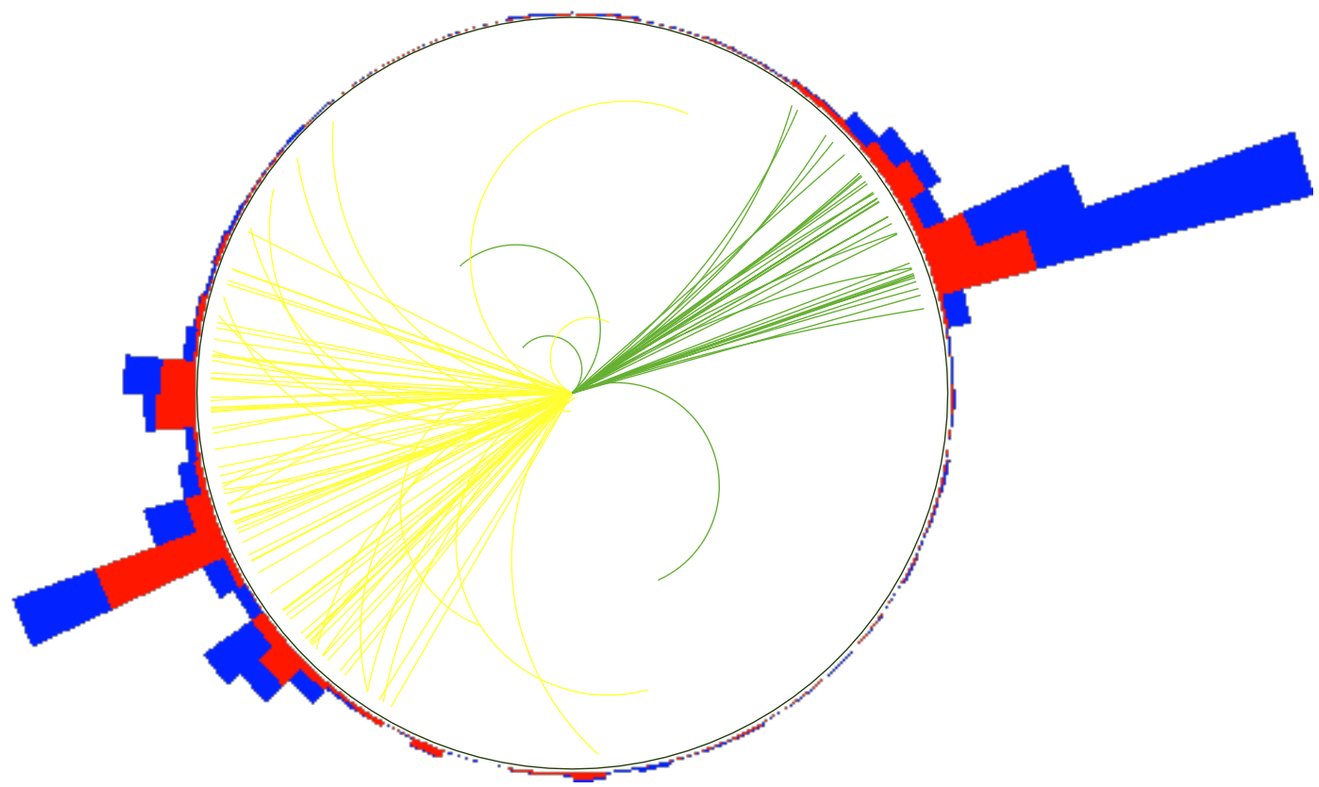
Jet algorithms and jet substructure HATS



Jim Dolen, Justin Pilot, Alexx Perloff, Ben Kreis, Sal Rappoccio, John Stupak, Nhan Tran

https://indico.cern.ch/event/394942/
https://twiki.cern.ch/twiki/bin/view/CMS/JetSubstructureHATS2016

Before we start the lecture, start the process outlined here:
https://twiki.cern.ch/twiki/bin/view/CMS/JetSubstructureHATS2016#Getting_started

For the hand-on portion of the tutorial, begin by logging into an slc6 LPC node. If needed, set the following:

```
source /cvmfs/cms.cern.ch/cmsset_default.csh

setenv SCRAM_ARCH slc6_amd64_gcc530
```

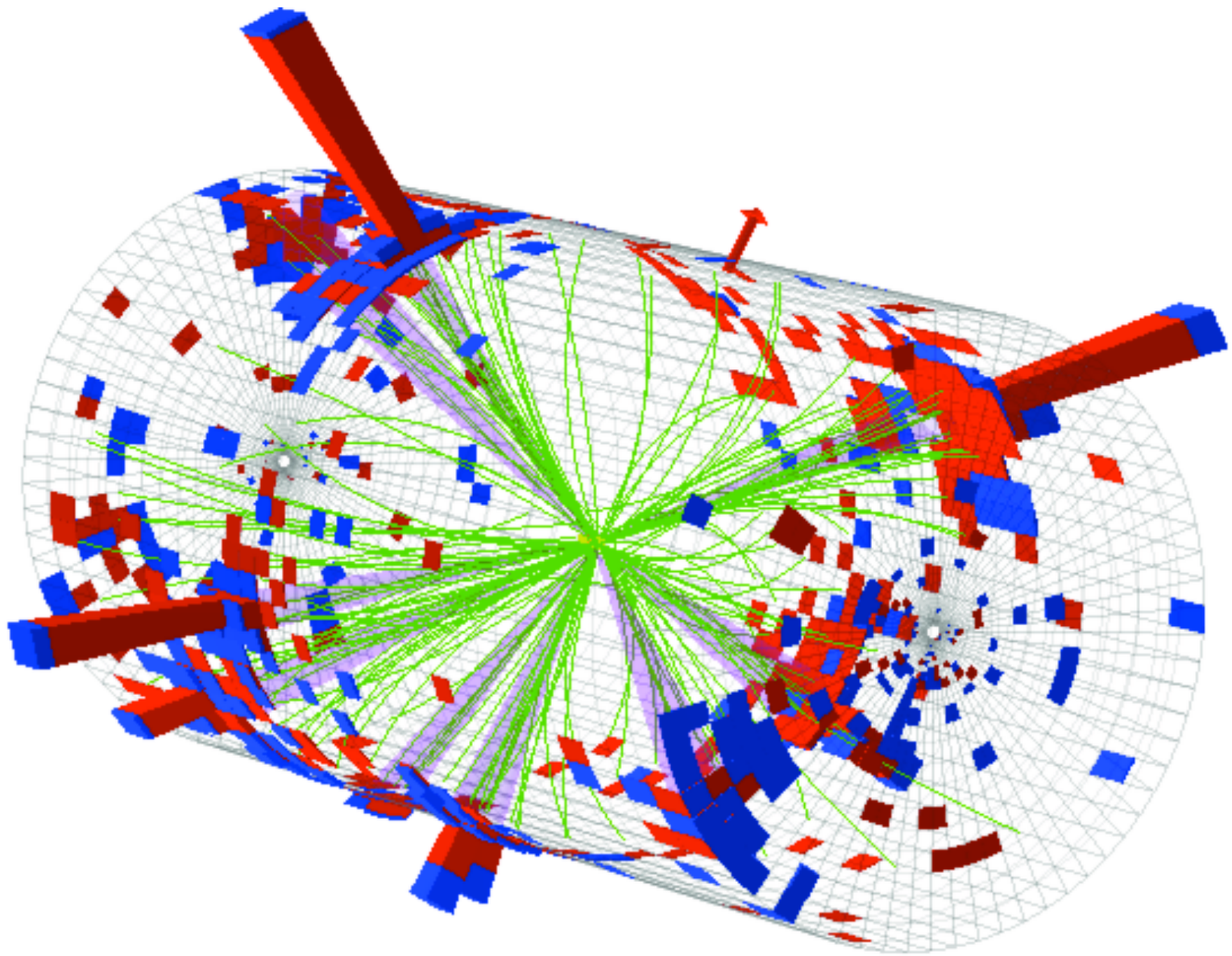
Then checkout the jet toolbox and some other code from github:

```
cmsrel CMSSW_8_0_12

cd CMSSW_8_0_12/src/

cmsenv

git-cms-init

mkdir Analysis

git-cms-addpkg JetMETCorrections/Modules

git clone git@github.com:cms-jet/JMEDAS.git Analysis/JMEDAS

git clone git@github.com:cms-jet/JetToolbox.git Analysis/JetToolbox -b jetToolbox_763

scram b -j 8
```

Same setup for JEC hats tomorrow: https://twiki.cern.ch/twiki/bin/view/CMS/JECHATSatLPC2016

Section 1: jet algorithms

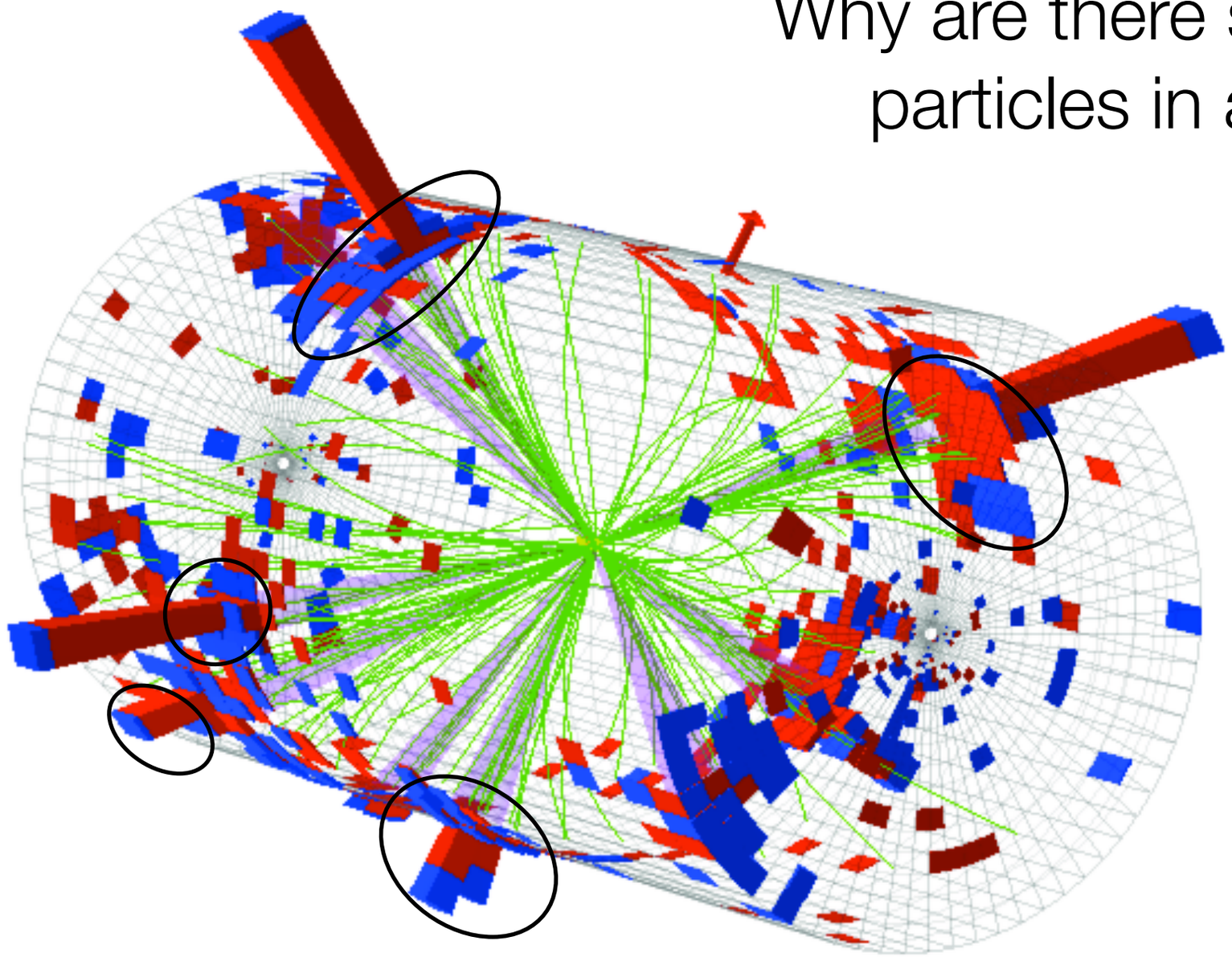
What is a jet?



CMS Experiment at LHC, CERN
Data recorded: Mon May 23 21:46:26 2011 EDT
Run/Event: 165567 / 347495624
Lumi section: 280
Orbit/Crossing: 73255853 / 3161

What is a jet?

Why are there so many particles in a jet?

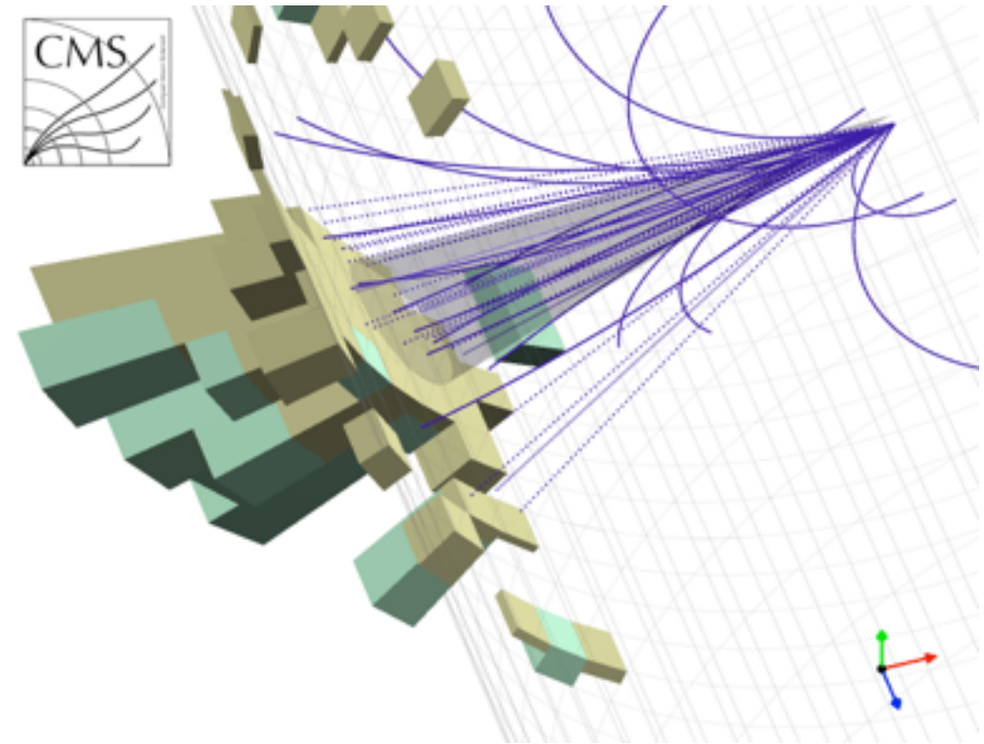


How do you determine which particles are included in a jet?

CMS Experiment at LHC, CERN
Data recorded: Mon May 23 21:46:26 2011 EDT
Run/Event: 165567 / 347495624
Lumi section: 280
Orbit/Crossing: 73255853 / 3161

What is a jet?

- Initial particle = quark or gluon
- Final object = collimated spray of hadrons and hadron decay products
- Need a “jet algorithm” to associate the particles in this spray



$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p})\, \Delta R_{ij}^2 / R^2$$

$$d_{iB} = p_{ti}^{2p}\ .$$

***p* = 1 → kt algorithm (KT)**

***p* = 0 → Cambridge Aachen algorithm (CA)**

***p* = -1 → anti-kt algorithm (AK)**

- Sequential recombination algorithm
  - Iteratively find the two particles in the event which are closest in some distance measure and combine them
    - Combine two particles if $d_{ij} < d_{iB}$
    - Stop when $d_{ij} > d_{iB}$

Fastjet

FastJet code documentation

These pages provide automatically generated documentation for the FastJet package.

www.fastjet.fr

The most useful classes

Many of the facilities of FastJet can be accessed through the three following classes:

- **PseudoJet**: the basic class for holding the 4-momentum of a particle or a jet.
- **JetDefinition**: the combination of a **JetAlgorithm** and its associated parameters. Can also be initialised with a **plugin**.
- **ClusterSequence**: constructed with a vector of input (**PseudoJet**) particles and a **JetDefinition**, it computes and stores the information on how the input particles are clustered into jets.

Selected more advanced classes

- **ClusterSequenceArea**: with the help of an **AreaDefinition**, provides jets that also contain information about their area.

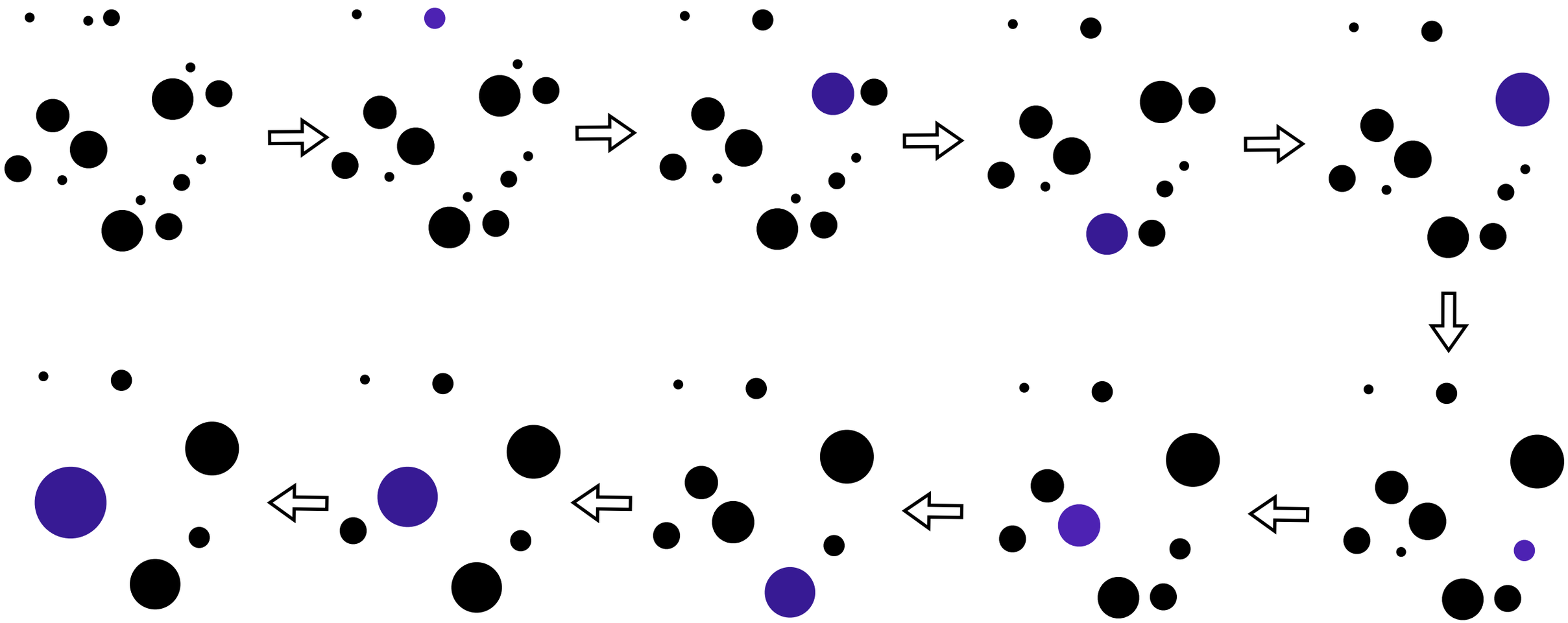
Selected additional tools

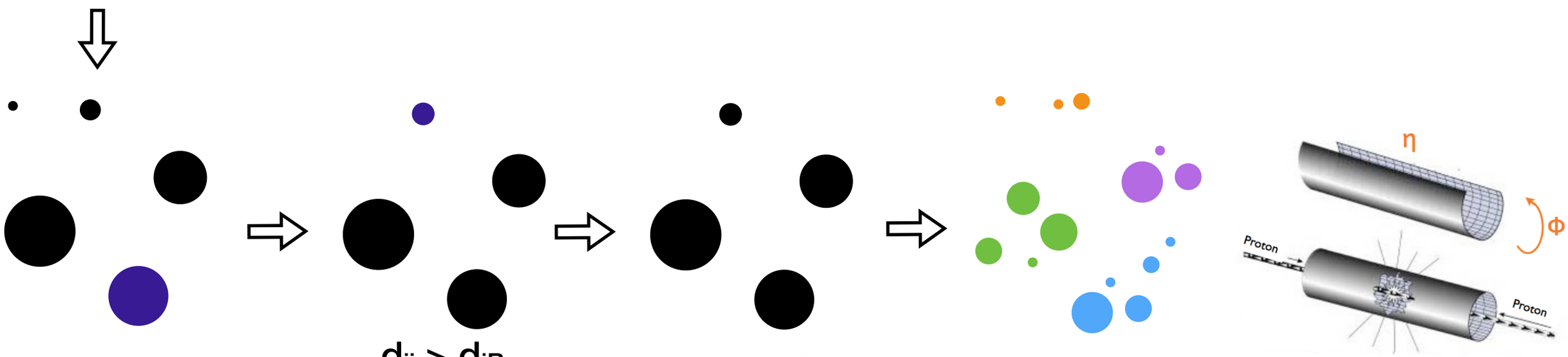
- **JetMedianBackgroundEstimator**: with the help of a **Selector**, a **JetDefinition** and an **AreaDefinition**, allows one to estimate the background noise density in an event; for a simpler, quicker, effective alternative, use **GridMedianBackgroundEstimator**
- **Transformer**: class from which are derived various tools for manipulating jets and accessing their substructure. Examples are **Subtractor**, **Filter**, **Pruner** and various taggers (e.g. **JHTopTagger** and **MassDropTagger**).

Initial particles

Combine the 2 particles with smallest d_{ij}

Continue iteratively combining particles (at each step combine the protojets with smallest d_{ij})



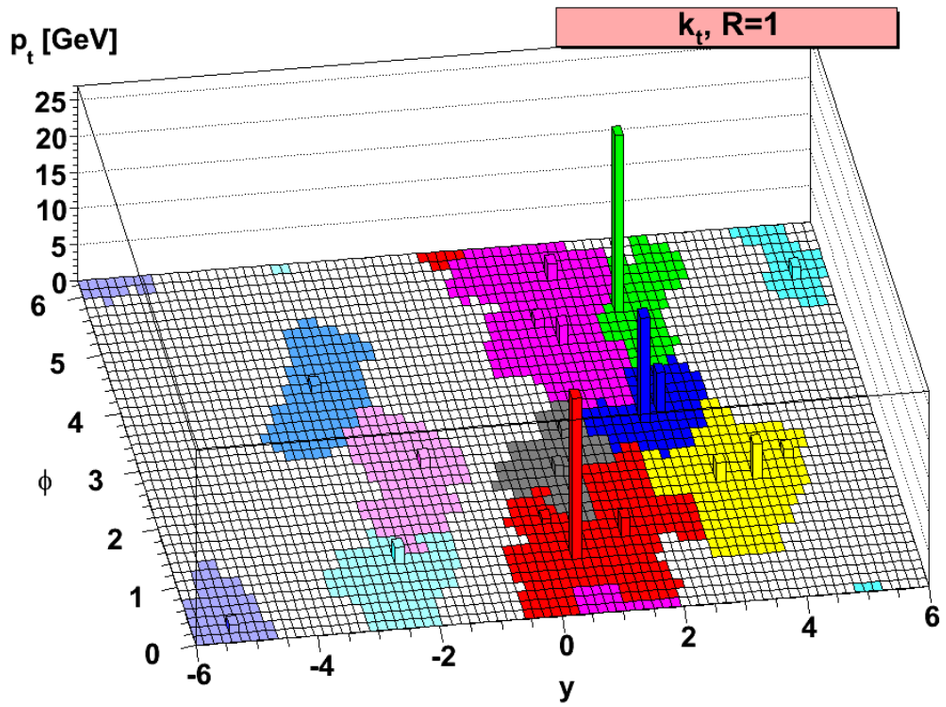


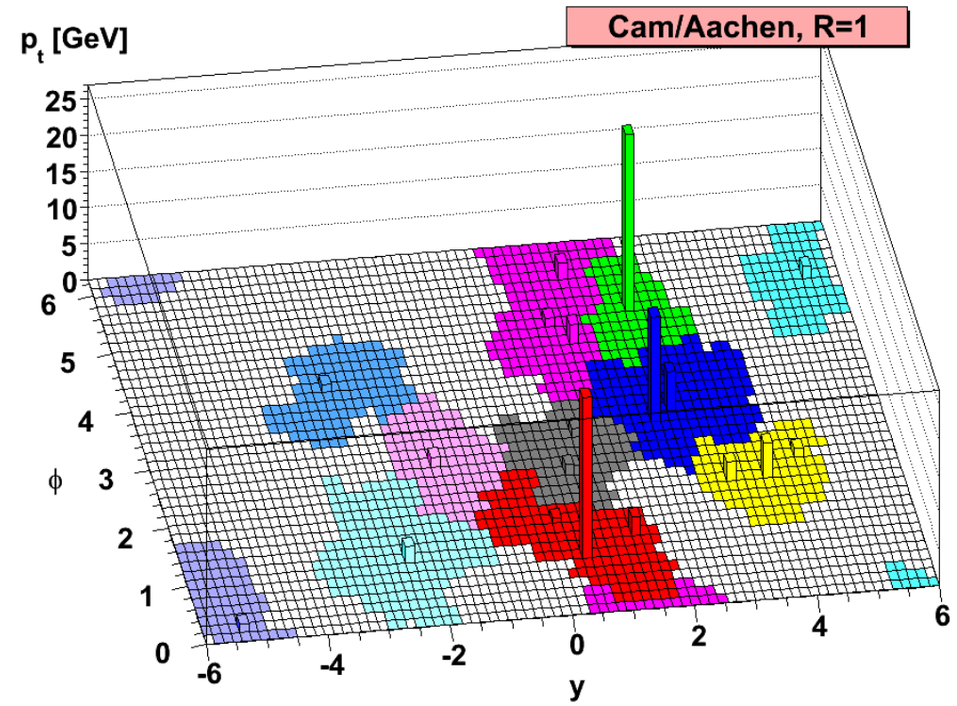
$d_{ij} > d_{iB}$ → stop clustering

Found 4 jets

4 jets, each with N constituents

Algorithm comparison

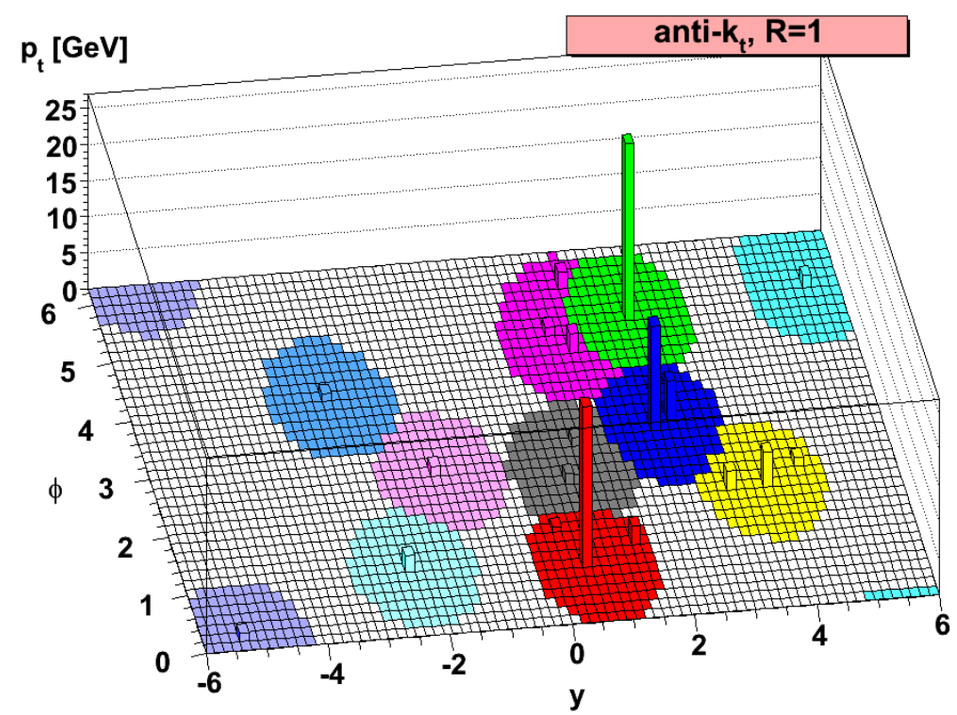




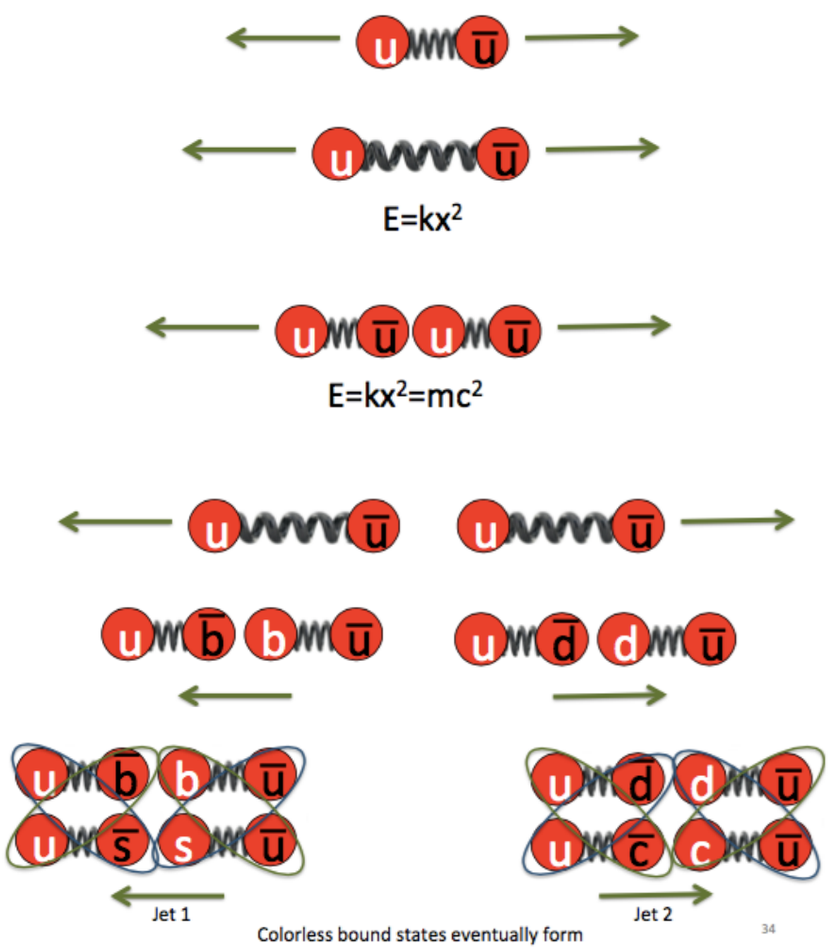
kT - soft particles clustered first

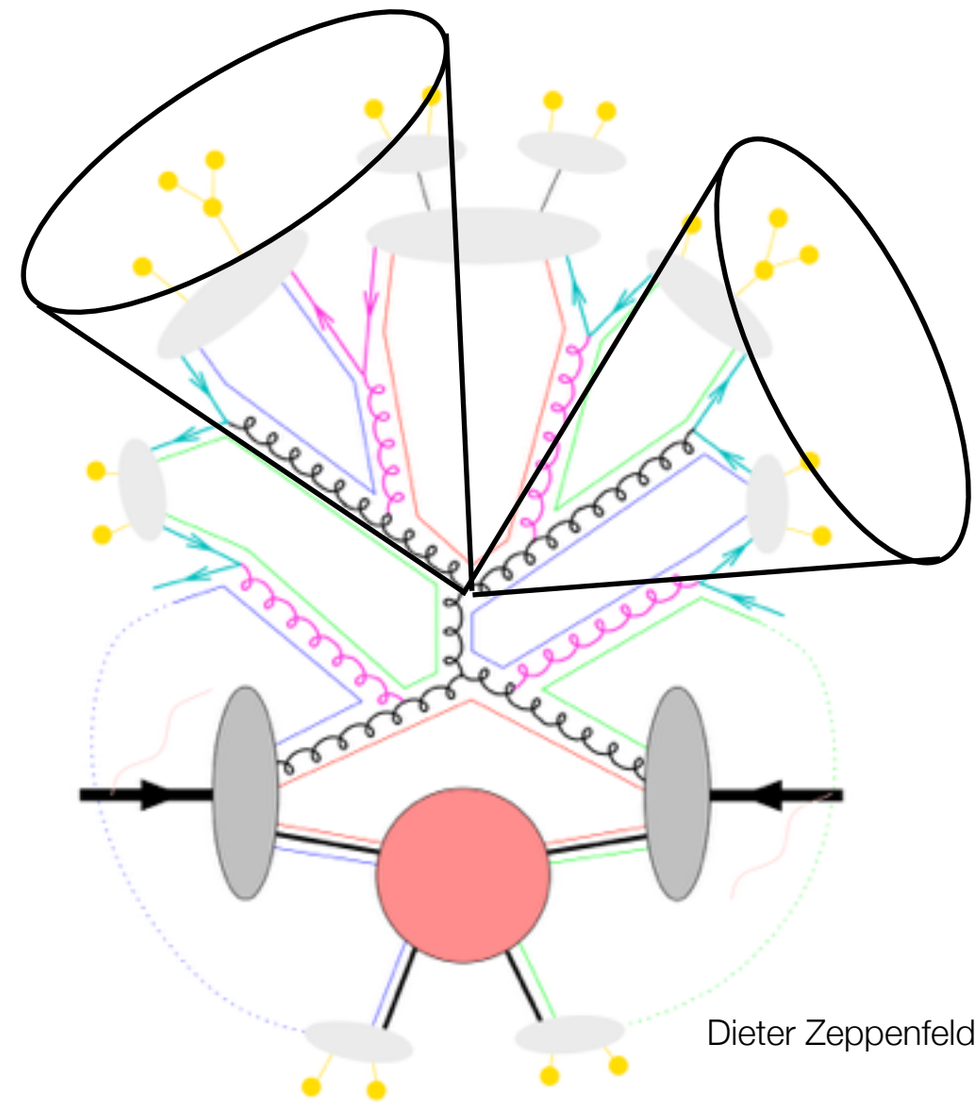
CA - angular clustering

anti-kT - hard particles clustered first



Jets are messy

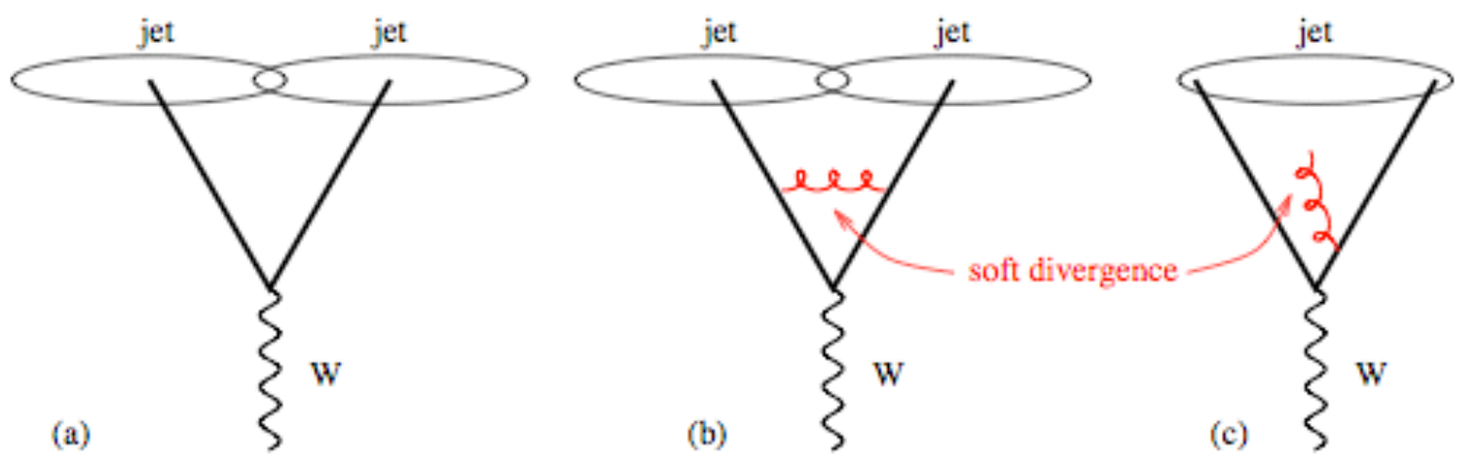




Dieter Zeppenfeld

Jet Algorithms

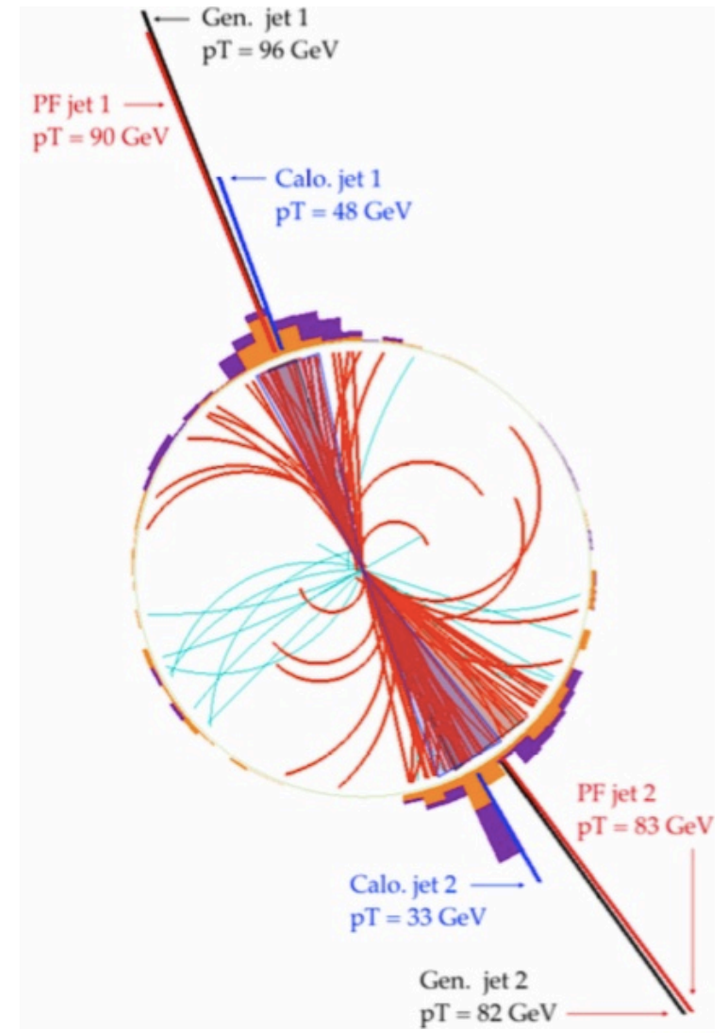
- A good jet algorithm is infrared and collinear safe
  - The set of hard jets should be unchanged by soft emission and collinear splitting

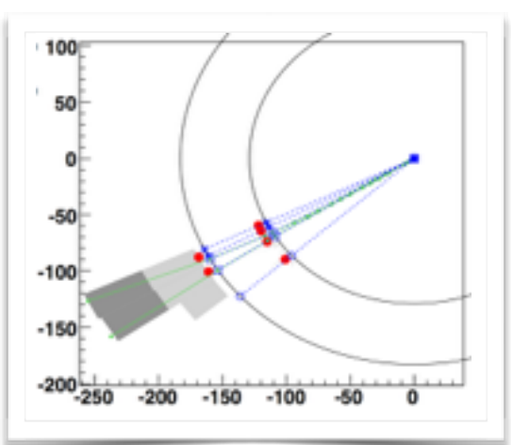


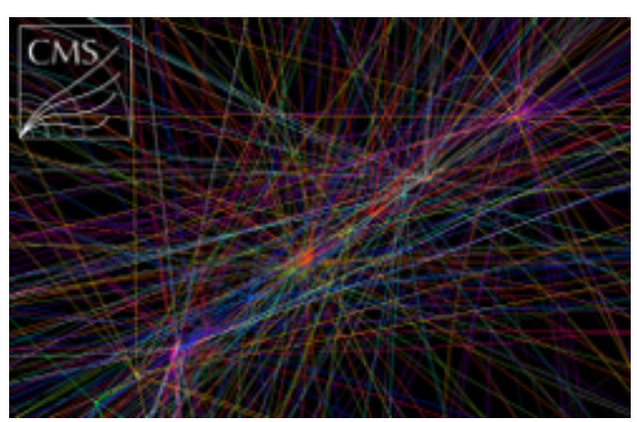
Graphics from G. Salam
arXiv:0906.1833v1

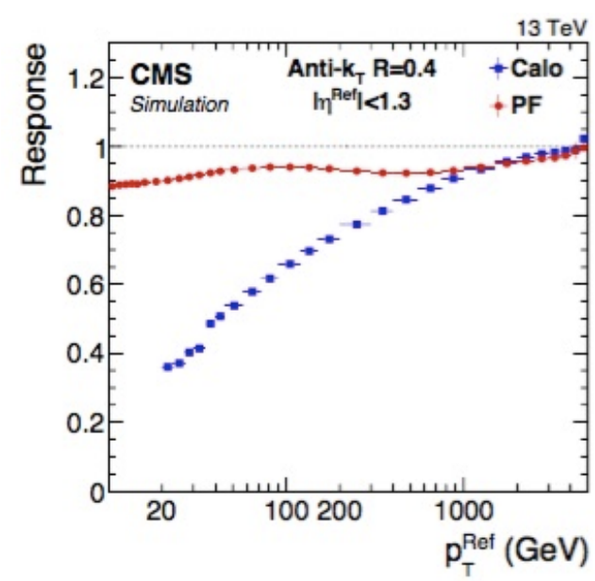
Jet inputs

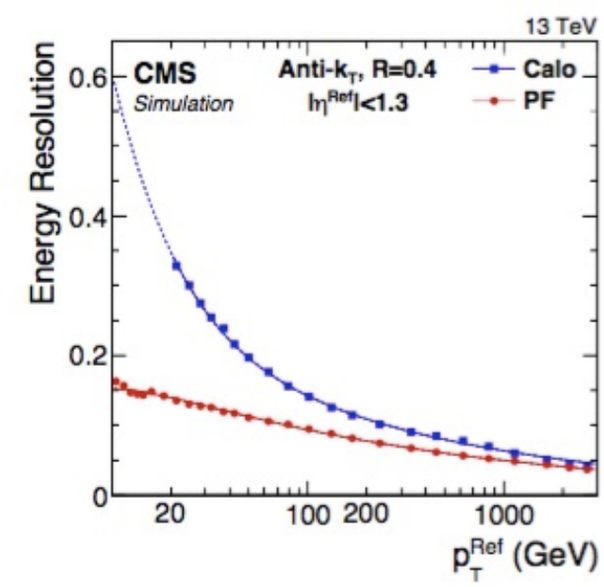
- Calorimeter towers or calorimeter clusters (CaloJets)
- Particle flow (PF) particles
  - PF Jets = jets clustered from all PF inputs
  - CHS jets = “Charge Hadron Subtracted” jets = remove charged PF particles associated to non-primary vertices (remove charged pileup)
  - PUPPI jets = PF constituents have been weighted/removed by an algorithm which is designed to remove pileup contamination (more info tomorrow)
- Generated particles (GenJets or GenJetsNoNu)





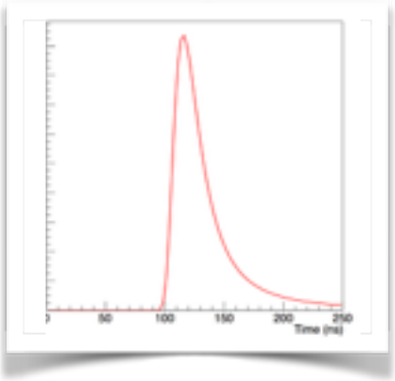




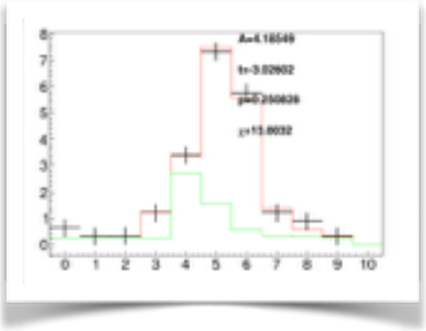


Jet and MET reconstruction

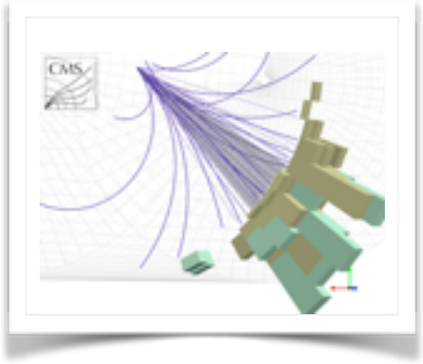
ECAL
pulse reconstruction



HCAL
pulse reconstruction

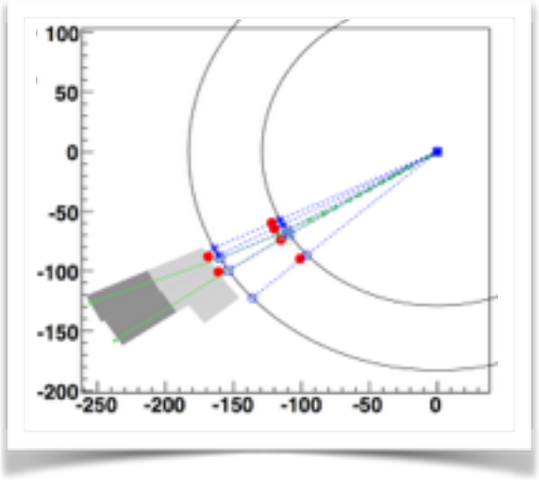


Track
reconstruction

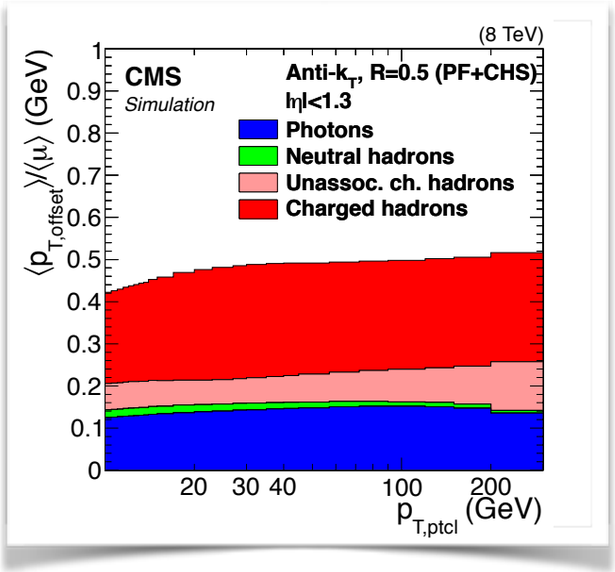


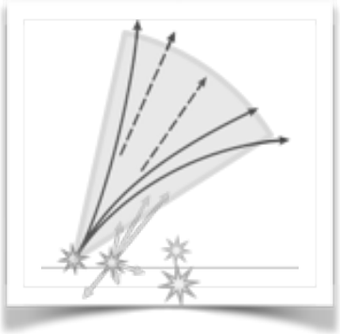
Particle Flow
cluster
reconstruction
and calibration

Particle Flow
linking
and particle
reconstruction

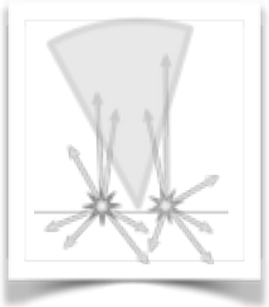


Per particle
pileup removal
(CHS, PUPPI)





Jet
Clustering



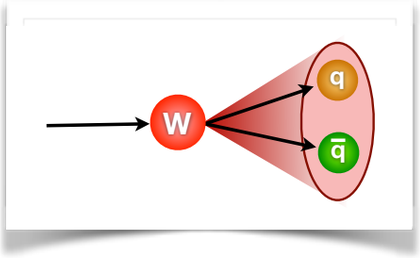
Jet & MET
corrections

Pileup
Jet ID

Quark /
gluon
jet ID

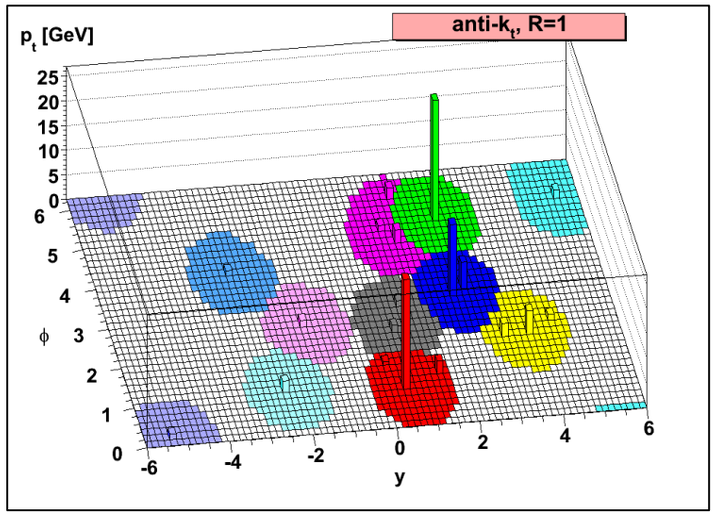
W/Z/H/t
tagging

Data
quality

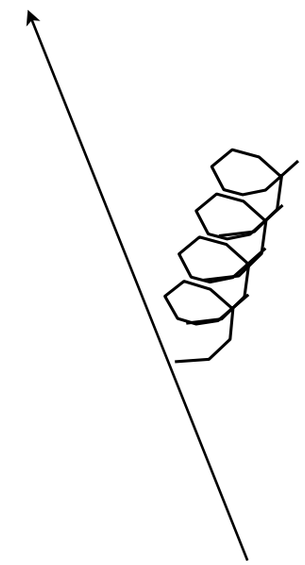


Jet properties

- Jet 4-vector = sum of all constituent particle 4-vectors
  - energy, pT, η, Φ
  - Jet mass $m = \sqrt{E^2 - p^2}$
  - Jet constituent multiplicities (PF)
    - ex. charged multiplicity
  - Jet constituent fractions
    - ex. charged hadron energy fraction
  - Jet area = area in η-Φ plain in which an infinitely soft particle will be clustered with the jet



what is jet mass?

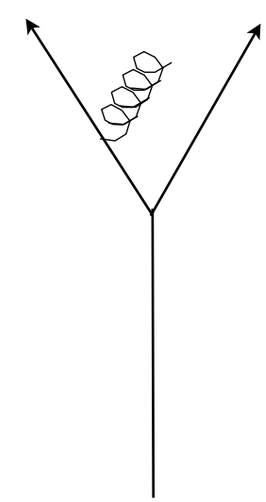


Jet mass is a perturbative quantity,
at LO the jet mass is the parton mass ~ 0.

$$\langle M^2 \rangle \simeq C \cdot \frac{\alpha_s}{\pi} p_t^2 R^2$$

Jet mass is proportional to R, pT
C is a form factor related to originating parton and clustering algorithm. For non-cone algorithms:

$$\langle M^2 \rangle \simeq \left.\begin{array}{ll} \text{quarks:} & 0.16 \\ \text{gluons:} & 0.37 \end{array}\right\} \times \alpha_s p_t^2 R^2$$



For heavy objects, the LO mass scale is the heavy object mass.

http://arxiv.org/abs/0906.1833

Exercise 1

Exercise -- Jet Kinematics

Start by plotting basic kinematics, jet pT, jet eta, jet phi, and jet mass, using the following command:

```
python -i plots1.py
```

Try to run over the different samples listed above. What do you expect to change in each case?

- What is the difference in the jet mass distributions between signal and background? What features do you see?

Next, we will examine how the **jet mass depends on jet momentum** (pT). For this exercise you should use the `plot2DQuantities.C` script.

- What is the expected relationship between pT and mass?
- How does QCD compare to ttbar or WW? What thresholds do you see in the 2D plots?

Exercise -- Jet Sizes

Analyses in CMS use various distance parameters in the clustering algorithms, notably 0.4 and 0.8. How do you expect the jet kinematics to change when using a clustering distance parameter of R = 0.4 vs R = 0.8?

Use the following command to draw comparison plots -- AK4 jets are shown in the black histogram; AK8 jets shown in the red histogram.

```
python -i compareAK4AK8.py
```

What about jet areas? After running the following command and using your knowledge of jet clustering, can you explain the shapes that you see? Use the following:

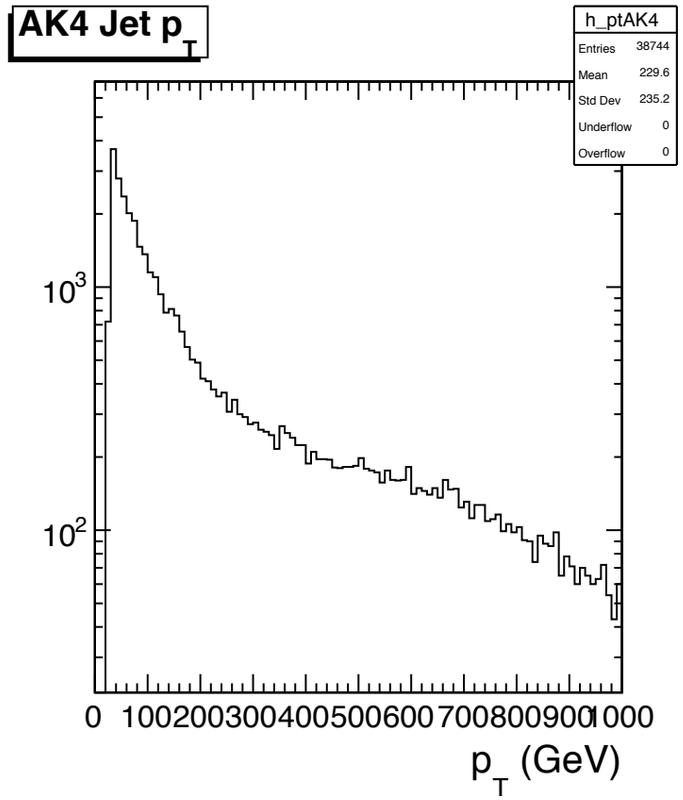
```
python -i plots3.py
```

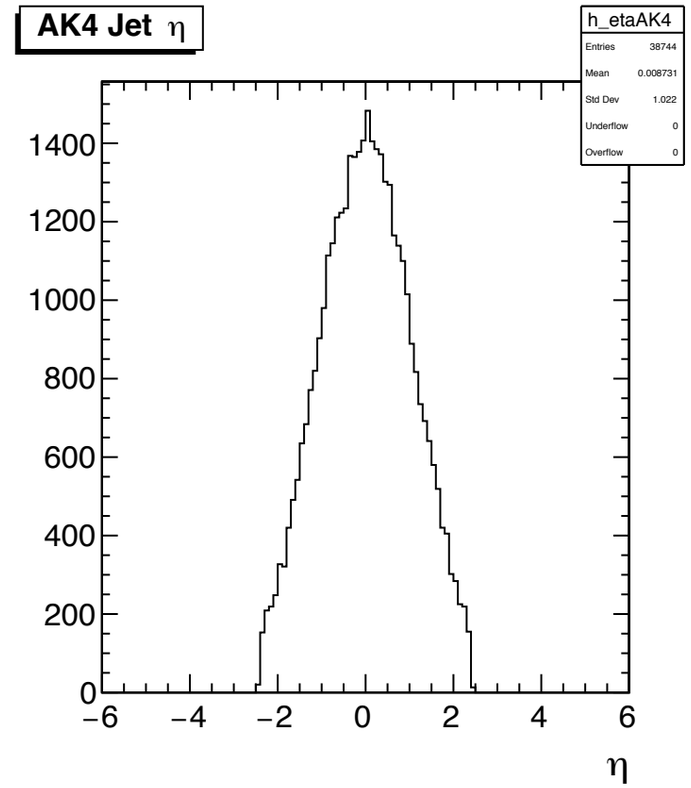
Exercise -- Generator-level Jets

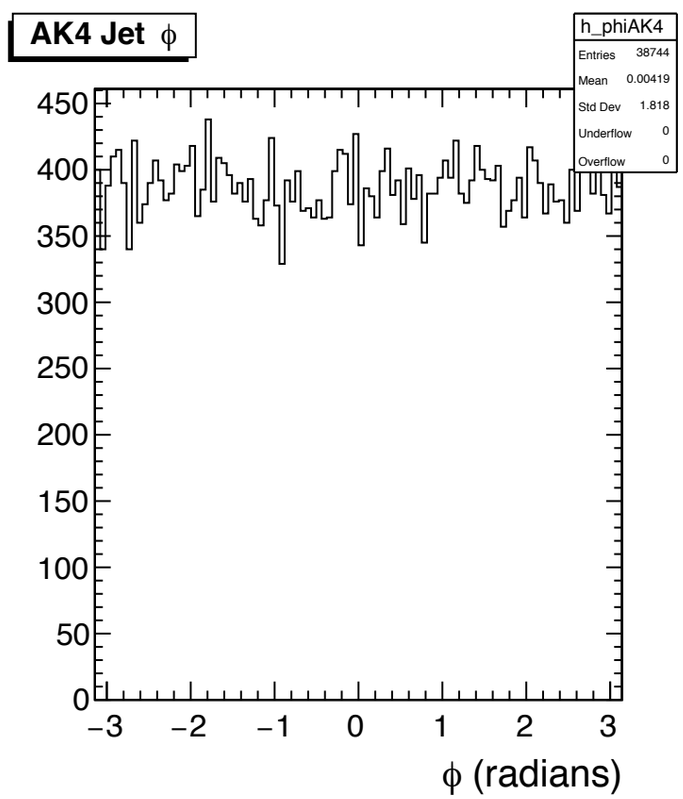
How does the CMS detector affect the reconstruction of jets? Compare the reconstructed jet kinematics with the generator-level jet kinematics:

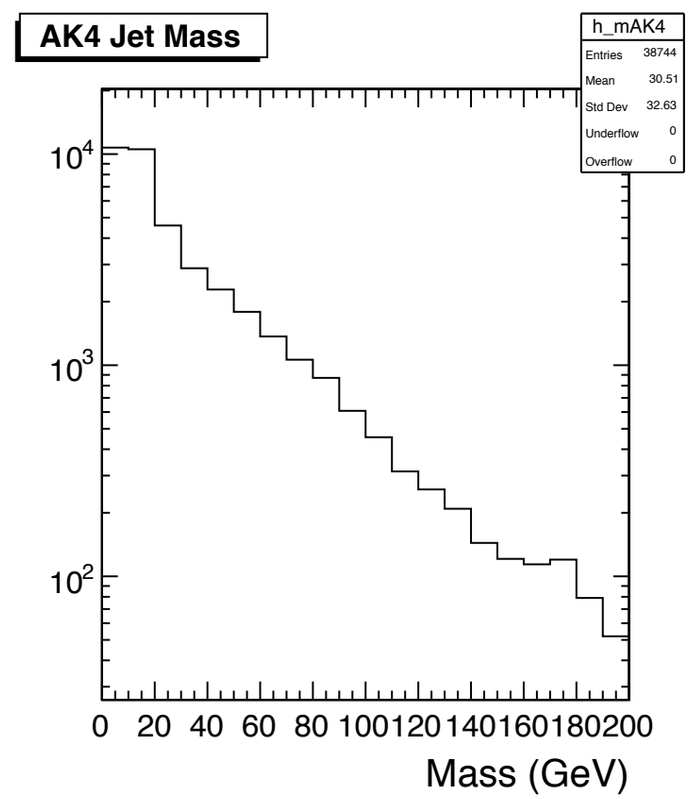
```
python -i plots2.py
```

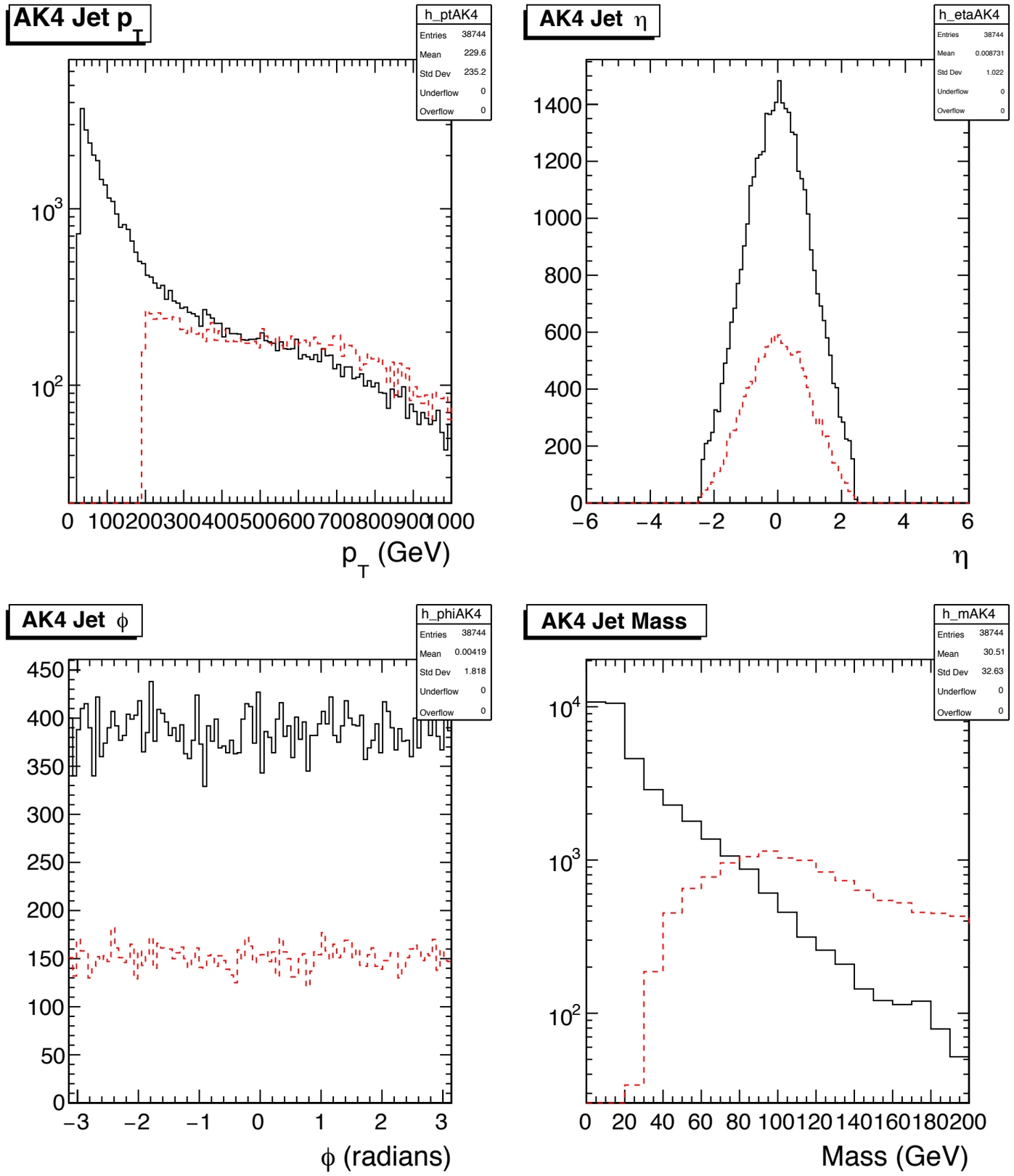
What effects can you see when comparing the two? Why do the mass peaks shift around?

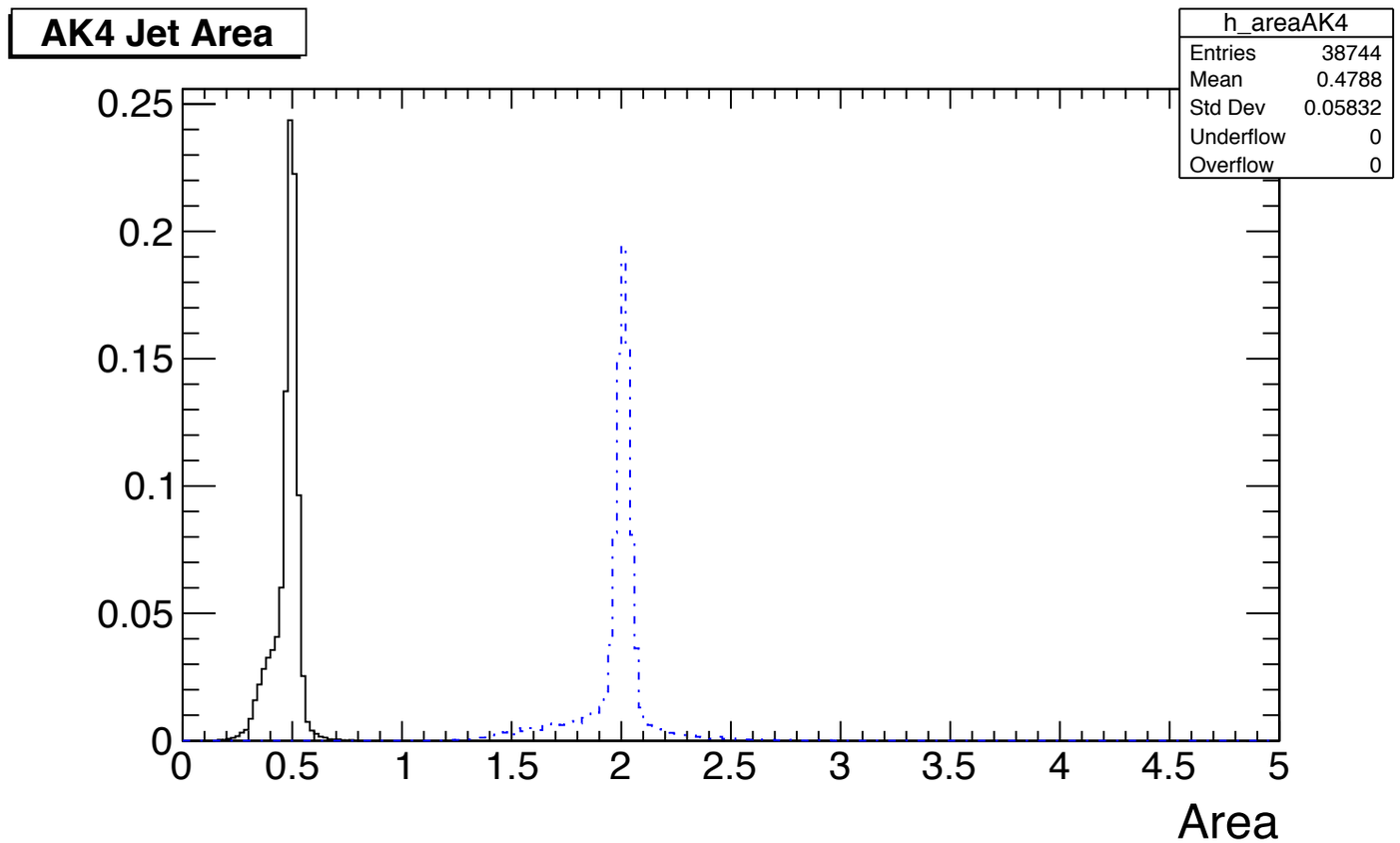

AK4 Jet p_T
h_ptAK4
Entries 38744
Mean 229.6
Std Dev 235.2
Underflow 0
Overflow 0
10^3
10^2
0 100200300400500600700800900 1000
p_T (GeV)

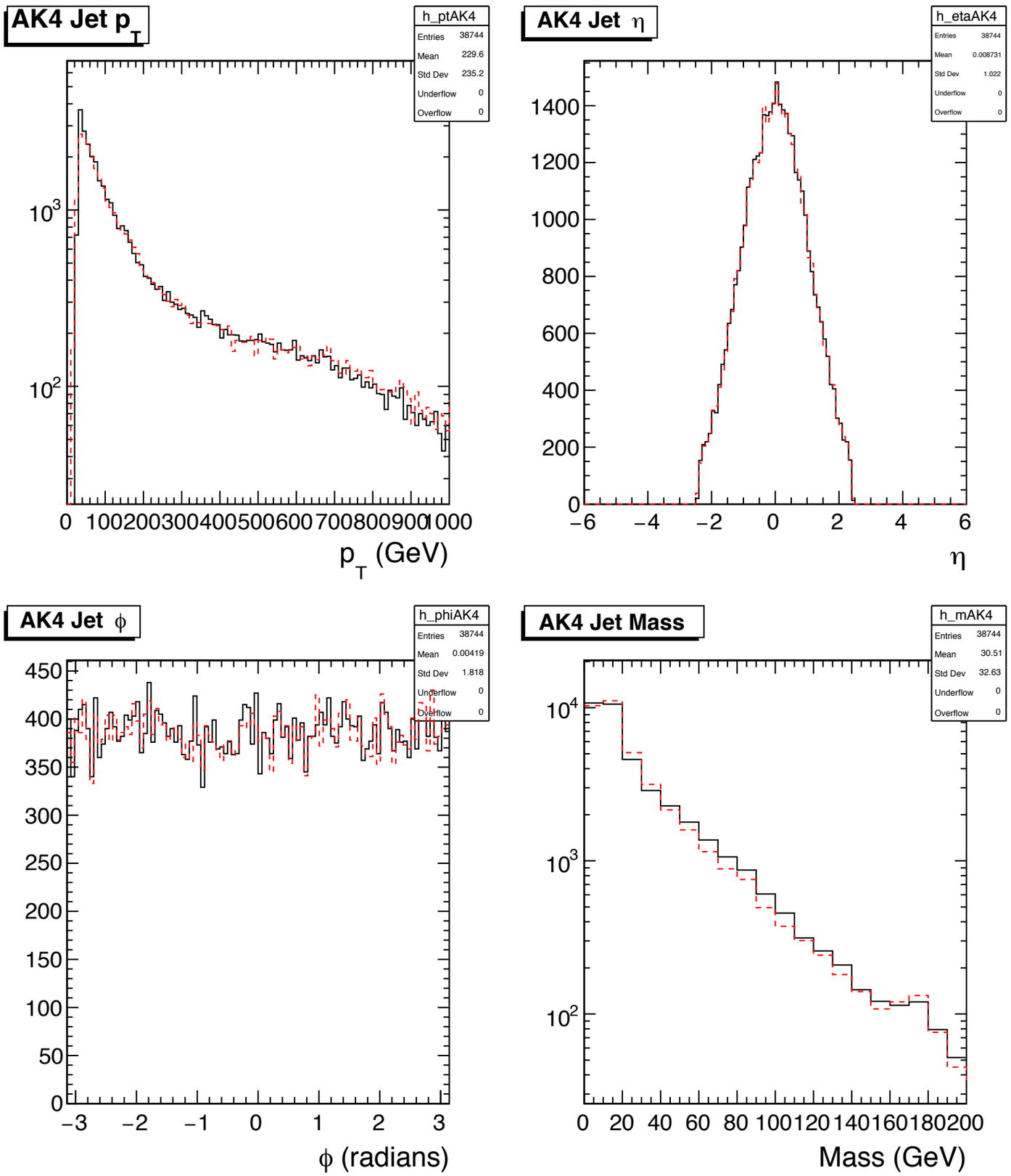

AK4 Jet η
h_etaAK4
Entries 38744
Mean 0.008731
Std Dev 1.022
Underflow 0
Overflow 0
1400
1200
1000
800
600
400
200
0
−6 −4 −2 0 2 4 6
η


AK4 Jet ϕ
h_phiAK4
Entries 38744
Mean 0.00419
Std Dev 1.818
Underflow 0
Overflow 0
450
400
350
300
250
200
150
100
50
0
−3 −2 −1 0 1 2 3
ϕ (radians)


AK4 Jet Mass
h_mAK4
Entries 38744
Mean 30.51
Std Dev 32.63
Underflow 0
Overflow 0
10^4
10^3
10^2
0 20 40 60 80 100120140160180200
Mass (GeV)


AK4 Jet p_T
h_ptAK4
Entries 38744
Mean 229.6
Std Dev 235.2
Underflow 0
Overflow 0
p_T (GeV)
AK4 Jet η
h_etaAK4
Entries 38744
Mean 0.008731
Std Dev 1.022
Underflow 0
Overflow 0
η
AK4 Jet ϕ
h_phiAK4
Entries 38744
Mean 0.00419
Std Dev 1.818
Underflow 0
Overflow 0
ϕ (radians)
AK4 Jet Mass
h_mAK4
Entries 38744
Mean 30.51
Std Dev 32.63
Underflow 0
Overflow 0
Mass (GeV)

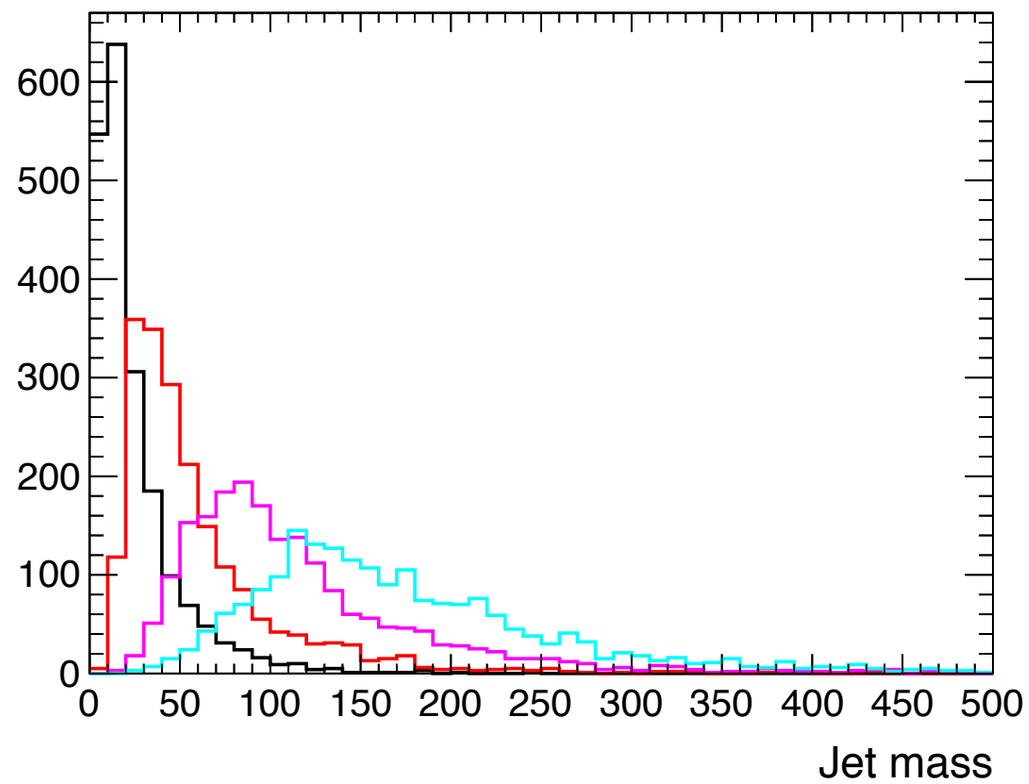

AK4 Jet Area
h_areaAK4
Entries 38744
Mean 0.4788
Std Dev 0.05832
Underflow 0
Overflow 0
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Area


AK4 Jet p_T
h_ptAK4
Entries 38744
Mean 229.6
Std Dev 235.2
Underflow 0
Overflow 0
p_T (GeV)
AK4 Jet η
h_etaAK4
Entries 38744
Mean 0.008731
Std Dev 1.022
Underflow 0
Overflow 0
η
AK4 Jet φ
h_phiAK4
Entries 38744
Mean 0.00419
Std Dev 1.818
Underflow 0
Overflow 0
φ (radians)
AK4 Jet Mass
h_mAK4
Entries 38744
Mean 30.51
Std Dev 32.63
Underflow 0
Overflow 0
Mass (GeV)

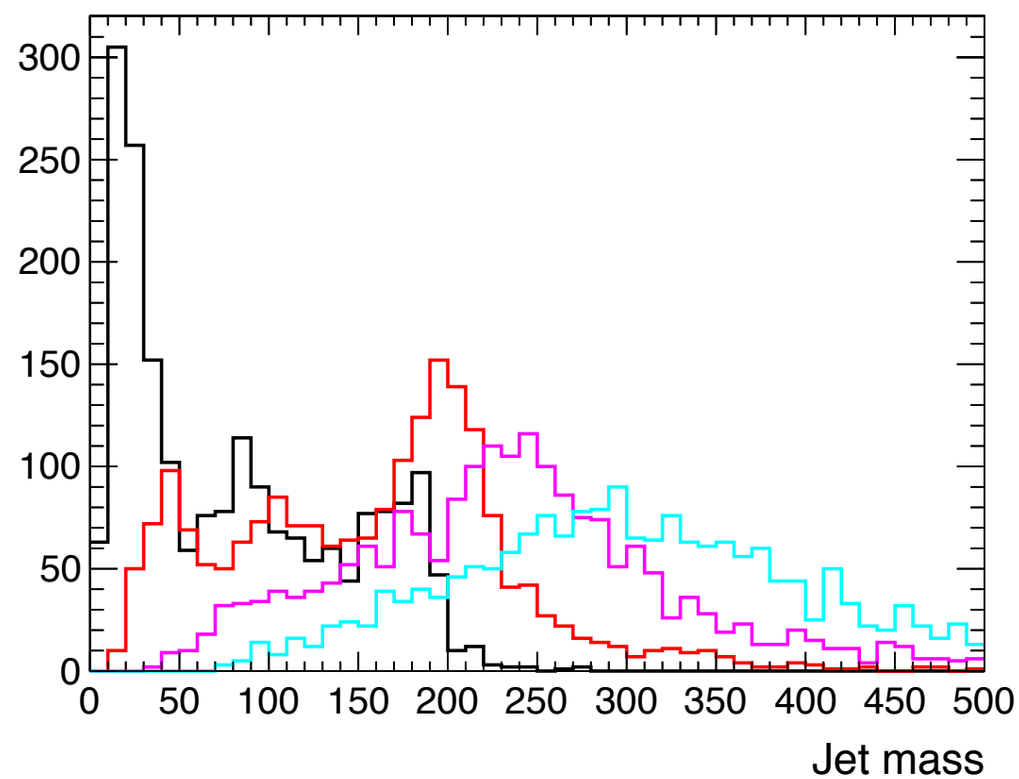
Additional informational plots
(made using the code from the “advanced” section of the HATS)
#homework

AK4, AK8, AK12, AK15

QCD

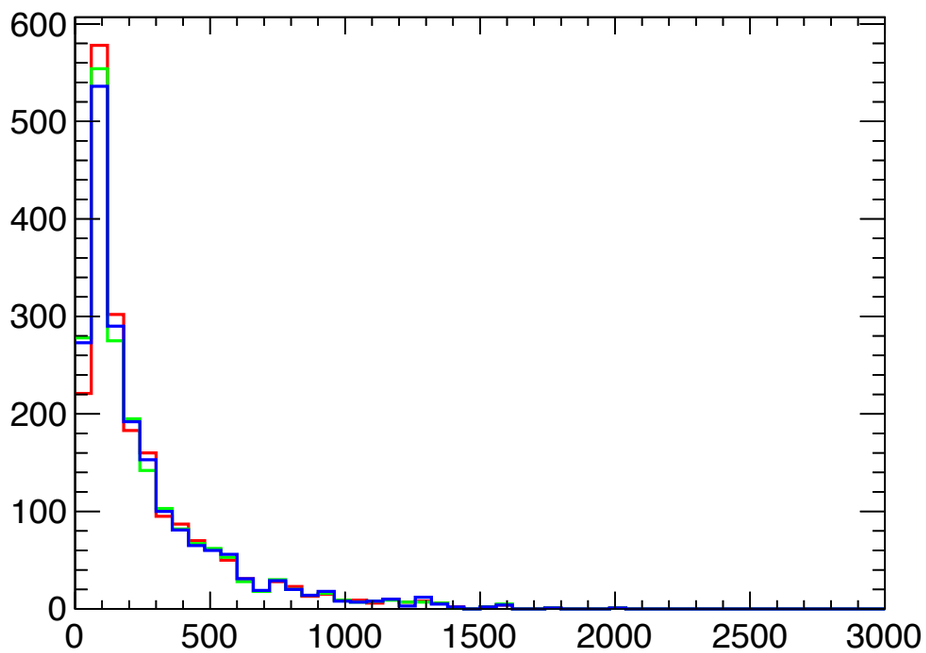


ttbar

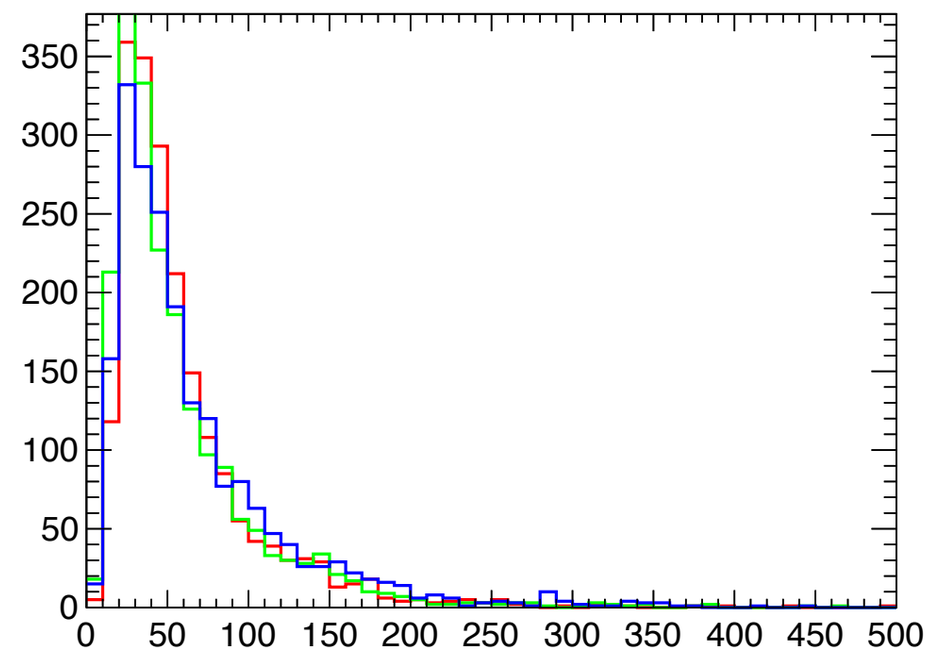


AK, CA, KT

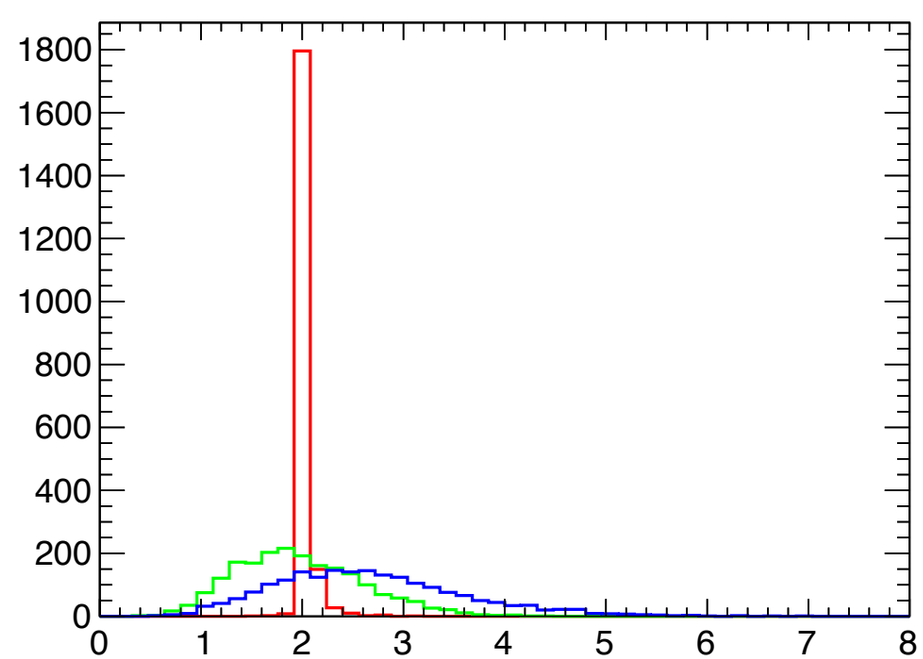
pT

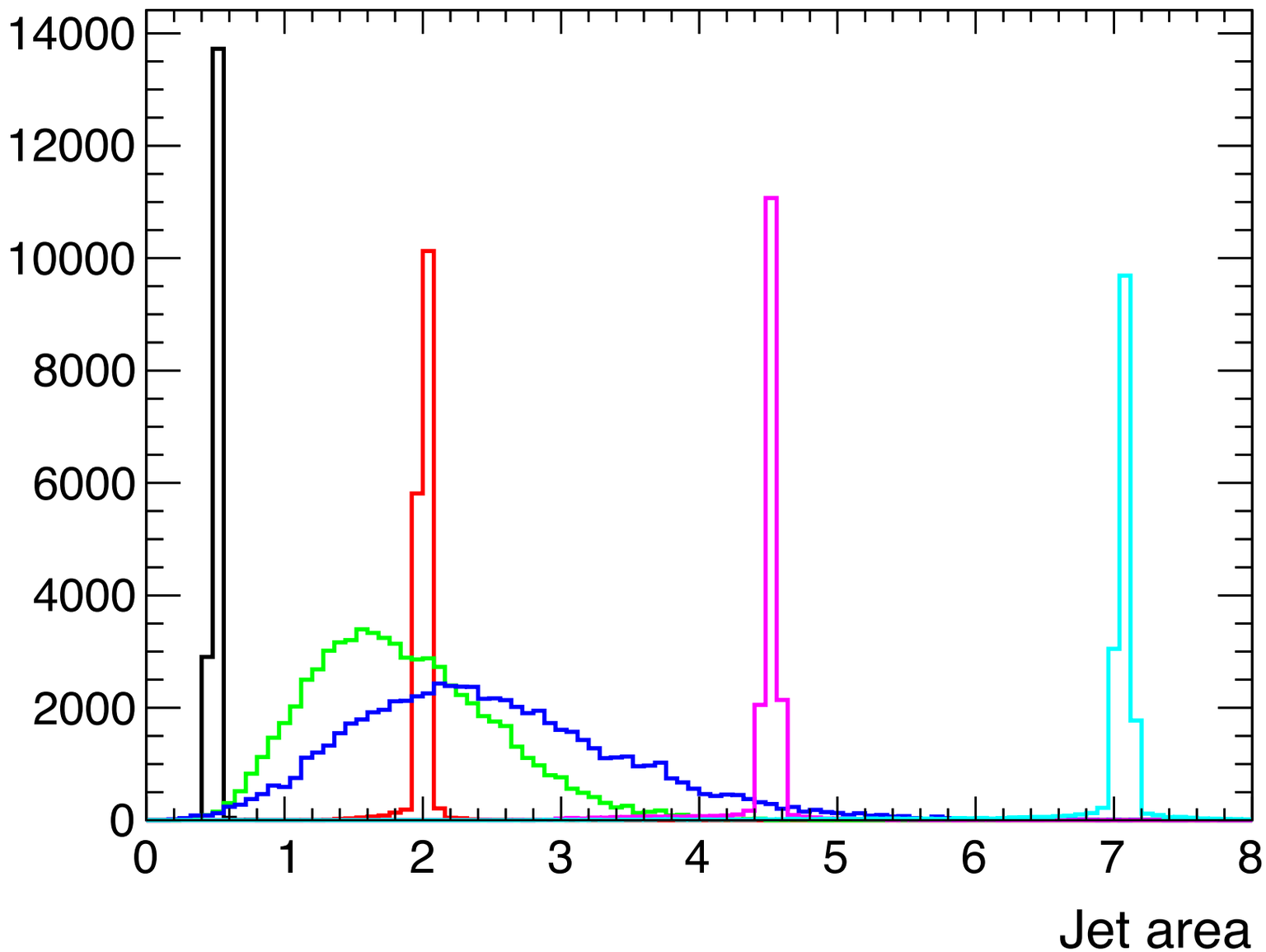


mass



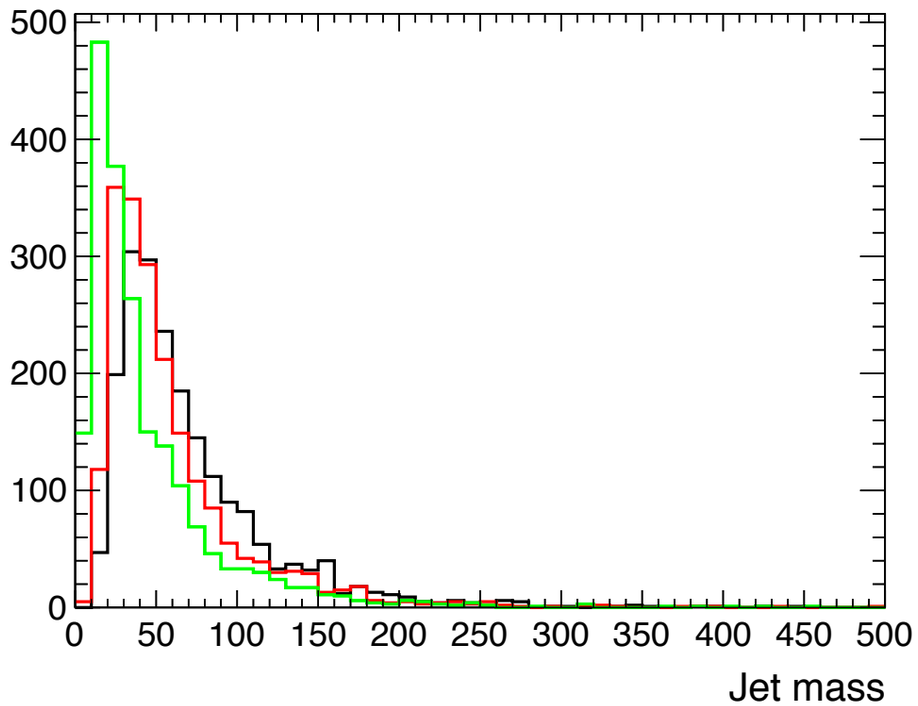
area

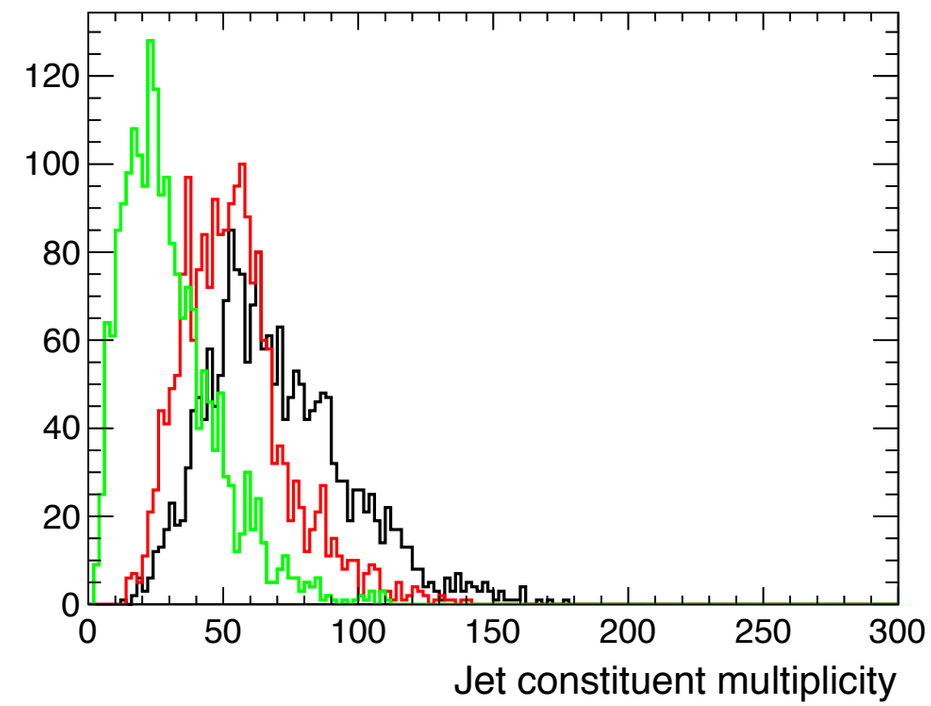


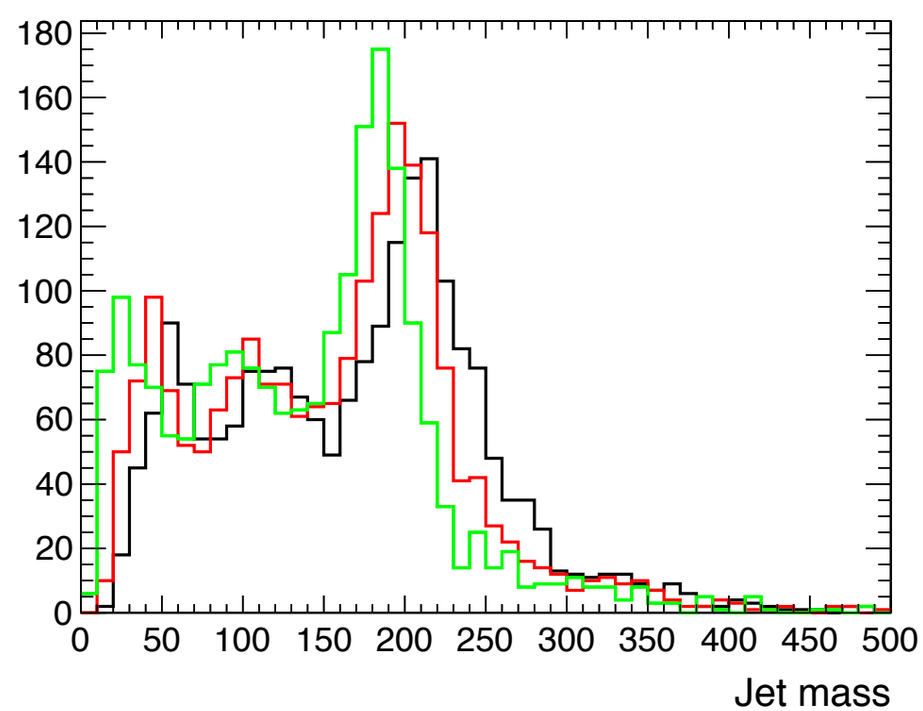


AK4, AK8, CA8, KT8, AK12, AK15

PF, CHS, PUPPI


500
400
300
200
100
0
0 50 100 150 200 250 300 350 400 450 500
Jet mass

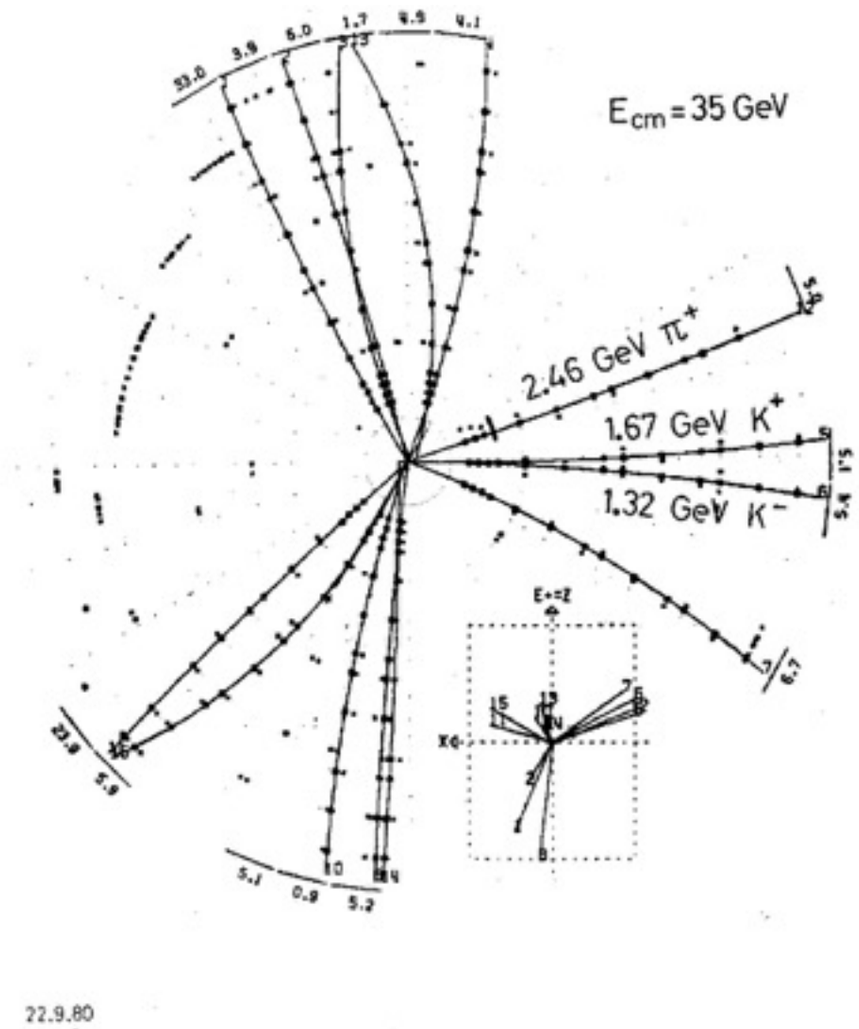

120
100
80
60
40
20
0
0 50 100 150 200 250 300
Jet constituent multiplicity

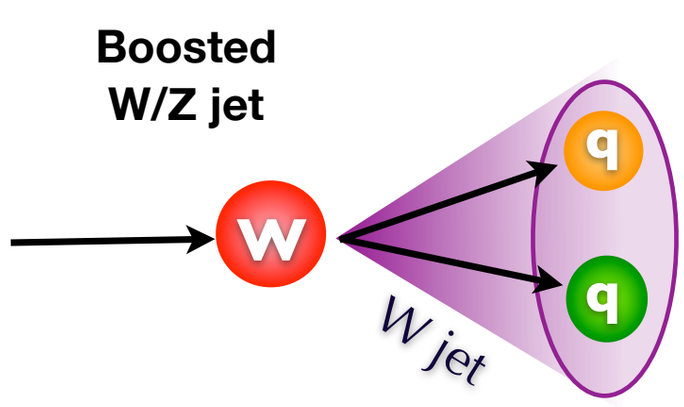

180
160
140
120
100
80
60
40
20
0
0 50 100 150 200 250 300 350 400 450 500
Jet mass

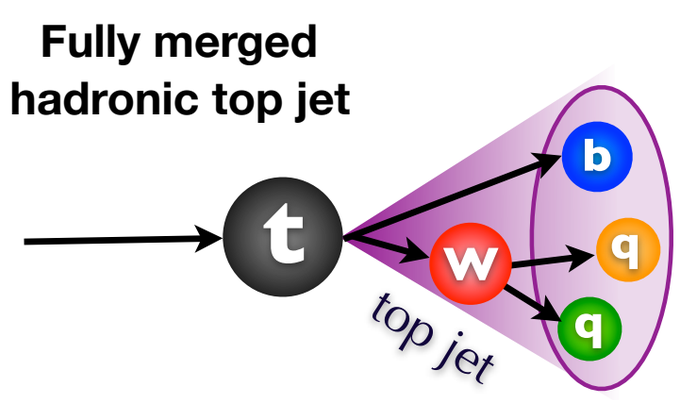
Section 2: types of jets

Types of jets

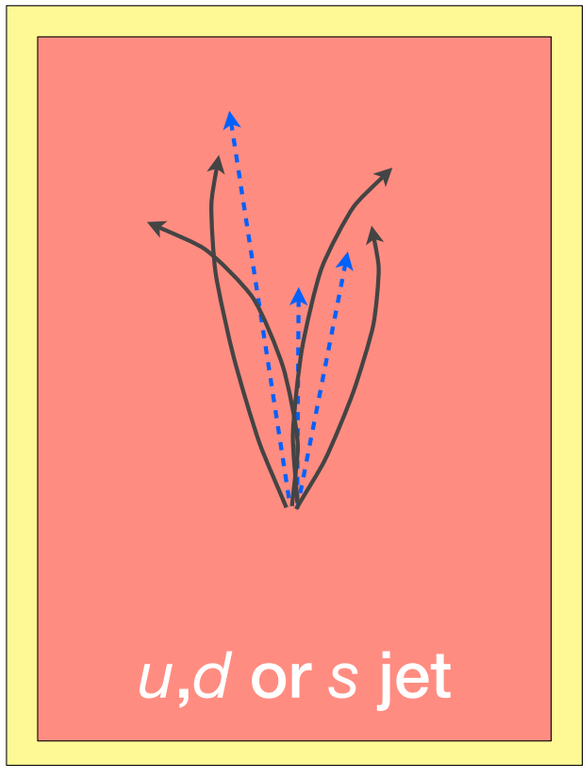
- Noise jets
- Light quark initiated jets and gluon initiated jets
  - Very similar, but we have some handles to discriminate (see later in the talk)
- B jets
  - Displaced vertex
- Pileup jets
  - Formed from overlapping soft PU particles
- Jets from merged boosted objects
  - W jet
  - Z jet
  - top jet
  - Higgs jet

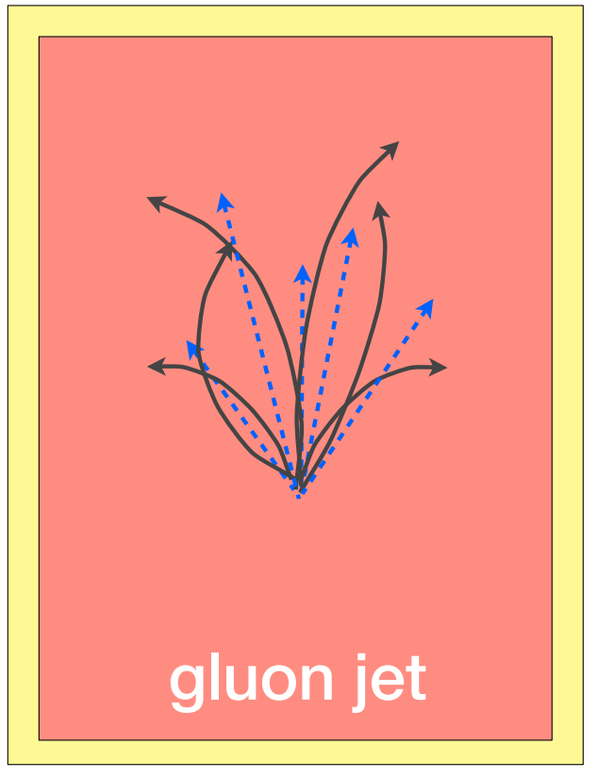


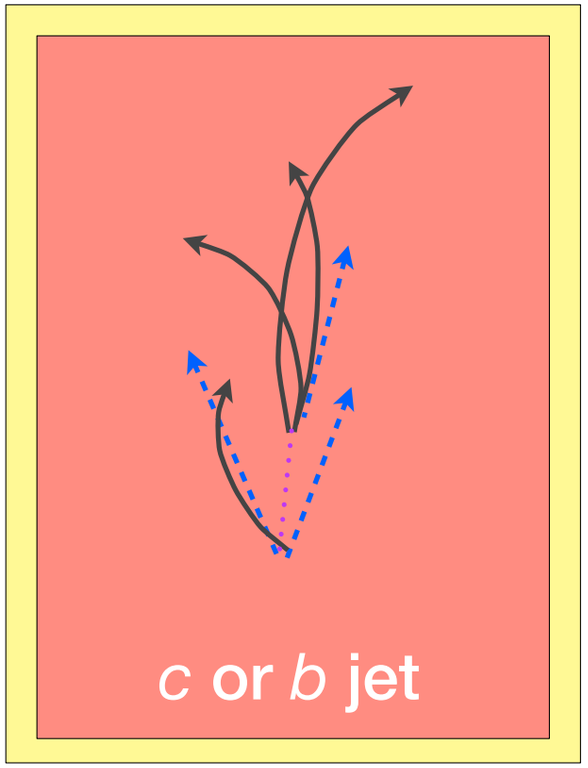


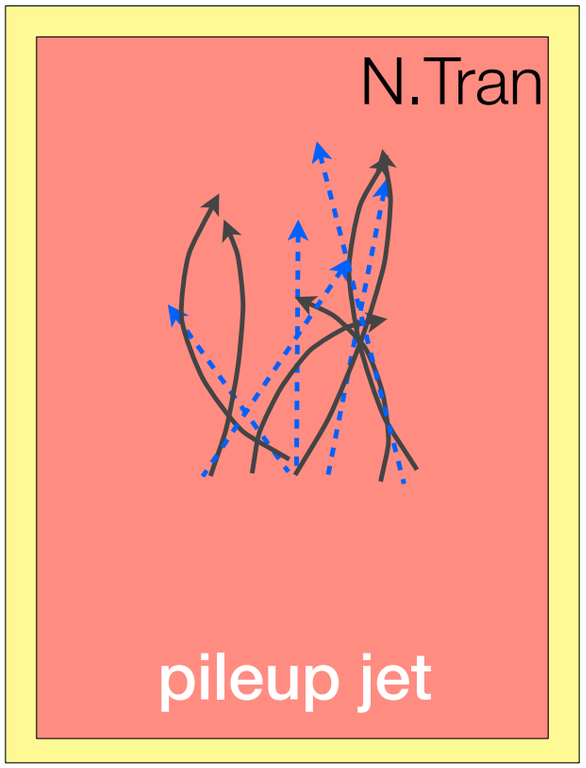


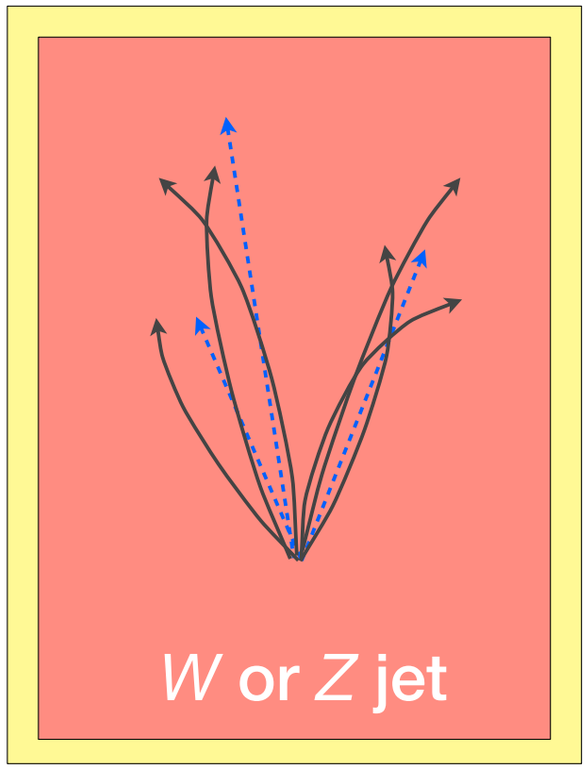
Types of jets

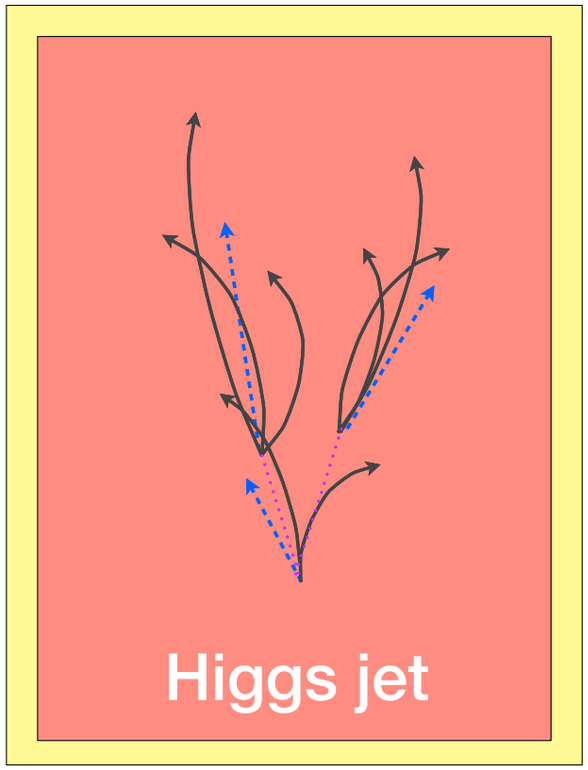


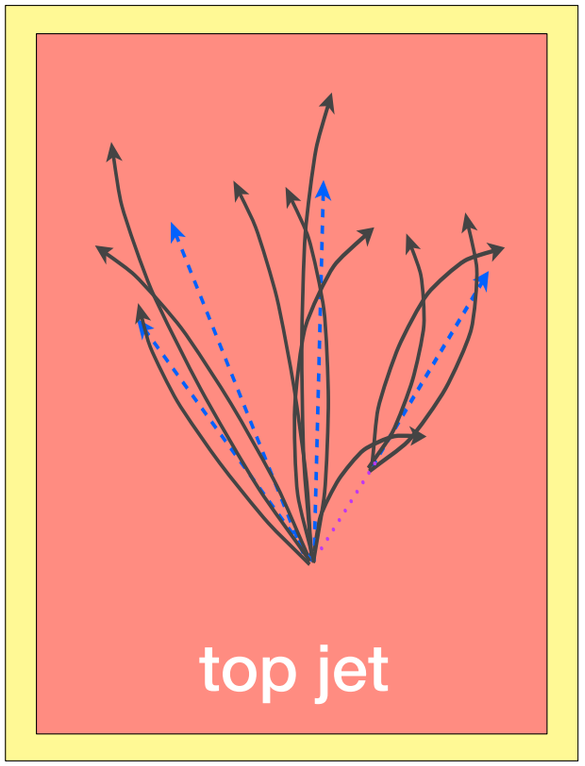


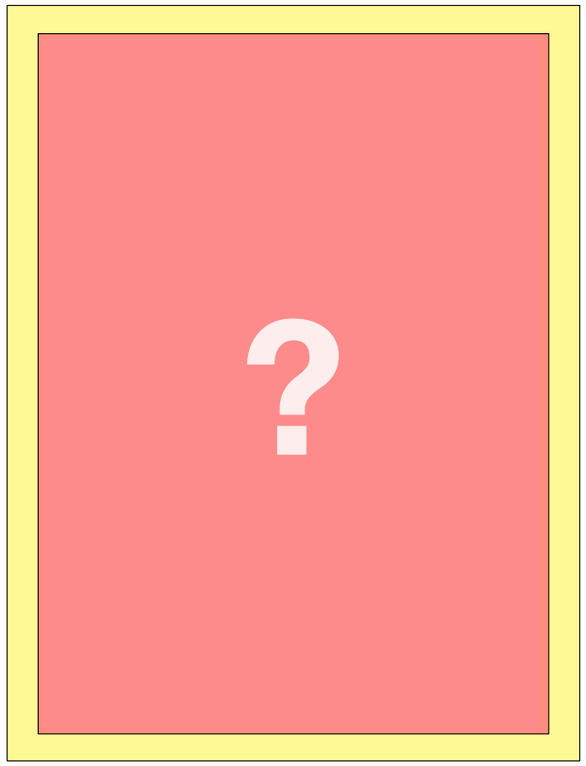






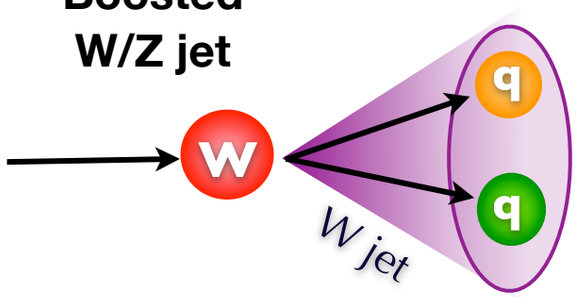




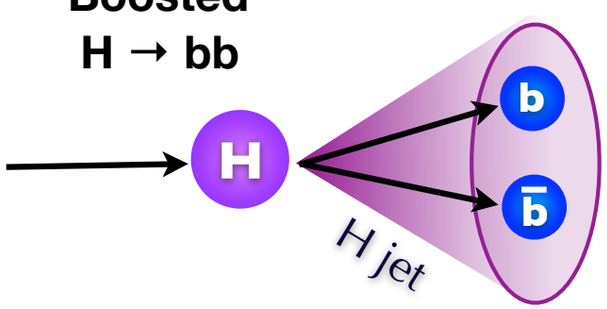


Boosted jets

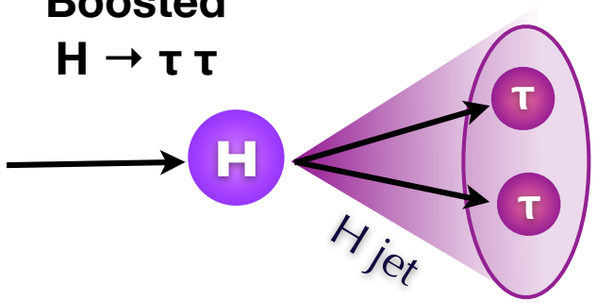
Boosted
W/Z jet



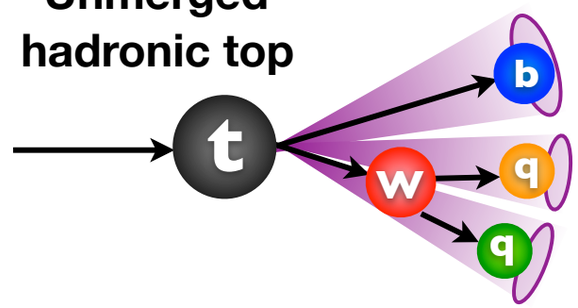
Boosted
H → bb



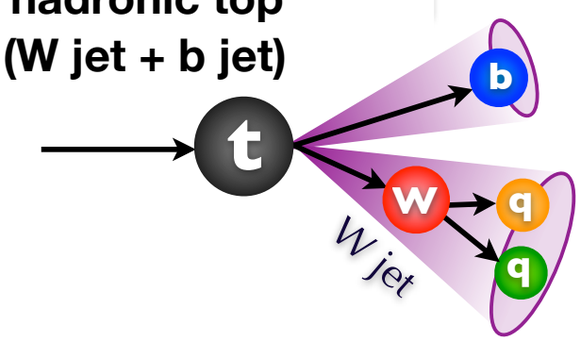
Boosted
H → τ τ



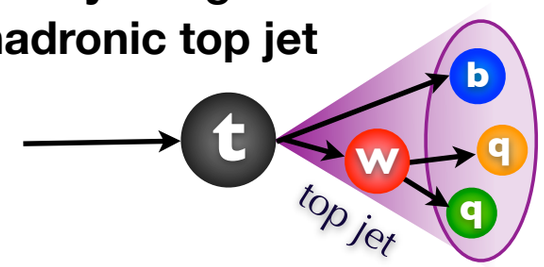
Unmerged
hadronic top



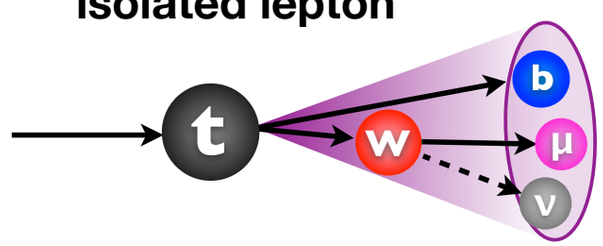
Partially merged
hadronic top
(W jet + b jet)



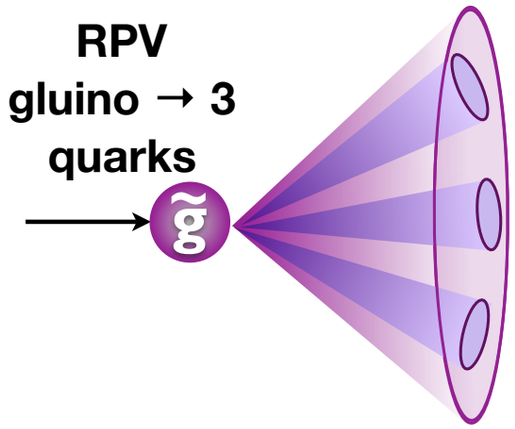
Fully merged
hadronic top jet



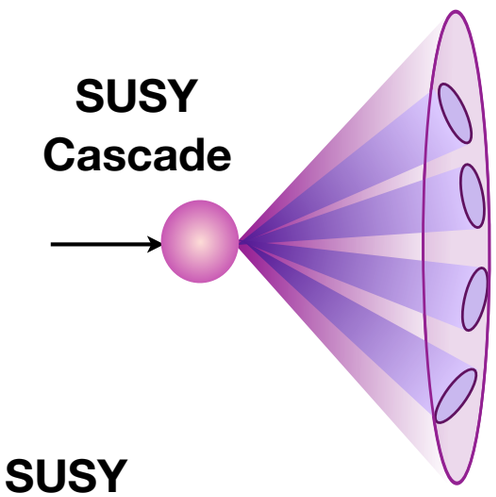
Leptonic top with non-
isolated lepton



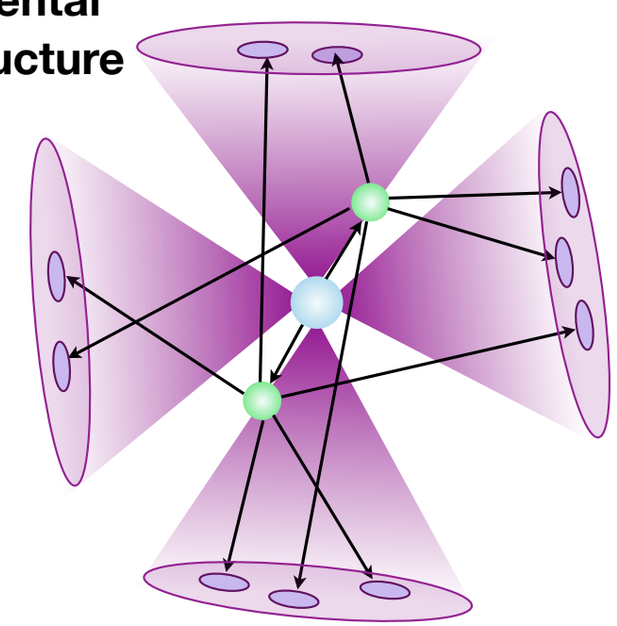
RPV
gluino → 3
quarks



SUSY
Cascade

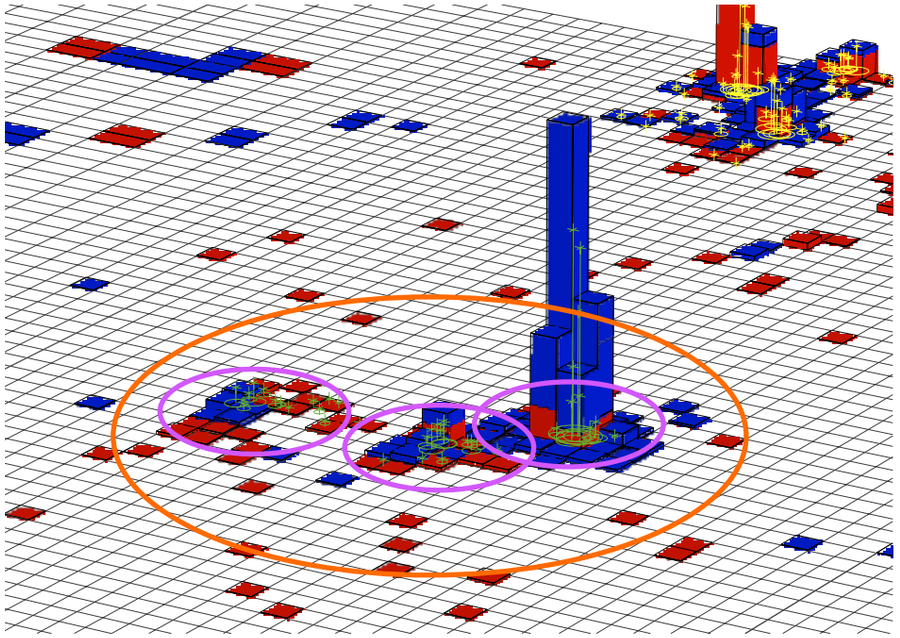


SUSY
Accidental
Substructure



Jet substructure

- The internal structure of the jet constituents give us handles to discriminate between:
  - Noise jets
  - Quark jets
  - Gluon jets
  - b-jets
  - W-jets
  - Z-jets
  - t-jets
  - H-jets
  - PU jets



Noise jet ID

CMS AN-15-269
Magda Diamantopoulou, Niki Saoulidou, Eirini Tziaferi

- Reject jets originating from noise, poor reconstruction, etc.
  - Remove noise at the object level
  - Necessary for most analyses
  - Complementary to “MET filters” (hit level noise rejection)
- Three working points:
  - Loose/tight: reject calorimetric noise
  - Tight lepton veto: reject mis-reconstructed leptons, effectively also rejecting jets clustered from a lepton (resolves ambiguity)
- Utilize PF jet energy fraction and particle multiplicities

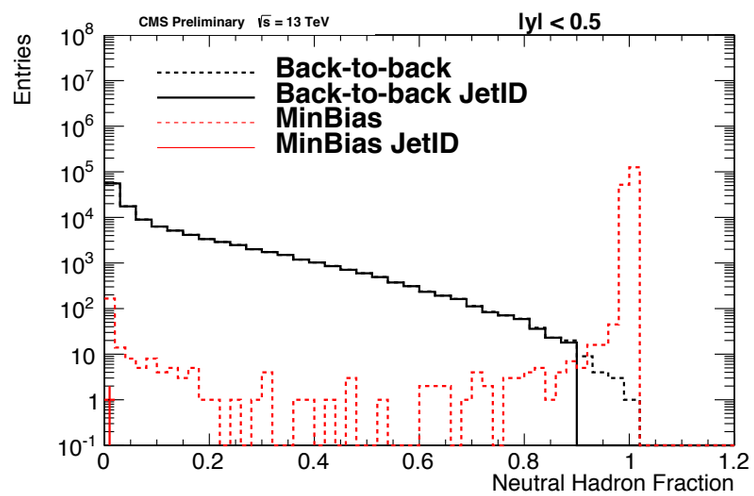
Jet Variables	|η| <2.7			|η| : 2.7-3	|η| : 3-5
	Loose PFJetID	Tight PFJetID	Tight Lepton Veto		
Charged Hadron Fraction	>0.0	>0.0	>0.0		
Neutral Hadron Fraction	<0.99	<0.9	<0.9		
Charged Multiplicity	>0.0	>0.0	>0.0		
Charged EM Fraction	<0.99	<0.99	<0.9		
Neutral EM Fraction	<0.99	<0.9	<0.9	<0.9	<0.9
Muon Fraction			<0.8		
Neutral Multiplicity				>2	>10

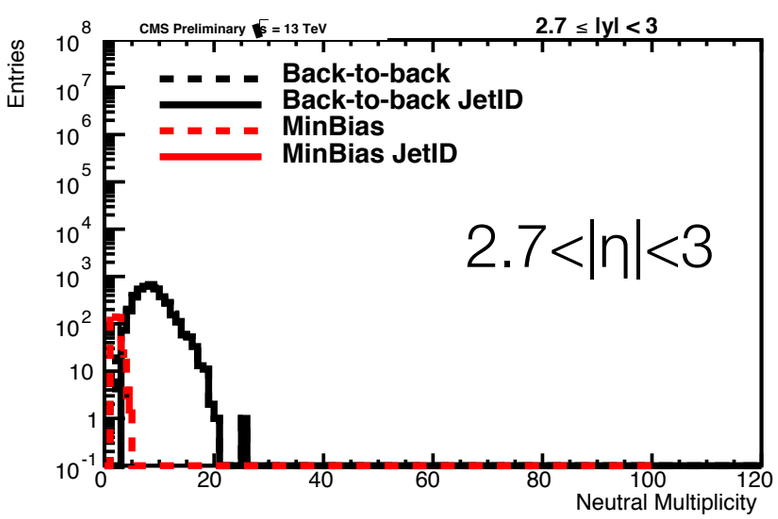
Noise jet ID

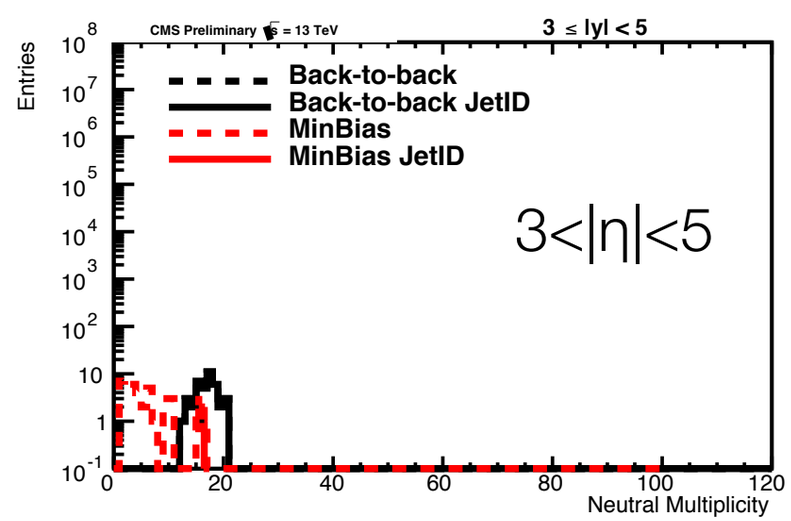
Jet Variables	|η| <2.7			|η| : 2.7-3	|η| : 3-5
	Loose PFJetID	Tight PFJetID	Tight Lepton Veto		
Charged Hadron Fraction	>0.0	>0.0	>0.0		
Neutral Hadron Fraction	<0.99	<0.9	<0.9		
Charged Multiplicity	>0.0	>0.0	>0.0		
Charged EM Fraction	<0.99	<0.99	<0.9		
Neutral EM Fraction	<0.99	<0.9	<0.9	<0.9	<0.9
Muon Fraction			<0.8		
Neutral Multiplicity				>2	>10

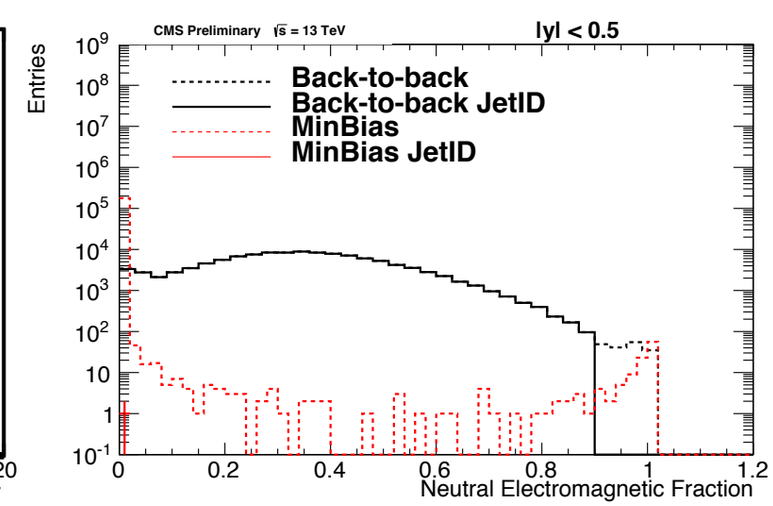
$\|\eta\|$ bin	Efficiency (%) of the PFJetID criteria	
	HLT_PFHT350	HLT_PFHT800
0 - 0.5	99.8160	99.7972
0.5 - 1.0	99.8919	99.7638
1.0 - 1.5	99.7604	99.7410
1.5 - 2.0	99.7498	99.6957
2.0 - 2.7	99.2339	99.1816
2.7 - 3.0	98.5816	96.4699
3.0 - 5.0	99.7665	100.0

$\|\eta\|$ bin	Background rejection (%)
0 - 0.5	99.9994
0.5 - 1.0	100.0
1.0 - 1.5	100.0
1.5 - 2.0	100.0
2.0 - 2.7	100.0
2.7 - 3.0	92.3513
3.0 - 5.0	92.8571



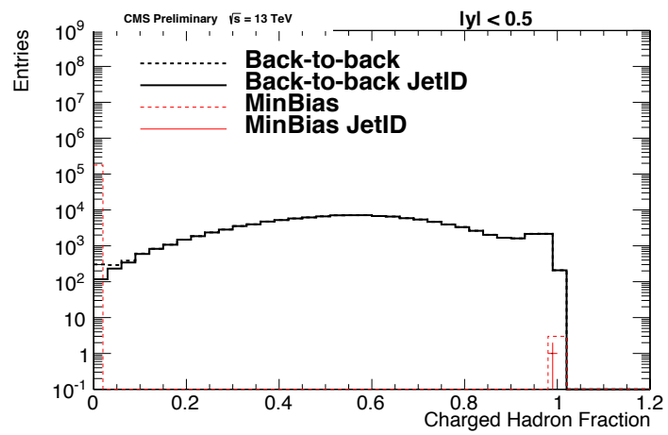


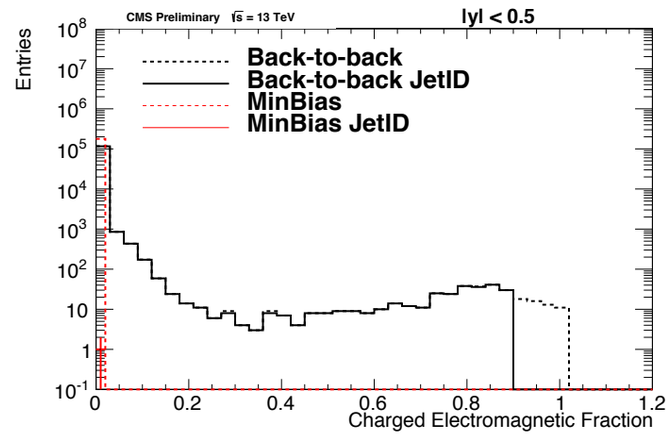


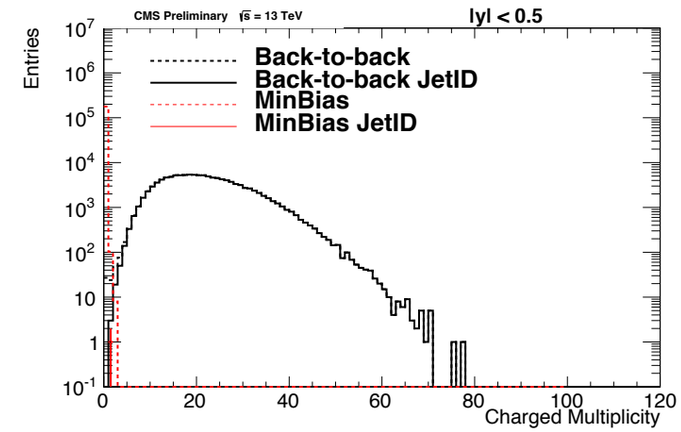


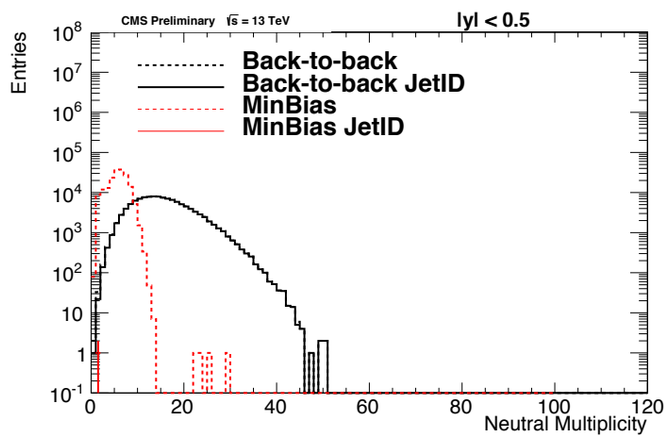
Noise jet ID

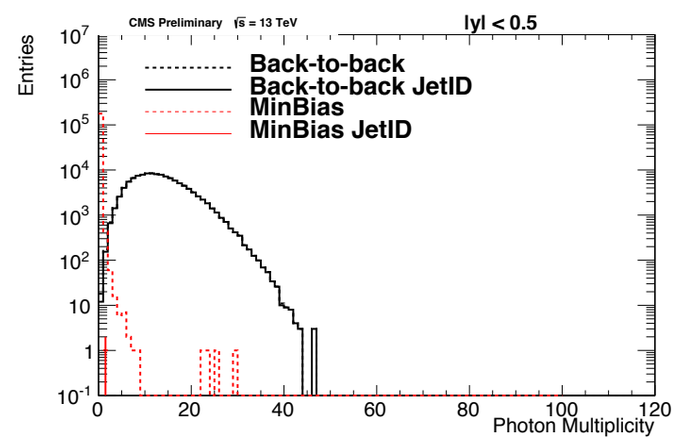
PAS

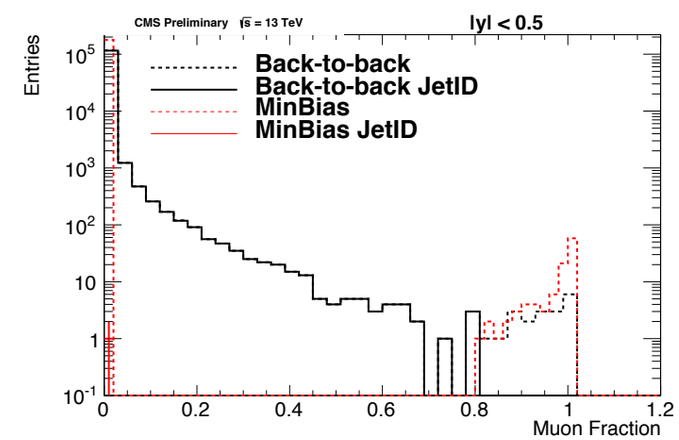


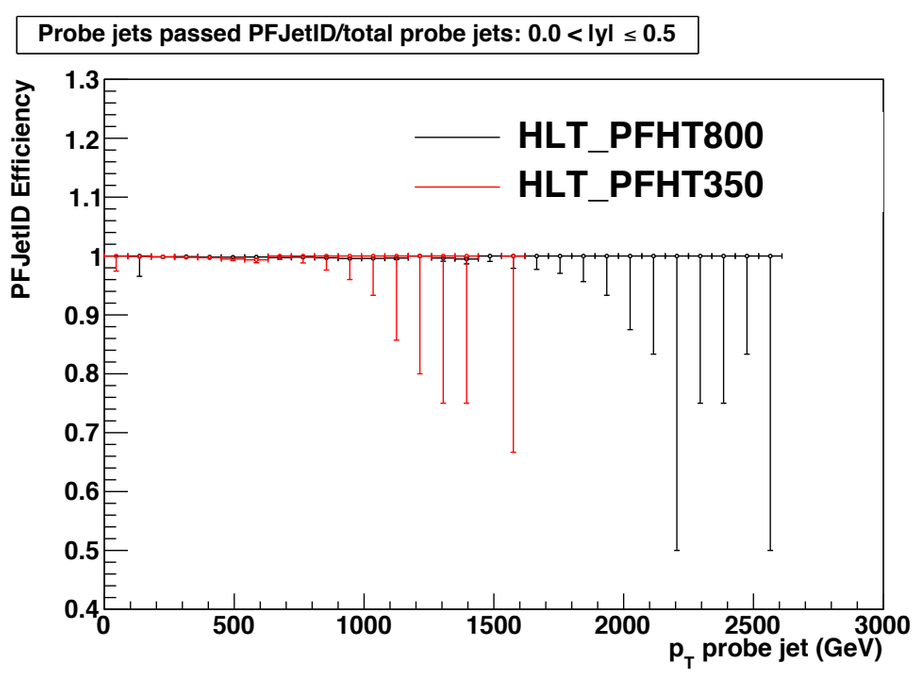


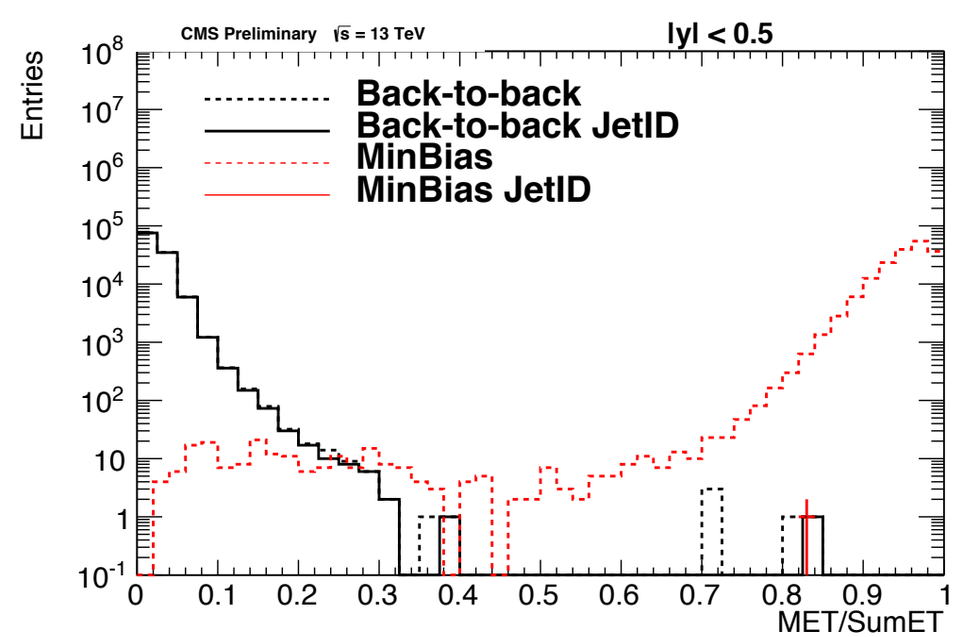




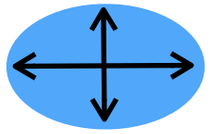


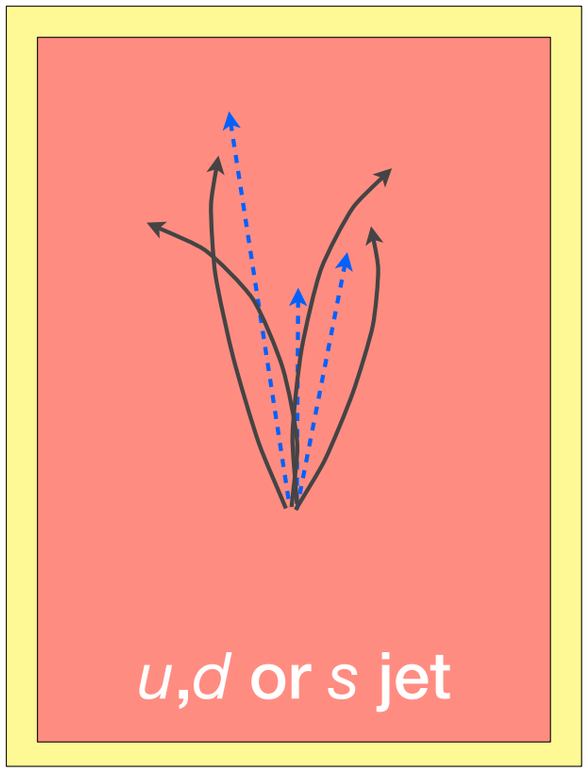


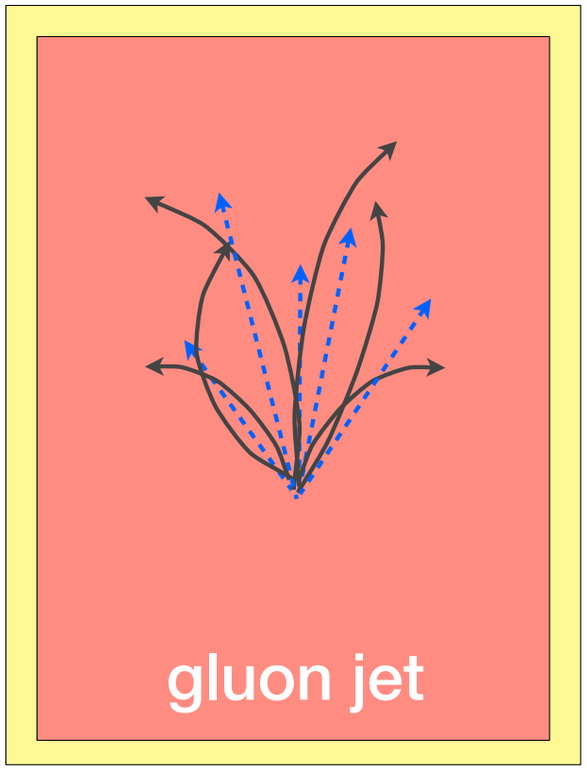


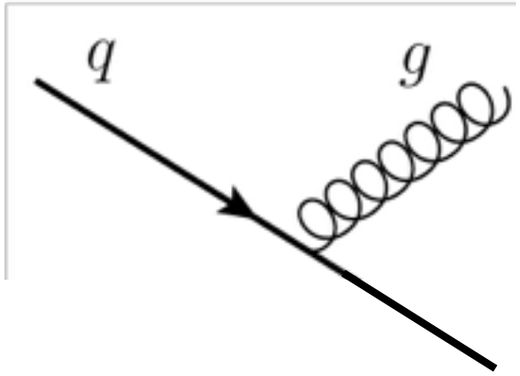


Quark/gluon discrimination

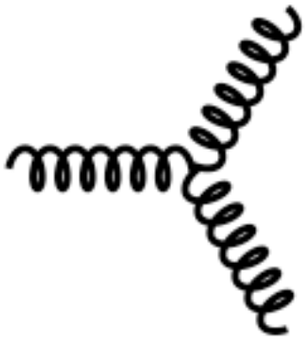
- Quarks and gluons have different QCD color factors
  - Gluon more likely to radiate a gluon
  - Gluon jets tend to be wider with larger multiplicities and correspondingly fewer hard particles
  - Quark jets tend to be narrow with smaller multiplicities and asymmetrical energy shared between constituents
- Discriminating quark/gluon jets
  - jet constituent multiplicity
  - jet shape (minor axis width) 
  - pTD (energy sharing)







$C_F = 4/3$



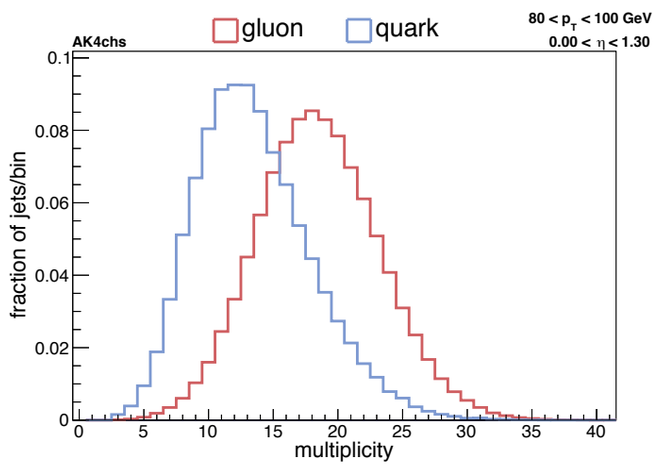
$C_A = 3$

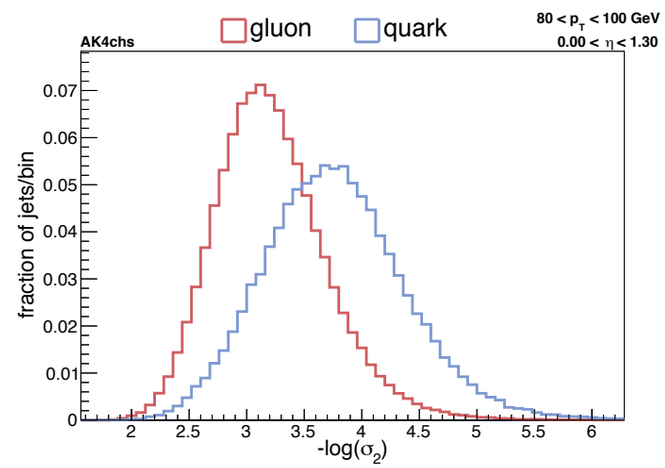
$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}}$$

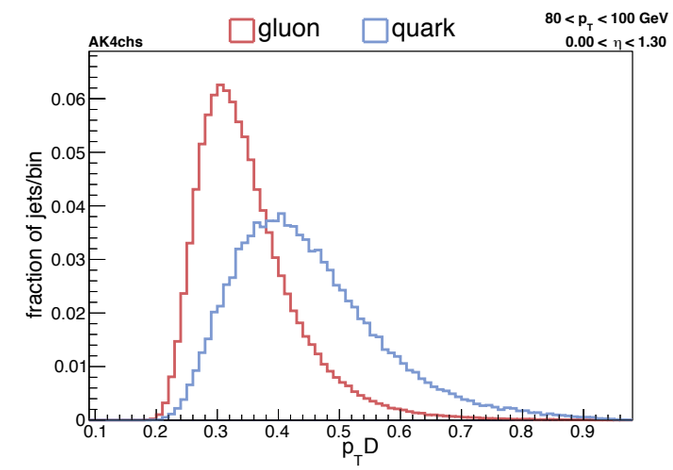
→ 1 if all momentum carried by one particle
→ 0 if jet has infinite number of particles

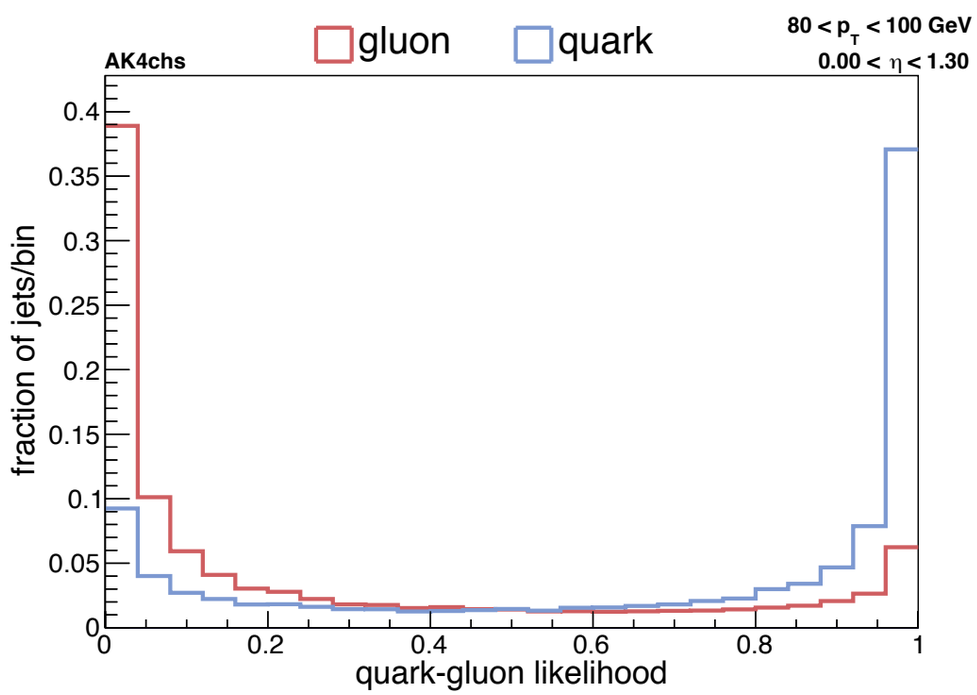
Quark/gluon jet discrimination

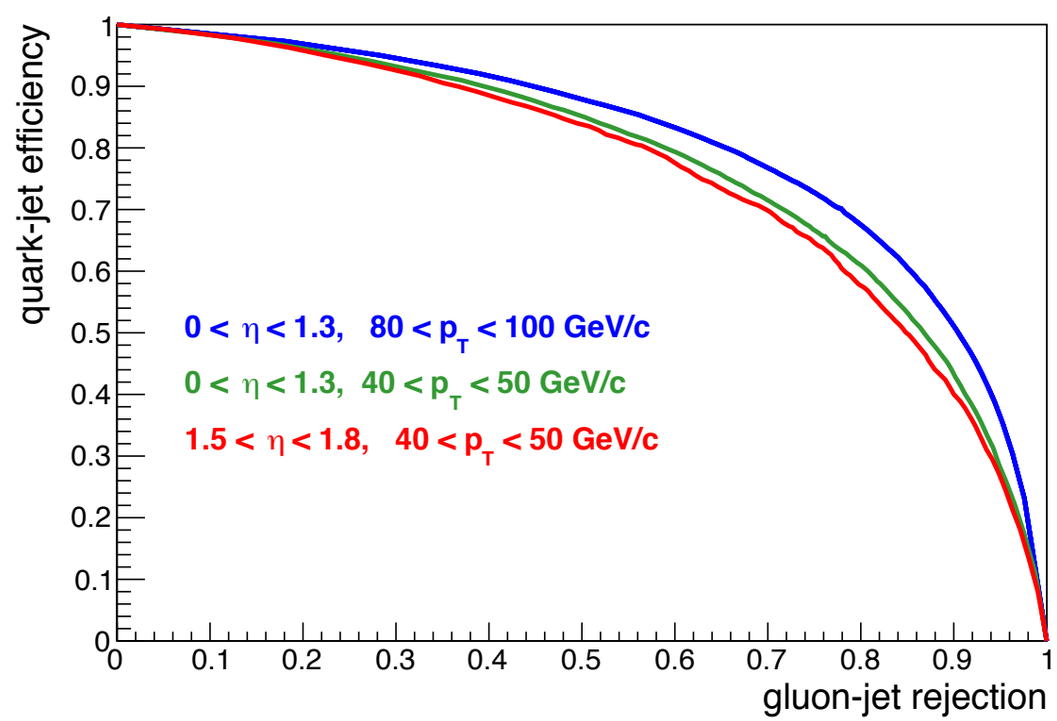
- See JME-13-002 and JME-16-003





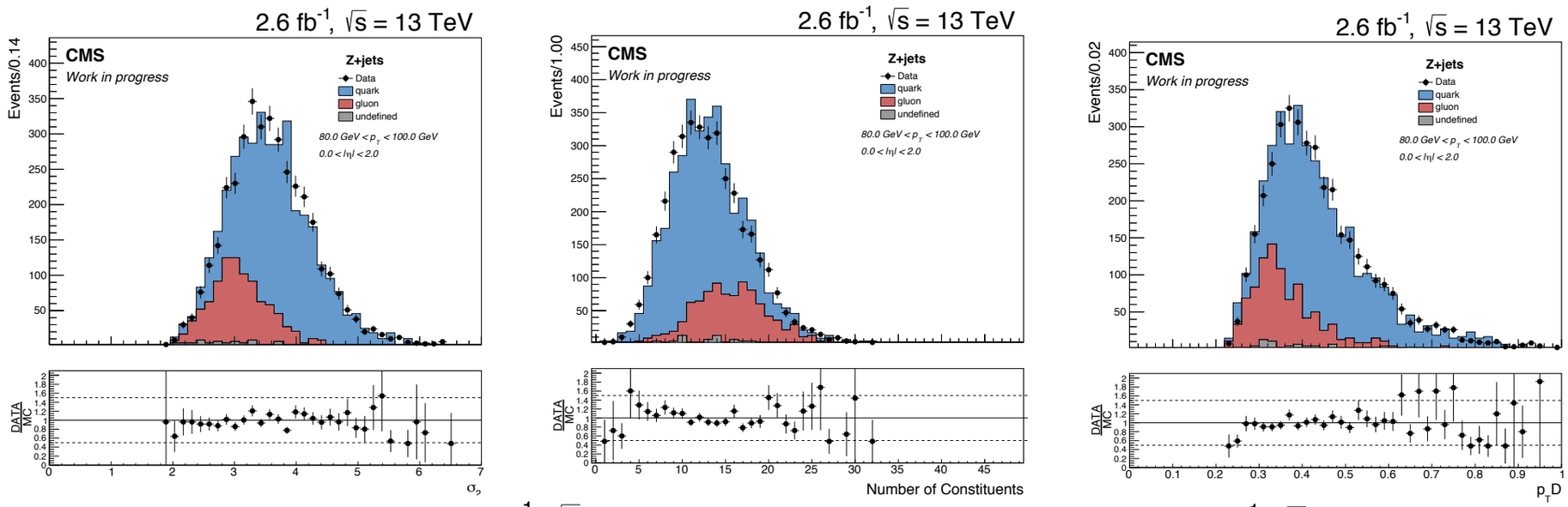


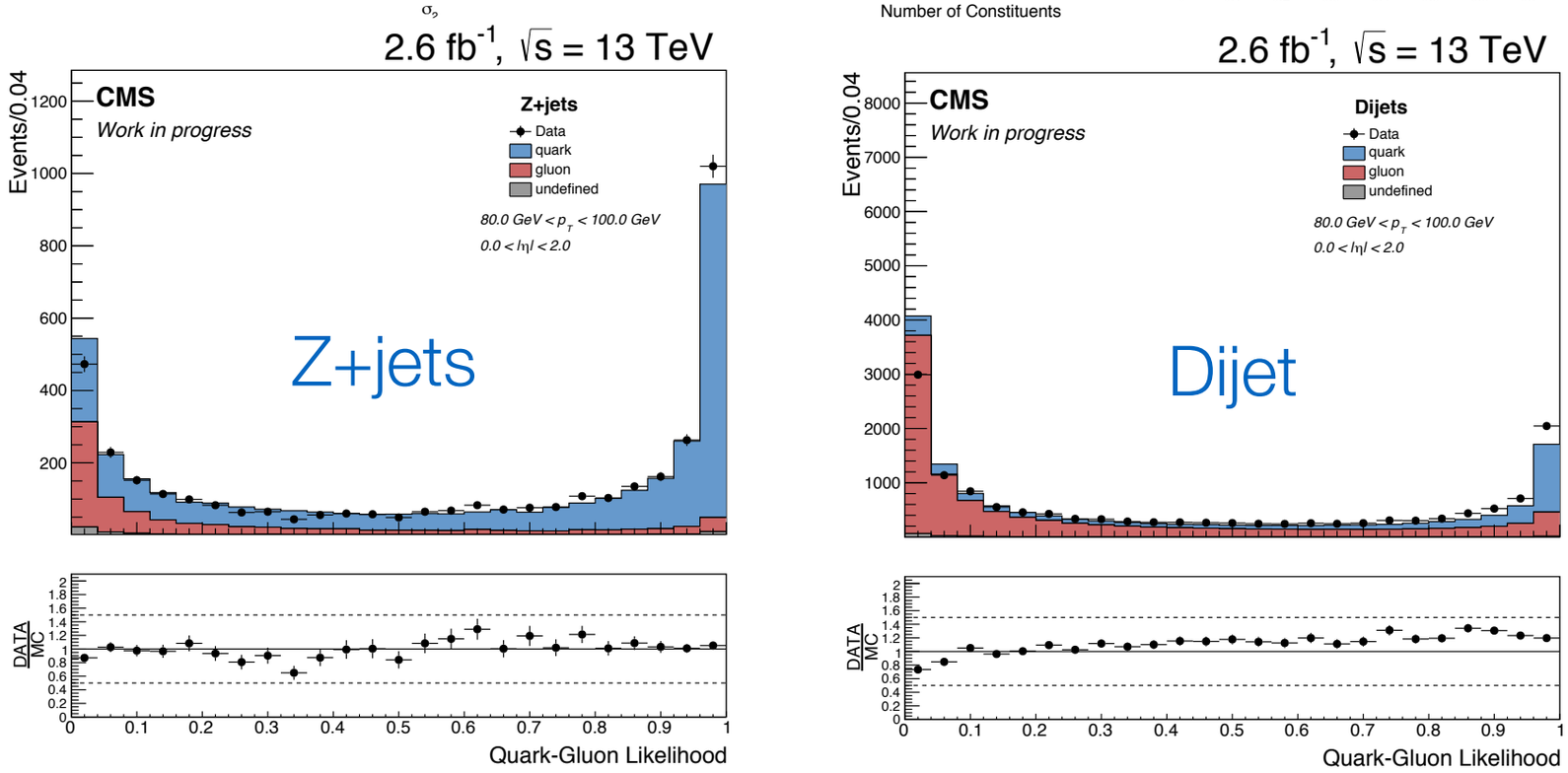




Quark/gluon validation

- Z+jets (quark enriched) & dijet (gluon enriched) selections

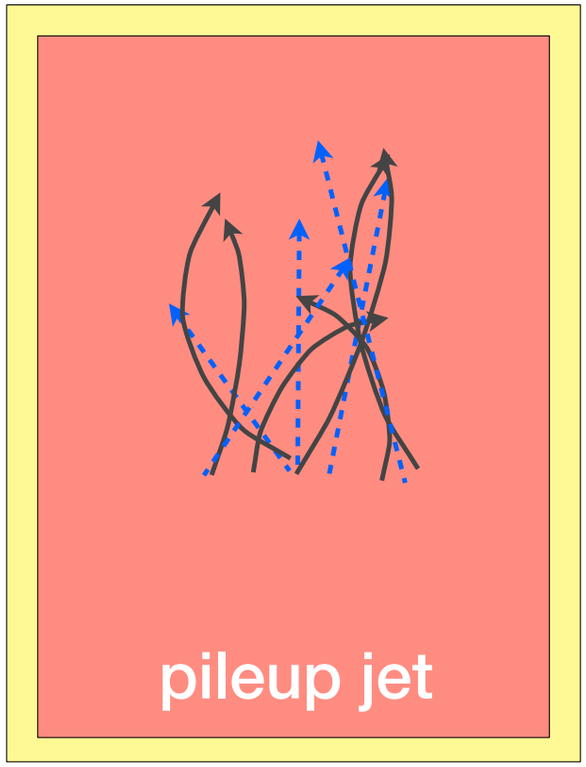


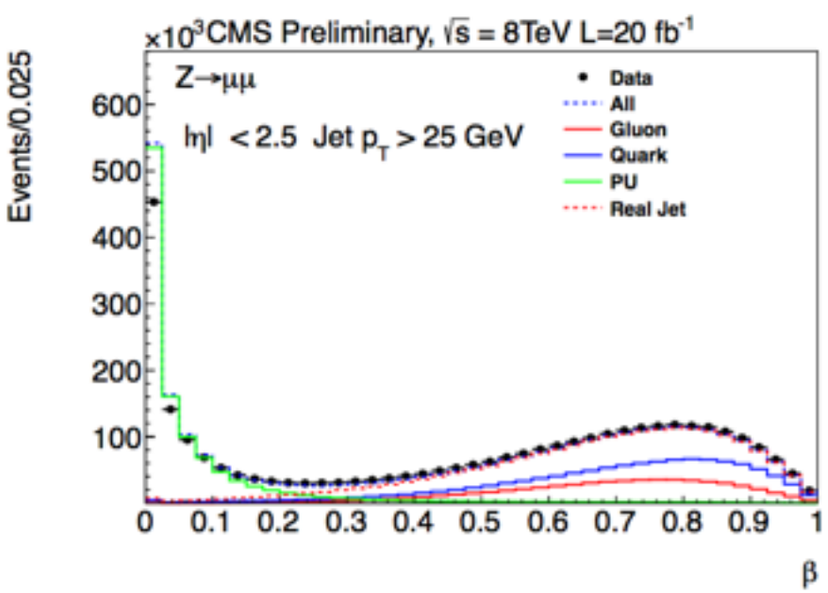


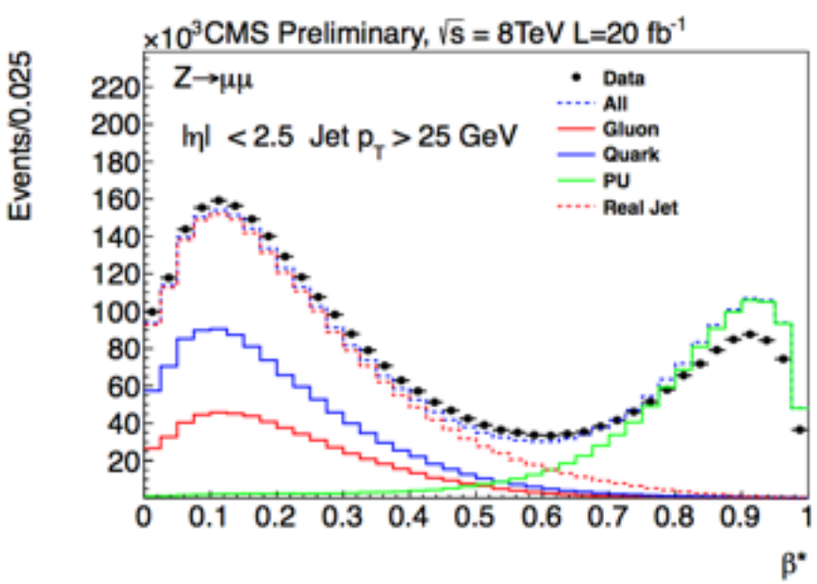
Pileup jets

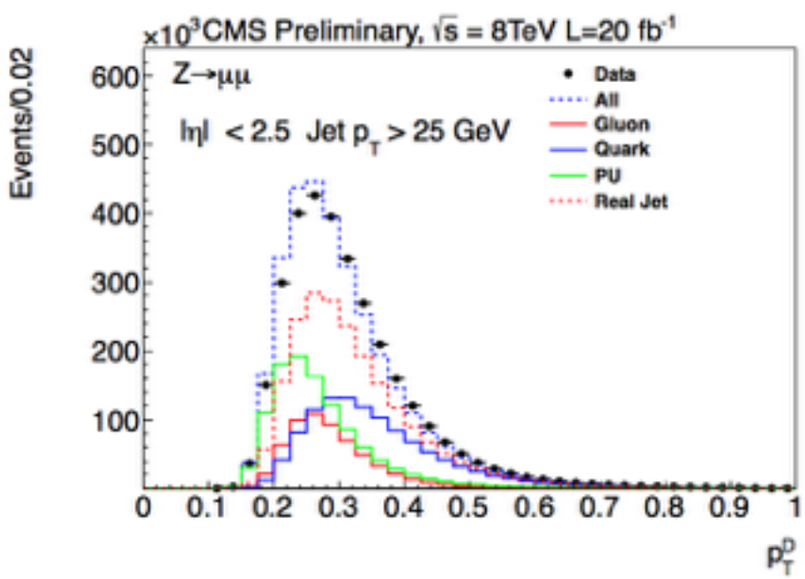
- Many overlapping soft particles from PU can add up to make a jet of substantial pT
- Identifying pileup jets relies on charged PU vertex information and jet shape

$$\beta = \frac{\sum_{i \in PV} p_{Ti}}{\sum_i p_{Ti}} \qquad \beta^* = \frac{\sum_{i \in otherPV} p_{Ti}}{\sum_i p_{Ti}}$$



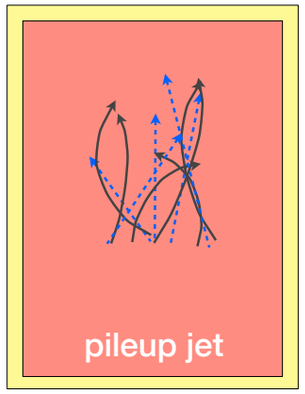




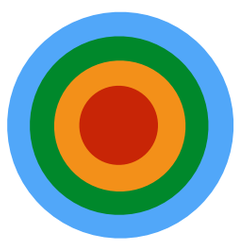


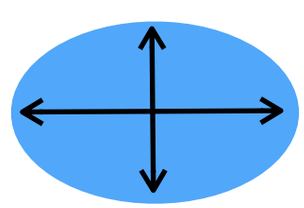
Pileup jet ID

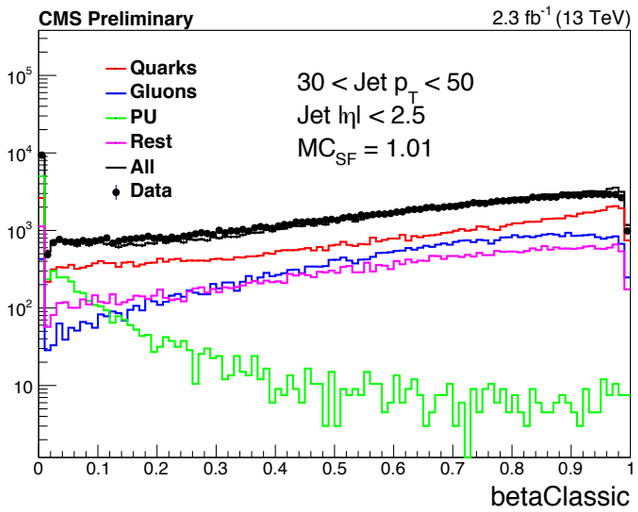
- Pileup jets originate from overlapping low energy pileup particles
- Pileup jet ID BDT: Identify and reject PU jets
  - Track variables→ pileup jet tracks are associated to non-primary vertices
  - Jet shape variables → pileup jets tend to be broad (pileup is randomly distributed)
- See: JME-13-005 and JME-16-003

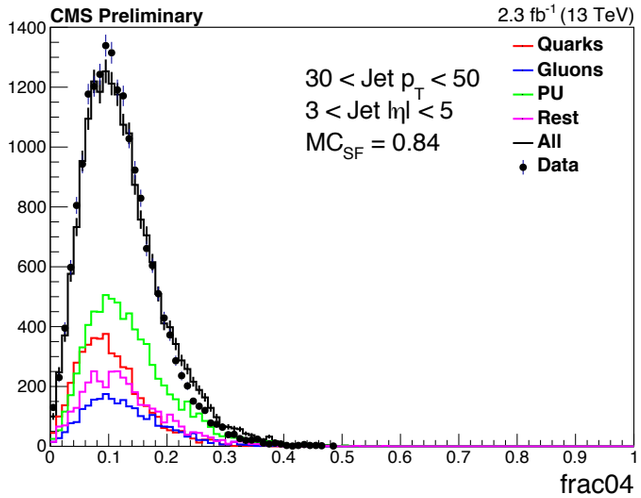


Tracking variables	Jet shape variables
$\beta = \frac{\sum_{i \in PV} p_{Ti}}{\sum_i p_{Ti}}$ $N_{vertices}$	$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R_i^2 p_{Ti}^2}{\sum_i p_{Ti}^2}$ f_{ring} (energy fraction within annulus) [0,0.1], [0.1,0.2], [0.2,0.3], [0.3,0.4] p_T leading constituent / p_T jet Total and charged multiplicity Major/minor axis p_T^D Pull $\quad$ Pull Vector $\vec{t} = \sum_{i \in \text{jet}} \frac{p_T^i \lvert r_i \rvert}{p_T^{jet}} \vec{r}_i$





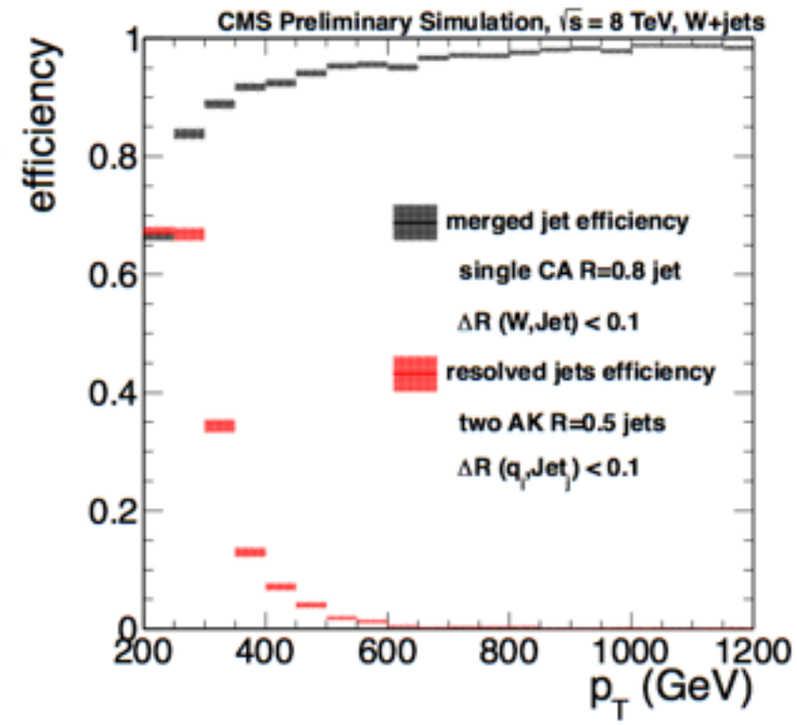




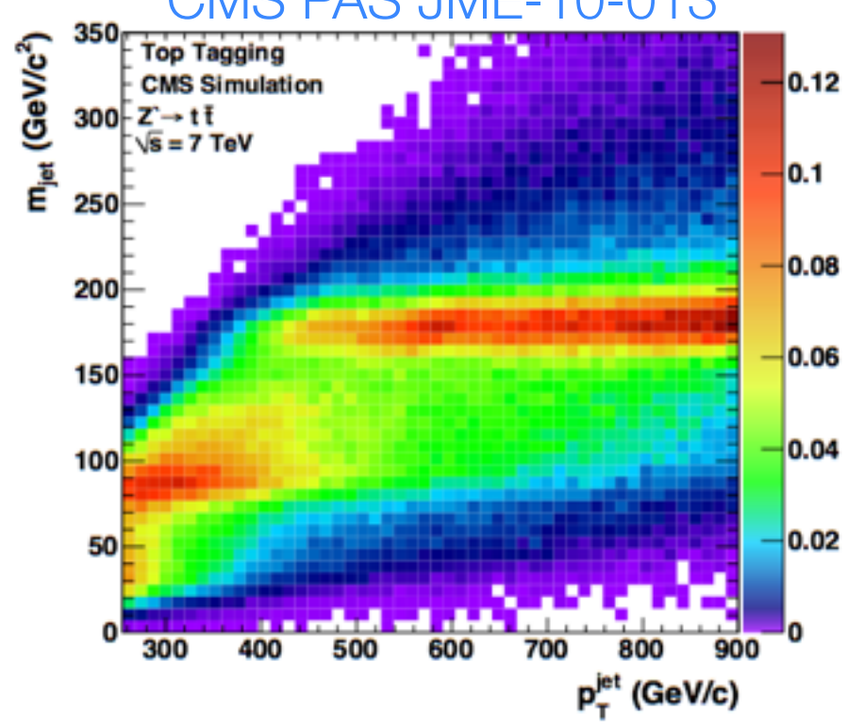
Boosted Object Examples

- W/Z
  - Above 200 GeV the efficiency to resolve two AK5 jets from a hadronic W decay decreases quickly
  - Most W/Zs are fully merged within a CA8 jet in this pt range
- Top
  - CA8 jet
    - pT>400 majority are merged
    - 250<pT<400 partially merged
- Higgs
  - CA8
    - pT>300 majority are merged
  - CA15
    - pT>170 majority are merged
- Accessible via jet substructure
  - Jet contains lots of information!

CMS PAS JME-13-006

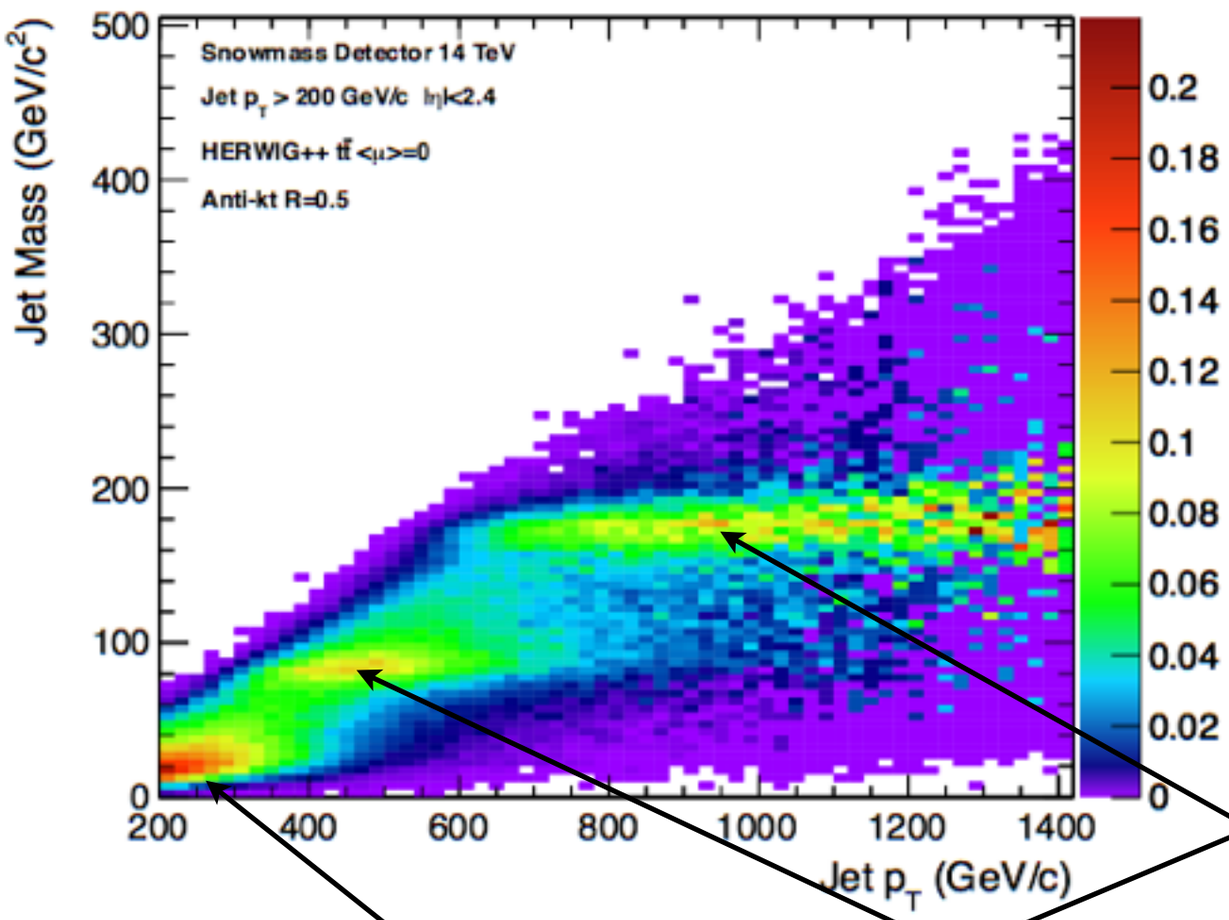


CMS PAS JME-10-013

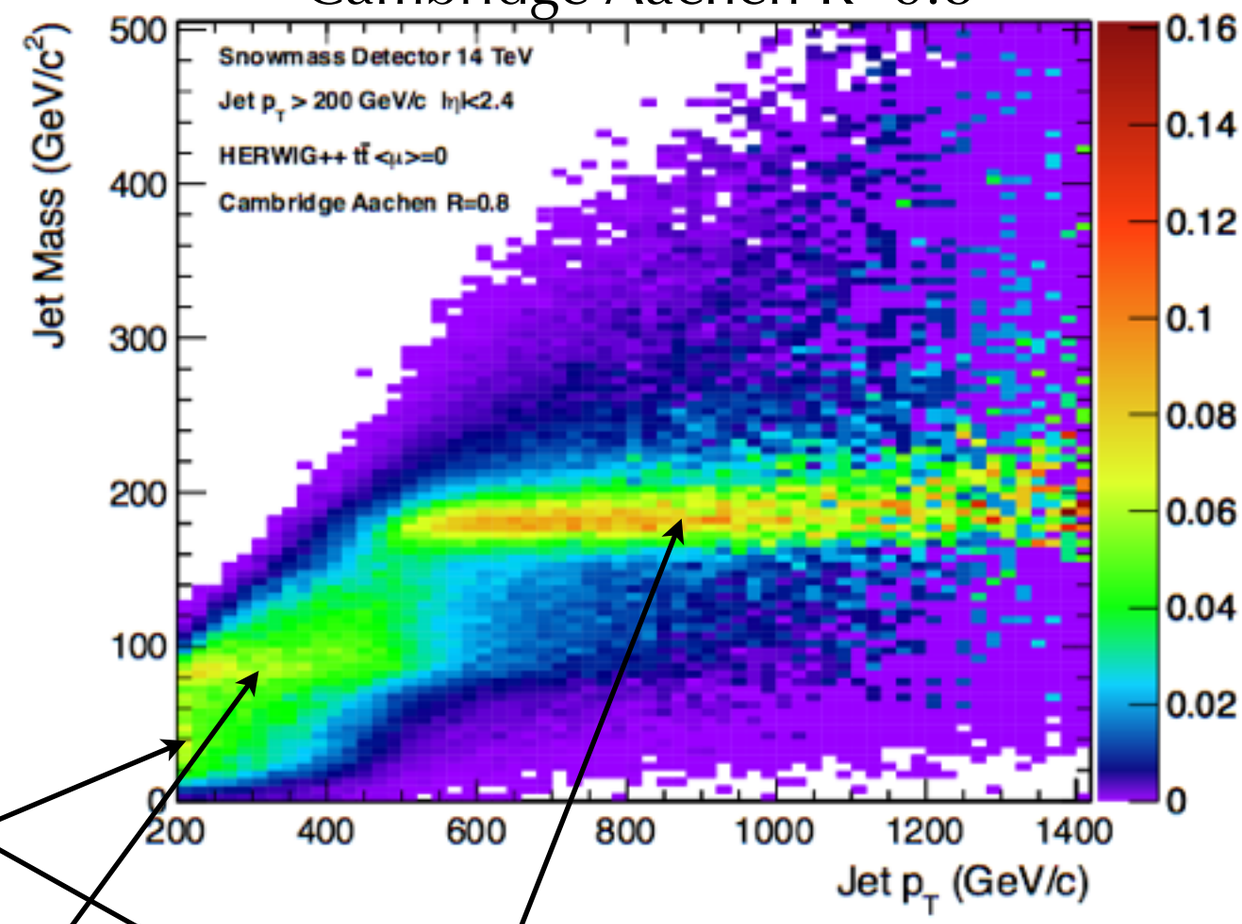


AK5 and CA8 comparison

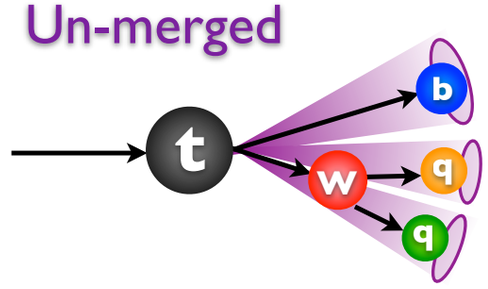
Anti-kt R=0.5



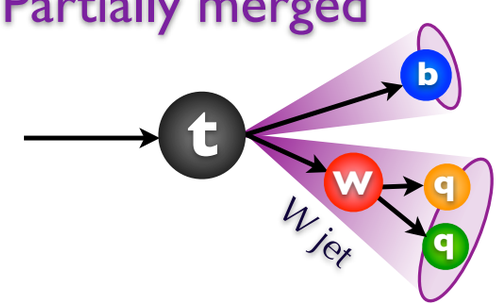
Cambridge Aachen R=0.8



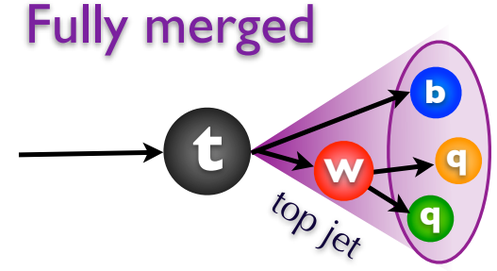
Un-merged



Partially merged



Fully merged

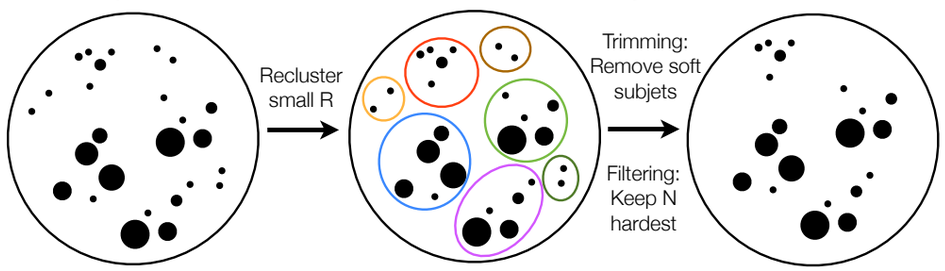


Jet grooming

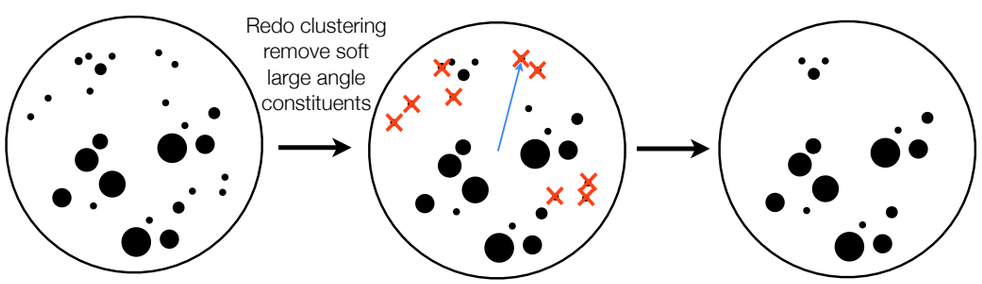
Algorithmic jet substructure techniques designed to remove isolated soft radiation in jets (contamination from ISR, UE, pileup)

J. Dolen

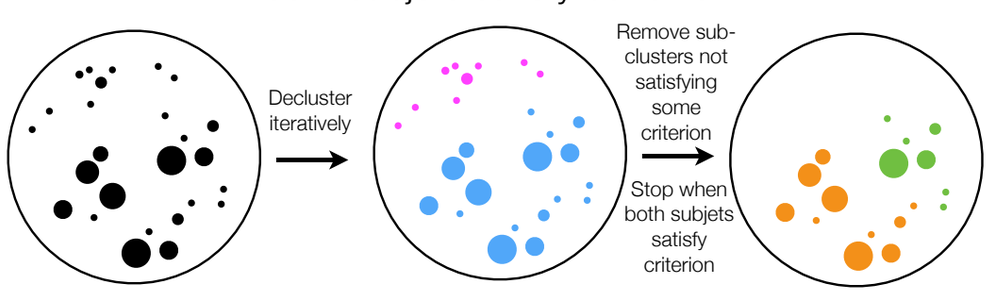
Trimming, Filtering - Recluster jet with smaller distance parameter. Condition based subjet removal.

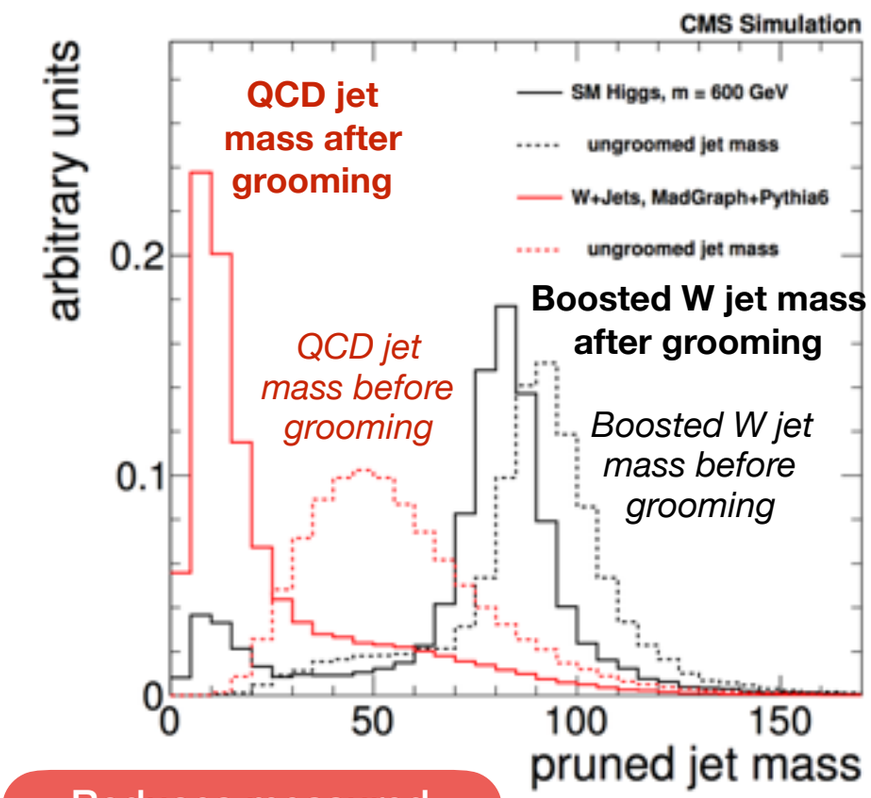


Pruning - Recluster jet. Remove soft large angle particles.



BDRS, MMDT, Soft Drop, JHU top tagger, CMSTT - Recursively decluster jet. Remove sub-clusters not satisfying algorithm condition. Stop declustering when both subjets satisfy condition.

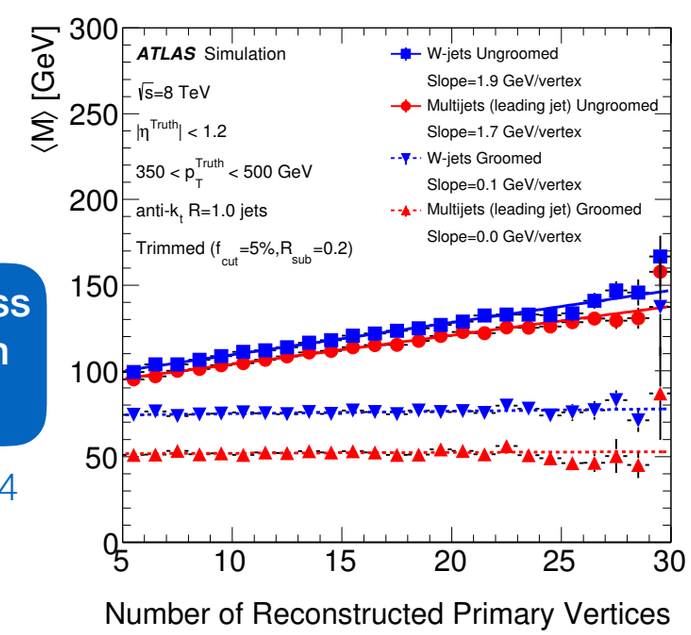




CMS HIG-13-008
H → WW → lνqq

Improved jet mass resolution for boosted heavy object

Reduces measured QCD jet mass (improves discrimination)



Reduces jet mass dependence on pileup

ATLAS CERN-PH-EP-2015-204
arxiv:1510.05821

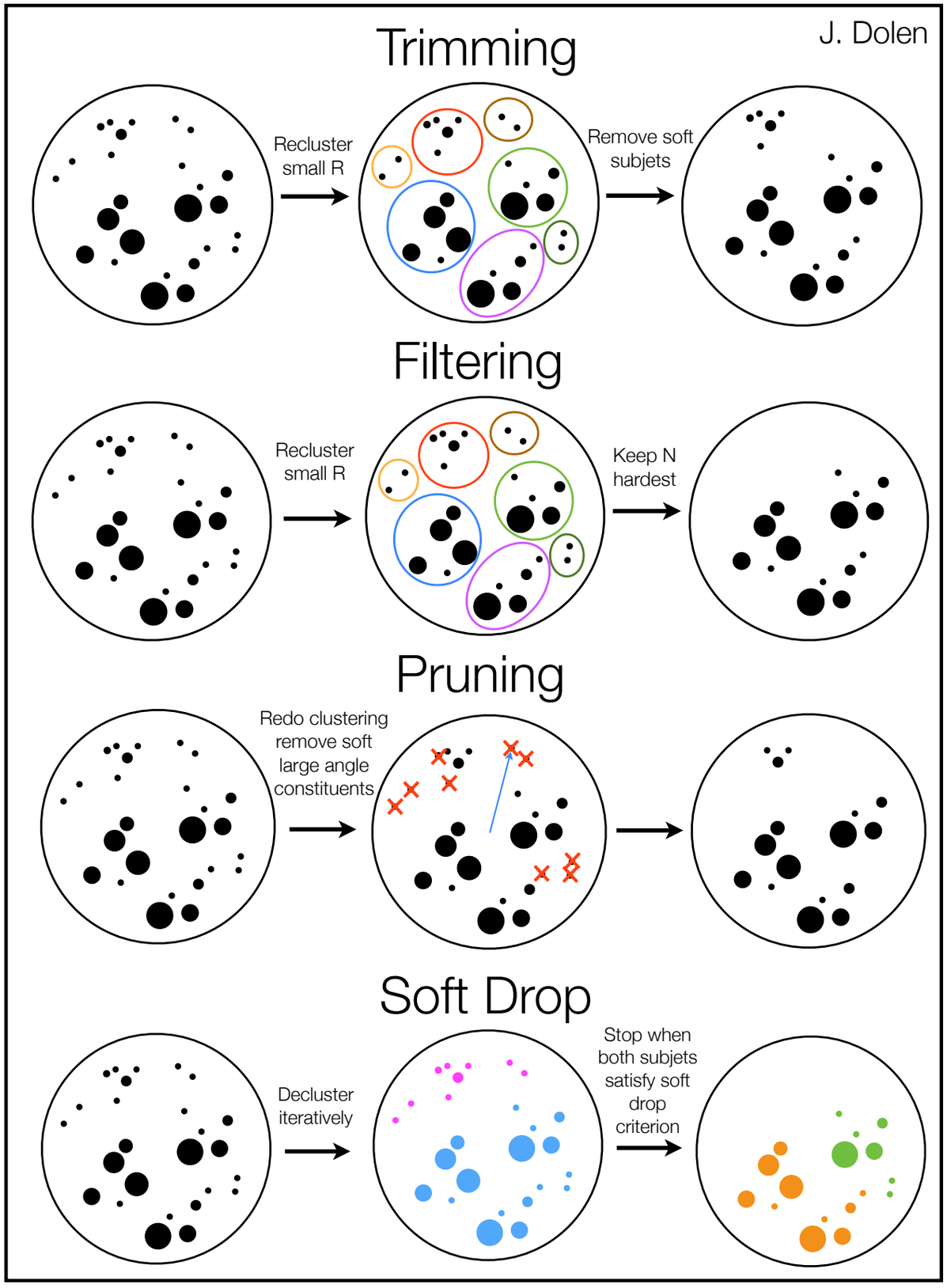
Jet grooming

- Trim - recluster with small distance parameter r_{trim}. Keep subjets which satisfy momentum fraction.
- Filter - recluster with small distance parameter r_{filt}. Keep n_{filt} subjets.
- Prune - recluster. Throw out subjets which are too soft ($z<z_{cut}$) or at large angle ($\Delta R_{12}>D_{cut}$)

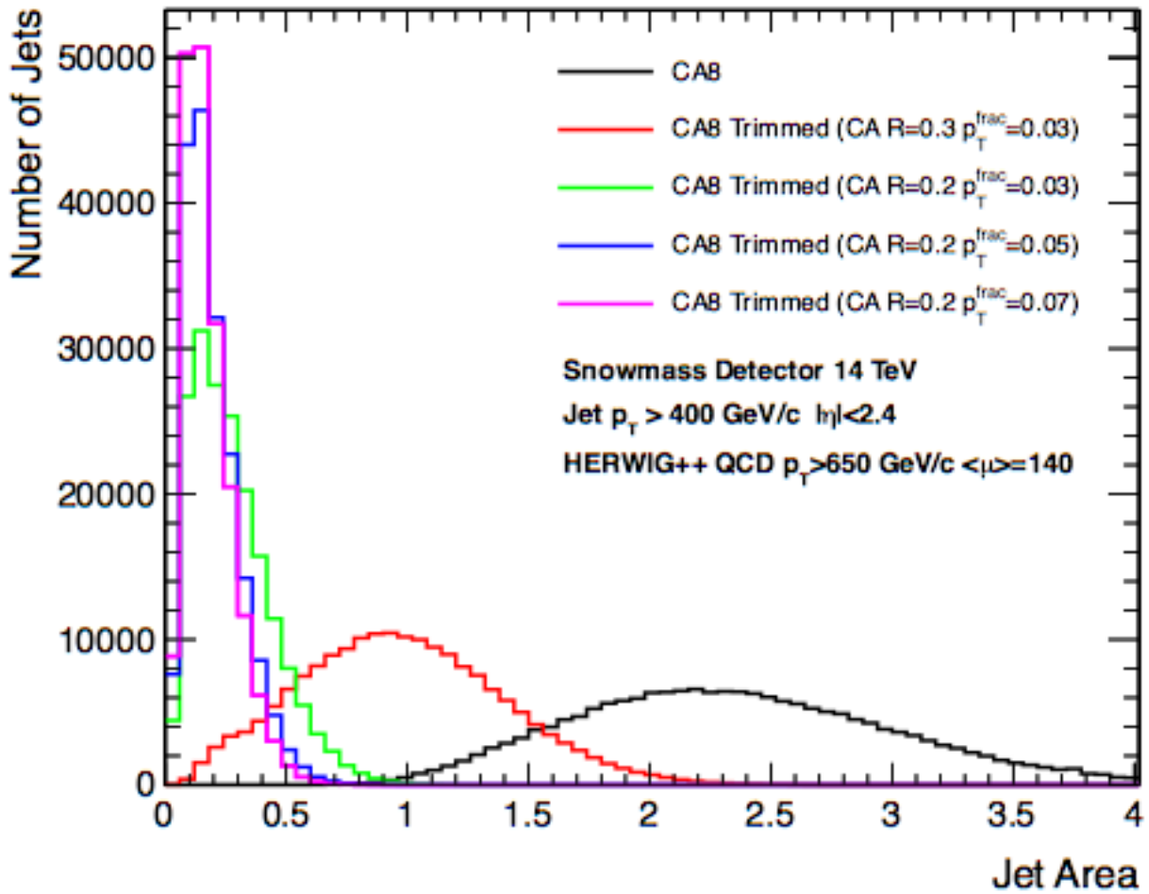
$$z \equiv \min(p_{T_1}, p_{T_2})/p_{T_p}$$

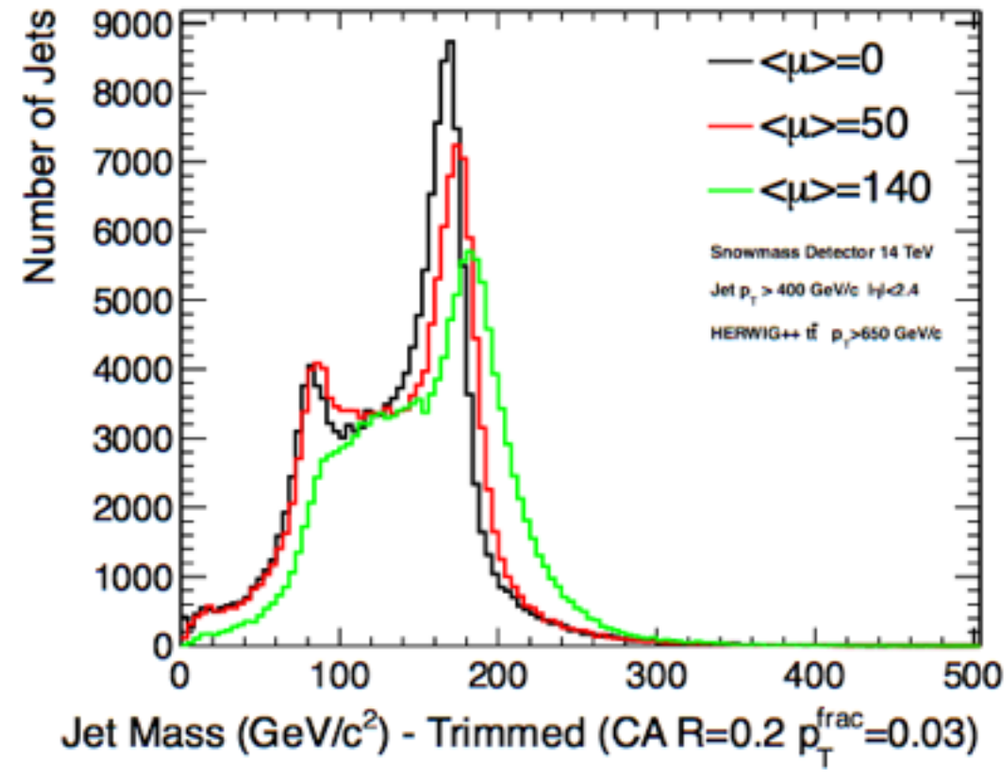
- Soft drop - decluster. Throw out subjets which do not satisfy the soft drop condition. Of $\beta>0$ larger angle subjets have a larger p_T fraction requirement.

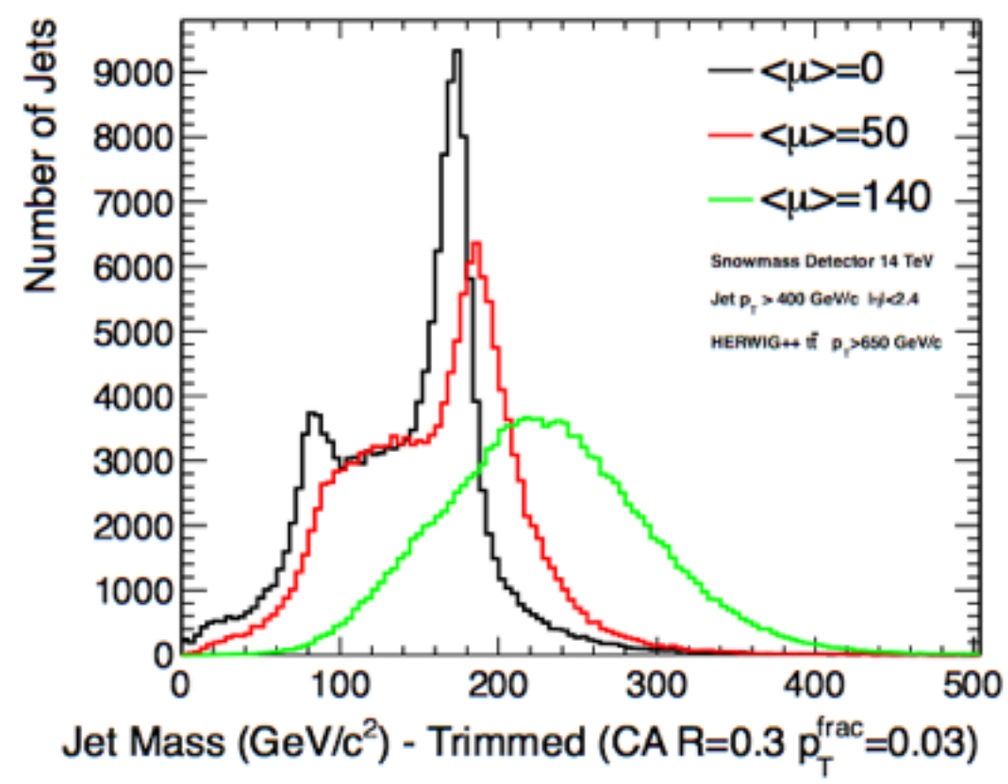
$$\text{Soft Drop Condition:} \quad \frac{\min(p_{T1}, p_{T2})}{p_{T1}+p_{T2}} > z_{\rm cut}\left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$



Grooming parameters



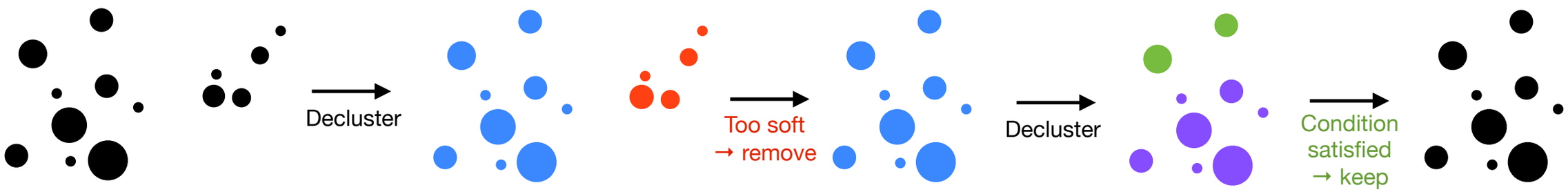




Soft drop

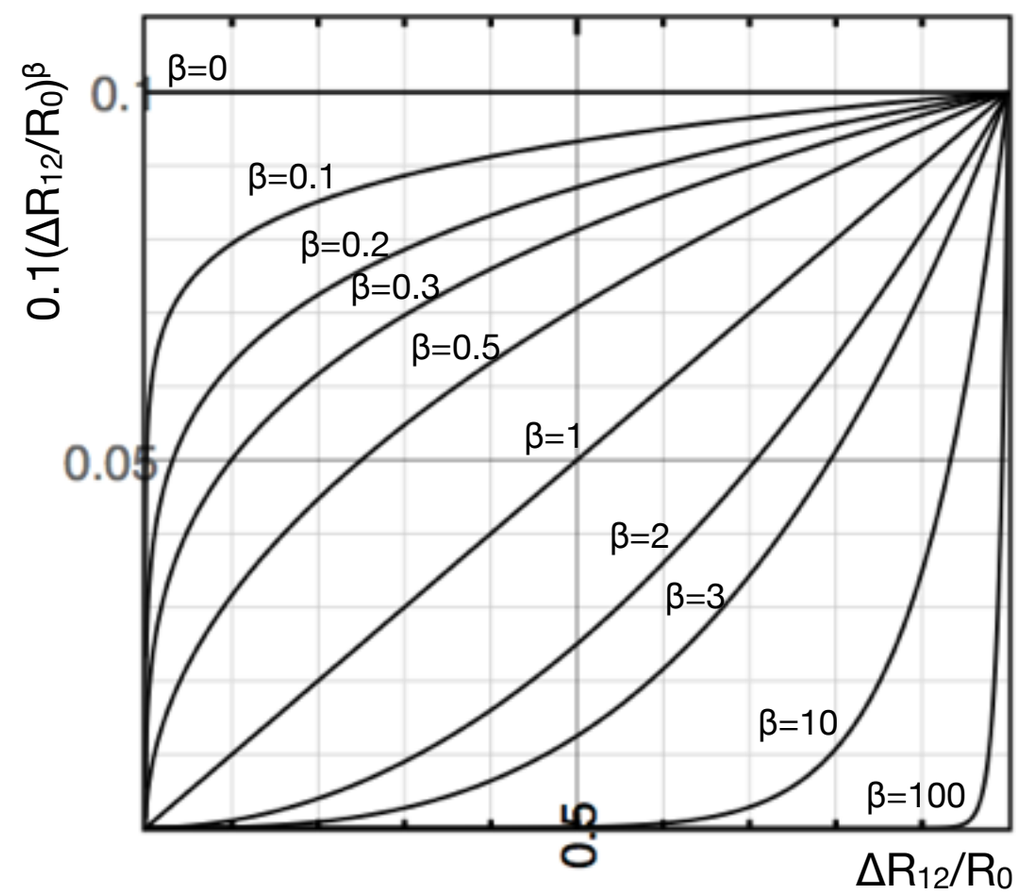
- Recursively decluster jet. Remove the softer component unless the soft drop condition is satisfied.

$$\text{Soft Drop Condition:} \quad \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$



J. Dolen

- Soft wide angle radiation fails the condition
  - As z_{cut} ⇑ ⇒ more aggressive grooming
  - As β ⇓ ⇒ more aggressive grooming
- Example (z_{cut} = 0.1) :
  - If β = 0, remove softer subjet if p_T fraction < 0.1 (~equivalent to MMDT)
  - If β > 0, remove softer subjet if p_T fraction < x, where x increases with ΔR and has maximum value 0.1
  - β → ∞ no grooming
  - β <0 soft drop becomes a tagger instead of a groomer (finds jets with hard, large angle subjets)



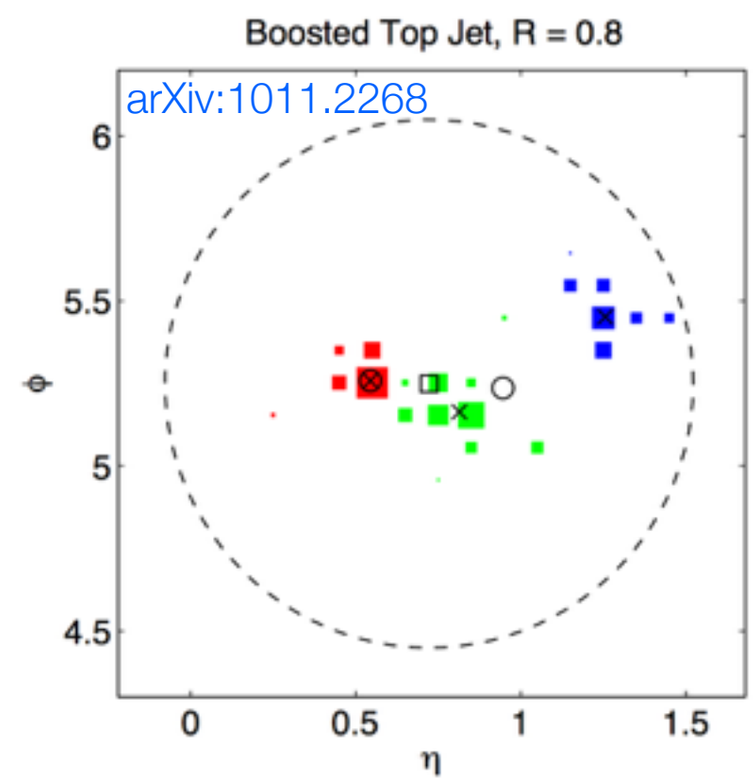
N-subjettiness

J. Thaler, K Van Tilburg
arXiv:1011.2268

- Jet shape variable which determines how consistent a jet is with having N or fewer subjets
  - $\tau_N \rightarrow 0$ if jet has N or fewer subjets (energy spread is close to the subjet axes)

$$\tau_N = \frac{\sum_{i=1}^{n_{\text{constituents}}} p_{\text{T},i} \min\{\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i}\}}{\sum_{i=1}^{n_{\text{constituents}}} p_{\text{T},i} R}$$

- More discrimination by using ratios (τ_3/τ_2 for top tagging, τ_2/τ_1 for W tagging, etc.)
- Need to find the subjet axis before calculating N-subjettiness
  - Multiple options:
    - kt axes
    - One-pass optimization (CMS default)
    - etc.



Energy Correlation Functions

$$\mathrm{ECF}(0,\beta) = 1,$$
$$\mathrm{ECF}(1,\beta) = \sum_{i\in J} p_{T\,i},$$
$$\mathrm{ECF}(2,\beta) = \sum_{i<j\in J} p_{T\,i}\, p_{T\,j} (R_{ij})^{\beta},$$
$$\mathrm{ECF}(3,\beta) = \sum_{i<j<k\in J} p_{T\,i}\, p_{T\,j}\, p_{T\,k}\, (R_{ij} R_{ik} R_{jk})^{\beta},$$
$$\mathrm{ECF}(4,\beta) = \sum_{i<j<k<\ell\in J} p_{T\,i}\, p_{T\,j}\, p_{T\,k}\, p_{T\,\ell}\, (R_{ij} R_{ik} R_{i\ell} R_{jk} R_{j\ell} R_{k\ell})^{\beta}$$

Jet constituent based observables sensitive to N subjet substructure

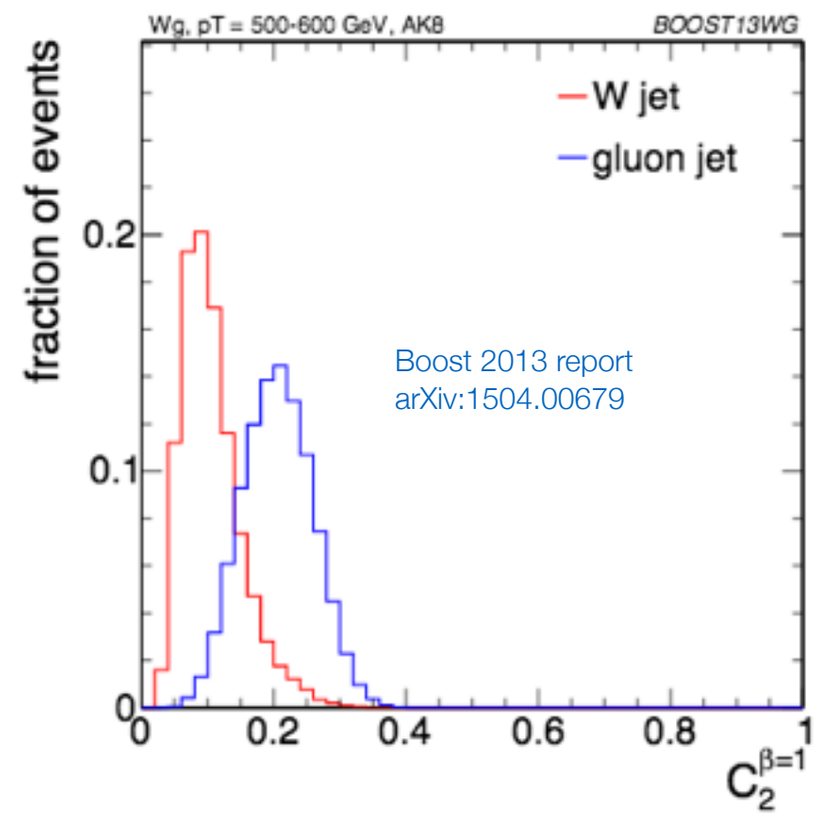
ECF ratios (C2,D2, etc.) for boosted object discrimination (Small C2 $\Longrightarrow$ 2 subjets)

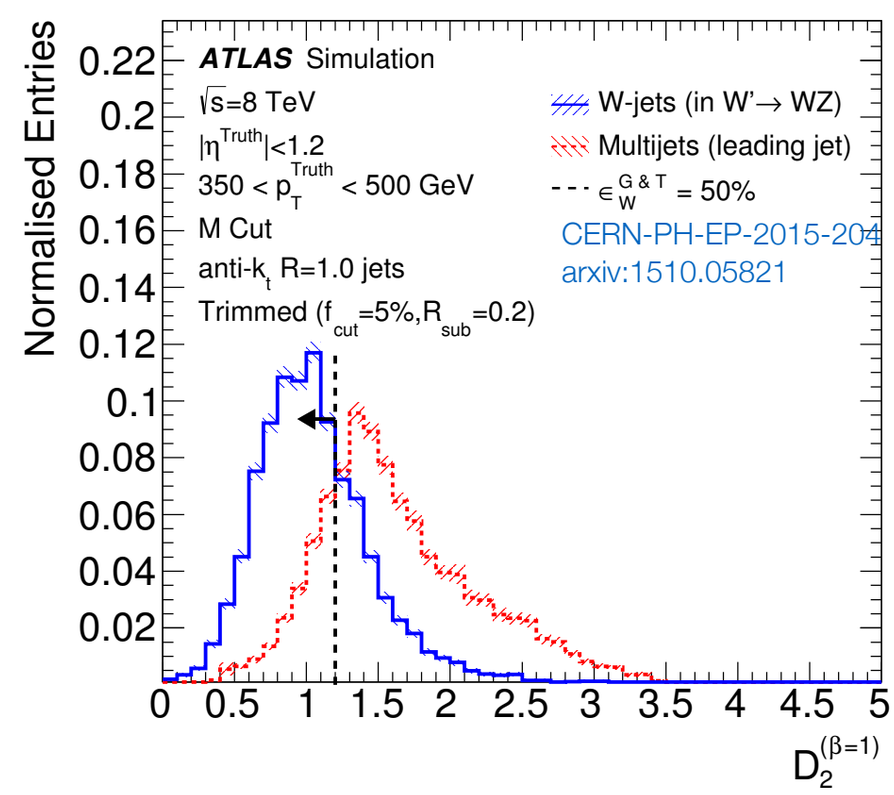
β parameter allows access to different angular scales

$$C_N^{(\beta)} = \frac{\mathrm{ECF}(N+1,\beta)\,\mathrm{ECF}(N-1,\beta)}{\mathrm{ECF}(N,\beta)^2}$$

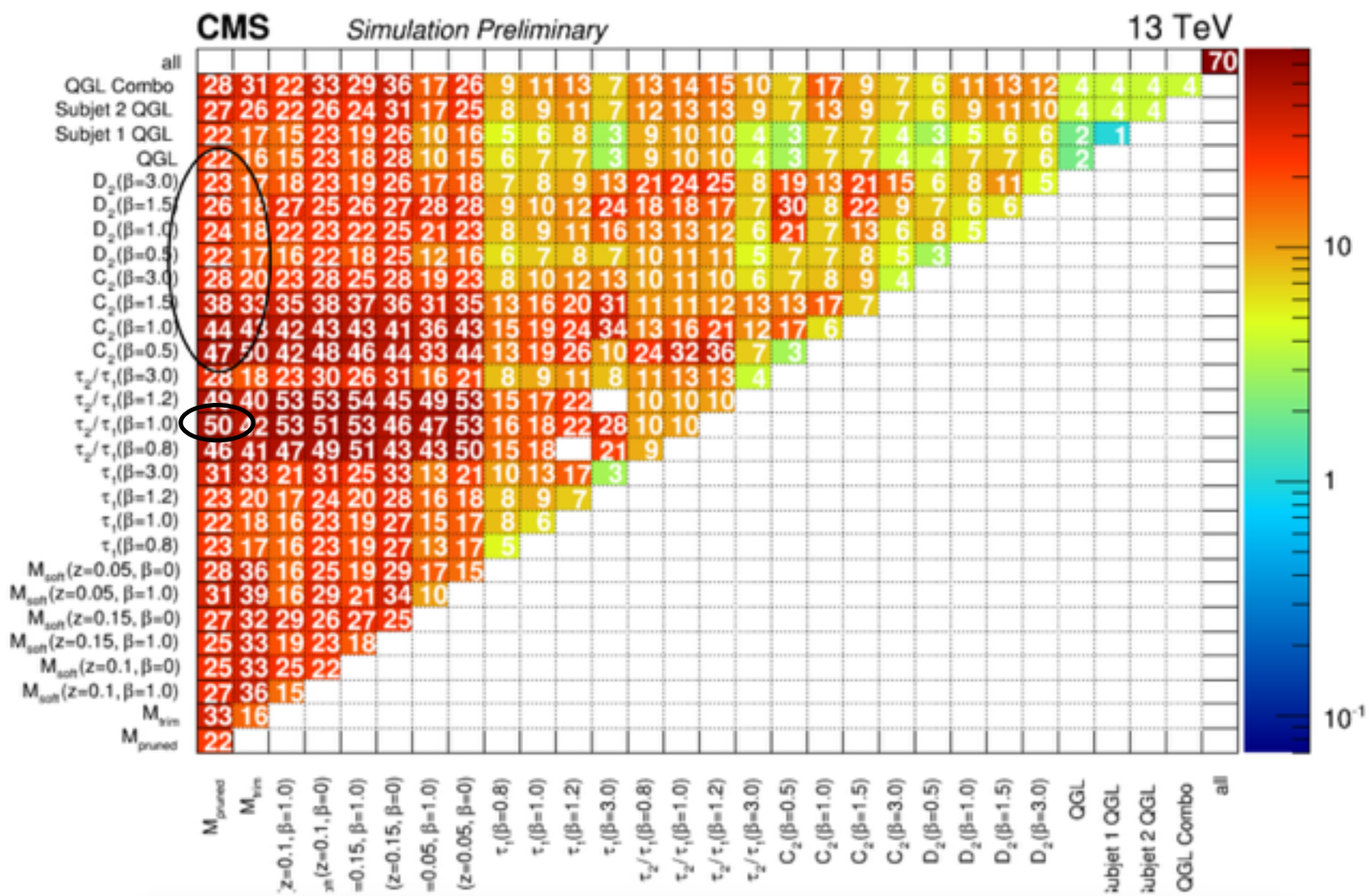
$$D_2(\beta) = \frac{\mathrm{ECF}(3,\beta)\mathrm{ECF}(1,\beta)^3}{\mathrm{ECF}(2,\beta)^3}$$

A. Banfi, G. P. Salam, and G. Zanderighi, JHEP 0503 (2005) 073
M. Jankowiak and A. J. Larkoski JHEP 1106 (2011) 057
A. Larkoski, G. Salam, J. Thaler JHEP 2013:108





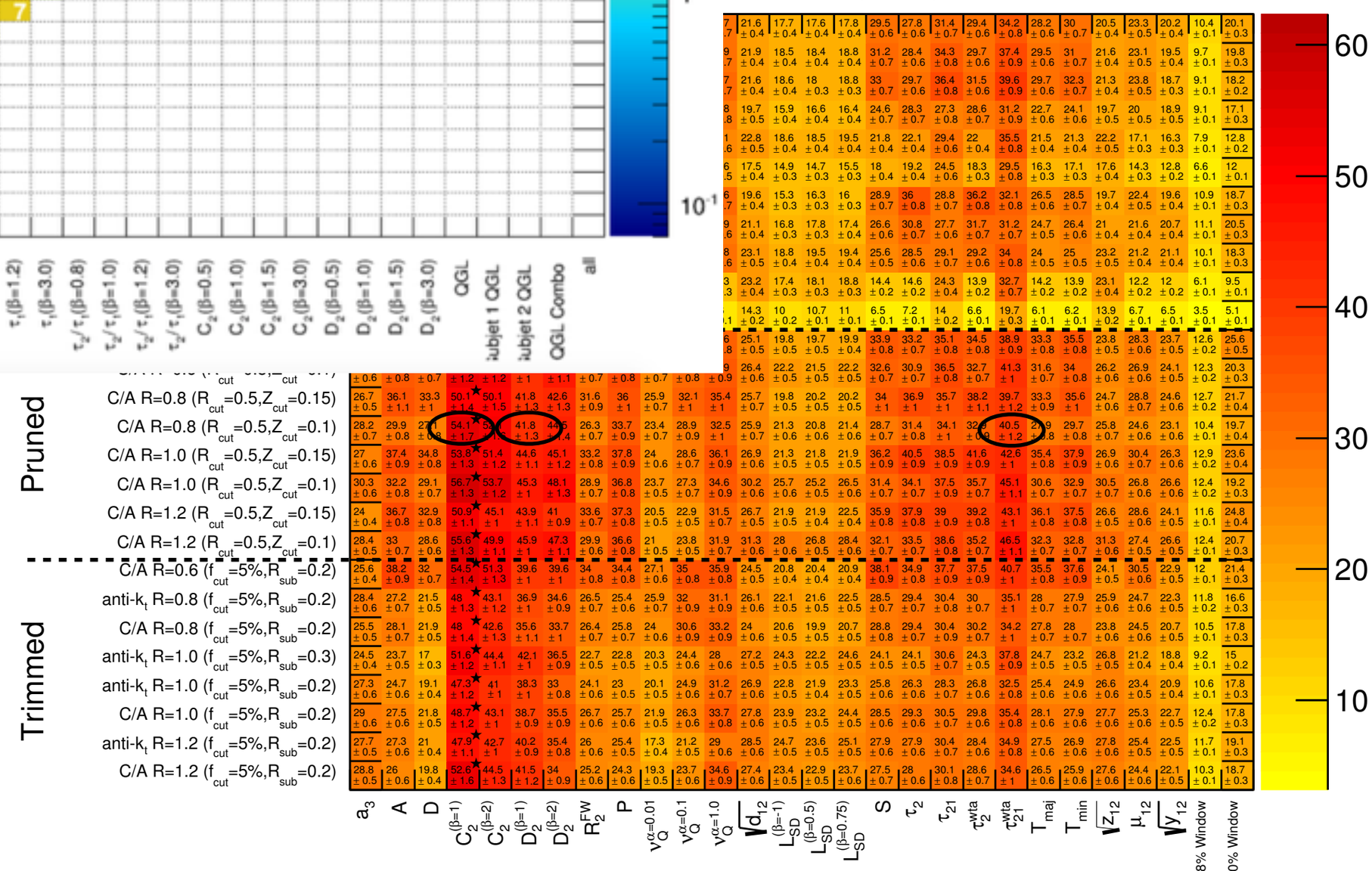
V-tagging optimization



t 4-momentum not calibrated

^{ruth} < 500 GeV , M Cut

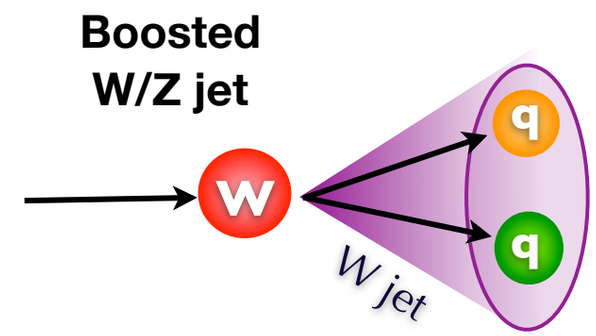
ariable for jet algorithm

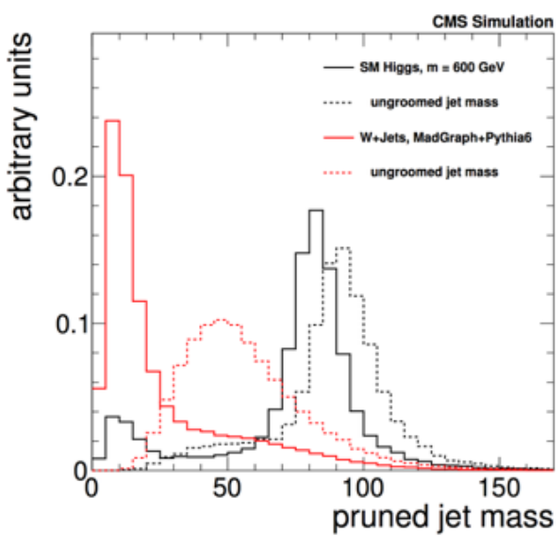


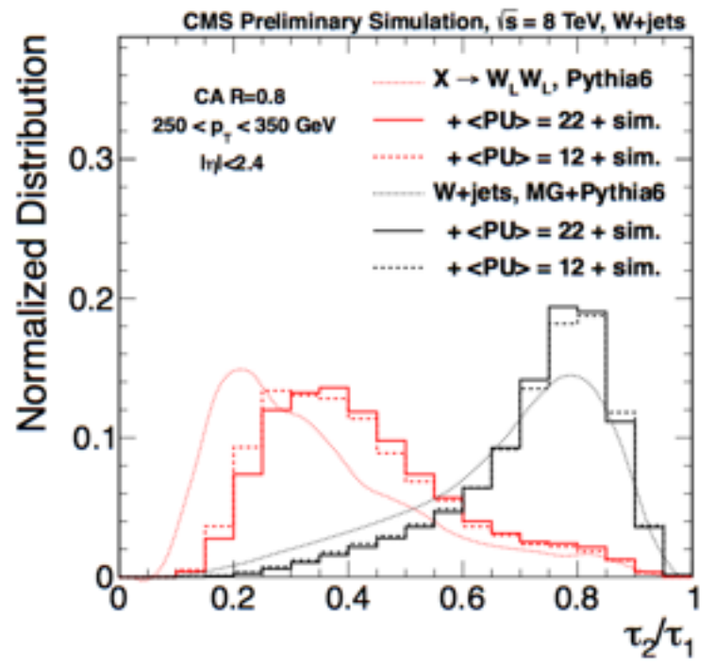
CMS and ATLAS come to different conclusions about the performance of τ_2/τ_1, C_2, D_2

2-prong tagging

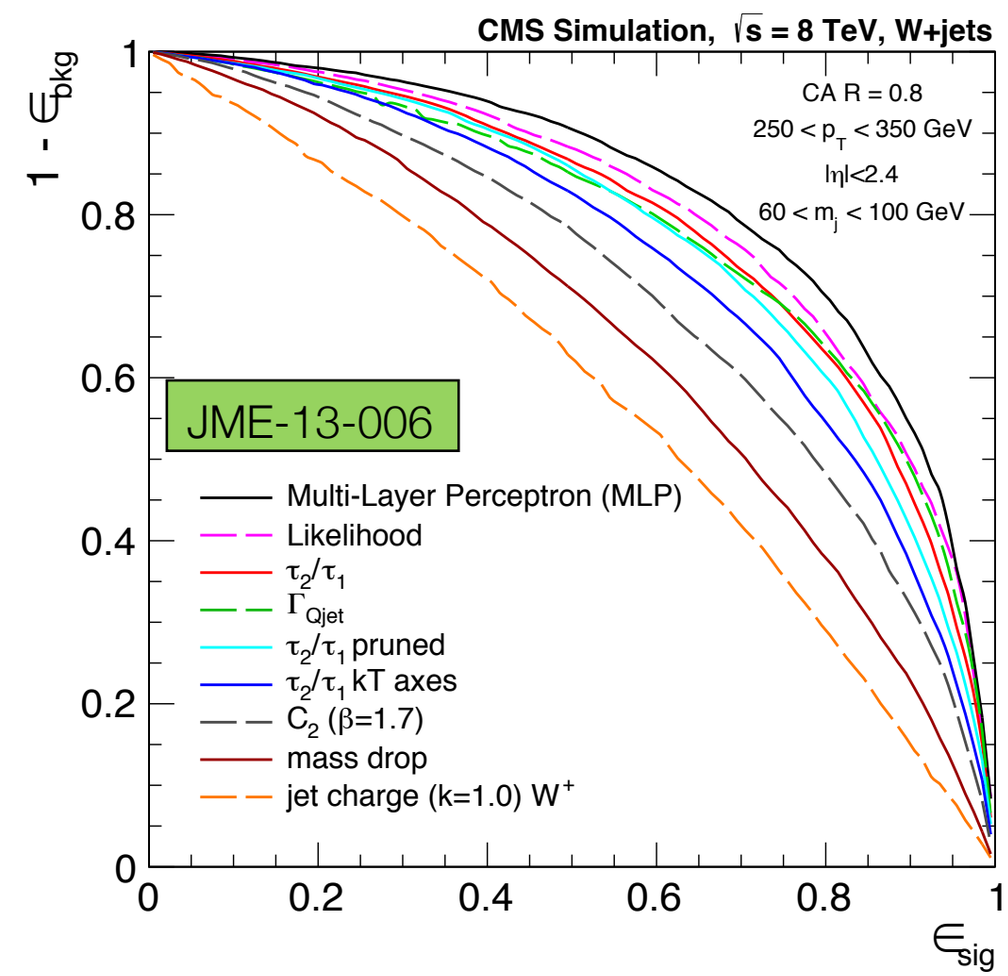
- Boosted W, Z
  - pruned mass + N-subjettiness
  - PUPPI soft drop AK8 jet + N-subjettiness
- Boosted Higgs → bb
  - Additionally subjet b-tagging







W tagging performance

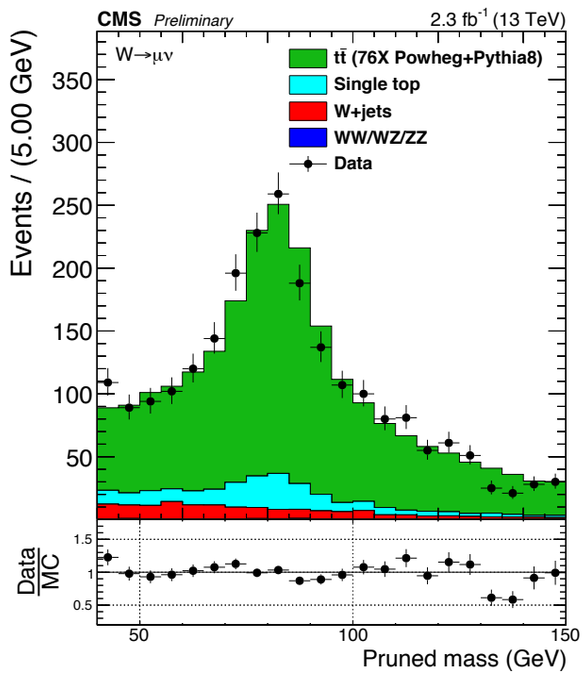


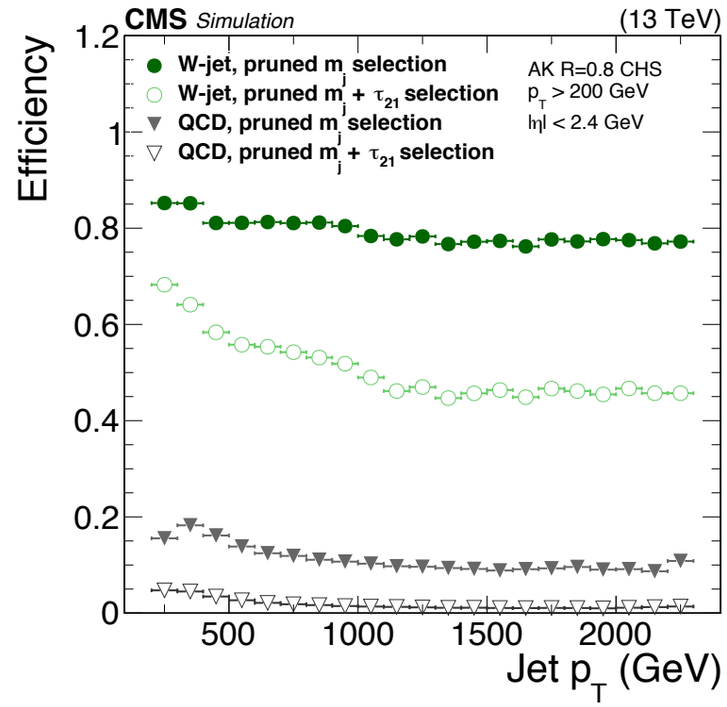
W-tagging

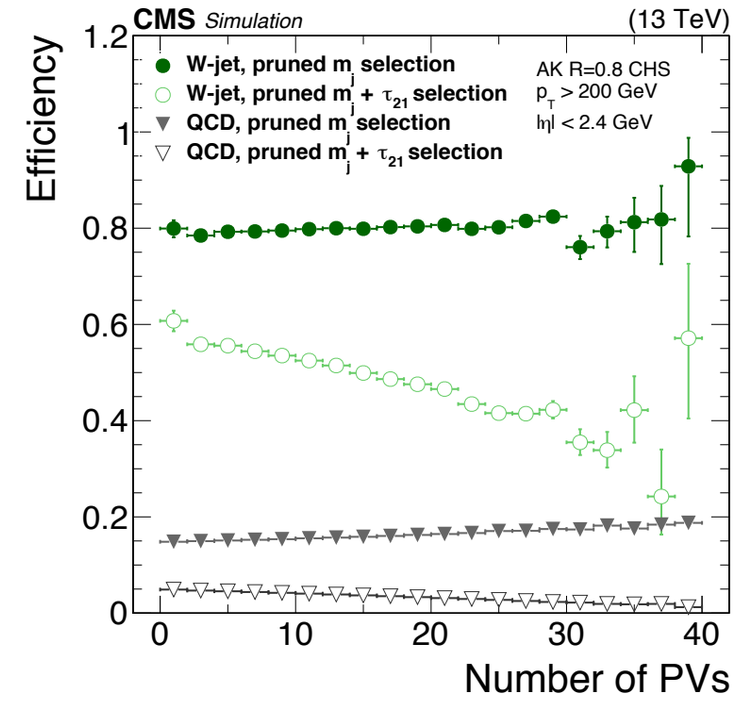
https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetWtagging

Thea Aarrestad

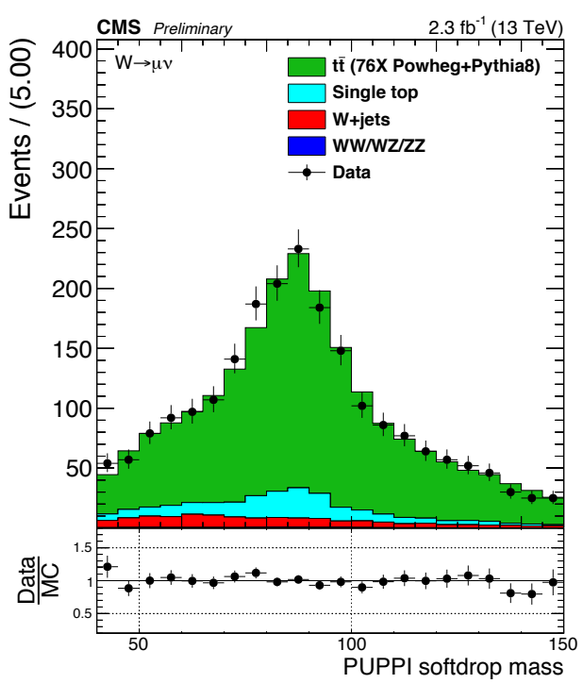
AK8 CHS pruned mass + τ_2/τ_1 (Run 1 baseline)

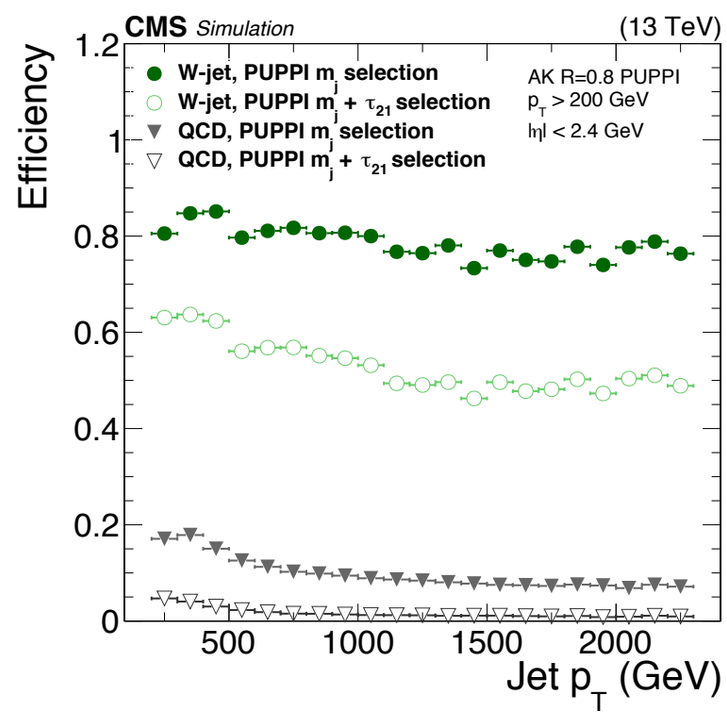


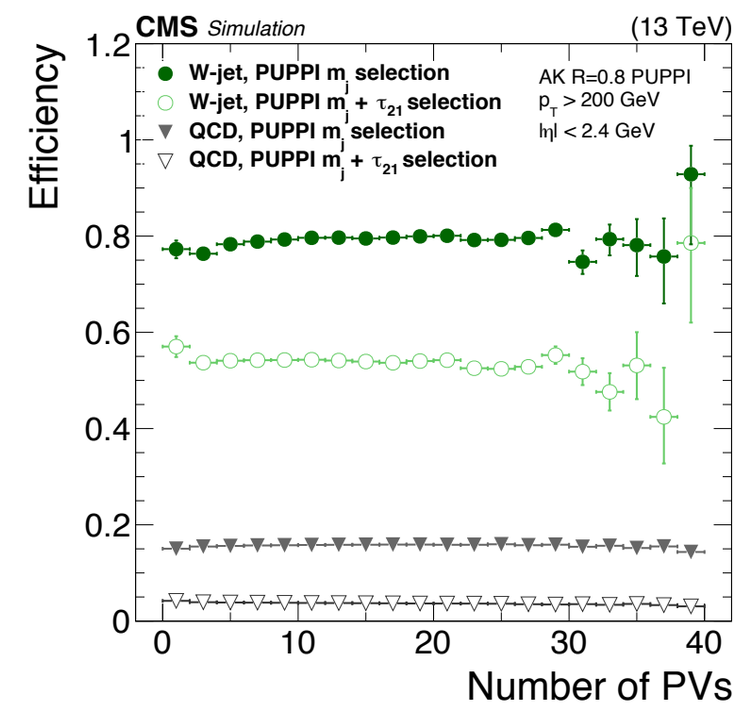




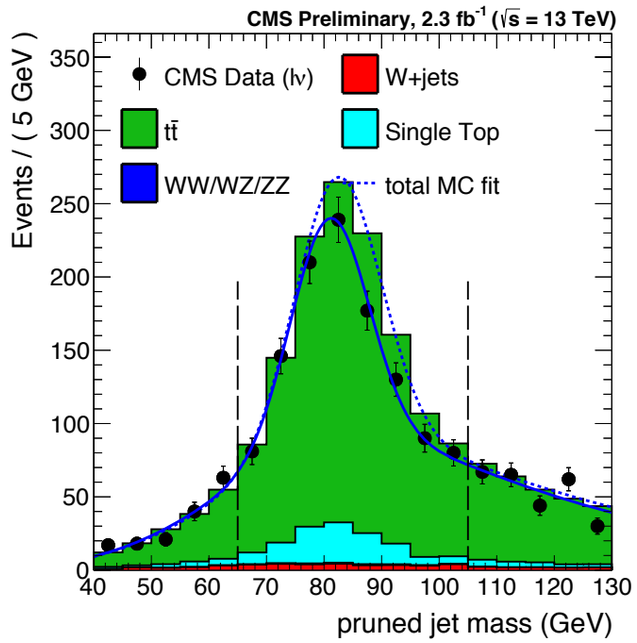
AK8 PUPPI SoftDrop mass + PUPPI τ_2/τ_1







W-tagging scale factor



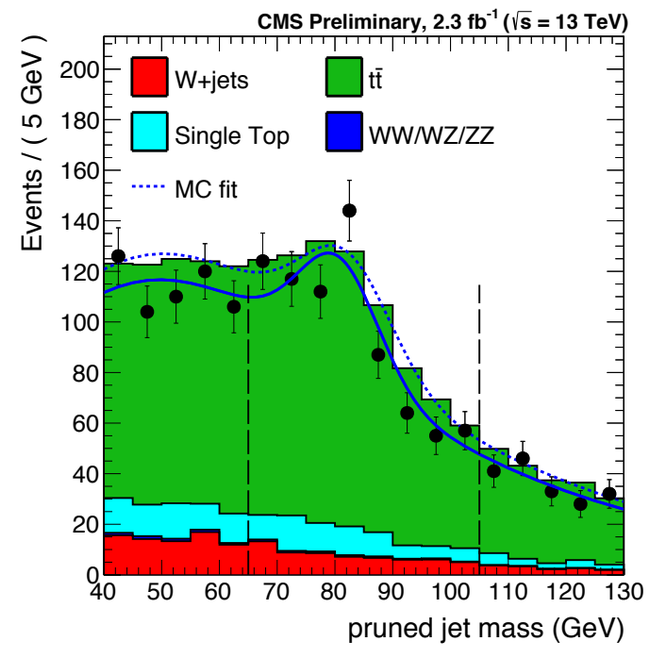
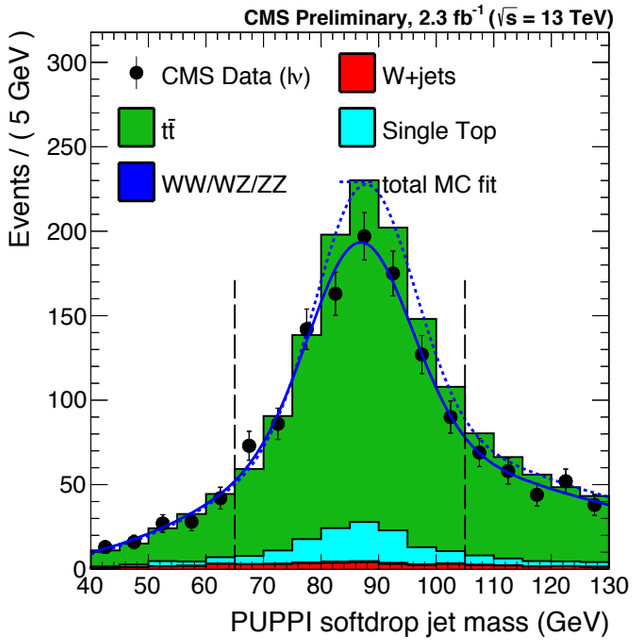


Table 3: Summary of the fitted W-mass peak fit parameters.

Parameter	Data	Simulation	Data/Simulation
pruning $\langle m \rangle$	80.9 ± 0.6 GeV	82.5 ± 0.1 GeV	0.980 ± 0.007
pruning σ	6.7 ± 0.7 GeV	7.5 ± 0.3 GeV	0.89 ± 0.10
PUPPI softdrop $\langle m \rangle$	86.8 ± 0.8 GeV	87.9 ± 0.2 GeV	0.988 ± 0.010
PUPPI softdrop σ	9.2 ± 1.0 GeV	8.7 ± 0.4 GeV	1.07 ± 0.09

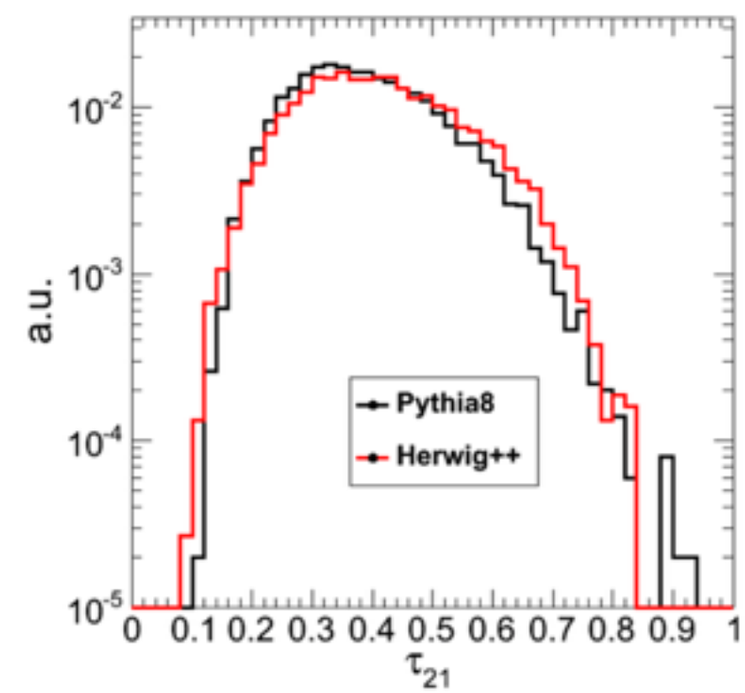
76X

Category	Definition	W scale factor
High-purity Pruning	$(\tau_2/\tau_1 < 0.45)$	0.94 ± 0.06 (stat) ± 0.03 (sys) ± 0.003 (sys)
Low-purity Pruning	$(0.45 < \tau_2/\tau_1 < 0.75)$	1.27 ± 0.33 (stat) ± 0.13 (sys) ± 0.008 (sys)
High-purity Pruning	$(\tau_2/\tau_1 < 0.6)$	0.98 ± 0.03 (stat) ± 0.003 (sys) ± 0.02 (sys)
High-purity PUPPI softdrop	$(\tau_2/\tau_1 < 0.4)$	0.99 ± 0.09 (stat) ± 0.04 (sys) ± 0.06 (sys)
Low-purity PUPPI softdrop	$(0.45 < \tau_2/\tau_1 < 0.75)$	1.06 ± 0.35 (stat) ± 0.17 (sys) ± 0.12 (sys)



Analysis implementation of SF

- Apply SF only to real W/Z/H-jets
  - Be careful not to to apply them to merged top quarks, combinatorial merged jets in ttbar, or non-merged jets
  - example implementation: B2G-16-002 (TT, T → bW, tH, tZ) [binning in N_{Vtags}]
    - SF applied only to jets matched to hadronic high pT W
    - JME recommendation in progress. In study: matching to W daughters, effect of ttbar b-merging.
- Systematic uncertainties
  - Jet mass calibration: vary jet mass calibration within JES uncertainty
  - SF uncertainty
    - Statistical + ttbar kinematic systematic + fit stability systematic
  - Jet mass resolution
    - If you are not JER smearing the jet mass distribution: smear jet mass with the measured SF resolution and vary within measured SF uncertainty
    - If you are applying JER smearing to the jet mass distribution: no additional smearing needed but propagate uncertainties
  - pT extrapolation uncertainty
    - SF only measured for 200-300 GeV W's
    - Extrapolation uncertainty measured from the double ratio of the selection efficiency in Pythia and Herwig samples



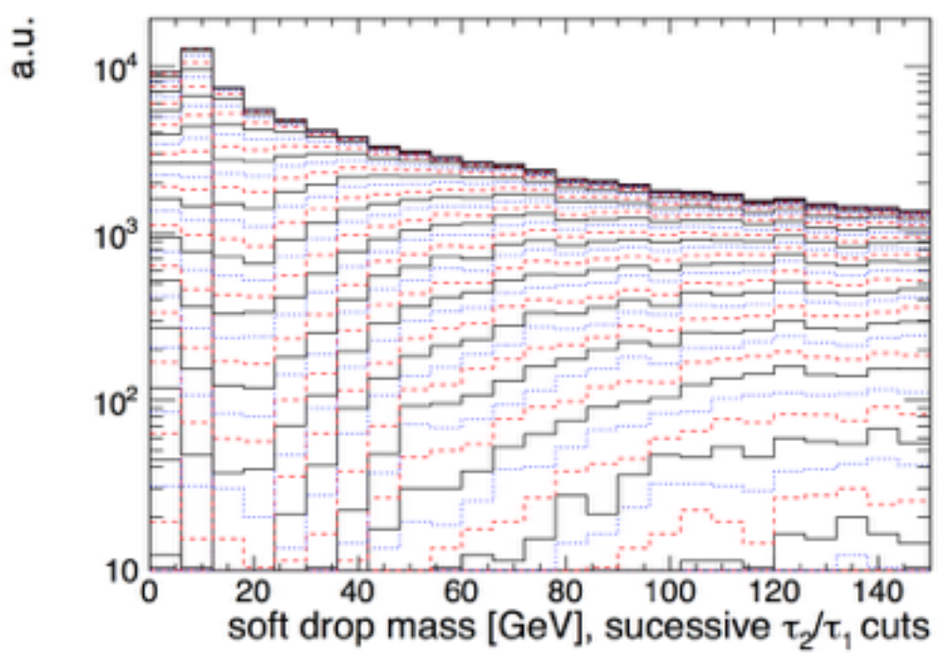
W-tagging τ_2/τ_1 scale factors

Category	Definition	W scale factor
High-purity Pruning	$(\tau_2/\tau_1 < 0.45)$	0.94 ± 0.06 (stat) ± 0.03 (sys) ± 0.003 (sys)
Low-purity Pruning	$(0.45 < \tau_2/\tau_1 < 0.75)$	1.27 ± 0.33 (stat) ± 0.13 (sys) ± 0.008 (sys)
High-purity Pruning	$(\tau_2/\tau_1 < 0.6)$	0.98 ± 0.03 (stat) ± 0.003 (sys) ± 0.02 (sys)
High-purity PUPPI softdrop	$(\tau_2/\tau_1 < 0.4)$	0.99 ± 0.09 (stat) ± 0.04 (sys) ± 0.06 (sys)
Low-purity PUPPI softdrop	$(0.45 < \tau_2/\tau_1 < 0.75)$	1.06 ± 0.35 (stat) ± 0.17 (sys) ± 0.12 (sys)

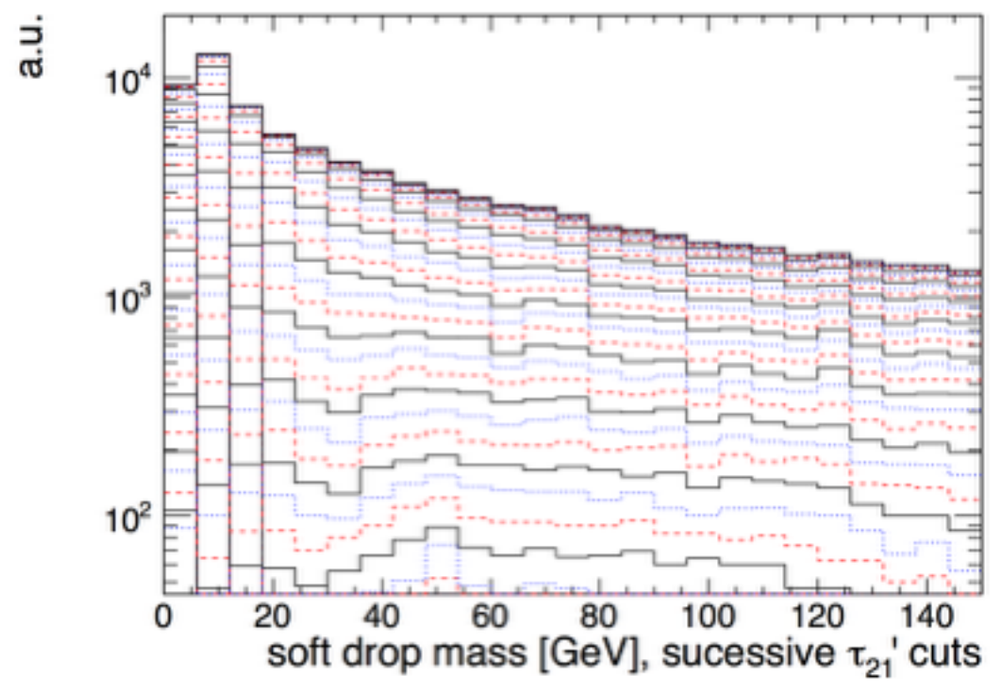
- Systematic uncertainties due to:
  - Description of ttbar kinematics (nearby jets, p_T spectrum)
    - Evaluate based on Powheg+Pythia8 (main) vs. Madgraph+Pythia8 (alternative)
  - Fit stability and signal (matching W) definition:
    - Assume Gaussian peak, absorbing tails into background function (more robust) vs.
    - Assume Gaussian peak with tails estimated from matched MC (more precise def. of real W, less robust fit)
  - Pythia8 vs. Herwig++ shower difference not accounted as systematic since measured with scale factor
  - Uncertainty for extrapolation to high pT based on Pythia8 vs. Herwig++ difference in pT-dependence modeling
    - for tau21<0.45 : additional 6%, 10%, 13% uncertainty for pT=500, 1000, 2000 GeV
    - for tau21<0.6 : additional 4%, 8%, 10% uncertainty for pT=500, 1000, 2000 GeV
  - Sources from lepton identification, b tagging, MET scale and underlying event are at < 0.5% level and negligible
  - Sources from JES/JER evaluated for each analysis and not listed

New ideas

- Designing Decorrelated Taggers (DDT) http://arxiv.org/abs/1603.00027
  - Example: transform $\tau_{21} \rightarrow \tau_{21}$` such that correlations with p_T and mass are reduced
  - Reduce sculpting of mass distribution after applying τ_{21} cuts

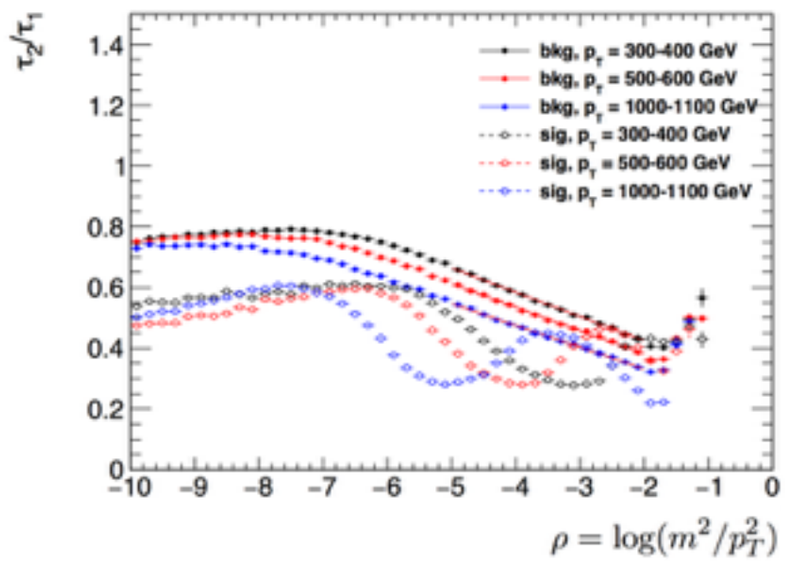


→



Using DDT

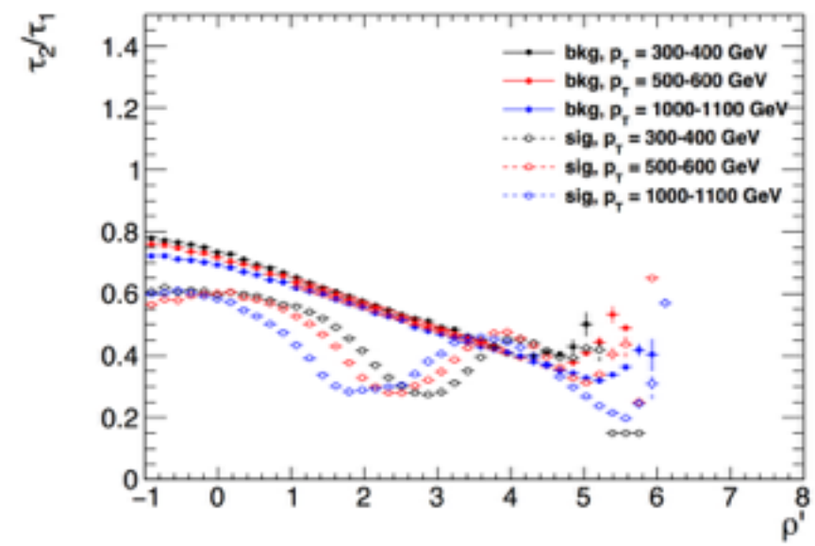
Utilize unique properties of soft drop:



→

$$\rho' = \rho + \log\frac{p_T}{\mu} = \log\left(\frac{m^2}{p_T\mu}\right)$$

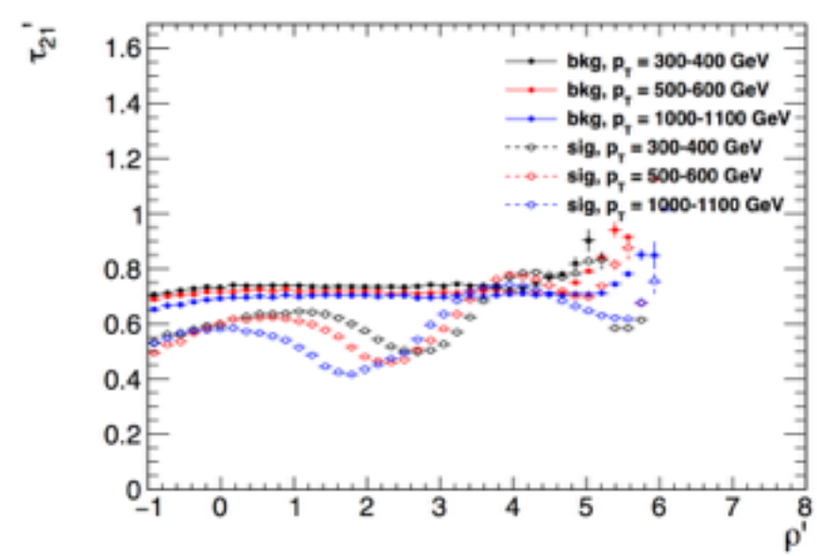
Change variable to remove pT dependence:



↓

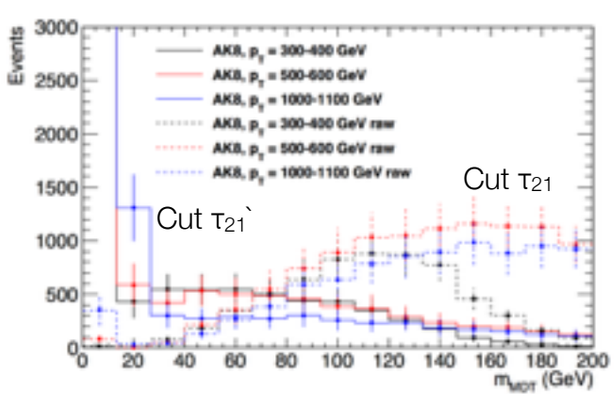
Measure slope and transform τ_{21} in order to remove dependence on mass and pT

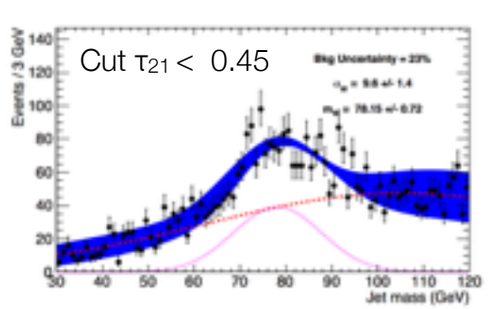
$$\tau_{21}' = \tau_2/\tau_1 - M \times \rho'$$

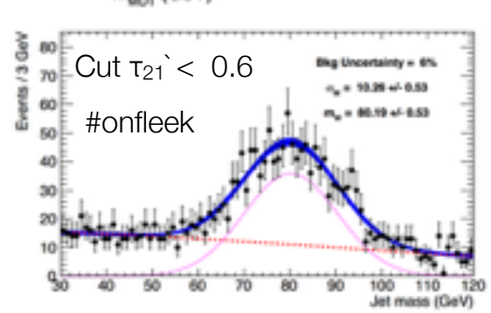


←

Cut on τ_{21}`. Removed jet mass sculpting. Improved background estimation (reduced uncertainties).



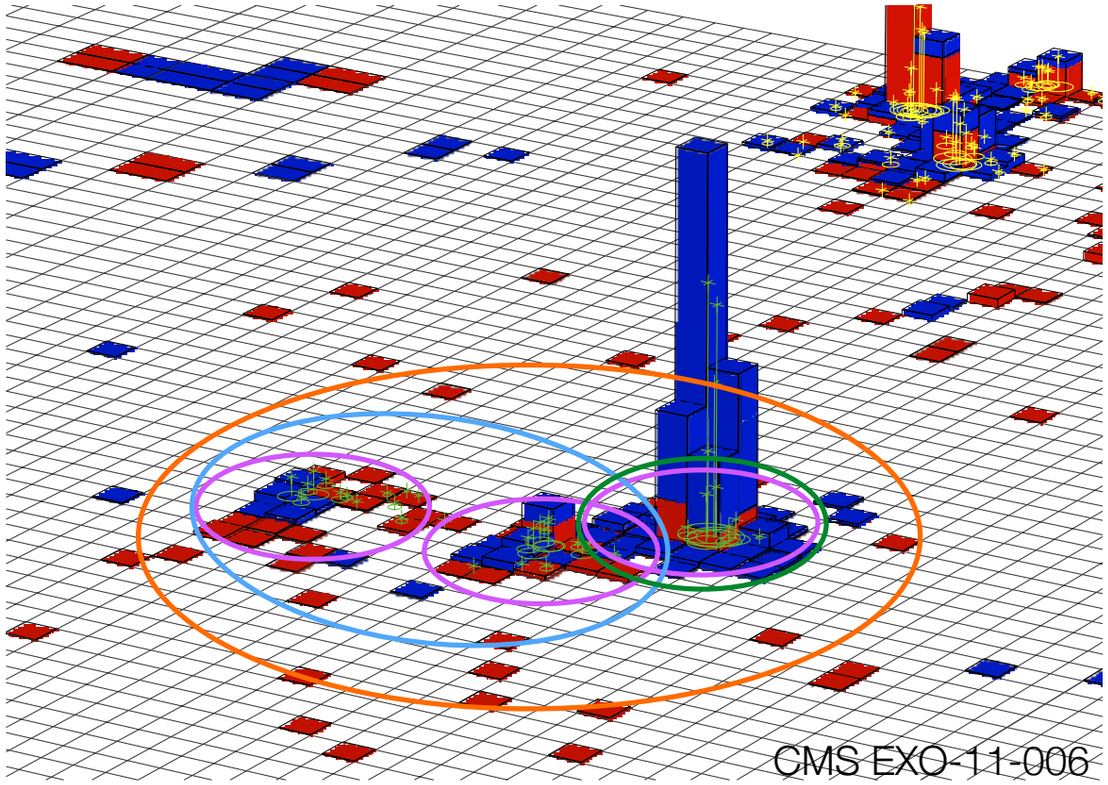




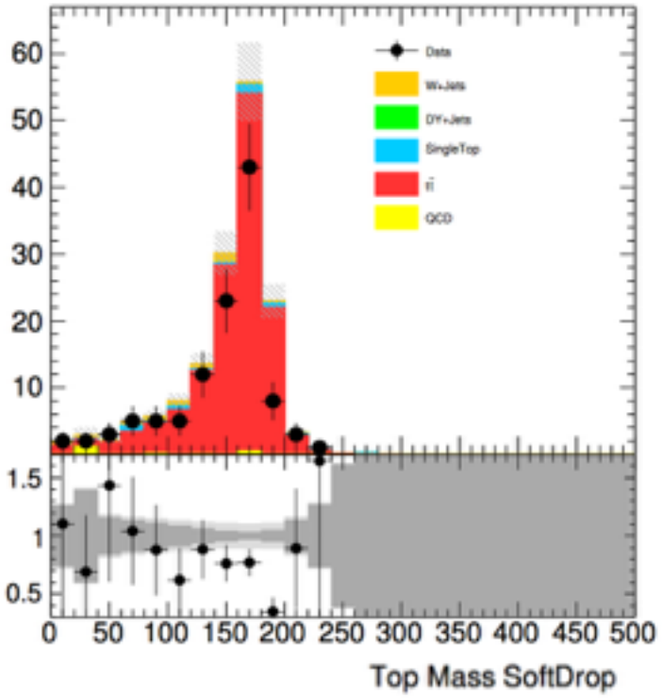
top tagging

https://twiki.cern.ch/twiki/bin/view/CMS/JetTopTagging

- Top tagging
  - Default algorithm = AK8 soft drop jet mass + τ_3/τ_2 + (subjet b-tag)
    - planned for 2016: PUPPI soft drop mass + PUPPI τ_3/τ_2

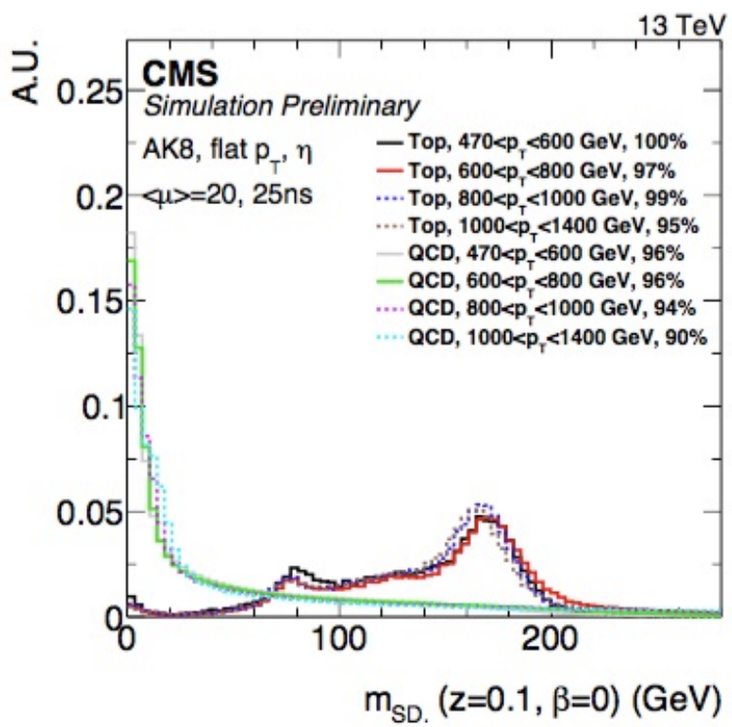


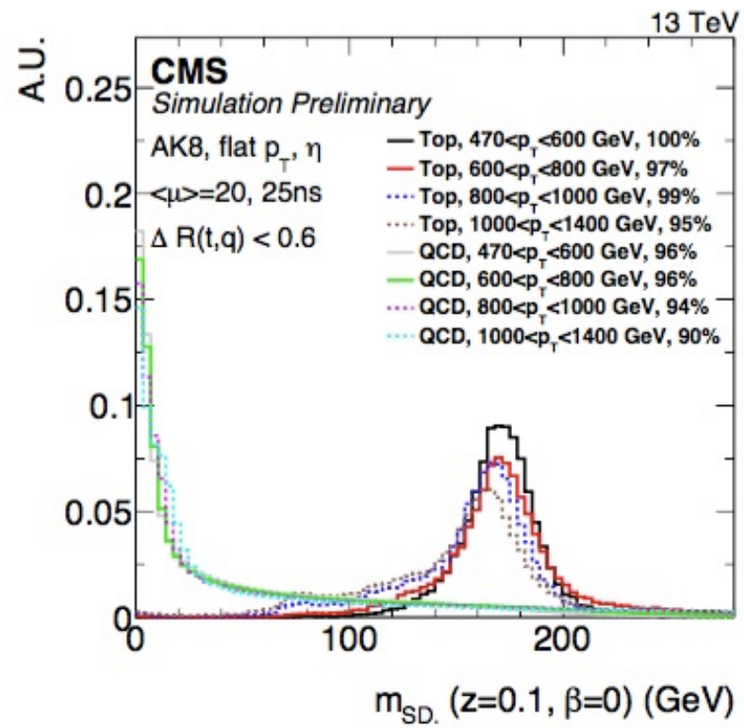
- 80X miniAOD v2: adding PUPPI substructure variables
  - Add PUPPI soft drop, PUPPI N-subjettiness (remove CMS top tagger)

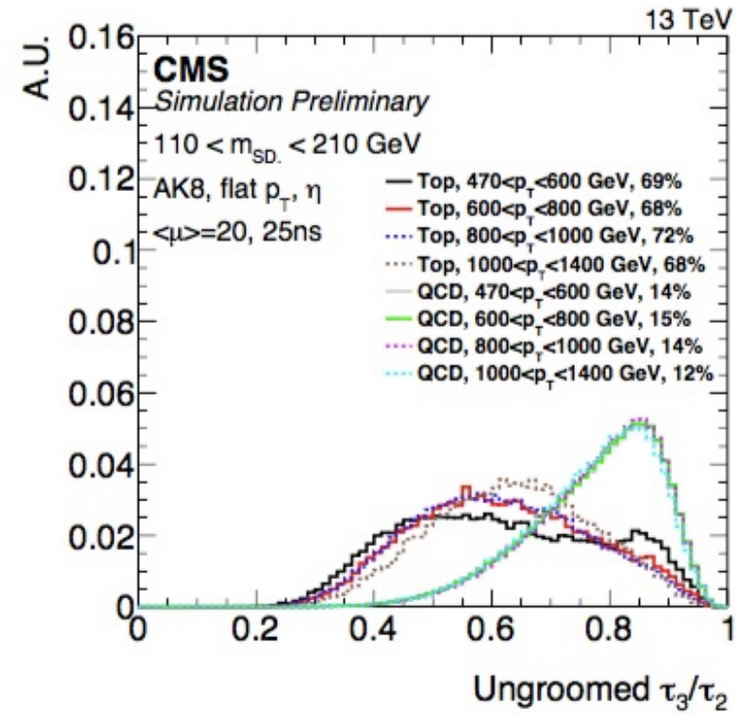


- Alternative algorithms:
  - CMS Top Tagger v1
  - HEP Top Tagger v2
  - Shower deconstruction etc.

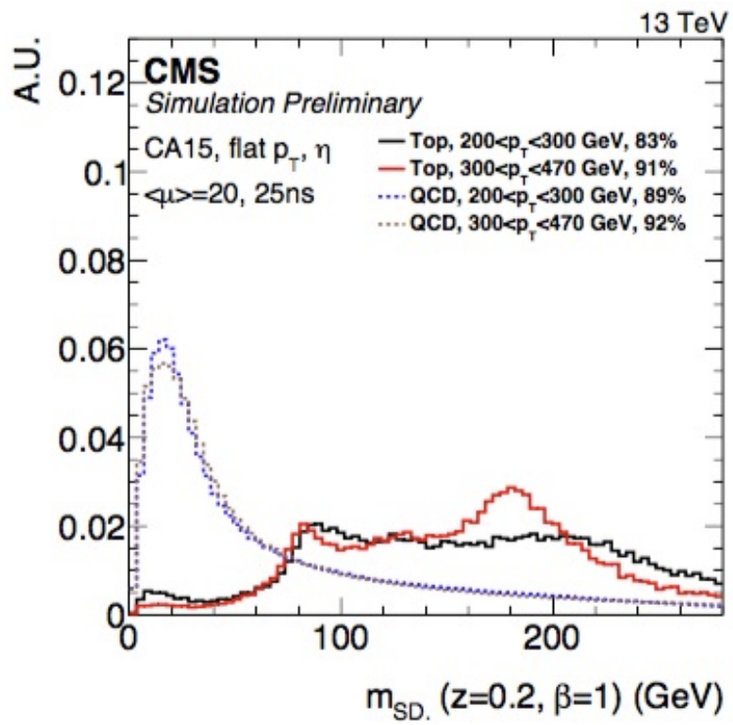
top-tagging - high pT (AK8)

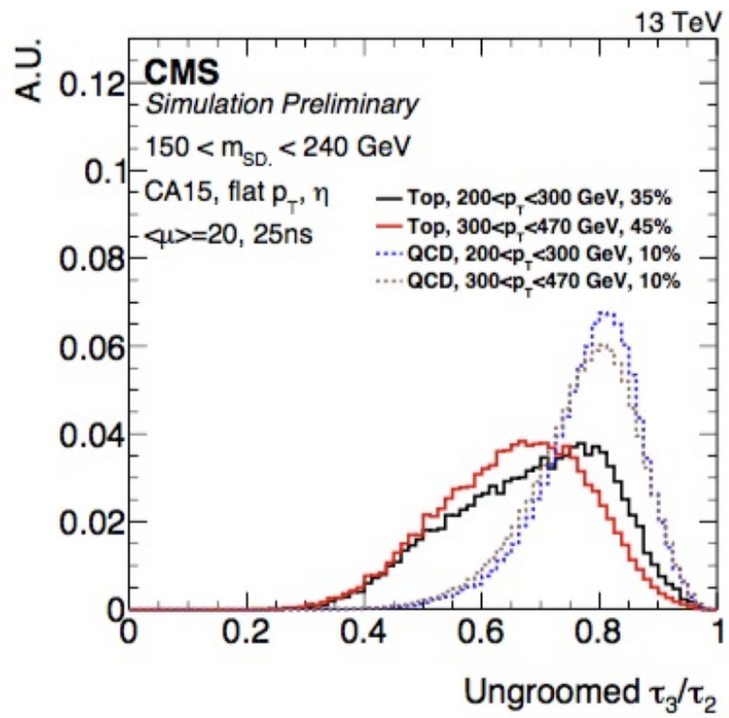


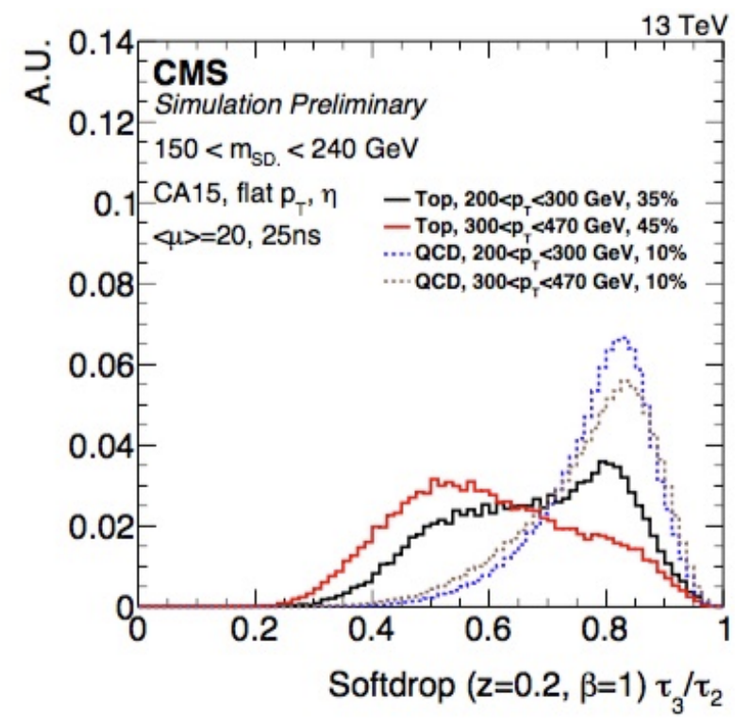


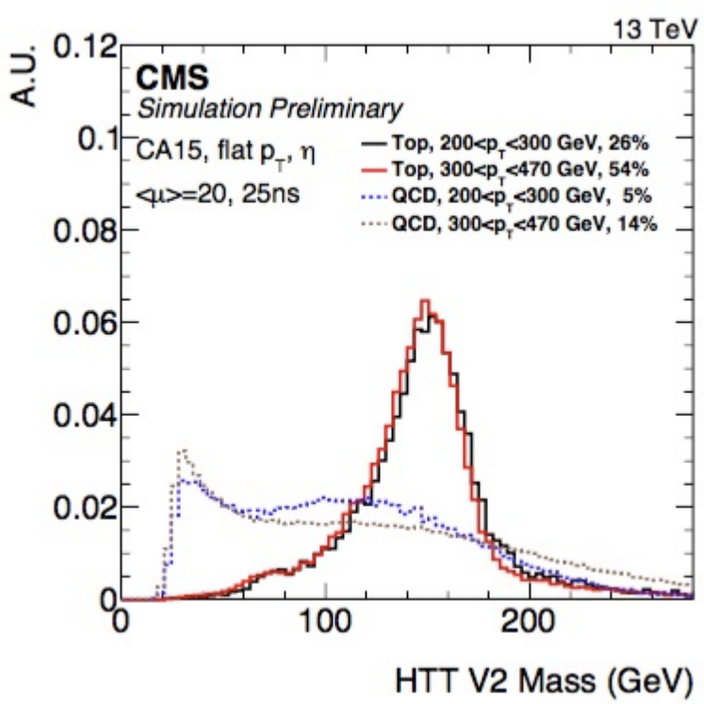


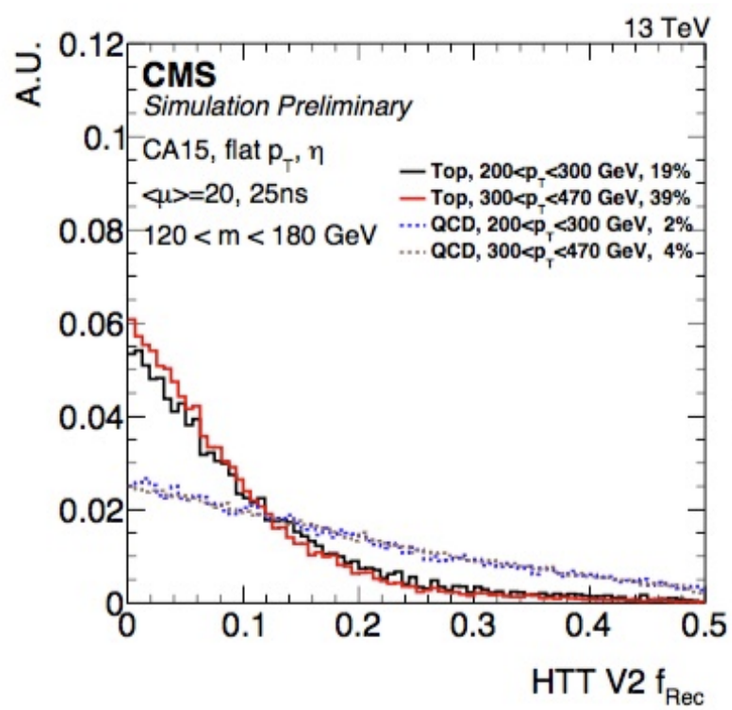
top-tagging - low pT (CA15)

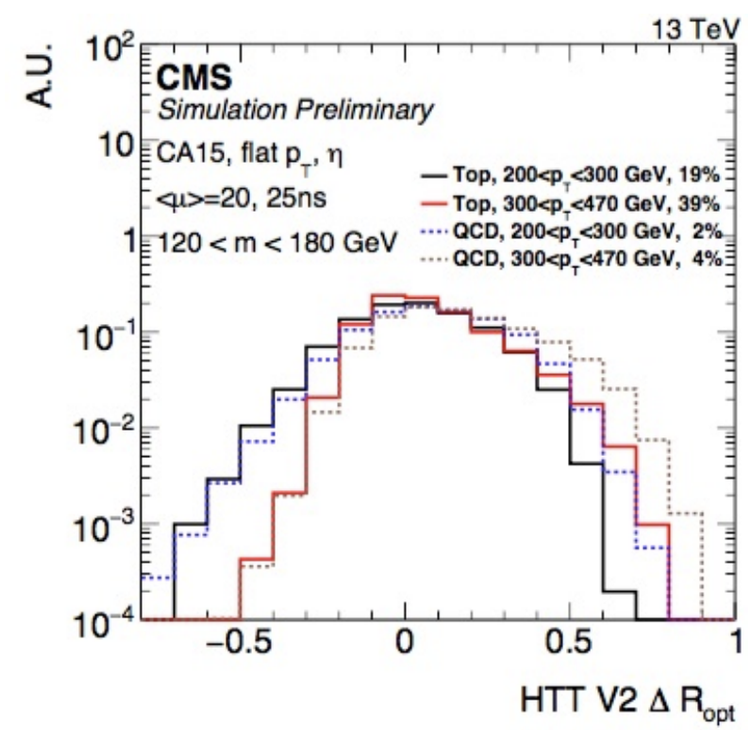




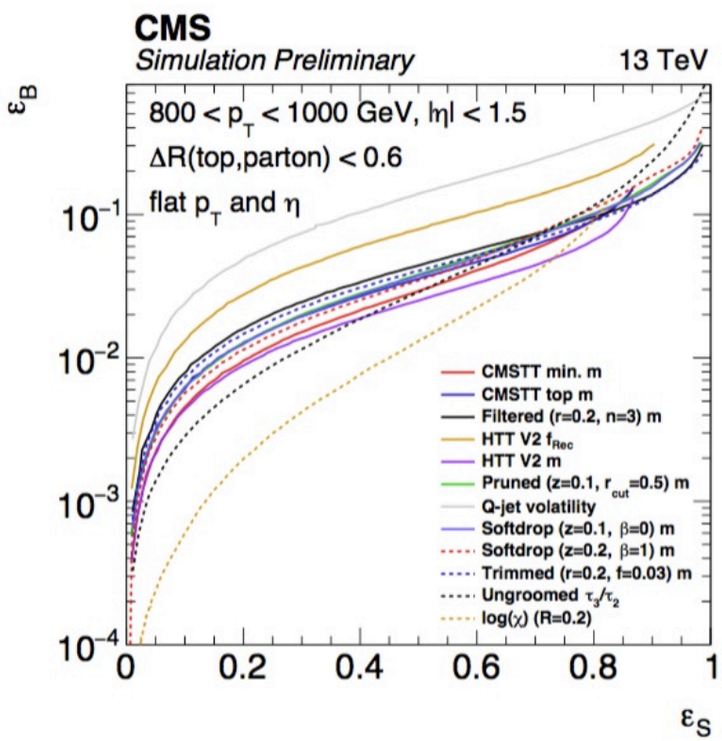


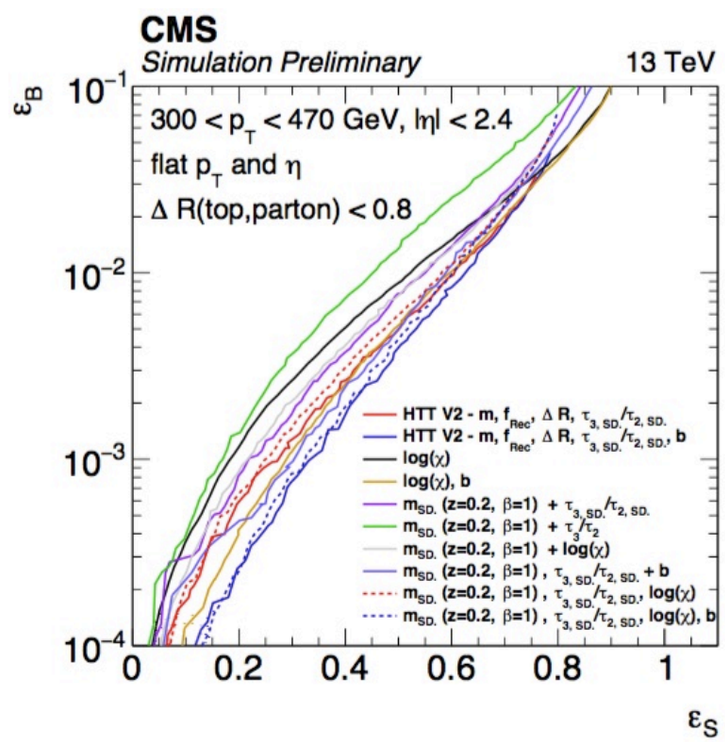


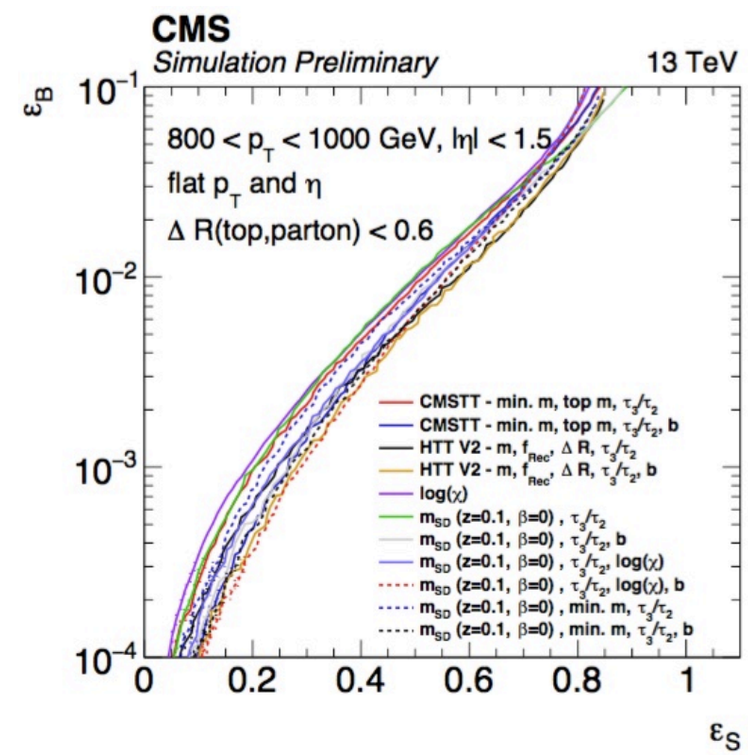




Algorithm comparison

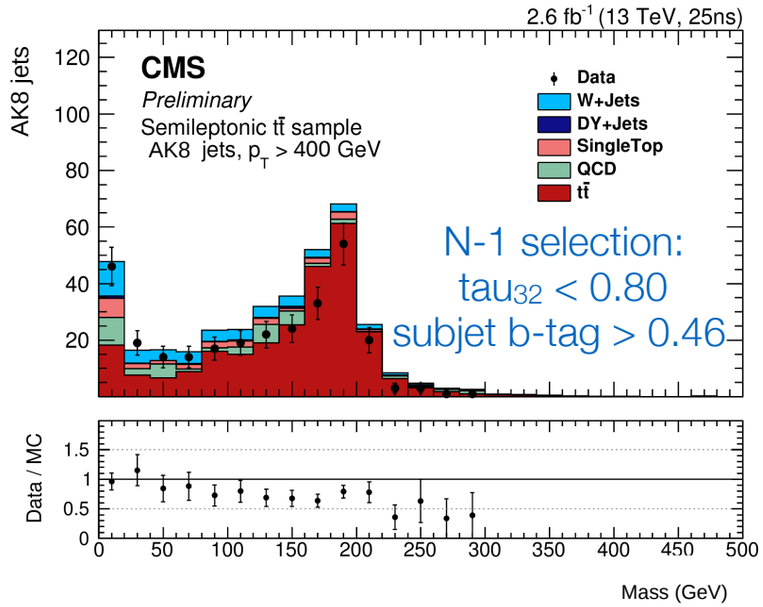


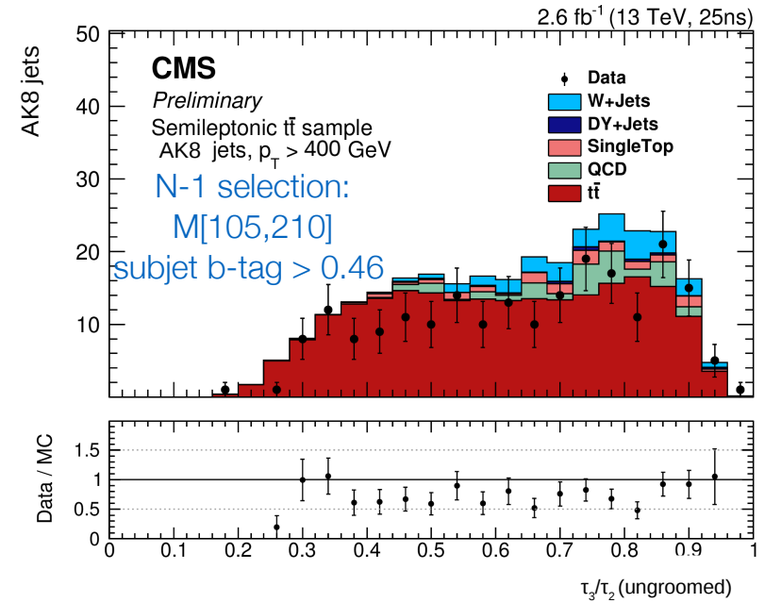




Top-tagging validation 76X

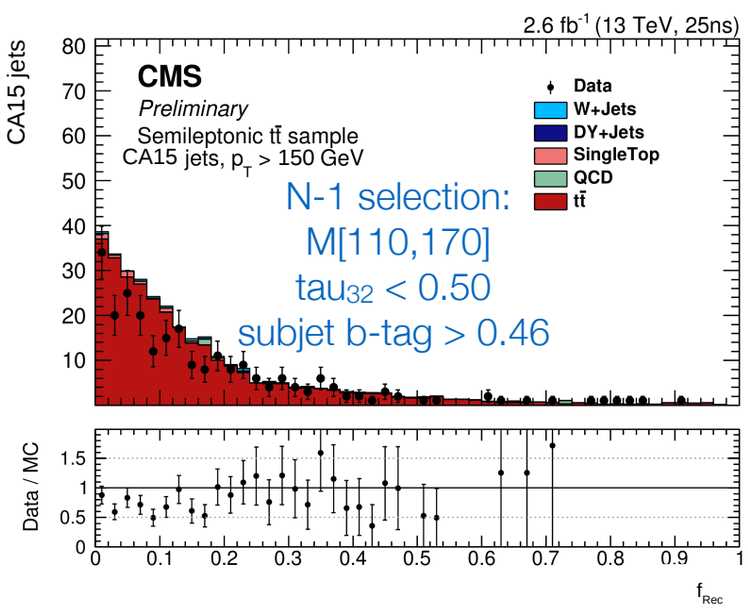
AK8 jets (SoftDrop + τ_3/τ_2)

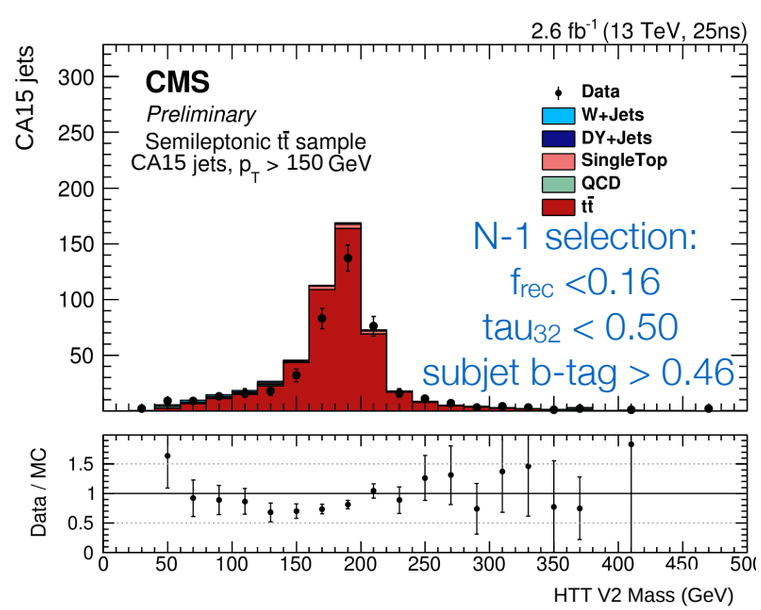


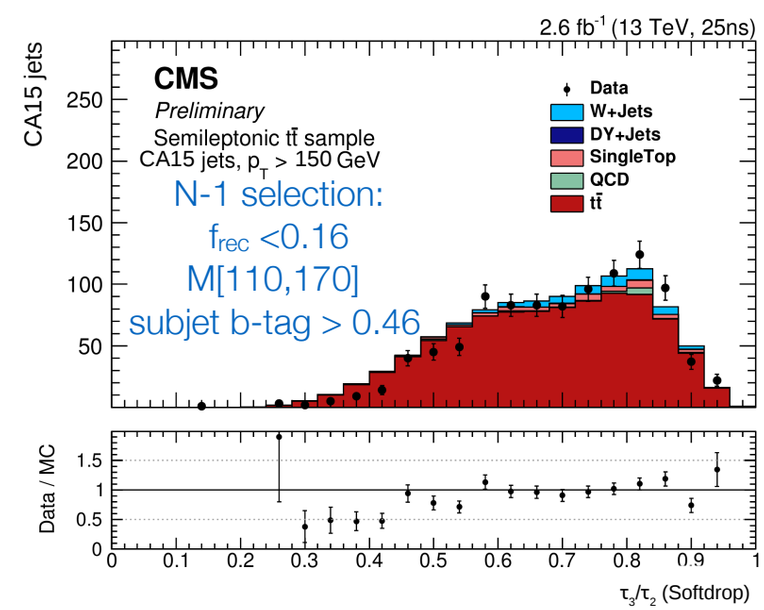


$$f_{Rec} = \min_{ij} \left| \frac{\frac{m_{ij}}{m_{123}}}{\frac{m_W}{m_t}} - 1 \right|$$

CA15 jets (HEPTopTaggerV2 + τ_3/τ_2)







Top tagging scale factor

- The scale factor is defined as the ratio of the efficiency in data and the efficiency in tt MC
- Residual background contamination is subtracted from data, before evaluating the scale factors.

AK8 PUPPI + SoftDrop

$\epsilon(B)$	p_T bin [GeV]	400 - 550	550 - ∞	incl. (400 - ∞)
0.1 %	$M_{SD}[105,210]$, $\tau_{32,SD} < 0.46$, subjet b-tag > 0.46	1.33±0.26	1.04±0.33	1.24±0.20
0.3 %	$M_{SD}[105,210]$, $\tau_{32,SD} < 0.54$, subjet b-tag > 0.46	1.25±0.22	0.89±0.26	1.13±0.17
1.0 %	$M_{SD}[105,210]$, $\tau_{32,SD} < 0.65$, subjet b-tag > 0.46	1.41±0.20	0.73±0.21	1.19±0.15
3.0 %	$M_{SD}[105,210]$, $\tau_{32,SD} < 0.80$, subjet b-tag > 0.46	1.21±0.16	0.82±0.20	1.09±0.13

AK8 CHS + SoftDrop

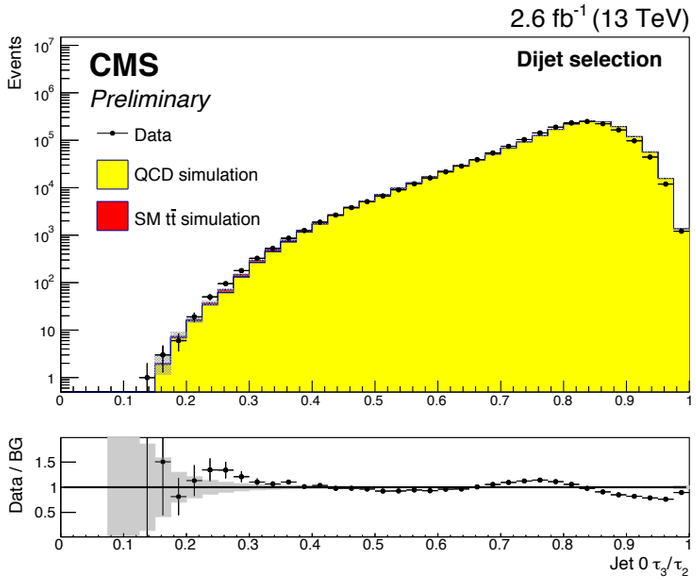
$\epsilon(B)$	p_T bin [GeV]	400 - 550	550 - ∞	incl. (400 - ∞)
0.1 %	$M_{SD}[105,220]$, $\tau_{32,SD} < 0.50$, subjet b-tag > 0.46	0.73±0.13	1.02±0.28	0.81±0.12
0.3 %	$M_{SD}[105,220]$, $\tau_{32,SD} < 0.57$, subjet b-tag > 0.46	0.79±0.12	0.94±0.23	0.83±0.11
1.0 %	$M_{SD}[105,220]$, $\tau_{32,SD} < 0.67$, subjet b-tag > 0.46	0.87±0.11	0.96±0.20	0.90±0.09
3.0 %	$M_{SD}[105,220]$, $\tau_{32,SD} < 0.81$, subjet b-tag > 0.46	0.86±0.09	1.03±0.18	0.90±0.08

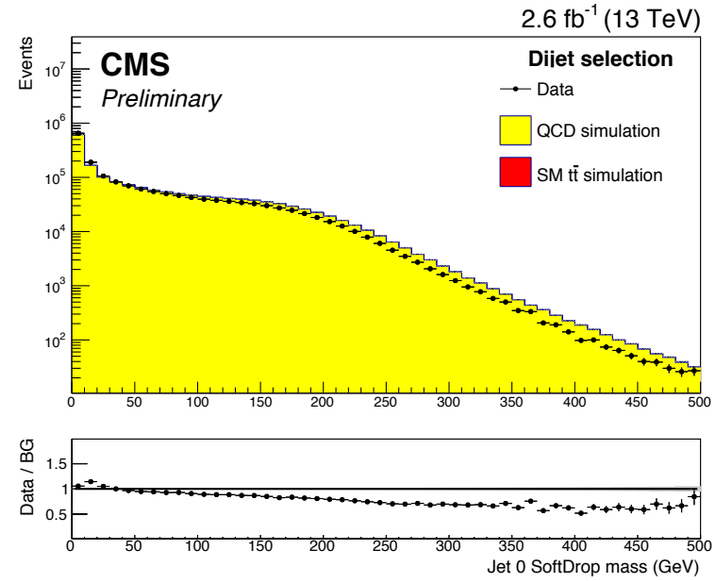
Statistical only uncertainty

Top-tagging - dijet measurements

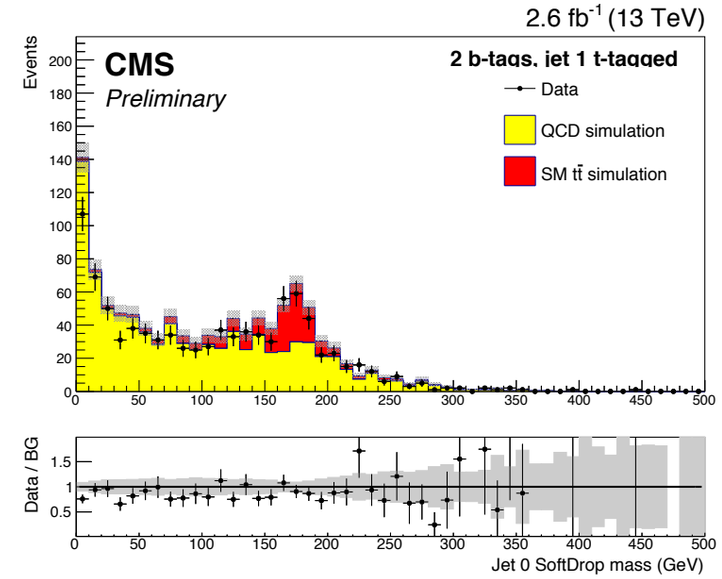
- Data MC comparisons

2 pT>400 jets, no tagging requirements

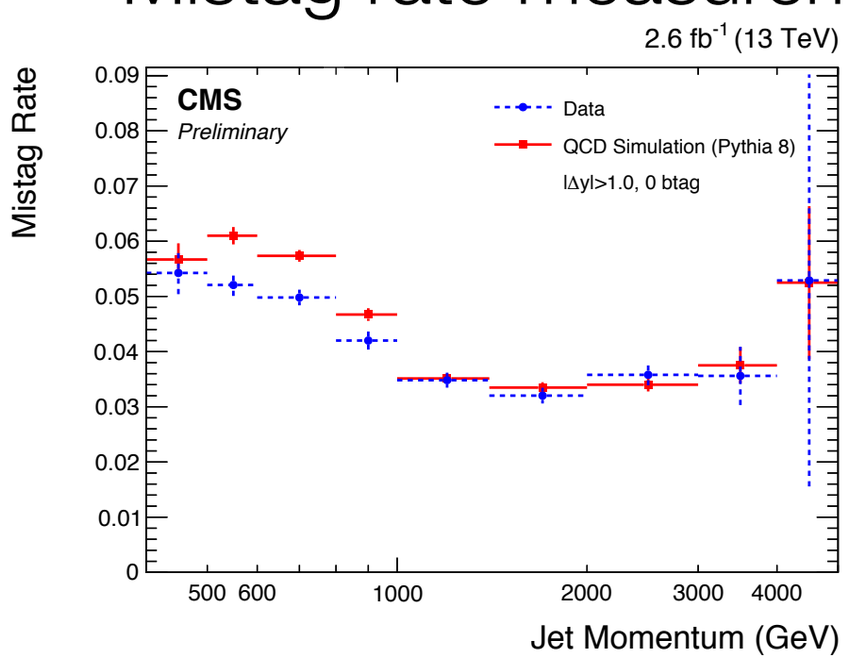




2 subjet b-tags + 1 top-tag



- Mistag rate measurement (anti-tag and probe)



$$\text{Mistag rate} = \frac{\text{\# of tagged jets}}{\text{\# of probe jets}}$$

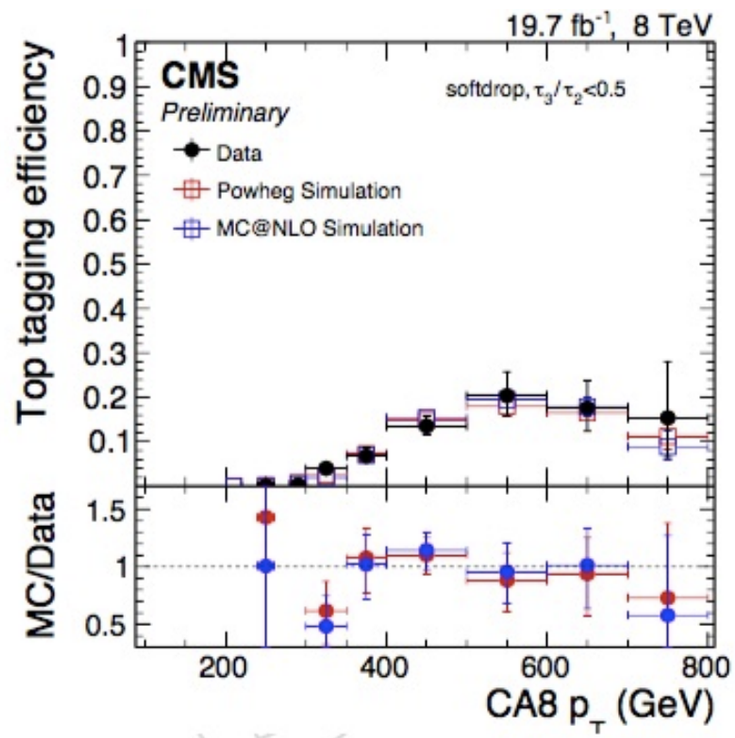
Jets in miniAOD

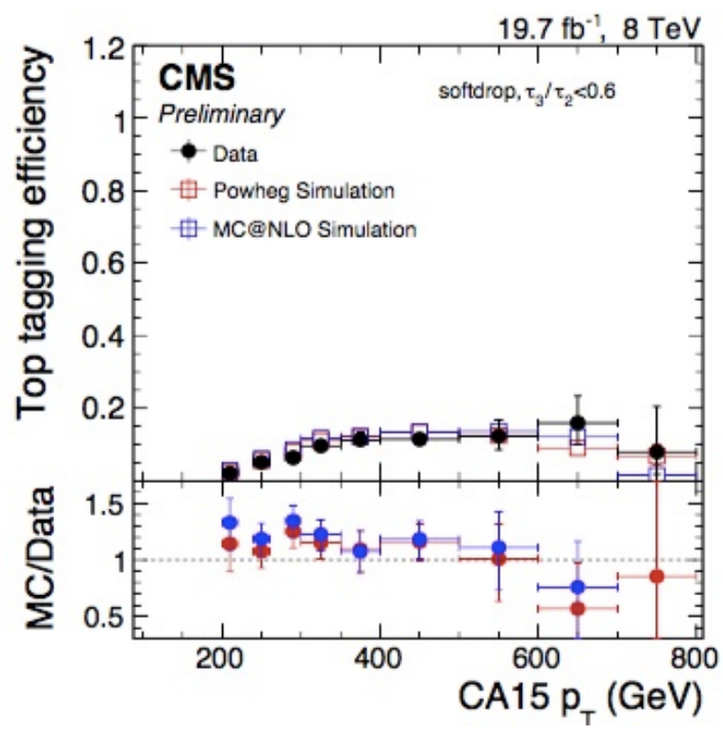
https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD2016

- slimmedJets = ak4PFJetsCHS
  - p_T >10 GeV (typically analysis cut will be at least p_T>20)
  - jet corrections applied
    - May not be the most recent corrections!
- slimmedJetsPuppi = ak4PFJetsPUPPI
- slimmedJetsAK8 = ak8PFJetsCHS
  - p_T >170 GeV (typically analysis cut will be at least p_T>200)
  - groomed masses, N-subjettiness etc. attached as userfloats
  - ak8PFJetPUPPI jet quantities attached as userfloats

Exercise 2

- Grooming
- W-tag
- top-tag
- How does jet mass change with pT/NVTX?
- How does N-subjettiness change with pT/NVTX?


19.7 fb^{-1}, 8 TeV
CMS
Preliminary
softdrop, τ_3/τ_2<0.5
Data
Powheg Simulation
MC@NLO Simulation
Top tagging efficiency
MC/Data
CA8 p_T (GeV)


19.7 fb^{-1}, 8 TeV
CMS
Preliminary
softdrop, τ_3/τ_2<0.6
Data
Powheg Simulation
MC@NLO Simulation
Top tagging efficiency
MC/Data
CA15 p_T (GeV)

Jet Toolbox

TWiki > CMS Web > JetMET > JetMETAlgorithmsReconstruction > JetToolbox (2016-04-25, AlejandroGomez)

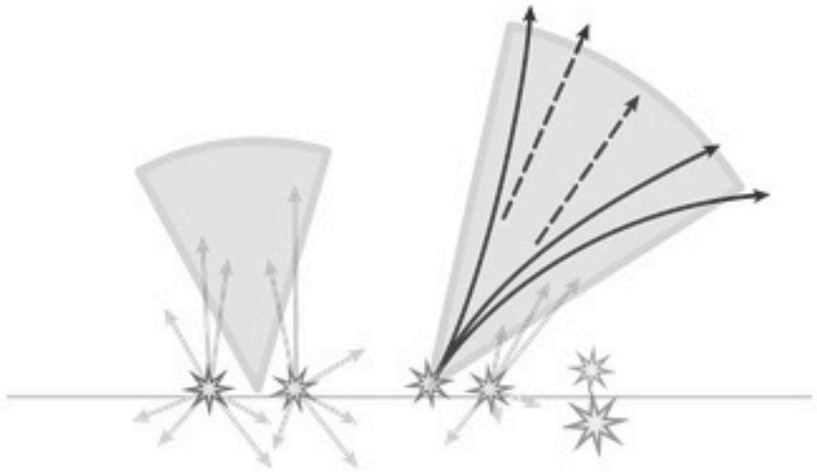
Edit Attach PDF

Jet Toolbox

Jet Toolbox in CMS

Contents

- Introduction
- How to run the jetToolbox
- Arguments
- Pile-up Mitigation Tools
  - CHS
  - Puppi
  - SoftKiller
  - Constituent Substraction
  - Plain PF Jets
  - New PF Collection
- Substructure Tools
  - Grooming ValueMaps
  - addSubjets (from prunning or softdrop)
  - N-subjettiness
  - QJets
  - QGTagger
  - Pileup JetID
  - Energy Correlation Functions
  - HEP and CMS Top Tagger
- Frequently Ask Questions
- Review status



https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetToolbox

```python
def jetToolbox( proc, jetType, jetSequence, outputFile,
                newPFCollection=False, nameNewPFCollection = '',
                PUMethod='CHS',
                miniAOD=True,
                runOnMC=True,
                JETCorrPayload='', JETCorrLevels = [ 'None' ], GetJetMCFlavour=True,
                Cut = '',
                bTagDiscriminators = None,
                bTagInfos = None,
                subJETCorrPayload='', subJETCorrLevels = [ 'None' ], GetSubjetMCFlavour=True,
                CutSubjet = '',
                addPruning=False, zCut=0.1, rCut=0.5, addPrunedSubjets=False,
                addSoftDrop=False, betaCut=0.0,  zCutSD=0.1, addSoftDropSubjets=False,
                addTrimming=False, rFiltTrim=0.2, ptFrac=0.03,
                addFiltering=False, rfilt=0.3, nfilt=3,
                addCMSTopTagger=False,
                addMassDrop=False,
                addHEPTopTagger=False,
                addNsub=False, maxTau=4,
                addNsubSubjets=False, subjetMaxTau=4,
                addPUJetID=False,
                addQJets=False,
                addQGTagger=False, QGjetsLabel='chs',
                addEnergyCorrFunc=False, ecfBeta = 1.0, maxECF=3,
                ):
```

https://github.com/cms-jet/JetToolbox/blob/jetToolbox_763/python/jetToolbox_cff.py

```python
# AK R=1.2 jets from CHS inputs with basic grooming, W tagging, and top tagging
jetToolbox( process, 'ak12', 'ak12JetSubs', 'out',
  PUMethod='CHS',
  addPruning=True, addSoftDrop=True ,          # add basic grooming
  addTrimming=True, addFiltering=True,
  addSoftDropSubjets=True,
  addNsub=True, maxTau=4,                      # add Nsubjettiness tau1, tau2, tau3, tau4
  JETCorrPayload = 'AK8PFchs', JETCorrLevels = ['L2Relative', 'L3Absolute']
)
```

Exercise 3

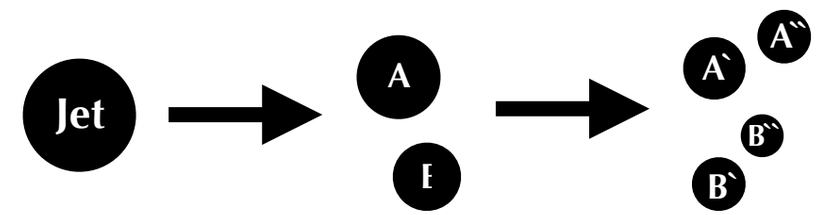
- Cluster jets with the jet toolbox and make some plots

Backup

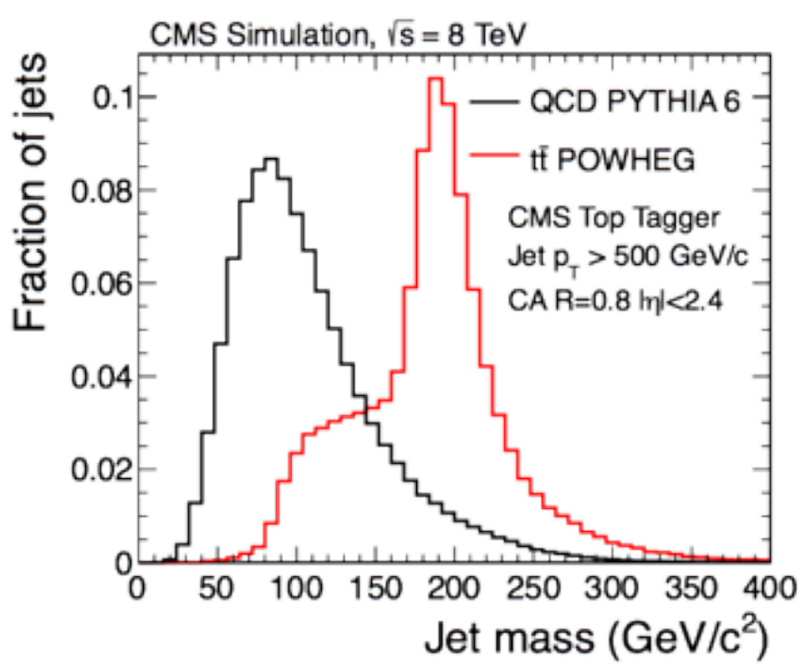
CMS Top Tagger

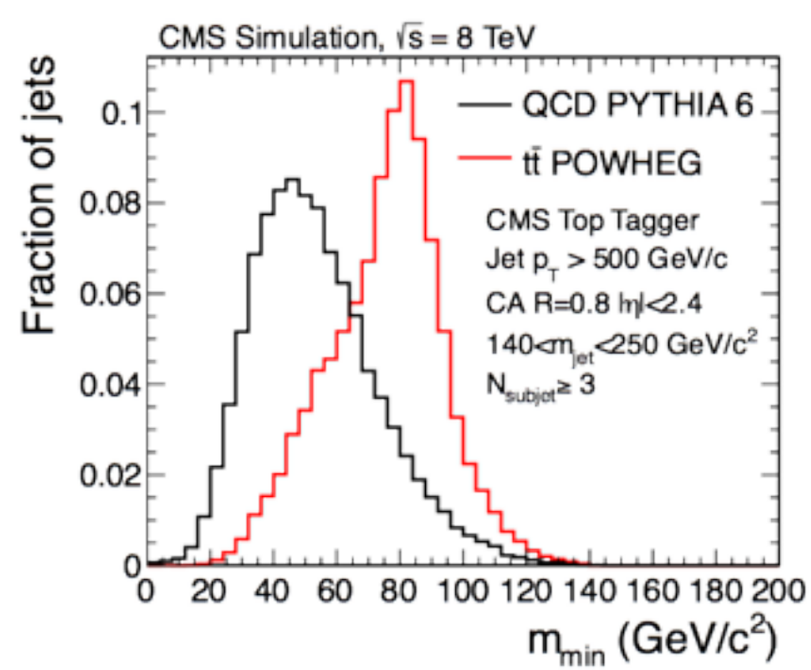
Based on JHU Top Tagger (Kaplan et al.)

- Cluster with CA8 - Decluster in two stages in order to find up to 4 subjets
- Subjets must satisfy two requirements
  - Momentum fraction criterion: $p_T^{subjet} > 0.05 \times p_T^{jet}$
  - Adjacency criterion: $\Delta R(C_1, C_2) > 0.4 - 0.0004 \times pT(C)$



- Iterative process - throw out subjets that fail momentum fraction cut and try to decluster again
- Tag top jets with selections on the CA8 jet mass, number of subjets, and minimum pairwise subjet mass ($m_{min} = \min[m_{12}, m_{13}, m_{23}]$)



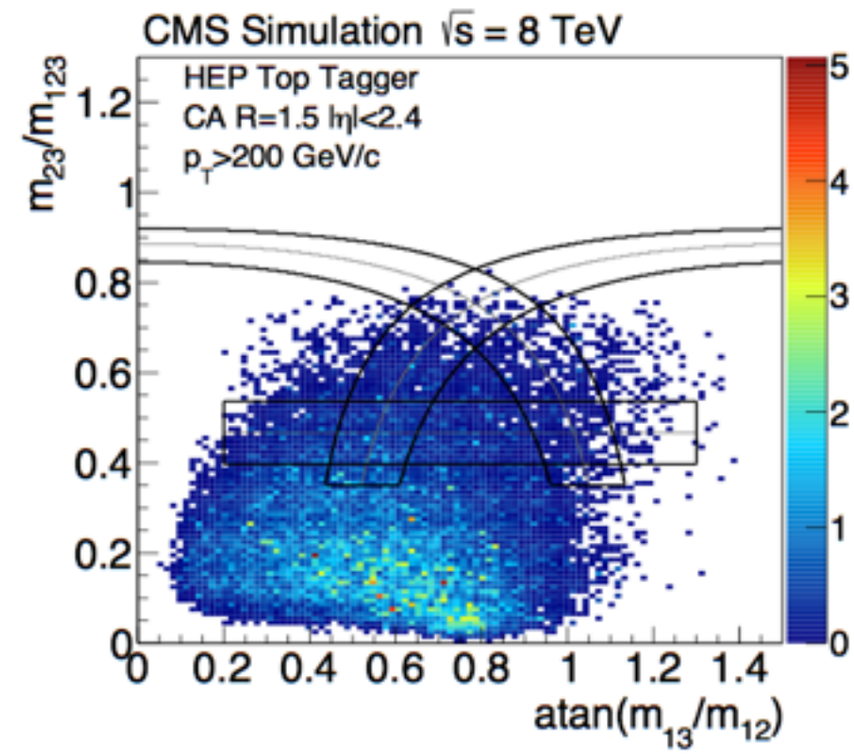


JME-13-007

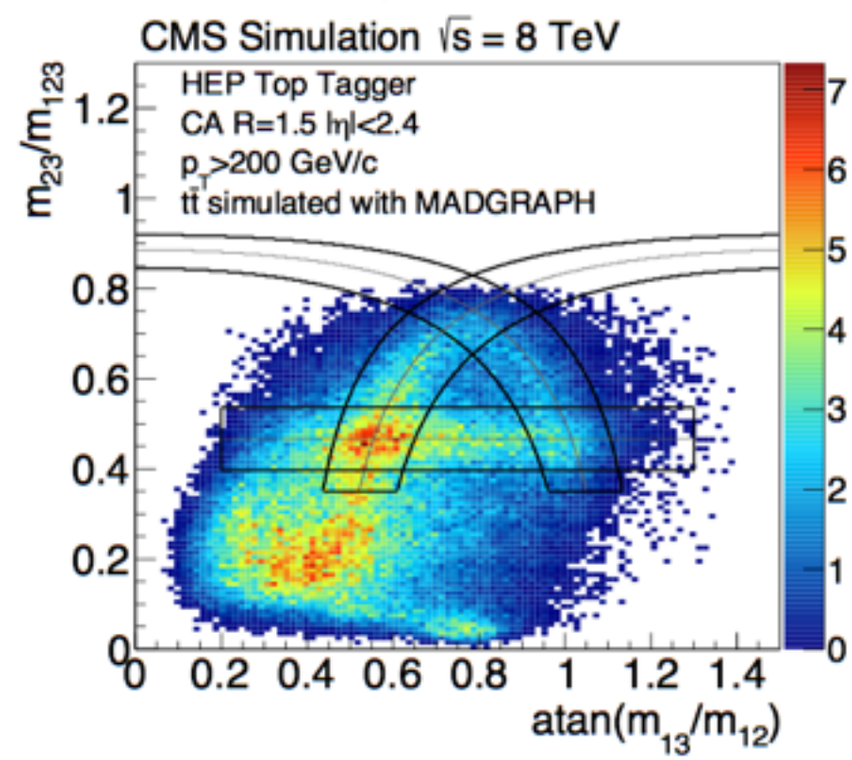
HEP Top Tagger v1

- Plehn et al. (arXiv:1006.2833)
- Very large jets (Cambridge Aachen R=1.5)
  - Increased efficiency at low p_T (p_T>200)
- Jet grooming is built in
- Top mass selection using triplet subjet mass m_{123}
- W mass selection using ratios of pairwise subjet mass

Background jet simulation

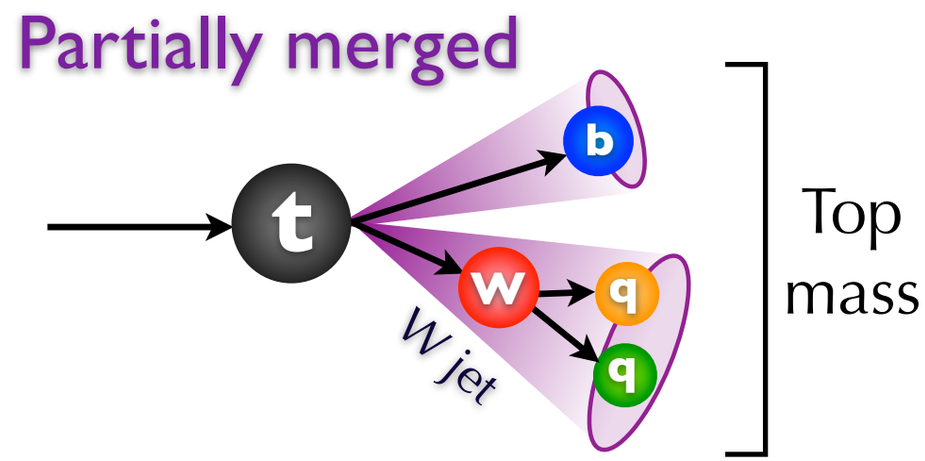


Top jet simulation

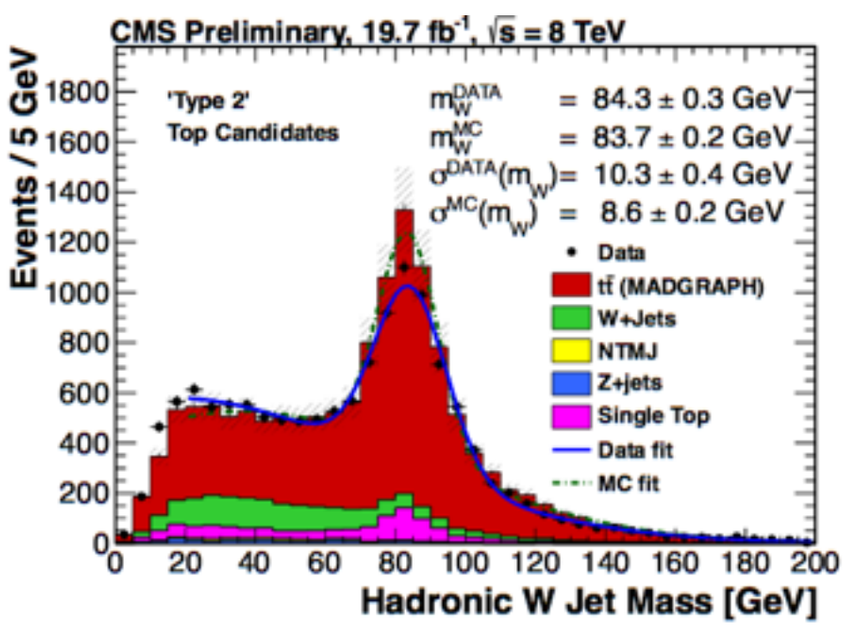


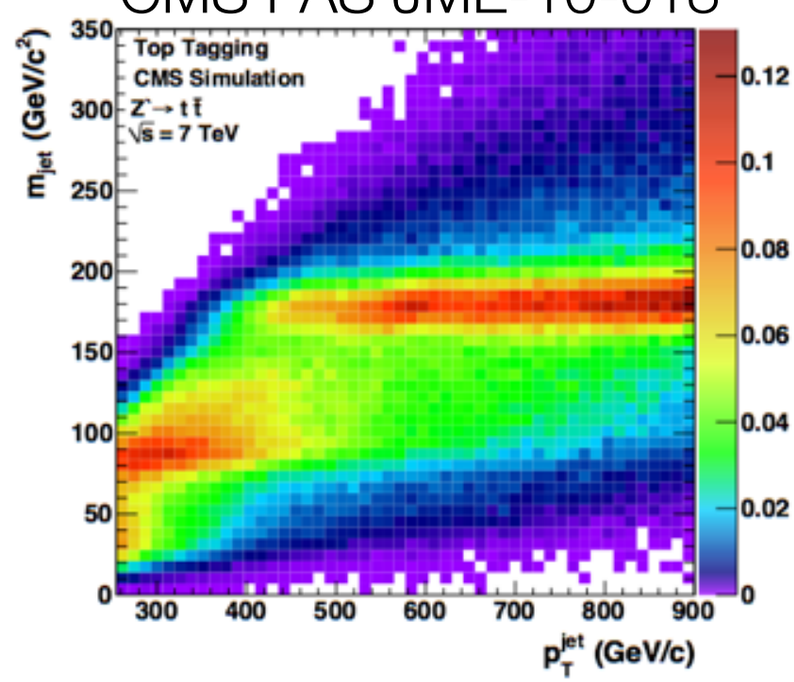
Type 2 top - partially merged

- Moderate boost: top → W jet + b jet
- W jet substructure found with jet pruning algorithm (Ellis, Vermillion, Walsh)
- Identify W jets with pruned jet mass selection and mass drop = mass of most massive subjet / mass of pruned jet
- Important substructure validation



CMS PAS JME-10-013

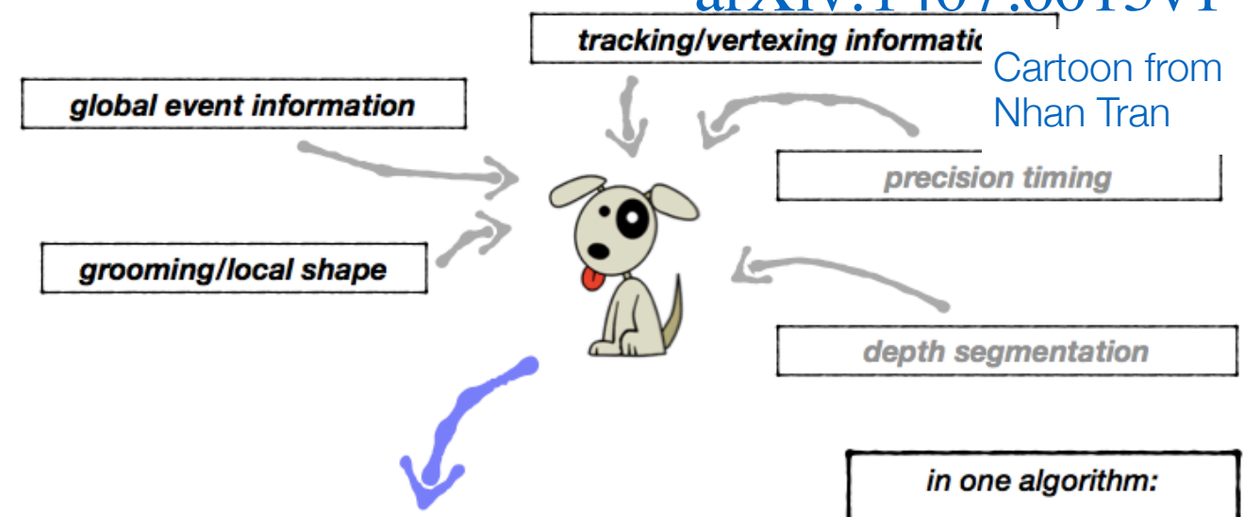




Pileup Per Particle Identification (PUPPI)

- Assign a weight to each particle equal to the probability that the particle originates from pileup
  - 0 very likely pileup → 1 very likely hard scatter
  - Multiply weight by particle 4-vector
  - Remove particles with small weight
  - Cluster jets with weighted particles
    - Also corrects the jet shape

arXiv:1407.6013v1



Cartoon from Nhan Tran

following the "jets without jets" paradigm...

Define on a ***per particle*** basis, ***before jet clustering***, a weight for ***how likely*** a particle (or jet constituent) is to be from pileup or the leading vertex, then rescale each particle four momentum by that likelihood

in one algorithm:

correct the jet pT

correct the jet mass/ shapes

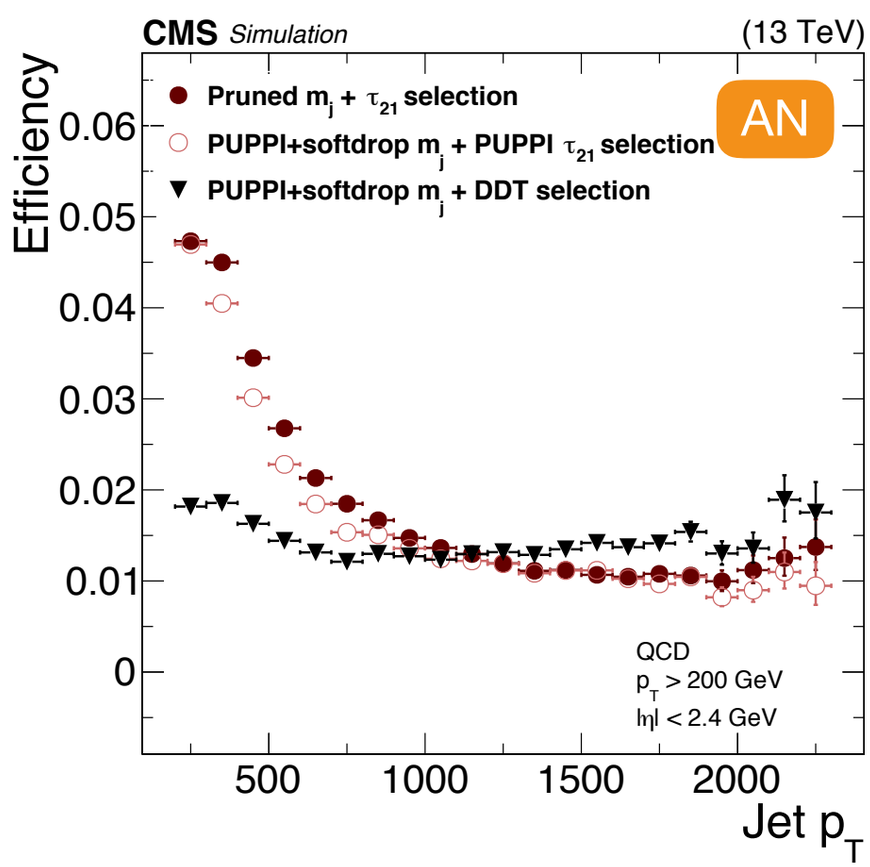
perform pileup jet ID

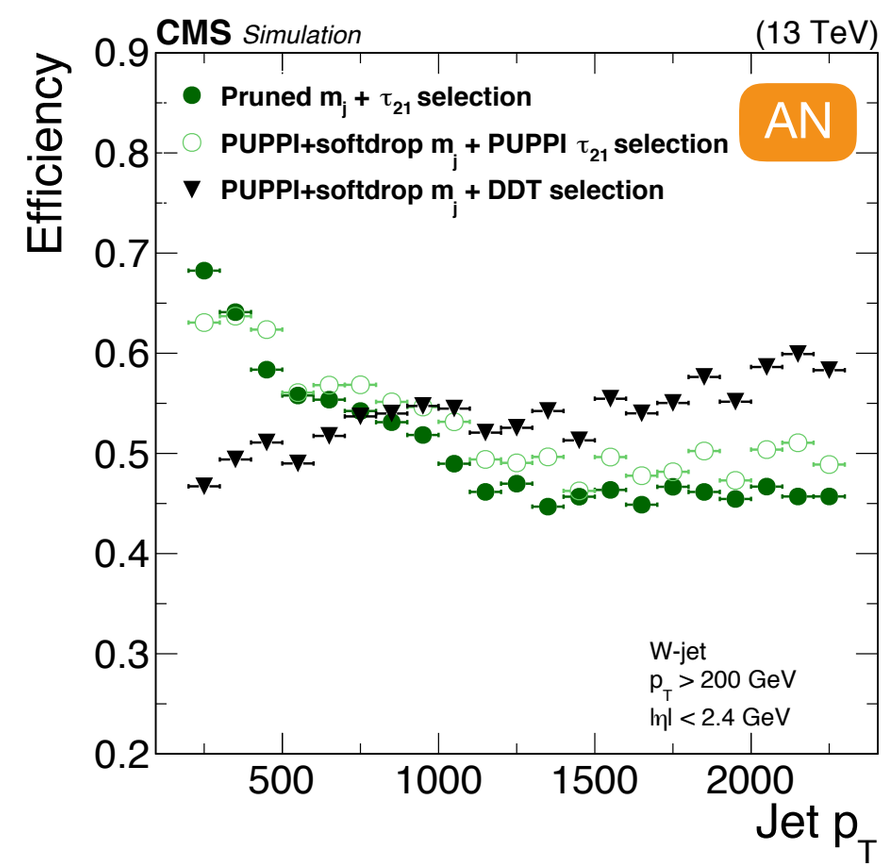
classify particles for MET determination

- PUPPI framework designed to utilize all handles to mitigate pileup
  - ρ (pileup energy density measured per event)
  - charged tracks from the hard scatter
  - charged tracks from pileup vertices
  - the local distribution of pileup with respect to particles from the leading vertex
    - LV particle radiates preferentially at small angle. Pileup from many vertices is randomly distributed in angle and its pT spectrum falls more rapidly

W-tagging efficiency in simulation

- DDT N-subjettiness: new method to transform a variable such that the correlations with mass/p_T are reduced
  - Leads to improved background estimation methods
    - * $\tau_{21,DDT} = \tau_{21,\mathrm{PUPPI}} + 0.063 \times \log\left(\frac{m^2_{\mathrm{PUPPI+SD}}}{p_{T,\mathrm{PUPPI+SD}}}\right)$
- Study PUPPI+softdrop + DDT* (see talk by Nhan) , for DDT* ≤ 0.52
- Background mis-tagging rate stable. Signal efficiency increases with p_T



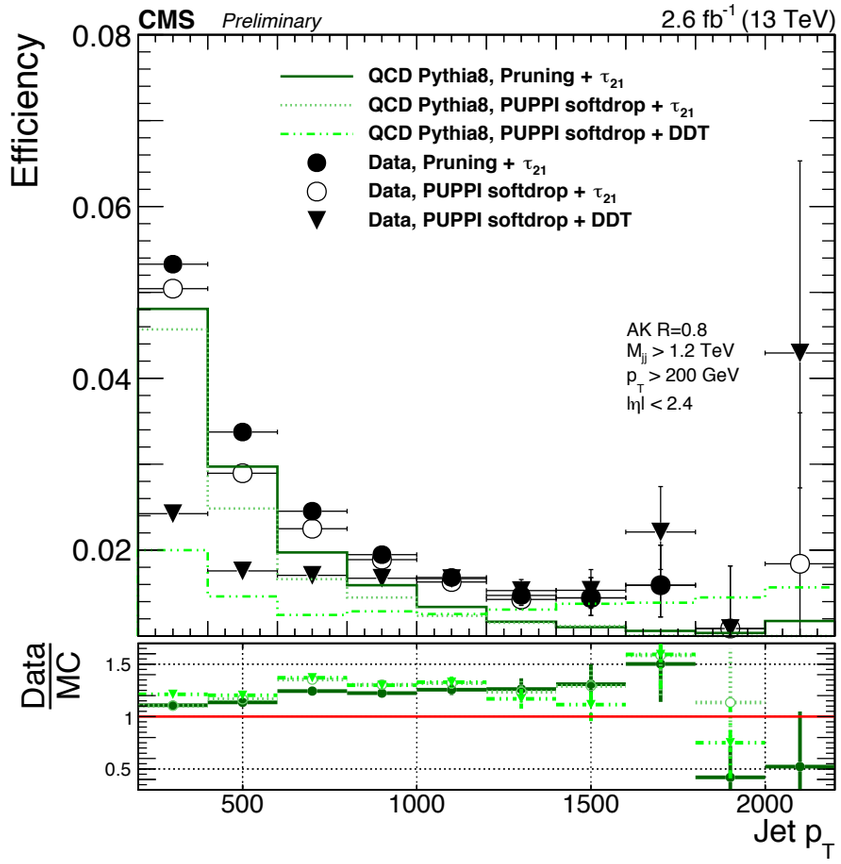


Plots not currently in PAS. To discuss. (DDT is used in analysis)

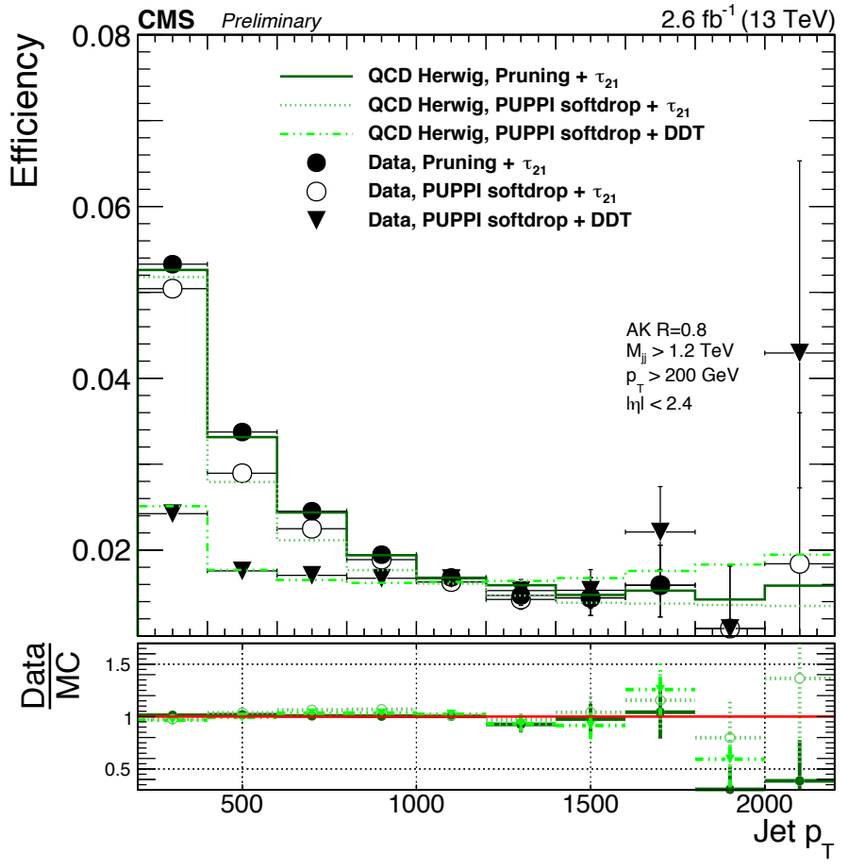
W-tagging τ_2/τ_1 fake rate

- Compare fake rate with QCD Pythia8 and Herwig++ simulation
- Pythia8 Tune CUETMP8 shows disagreement similar to Pythia6 Tune Z2* in Run 1, but worse than Pythia8 Tune 4C in Run 1
- Herwig++ Tune CUETHS1 shows perfect agreement, better than Tune 23 in Run 1

Pythia 8



Herwig++



PAS

PUPPI for jets and MET

arXiv:1407.6013

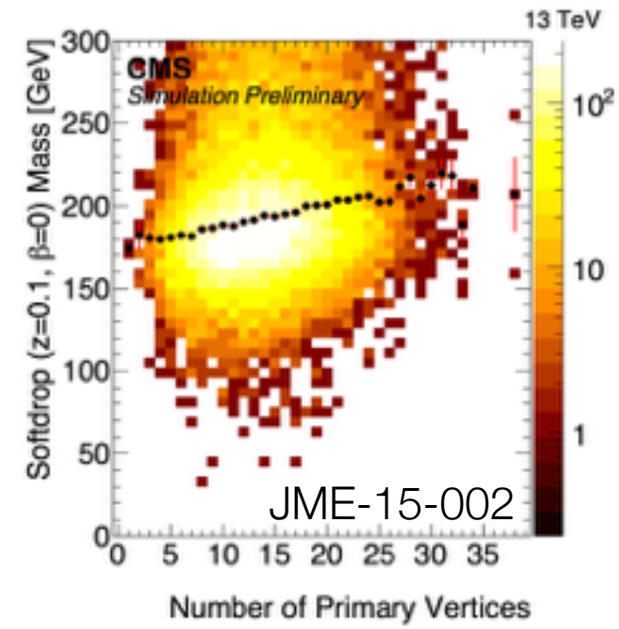
JME-14-001

https://twiki.cern.ch/twiki/bin/view/CMS/PUPPI

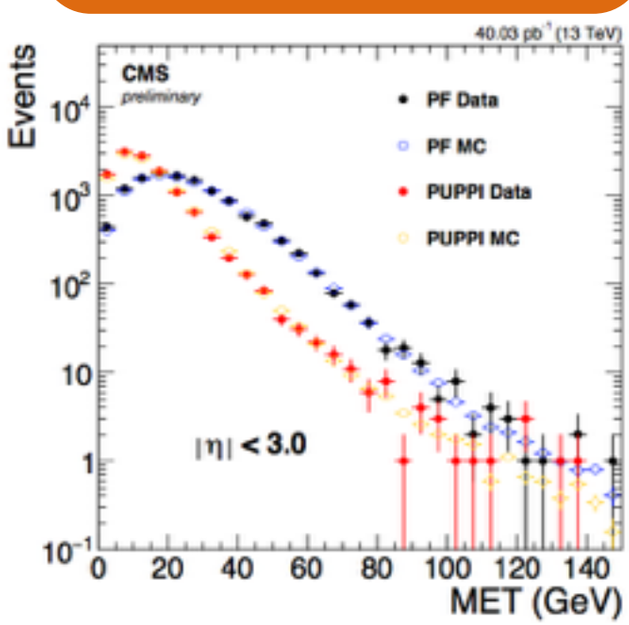
Daniele Bertolini, Philip Harris,
Matthew Low, Nhan Tran
Satoshi Hasegawa

- Remove pileup before clustering jets
  - Assign each particle a weight based on the likelihood it originated from pileup
    - Pileup tends to be soft and high angle
- 76X tuning integrated into 80X ✅
  - Improved performance in forward region

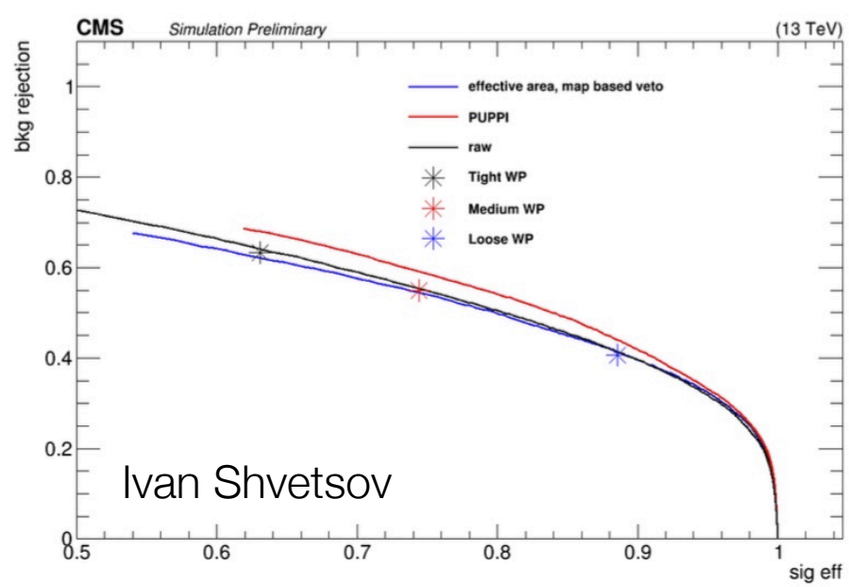
PUPPI jet mass stable with pileup



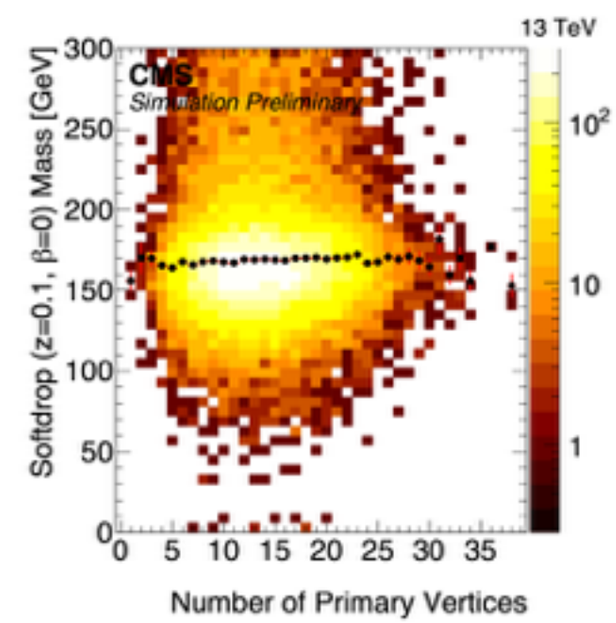
Improved MET resolution



Best performing algorithm for lepton and photon isolation



Ivan Shvetsov

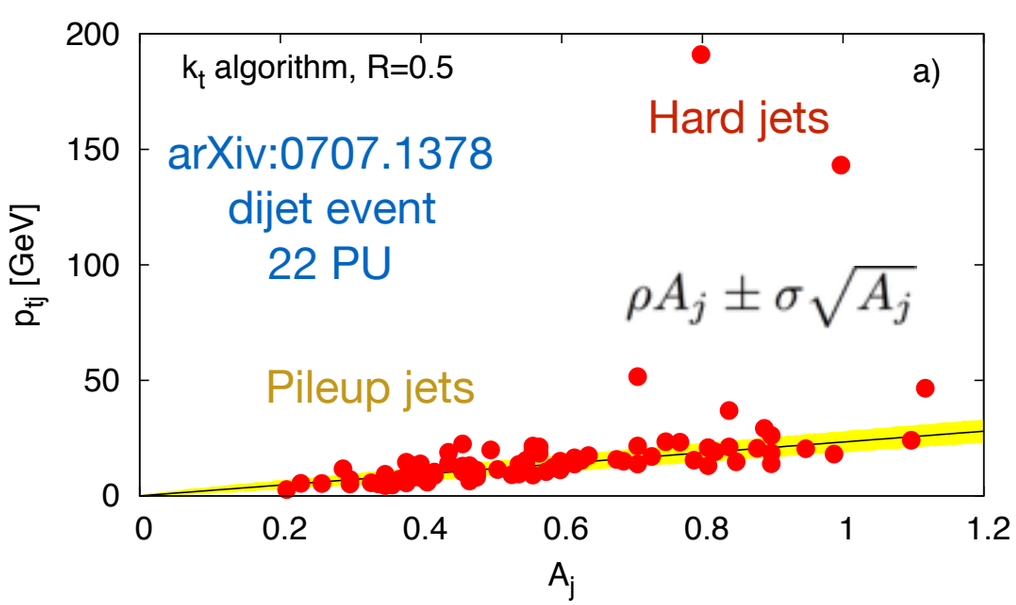


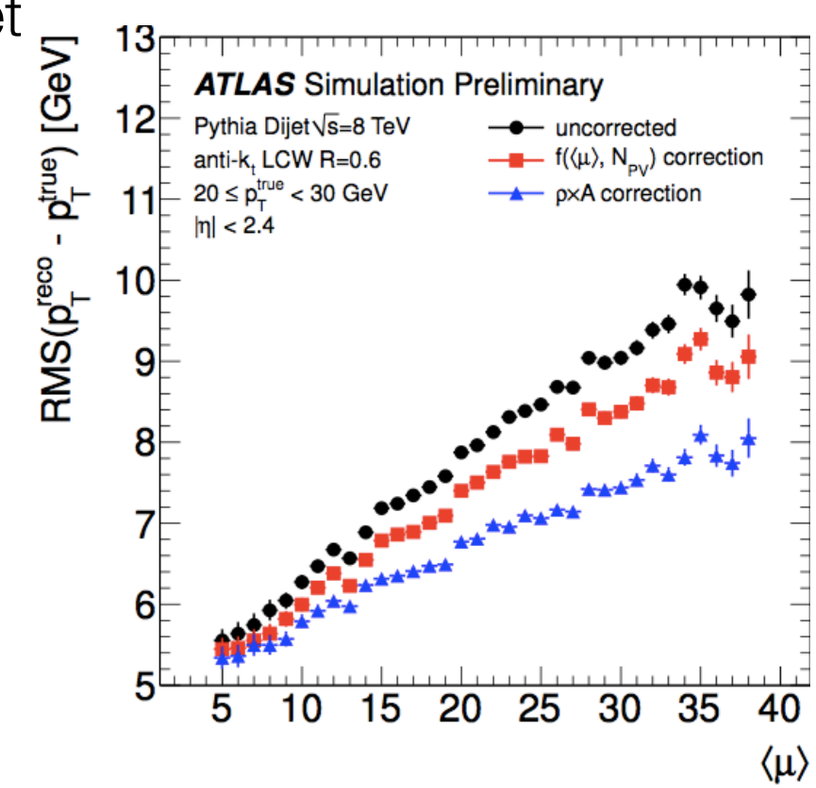
Pileup Subtraction

- Charged particles from non primary vertices can be identified using tracking/particle flow
  - The energy contribution of these particles to jets can be subtracted (*charged hadron subtraction*)
- Pileup contribution from jets can be subtracted by measuring the energy density of pileup per event
  - *Jet area method*
  - Pileup fluctuations within the event reduce the jet pT resolution

Jet area method
Cacciari, Salam, Soyez

$$p_{\mathrm{T}}^{\mathrm{sub}} = p_{\mathrm{T}} - \rho A$$
$$p_{\mu}^{sub} = p_{\mu} - \rho A_{\mu}$$
$$\rho = \mathrm{median}[p_{\mathrm{T}j}/A_j]$$





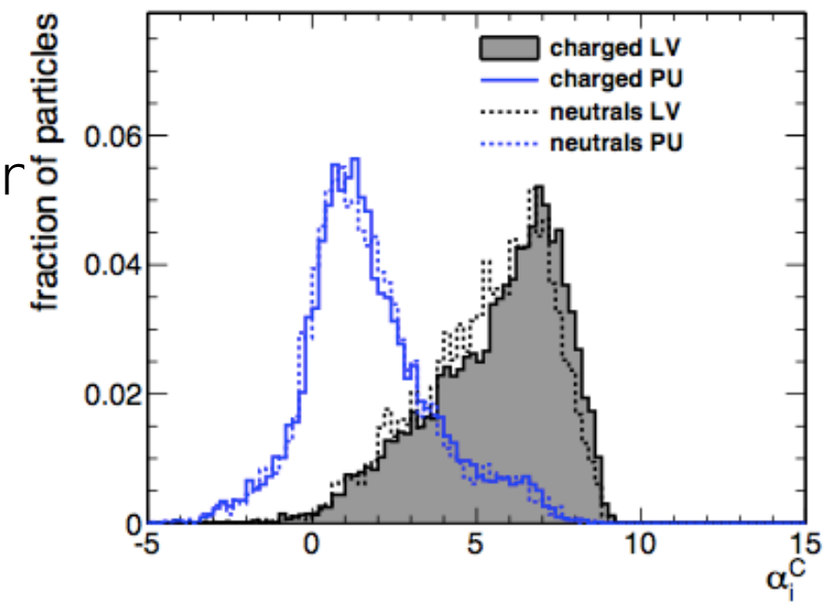
PUPPI Algorithm

1. Calculate α_i for each particle i

$$\alpha_i = \log \sum_{j \in \text{event}} \xi_{ij} \times \Theta(R_{\min} \leq \Delta R_{ij} \leq R_0),$$

where $\xi_{ij} = \frac{p_{Tj}}{\Delta R_{ij}}$.

- Sum of $p_{Tj}/\Delta R_{ij}$ for each particle j within an annulus around particle i
  - R_0 determines the outer cone size
  - R_{min} chosen based on detector resolution



- Choice of ξ_{ij}
  - ~1/ΔR : Collinear radiation from a particle from the hard scatter is mostly radiated at small angles while pileup has no angular preference (bigger ΔR → smaller α → less likely to be pileup)
  - ~ p_{Tj} : pT spectrum of pileup pileup falls off faster than pT of particles from the hard scatter (smaller pT → smaller α → less likely to be pileup)

- In tracking region sum over charged particles from the LV, in forward region sum over all particles

PUPPI Algorithm

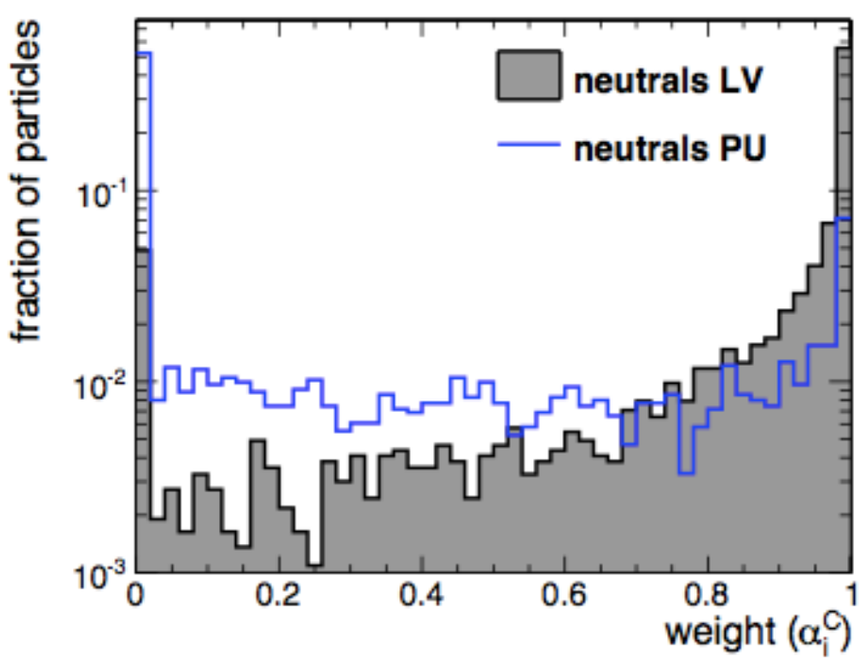
2. All charged pileup particles are assigned a weight $w_i = 0$ and all charged leading vertex particles are assigned a weight $w_i = 1$

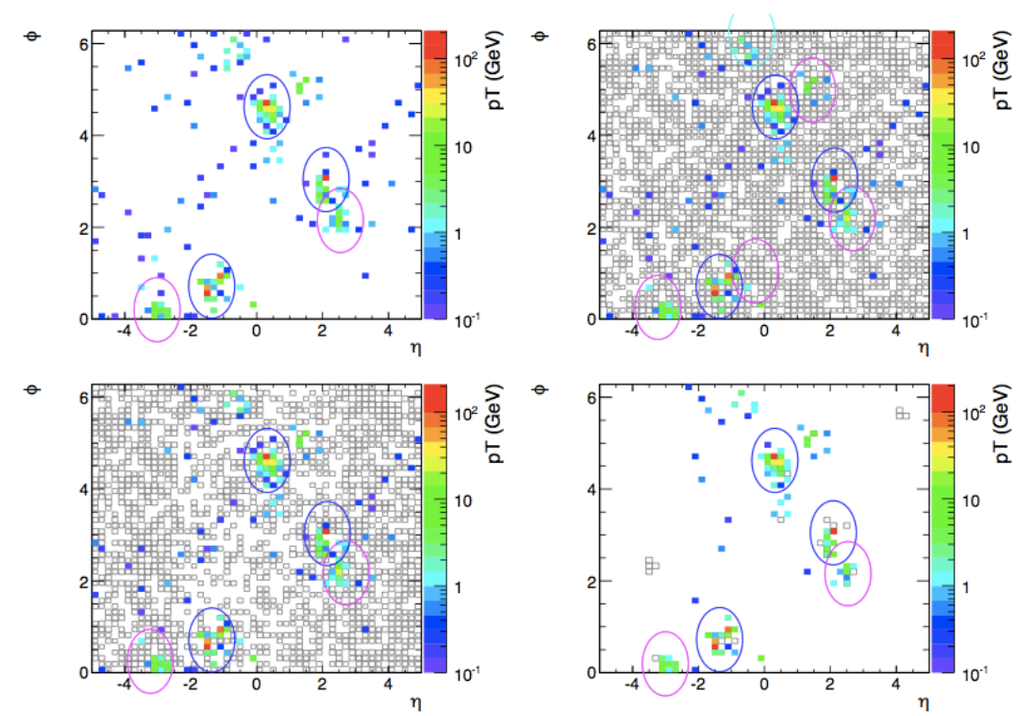
3. The weights of all other particles are calculated using:

$$\chi_i^2 = \Theta(\alpha_i - \bar{\alpha}_{\mathrm{PU}}) \times \frac{(\alpha_i - \bar{\alpha}_{\mathrm{PU}})^2}{\sigma_{\mathrm{PU}}^2}$$

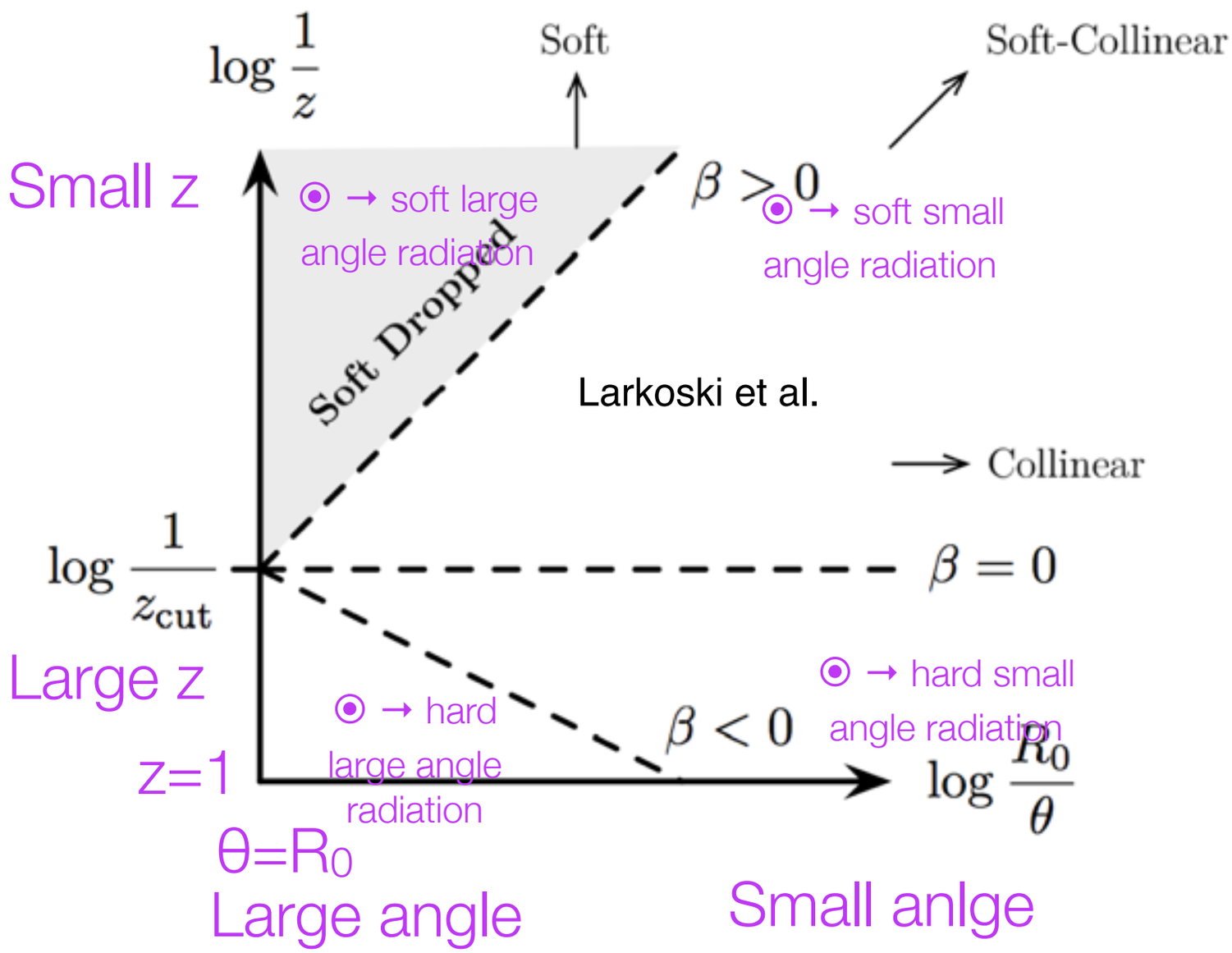
4. The four-momentum of each particle is rescaled by its weight $p_i^\mu \to w_i \times p_i^\mu$

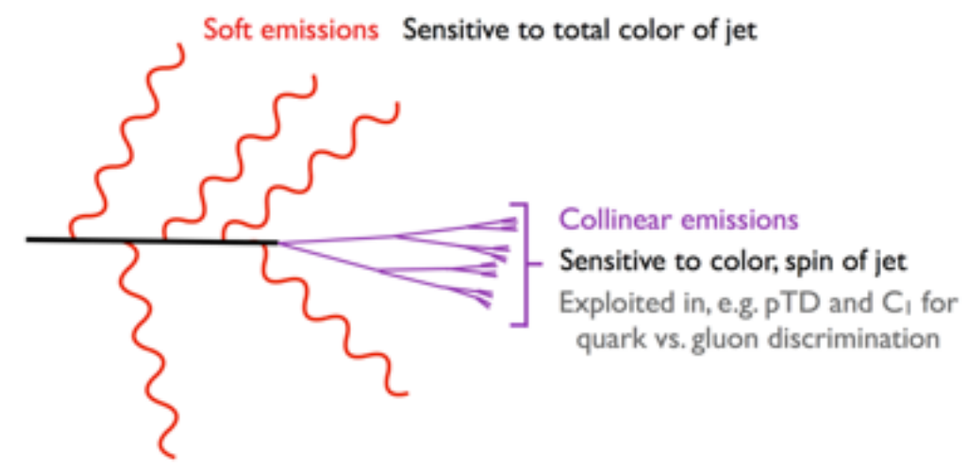
5. Particles with small weights $w_i < w_{cut}$ or with low (rescaled) transverse momentum $p_{Ti} < p_{T,cut}$ are discarded.





Soft drop



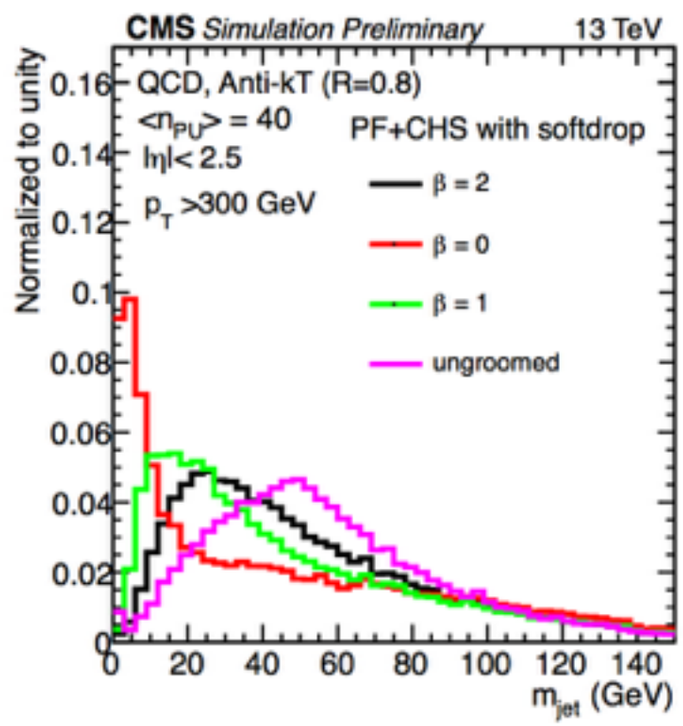


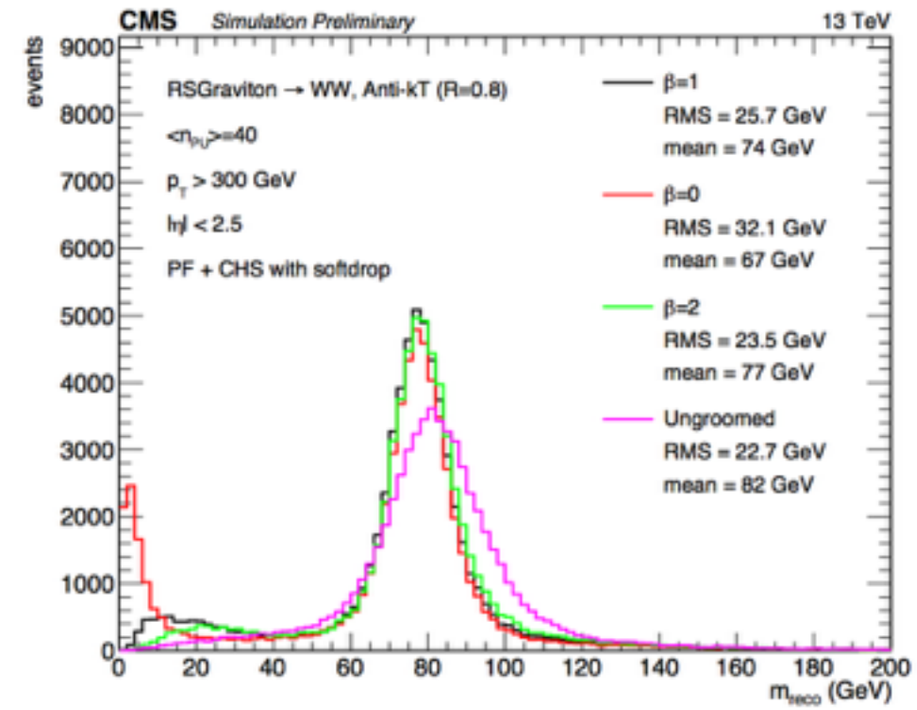
$0 \leq z \leq 1$ is the energy fraction

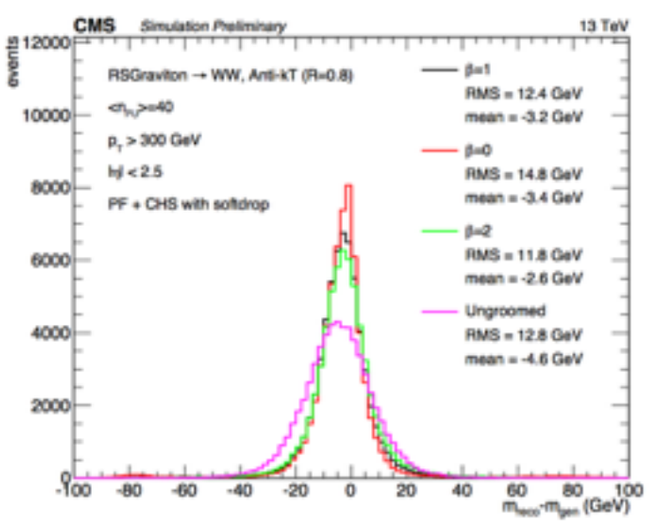
$\theta \leq R_0$ is the angle of the emission

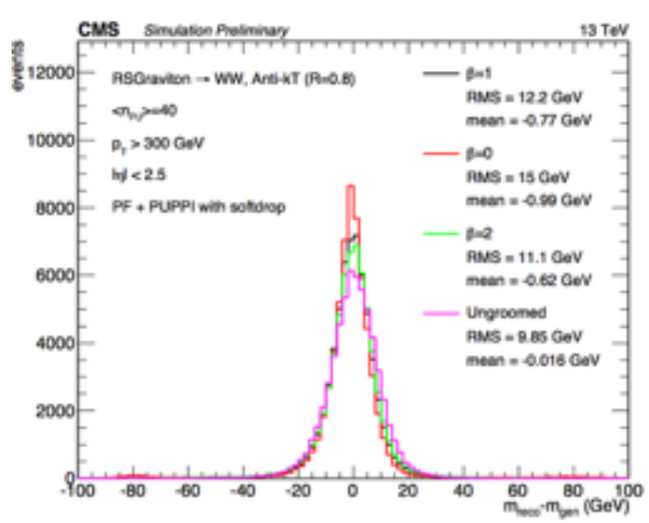
JME-14-001/002

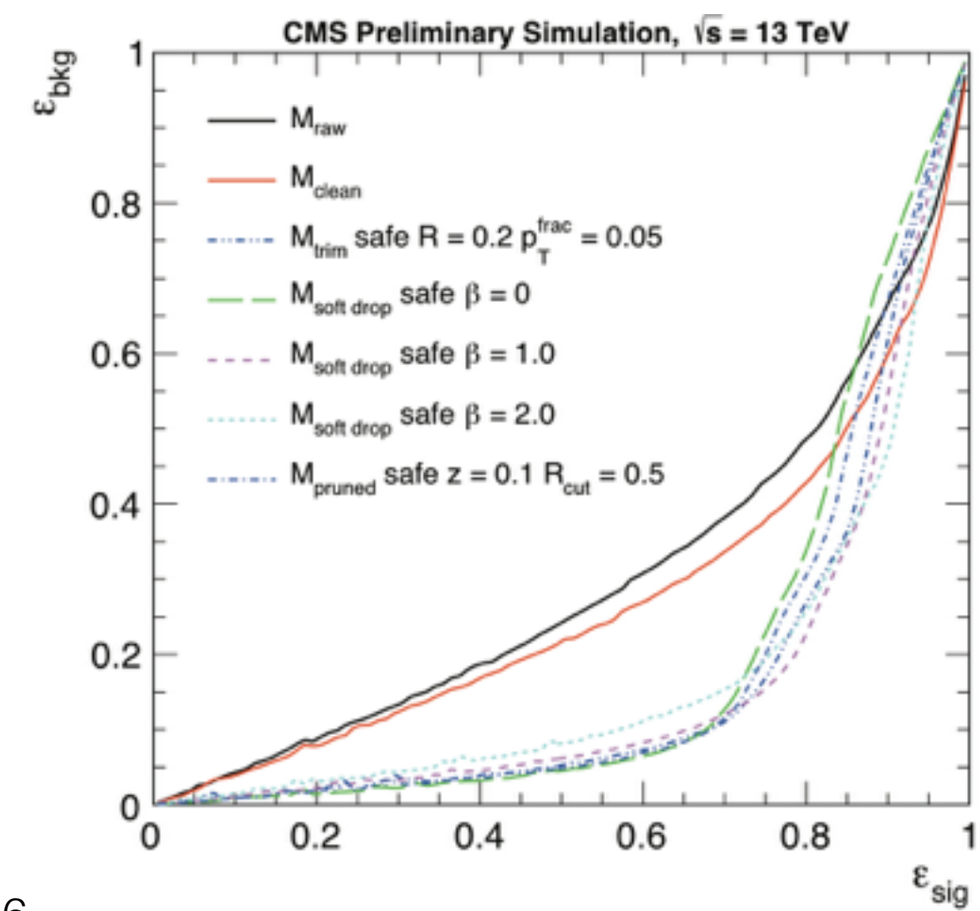
- Soft drop studied with CHS+ safe subtraction, PUPPI etc.
- No z_{cut} variation (fixed 0.1)
- For efficiencies <0.7, β=0 showed best performance
  - Current CMS default β=0, z_{cut} =0.1







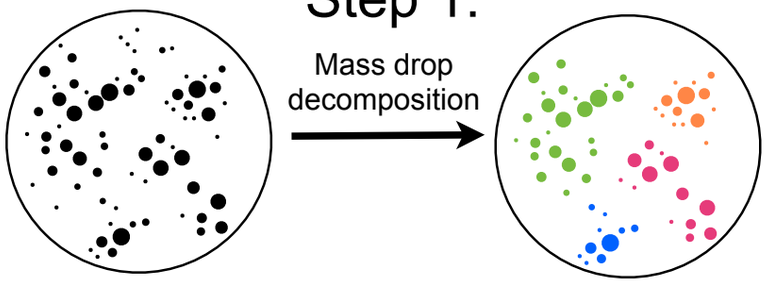




HEP Top Tagger algorithm

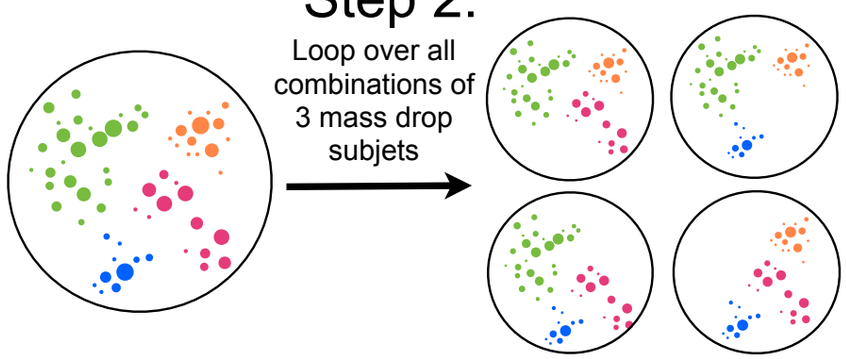
Step 1:

Mass drop decomposition



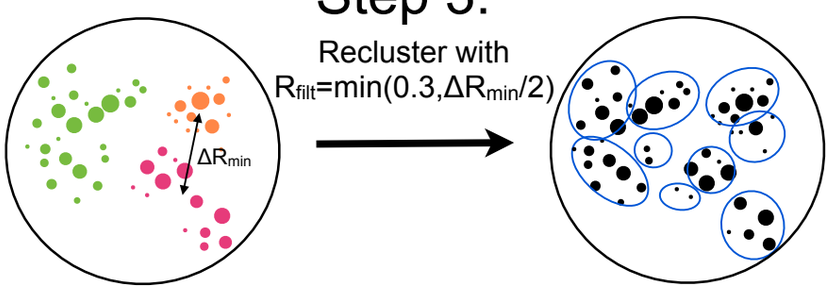
Step 2:

Loop over all combinations of 3 mass drop subjets



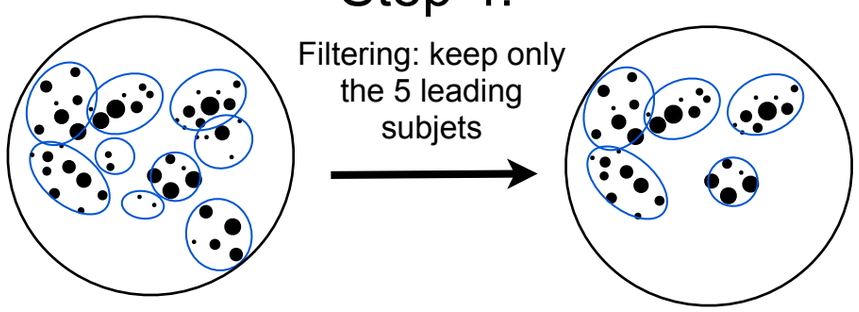
Step 3:

Recluster with $R_{filt}=\min(0.3,\Delta R_{min}/2)$



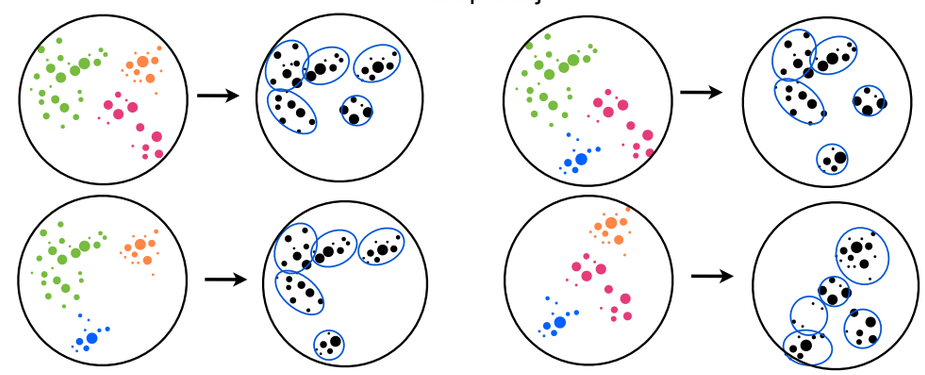
Step 4:

Filtering: keep only the 5 leading subjets



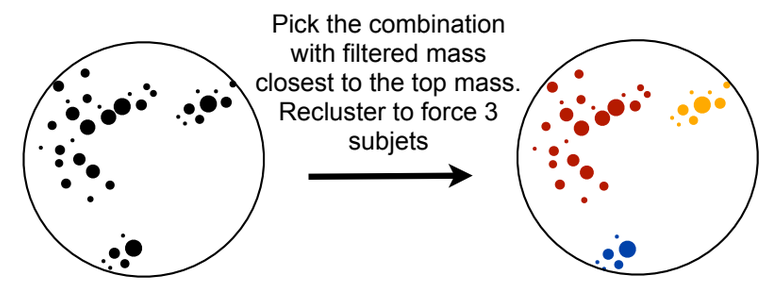
Step 5:

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



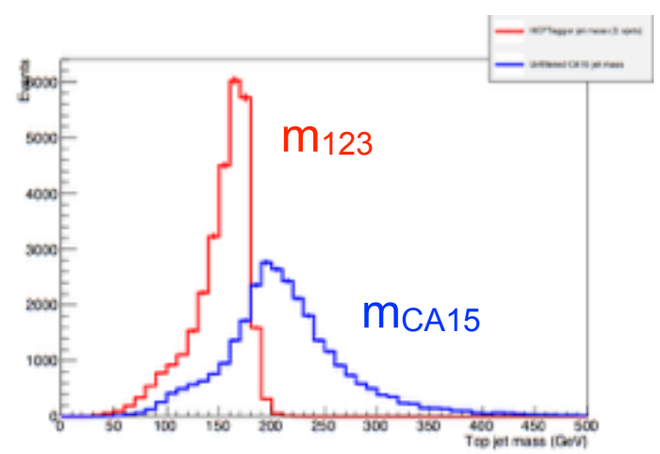
Step 6:

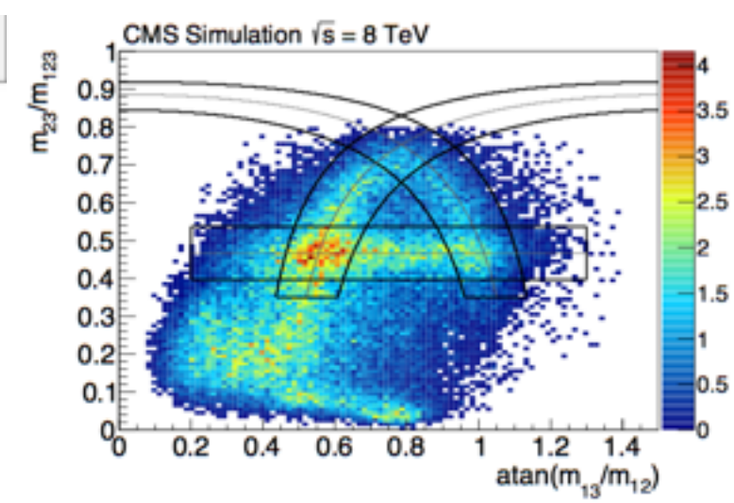
Pick the combination with filtered mass closest to the top mass. Recluster to force 3 subjets



Step 7:

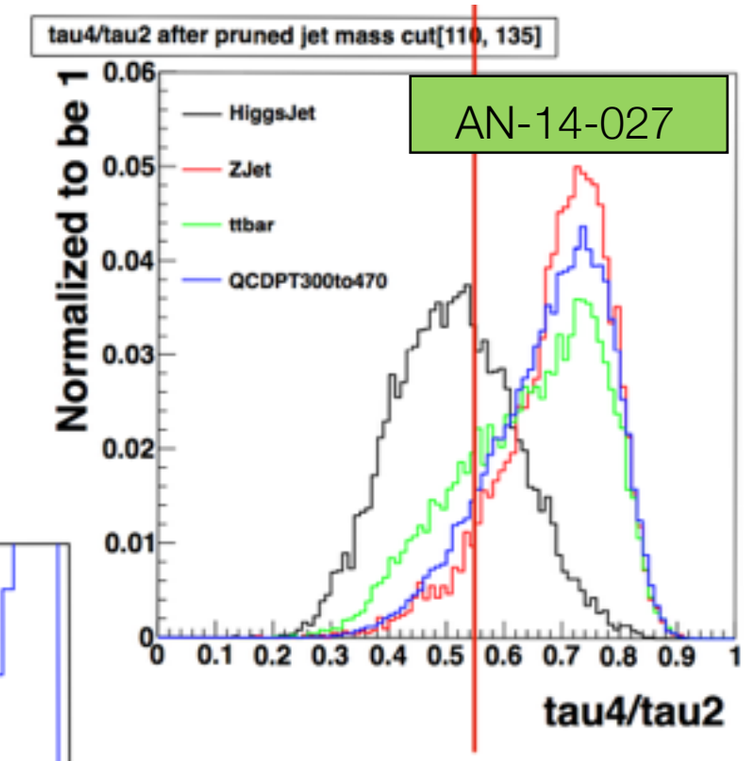
Apply cuts



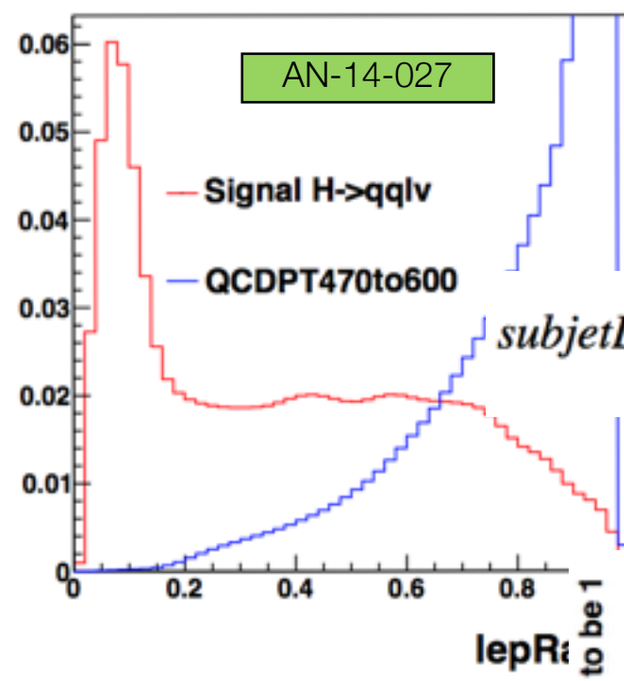


4-prong tagging

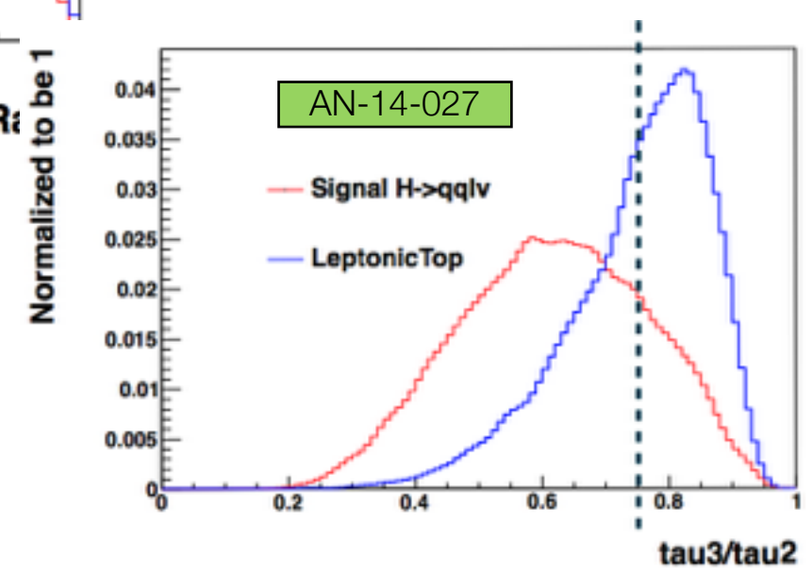
- Example: Boosted H → W*W → qqqq or qqlv
- Hadronic case
  - Pruned Higgs mass selection used to reduce QCD and boosted top/Z
  - τ_4/τ_2 shows promise to remove QCD, Z jets and unmerged or partially merged tops



- If real W decays leptonically
  - New ideas from Petar, Yongjie and JHU theorists
  - subjetLepRatio and N-subjettiness with clustered leptons
    - If real W, lepton will be hard and it will have a significant fraction of one subjets momentum:
    - N-subjettiness with clustered leptons
      - τ_3/τ_2 can be used to separate H → W*W → qqlv from boosted leptonic tops



$$subjetLepRatio = \frac{(closetSubjet - lepton).pt()}{closetSubjet.pt()}$$

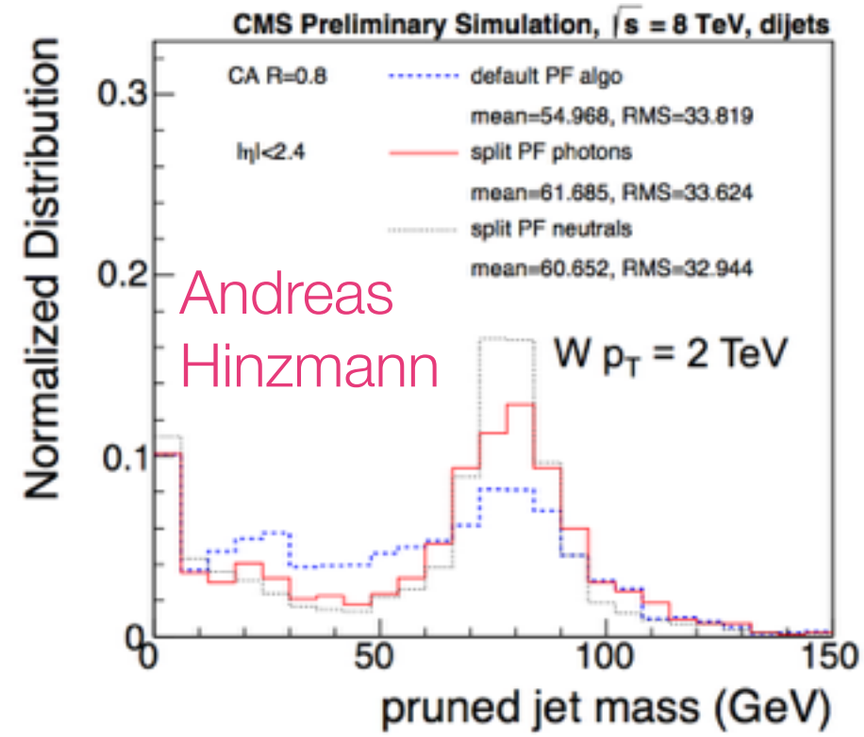


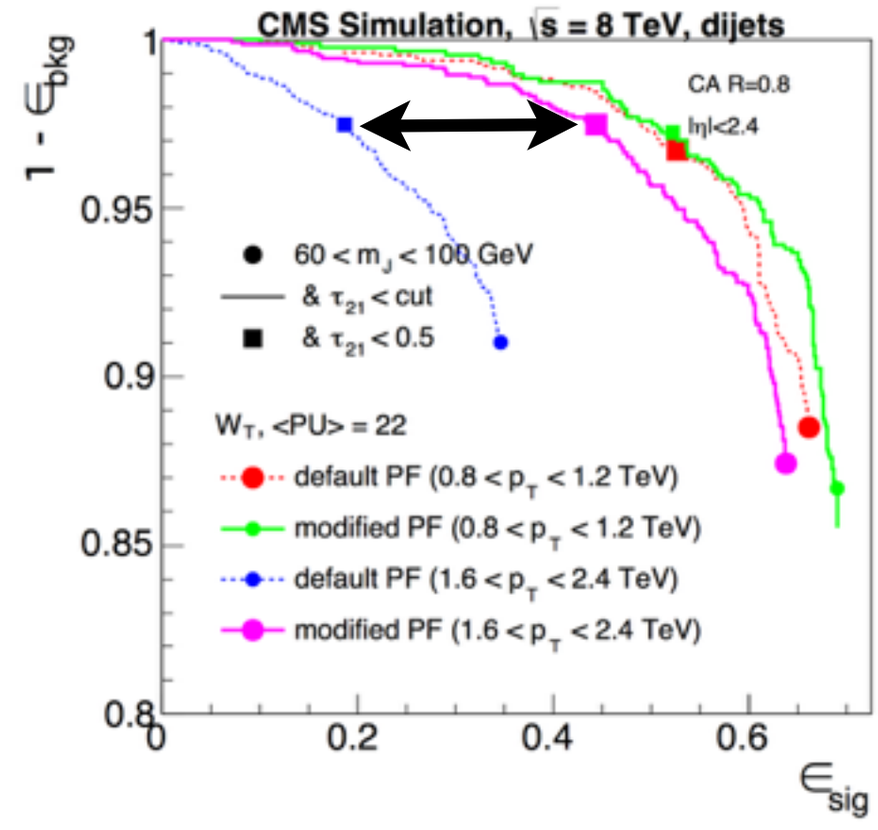
Jet substructure challenges in Run 2

- High pT
  - Close tracks, granularity limits
  - Improvements to particle flow provide new confidence in our tools!
- High pileup
  - Grooming algorithms
  - Per-particle pileup removal tools
    - CHS, PUPPI, soft killer
  - Rho-based pileup removal
    - Jet shape subtraction

High p_T tagging

- Run 1 - problems observed with substructure variables at very high p_T (>1.5 TeV)
  - Pruned mass, N-subjettiness, minMass, etc.
- Numerous changes to particle flow over the past year have recovered these tools
  - Great work by Andreas and others
  - Biggest improvement: splitting “merged photons” and “merged neutral hadrons” into ECAL clusters
    - “Merged photon” - If there is energy in ECAL clusters after track energy removed and there is no HCAL energy, all local ECAL clusters are formed into one photon
    - Fix: Each ECAL cluster is one photon
    - With these changes pruned jet mass and N-subjettiness regain good performance above 3.5 TeV
    - To be included in 72X





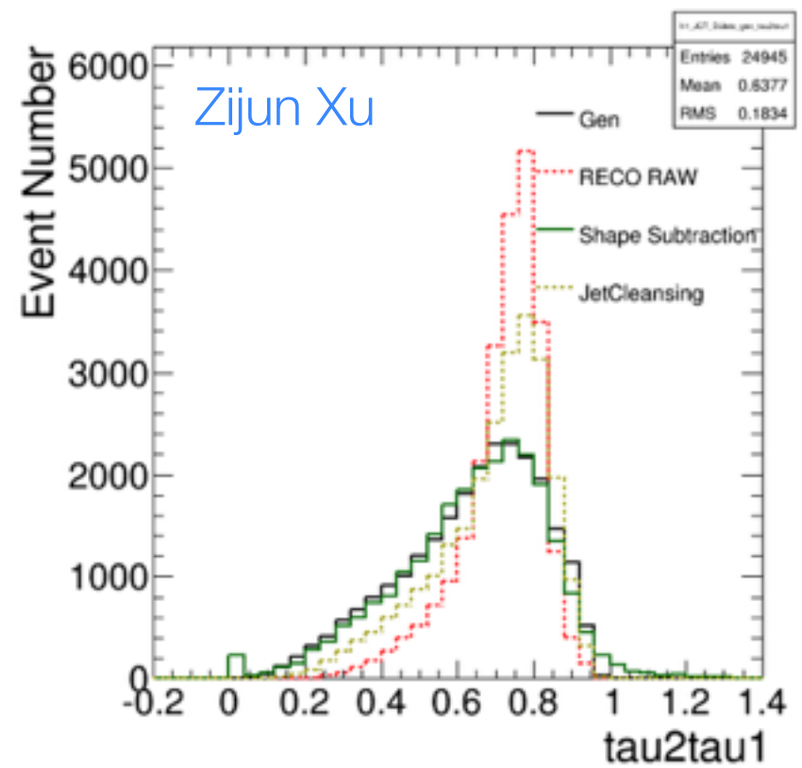
Dealing with pileup

- Run 1
  - rho-based area subtraction

Cacciari, Salam, Sovez

$$p_{\mathrm{T}}^{\mathrm{sub}} = p_{\mathrm{T}} - \rho A$$
$$p_{\mu}^{sub} = p_{\mu} - \rho A_{\mu}$$
$$\rho = \mathrm{median}[p_{\mathrm{T}j}/A_j]$$

- Shape subtraction
  - Improves area subtraction method to not just correct the jet 4-vector but also jet shape variables such as N-subjettiness
  - Shape subtraction estimates the sensitivity of a variable to pileup and then corrects it
  - Should be useful for jet mass and N-subjettiness

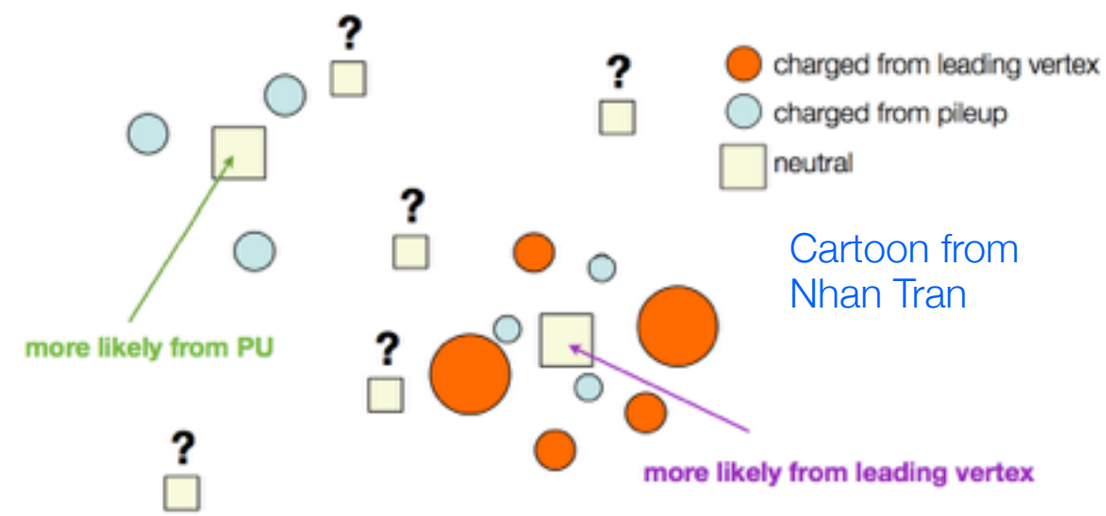


- Per particle pileup removal
  - CHS, PUPPI, soft killer

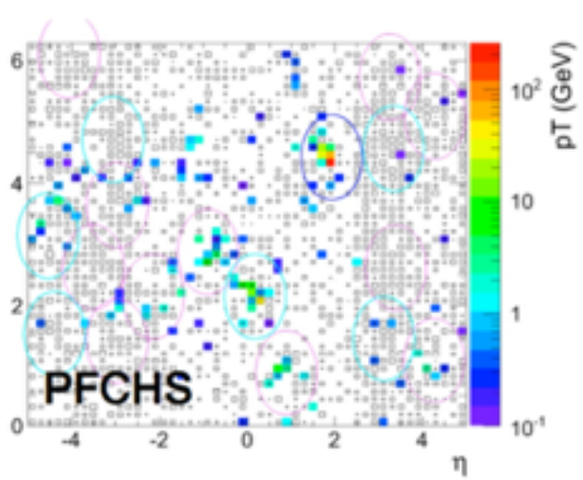
PUPPI

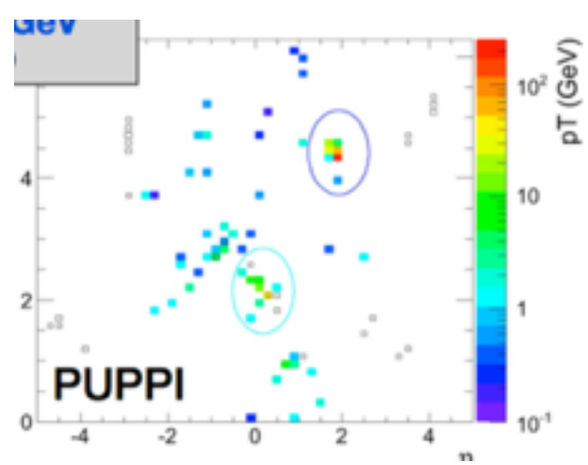
Daniele Bertolini, Phil Harris, Matt Low, Nhan Tran

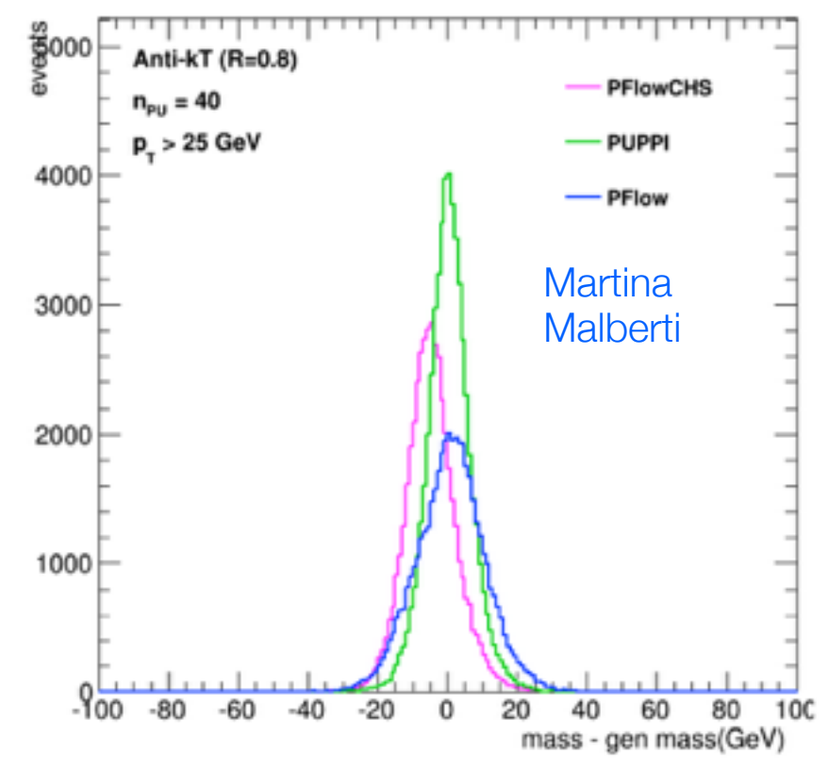
- PileUp Per Particle Identification
  - On a per particle basis (before jet clustering), use the charge track information to calculate the probability that a particle is from pileup and assign a weight
- Improves jet mass resolution!
- Work ongoing to tune PUPPI parameters and apply JEC (Martina Malberti)
- Studies aiming for BOOST 2014



Cartoon from Nhan Tran



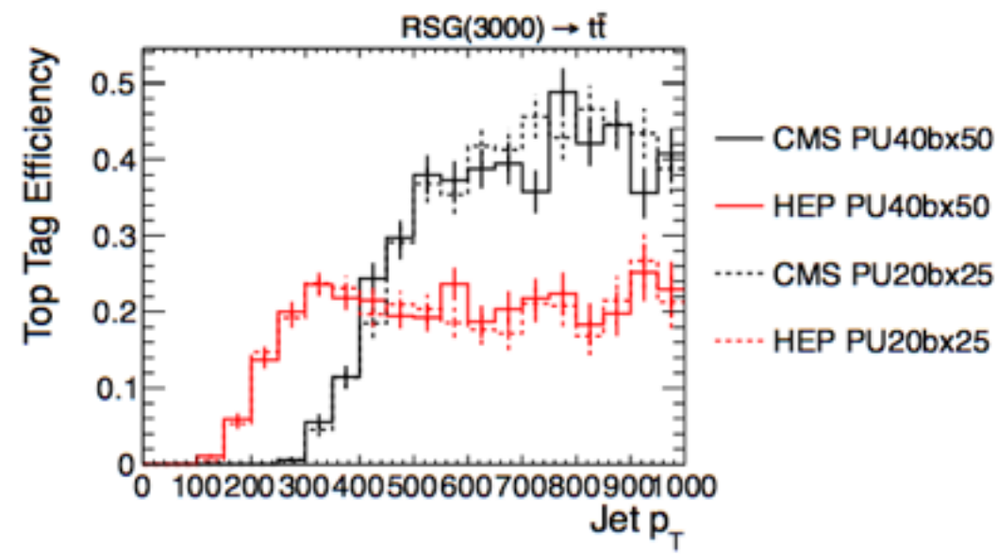




Martina Malberti

Jet substructure studies for B2G CSA14

- B2G needs to evaluate how changes to physics objects (ex. CA8 → AK8), updated particle flow, and high pileup scenarios affect our analyses
  - Redevelop B2G workflow
    - CMS top tagger etc. now in RECO
    - Use jet toolbox
  - Develop miniAOD workflow



- Work in parallel with JetMET group
  - JetMET is using CSA14 samples for current studies of jet grooming, jet tagging observables, and PUPPI
    - “expert Ntuples” - jets are clustered by hand with fastjet
  - B2G’s role is to test the standard tools in an analysis context

Resources

- Jet toolbox (https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetToolbox)
  - easy to access substructure tools
    - Pileup Jet ID
    - q/g Tagging
    - Groomed mass
    - Pruning, trimming, filtering
    - Top-Tagging
    - N-subjettiness
    - Q-jets
- CMS tutorials
  - CMS DAS (https://twiki.cern.ch/twiki/bin/viewauth/CMS/WorkBookExercisesCMSDataAnalysisSchool)
  - LPC HATS
    - https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHATSJetSubstructure2014
    - https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHATSJetSubstructure

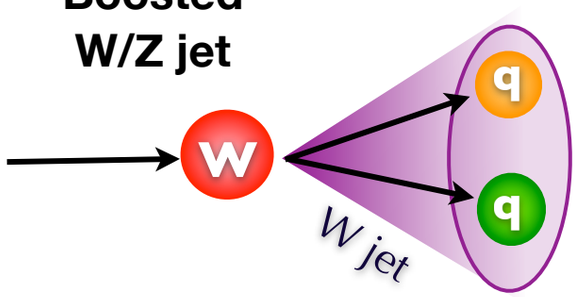
Conclusion

- Jet substructure was an important component of the success of B2G in run 1 and will be even more important in run 2
- It is important to test these tools and prepare for data
- New recommended tools are coming soon

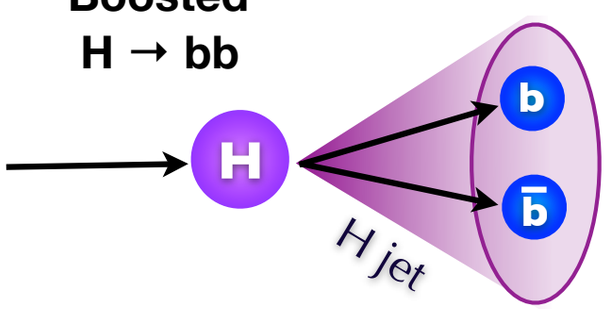
Backup

Boosted Menu (only a sample)

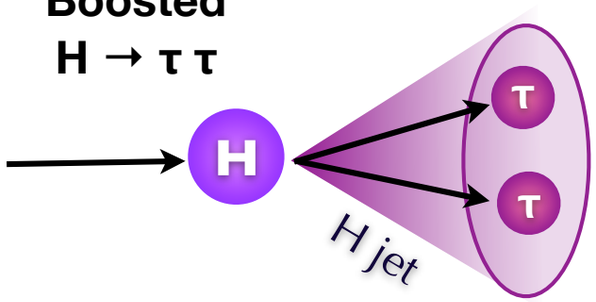
Boosted
W/Z jet



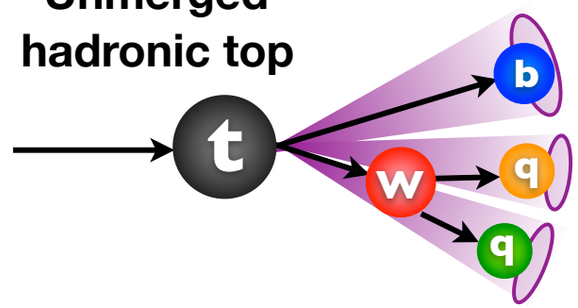
Boosted
H → bb



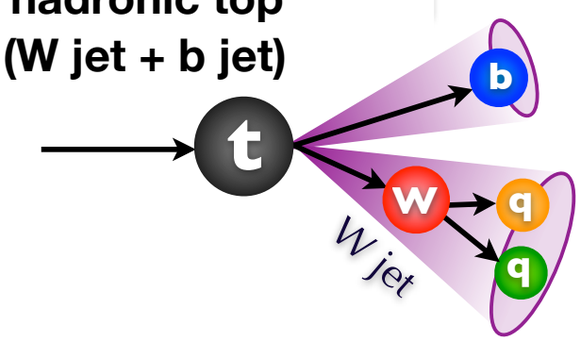
Boosted
H → τ τ



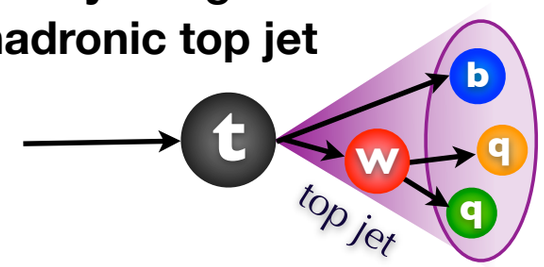
Unmerged
hadronic top



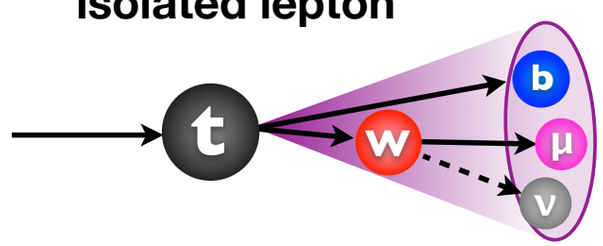
Partially merged
hadronic top
(W jet + b jet)



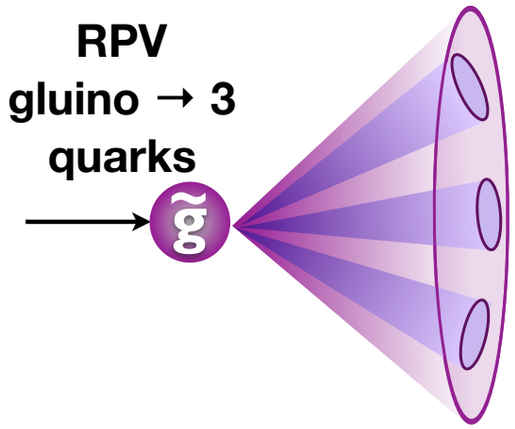
Fully merged
hadronic top jet



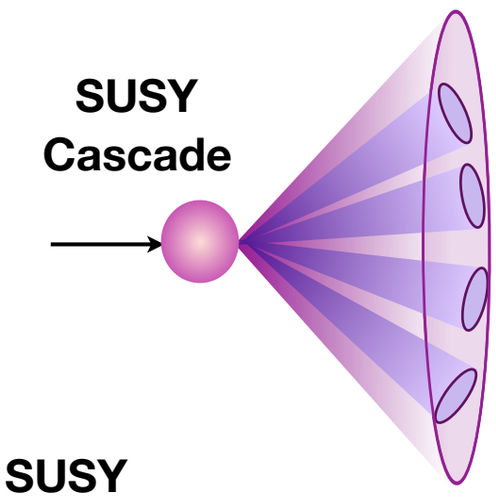
Leptonic top with non-
isolated lepton



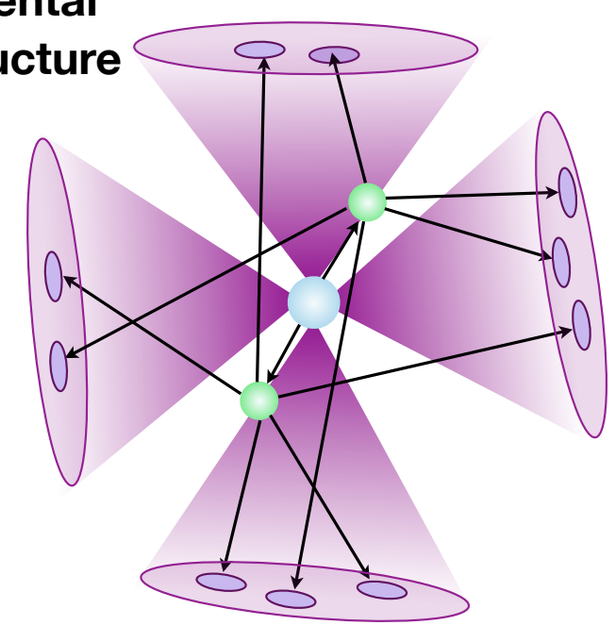
RPV
gluino → 3
quarks



SUSY
Cascade



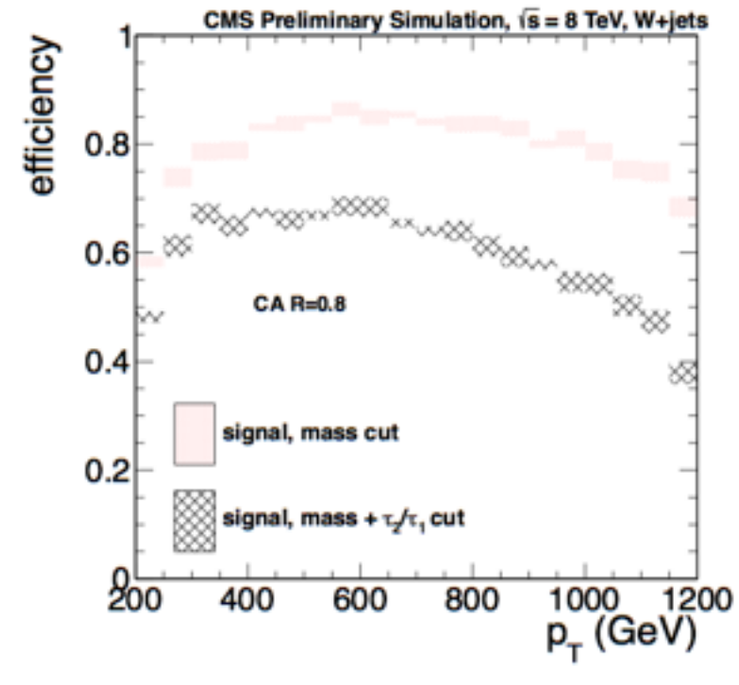
SUSY
Accidental
Substructure

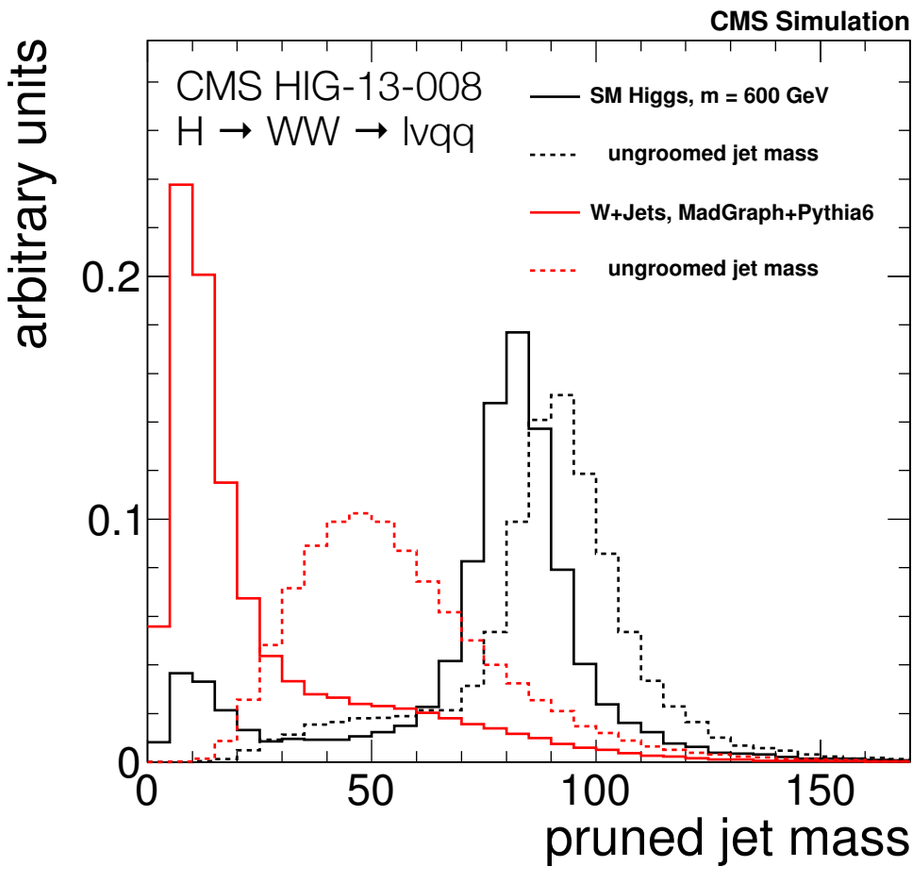


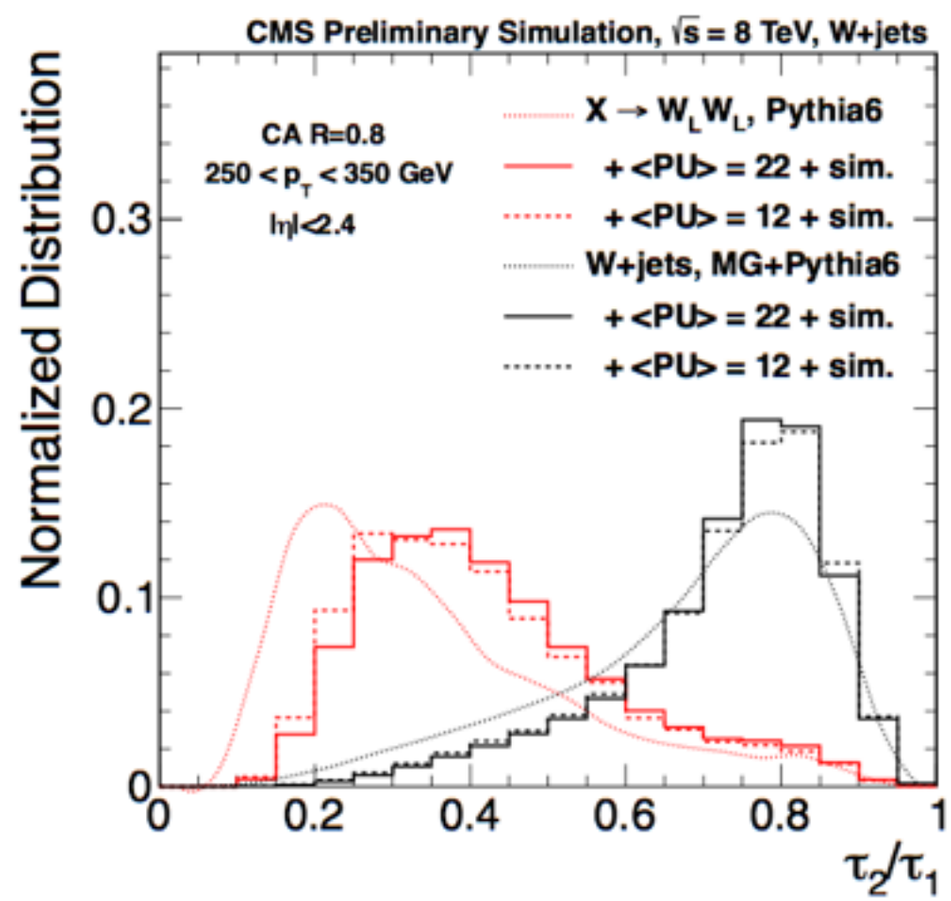
W-jet tagging

JME-13-006

- Pruned W-mass selection + N-subjettiness
  - $60 < m_{pruned} < 100$ GeV
  - $\tau_2/\tau_1 < 0.5$







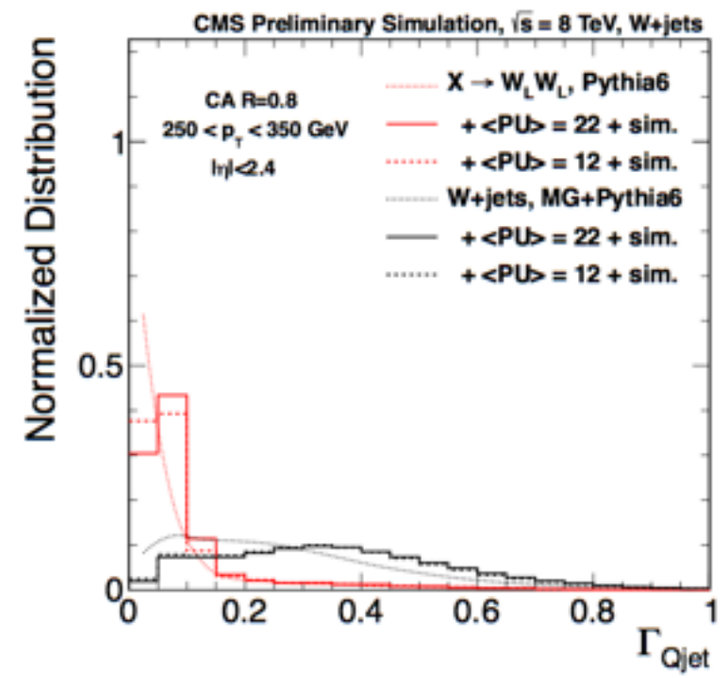
Qjets

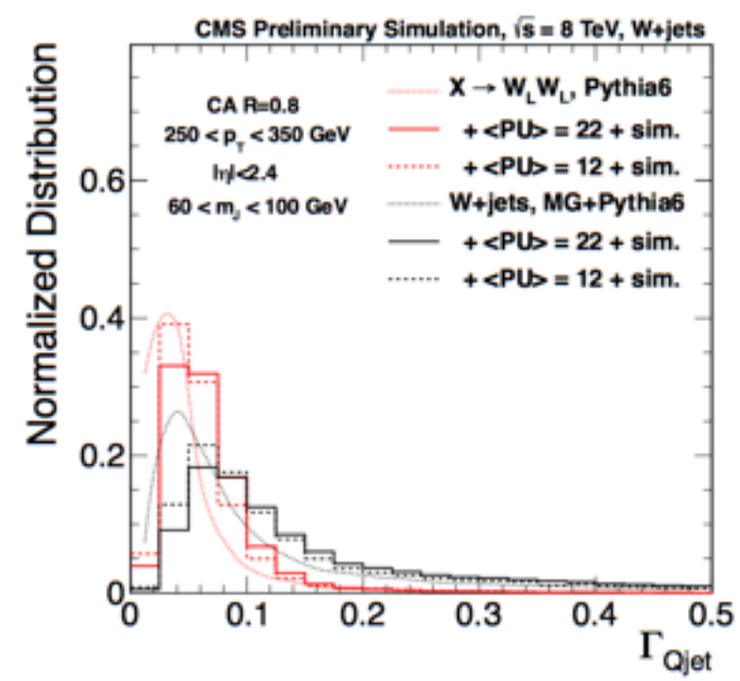
Krohn et al.
arXiv:1201.1914v2

http://jets.physics.harvard.edu/Qjets/html/Welcome.html

- Add an element of randomness to jet clustering with pruning
  - Weight determines how likely two particles are to be clustered $\omega_{ij}^{(\alpha)} \equiv \exp\left\{-\alpha \frac{(d_{ij} - d^{\min})}{d^{\min}}\right\}$
- Cluster many times
- Top/Higgs/W/Z jets have a hard mass scale
  - tend to get the same mass every time you cluster
- QCD
  - Large variation in measured jet mass depending on which parts of the jet get pruned away
- Tag jets based on their volatility $\nu = \Gamma / \langle m \rangle$
  - QCD shows larger variation $\Gamma \equiv \sqrt{\langle m^2 \rangle - \langle m \rangle^2}$

W-tagging with Q jets volatility





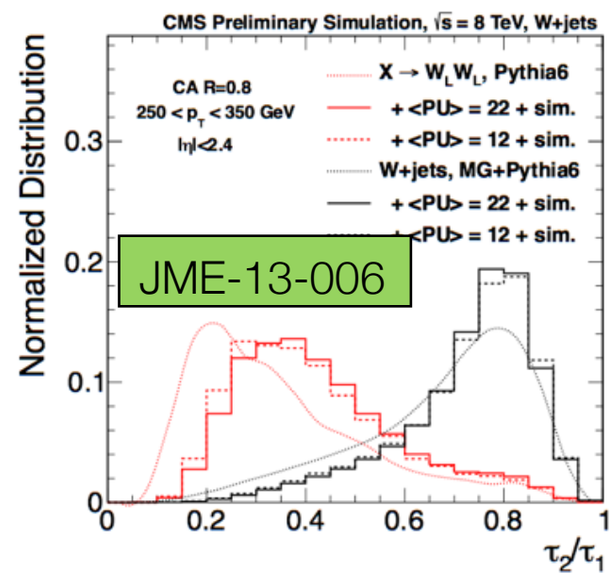
JME-13-006

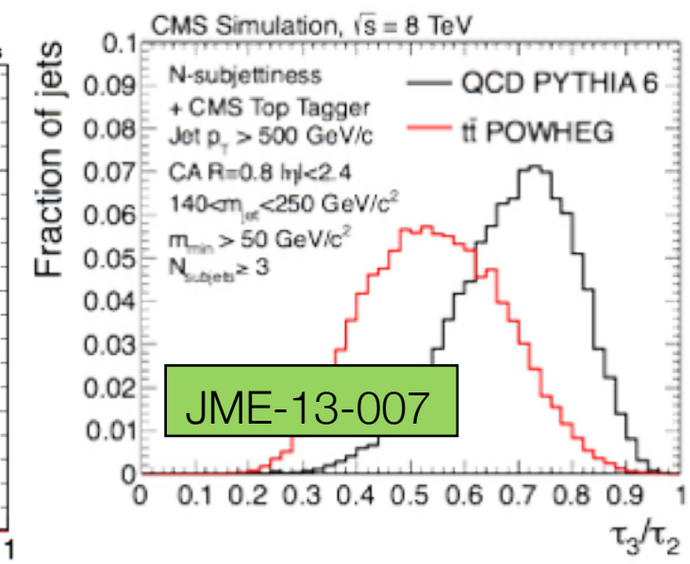
N-subjettiness

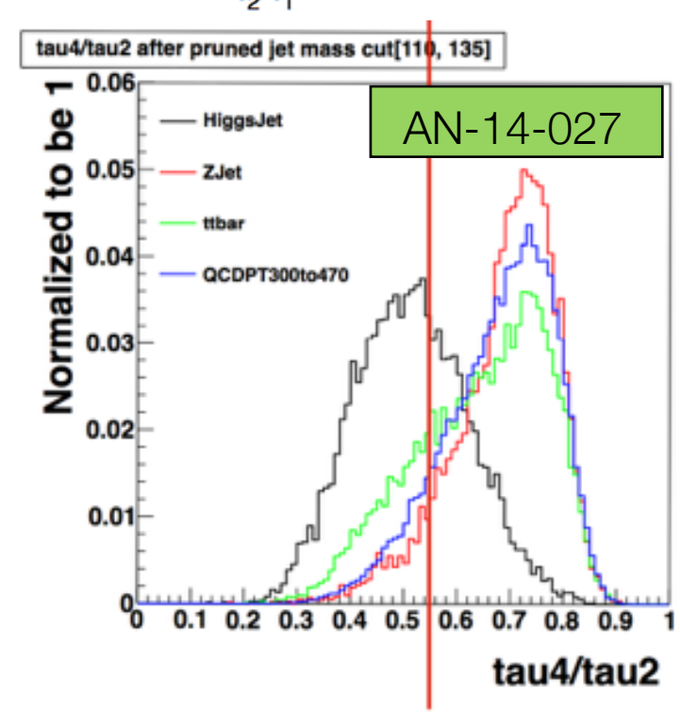
J. Thaler, K Van Tilburg
arXiv:1011.2268

$$\tau_N = \frac{\sum_{i=1}^{n_{\text{constituents}}} p_{T,i} \min\{\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i}\}}{\sum_{i=1}^{n_{\text{constituents}}} p_{T,i} R}$$

- Jet shape variable which determines how consistent a jet is with having **N or fewer subjets**
  - $\tau_N \rightarrow 0$ if jet has N or fewer subjets (energy spread is close to the subjet axes)
- Recommended subjet axes: one-pass optimized kt
- W/H(bb)/Z-tagging: τ_2/τ_1
- Top-tagging: τ_3/τ_2
- $H \rightarrow WW^* \rightarrow qqqq$ tagging: τ_4/τ_2

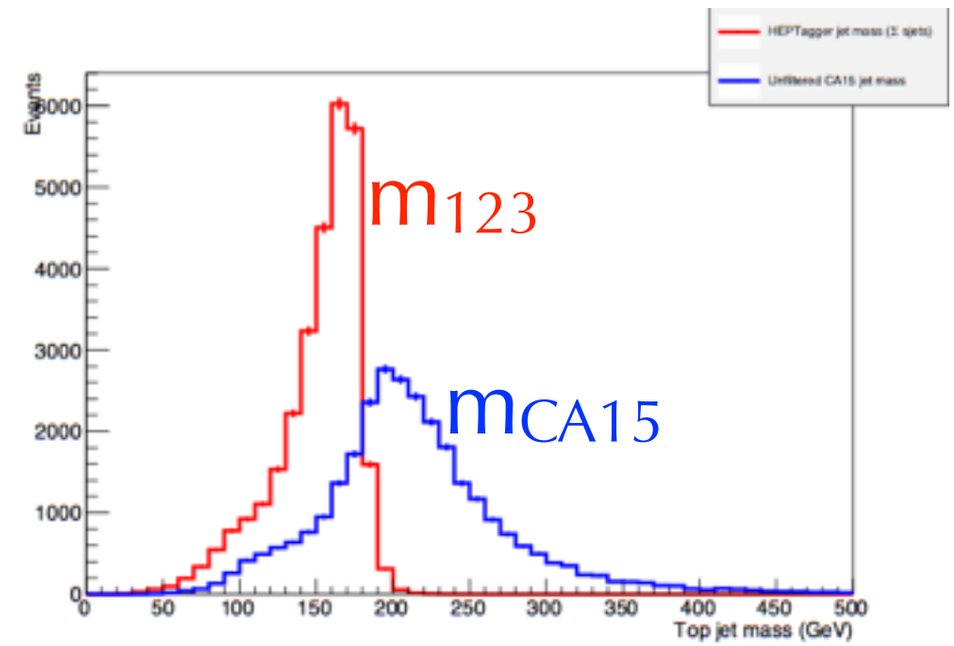


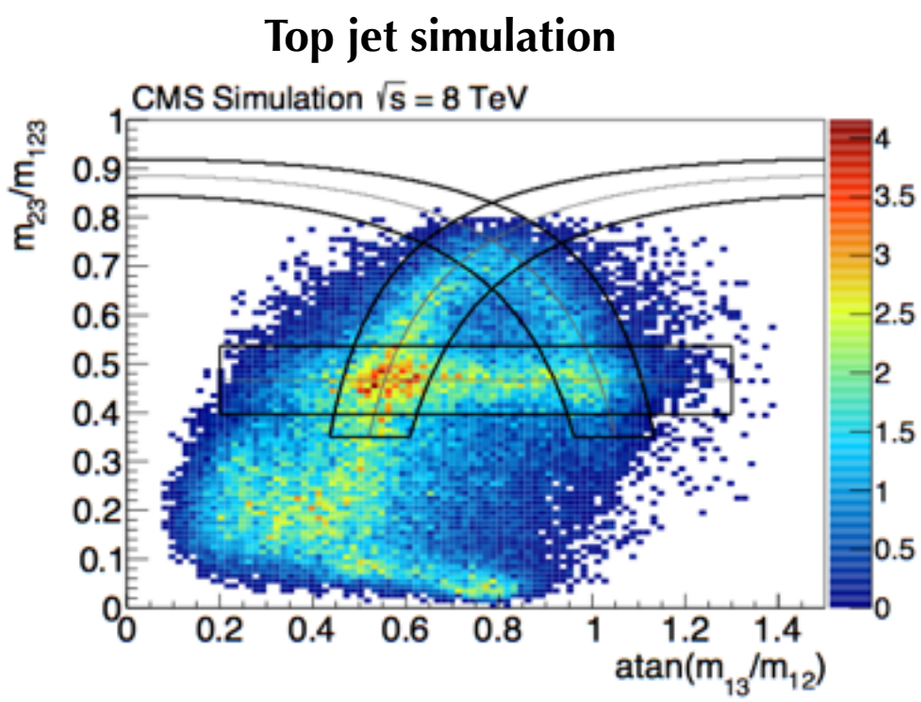


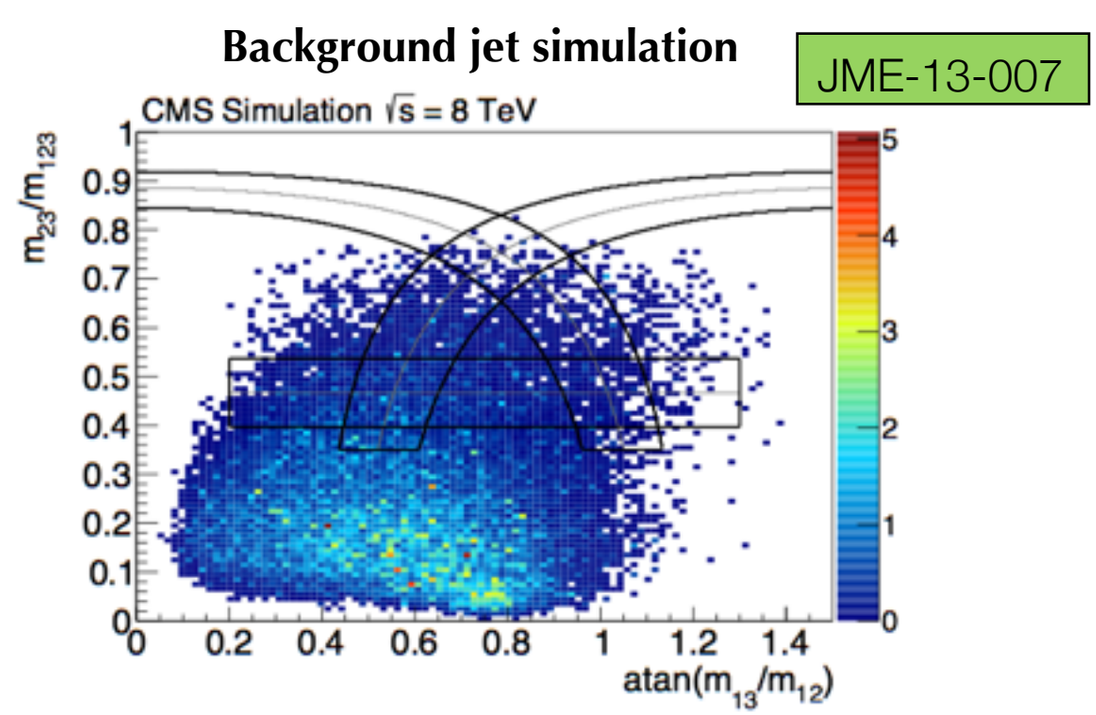


HEP Top Tagger

- Plehn et al. (arXiv:1006.2833)
- Very large jets (Cambridge Aachen R=1.5)
  - Increased efficiency at low p_T (p_T>200)
- Jet grooming is built in:
  - A subset of the jet constituents are used to tag tops
- Top mass selection using triplet subjet mass m_{123}
- W mass selection using pairwise subjet mass





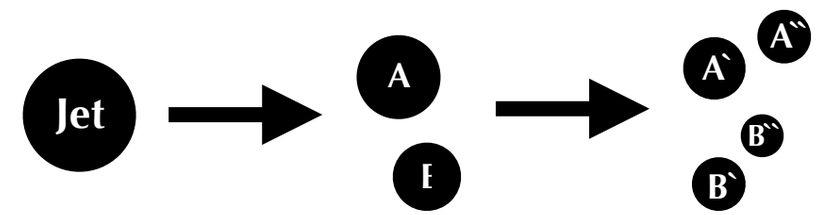
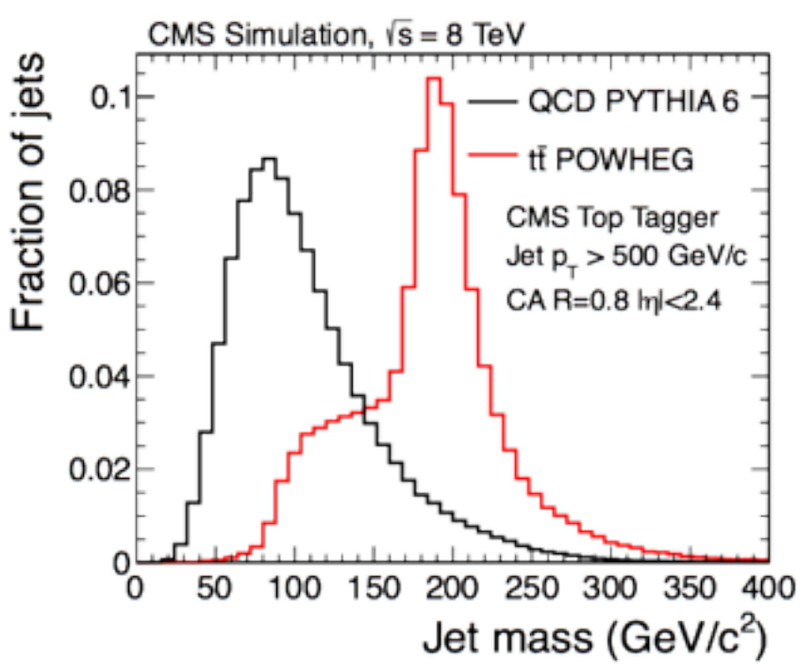
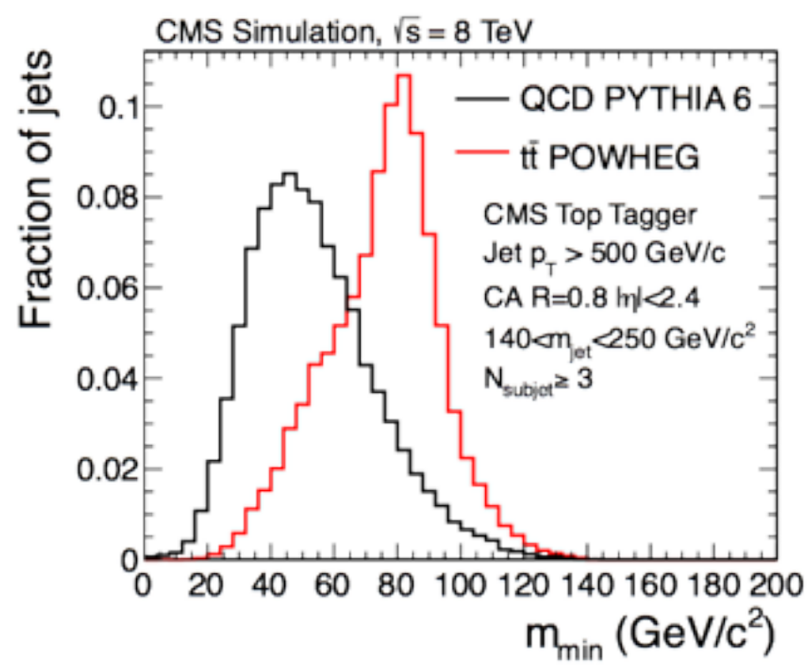
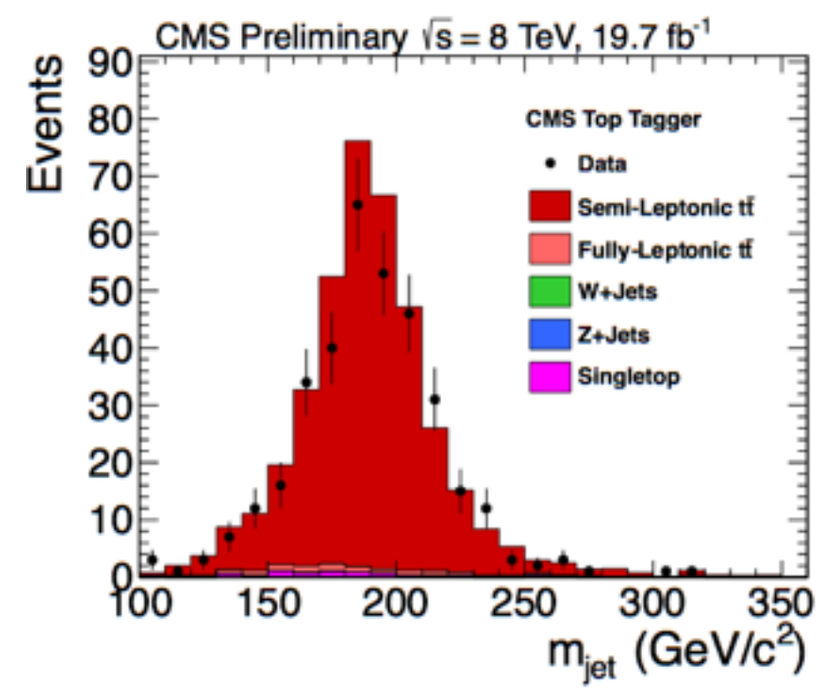


CMS Top Tagger

Based on JHU Top Tagger (Kaplan et al.)

JME-13-007

- Cluster with CA8 - Decluster in two stages in order to find up to 4 subjets
- Subjets must satisfy two requirements
  - Momentum fraction criterion: $p_T^{subjet} > 0.05 \times p_T^{jet}$
  - Adjacency criterion: $\Delta R(C_1, C_2) > 0.4 - 0.0004 \times pT(C)$



- Iterative process - throw out subjets that fail momentum fraction cut and try to decluster again
- Tag top jets with selections on the CA8 jet mass, number of subjets, and minimum pairwise subjet mass ($m_{min} = \min[m_{12}, m_{13}, m_{23}]$)