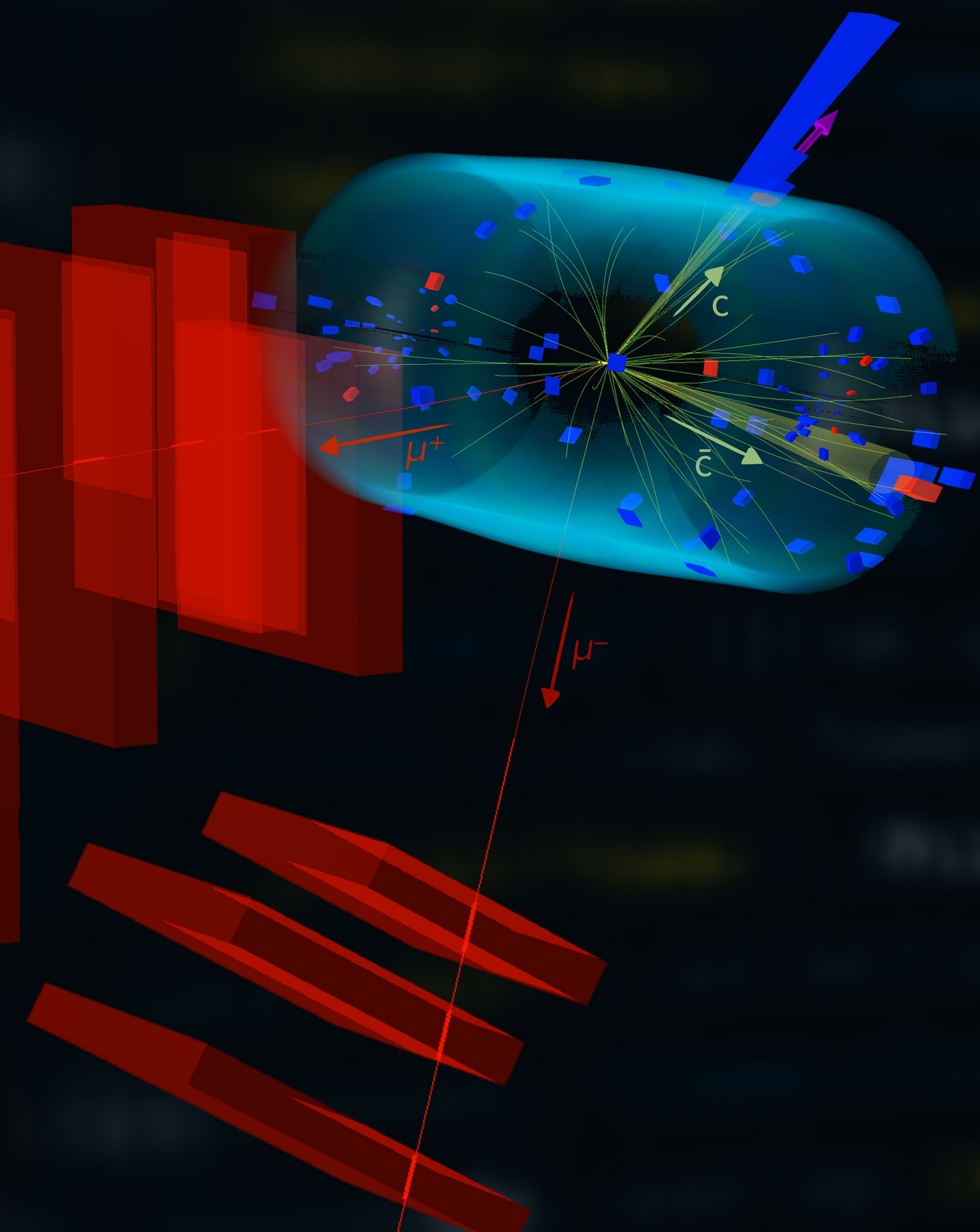
Deep boosted-jet taggers and Hcc measurement



李聪乔 (Congqiao Li), *Peking University*

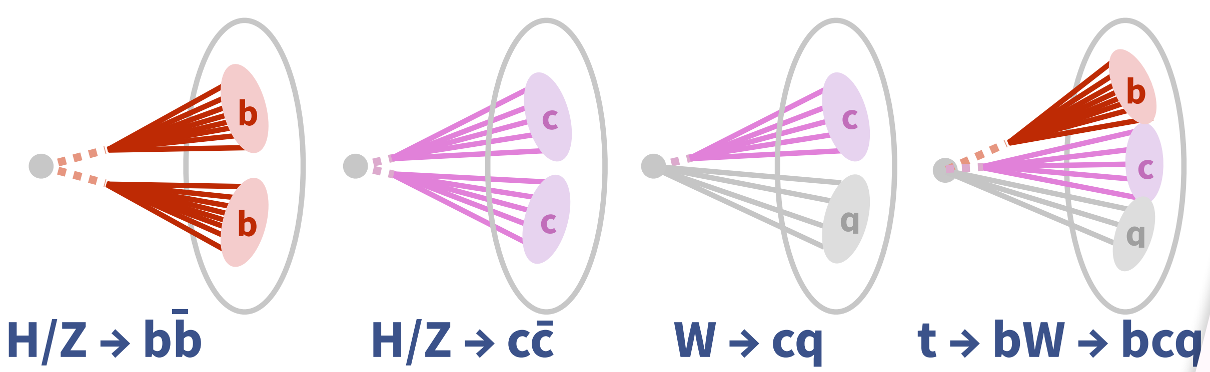
on behalf of the CMS Collaboration

CLHCP 2021 · Nanjing, China
27 November, 2021

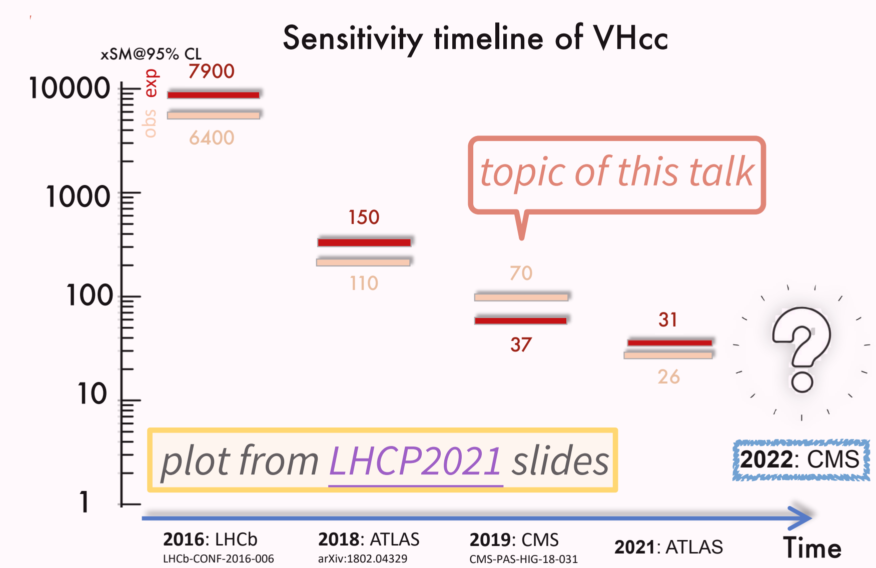
Introduction

- ➔ ***Deep boosted-jet tagging*** is a new yet promising technique in the LHC experiment
  - ❖ deal with traditional jet classification task with the deep neural network
  - ❖ boosted jet (*R*=0.8 or 1.5) explores the rich phase-space from the boosted region when decay particles of a resonance merge into one jet
  - ❖ able to capture the full correlation of the large-R jet constituents
- ➔ Boosted-jet tagging in CMS
  - ❖ include the tagging of t/W/Z/H and BSM particles, decaying to hadrons with different flavours

different types of the large-R jet structure



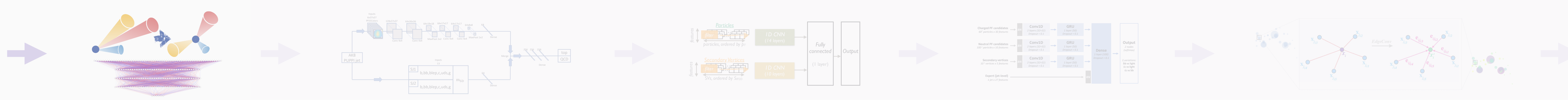
...



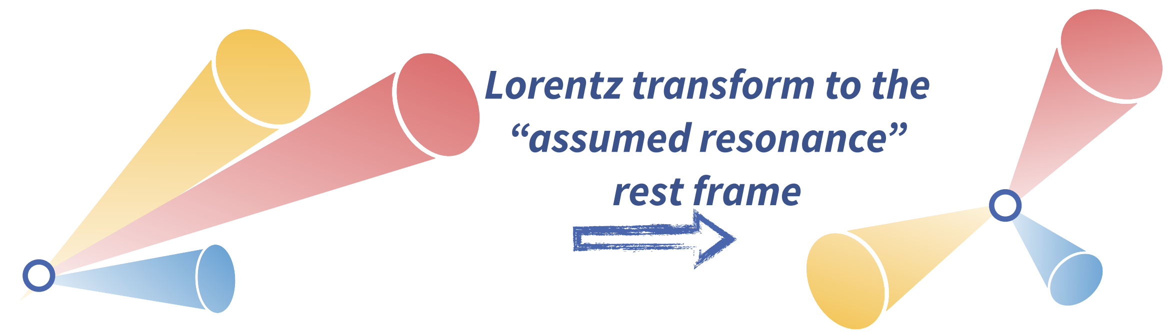
- ➔ ***Higgs to charm coupling measurement***
  - ❖ measuring Higgs couplings with 2nd generation fermions are the next milestone
  - ❖ the main difficulty to probe the H→cc signal is the charm jet identification
  - ❖ ***boosted H→cc jet tagging technique is first explored in CMS*** *and improves the measurement sensitivity*

- ➔ In this talk, we will
  - ❖ introduce various deep boosted-jet taggers developed in CMS (main focus)
  - ❖ present an overall image of the H→cc analysis in CMS, and explain how deep boosted-jet tagger brings the improvement

Boosted event shape tagger (BEST)

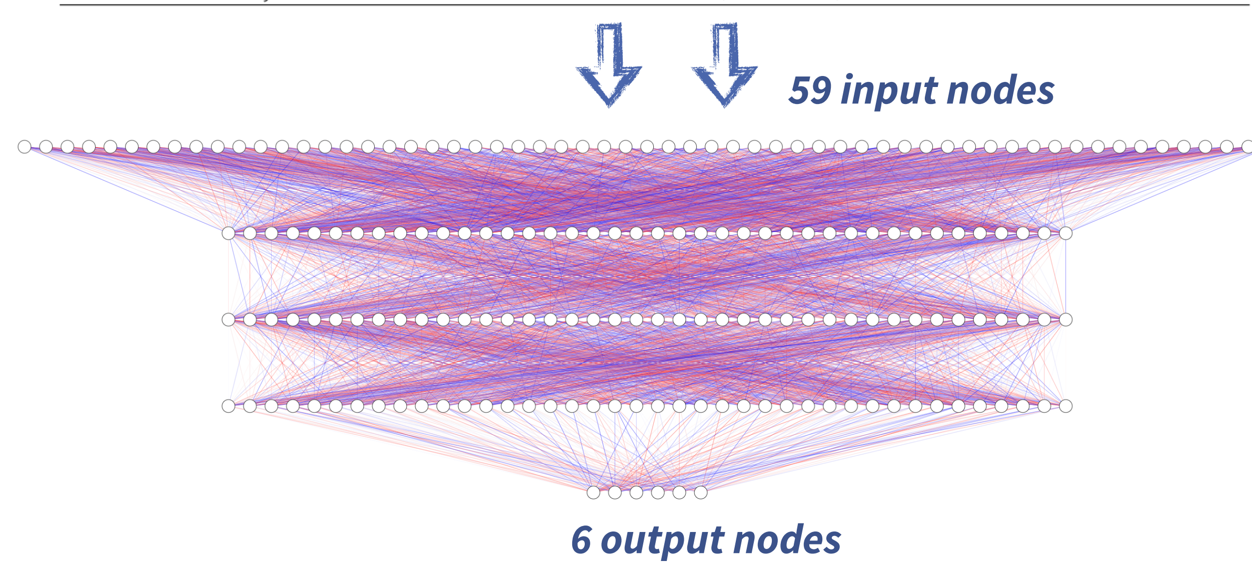


- ➔ ***BEST***: a multi-class tagger to discriminate hadronic decays of high-p_T t/W/Z/H bosons from jets arising from b/light quarks, and gluons [Phys. Rev. D 94, 094027]
- ➔ Architecture:
  - ❖ feed-forward NN with 3 hidden layers; 59 nodes as input, 6 nodes as output
- ➔ Input:
  - ❖ 59 input features as "***boosted event shapes***": high-level jet quantities + global features
    - ‣ include advanced "event shape" variables: Fox-Wolfram moments; sphericity; aplanarity; thrust …
- ➔ Performance for all taggers summarized in p.7

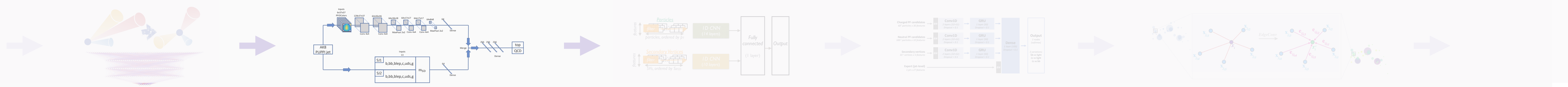


summary of input variables [JINST 15 (2020) P06005]

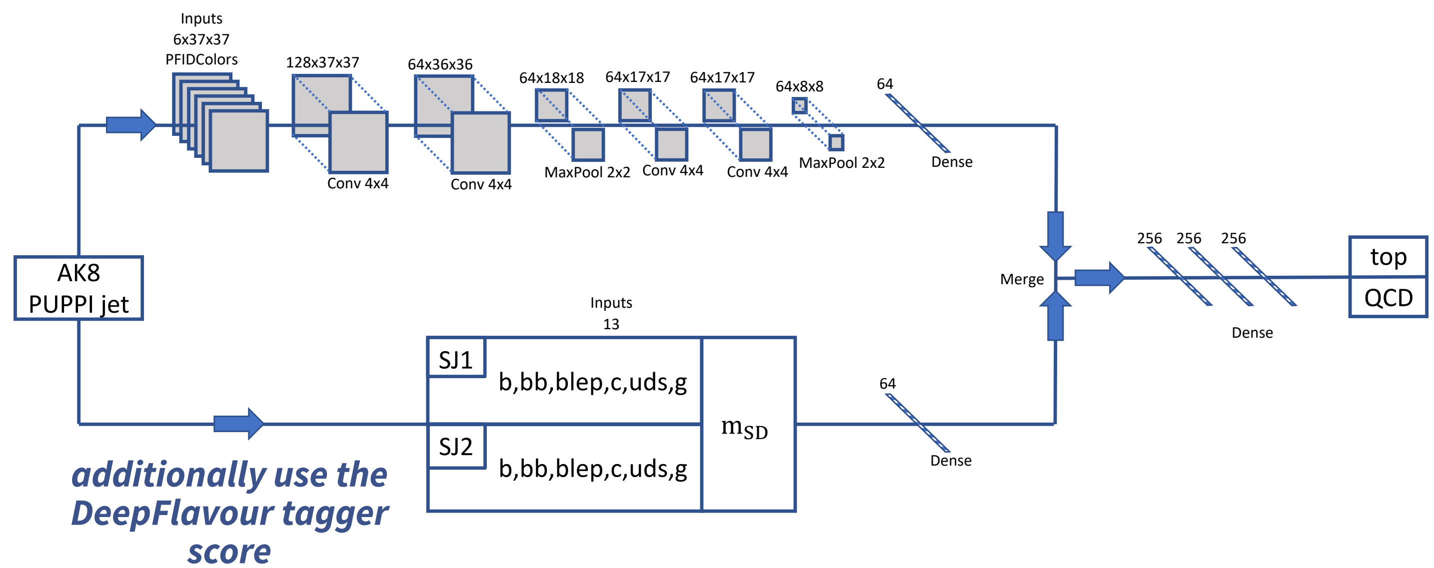
BEST training quantities		
Jet charge	Fox–Wolfram moment H_1/H_0 (t,W,Z,H)	m_{12} (t,W,Z,H)
Jet η	Fox–Wolfram moment H_2/H_0 (t,W,Z,H)	m_{23} (t,W,Z,H)
Jet τ_{21}	Fox–Wolfram moment H_3/H_0 (t,W,Z,H)	m_{13} (t,W,Z,H)
Jet τ_{32}	Fox–Wolfram moment H_4/H_0 (t,W,Z,H)	m_{1234} (t,W,Z,H)
Jet soft-drop mass	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)	



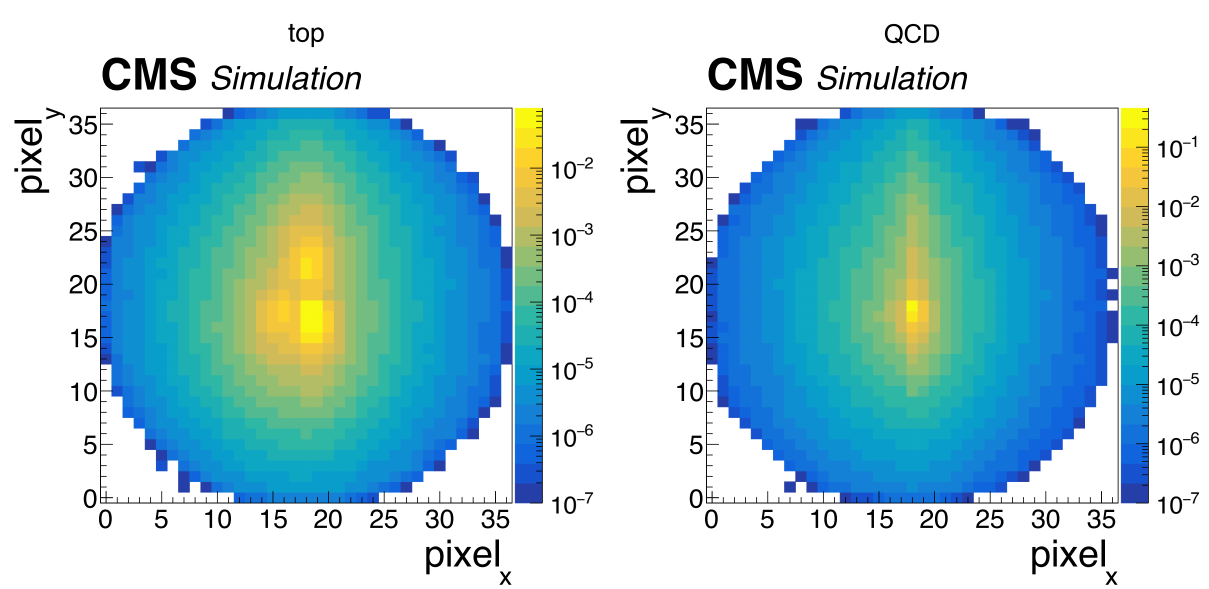
ImageTop



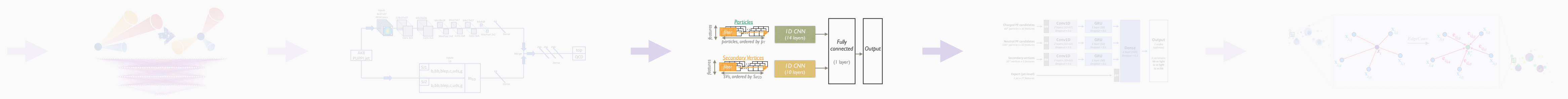
- ➔ ***ImageTop***: discriminate top vs. QCD jets using the ***2D CNN image recognition*** techniques
- ➔ Architecture:
  - ❖ a 2D CNN model: preprocess on the low-level input and create a jet image to pass the 2D CNN chain
- ➔ Input: pixelized jet image after preprocessing
- ➔ ***ImageTop-MD***: a mass decorrelated version
  - ❖ decorrelate mass dependency by probabilistically removing QCD events to achieve ***a same mass spectrum*** for the top & QCD input sample



average jet images after preprocessing for ImageTop [JINST 15 (2020) P06005]



DeepAK8



➔ ***DeepAK8***: multi-class classifier for t/W/Z/H tagging based on 1D CNN in ResNet architecture

- details in [JINST 15 (2020) P06005]; a widely-used boosted jet tagger in CMS

➔ Architecture:

- two individual ***1D CNN chains in ResNet architecture*** (adding shortcuts across layers) to process low-level features

➔ DeepAK8-MD: the mass decorrelated version trained with ***an "adversarial" architecture***

- added a mass prediction network to predict the jet mass from the learned features
- adversarial training strategy: minimize the joint loss will ***improve classification accuracy*** while ***prevent mass correlation***

Inputs

Substructure — Particles

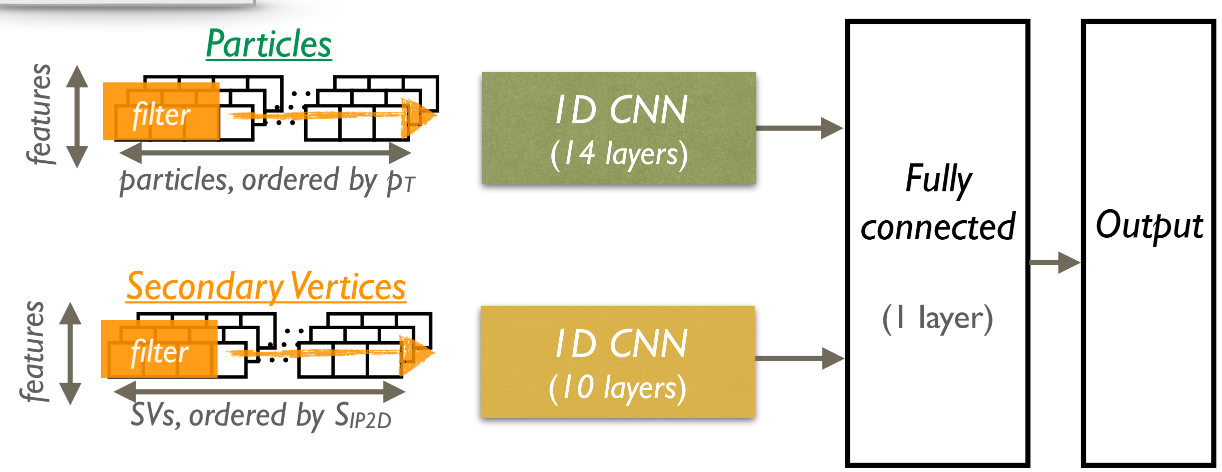
- Up to 100 PF candidates(*)
- Sorted in descending p_T order
- Uses basic kinematic variables, Puppi weights, and track properties (quality, covariance, displacement, etc.)

Flavour — Secondary vertices

- Up to 7 SVs(*) (inside jet cone)
- Sorted in descending S_{IP2D} order
- Uses SV kinematics and properties (quality, displacement, etc.)

(*) Number chosen to include all candidates for ≥ 90% of the events

Architecture



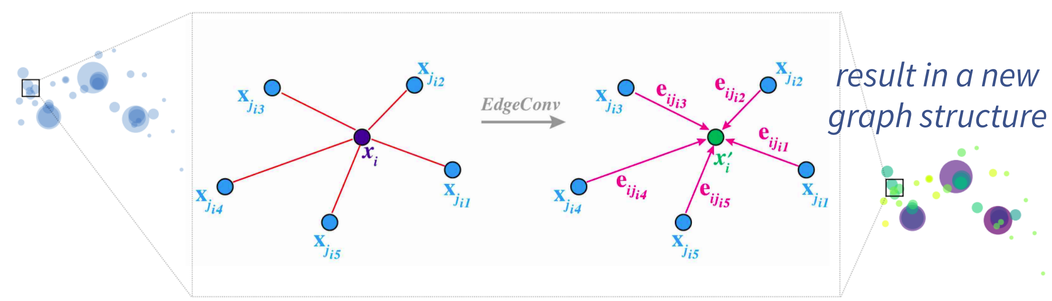
Output

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)

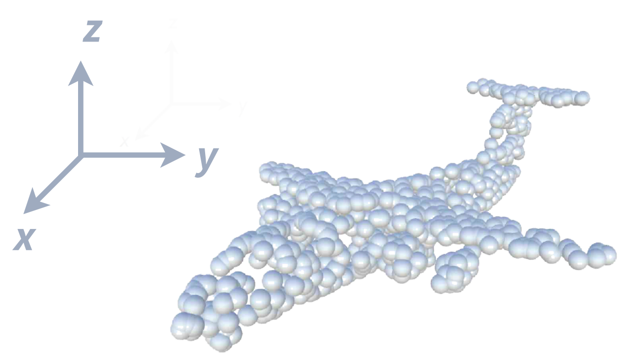
ParticleNet

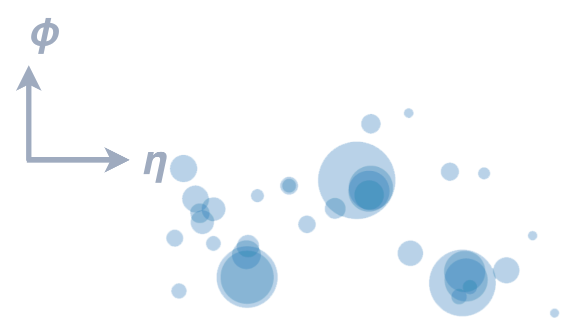
- ➔ ***ParticleNet***: A multi-class jet classifier for t/H/W/Z tagging based on graph NN [[Phys.Rev.D 101, 056019 (2020)]]
  - ❖ achieve ***state-of-the-art performance*** for large-*R* jet tagging at CMS [[CMS-DP-2020-002]]
  - ❖ ***ParticleNet-MD***: The mass-decorrelated version trained with flat (p_T, mass) distribution
- ➔ Architecture:
  - ❖ treat a jet as an ***unordered set of particles*** in the η–ϕ space
  - ❖ use graph NN which maintains the ***permutation-invariant symmetry*** (model based on Dynamic Graph CNN (DGCNN) architecture with EdgeConv operation)
- ➔ Input: low-level features of PF candidates / SVs

EdgeConv: an analogous convolution, operated on the graph by finding k-nearest neighbours for each node



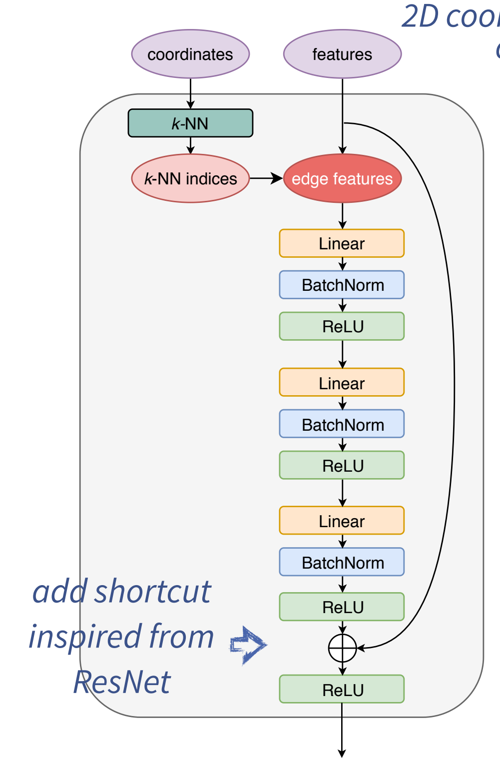
DGCNN explained



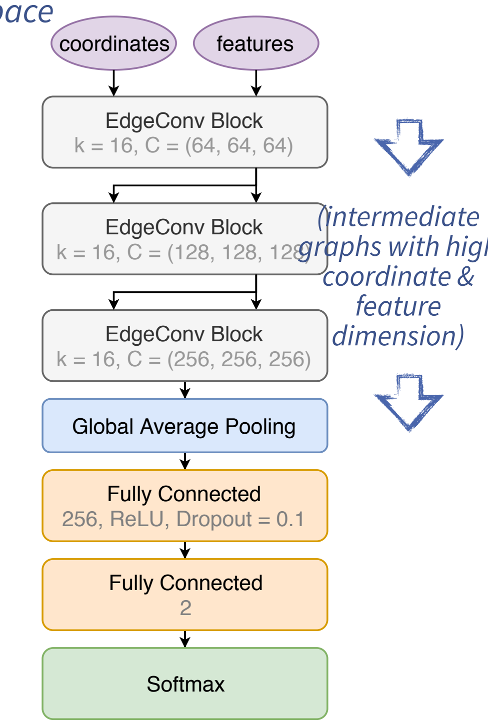


typical top jet image as a "point cloud"

2D coordinate space on η-ϕ



EdgeConv block

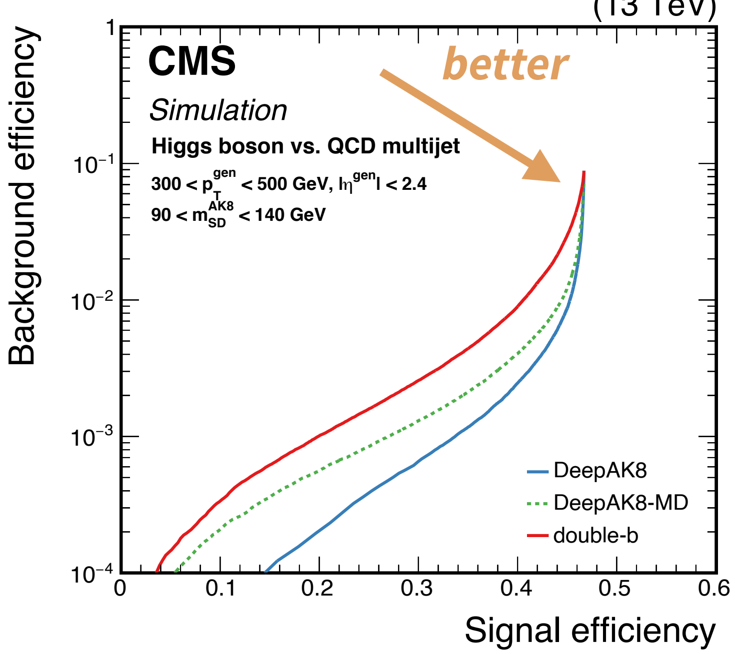


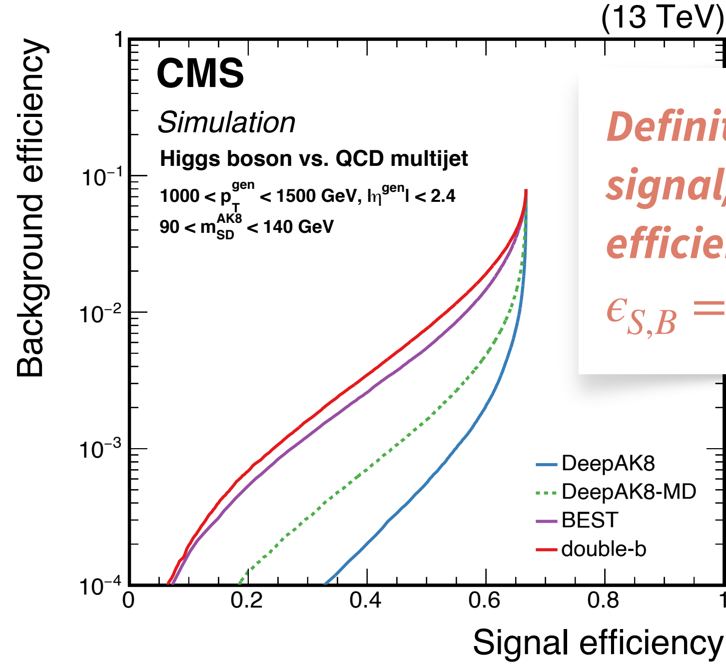
ParticleNet full architecture

Performance in boosted-jet tagging

ROC comparisons on various taggers [JINST 15 (2020) P06005]

Higgs vs. QCD

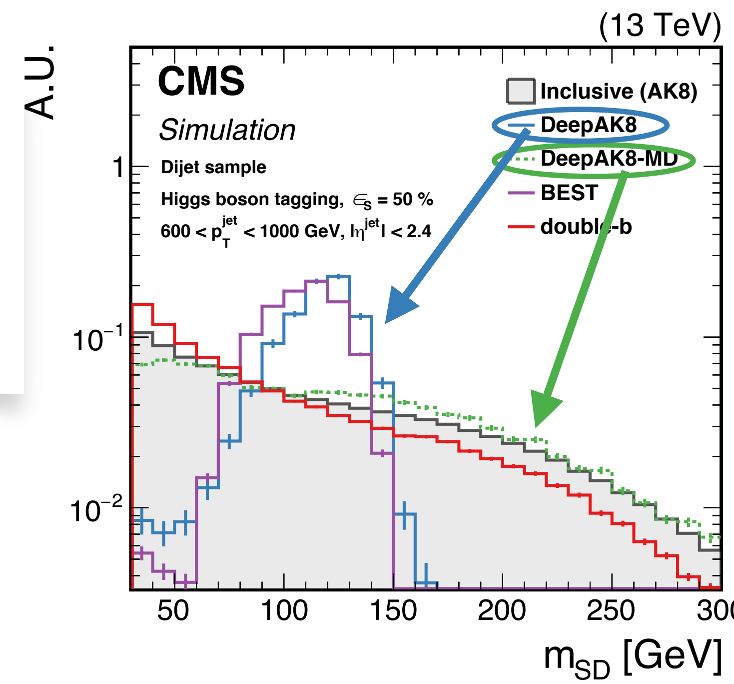




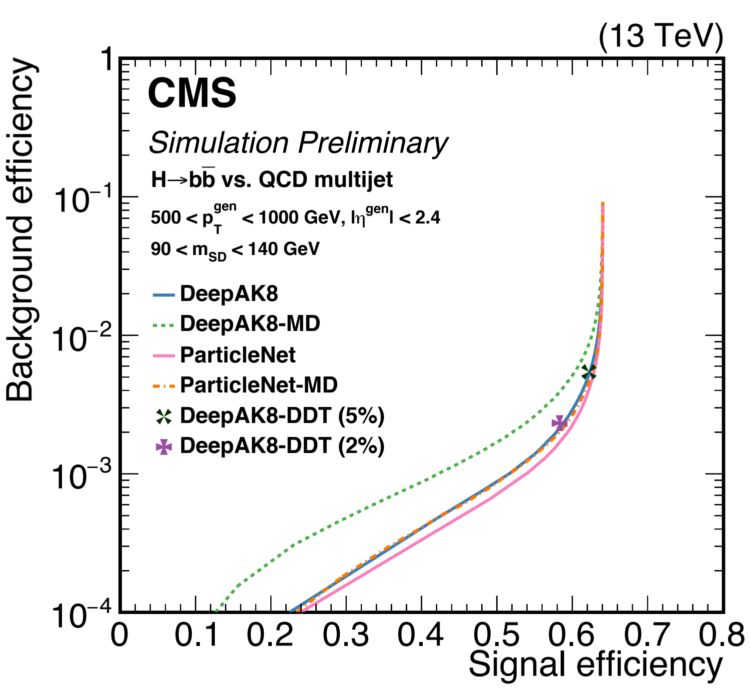
Definition of the signal/background efficiency:

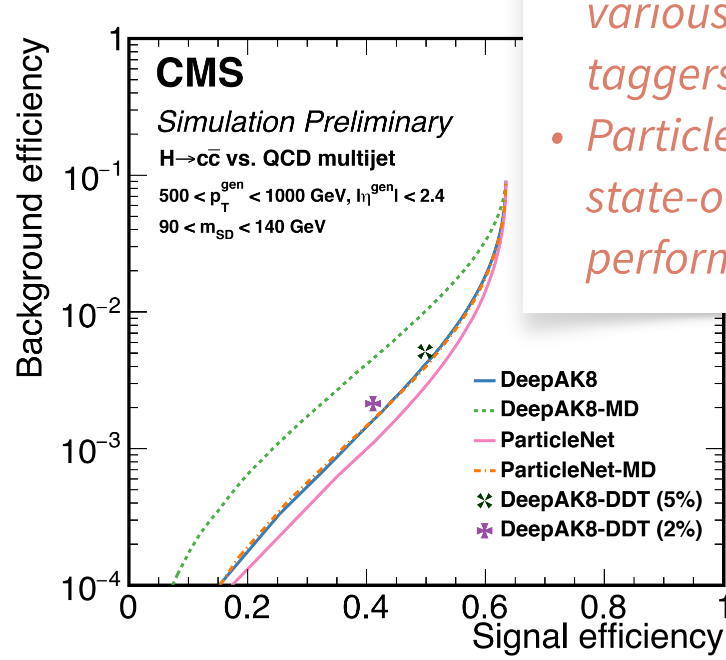
$$\epsilon_{S,B} = N_{S,B}^{\text{tagged}} / N_{S,B}^{\text{total}}$$

mass sculpting effect in various taggers [JINST 15 (2020) P06005]



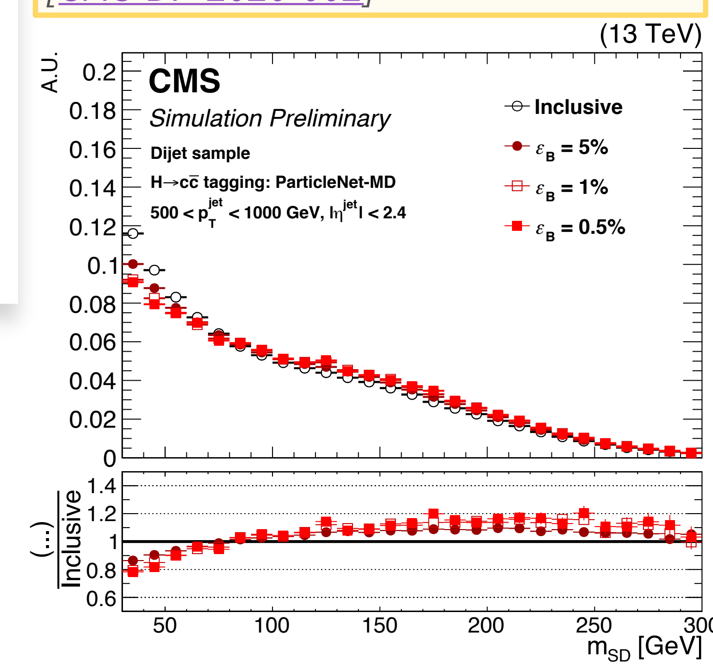
ROC comparisons on ParticleNet vs. DeepAK8 [CMS-DP-2020-002]





- *DeepAK8 surpasses various previous DNN taggers*
- *ParticleNet shows the state-of-the-art performance*

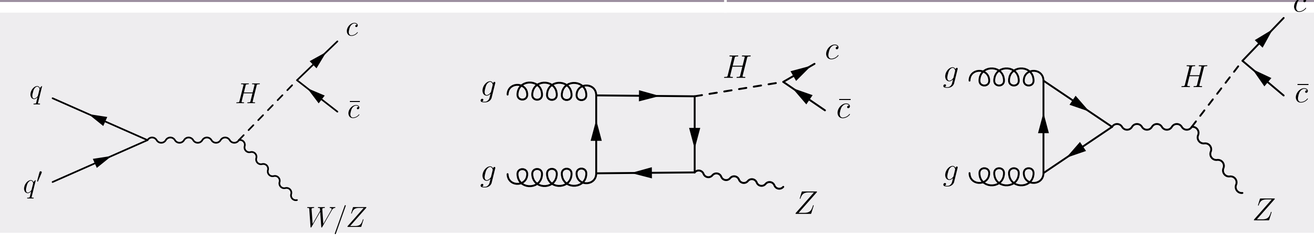
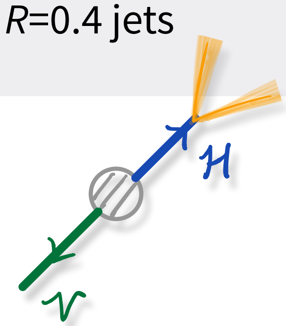
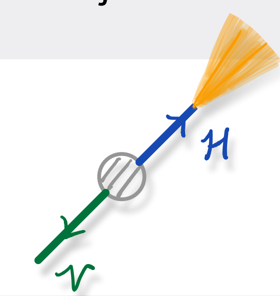
ParticleNet mass-decorrelation effect [CMS-DP-2020-002]



VH→cc: first channel for Hcc measurement

GENERAL STRATEGY

- *CMS analysis with 2016 data (35.9 fb^{-1}): [CMS-PAS-HIG-18-031]*
- *full Run-2 analysis ongoing*

	Resolved-jet topology	*Merged-jet topology*
Feynman diagrams		
targeted channels	3 full leptonic channels: **0L**: Z→νν; **1L**: W→ℓν; **2L**: Z→ℓℓ	
Higgs reconstruction	reconstruct H→cc decay with two R=0.4 jets	reconstruct H→cc decay with one R=1.5 fatjet
motivation	• probe **larger fractions of the available phase space** 	• **improve signal purity** with large p_T(H) • allow to better exploit the **correlations of two charms** when they are contained in the same jet 
charm tagging	require two c-jets from DeepCSV score	require one cc-jet from the DeepAK8-MD score (R=1.5 version)
general strategy	• define SR: *two jets with the high c-tagging score* • further train *BDT* to separate VHcc signal vs. backgrounds • CRs defined for V+b/c/light jets, and tt̄ • extract VHcc signal by *the simultaneous fit on SR (on BDT) and all CRs (on c-tagging discriminant)*	• train *kinematics BDTs* to separate VHcc signal vs. BKG, using only kinematics properties - no mass/flavour information • CR defined by inverting the BDT cut / N_j for tt̄ BKG • *apply cc-tagging selection on three WPs* • finally extract VHcc by *fitting on the fatjet soft-drop mass*
phase-space separation	event with p_T(V)<300 GeV	event with p_T(V)>300 GeV

DeepAK8 for Hcc measurement

- *CMS analysis with 2016 data (35.9 fb^{-1}): [CMS-PAS-HIG-18-031]*
- *full Run-2 analysis ongoing*

→ The DeepAK8 tagger explores the merged di-charm phase-space for the first time in analyses

- identify H→cc jet while vetoing while bb-/light-flavour jets
- use a re-trained tagger adapted to R=1.5 jets
- eventually **fit the soft-drop jet mass** to extract the H→cc signal

→ Calibration of the H→cc tagger is crucial to analysis

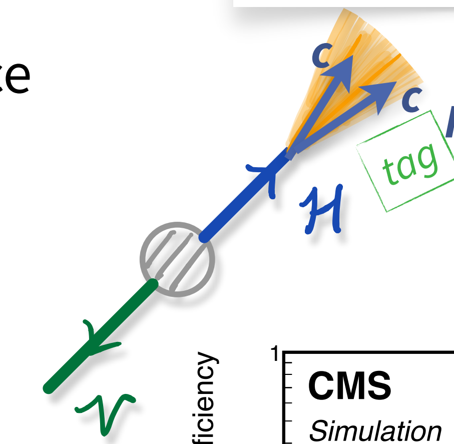
- H→cc/H→bb jet calibrated with the g→cc/g→bb proxy jet, using the QCD multijet sample
- background jets assigned with a rate parameter extracted from the CR fit

→ Stringent limit on H→cc, with 35.9 fb^{-1} data

95% CL exclusion limit on $\mu_{\mathrm{VH(H\to c\bar{c})}}$	Resolved-jet $(p_{\mathrm{T}}(\mathrm{V}) < 300\,\mathrm{GeV})$	Merged-jet $(p_{\mathrm{T}}(\mathrm{V}) \geq 300\,\mathrm{GeV})$	Combination 0L	Combination 1L	Combination 2L	Combination All channels
Expected	45^{+18}_{-13}	73^{+34}_{-22}	79^{+32}_{-22}	72^{+31}_{-21}	57^{+25}_{-17}	37^{+16}_{-11}
Observed	86	75	83	110	93	70

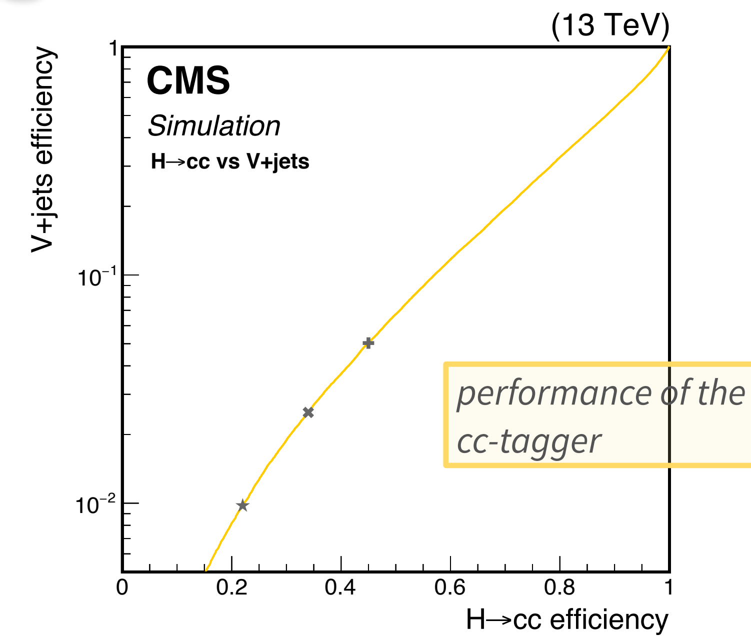
- cf. ATLAS [ATLAS-CONF-2021-021], with 139 fb^{-1} data
  - μ_{VH}(H→cc) < 26 (31) obs. (exp.)

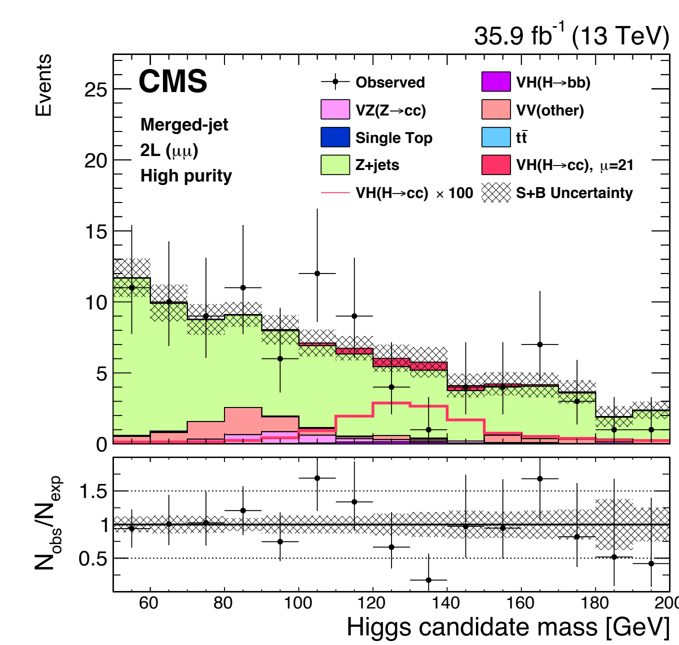
→ Expect better limit with full Run 2 (139 fb^{-1}) data, utilizing a more competent H→cc tagger



H(cc) tagging discriminant

$$\frac{score(H \to cc)}{score(H \to cc) + score(QCD)}$$



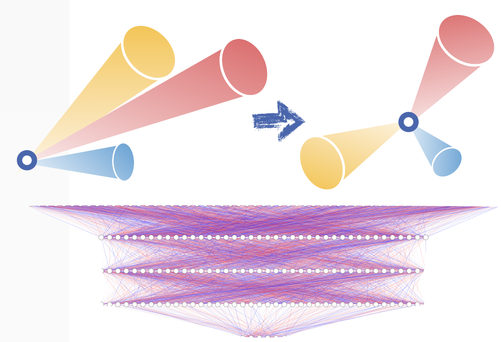


signal extraction on the large-R jet mass

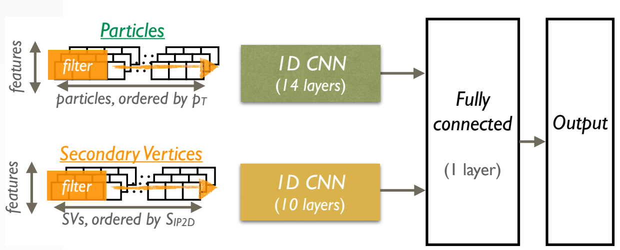
Summary

BDT (high-level inputs)

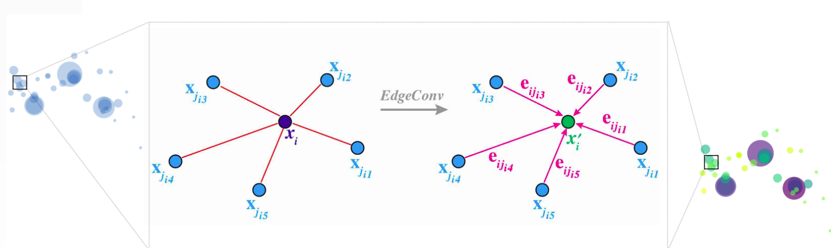
feed-forward NN (high-level inputs)



1D/2D CNN, RNN (low-level inputs)



graph NN (low-level inputs)

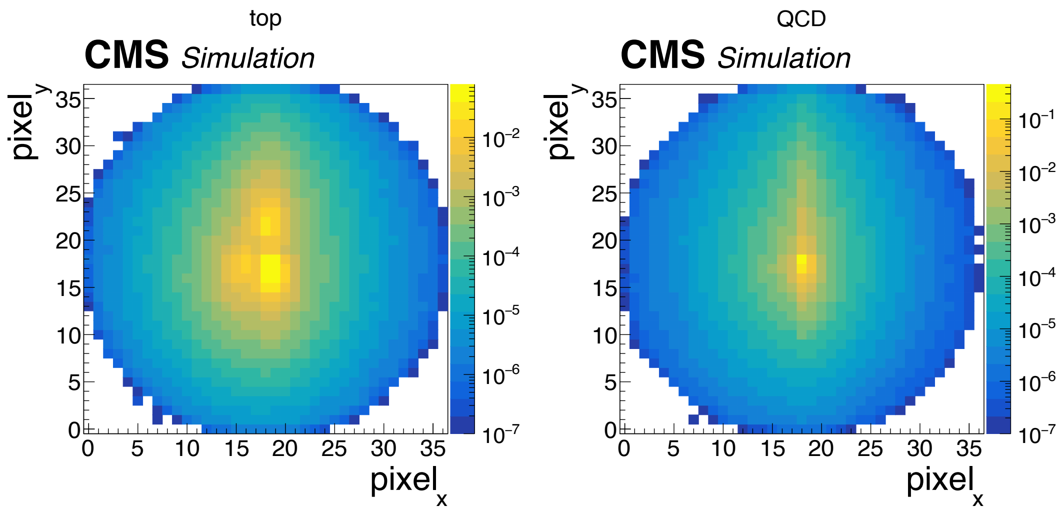


- ➔ Novel DNN approaches for the boosted-jet tagging open a new era
  - allow direct use of high-dimensional low-level inputs and output multi-class scores
  - can be designed to explore jet substructure and flavour information simultaneously
  - capture the underlying symmetry and physics principles with the dedicated NN model
- ➔ Deep taggers are deployed to an increasing number of CMS analyses
  - DeepAK8 successfully used in the H→cc measurement, exploring the untouched di-charm phase-space to improve the sensitivity
  - achieve impressive results in various ongoing CMS analyses

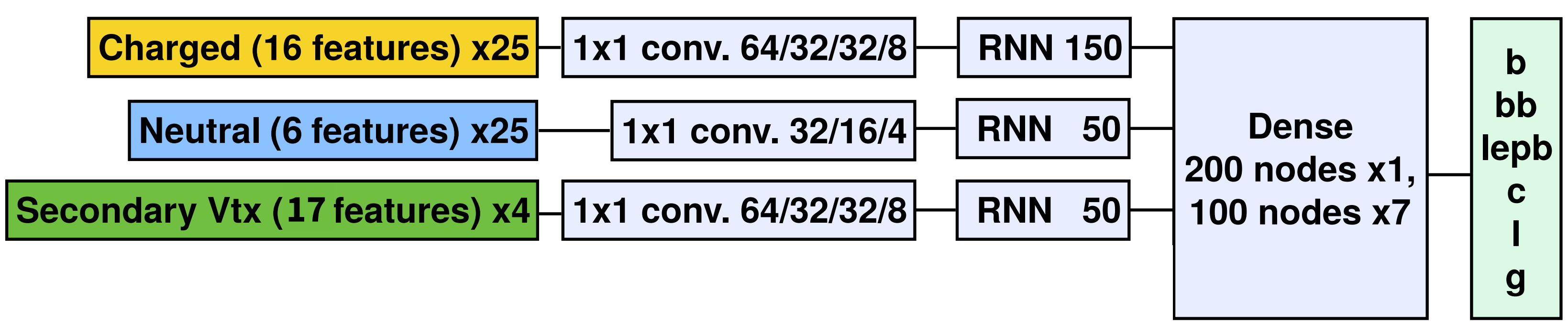
Backup

ImageTop: details

- Standardisation of the jet image
  - shift the jet to the origin
  - rotate and flip: major axis in the vertical & maximum intensity is in the lower-left quadrant



  - pixelize into the 37 × 37 grid, with $\Delta\eta = \Delta\phi = 3.2$
- DeepFlavour
  - designed for AK4 (*R*=0.4) b-jet tagging
  - 1D CNN + RNN (LSTM) network based on the low-level inputs from the charged PF candidates / neutral PF candidates / SVs
  - output six scores



DeepAK8(-MD): details

summary of DeepAK8 input variables [CMS-TS-2019-017]

Variable	Definition
	For both charged and neutral particles.
$\log p_T$	logarithm of the particle's p_T
$\log E$	logarithm of the particle's energy
$\log(p_T/p_T(\text{jet}))$	logarithm of the particle's p_T relative to the jet p_T
$\lvert\eta\rvert$	absolute value of the particle's pseudorapidity
$\Delta\eta(\text{jet})$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi(\text{jet})$	difference in azimuthal angle between the particle and the jet axis
$\Delta R(\text{jet})$	angular separation between the particle and the jet axis
$\Delta R(\text{subjet 1})$	angular separation between the particle and the subjet leading in p_T
$\Delta R(\text{subjet 2})$	angular separation between the particle and the subjet subleading in p_T
$\min \Delta R(\text{SV})$	angular separation between the particle and closest secondary vertex
w_{PUPPI}	PUPPI weight of the particle
q	electric charge of the particle
`isMuon`	if the particle is identified as a muon
`isElectron`	if the particle is identified as an electron
`isPhoton`	if the particle is identified as a photon
`isChargedHadron`	if the particle is identified as a charged hadron
`isNeutralHadron`	if the particle is identified as a neutral hadron
f_{HCAL}	fraction of energy deposited in HCAL
	For charged particles only. A default value of 0 is assigned for neural particle.
`pvAssociationQuality`	flag related to the association of the track to the primary vertices
`lostInnerHits`	quality flag of the track related to missing hits on the pixel layers
d_{xy}	transverse impact parameter of the track
d_z	longitudinal impact parameter of the track
$d_{xy}/\sigma_{d_{xy}}$	significance of the transverse impact parameter
d_z/σ_{d_z}	significance of the longitudinal impact parameter
χ^2/dof	χ^2 value of the trajectory fit normalized to the number of degrees of freedom
`qualityMask`	quality flag of the track
$\text{cov}(q/p, q/p)$	variance of the track parameter q/p
$\text{cov}(\lambda, \lambda)$	variance of the track parameter λ
$\text{cov}(\phi, \phi)$	variance of the track parameter ϕ
$\text{cov}(d_{xy}, d_{xy})$	variance of the track parameter d_{xy}
$\text{cov}(d_z, d_z)$	variance of the track parameter d_z
$\text{cov}(d_{xy}, d_z)$	covariance of the track parameter d_{xy} and d_z
$\text{cov}(\phi, d_{xy})$	covariance of the track parameter ϕ and d_{xy}
$\text{cov}(\lambda, d_z)$	covariance of the track parameter λ and d_z
η_{rel}	pseudorapidity of the track relative to the jet axis
$p_{\text{T,rel}}$ ratio	track momentum perpendicular to the jet axis, divided by the magnitude of the track momentum
$p_{\text{par,rel}}$ ratio	track momentum parallel to the jet axis divided by the magnitude of the track momentum
d_{2D}	signed 2D impact parameter (i.e., in the transverse plane) of the track
$d_{\text{2D}}/\sigma_{\text{2D}}$	signed 2D impact parameter significance of the track
d_{3D}	signed 3D impact parameter of the track
$d_{\text{3D}}/\sigma_{\text{3D}}$	signed 3D impact parameter significance of the track
`trackDistance`	distance between the track and the jet axis at their point of closest approach

Table 5.2: Input variables of each jet constituent particle.

Variable	Definition
$\log p_T$	logarithm of the SV's p_T
$\log E$	logarithm of the SV's energy
$\log(p_T/p_T(\text{jet}))$	logarithm of the SV's p_T relative to the jet p_T
$\lvert\eta\rvert$	absolute value of the SV's pseudorapidity
$\Delta\eta(\text{jet})$	difference in pseudorapidity between the SV and the jet axis
$\Delta\phi(\text{jet})$	difference in azimuthal angle between the SV and the jet axis
$\Delta R(\text{jet})$	angular separation between the SV and the jet axis
m_{SV}	mass of the SV
N_{tracks}	number of tracks associated with the SV
χ^2/dof	χ^2 value of the SV fit normalized to the number of degrees of freedom
d_{2D}	signed 2D impact parameter (i.e., in the transverse plane) of the SV
$d_{\text{2D}}/\sigma_{\text{2D}}$	signed 2D impact parameter significance of the SV
d_{3D}	signed 3D impact parameter of the SV
$d_{\text{3D}}/\sigma_{\text{3D}}$	signed 3D impact parameter significance of the SV
$\cos(\vec{p}_{\text{SV}}, \overrightarrow{\text{(PV,SV)}})$	cosine of the angle between the SV momentum and the vector pointing from the primary vertex to the SV

Table 5.3: Input variables for each secondary vertex (SV) inside the jet.

DeepAK8(-MD) architecture (full scale) [CMS-TS-2019-017]

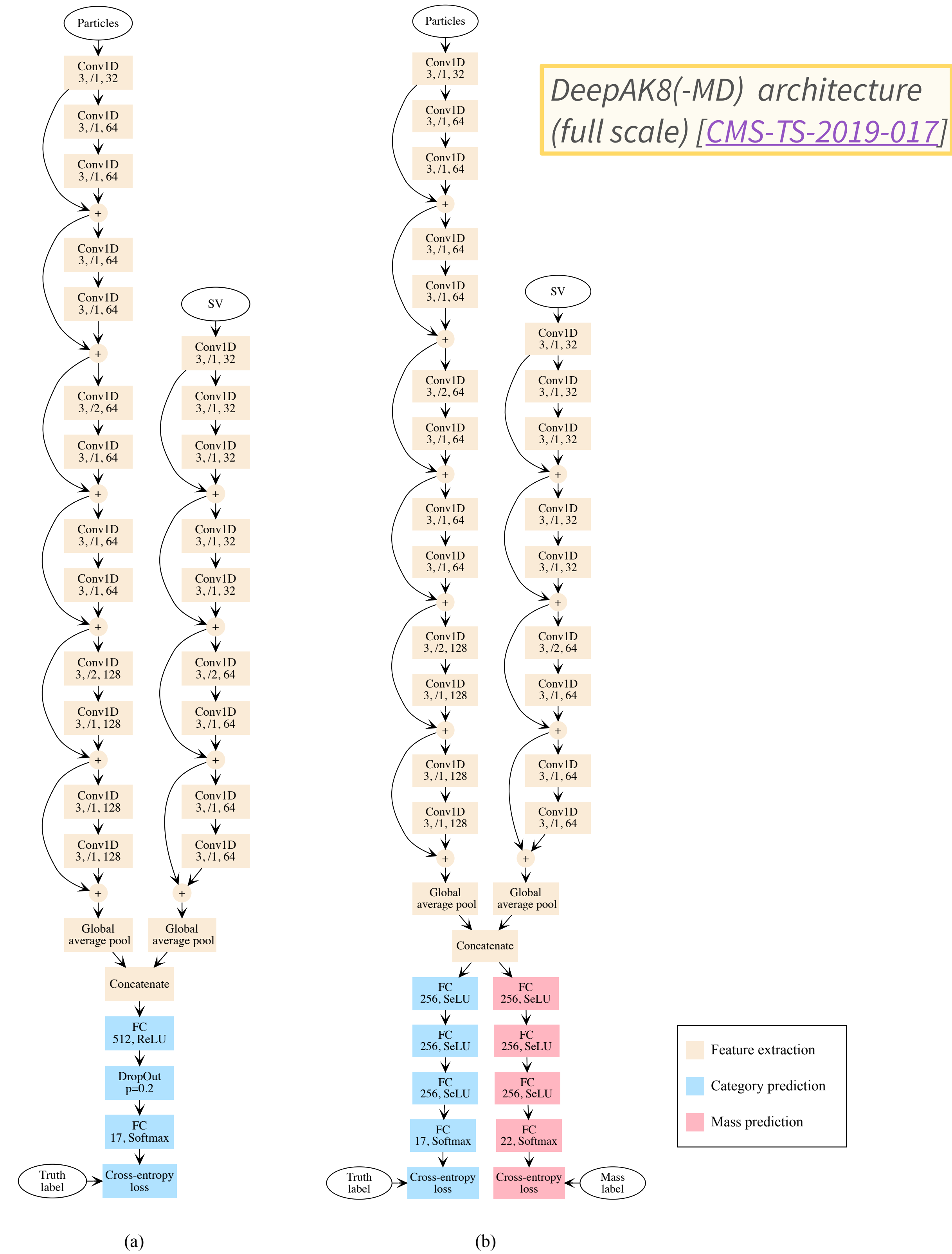
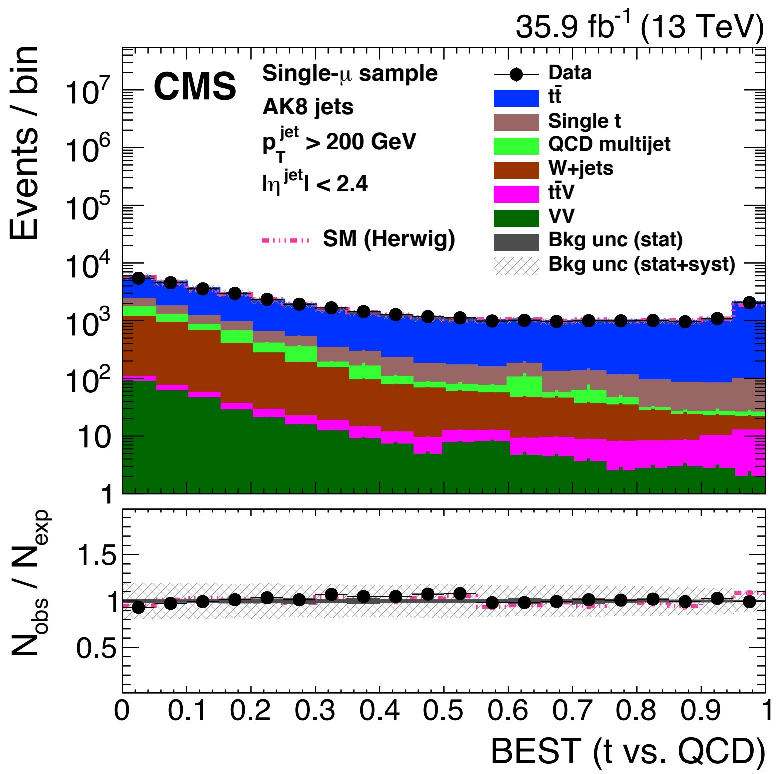


Figure 5.4: The network architecture of (a) DeepAK8 and (b) DeepAK8-MD.

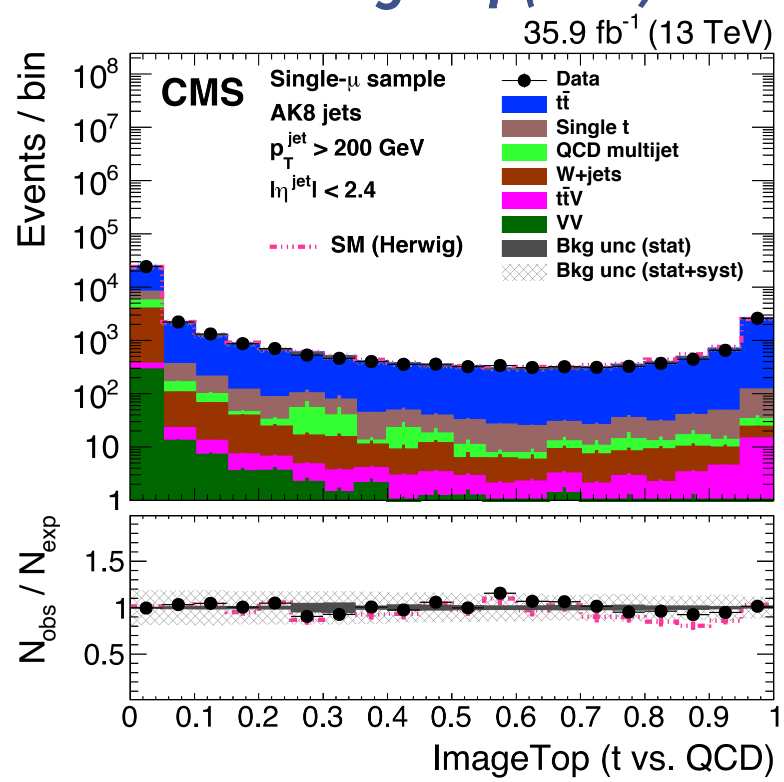
Data/MC comparison

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

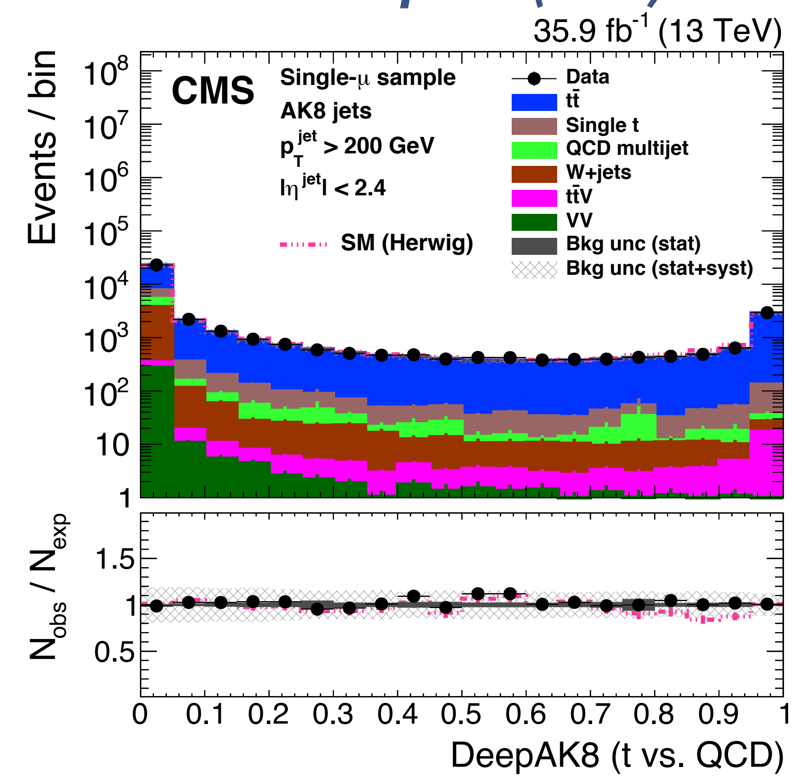
BEST



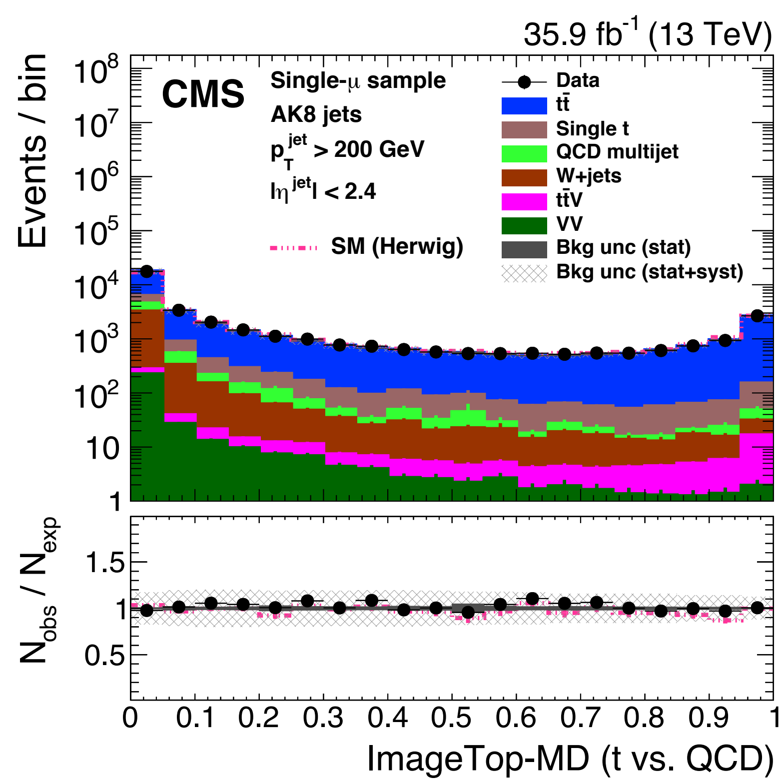
ImageTop(-MD)

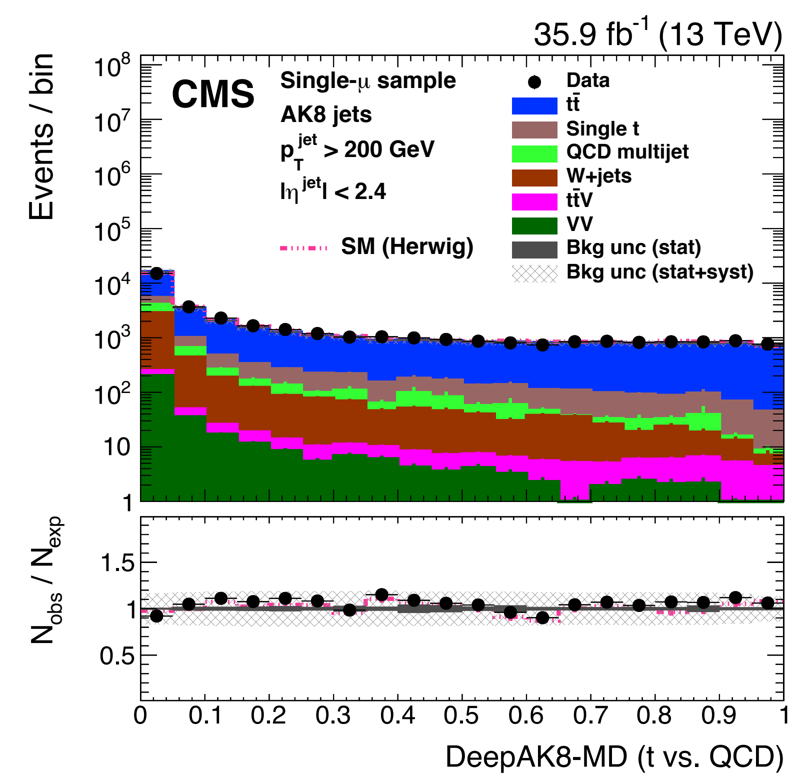


DeepAK8(-MD)

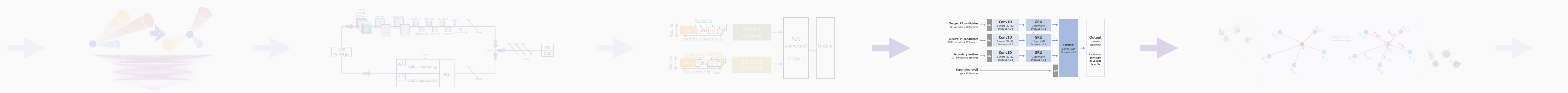


- ➔ using 2016 single-μ data
- ➔ SM (Herwig) shows the MC prediction using Herwig (instead of Pythia) for hadronization



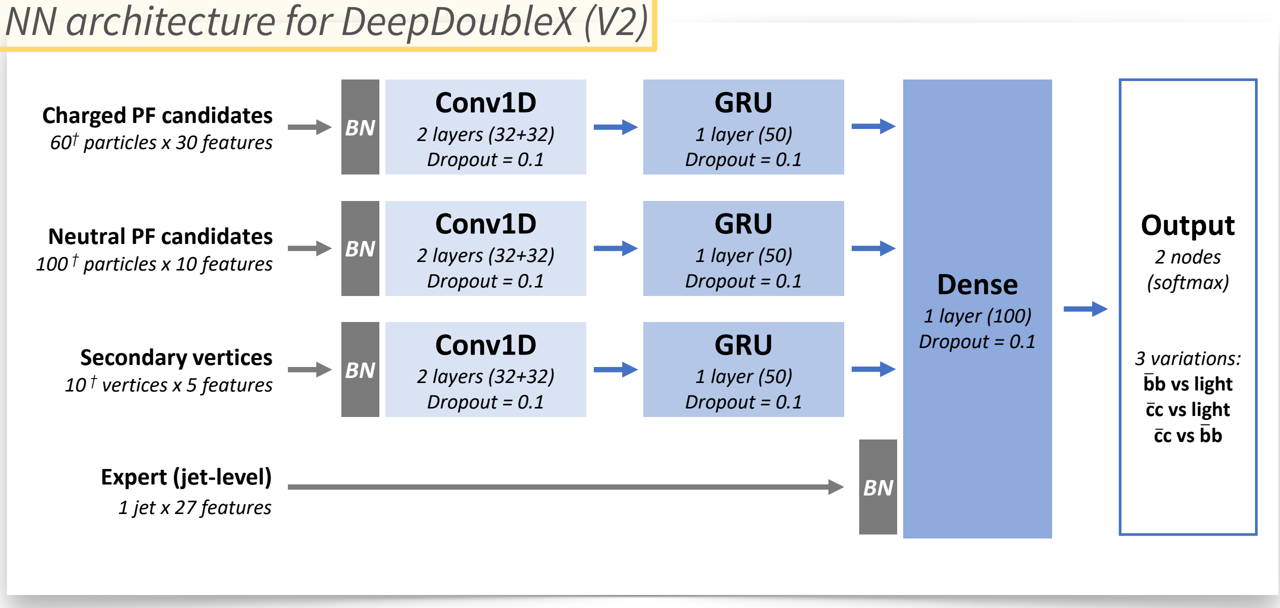


DeepDoubleX(-MD)



- ***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
- 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

NN architecture for DeepDoubleX (V2)



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