Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

New jet tagging techniques at CMS

Dennis Schwarz

on behalf of the CMS Collaboration

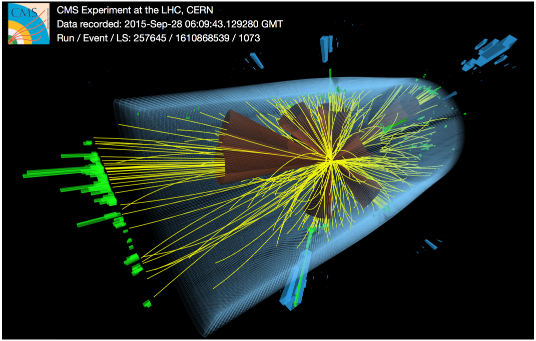
ICHEP 2020

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

Introduction



- Abundant jet production at the LHC with manifold origins:
  - Gluons
  - Light quark flavors
  - Boosted heavy objects (Top, W, Z, Higgs)
  - Pile up
- Many LHC analyses rely on efficient and stable identification

Boosted heavy objects

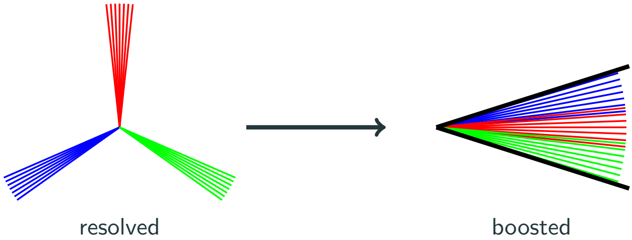
- Search for new heavy particles

→ Decays into high-p_T Top/W/Z/Higgs

→ Boosted decays

→ Reconstruction of hadronic decays in a single jet

Example: top quark



Jet substructure

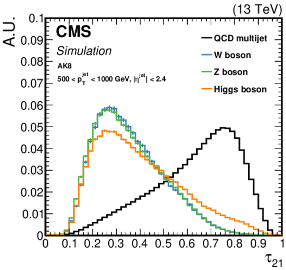
JINST 15 (2020) P06005

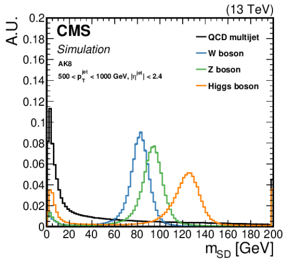
- Identification of jet origin using jet substructure

- N-prong structure of the jet
  $\rightarrow$ Energy distribution functions
  (e.g N-subjettiness ratio τ_{21} or τ_{32})

- Sensitivity to origin particle mass
  $\rightarrow$ Jet mass m_{jet} or m_{SD}

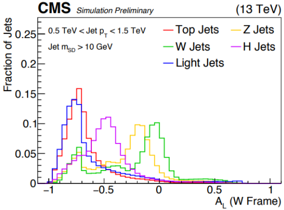
$\rightarrow$ Can we do better?

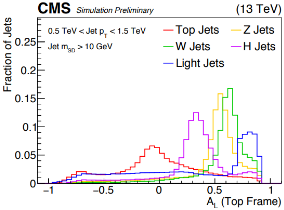




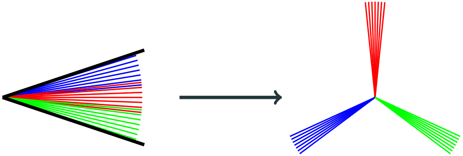
BEST (Boosted Event Shape Tagger)

- Boost jet into rest frame assuming top/W/Z/H mass
- If boost is correct
  - → Isotropic angular distribution
  - → Momentum symmetry in rest frame
- Fully connected neural network with additional jet properties (N-subjettiness, mass, ...)





$$A_L = \frac{\sum_{\text{jet}} p_z^{\text{jet}}}{\sum_{\text{jet}} p^{\text{jet}}}$$

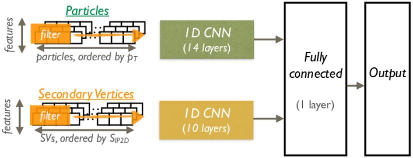


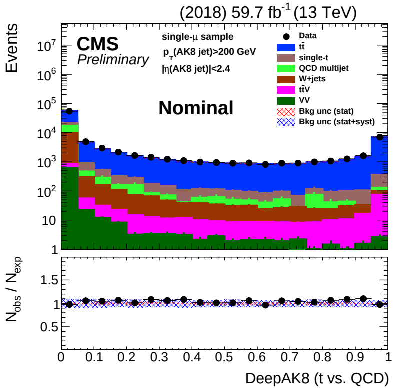
lab frame   rest frame

DeepAK8

CMS DP-2020-025, JINST 15 (2020) P06005

- Multi-class classifier
- Machine learning approach using particle candidates
- Properties of up to 100 jet constituents and 7 secondary vertices
- Two separate networks, later combined with fully connected layer



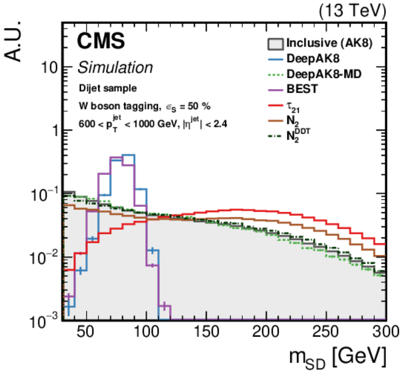


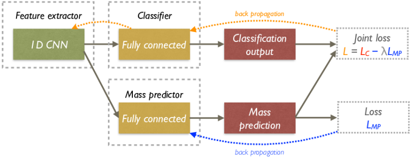
Stability against m_{jet}

JINST 15 (2020) P06005

- DeepAK8 learns jet mass as feature
  - → Sculpts m_{jet} distribution
  - → Complicates background estimation
- De-correlation with mass predictor network
  - → Penalty if outputs largely correlated with mass

→ Efficiency stable against jet mass after de-correlation

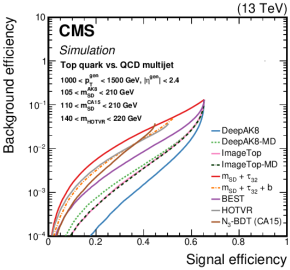


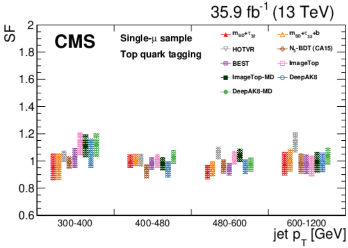


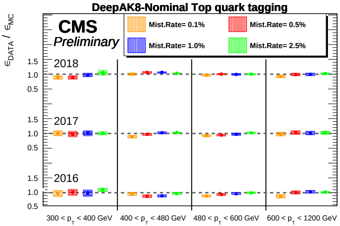
Performance and validation

JINST 15 (2020) P06005, CMS DP-2020-025

- Large machinery validates tagger performance in CMS
- Machine learning approaches already outperform classical taggers
- Correction factors are derived via tag and probe method



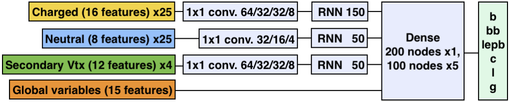


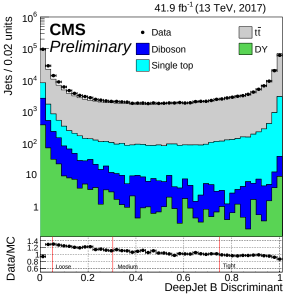


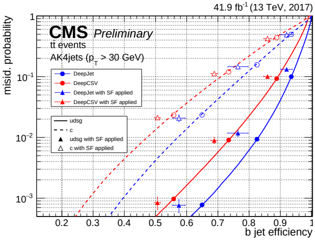
DeepJet

CMS DP-2018-058

- Identification of small radius jets (AK4)
- Distinguish quark flavors and gluons
- Deep neural network approach
- Inputs: Charged, neutrals, secondary vertices, global jet variables



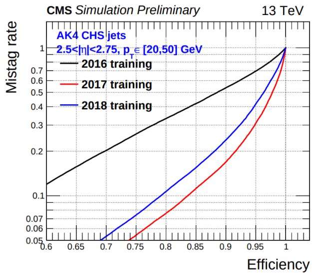


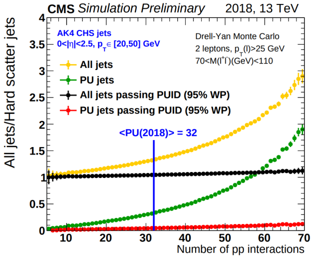


Pile up mitigation

CMS DP-2020-020

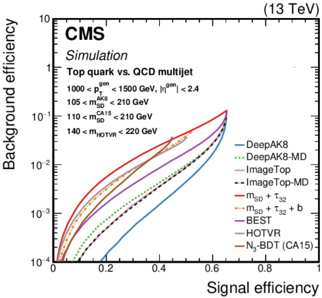
- Pile up jet identification (PU ID)
- Boosted decision tree
- Jet and event variables
- Huge improvement in $2.5 < |\eta| < 2.75$ region with new pixel layer!





Summary

- Jet identification crucial for LHC analyses
- Large machinery within CMS validates performance
- Categorization of jets profits from machine learning techniques and lots of new taggers are developed



And there is more:

- ImageTop [JINST 15 (2020) P06005]
- Particle Net [Phys. Rev. D 101, 056019 (2020), CMS DP-2020-002]
- ...