Jet ID in CMS

Philip Harris (CERN)
CMS collaboration

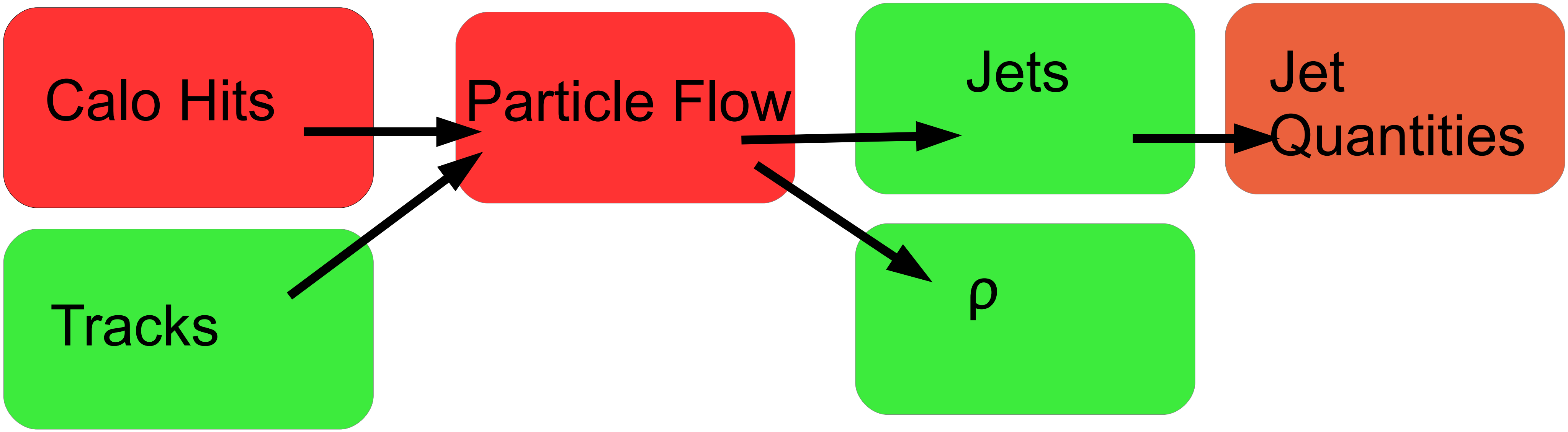
WAT is the Event?

- The event is composed of many jets
  - Correcting jets not necessarily correcting the event

Origins of PU Jet Id in CMS:
Approach to remove PU to the whole event

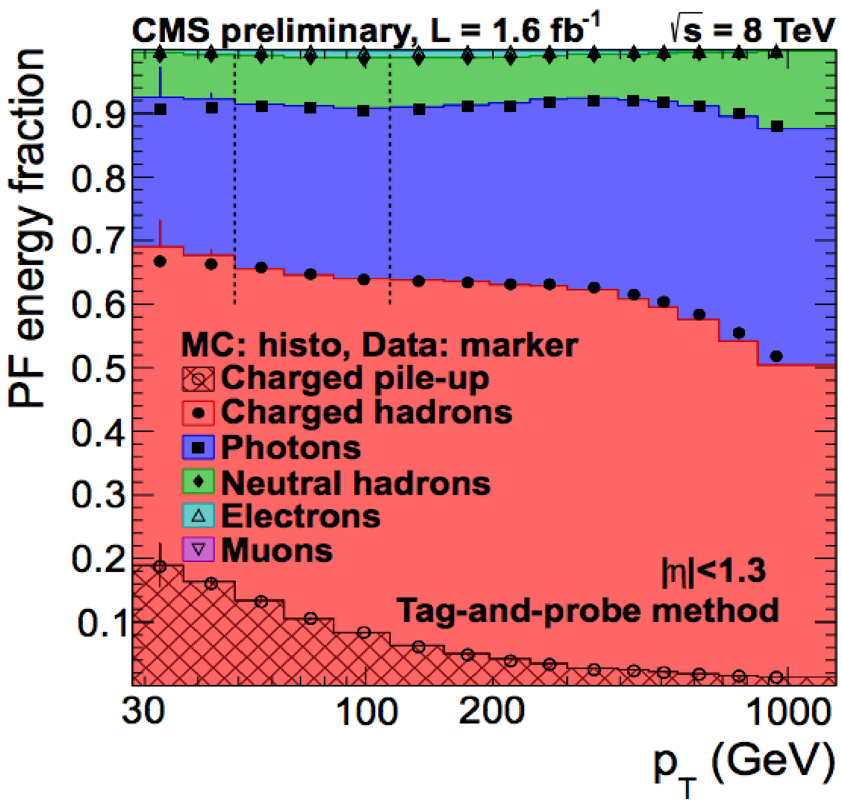
Both *MET* and Jets

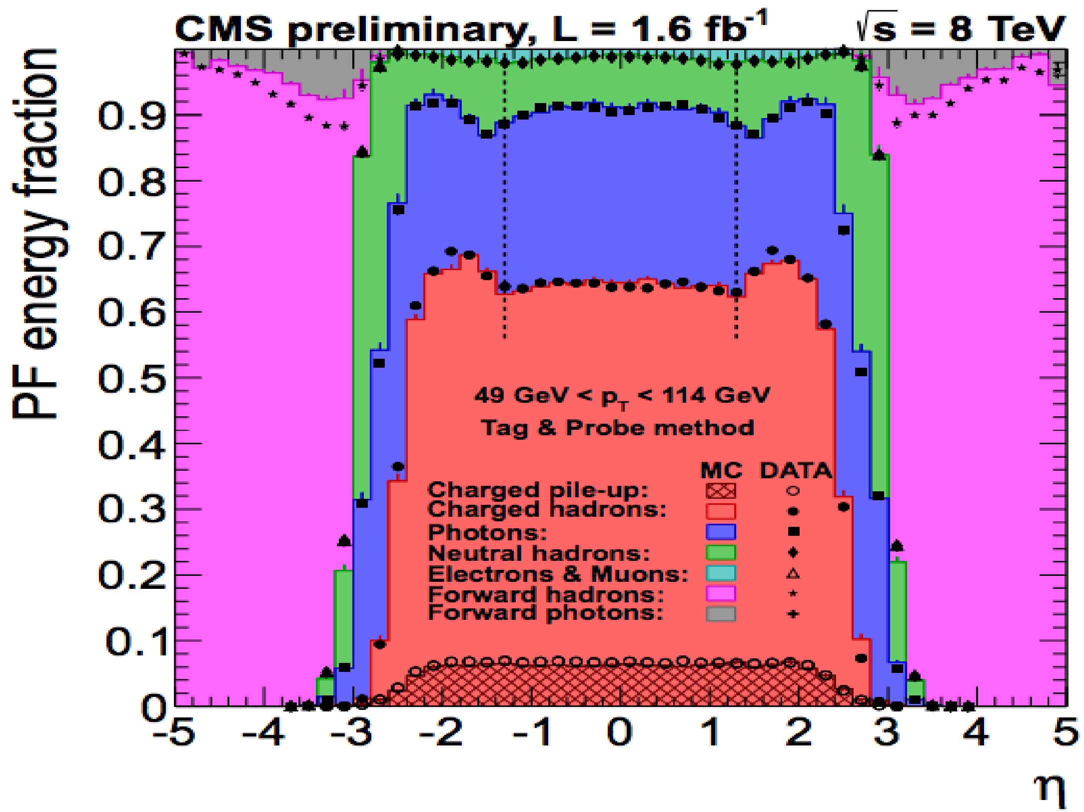
Stages of Jet ID



- At the lowest level perform basic noise cleaning
  - Each of the red boxes use noise cleaning
- Calo Hits : Basic cleaning
- PF : Cleaning on track quality + matching
- Jets : See next slide

Basic Jet ID : Cleaning Noise





Fraction	Loose	Medium	Tight
Neut. Had	< 0.99	< 0.95	<0.9
Neut. EM	< 0.99	< 0.95	<0.9
#constitsX	> 0	> 0	> 0
Chrgd. H	> 0	> 0	> 0
Neut. EM	< 0.99	< 0.95	0.9
#constitsX	> 0	> 0	> 0

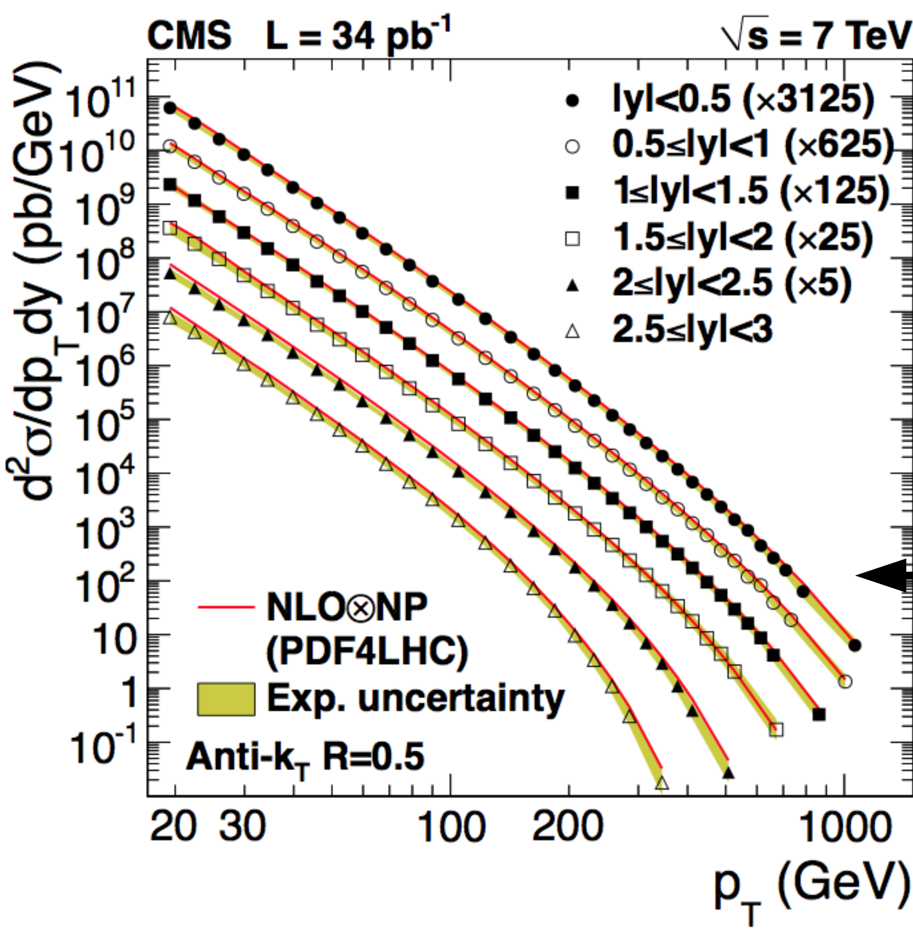
All Jets in CMS start with these cuts

Red : inside tracker

What is Pileup?

- Any object from another collision
  - Some resulting QCD from UE/jet production
  - Consider instance where objects are clustered

What is the rate of overlapping pileup?



Phys.Rev.Lett.107:132001,2011

Two jets w/Sum $pT = pT_1 + pT_2$

Consider a simplified model

$$P(\text{overlap}|pT) \approx N^2_{pu}\, a^2_{jet} \int A$$

$$\int A = \int d pT \frac{d\sigma_1}{dpT} \frac{d\sigma_2}{dpT} \delta(pT - pT_1 + pT_2)$$

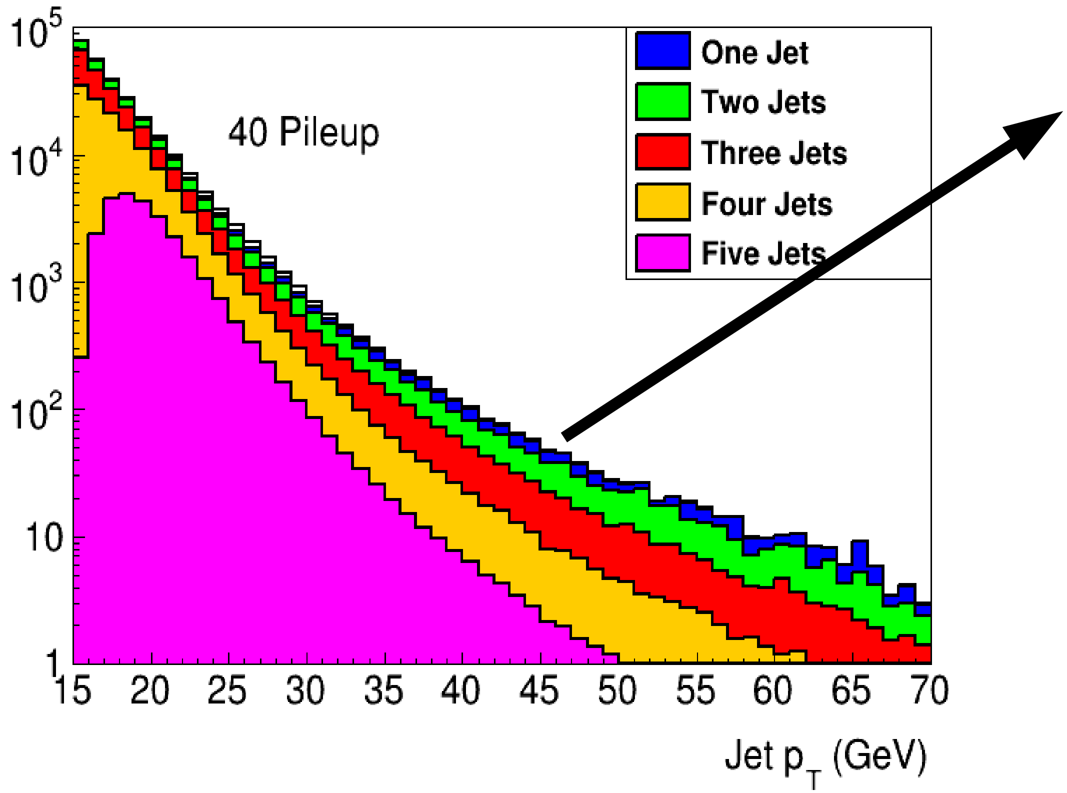
$$\frac{d\sigma}{dpT} = pT^{-5} \text{ (NLO/measurement)}$$

$$\int A = C\, pT^{-6.2} \text{ (C a constant)}$$

What is Pileup?

- Any object from another collision
  - Some resulting QCD from UE/jet production
  - Consider instance where objects are clustered

What is the rate of overlapping pileup?



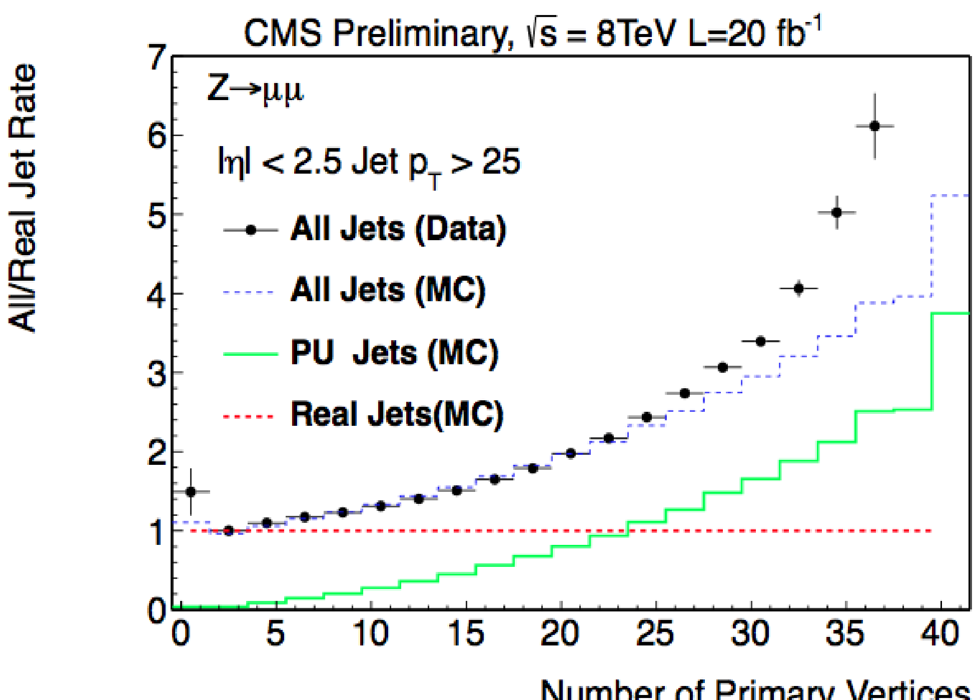
Contribution of overlapping jets To pileup jet spectrum

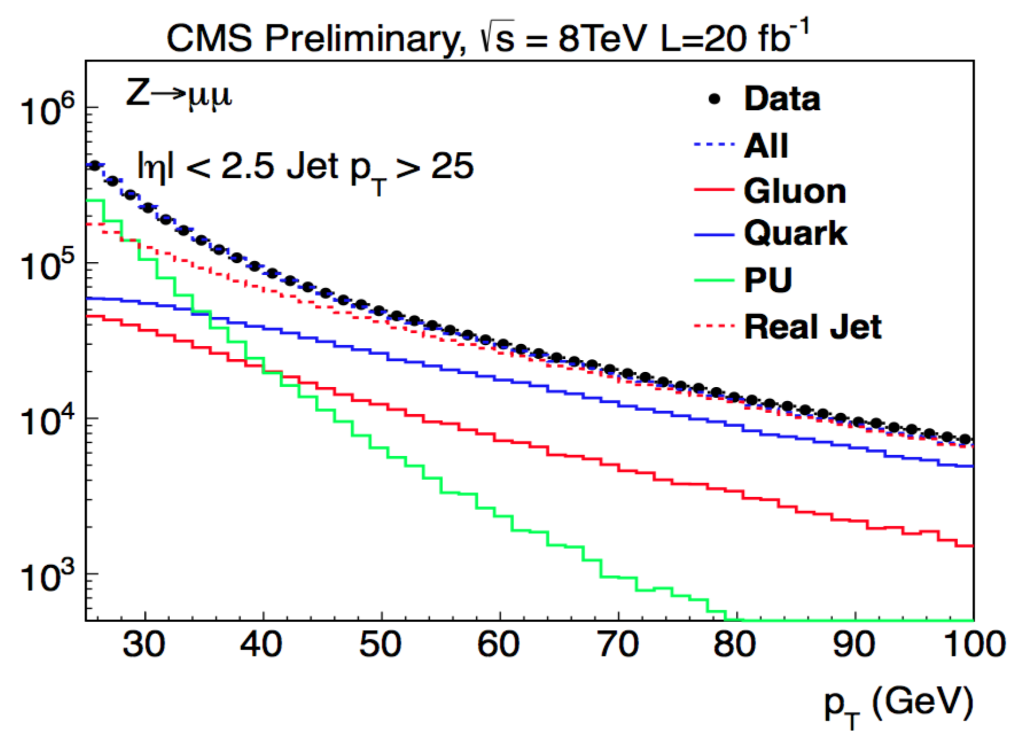
Two jets are dominant contribution

What does it look like in Data?

$$P(\text{overlap}|pT) \approx C\, N^2_{pu}\, a^2_{jet}\, pT^{-6.2}$$

- Expect pileup to grow quadratically
- Expect pileup jets to fall off more rapidly
  - Pileup jets remain a few% level problem up to 70 GeV





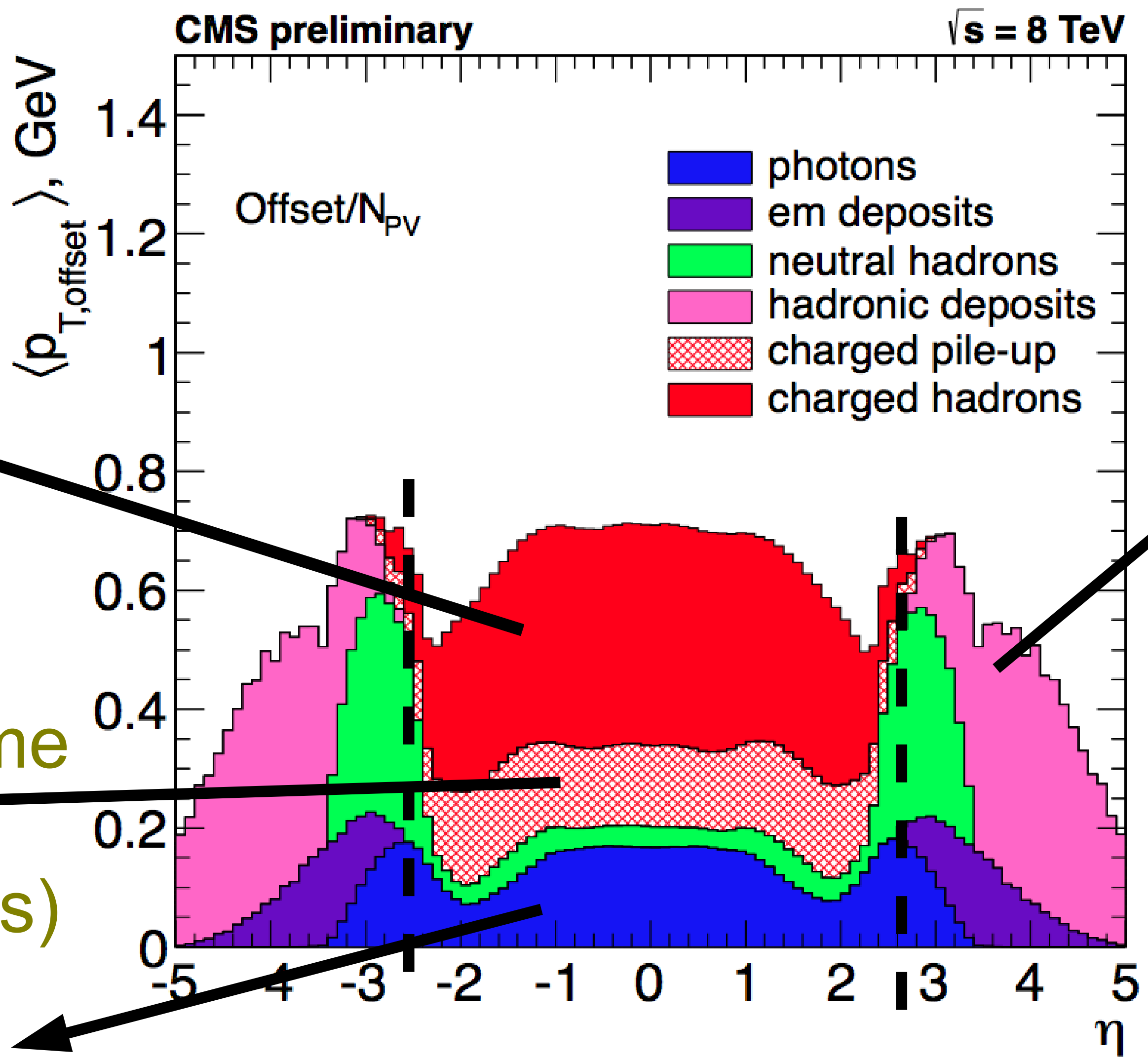
Number of PV = 0.7 x N_{pu}

Pileup Composition in CMS

Flag with Vertexing

Flag with some uncertainty (vtx + shapes)

Guilty by Association (rely on clustering)



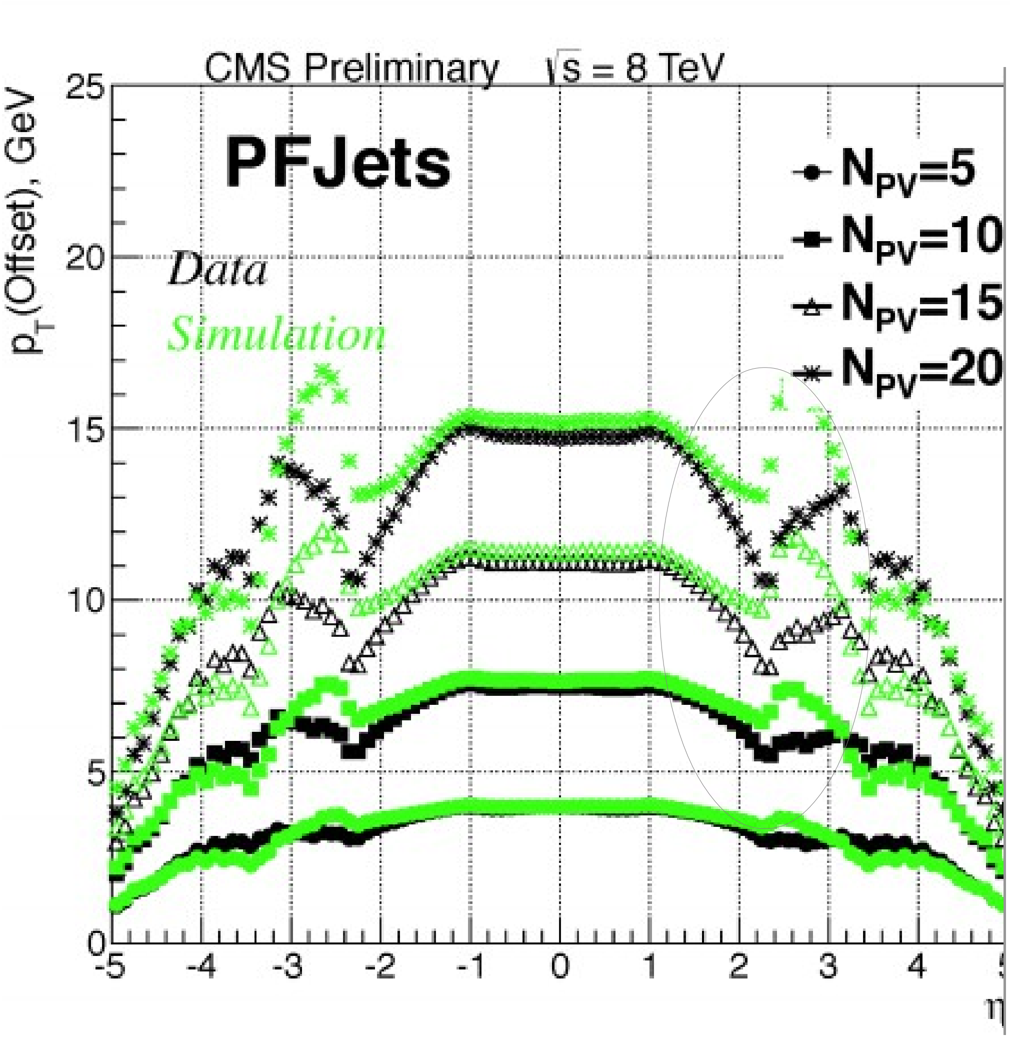
No tracker Must rely on shapes to identify

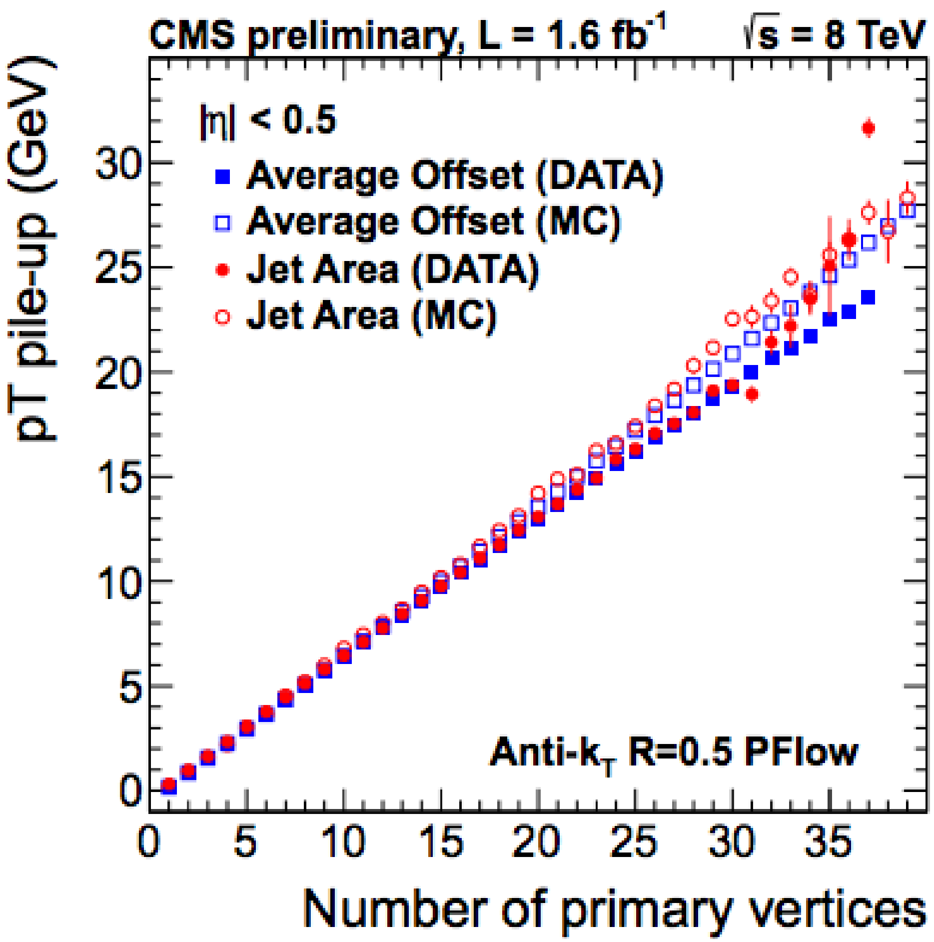
Charge Hadron Subtraction:
Flag from another vertex and remove it

Cleaning up the Pielup

- Pileup Subtraction

  Pileup measured in zero bias events : effectively is 1 2D Plot





Pileup Removal in CMS

- PF Jet reconstruction
  - Take all particle flow (PF) candidates
  - Cluster
  - Apply ρ correction
- Charged Hadron Subtraction(CHS) Jet reconstruction
  - Remove PF candidates assigned to another vertex
  - Cluster
  - Apply modified ρ correction (modified in TK volume)
  - Baseline for substructure and shape variables

Pileup Removal in CMS

- PF Jet reconstruction
  - Take all particle flow (PF) candidates
  - Cluster
  - Apply ρ correction

- Charged Hadron Subtraction(CHS) Jet reconstruction
  - Remove PF candidates assigned to another vertex
  - Cluster
  - Apply modified ρ correction (modified in TK volume)
  - Baseline for substructure and shape variables

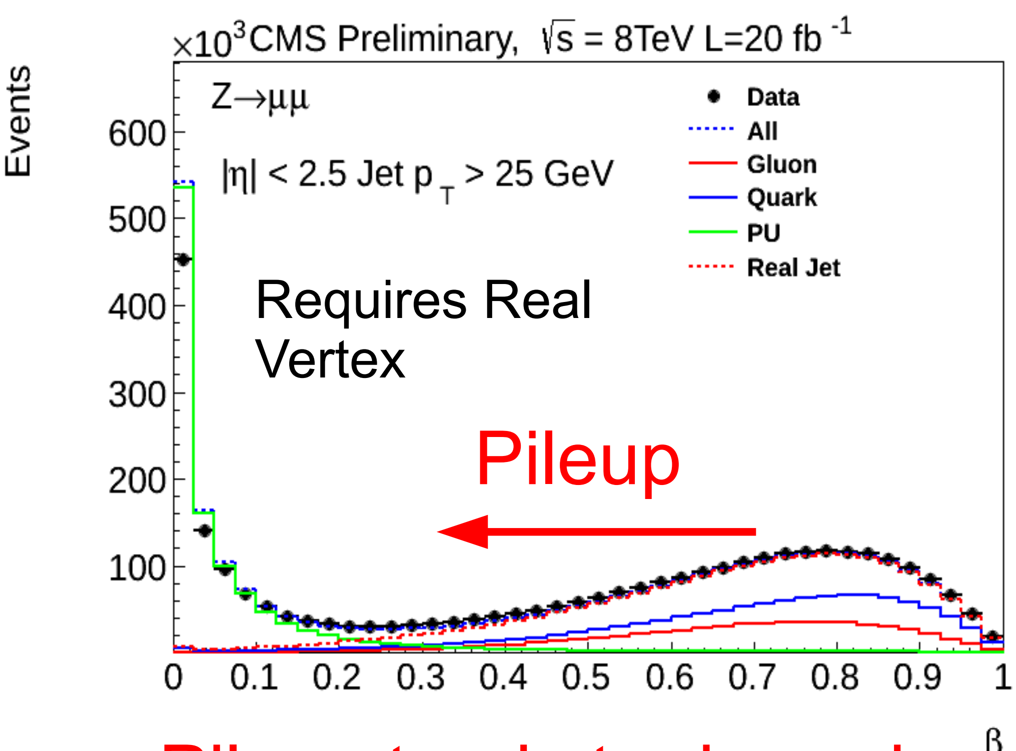
Apply Pileup jet Id on either (separate for each algo)

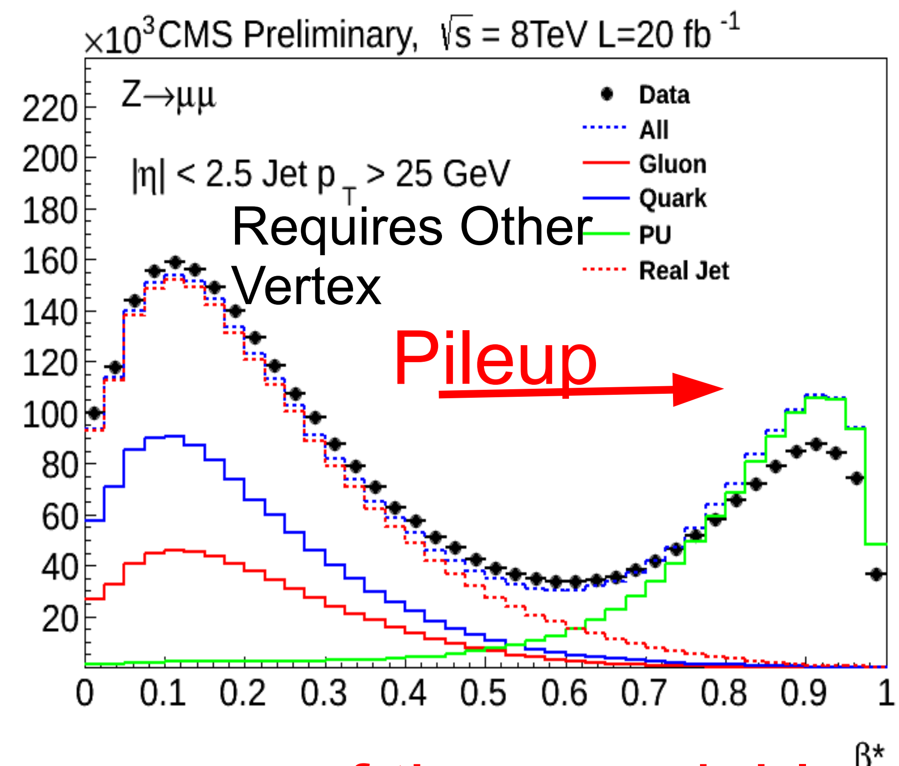
Pileup Jet Id Algorithm:Tracking

- 13 variables for the full discrimination
  - 4 Vertexing related variables (2 most impt shown):

#vertices, dZ of leading track in jet +

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

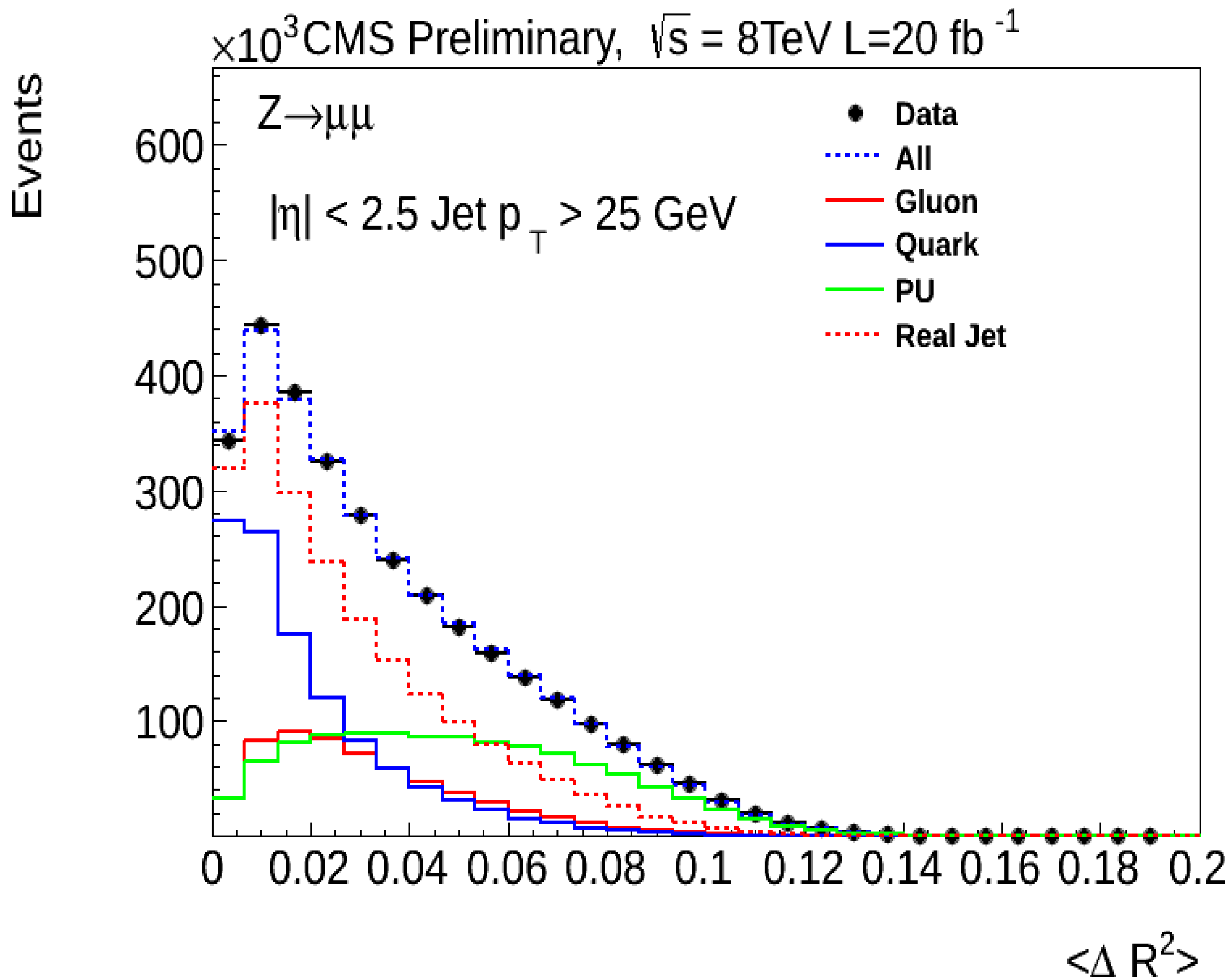




Pileup tends to degrade performance of these variables

Pileup Jet Id Algorithm:Shapes

- 13 variables for the full discrimination
  - QG disc: *pTD*,#Charged particles,#Neutral particles
  - 6 Shape variables

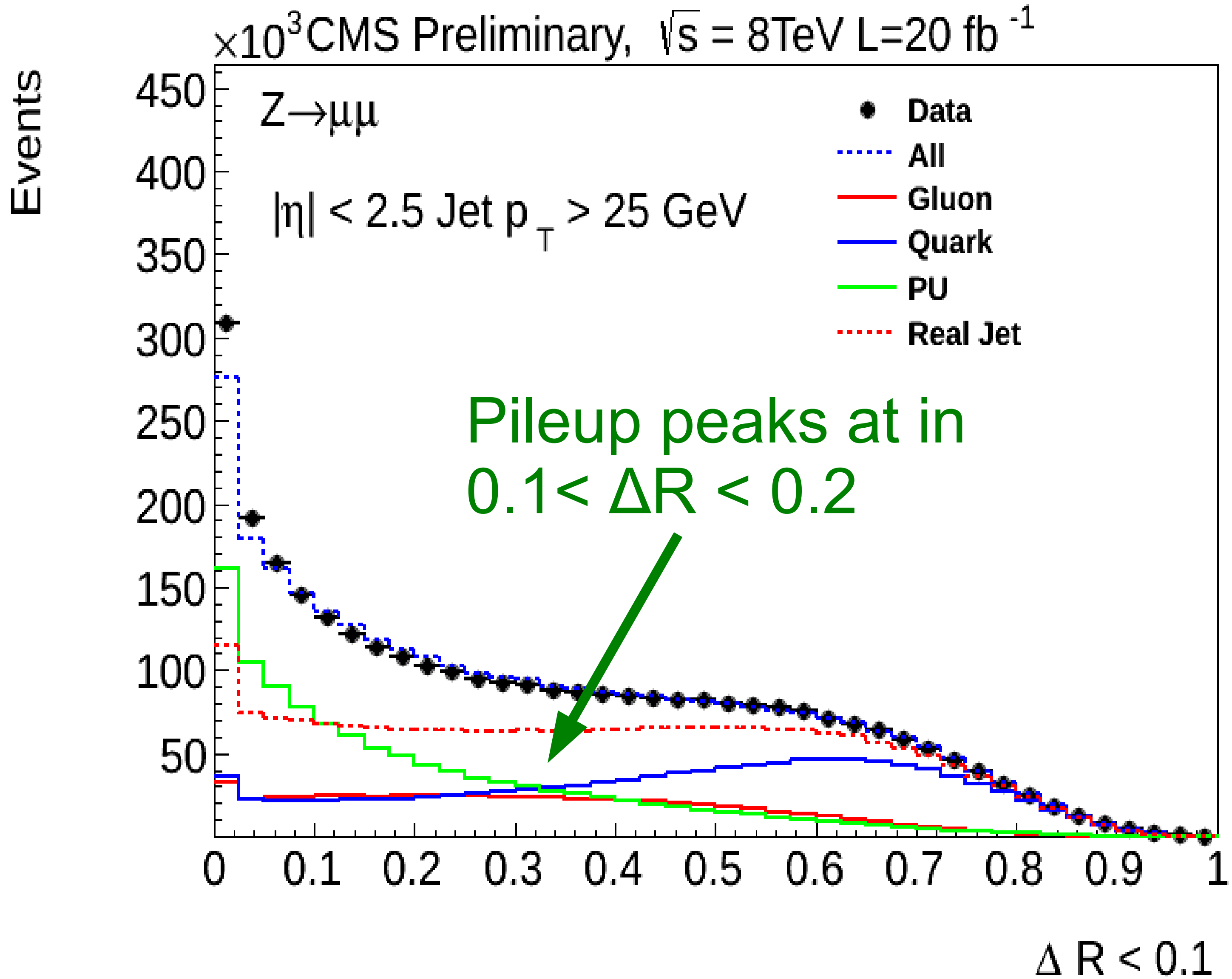


Most Effective
Jet shape variable

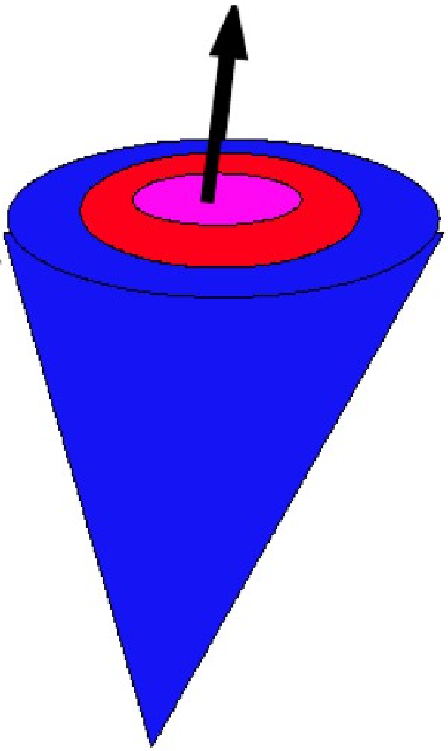
$$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R_i^2 p_{Ti}^2}{\sum_i p_{Ti}^2}$$

Pileup Jet Id Algorithm:Cones

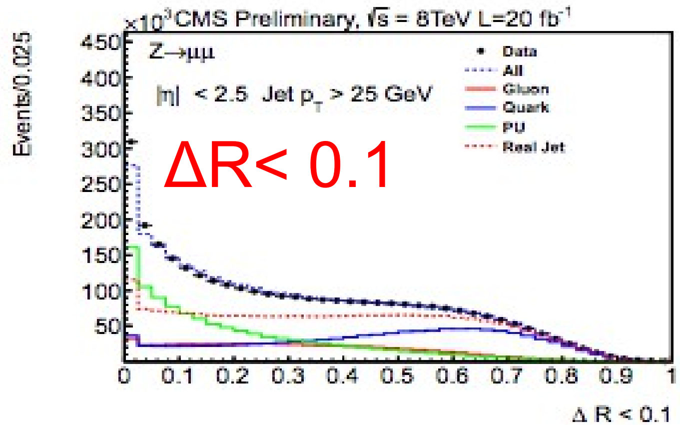
- Additional shape variables : ΔR annuli

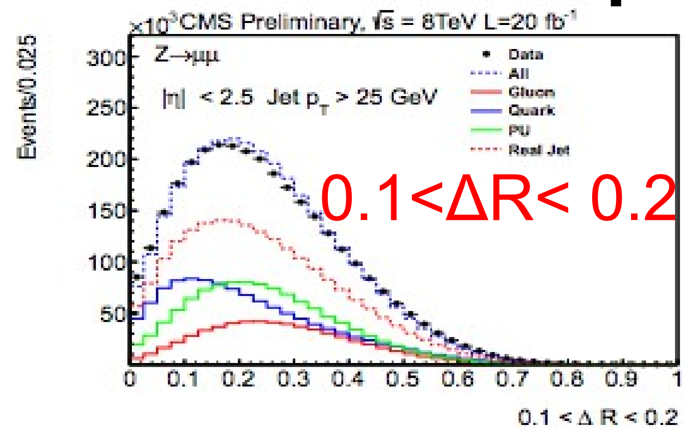


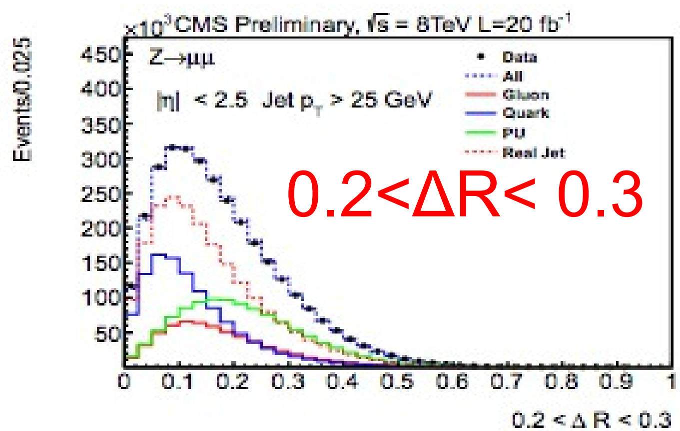
ΔR 0.1=>0.5

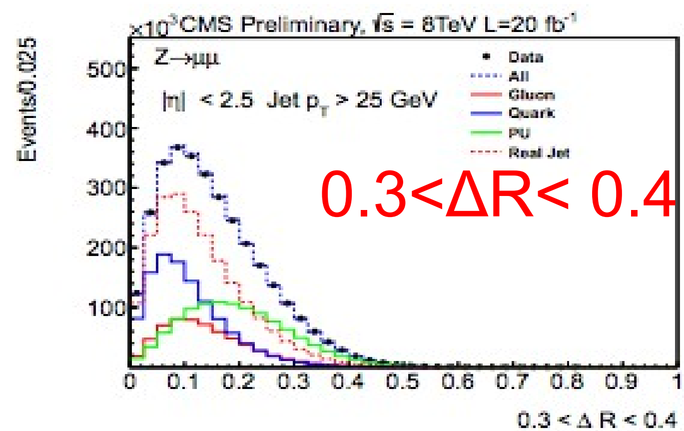


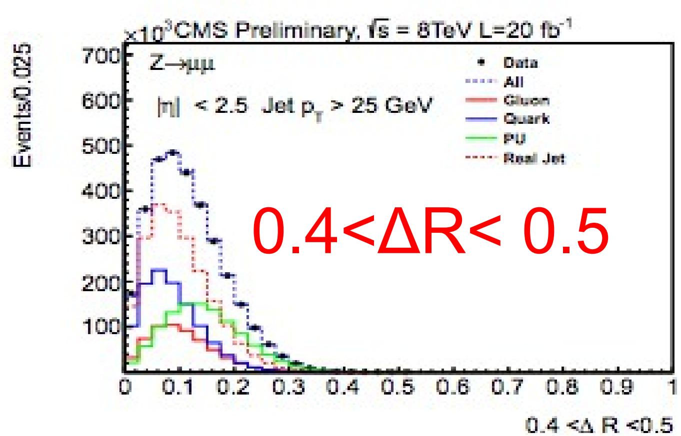
Evolution of the Jet Shape











Separation of PU clear when looking at the rings

Algorithm Construction

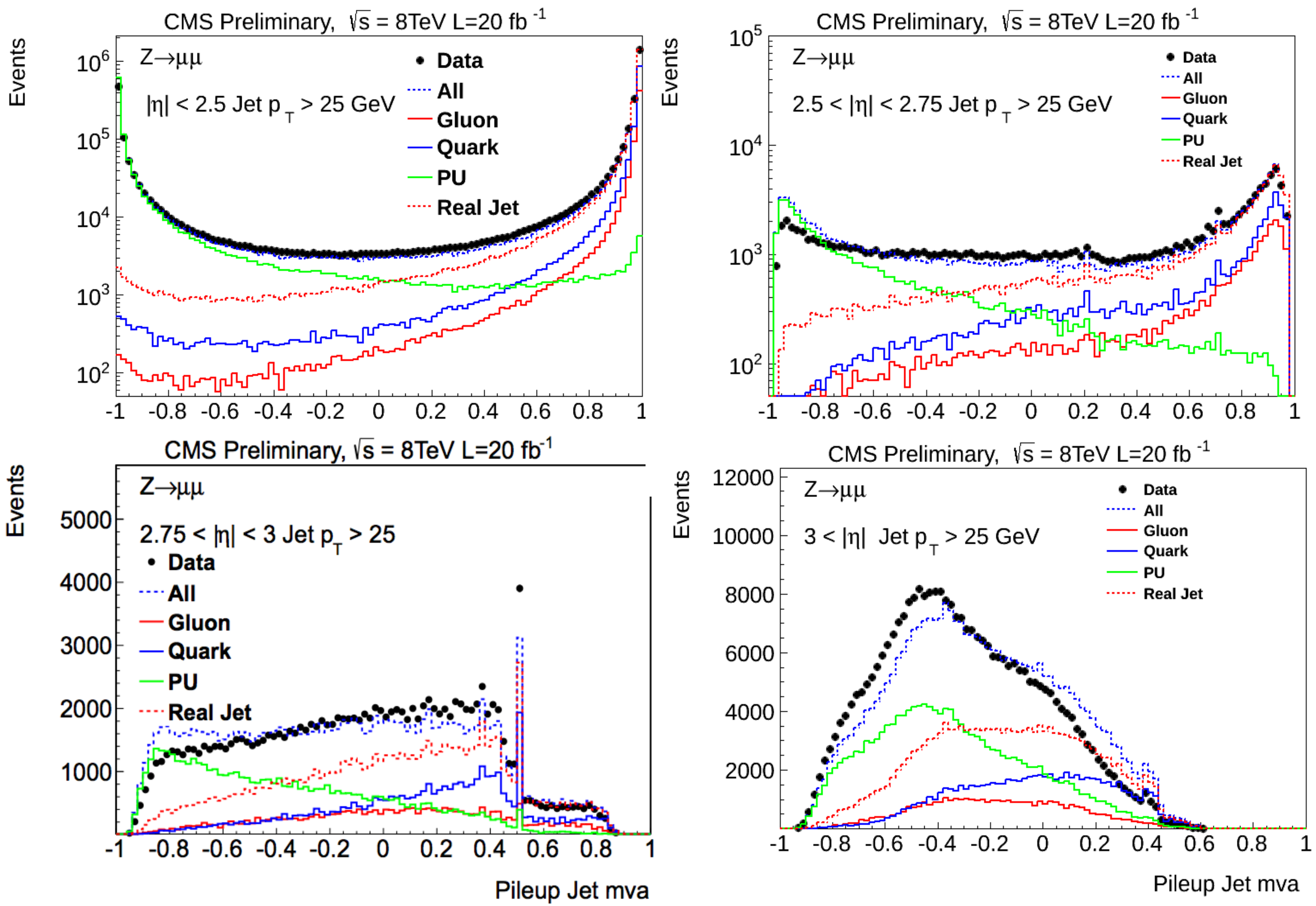
- Construct a Boosted decision tree real vs PU Jets
  - Train in four separate regions of η

$\|\eta\| < 2.5$ tracking Shape variables	$2.5 < \|\eta\| < 2.75$ Weak tracking (tracking ends at 2.5) Shape variables
$2.75 < \|\eta\| < 3.0$ Shape variables	$3.0 < \|\eta\| < 5.0$ Forward HCAL Shape variables

Construct a Boosted decision tree (trained on Z+jets for each)

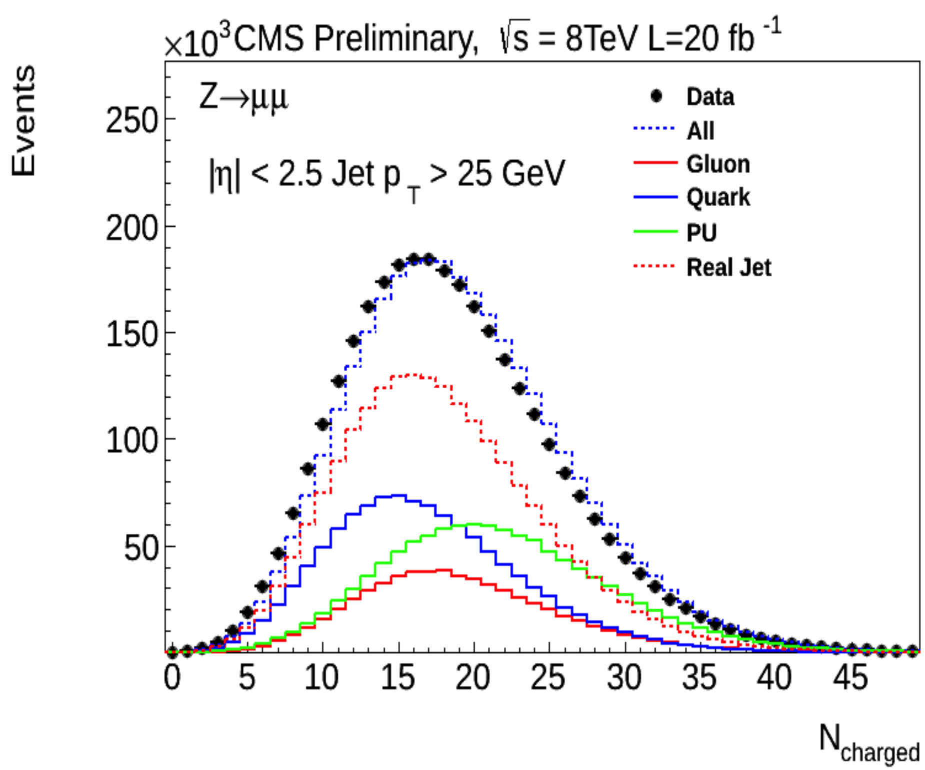
Pileup Jet Id in Data

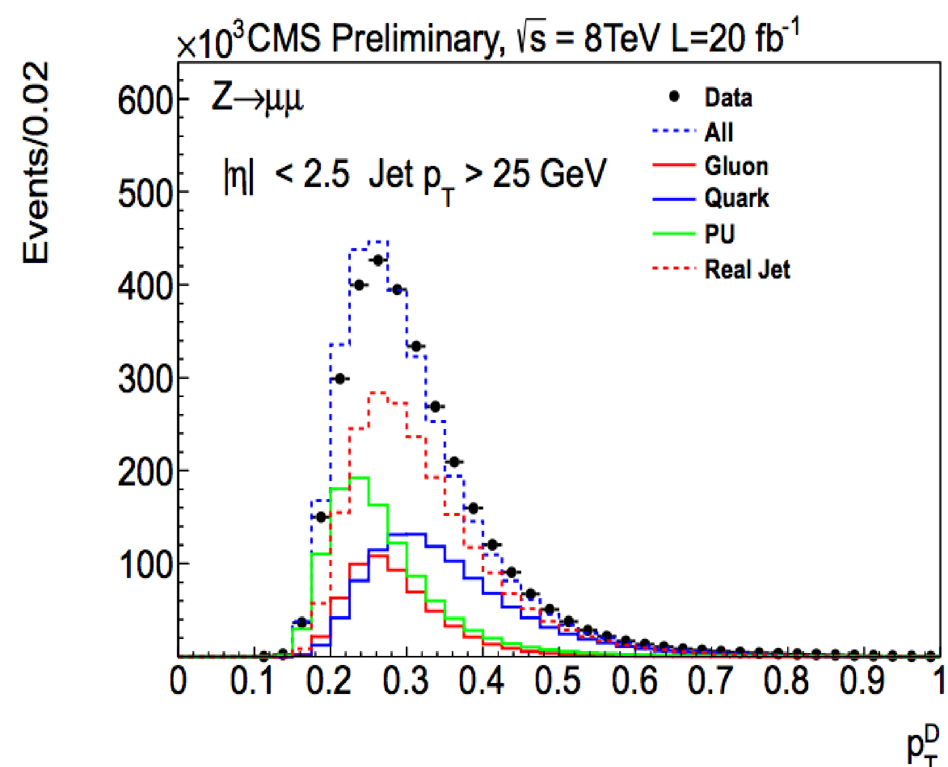
- Fraction of pileup grows with higher |η|



Quark/Gluon Discrimination in PU

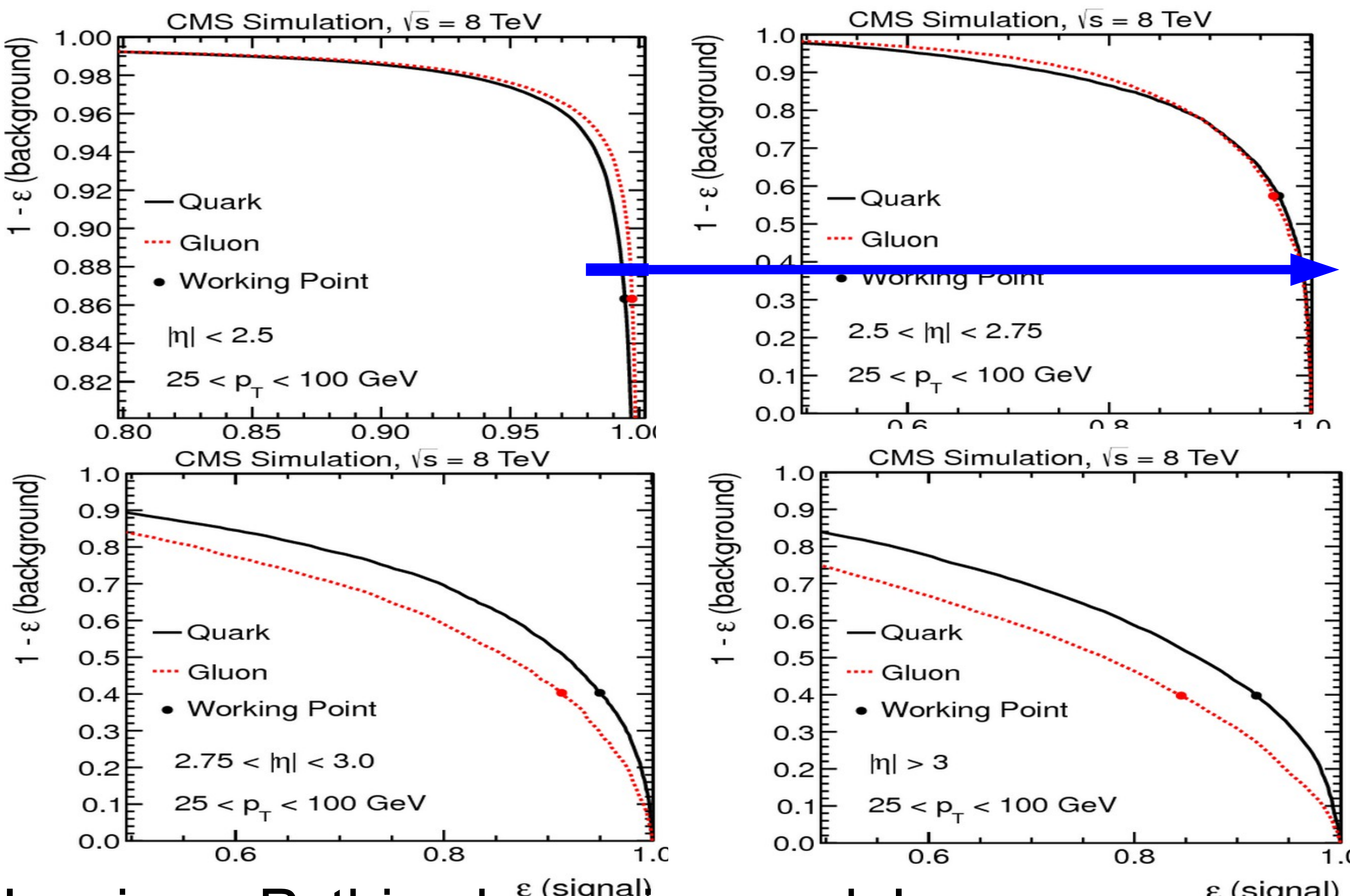
- Using jet shapes gives added advantage to PU
  - However: gives differences between Quarks/Gluons
- To improve Q vs PU and G vs PU
  - We add the quark gluon discrimination varaibles in





Pileup Jet Id Performance

- Largest systematic : from quark/gluon variation



CHS results consistent w/specific working point

Herwig vs Pythia showering models bounded by quark/gluon variation

Full Summary of Variables

$\langle \Delta R^2 \rangle$

$A < (\Delta R) < A + 0.1$

Shape

$$\longrightarrow \frac{1}{p_T^{jet}} \sum_{i \in A < \Delta R < A+0.1} p_{Ti}$$

$N_{charged}$

Q/G

$N_{neutrals}$

p_T^D

$$\longrightarrow p_T^D = \frac{\sqrt{\sum_i p_{Ti}^2}}{\sum_i p_{Ti}}$$

$$\beta = \frac{\sum_{i \in PV} p_{Ti}}{\sum_i p_{Ti}}$$

Tracking

We rely on tracks associate to vertices

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

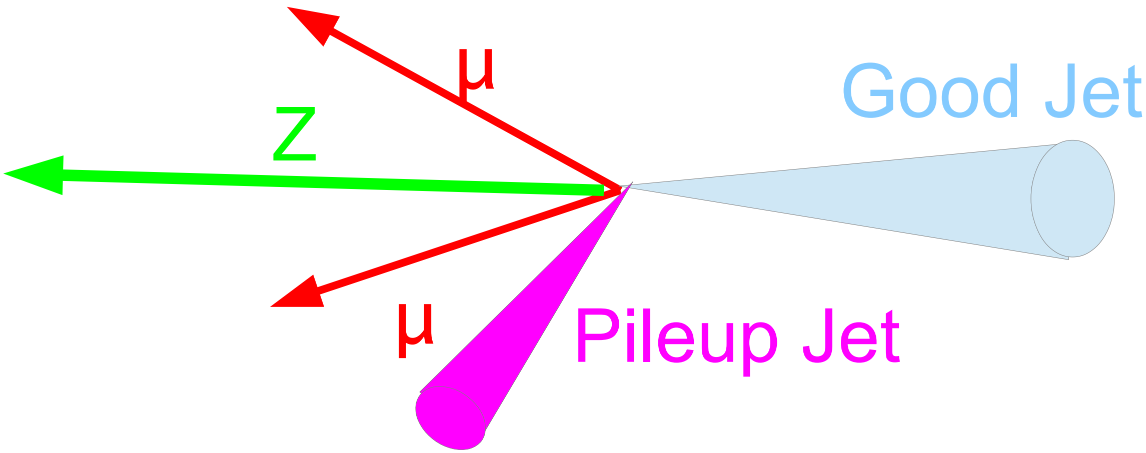
d_Z

$n_{vertices}$

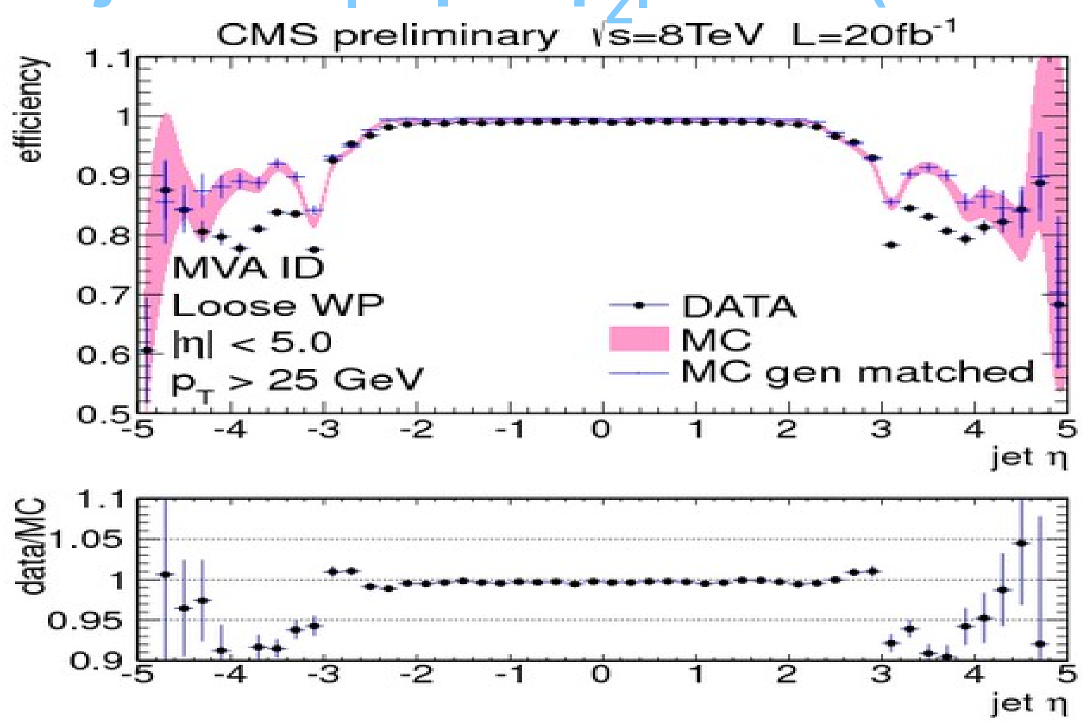
Event → Measure of the event scale

Pileup Jet Id: Efficiency in Data

- Use Z+Jet balance to measure efficiency in data



Good jet eff: |Δφ – φ_Z| > 3.0 (subtract scaled PU)



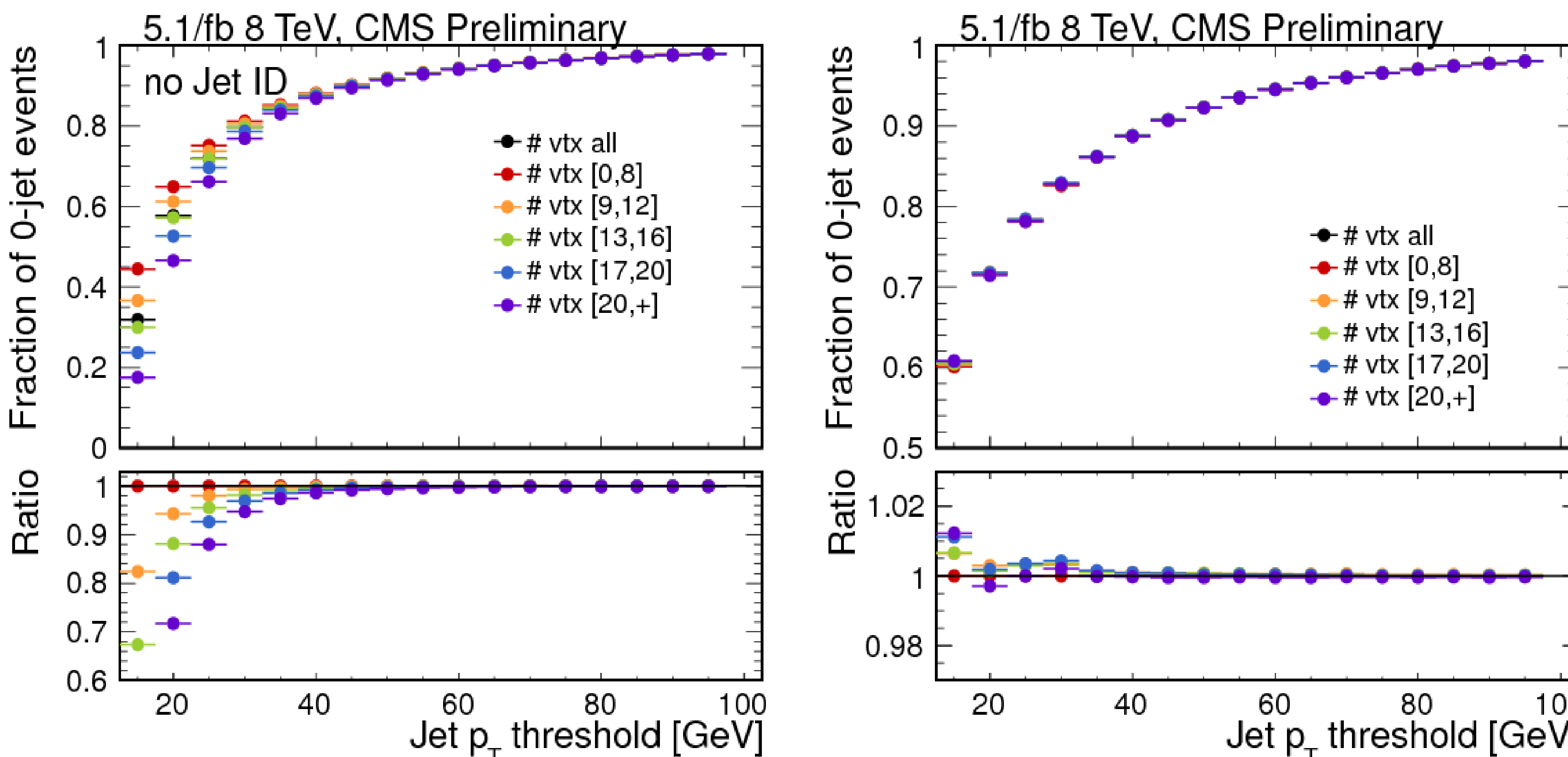
Example VBF:

VBF Eff: 90%

VBF Fake rate 15%

Usage Examples: Jet Vetos

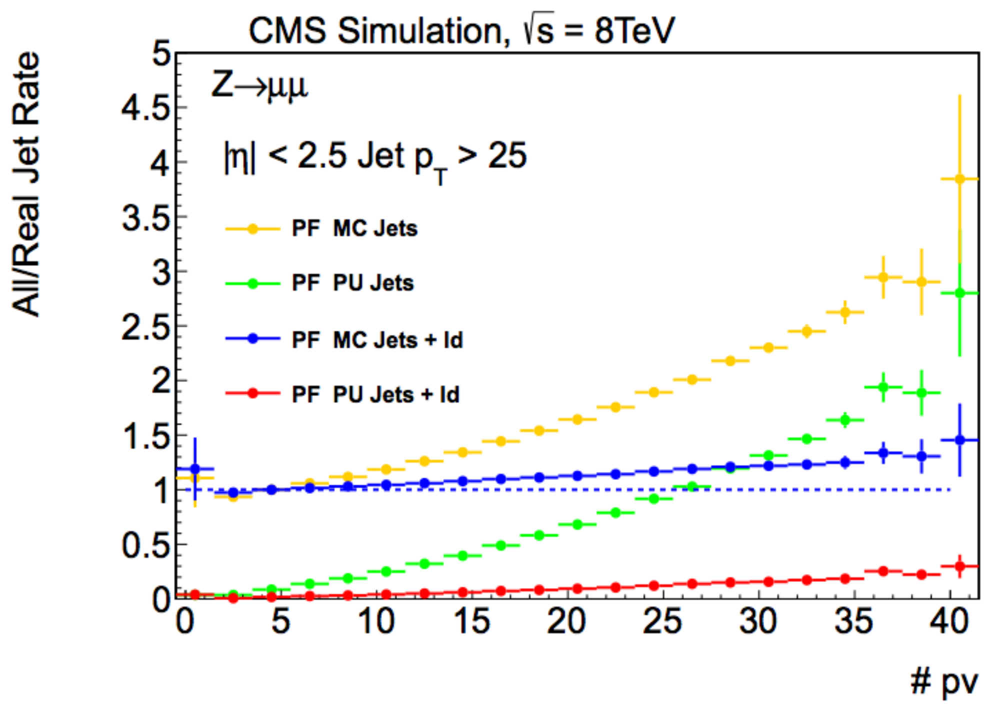
- Pileup jet id allows extension of jet vetos to low *pT*
  - Critical for b-tag veto (requires jets with *pT* > 10 GeV)

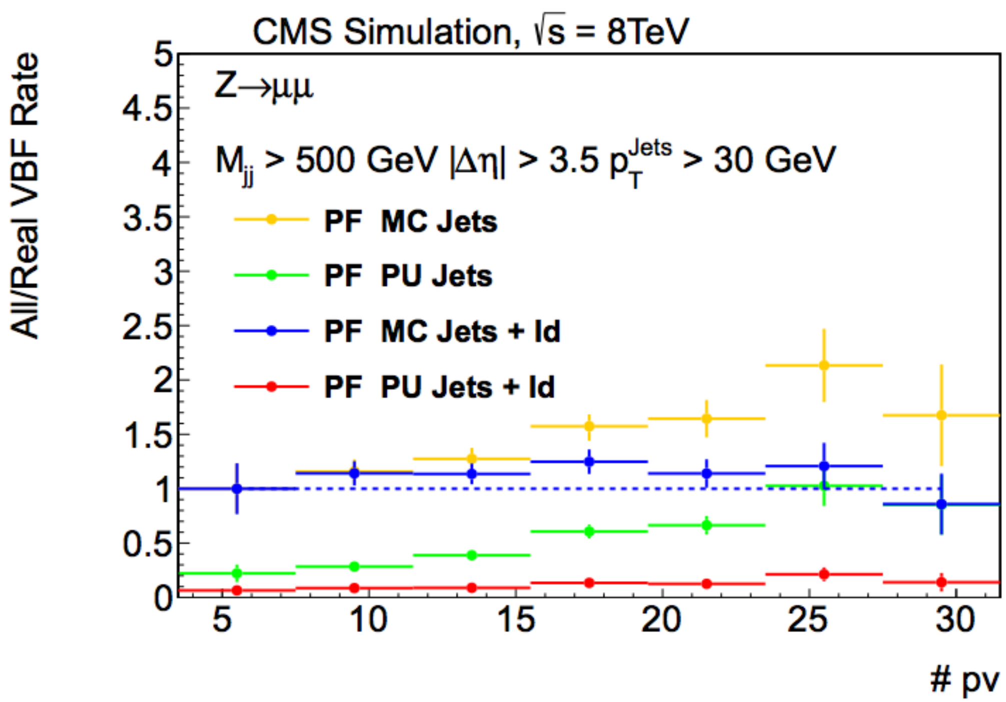


Vector boson fusion background reduced by a factor of 2 (90% good jet eff)

Reducing VBF backgrounds

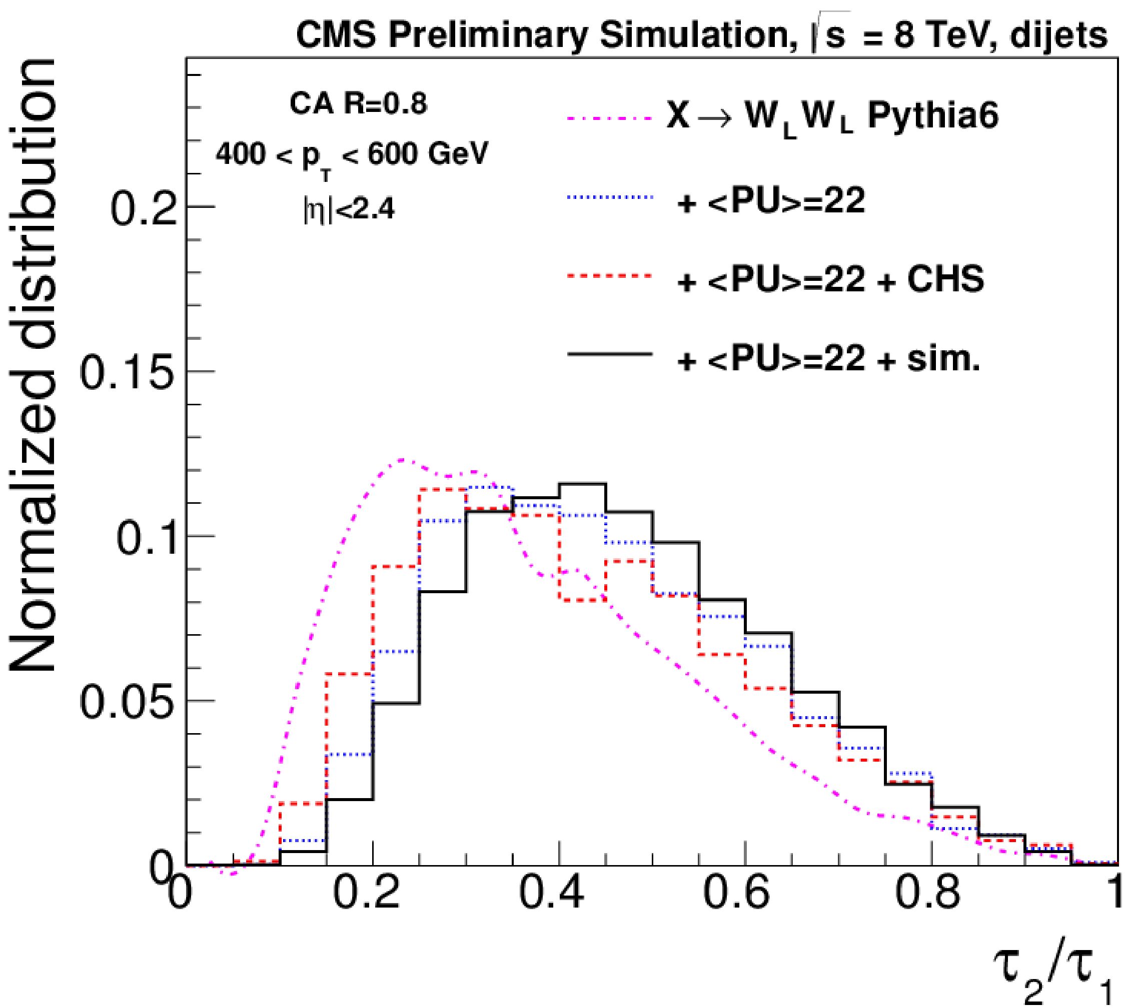
- Critical area for the PU Jet Id
  - Rely on both tracking and non-tracking pu rejection
  - PU Jet id was used in every VBF analysis
- Forward jets play an important role in *MET*
  - big advantage of using PU Jet Id over CHS





W Substructure : CHS

- Charge hadron subtraction clustering default
  - Jet shape effects minizmied with the CHS



Charged hadron subtraction reduces the Pileup effect by 2

Conclusions

- Pileup Jet Id has been a success
  - However it is just one approach to pileup rejection
  - Discrimination in both forward/central region
  - Critical element for reducing *MET* resolution
    - See Mathieu's talk on *MET* performance
- CHS has also worked well
  - A complementary approach to PU Jet Id
  - Pileup Jet Id can be run on CHS jets
  - Used in about 60% of jets in CMS
  - CHS gives an improvement in the jet shapes
- Look forward to extending with new approaches

Backup

Efficiency vs Kinematics