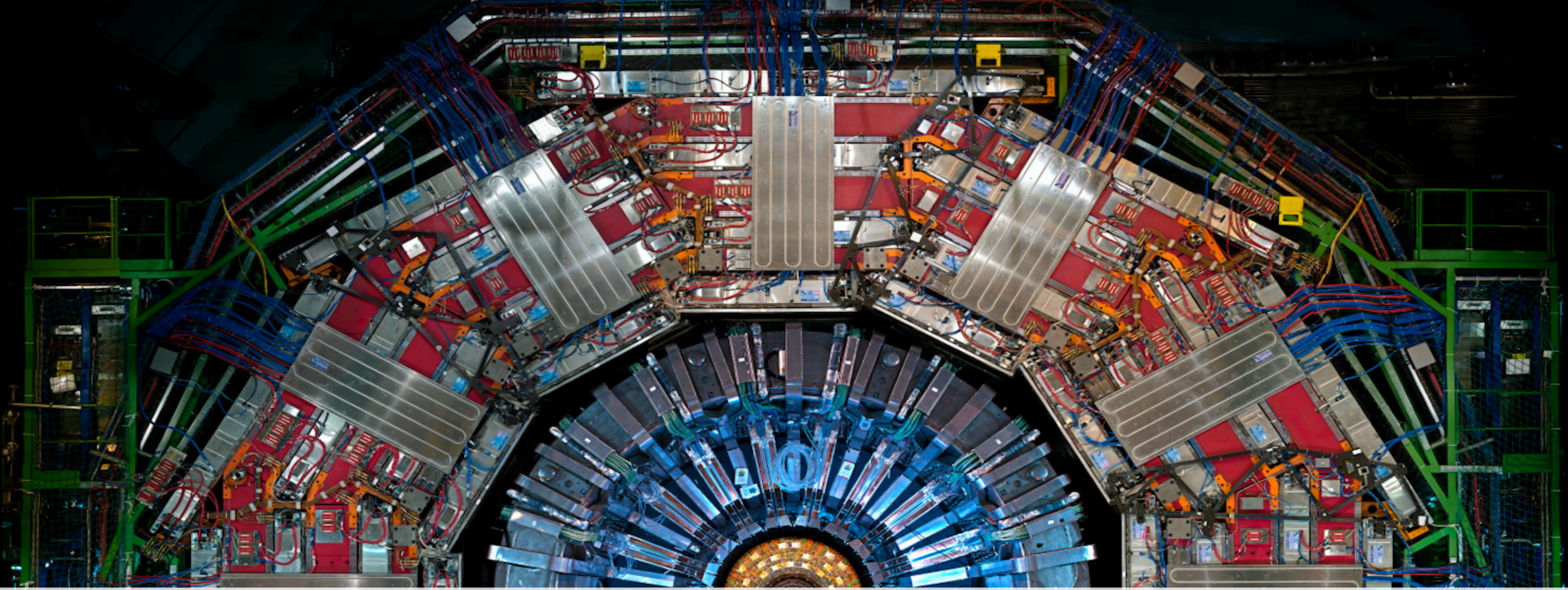


New(*) jet taggers and algorithms as tools for SM and BSM physics at CMS

Loukas Gouskos (CERN)
on behalf of the CMS Collaboration

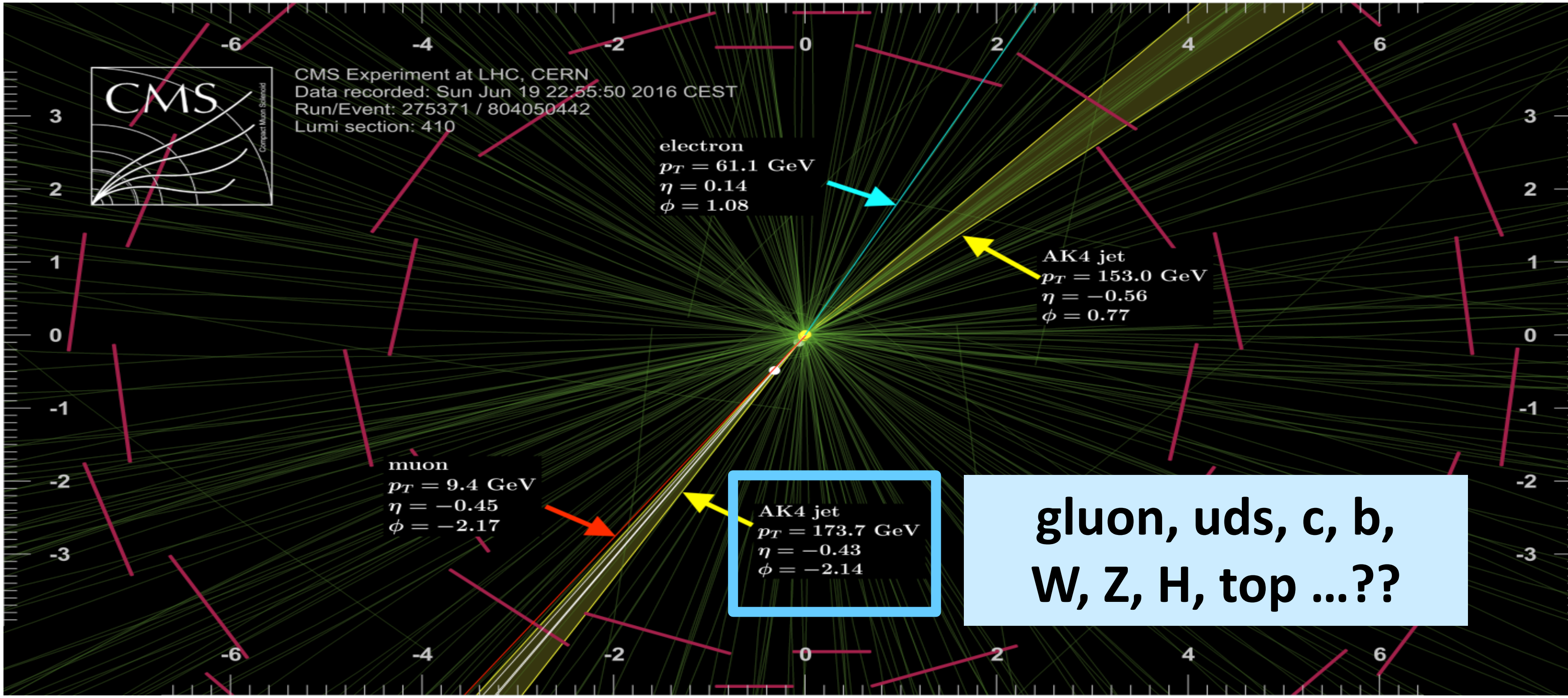
CERN-LHC Seminar
Nov 12th, 2019

(*)New = focus on the most recent results: CMS-PAS-JME-001, CMS-PAS-JME-002

Outline

- Introduction
- Jet reconstruction in CMS
  - Jet calibration
  - PU Mitigation techniques
  - Interim: highlights from physics results using these tools
- High-p_T Jet identification (”Boosted jet tagging”)
  - Design and performance in simulation
  - Calibration in data
  - Interim: highlights from physics results using these tools
- Summary & outlook

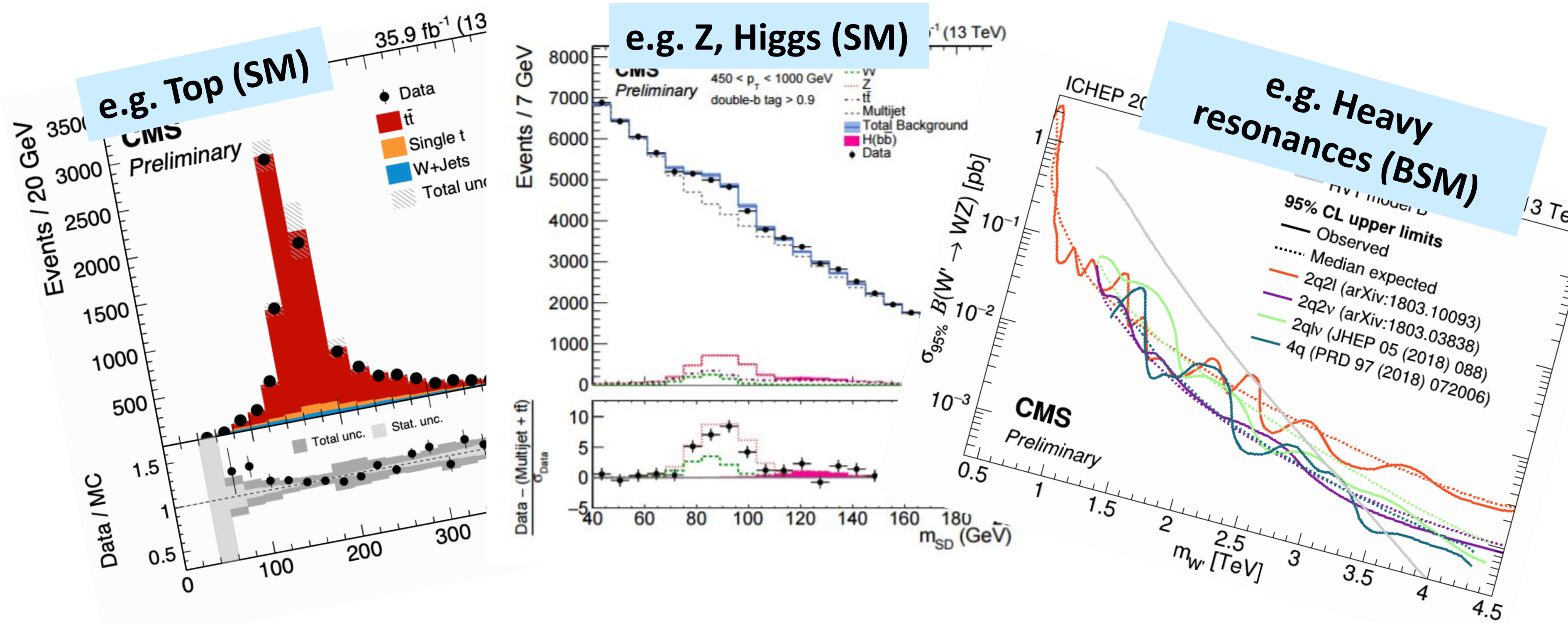
Introduction



- People have been tagging jets for almost 30 years at colliders
  - starting with b jets at LEP and Tevatron, then top, W/Z and Higgs jets at the LHC.
- But it is only now that we have begun to develop powerful and multi-object tagging capabilities.
  - potential to open access to many new physics topics that had been written off previously

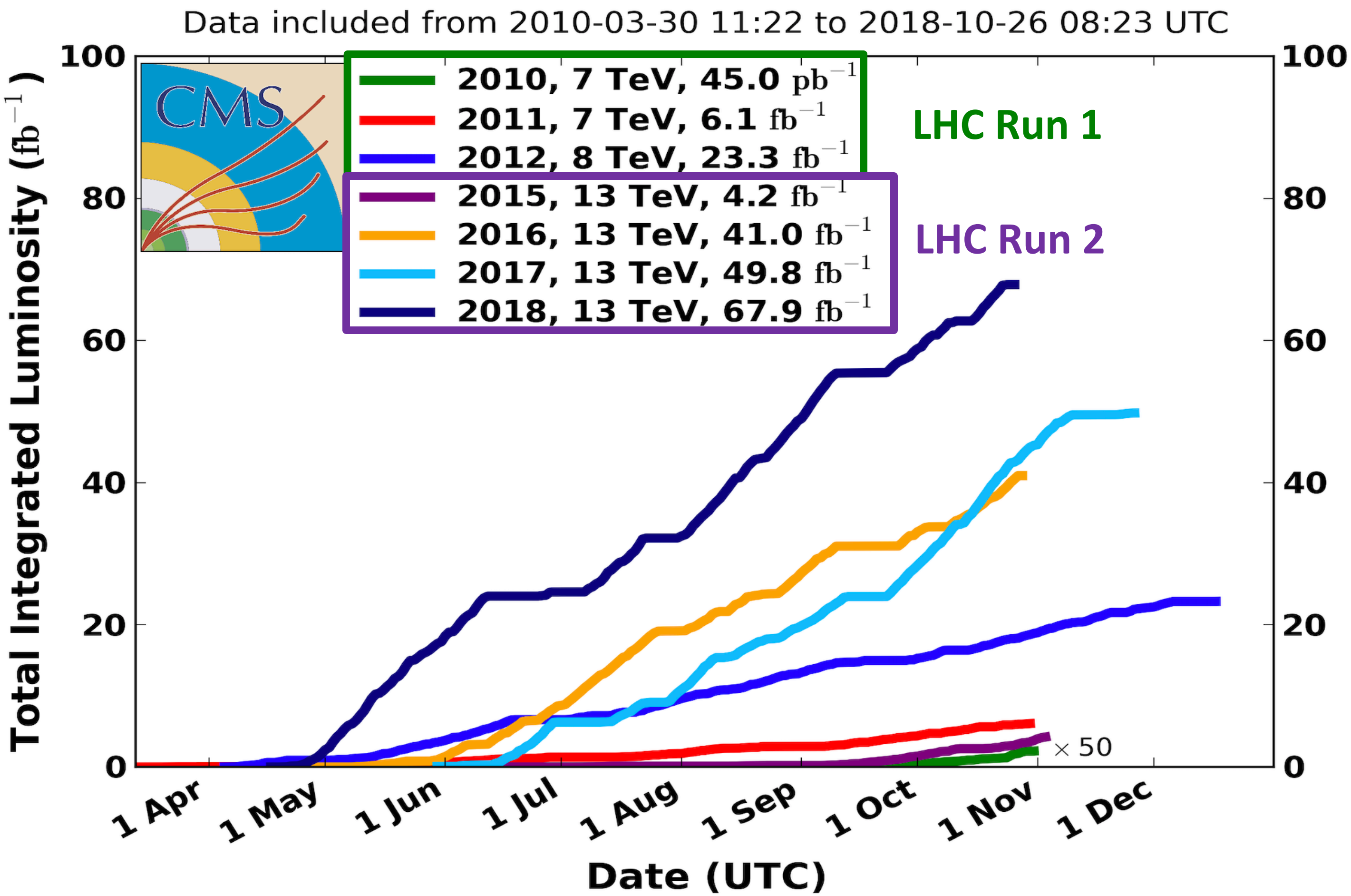
Introduction (II)

- Jets are essential for the majority of physics analyses at the LHC
  - both for **standard model (SM)** and **beyond SM (BSM)** physics
    - typically largest signal acceptance (compared to leptonic final states), improved signal purity at high-p_T regime, …



- **Key for success:** Well calibrated jets & constantly improving our "JetToolBox"
  - particularly important these days that LHC integrated luminosity increases only ~linearly with time & do not expect big jumps in collision energy

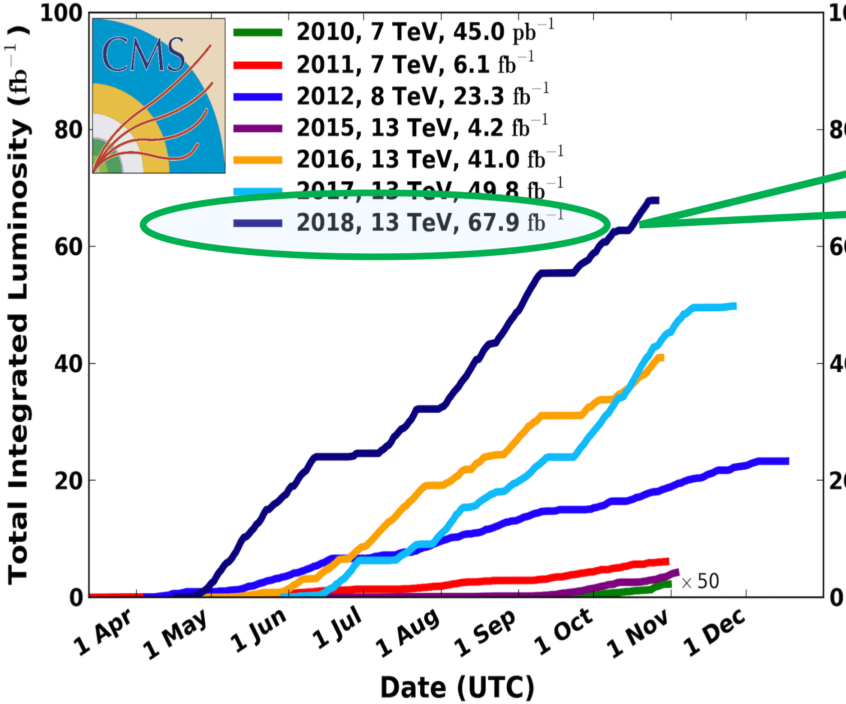
LHC performance after 10 yrs of operation



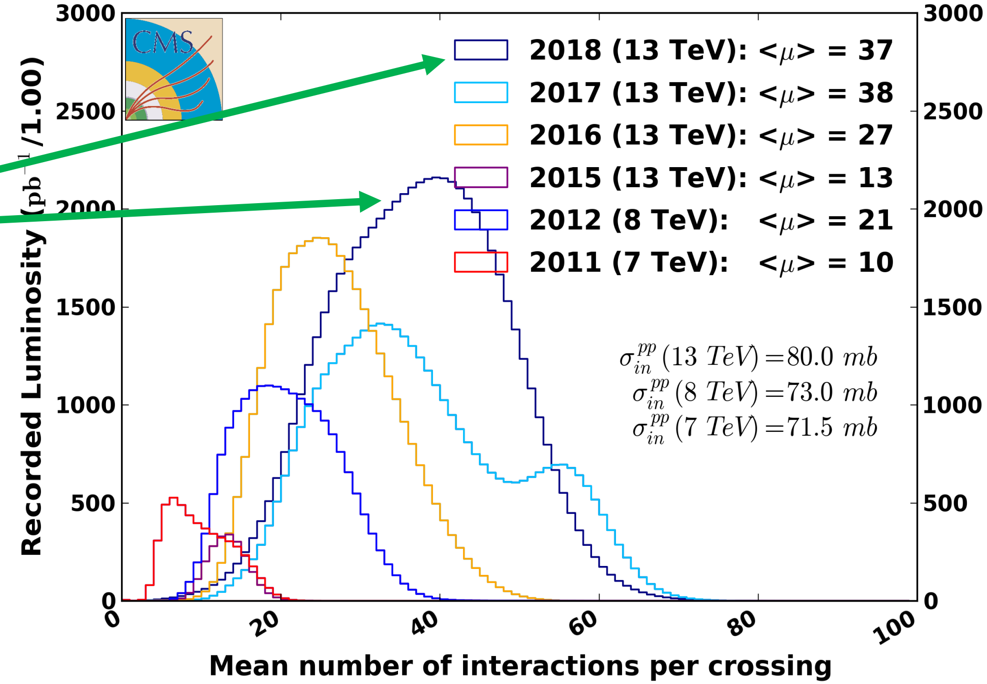
- **Remarkable performance**; exceeding initial expectations
  - We are just at the beginning; ~5% of the total expected data collected so far
- Great opportunity for sensitive SM measurements & BSM searches

Price to pay: “pileup”

Integrated luminosity



Average Pileup

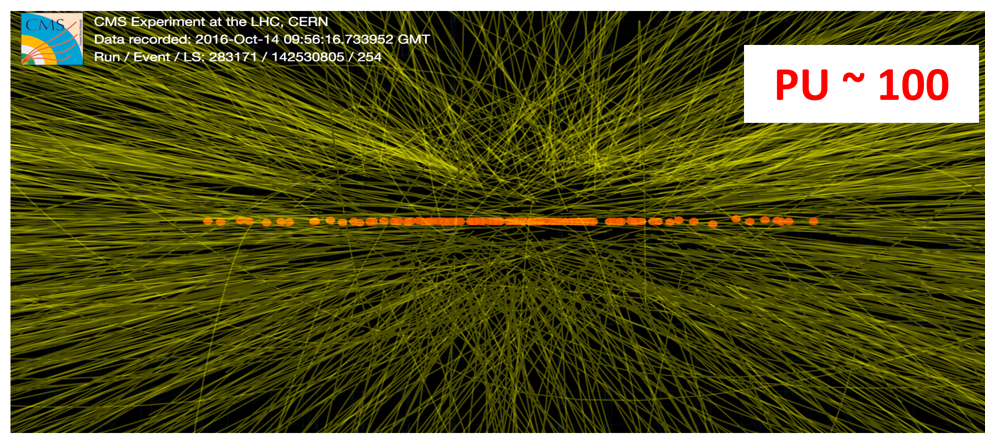


→ **LHC-Run2**: ~40

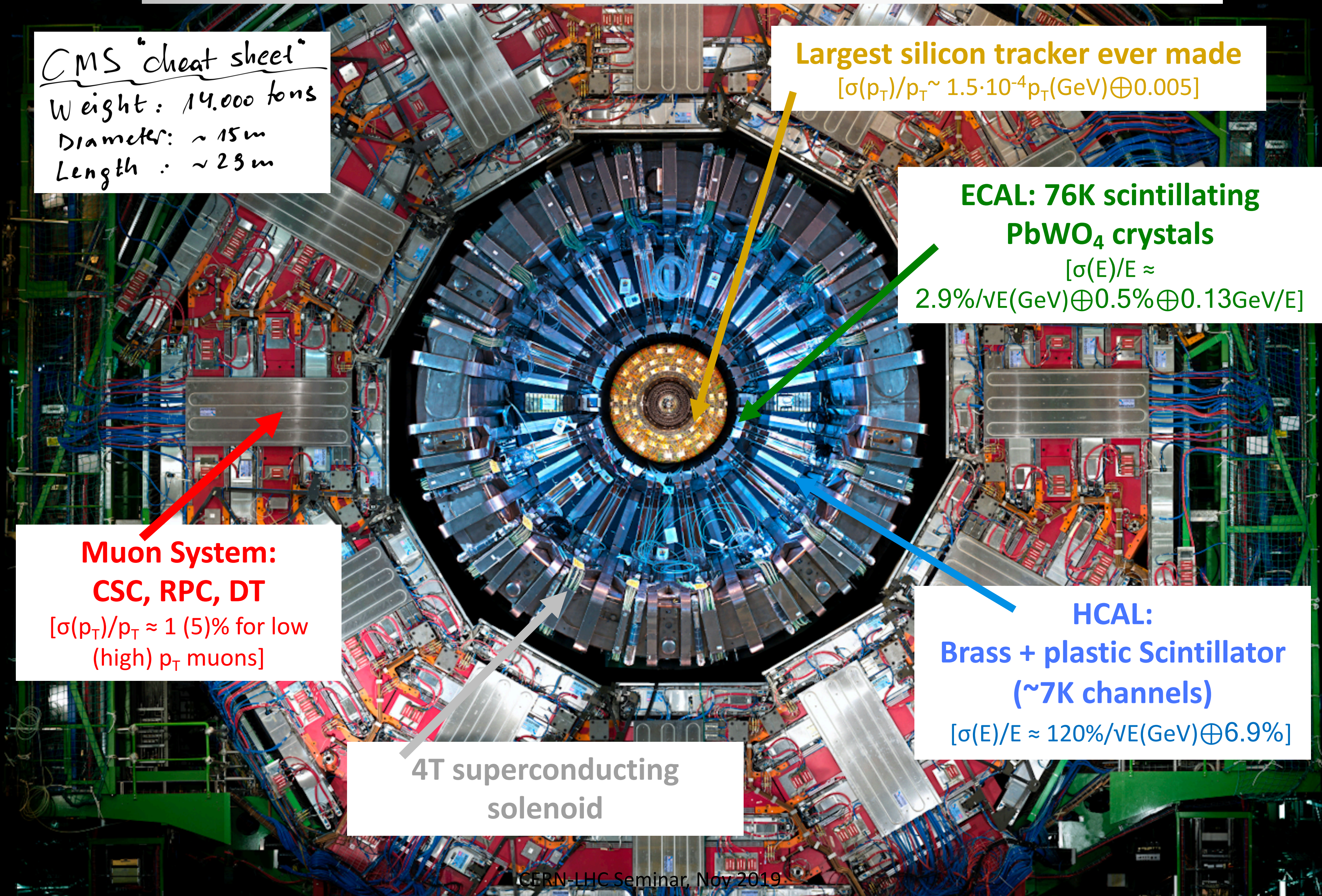
→ Expect: ~50 for **Run 3**;

~150-200 for **HL-LHC**

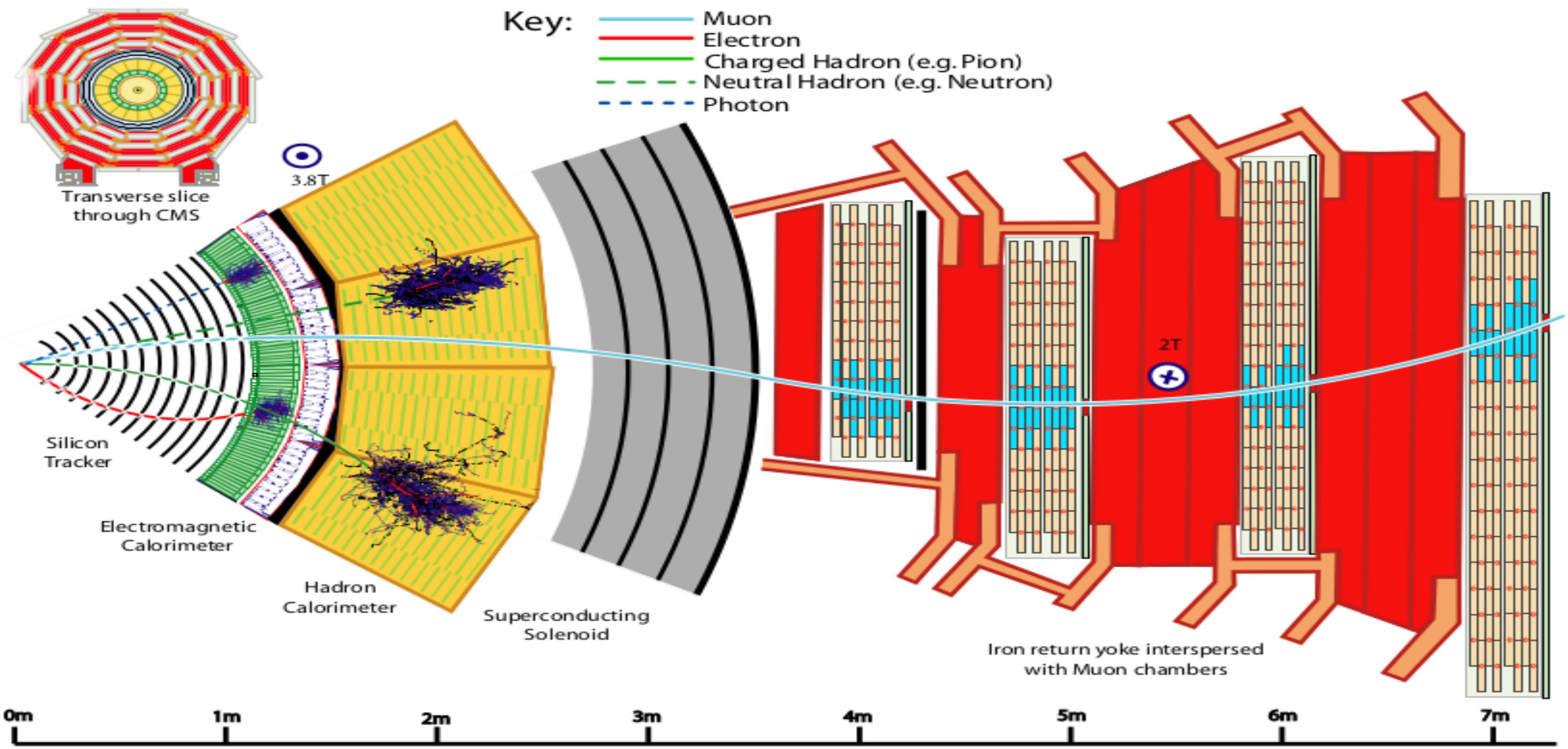
Major challenge for the experiments



The Compact Muon Solenoid experiment



Event reconstruction in CMS



- CMS Event Reconstruction using Particle Flow (PF) algorithm
  - Combines information from all subdetectors
  - Mutually exclusive list of particles
    - Then build higher level objects: e.g. jets, ME_T, …

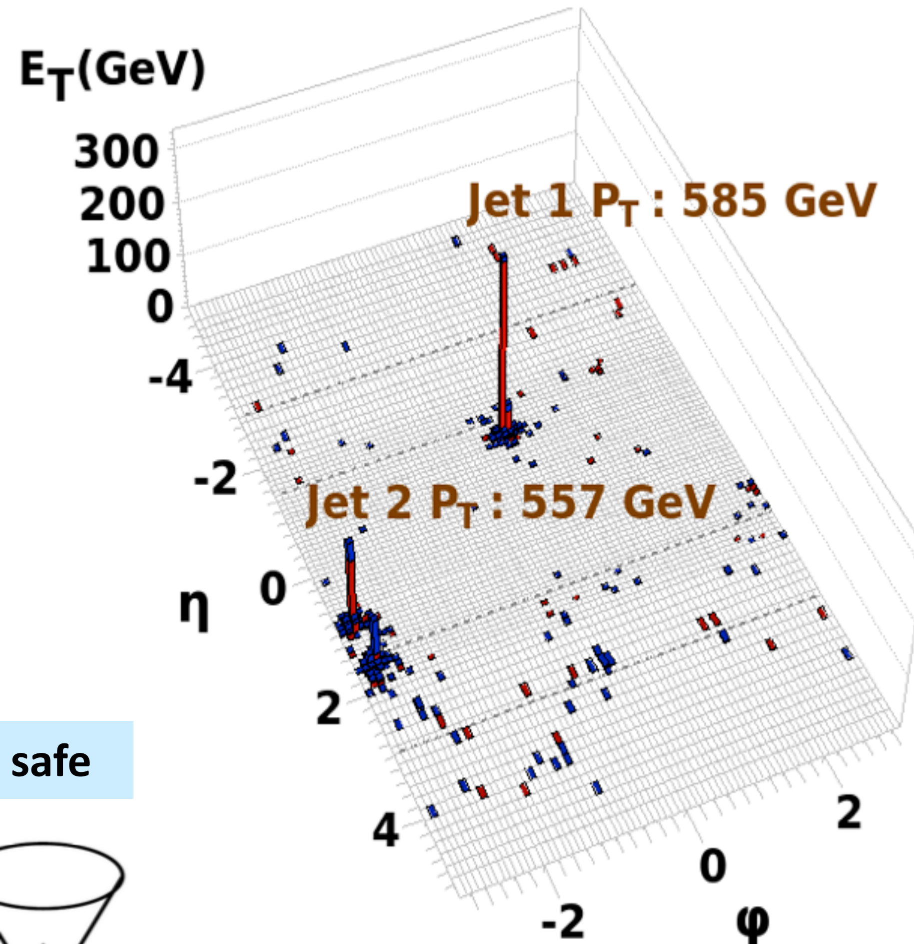
Significant improvement in object performance wrt traditional approaches

Jet reconstruction in CMS

[NB. “**New !**” : results from 2019]

From PF particles → PF jets

- **Jets:**
  reconstructed by clustering the particles returned by PF algorithm

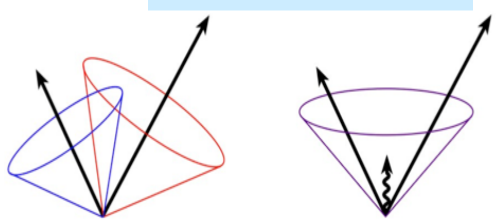


- **Anti-k_T (AK) clustering algorithm:**

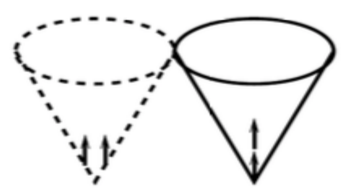
$$d_{ij} = min\left(\frac{1}{p_{ti}^2}, \frac{1}{p_{tj}^2}\right) \times \frac{R_{ij}^2}{R}$$

  - Clusters "around" the harder particle
  - Infrared & collinear (IRC) safe

Infrared safe



Collinear safe



- Typical jet collections in CMS: R=0.4 (AK4 jets), R=0.8 (AK8 jets)

Jet calibration: energy scale

- Follow a factorized approach:

Pileup correction

- Derived by comparing same jets with and w/o PU
- parametrized in: p_T, η, ρ

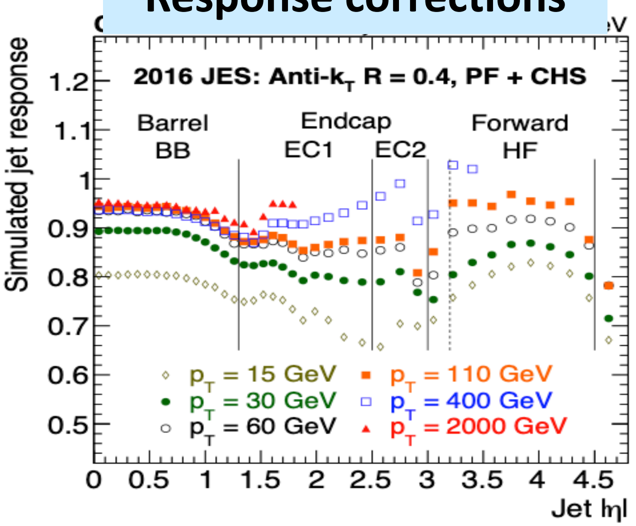
Response correction

- Particle level simulation-based corrections for jet p_T
- Extracted as a function of η, p_T

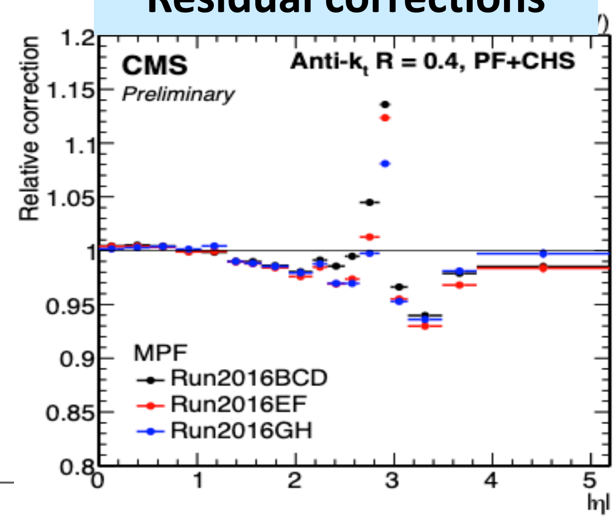
Residual corrections

- account for residual differences between data & simulation
- data-driven approach

Response corrections



Residual corrections



CMS-DPS-2018-028

Jet response very close to unity after all corrections

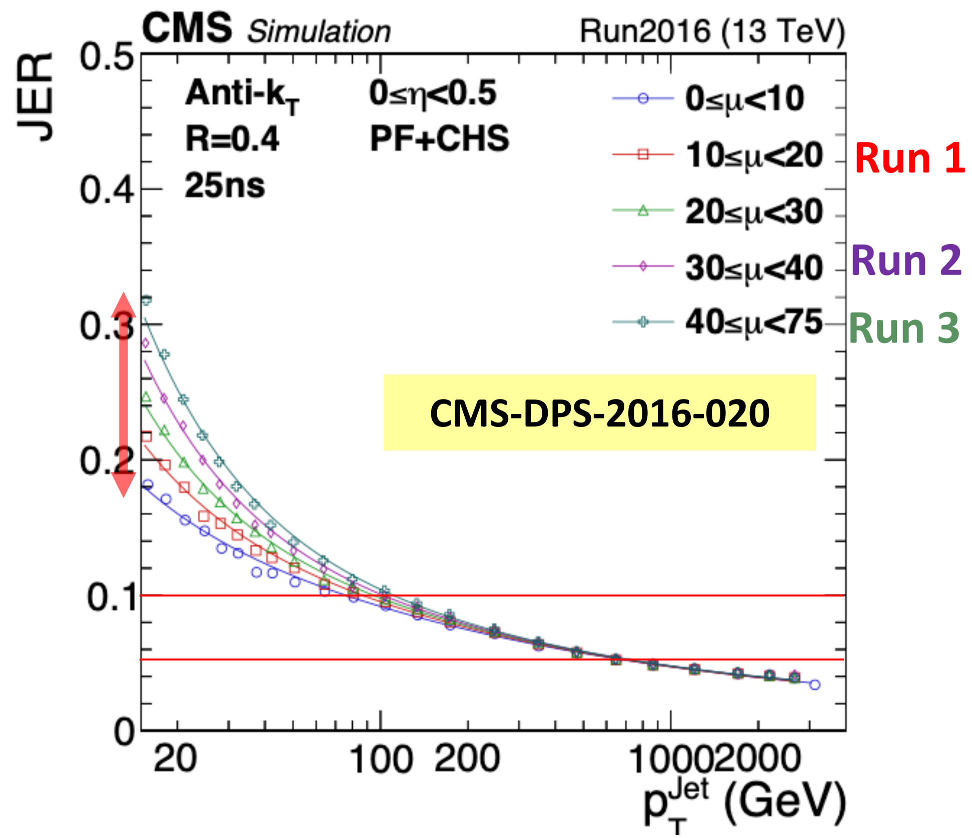
Jet calibration: energy resolution

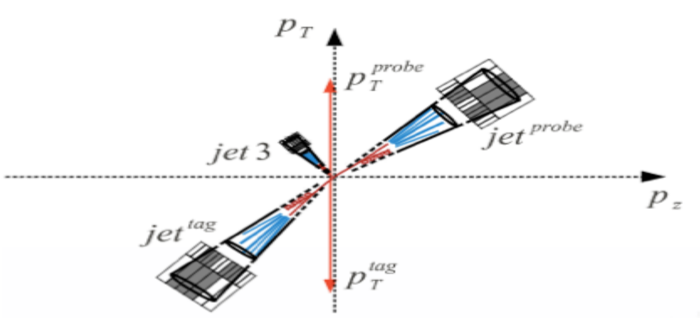
- Jet energy resolution (JER):

$$\mathrm{JER} = \sigma\left(\frac{\langle p_\mathrm{T}\rangle}{\langle p_{\mathrm{T,ptcl}}\rangle}\right)$$

  - Better than 10% (5%) for p_T(jet)>100 (1000) GeV

- Calibration in data using dijet, Z+jets, and γ+jets samples
  - Data/MC correction factors: ~1.1-1.2



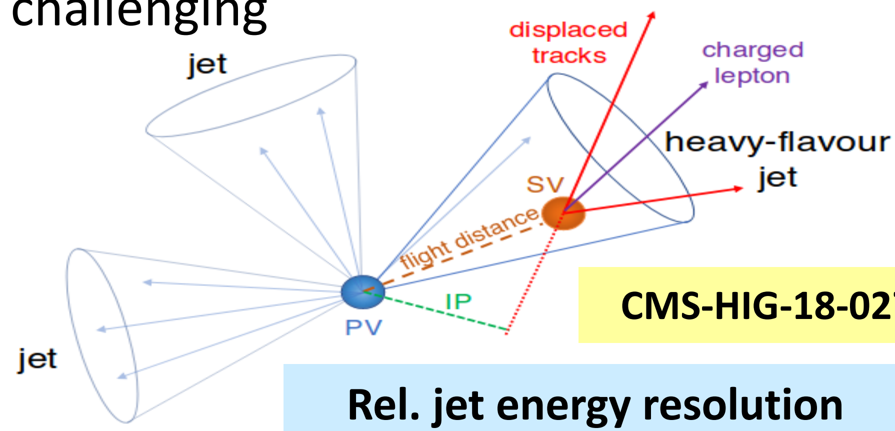


Strong dependence on PU for p_T(jet)< O(100) GeV
→ Almost 2x resolution degradation
→ Less dramatic for high-p_T jets

New !

Jet calibration: Improving b quark energy & resolution

- Precise estimation of b-quark energy challenging
  - mainly due to energy loss via (undetected) neutrinos from semileptonic decays (~20%)



CMS-HIG-18-027

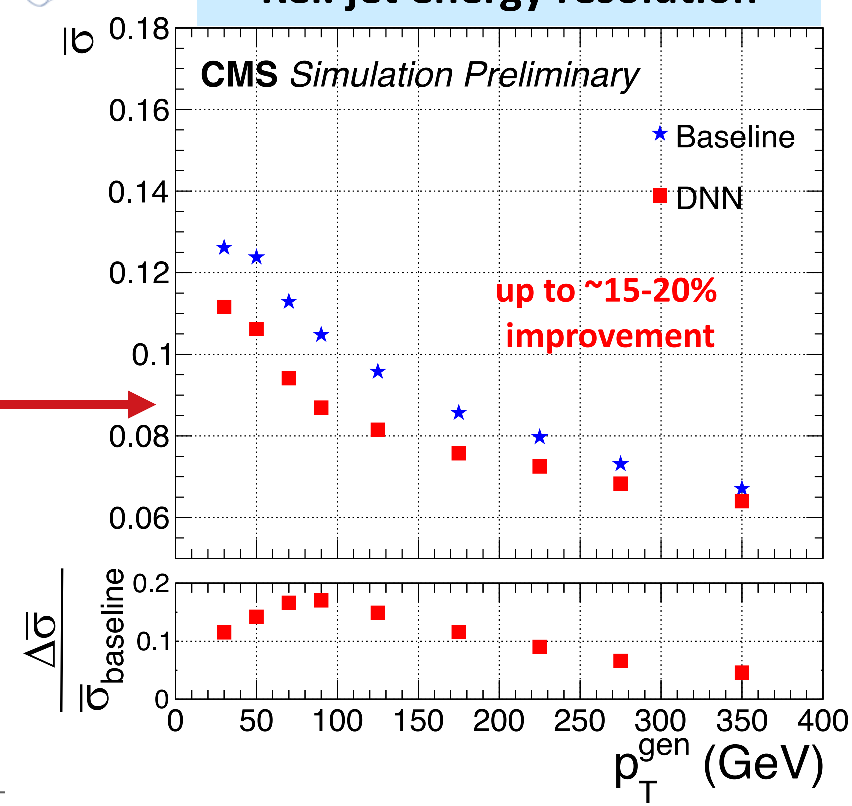
- **Goal:** Improve b-jet energy scale and resolution using a DNN-based regressor

Inputs:

- Jet kinematics
- Jet composition
- PU information
- Info about semi-leptonic decays
- Secondary vertex (SV) properties

Output: energy & resolution corrections

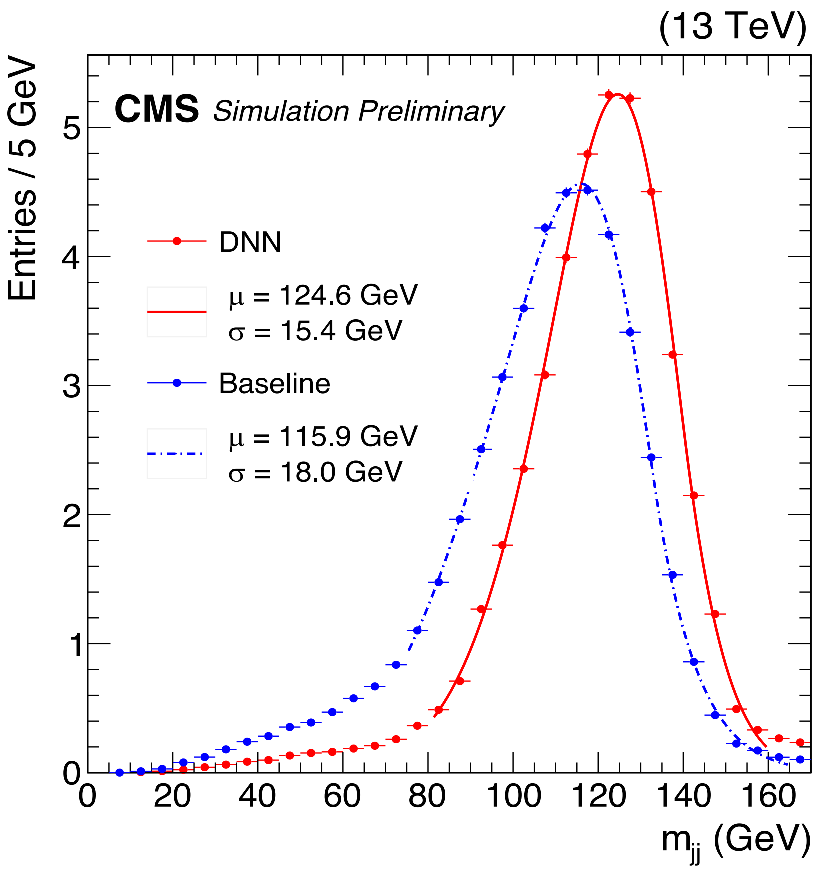
Rel. jet energy resolution

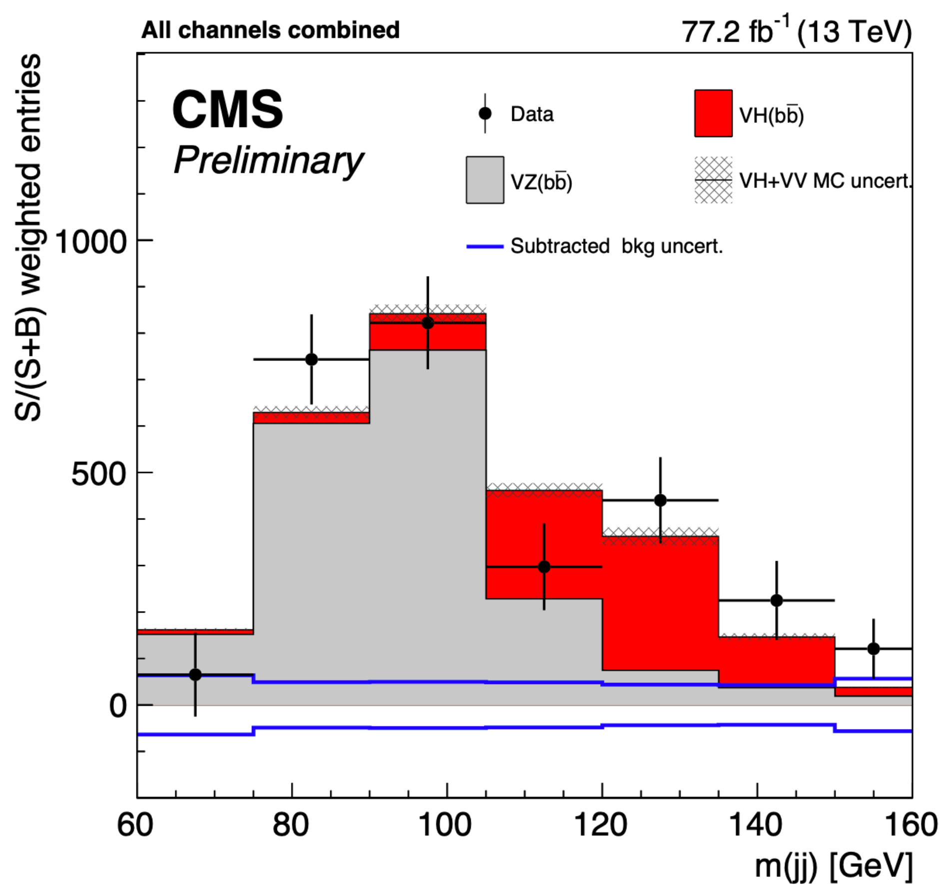


Jet calibration: Improving b quark energy & resolution (II)

- First application:
  - Observation of the Higgs boson decaying a b-quark pair

CMS-HIG-18-016



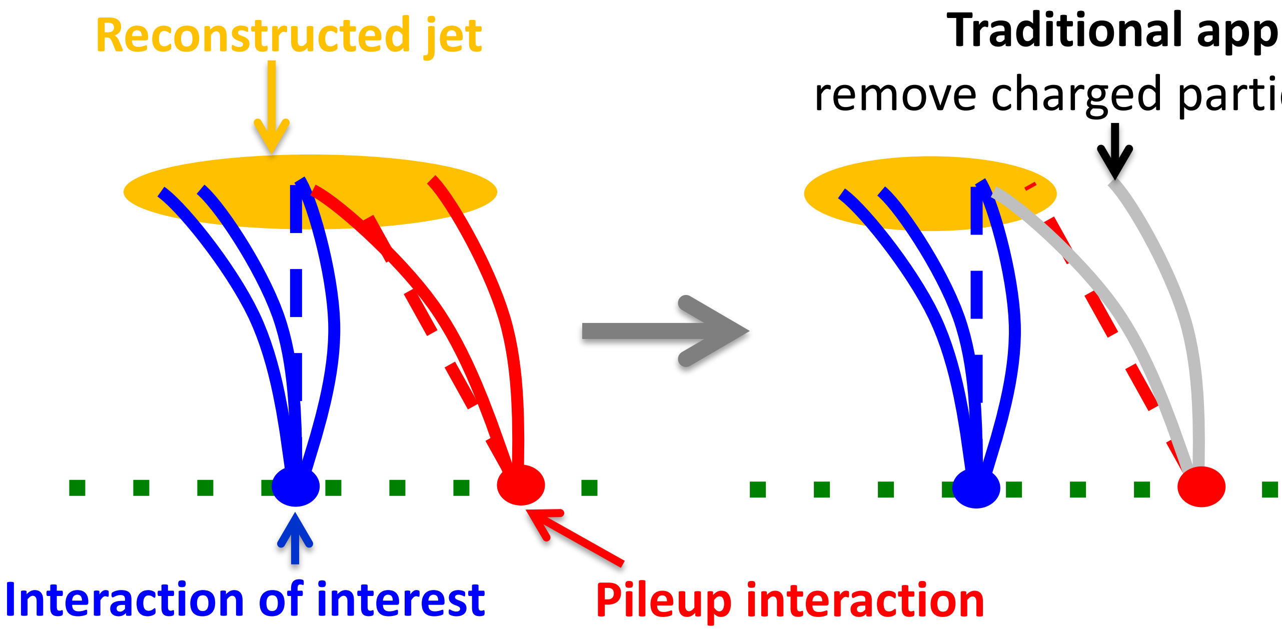


Improvement ~20-25% in mass resolution

~10% gain in sensitivity with the DNN-based b-jet energy regression tool

A big challenge: pile-up mitigation

- Contributions from pileup particles (charged and neutral)
  - impact jet performance [but also missing transverse energy, lepton isolation…]



Traditional approach: <u>"Charged hadron subtraction (CHS)"</u>

→ **Charged contribution in jet ~60%:** Remove charged particles associated to PU vertices
→ **Correct energy from neutral contributions in jet (~40%):** based on the average PU energy density in the event

Tuned for Run 1 & early Run 2 conditions

A big challenge: pile-up mitigation (II)

New !

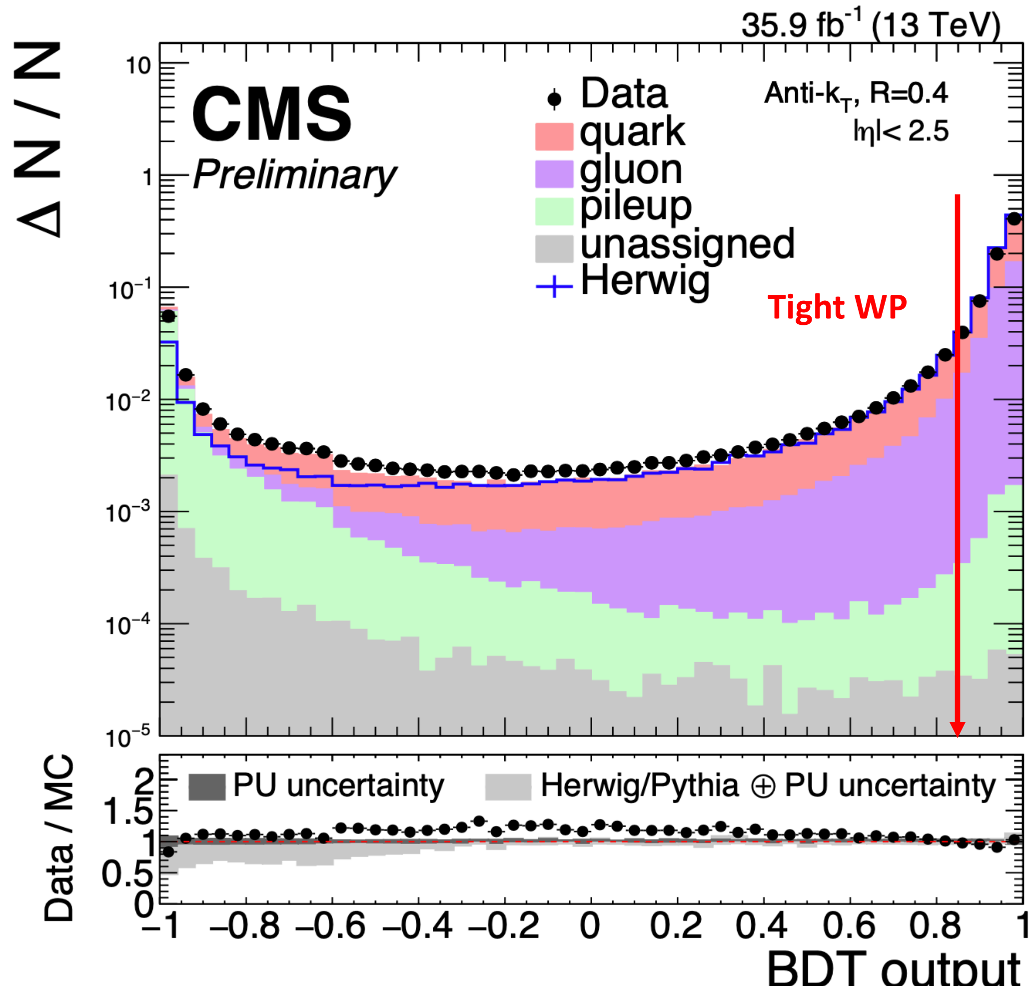
CMS-JME-18-001

- Additional handle to suppress PU: **PileUp Jet ID**
  - designed to reject jets originating from PU

→ Deploys multivariate technique (i.e. BDT) to reject pileup jets.

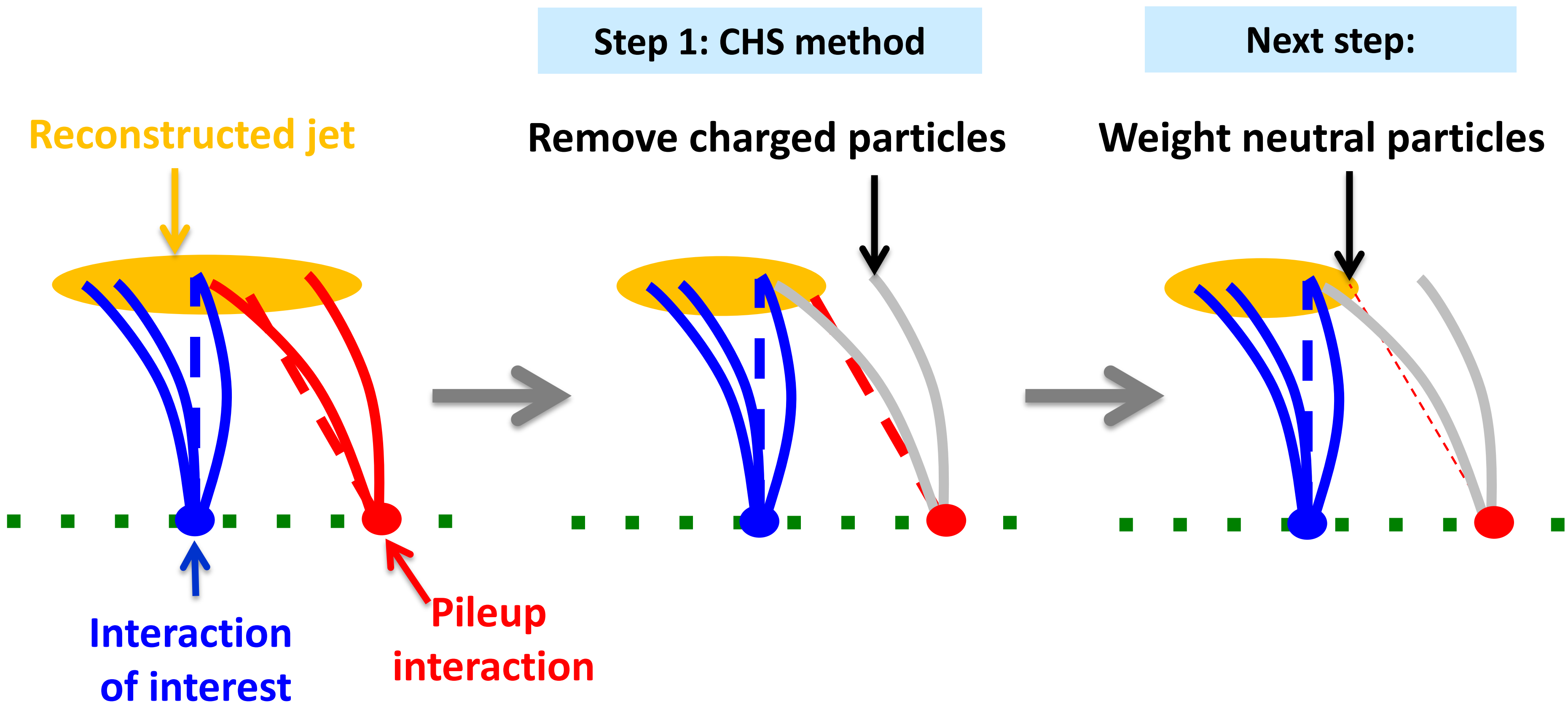
→ **Inputs:**
tracking-related variables, jet shapes, quark/gluon variables

→ Overall reasonable agreement between data and MC



Pile-up mitigation: What about neutral particles?

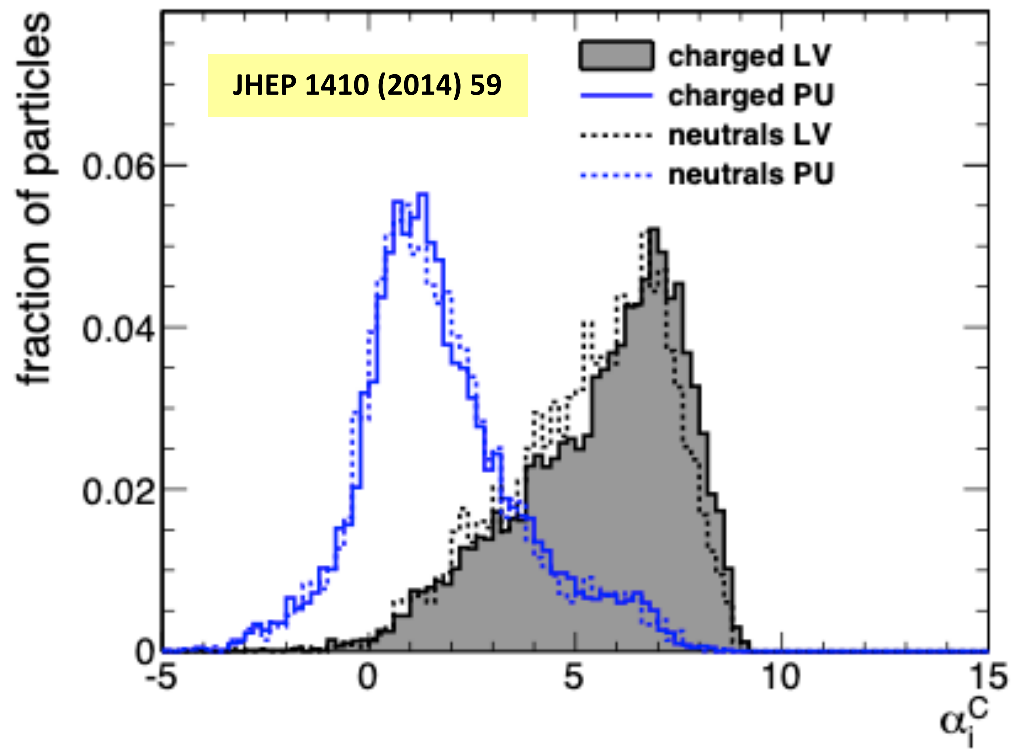
- Ideally we want to follow a similar approach as for charged PU particles
  - remove/correct **"particles"** <u>not</u> an **"average energy"** → no loss of information



Pile-up mitigation: PUPPI

- **P**ile**U**p **P**er **P**article **I**dentification (**PUPPI**):
  - Calculate a ***weight*** for each neutral particle related to the likelihood this particle to originate from PU or not
    - Then rescale each particle's momentum using this weight.
  - **Exploits information:**
    (a) Local particle distribution
    (b) Event PU properties
    (c) Tracking

Assumption:
PU properties between charged and neutral particles similar



Advantages:
correct p_T(jet), jet mass & shape
pileup-Id, ME_T, photon/lepton isolation .. in one go

PUPPI algorithm in a nutshell

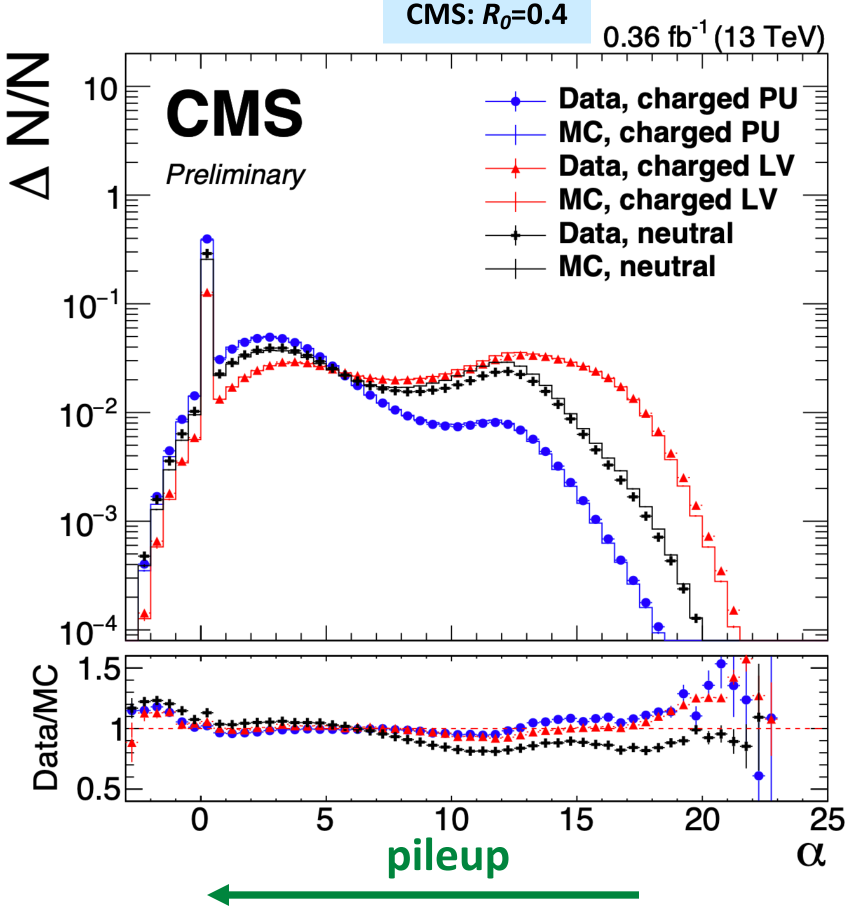
New !

CMS-JME-18-001

- For each particle "*i*" define a local metric "*a*", that differs between pileup (**PU**) and leading vertex (**LV**) particles:

$$\alpha_i = \log \sum_{j \neq i, \Delta R_{ij} < R_0} \left(\frac{p_{Tj}}{\Delta R_{ij}} \right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from LV,} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles} \end{cases}$$

CMS: R_0=0.4



→ **ΔR_{ij}: distance in η-ϕ**
smaller for particles from LV
larger for PU particles [i.e. particles distributed more democratically in ΔR]

→ **Also PU particles typically softer**

→ **PU-like particles populate smaller values of a**

→ **Good overall data-MC agreement**

PUPPI algorithm in a nutshell (II)

New !

CMS-JME-18-001

- For each particle "*i*" define a local metric "*a*", that differs between pileup (**PU**) and leading vertex (**LV**) particles:

$$\alpha_i = \log \sum_{j\neq i, \Delta R_{ij}<R_0} \left(\frac{p_{\mathrm{T}j}}{\Delta R_{ij}}\right)^2 \begin{cases} \text{for } |\eta_i| < 2.5, & j \text{ are charged particles from LV,} \\ \text{for } |\eta_i| > 2.5, & j \text{ are all kinds of reconstructed particles} \end{cases}$$

- **Next:** Translate "*a*" to a PU probability
  - Use **charged PU particles** to estimate the expected PU distribution in an event:

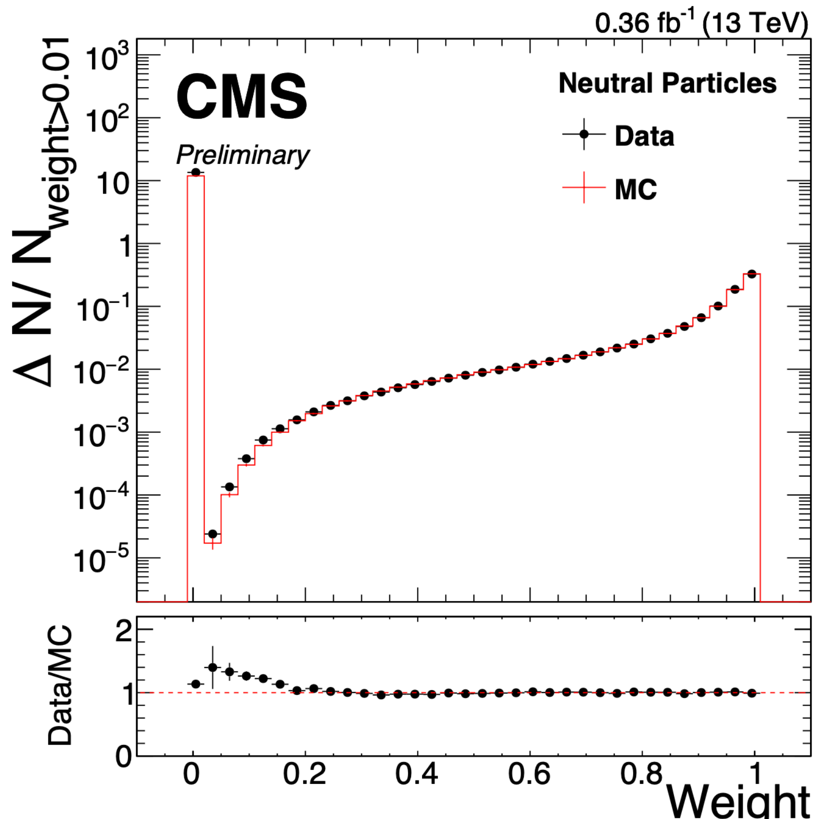
$$\bar{\alpha}_{PU},\ RMS_{PU}$$

  - Compare the "*a*" of each neutral particle with the expected values using a ***χ²*** approximation:

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})\,|\alpha_i - \bar{\alpha}_{PU}|}{RMS_{PU}^2}$$

  - Compute the weight:

$$w_i = F_{\chi^2, NDF=1}(\chi_i^2)$$



pileup

- **Finally:** Reweight each particle's 4-vector with "w_i"

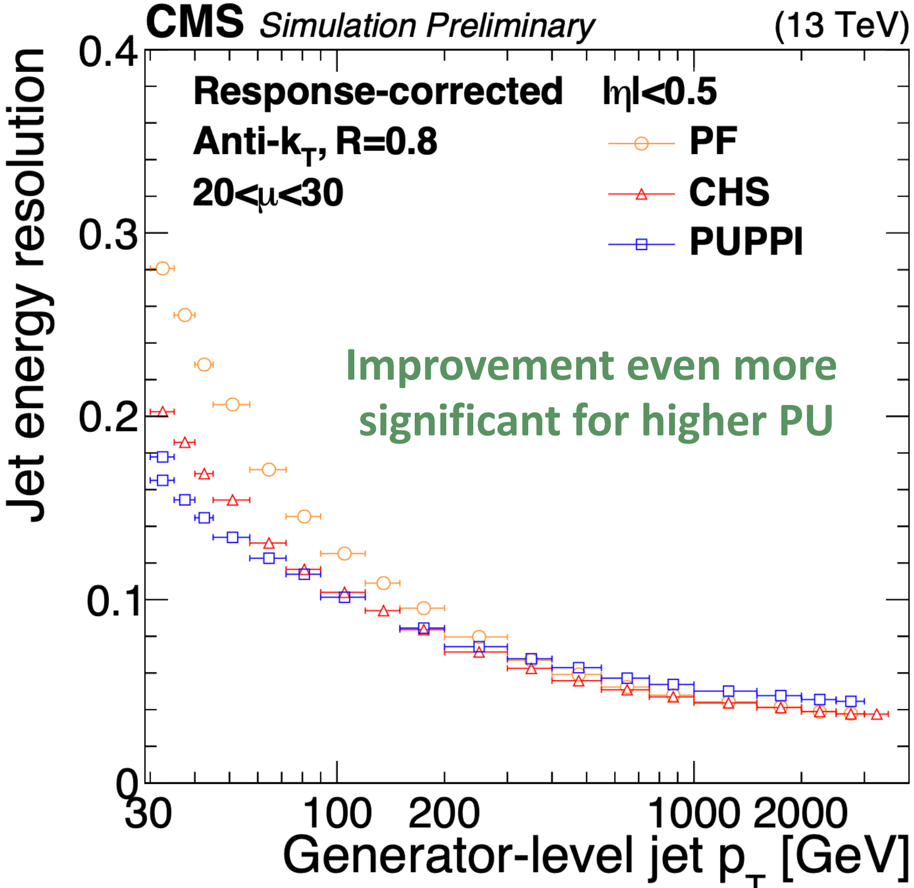
Procedure performed on an event-by-event level

PUPPI performance: jets

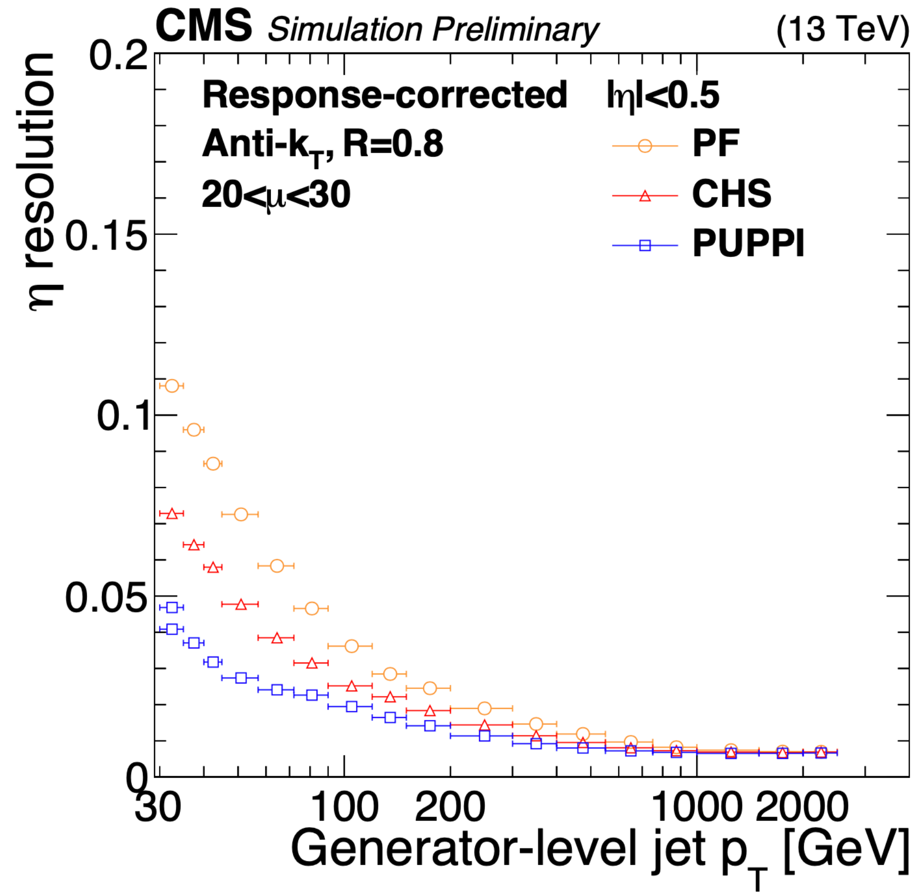
New !

CMS-JME-18-001

Jet energy resolution



Angular resolution



→ **Significant gain with PUPPI particularly at lower p_T(jet);**
[Same for slim (R=0.4) jets]

→ **Significantly improved ME_T stability and Lep. Isolation with PU [back-up]**

High-p_T "Boosted" jet tagging in CMS

i.e. using large cone jets

[Mainly focus on top and W tagging]

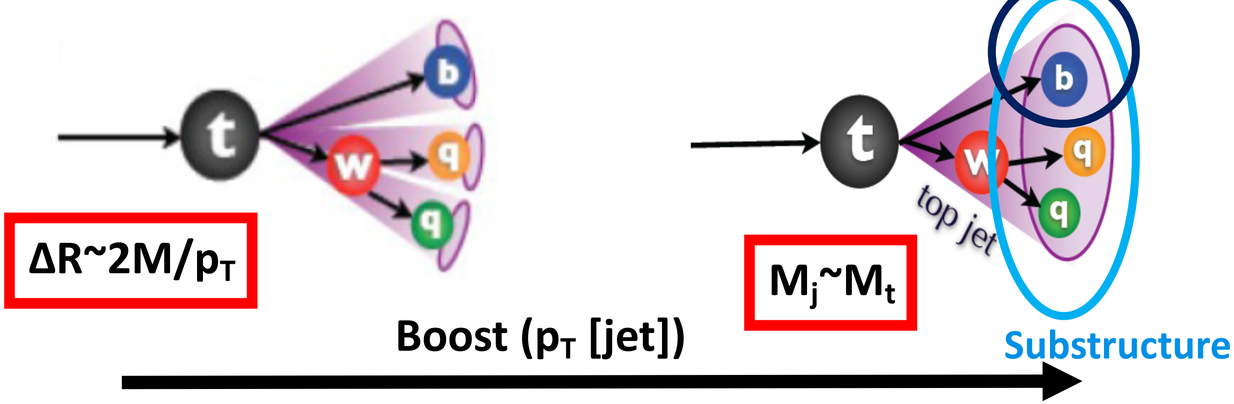
Boosted jet tagging: basics

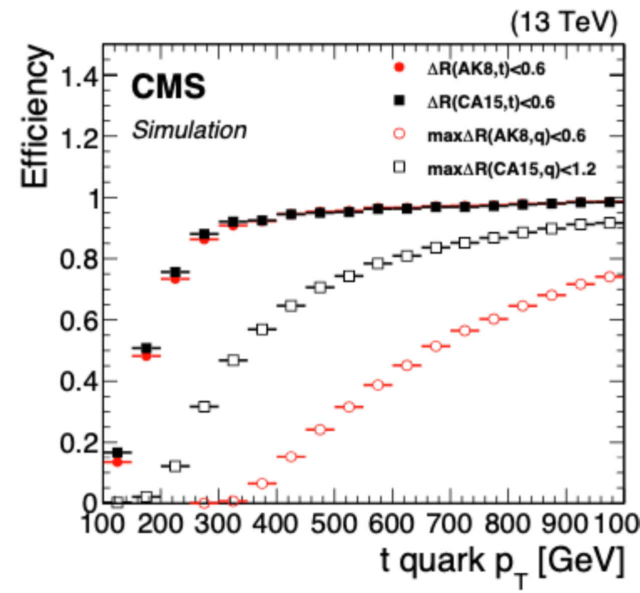
CMS-JME-18-002

- Use boosted top quark tagging as an example:

Low boost:
unmerged decay products
→ top decay products resolved as three R=0.4 jets

Large boost:
Top products merged





- **Hadronic top quark decay:** a bottom quark and a W boson → 3 quarks in total

- **Main handles:**
  - **Jet mass**
  - **Jet substructure:** identify the 3-prong structure in a single "wide" jet
  - **Flavour:** Identify the b quark [or even W → cX]
  - **Challenges:** soft radiation, Pile-up, etc..

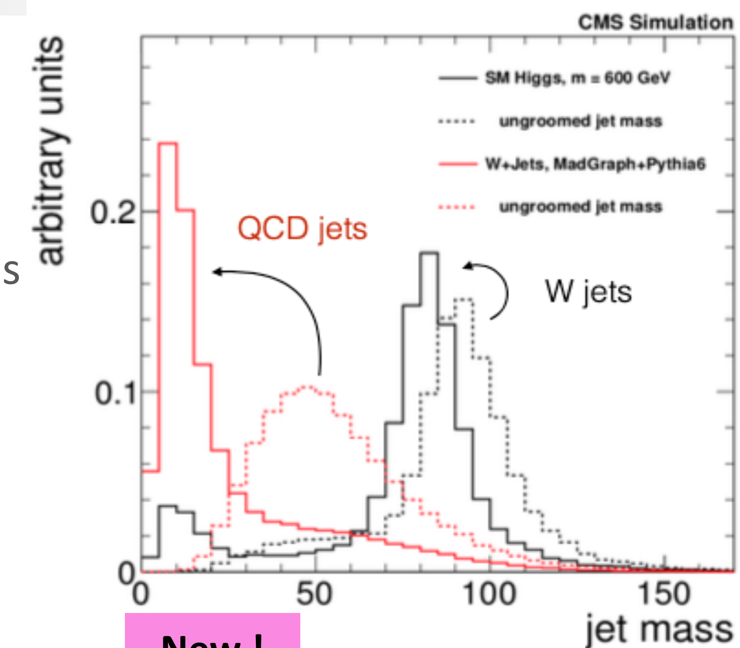
Not clear if these effects factorize [or need to be exploited simultaneously]

Jet mass

- **Jet grooming**
  - i.e. remove soft and wide angle radiation [e.g. from ISR, UE, …]
    - **otherwise** mass of BKG jets → high values
  - CMS default: **Softdrop method**

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

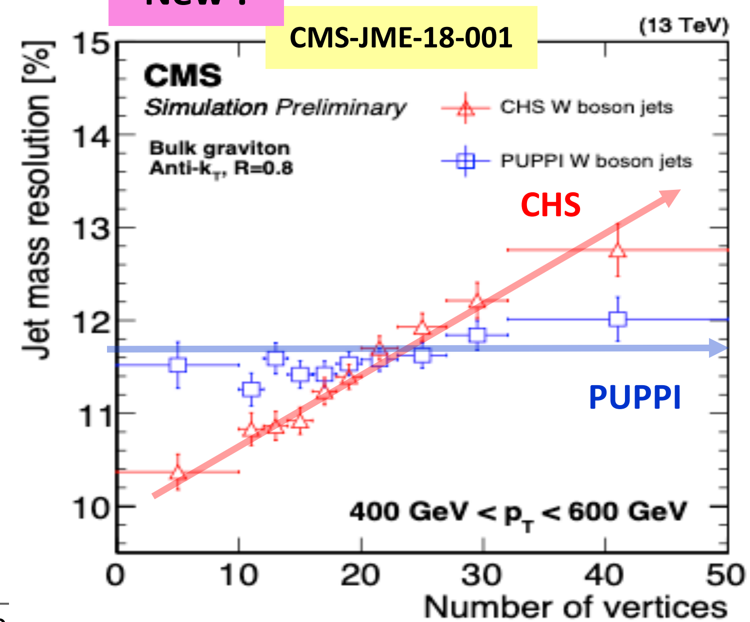
CMS: z=0.1, β=0



- **Pile-up mitigation** [i.e. PUPPI]
  - very stable mass resolution and mass peak [backup] with PUPPI
  - **PUPPI:** Default tool in CMS for PU mitigation in large radius jets

New !

CMS-JME-18-001



Jet substructure

New !

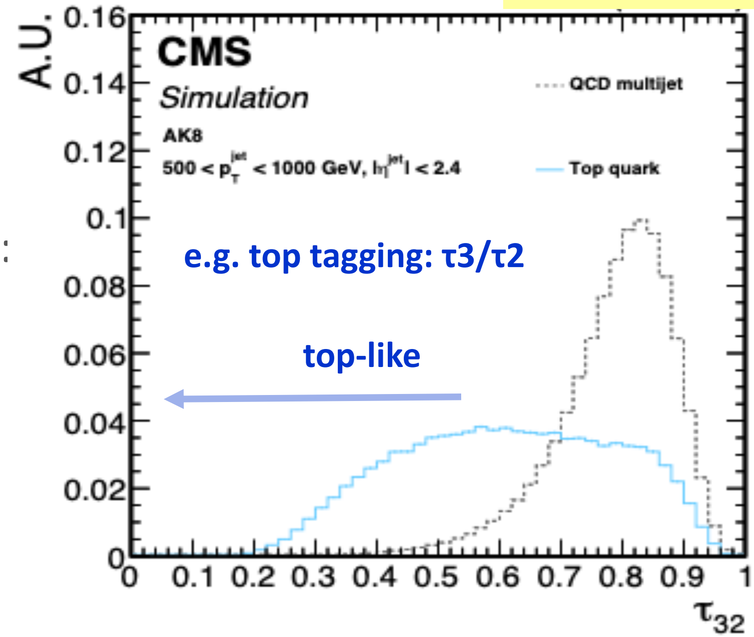
CMS-JME-18-002

- Explore energy distributions inside the jet
  - **N-subjettiness:**
    - Identify ***"N"*** candidate subjets using k_T
    - Check compatibility with N-subjets (prongs):

$$\tau_N = \frac{1}{d_0} \sum_i p_{\mathrm{T},i} \min\left[\Delta R_{1,i}, \Delta R_{2,i}, \ldots, \Delta R_{N,i}\right]$$

[small when compatible with N-prongs]

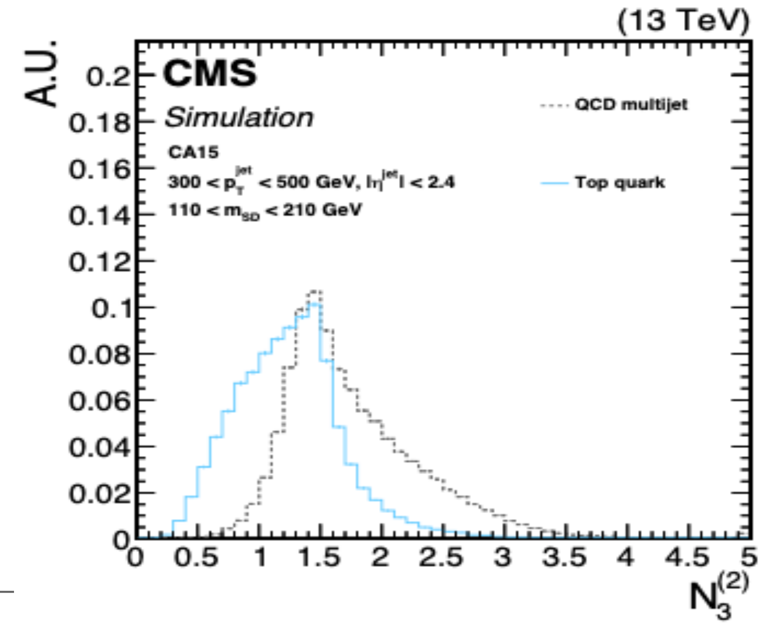
    - Usually take ratios → more powerful



  - **Energy correlation functions [ECF]:**
    - similar to $\boldsymbol{\tau_N}$; but uses an axis-free approach

$${}_{o}e_N^{\beta} = \sum_{1 \le i_1 < i_2 < \cdots < i_N \le N_C} \left[\prod_{1 \le k \le N} \frac{p_{\mathrm{T}}^{i_k}}{p_{\mathrm{T}}^{J}}\right] \prod_{m=1}^{o} \min_{i_j < i_k \in \{i_1, i_2, \cdots, i_N\}}^{(m)} \left\{\Delta R_{i_j, i_k}^{\beta}\right\}$$

    - Weighted by the angular separation of the constituents
    - N–prong jet: $e_M \ll e_N$ for M>N; e(N=4) / e(N=3) is the analog of τ_{32}



Overview of CMS tagging algorithms

New !

CMS-JME-18-002

- A very diverse set of jet tagging algorithms:

  Traditional approaches: ”cut-based” selection based on these “theory-inspired” variables; developed mainly during Run 1 and early Run 2 [serve as baseline]

Advantage(s) of traditional methods:
- Extensively studied → robustness
- Easy interpretation
- Serve as baseline for more advanced tools

Algorithm	t quark	W boson	Z boson	Higgs boson	decay modes
$m_{SD} + \tau_{32}$	✓				
$m_{SD} + \tau_{32} + \mathrm{b}$	✓				
$m_{SD} + \tau_{21}$		✓	✓		
HOTVR	✓				
$N_3 - \mathrm{BDT}$ (CA15)	✓				
N_2		✓	✓	✓	
BEST	✓	✓	✓	✓	
ImageTop	✓				
DeepAK8	✓	✓	✓	✓	✓
Jet mass decorrelated algorithms					
N_2^{DDT}		✓	✓	✓	
double–b			✓	✓	
ImageTop–MD	✓				
DeepAK8–MD	✓	✓	✓	✓	✓

New !

Overview of CMS tagging algorithms

CMS-JME-18-002

- A very diverse set of jet tagging algorithms:
  - **Traditional approaches:** ”cut-based” selection based on these “theory-inspired” variables; developed mainly during Run 1 and early Run 2 [serve as baseline]
- New generation of tools using:
  - **Shallow Machine Learning [ML]** + High-level (HL) variables
  - **Advanced ML** on HL vars or even low-level information [e.g. PF candidates]

Algorithm	t quark	W boson	Z boson	Higgs boson	decay modes
$m_{SD} + \tau_{32}$	✓				
$m_{SD} + \tau_{32} + b$	✓				
$m_{SD} + \tau_{21}$		✓	✓		
HOTVR	✓				
$N_3 - \mathrm{BDT}$ (CA15)	✓				
N_2		✓	✓	✓	
BEST	✓	✓	✓	✓	
ImageTop	✓				
DeepAK8	✓	✓	✓	✓	✓
Jet mass decorrelated algorithms					
N_2^{DDT}		✓	✓	✓	
double–b			✓	✓	
ImageTop–MD	✓				
DeepAK8–MD	✓	✓	✓	✓	✓

Overview of CMS tagging algorithms

- A very diverse set of jet tagging algorithms:
  - **Traditional approaches:** "cut-based" selection based on these "theory-inspired" variables; developed mainly during Run 1 and early Run 2 [serve as baseline]
- New generation of tools using:
  - **Shallow Machine Learning [ML]** + High-level (HL) variables
  - **Advanced ML** on HL vars or even low-level information [e.g. PF candidates]

Algorithm	t quark	W boson	Z boson	Higgs boson	decay modes
$m_{SD} + \tau_{32}$	✓				
$m_{SD} + \tau_{32} + b$	✓				
$m_{SD} + \tau_{21}$		✓	✓		
HOTVR	✓				
N_3 − BDT (CA15)	✓				
N_2		✓	✓	✓	
BEST	✓	✓	✓	✓	
ImageTop	✓				
DeepAK8	✓	✓	✓	✓	✓
Jet mass decorrelated algorithms					
N_2^{DDT}		✓	✓	✓	
double–b			✓	✓	
ImageTop–MD	✓				
DeepAK8–MD	✓	✓	✓	✓	✓

Disclaimer:

Impossible to give justice to all tools developed during Run2

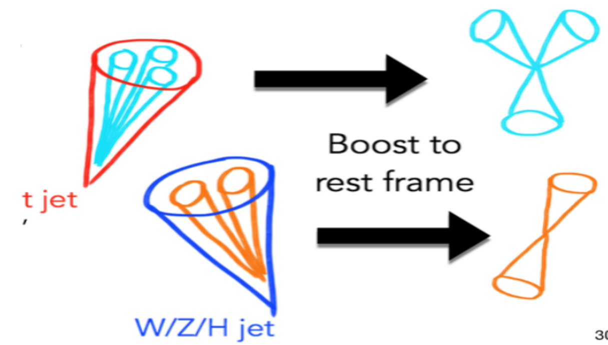
Focus more on the latest developments using advanced ML techniques

New !

Boosted event shape tagger: BEST

- Discriminate between particle species [top, H, Z, W, b, light] simultaneously

  - Use different hypothesized reference frames corresponding to the heavy particle masses
    - When boosting to the "correct" rest frame, jet constituents should show the expected N-prong structure



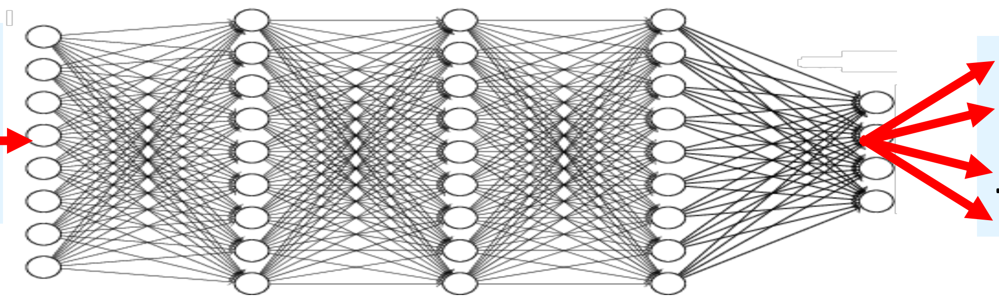
  - Then, for each case calculate the quantities:

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)	

New ! Boosted event shape tagger: BEST (II)

- Use a DNN to classify jets via many kinematic quantities (59 in total)
  - DNN: dense network, 3 hidden layers, 40 nodes/layer

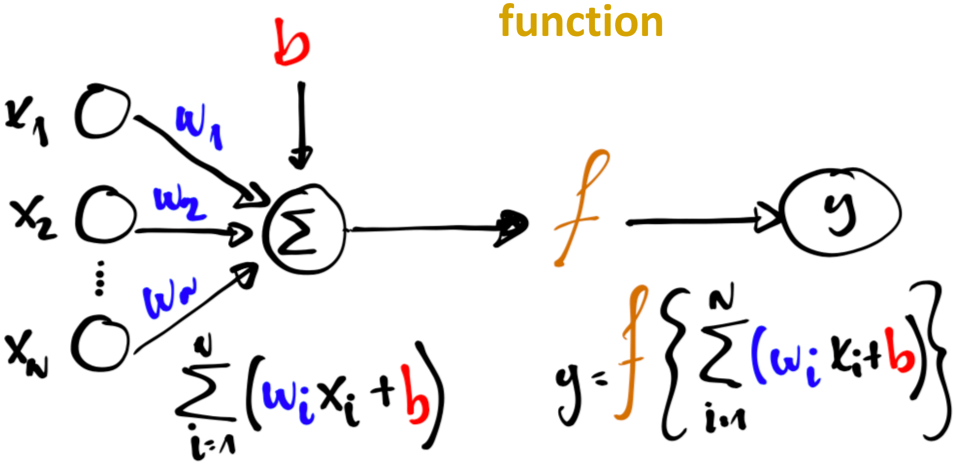
Inputs
Use "human-made" (high-level) variables



Output
Multiclass classifier
top, H, Z, W, b, light

Reminder:

inputs **bias** **Activation function** **output**



In a nutshell; in each node:
→ Multiply each input (x_i) with a weight (w_i)
→ Introduce a bias (b) and sum everything
→ Add non-linearity to the output with the activation function f (e.g. *ReLu*)
→ Compare output with target: *loss function* e.g. Cross Entropy:

$$\mathcal{L} = \sum_{x_i}\{y \ln p + (1-y)\ln(1-p)\}$$

→ Minimize loss by tuning w_i and b

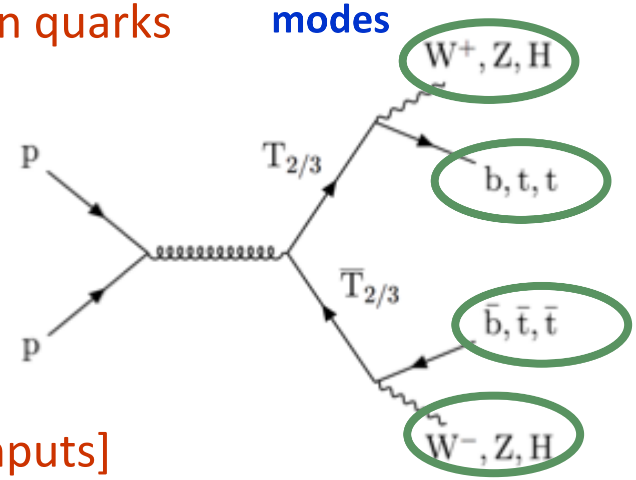
Search for Vector Like Quarks (VLQ)

New !

CMS-B2G-18-005

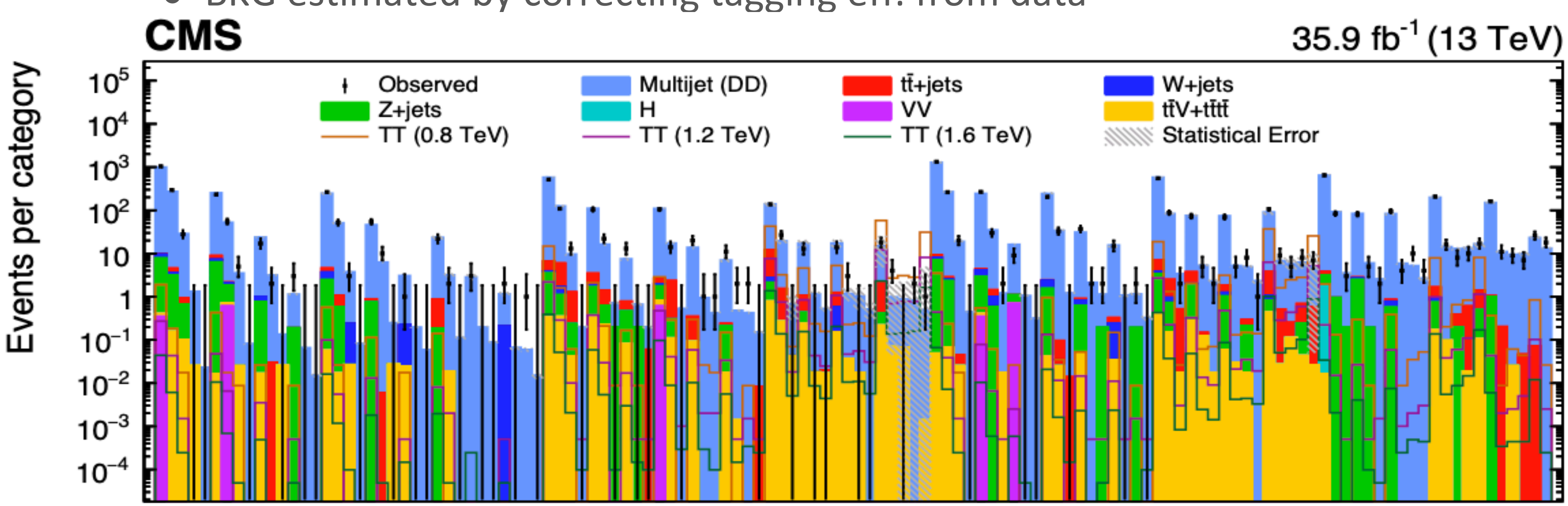
- VLQ (T, B) transform under same group as SM EWK bosons
  - leads to a variety of decay modes; mainly to 3rd-Gen quarks
    - relative fractions depends on the model

main decay modes



Jet tagging algorithms with multi-tagging capabilities were developed exactly for this

- Analysis targets the all-hadronic channel:
  - Key player: **new BEST** multiclass tagger [DNN+HL inputs]
    - BKG estimated by correcting tagging eff. from data



The 126 search regions

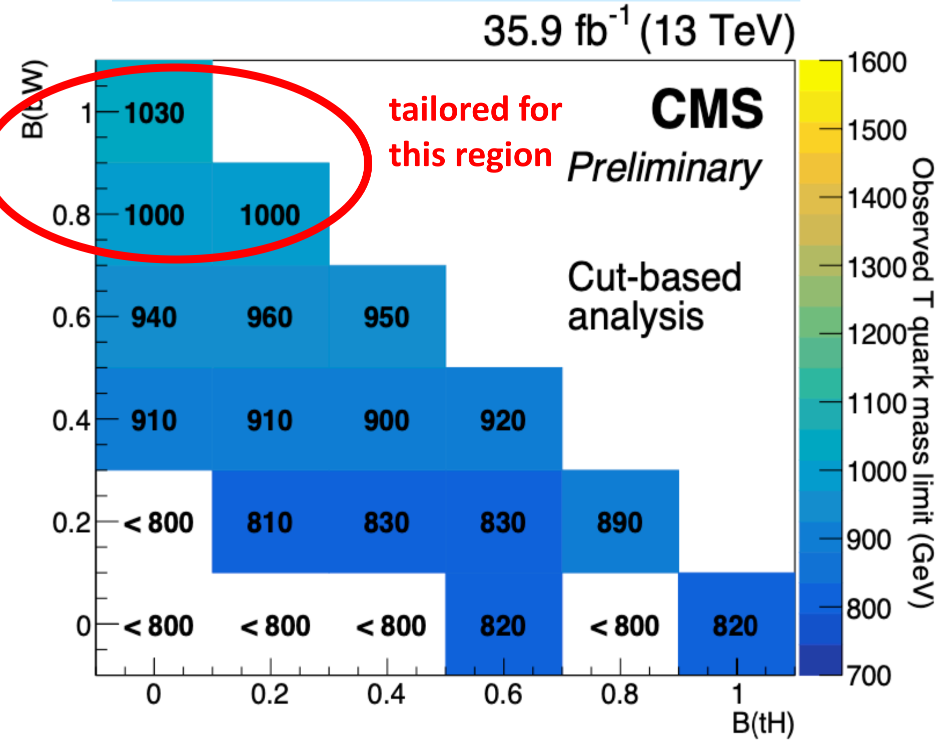
New !

Search for Vector Like Quarks (VLQ) (II)

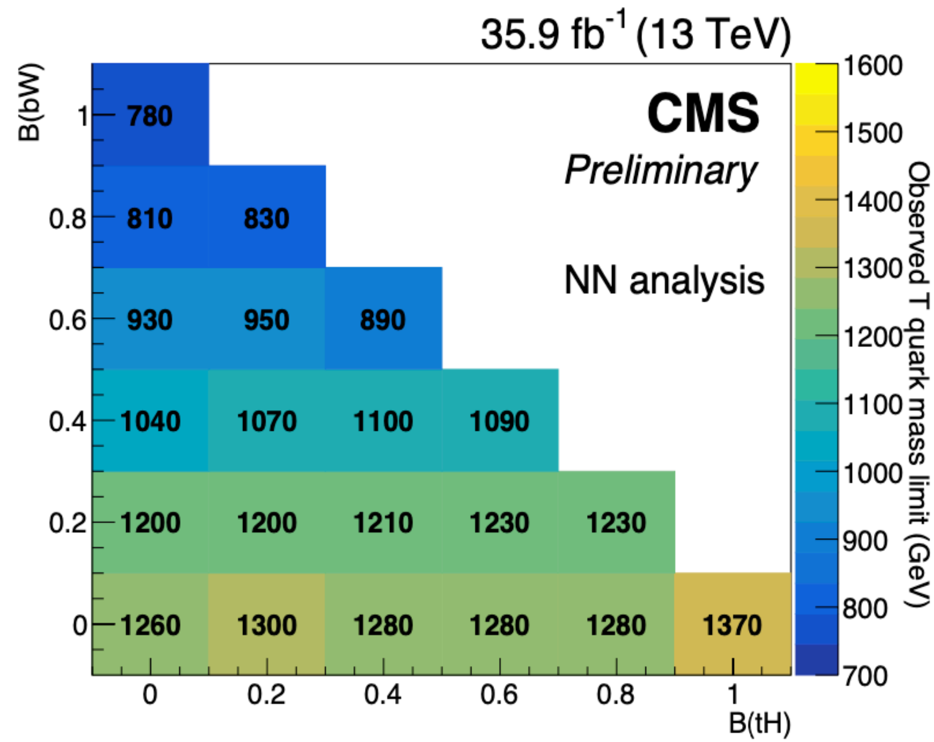
CMS-B2G-18-005

- **Traditional approach:** [also included in this analysis]
  - cut-based selection; tailored to a single decay mode / single category [i.e. T→bW]
    - Categorize events based on W-tags [SD+τ_{21}] and b-tags; fit H_T shape

Traditional approach



With new tools

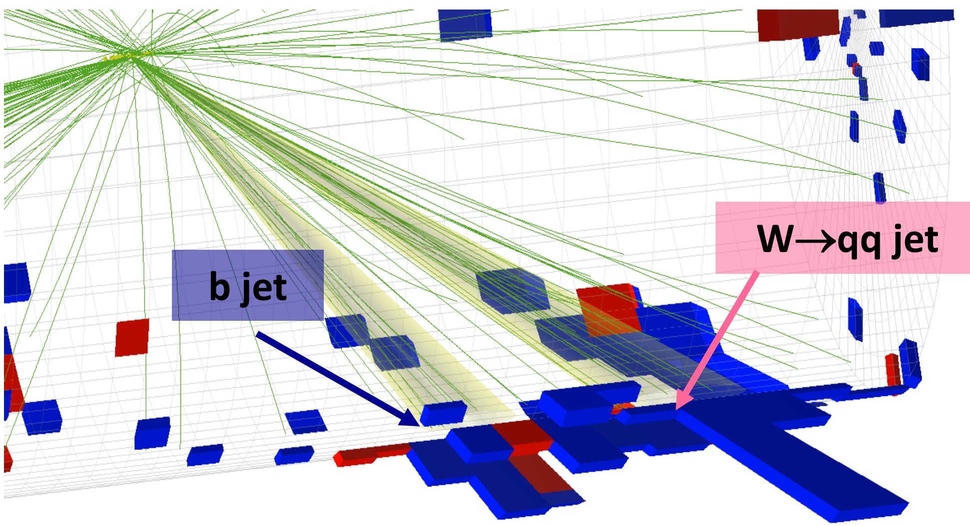


The new tool (i.e. BEST tagger) allows for a much more comprehensive approach
Stronger results over a wider variety of models

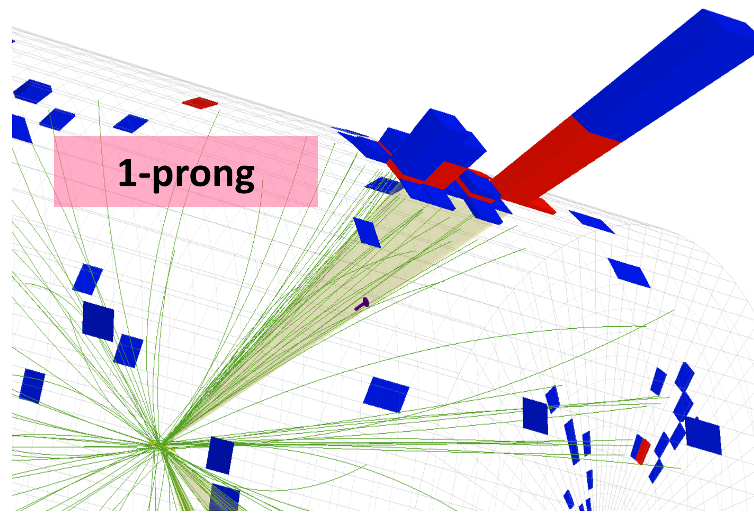
A step further: Particle based jet tagging

- **In theory:** A jet is a spray of particles produced by the hadronization of quarks and gluons
  - QCD confinement allows only colorless states
- **In practice:** A jet is a cone **<u>of reconstructed particles</u>** in the detector
  - With mass and kinematics consistent with e.g. the top decay

Top decay in real life



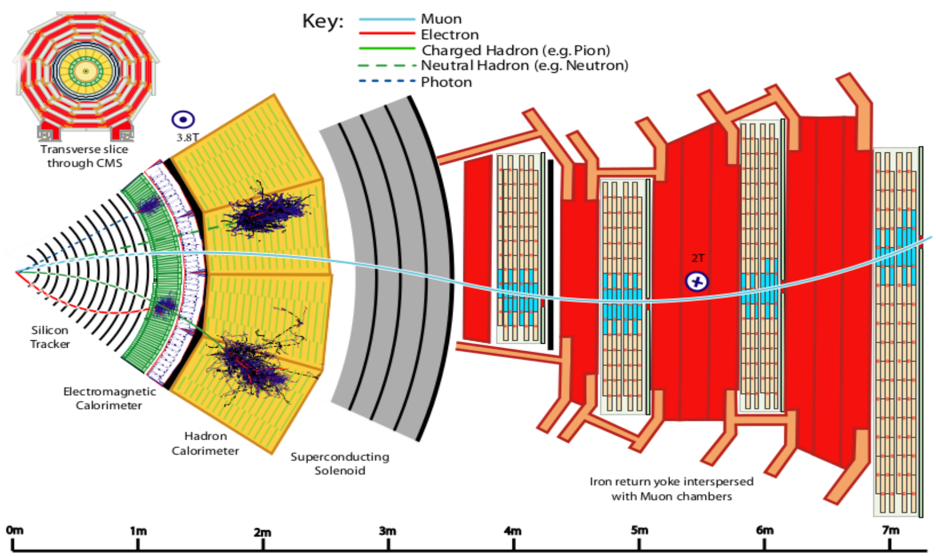
BKG [gluon/quark]



- Can we gain by moving to **<u>particle based jet tagging</u>**?

Exploring [more] CMS potential

- Reminder: CMS PF algorithm
  - Combines information from all subdetectors
  - Mutually exclusive list of particles



- Rich information for each particle
  - Energy/momentum
  - Position
  - Particle category
  - Displacement from the PV
  - Reconstruction quality
  - ....

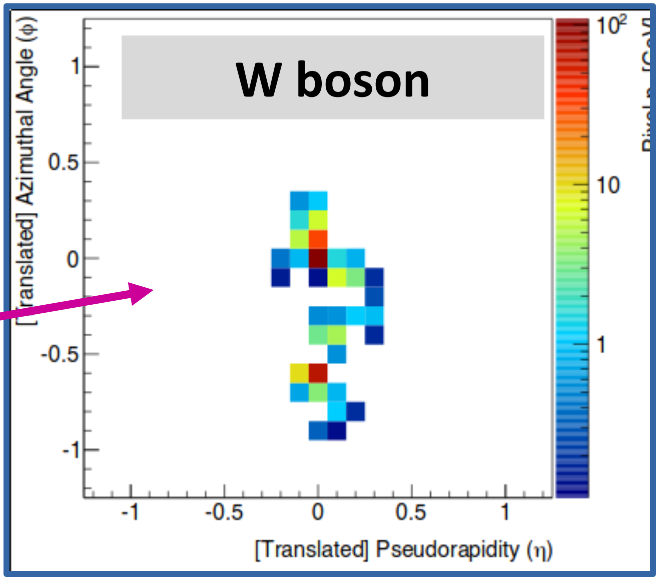
Inputs to jet substructure

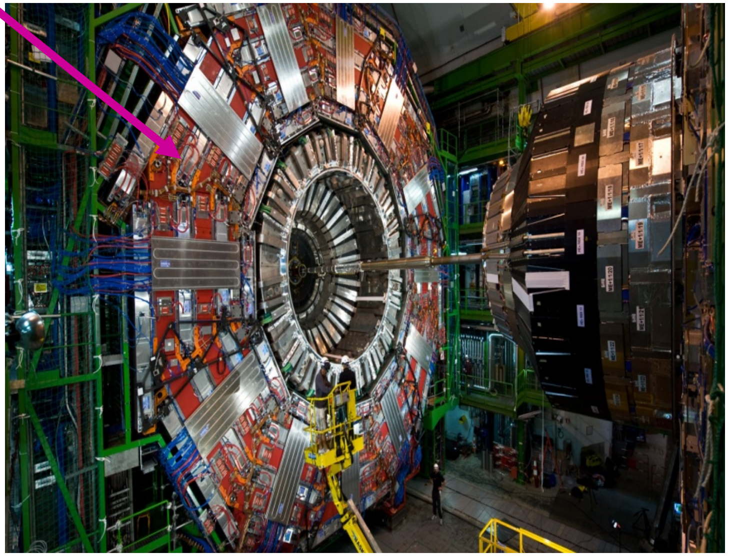
Inputs for flavour tagging

- [O(50) properties/particle] x [~50-100 particles/jet] ~O(1000) inputs/jet
- **Perfect case** for DNN with “complex” architecture

Deep learning approaches for jet tagging

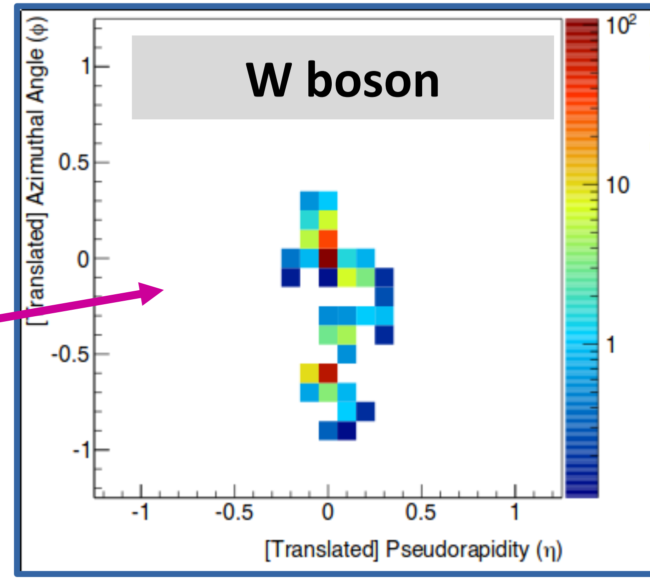
- Based on **jet image**:
  - Treat **detector** (i.e. calorimeters) as a **camera** & the **jet** as an **image**
  - Apply techniques used for image recognition (i.e. Convolutional Neural Networks – CNN)
  - **But:** jet images are very sparse
  - **Also:** LHC are very heterogeneous/complex
    not “image-like”
    - difficult to include information from other subdetectors (e.g. tracking)

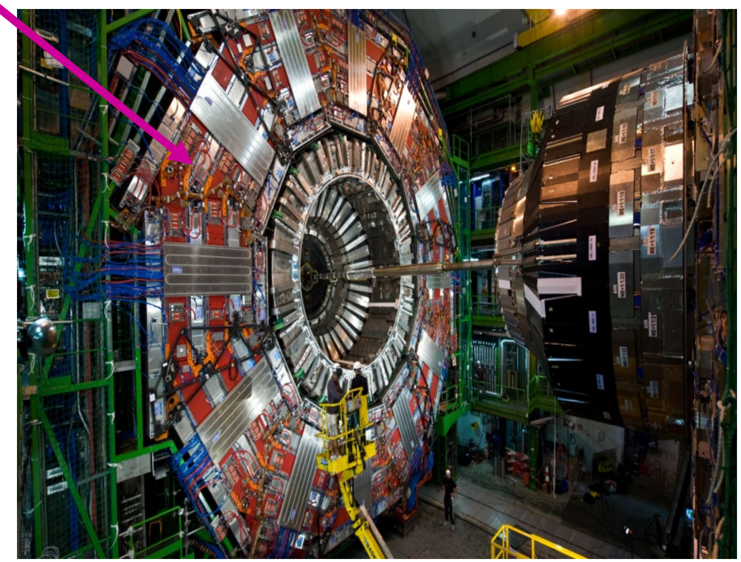




Deep learning approaches for jet tagging

- Based on **jet image**:
  - Treat **detector** (i.e. calorimeters) as a **camera** & the **jet** as an **image**
  - Apply techniques used for image recognition (i.e. Convolutional Neural Networks – CNN)
  - **But:** jet images are very sparse
  - **Also:** LHC are very heterogeneous/complex **not "image-like"**
    - difficult to include information from other subdetectors (e.g. tracking)

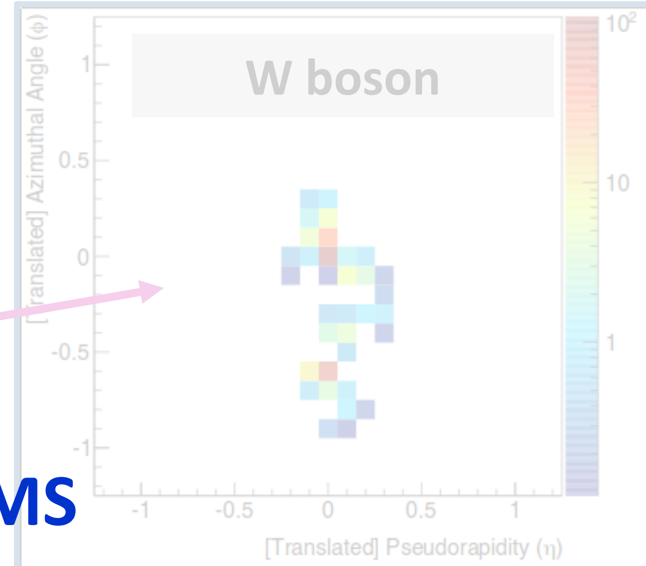




- Based on **particle sequence**:
  - **Jet** as a **sequence of constituent particles**
  - Apply techniques used for **natural language processing** [e.g. CNN-1D,..]
  - Inclusion of more information straight forward
  - Explore **more** of the **CMS detector & CMS event reconstruction potential**

Deep learning approaches for jet tagging

- Based on **jet image**:
  - Treat **detector** (i.e. calorimeters) as a **camera** & the **jet** as an **image**
  - Apply techniques used for image recognition (i.e. Convolutional Neural Networks – CNN)
  - But: jet images are very sparse
  - Also: LHC are very heterogeneous/complex not "image-like"
    - difficult to include information from other subdetectors (e.g. tracking)

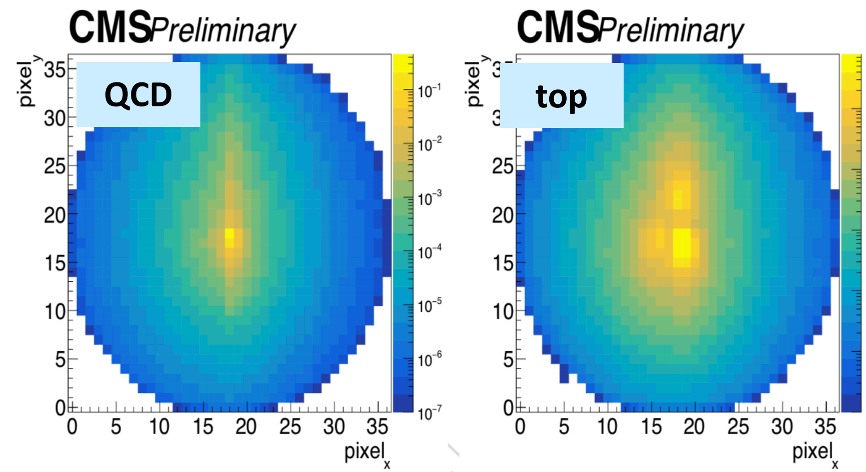


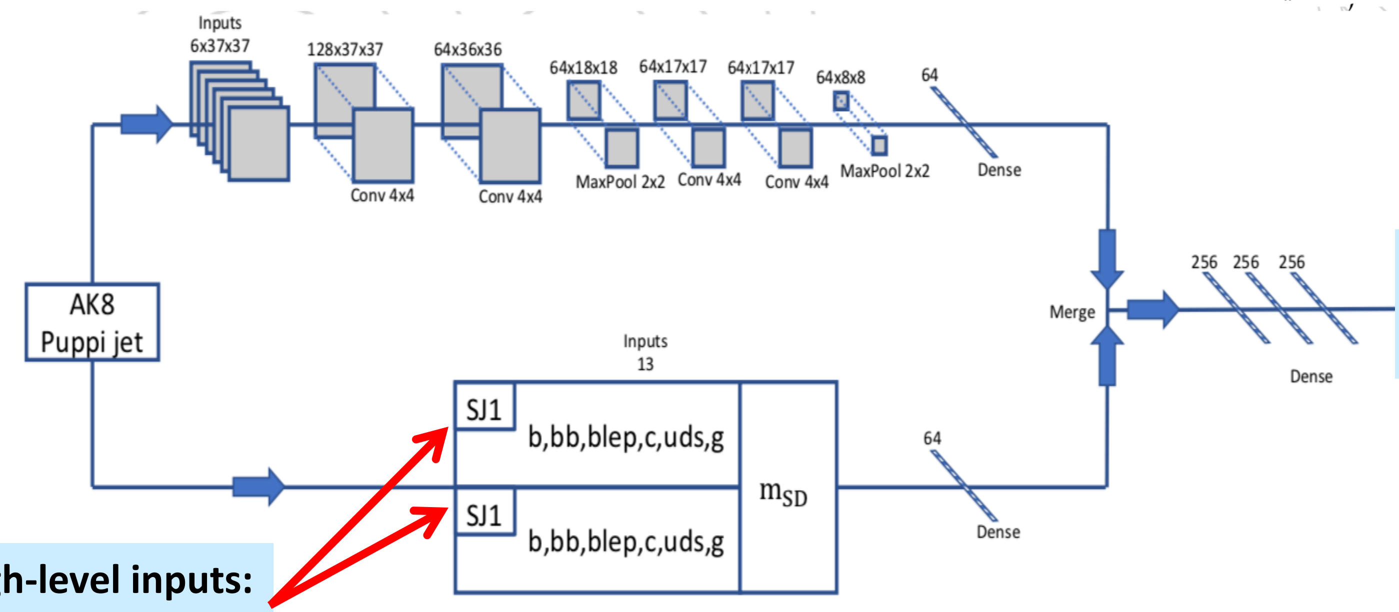
Explore both in CMS

- Based on **particle sequence**:
  - **Jet** as a **sequence of constituent particles**
  - Apply techniques used for **natural language processing** [e.g. CNN-1D,..]
  - Inclusion of more information straight forward
  - Explore much more of the **CMS detector & CMS event reconstruction potential**

New ! Particle based jet tagging: ImageTop

- DNN-based top tagger using PF candidates & apply DNN-based flavor tagging on the subjets
  - **PF candidates:** split into relevant "channels" based on the PF candidate flavor
    - neutral p_T, track p_T, muon p_T, …
- Architecture:
  - inspired by TH paper 1803.00107





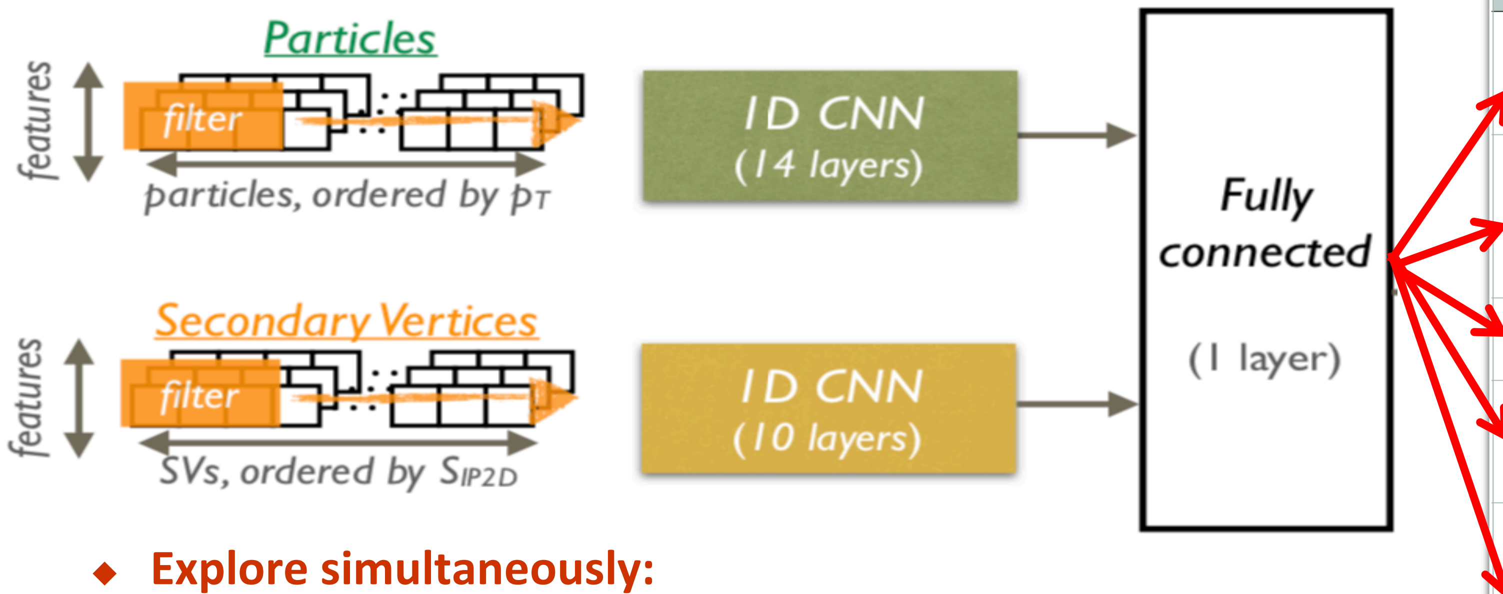
Output: top vs. QCD

High-level inputs: jet flavour

New ! Particle based jet tagging: DeepAK8

CMS-DPS-2017-049, NIPS 2017 paper, CMS-JME-18-002

- Advanced boosted jet tagger: DeepAK8
  - multi-class classifier for top, W, Z, Higgs, and QCD jets
  - inspired by ResNeXt50



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)

  - **Explore simultaneously:** jet substructure, flavor, PU…

Inputs:
→ **Up to 100 inclusive particles (42 features/particle)**
→ **Up to 7 Secondary vertices (15 features/SV)**

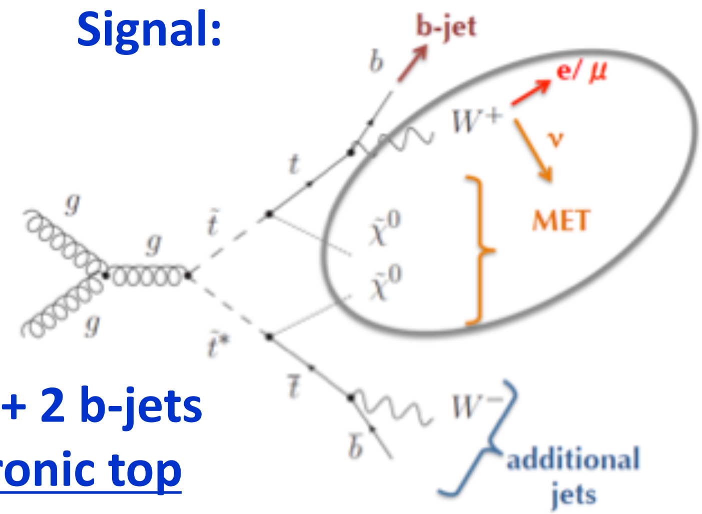
A very versatile boosted jet tagger
→ **various decay modes with different flavor content**

SUSY searches: e.g. Top squarks in 1Lep

CMS-SUS-19-009

- Top squark decays: very rich phenomenology
  - decays via on-shell top quarks / W bosons can significantly benefit from new developments in object tagging

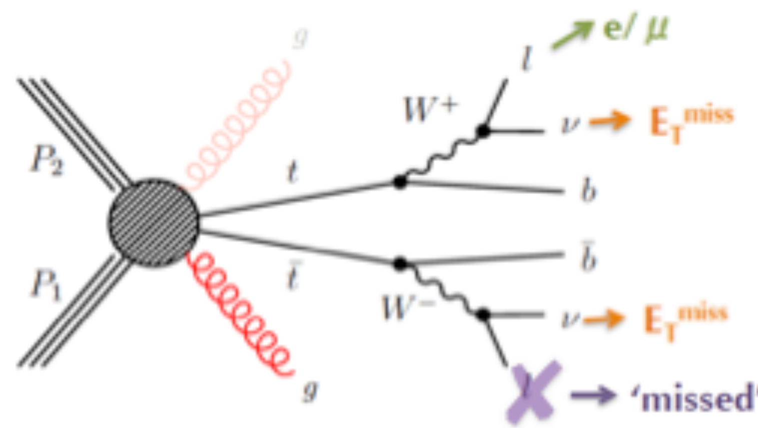
Signal:



1L + ME_T + 2 b-jets + <u>1 Hadronic top</u>

- Events categorized based on:
  - N_j, M_{Lb}, M_T, ME_T
  - Main BKGs: tt(2L)
    - to smaller extend W+jets & ttZ

Main BKG: tt(2L)



- tt(2L) killers:
  - "modified topness":

$$\boxed{S(\vec{p}_W, p_{\nu,z}) = \frac{(m_W^2 - (p_\nu + p_\ell)^2)^2}{a_W^4} + \frac{(m_t^2 - (p_b + p_W)^2)^2}{a_t^4}}$$

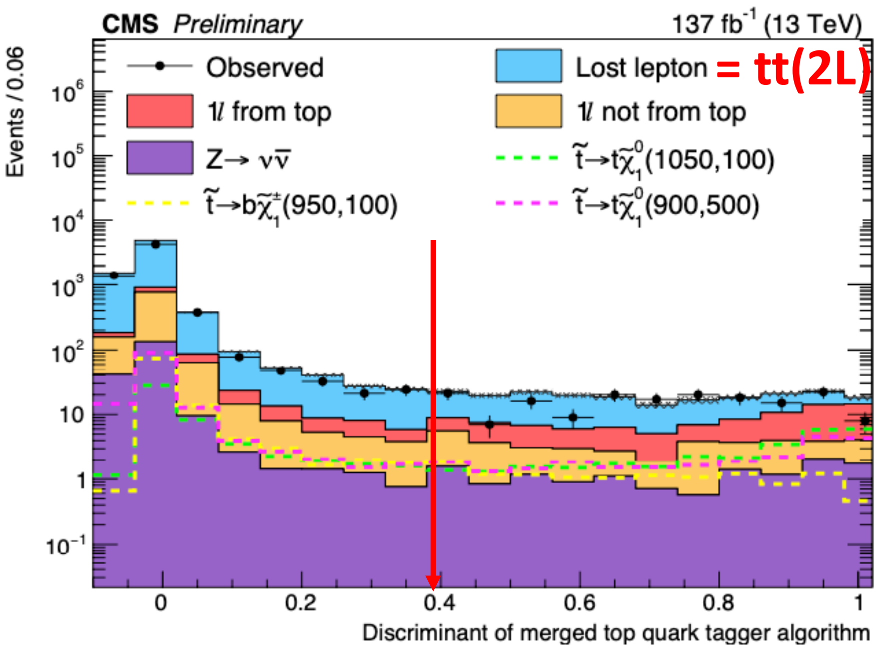
Inspired from: arXiv: 1212.4495

New ! SUSY searches: e.g. Top squarks in 1Lep (II)

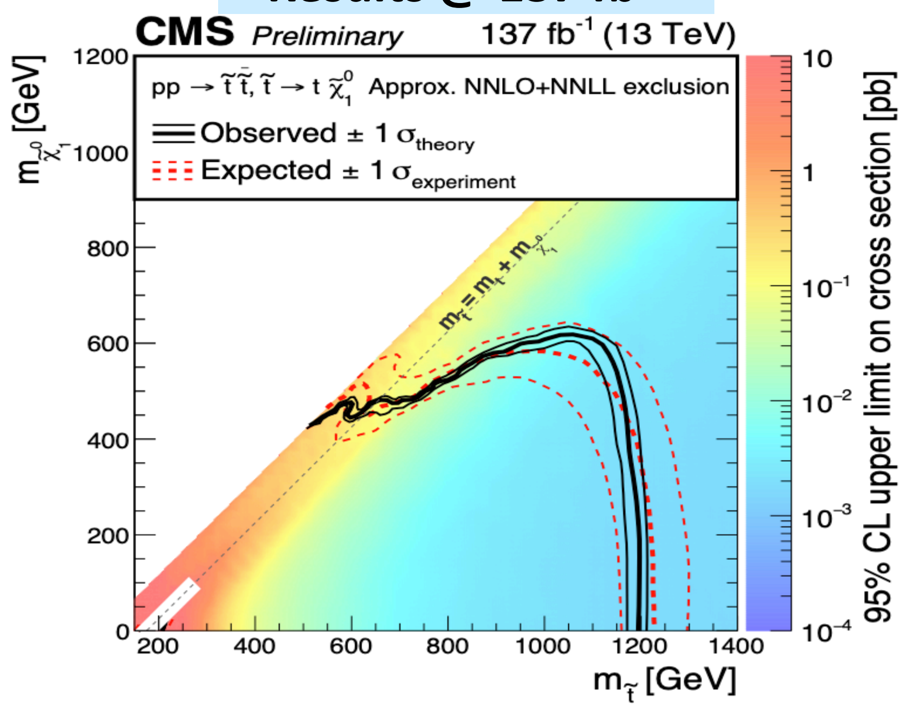
CMS-SUS-19-009

- **New** for last version of the search [with full Run2 data]:
  - Introduce categories with hadronic top quarks [improve tt(2L) rejection]
    - using **DeepAK8** (top score) for high-p_T & resolved top-tagger for moderate p_T

DeepAK8 (top score) in data & MC



Results @ 137 fb^{-1}

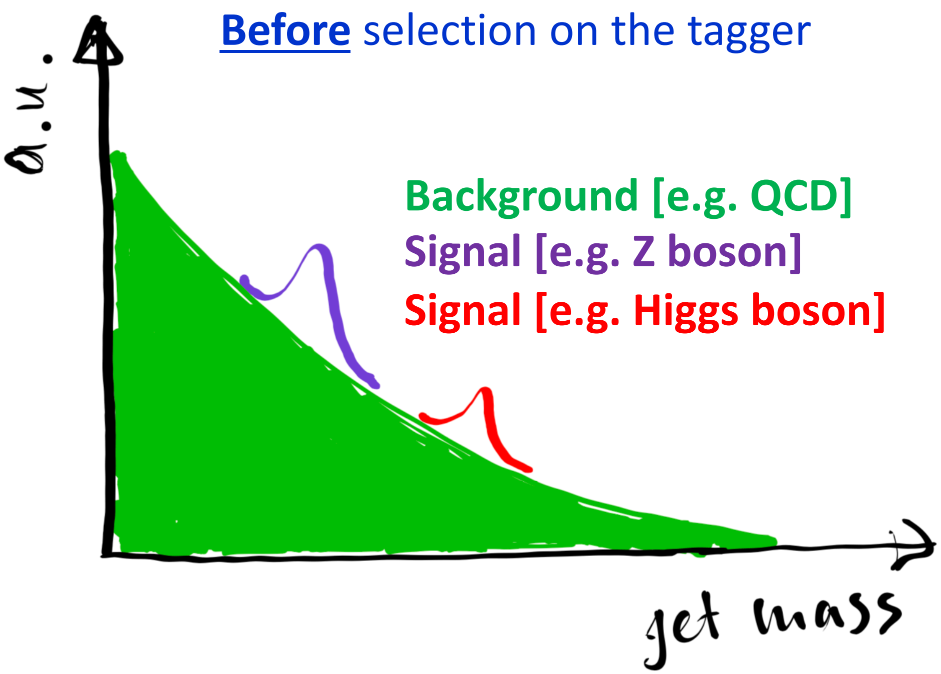


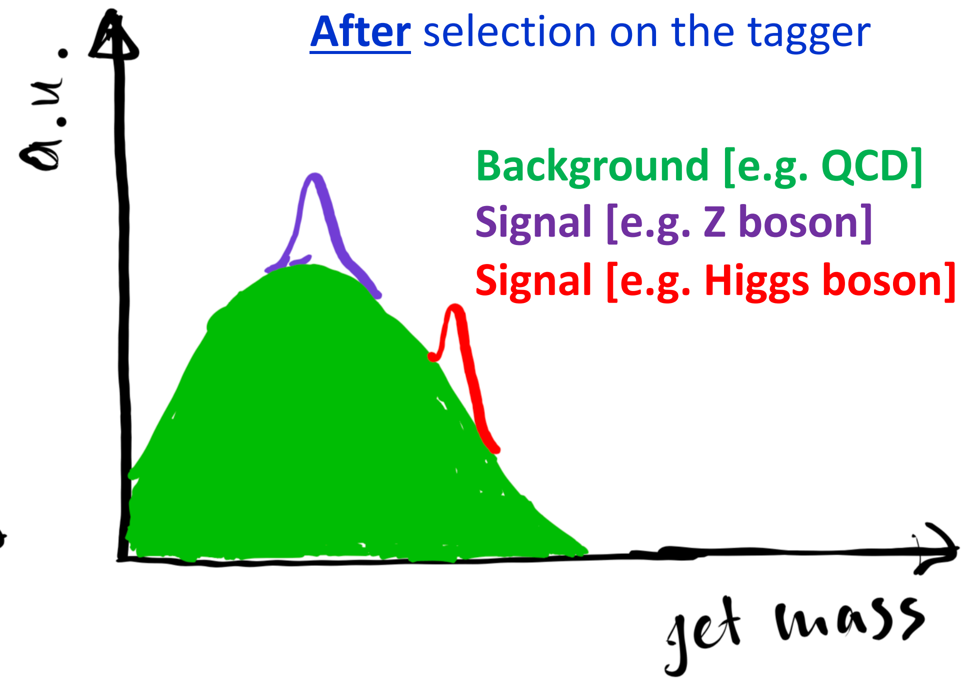
Pre-fit distributions
Overall very good agreement

Top tagging improves reach up to ~50 GeV in m_{stop}

Mass (de-)correlation

- **But:** Taggers are usually correlated with jet mass [i.e. mass-dependent]



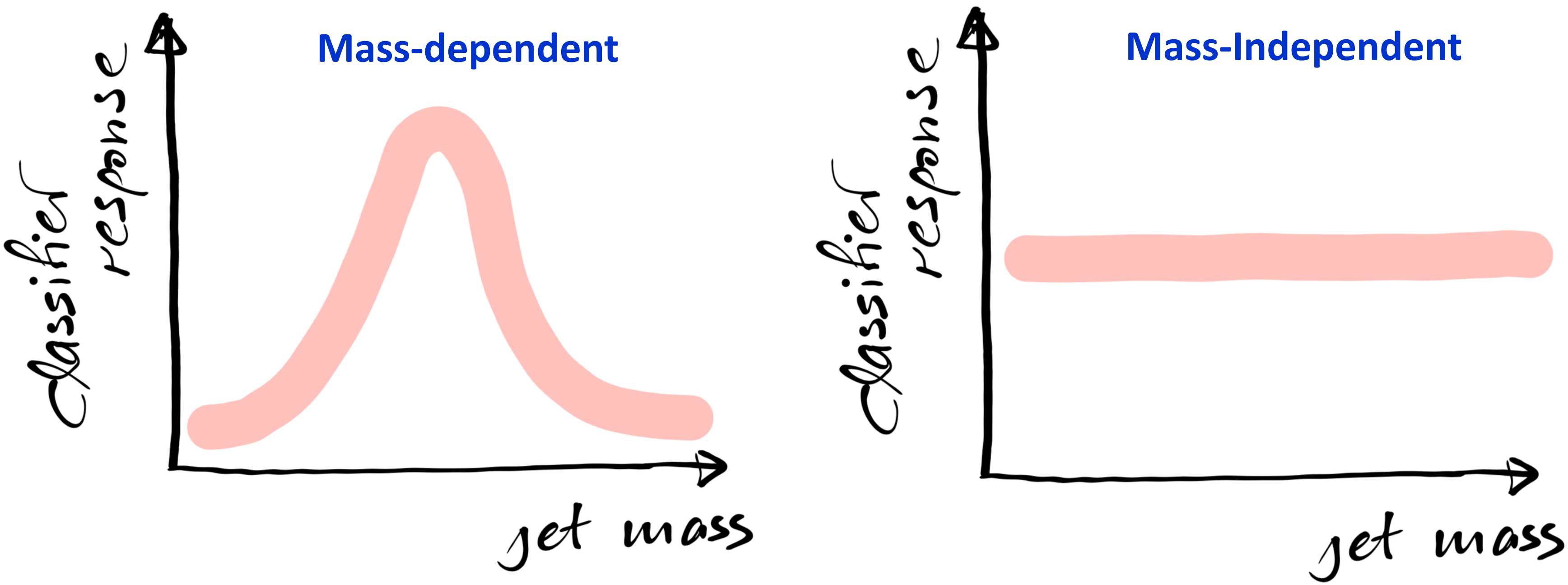


Jet Mass distribution in **Background** jets becomes similar to that from signal: **"Mass sculpting"**

- Depending on the analysis this may not be welcome

Mass (de-)correlation (II)

- What "mass-dependence" mean for the tagging algorithm?



- Different methods explored in CMS to mitigate mass sculpting

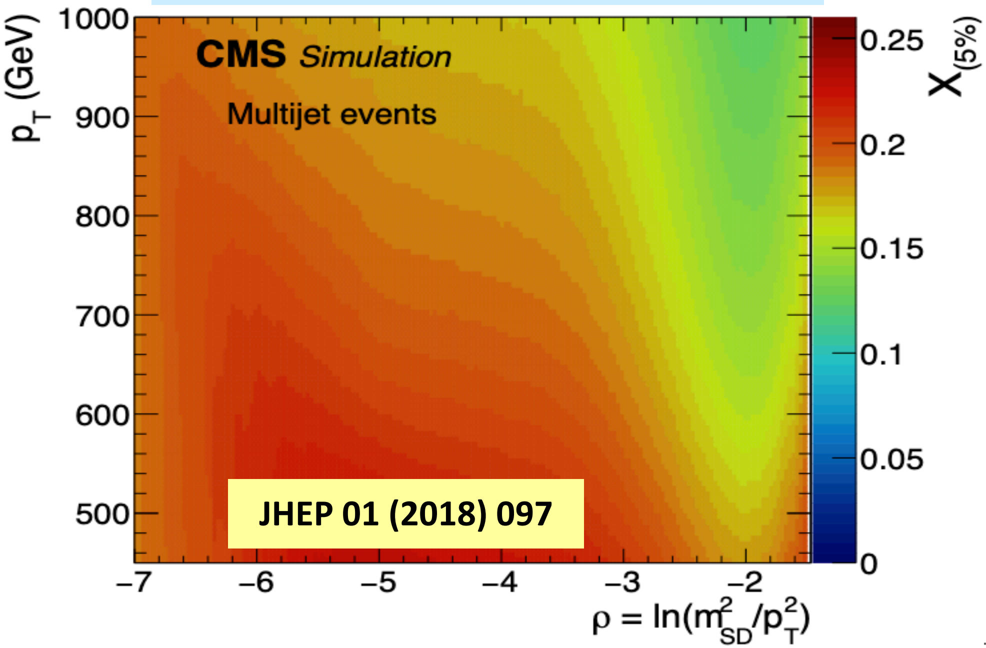
Mass de-correlation in CMS

- The “simpler” approach: **Sample reweighting**
  - Reweight for the training the QCD jet mass distribution to match the signal one

Mass de-correlation in CMS (II)

- <u>The "simpler" approach:</u> **Sample reweighting**
  - Reweight for the training the QCD jet mass distribution to match the signal one
- <u>The "brute force" approach:</u> "**D**esigning **D**ecorrelated **T**agger" (DDT)
  - Define a metric e.g., $\boxed{\rho = \ln(m_{SD}^2/p_T^2)}$ to capture tagger's correlation with m(jet)
  - **Then:** transform tagger's response to preserve constant BKG rejection across **jet mass** and $\mathbf{p_T}$: $\mathbf{Tagger^{DDT} = Tagger - X_{(\#\%)}}$

DDT map for 5% BKG rejection



Caveat:
Need a different transformation (map) for different working points

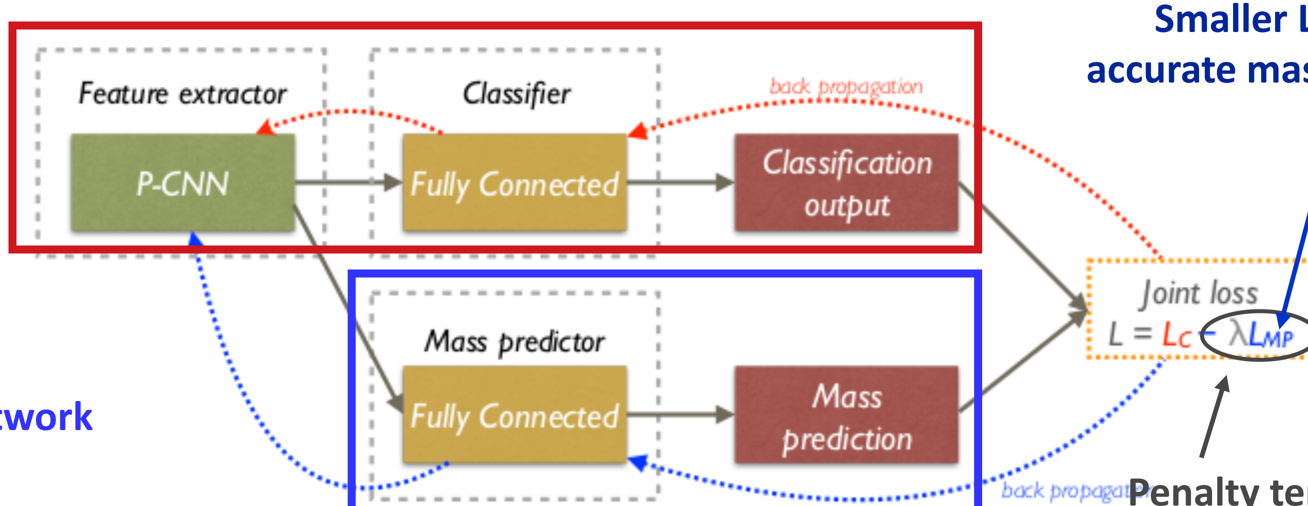
Mass de-correlation in CMS (III)

New !

- The "simpler" approach: **Sample reweighting**
  - Reweight for the training the QCD jet mass distribution to match the signal one
- The "brute force" approach: "**D**esigning **D**ecorrelated **T**agger" (DDT)
  - Define a metric e.g., $\boxed{\rho = \ln(m_{\mathrm{SD}}^2/p_{\mathrm{T}}^2)}$ to capture tagger's correlation with m(jet)
  - **Then:** transform tagger's response to preserve constant BKG rejection across **jet mass** and **p_T**: **$\mathrm{Tagger}^{\mathrm{DDT}}$ = Tagger – $X_{(\#\%)}$**
- The "painful" approach: **Adversarial networks**
  - Introduce a mass prediction network to predict m(jet) from the features extracted by the algorithm (DNN)

Nominal algorithm

+ mass prediction network



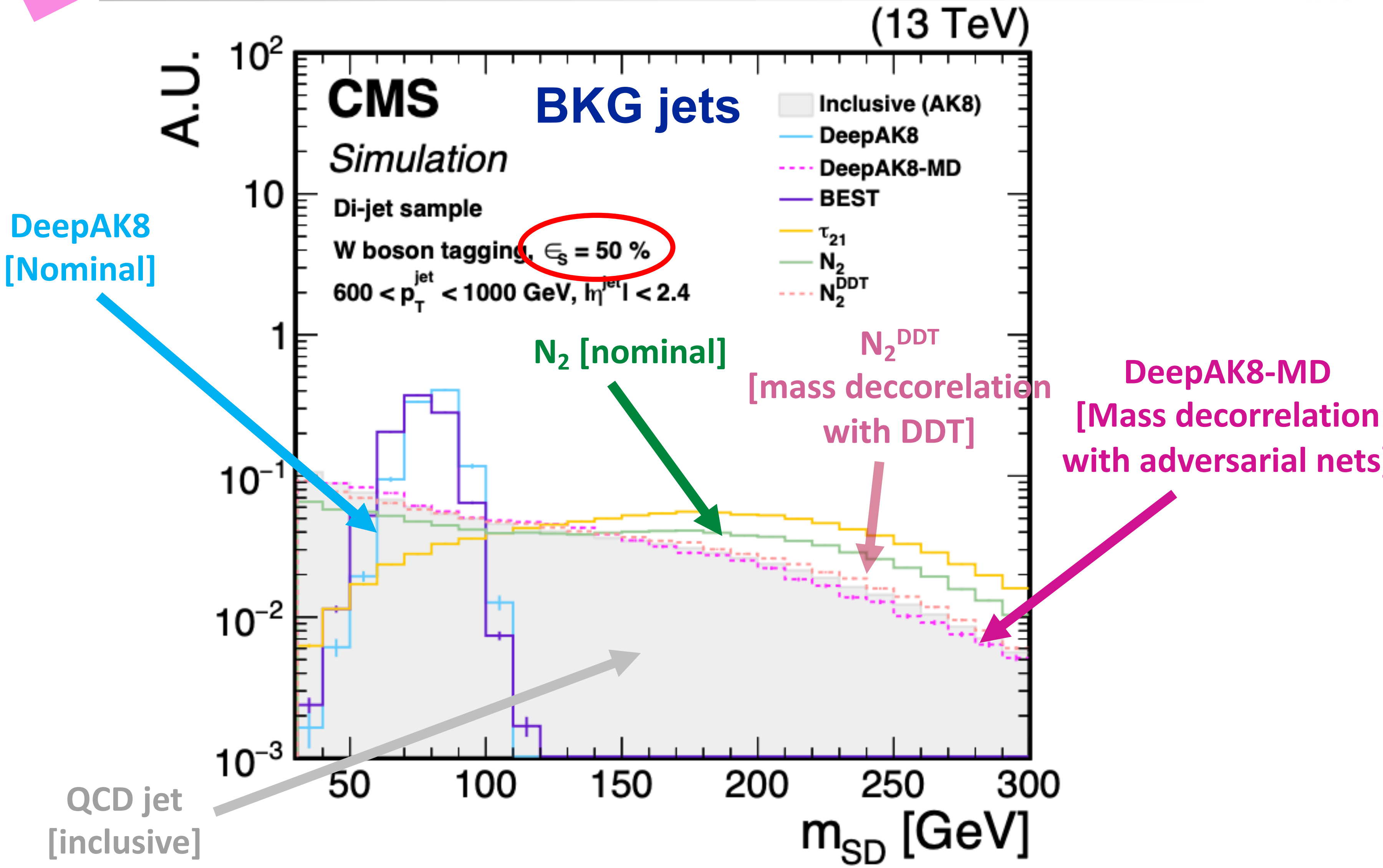
Smaller L_{MP} more accurate mass prediction

Penalty term

[λ: balancing between performance and mass decorrelation]

Mass (de-)correlation: Results

New !

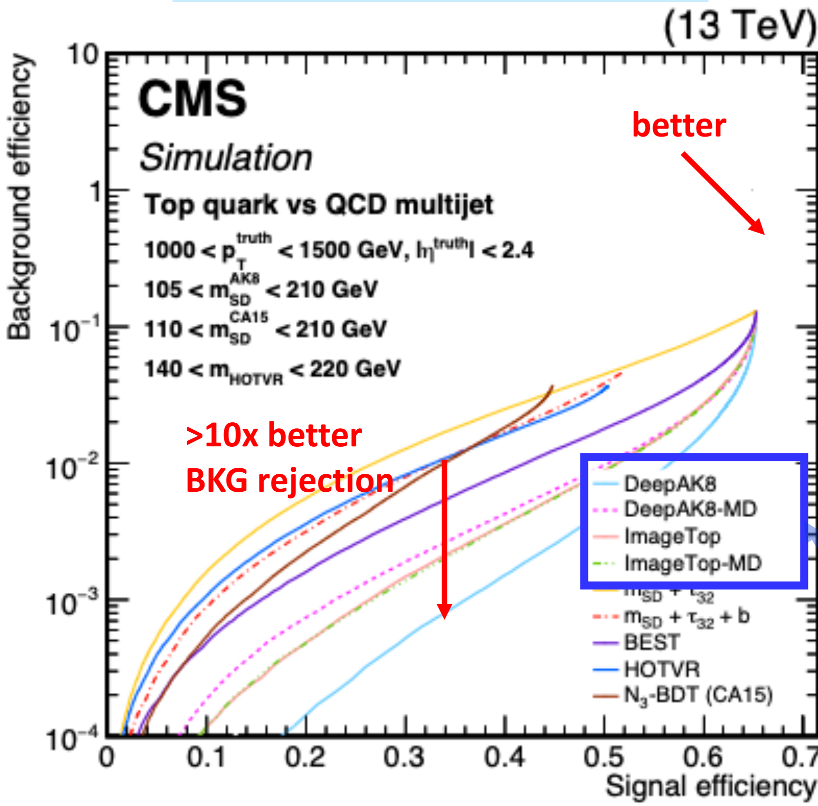


Mass sculpting significantly reduced with mass decorrelated versions

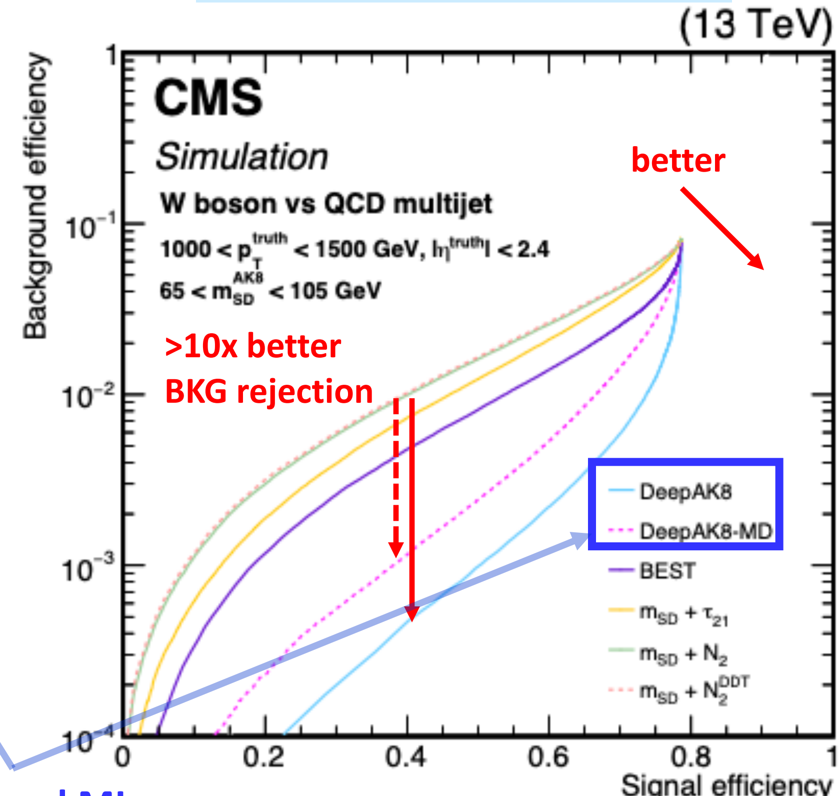
Top & W tagging performance

New !

Top quark tagging



W boson tagging



Advanced ML
+ Low-level inputs

New generation taggers show significantly improved performance
Performance stable wrt jet kinematics & data taking conditions (e.g. PU)

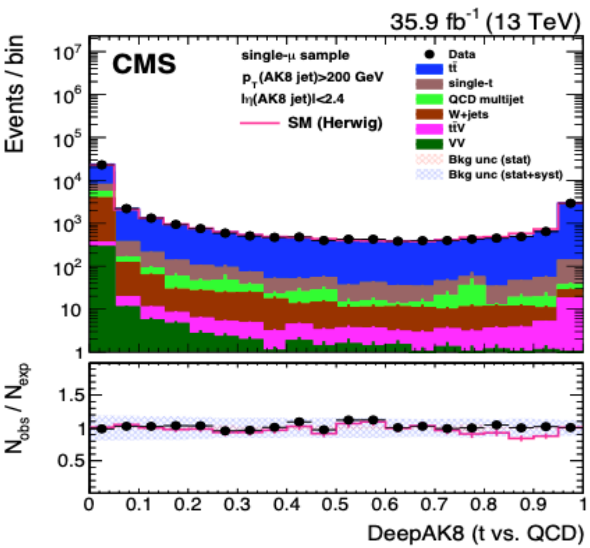
Performance in data [top & W]

New !

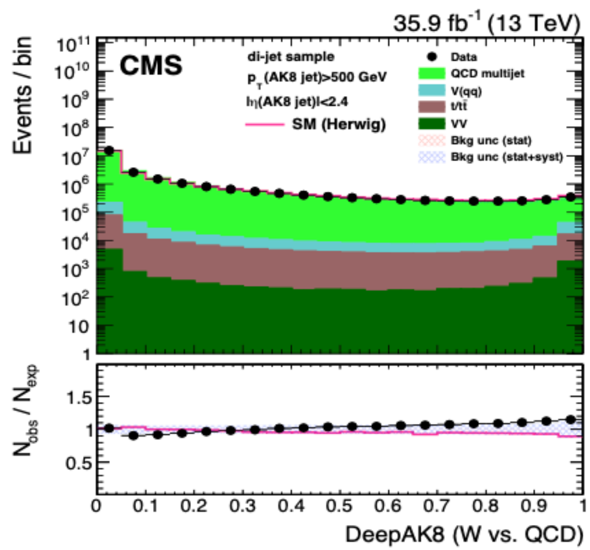
- Three orthogonal data samples used for the validation of top/W performance
  - **Signal sample:** single-mu events; ttbar(1L) dominated;
    - 2 regions: $p_T(j)>200$ [W-enhanced] and $p_T(j)>500GeV$ [top-enhanced]
  - **Background samples:** dijet ($H_T>1000$ GeV) and γ+jet events ($p_T(\gamma)>200$ GeV)
    - different quark – gluon fraction

e.g., DeepAK8 [i.e. advanced ML + low-level inputs] in data

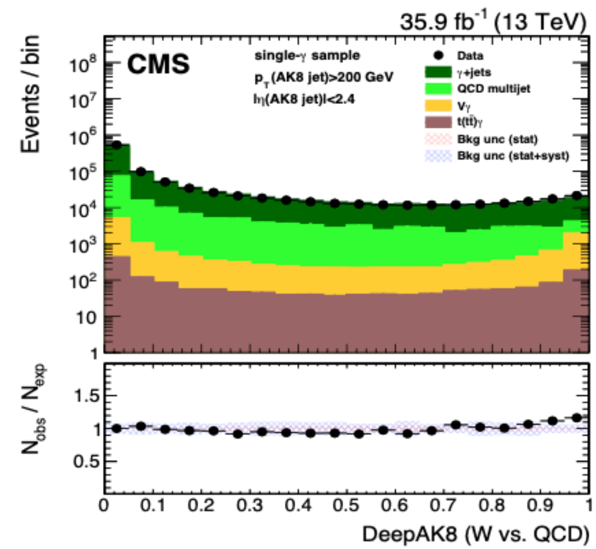
Single-μ



Dijet



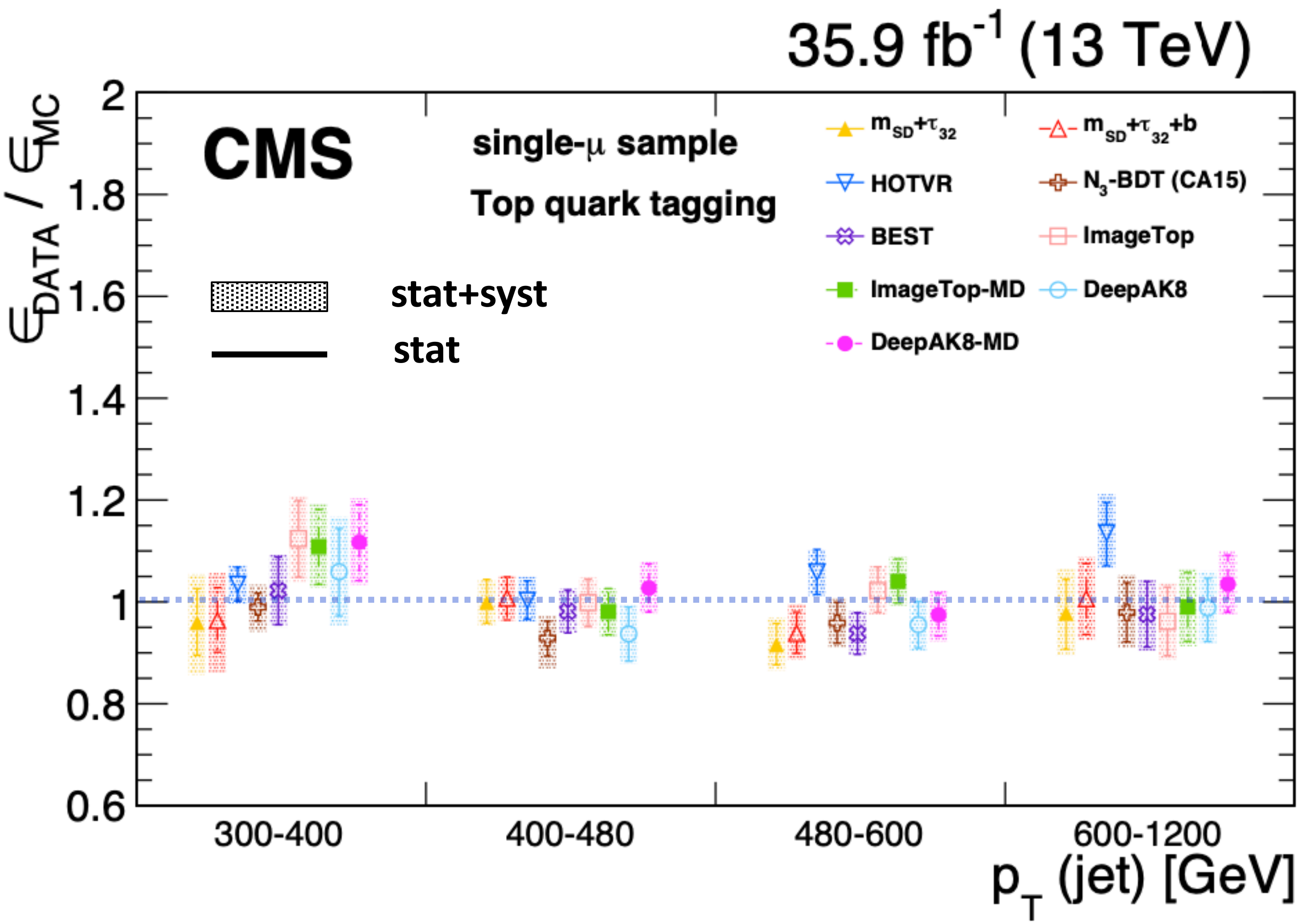
Single-γ



Very good overall data – MC agreement

Corrections to simulation: Top tagging eff.

New !

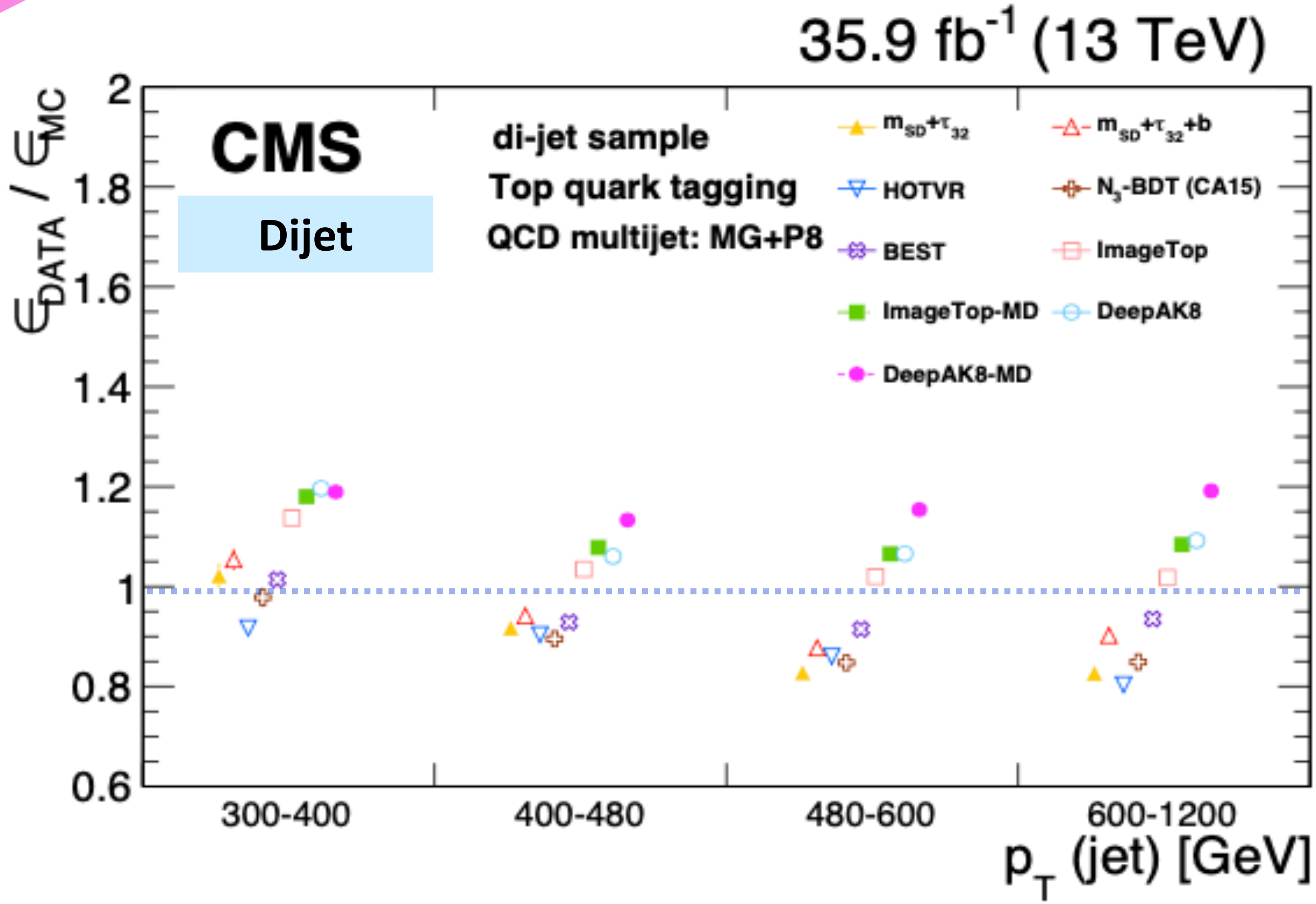


Performance translates in data
[similar conclusions for the W tagging case]

Systematic uncertainties on more advanced taggers under control

Corrections to simulation: Top mistagging

New !



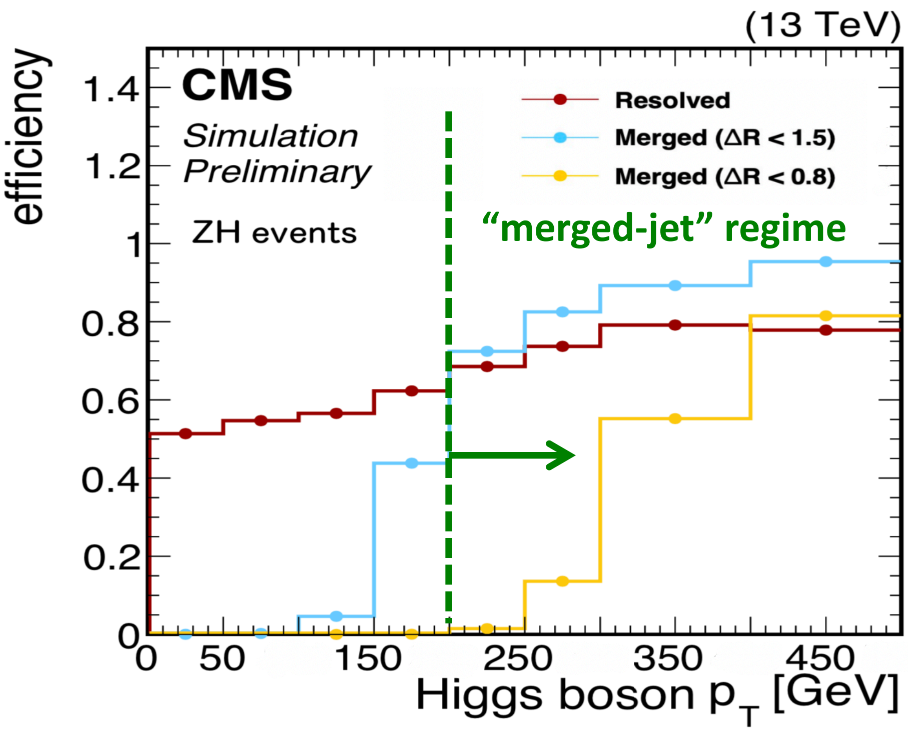
Overall very reasonable agreement between data and MC
SF range 0.8—1.2 [typically larger for tagger using low-level features]
[more studies on sample dependence in back up]

Direct search for H→cc

New !

CMS-HIG-18-031

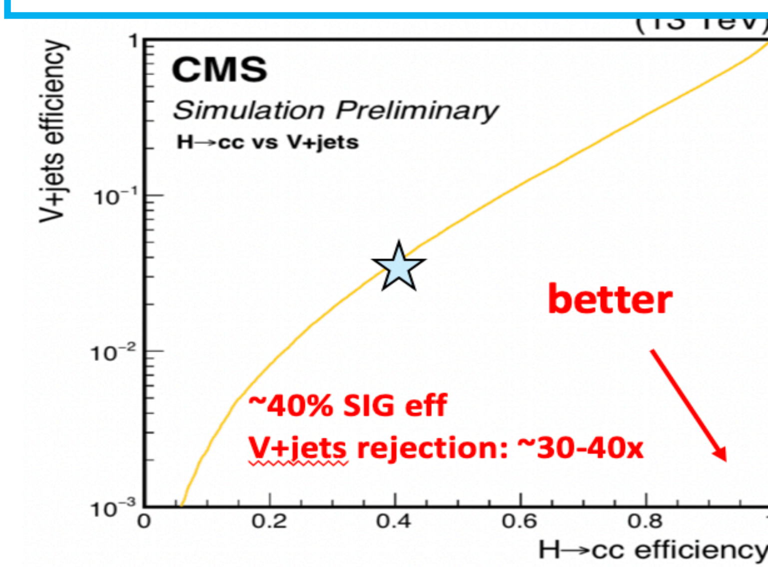
- New tagging tools used in the first CMS search for VH(→cc)
  - Very challenging to hunt at the LHC
    - very small expected signal & charm tagging more challenging than b-tagging
- **Cornerstone of the search:** reconstruction of two charm quarks



Merged-jet topology

→ A single **AK15** (R=1.5) jet to reconstruct the Higgs H→cc decay

→ Use new **DeepAK8-MD** algorithm [First cc-tagger at the LHC]



Exploit as much as possible of Higgs decay topology by combining the resolved and merged topologies

Calibration in data using jets from gluon splitting

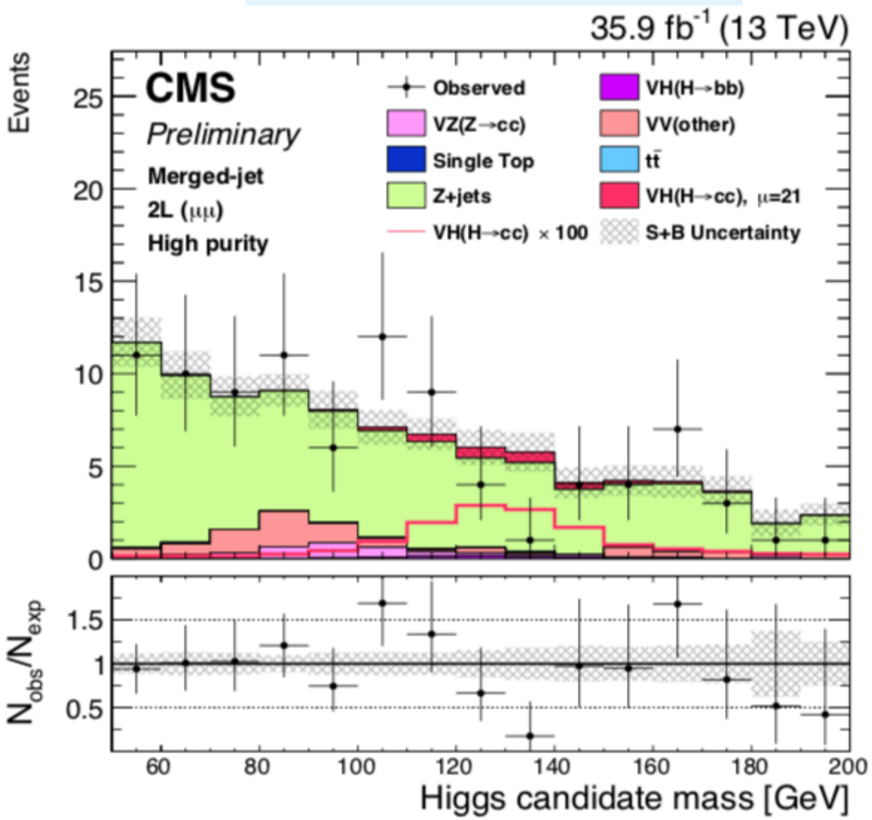
New !

Direct search for H→cc

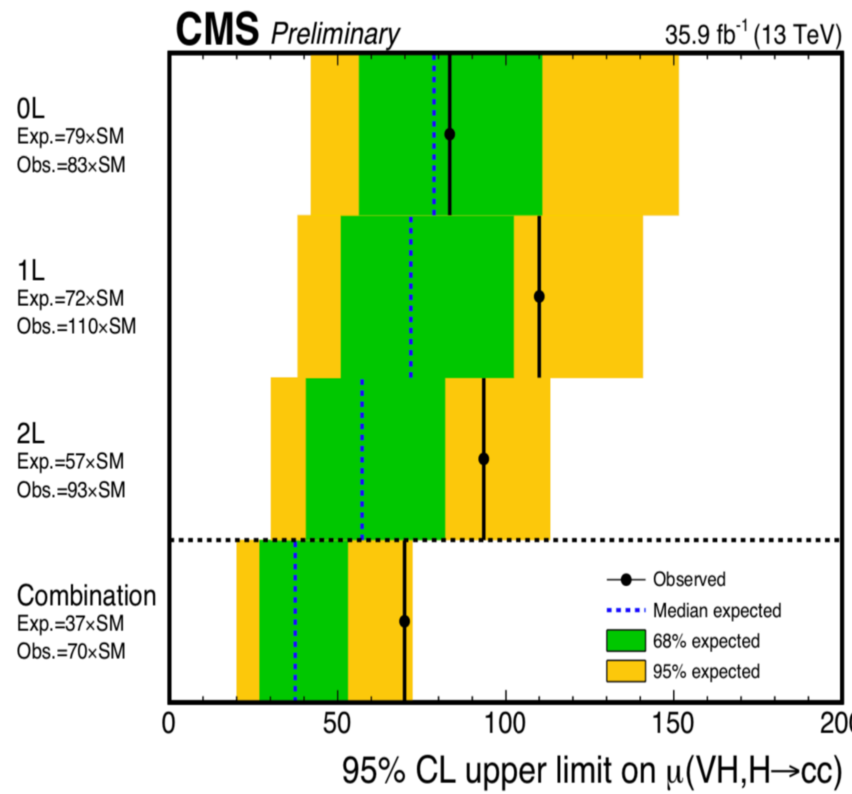
CMS

CMS-HIG-18-031

DeepAK8-MD; 2L (μμ)



95% C.L. exclusion limit on $\mu_{VH(cc)}$:



95% C.L. exclusion limit on $\mu_{VH(\to cc)}$:
70 (37) observed (expected)

VH(→cc) results with 35.9 fb^{-1} (2016) are the tightest limits to date

Summary and Outlook

Summary & outlook

- Jet tagging is essential for the success of the LHC Physics program
  - Large effort at the LHC to improve existing / develop new jet tagging methods

- **Key player in these developments:** Advanced machine learning algorithms
  - Allows us see much more of the true potential of the CMS apparatus
  - Still room for improvement / other ideas to try
    - Strong interest by the theory and experiment communities

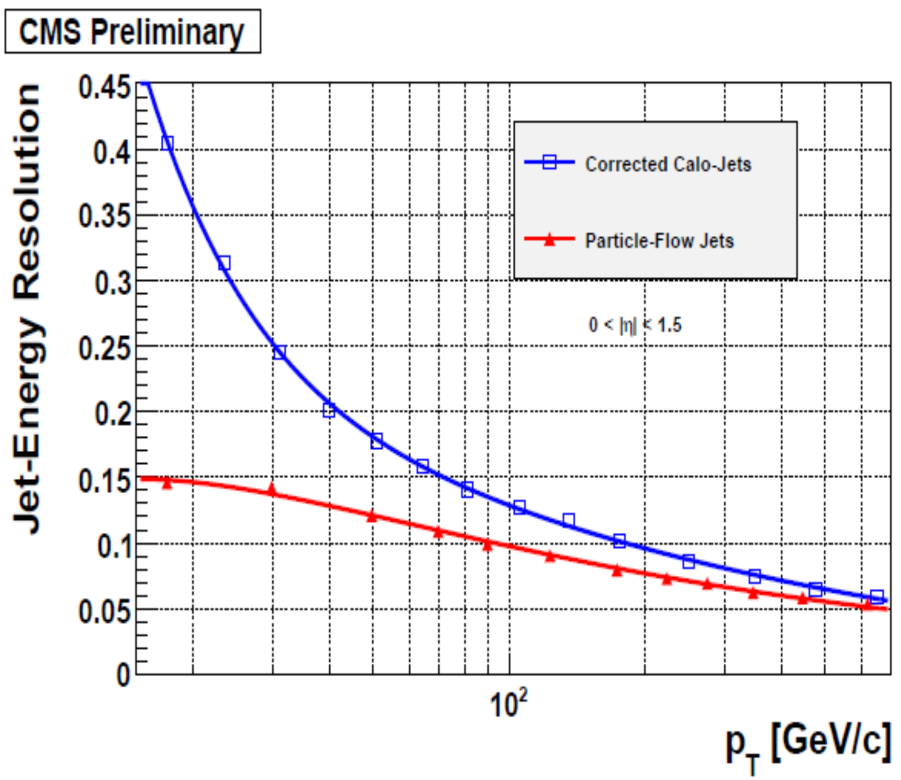
- **Effort pays off:** Large gain in performance wrt traditional approaches
  - Performance translates in data
  - Lots of effort to better understand what the DNN learns
  - Application of the new tools in physics analyses started
    - Already seeing important gain in physics reach

Many more (exciting) results using these tools in the pipeline
Stay tuned !

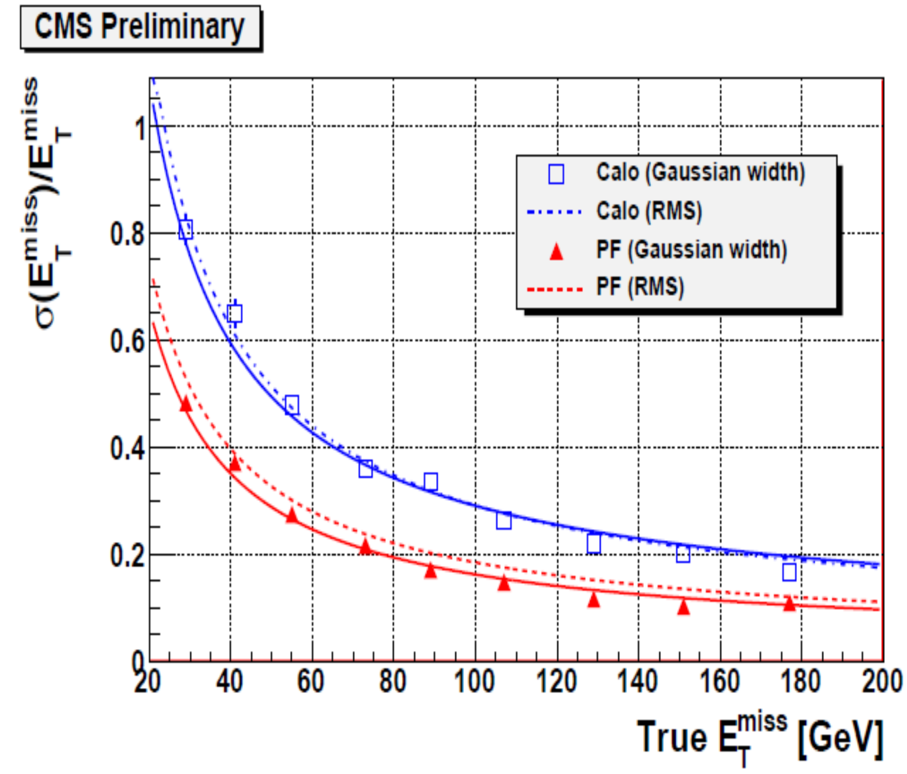
Backup

From Calo → PF

Jet resolution



ME_T resolution

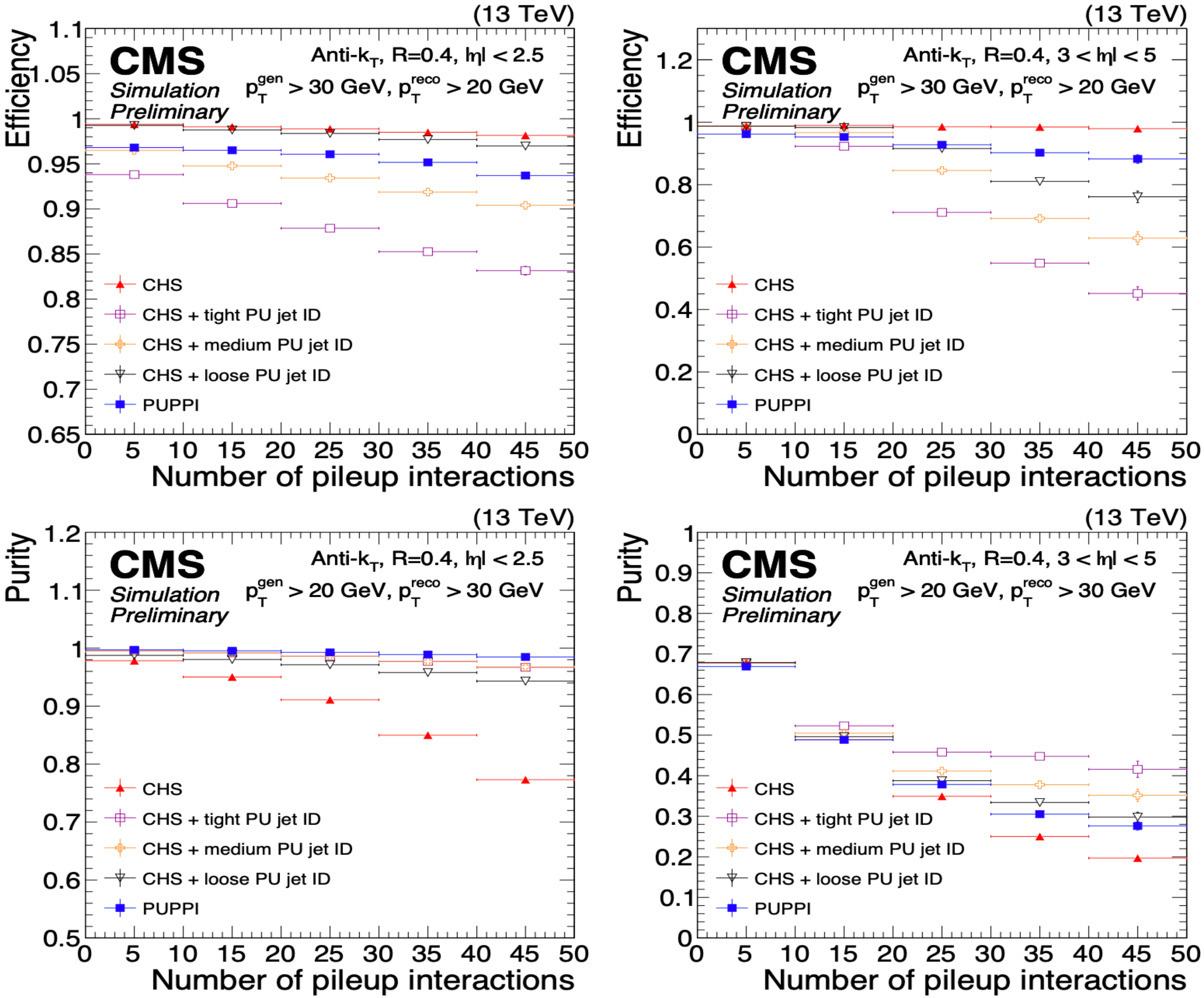


Significant improvement in object performance wrt traditional approaches

Pileup jet Id inputs

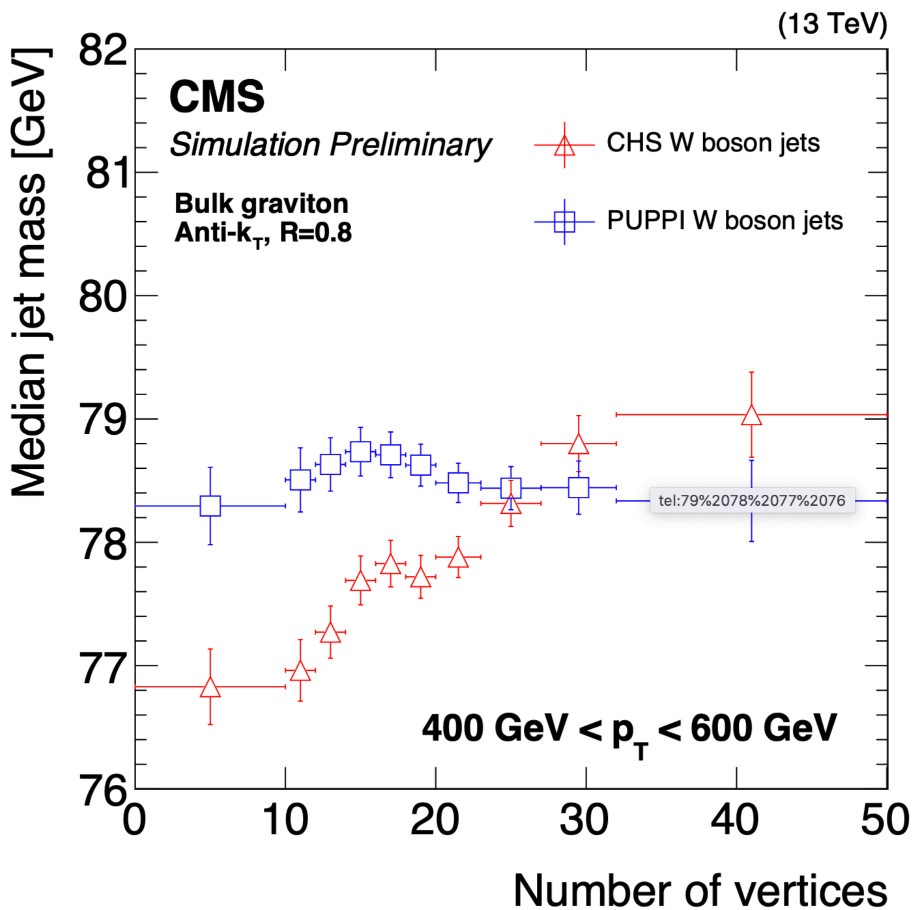
β	fraction of transverse momentum of charged particles associated to the primary vertex, defined as $\frac{\sum_{i\in V} p_{Ti}}{\sum_i p_{Ti}}$ where i iterates over all the PF particles in the jet
$n_{vertices}$	number of vertices in the event
$\langle \Delta R^2 \rangle$	p_T^2 average weighted by square distance of jet constituents from the jet axis : $\frac{\sum_i \Delta R^2 p_{Ti}^2}{\sum_i p_{Ti}^2}$
$f_{\mathrm{ring}X}$, $X =$ 1, 2, 3, and 4	fraction of p_T of the constituents ($\sum p_{Ti}/p_T^{\mathrm{jet}}$) in the region $R_i < \Delta R < R_{i+1}$ around the jet axis, where $R_i = 0, 0.1, 0.2,$ and 0.3 for X=1, 2, 3, and 4
$p_T^{\mathrm{lead}}/p_T^{\mathrm{jet}}$	transverse momentum fraction carried by the leading PF candidate
$p_T^{\mathrm{l.\,ch.}}/p_T^{\mathrm{jet}}$	transverse momentum fraction carried by the leading charged PF candidate
$\lvert\vec{m}\rvert$	pull magnitude, defined as $\lvert(\sum_i p_T^i \lvert r_i \rvert \vec{r}_i)\rvert / p_T^{\mathrm{jet}}$ where $\vec{r}_i$ is the direction of the particle i from the direction of the jet
N_{total}	number of PF candidates
N_{charged}	number of charged PF candidates
σ_1	major axis of the jet ellipsoid in the η-ϕ space
σ_2	minor axis of the jet ellipsoid in the η-ϕ space
p_T^D	jet fragmentation distribution, defined as $\sqrt{\sum_i p_{T_i}^2} / \sum_i p_{Ti}$

PUID vs. PUPPI

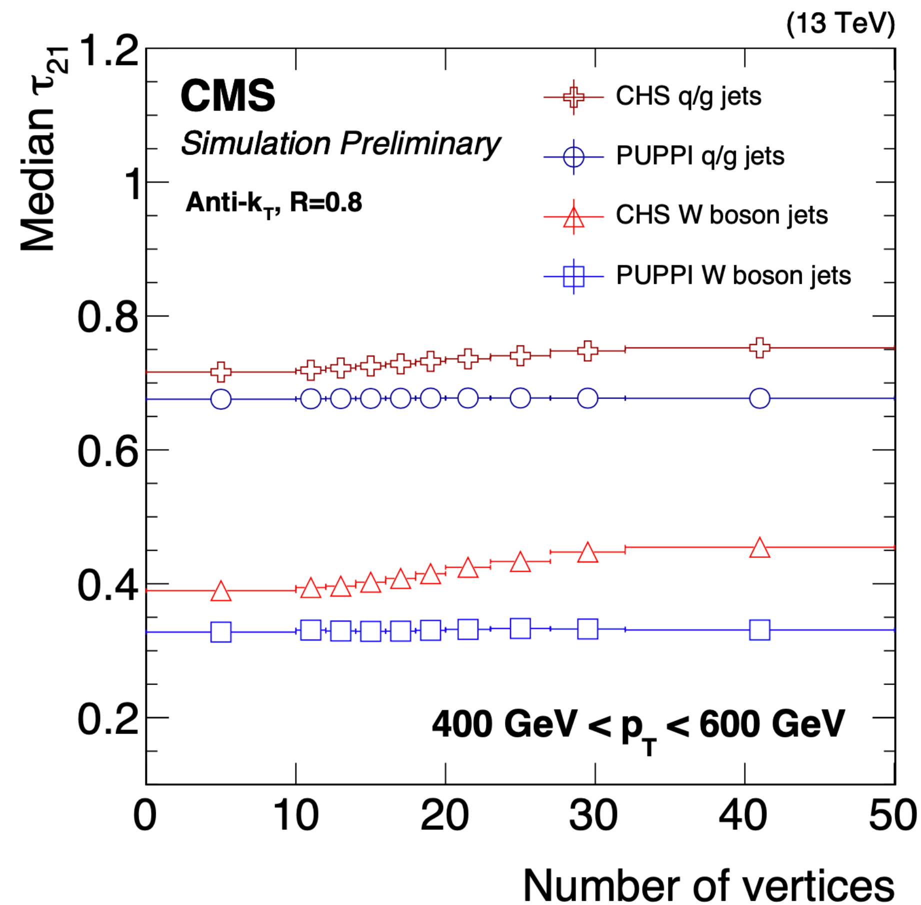


PUPPI performance: substructure

jet mass

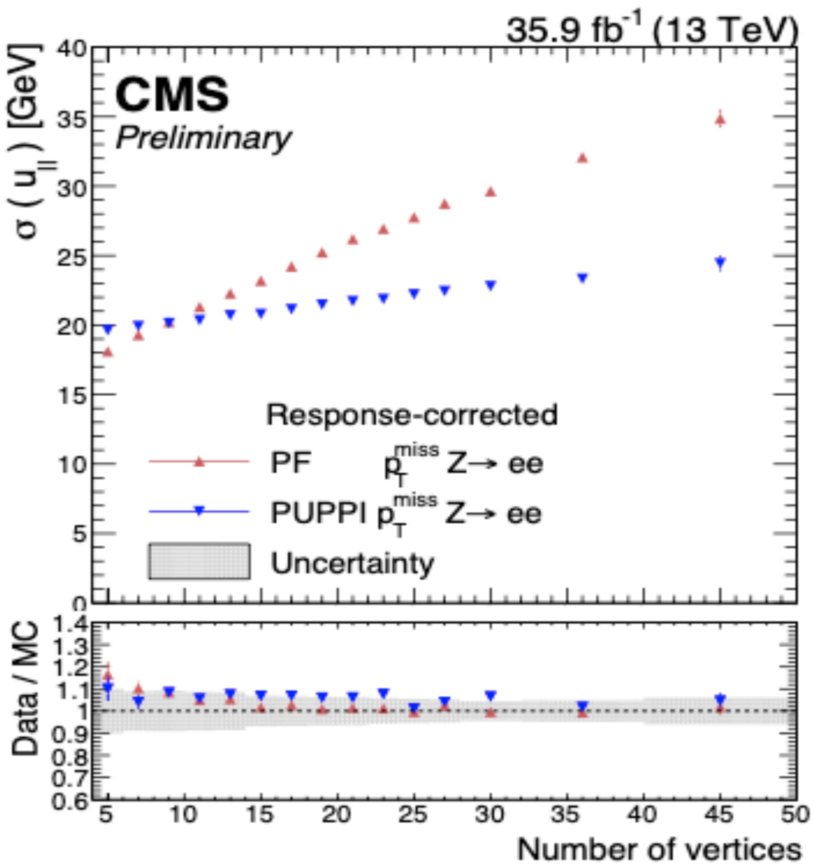


τ_{21}

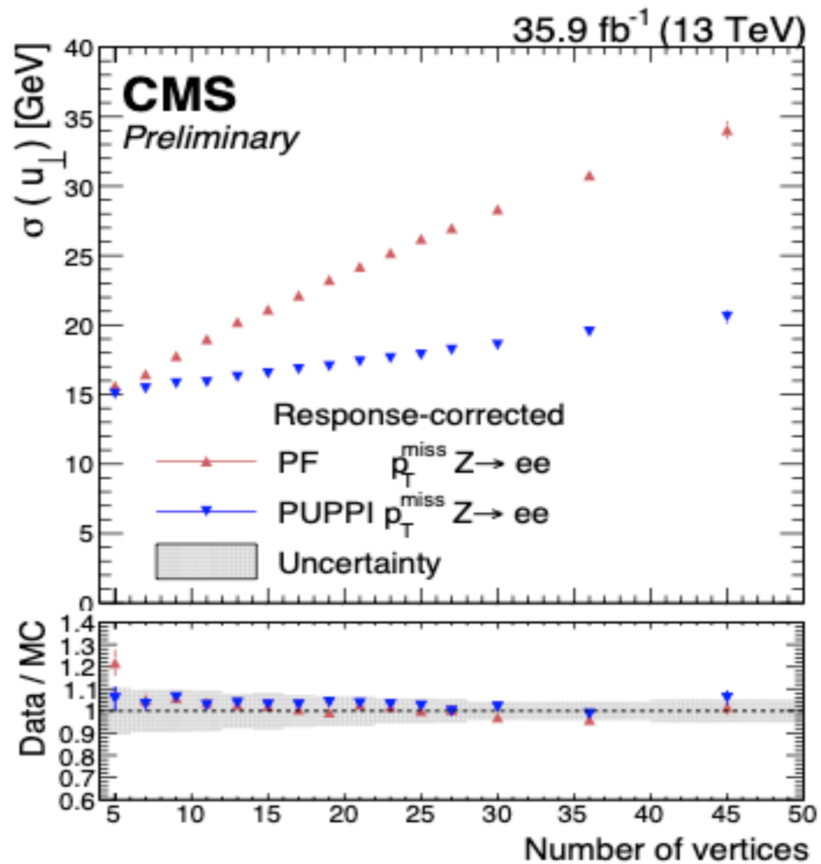


PUPPI performance: ME_T

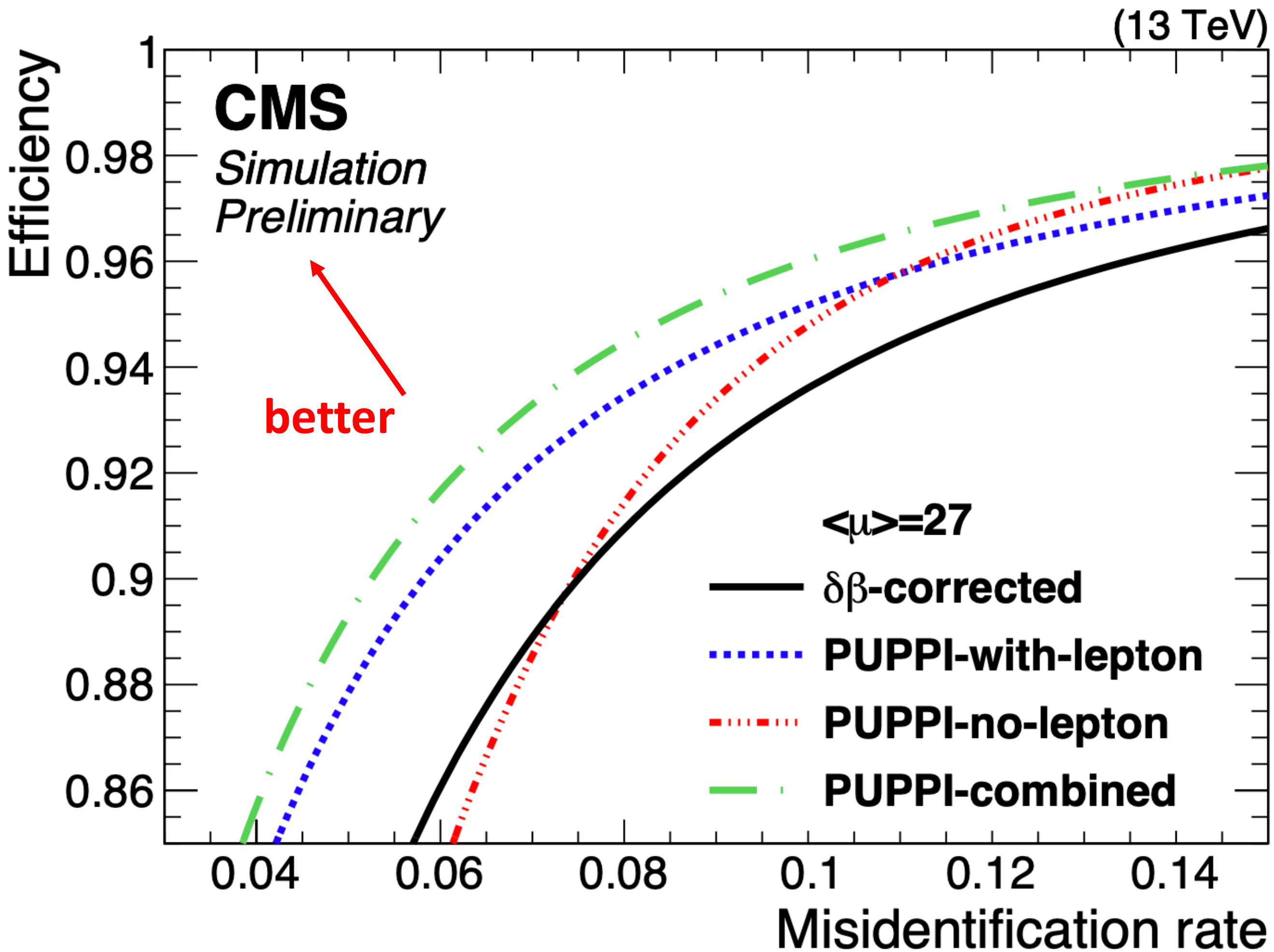
parallel component



perpendicular component



PUPPI performance: Lepton Isolation



similar conclusions for electrons

$$\delta\beta\text{-Iso}^{\mu^i} = \sum_{\Delta R(i,j)<0.4}^{\text{CH-LV}} p_T^j + \max(0, \sum_{\Delta R(i,j)<0.4}^{\text{NH}} p_T^j + \sum_{\Delta R(i,j)<0.4}^{\text{PH}} p_T^j - \frac{1}{2} \sum_{\Delta R(i,j)<0.4}^{\text{CH-PU}} p_T^j)$$

Jet calibration: Improving b quark energy & resolution

- Inputs:

- jet kinematics: jet p_{T}, η, mass, and transverse mass, defined as $\sqrt{E^2 - p_z^2}$;
- information about pileup interactions: the median energy density in the event ρ, corresponding to the amount of transverse momentum added to the event per unit area, for example by minimum bias particles [33].
- information about semileptonic decays of b hadrons (if an electron or muon candidate is clustered within a jet): the transverse component of lepton momentum perpendicular to the jet axis, the distance $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ (where ϕ is the azimutal angle in radians) between the lepton candidate direction and the jet axis, and a variable that encodes information about the lepton candidate's flavor;
- secondary vertex: transverse momentum, mass and number of charged tracks associated with the highest p_{T} secondary vertex linked to the jet (pion mass is assigned to all reconstructed tracks forming the secondary vertex); the distance between the collision vertex and the secondary vertex computed in the three dimensional space with its associated uncertainty [34, 35];
- jet composition: the largest p_{T} value found among those of charged tracks associated to the jet; and fractions of energy carried by jet constituents split in electrons and photons, charged hadrons, neutral hadrons, and muons. These fractions are computed for the whole jet, and separately in five rings of ΔR around the jet axis; $\Delta R = 0$–0.05, 0.05–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4.
- multiplicity of the PF candidates clustered into the jet; and
- information about jet energy sharing among the jet constituents : $p_T D = \frac{\sum_i p_{T,i}^2}{\sum_i p_{T,i}}$ where i runs over all jet constituents.

Jet calibration: Improving b quark energy & resolution

- Huber loss function:

$$H_\delta(z) = \begin{cases} \frac{1}{2}z^2 & \text{, if } |z| < \delta; \\ \delta|z| - \frac{1}{2}\delta^2 & \text{, otherwise.} \end{cases}$$

  - z: difference between target and prediction

- Quantile loss function:

$$\rho_\tau(z) = \begin{cases} q_\tau \cdot z & \text{, if } z > 0; \\ (q_\tau - 1) \cdot z & \text{, otherwise,} \end{cases}$$

- Total loss:

$$\text{loss}(x,y) = E_{(x,y)\sim p(x,y)}[H_1(y - \hat{y}_0(x)) + \rho_{75\%}(y - \hat{y}_1(x)) + \rho_{25\%}(y - \hat{y}_2(x))]$$

- Outputs:
  - Energy
  - Resolution: half of the difference between the 75% and 25% quantiles of the target distribution

Example of ECF calculation

- General expression:

$$
{}_{o}e_{N}^{\beta} = \sum_{1 \le i_1 < i_2 < \cdots < i_N \le N_C} \left[\prod_{1 \le k \le N} \frac{p_{\mathrm{T}}^{i_k}}{p_{\mathrm{T}}^{J}} \right] \prod_{m=1}^{o} \overset{(m)}{\min_{i_j < i_k \in \{i_1, i_2, \cdots, i_N\}}} \left\{ \Delta R_{i_j, i_k}^{\beta} \right\}
$$

- Example:

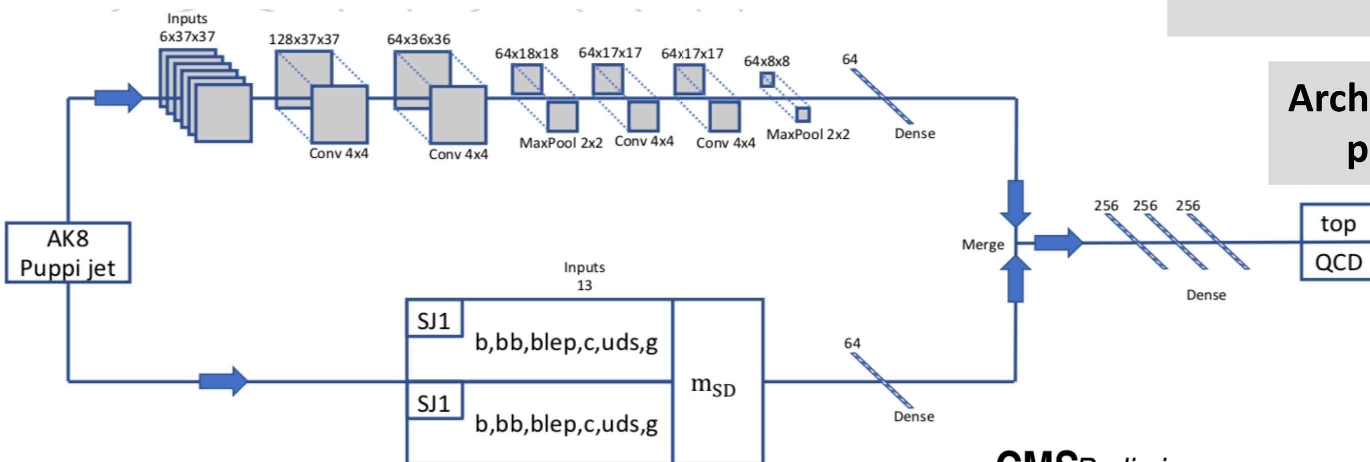
$$
{}_{2}e_{3}^{1} = \sum_{1 \le a < b < c \le M} \frac{p_{\mathrm{T}}^{a}}{p_{\mathrm{T}}^{J}} \frac{p_{\mathrm{T}}^{b}}{p_{\mathrm{T}}^{J}} \frac{p_{\mathrm{T}}^{c}}{p_{\mathrm{T}}^{J}} \min\{\Delta R_{ab} \Delta R_{ac}, \Delta R_{ab} \Delta R_{bc}, \Delta R_{bc} \Delta R_{ac}\}
$$

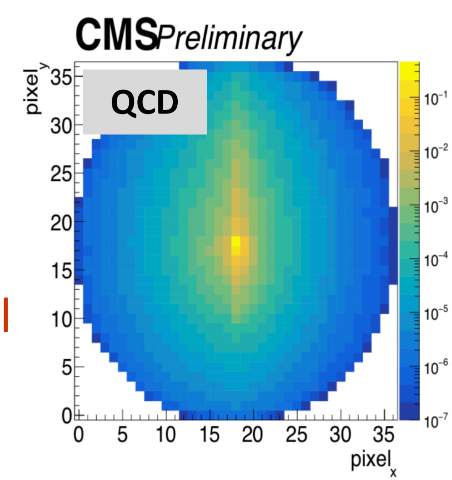
Particle based jet tagging: ImageTop

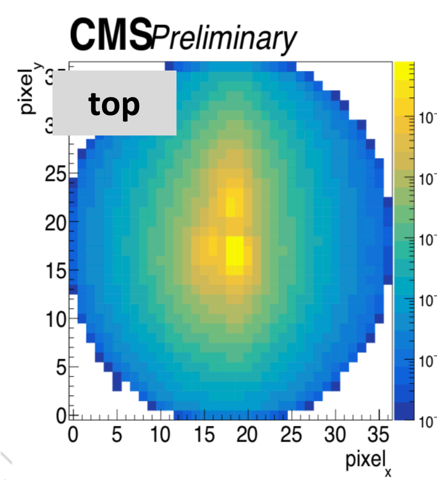
- DNN-based top tagger using PF candidates & apply DNN-based flavor tagging on the subjets
  - PF candidates: split into relevant "channels" based the PF candidate flavor
- 2D-CNN: 64-128 maps; 4x4 kernel.

channels: neutral p_T, track p_T, mu p_T, ..

Architecture based on TH paper: 1803.00107





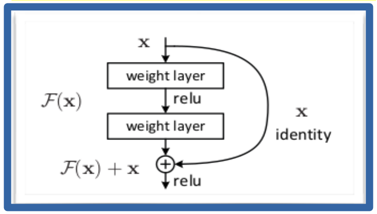


- Preprocessing:
  - Shift so that centroid at origin
  - Rotate such that principal axis is vertical
  - Flip to orient third maxima

Highlights from the network architecture

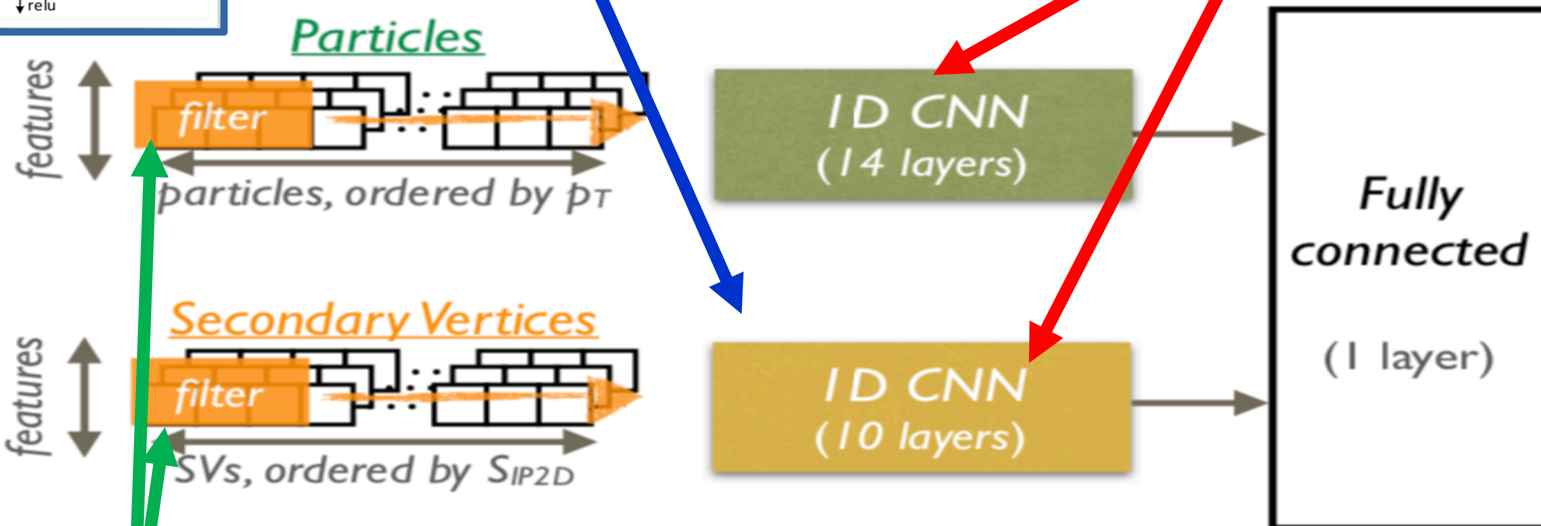
- Inspired by ResNeXt50:

1512.03385,1603.05027



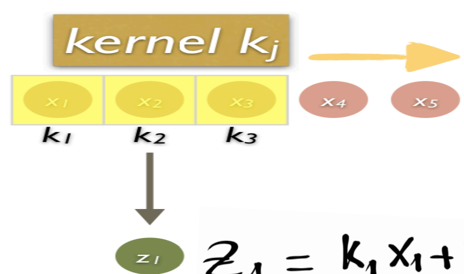
Residual connections
[improves performance & makes training easier]

Deep network to better exploit particle correlations



"Move" in particle triplets
Exploit correlation between nearby particles faster

Particle sequence:

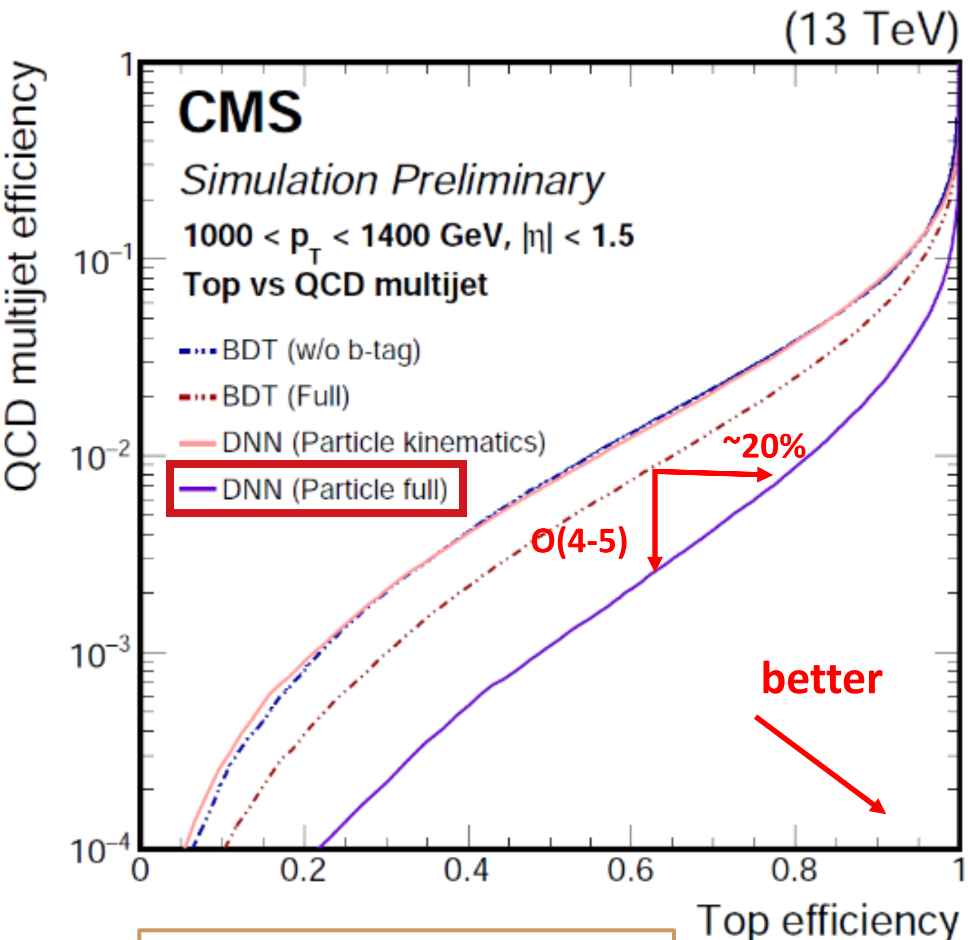


$$z_1 = k_1 x_1 + k_2 x_2 + k_3 x_3$$

e.g. for one feature:
[e.g. p_T of particle]

Performance in MC

e.g. Top vs. QCD



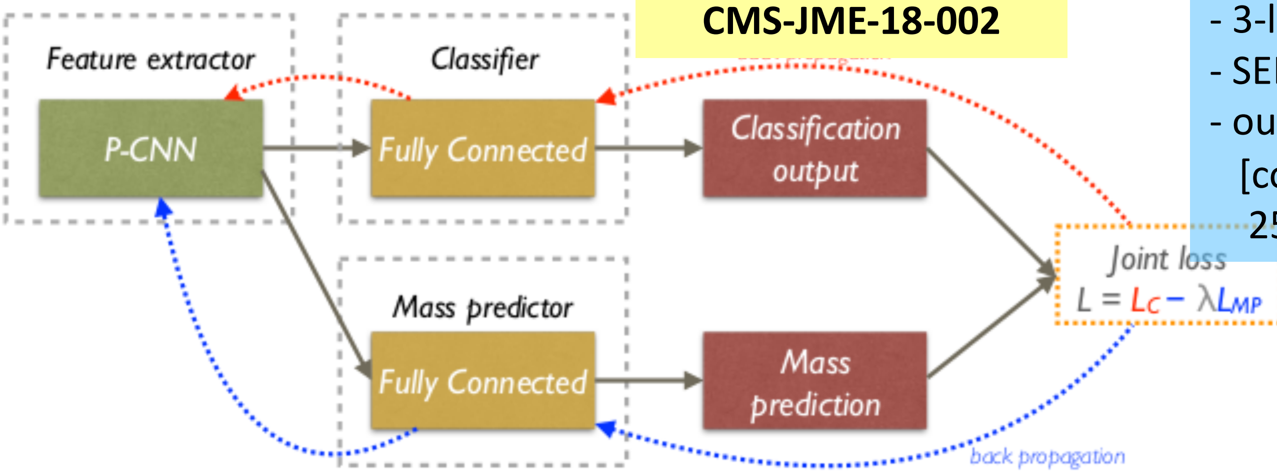
DPS-2017-049,
NIPS 2017 paper

- Compare performance vs. a BDT-based approach which uses state of the art high-level features [ref. 10.1007/JHEP10(2017)005]
  - inputs: jet substructure, b/c flavour tagging, quark-gluon tagging
- Investigate impact of flavour tagging in performance
- Similar performance between the DNN and the BDT approach without flavour info
- Very significant gain [~4-5x] for the Particle-based DNN approach after including flavour

Merged-jet: mass decorrelation

- Use adversarial training to regulate the behaviour of the network
  - Introduce a mass prediction network to predict the jet mass from the features extracted by the CNNs
  - It's loss (L_{MP}) is an indicator for mass correlation; network trained to maximize L_{MP} ; making it harder for the adversary to learn the mass
    - Smaller L_{MP} more accurate mass prediction ; the features extracted by the CNNs have a higher correlation with jet mass
  - Introduce a joint loss: $L = L_C - \lambda L_{MP}$, second term a penalty on mass correlation
    - Minimizing L -> simultaneously improve classification & reduce mass correlation
    - λ: hyperparameter balancing between performance and mass independence

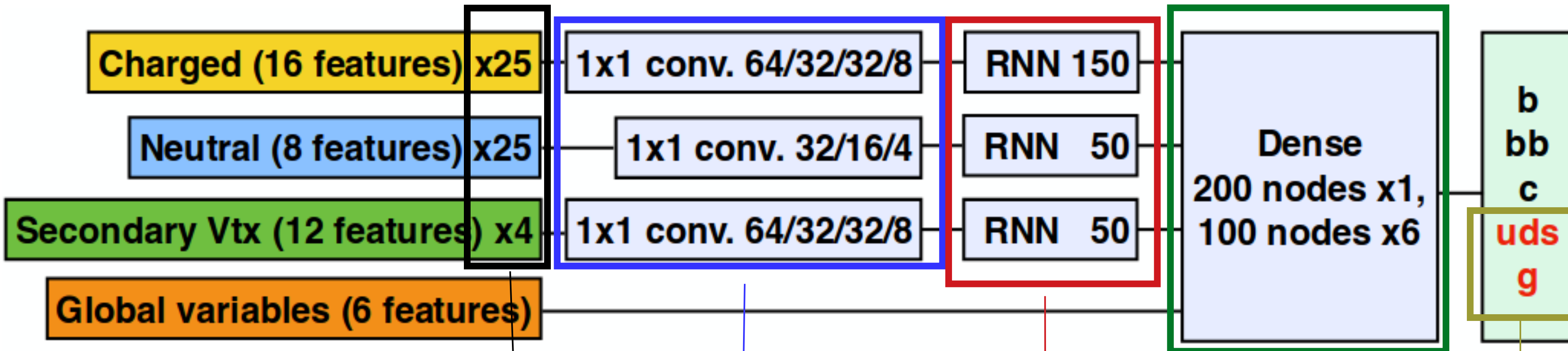
CMS-JME-18-002



- 3-layer FC network, 256 units/ layer
- SELU activation function
- output: 22 classes
  [corresponding to 22 bins from 30 to 250 GeV]

Particle-based "AK4" jet tagging: DeepJet

- A multiclass classifier for: b, bb, c, uds, gluons
- Highlights from the architecture:



Number of particles/SV

LSTM layers
Builds a summary of the information extracted in each set of features

New wrt DeepCSV

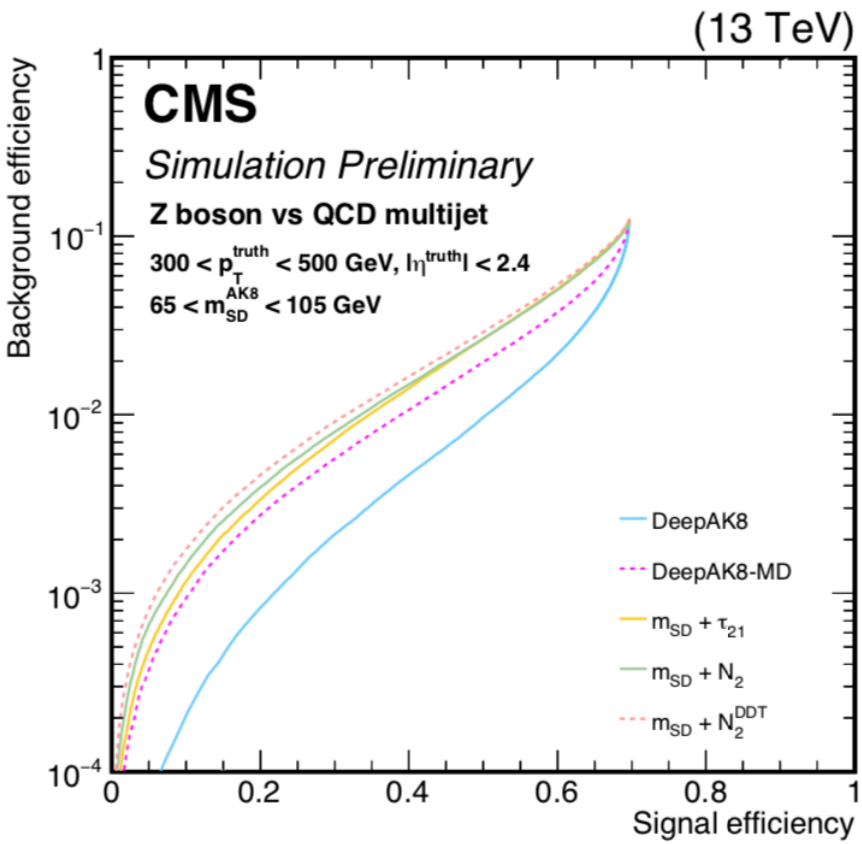
Feature extractor
convolution performed on each particle / SV [1x1]

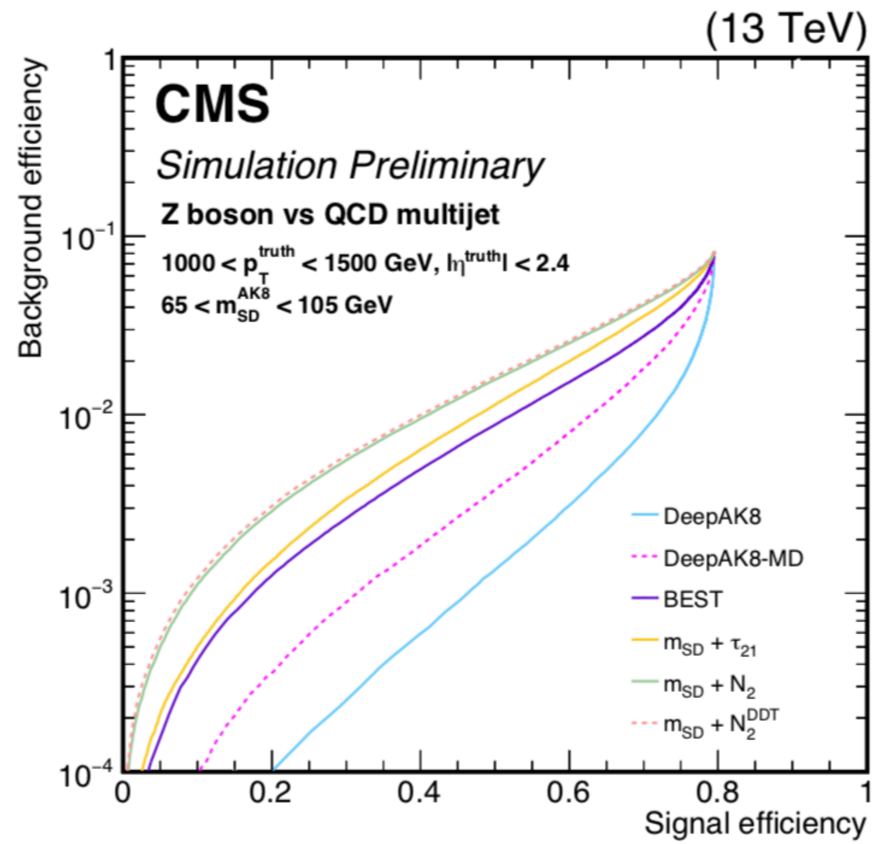
Correlations and classification

Meged-Jet tagger: Performance Z

- DeepAK8 [CMS-JME-18-002]:

CMS-JME-18-002

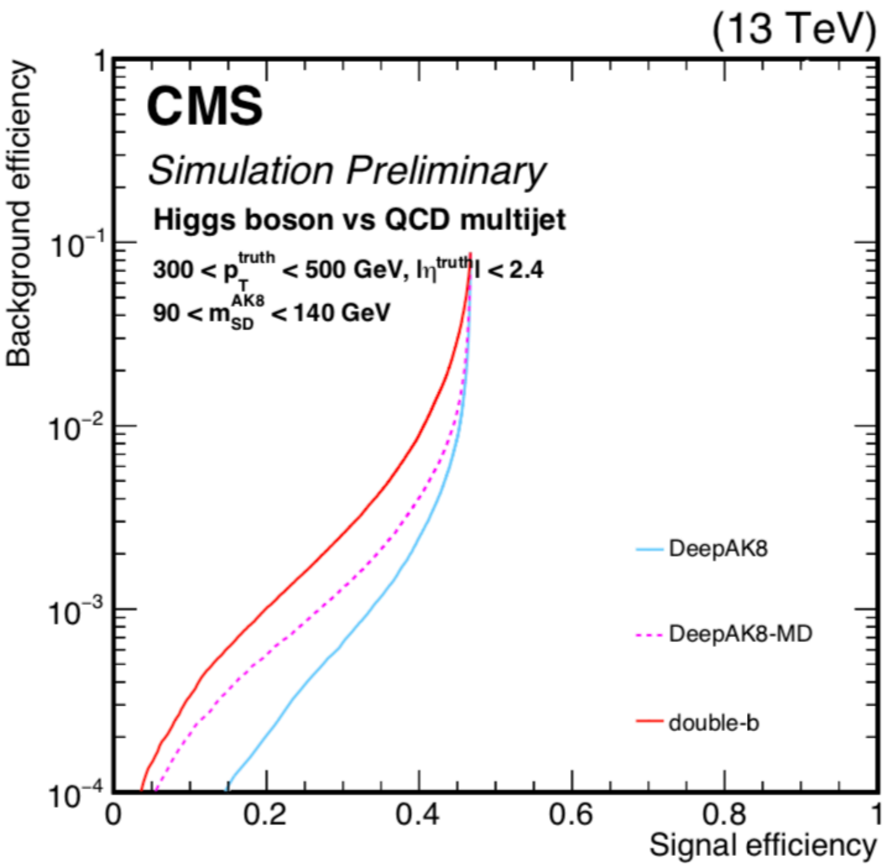


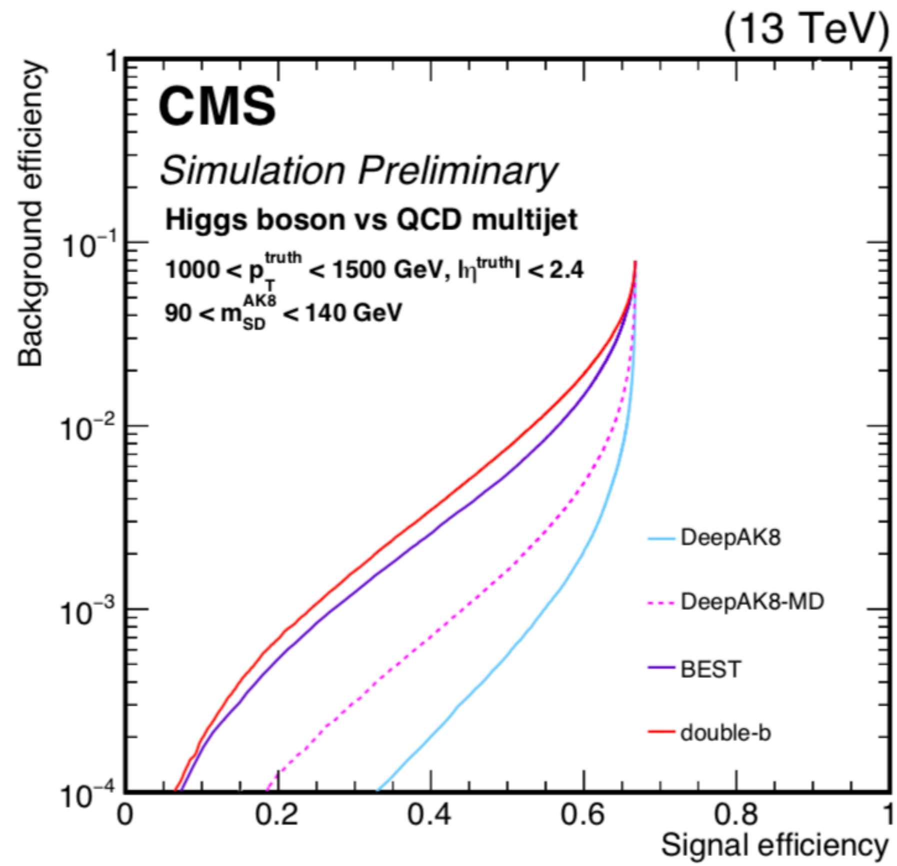


Merged-Jet tagger: Performance Higgs

- DeepAK8 [CMS-JME-18-002]:

CMS-JME-18-002





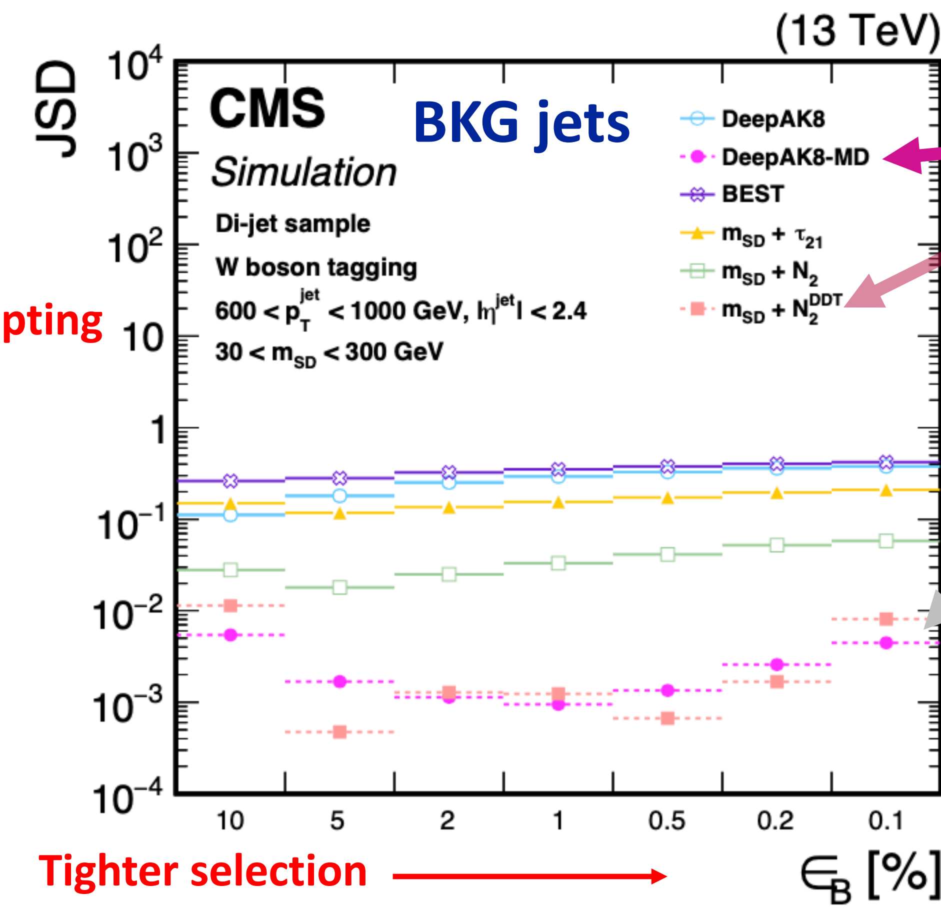
Mass (de-)correlation: Results (II)

- Quantify mass sculpting using the Jensen-Shannon divergence (JSD):

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

  - symmetrized version of Kullback-Leibler divergence (K $\mathrm{KLD}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$

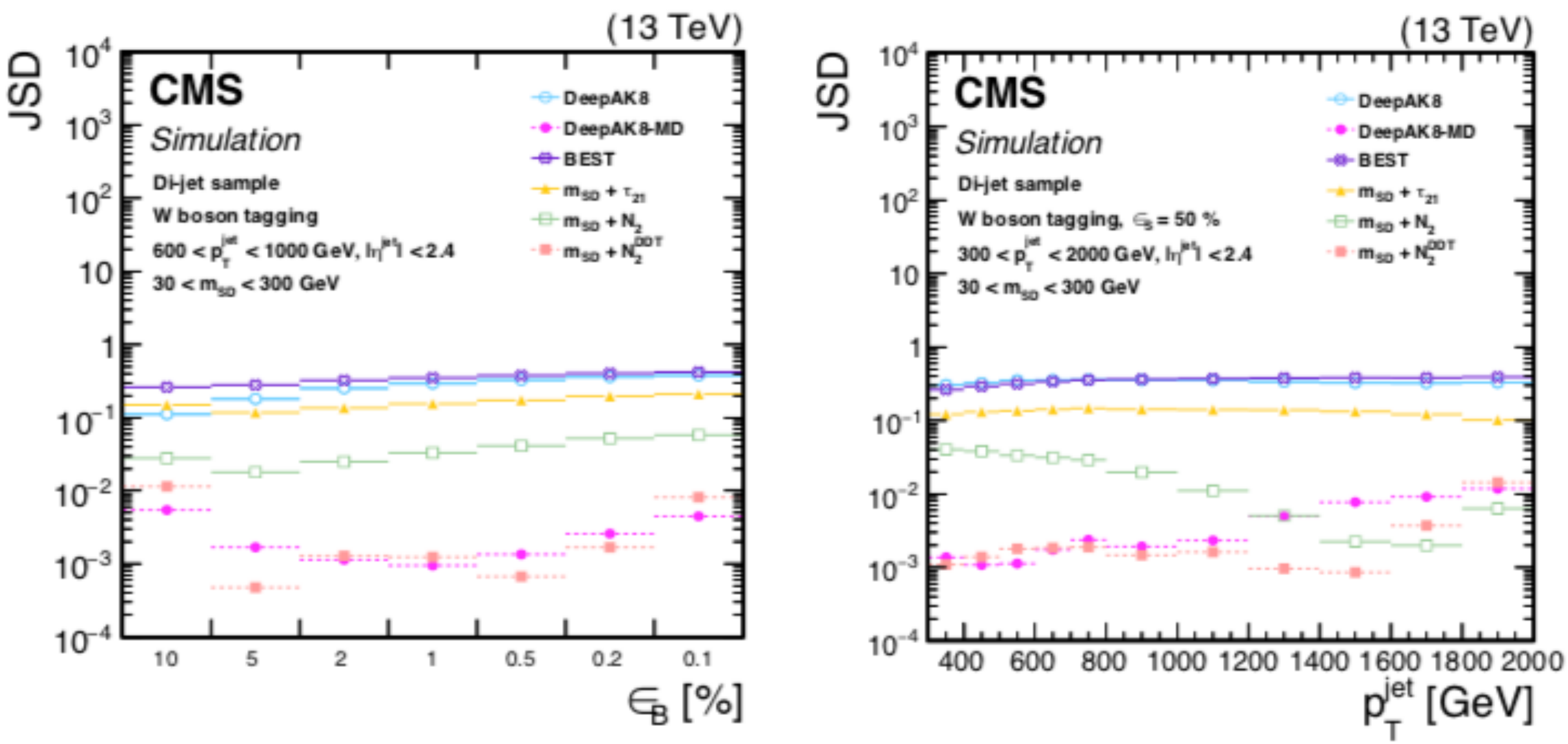
Less
mass sculpting



Mass decorrelated versions

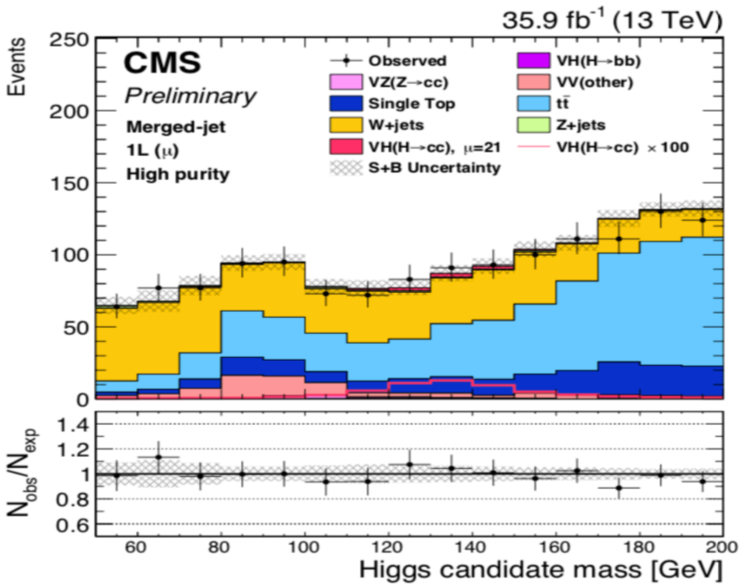
- Mass sculpting under control over a broad range of selections
- Small dependence on jet p_T and pileup

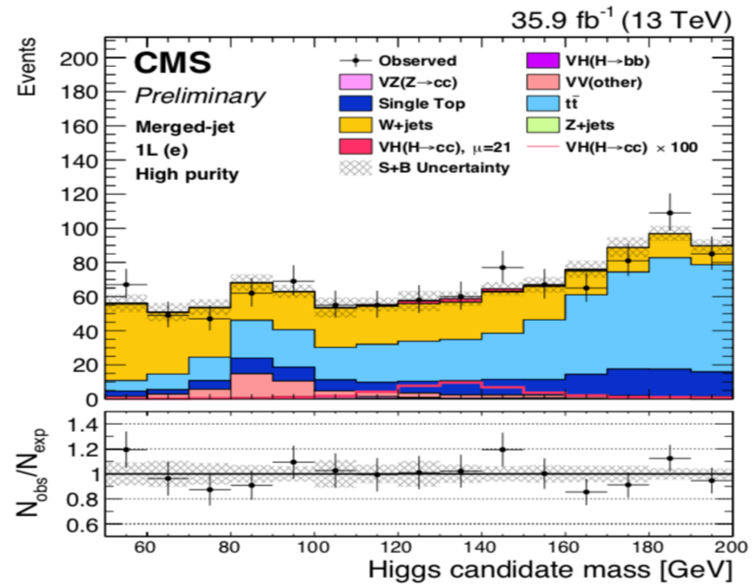
Mass (de-)correlation (II)

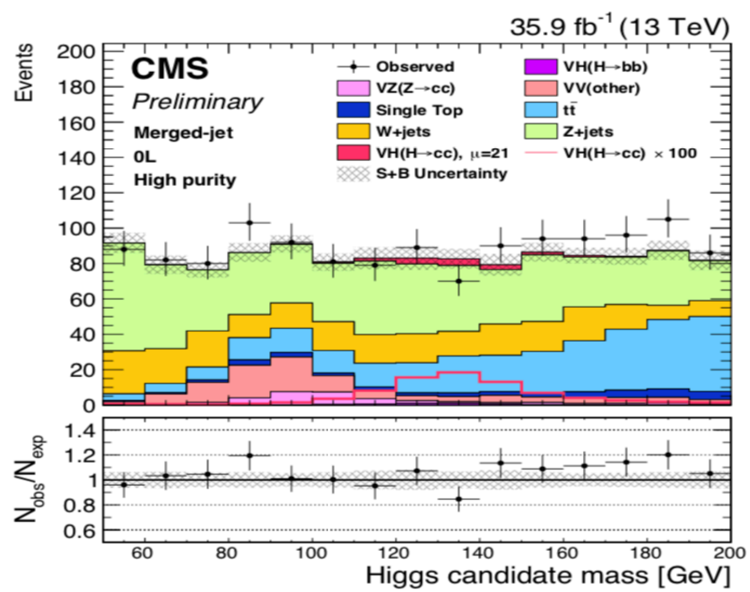


better

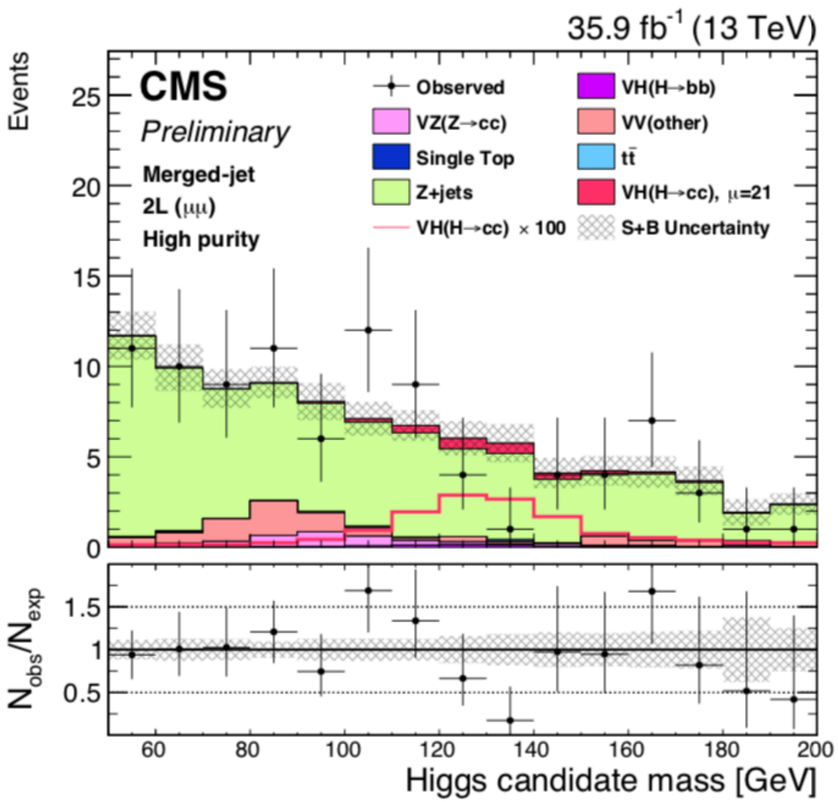
Merged-jet: postfit 0L, 1L

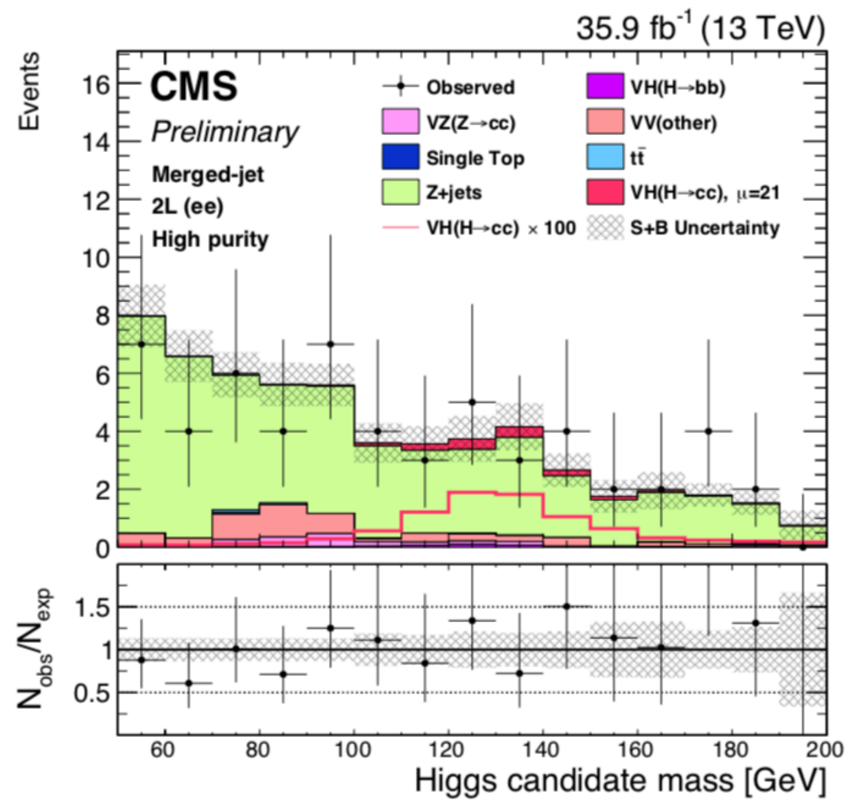






Merged-jet: postfit 2L





VH(→cc) results

- **First:** validate search by measuring the VZ(→cc) process
  - Same procedure as VH(→cc) but extract the VZ(→cc) signal

Topology	Significance obs (exp)	$\mu_{VZ(\to cc)}$
Resolved-jet	1.5 (1.2)	$1.35^{+0.94}_{-0.95}$
Merged-jet	0.9 (1.3)	$0.69^{+0.89}_{-0.75}$

Results consistent with SM expectation within uncertainties

- **Next:** VH(→cc) results in each topology:

95% C.L. exclusion limit on $\mu_{VH(\to cc)}$

	Resolved-jet (inclusive)				Merged-jet (inclusive)			
	0L	1L	2L	All channels	0L	1L	2L	All channels
expected UL	84	79	59	38	81	88	90	49
observed UL	66	120	116	75	74	120	76	71

$\mu < 75$ obs. ($38 \left(^{+16}_{-11}\right) \left[^{+35}_{-18}\right]$ exp. (1σ) $[2\sigma]$)

$\mu < 71$ obs. ($49 \left(^{+24}_{-15}\right) \left[^{+59}_{-24}\right]$ exp. (1σ) $[2\sigma]$)

Best fit signal strength

Topology	$\mu_{VH(\to cc)}$
Resolved-jet	41^{+20}_{-20}
Merged-jet	21^{+26}_{-24}

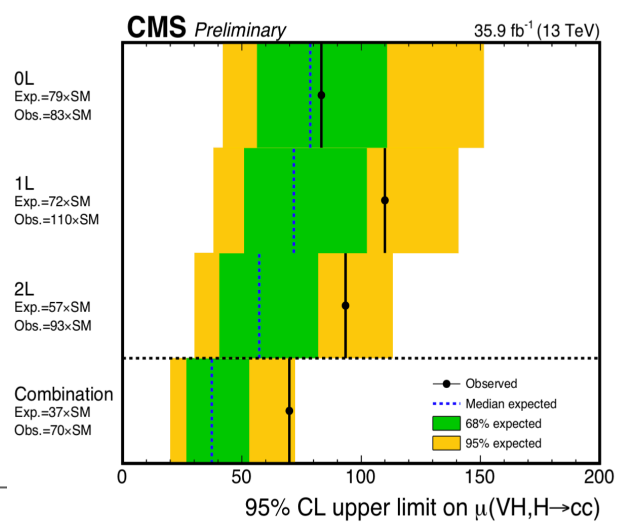
Combination

- Combination: **resolved-jet:** $p_T(V) < 300$ GeV / **merged-jet:** $p_T(V) > 300$ GeV
  - Systematics: correlated, but: c/cc-tagging efficiency & PDF, μ_R, μ_F for V+jets

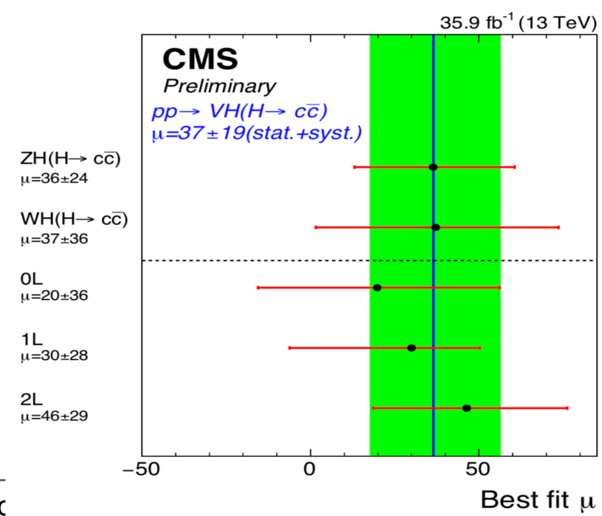
VH(→cc) results with 35.9 fb^{-1} (2016):

95% CL exclusion limit						
	resolved-jet ($p_T(V) < 300$ GeV)	merged-jet ($p_T(V) \geq 300$ GeV)	combination 0L	1L	2L	All channels
expected	45^{+18}_{-13}	73^{+34}_{-22}	79^{+32}_{-22}	72^{+31}_{-21}	57^{+25}_{-17}	37^{+16}_{-11} $\left(^{+35}_{-17}\ 2\sigma\right)$
observed	86	75	83	110	93	70

95% C.L. exclusion limit on $\mu_{VH(cc)}$



Best fit signal strength

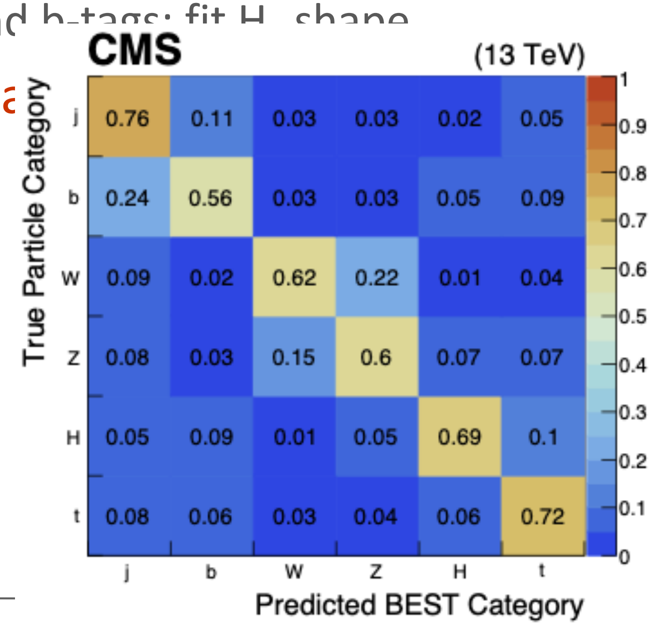


Search for Vector Like Quarks (VLQ) (II)

- VLQ transform under same group as SM EWK bosons
  - leads to a variety of decay modes; mainly to 3rd-Gen quarks
    - relative fractions depends on the model

Jet tagging algorithms with multitagging capabilities are developed exactly for this

- Analysis targets the all-hadronic channel with two approaches:
  - (A) Traditional cut-based approach; tailored to a single decay mode [T→bW]
    - Categorize events based on W-tags [SD+τ_{21}] and b-tags; fit H_T shape
  - (B) An approach using the **new BEST** multicla
    - Count object multiplicities according to BEST classification; form all possible combinations
    - Fit H_T shape for signal extraction
    - BKG estimated from data by correcting

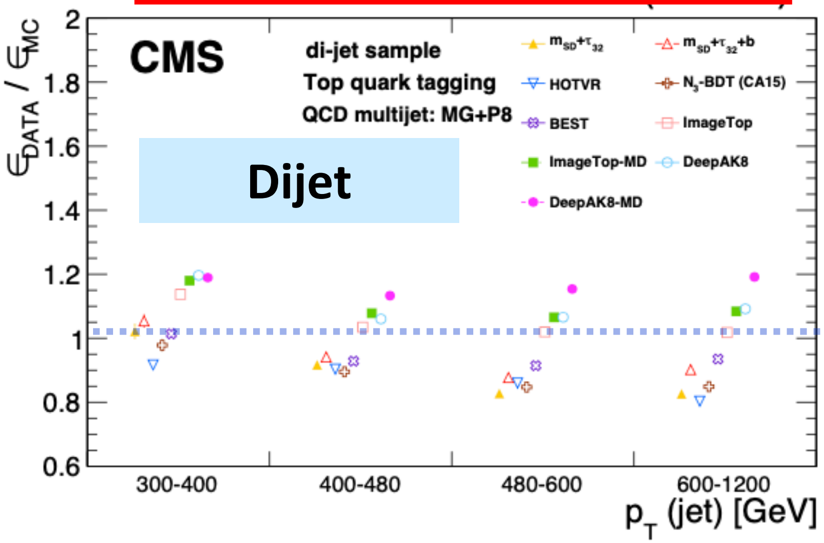


A more comprehensive approach
Stronger limits over a wider variety of models

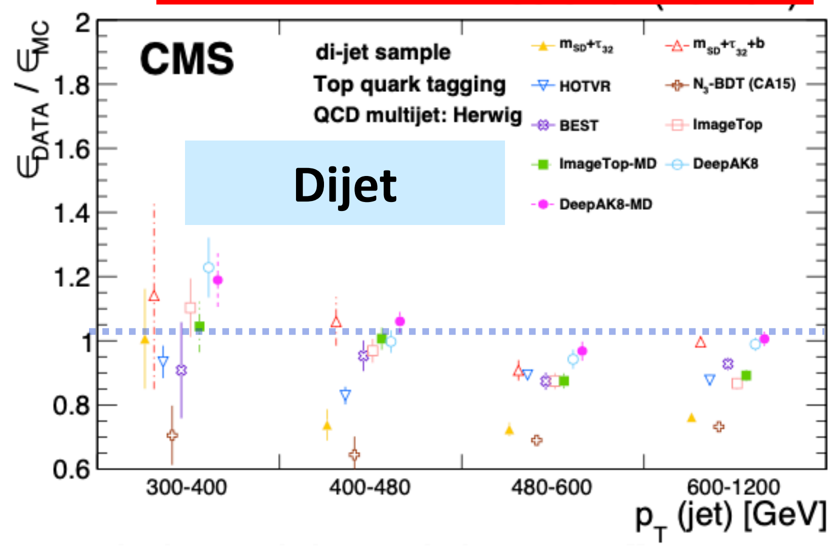
Corrections to simulation: Top mistagging

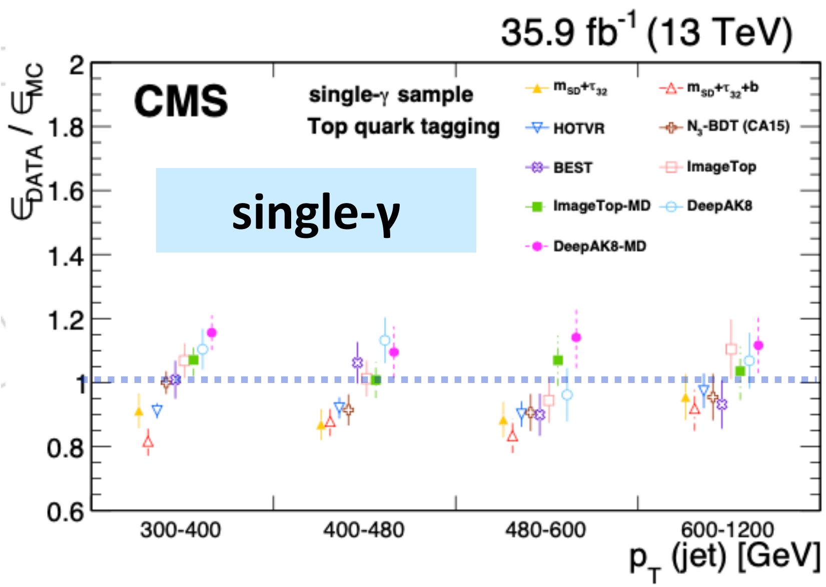
New !

QCD MC: MG+Pythia



QCD MC: Herwig





→ **Dijet: SF range 0.8 – 1.2**
generally larger for taggers using low level features

→ **Dijet: dependence on MC sample**
differences with QCD Herwig [to some extend due to differences in Q/G modeling]

→ **Single+γ: SF much more consistent with 1**
quark dominated sample

[Similar conclusions for W-tagging]