National and Kapodistrian
UNIVERSITY OF ATHENS

ΙΣΝ SNF ΙΔΡΥΜΑ ΣΤΑΥΡΟΣ ΝΙΑΡΧΟΣ
STAVROS NIARCHOS FOUNDATION

CMS

Identification of hadronically decaying heavy objects (W, Z, t, H) using Machine Learning techniques

Pantelis Kontaxakis
on behalf of the CMS Collaboration

HEP 2019 - NCSR “DEMOKRITOS”

April 20, 2019

Introduction

- **Many new physics processes produce boosted top quarks or gauge bosons (W/Z/Higgs) as signature**

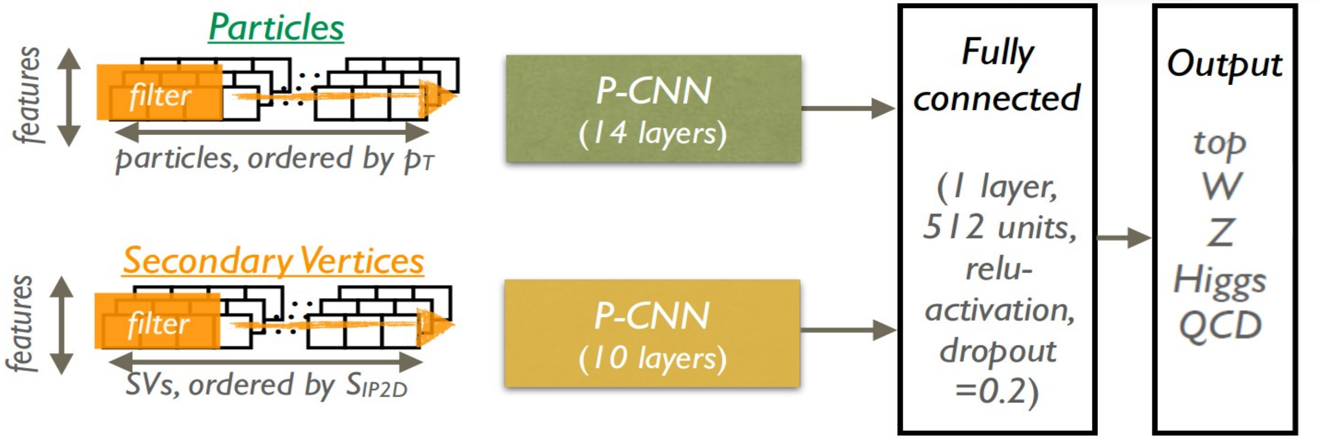
- **Reconstruction and identification of these boosted objects (top/W/Z/Higgs) from their hadronic decays provide powerful handles to probe new physics and SM measurements at the LHC**
  - Strong interest by the theory and the experiment communities

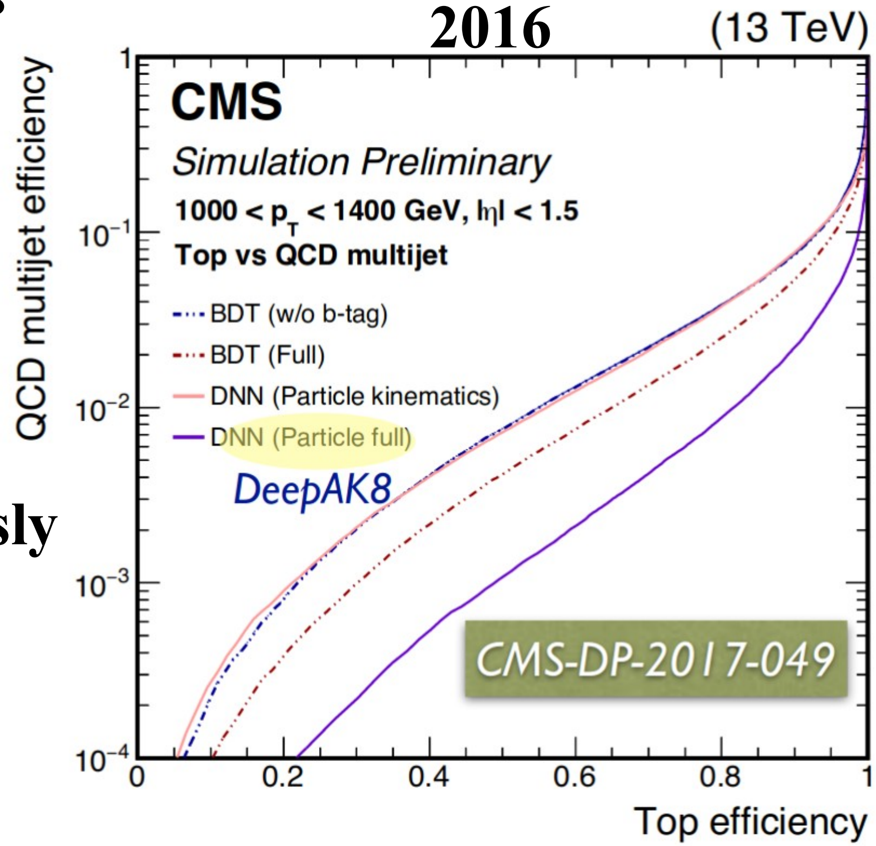
- **Heavy object tagging very challenging**
  - Large background from QCD jets, difficult to distinguish

- **DeepAK8 tagger:**
  - Heavy object tagger based in standard AK8 jets
  - Deep neural network (DNN) using 1D convolution technique
  - Two versions of DeepAK8:
    - → Nominal version: Aims at best possible performance but features mass sculpting
    - → Mass decorrelated version: Focus on minimizing the mass sculpting

DeepAK8

- **DeepAK8 tagger:** Multi-class classifier for top, W, Z, Higgs and QCD jets based on standard anti-k_T R=0.8 (AK8) jets
- **Use of low-level inputs (PF candidates, secondary vertices)**
  - Up to 100 PF candidates | 40 features/candidate
  - Up to 5 SV | 14 features/SV
- **Based solely on 1D convolution networks (P-CNNs)**
  - Residual networks
  - Move on particle triplets
- **Exploits substructure and flavor information simultaneously**





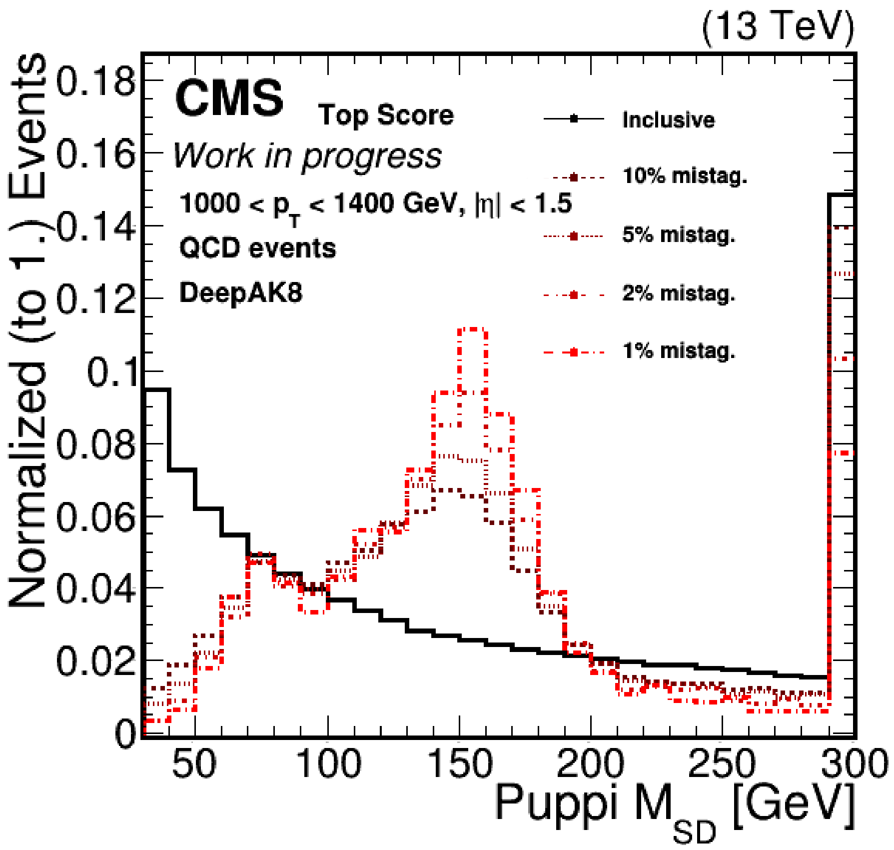
DeepAK8: Labels

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 features a fine-grained label definition**
  - Major categories defined for Higgs, top, W, Z by requiring the quarks from the heavy particle decay to be contained within the jet cone ($\Delta R(\text{jet},q)<0.8$)
    - ➔ Minor categories are further defined based on the flavor content of the matched quarks
  - For QCD jets, the minor categories are defined based on the number of b- or c-hadrons inside the jet
  - Full list shown in the table, with decreasing priority from top to bottom
- **This makes DeepAK8 a versatile tagger**
  - Prediction scores can be thought of as “probabilities” and can be easily transformed/aggregated for various needs, without the need of a dedicated re-training

Jet Mass Correlation

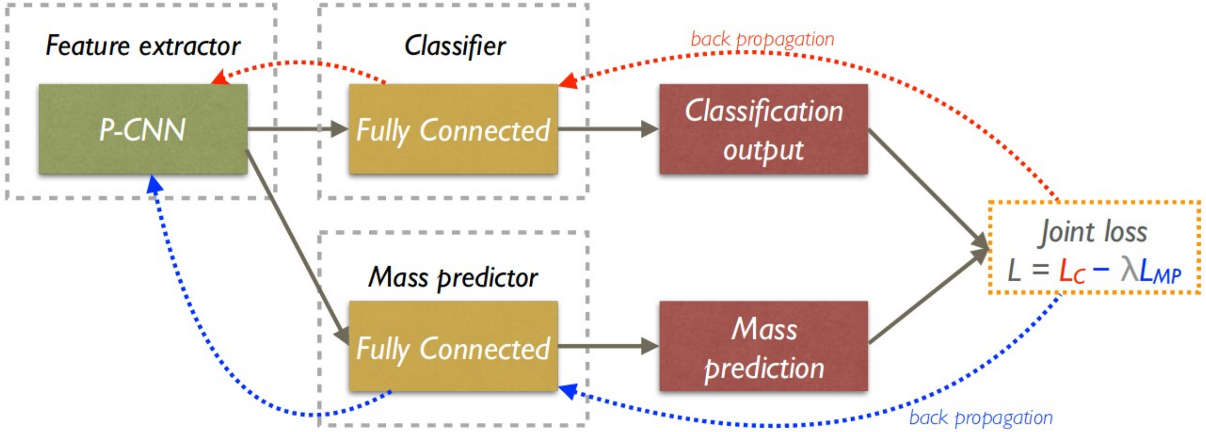
- **The base version of DeepAK8 shows significantly improved performance**
- **But bkg jets selected by the DeepAK8 algorithm show a modified mass distribution similar to that of the signal**
  - The mass of a jet is one of the most discriminating variables
  - P-CNNs are able to extract features that are correlated to the mass
- **Mass sculpting itself is not necessarily a problem but such modification may be undesirable if the mass variable is used in the signal/bkg separation**



- **An alternative DeepAK8 algorithm is developed to be largely decorrelated with the mass of the jet, while preserving the discrimination power as much as possible by applying a penalty to reduce the mass sculpting.**

De – correlating the mass

- **We can regulate the P-CNNs so that the extracted features are not correlated with the jet mass**
  - The subsequent classifier will naturally inherit the mass independence
- **We use adversarial training to balance the behavior of the P-CNNs**
  - A mass prediction network is introduced with the goal of predicting the jet mass from the features extracted by the P-CNNs
  - Its loss, L_{MP}, is an indicator for mass correlation
    - Smaller $L_{MP} \leftrightarrow$ more accurate mass prediction (the inputs have a higher correlation with the jet mass)
  - A joint loss is introduced as $\mathbf{L}=\mathbf{L_C} - \boldsymbol{\lambda}\mathbf{L_{MP}}$, with the 2nd term as a penalty on mass correlation
    - Minimizing L $\leftrightarrow$ Simultaneously improving classification and reducing mass correlation
    - $\boldsymbol{\lambda}$: a hyperparameter that controls the relative importance of the loss of the classifier ($\mathbf{L_C}$) and the loss of the adversary ($\mathbf{L_{MP}}$)



Kullback – Leibler (KL) Divergence

- **Used to measure the difference between two probability distributions over the same variable**
  - The KL divergence of q(x) from p(x) is a measure of the information lost when q(x) is used to approximate p(x)
- **The KL divergence of q(x) from p(x) (x discrete random variable) is defined as:**

$$D_{KL}(p(x)\|q(x)) = \sum_{x\in X} p(x)\ln\left(\frac{p(x)}{q(x)}\right)$$

- **KL divergence actually measures the expected number of extra bits requested to code samples from p(x) when using a code based on q(x), rather than using a code based on p(x)**

- **p(x) represents the “true” distribution of data, observations etc**

- **q(x) typically represents a model or approximation of p(x)**

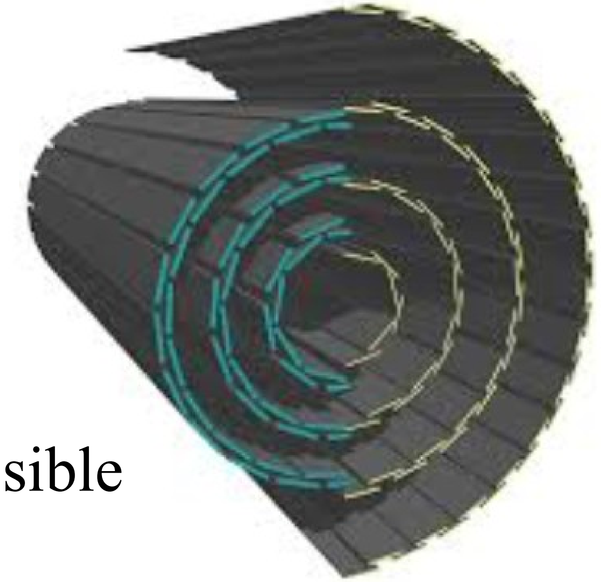
- **Usage of KL Divergence for DeepAK8 performance evaluation:**
  - Good figure of merit to measure the “distance” between two distributions and evaluate the mass sculpting of the Background jets

DeepAK8 for 2017 and 2018

- **For 2016:** DeepAK8 algorithm showed very good performance in Top/Z/W/Higgs tagging

- **For 2017 onwards:**
  - A new layer (new pixel detector) has been added in the tracker of the CMS detector
    - ➔ Potentially improvement in tagging with DeepAK8 as well
  - A dedicated retraining is needed to best exploit the new pixel detector

- **Retraining using 2017 Monte Carlo samples:**
  - Architecture same as 2016, but tuning the λ factor to get the best possible mass decorrelation



Pixel Detector

DeepAK8 Performance

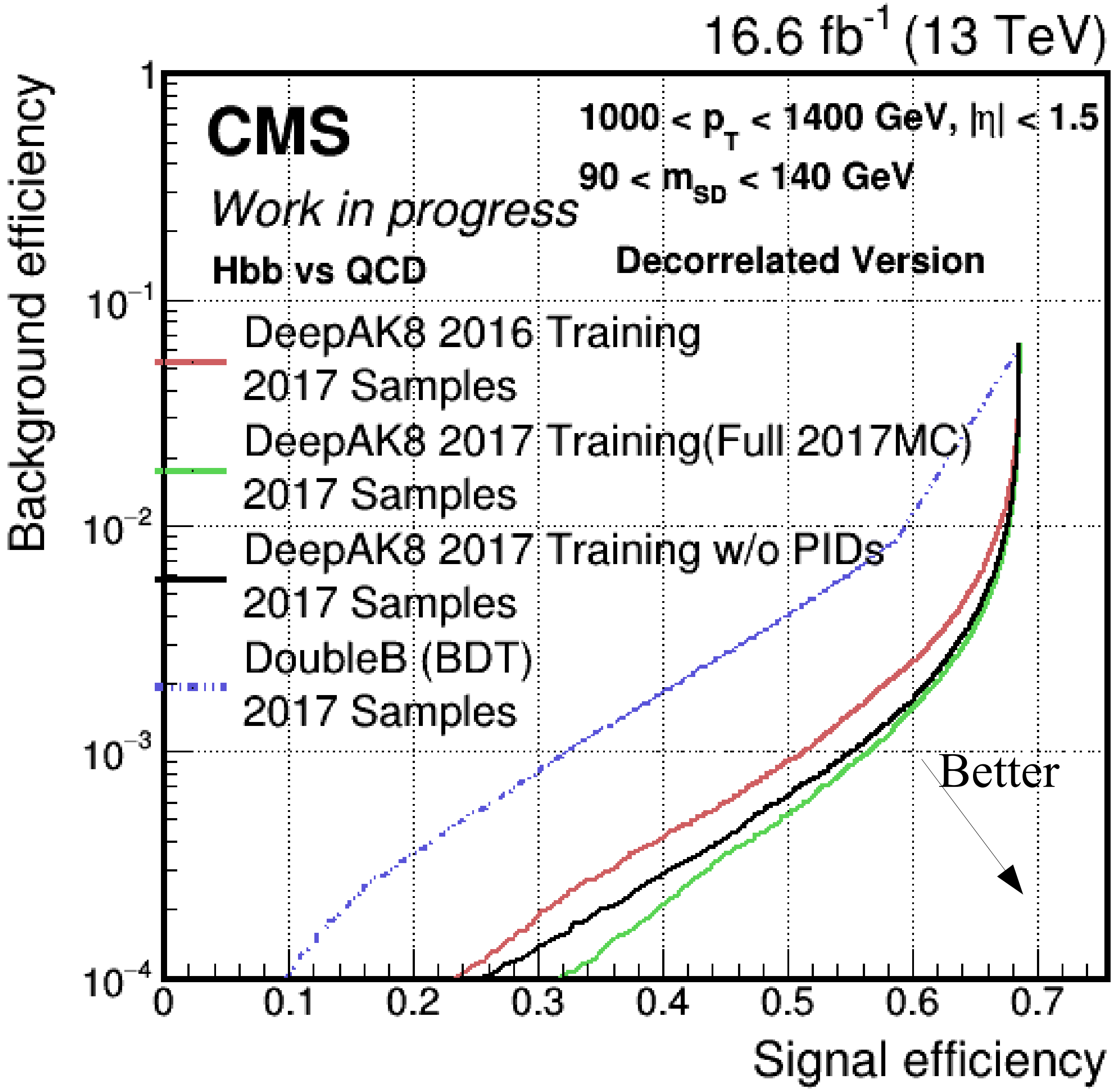
Description on versions used

- **DeepAK8 2016 Training | 2017 Samples:**
  Training done using the 2016 (phase0) samples, applied on 2017 (phase1) samples

- **DeepAK8 2017 Training (Full 2017MC) | 2017 Samples:**
  Training done using all the phase1 samples, applied on 2017 samples

- **DeepAK8 w/o PIDs | 2017 Samples:**
  Training using the phase1 samples without including the Particle Flow IDs as variables, applied on 2017 samples
  - Easier calibration + consistent with calibration strategy of other tagging algorithms

Performance: H→bb (“ZHbbvsQCD” score) #1

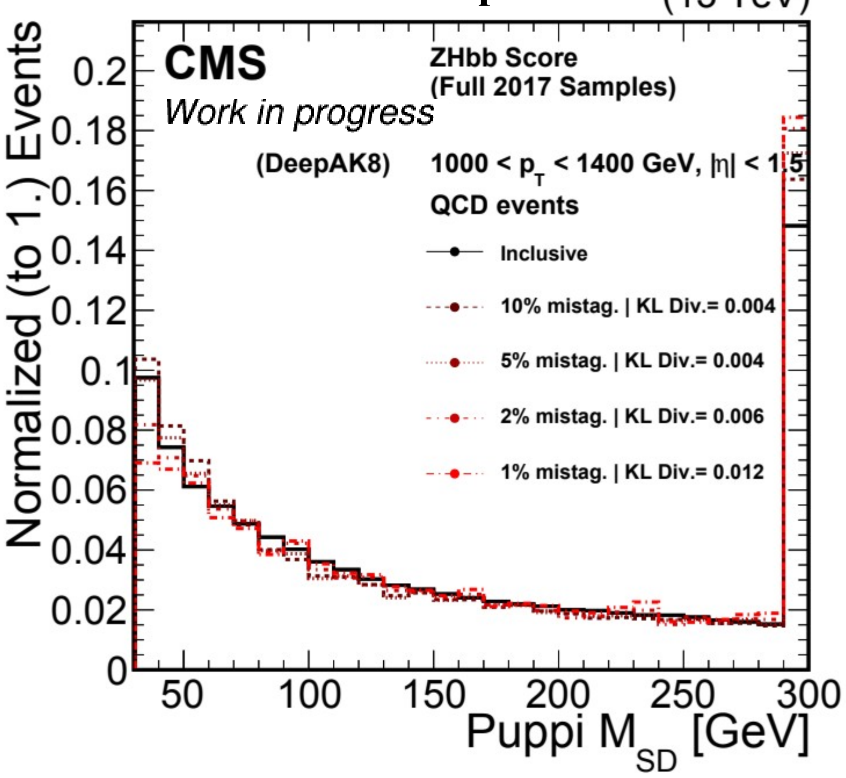


- **Performance for H→bb tagging**
  - Significant gain (10-20%) when training with 2017 samples, compared to training with 2016 samples
  - Removing Particle IDs from DeepAK8 training: Minimum loss in performance compared with the one when the Particle IDs are included
  - Way better performance for DeepAK8 compared with the DoubleB (BDT) algorithm

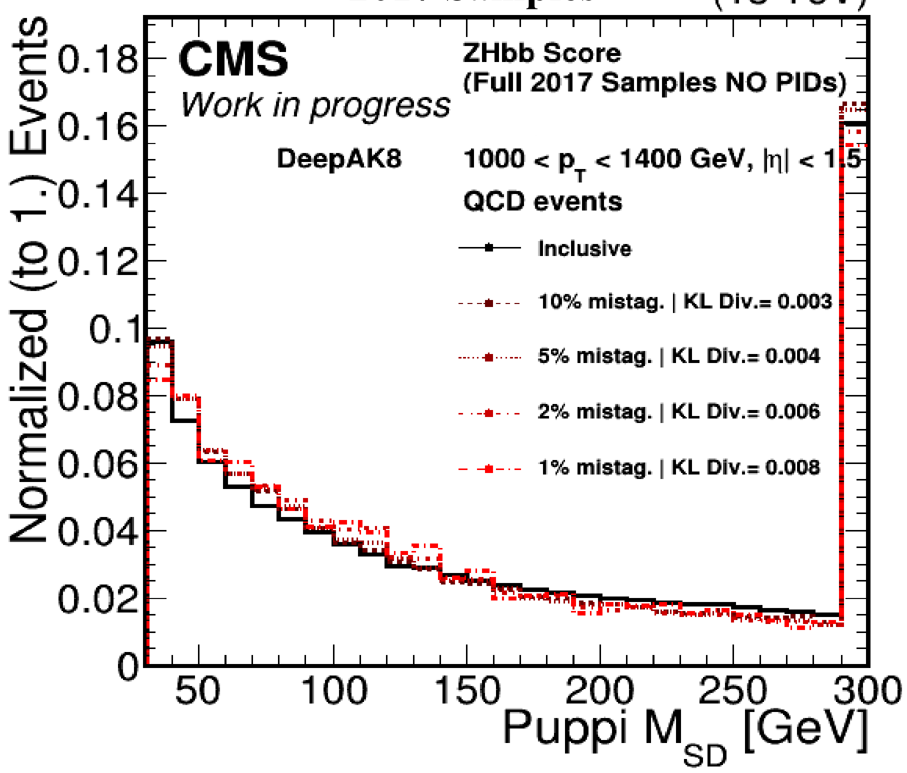
Performance: H→bb (“ZHbbvsQCD” score) #2

- **Mass sculpting for H→bb tagging (ZHbbvsQCD)**
  - Very good performance
  - Minimal Mass Sculpting for mistagging ~ 1%

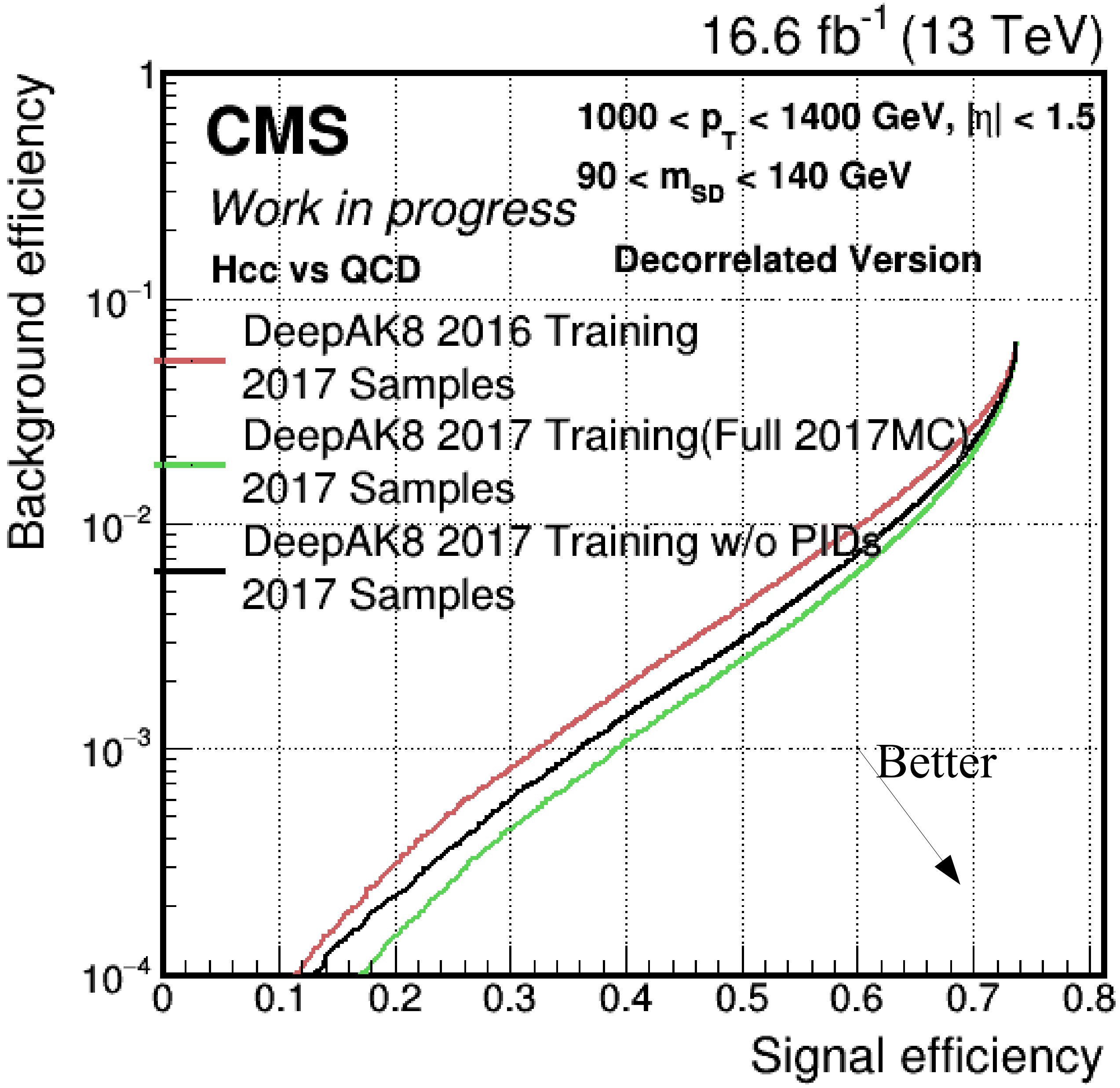
2017 Training
2017 Samples



2017 Training no PIDs
2017 Samples



Performance: H→cc ("ZHccvsQCD" score) #1

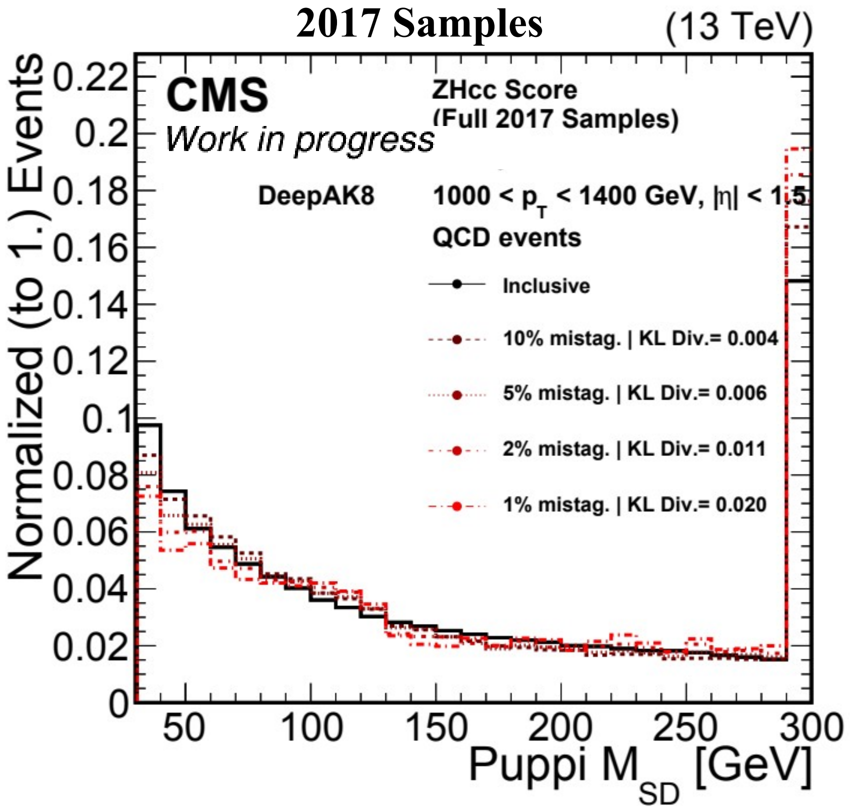


- **Performance for H→cc tagging**
  - Significant gain when training with 2017 samples, compared to training with 2016 samples
  - Removing Particle IDs from DeepAK8 training: Minimum loss in performance as the one when the Particle IDs are included

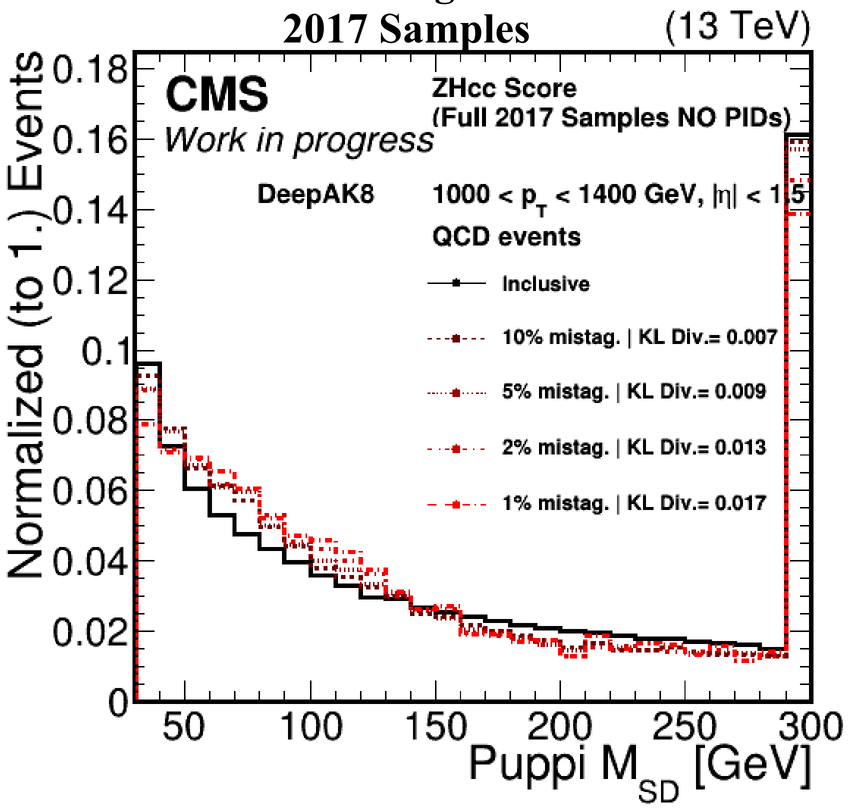
Performance: H→cc (“ZHccvsQCD” score) #2

- **Mass sculpting for H→cc tagging (ZHccvsQCD)**
  - Very Good performance
  - Minimum Mass Sculpting for DeepAK8 trained w/ or w/o Particle IDs

2017 Training
2017 Samples

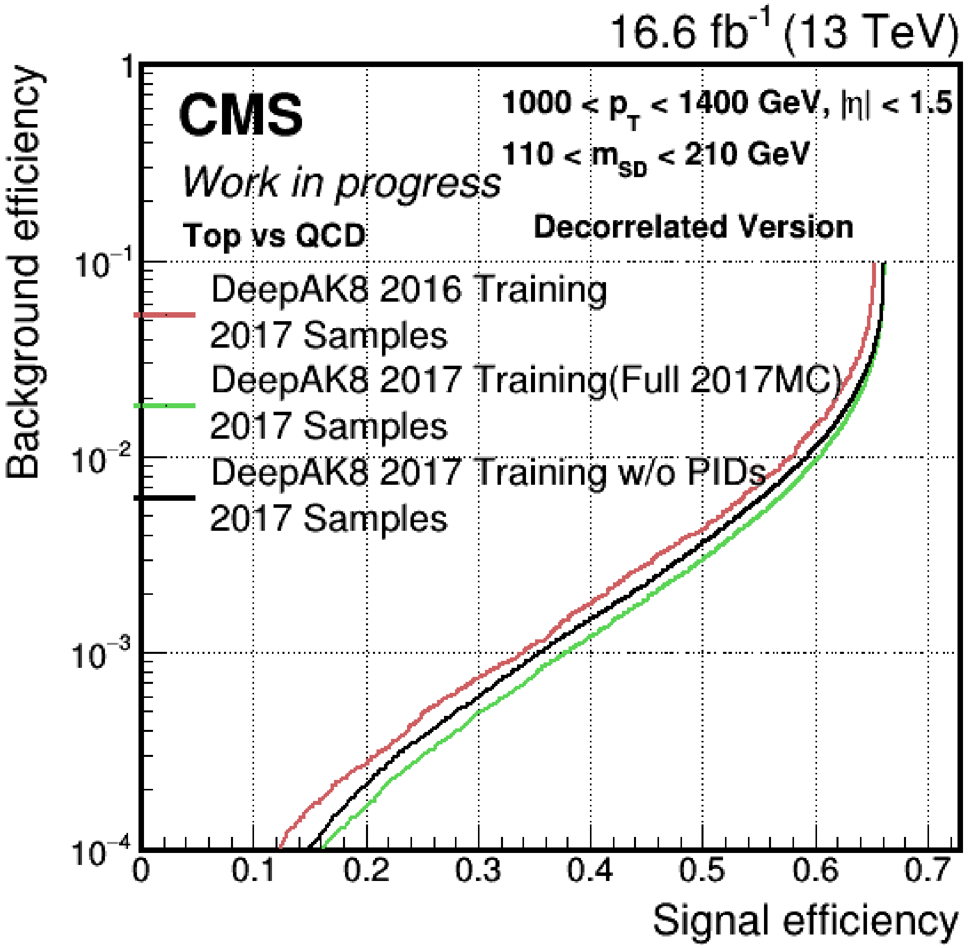


2017 Training no PIDs
2017 Samples

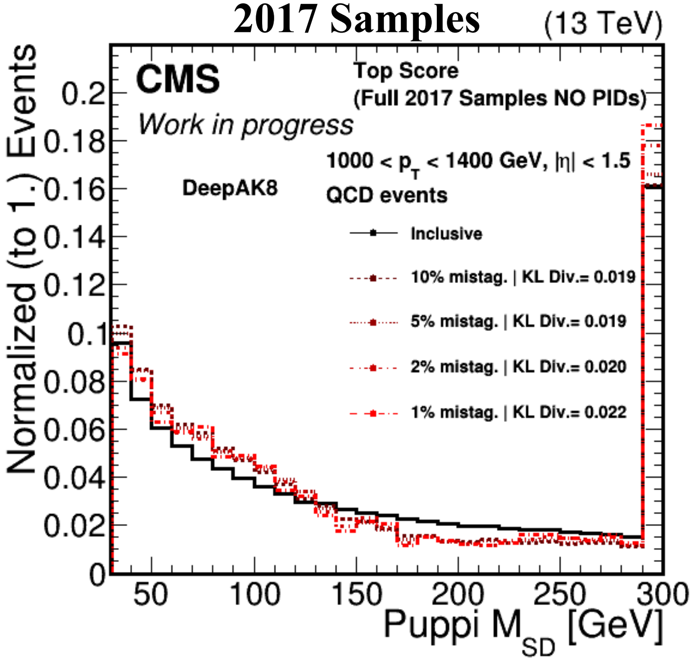


Performance: Top Tagging

- **Performance for top tagging**
  - Better performance when training with 2017 samples, compared to training with 2016 samples
  - Negligible mass sculpting
  - Smooth mass distribution for QCD bkg

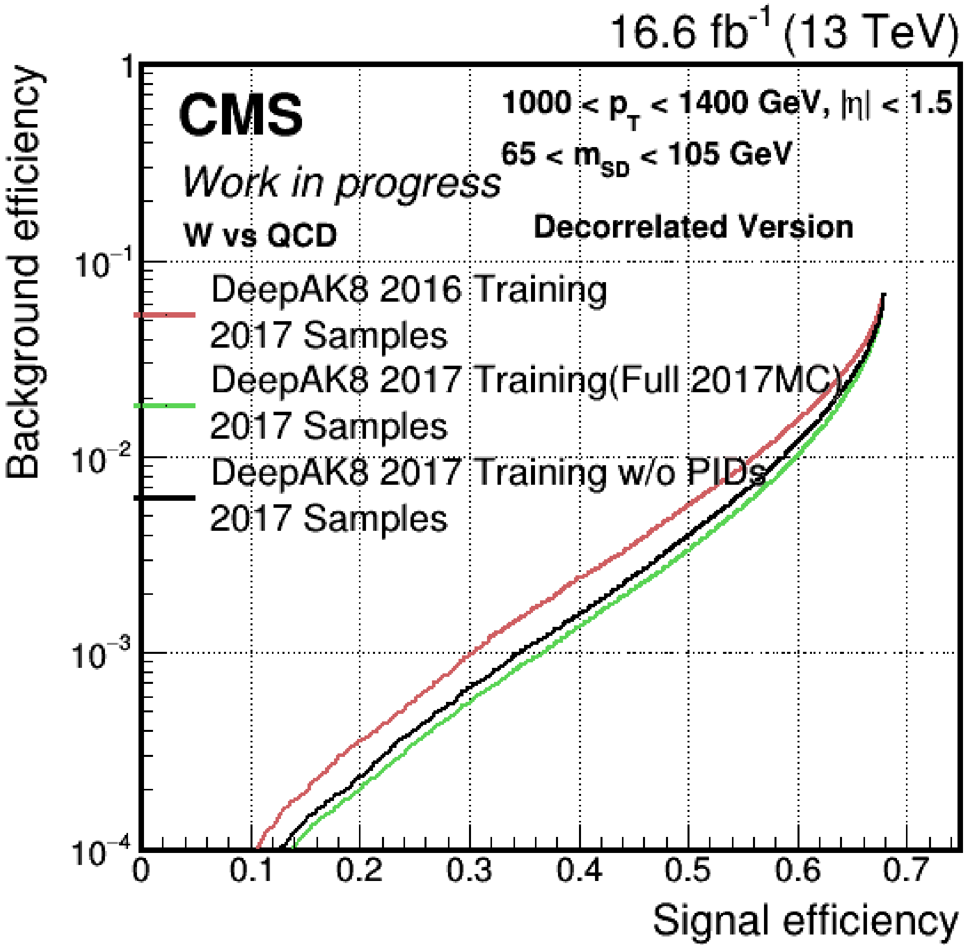


2017 Training no PIDs
2017 Samples

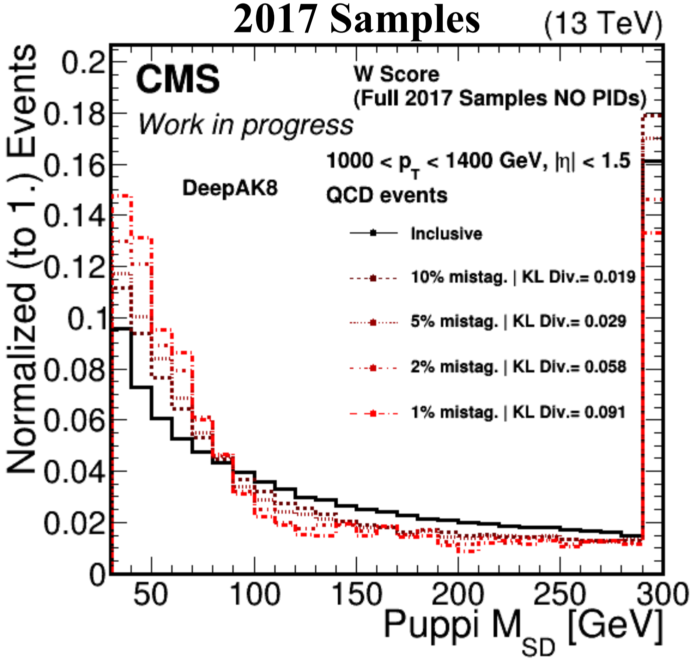


Performance: W Tagging

- **Performance for W tagging**
  - Very good performance with 2017 samples
  - Small mass sculpting for mistaging > 5% for 2017 retraining
  - Working to improve behaviour for tight WPs



2017 Training no PIDs
2017 Samples



Summary & Outlook

- **Check the performance of DeepAK8 multiclass tagger using the new pixel detector**
  - **Focused on the mass decorrelated version**
  - **Compared the performance of:**

a) Apply 2016 training on 2017 samples

b) Dedicated training using 2017 sample to fully exploit the upgraded detector

c) Removed the Particle ID to make the calibration easier and consistent with calibration strategy of other tagging algorithms

➢ Small loss in performance

- **Show significant improvement in performance (~10-20%) for identifying decay modes with heavy flavor quarks [cc, bb]**
- **Mass sculpting looks reasonable:**
  - Need some more on controlling mass sculpting in the very tight working points
  - Working on further improving the training

The research presented was carried out within the framework of a Stavros Niarchos Foundation grant to the National and Kapodistrian University of Athens

DeepAK8 Group

- Paris Sphicas[1,3]
- Pantelis Kontaxakis[1]
- Huilin Qu[2]
- Loukas Gouskos[3]
- Joe Incandela[2]

1) University of Athens, 2) UCSB, 3) CERN

BackUp Slides

Samples for 2017 DeepAK8 training

- **Signal:**
  - **TT:** - /ZprimeToTT_M*_TuneCP2_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM

    - /ZprimeToTT_M*_TuneCP2_13TeV-madgraphMLM-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14_**ext1**-v1/MINIAODSIM
  - **WW:** /BulkGravToWW_*_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
  - **ZZ:** /BulkGravToZZToZhadZhad_*_13TeV-madgraph/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
  - **H→VV:**/BulkGravTohhTohVVhbb_*_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
  - **HH→4b:** /GluGluToBulkGravitonToHHTo4B_*_13TeV-madgraph_correctedcfg/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
  - **HH→4c:**/GluGluToBulkGravitonToHHTo4C_*_13TeV-madgraph-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
- **Background:**
  - QCD_Pt_*_TuneCP5_13TeV_pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM

Score Definition

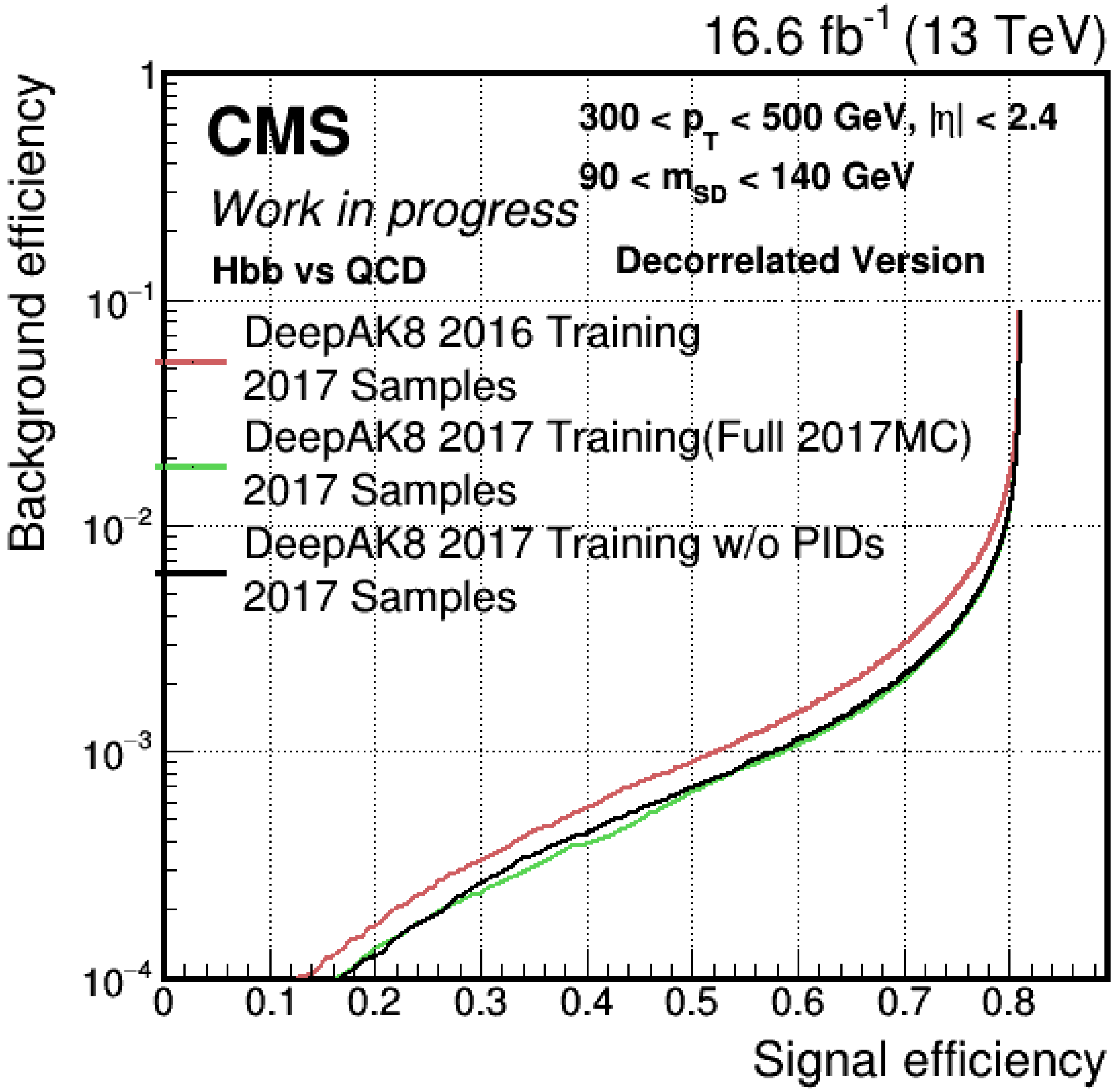
$$ZHbbvsQCD := \frac{Zbb + Hbb}{Zbb + Hbb + QCD_{b,bb,c,cc,others}}$$

$$ZHccvsQCD := \frac{Zcc + Hcc}{Zcc + Hcc + QCD_{b,bb,c,cc,others}}$$

DeepAK8 Performance

$300 < p_T^{AK8Jet} < 500$ GeV, $|\eta| < 2.4$

Performance: H→bb (“ZHbbvsQCD” score) #1

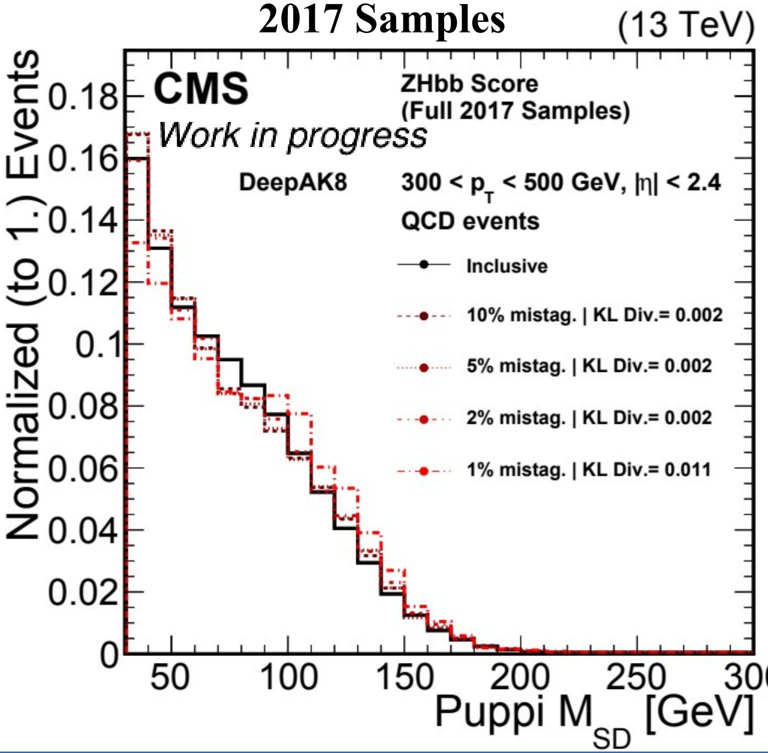


- **Performance for H→bb tagging**
  - Significant gain when training with 2017 samples
  - No loss in performance when Particle IDs are not included for training

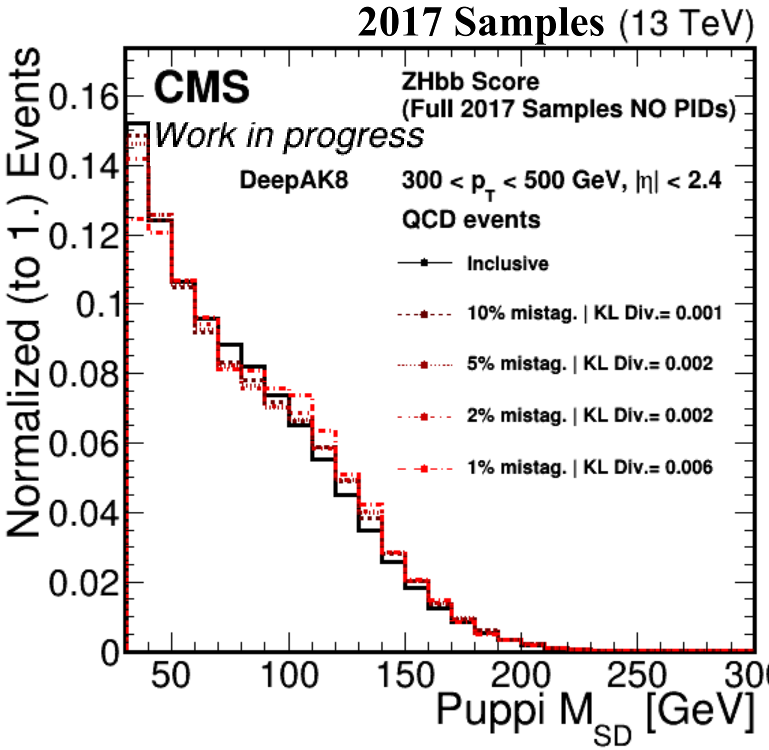
Performance: H→bb ("ZHbbvsQCD" score) #2

- **Mass sculpting for H→bb tagging (ZHbbvsQCD)**
  - Very Good performance
  - Minimal Mass Sculpting for 2017 training
  - Better behavior compared to 2016 training

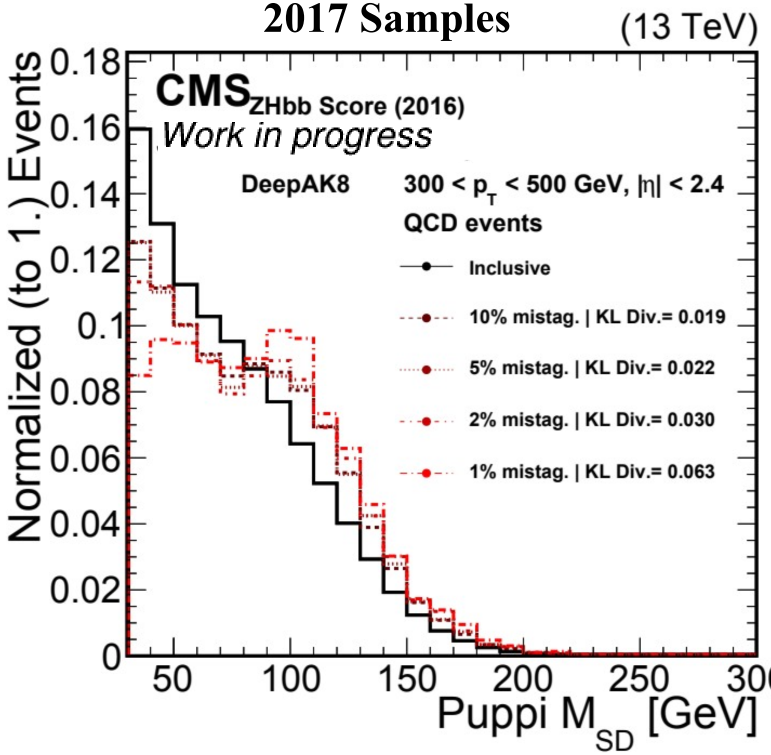
2017 Training
2017 Samples



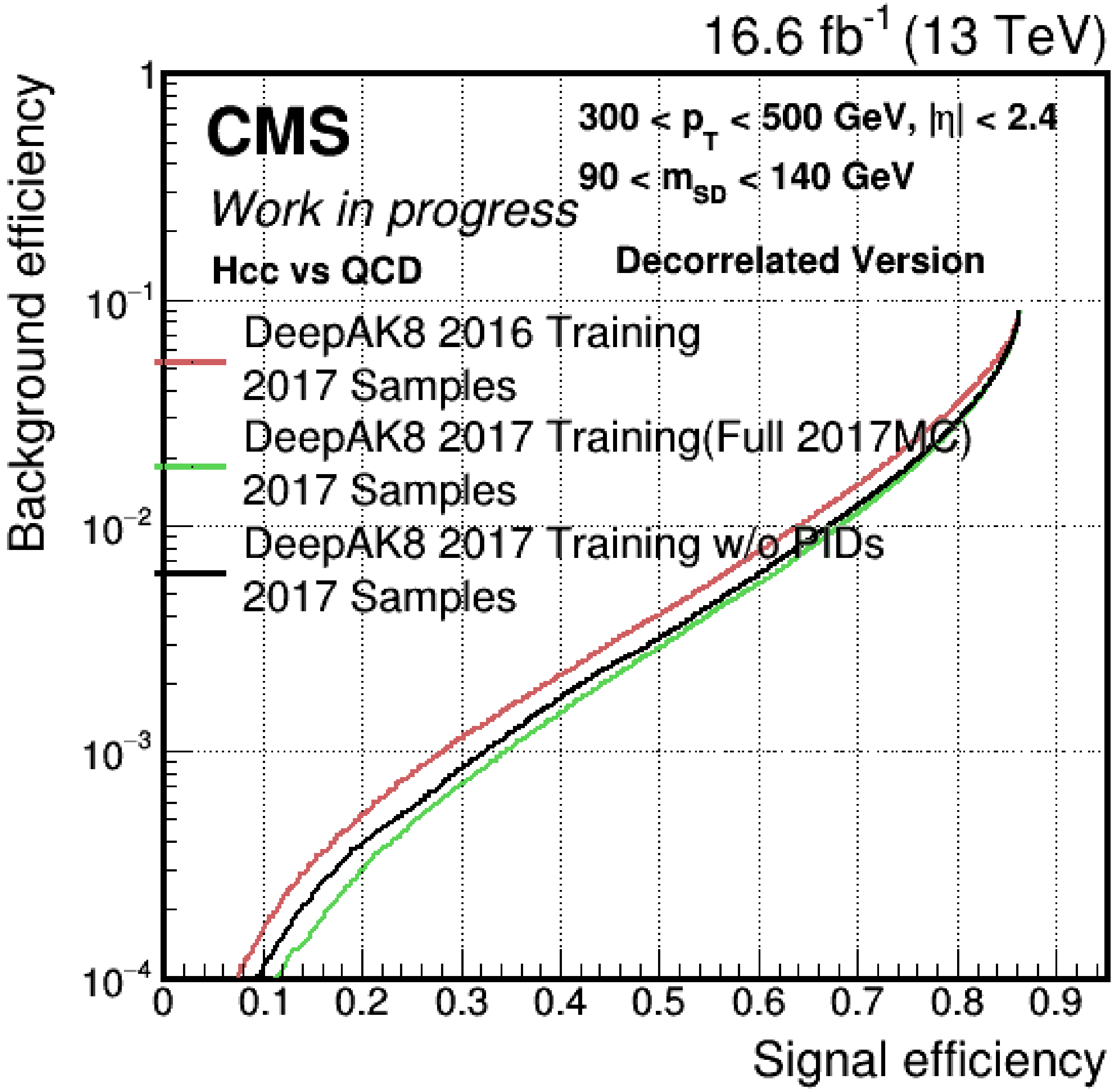
2017 Training no PIDs
2017 Samples



2016 Training
2017 Samples



Performance: H→cc (“ZHccvsQCD” score) #1

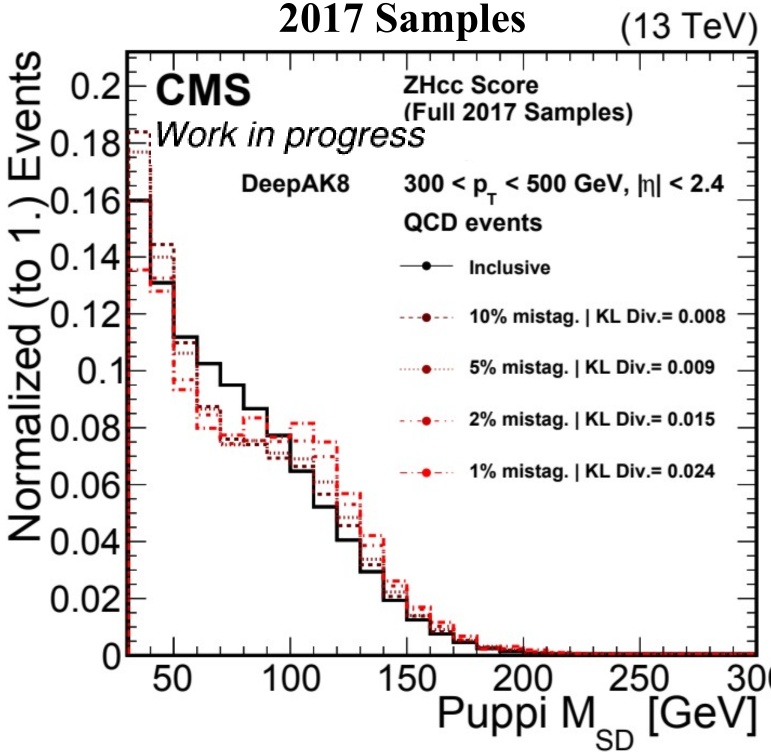


- **Performance for H→cc tagging**
  - Significant gain when training with 2017 samples and for low p_T region
  - Minimal loss in performance when Particle IDs are not included for training

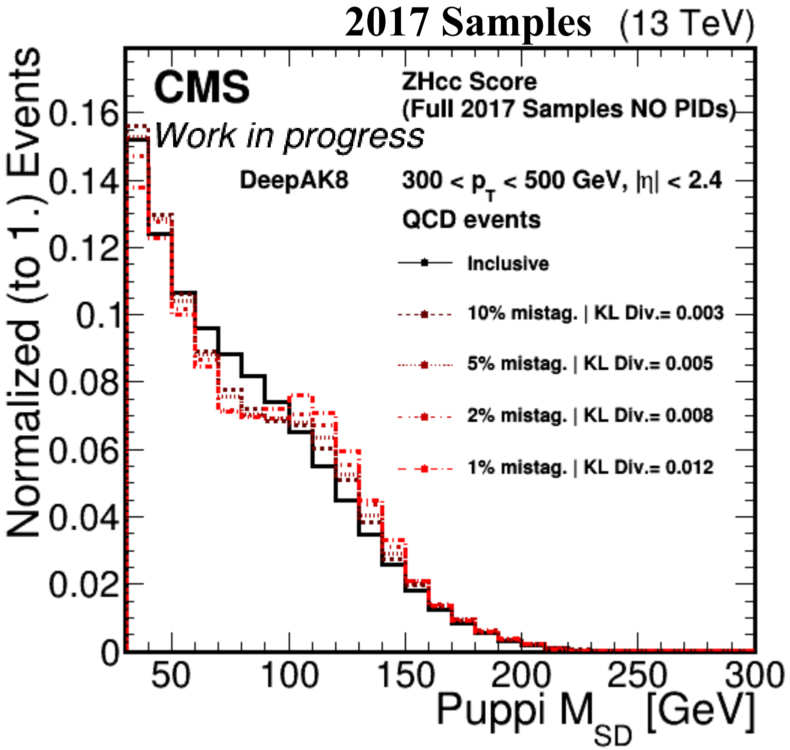
Performance: H→cc (“ZHccvsQCD” score) #2

- **Mass sculpting for H→cc tagging (ZHccvsQCD)**
  - Reasonable performance
  - Lower mass sculpting compared with 2016 training
  - Investigation to improve mass sculpting

2017 Training
2017 Samples



2017 Training no PIDs
2017 Samples



2016 Training
2017 Samples