UNIVERSITY OF HELSINKI

HELSINKI INSTITUTE OF PHYSICS

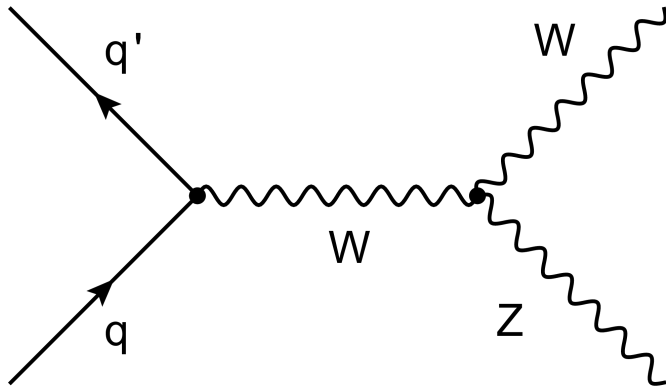
CMS Compact Muon Solenoid

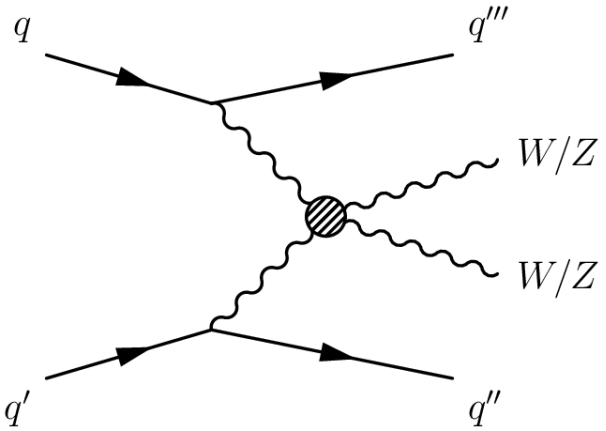
Jet Substructure in CMS

Nurfikri Norjoharuddeen
nurfikri.bin.norjoharuddeen@cern.ch

COMETA WG3
Topical Meeting on Jet Substructure
22 April 2024

Motivation for Jet Substructure in Multi-Boson Analyses

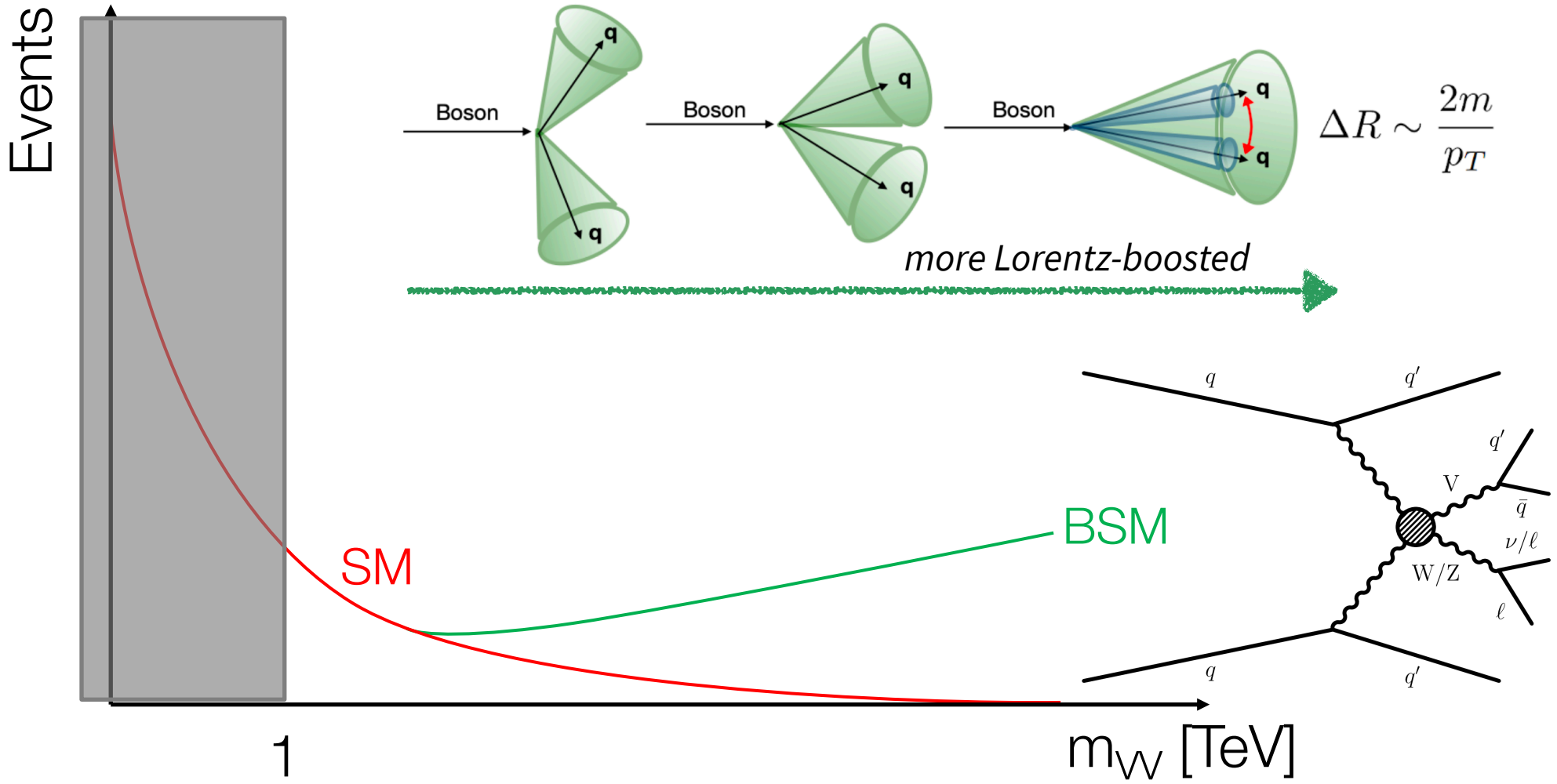




Motivation for Jet Substructure in Multi-Boson Analyses

TeV-scale m_{VV}

"boosted" boson decays -> collimated decay products



Jet Algorithms in CMS

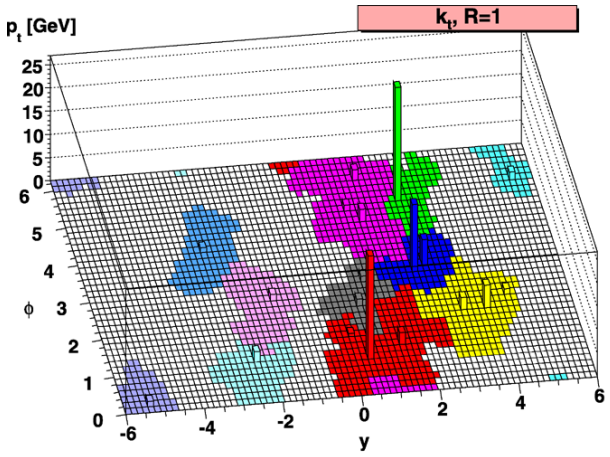
EPJ C 67, 637–686 (2010)

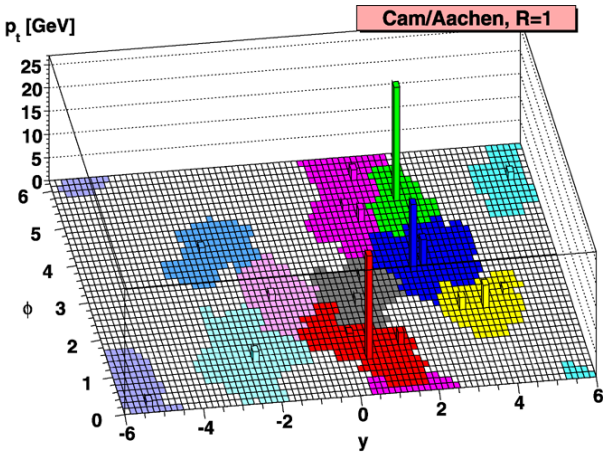
Sequential clustering algorithms with distance parameter R

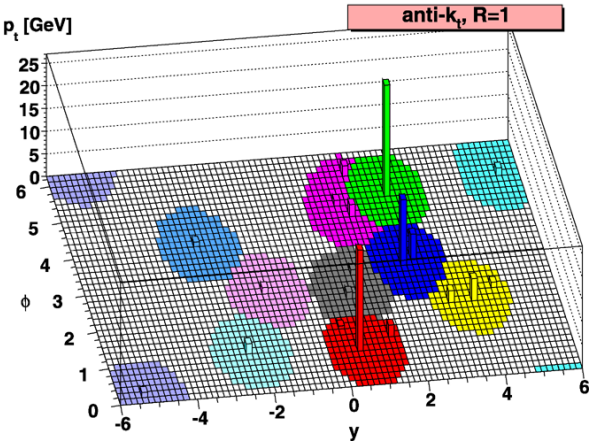
1) k_T
2) Cambridge-Aachen (CA)
3) anti-k_T (AK)

$$d_{ij} = \min\left(p_{ti}^{2p}, p_{tj}^{2p}\right)\frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = p_{ti}^{2p},$$







"Standard" jet algorithms for **CMS**

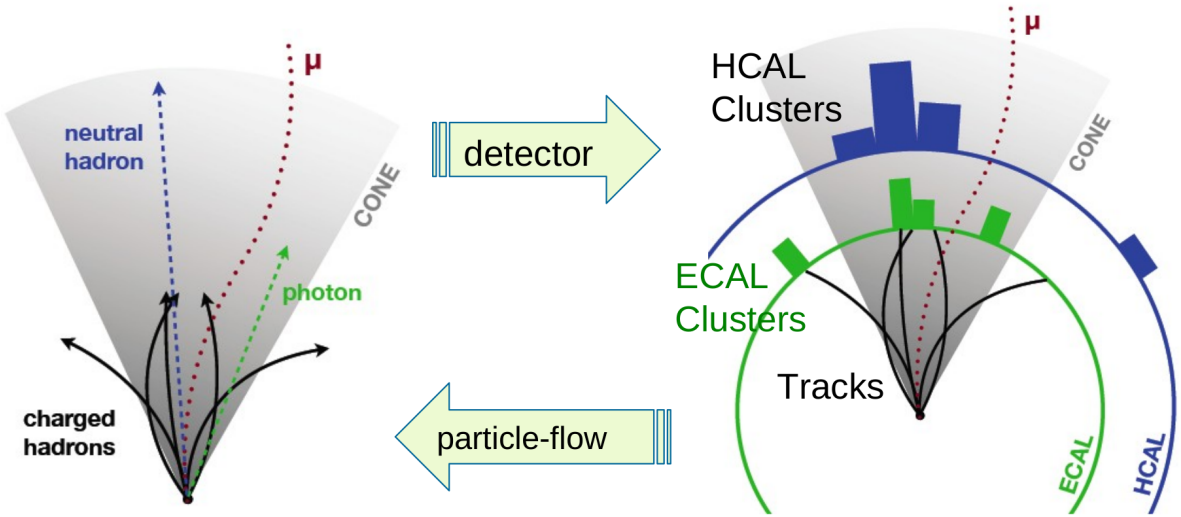
- Small-R jets: anti-k_T R = **0.4 [AK4]**
- Large-R jets: anti-k_T R = **0.8 [AK8]**

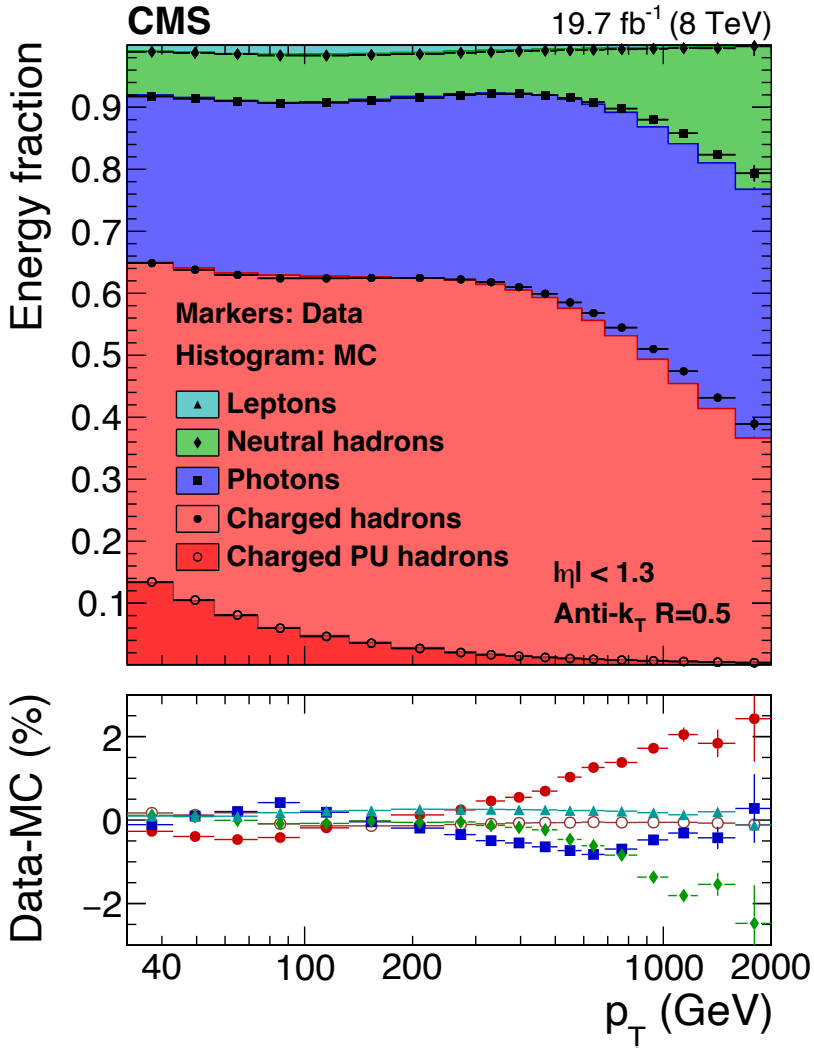
- AK or CA R=1.5 are sometimes used. Analysis specific.
- "Non-standard" jets used also (e.g Variable-R jets) JINST 15 (2020) P06005

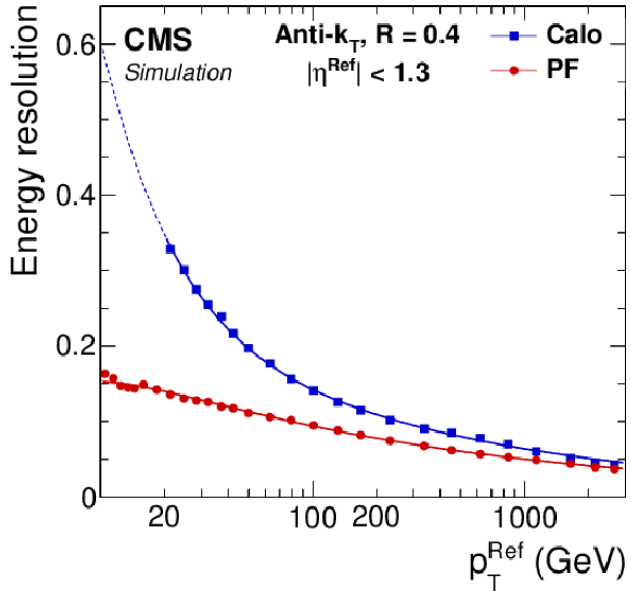
Input for Jet Reconstruction

Particle Flow (PF) Algorithm

JINST 12 (2017) P10003







Pileup Mitigation for Jet Reconstruction

JINST 15 (2020) P09018
CMS-DP-2021-001

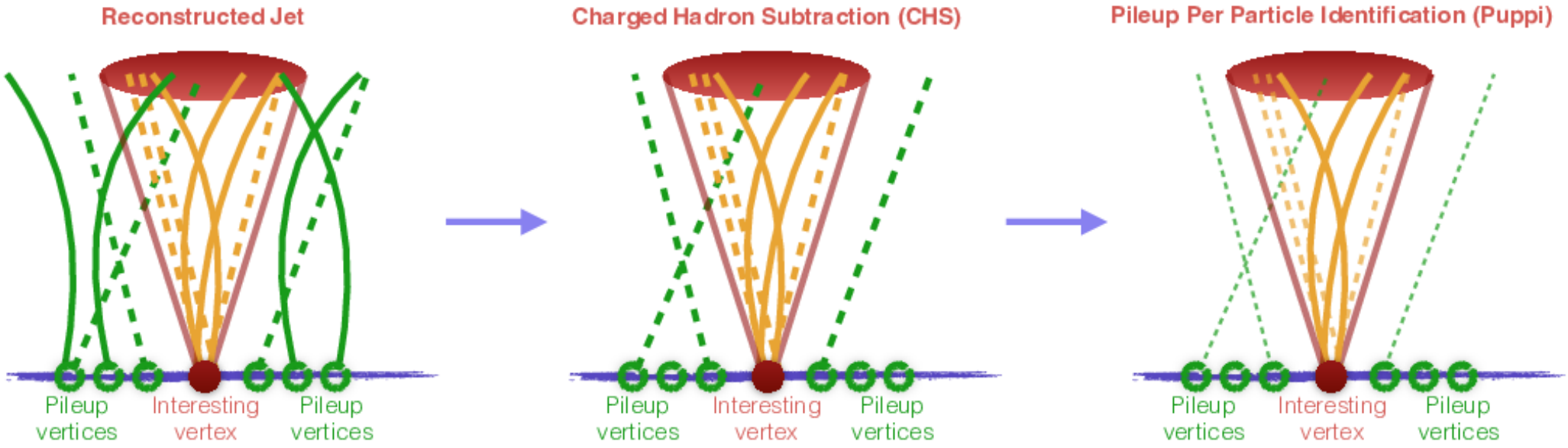


Figure by Andrea Malara

Constituent-level pileup mitigation
Charged Hadron Subtraction (CHS)
Pileup per Particle Identification (Puppi)

(Large-R) Jet Grooming

JHEP 1405 (2014) 146
JINST 15 (2020) P06005

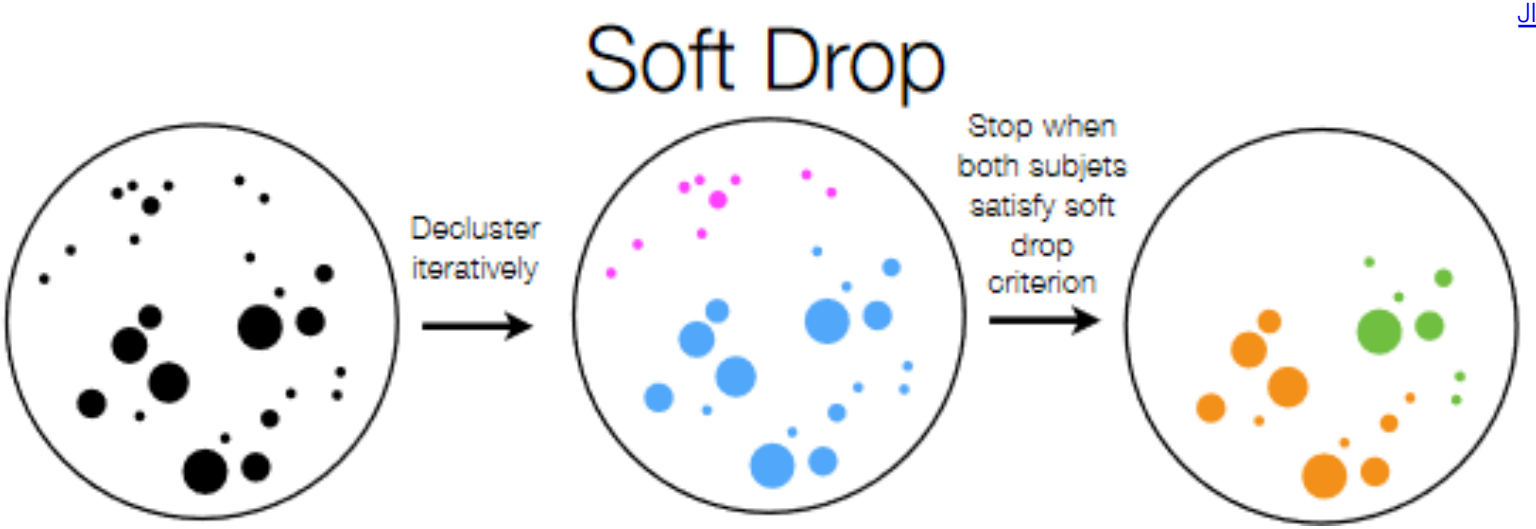


Figure by J. Dolen

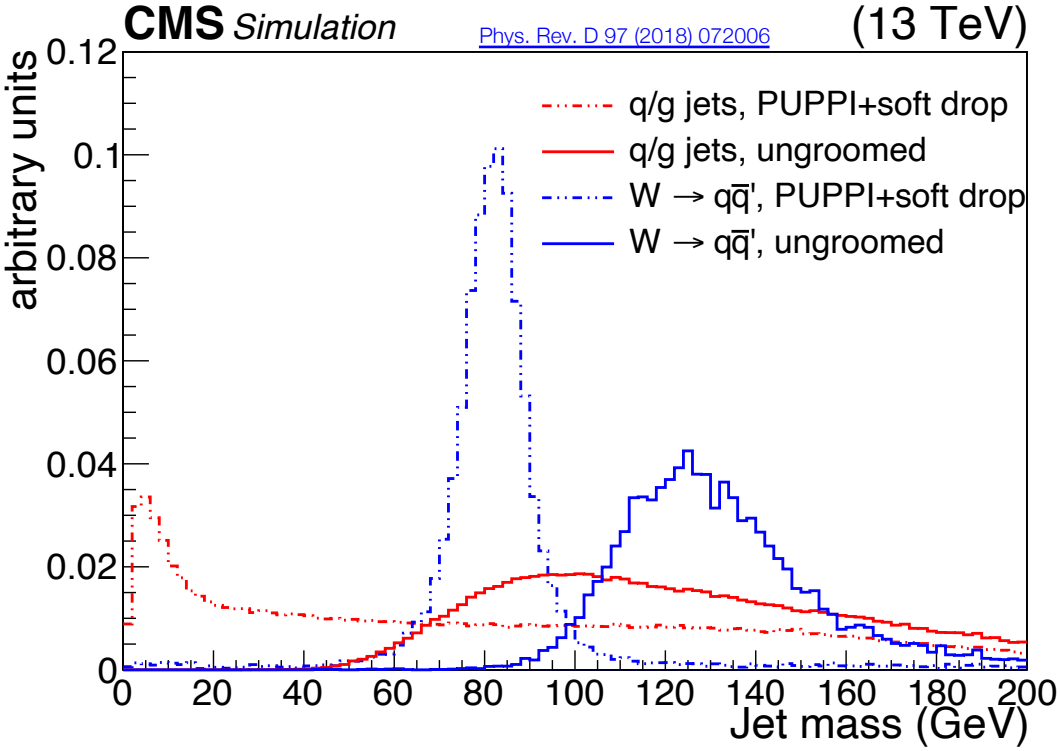
1. Recluster with Cambridge-Aachen algorithm.
2. Reverse clustering history.
3. Check criterion:

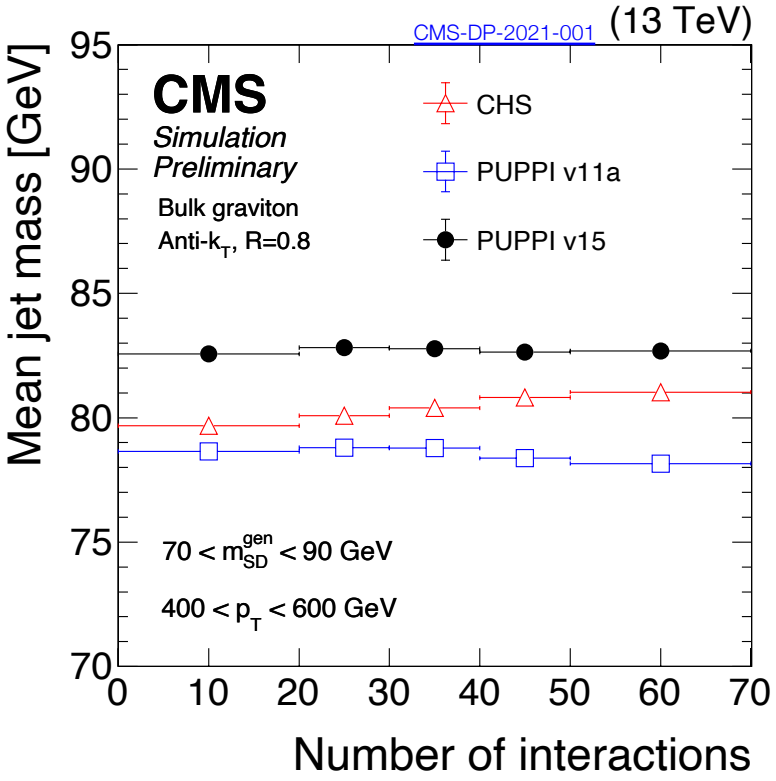
$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta} \qquad z_{cut} = 0.1 \qquad \beta = 0$$

4. Pass: two subjets are final.
   Fail: remove sub-leading subjet & repeat (1).

(Large-R) Jet Grooming

Jet grooming (+ Pileup Mitigation) “cleans” up large-R jets





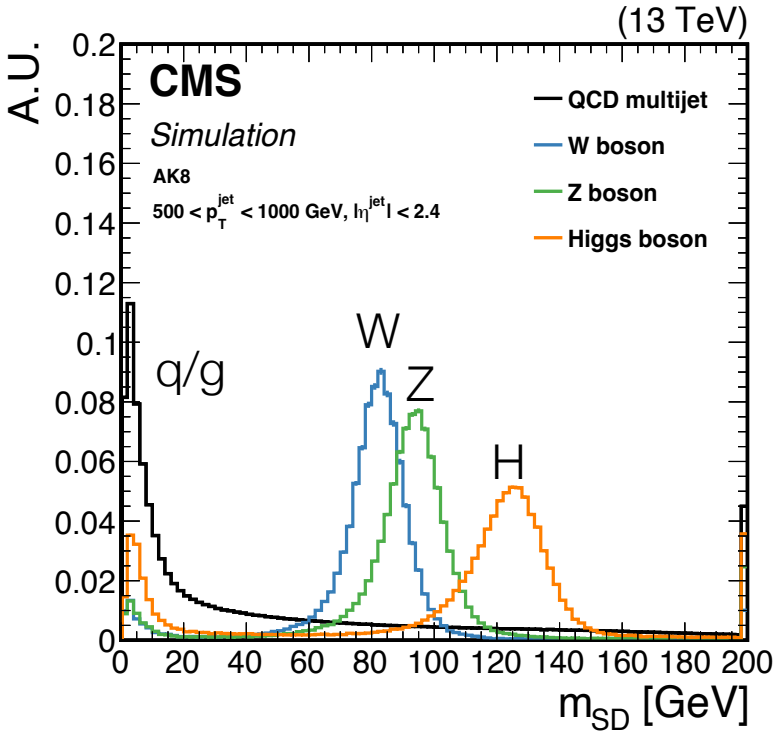
Jet Tagging (Jet Mass & Substructure Observable)

JINST 15 (2020) P06005

Jet mass

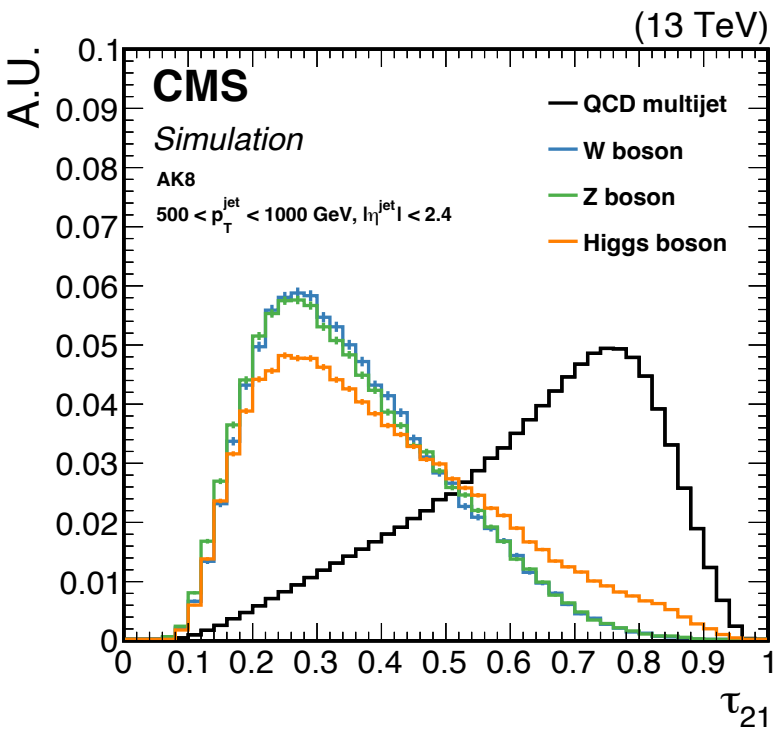
Measures of energy distributions

Soft-drop mass

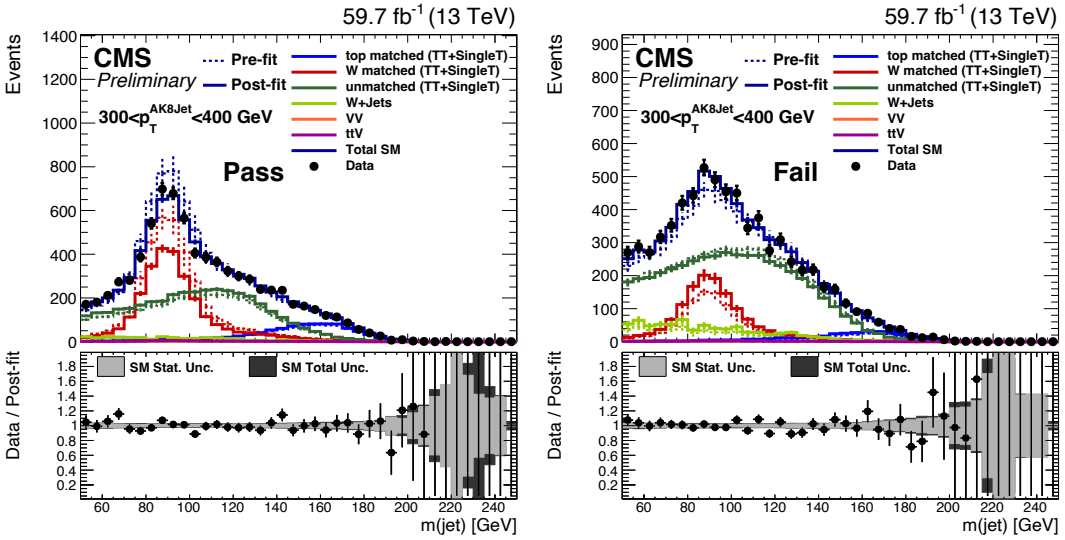


N-subjettiness

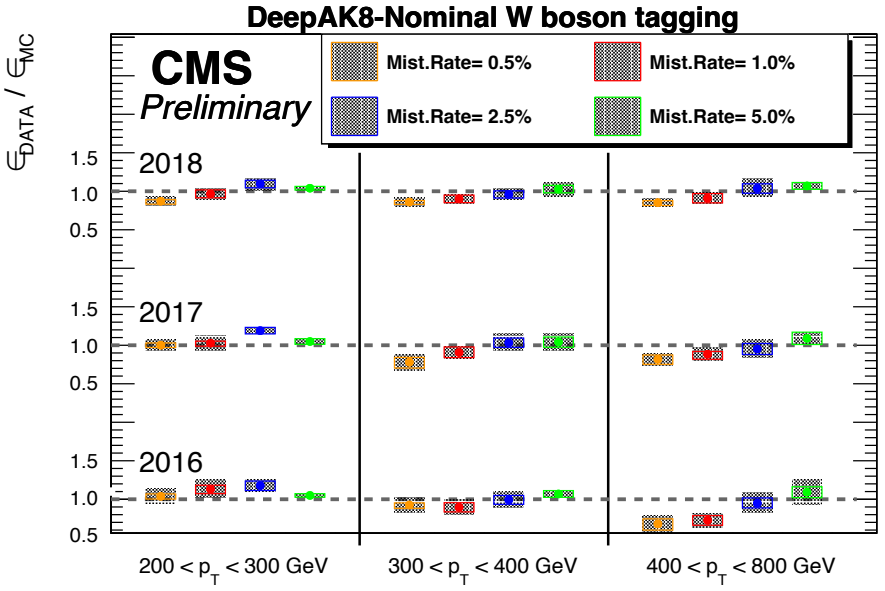
$$\tau_N = \frac{1}{d_0} \sum_k p_T^k \min(\Delta R_{1,k}, \ldots, \Delta R_{N,k}) \qquad d_0 = \sum_k p_T^k R_0$$



Jet Tagging Efficiency Calibration



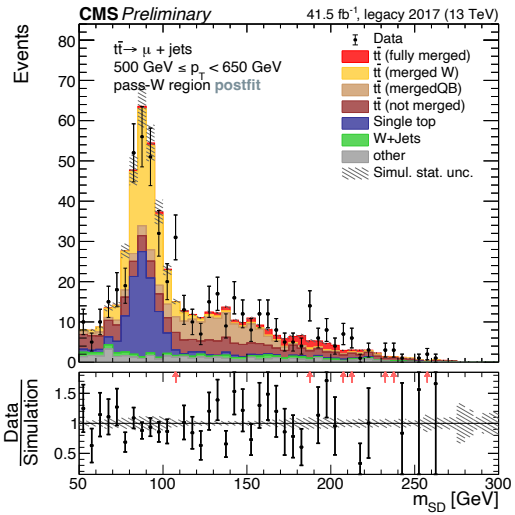
[CMS-DP-2020-025]

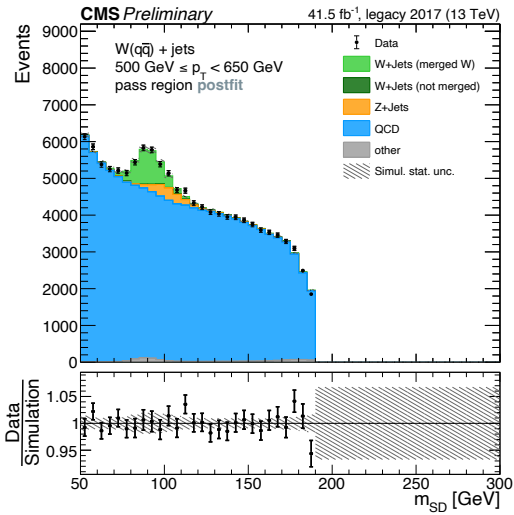


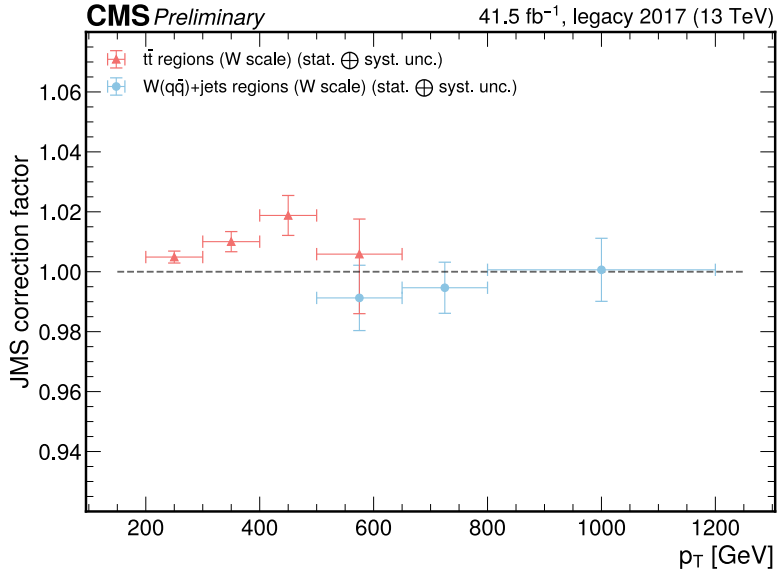
- Measure the tagging efficiency in simulation & data.
- Often done with samples enriched with W→qq (semi-leptonic tt decay).
- Procedure:
  - Define a tagging selection
  - Create pass/fail regions.
  - Simultaneous likelihood fit of both regions to obtain relative normalizations.

Jet Mass Scale Calibration

- Adapt efficiency measurement analyses to measure the jet mass scale (JMS).
- Done with samples enriched with W→qq (tt & W+jets).
- Fit the JMS in data & simulation.







CMS-DP-2023-044

Recent developments of jet tagging & substructure

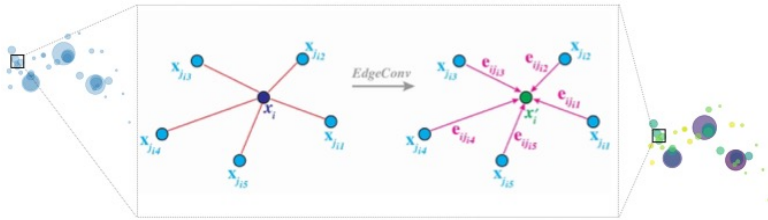
Constituents-based Jet Tagging

Can we improve boson tagging using lower-level jet information (i.e constituents, secondary vertices) as inputs to Machine Learning techniques?

ParticleNet (Latest) arxiv:1902.08570

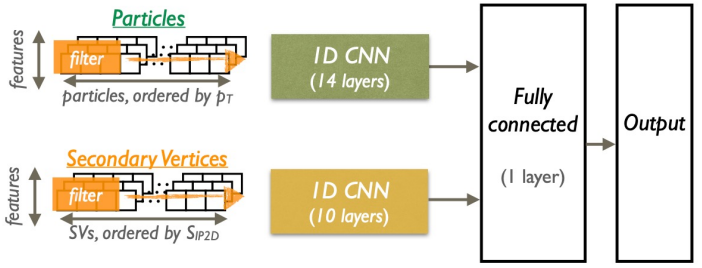
- "Particle cloud" representation of jets.
- Graph CNN architecture.
- Output: Multi classification scores for W / Z / H / top/QCD + decays.

Figure by Huilin Qu

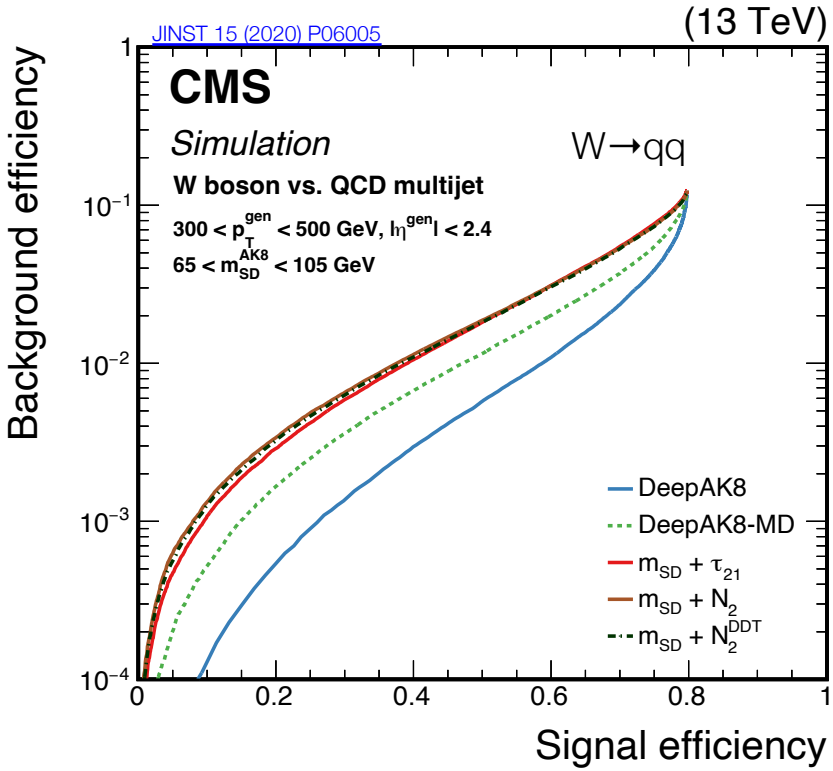


DeepAK8 arXiv:2004.08262

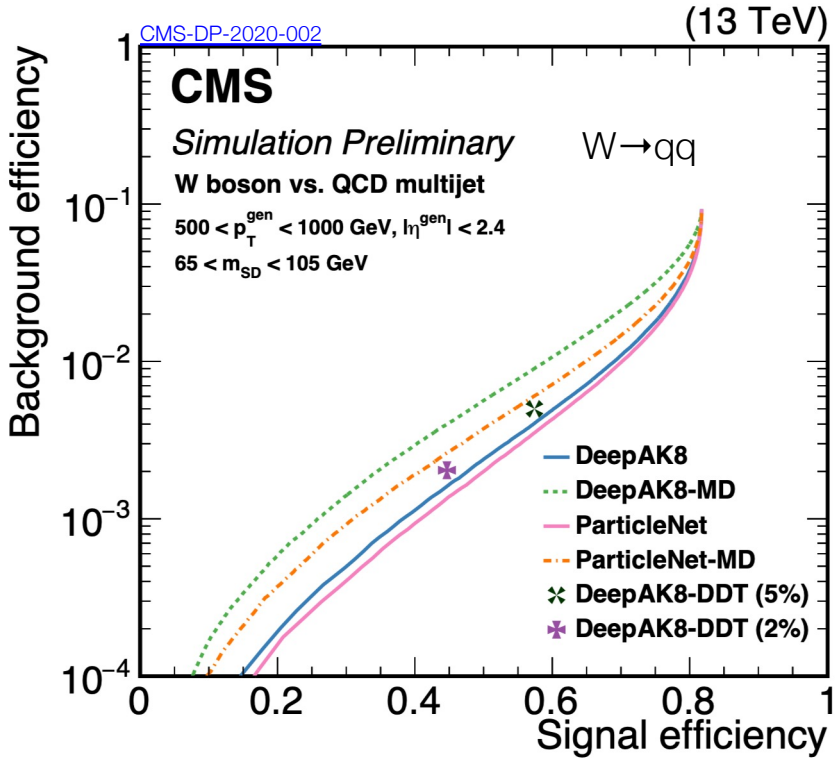
- 1D Convolutional Neural Network (CNN)
- Based on ResNet architecture.arxiv:1512.03385



Constituents-based Jet Tagging: Performance

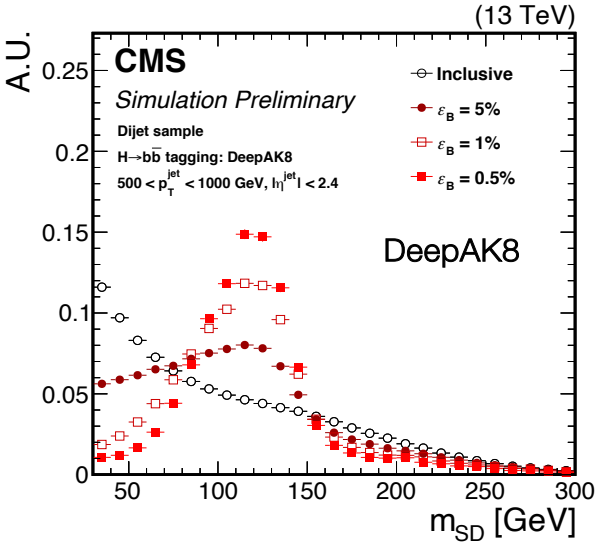


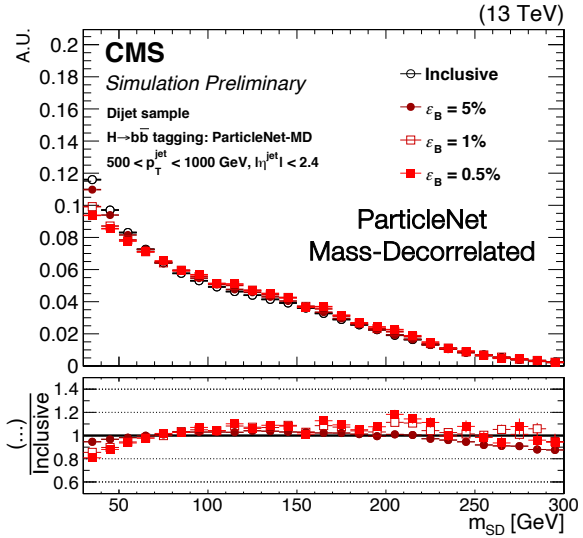
DeepAK8 significantly better at tagging jets compared to substructure variables.



ParticleNet improves on the performance of DeepAK8.

Constituents-based Jet Tagging (Mass-Decorrelated)

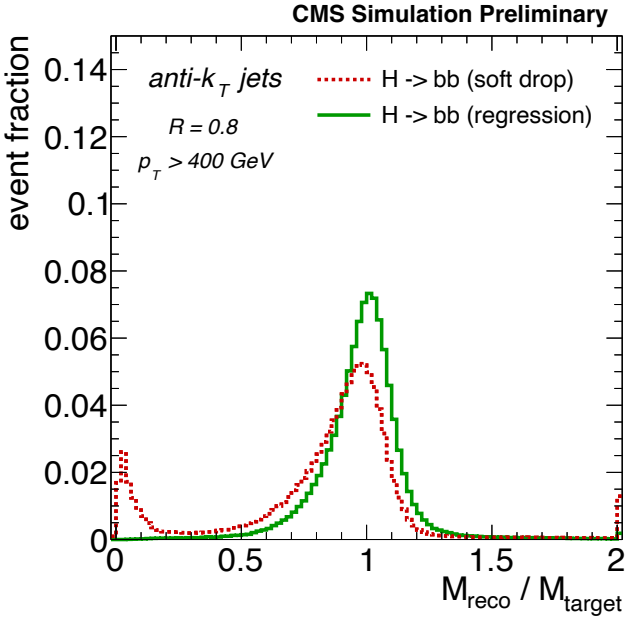


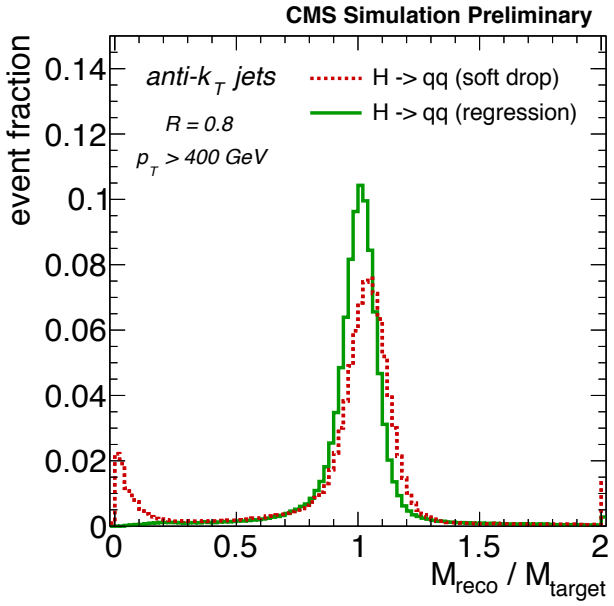


- Substructure variables / tagger outputs are correlated with jet mass.
- Selections leads to sculpting of jet mass distribution of QCD background processes.
  - ➢ Undesirable for analyses using jet mass to predict background processes.
- Mitigate mass-sculpting through Mass-Decorrelation.
- ParticleNet Mass-Decorrelated version (ParticleNet-MD)
  - Training with dedicated signal sample of hadronic decays of spin-0 particle X with flat mass spectrum m_X.
  - Smooth mass shapes achieved.

Constituents-based Jet Mass Regression

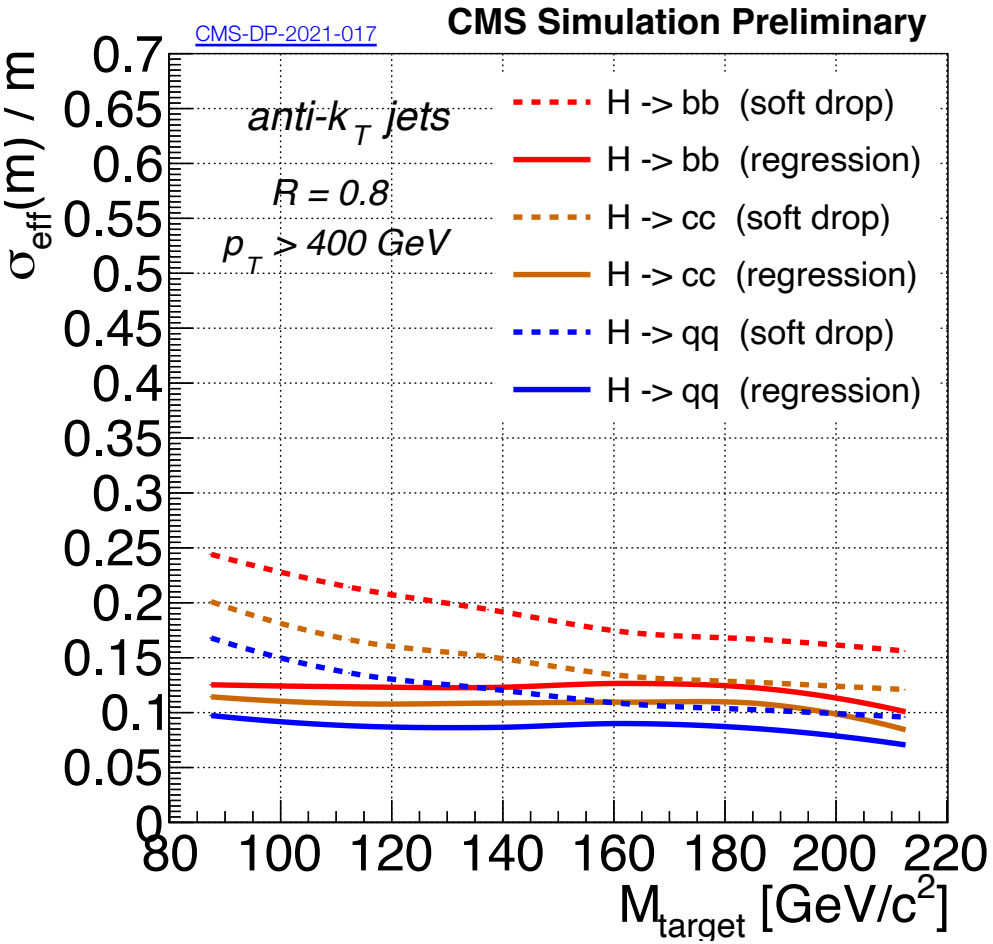
CMS-DP-2021-017





- Develop algorithm to reconstruct jet mass with the best possible jet mass scale and resolution
- Exploit ParticleNet architecture to predict the mass of AK8 jets.
  - Same inputs & training configuration as tagging.
- Additional output of ParticleNet: “regressed mass”.
- Substantial improvement in the jet mass scale & resolution.

Constituents-based Jet Mass Regression

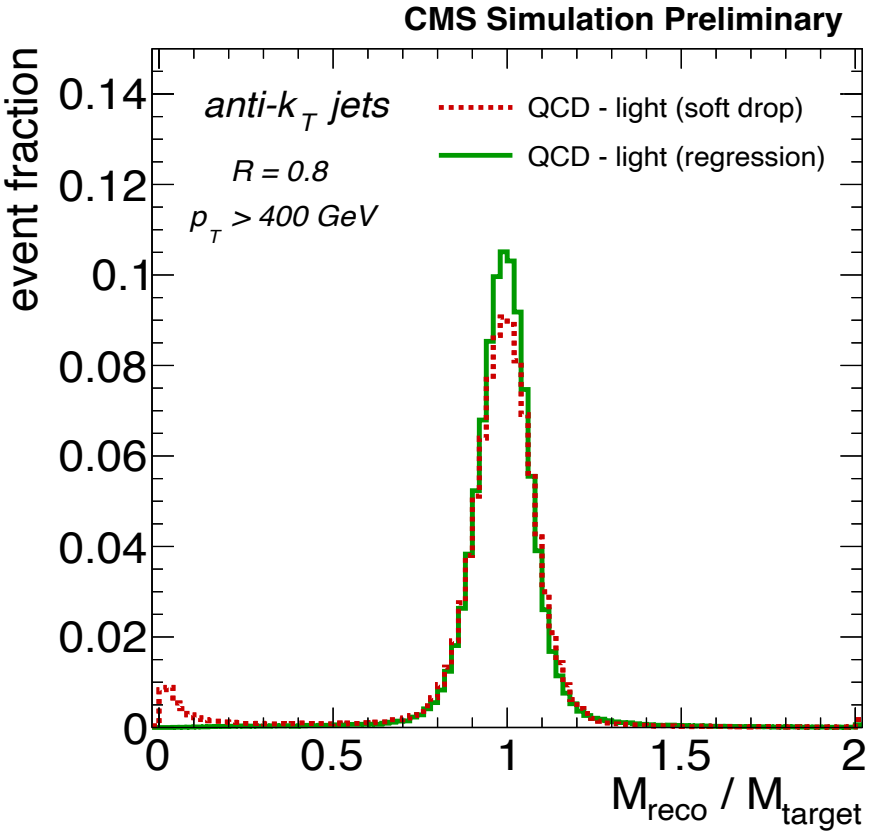


Improvement of signal jets' regressed mass resolution & stable across mass.

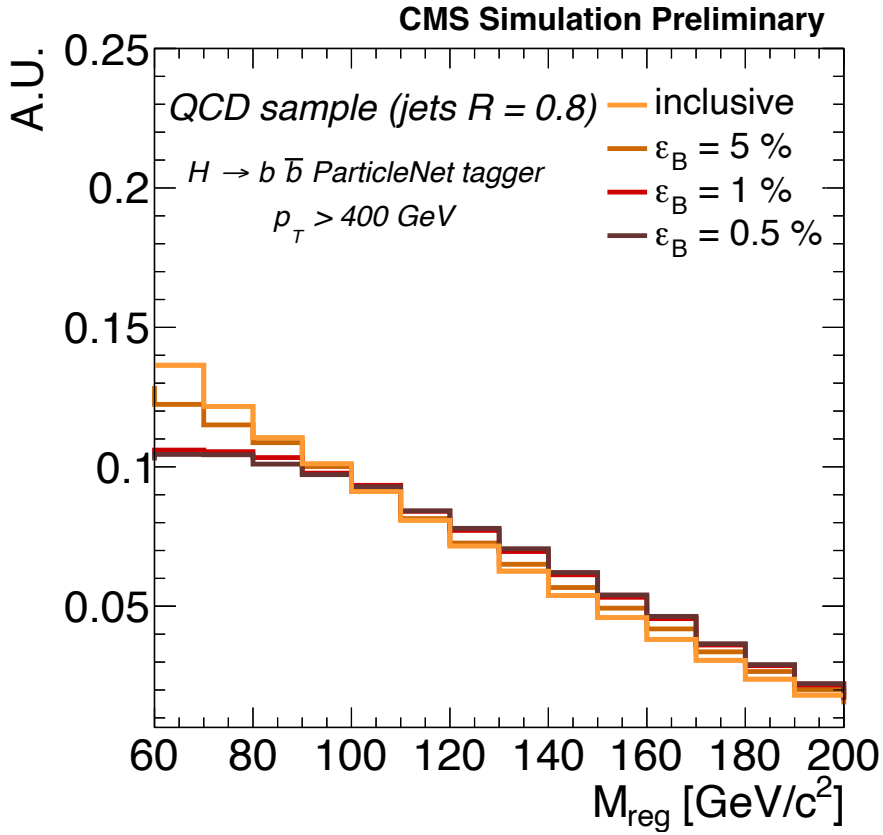
Constituents-based Jet Mass Regression

CMS-DP-2021-017

Impact on QCD-jets



Improves resolution



Limited sculpting after tagging

Summary

- CMS uses Anti-k_T R=0.8 jets to tag boosted hadronically decaying bosons.
  - Inputs: Particle Flow + Puppi.
  - Grooming: Soft-drop.

- Tag jets with mass & variables quantifying substructure of the jets.

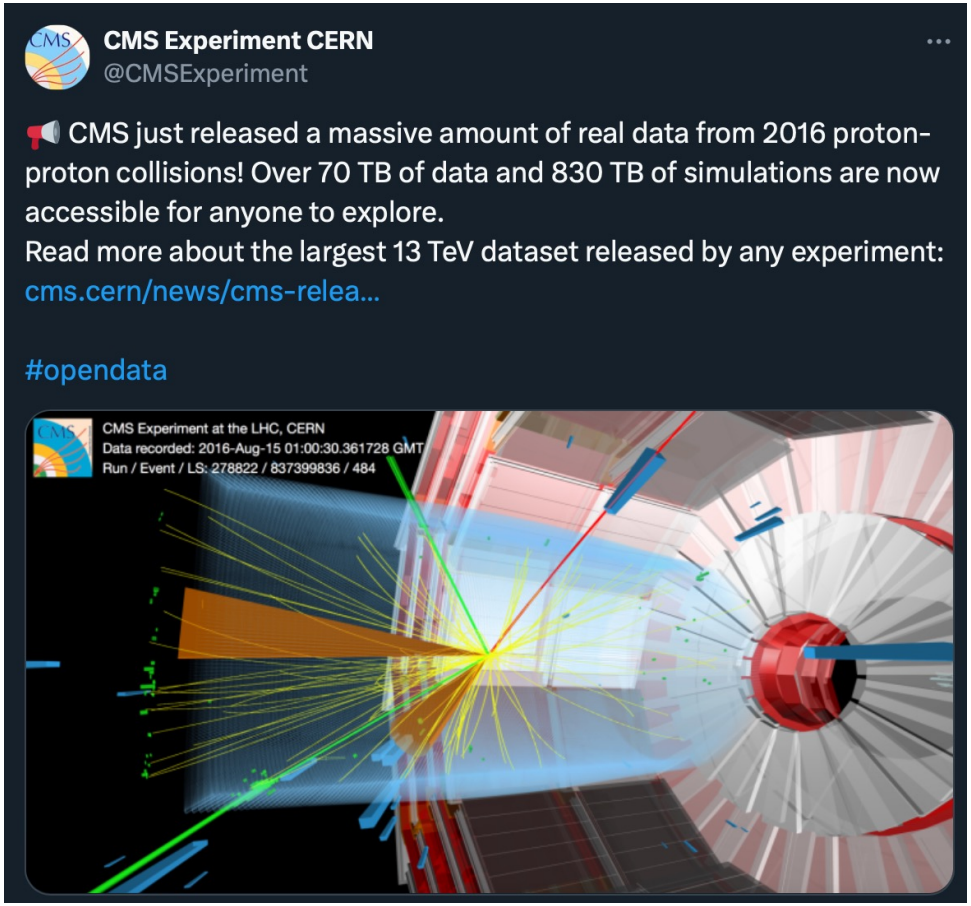
- Recent developments with Machine Learning techniques for substructure.
  - Latest tagger: ParticleNet architecture to identify jets from hadronic boson decays & the mass of the jets.
  - Substantial gain in performance with respect to substructure variables & DeepAK8.

CMS Open Data

- CMS recently released (half of) 2016 datasets & simulations.

- New data format: NanoAOD
- Light-weight.
- Readable with bare ROOT or python with uproot.

- NanoAOD with Particle Flow objects.
- Available for subset of datasets [1].
- Example provided to produce them by yourself [2].



https://opendata.cern.ch/docs/cms-releases-2016data-2024

[1] https://opendata.cern.ch/search?q=&f=experiment%3ACMS&f=file_type%3Ananoaod-pf
[2] https://opendata.cern.ch/record/12504

5th CMS Open Data Workshop (& Hackathon)

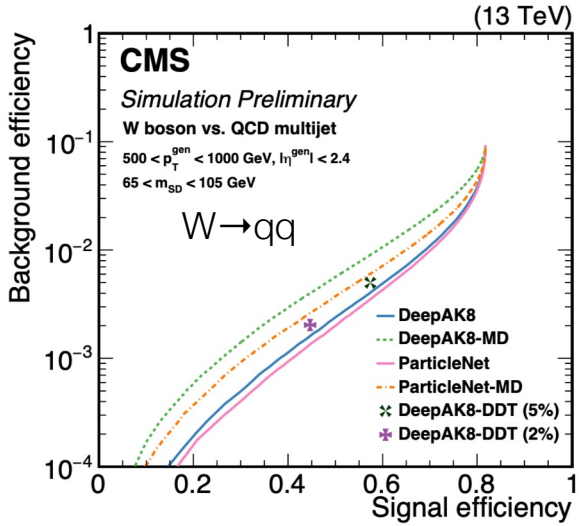
Bridge the technical gap between external analysts and CMS analysis machinery.

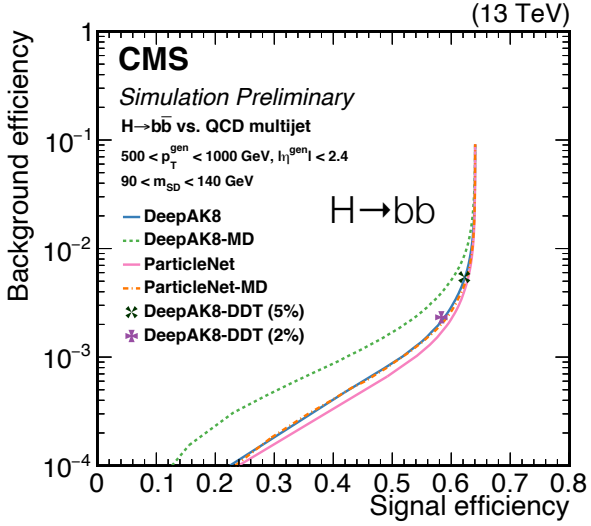
https://indico.cern.ch/event/1387505

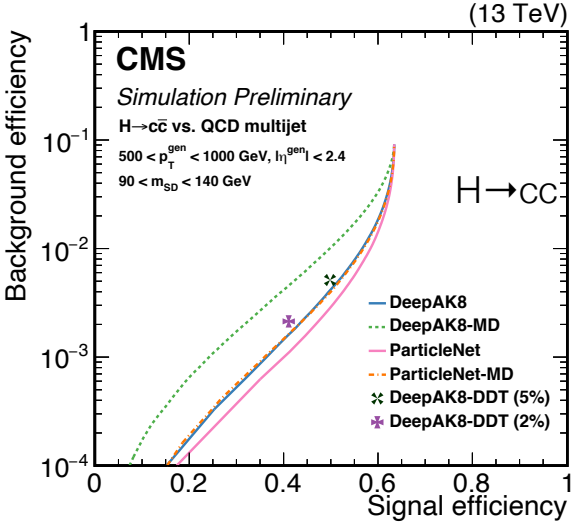
EXTRA SLIDES

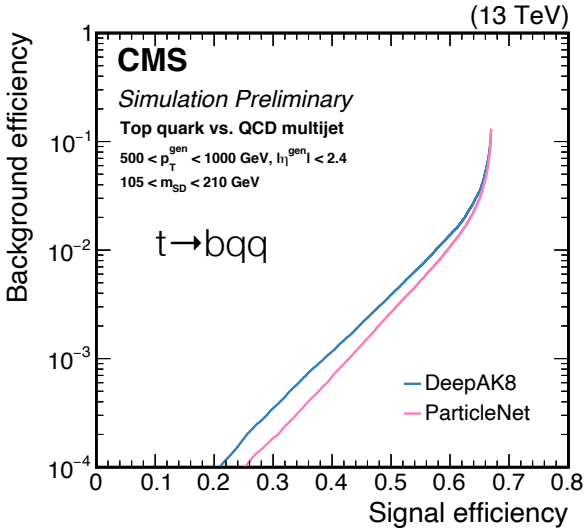
ParticleNet performance in simulation

CMS-DP-2020-002



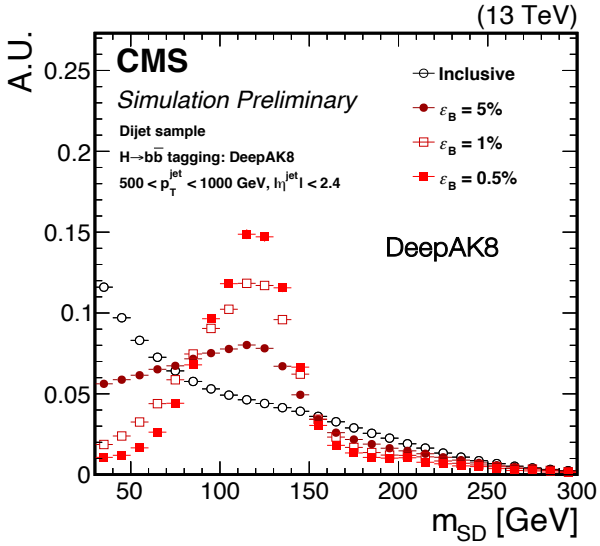


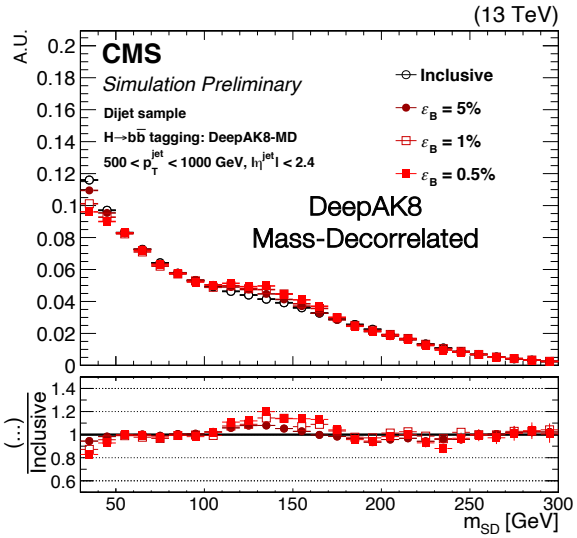


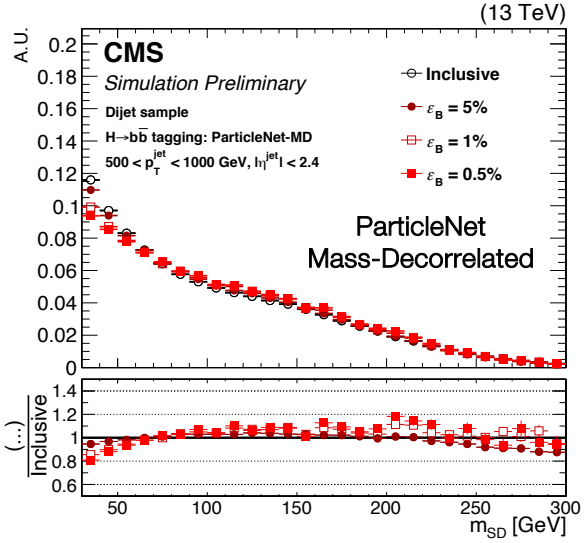


ParticleNet best performing tagger for W / Z / H / Top.

Jet tagging: Mass-decorrelation

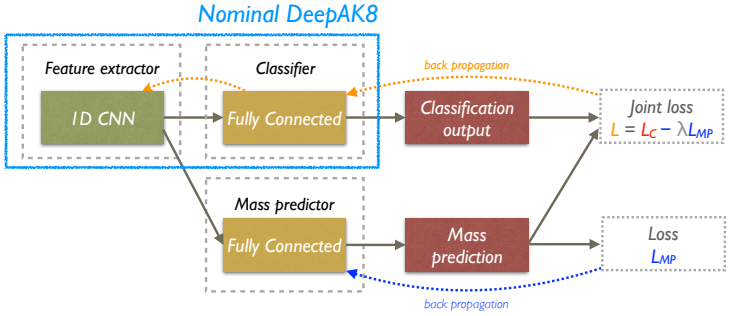






DeepAK8-MassDecorrelated version (DeepAK8-MD)

- Use adversarial training arXiv:1611.01046
  1) Add a mass prediction network to predict jet mass.
  2) Accuracy of prediction included in the loss function.
  3) Minimizing joint loss prevent mass correlation while improving classification accuracy.
- Signal/background samples reweighted to a flat (pT, mass) distribution.



Motivation for jet substructure in VBS

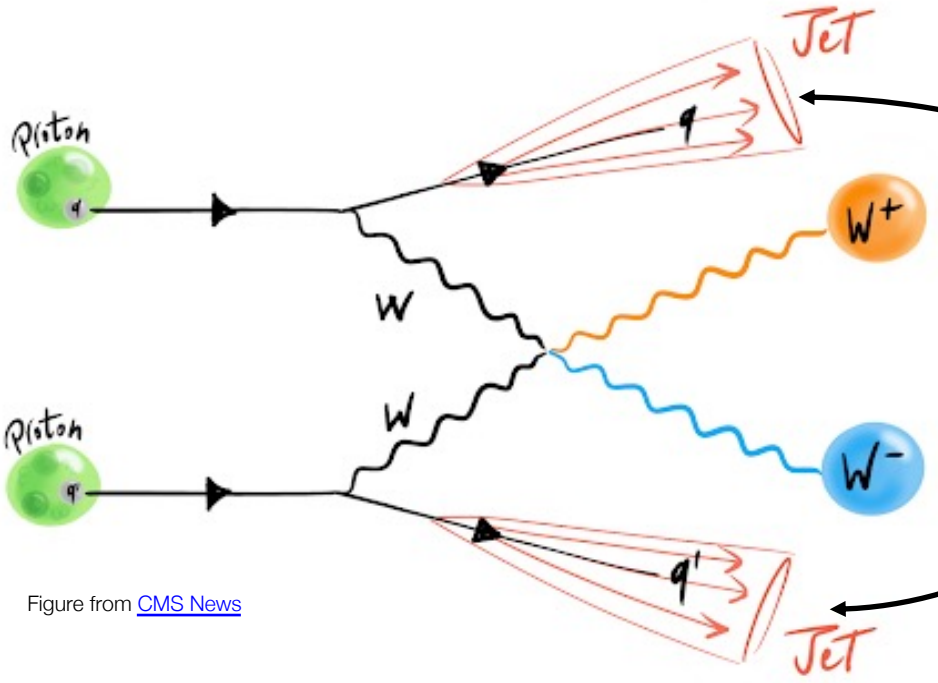


Figure from CMS News

2 Forward Jets

Pile-up Jet Suppression

Quark vs Gluon Discrimination

Pileup Mitigation

JINST 15 (2020) P09018

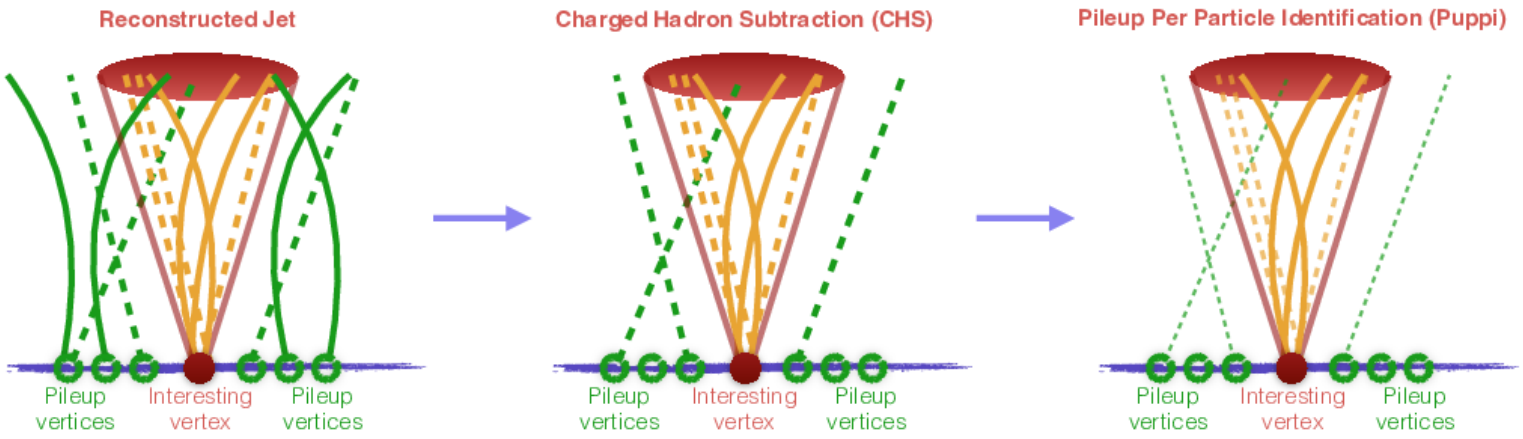
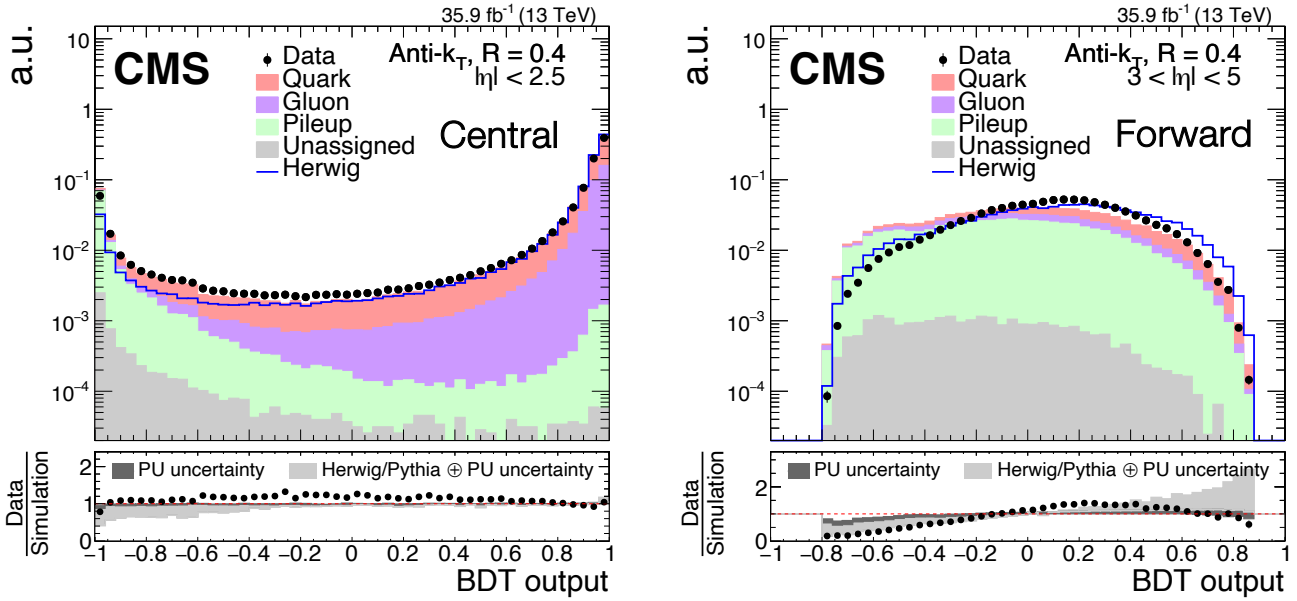


Figure by Andrea Malara

Constituent-level pileup mitigation

Pileup Jet Identification with BDT

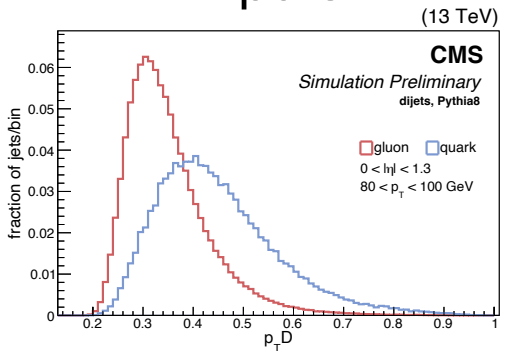


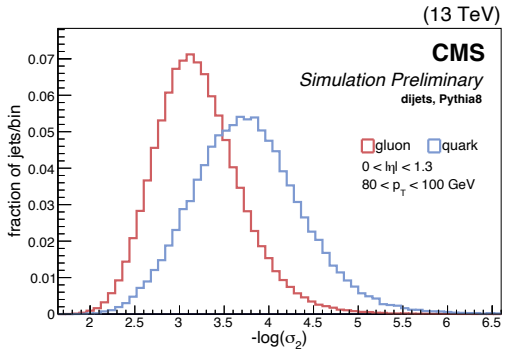
Quark-Gluon Jet Tagging

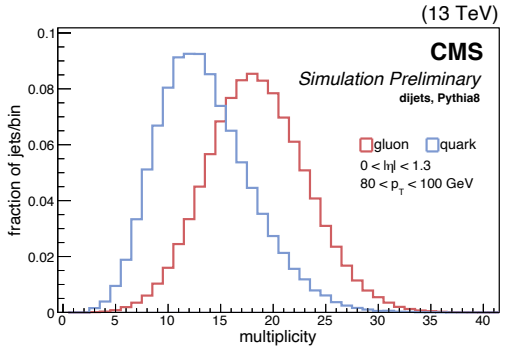
CMS-PAS-JME-16-003

Quark-Gluon Likelihood (QGL)

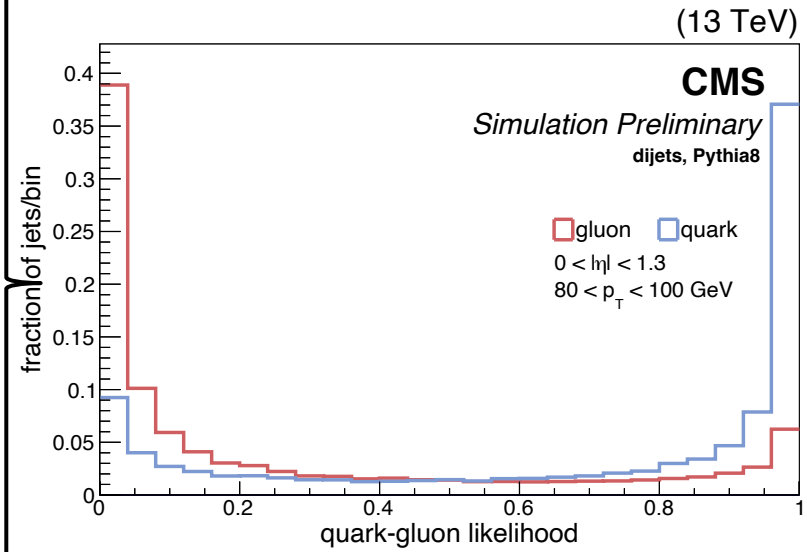
Inputs







QGL discriminant



Validation with Data