Boosted W/Z tagging in CMS

Luca Brianza[1]
On behalf of the CMS collaboration

Boost 2016 – Zurich

[1]**INFN & University of Milano-Bicocca**

Outlook

- ***Introduction***
- ***Brief panoramic of the main grooming techniques and substructure observables***
- ***Performances at 13 TeV (simulation)***
- ***W/Z tagging in some run II analyses***

Introduction

LHC Run II: a new era

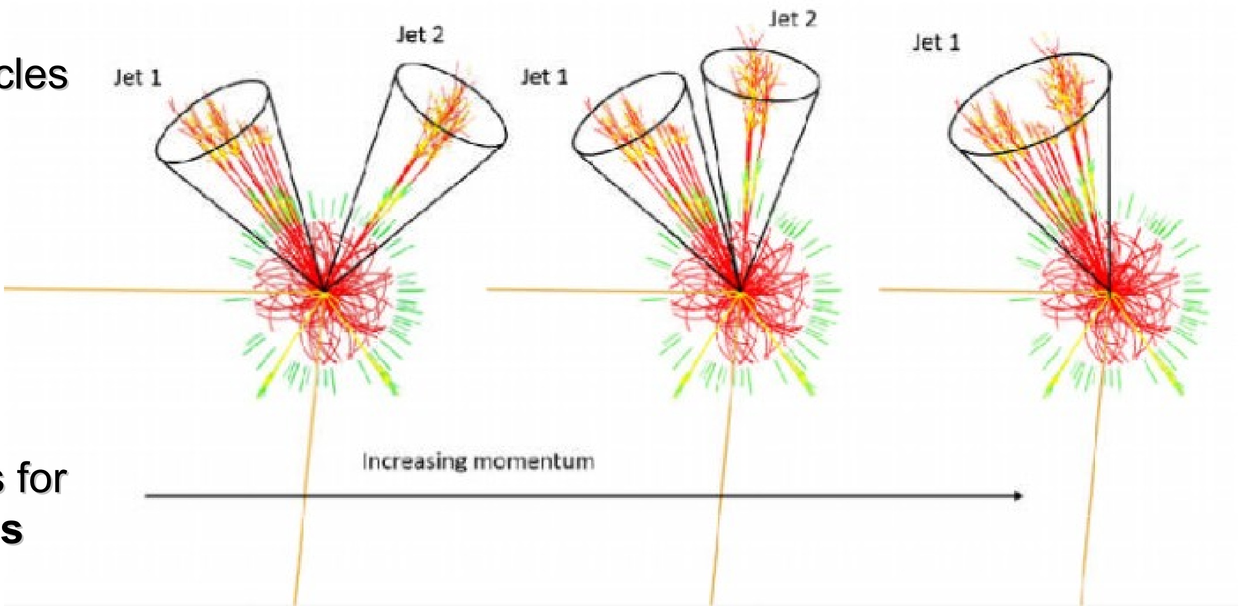
Higher energies involved, can probe the very high-mass part of the spectrum

Several searches for heavy particles (TeV scale):

- Gravitons
- W'/Z'
- Heavy Higgs

…

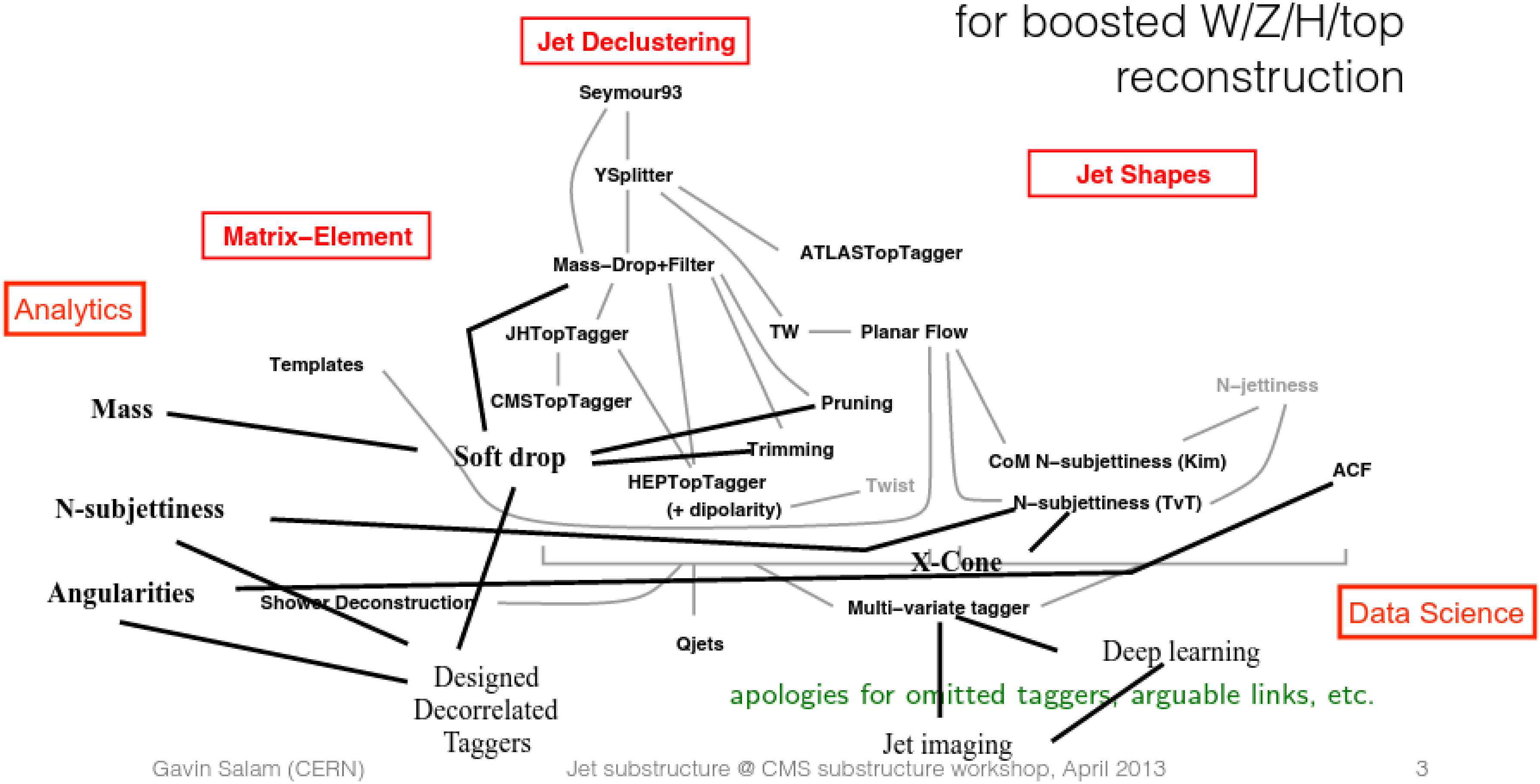
Models usually predict large BRs for these particles into **W/Z bosons**



→ Decay products from hadronically W/Z decaying bosons highly boosted

The market

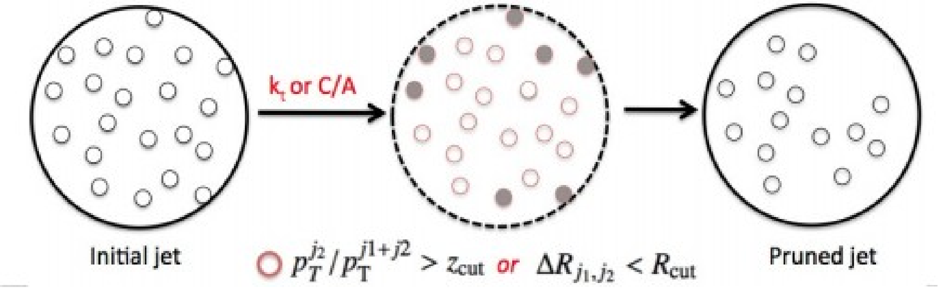
Some of the tools developed
for boosted W/Z/H/top
reconstruction



Original Slide from Gavin Salam

Pruning

Pruning:



***"Usual" direction**: start from protojets, recluster constituents*

$$z < z_{\rm cut} \quad \text{and} \quad \Delta R_{12} > D_{\rm cut}.$$

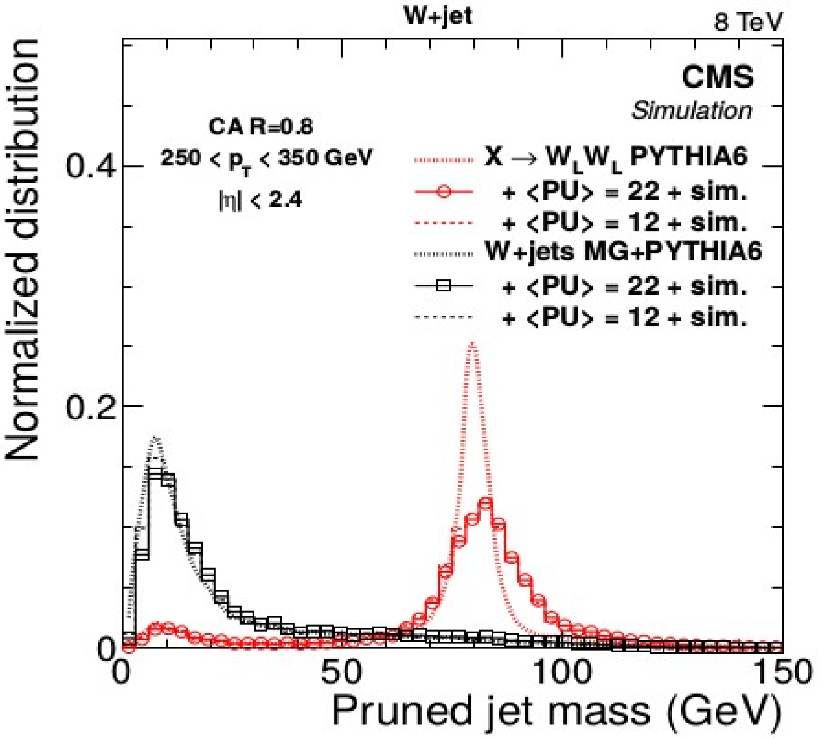
$$z \equiv \min(p_{T_1}, p_{T_2})/p_{T_p}, \qquad \theta \equiv \Delta R_{12}.$$

→ Remove softer and wide-angle constituents

CMS:

D_0 = 0.5

Z_{cut}=0.1



Ellis et al. http://arxiv.org/abs/0912.0033

Softdrop/MMDT

CMS-PAS-JME-14-002

Softdrop:

Proceed in the "opposite" direction

→ Start from the final jet

→ Decluster, removing at each step wide-angle and soft radiation contributions

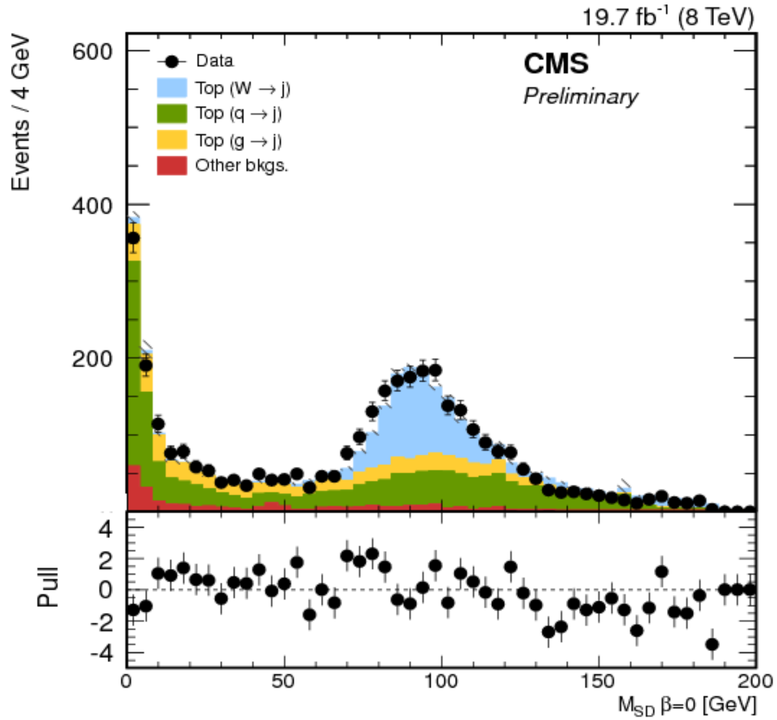
$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\mathrm{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$

CMS:

R_0 = 0.8

β=0

Z_{cut}=0.1



Dasgupta et al. arXiv:1307.0007
Larkoski et al. ArXiv:1402.2657

N-subjettiness

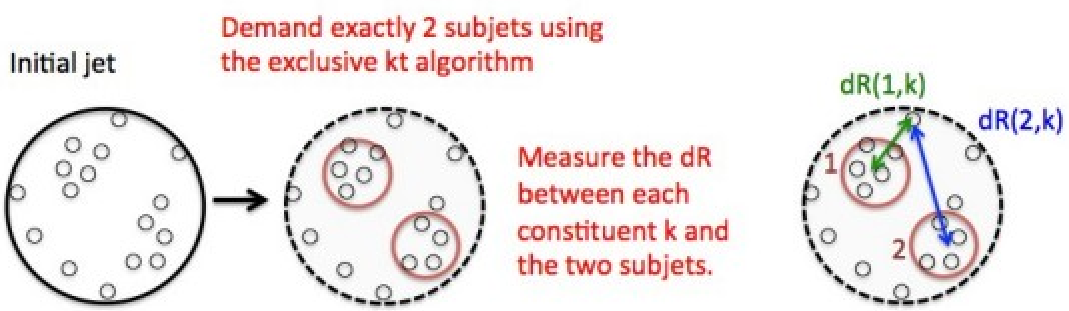
Quantify how well a jet can be divided into N subjets

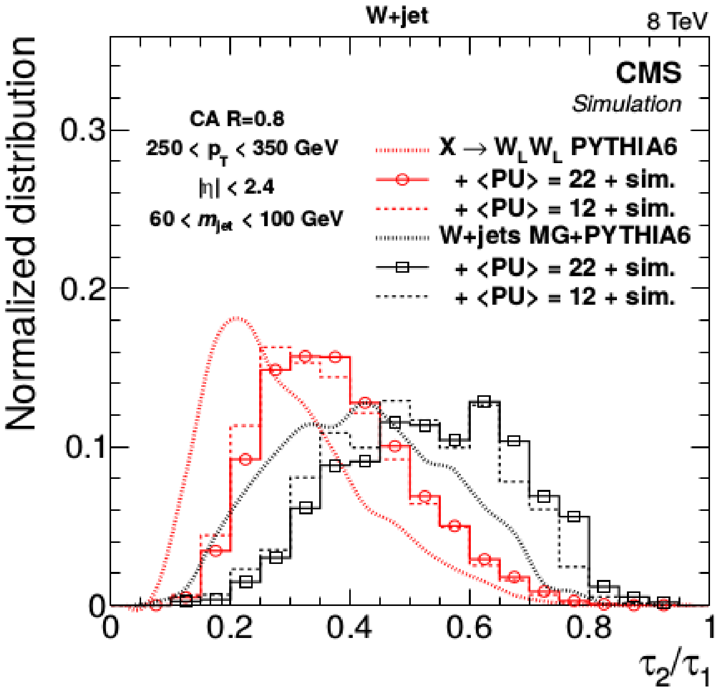
Recluster the jet with kT-algorithm until N subjets are left

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k}\}$$

Computed using ungroomed jet

Example: 2-subjettiness





Thaler et al.
http://arxiv.org/abs/1108.2701

Performances in MC

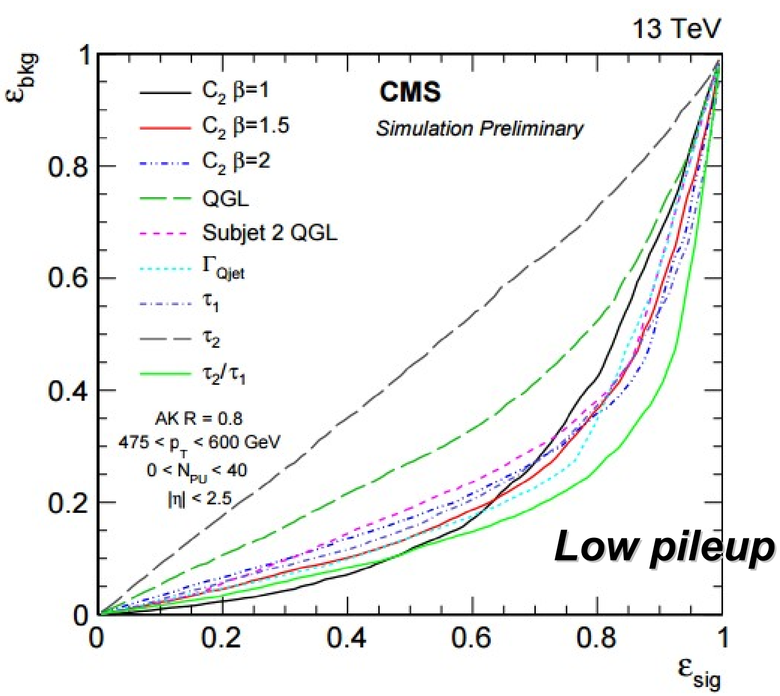
Performances in MC (early studies)

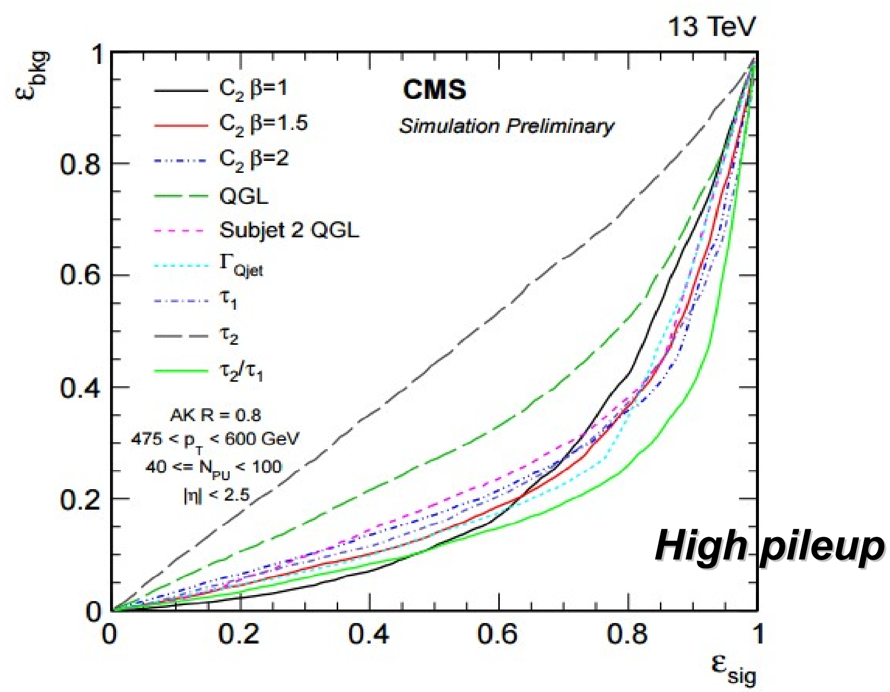
CMS-PAS-JME-14-002

Performances of various taggers estimated from simulation

Best performance: ***N-subjettiness*** ***(green curve)***

Run I and 2015 analyses
Baseline tagger
→ pruning + N-subjettiness





Interlude: PUPPI

→ ***PileUp Per Particle Identification (PUPPI)***

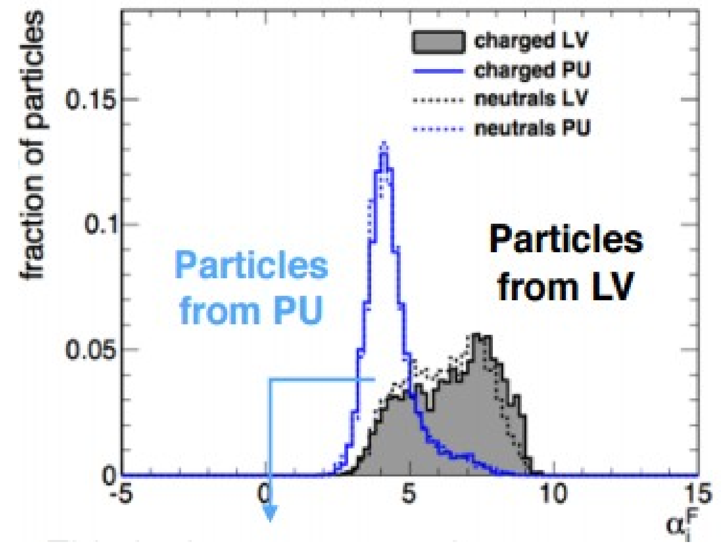
→ *It takes as input particle flow particles (charged/neutral hadrons, photon, charged leptons)*

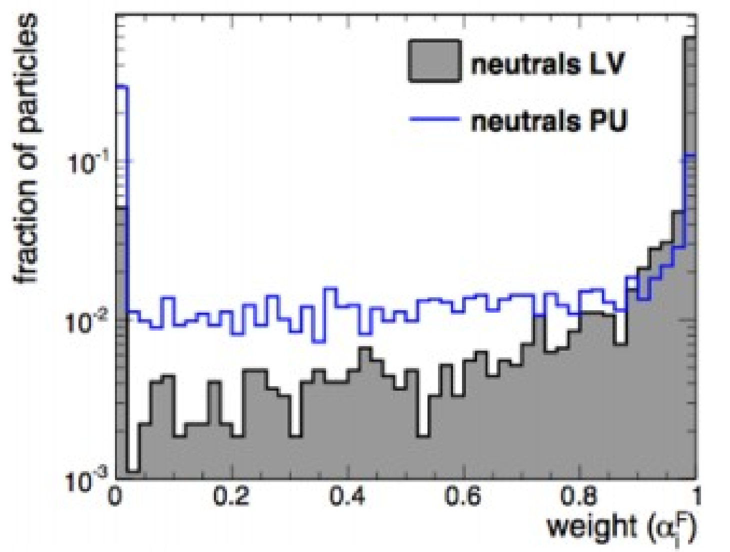
→ *Define **α** of each charged particle (i)*
using other particles (j) around it

$$\alpha_i = \log \sum_{j \in \text{event}} \underbrace{\frac{P_T^j}{\Delta R_{ij}}}_{\text{PT sum weighted with distance}} \underbrace{\Theta(R_{\min} < \Delta R_{ij} < R_0)}_{\text{Step function to take into account only particles around it.}}$$

→ *Transform the distribution of **α** in a weight (1 for particle from LV, 0 for particles from PU)*

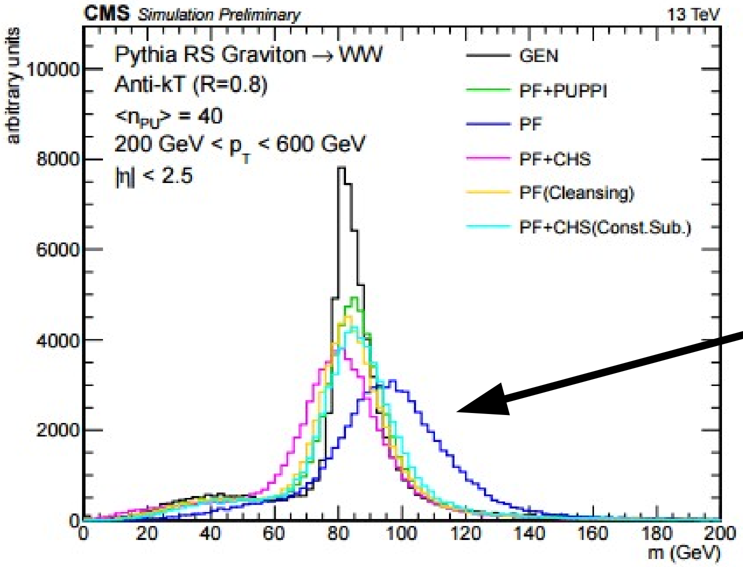
→ *Jet reconstruction algorithm an run on particles with taking into account the weight →* ***PUPPI jet***





Performances in MC (early studies)

CMS-PAS-JME-14-001

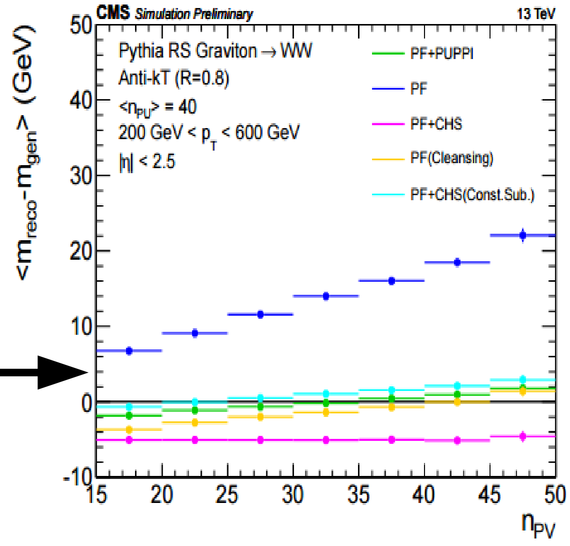


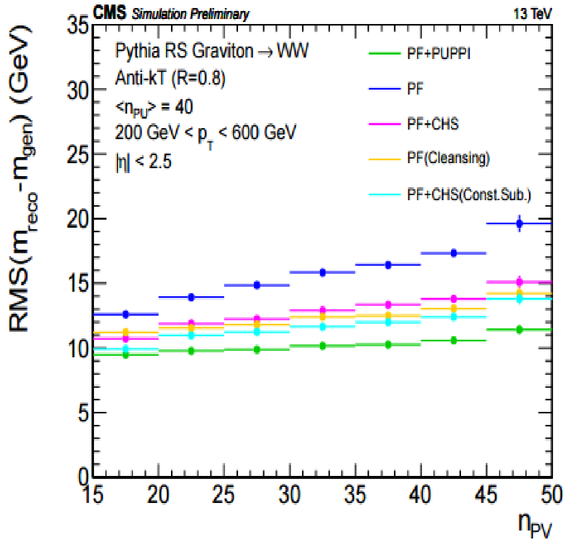
Boosted W also used to validate performances of different pile-up mitigation techniques

Check mass peak position and jet mass resolution

Stability vs Number of vertices

In general, PUPPI shows the best mass resolution and best stability against pile-up





Performances in MC

NEW!

CMS DP-2016/039

Signal: mixture of G → WW
Background: QCD jets

New studies: compare 3 different w-tagger

→ pruning + τ_{21}

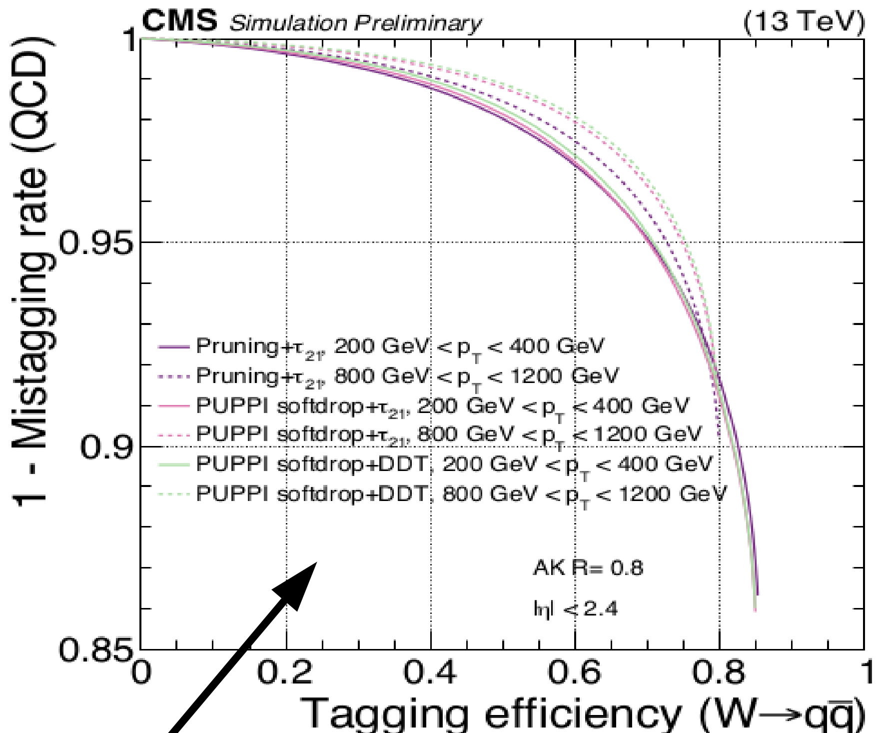
→ softdrop + τ_{21} (PUPPI)

→ softdrop + DDT (PUPPI)

DDT: re-definition of N-subjettiness in order to make it independent from jet pT and mass

(see Salvatore's talk)

$$DDT = \tau_{21,PUPPI} + 0.063*\log(M^2_{PUPPI}/p_{T,PUPPI})$$



Similar performances, PUPPI (DDT) slightly better

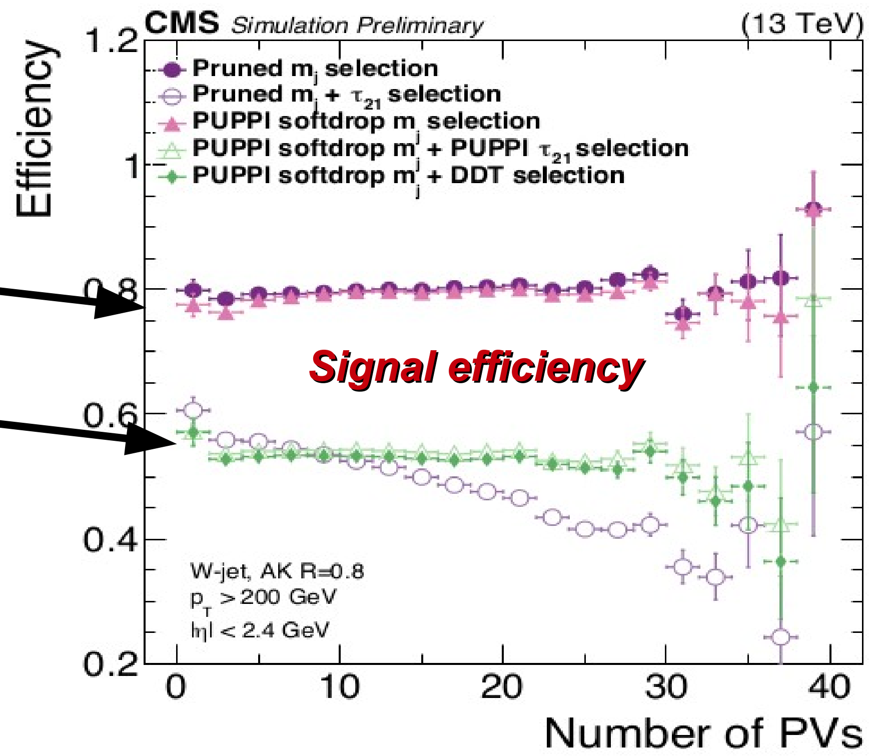
Performances in MC – vs pileup

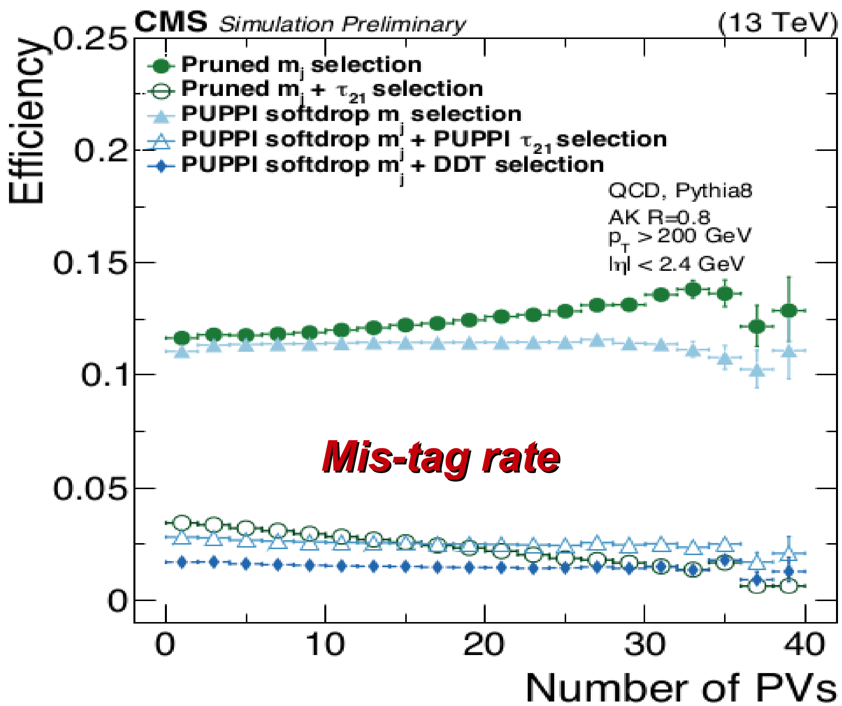
NEW!

CMS DP-2016/039

Top curves: before N-subjettiness cut

Bottom curves: after N-subjettiness cut





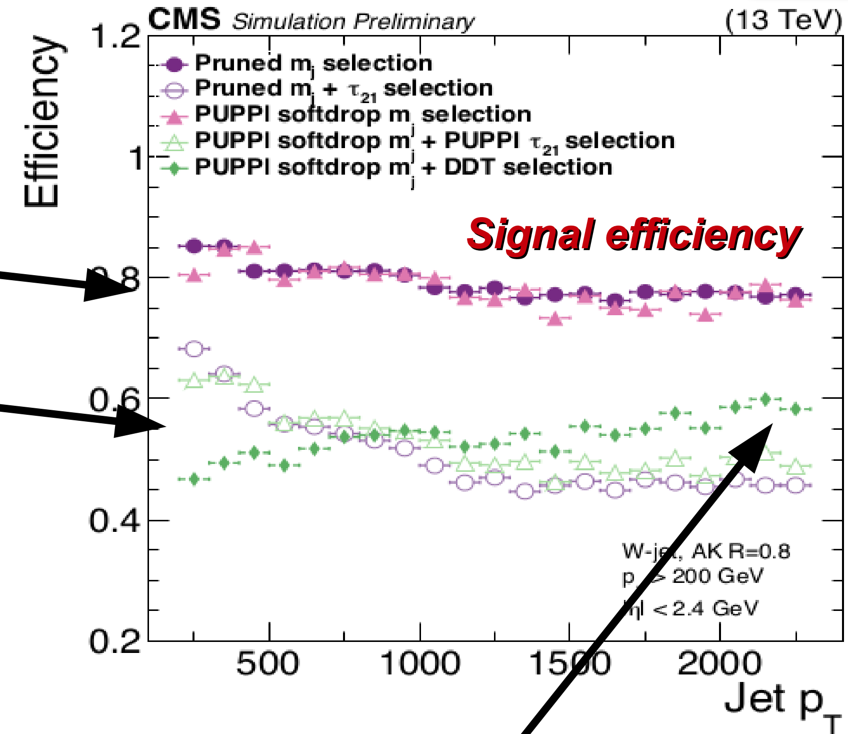
PUPPI: best stability vs pile-up

DDT: lowest mis-tag rate

Performances in MC – vs jet pT

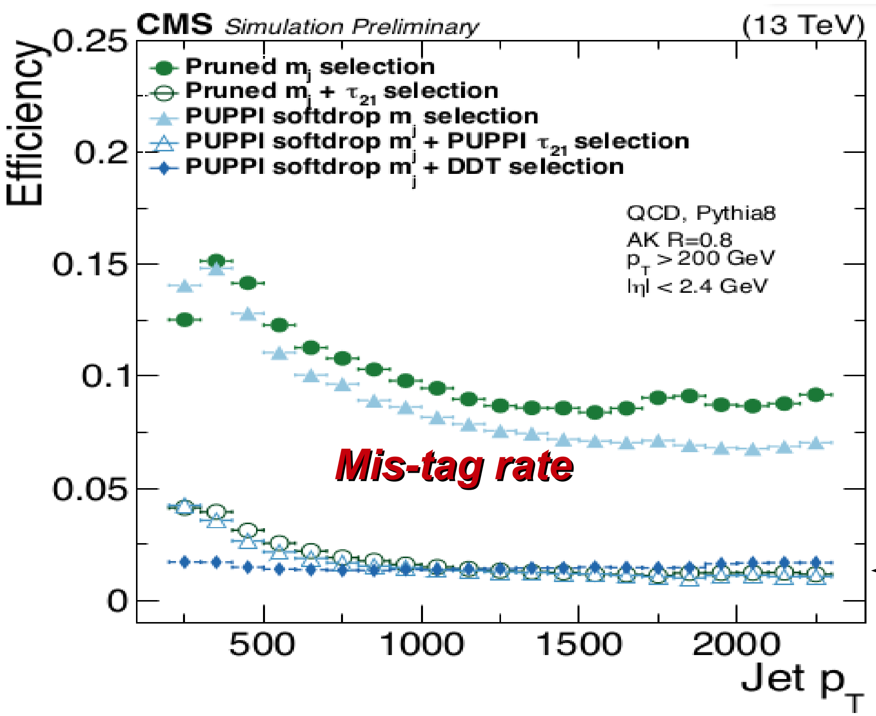
NEW!

CMS DP-2016/039



Top curves: before N-subjettiness cut

Bottom curves: after N-subjettiness cut



DDT: higher efficiencies at higher pT

DDT: flat mis-tag rate vs jet pT

Performances in data

Reconstruction of boosted W/Z tagger exploited in several analyses

Full overview of some analyses in Petar's talk

Here only focus on W/Z-tagger performance

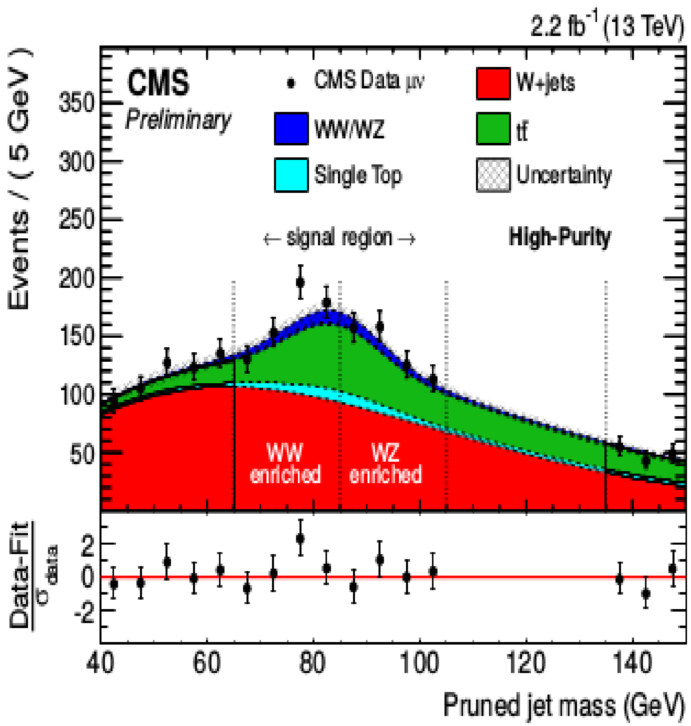
Performance in data: X → WV → lvqq

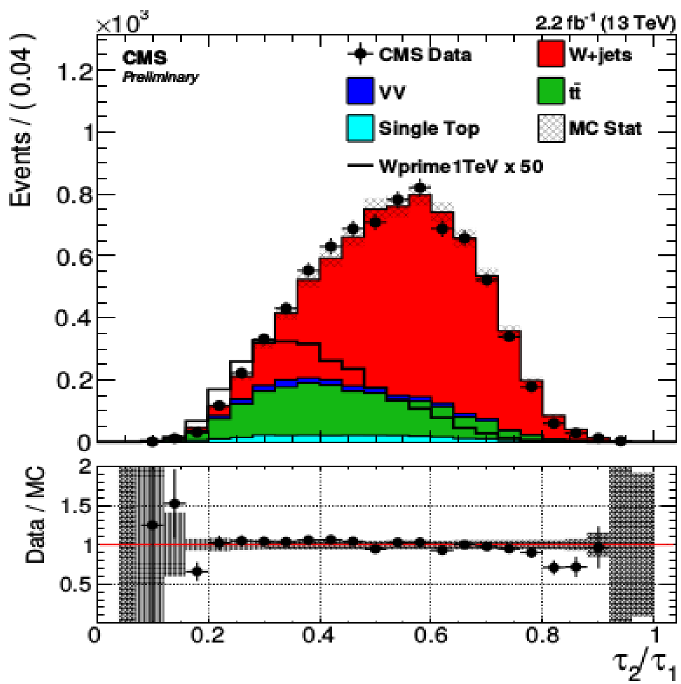
CMS-PAS-EXO-15-002
CMS-PAS-B2G-16-004

X → WV → lvqq (lvJ)
Bump search using M(lvJ) spectrum

Ungroomed jet pT > 200 GeV
Leptonic W pT > 200 GeV
Lepton pT > 40(45) GeV for mu (ele)
MET > 40 (80) GeV for mu (ele)

Categorization using jet mass *W: [65-105] GeV, Z: [85-105] GeV → 2 categories (WW,WZ)*
Categorization in N-subjettiness: $\tau_{21}<0.6$ and $0.6<\tau_{21}<0.75$





W-tagger performance in data

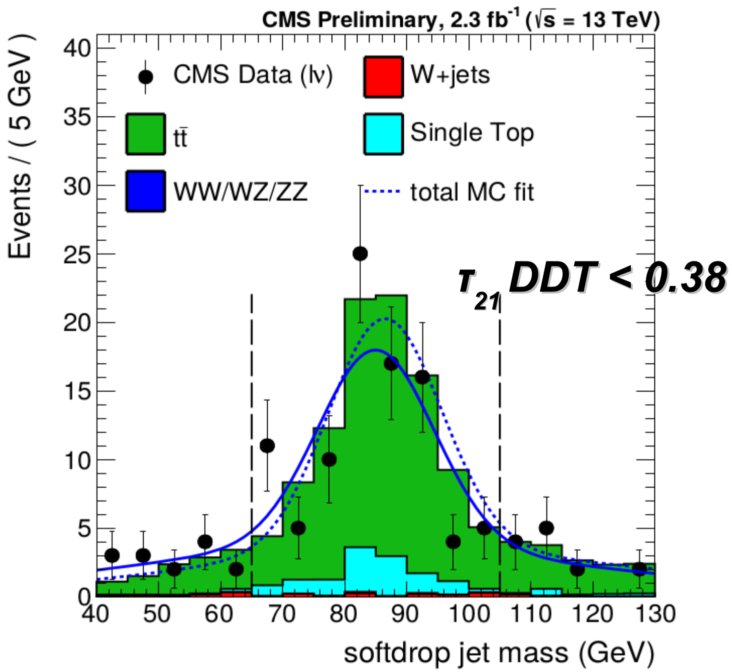
CMS-EXO-16-030
(see Nhan's talk from tomorrow)

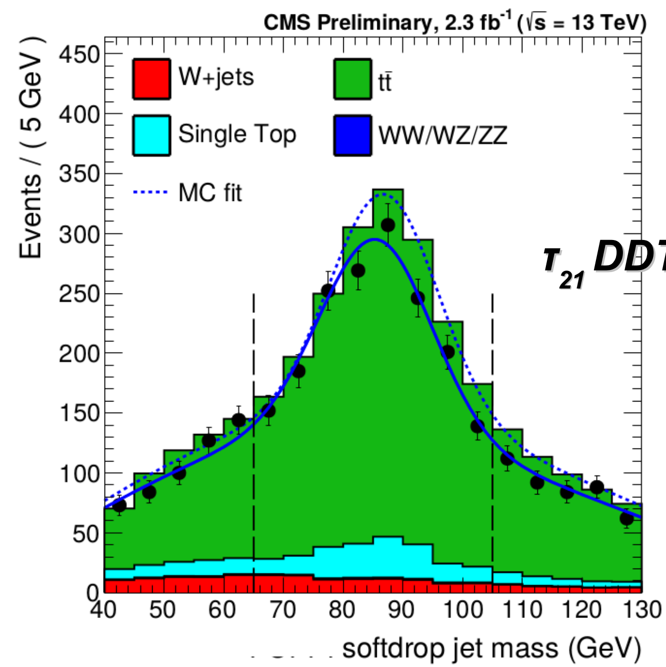
Top-enriched control samples used to extract W-tagger data/MC scale factors

Perform a simultaneous fit on jet mass in pass/fail regions to get τ_{21} efficiency in data and MC

Peak mass position and resolution corrected to account for data/MC differences

N-subjettiness scale factor:	_0.95 ± 0.20_
Jet mass peak correction:	_-0.59 ± 0.87 GeV_
Jet mass resolution correction:	_1.10 ± 0.12 GeV_





Conclusions

Boosted W/Z: more important now than ever

A lot of tools to reconstruct boosted W/Z

Early simulation studies showed
pruning/softdrop and N-subjettiness best groomers/substructure observables

Baseline tagger for analyses in 2015: ***pruning (with CHS) + τ_{21}***
Extensive validation with 2015 data done in several analyses

New studies made to assess performances of V-tagger with different pile-up algorithms
→ ***PUPPI shows best performances in terms of efficiency and stability,***
in combination with softdrop
<u>Particularly promising: softdrop + DDT cut</u>

Backup

Softdrop/trimming

Softdrop: *Proceed in the opposite direction:*

Starting from the final jet

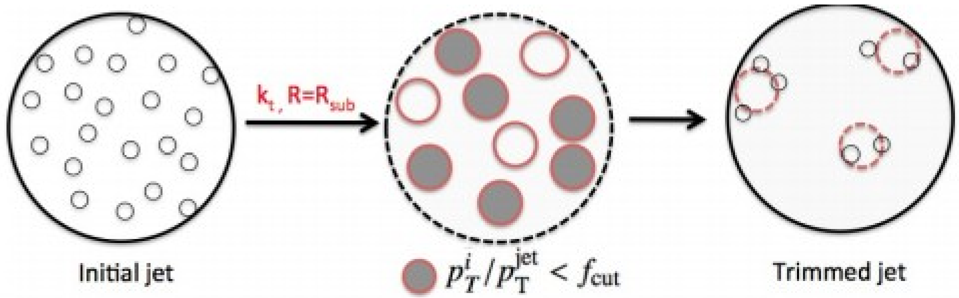
Declustering, removing at each step wide-angle and soft radiation contributions:

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

Larkoski et al.
https://arxiv.org/pdf/1402.2657

Trimming: *Use Kt algorithm to create subjets of size R_{sub}<R*

Remove subjets whose pT does not pass a threshold



Khron et al.
http://arxiv.org/abs/0912.1342

Mass drop / filtering

Mass drop/ filtering:

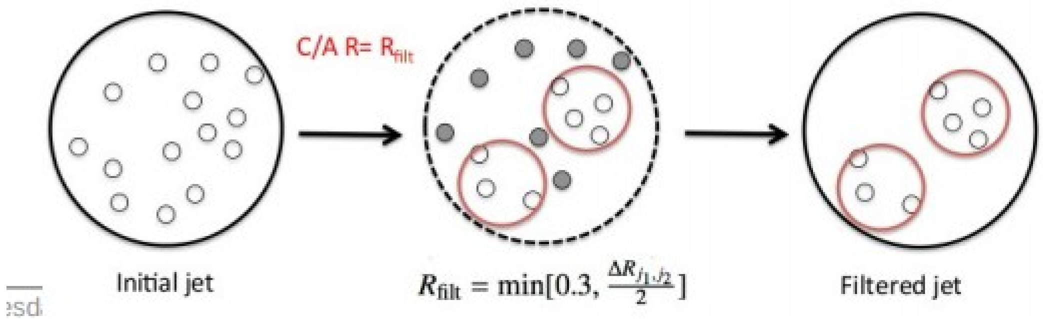
Decluster the jet into n subjets

Until a significant mass drop is obtained and the splitting is not too much asymmetric

$$\frac{\min[(p_T^{j_1})^2,(p_T^{j_2})^2]}{(M^{\text{jet}})^2}\times\Delta R^2_{j_1,j_2}>y_{\text{cut}}$$

$$\mu = \frac{m_{j1}}{m_j}$$

Then recluster subjets using smaller R



Butterworth et al
http://arxiv.org/abs/0802.2470

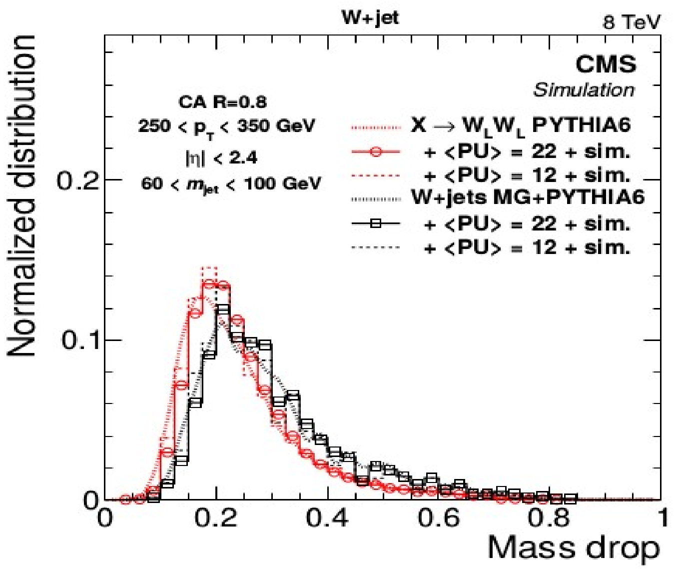
Mass drop, Q-jet volatility

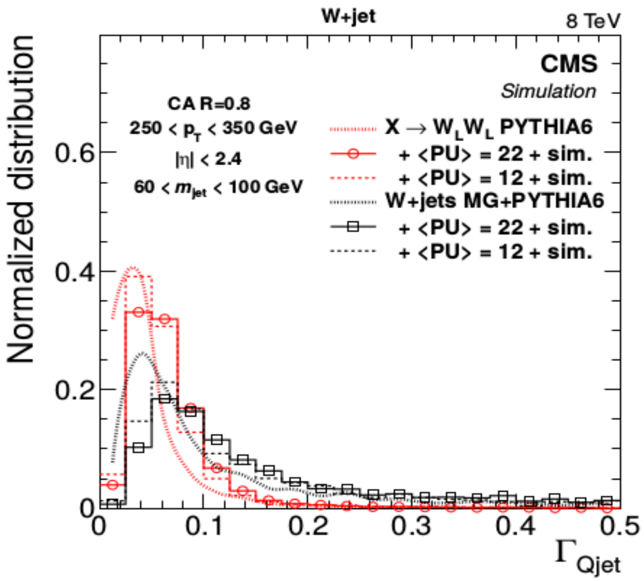
Mass drop:

ratio between the higher mass subjet and the total pruned jet

$$\mu = \frac{m_{j1}}{m_j}$$

Butterworth et al
http://arxiv.org/abs/0802.2470





Q-jet volatility:

The jet is reclustered several times, Using a random sequence

→ from distribution of the jet mass:

$$\Gamma_{\mathrm{Qjet}} = \mathrm{RMS}/\langle m \rangle$$

Ellis et al.
arXiv:1201.1914.

Energy correlation function, planar flow, jet charge

Energy correlation function:

$$C_2^{\beta} = \frac{\sum_{i<j<k} p_{\mathrm{T}i} p_{\mathrm{T}j} p_{\mathrm{T}k} (R_{ij} R_{ik} R_{jk})^{\beta} \sum_i p_{\mathrm{T}i}}{(\sum_{i<j} p_{\mathrm{T}i} p_{\mathrm{T}j} (R_{ij})^{\beta})^2}$$

Similar to N-subjettiness, the numerator quantifies how likely a jet is composed of 2 subjets, the denominator how likely of 1 subjet

Larkoski et al. arXiv:1305.0007.

Planar flow:

Characterises the geometric distribution of energy deposition from a jet (QCD jets: more isotropic)

Cui et al. arXiv:1012.2077.

Jet charge:

Measure of the electric charge of the original parton

$$Q^{\kappa} = \frac{\sum_i q_i (p_{\mathrm{T}}^i)^{\kappa}}{(p_{\mathrm{T}}^{\mathrm{jet}})^{\kappa}}$$

Khron et al. arXiv:1209.2421.