# Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment

Speaker: Cheng Chen, Peking University

on behalf of the analysis team

China LHC Physics Workshop 2019

October 25th, Dalian University of Technology

# Introduction

# INTRODUCTION

- Reconstruction and identification of boosted hadronically decaying heavy particles (top/W/Z/H) is an important topic at the LHC
  - provides powerful handles for both searches for new physics and measurements of standard model processes
- Jet substructure
  - high boost leads to highly collimated decay products: can be clustered to form a single large-R jet
  - jets from heavy particles exhibit different radiation patterns, therefore are distinguishable from jets from QCD radiations
  - lots of progresses from both theory and experiment communities
  - most recently: machine learning (ML) techniques bring new improvements and insights in jet substructure

# Introduction

- CMS has been very active in jet substructure studies
  - exploiting jet substructure techniques in analyses
    - searches for heavy resonances / SUSY / etc.
    - measurements of Higgs properties and other SM processes, etc.
  - dedicated measurements of jet substructure observables
  - development of new boosted jet identification algorithms
- JME-18-002: first CMS paper on heavy resonance tagging at 13 TeV
  - comprehensive overview of a large number of heavy resonance tagging algorithms used (and developed) in CMS
  - detailed comparison of the performance in simulated samples
    - for top / W / Z / H tagging
  - validation of the performance in data and evaluation of systematic uncertainties
    - focused on top/W tagging
    - performance of H(Z) -> bb tagging in data to come in BTV-18-001

# Samples and Datasets

- Simulated samples
  - for performance comparison in simulation
    - signals: spin-1 Z'-> tt / WW; spin-2 BulkGraviton -> ZZ / HH [MadGraph + Pythia]
      - for Higgs boson, only H->bb decay considered
    - background: QCD multijet [Pythia]
    - signal samples reweighted to match the $p_T$ spectrum of the background sample
  - for performance studies in data
    - ttbar: POWHEG (NLO) + Pythia [ POWHEG + Herwig++ for systematics ]
    - W/Z/γ+jets: MadGraph (LO) + Pythia
    - VV/ttV/single-top: MadGraph/POWHEG (NLO) + Pythia
    - QCD multijet: MadGraph + Pythia [ Herwig++ for systematics ]
- Datasets: 35.9 $fb^{-1}$ of 2016 data
- Corrections
  - JEC/JER, lepton/b-tagging efficiencies, pileup reweighting, etc.
  - ttbar: top-$p_T$ reweighting
  - QCD multijet / γ+jets: reweight jet $p_T$ to match data

# Overview of the Algorithms

# Summary of the Algorithms

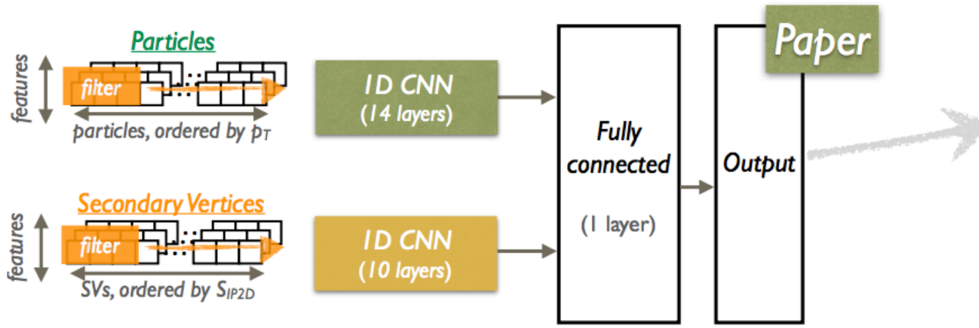
- A variety of jet tagging algorithms studied:
  - "cut-based" algorithms using theory-inspired (high-level) observables
    - groomed mass, N-subjettiness, energy correlation functions (ECF), etc.
  - ML-based algorithms using high-level observables
    - $N_3$-BDT (CA15), BEST
  - ML-based algorithms using low-level observables
    - ImageTop, DeepAK8

Paper

| Algorithm | $p_T$ (jet) [ GeV ] | t quark | W boson | Z boson | Higgs boson | decay modes |
|---|---|---|---|---|---|---|
| $m_{SD} + \tau_{32}$ | 400 | ✓ | | | | |
| $m_{SD} + \tau_{32} + \mathrm{b}$ | 400 | ✓ | | | | |
| $m_{SD} + \tau_{21}$ | 200 | | ✓ | ✓ | | |
| HOTVR | 200 | ✓ | | | | |
| $N_3 - \mathrm{BDT}$ (CA15) | 200 | ✓ | | | | |
| $m_{SD} + N_2$ | 200 | | ✓ | ✓ | ✓ | |
| BEST | 500 | ✓ | ✓ | ✓ | ✓ | |
| ImageTop | 600 | ✓ | | | | |
| DeepAK8 | 200 | ✓ | ✓ | ✓ | ✓ | ✓ |
| Jet mass decorrelated algorithms | | | | | | |
| $m_{SD} + N_2^{DDT}$ | 200 | | ✓ | ✓ | ✓ | |
| double–b | 300 | | | ✓ | ✓ | |
| ImageTop–MD | 600 | ✓ | | | | |
| DeepAK8–MD | 200 | ✓ | ✓ | ✓ | ✓ | ✓ |

# DEEPAK8

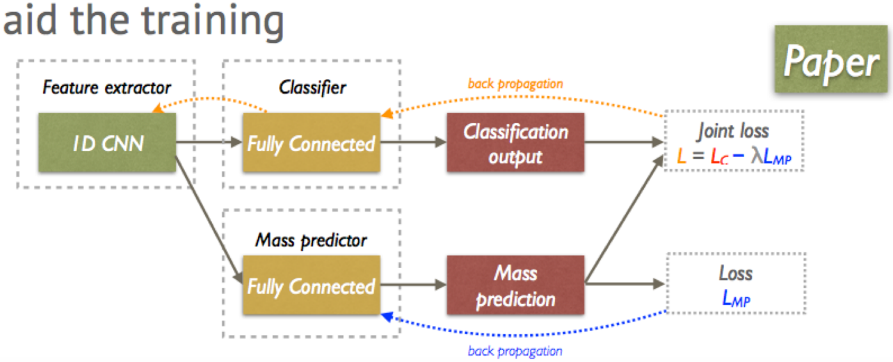
- DeepAK8:
  - multi-class classifier for t/W/Z/H tagging
    - categories subdivided based on decay modes (e.g., $Z\to bb$, $Z\to cc$, $Z\to qq$)
  - directly uses jet constituents (PF candidates / secondary vertices)
  - 1D CNN based on the ResNet [arXiv: 1512.03385] architecture




Output

| Category | Label |
|---|---|
| Higgs | H (bb) |
| | H (cc) |
| | H (VV*→qqqq) |
| Top | top (bcq) |
| | top (bqq) |
| | top (bc) |
| | top (bq) |
| W | W (cq) |
| | W (qq) |
| Z | Z (bb) |
| | Z (cc) |
| | Z (qq) |
| QCD | QCD (bb) |
| | QCD (cc) |
| | QCD (b) |
| | QCD (c) |
| | QCD (others) |

- DeepAK8-MD
  - mass-decorrelated version using adversarial training techniques
  - signal and background samples reweighted to yield flat distributions in both $p_T$ and $m_{SD}$ to aid the training

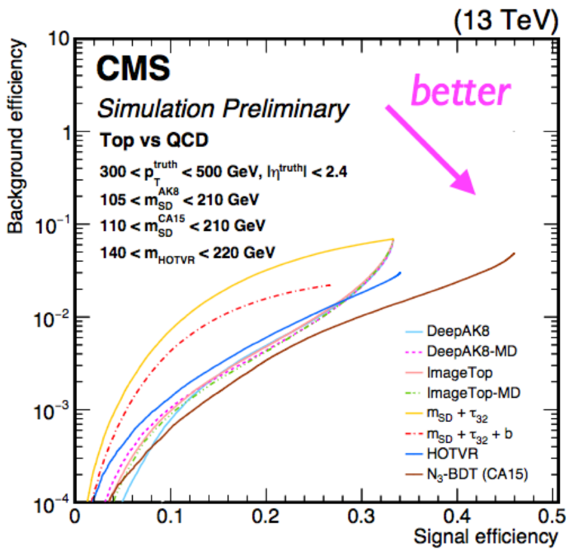


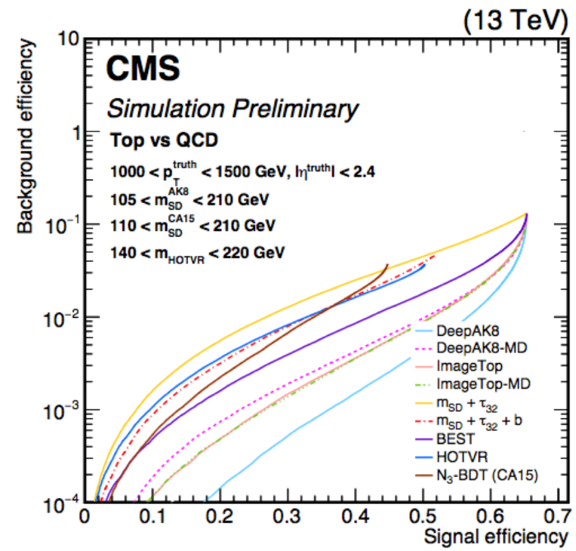
# Performance in Simulation

# ROC (Top/W)

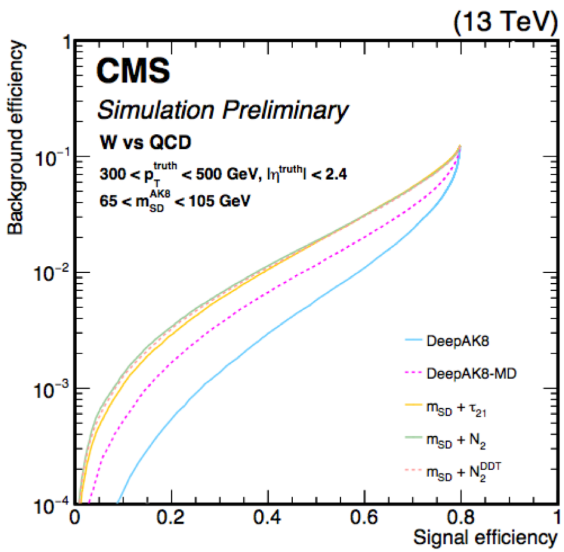
Paper

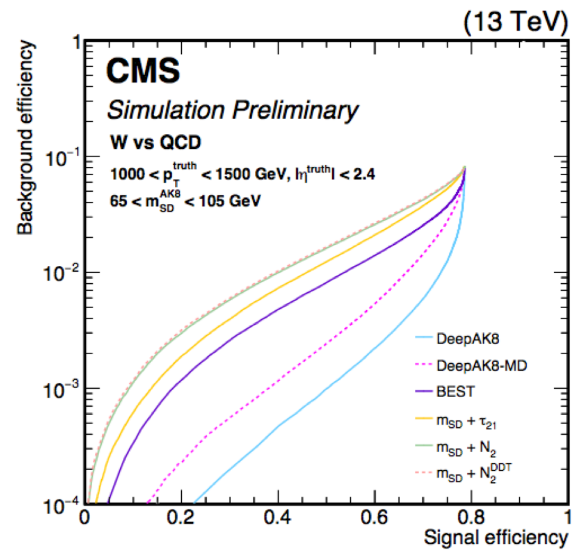
*top vs QCD*




*W vs QCD*



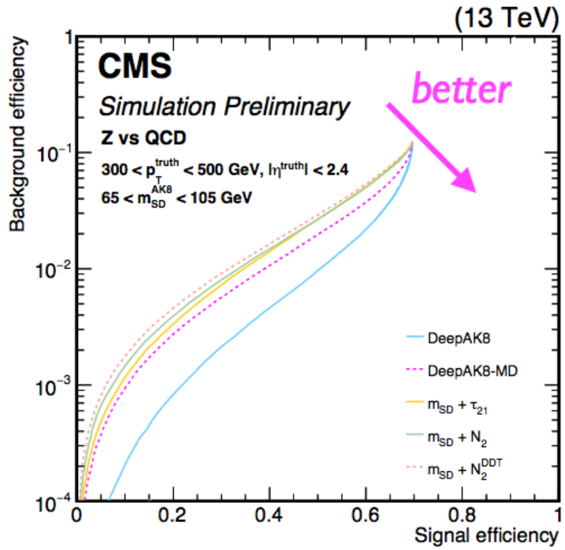



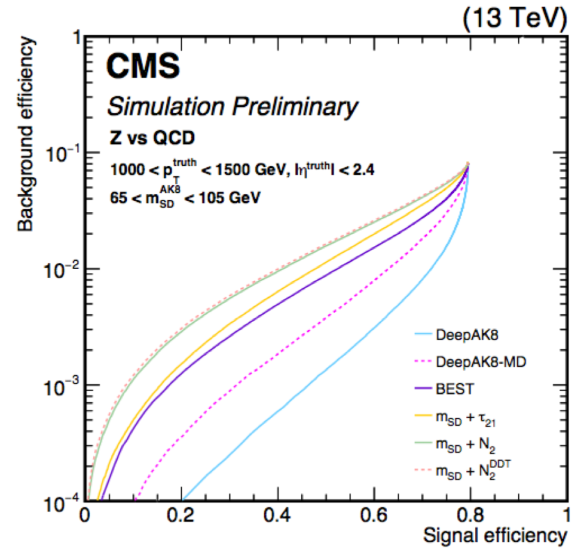

*Significant improvement from the new developments*

# ROC (Z/H)

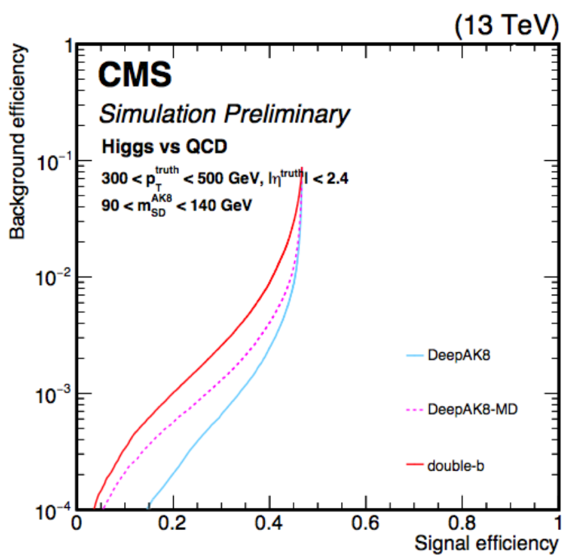
Paper

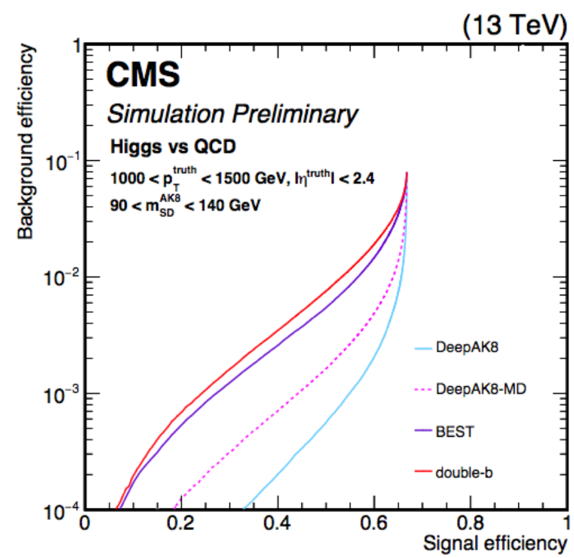
*Z vs QCD*




*H vs QCD*



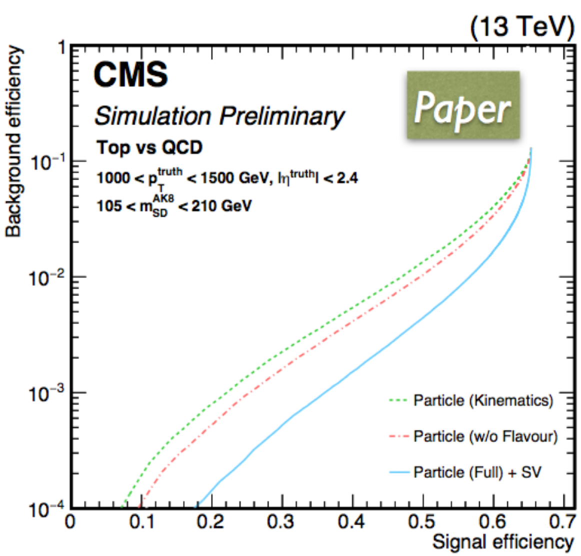



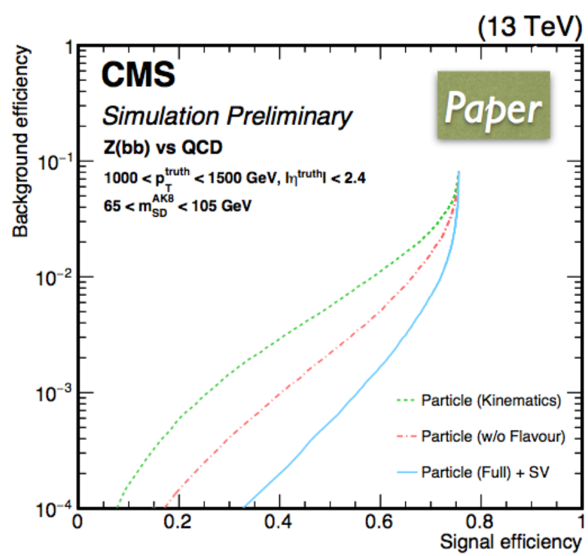

*Significant improvement from the new developments*

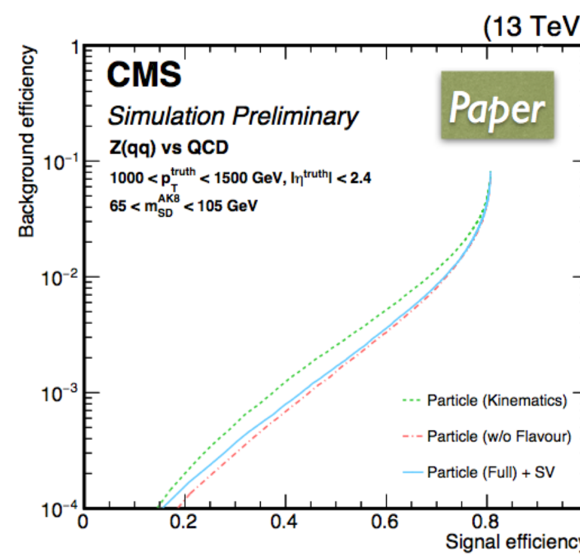
# Ablation Study of DeepAK8

- DeepAK8 shows substantial gain compared to traditional approaches
- To understand the main sources of the improvement, alternative versions of DeepAK8 were trained using a subset of the input features
  - Particle (kinematics): only kinematic info of PF candidates
    - four momenta, distances to the jet and subjet axes, etc.
  - Particle (w/o Flavour): adding experimental info
    - charge, particle identification, track quality, etc.
  - Particle Full + SV (the full DeepAK8): adding features related to heavy-flavour tagging
    - track displacement, track-vertex association, SV features, etc.

Update



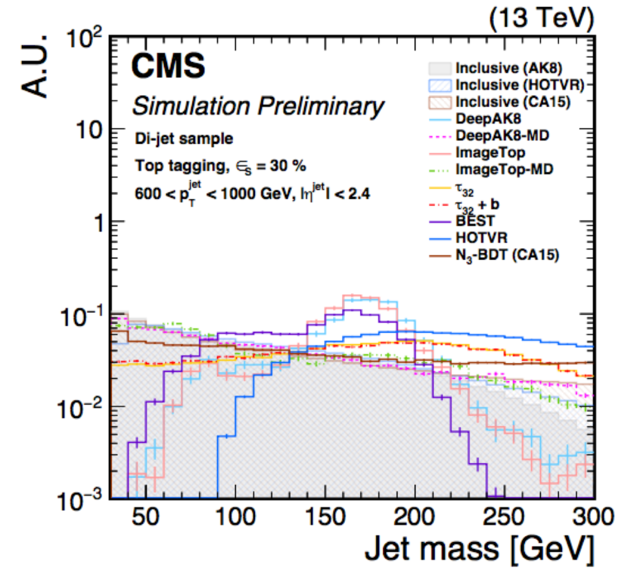



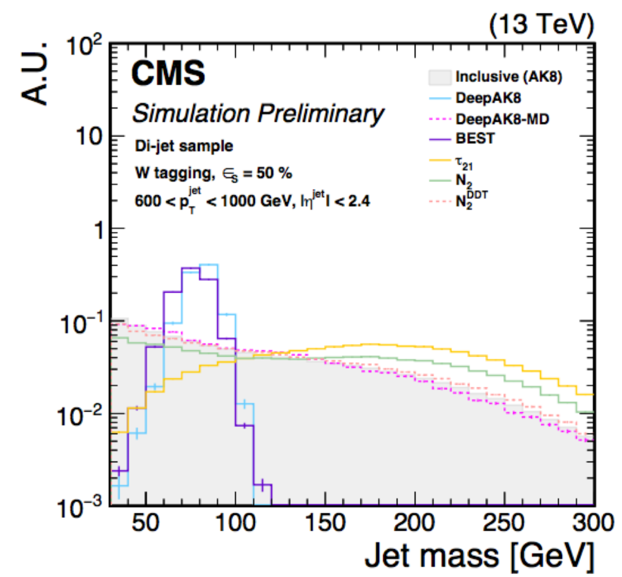



# Correlation with Jet Mass

Paper

- Jet substructure variables / ML taggers typically correlated with jet mass
  - selection on substructure variables / ML taggers changes the mass shapes: "mass sculpting"
  - whether mass sculpting is a problem depends on the analyses...
- A number of mass decorrelation techniques explored in CMS
  - $N_2^{DDT}$:
    - transform $N_2$ such that a selection on $N_2^{DDT}$ yields a constant eff(B) across $p_T$ and mass
  - ImageTop-MD:
    - reweight QCD sample to match the mass distribution of top sample in the training
  - DeepAK8-MD:
    - exploit adversarial training to reduce the mass dependency
    - reweight all signal and background samples to have a flat mass distribution to aid the training



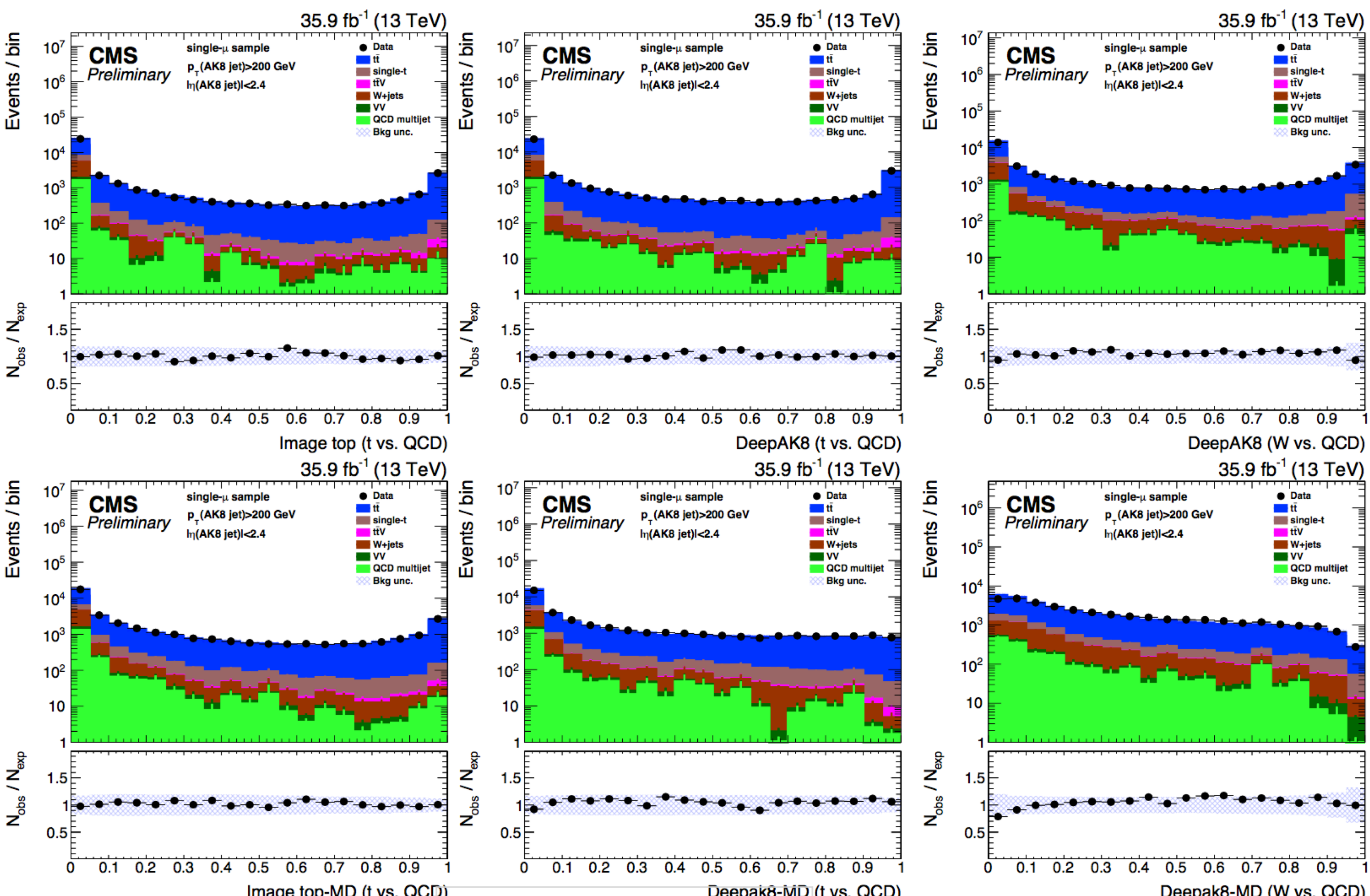
# Performance in Data and systematic uncertainties

# Event Selection

- Three data samples used for the validation of top/W tagging performance
  - signal sample
    - single-μ sample: ttbar(1L) dominated
    - 2 regions: $p_T$(jet) > 200 GeV [W-enhanced] and $p_T$(jet) > 500 GeV [top-enhanced]
    - used for extracting simulation-to-data corrections of top/W signal efficiency
  - background samples
    - di-jet sample (HT>1000 GeV)
    - single-photon sample ($p_T(\gamma)$>200 GeV)
    - different quark-gluon fractions
- Systematics
  - parton showering [up to 50%], PDF & scale [~5-15%]
  - JES/JER [~%], MET [~%], PU[<5%], lumi [2.5%]
  - MC stats
  - b-tag, lep & trigger eff. very small -> neglected

# SINGLE-MUON SAMPLE: DEEPAK8/IMAGETOP

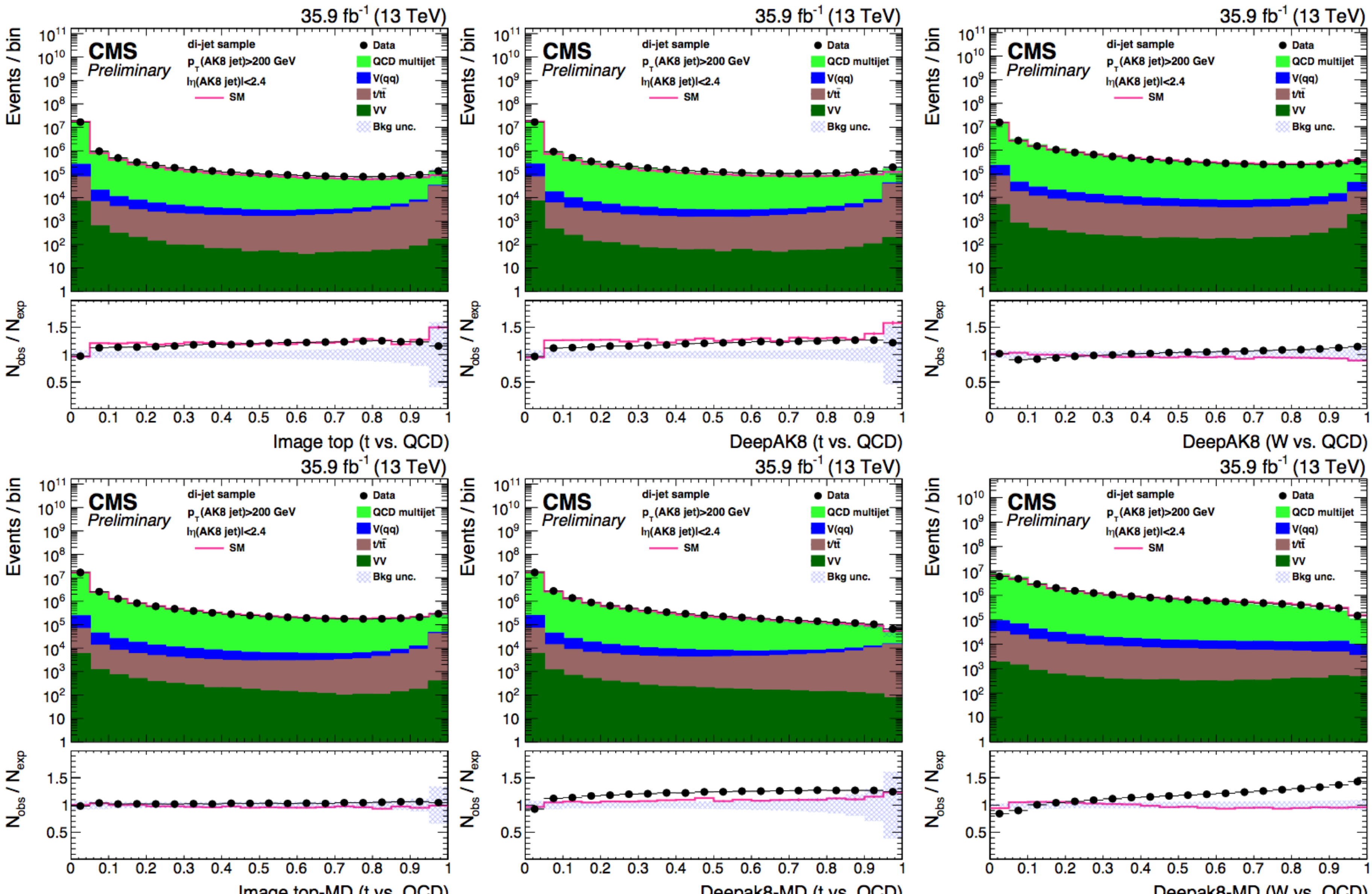
Paper




*Good data/MC agreement*

# Di-Jet Sample: DeepAK8/ImageTop

Update Paper

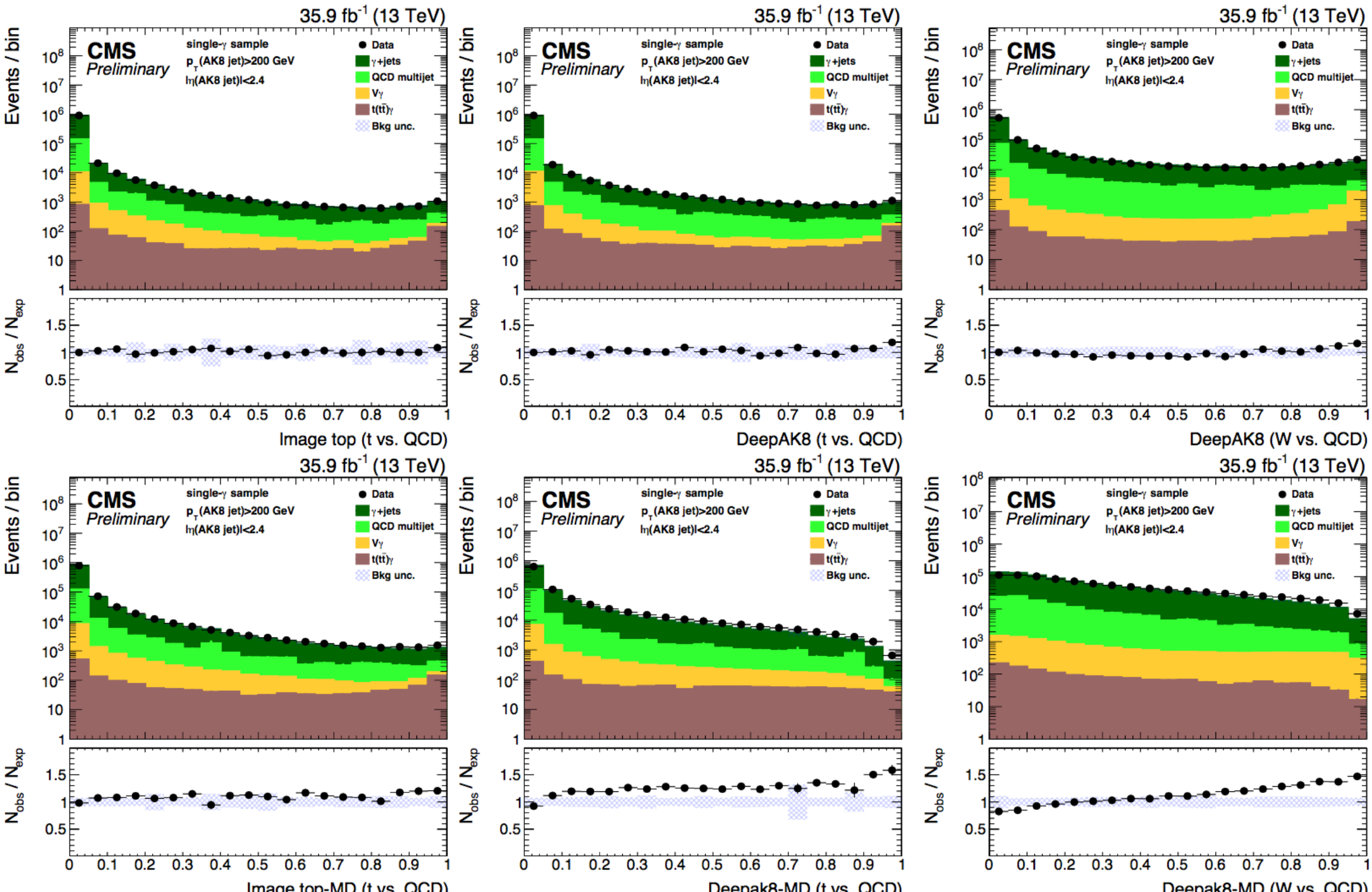



*Jet $p_T$ reweighted to match data*

Update

Paper

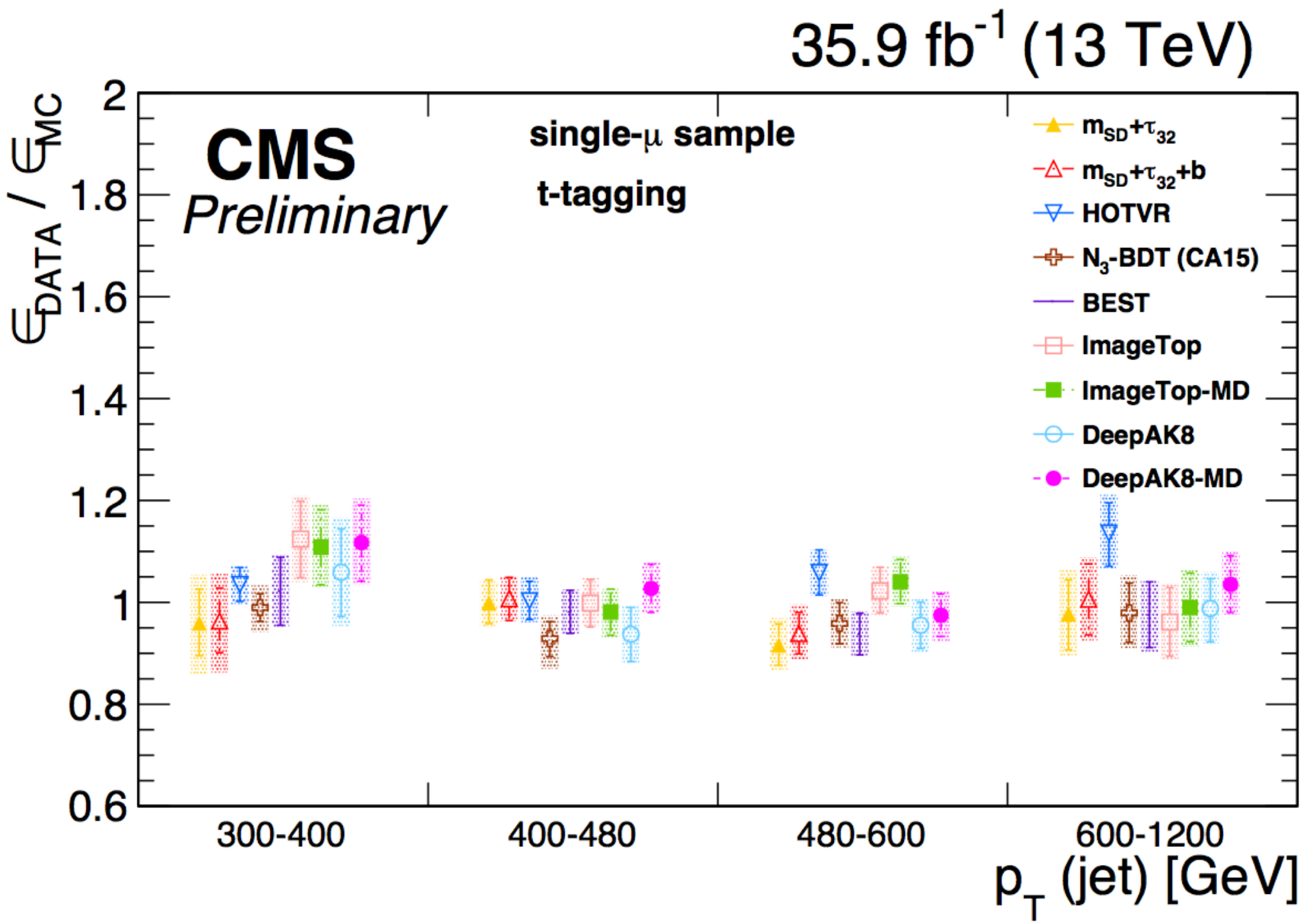
# Single-Photon Sample: DeepAK8/ImageTop




*Jet $p_T$ reweighted to match data*

# Top-tagging Scale Factors

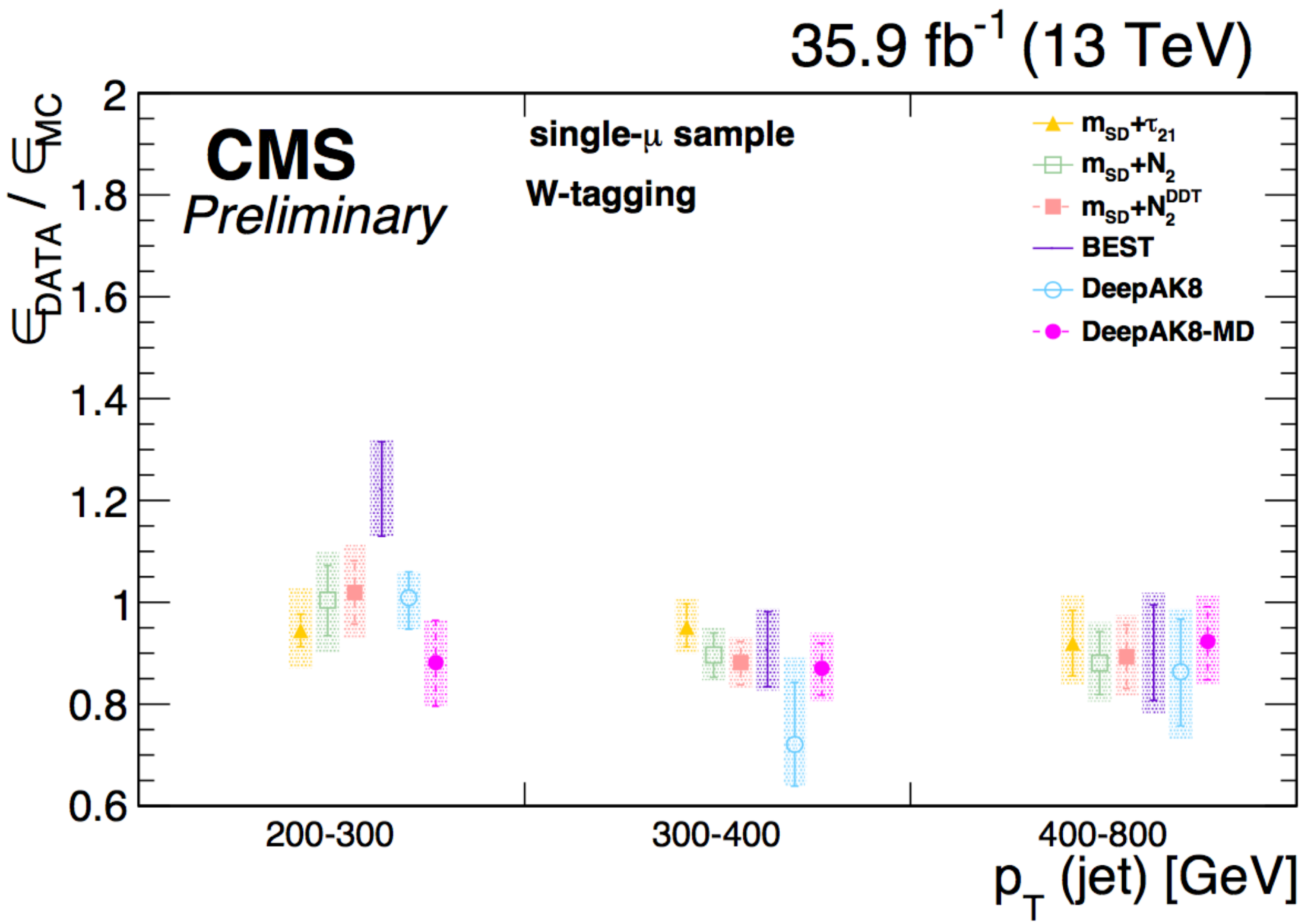
Paper




*Typically consistent with 1 within 10-20%*

# W-tagging Scale Factors

Paper




*Typically consistent with 1 within 10-20%*

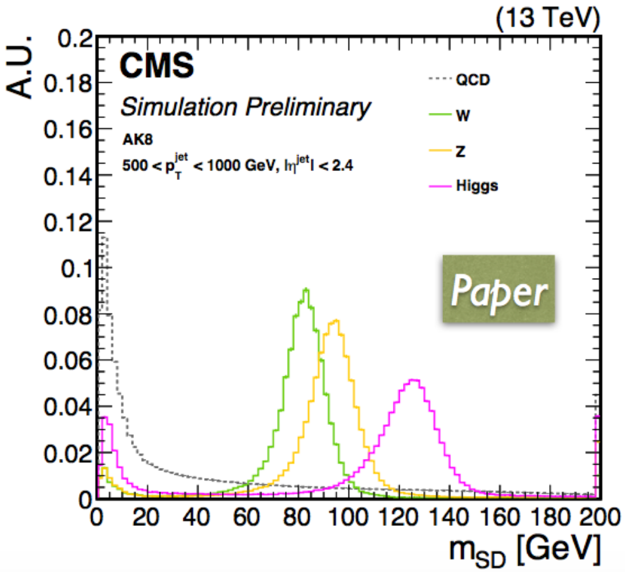
# SUMMARY

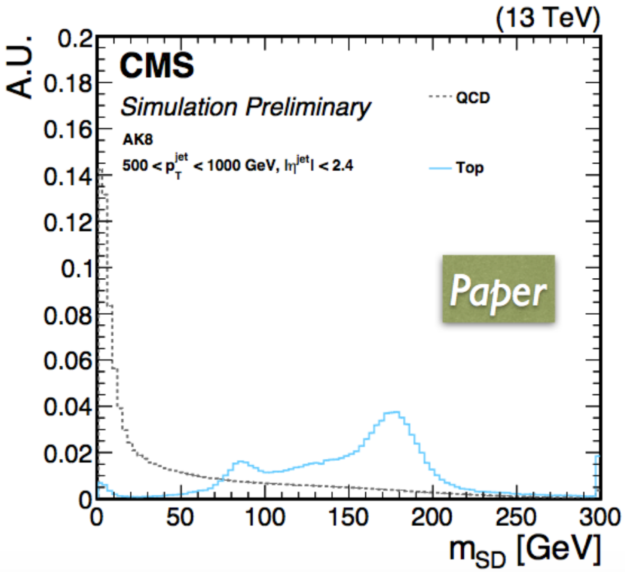
- Lots of effort in CMS to improve existing methods and develop new approaches for boosted jet tagging
  - "cut-based" algorithms using theory-inspired observables
  - ML-based algorithms using high-level / low-level observables
  - new development brings substantial gain in performance
- Heavy resonance tagging paper
  - comprehensive overview of a large number of heavy resonance tagging algorithms used / developed in CMS
  - detailed evaluation of the performance in simulation and data in a fully coherent way
  - reference for analyzers to decide the best approach for their use cases

# Backups

# CUT-BASED ALGORITHM: $M_{SD}$

- Cut-based algorithms
  - based on theory-inspired variables; extensively studied from both theory and experiment side
  - robust and easily interpretable
  - baseline for comparison with new algorithms
- Jet grooming with the soft drop (SD) algorithm
  - SD condition: $\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$ ( $\beta$=0, $z_{cut}$=0.1 used in CMS )
    - removes soft and wide-angle radiations
  - groomed mass ($m_{SD}$) computed from the two subjets returned by the SD algorithm
  - strongly reduces the "Sudakov" peak structure in the jet mass distribution



# CUT-BASED ALGORITHM: N-SUBJETTINESS

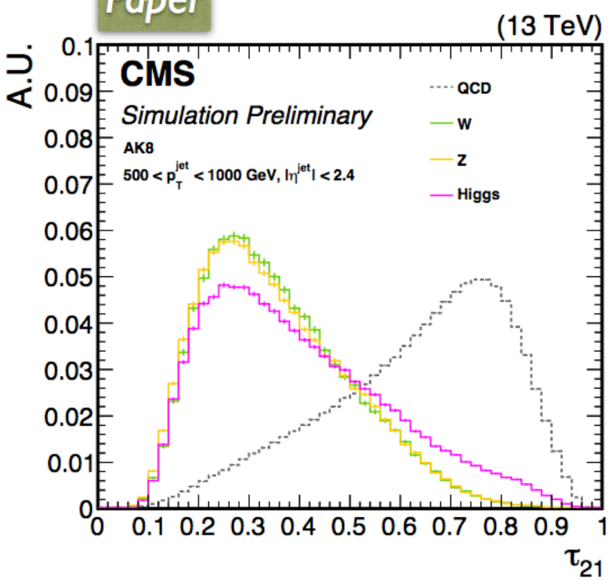
- N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_i p_{\mathrm{T},i} \min \left[ \Delta R_{1,i}, \Delta R_{2,i}, \ldots, \Delta R_{N,i} \right]$$

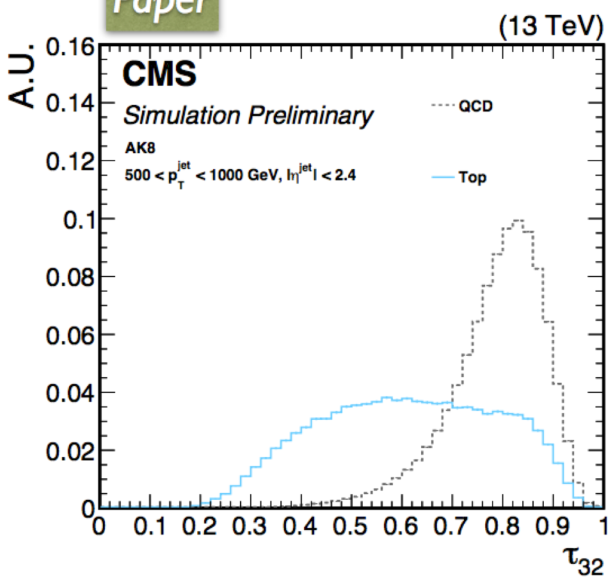
*ΔR to the N subjet axes*
*(subjets found w/ exclusive $k_T$ algo)*

- quantifies the compatibility of a jet with having N subjets
- more discriminating power with the ratios
  - $\tau_{21} := \tau_2/\tau_1$ : 2-prong (W/Z/H) tagging
  - $\tau_{32} := \tau_3/\tau_2$ : 3-prong (top) tagging
    - subjet b-tagging can be added to further improve the performance

Paper




Paper



# ECF: $N_2$

- Generalized energy correlation functions (ECF)

$$ {}_o e_N^{\beta} = \sum_{1 \le i_1 < i_2 < \cdots < i_N \le N_C} \left[ \prod_{1 \le k \le N} \frac{p_T^{i_k}}{p_T^J} \right] \prod_{m=1}^{o} \min_{i_j < i_k \in \{i_1, i_2, \cdots, i_N\}}^{(m)} \left\{ \Delta R_{i_j, i_k}^{\beta} \right\} $$

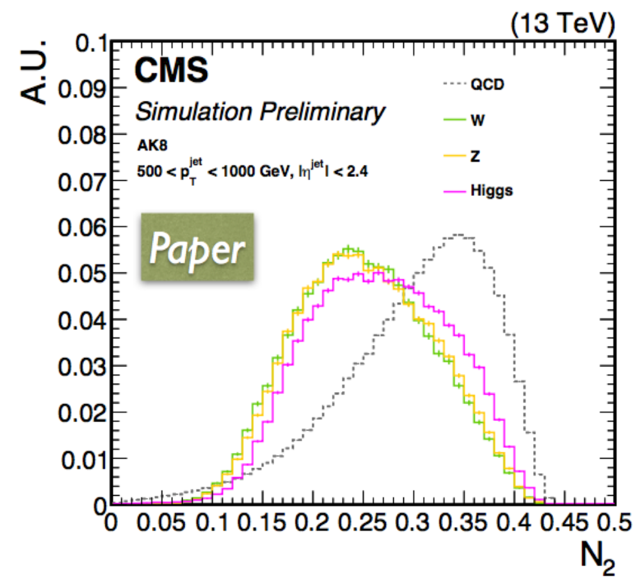
  - tests the compatibility with having N radiation centers
    - similar to N-subjettiness, but w/ an axis-free approach
    - for a N-prong jet, $e_N >> e_M$ for M > N
- For 2-prong tagging (W/Z/H): ECF ratio $N_2$

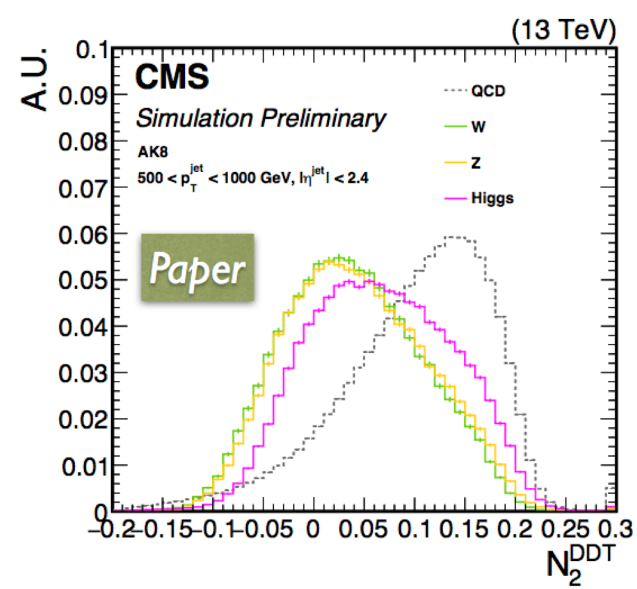
$$ N_2^1 = \frac{{}_2 e_3^1}{({}_1 e_2^1)^2} $$

- Mass decorrelated version: $N_2^{DDT}$
  - "designed decorrelated tagger" approach:

$$ N_2^{\mathrm{DDT}}(\rho, p_T) = N_2(\rho, p_T) - N_2^{(X\%)}(\rho, p_T) $$

    - $\rho = \ln(m_{SD}^2 / p_T^2)$ is a dimensionless scaling variable
    - $N_2^{\mathrm{DDT}} < 0$ yields a constant QCD background efficiency of X% across the mass and $p_T$ range with no loss of performance
    - X=5 used in this paper



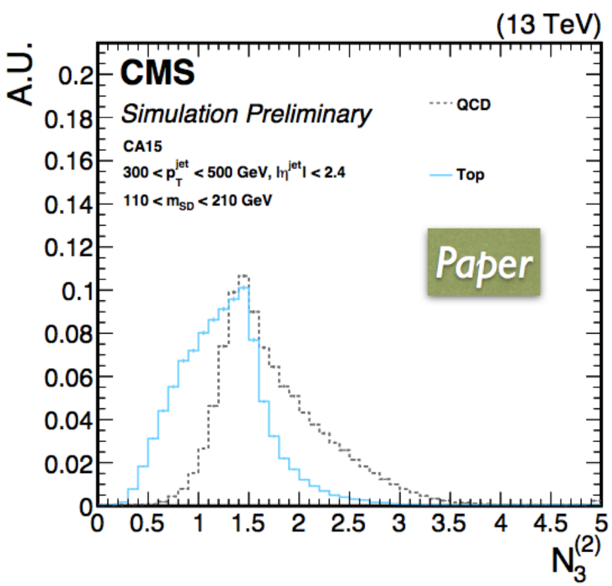



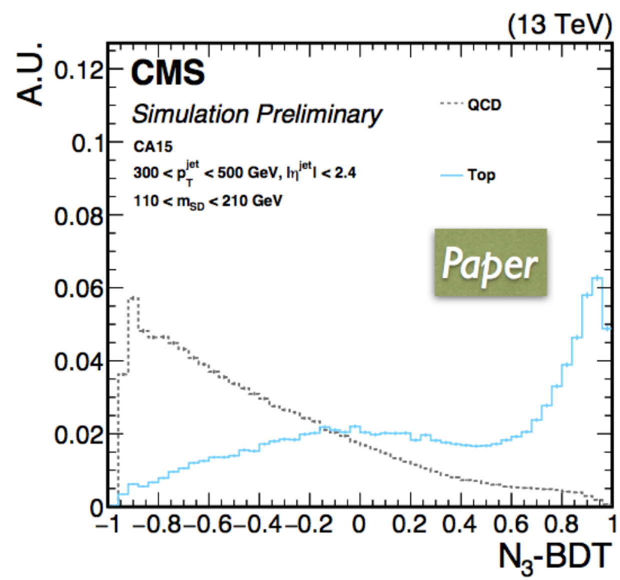
# ECF: $N_3$-BDT (CA15)

- ECFs also exploited for top-tagging in CMS
  - using CA15 jets to cover lower $p_T$ regions ($p_T$>200 GeV)
  - boosted decision trees (BDT) trained with a set of 11 ECF ratios + $\tau_{32}$ + HTTv2 $f_{rec}$

$$\frac{{}_1e_2^{(2)}}{\left({}_1e_2^{(1)}\right)^2}, \frac{{}_1e_3^{(4)}}{{}_2e_3^{(2)}}, \frac{{}_3e_3^{(1)}}{\left({}_1e_3^{(4)}\right)^{3/4}}, \frac{{}_3e_3^{(1)}}{\left({}_2e_3^{(2)}\right)^{3/4}}, \frac{{}_3e_3^{(2)}}{\left({}_3e_3^{(4)}\right)^{1/2}},$$

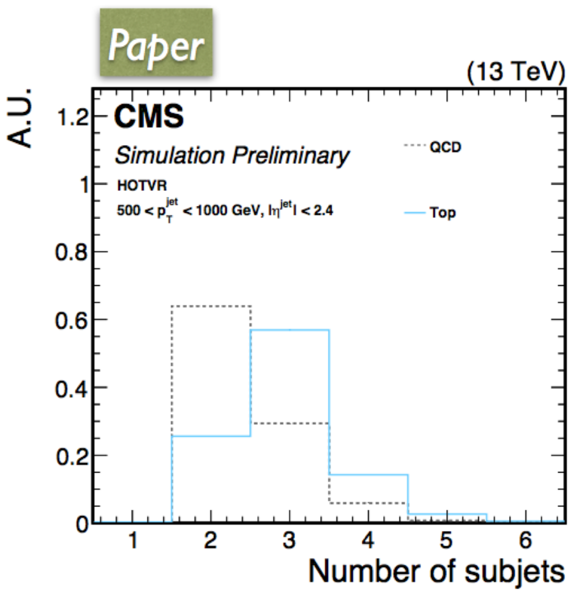
$$\frac{{}_1e_4^{(4)}}{\left({}_1e_3^{(2)}\right)^2}, \frac{{}_1e_4^{(2)}}{\left({}_1e_3^{(1)}\right)^2}, \frac{{}_2e_4^{(1/2)}}{\left({}_1e_3^{(1/2)}\right)^2}, \frac{{}_2e_4^{(1)}}{\left({}_1e_3^{(1)}\right)^2}, \frac{{}_2e_4^{(1)}}{\left({}_2e_3^{(1/2)}\right)^2}, \frac{{}_2e_4^{(2)}}{\left({}_1e_3^{(2)}\right)^2}$$

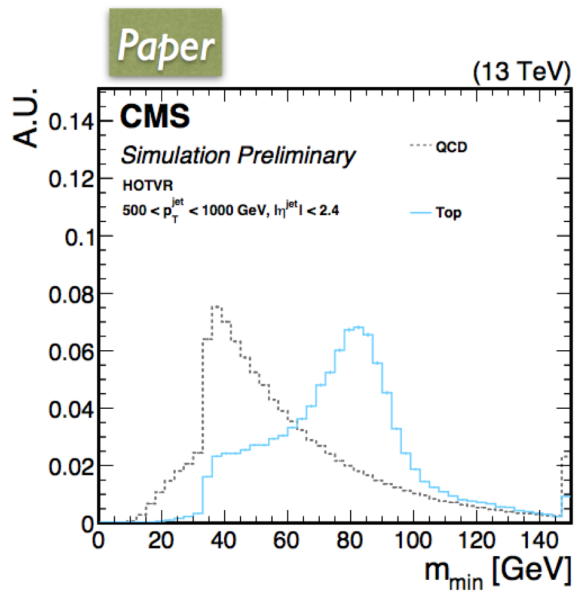


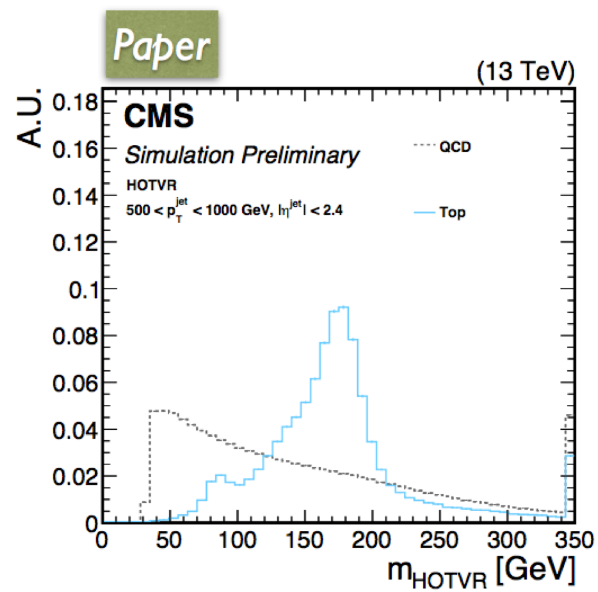



# HOTVR

- Heavy Object Tagger with Variable R (HOTVR): jet algo + heavy resonance tagger
  - jet clustering with ($p_T$-dependent) variable distance parameter R
    - Puppi corrected PF candidates used in the CMS implementation
  - soft clusters removed during clustering
    - stable jet mass distribution; prevent additional radiation into jets
  - can be used for tagging different heavy resonances (t/W/Z/H)
    - only top tagging studied in this paper

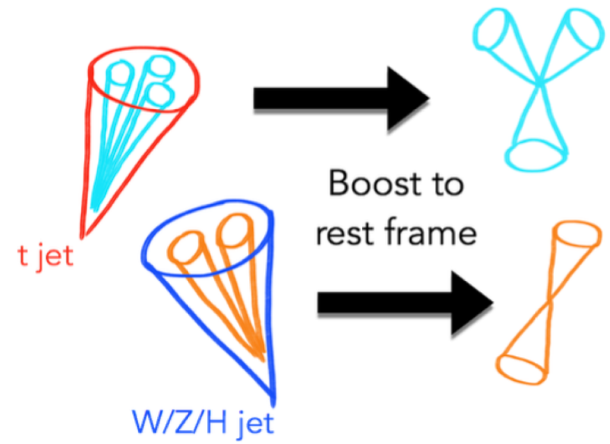


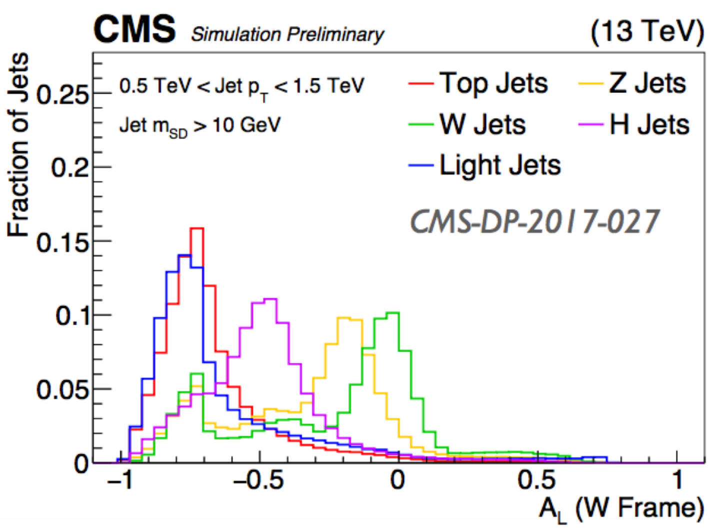



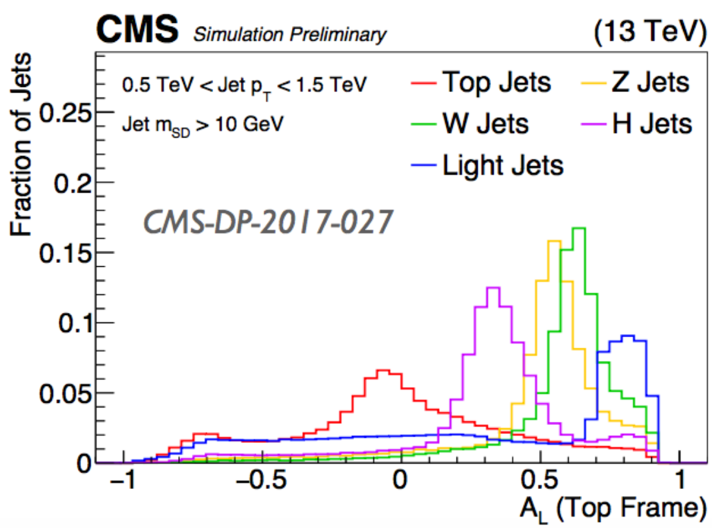



# Boosted Event Shape Tagger

- Boosted Event Shape Tagger (BEST)
  - multi-class tagger for t/W/Z/H
  - compute jet kinematic/shape variables in four reference frames: corresponding to the rest frames if the jet originates from t/W/Z/H
    - when boosting to the “correct” rest frame, jet constituents should be isotropic and show the expected N-prong structure
  - neural network (NN) trained with these kinematic variables as well as subjet b-tagging discriminants
    - 3 fully-connected layers w/ 40 nodes each

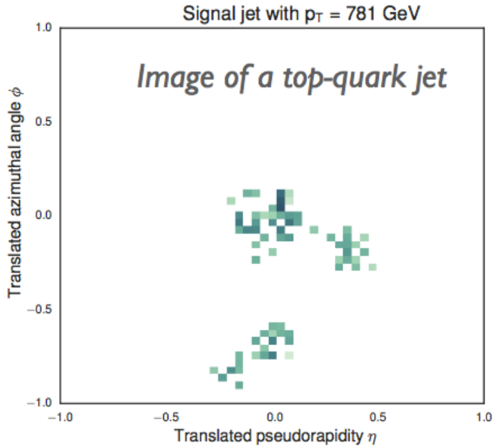


# Deep Learning Approaches

- New approaches based on deep learning have been proposed in recent years and attracted lots of attention
  - low-level inputs + deep neural networks
- Two types of deep learning approaches for jet tagging
  - image-based
    - convert jet to an image using energy depositions on the calorimeters
    - exploiting computer vision techniques – typically 2D convolutional neural networks (CNN)
    - image sparsity, heterogeneous detector can be a challenge
  - particle-based
    - treat jet as a collection of its constituent particles
    - recurrent NN, 1D CNN, graph NN, etc., can be exploited
    - more natural idea due to particle-flow reconstruction at CMS
      - incorporate information from all sub-detectors (e.g., tracker) and exploit the full granularity
- Both approaches have been explored in CMS



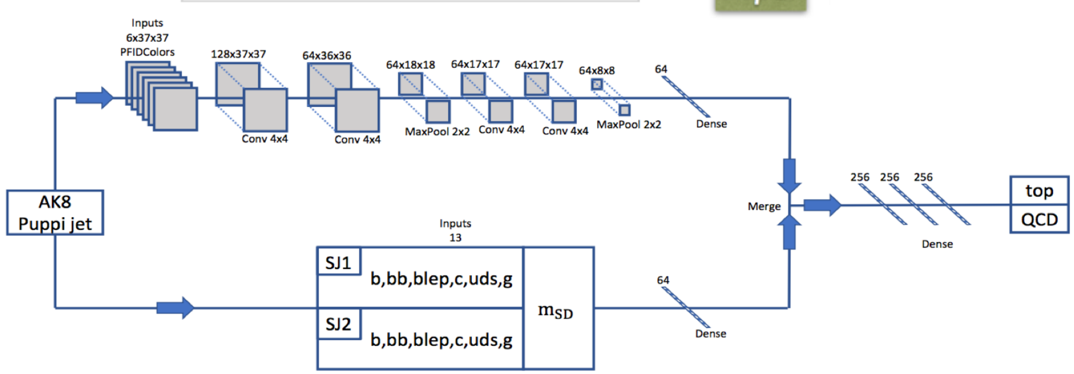

arXiv:1704.02124

arXiv: 1702.00748, arXiv:1711.02633

# IMAGETOP

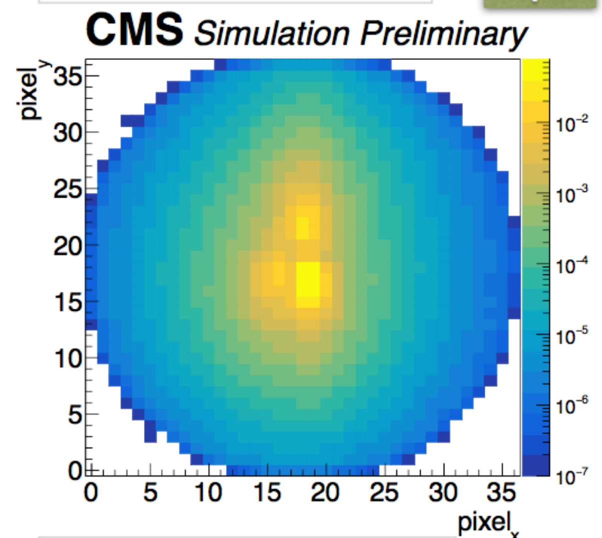
- ImageTop
  - top tagging algorithm based on jet images
  - jet image adaptively zoomed based on jet $p_T$ to account for the increased collimation at high $p_T$
  - four “color” channels for the jet image:
    - neutral $p_T$, track $p_T$, N(muon), N(track)
  - DeepJet b-tagging discriminants of the subjets also used

*ImageTop network architecture* Paper

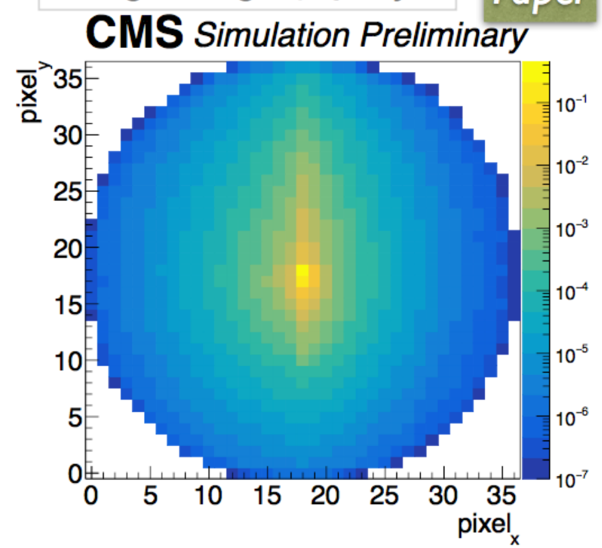



- ImageTop-MD
  - mass decorrelated version by reweighting the QCD sample to match the mass shape of the top sample in the training

*Averaged image of top jets* Paper

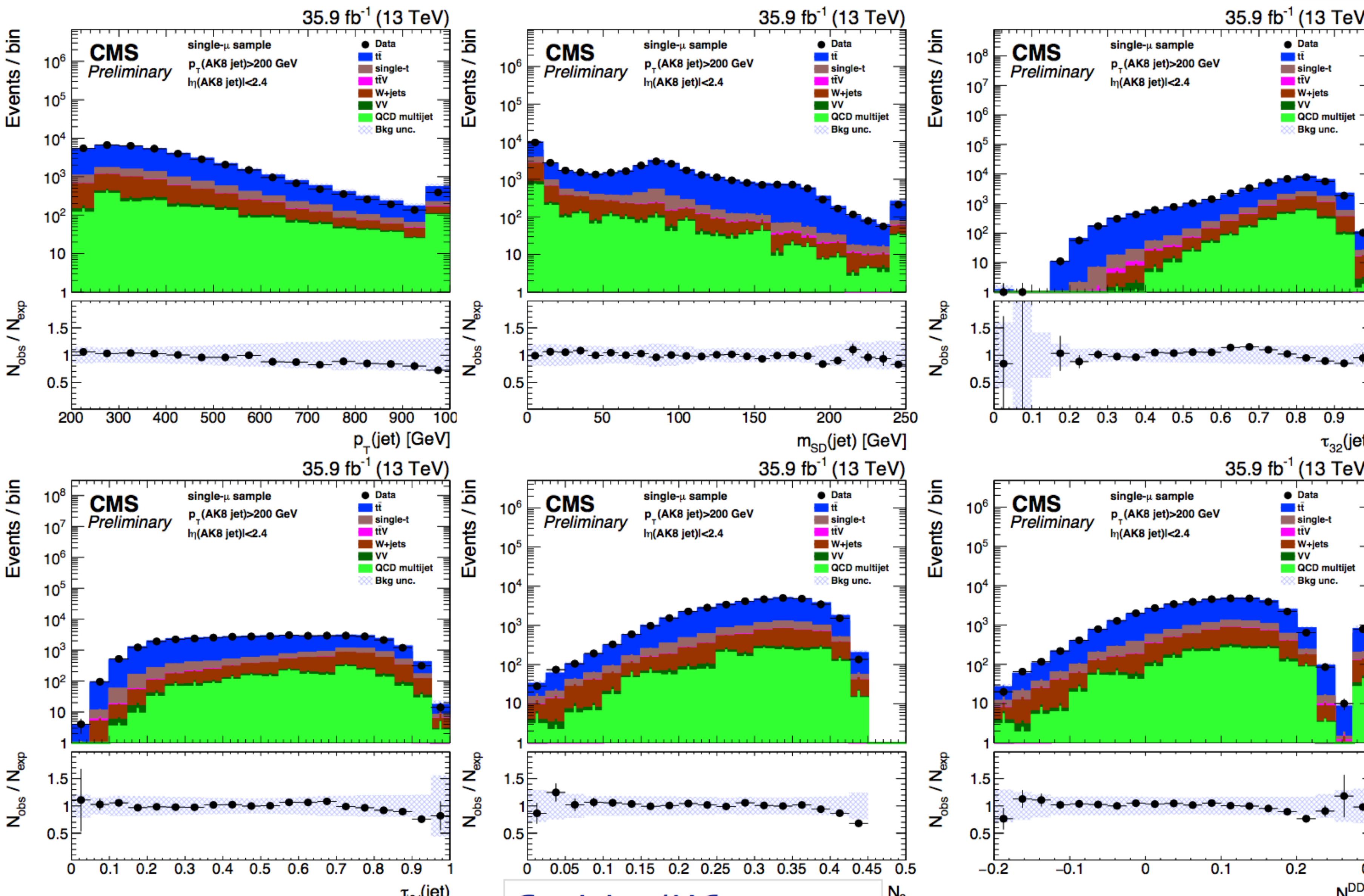



*Averaged image of QCD jets* Paper



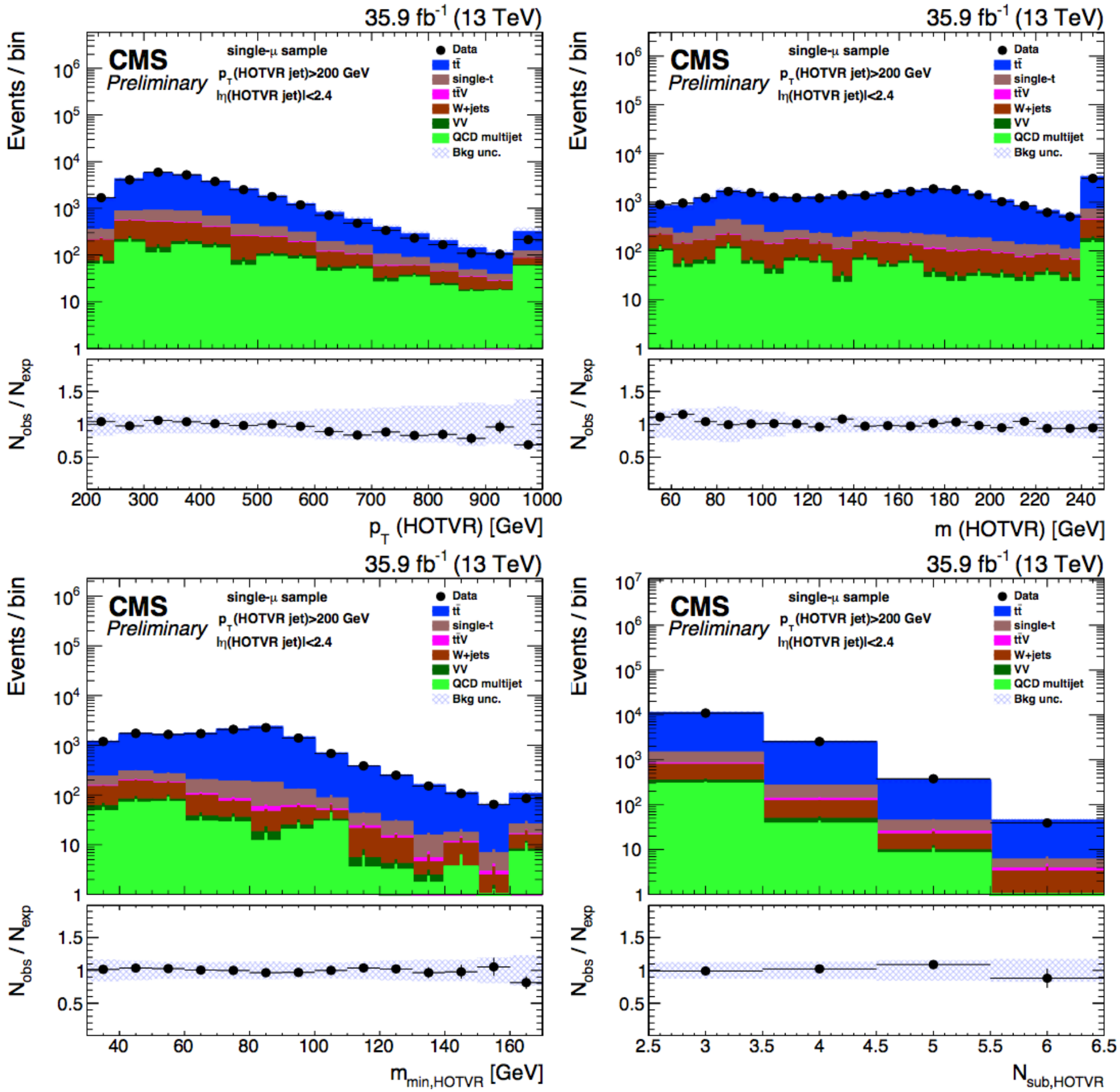
# Single-Muon Sample: Basic Vars

Paper




Good data/MC agreement

# Single-Muon Sample: HOTVR

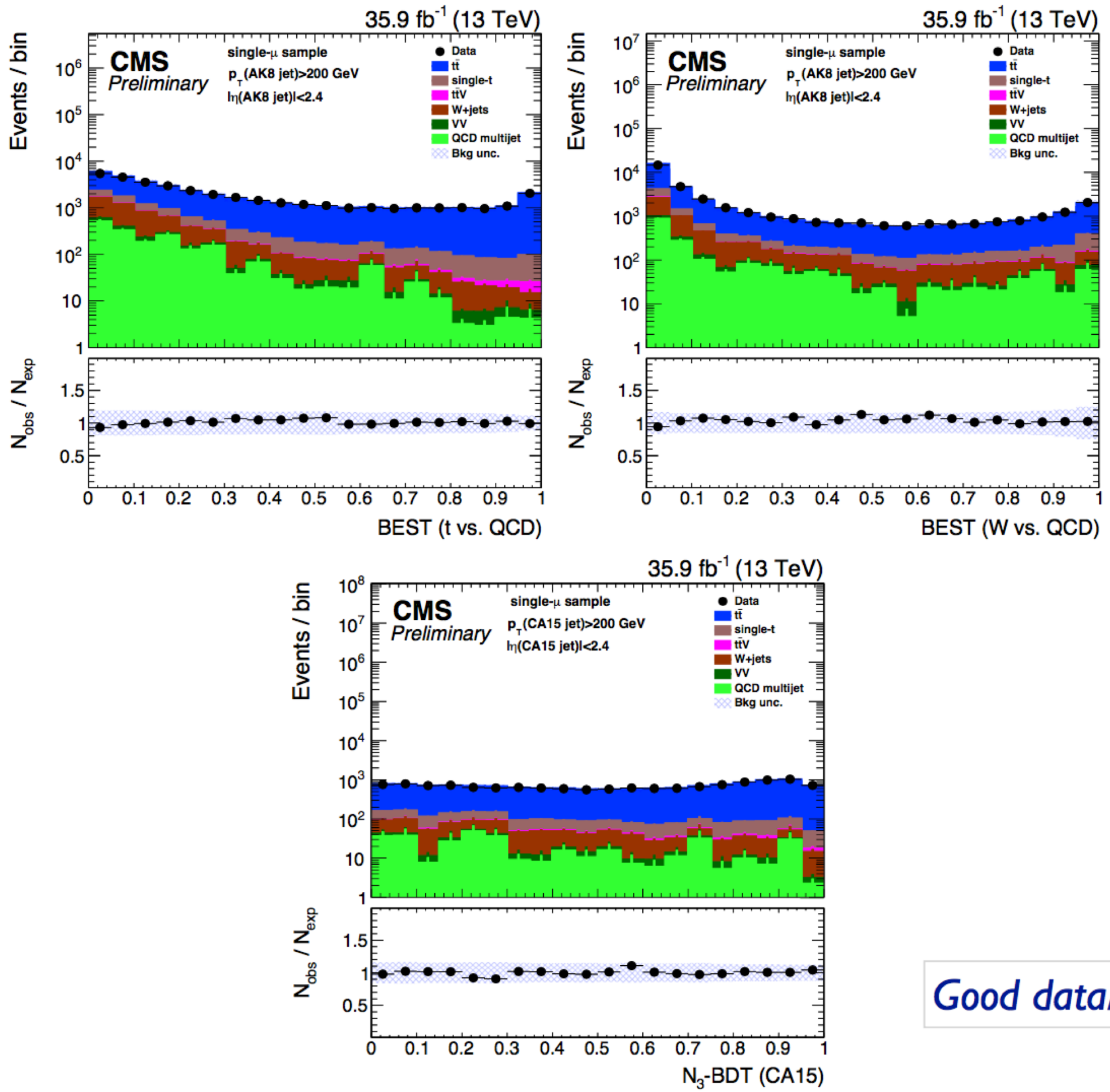



Paper

Good data/MC agreement

# Single-Muon Sample: BEST/N3-BDT

Paper

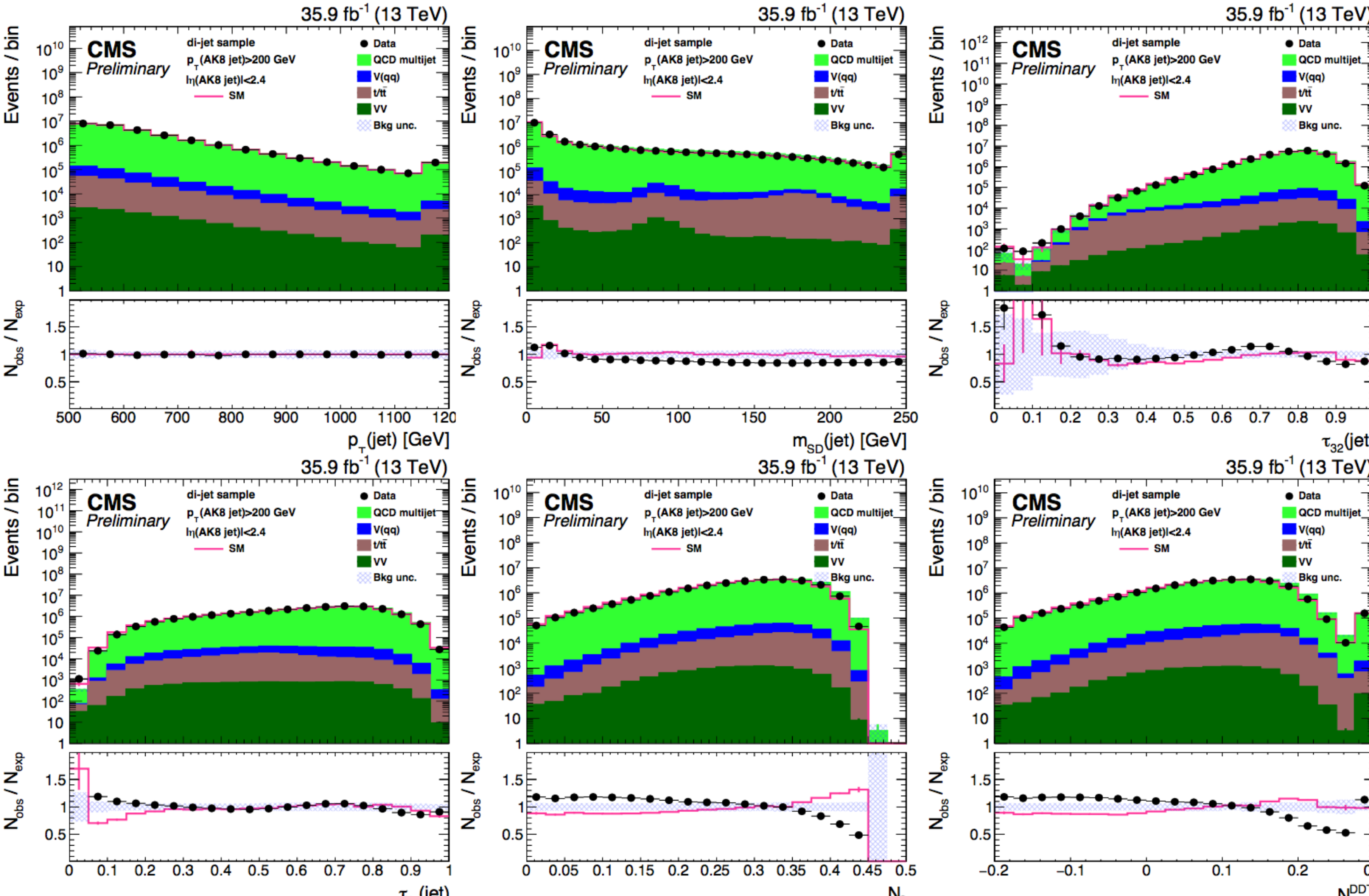



Good data/MC agreement

# DI-JET SAMPLE: BASIC VARS

Update

Paper

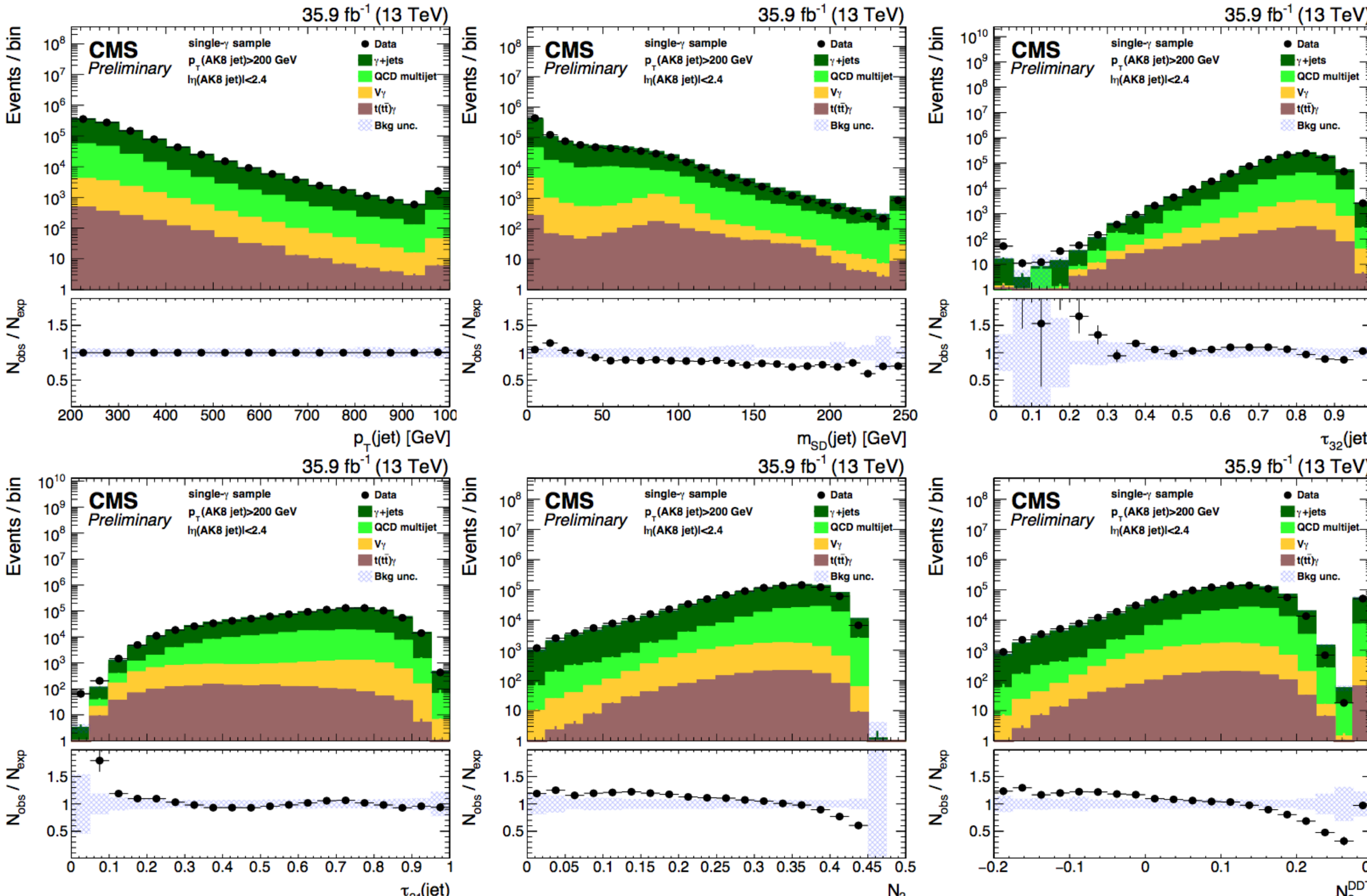



*Jet $p_T$ reweighted to match data*

# Single-Photon Sample: Basic Vars

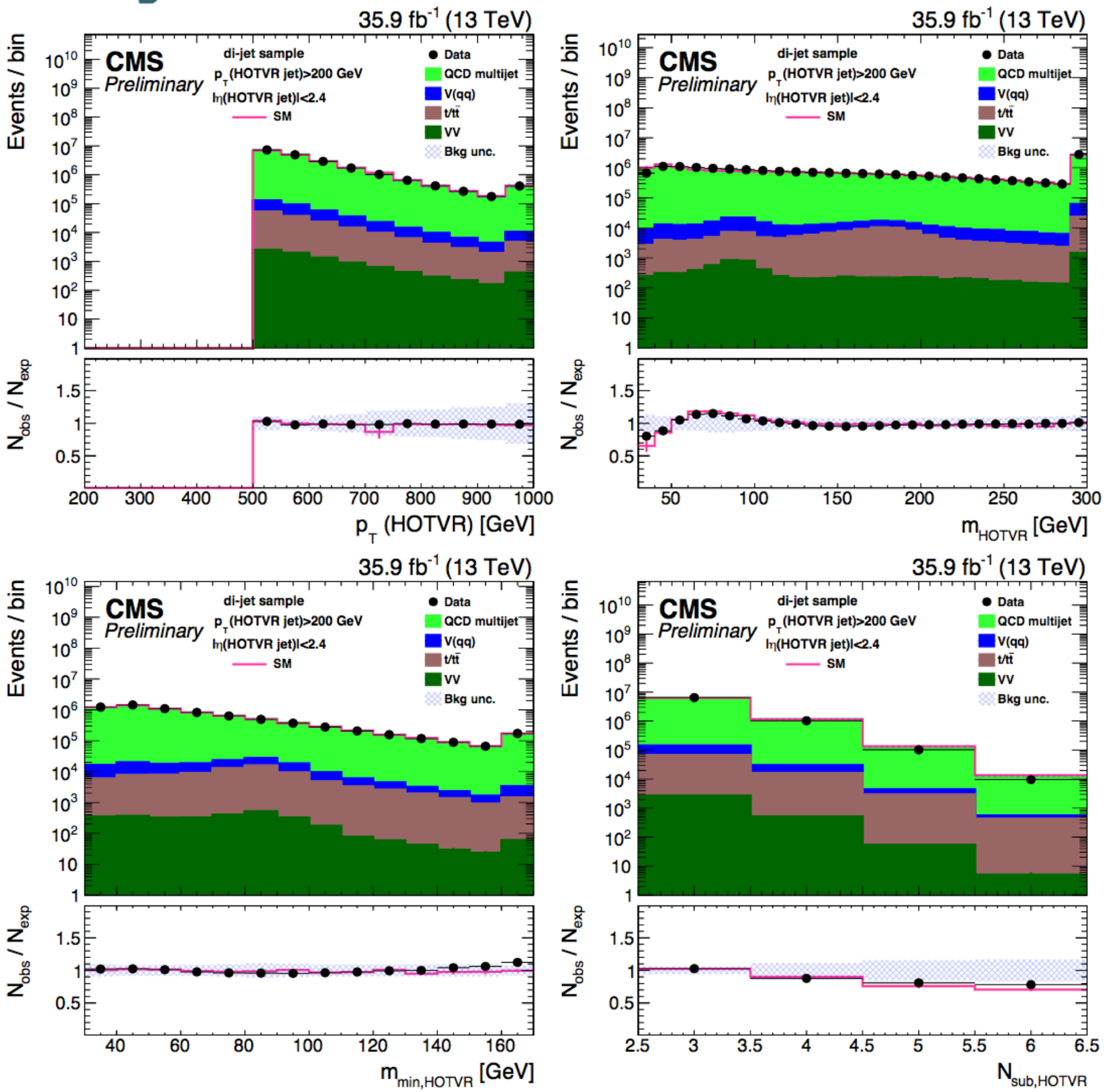
Update

Paper




*Jet $p_T$ reweighted to match data*

# Di-Jet Sample: HOTVR

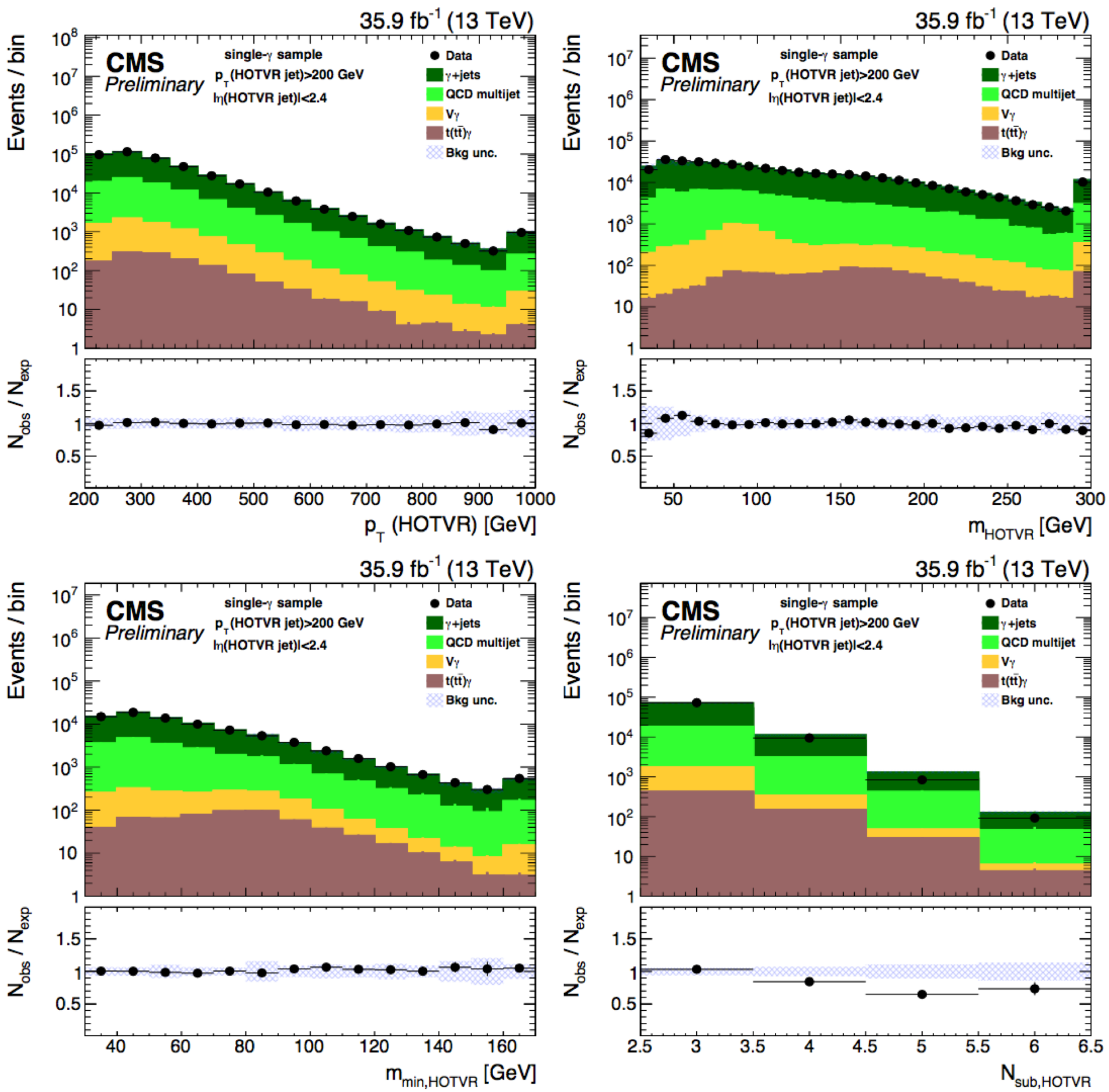



Update

Paper

*Jet $p_T$ reweighted to match data*

# Single-Photon Sample: HOTVR




Update

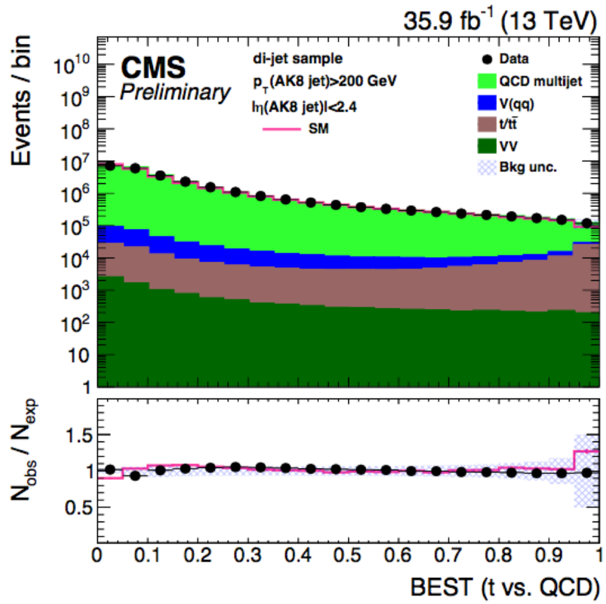
Paper

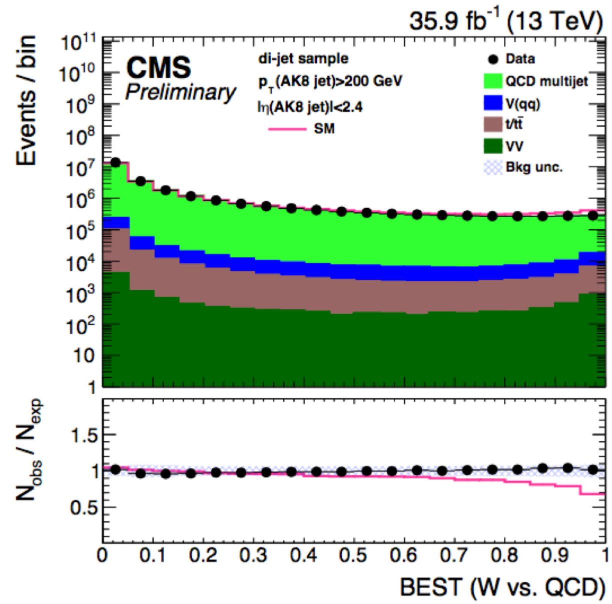
*Jet $p_T$ reweighted to match data*

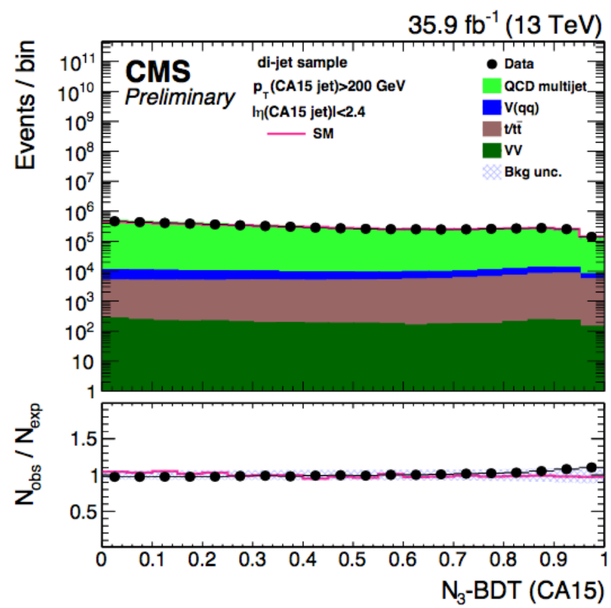
# Di-Jet Sample: BEST/N3-BDT

Update

Paper



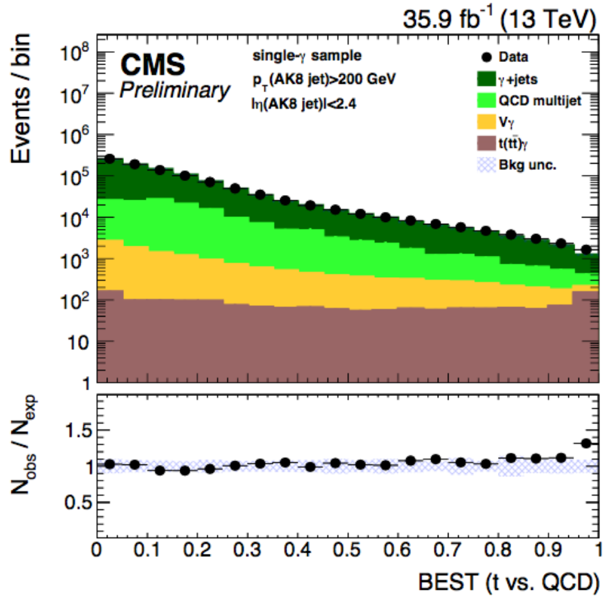



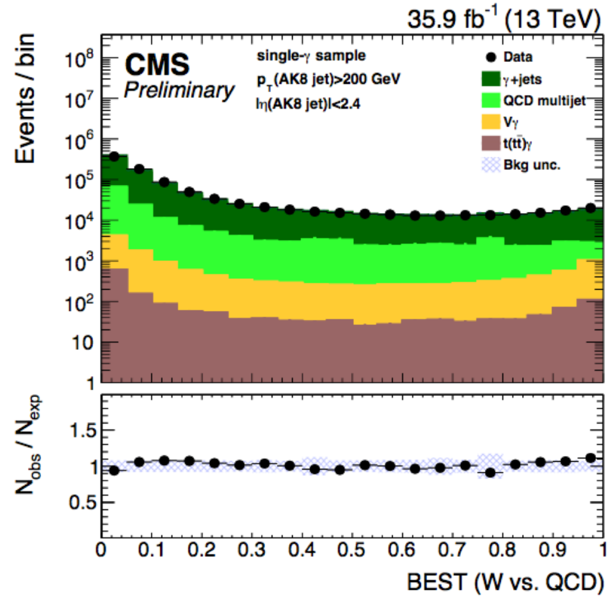

*Jet $p_T$ reweighted to match data*

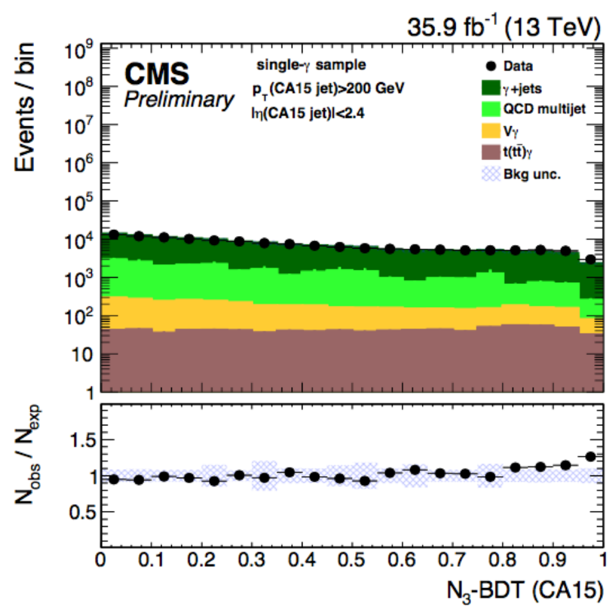
# Single-Photon Sample: BEST/N$_3$-BDT

Update

Paper




*Jet $p_T$ reweighted to match data*