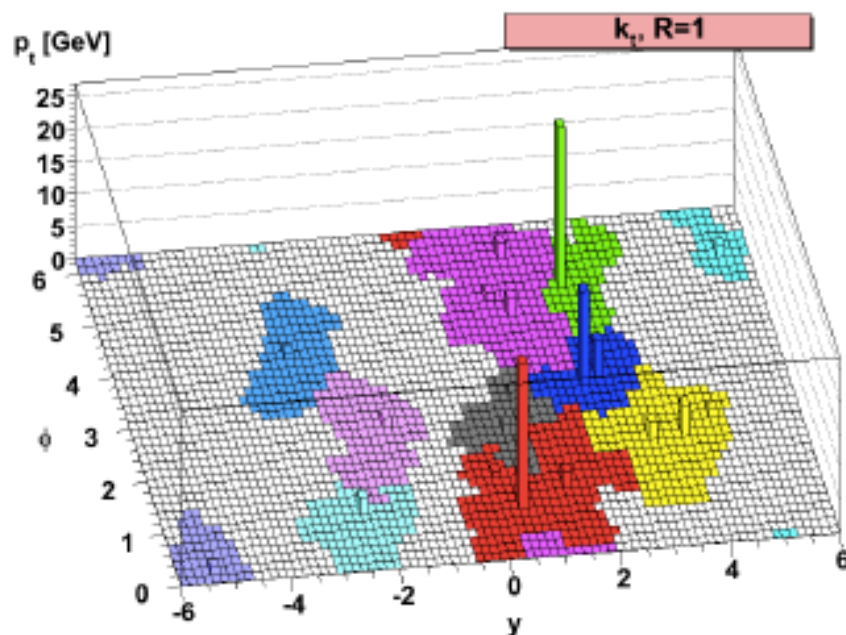
NB: in anti- k_t pairs with a **hard** particle will cluster first: if no other hard particles are close by, the algorithm will give **perfect cones**

Jet's "reach"



Algorithmically, a jet is simply a collection of particles

For a number of reasons, it is however useful to consider its **spatial extent**, i.e., given the position of its axis, up to where does it collect particles? What is its shape?



These details are important for a number of corrections of various origin: perturbative, hadronization, pileup, detector-related etc.

Note that the intuitive picture of a jet being a cone (or radius R) is **wrong**. This is what k_t jets look like.

From jet "reach" to jet areas



Not one, but three **definitions** of a jet's size:

Cacciari, Salam, Soyez
arXiv: 0802.1188

▶ **Passive area**

Place a single very soft particle (a '**ghost**') in the event, measure the extent of the region where it gets clustered within a given jet

Reach of jet for **pointlike** radiation



Active area

Fill the events with many very soft particles ('**ghosts**'), cluster them together with the hard ones, see how many get clustered within a given jet

Reach of jet for **diffuse** radiation

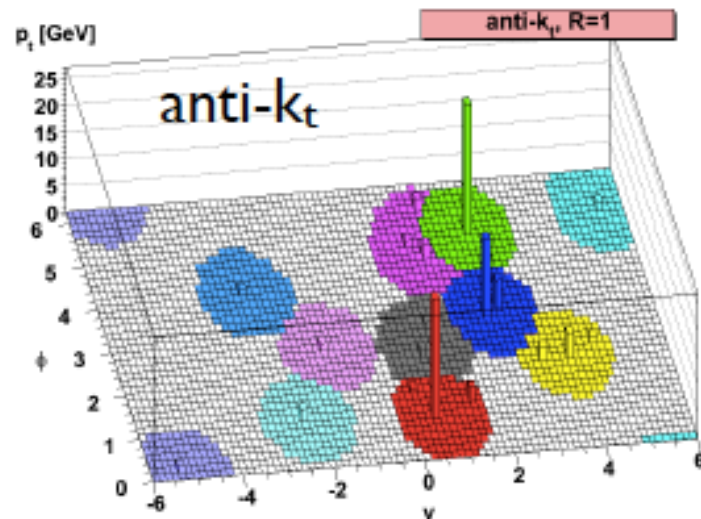
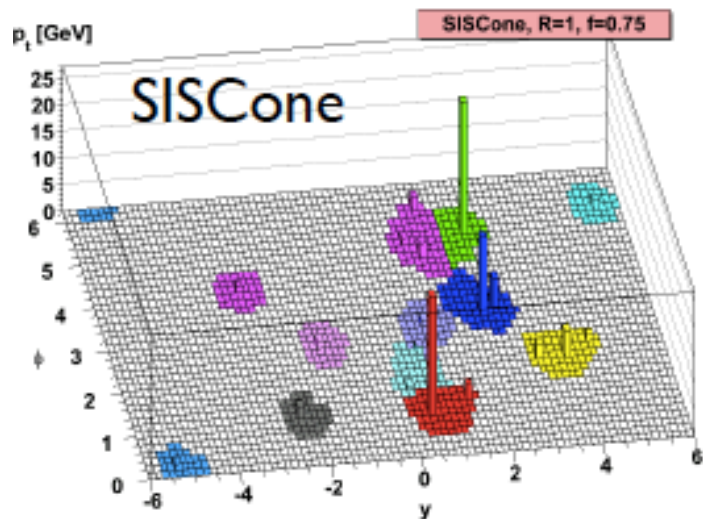
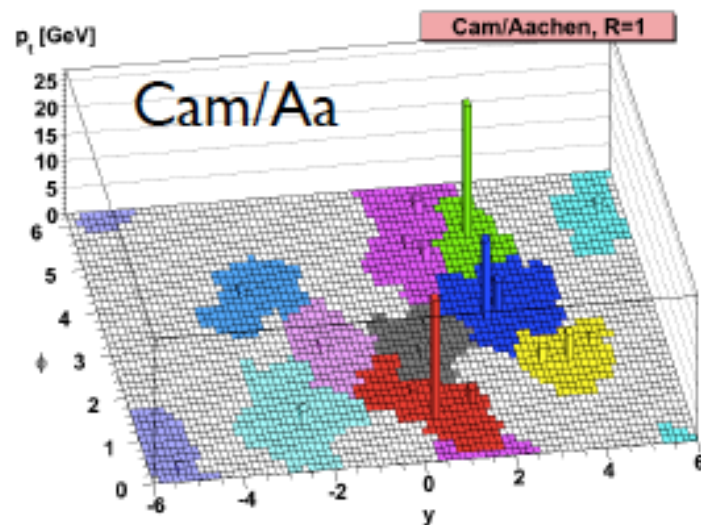
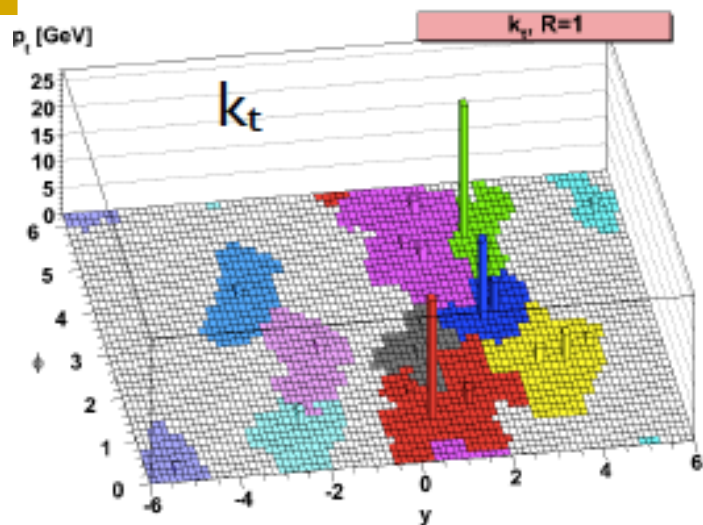
▶ **Voronoi area**

Sum of areas of intersections of Voronoi cells of jet constituents with circle of radius R centred on each constituent

Coincides with passive area for k_t algorithm

In the large number of particles limit all areas converge to the same value.

Jet areas for the most widely used algorithms

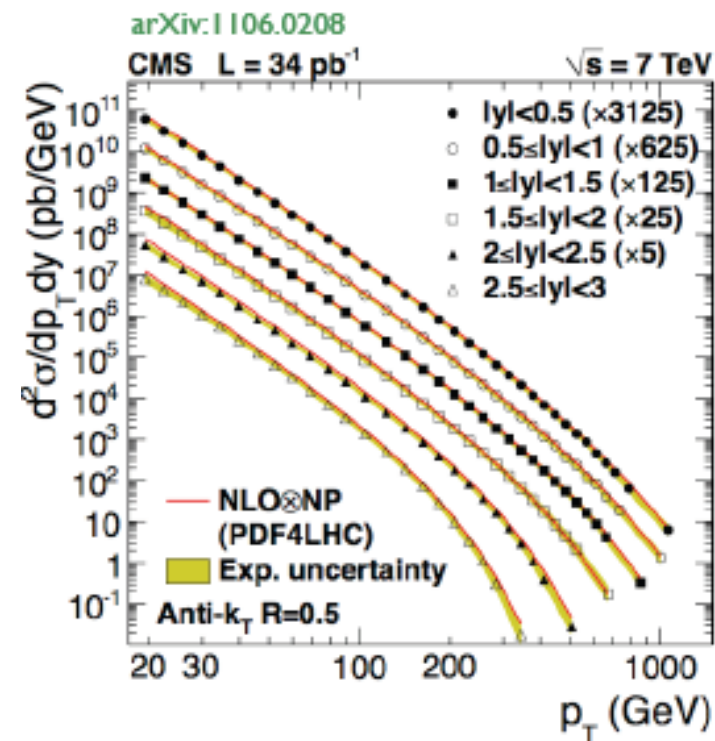
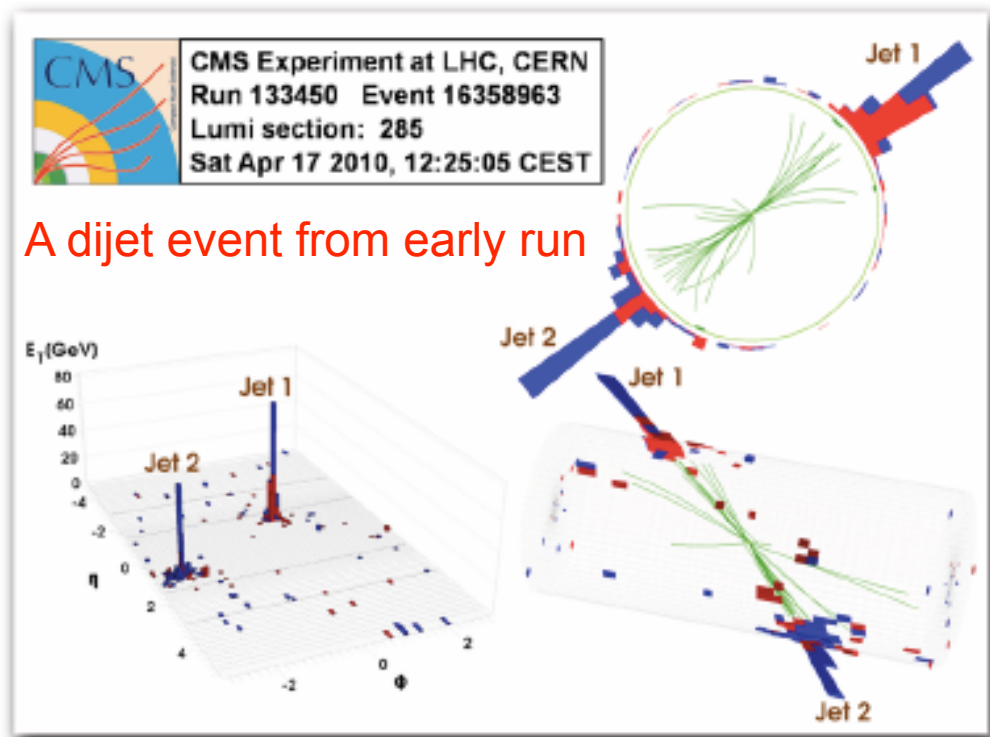


Jets in CMS



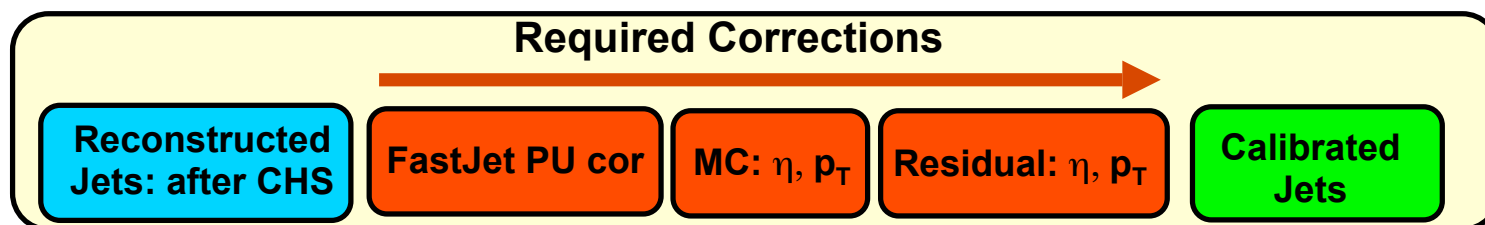
- Use charged hadron subtracted (CHS) Particle Flow jets
 - Remove all charged hadrons not from the primary vertex
- Neutrals are subtracted via jet area ρ correction where ρ is energy density computed with the KT6 algorithm
 - **Active area** used in all cases including groomed
- Non-linearities in η and p_T are corrected for using data & MC
 - AK5, AK7 jets have dedicated corrections while other large radii jets use AK7 corrections
- Jet finding & grooming algorithms implemented **via FastJet3**

Example of jets and jet observables in CMS



Jets extensively measured in hadronic collisions.
Very good agreement with pQCD predictions over 10 orders of magnitude.

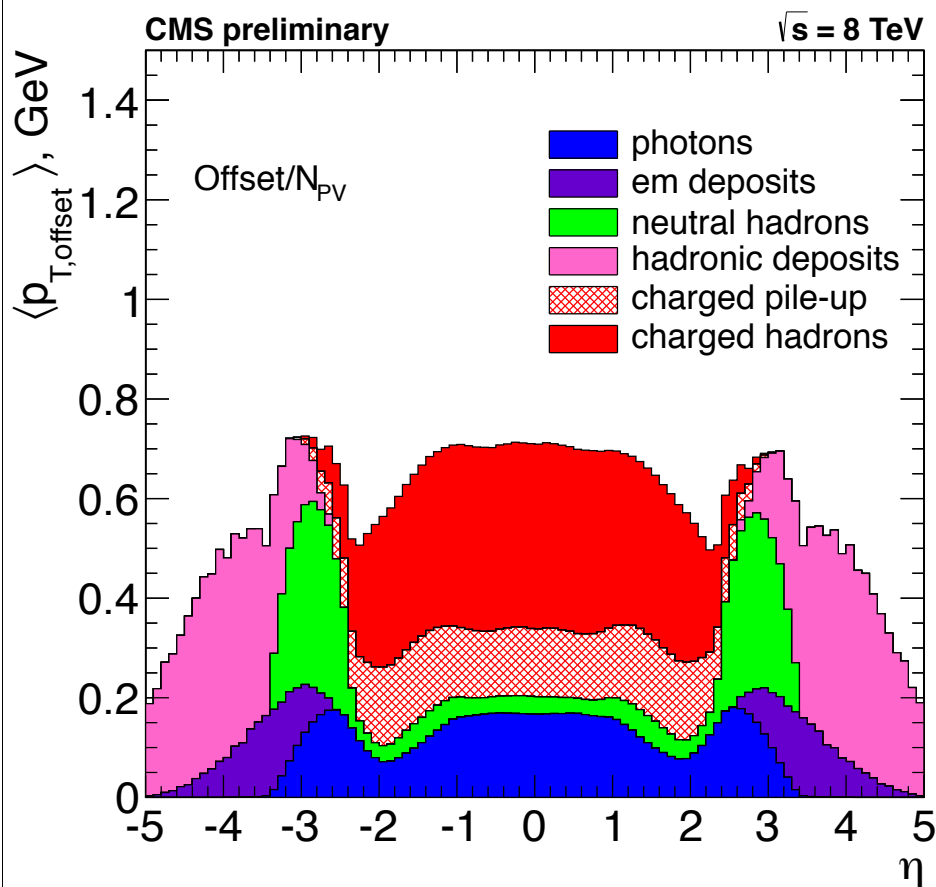
Jet energy calibration: overview



- ✓ Factorization facilitates the use of data-driven corrections
 - Breaking the correction into pieces that are naturally measured in collider data:
 - **Pileup subtraction**: pile-up density x area
 - **MC**: jet response vs. η , P_T using MC truth.
 - **Residual**: jet response vs. η , P_T using dijet balance and γ/Z +jet in data.

In CMS the most widely used jet is anti- k_T 0.5 (0.7 for QCD measurements). Jet substructure studies done with anti- k_T 0.5, 0.7, 0.8 and CA 0.8 with various grooming techniques.

Pileup contribution to jet energy



◆ Pileup (PU) measured with Zero Bias data

- Most charged hadrons can be associated to pileup vertices and removed

- Part that can be removed is labeled “charged hadrons”

- Part that remains as PU needs to be subtracted

PU density x effective area
(FastJet- ρ)

More details on jet calibration in CMSDAS short exercise

<https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideCMSDataAnalysisSchoolJetAnalysis>

Introduction to Jet Substructure & jet grooming



□ Jet substructure is currently a very popular tool to analyze jets that are produced from heavy objects such as top quarks, Higgs bosons, W/Z bosons, or any beyond-Standard-Model hadronically-decaying object.

□ An excellent overview of jet substructure techniques can be found here:

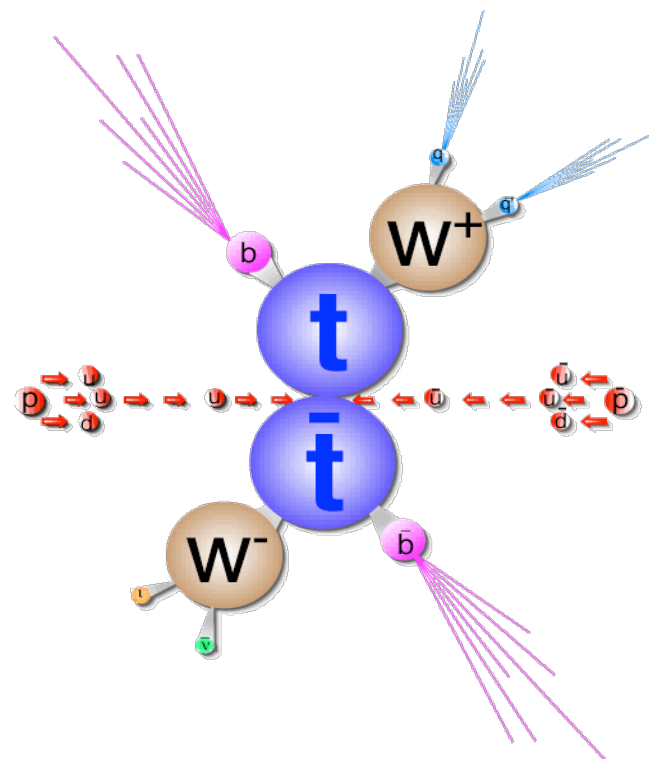
[Boosted objects: a probe of beyond the Standard Model physics](#) by Abdesselam et al.

Designed to find heavy objects within a jet, and in the process, removes soft and wide-angle clusters from the clustering sequence.

Jet substructure: motivation

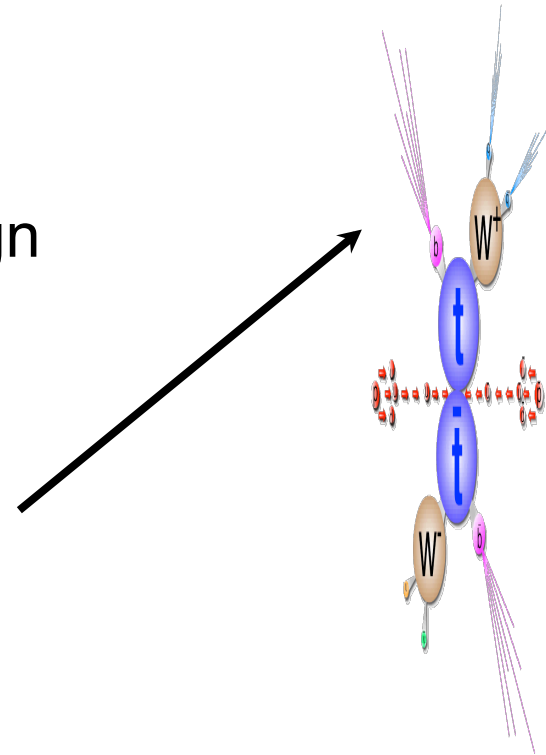


- Problem! Traditional techniques start to lose sensitivity (in part) due to jet merging at higher masses!
- Cannot rely on methods to assign partons to jets anymore
- Have to consider cases where partons merge into a single jet



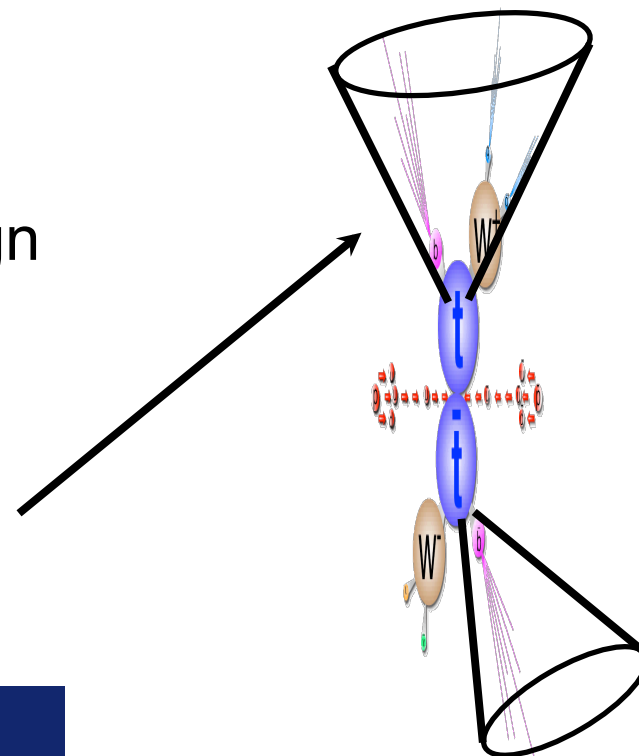
Jet substructure: motivation

- Problem! Traditional techniques start to lose sensitivity (in part) due to jet merging at higher masses!
- Cannot rely on methods to assign partons to jets anymore
- Have to consider cases where partons merge into a single jet



Jet substructure: motivation

- Problem! Traditional techniques start to lose sensitivity (in part) due to jet merging at higher masses!
- Cannot rely on methods to assign partons to jets anymore
- Have to consider cases where partons merge into a single jet



To understand how to break up a jet, first we need to understand what it is!

Jet grooming techniques I

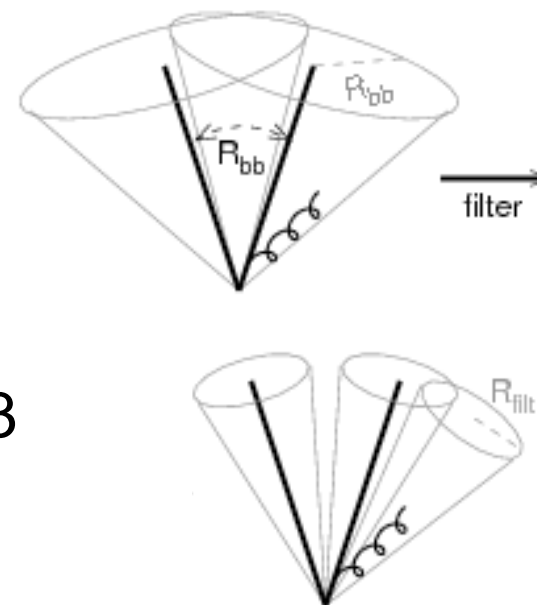


◆ Study jet mass properties with three grooming techniques

- Filtering: [arxiv: 0802.2470](https://arxiv.org/abs/0802.2470) Butterworth, Davison, Rubin, Salam
- Trimming: [arxiv: 0912.1342](https://arxiv.org/abs/0912.1342) D. Krohn, Jesse Thaler, LianTao Wang
- Pruning: [arxiv: 0903.5081](https://arxiv.org/abs/0903.5081) Steve Ellis, Chris Vermilion, Jon Walsh
- This round of analysis uses **default parameters** from each of the references.

◆ Filtering

- reclustering jet constituents with smaller radius, R_{filt} , **keeping n_{filt} hardest sub-jets**
- default parameters: $R_{\text{filt}} = 0.3$, $n_{\text{filt}} = 3$
- sub-jet clustering algorithm: Cambridge-Aachen

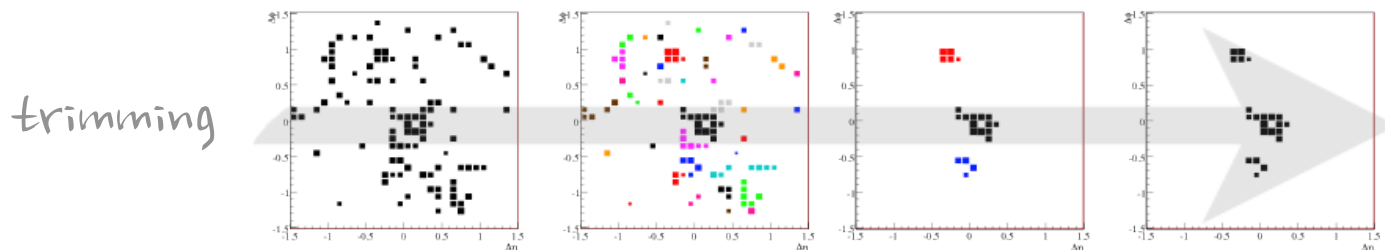


Jet grooming techniques II



◆ Trimming

- reclustering with smaller radius, R_{filt} , **keeping sub-jets with a fraction, $pT_{\text{frac,min}}$, of original jet p_T**
- default parameters: $R_{\text{filt}} = 0.2$, $pT_{\text{frac,min}} = 0.03$
- sub-jet clustering algorithm: kT



◆ Pruning

- reclustering with sequential recombination algorithm, **veto soft and large-angle recombinations** between pseudojets i and j
 - veto: $\Delta R_{ij} > r_{\text{cut}} \times 2m/p_T$; $z = \min(p_{T_i}, p_{T_j})/p_{T_{i+j}} < z_{\text{cut}}$
- default parameters: $z_{\text{cut}} = 0.1$, default $r_{\text{cut}} = 0.5$
- sub-jet clustering algorithm: Cambridge-Aachen

Jet mass and substructure



arXiv: 0802.2470

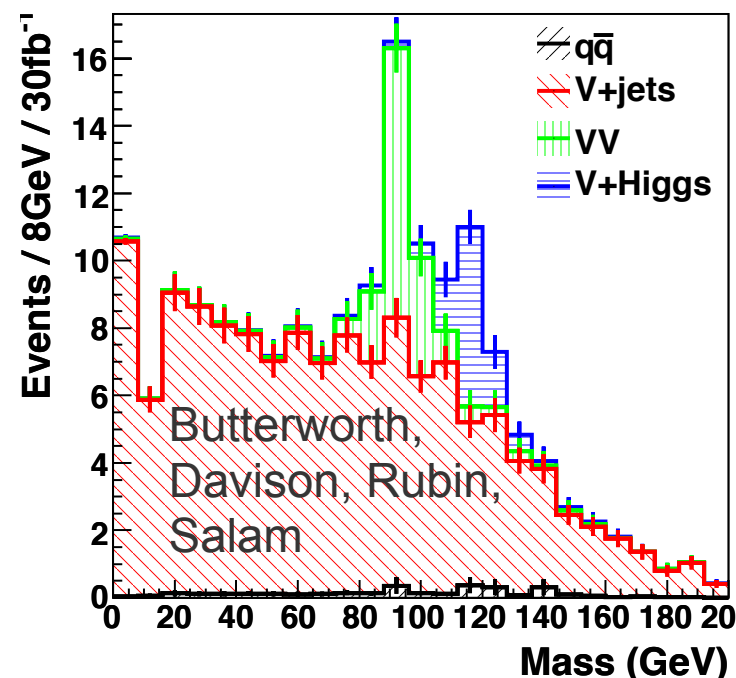
- Jet substructure can be used to improve sensitivity of hadronic decays of **boosted heavy particles such as Higgs, W/Z, and top**

- Requires understanding of QCD radiation inside jet, structure of constituent particles

- In CMS, have analyzed jet mass properties

 - **inclusive & systematic comparison of jet grooming algorithms**

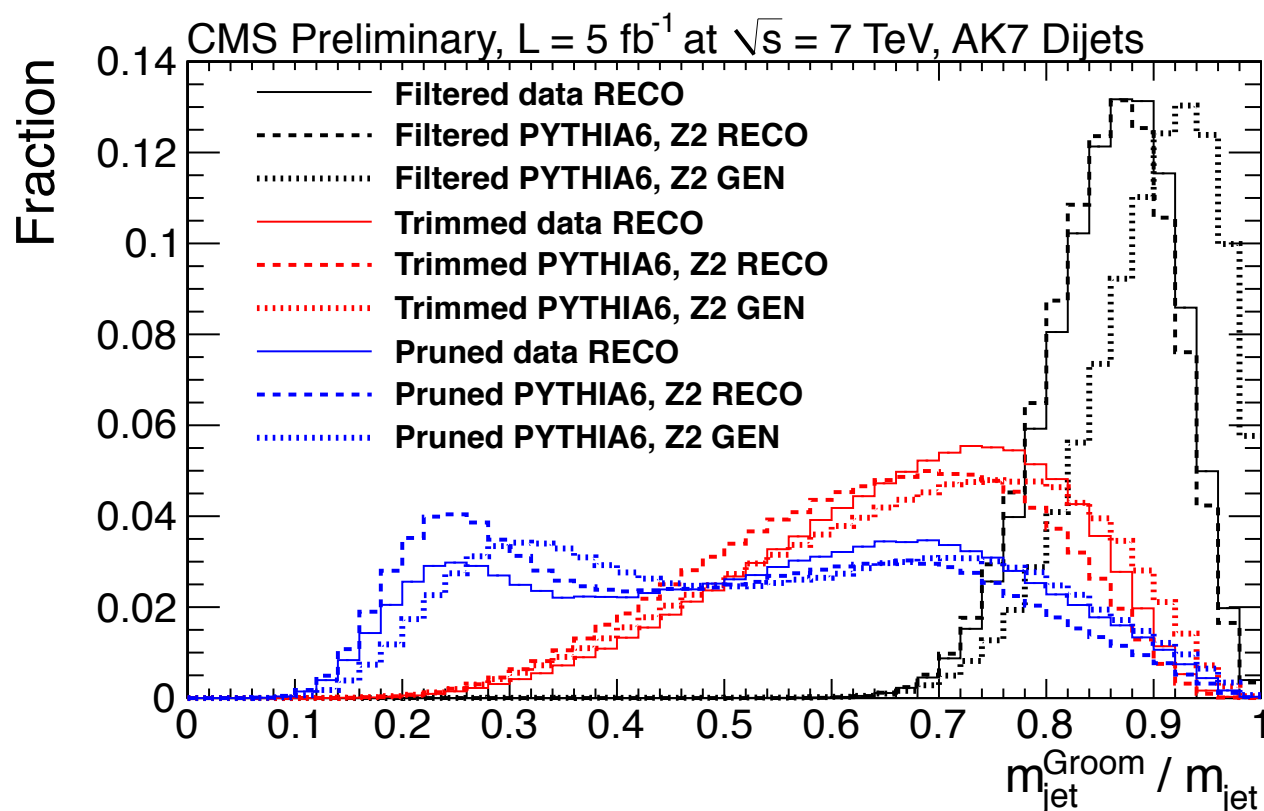
- Use **di-jet and V+jets events** as complimentary probes of gluon- and quark-enriched environments, respectively



Jet mass for groomed jets



Comparison of grooming algorithms at particle level (GEN), reconstructed simulation (RECO) level, and in data



Pruning is the most aggressive, filtering is the least aggressive

[arXiv:1303.4811](https://arxiv.org/abs/1303.4811)

Understanding the “background”



How the jet mass distribution looks like for garden-variety boosted jets (either in multi-jet or W/Z+jets data)

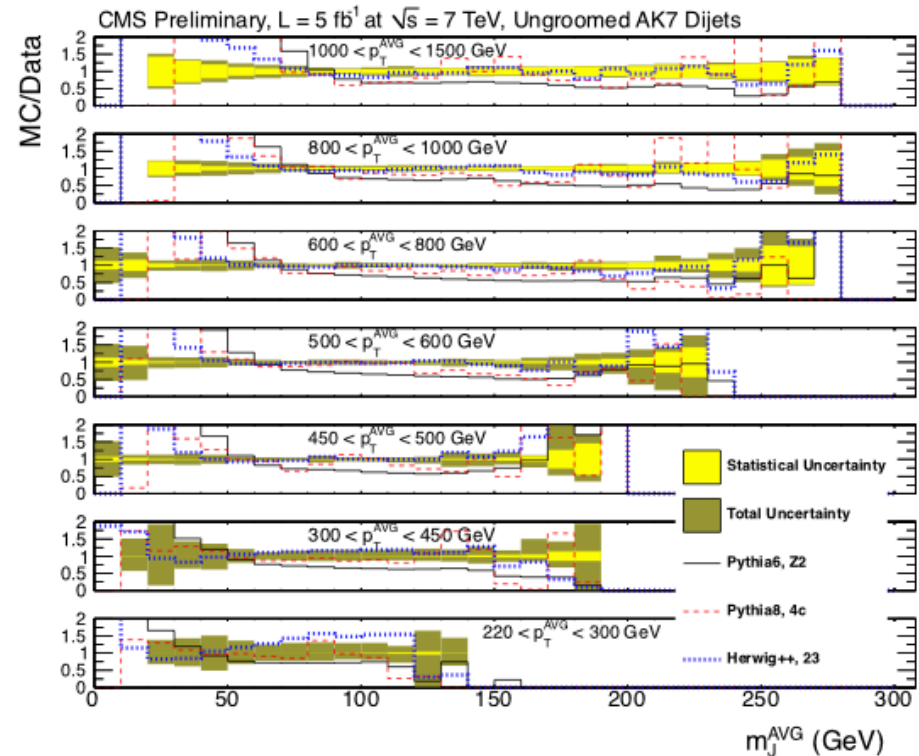
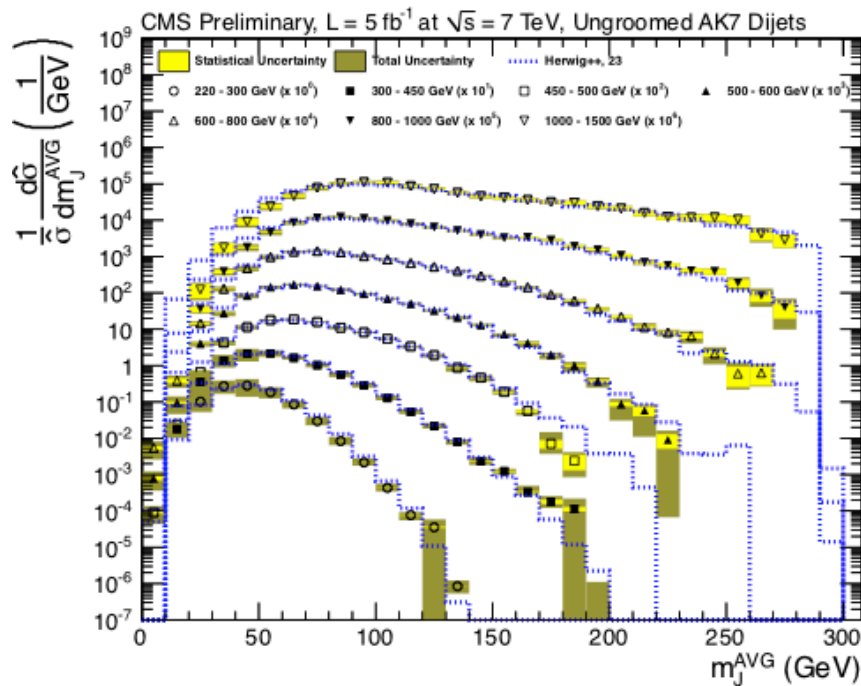
- Perform an inclusive comparative study of jet mass for available jet grooming methods
- Make a comparison of data to simulation as a validation of parton showering models

Plots are presented in increasing aggressiveness of grooming [Ungroomed, Filtered, Trimmed, Pruned] and best digested by flipping through.



How the m_j looks like for QCD jets: Ungroomed

MC to data agreement is reasonable using both Pythia Z2 and Herwig++, though **Herwig++ seems to have better agreement.**



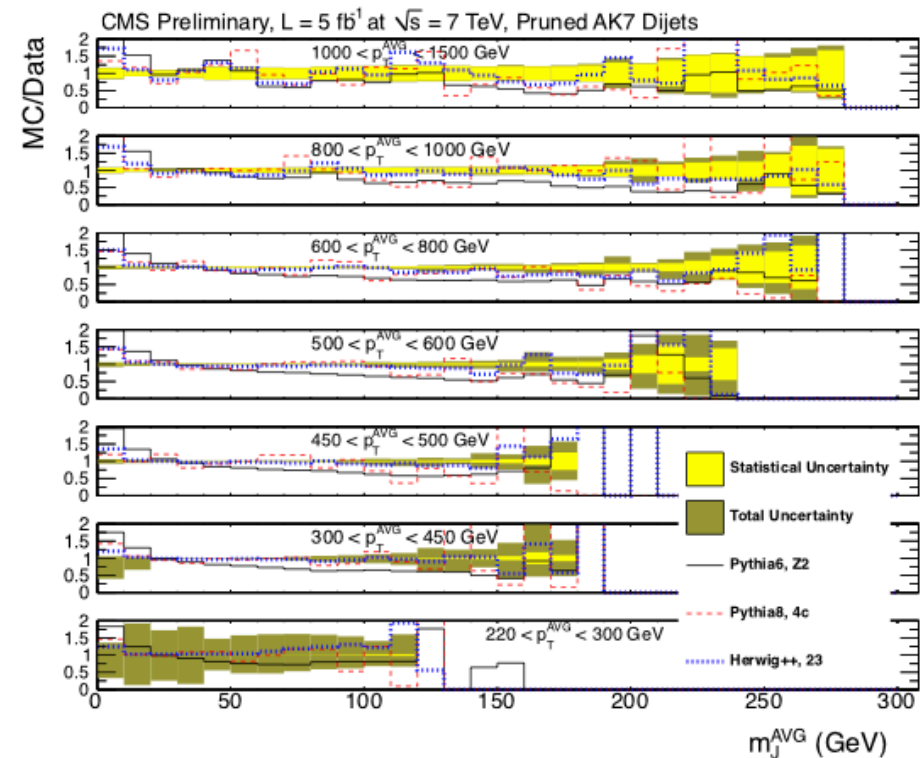
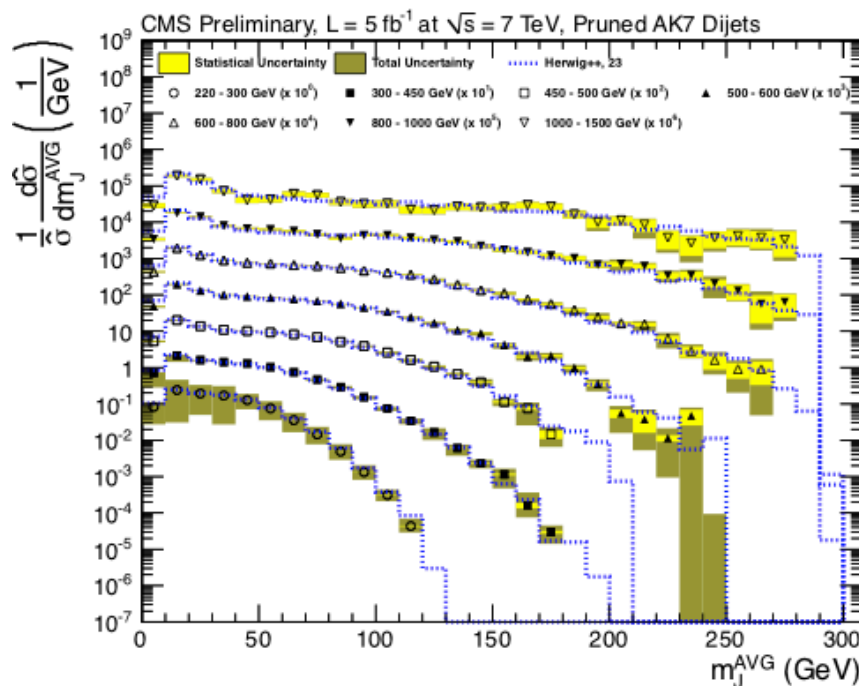
[arXiv:1303.4811](https://arxiv.org/abs/1303.4811)

Pruned

arXiv:1303.4811



As the grooming algorithm becomes more aggressive, Herwig++ agreement is better with data.



Turnover region moves to the left with harder grooming while tails remain similar in all cases – important feature in new physics searches.

Jet substructure variables

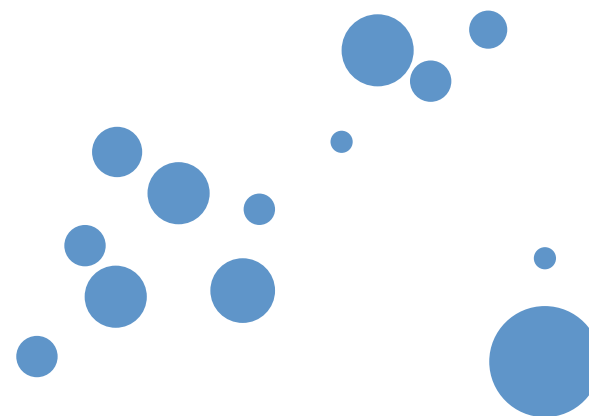
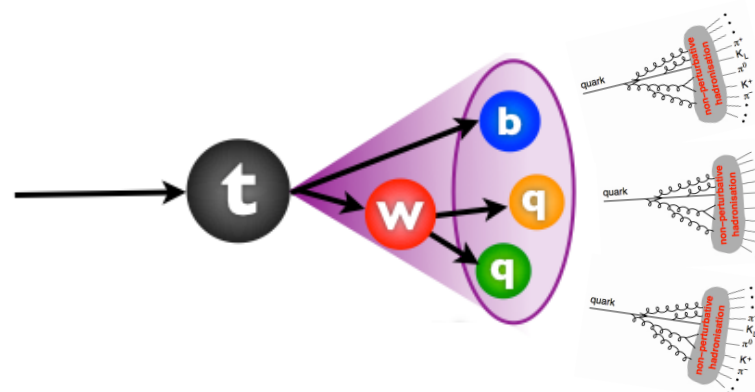


- There are several jet substructure variables that can be used to discriminate between hadronic decays of **boosted heavy particles (Higgs, W/Z, top,...)** and the QCD jets
 - Already discussed some of them, e.g., jet mass before and after grooming, jet p_T , jet size, etc.
 - Others (# subjets, subjet mass/kinematics, ...) introduced in the following slides in the context of W & top tagging.
- We will exclusively focus on these in our second session today.

Example of boosted signal: Top jets

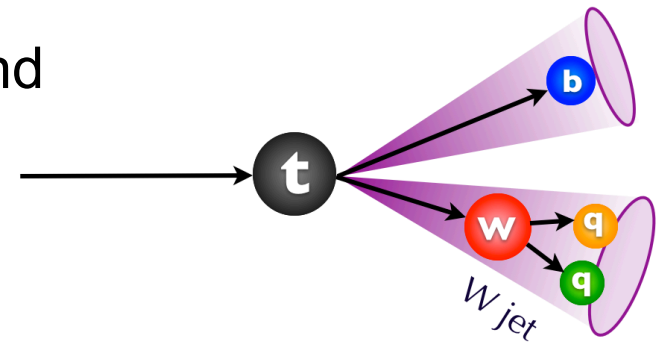


- If a top quark or other heavy object is produced with enough energy its decay products can be reconstructed within one jet
- Substructure can be used to identify these jets
 - Top jets produce at least 3 subjects
 - Two of the subjects should have the W mass and three should have the top mass



Example of boosted signal: W jets

- A partially merged boosted top results in a W jet and a b jet
- W jets can be identified using jet and subjet properties:
 - Jet mass = W mass
 - Two subjets
 - Subjet mass \ll jet mass
 - Both subjets should have similar momentum



Summary



- ☑ The Jet substructure is coming up as a very popular tool to analyze jets that are produced from heavy objects
- ☑ First systematic study of jet grooming techniques on jet substructure performed at CMS
 - an important benchmark in understanding various algorithms for searches for new physics
 - With more data, probing deeper into boosted regime
- ☑ This hands-on tutorial should get you started
 - Please find us during the breaks or email us in case you come up with a question later
- ☑ Lot of useful material discussed in jet-substructure workshop

<https://indico.cern.ch/conferenceDisplay.py?confId=242845>

BACKUP SLIDES



Jet clustering in FastJet

```
/// define a jet definition
JetDefinition jet_def(JetAlgorithm jet_algorithm,
                      double R,
                      RecombinationScheme rec_sch = E_scheme);
```

Can be one of the algorithms described in the next slides

Can be particle-flow particles or gen-particles or tracks or calo-towers

```
/// create a ClusterSequence, extract the jets
ClusterSequence cs(input_particles, jet_def);
vector<PseudoJet> jets = sorted_by_pt(cs.inclusive(jets));
...
// pt of hardest jet
double pt_hardest = jets[0].pt();
...
// constituents of hardest jet
vector<PseudoJet> constits = jets[0].constituents();
```



How to compute jet areas in FastJet?

```
/// specify where to place the ghosts, their area, how many times
/// to repeat the clustering
double rapmax = 2.5;
int nrepeat = 1;
double ghost_area = 0.01;
GhostedAreaSpec gas(rapmax, nrepeat, ghost_area);

/// use this configuration to define the area
/// A sensible default for gas is provided
AreaDefinition area_def(active_area, gas);

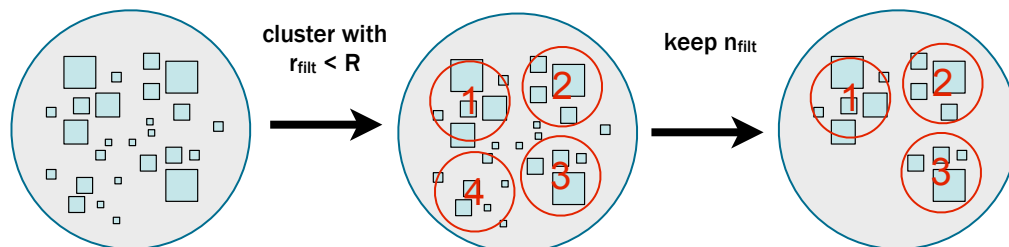
/// construct cluster sequence with areas
ClusterSequenceArea(input_particles, jet_def, area_def);
....
jets[0].area()
```

Active areas for some of the most popular jet algorithms are shown on the next slide

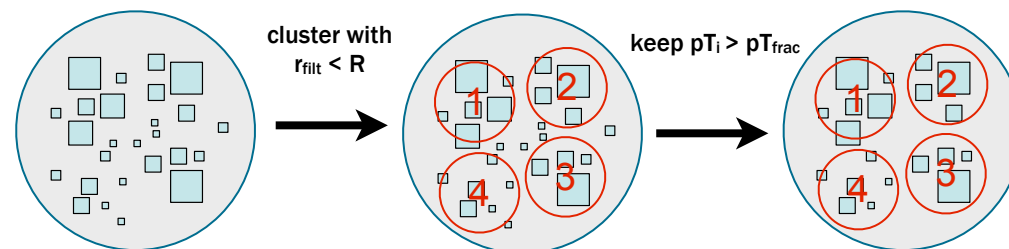
Comparison of the three grooming techniques



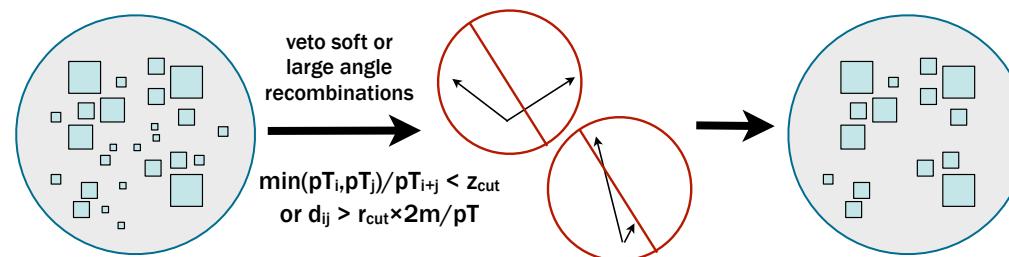
filtering



trimming



pruning

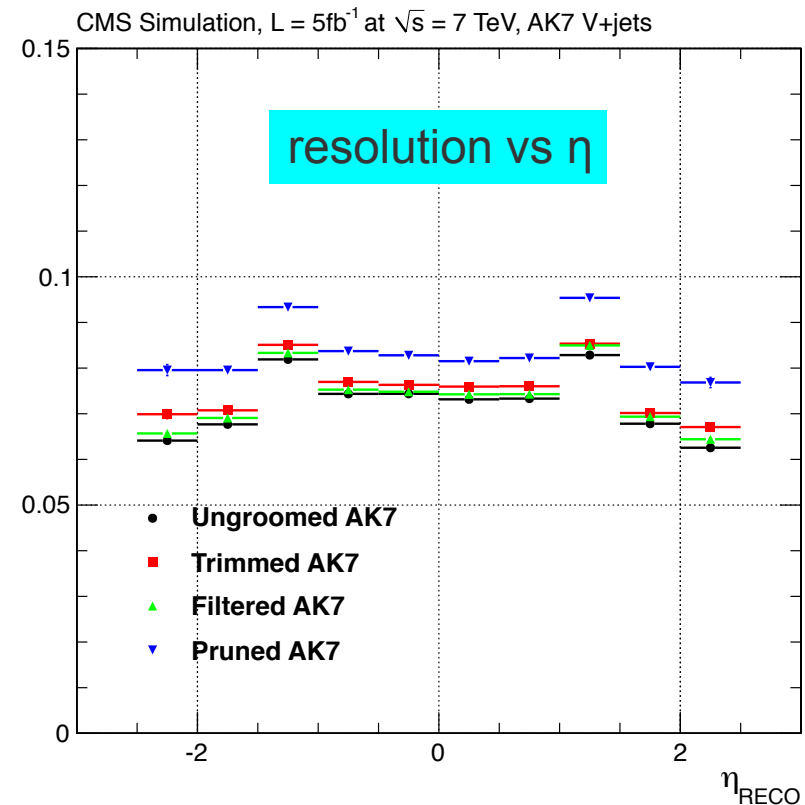
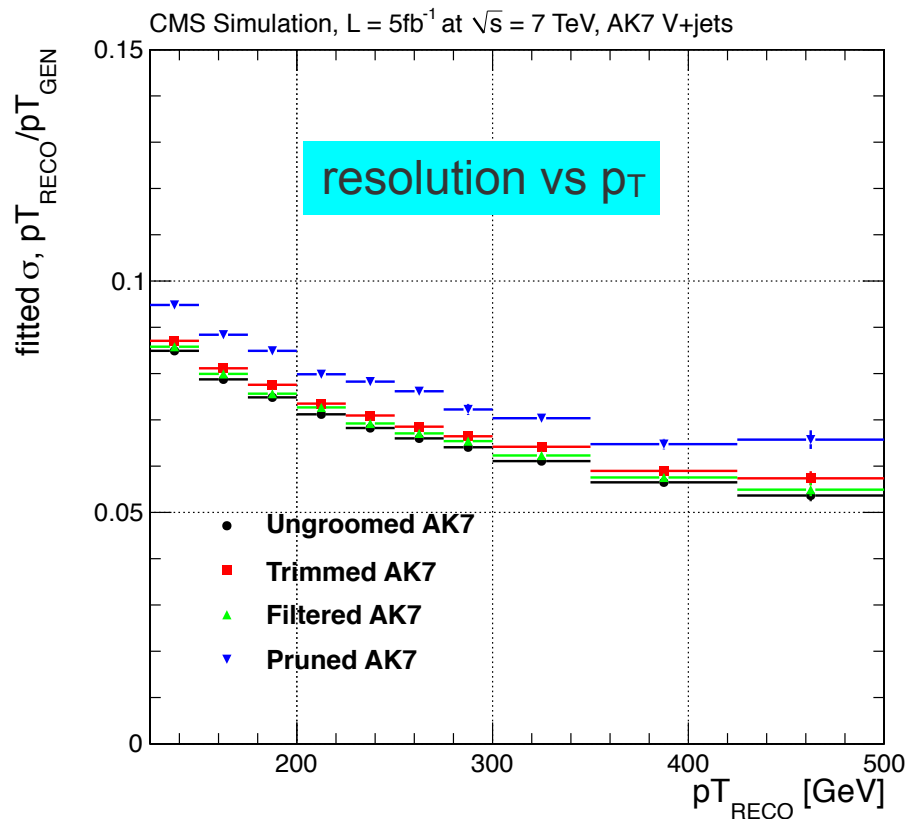




Jet p_T resolution for groomed jets

Jet p_T resolution for various grooming algorithms also shows good agreement to within a few percent.

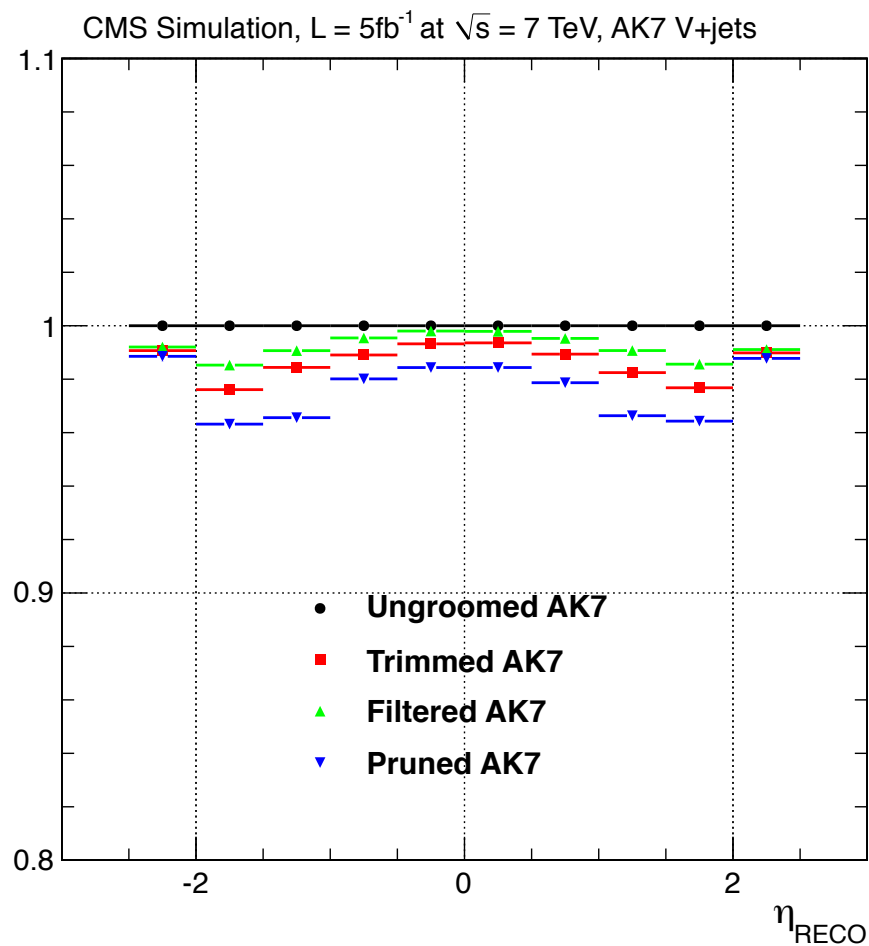
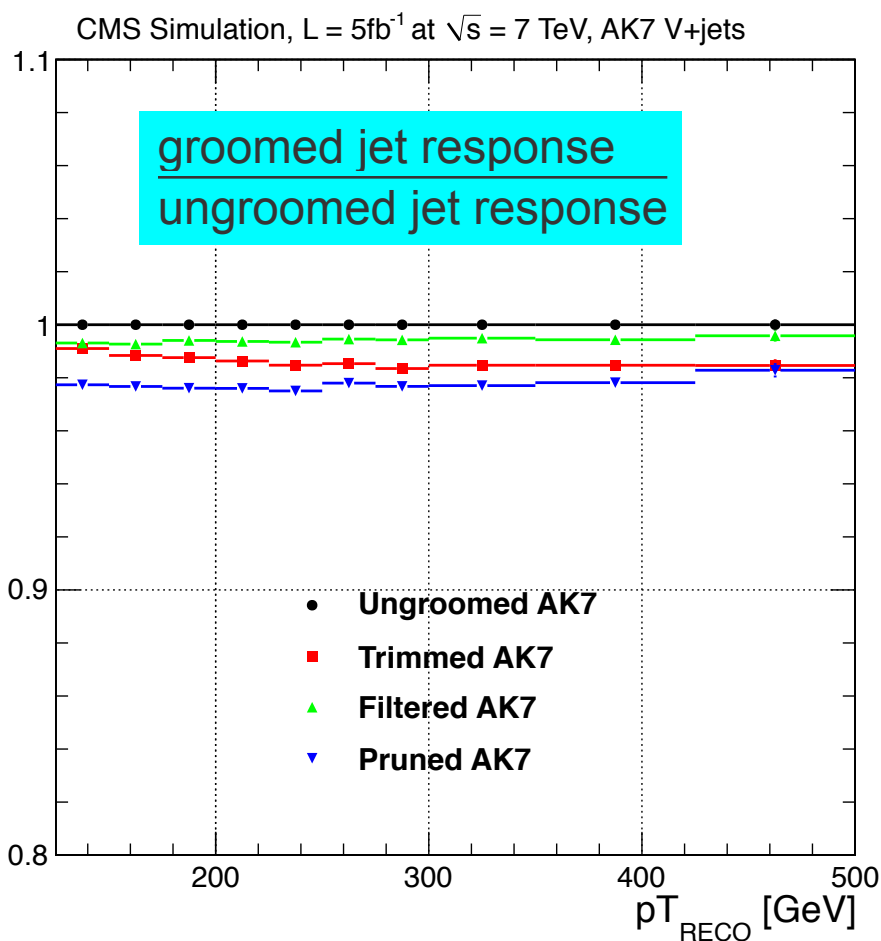
• **Groomed jet p_T resolution degraded slightly.**



Jet p_T response for groomed jets



Groomed jet response within a few % of ungroomed case.



Performance versus pileup

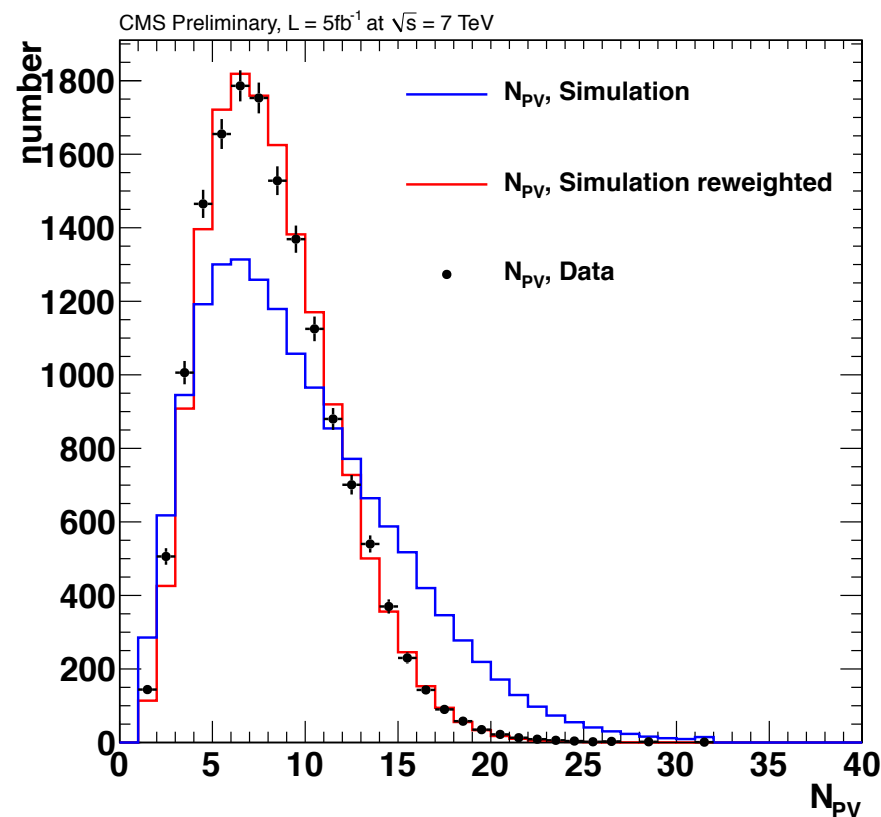


- It is of particular interest to understand the sensitivity of large size jets in presence of PU

-Grooming techniques may serve to mitigate pileup sensitivity by effectively reducing the jet area

- Understand performance of mean jet mass as a function of number of primary vertices

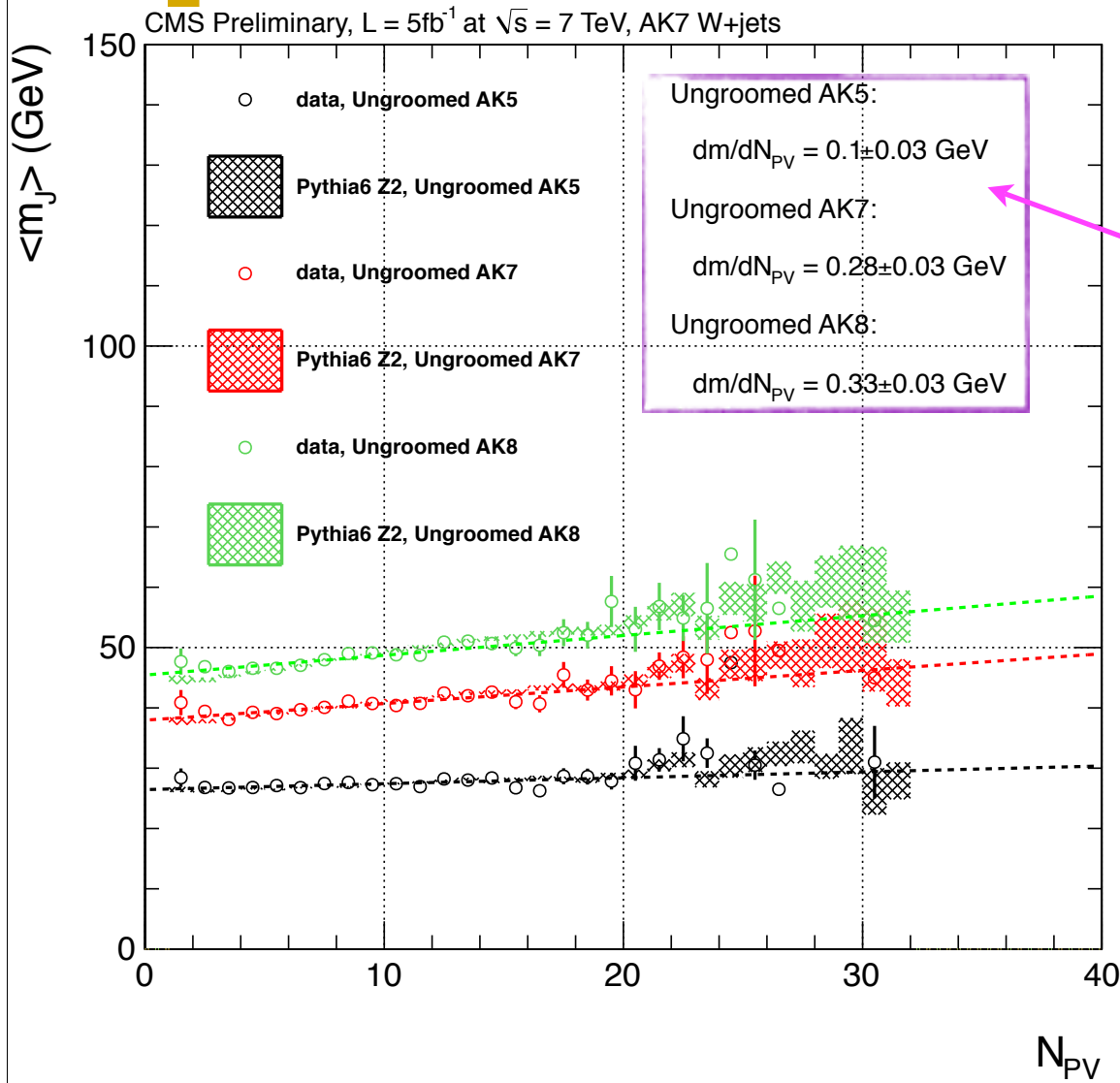
Pileup profile in 7 TeV data



Size matters here: larger jet size \Rightarrow larger effect



arXiv:1303.4811



◆ Ungroomed jet mass is very sensitive to PU

• $\langle m_J \rangle$ increases linearly as a function of the number of primary vertices

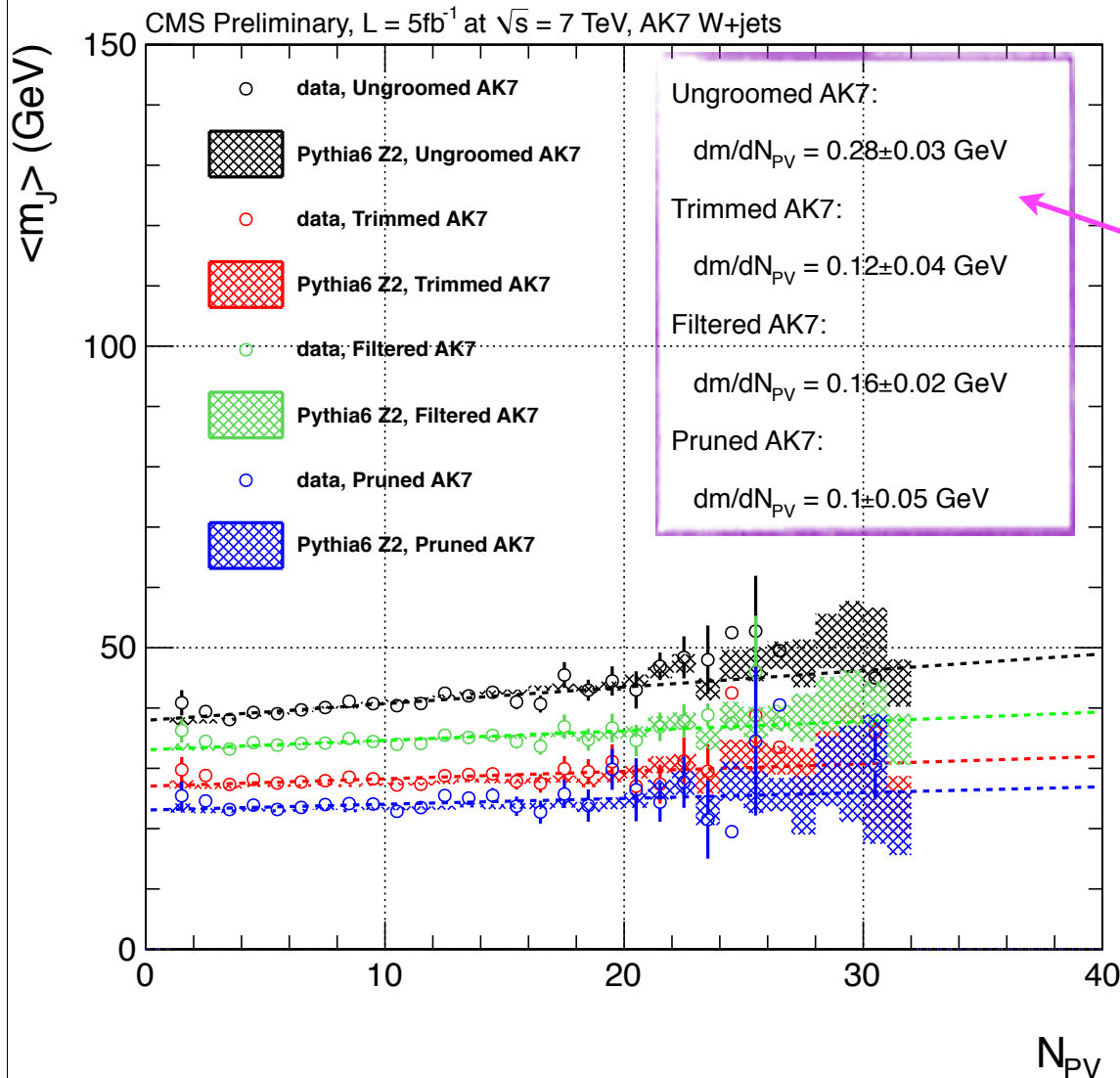
◆ Effect becomes more pronounced as the jet size increases

• AK8 shows much worse effect than AK5

Grooming helps to ameliorate the situation



arXiv:1303.4811



◆ Groomed jets are less sensitive to PU

• $\langle m_J \rangle$ vs NPV slope becomes flatter

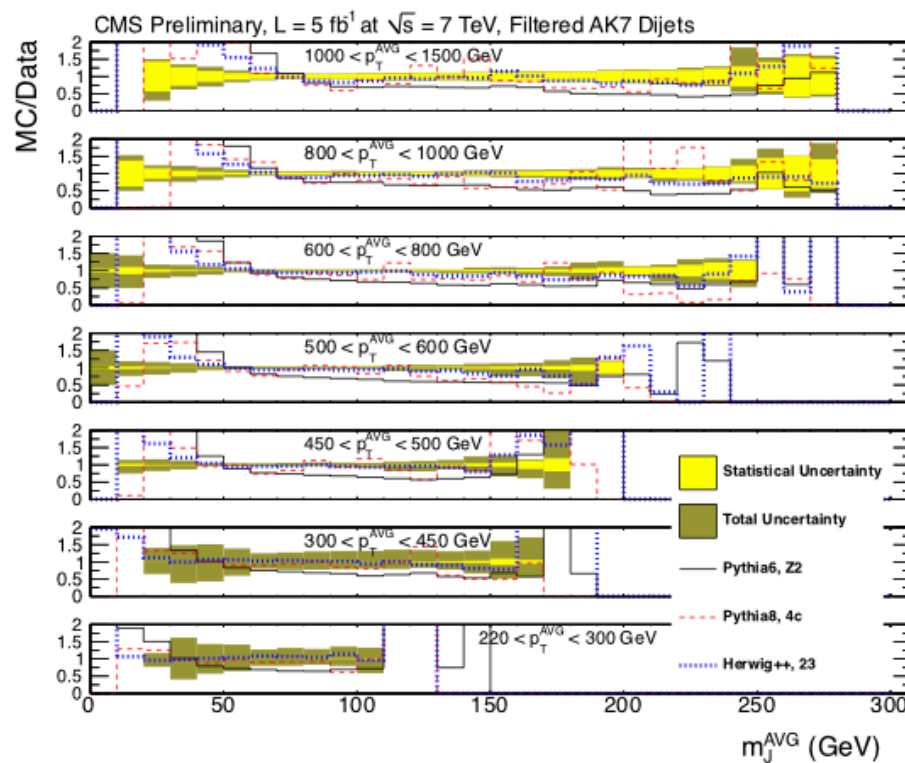
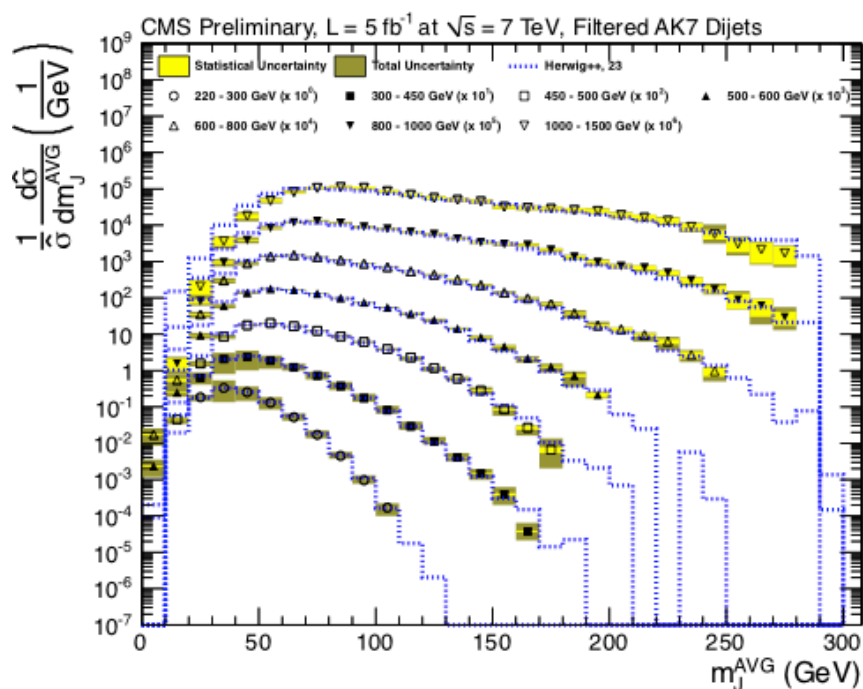
◆ Observe the expected behavior that $\langle m_J \rangle$ typically scales as R^3

$s_{0.7}/s_{0.5}$	$= 2.7 \pm 0.9$	$((0.7/0.5)^3 = 2.74)$,
$s_{0.8}/s_{0.5}$	$= 3.3 \pm 1.0$	$((0.8/0.5)^3 = 4.10)$,
$s_{0.8}/s_{0.7}$	$= 1.2 \pm 0.2$	$((0.8/0.7)^3 = 1.49)$

Filtered



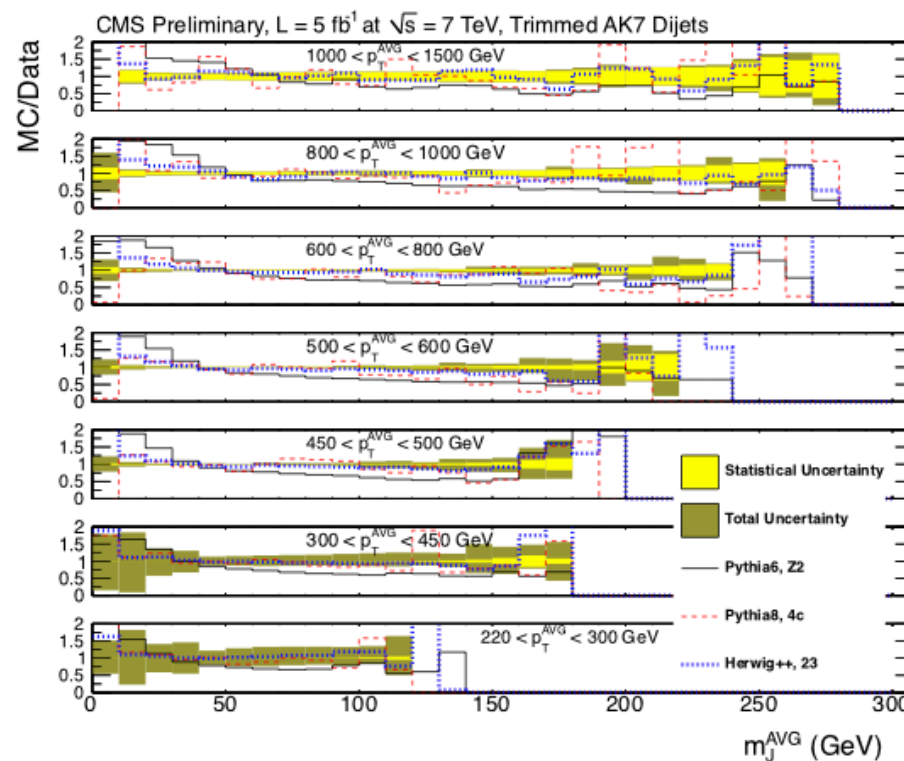
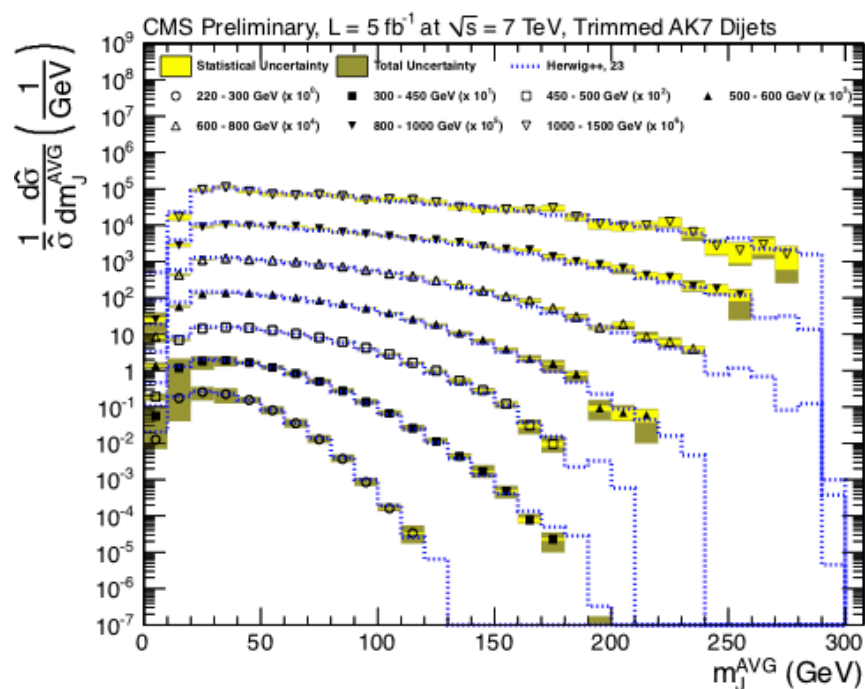
arXiv:1303.4811



Trimmed



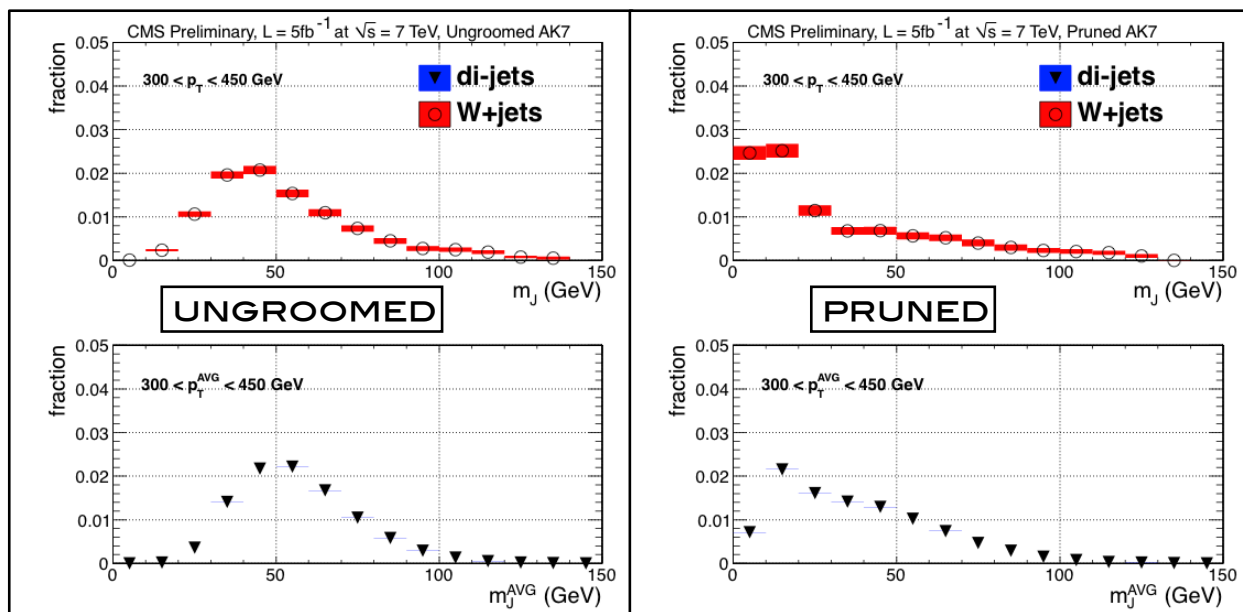
arXiv:1303.4811



Quark vs gluon differences



For visualization, present di-jet and V+jet distributions with one above the other.



p_T bin:
300–450 GeV

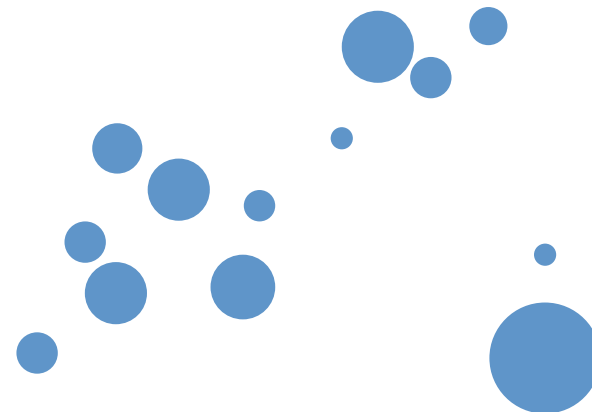
N.B. not the exact same
observables
 m_J vs $\langle m_J \rangle$ in bins of p_T ,
 $\langle p_T \rangle$

Shows the difference in jet mass for di-jet and V+jet processes where the former (latter) radiates more (less) due to larger composition of gluon (quark)-initiated jets



CMS top tagging algorithm

- Based on Kaplan et al. (arXiv:0806.0848)
- CMS PAS JME-09-001 and CMS AN-2010/080
- Cluster a jet using a sequential recombination algorithm and then decluster in order to find substructure
- Subjets must satisfy two requirements
 1. Momentum fraction criterion - Subjets should carry a significant fraction of the jets momentum ($p_T^{\text{subjet}} > 0.05 \times p_T^{\text{hard jet}}$)
 2. Adjacency criterion - subjets should be separated by some minimum distance ($\Delta R(C_A, C_B) > 0.4 - 0.0004 \times p_T$)
- Iterative process - throw out subjets that fail momentum fraction cut and try to decluster again
 - Some jet grooming is built in



Top tagging variables



Three Variables are used to tag tops:

1. Jet mass - top jets have mass within a top mass window
2. Number of subjets - top jets have 3 or 4 subjets
3. Minimum Pairwise Mass - Pairwise mass of two of the subjets should reconstruct the W mass

Define MinMass = minimum pairwise mass of subjets

$$m_{\min} = \min[m_{12}, m_{13}, m_{23}]$$

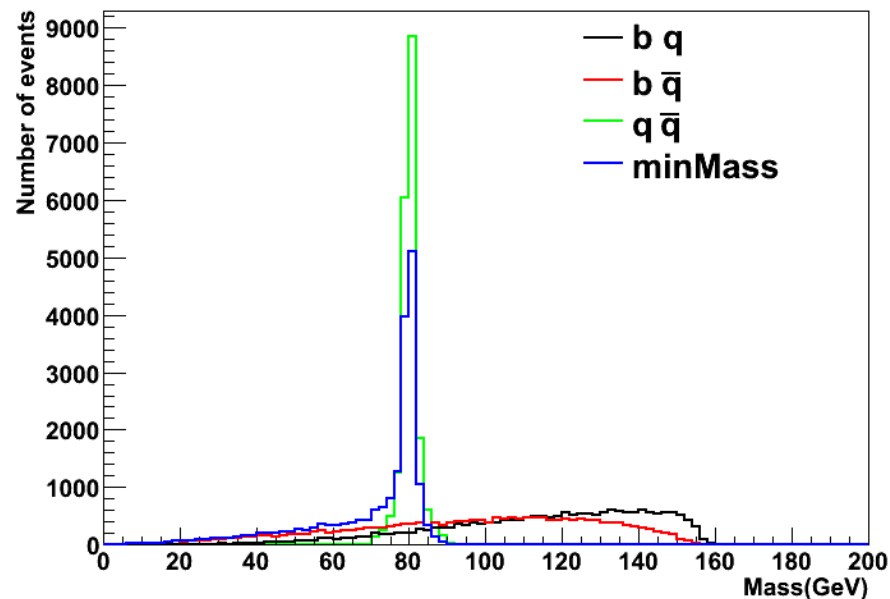
Define Top Tag Cuts

$$140 < m_{\text{jet}} < 250$$

$$N_{\text{subjets}} \geq 3$$

$$m_{\min} > 50$$

Generated top decay products: pairwise mass





Example: W-tagging with pruning

- Ellis et al. (arXiv:0903.5081)
- Improves mass resolution by removing soft, large angle particles from the jet
- Recluster each jet, requiring that each recombination satisfy the following:

$$\frac{\min(p_{T1}, p_{T2})}{p_{Tp}} > 0.1$$

$$\Delta R_{12} < 0.5 \times \frac{m_{jet}}{p_T}$$

- For W tagging, require:

Jet mass in 60-100 GeV/c²

$$\text{Mass drop } (\mu) < 0.4 \quad \mu = \frac{m_{j1}}{m_j}$$

Conway, Erbacher, Dolen, Hu, Maksimovic, Rappoccio, Sierra-Vasquez

