Jet substructure tagging and pileup mitigation

ML4Jets 2020

Alejandro Gómez Espinosa[1]
On behalf of the CMS Collaboration

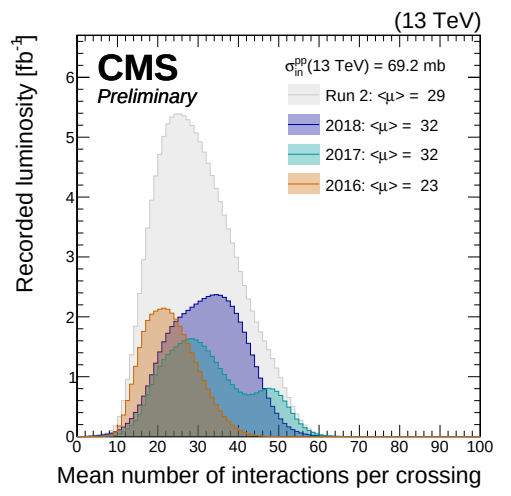
January 17, 2020

[1] ETH-Zurich

Contents

1. Pileup mitigation - JME-018-001
   - Techniques used in CMS
   - Efficiency and purity
   - MET resolution
2. Jet substructure - JME-018-002
   - Summary of algorithms studied
   - DeepAK8
   - BEST
   - ImageTop
   - Performance in simulation
3. Summary

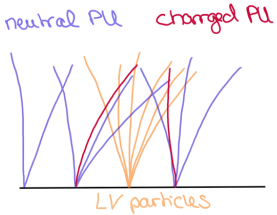
Pileup mitigation - JME-018-001



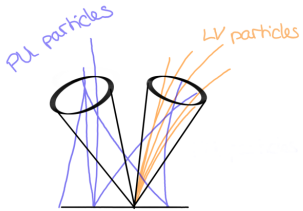
Pileup is a big challenge in LHC physics. In Run II we got approximately 29 interactions per event, while in Run III we expect up to 50.

PU mitigation techniques in CMS

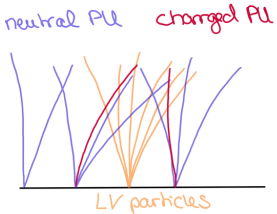
Charged Hadron Subtraction (CHS)



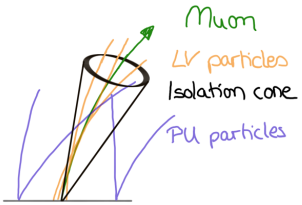
Pileup Jet ID



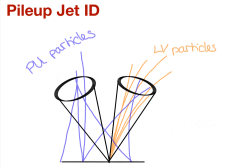
PileUp Per Particle Identification (PUPPI)



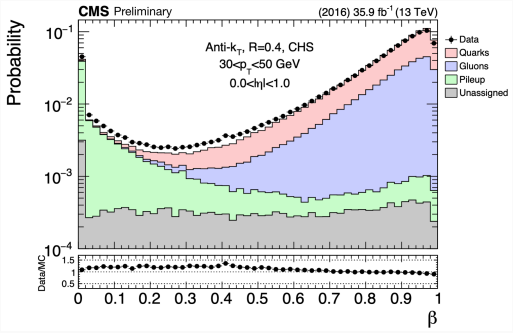
$\delta\beta-$corrected Isolation

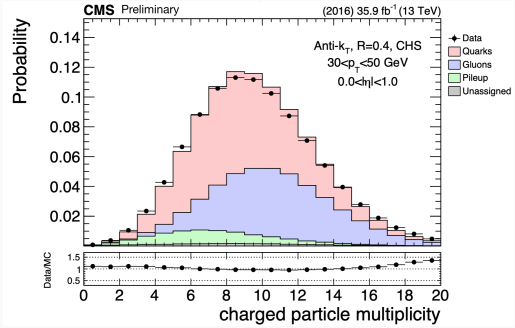


From Anna's talk.

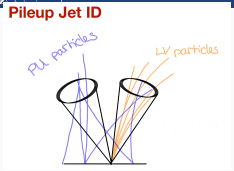


- *After CHS*, one can still find additional jets not originated from leading vertex.
- A boosted desicion tree (BDT) is trained, in bins of eta and pt, to identify those jets.
- 12 variables are used in the training (complete list in backup).

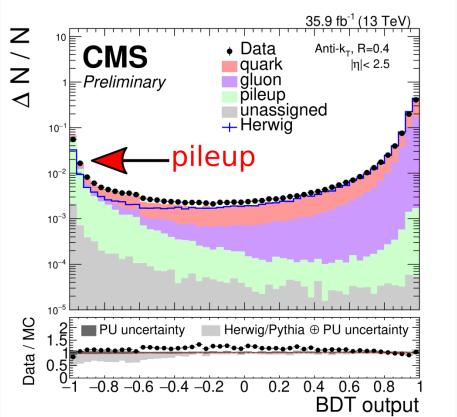


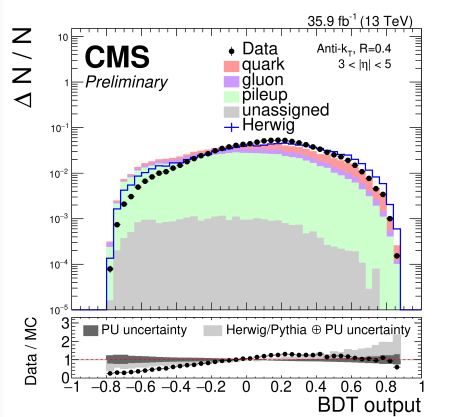


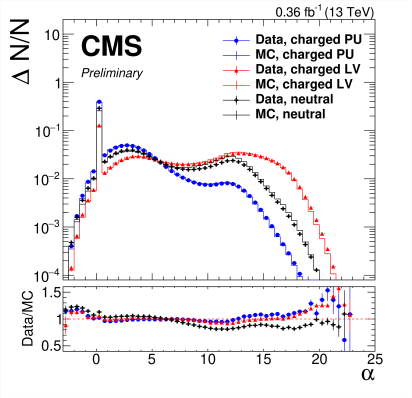
Pileup Jet ID



- *After CHS*, one can still find additional jets not originated from leading vertex.
- A boosted desicion tree (BDT) is trained, in bins of eta and pt, to identify those jets.
- BDT outputs shows good agreement in central region.



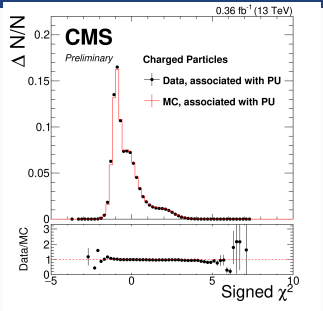


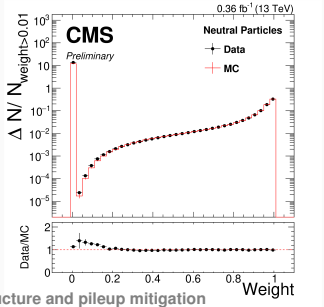


- PUPPI uses information related to local particle distribution, event PU properties and tracking information together to mitigate the effect of PU.
- The variable α,

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

  is used on an event-by-event basis to calculate a per-particle weight.





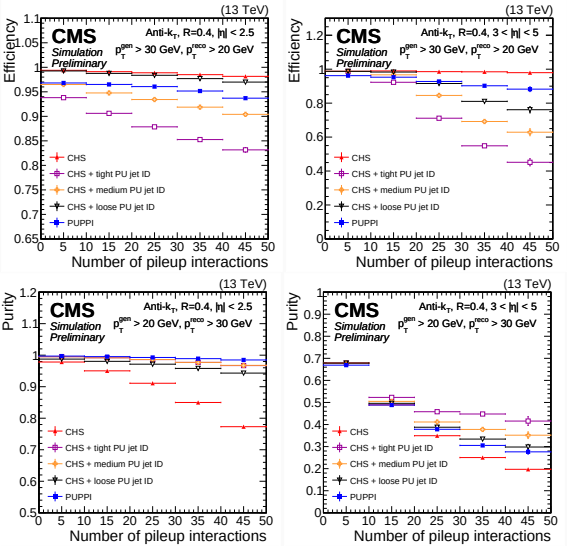
- PUPPI uses information related to local particle distribution, event PU properties and tracking information together to mitigate the effect of PU.
- The variable α,

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

  is used on an event-by-event basis to calculate a per-particle weight.

- Per particle, *signed* χ^2_i: $\text{signed } \chi^2_i = \dfrac{(\alpha_i - \bar{\alpha}_{\text{PU}})|\alpha_i - \bar{\alpha}_{\text{PU}}|}{\text{RMS}^2_{\text{PU}}}$, This variable and the weight are well modeled in data.

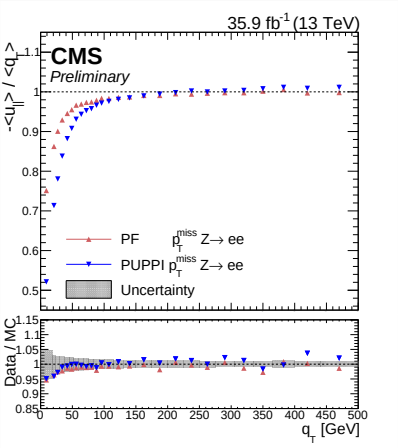
Efficiency and purity

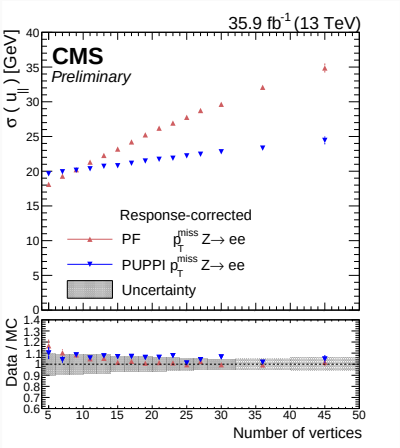


- We compute the efficiency and purity of the PU methods using simulations.
- Efficiency: fraction of generator-level jets with $p_T > 30$ GeV that match within $\Delta R < 0.4$ with a reconstruction-level jet with $p_T > 20$ GeV.
- Purity: fraction of reconstruction-level jets with $p_T > 30$ GeV that match within $\Delta R < 0.4$ with a generator-level jet with $p_T > 20$ GeV from the main interaction.
- PUPPI has an improved efficiency and purity performance overall in the central region.

MET resolution

- PUPPI also improves the performance of MET, compared to the previous method (PF MET).
- Hadronic response as a function of q_T (left) is stable, while the perpendicular resolution of the recoil against PU vertices (right) improves.

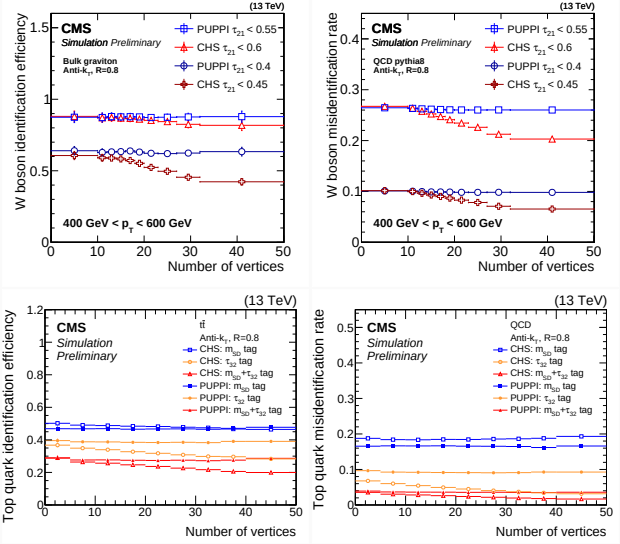




Jet substructure - JME-018-002

Tagging performance

- The performance of non-ML taggers is studied as a function of primary vertices.
- W tagging: CHS/PUPPI + τ_{21}, Top tagging: CHS/PUPPI + τ_{32}.
- In both cases, PUPPI shows a stable performance compared to taggers using CHS.



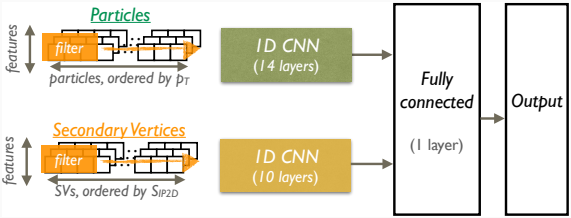
Summary of the algorithms

Algorithm	p_T (jet) [GeV]	t quark	W boson	Z boson	Higgs boson	decay modes
$m_{SD} + \tau_{32}$	400	✓				
$m_{SD} + \tau_{32}$ + b	400	✓				
$m_{SD} + \tau_{21}$	200		✓	✓		
HOTVR	200	✓				
N_3 − BDT (CA15)	200	✓				
$m_{SD} + N_2$	200		✓	✓	✓	
BEST	500	✓	✓	✓	✓	
ImageTop	600	✓				
DeepAK8	200	✓	✓	✓	✓	✓
Jet mass decorrelated algorithms						
$m_{SD} + N_2^{DDT}$	200		✓	✓	✓	
double–b	300			✓	✓	
ImageTop–MD	600	✓				
DeepAK8–MD	200	✓	✓	✓	✓	✓

Summary of the algorithms

Algorithm	p_T (jet) [GeV]	t quark	W boson	Z boson	Higgs boson	decay modes
$m_{SD} + \tau_{32}$	400	✓				
$m_{SD} + \tau_{32} + b$	400	✓				
$m_{SD} + \tau_{21}$	200		✓	✓		
HOTVR	200	✓				
$N_3 - \text{BDT (CA15)}$	200	✓				
$m_{SD} + N_2$	200		✓	✓	✓	
BEST	500	✓	✓	✓	✓	
ImageTop	600	✓				
DeepAK8	200	✓	✓	✓	✓	✓
Jet mass decorrelated algorithms						
$m_{SD} + N_2^{DDT}$	200		✓	✓	✓	
double–b	300			✓	✓	
ImageTop–MD	600	✓				
DeepAK8–MD	200	✓	✓	✓	✓	✓

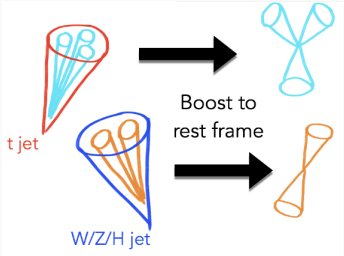
DeepAK8

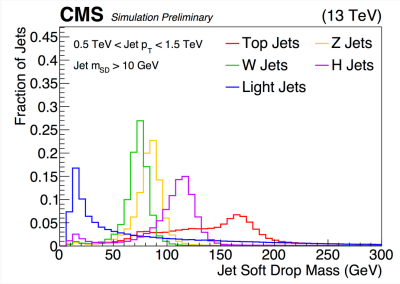


- DeepAK8 is a multiclass classifier for the identification of hadronically decaying particles, with five main categories, W/Z/H/t/other.
- PF candidates and secondary vertex information as inputs.
- It is a one-dimensional CNN based on the ResNet architecture.
- DeepAK8 was described in more detailed in Huilin's talk or in CMS-DP-2020/002

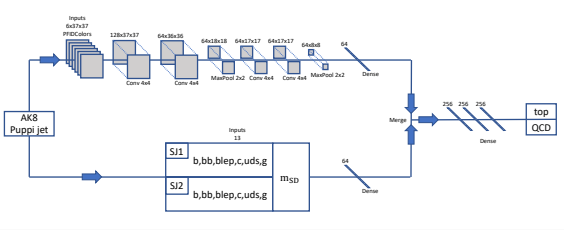
Boosted Event Shape Tagger (BEST)

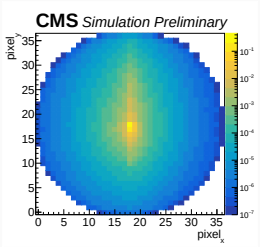
- BEST is a multi-classification algorithm designed to discriminate hadronic decays of high-pT top/W/Z/H bosons from jets arising from b/light quarks, and gluons. It is based on Phys. Rev. D 94, 094027.
- Main idea is transforming the entire set of jet constituents four times, each with a different boost vector.
- Architecture: fully-connected NN, consists of 3 hidden layers with 40 nodes in each layer using a ReLU activation function.
- 59 quantities are used to train a NN. (variables in backup)

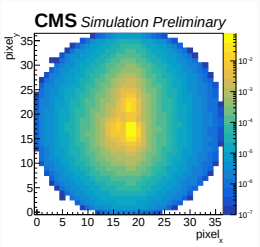




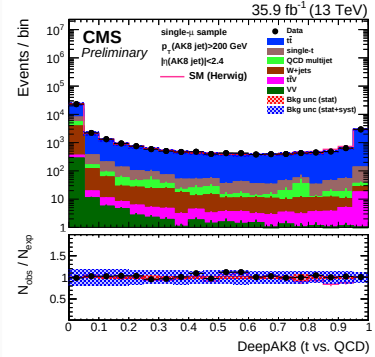
- ImageTop: uses standard image recognition techniques based on two dimensional CNN to discriminate boosted top jets from QCD jets.
- After preprocessing the jet to increase the collimation of the high pt particles, the image jet contains four *colors*: neutral pT, track pt, number of muons, and number of tracks.
- Architecture cartoon is shown on the right, as well as the average image for tops (left) and QCD jets (right).

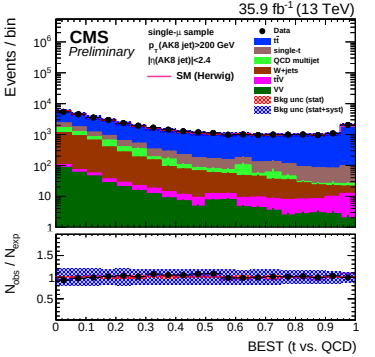


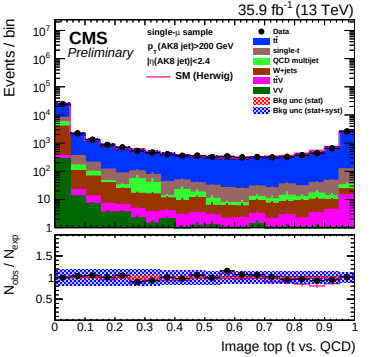


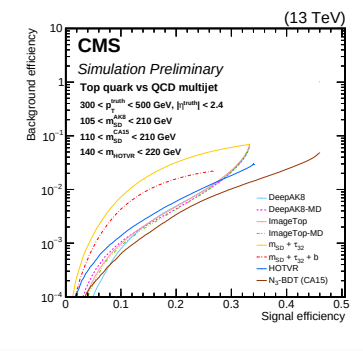


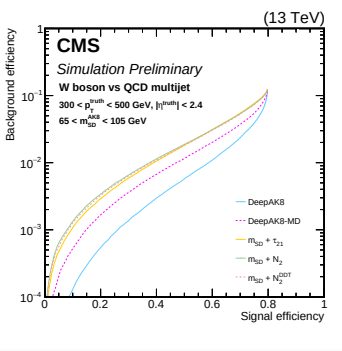
DeepAK8	BEST	ImageTop

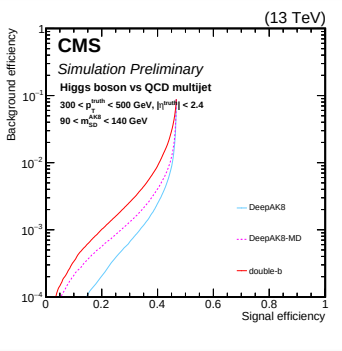






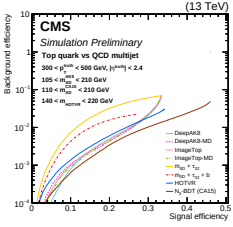


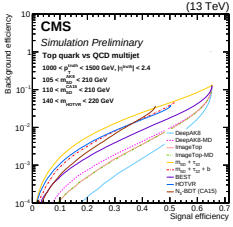


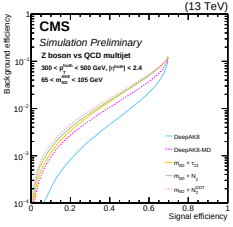


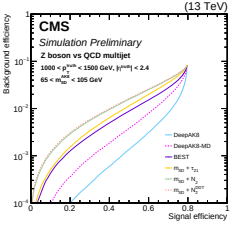
Efficiency: number of events passing the selection and algorithm divided by the total number of events generated in signal/background.

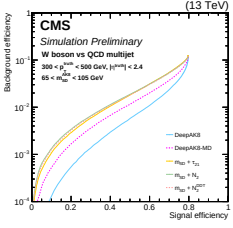
Performance in simulation

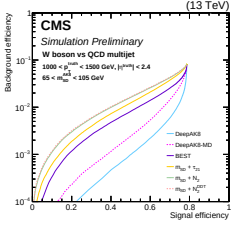


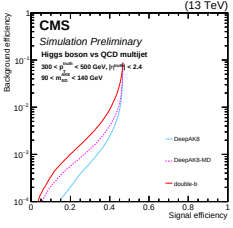


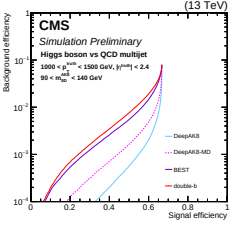




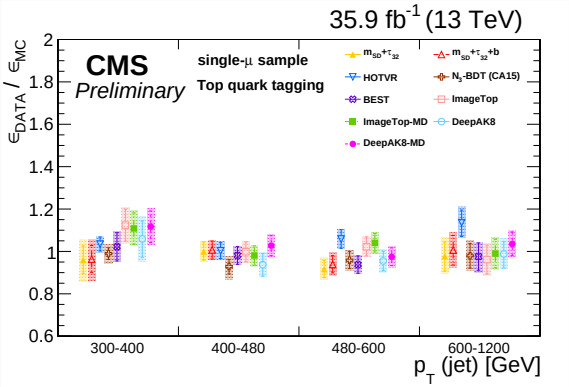


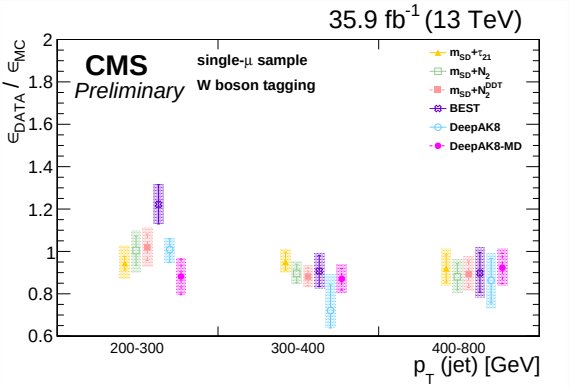






Efficiency: number of events passing the selection and algorithm divided by the total number of events generated in signal/background.





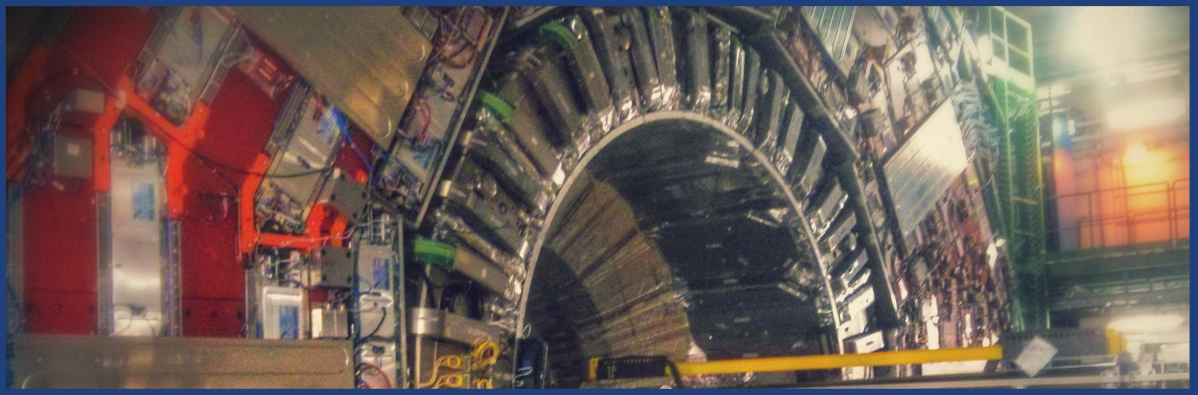
Scale factor: ratio of efficiency in data over simulations.

Summary

Summary

- Pileup is a big challenge encounter in the present conditions of the LHC. However, the PUPPI algorithm shows to be very robust against higher number of primary interactions.
- CMS will use PUPPI as the default PU mitigation algorithm in Run III.
- CMS studied a variety of ML-based algorithms for jet tagging. Many of these algorithms are already applied in public results.

Thank you for the attention...

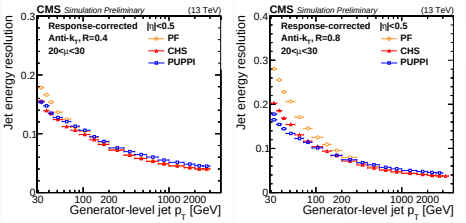


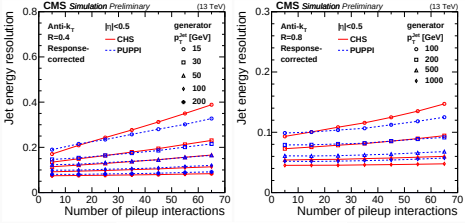
... questions?

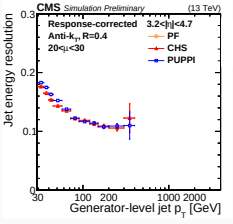
PU JetID

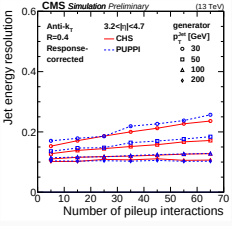
β	fraction of transverse momentum of charged particles associated to the primary vertex, defined as $\frac{\sum_{i \in \mathrm{LV}} p_{\mathrm{T}i}}{\sum_i p_{\mathrm{T}i}}$ where i iterates over all the PF particles in the jet
$n_{vertices}$	number of vertices in the event
$\langle \Delta R^2 \rangle$	p_{T}^2 average weighted by square distance of jet constituents from the jet axis : $\frac{\sum_i \Delta R^2 p_{\mathrm{T}i}^2}{\sum_i p_{\mathrm{T}i}^2}$
$f_{\mathrm{ring}X}$, X = 1, 2, 3, and 4	fraction of p_{T} of the constituents ($\sum p_{\mathrm{T}i} / p_{\mathrm{T}}^{\mathrm{jet}}$) 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
$p_{\mathrm{T}}^{\mathrm{lead}} / p_{\mathrm{T}}^{\mathrm{jet}}$	transverse momentum fraction carried by the leading PF candidate
$p_{\mathrm{T}}^{\mathrm{l.\,ch.}} / p_{\mathrm{T}}^{\mathrm{jet}}$	transverse momentum fraction carried by the leading charged PF candidate
$\lvert\vec{m}\rvert$	pull magnitude, defined as $\lvert(\sum_i p_{\mathrm{T}}^i \lvert r_i \rvert \vec{r}_i)\rvert / p_{\mathrm{T}}^{\mathrm{jet}}$ 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 $\sqrt{\sum_i p_{T_i}^2} / \sum_i p_{\mathrm{T}i}$

Jet energy resolution

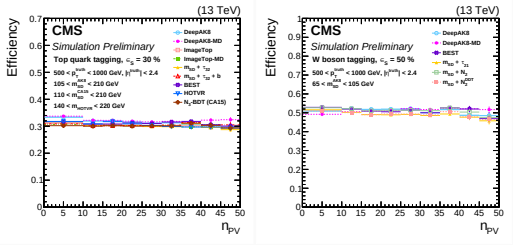
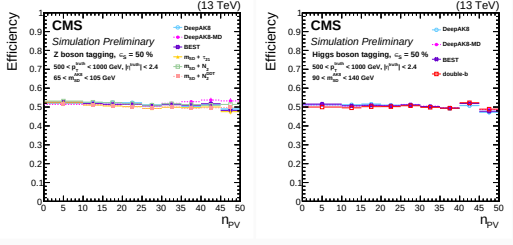
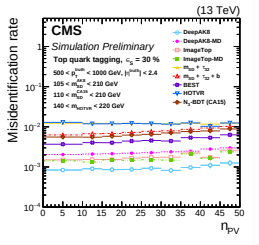
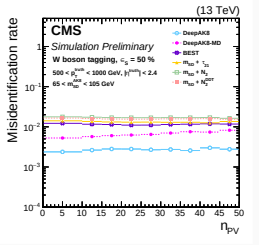
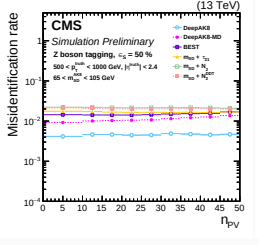
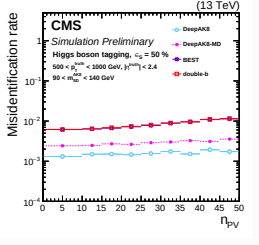








BEST

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)	

Performance in simulation