# Jet Identification with CMS detector in proton-proton collisions at $\sqrt{\sigma}$ =13TeV

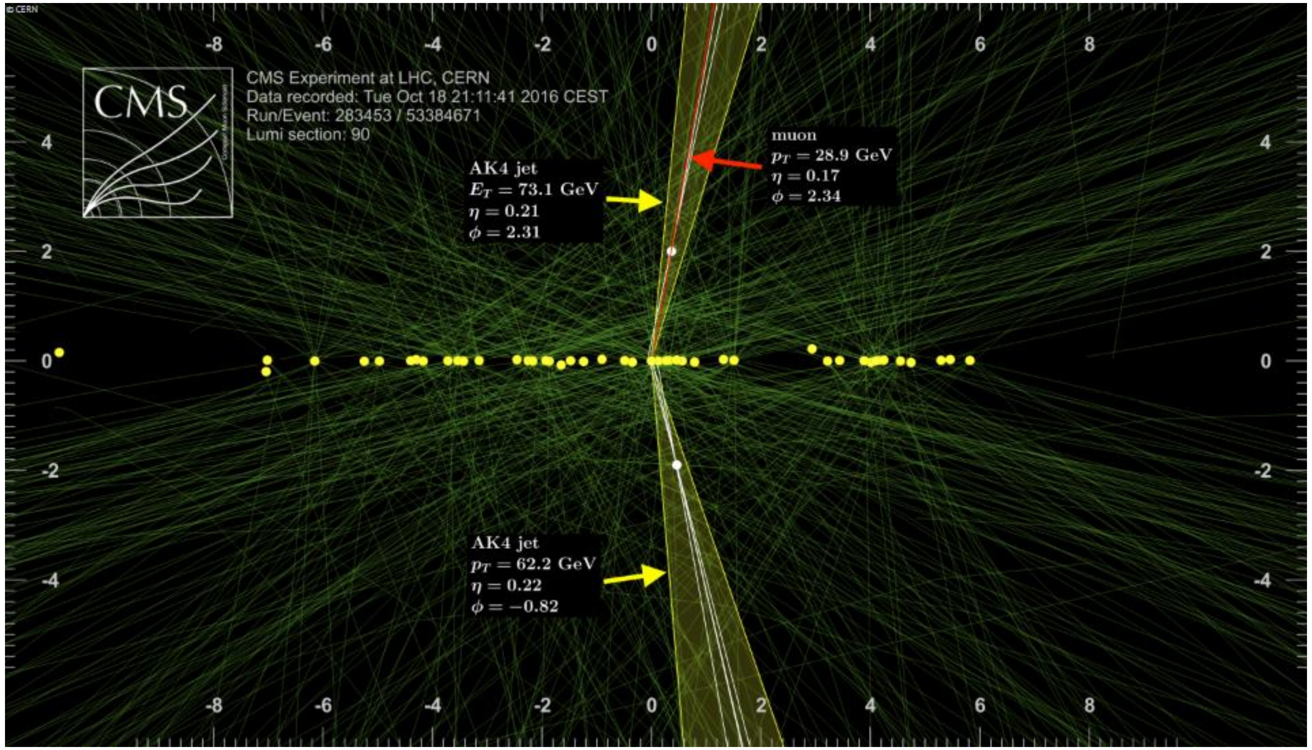
National Technical University of Athens

Conference on Recent Developments in High Energy Physics and Cosmology

K.Kousouris, G.Tsipolitis, A.Zacharopoulou (NTUA)

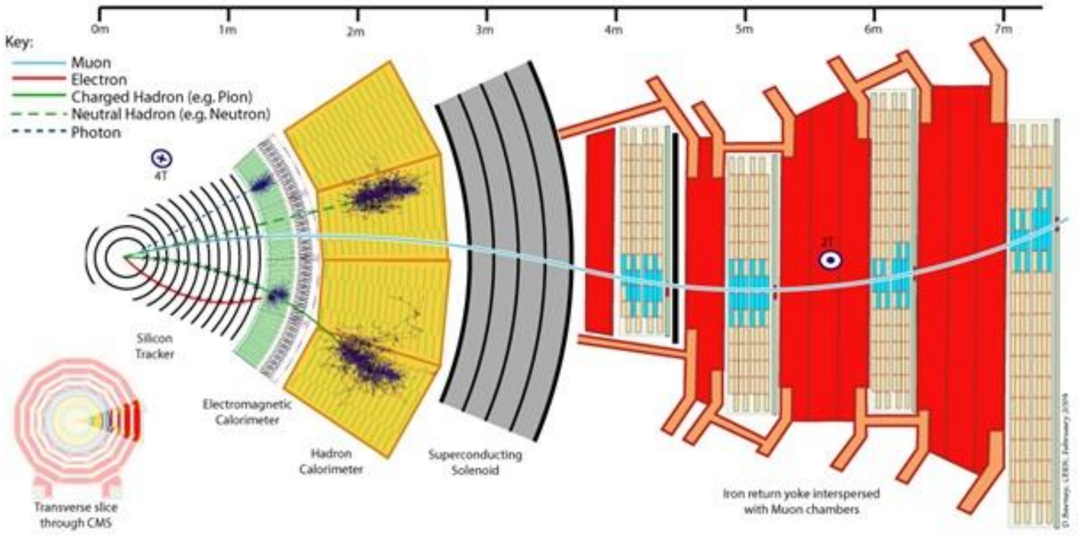
# Outline

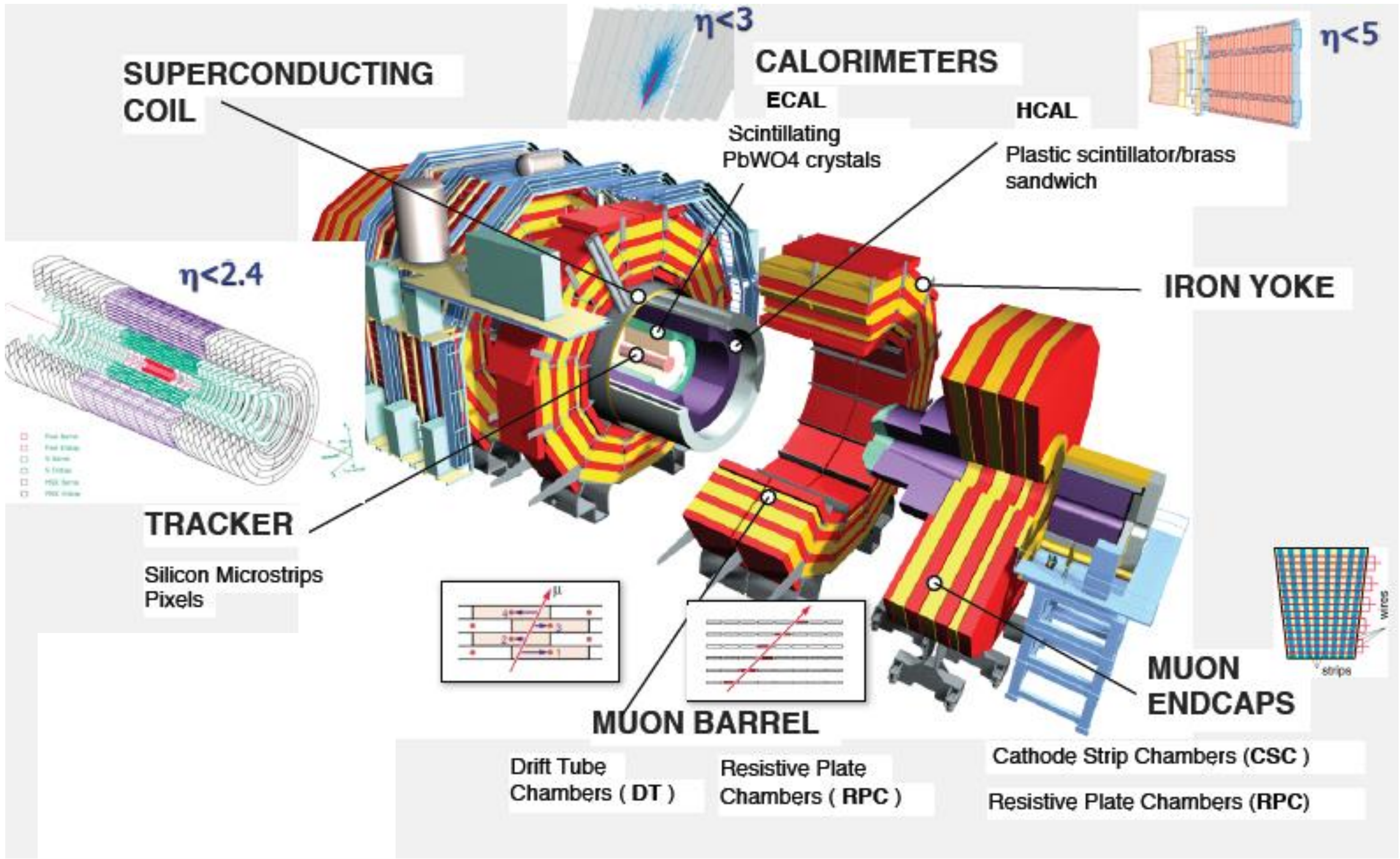
- Introduction
  - CMS detector
  - Hadronic Jets & pileup
- Analysis
  - Training
  - Performance in data
- Results



# CMS Detector

- Compact Muon Solenoid
- General interest
- 100m under the surface
- Total weight: 14.000 t
- Total length: 28.7m
- Total diameter: 15m
- Magnetic field: 3.8T



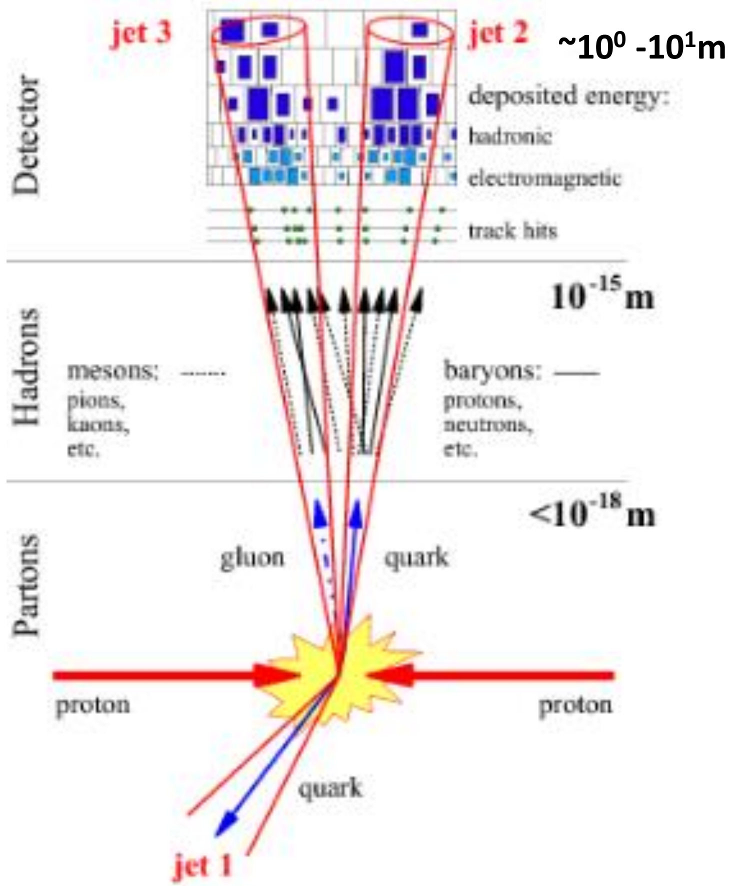



# Hadronic Jets & Pile Up

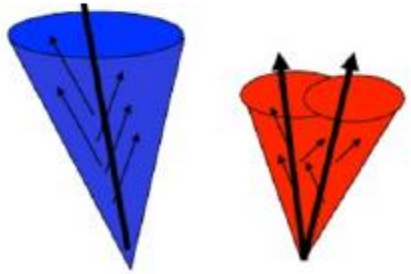
LHC: high frequency of bunch crossing
1 every 25ns
each bunch ~$10^{11}$ p

→ Pile up contamination → Correction is necessary

For this analysis PUPPI method was used




**Jets** ≈ local energy clusters accompanied by particle trajectories



Prompt jet

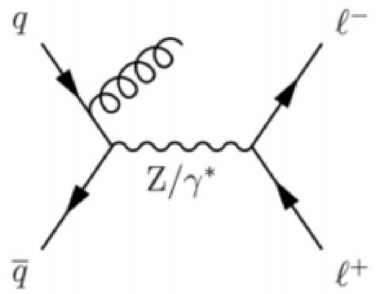
Pile up jet

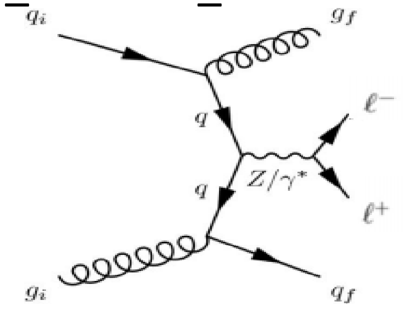
Prompt (typical) jet -> Primary vertex

Pile up jet -> Secondary Primary vertex

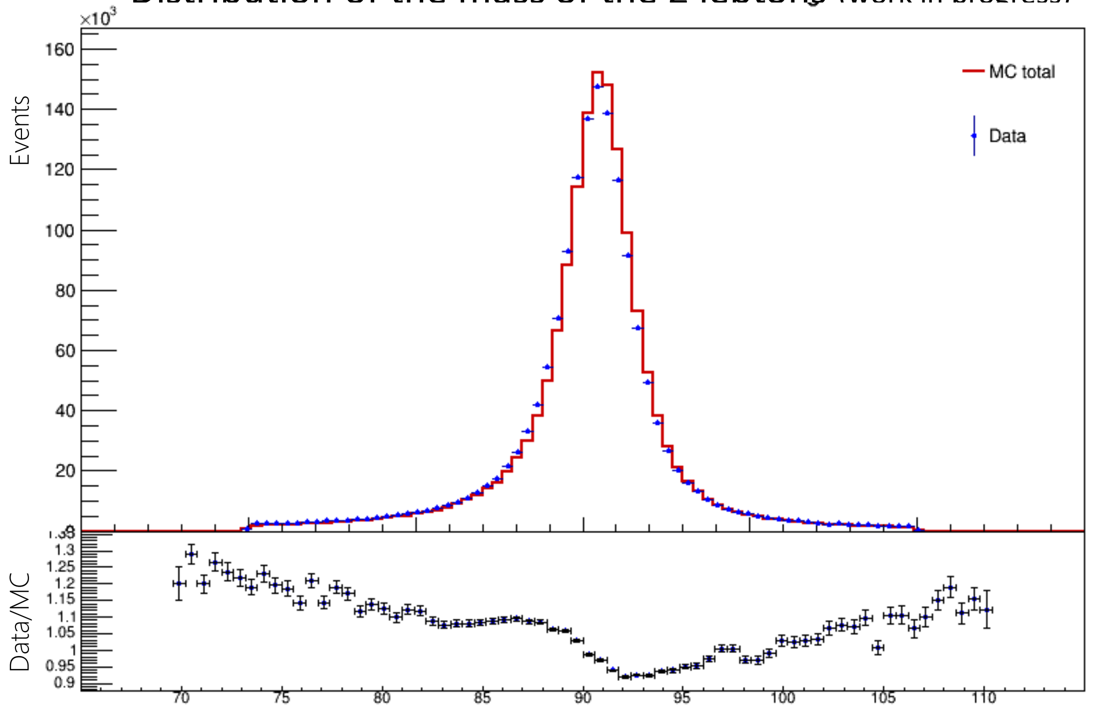
# Data and MC samples

- Study performed in Z(μμ)+jets sample
  - clean sample that allows to define prompt-jets and pileup-jets enriched regions
- Dataset (2017)
  - /DoubleMuon/Run2017B-31Mar2018-v1/MINIAOD
  - /DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD
  - /DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD
  - /DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD
  - /DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD
- MC samples
  - DY1JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8
  - DY2JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8
  - DY3JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8
  - DY4JetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8




Distribution of the mass of the 2 leptons (work in progress)

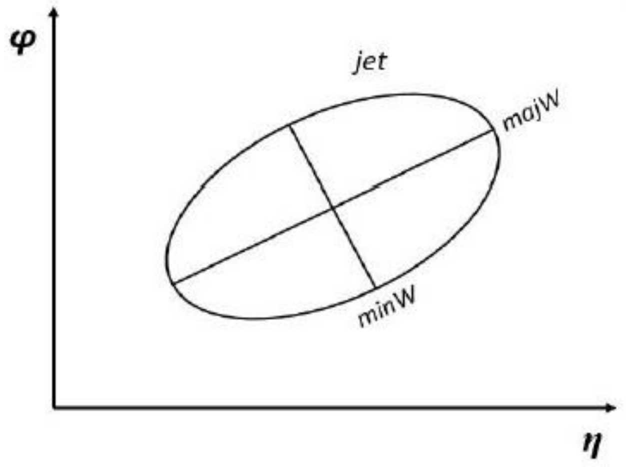


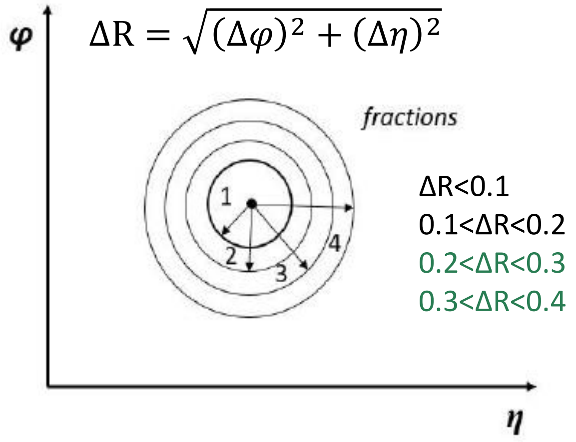

Mass of 2 leptons (GeV)

# Multivariate discriminant - Input Variables

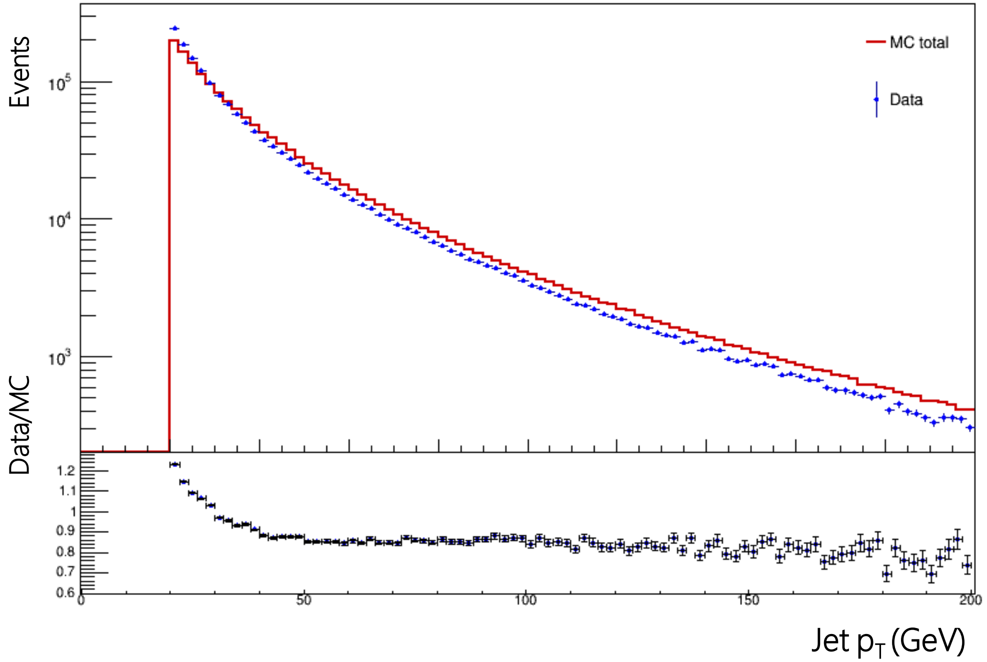
**For central |η|<3 and forward 3<|η|<5 region**

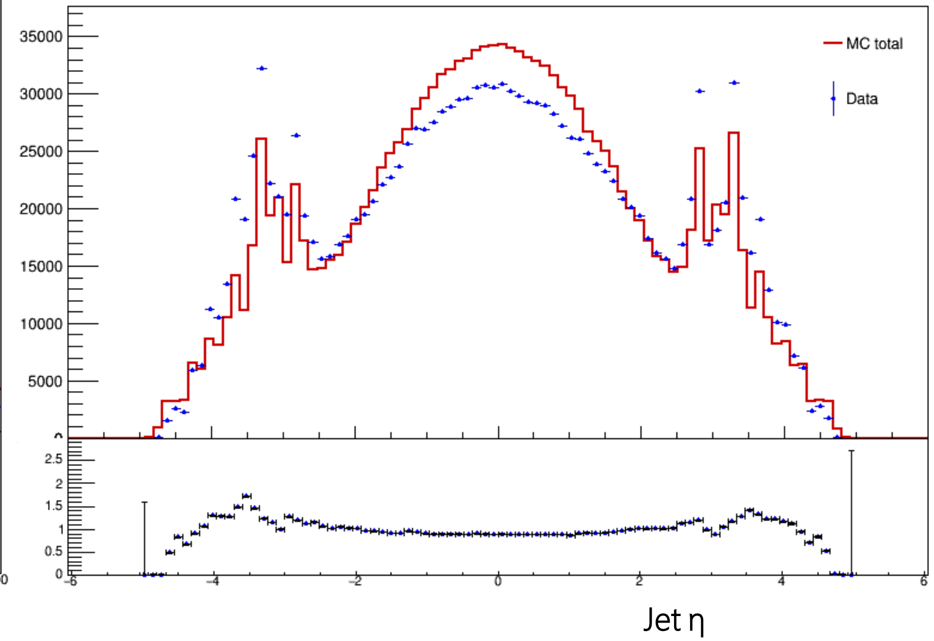
- Number of charged particles
- Number of constituents
- Transverse momentum of jet
- Pseudorapidity of jet
- Fraction of jet tracks associated with the primary vertex
- $p_T$-weighted average $p_T$ of constituents
- Magnitude of the pull vector
- Fraction of constituents' $p_T$ contained in the region ΔR<0.1 around the jet axis
- Fraction of constituents' $p_T$ contained in the region 0.1<ΔR<0.2 around the jet axis
- Fraction of constituents' $p_T$ contained in the region 0.2<ΔR<0.3 around the jet axis
- Fraction of constituents' $p_T$ contained in the region 0.3<ΔR<0.4 around the jet axis
- Minor axis of the jet ellipsoid in the η-ϕ plane
- Major axis of the jet ellipsoid in the η-ϕ plane
- Fraction of the $p_T$ carried by the leading charged constituent



# Multivariate discriminant - Input Variables

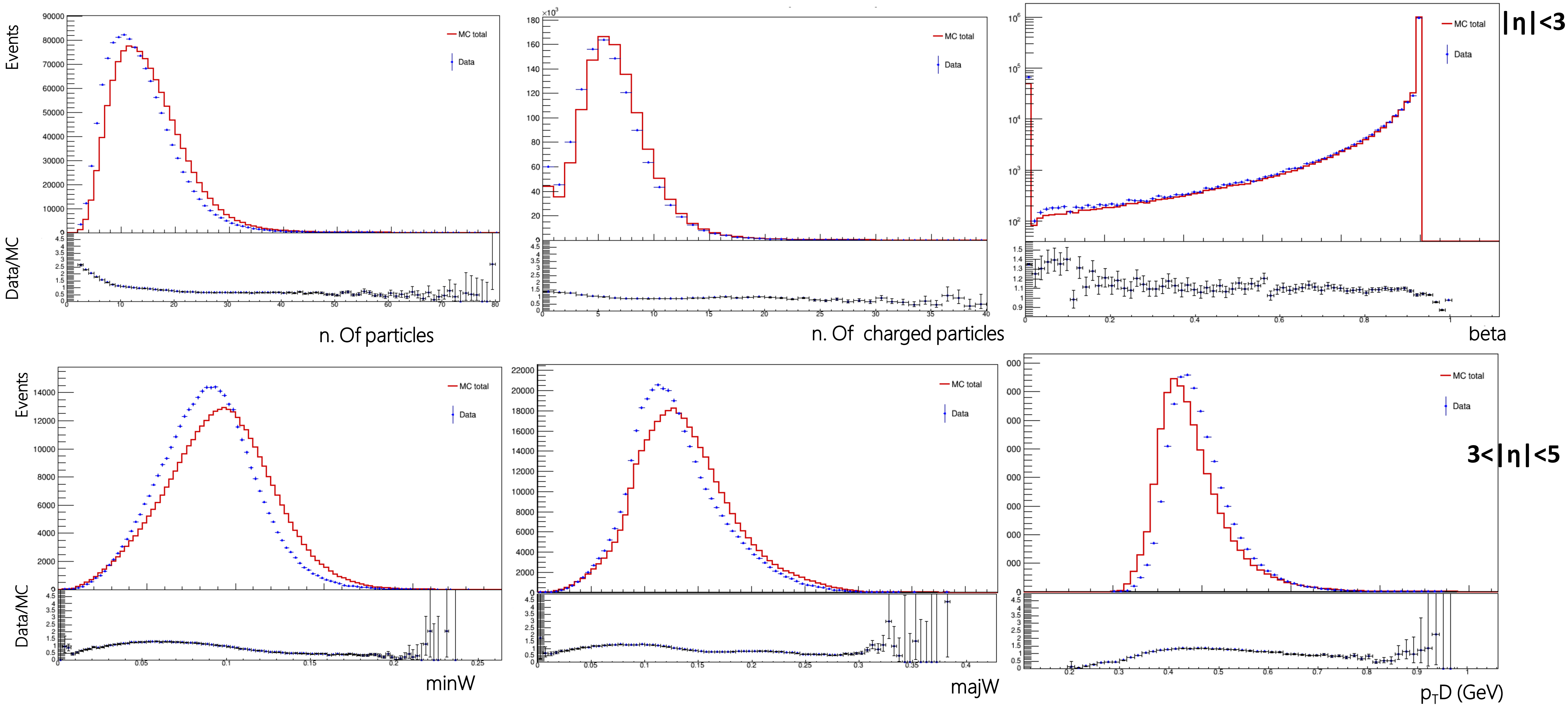



**Jet Collection**

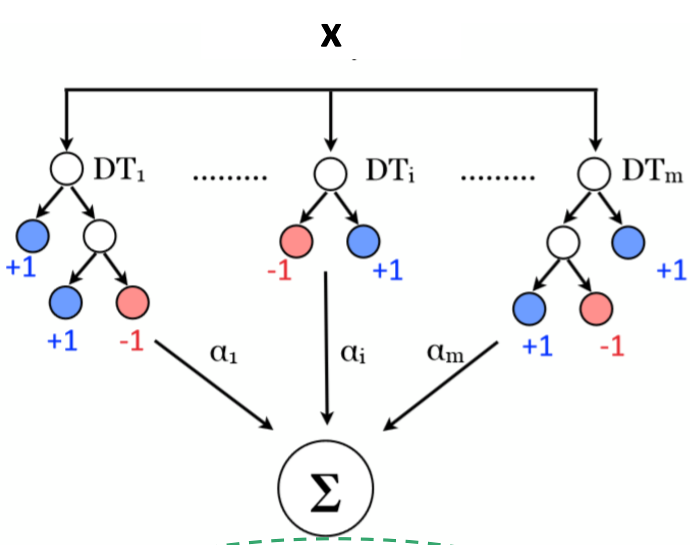
- $p_T$ >20GeV
- $|\eta|<5$

# Multivariate discriminant - Input Variables (central region )



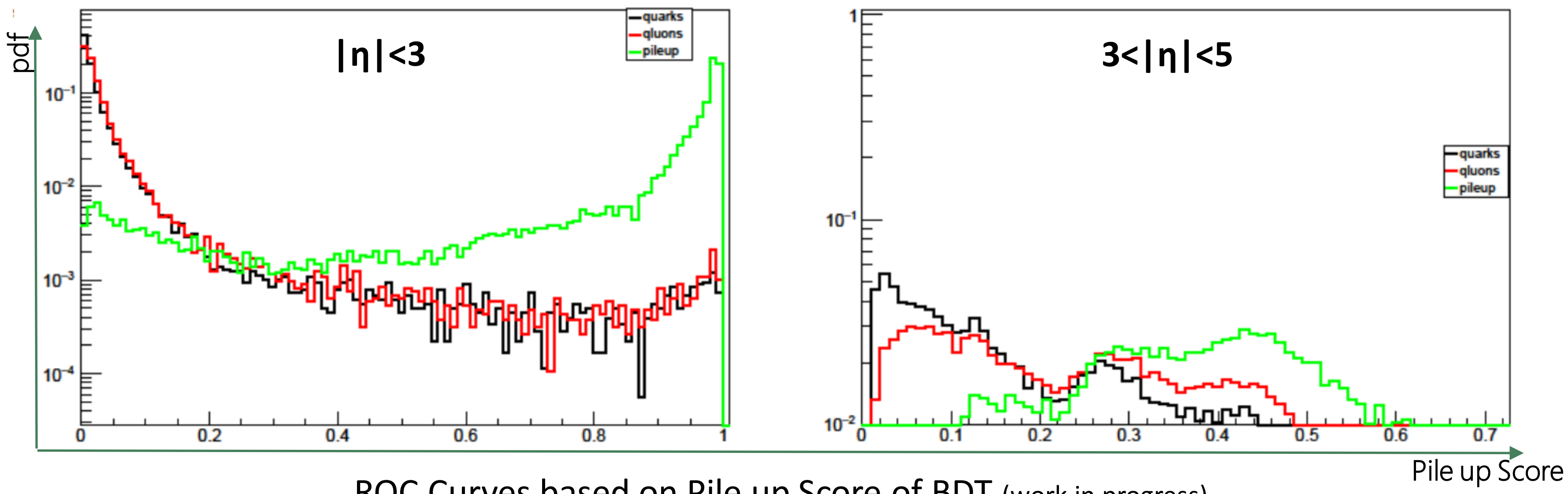
# Multivariate discriminant - Training

## Boosted Decision Trees

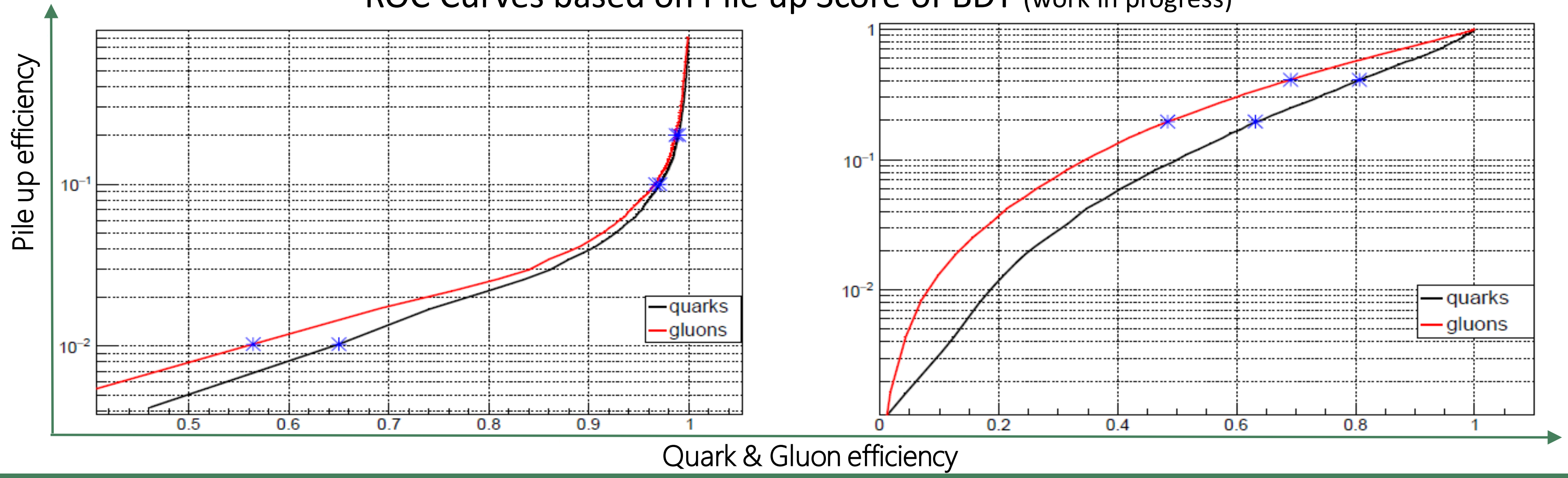



$$F(\vec{x}) = \sum_m a_m \varphi(\vec{x};\theta_k)$$

Distributions of BDT output Pile up Score for each class (work in progress)




ROC Curves based on Pile up Score of BDT (work in progress)



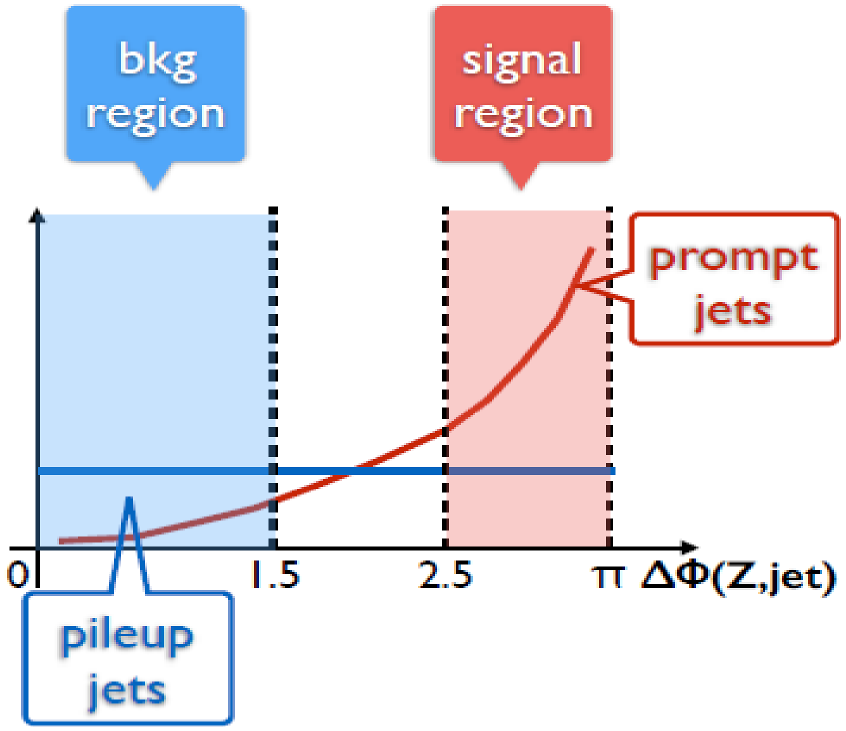
# Efficiency of Data & MC

- **Tag and probe method**
  - leading jet (probe) recoils against Z
  - Strong requirements on the Z+jets back-to-back topology ensure that the probe is most probably a prompt jet
  - Pile up jets have no azimuthial correlation with Z (flat $\Delta\Phi$ distribution)
- **Jet Selection**
  - Z event
  - $p_T$ balance: $0.5 < jetP_T / Zp_T < 1.5$
  - $|\Delta\phi(Z, jet)| > 2.5$->prompt (signal) region
  - $|\Delta\phi(Z, jet)| < 1.5$->pileup dominated (bkg) region
- **Efficiency**

$$\epsilon = \frac{N_{\text{pass Id, sig}} - k N_{\text{pass Id, bkg}}}{N_{\text{all, sig}} - k N_{\text{all, bkg}}}$$

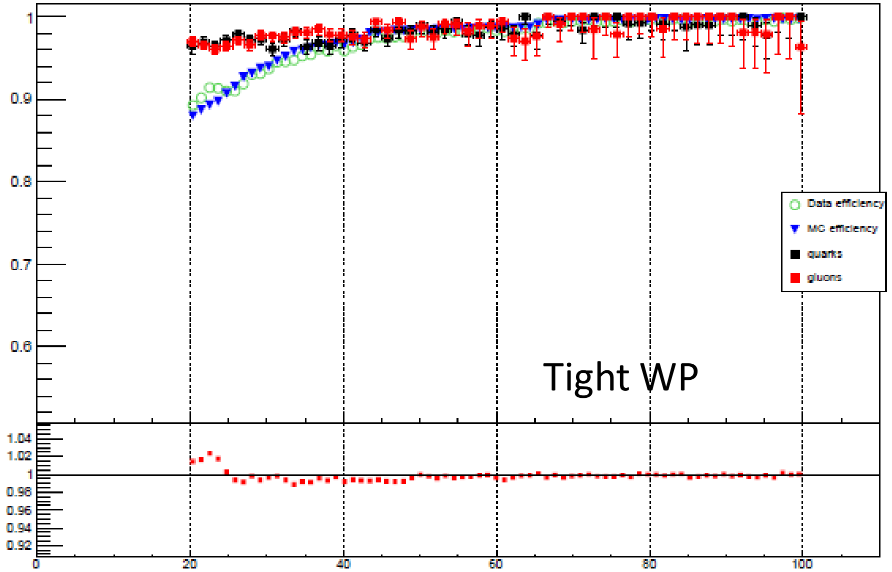
k:scaling factor
$k=(\pi-2.5)/1.5$

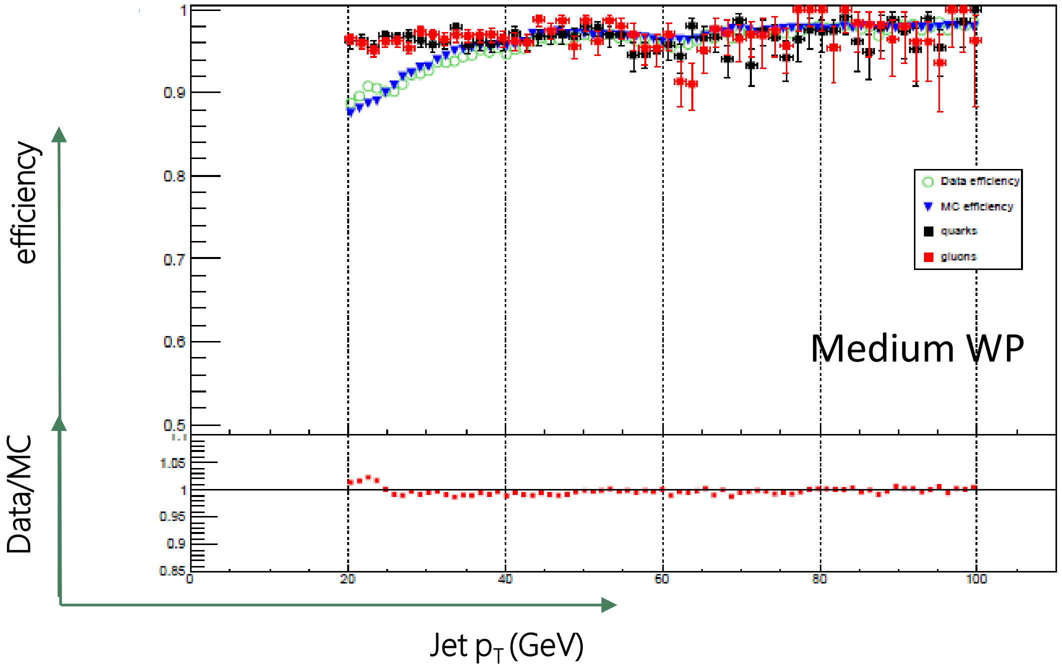
**pass,Id** ⟶ signal or background events which pass the cut criteria

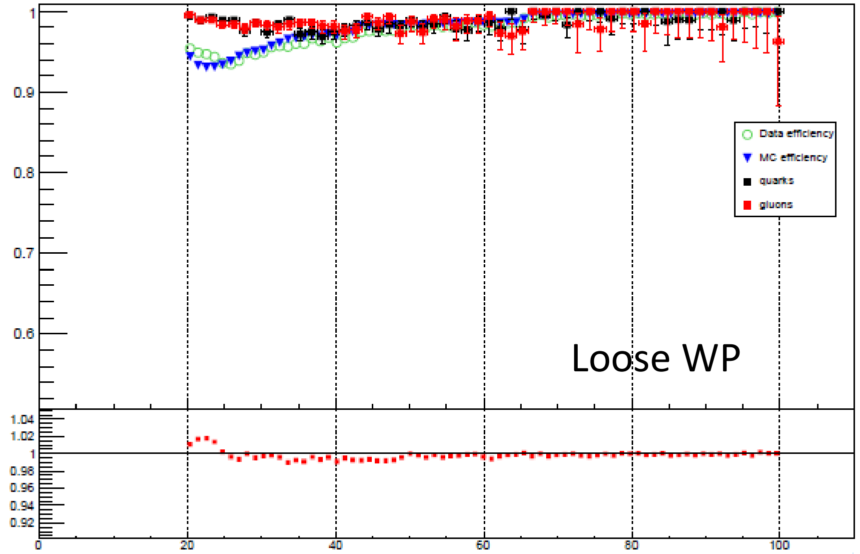


# Distributions of Efficiency of Data and MC as a function of jet $p_T$

(work in progress)



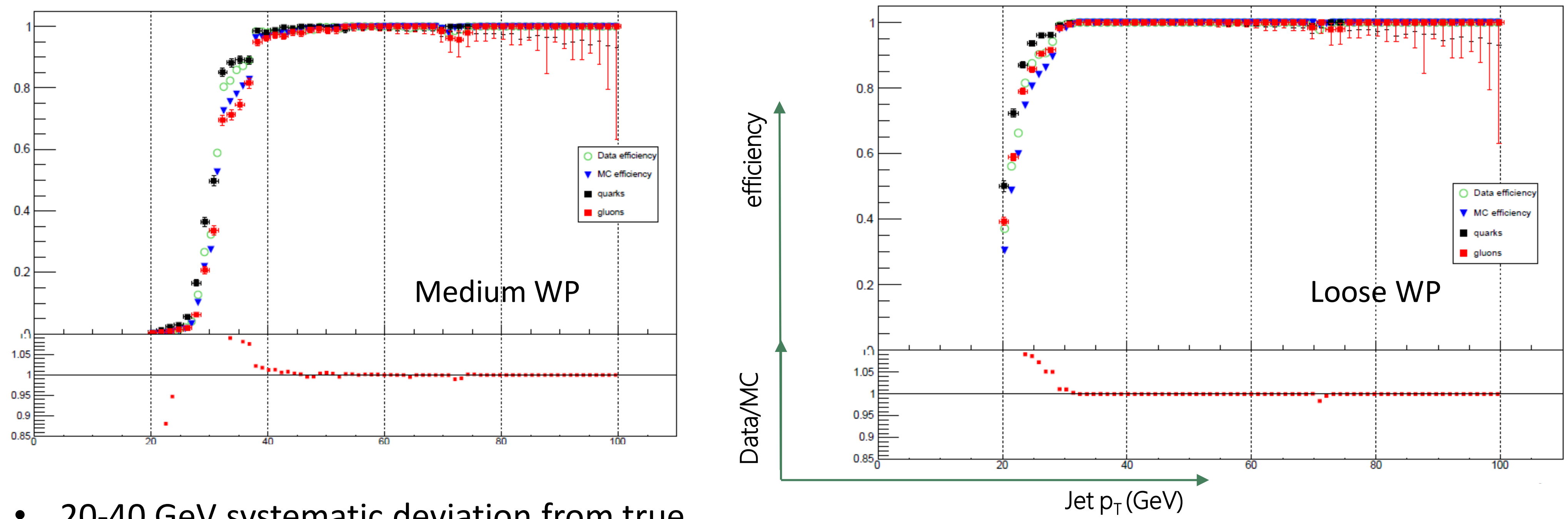

**|η|<3**

Quark – Gluons
"True efficiency"

- 20-30 GeV systematic deviation from true efficiency, discrepancy between data and MC
- ~30 -100 GeV reliable method
- 40-100 GeV agreement with true efficiency and between Data and MC

# Distributions of Efficiency of Data and MC as a function of jet $p_T$

(work in progress)




**3<|η|<5**

Quark – Gluons
"True efficiency"

- 20-40 GeV systematic deviation from true efficiency, discrepancy between data and MC
- Consistent trend of efficiency
- 40 -100 GeV reliable method
- 40-100 GeV good agreement with true efficiency and between Data με MC

# Summary & Results

- A method of separation and identification of prompt jet over pile up jet was developed
- The training of the pile up mva discriminant (BDT) was performed for 3 classes: Quark, Gluons and Pile up instead of 2 (Signal and Background) in both regions of the detector
- Working points were declared for each region as a function the transverse momentum of the jet
- Estimation of the "true efficiency" from the simulations MC
- Validation with the 2017 Data

- This data driven method that was developed seems to be biased in a part of the phase space
- In the region 20-30 GeV of $p_T$ there is about 9-10% deviation between data and MC
- Also in this region (20-30 GeV ) the efficiency of data and MC is systematically less than expected as calculated by the simulations
- For $p_T$ >30 GeV less discrepancy
- For the $p_T$ region 40-100 GeV this method seems to be reliable since the agreement between calculated efficiencies are greater

- In low $p_T$ region (20-40 GeV) the discrepancies are due to the fact that jet $p_T$ is used in the training of BDT and is the most discriminant variable in this region so in order to mitigate pile up signal is also reduced.
- It is not the desired feature and we are trying to improve.

Thank you for your attention!!

# Pile Up Per Particle Identification (PUPPI) Method

- **Goal:** mitigation of pile up
- To identify pile up uses both:
  - global information from the event
  - local information in particle level
- For each particle i: local variable $a_i = \log \sum_{j \in event} \frac{p_{T\,j}}{\Delta R_{ij}} \times \Theta(R_{min} \leq \Delta R_{ij} \leq R_0)$

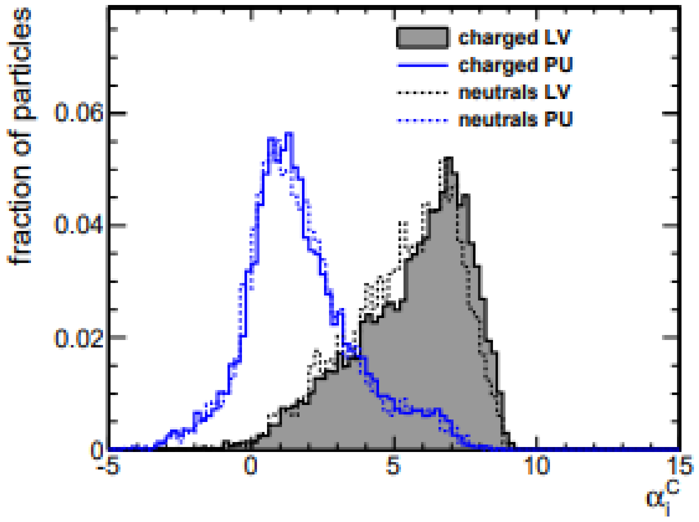
  hard scatter — pile up

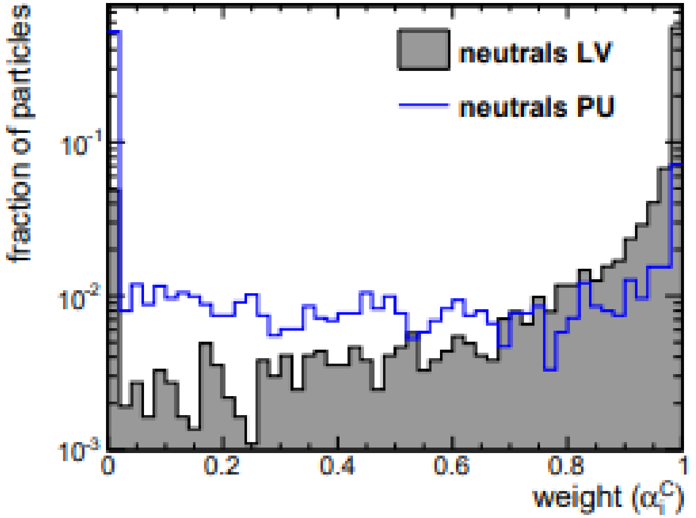
  distinguishes **parton shower-like radiation** over **pileup-like radiation**

- |η|<3 → charged particles from PV or PU vertex
- Distribution of α for charged pile up $\alpha^C_{PU}$ → proxy for total pile up
- Assign a weight to each particle by comparing its value $a_i$ to $\langle a \rangle$

  (pile up) **0<weight<1** (hard scatter)

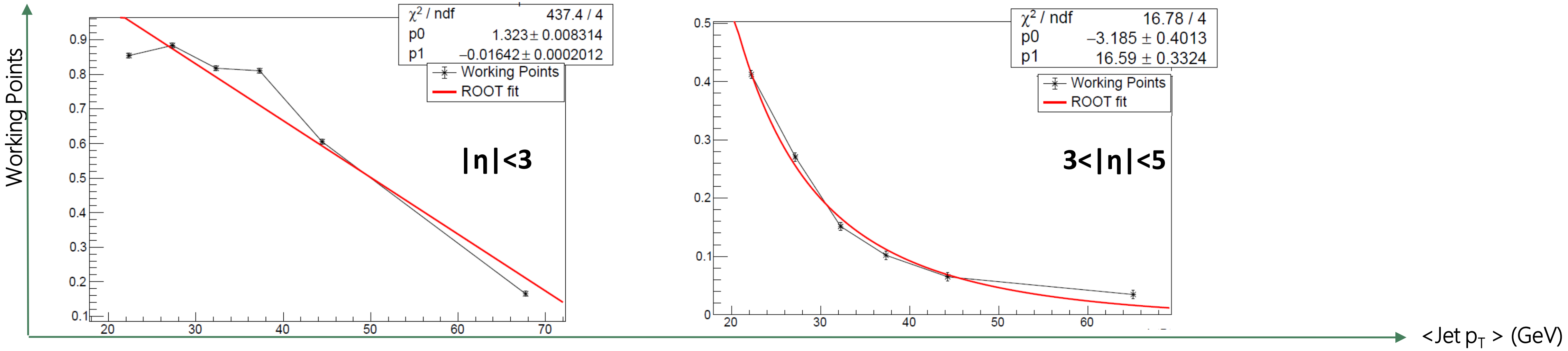
- Apply these weights to rescale the particle's four-momentum
- Particles with small weight or with small rescaled $p_T$ are discarded



# Multivariate discriminant - Performance

Working points in each region as a function of jet $p_T$ in order to maximize quark and gluon efficiency




Efficiency of loose WP as a function of jet $P_T$