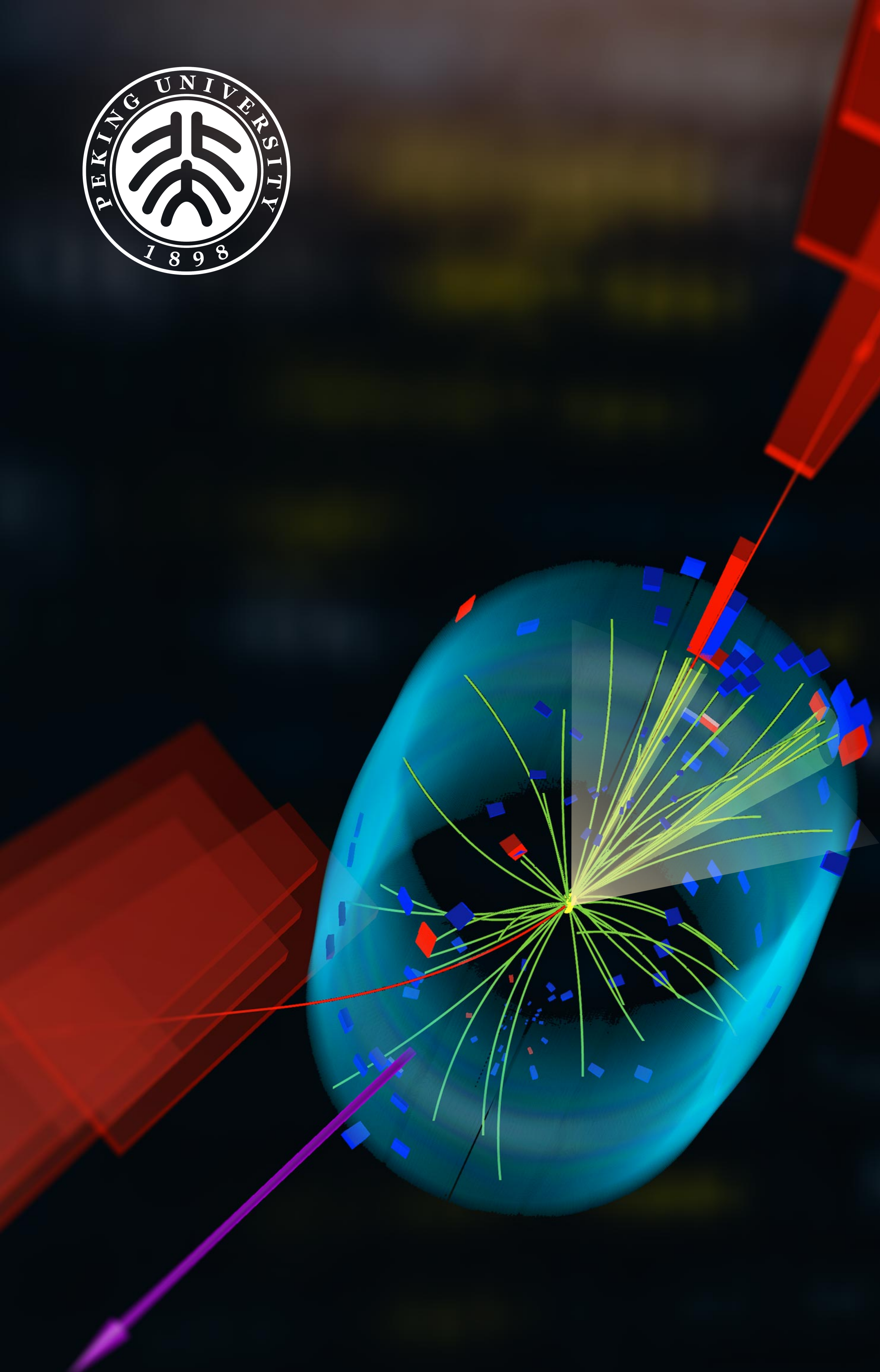


Boosted algorithms for searches

Congqiao Li *(Peking University)*

on behalf of the ATLAS and CMS Collaboration

LHCP2022 · Taipei
19 May, 2022

Introduction

➔ ***Boosted jet tagging algorithm*** explores phase space where the resonance jet is Lorentz-boosted and decay products are collimated

- all decay products can be collected within a large-*R* jet
- ***goals***: tagging resonances (W/Z/H/top) with hadronic decays and/or different flavour contents (X→$b\bar{b}$/$c\bar{c}$)
- ***technique***: rule-based jet substructure variables, BDT/DNN w/ jet observables, DNN w/ low-level constituent input

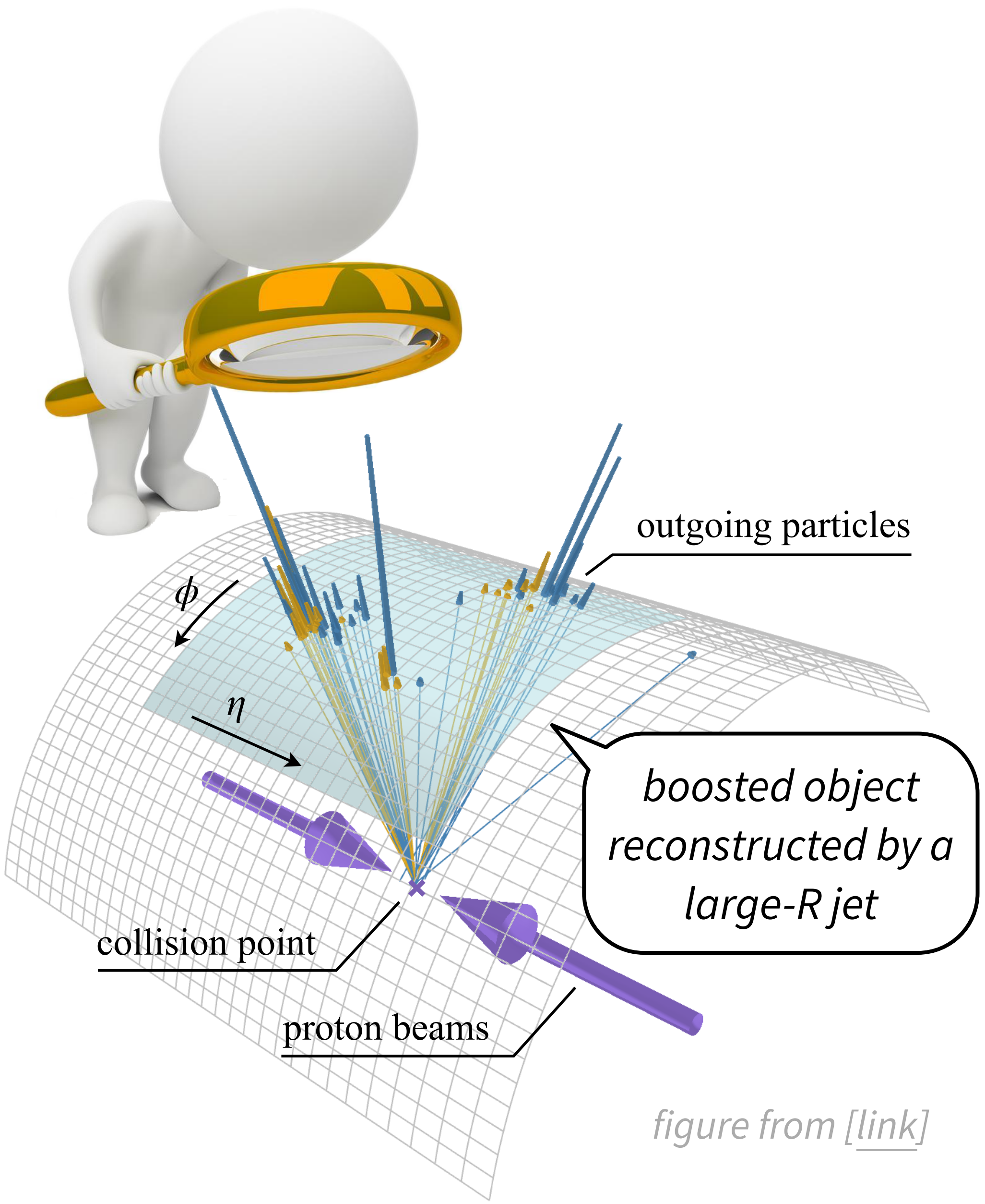


figure from [link]

Introduction

➔ ***Boosted jet tagging algorithm*** explores phase space where the resonance jet is Lorentz-boosted and decay products are collimated

- all decay
  large-*R* je
- ***goals***: ta
  hadronic
  contents
- ***techniqu***
  variables
  w/ low-le

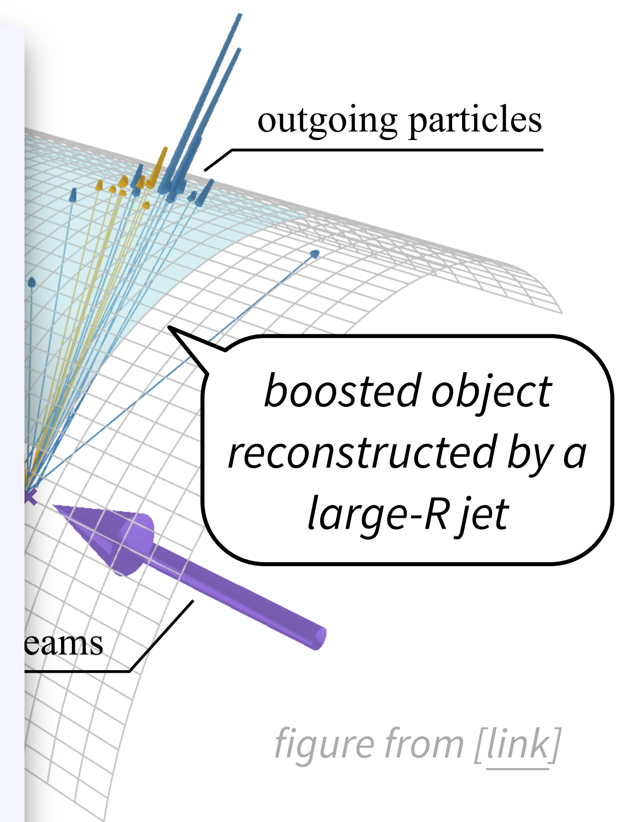


figure from [link]

Aim of this talk

- Provide an overview of recent boosted algorithms developed in ATLAS and CMS
- Highlight advanced techniques: new DNN architecture, mass decorrelation methods, calibration methods, …
- Showcase applications in new physics searches and standard model measurements
- Thoughts on future perspectives

How to reconstruct large-R jets?

Eur. Phys. J. C 81 (2021) 334

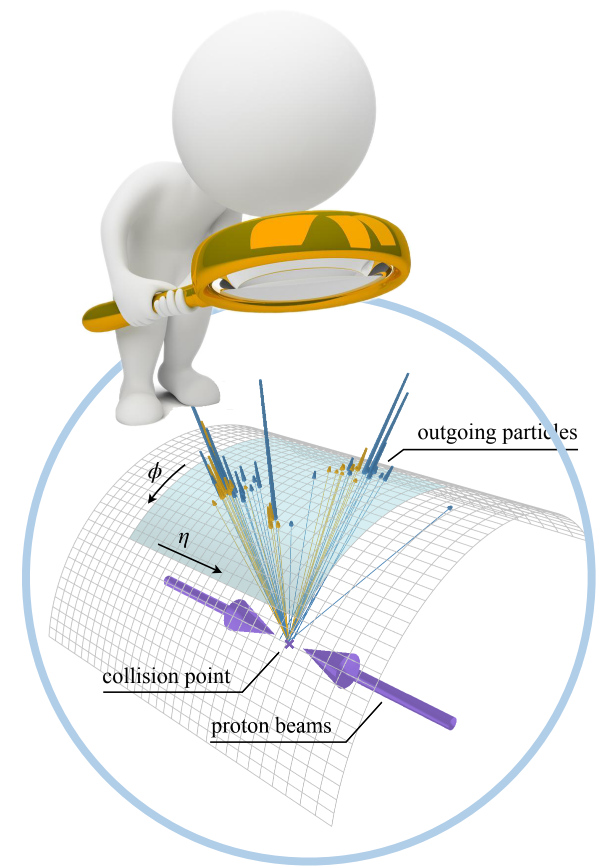
PF: JINST 12 (2017) P10003
PUPPI: JINST 15 (2020) P09018

→ ATLAS

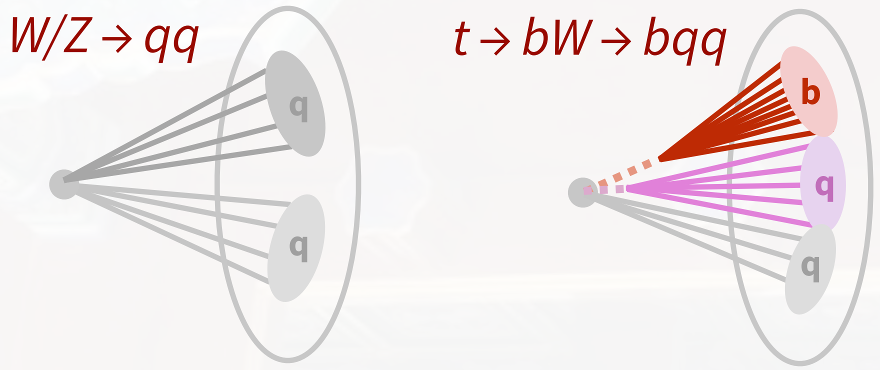
- ***LCTopo jets***: **topological cluster** ▶ clustered with by anti-k_T algo, R=1.0 ▶ groomed with trimming algo
- NEW ***UFO jets***: **Unified Flow Objects** (a combination of particle-flow objects (PFO) and Track-CaloClusters (TCC)) ▶ pile-up mitigation by Constituent Subtraction (CS)/ SoftKiller (SK)/PUPPI algo ▶ clustered by anti-k_T algo, R=1.0 ▶ groomed by soft-drop algo
  - PFO better at low p_T region; TCC benefits high p_T—UFO jets has better resolution across all p_T range
  - **latest tagging method applied to UFO jets** and see further improvements

→ CMS

- ***Large-R PUPPI jets***: **particle-flow (PF) candidates** ▶ pile-up suppressed by PUPPI algo (assign each PF candidate a factor to scale its 4-vec) ▶ clustered by anti-k_T algo, R=0.8 ▶ groomed by soft-drop algo



Heavy resonance tagging



W/Z→qq tagging (I): theory-inspired variables

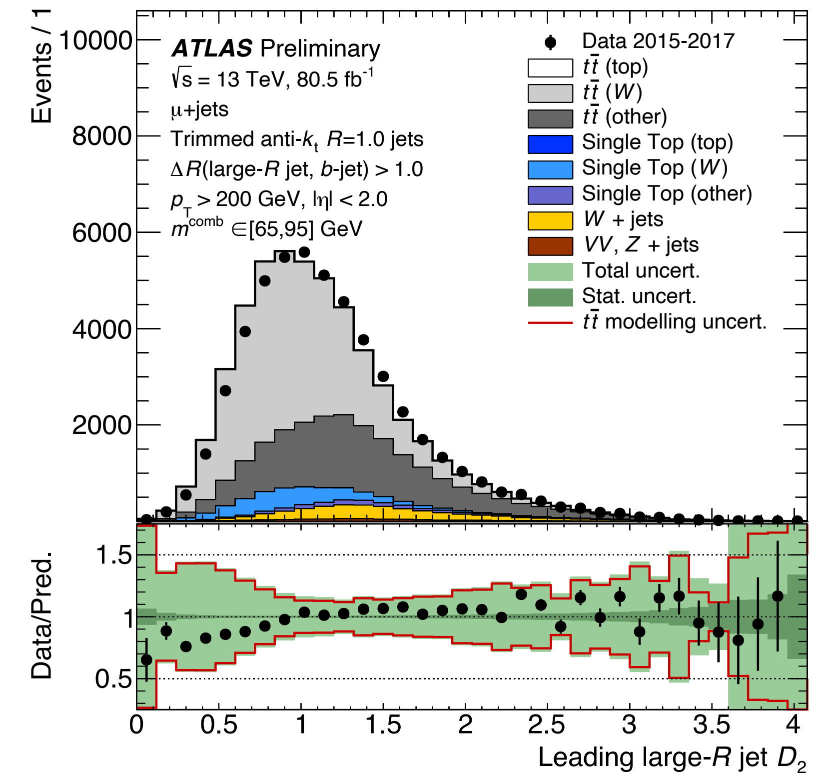
- ➔ Hadronic W/Z-tagged jets distinguished from QCD jets by their two-prong structure

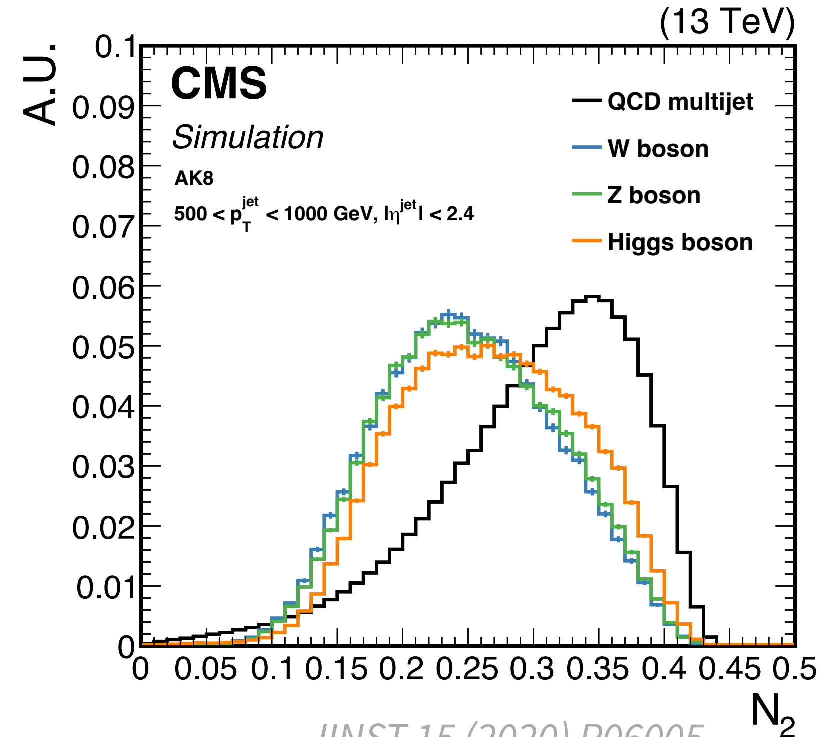
$$e_2^{(\beta)} = \sum_{1\le i<j\le n_J} z_i z_j \,\theta_{ij}^{\beta}\,, \qquad D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}\,,$$
$$e_3^{(\beta)} = \sum_{1\le i<j<k\le n_J} z_i z_j z_k \,\theta_{ij}^{\beta}\theta_{ik}^{\beta}\theta_{jk}^{\beta}\,,$$

JHEP 06 (2013) 108

- ➔ Theory-inspired ***jet substructure variable***
  - ❖ [ATLAS] **energy-correlation function (ECF) ratio**: D_2 (to identify two-prong structure)
    + jet mass (m_J) (trimmed mass for LCTopo jet; soft-drop mass for UFO)
    + # of inner detector track n_{trk}
    ATL-PHYS-PUB-2020-017
    ATL-PHYS-PUB-2021-029
  - ❖ [CMS] **N-subjettiness** variable τ_{21} or N-series of **ECF ratio**: N_2
    + soft-drop jet mass (m_{SD})
    JINST 15 (2020) P06005
  - ❖ hand-crafted variables have highlights in design (e.g. **IRC safety, axis independence**…), but performance cannot reach the multivariate approach

ATL-PHYS-PUB-2020-017





JINST 15 (2020) P06005

W/Z→qq tagging (II): ML with high-level features

➔ *BDT/DNN using high-level jet observables as input*

- [ATLAS] train a ***BDT/feedforward NN*** with jet observables as input
- [CMS] ***BEST***: 59 jet inputs as "boosted event shape", obtained by boosting the jet four times with a resonance assumption

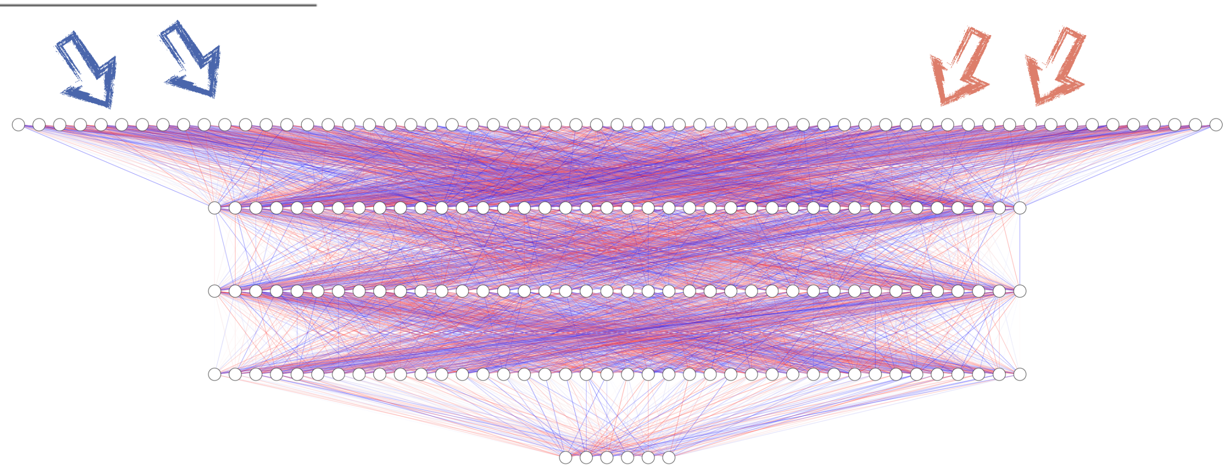
ATL-PHYS-PUB-2021-029

Table 1: List of substructure variables used in the DNN tagger training.

Variable	Description	Reference
D_2, C_2	Energy correlation ratios	[30]
τ_{21}	N-subjettiness	[41]
R_2^{FW}	Fox-Wolfram moment	[42]
$\mathcal{P}$	Planar flow	[43]
a_3	Angularity	[44]
A	Aplanarity	[45]
Z_{cut}, $\sqrt{d_{12}}$	Splitting scales	[33, 46]
$Kt\Delta R$	k_t-subjet ΔR	[47]

BEST algo: JINST 15 (2020) P06005

BEST training quantities		
Jet charge	Fox–Wolfram moment H_1/H_0 (t,W,Z,H)	m_{12} (t,W,Z,H)
Jet η	Fox–Wolfram moment H_2/H_0 (t,W,Z,H)	m_{23} (t,W,Z,H)
Jet τ_{21}	Fox–Wolfram moment H_3/H_0 (t,W,Z,H)	m_{13} (t,W,Z,H)
Jet τ_{32}	Fox–Wolfram moment H_4/H_0 (t,W,Z,H)	m_{1234} (t,W,Z,H)
Jet soft-drop mass	Sphericity (t,W,Z,H)	A_L (t,W,Z,H)
Subjet 1 CSV value	Aplanarity (t,W,Z,H)	
Subjet 2 CSV value	Isotropy (t,W,Z,H)	
Maximum subjet CSV value	Thrust (t,W,Z,H)	

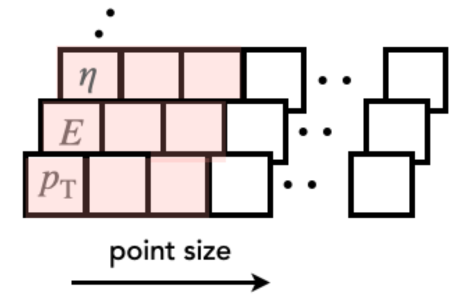


W/Z→qq tagging (III): ML with low-level input

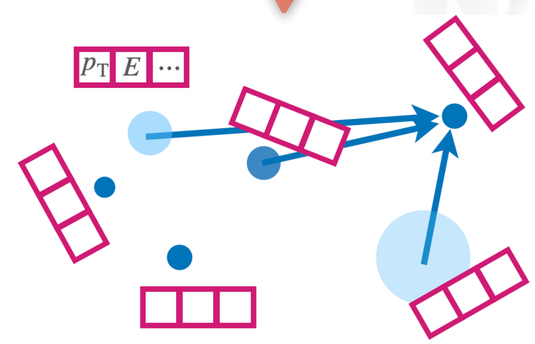
→ ***DNN with low-level constituent inputs***

- Why low-level input?—empowered by recent ML achievements, we are able to explore the full correlation of jet constituents by a network!
- [CMS] ***DeepAK8*** [*JINST 15 (2020) P06005*]
  organize "PF candidates" and "secondary vertices (SV)" as two sequences ▶ input to two 1D CNNs ▶ concatenate, pass to dense layer, output multiple (17) scores (**multi-classification**)
- [CMS] ***ParticleNet*** (**current state-of-the-art in CMS**) [*CMS-DP-2020-002*]
  NEW
  represent PF candidates and SVs in a **point cloud** ▶ use GNN architecture, apply edge convolutions to **exploit geometric features** ▶ output multiple scores

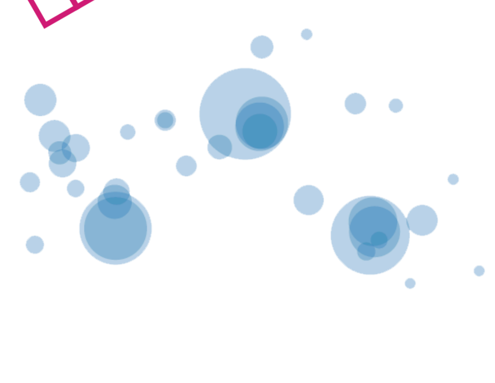
✓ permutational invariant: more effective representation of input data
✓ enable message passing to neighbouring nodes



sequence representation (by some order)



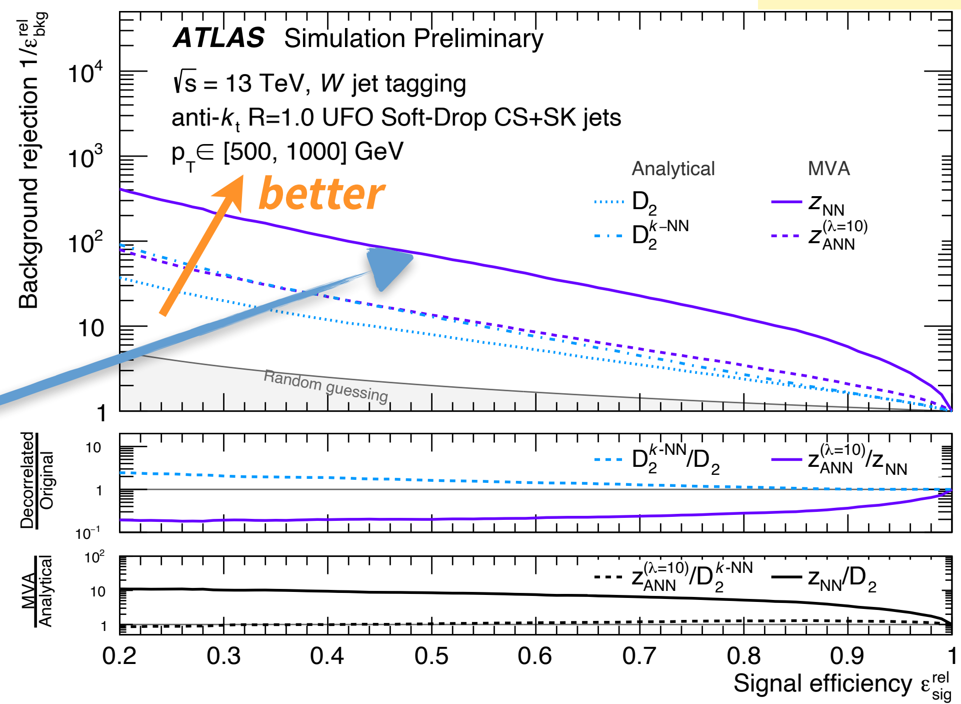
point cloud representation

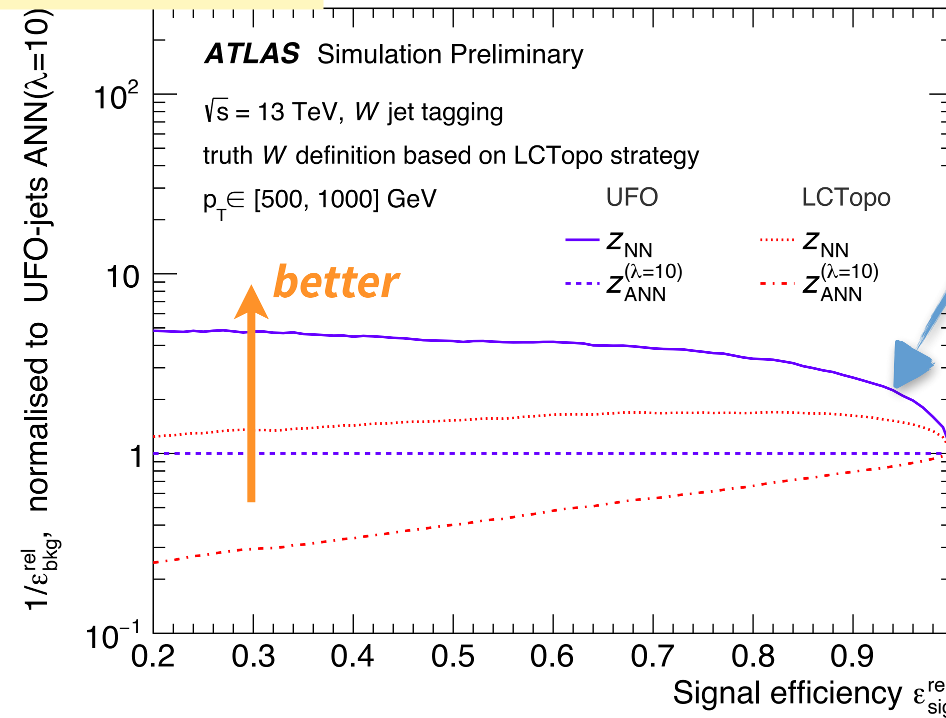


Performance of boosted W→qq taggers

ATLAS: UFO jets performs better then LCTopo jets

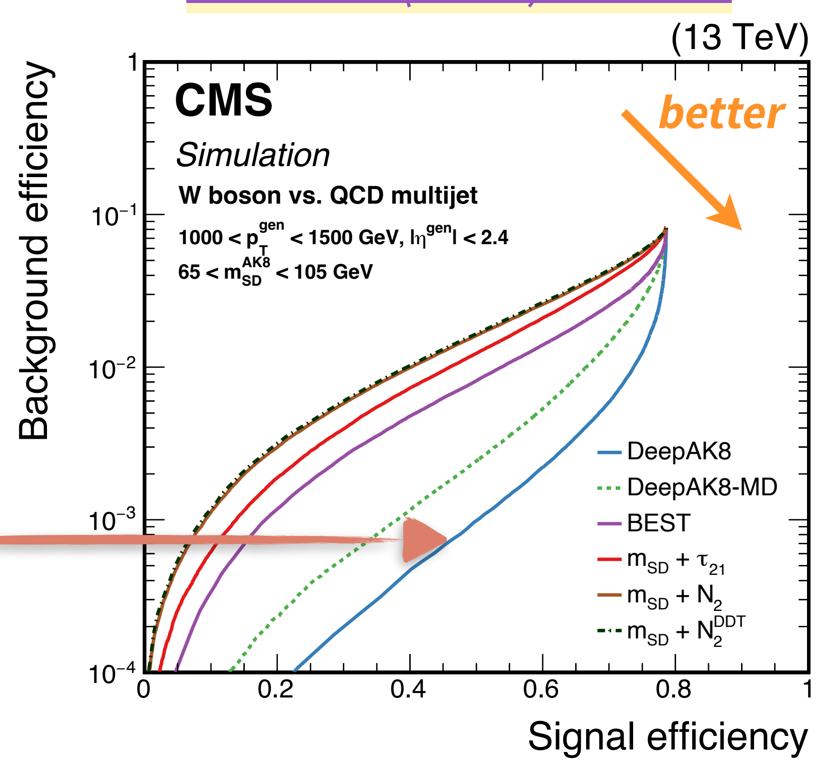
ATL-PHYS-PUB-2021-029



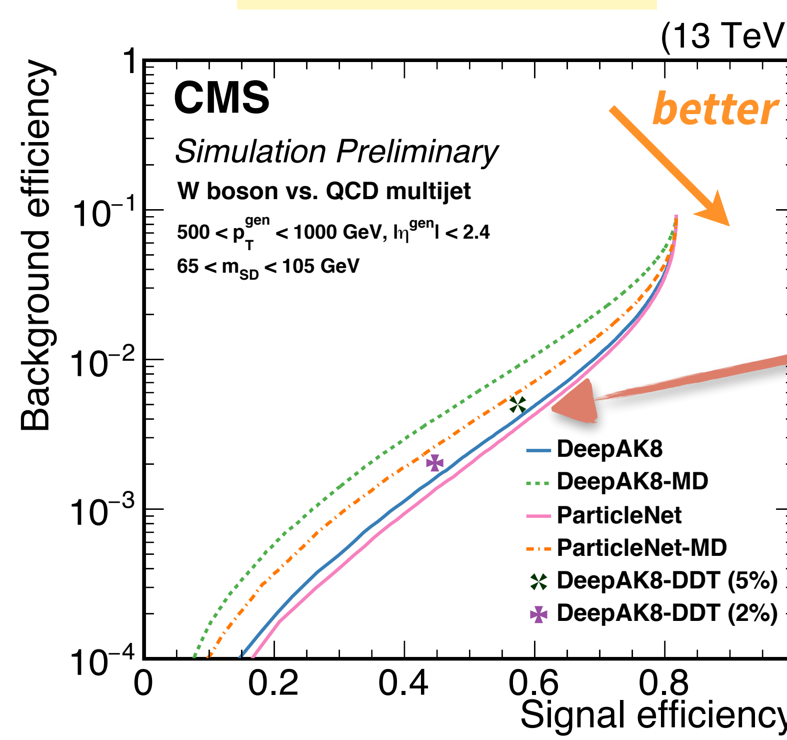


DNN performed on high-level inputs performs the best

JINST 15 (2020) P06005



CMS-DP-2020-002



CMS: benefits of using low-level inputs:
→ DeepAK8 has a huge improvement
→ **x10 BKG rejection** compared to DNN for high-level inputs approach, e.g. BEST (applying no mass decorrelation)

CMS: further improvement in ParticleNet
→ **additional x1.2 BKG rejection**

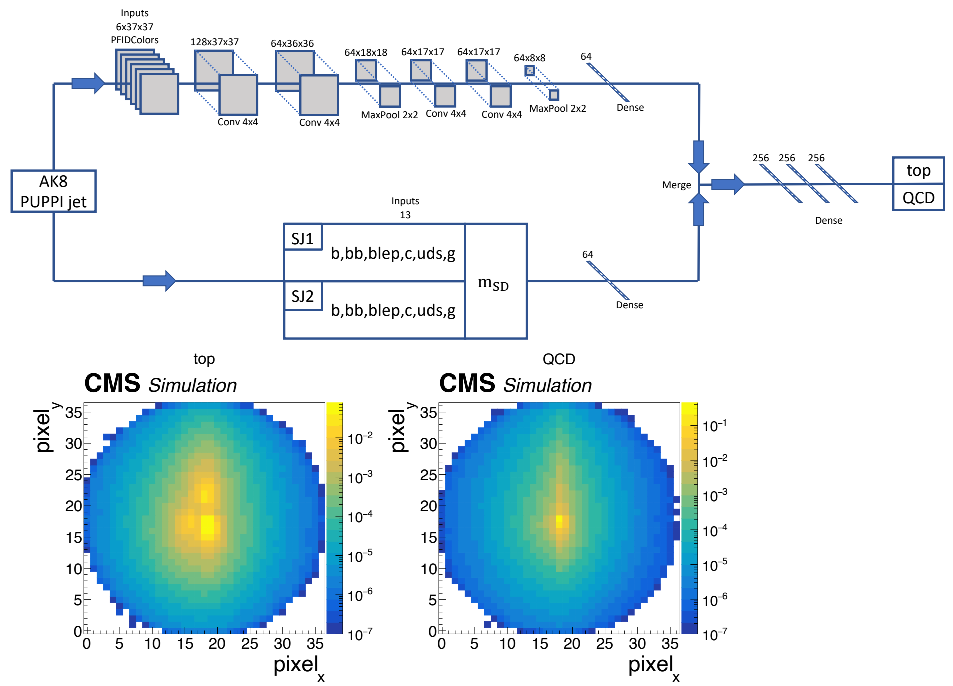
Boosted top tagging

- Hadronically decayed top jets are distinguished by their three-prong structure
- Summarise only the baseline taggers in ATLAS and CMS
  - [ATLAS] ***DNN-based tagger for UFO jets***: pass jet substructure observables as input to feedforward DNN   *ATL-PHYS-PUB-2021-028*
  - [CMS] ***DeepAK8 & ParticleNet***: same tagger for W/Z applies to top tagging as well
    - the model is designed to **output multiple (17) scores** covering W/Z/top/H decay modes   *CMS-DP-2020-002*

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)

- Previous top taggers include   *Eur. Phys. J. C 79 (2019) 375*
  - [ATLAS] ***TopoDNN*** (on LCTopo jets): up to 10 topoclusters with highest p_T as input ▶ feed to feedforward NN ▶ binary classification for top vs. QCD
  - [CMS] ***ImageTop***: create a jet image from PF candidates ▶ feed to 2D CNN (as image recognition task) ▶ also uses a *DeepFlavour* score which passes PF candidates and SVs to 1D CNN+LSTM to infer flavour scores ▶ concatenate and output two scores for top vs. QCD

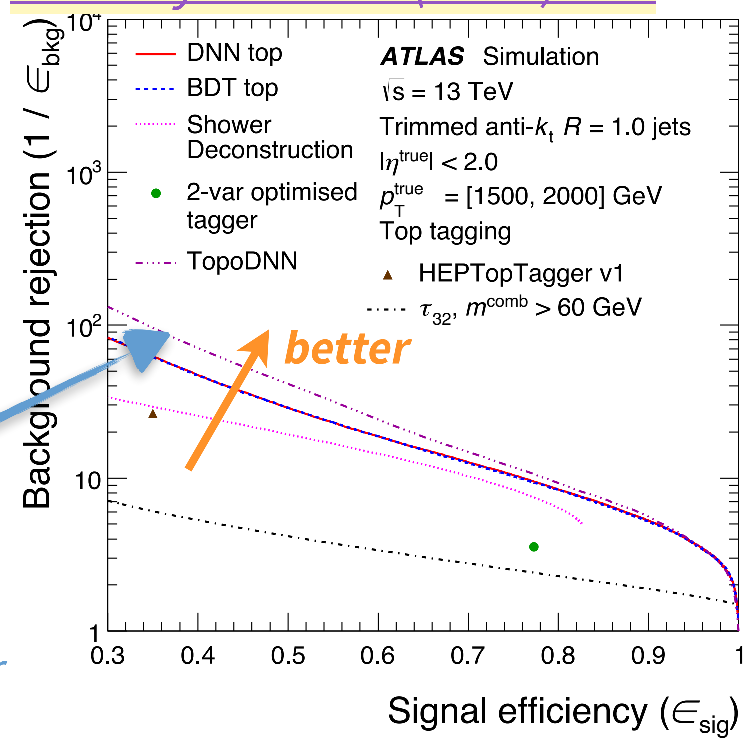


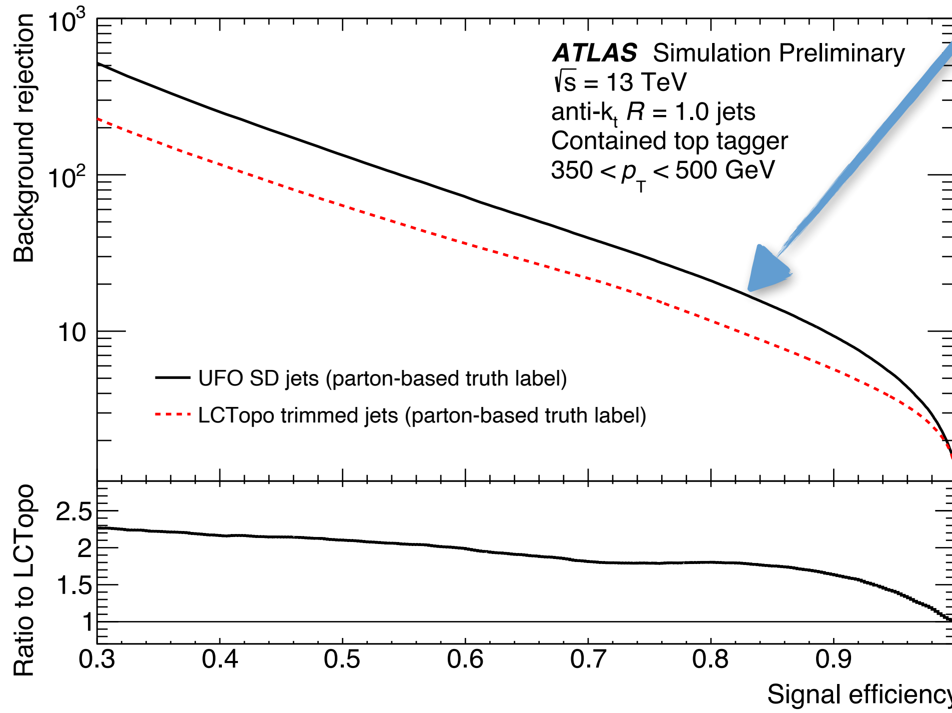
JINST 15 (2020) P06005

Performance of boosted top taggers

ATLAS: improved performance with the use of UFO jets

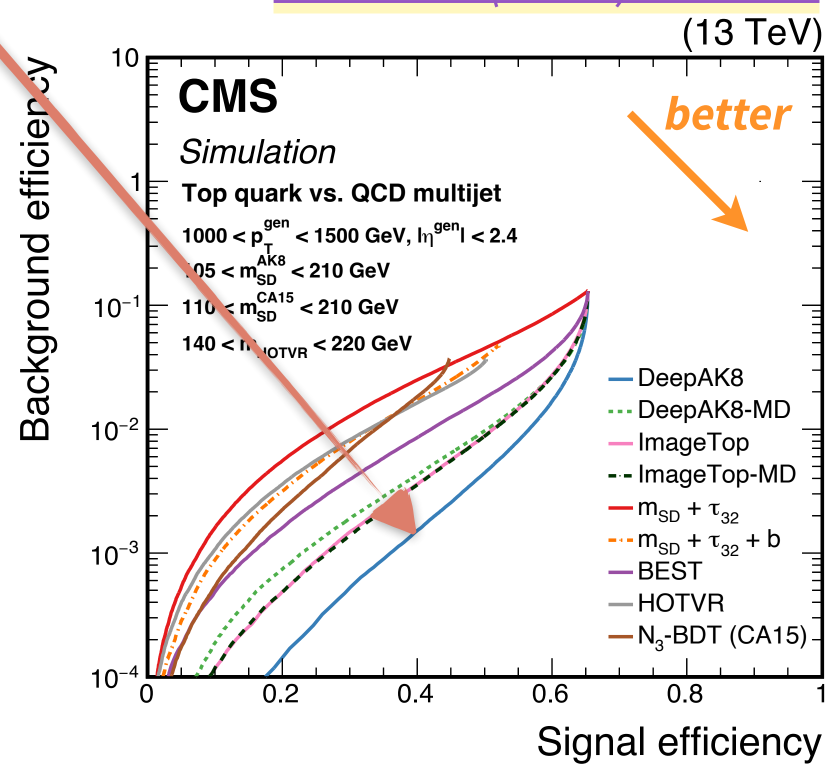
Eur. Phys. J. C 79 (2019) 375



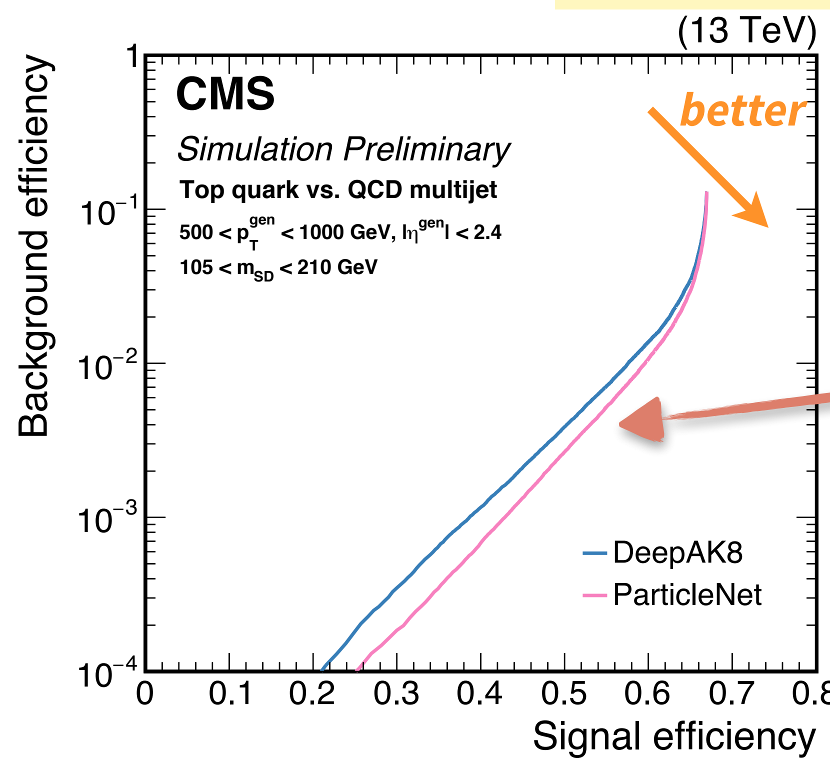


Using low-level inputs in training outperforms all other approaches

JINST 15 (2020) P06005



CMS-DP-2020-002

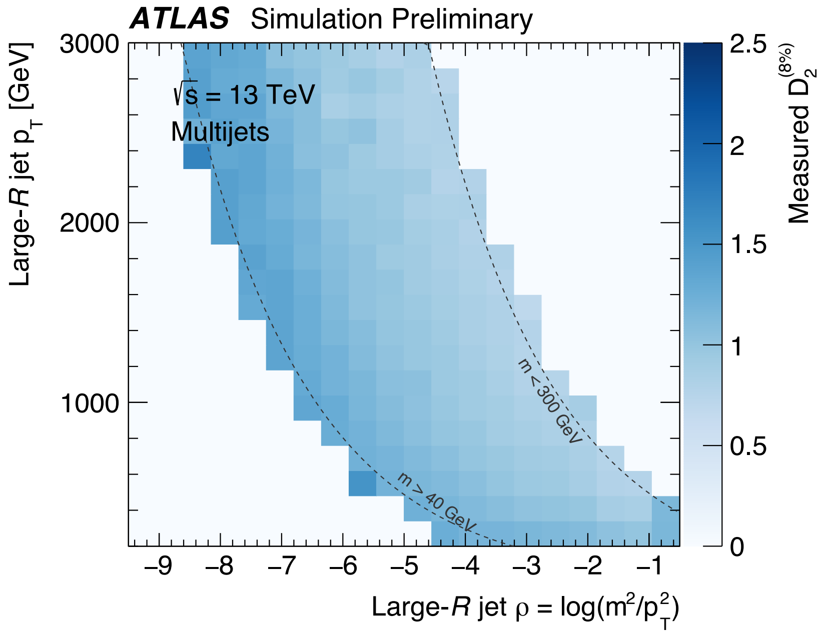


CMS: further improvement in ParticleNet → additional x1.5 BKG rejection

Mass decorrelation (I)

ATLAS: ATL-PHYS-PUB-2021-029
CMS: JINST 15 (2020) P06005

- ➔ Crucial to decorrelate with jet mass
  - ❖ as the DNN would learn from the jet kinematics and "sculpt a peak structure" in the background mass spectrum
- ➔ ***By manual decorrelation***: spirit is to adopt different tagger working points for different bins
  - ❖ [ATLAS] decorrelation of D_2 tagger: define jet bins on $(\rho = \ln(m_{\mathrm{SD}}^2/p_{\mathrm{T}}^2), p_{\mathrm{T}})$ ▶ manual bin-dependent working point $D_2^{8\%}$ at BKG eff = 8% ▶ define new tagger $D_2^{k-NN} = D_2 - D_2^{8\%}$
  - ❖ [CMS] same method, **denoted as "designed decorrelated tagger (DDT)"**: e.g. $N_2^{\mathrm{DDT}}(\rho, p_{\mathrm{T}}) = N_2(\rho, p_{\mathrm{T}}) - N_2^{5\%}(\rho, p_{\mathrm{T}})$

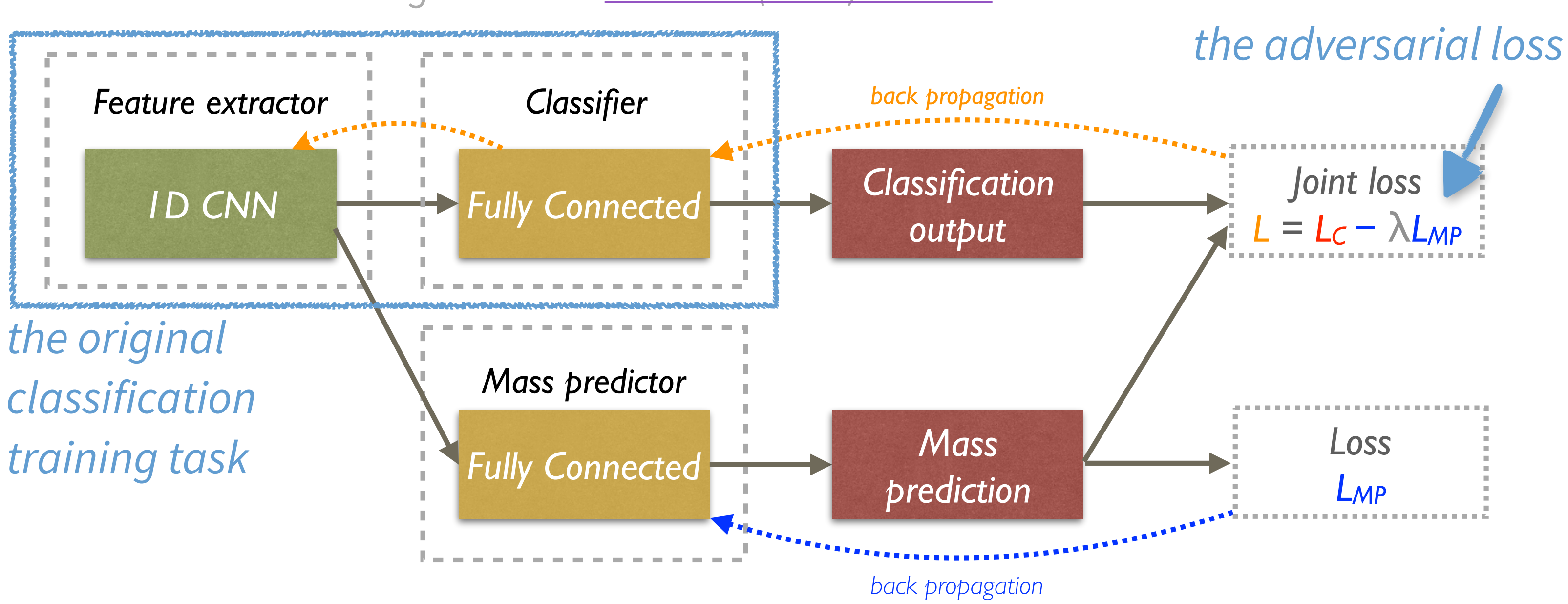


ATLAS: measured working point $D_2^{8\%}$ ATL-PHYS-PUB-2021-029

NEW

- ➔ ***By adversarial training***
  - ❖ [ATLAS] decorrelate the DNN score with mass by **adding an additional adversarial network** which contributes an adversarial loss
  - ❖ [CMS] same method adopted for DeepAK8 tagger (denoted DeepAK8-MD)

CMS: adversarial training workflow JINST 15 (2020) P06005



Mass decorrelation (II)

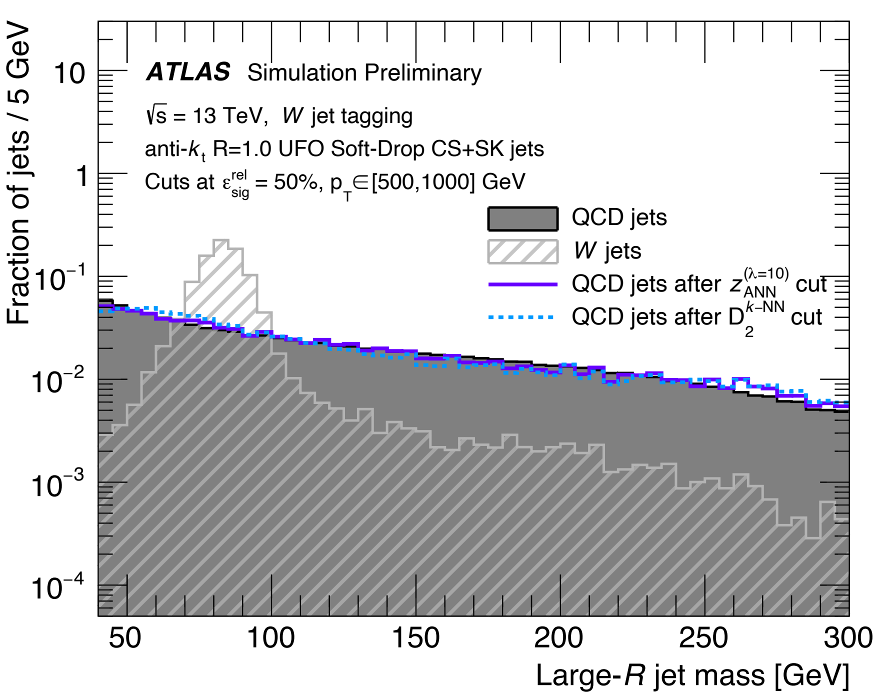
NEW → ***By training with flat-mass sample***

- [CMS] mass decorrelation approach for ***ParticleNet-MD***:
  construct X→bb/cc/qq sample for training: X=spin-0 scalar with variable-mass ▶ dedicated reweighting on $(p_\mathrm{T}, m_\mathrm{SD})$ from signal to QCD jets ▶ training performed on same ParticleNet model
  - **fewer performance loss w.r.t. adversarial training approach**

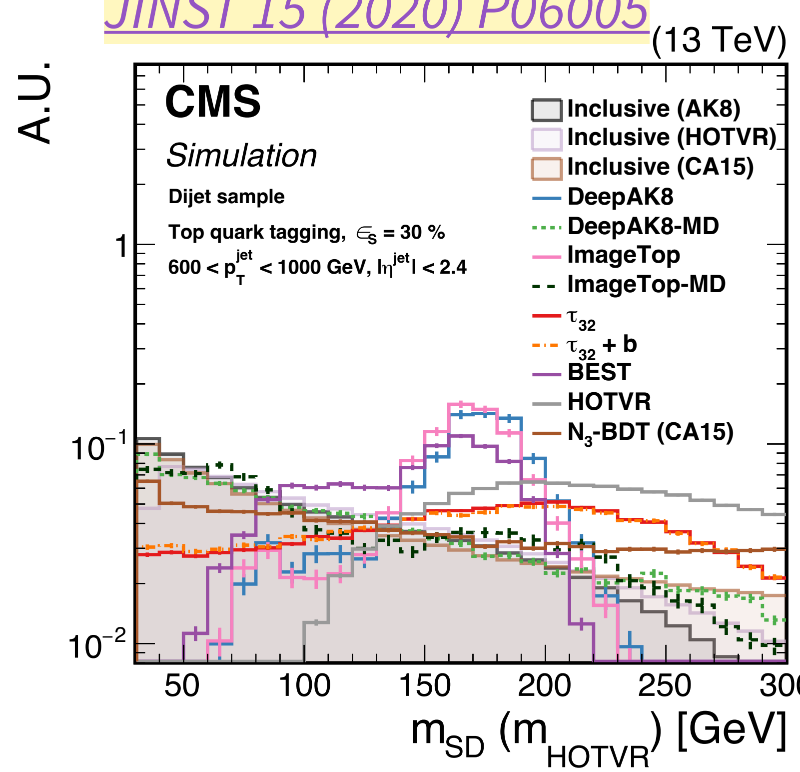
Performance

flat mass shape after the MD tagger selection

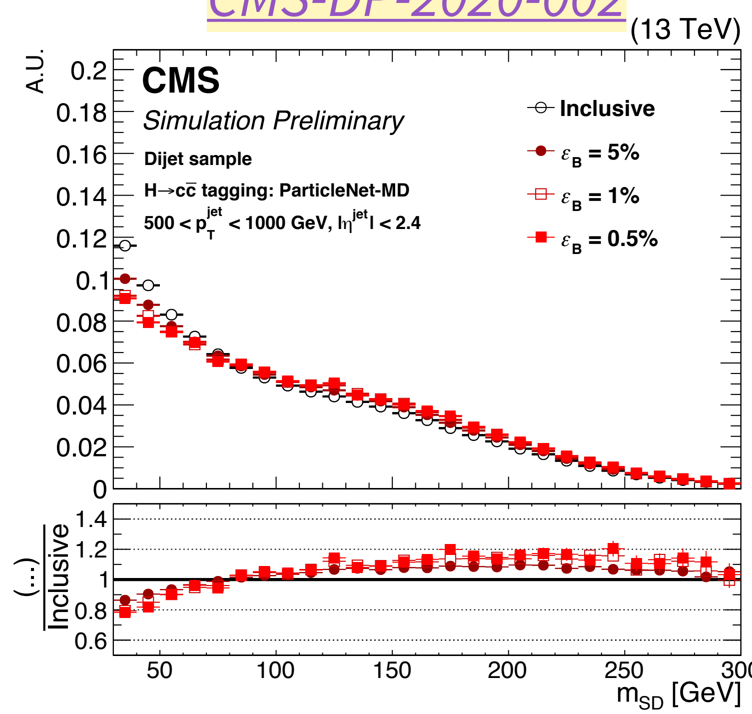
ATL-PHYS-PUB-2021-029



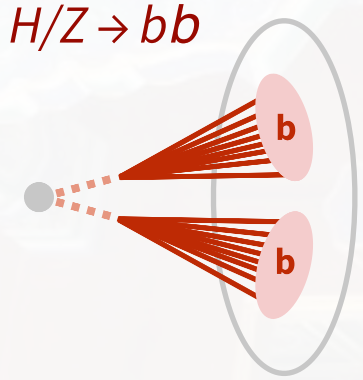
JINST 15 (2020) P06005

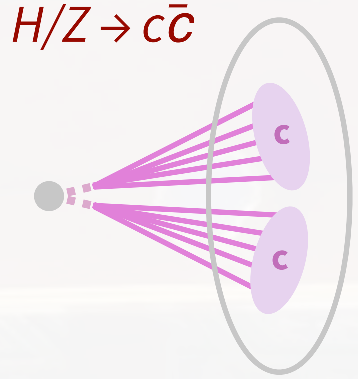


CMS-DP-2020-002



Heavy flavour tagging

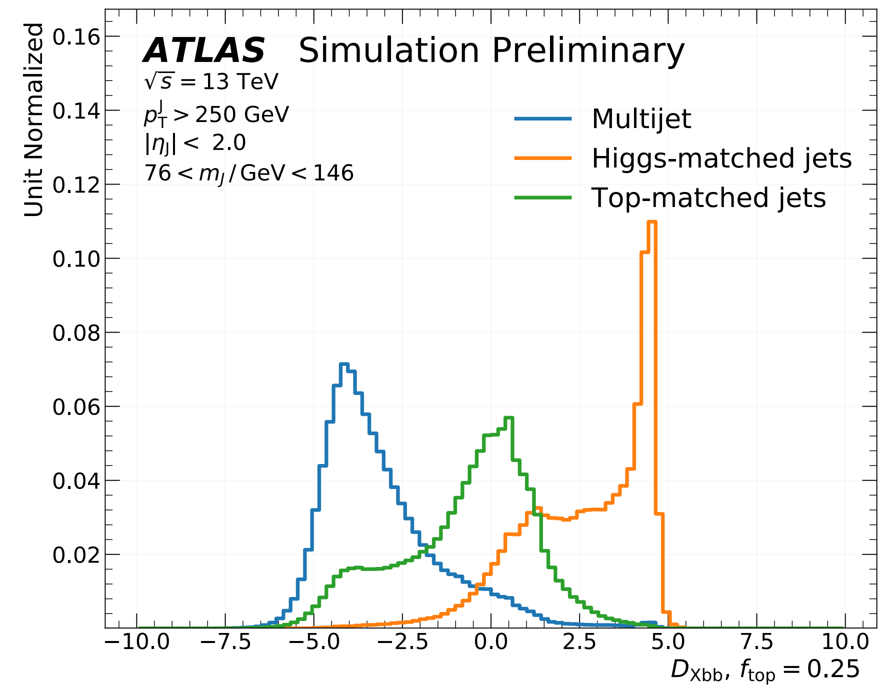




X→bb̄/cc̄ tagging

- ➔ Double-b/c flavour tagging techniques are crucial to recover sensitivity in boosted X→bb̄/cc̄ phase-space (X=H/Z/BSM particles)
  - only includes recent advanced developments

ATL-PHYS-PUB-2020-019



NEW

- ➔ [ATLAS] ***double b-tagger*** (for LCTopo jets associated to up to 3 variable-radius (VR) track-jets): use flavour tagging info ***DL1r*** of 3 track-jets + jet kinematics ▶ feedforward NN ▶ produce three scores: p(Higgs), p(multijet), p(top) ATL-PHYS-PUB-2020-019
  - N.B. ***DL1r***: track inputs passed to feedforward NN to output three scores p(b), p(c), p(light) ATL-PHYS-PUB-2017-013
  - final score $D_{\mathrm{Xbb}} = \ln \dfrac{p_{\mathrm{Higgs}}}{f_{\mathrm{top}} \cdot p_{\mathrm{top}} + (1 - f_{\mathrm{top}}) \cdot p_{\mathrm{multijet}}}$
- ➔ [CMS] ***DeepAK8-MD***: as detailed, flavour category also included (H→bb/cc/qq scores)

NEW

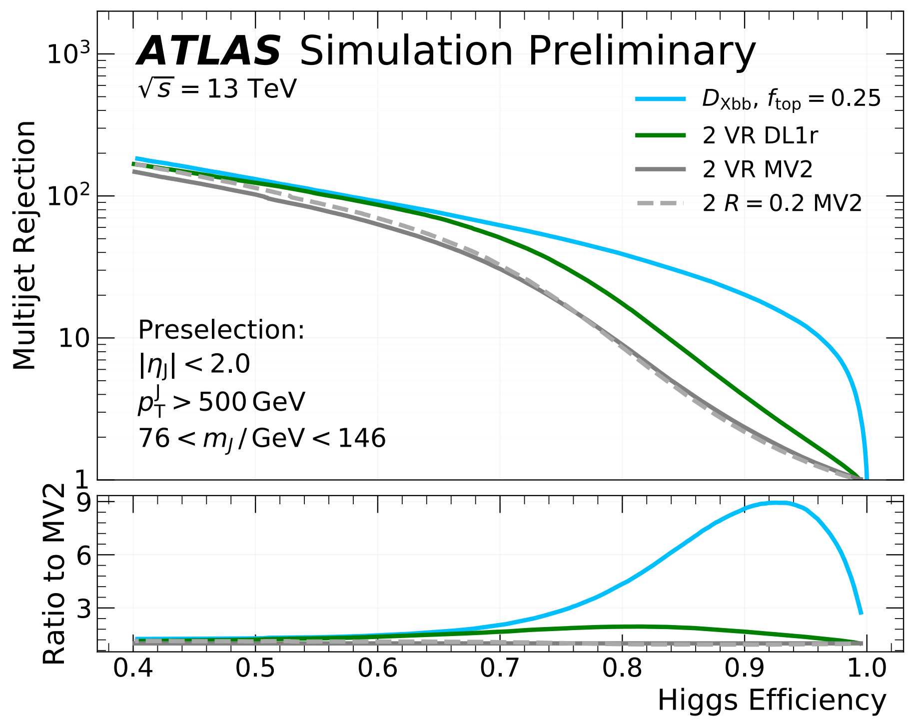
- ➔ [CMS] ***DeepDoubleX***: PF candidates, SVs (organised as sequences) and jet-level inputs ▶ 1D CNN+GRU ▶ two scores in 3 schemes (BvsL, CvsL, CvsB) CMS-DP-2018-046 *(for v1 tagger)*

NEW

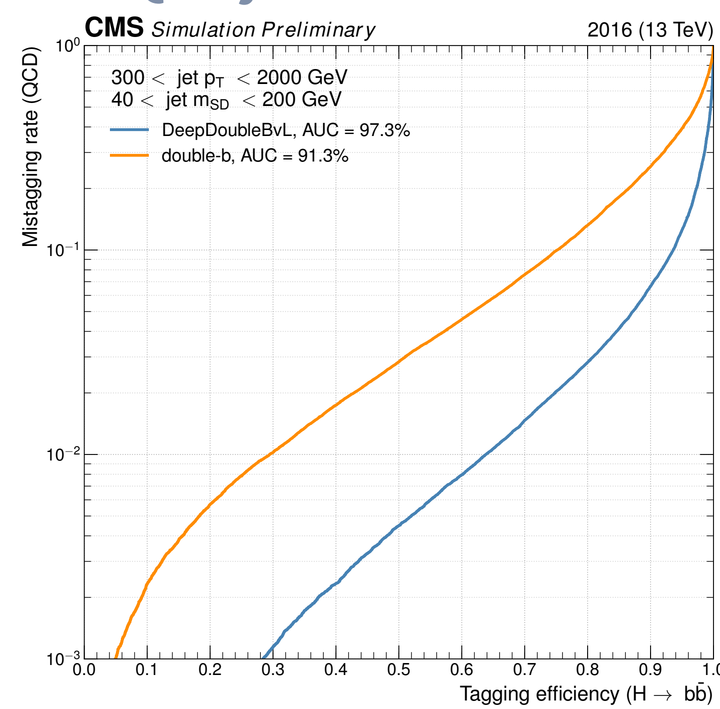
- ➔ [CMS] ***ParticleNet-MD***: reweight variable-mass Higgs signal & QCD backgrounds ▶ use PF candidates and SV inputs as point cloud ▶ GNN with edge convolution ▶ X→bb/cc/qq scores and 5 QCD scores CMS-DP-2020-002

Performance of X→$b\bar{b}$/$c\bar{c}$ taggers

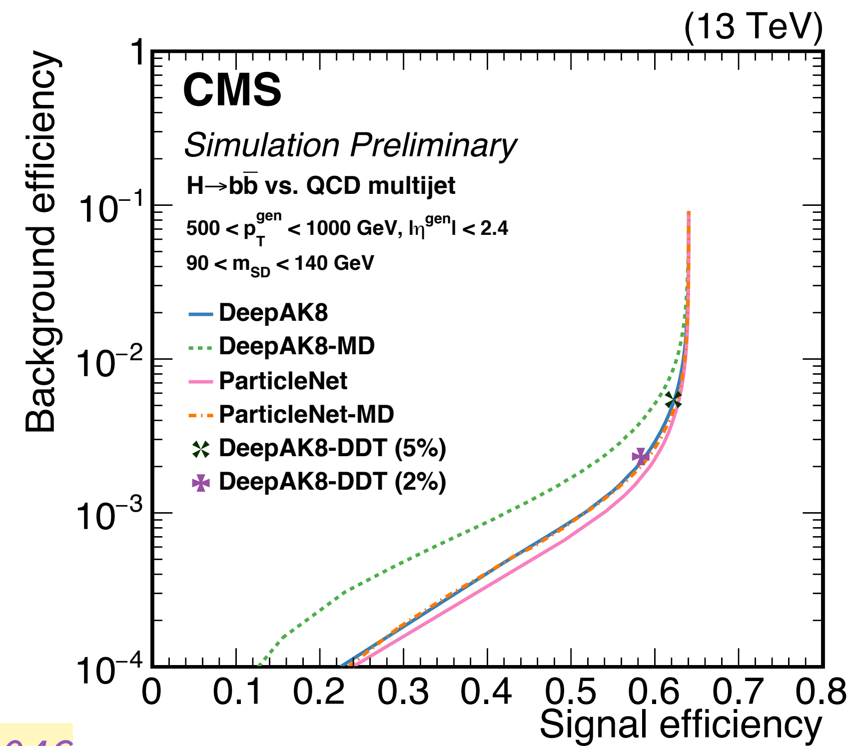
H→$b\bar{b}$ vs QCD jet discrimination



[ATL-PHYS-PUB-2020-019]

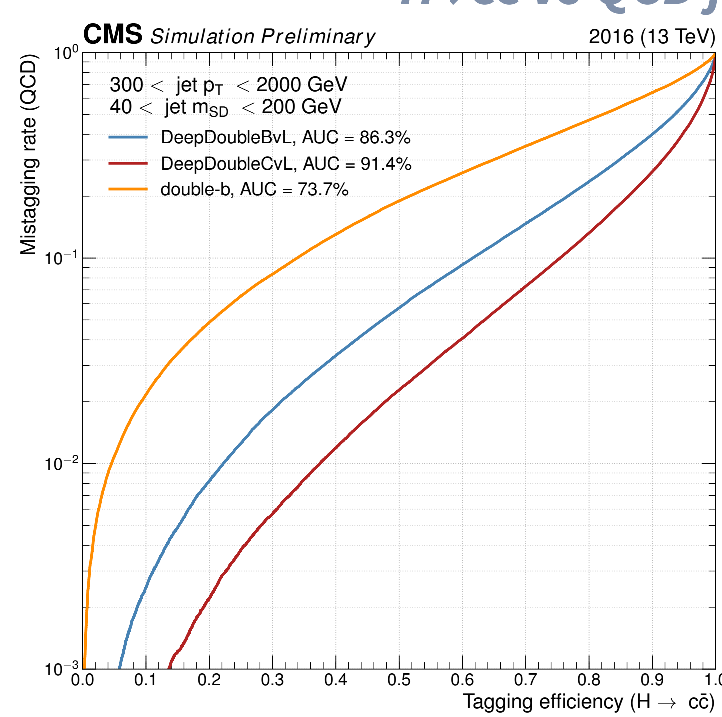


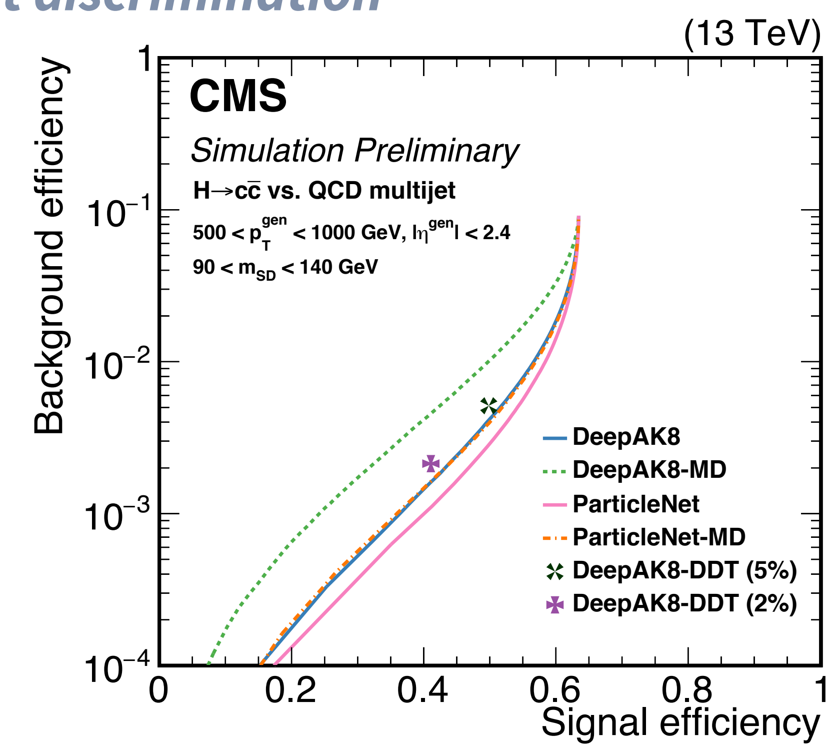
[CMS-DP-2018-046]



[CMS-DP-2020-002]

H→$c\bar{c}$ vs QCD jet discrimination





Applications and beyond

Calibration of boosted W/top taggers

➔ Deriving scale factors (SF) on tagging efficiency crucial in the real application

❖ $SF = \epsilon_{\text{data}}/\epsilon_{\text{MC}}$, i.e., ratio of the tagger efficiency passing a specific working point between data and MC, usually binned by p_T

➔ ***hadronic top/W taggers calibrated with $t\bar{t}$ events***

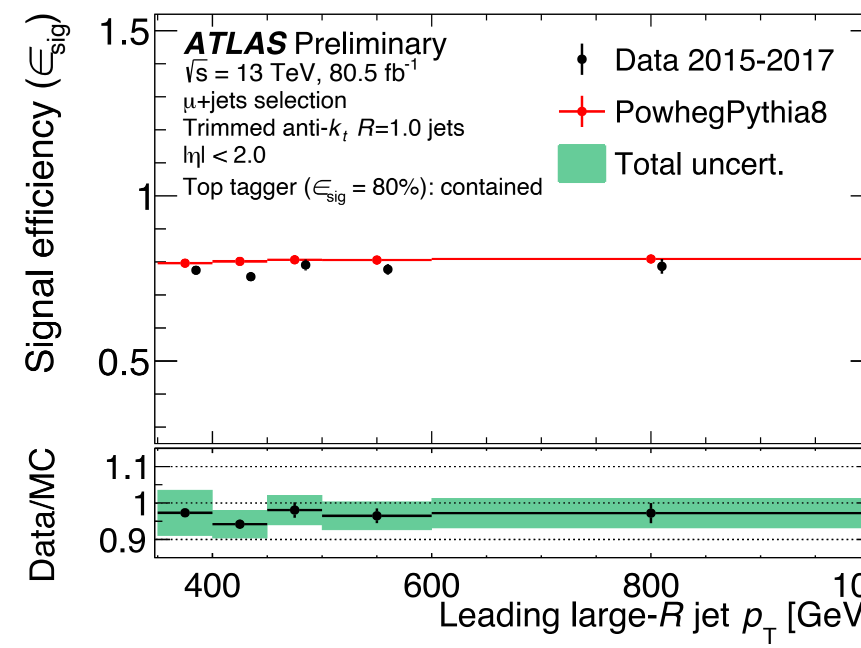
ATL-PHYS-PUB-2020-017

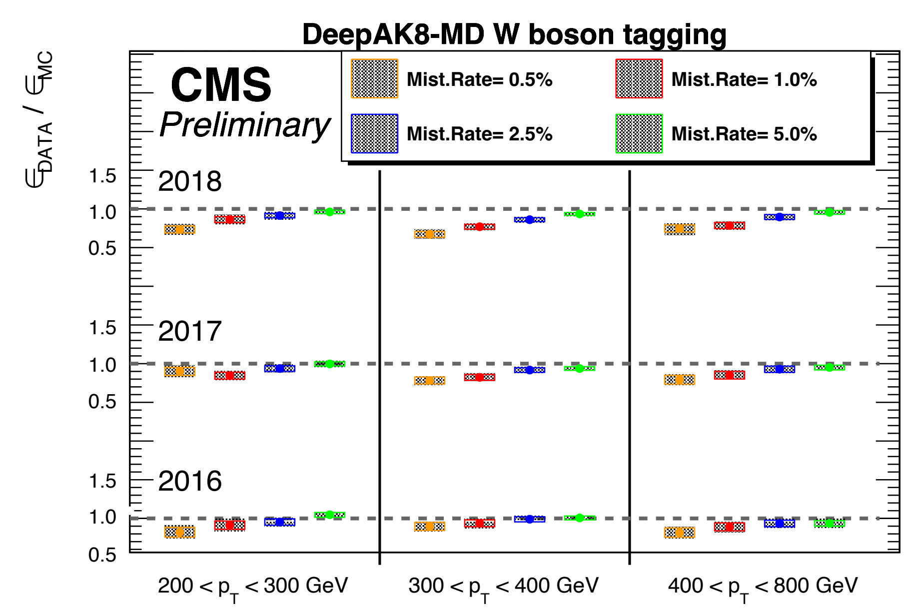
❖ [ATLAS] (separate for top/W tagging)
→ decompose MC jets into **"$t\bar{t}$ top-matched", "$t\bar{t}$ top-unmatched", "others"** (for top tagging) or **"$t\bar{t}$ W-matched", "others"** (for W tagging)
▶️ **simultaneous fit** on mass for pass/fail tagger region

- extrapolate SF with its uncertainties to higher p_T
- calibrate the BKG tagging efficiency (rejection) for QCD/γ+jet events as well

❖ [CMS] similar method: categorize MC jets to **"top-matched", "W-matched", "others"** and apply simultaneous fit

[CMS-DP-2020-025](CMS-DP-2020-025)





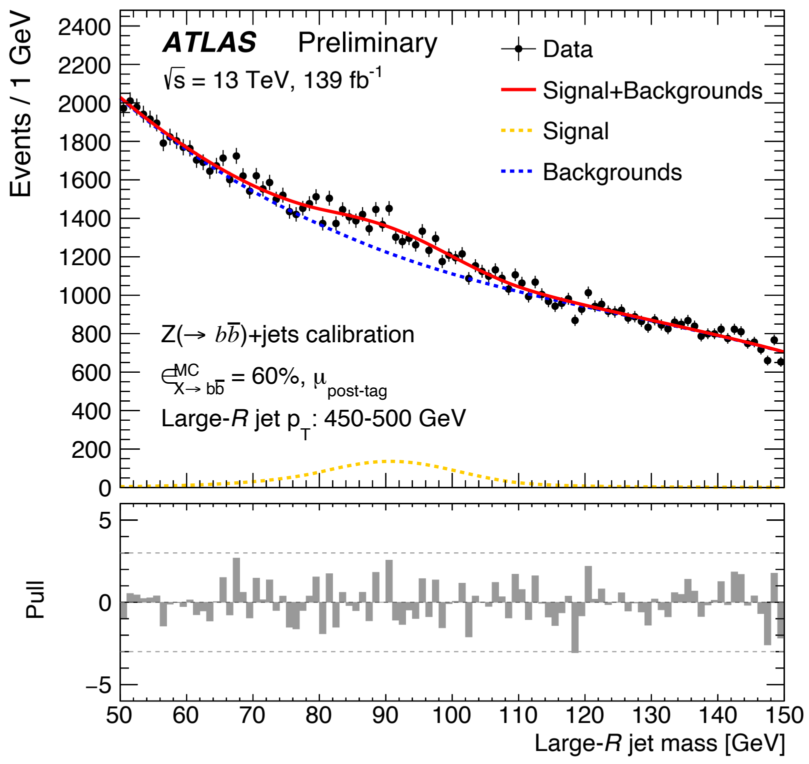
Calibration of boosted flavour taggers

➔ ***hadronic X→b$\bar{b}$/c$\bar{c}$ taggers calibrated with "proxy"***

ATL-PHYS-PUB-2021-035

NEW

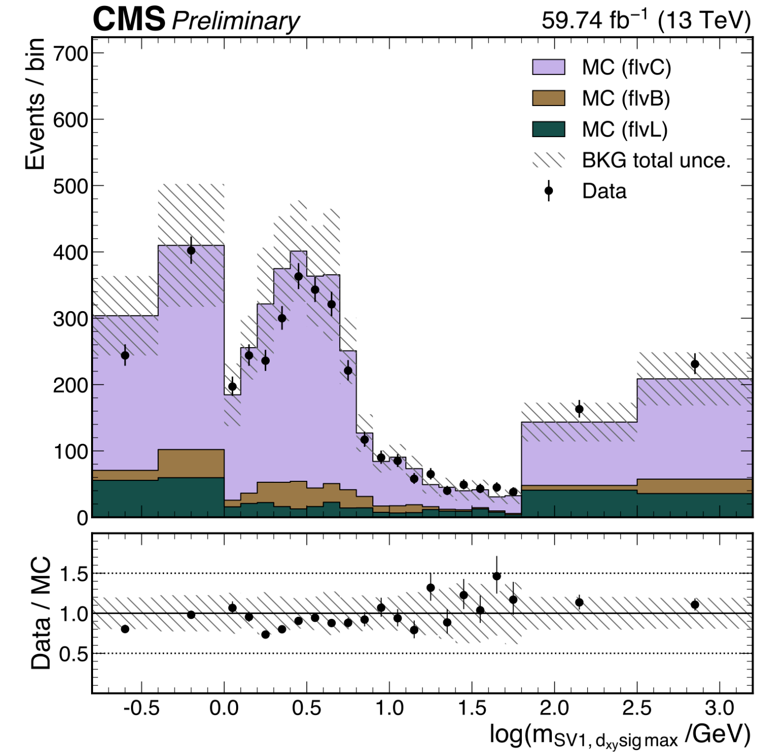
- [ATLAS] use **Z→b$\bar{b}$** jets as a proxy to H→b$\bar{b}$ jets
  Z→b$\bar{b}$ events with additional γ or jet ▶ data-driven estimation of QCD/γ+jet shape from mass sideband ▶ simultaneous fit on mass for pass & fail tagger region



NEW

- [CMS] use **"BDT selected g→b$\bar{b}$/c$\bar{c}$ jets"** as a proxy to H→b$\bar{b}$/c$\bar{c}$
  QCD jets categorised to b, c, light flavour ▶ simultaneous fit on $\ln(m_{SV})$ for pass & fail tagger region
  - BDT trained on QCD jets to veto jets with large gluon contamination, so as to select more H→b$\bar{b}$/c$\bar{c}$-like jets

CMS-DP-2022-005

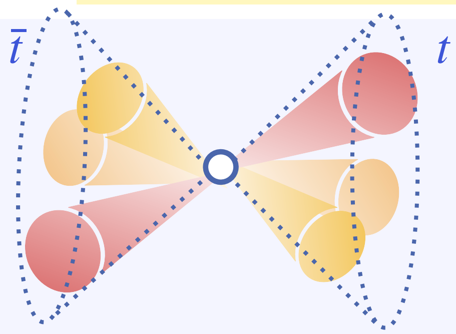


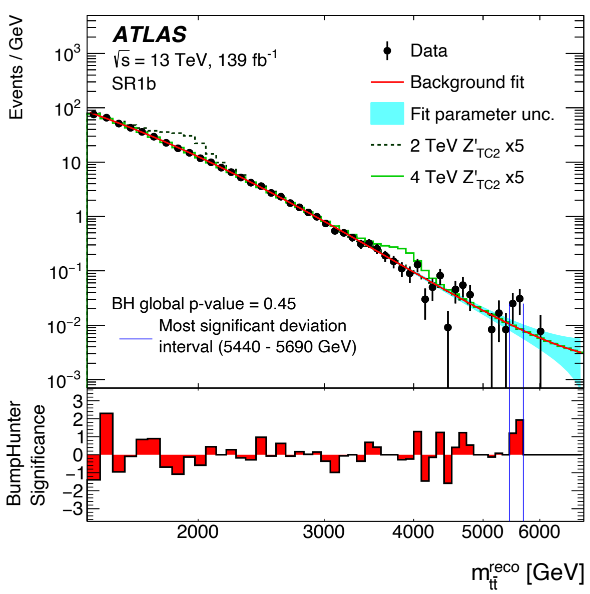
Applications

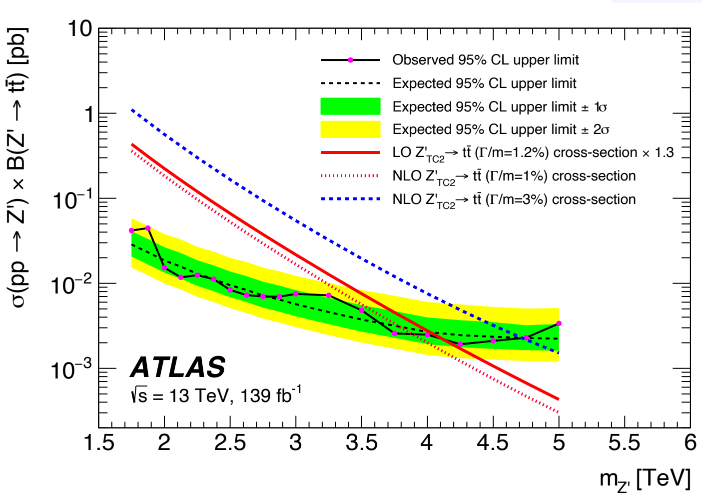
➔ ***Highlight only a few from many recent analyses*** that benefit from the advanced boosted tagging techniques

JHEP 10 (2020) 061

Resonance $t\bar{t}$ search in fully hadronic mode



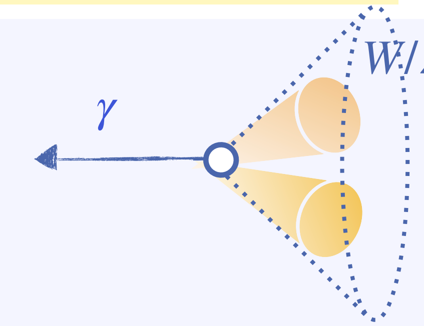


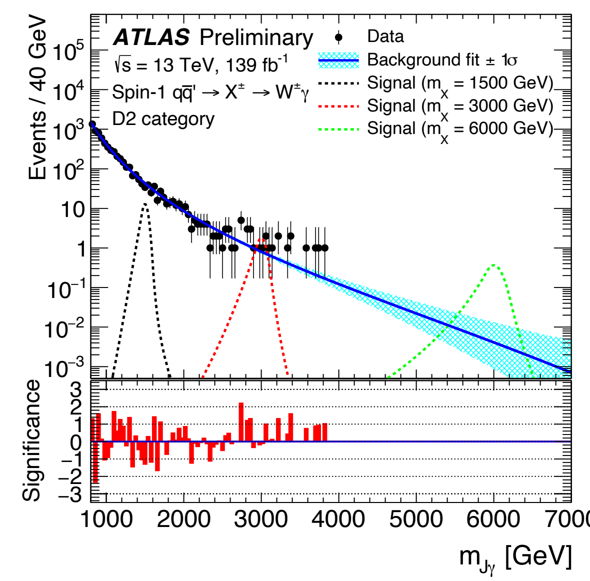


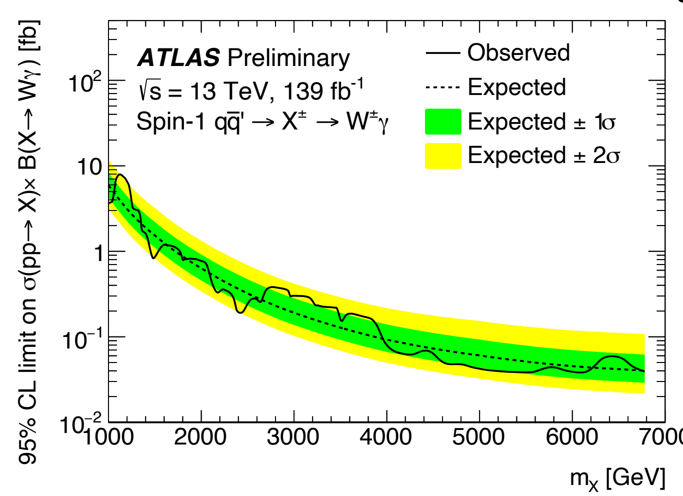
- boosted top tagging with DNN trained on jet-level features + *DL1* b-tagging on VR track-jets → ***improve σ limit by 65%!***
- smooth background template estimated from data
  - Z'($\to t\bar{t}$) mass excluded up to 3.9 TeV for decay width=1%

ATLAS-CONF-2021-041

Resonance Wγ/Zγ search







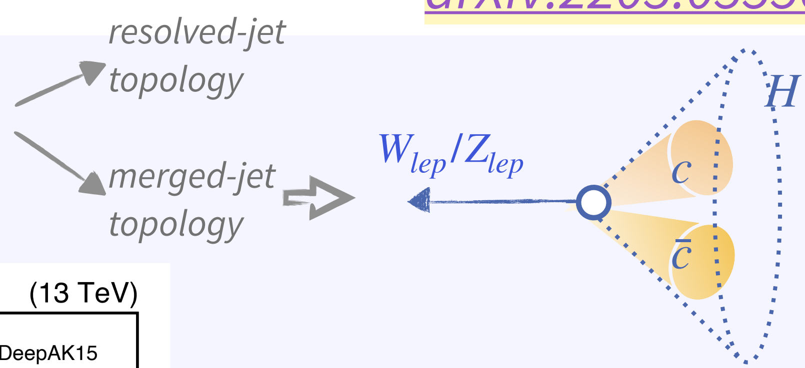
- boosted W/Z tagging with D_2 variable + b-tagging on VR track-jets (for $Z\to b\bar{b}$)
- data-driven BKG modelling
- results: upper limit on σ: 10–0.05 fb in the range 1.0–6.8 TeV

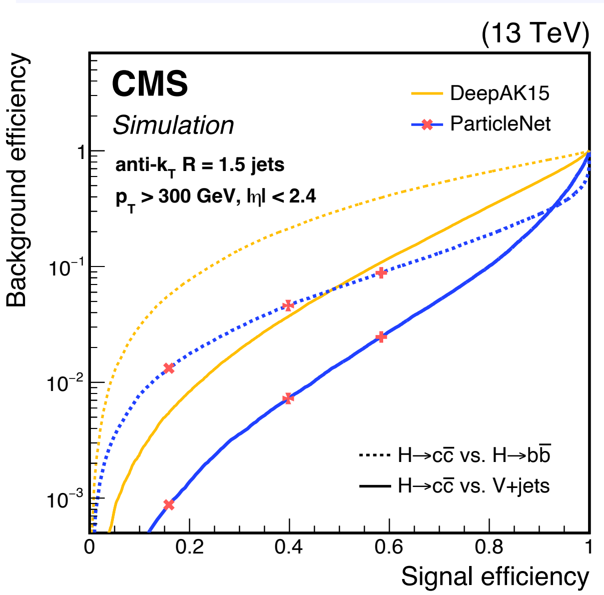
Applications (II)

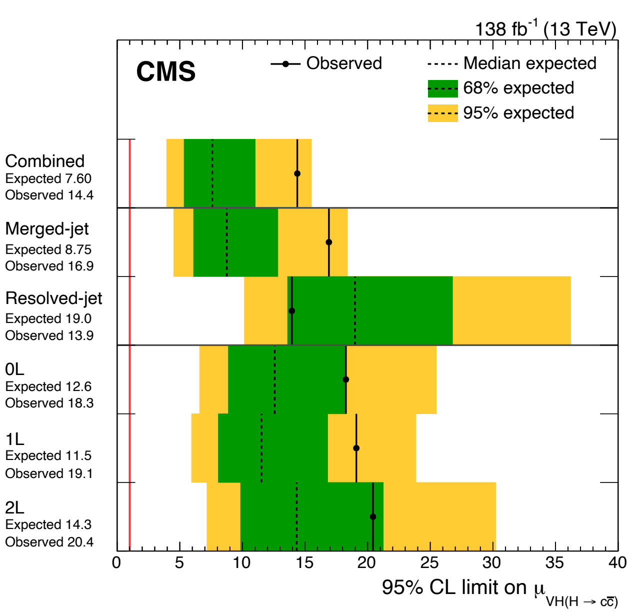
➔ ***Highlight only a few from many recent analyses*** that benefit from the advanced boosted tagging techniques

arXiv:2205.05550

VH→cc search



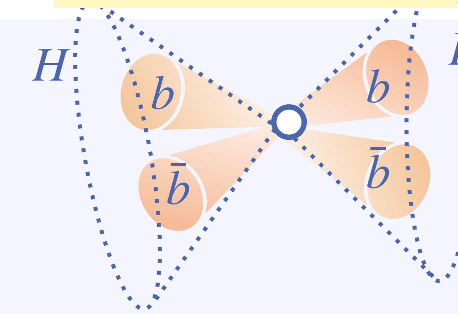




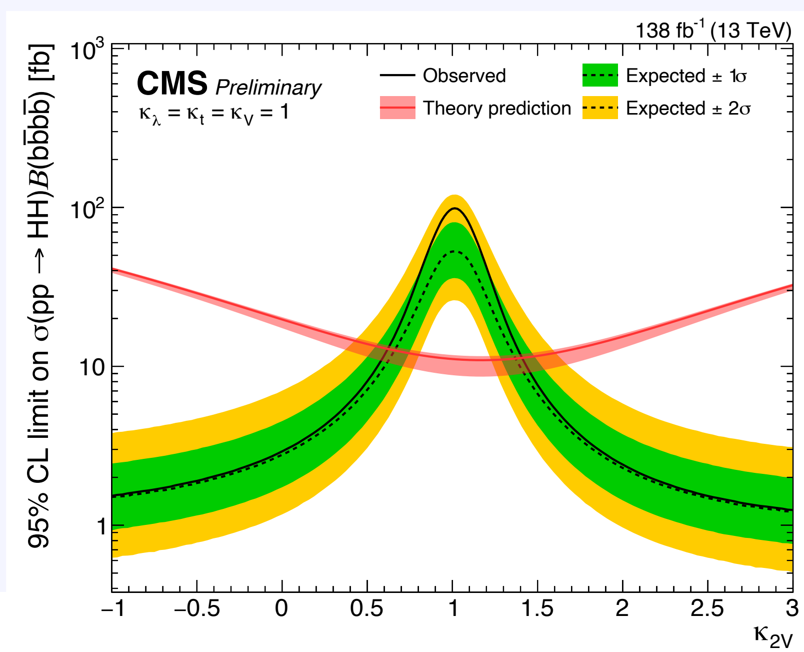
- Boosted H→cc̄ jet jet tagged by ParticleNet-MD → ***x5 improvement*** in BKG (QCD & V+jets) rejection!
- Fit on "jet mass" (merged topology) and an event BDT variable (resolved topology)
- Most stringent limit on H-c coupling to date: $1.1<|\kappa_c|<5.5$

arXiv:2205.06667

Boosted HH→4b search



- H→bb̄ jet jet tagged by ParticleNet-MD → ***x2 improvement*** in BKG rejection
- Regression on H→bb̄ jet mass based on ParticleNet → ***40% improvement*** in resolution
- Most stringent limit on κ_{2V} to date: $0.6<\kappa_{2V}<1.4$



More studies and beyond

- ➔ More developments in boost algorithms
  - ❖ boosted di-τ tagging — JHEP 11 (2020) 163
  - ❖ boosted di-gluon tagging — ATL-PHYS-PUB-2021-027
  - ❖ boosted jet mass regression — CMS-DP-2021-017
  - ❖ application of DNN-based boost tagging to trigger-level, …
- ➔ ***Where to seek for more improvement for future taggers?***
  - ❖ learn from known ML experiences which bring benefits: training with low-level inputs, end-to-end training & optimisation, multi-classification
  - ❖ cooperate with physics inspiration—latest pheno studies post interests on: jet symmetries [Shimmin. arXiv:2107.02908; Gong et al. 2201.08187; Murnane et al. 2202.06941], pairwise features [Qu et al. 2202.03772], …
  - ❖ borrow new advancements from ML: GNN/Transformer-based model [Qu et al. 2202.03772], training data engineering, …

Summary & outlook

- ➔ ***Recent advances in boost algorithms start to impose huge impact on analyses at LHC***
  - ❖ ATLAS and CMS explore new possibilities in the boosted phase-space
    - in context of **W/Z/top/H resonance tagging**, and/or with **flavour contents**
  - ❖ novel ML approaches greatly improve the sensitivity
    - developing path: **single/few rule-based jet observables → "shallow ML" using jet inputs → directly using low-level input to train deep NN**
    - results in more precise SM measurements, more stringent limit; or **even accelerate the finding of a new particle!**
  - ❖ correction of performance between data and MC still tractable
- ➔ ***…while facing new challenges in future developments***
  - ❖ model training will be more data thirsty
  - ❖ real deployment requires fast/on-the-fly tagger inference
  - ❖ eager for more precise and robust calibration methods
- ➔ ***Long but optimistic journey ahead!***

Backup

ParticleNet: details

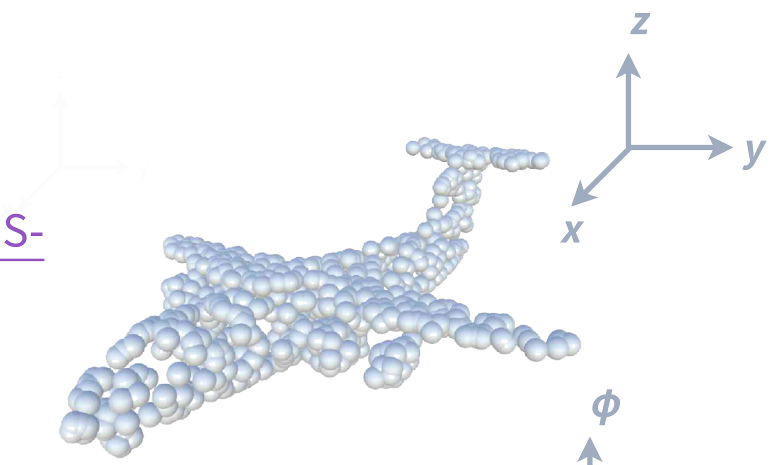
➔ ***ParticleNet***: A multi-class jet classifier for t/H/W/Z tagging based on graph NN [[Phys.Rev.D 101, 056019 (2020)](#)]

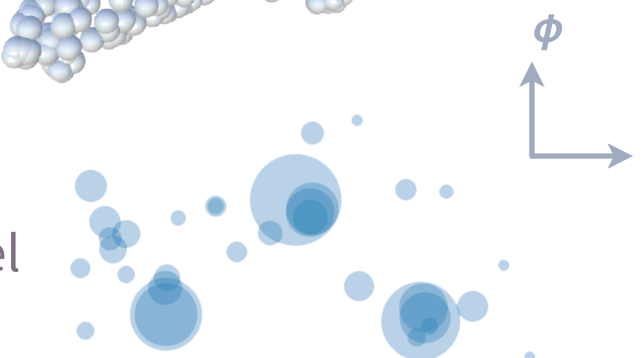
- ❖ achieve ***state-of-the-art performance*** for large-*R* jet tagging at CMS [[CMS-DP-2020-002](#)]

➔ Architecture:

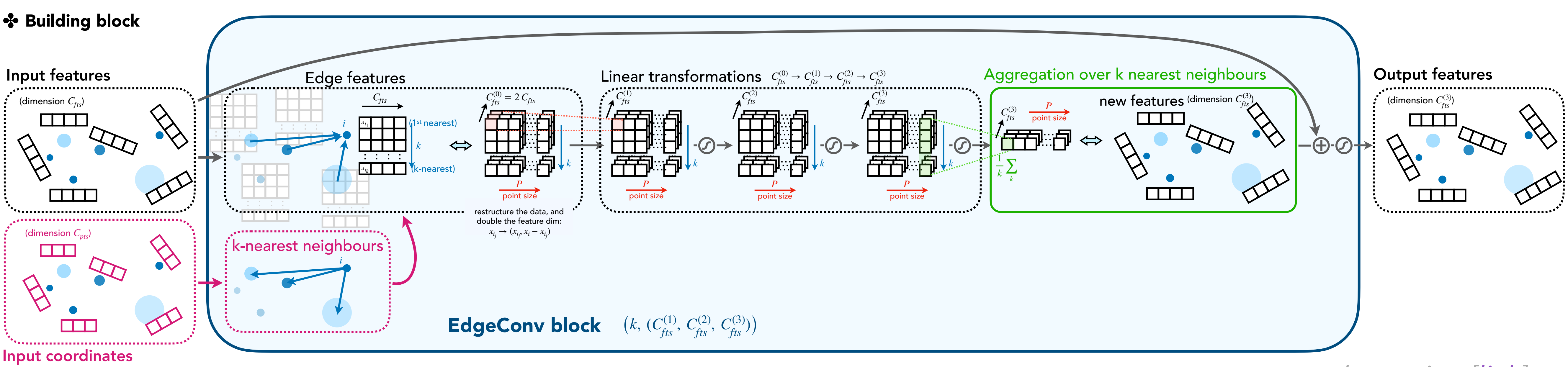
- ❖ treat a jet as an ***unordered set of particles*** in the η–ϕ space
- ❖ use graph NN that maintains the ***permutation-invariant symmetry:*** model based on Dynamic Graph CNN (DGCNN) architecture with EdgeConv operation

➔ Input: low-level features of PF candidates / SVs





typical top jet image as a "point cloud"

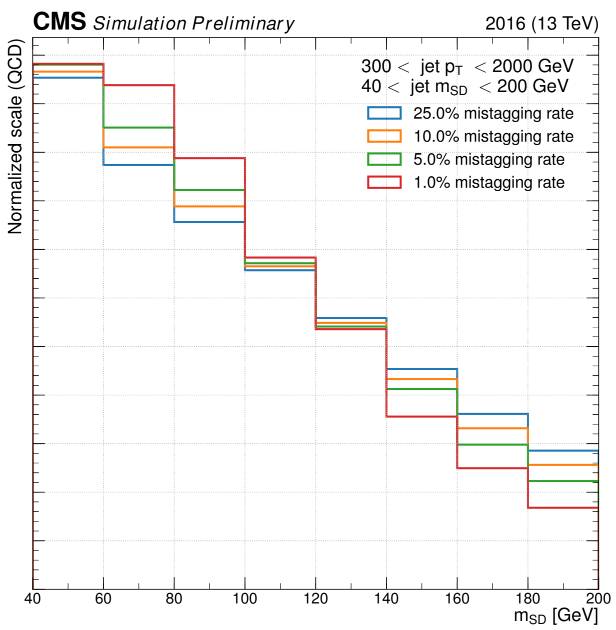


network overview [[link](#)]

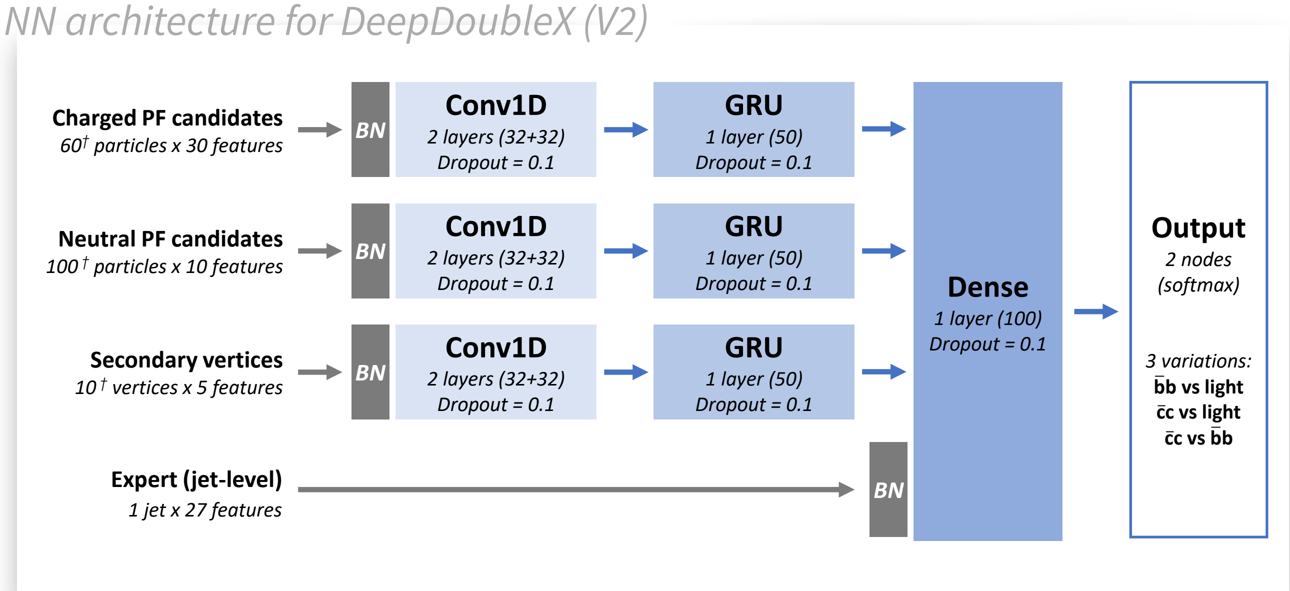
DeepDoubleX(-MD): details

- ➔ ***DeepDoubleX*** (V1): a bb/cc-flavour tagger based on 1D CNN+GRU [CMS-DP-2018-046]
  - ❖ NN similar with DeepJet (for R=0.4 jet tagging) architecture [JINST 15 (2020) P12012]
  - ❖ **MD version**: introduce additional "adversarial loss" in training: use KL divergence to quantify the shape difference
- ➔ Architecture:
  - ❖ separate 1D CNNs to process low-level features
  - ❖ gated recurrent units (GRU) applied after CNNs to handle the variable-length sequence
  - ❖ additional path to process the global features then concatenate all paths in a fully connected layer
- ➔ Inputs: low-level features from PF candidates / SVs and global features
- ➔ **Model upgraded to V2**:
  - ❖ optimize and add more input features; drop irrelevant features to shorten inference time
  - ❖ achieve up to 40% improvement from the V1 performance

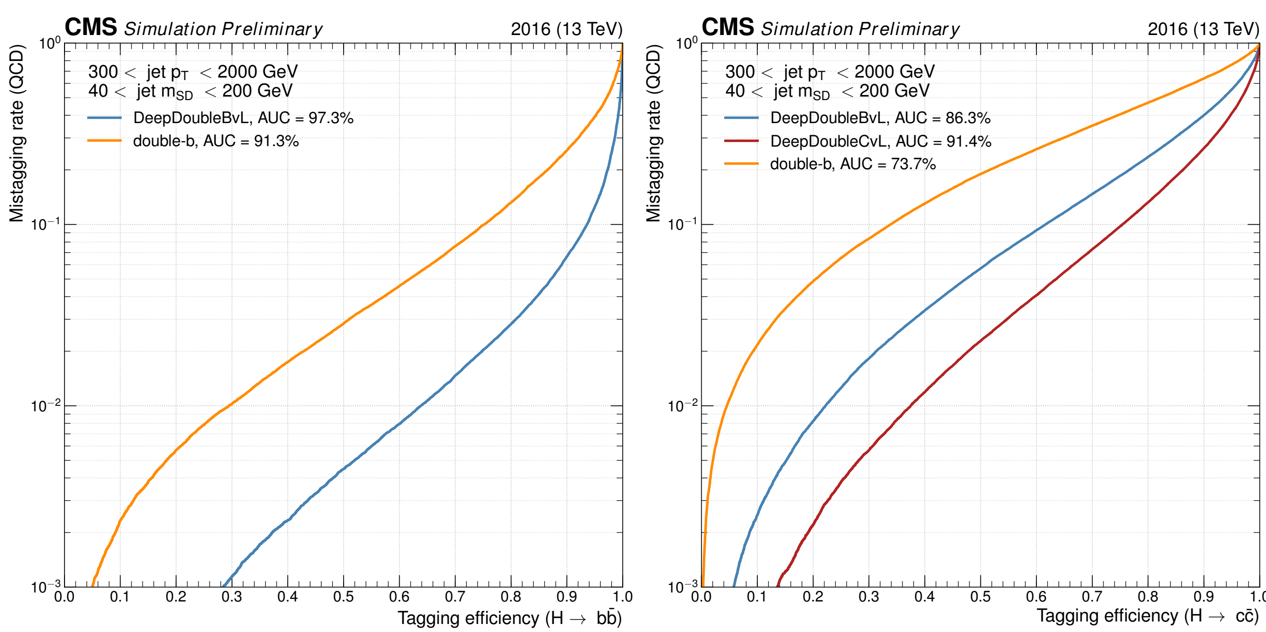
mass sculpting effect for the bb vs. light tagger



NN architecture for DeepDoubleX (V2)



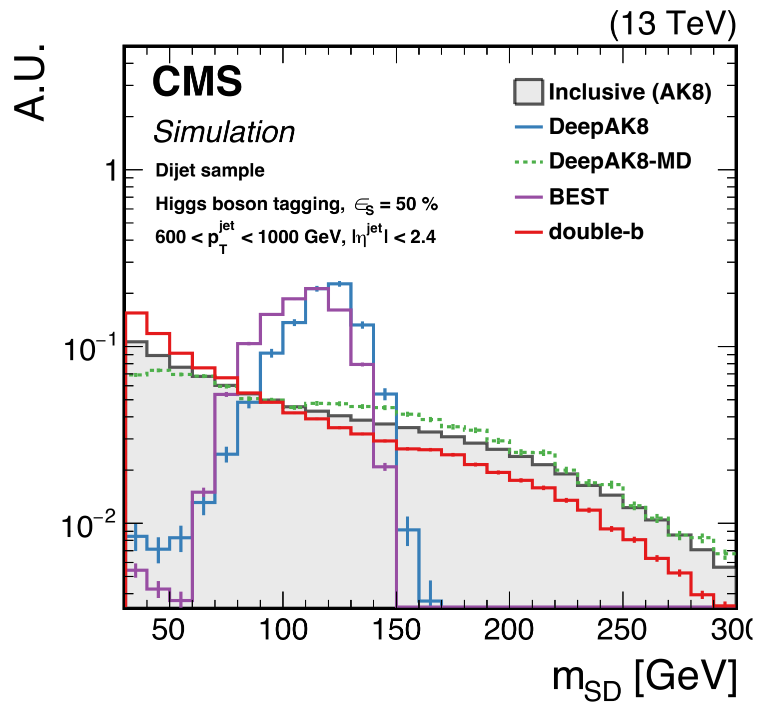
ROC for DeepDoubleX (V1) [CMS-DP-2018-046]

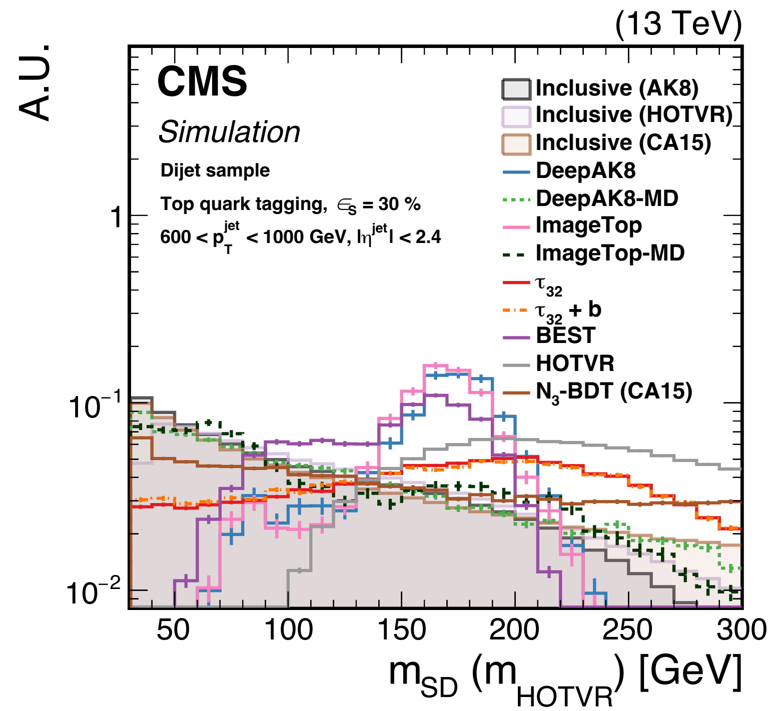


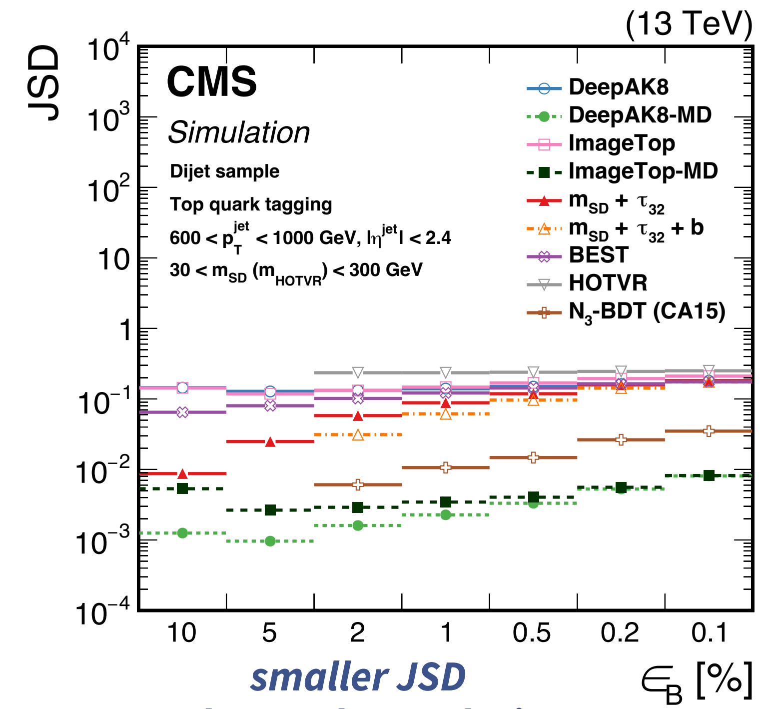
Mass decorrelation plots

mass sculpting effect in various taggers
[[JINST 15 (2020) P06005](https://doi.org/10.1088/1748-0221/15/06/P06005)]

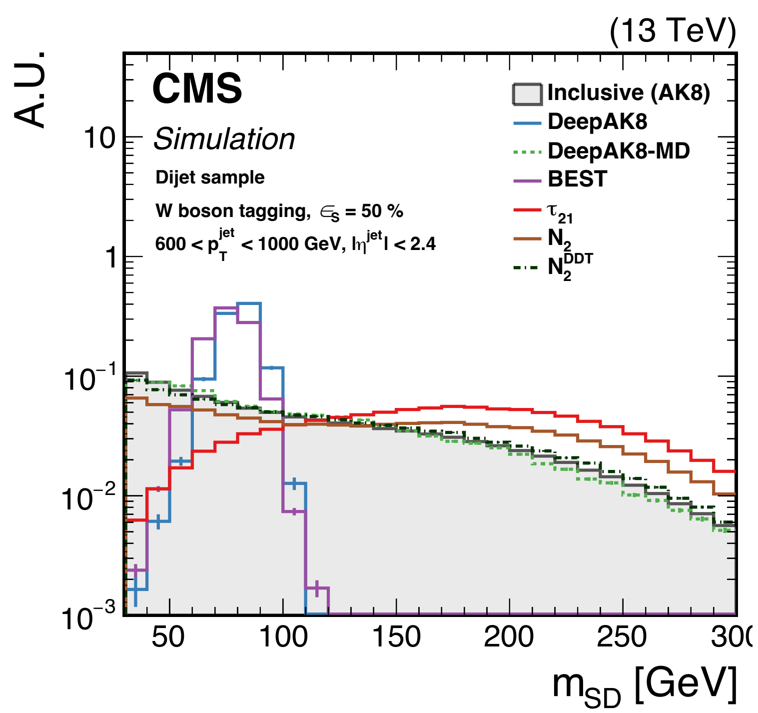
Jensen–Shannon divergence (JSD) as a function of BKG efficiency
[[JINST 15 (2020) P06005](https://doi.org/10.1088/1748-0221/15/06/P06005)]

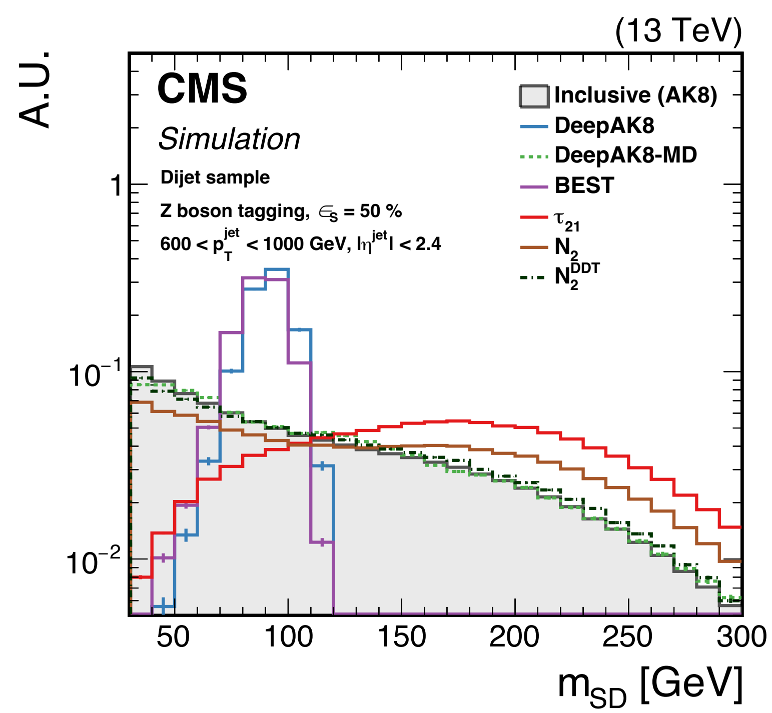


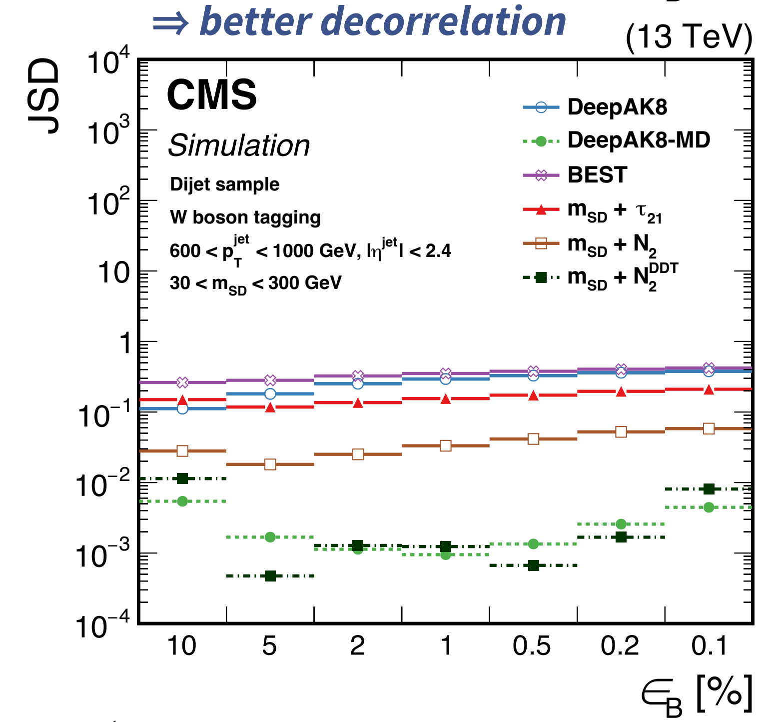




smaller JSD
⇒ better decorrelation







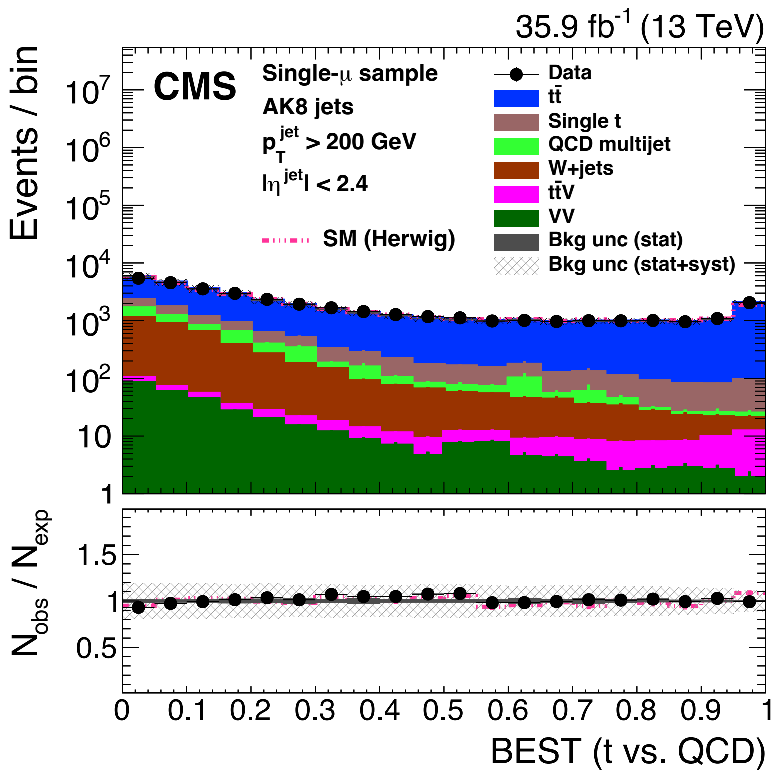
$$\mathrm{JSD}(P\|Q) = \frac{1}{2}\left(\mathrm{KLD}\left(P\|M\right) + \mathrm{KLD}\left(Q\|M\right)\right), \text{ where } M = \frac{P+Q}{2}$$

$$\mathrm{KLD}(P\|Q) = \sum_i P(i)\log_{10}\frac{P(i)}{Q(i)}$$

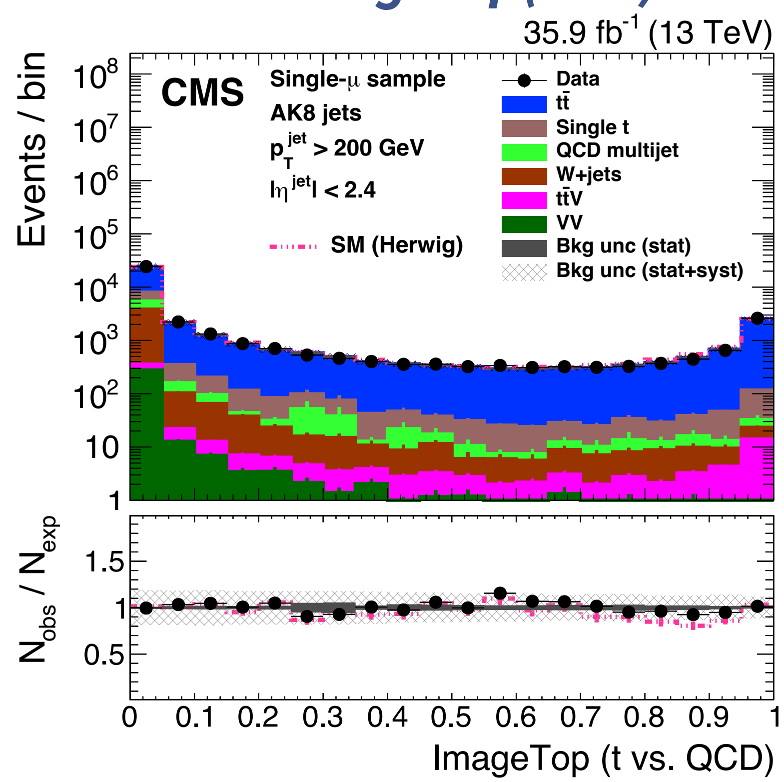
Data/MC comparison

data/MC comparison on single-μ samples [JINST 15 (2020) P06005]

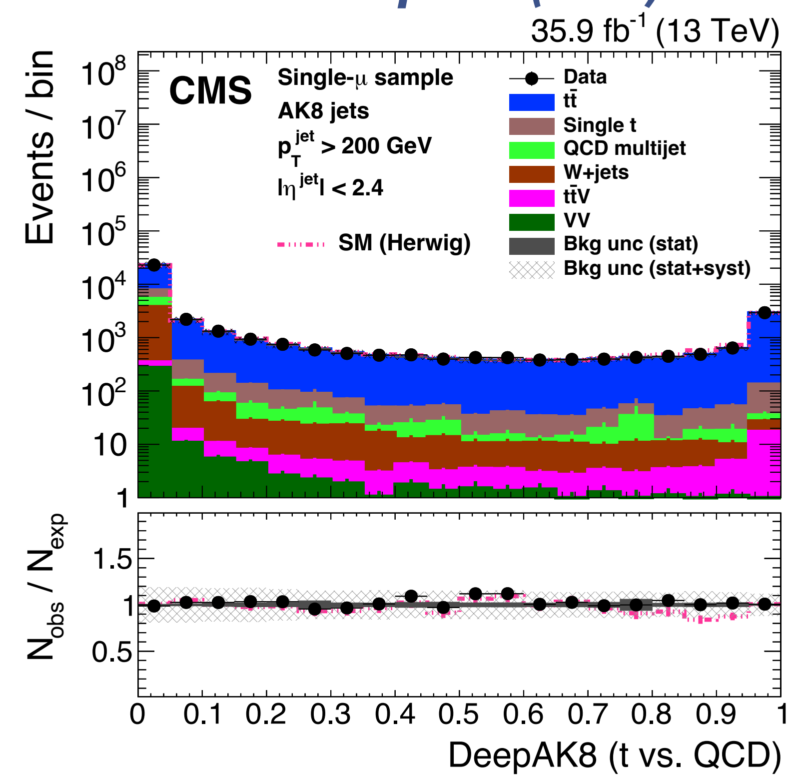
BEST



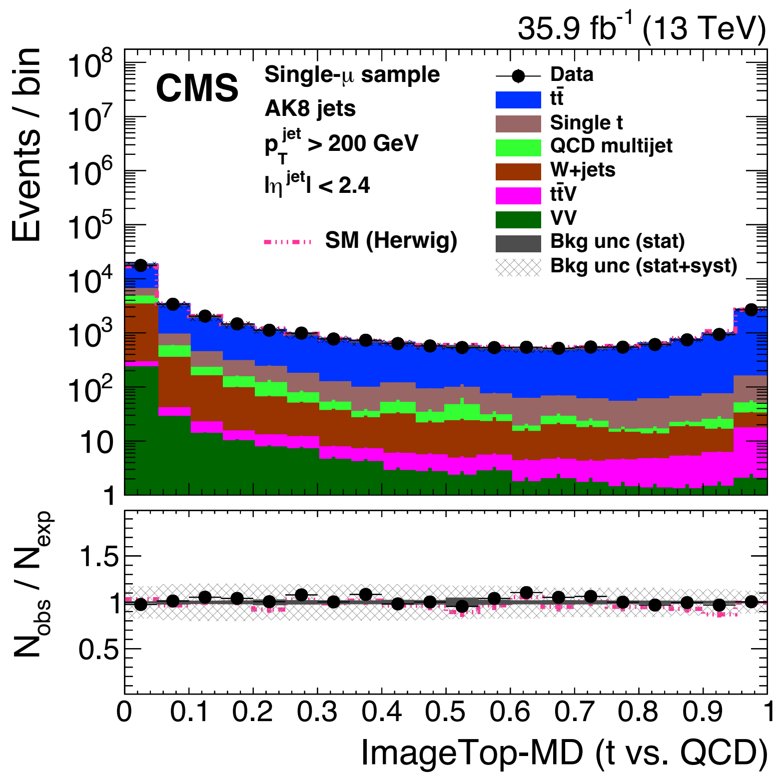
ImageTop(-MD)

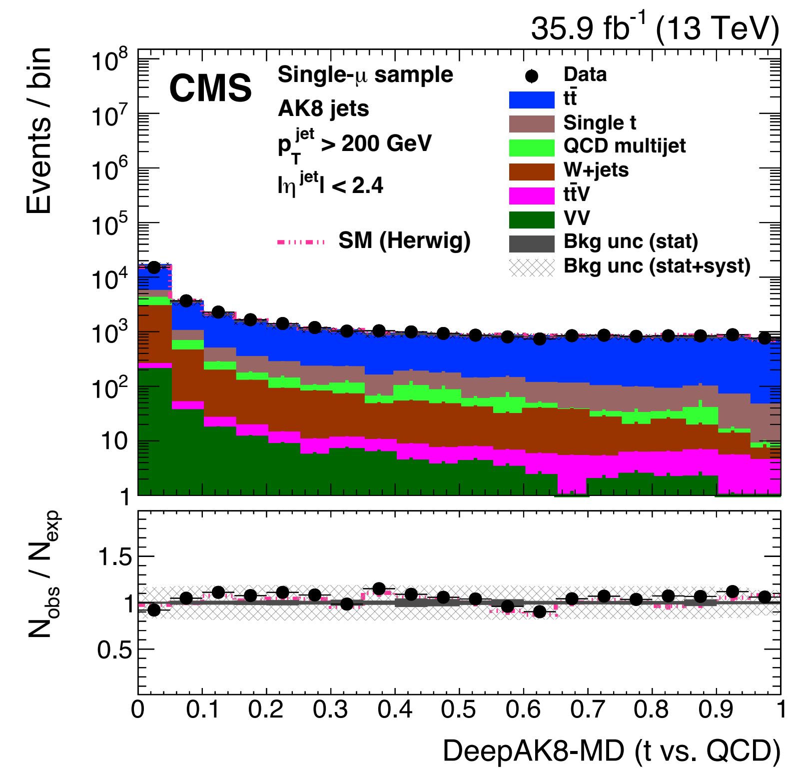


DeepAK8(-MD)



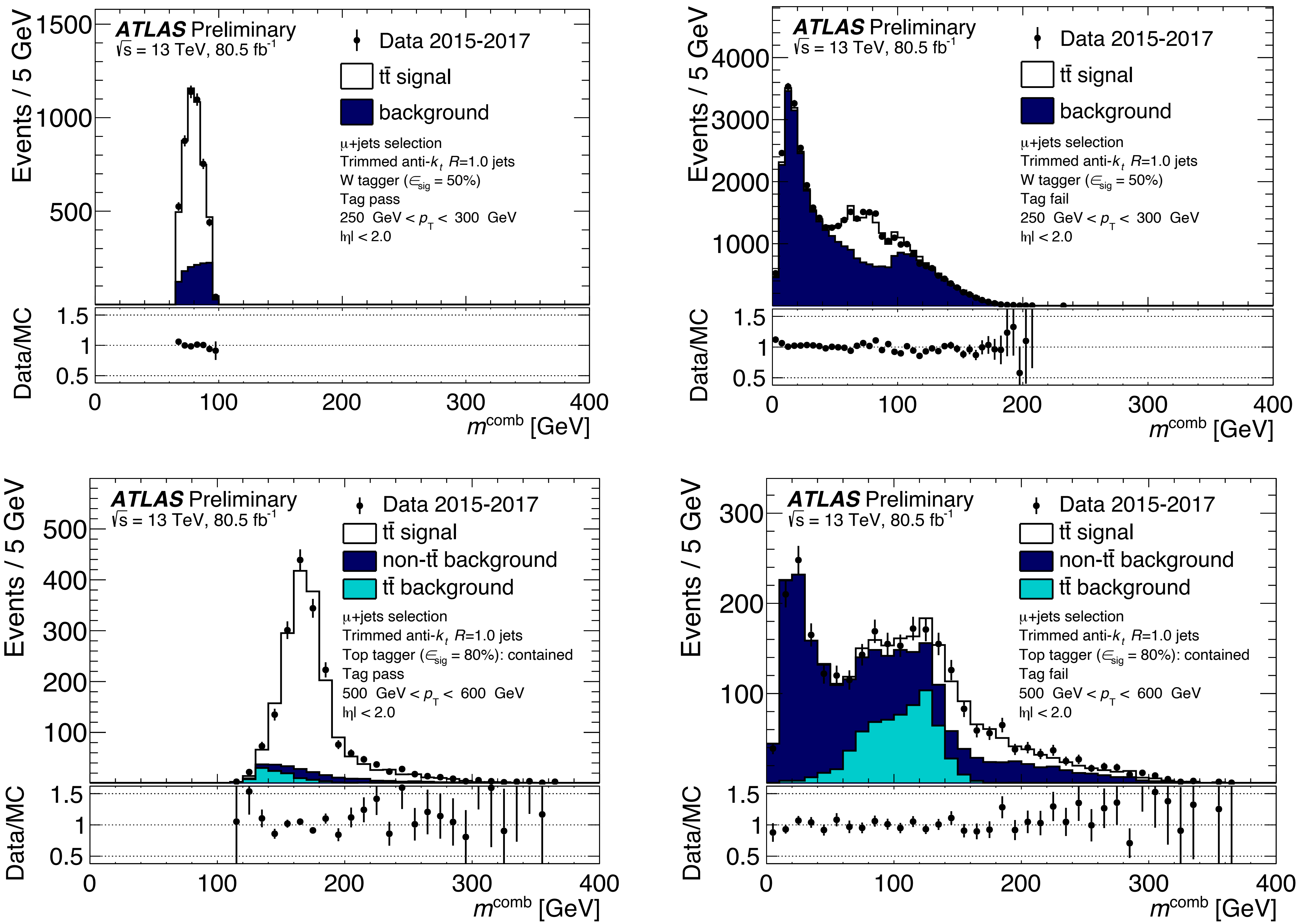
- ➔ using 2016 single-μ data
- ➔ SM (Herwig) shows the MC prediction using Herwig (instead of Pythia) for hadronization





Calibration of W/top taggers

ATL-PHYS-PUB-2021-035



Calibration of W/top taggers

CMS-DP-2020-025

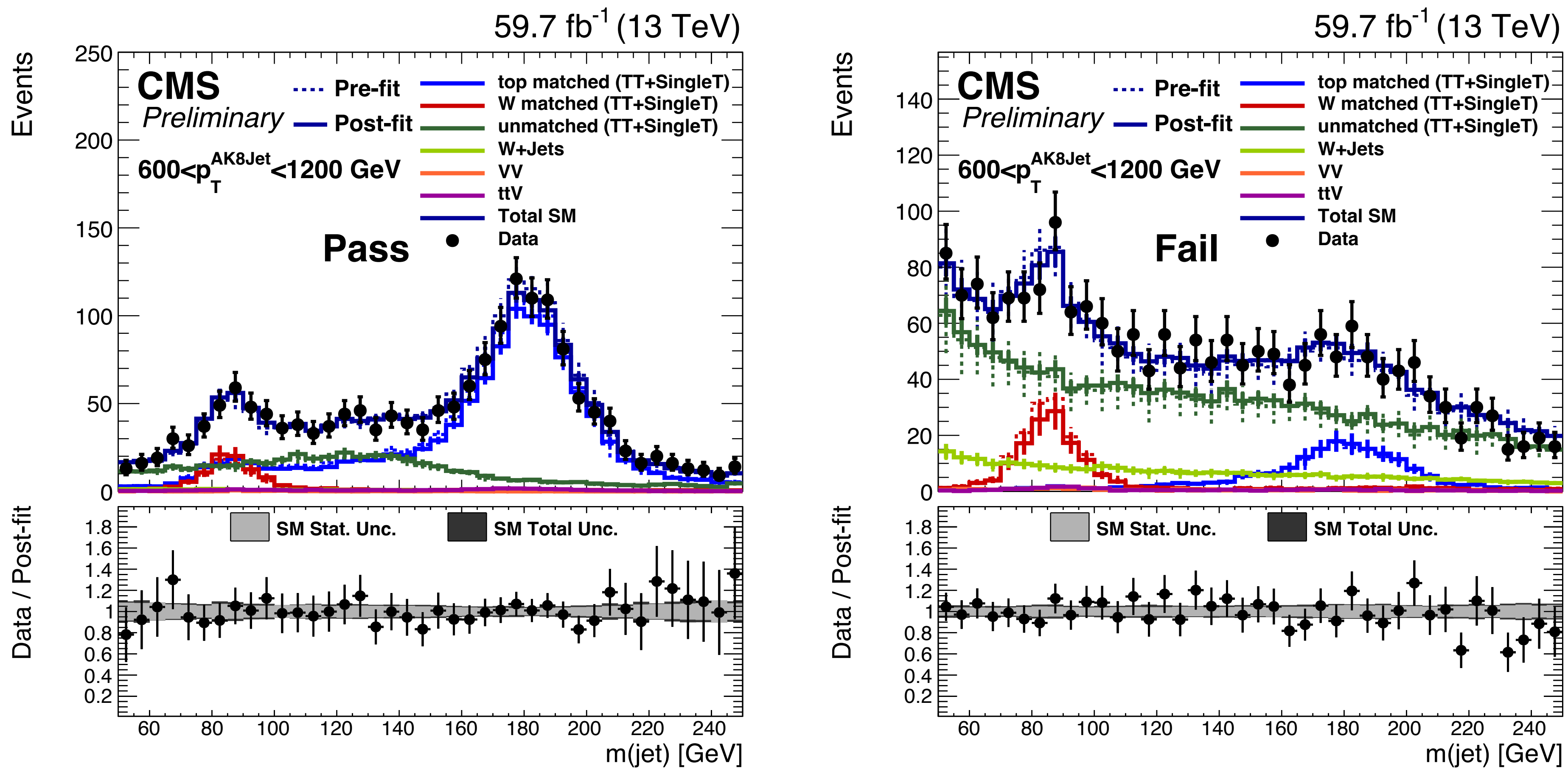
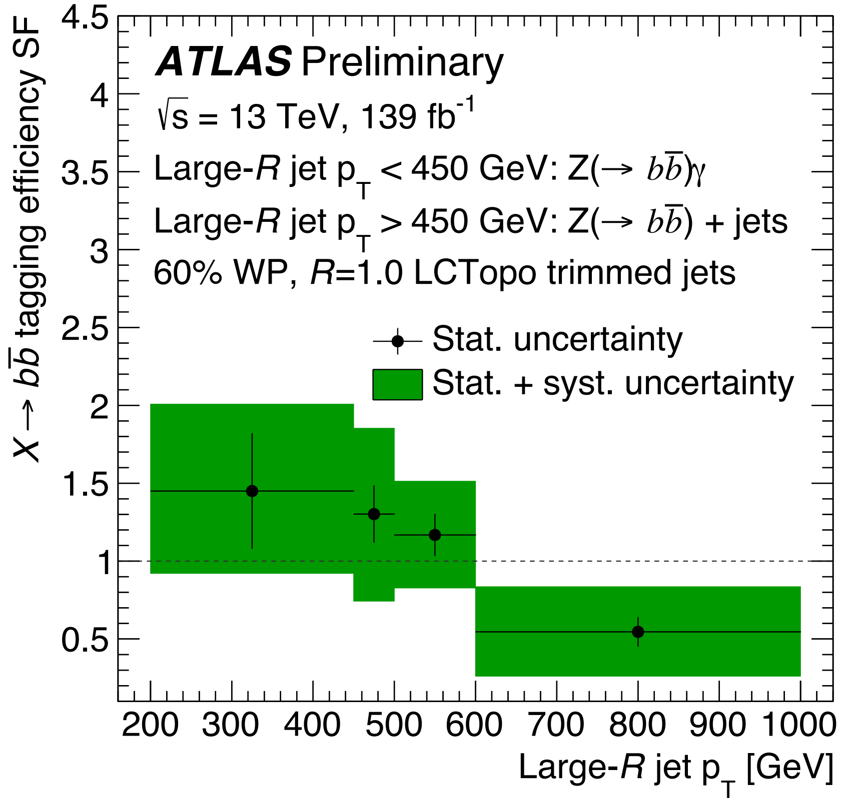
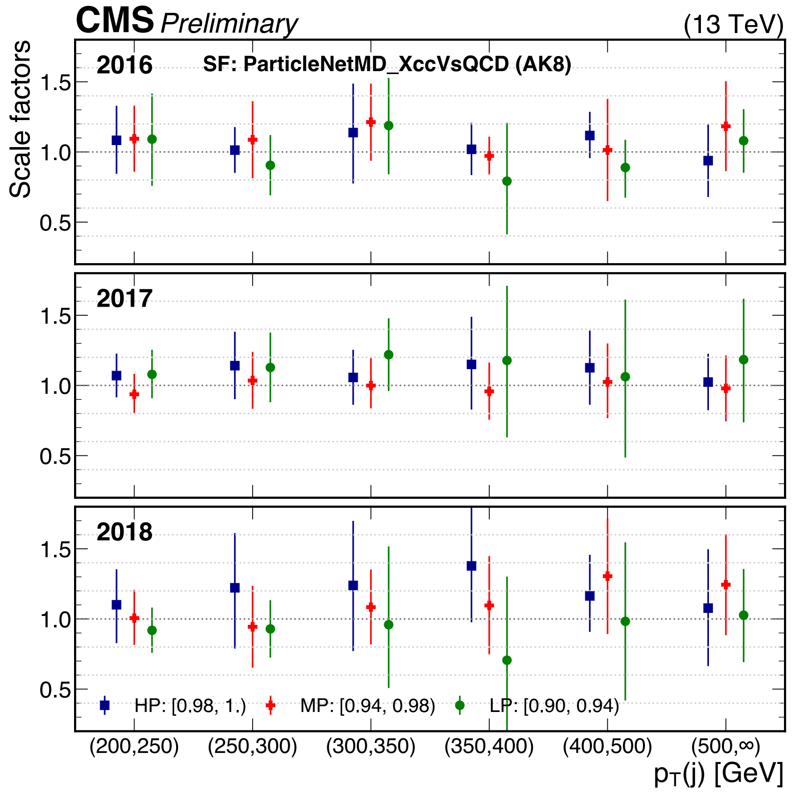


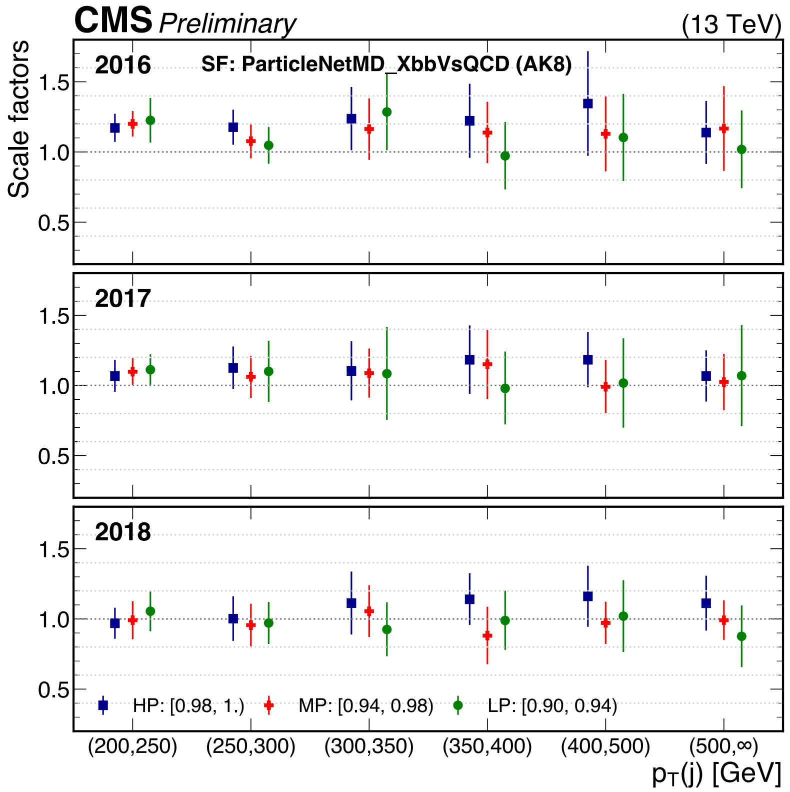
Figure 12: The m_{jet} distribution for data and simulation in the passing (left) and failing (right) categories for the mass decorrelation version of the top tagging (1% mis-identification rate) on the p_T window $600 < p_T^{\text{AK8Jet}} < 1200$ GeV. The solid lines correspond to the contribution of each category after performing maximum likelihood fit. The contribution from QCD multijet events is included in the total SM. The dashed lines are the expectation from simulation before the fit. The lower panel shows the data-to-simulation ratio. The "top/W matched" convention used here indicate that a simulated top quark/W boson is overlapping with the large-radius jet, but not necessarily all of its decay products.

Calibration of $X \rightarrow b\bar{b}/c\bar{c}$ taggers



ATL-PHYS-PUB-2021-035





CMS-DP-2022-005