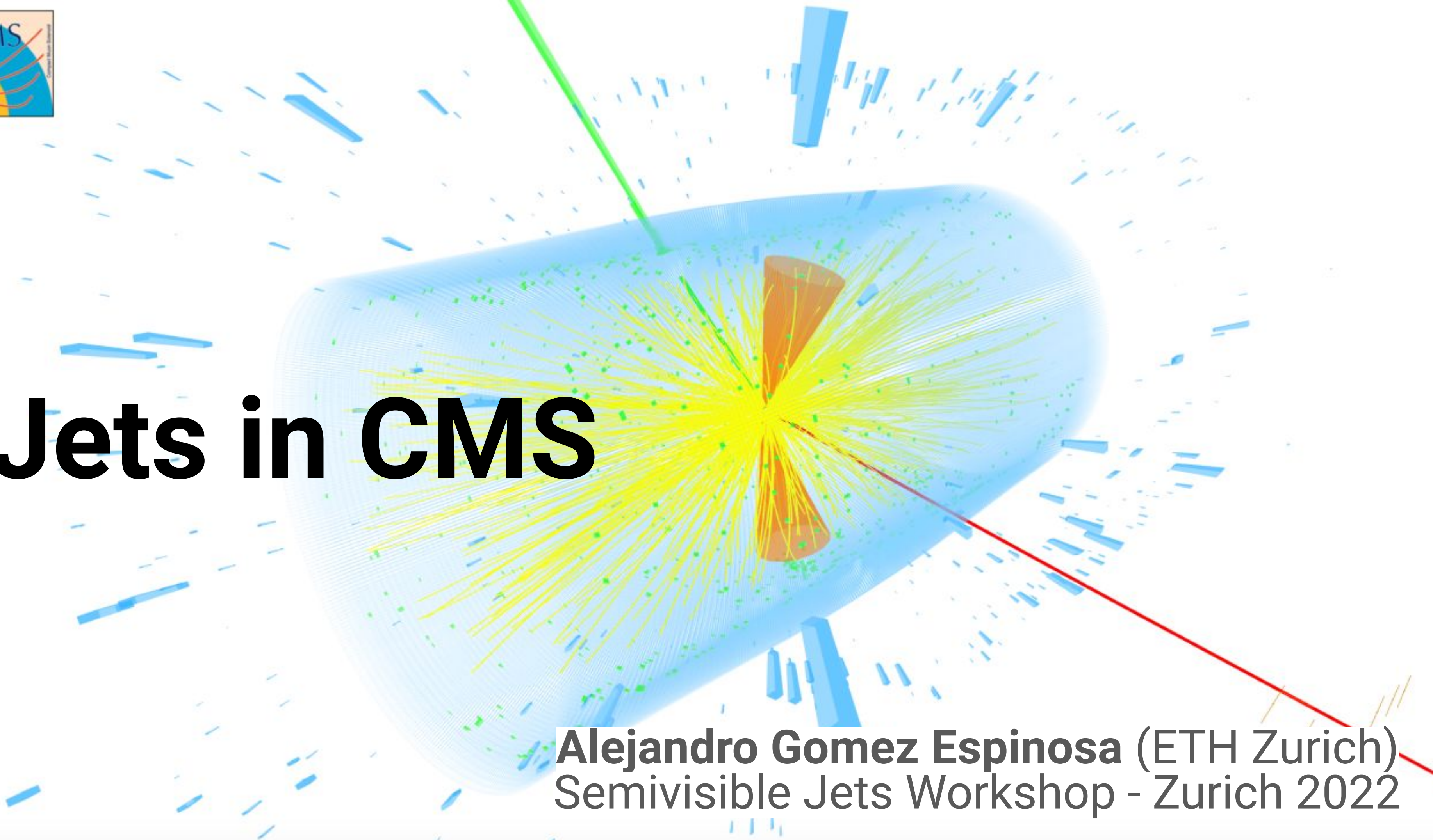
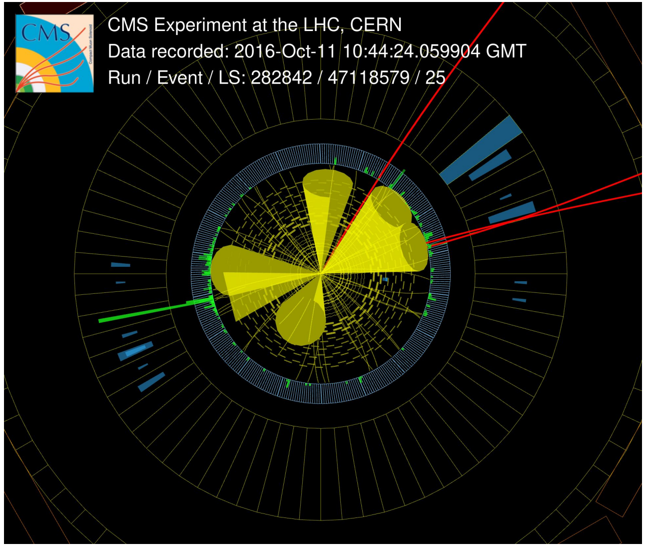


Jets in CMS

Alejandro Gomez Espinosa (ETH Zurich)

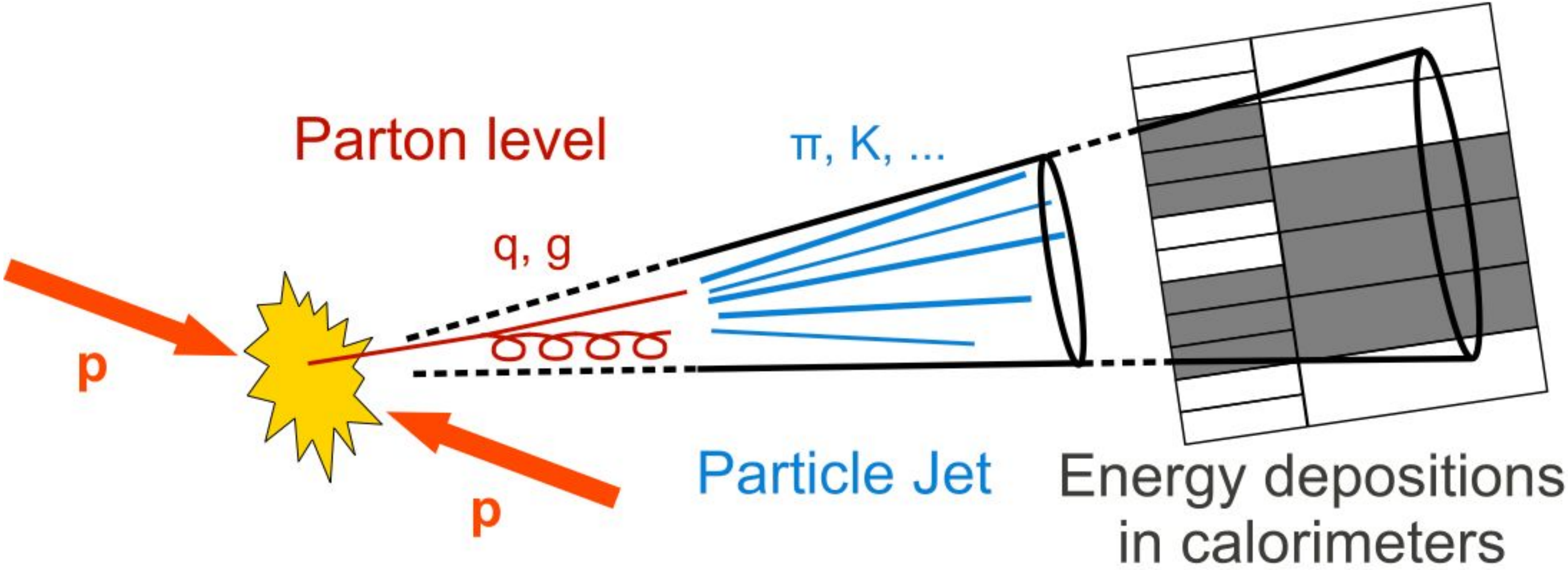
Semivisible Jets Workshop - Zurich 2022

Outline

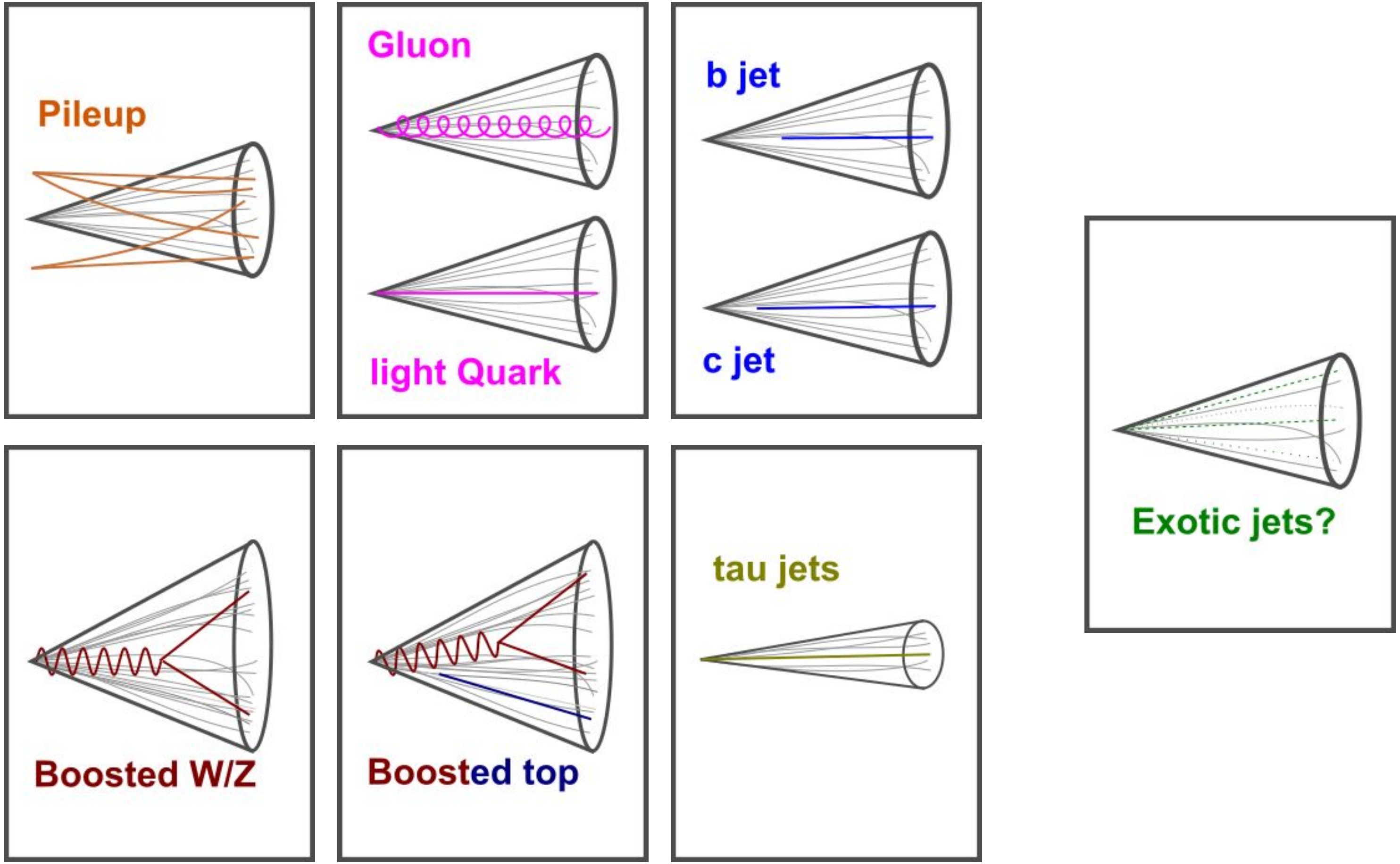


1. Intro
2. Pileup
3. Jets, jets, jets
4. Boosted jets

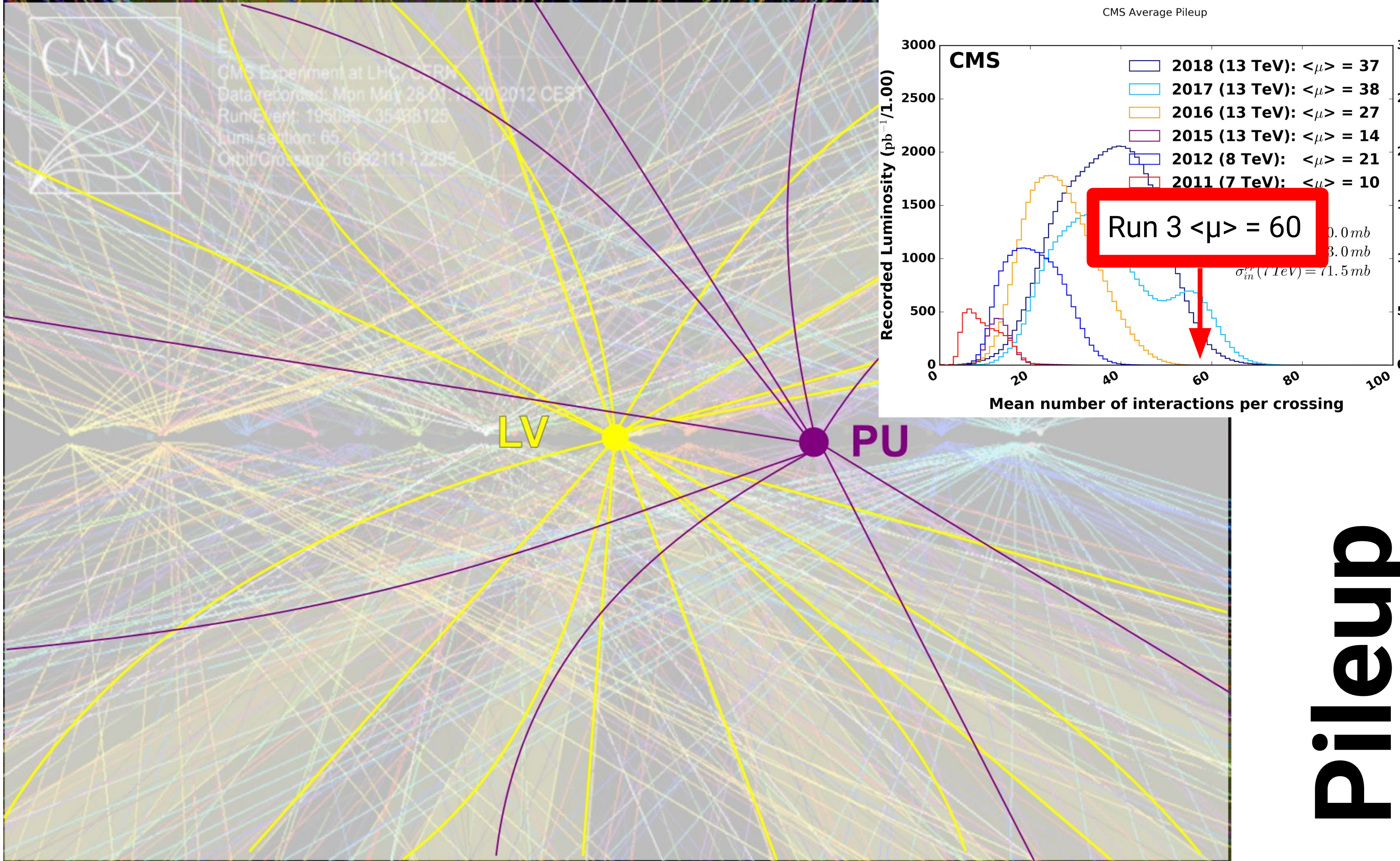
What is a jet?


Parton level
π, K, ...
q, g
p
p
Particle Jet
Energy depositions
in calorimeters

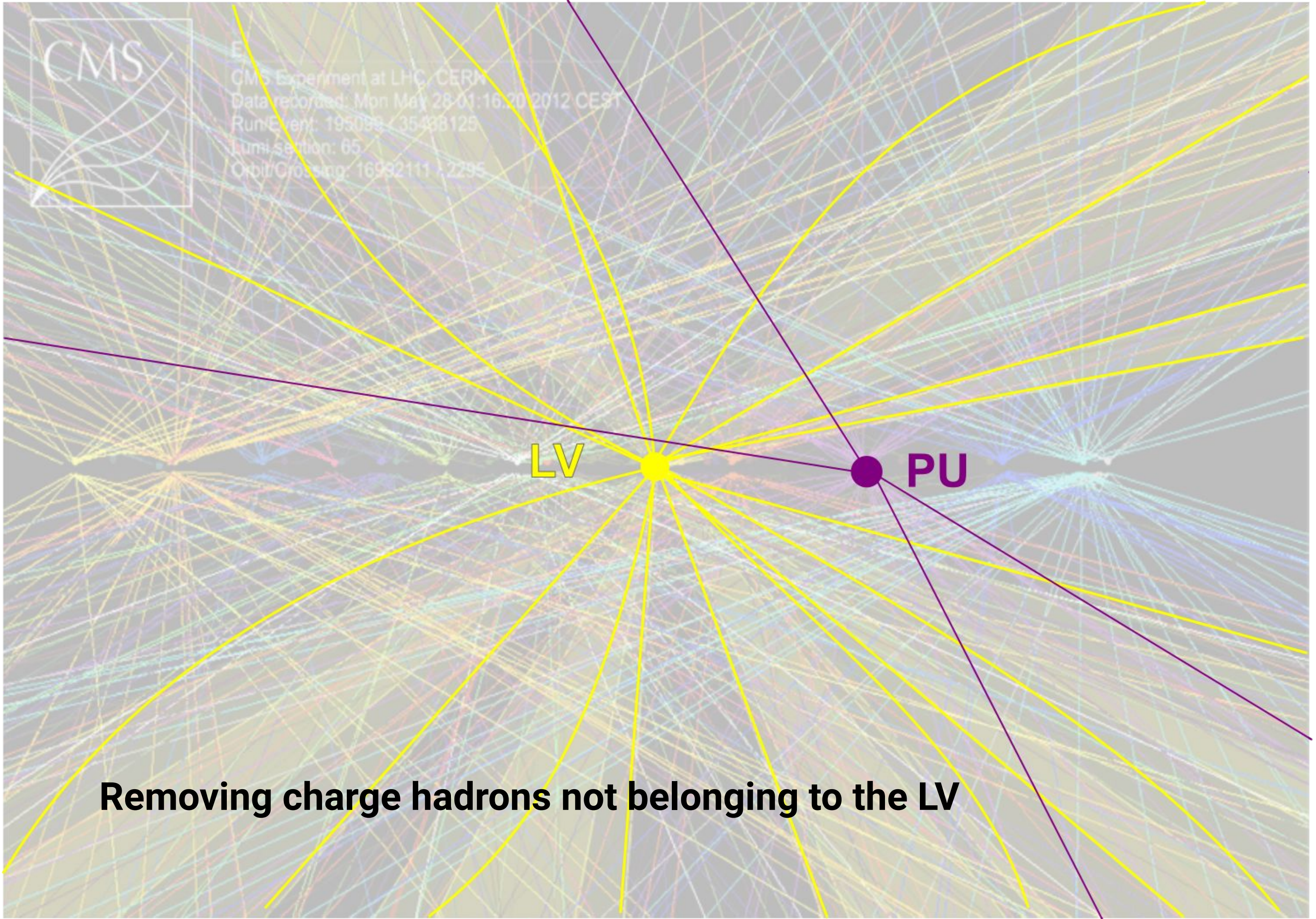
Different types of jets



Pileup

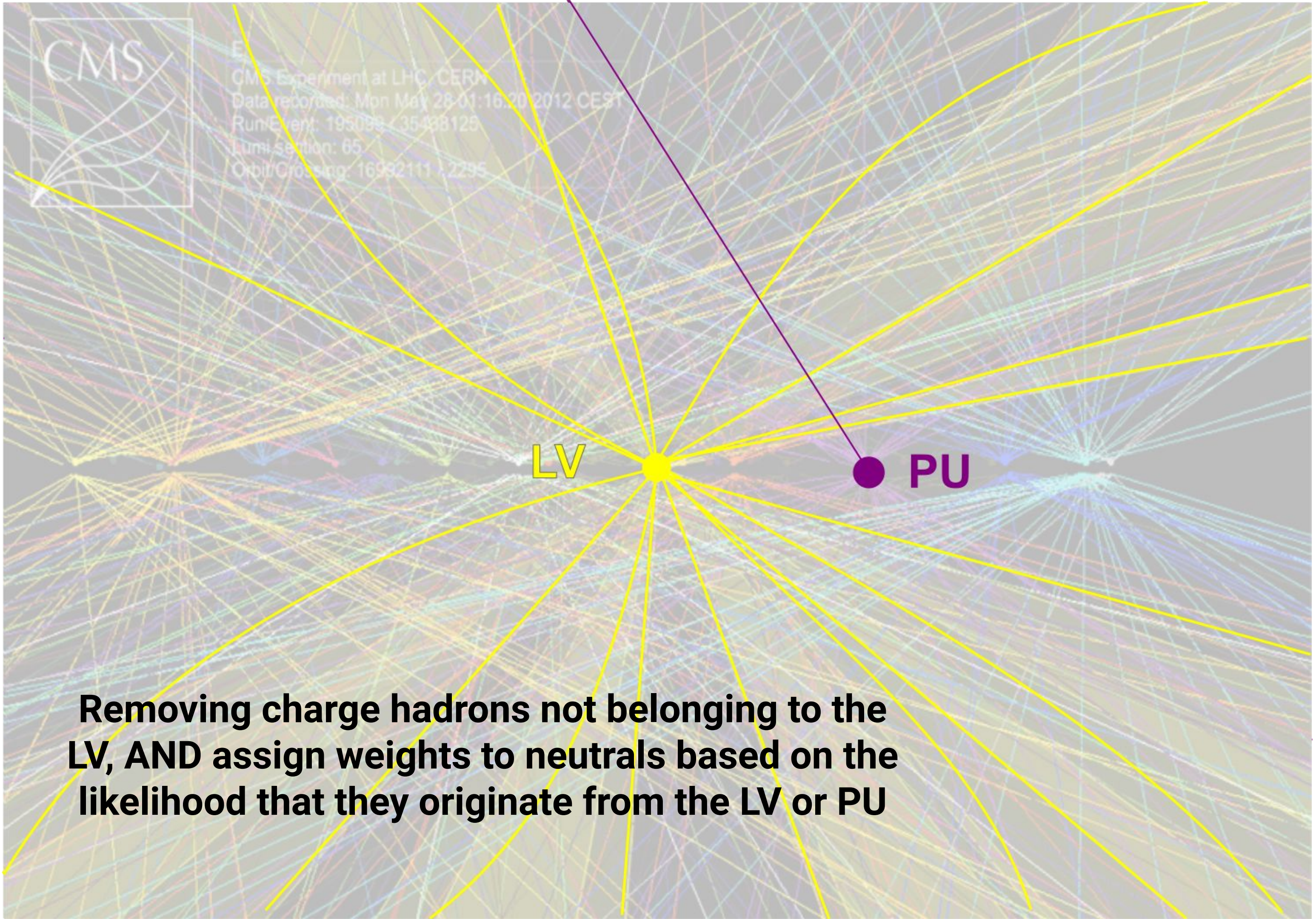

LV
PU
CMS Average Pileup
CMS
2018 (13 TeV): <μ> = 37
2017 (13 TeV): <μ> = 38
2016 (13 TeV): <μ> = 27
2015 (13 TeV): <μ> = 14
2012 (8 TeV): <μ> = 21
2011 (7 TeV): <μ> = 10
Run 3 <μ> = 60
Recorded Luminosity (pb⁻¹/1.00)
Mean number of interactions per crossing

Charge Hadron Subtraction (CHS)



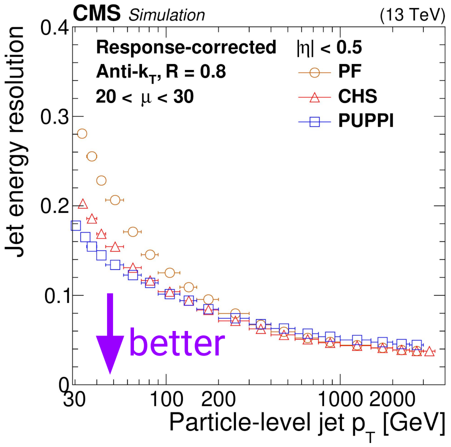
Removing charge hadrons not belonging to the LV

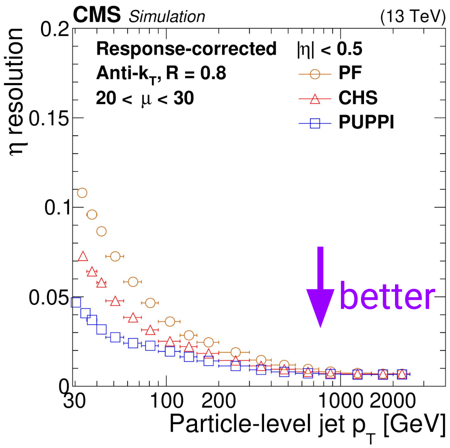
PUPPI

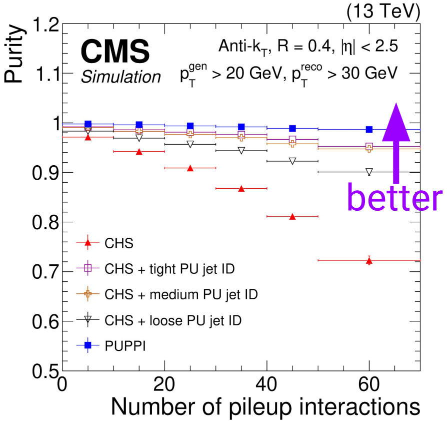


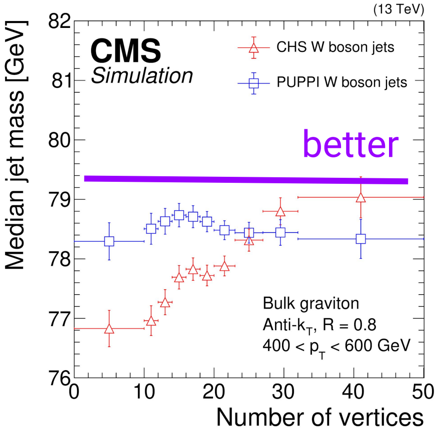
Removing charge hadrons not belonging to the LV, AND assign weights to neutrals based on the likelihood that they originate from the LV or PU

CMS strategy for pileup mitigation









arXiv: 2003.00502

Run I

CHS

Charge hadron subtraction was the pileup mitigation method by default, for both AK5 and AK7 jets.

Run II

CHS and PUPPI

CHS was the default algorithm for AK4 jets. For AK8 jets, we started with CHS but migrate quickly to PUPPI by the end of the Run.

Run III

PUPPI

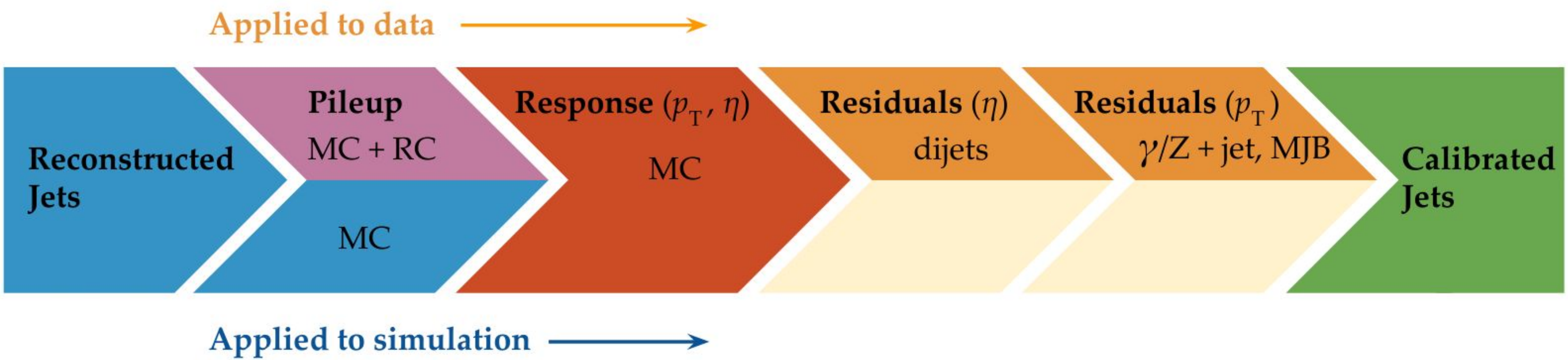
The default algorithm for AK4 and AK8 jets. It has prove to have a better overall performance.

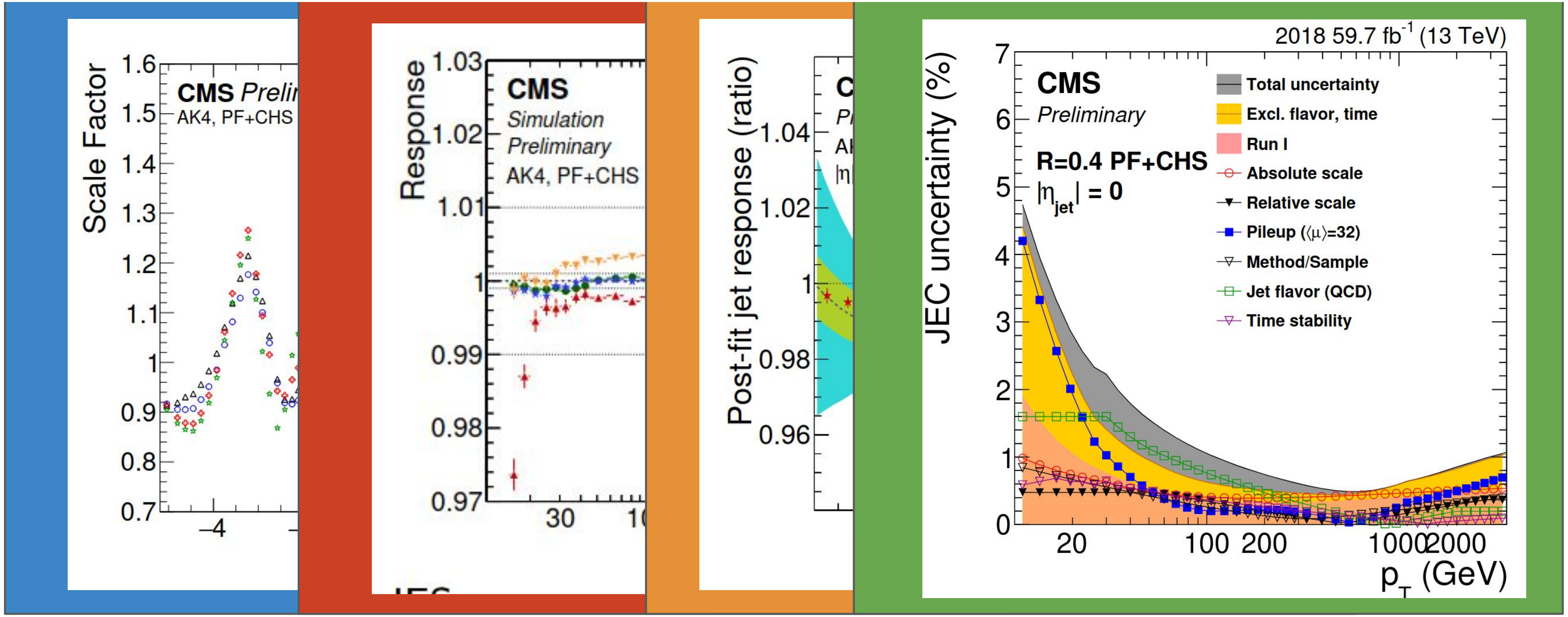
Run IV

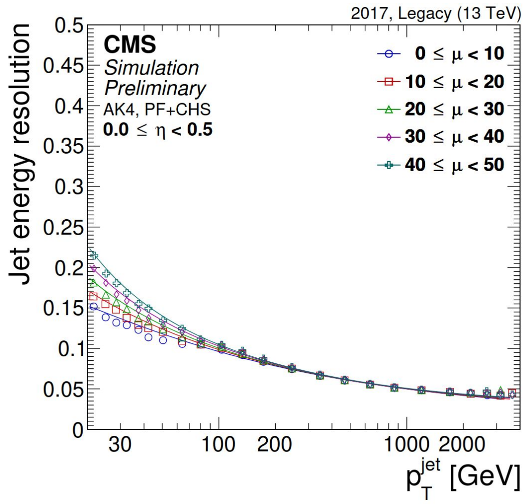
PUPPI ML?

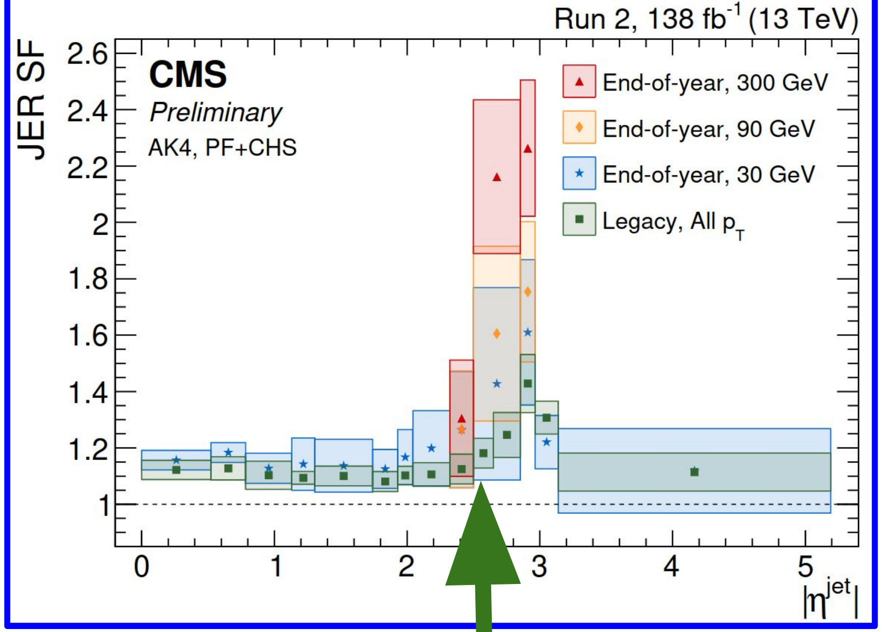
ML algorithms will take the stage.

Jet Energy Corrections

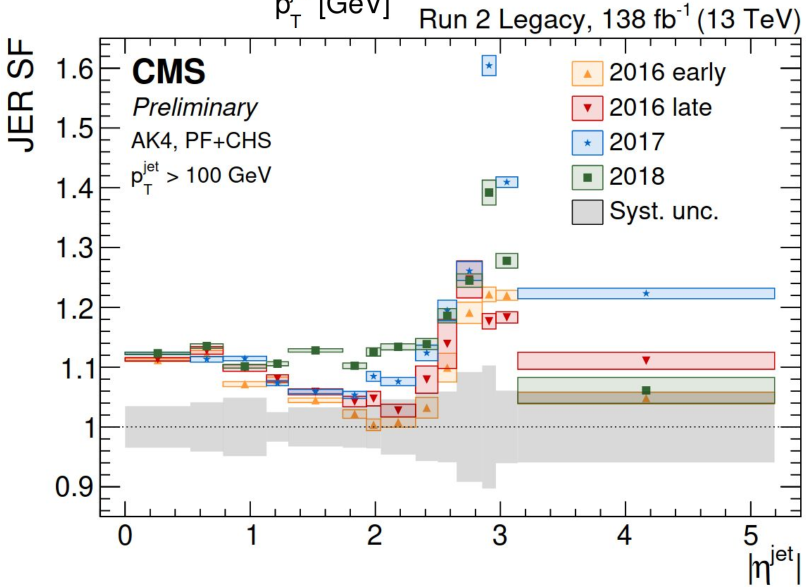




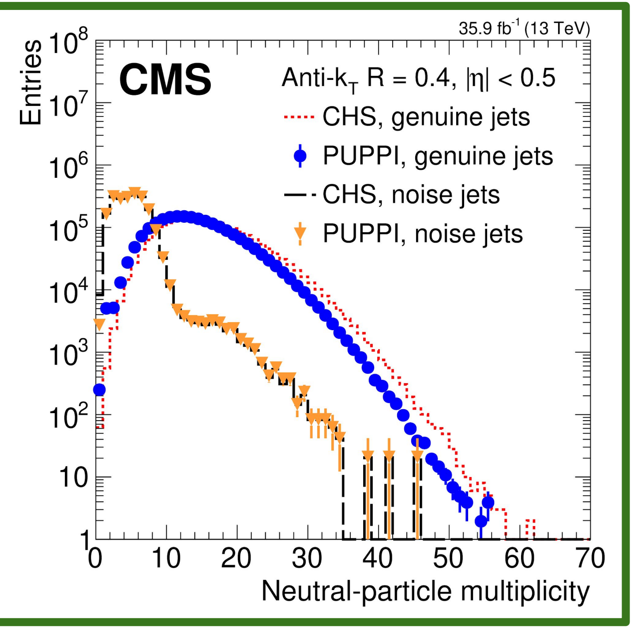


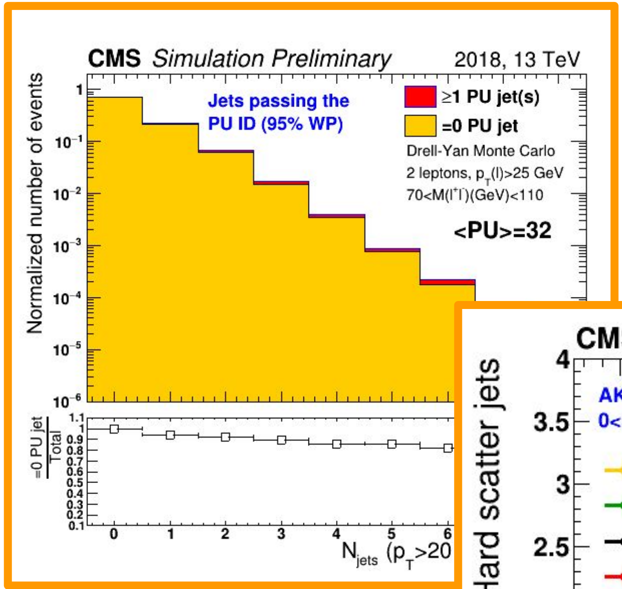


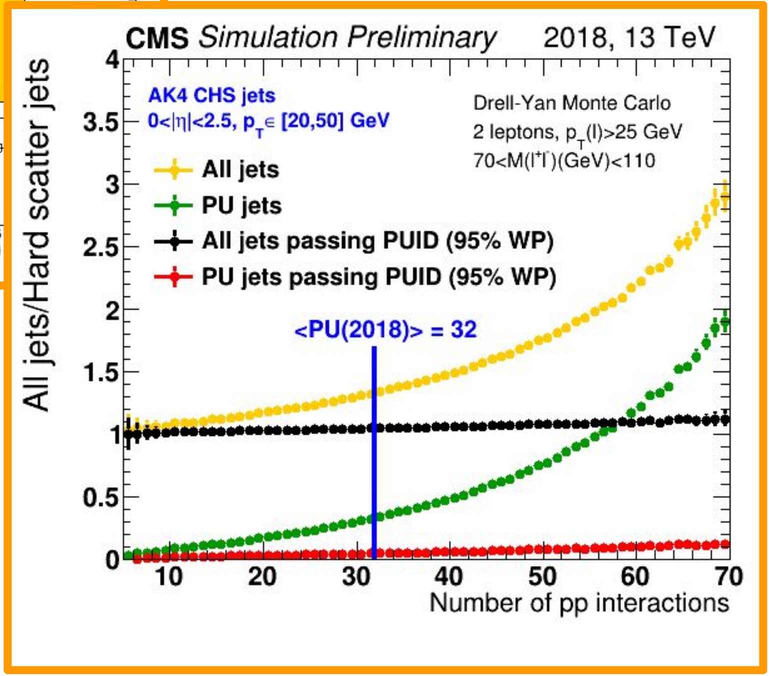
Improved with better reconstruction



Jet Energy Resolution



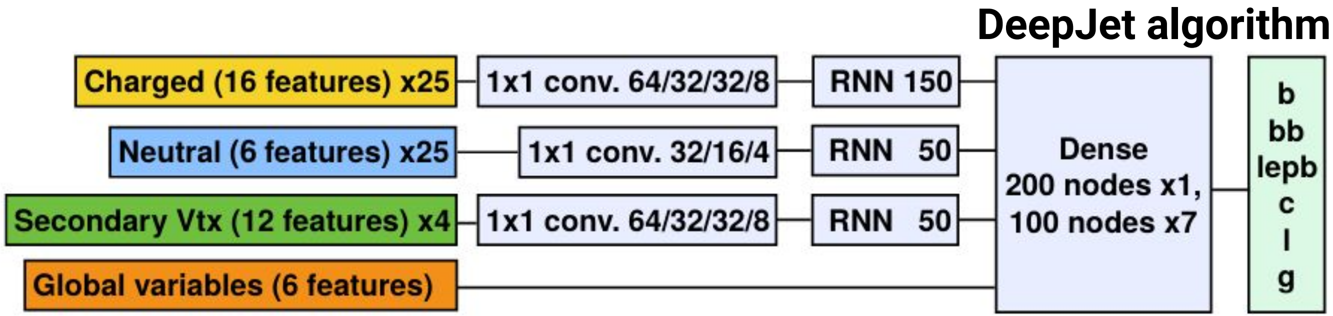


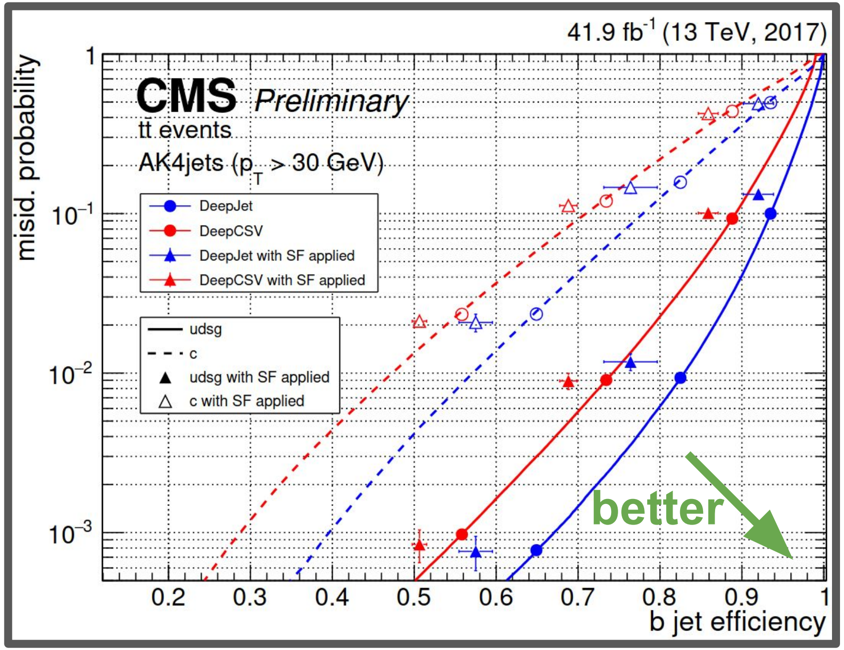


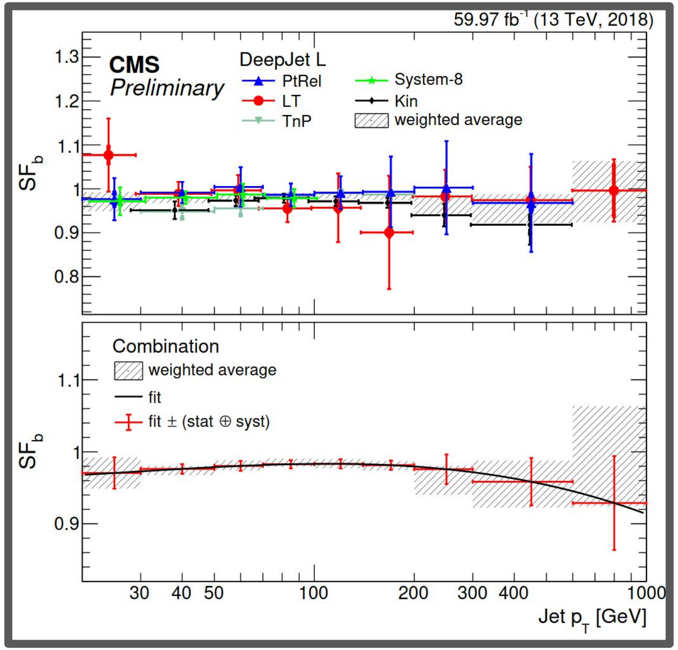
Jet Cleaning

Even though CHS/PUPPI do a good job in removing pileup, we use two extra steps to remove noise jets (**jetID**) and remaining PU jets (**PUjetID**).

b tagging

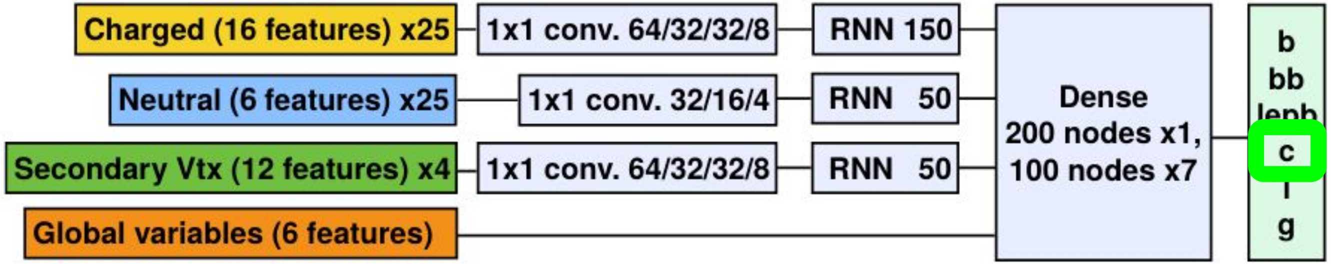

DeepJet algorithm
Charged (16 features) x25
1x1 conv. 64/32/32/8
RNN 150
Neutral (6 features) x25
1x1 conv. 32/16/4
RNN 50
Secondary Vtx (12 features) x4
1x1 conv. 64/32/32/8
RNN 50
Global variables (6 features)
Dense
200 nodes x1,
100 nodes x7
b
bb
lepb
c
l
g

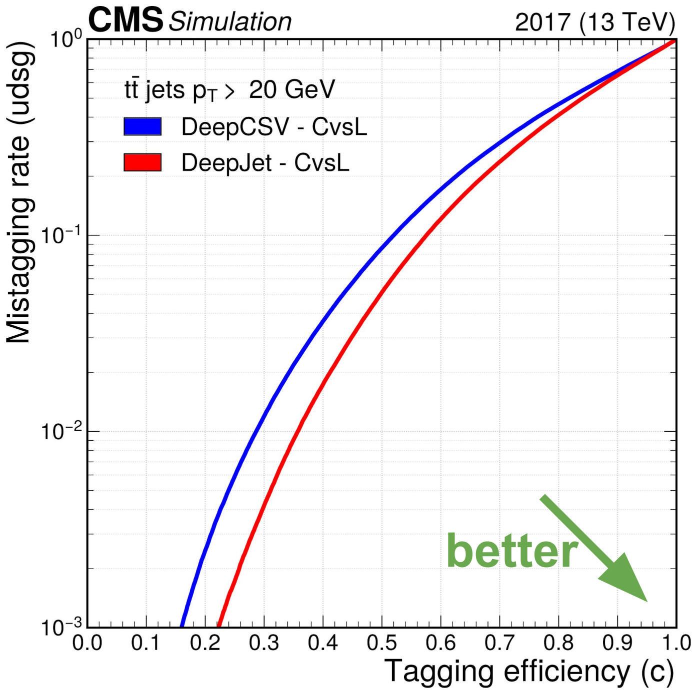

41.9 fb-1 (13 TeV, 2017)
CMS Preliminary
tt̄ events
AK4jets (pT > 30 GeV)
DeepJet
DeepCSV
DeepJet with SF applied
DeepCSV with SF applied
udsg
c
udsg with SF applied
c with SF applied
misid. probability
b jet efficiency
better

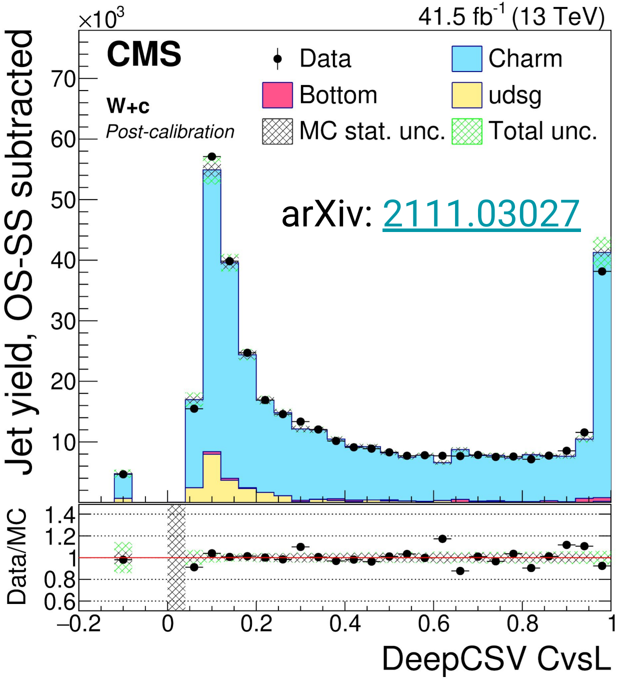

59.97 fb-1 (13 TeV, 2018)
CMS Preliminary
DeepJet L
PtRel
LT
TnP
System-8
Kin
weighted average
SFb
Combination
weighted average
fit
fit ± (stat ⊕ syst)
Jet pT [GeV]

c tagging

DeepJet algorithm

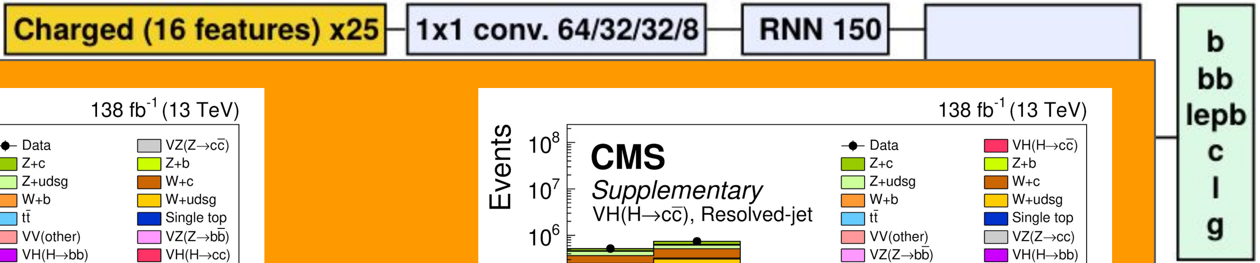




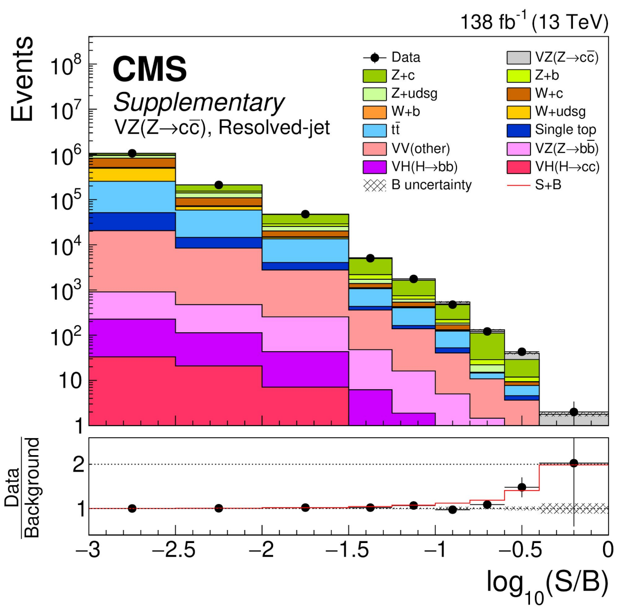


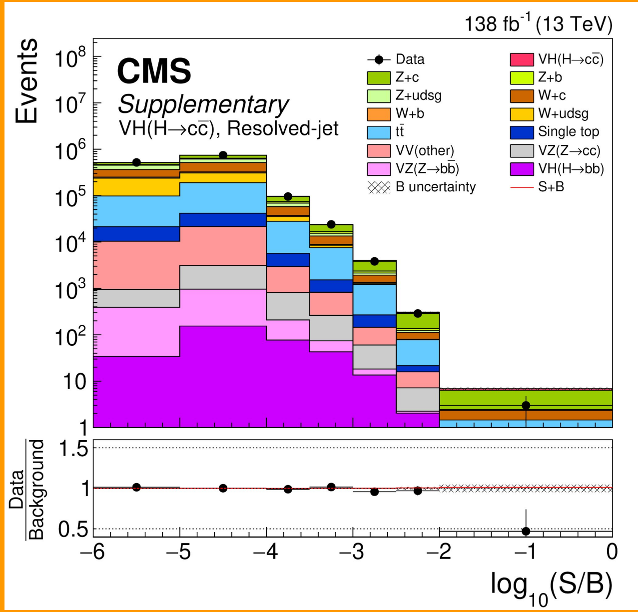
arXiv: [2111.03027](https://arxiv.org/abs/2111.03027)

DeepJet algorithm

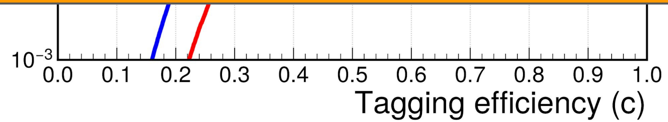


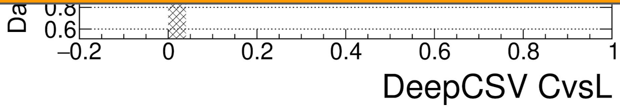
c tag



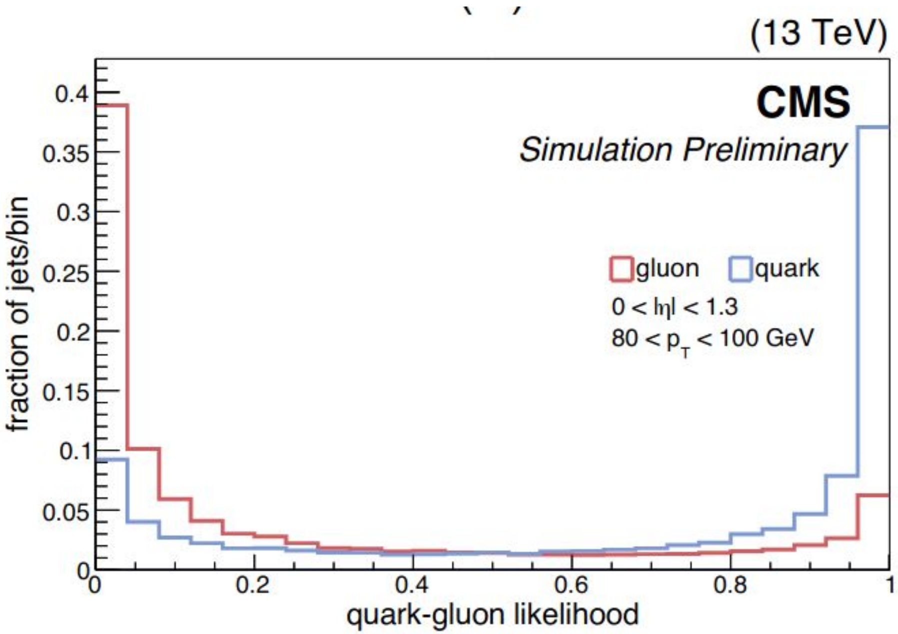


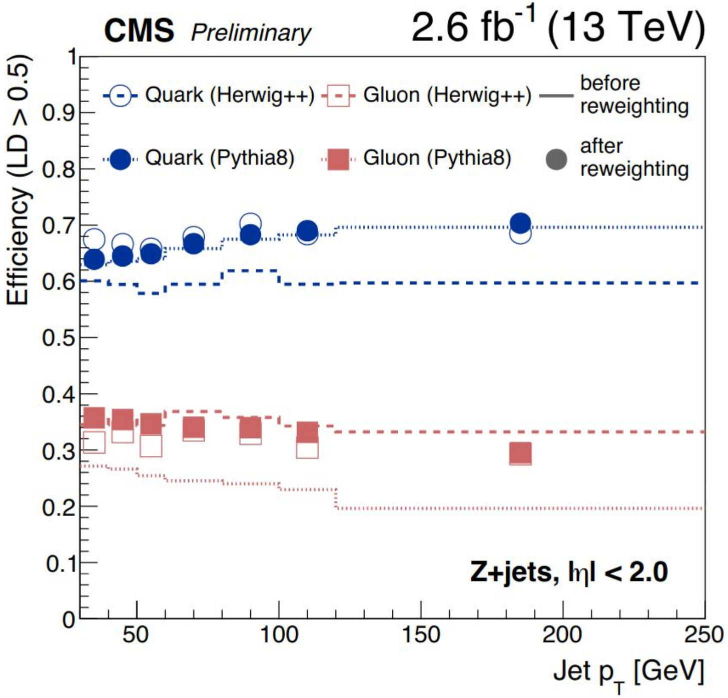
Allowed us to observed Z->cc and set strongest limits in H-> cc!
(arXiv:2205.05550)





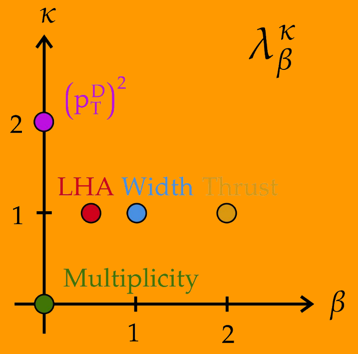
Quark/gluon tagging

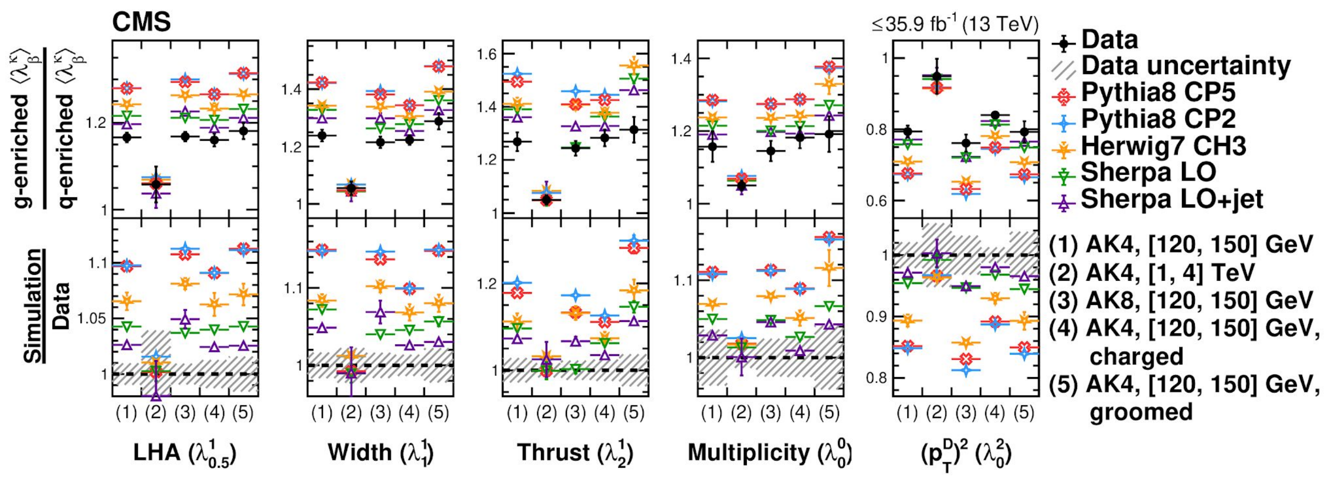




Quark Gluon Likelihood (QGL, a BDT based approach) has been used in CMS for many years.
We have some promising ML-based approaches for Run 3, but unfortunately no public results yet.

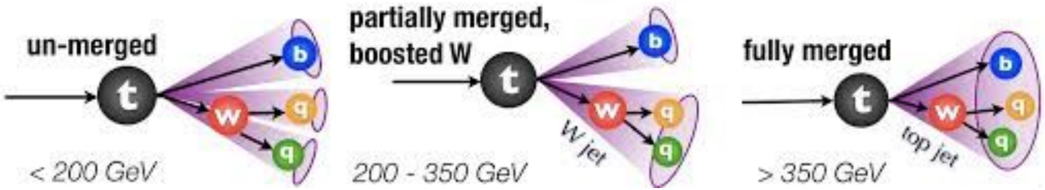
Quark/gluon tagging



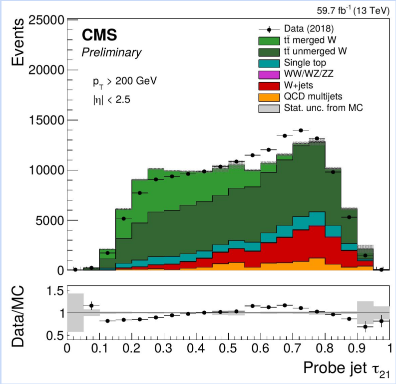


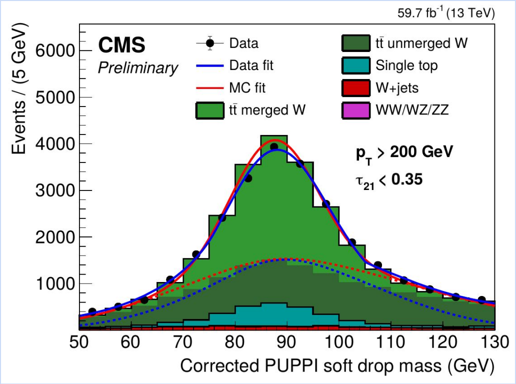
A detailed study of substructure variables between quark and gluon.
(arXiv:2109.03340)

We have some promising ML based approaches for Run 3, but unfortunately no public results yet.



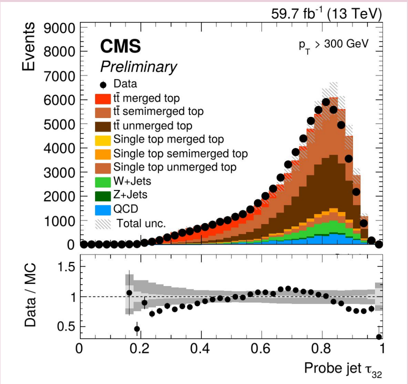
Boosted objects

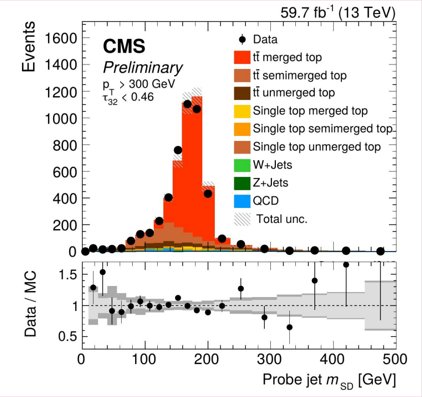




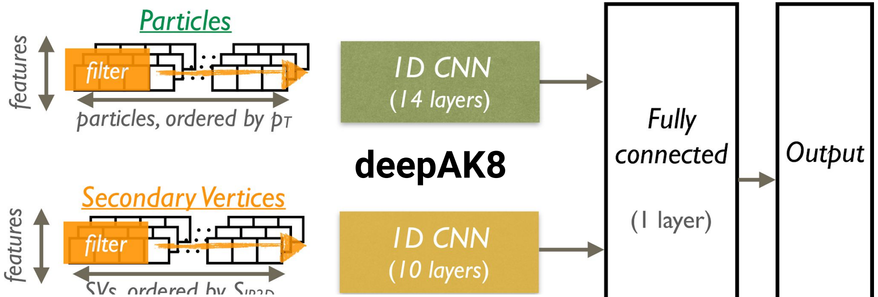
In CMS, the default jet cone size is **0.8** and the grooming technique (jet cleaning) is **softdrop**.

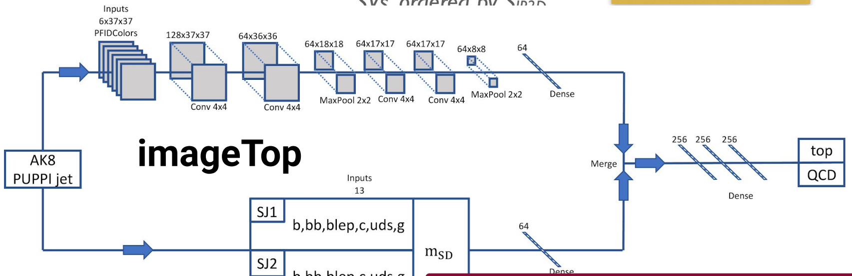
In CMS, we use the ratio of **nsubjettiness** as our default analytical discriminant. (τ_{21} for 2 prong decays, and τ_{32} for three prong decays).

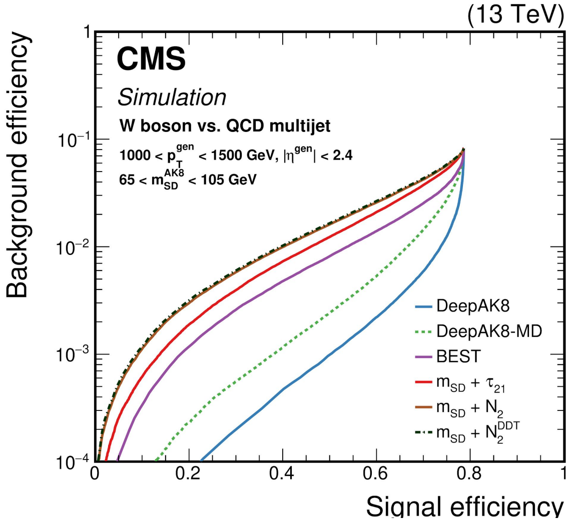


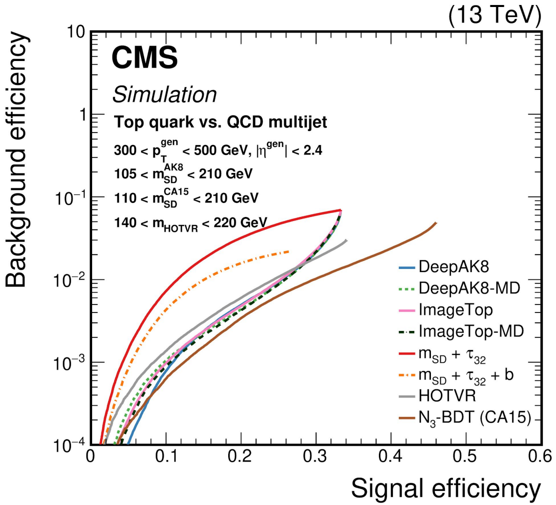


Boosted objects - The ML revolution







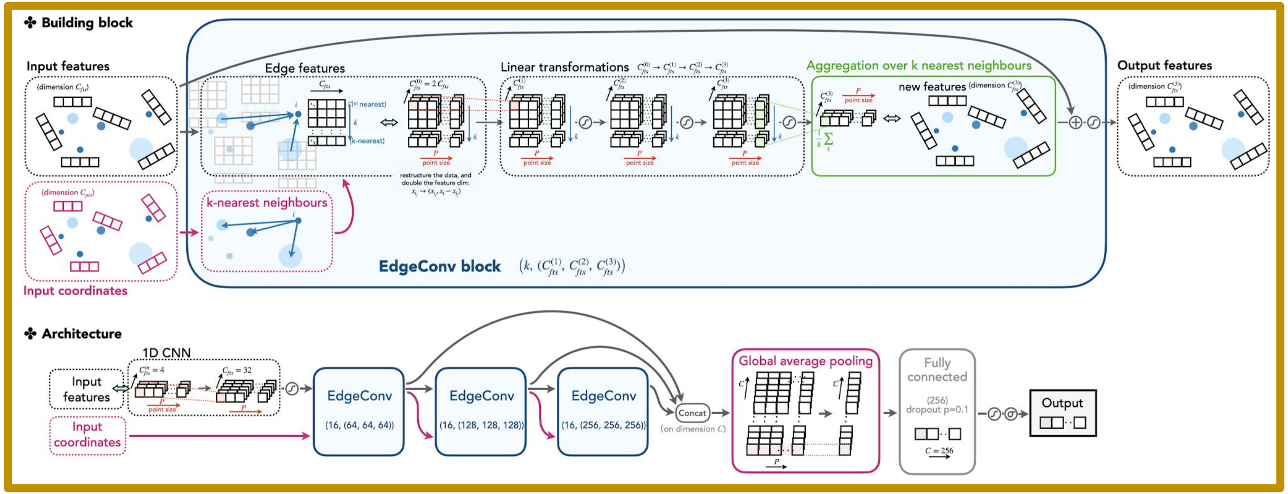


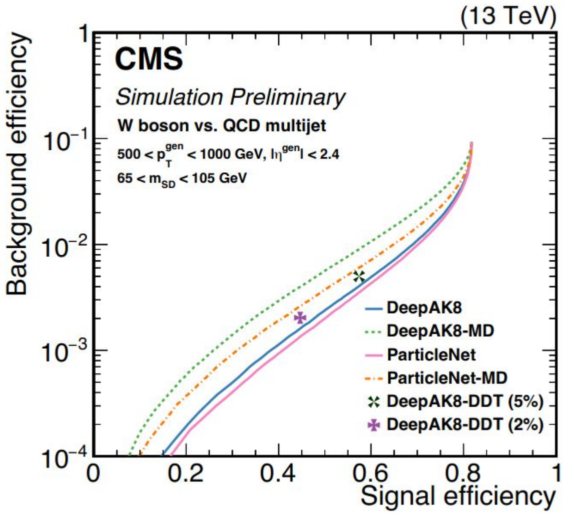
Algorithm	Subsection	jet p_T [GeV]	t quark	W boson	Z boson	H boson
$m_{SD} + \tau_{32}$	6.1	400	✓			
$m_{SD} + \tau_{32} + \mathrm{b}$	6.1	400	✓			
$m_{SD} + \tau_{21}$	6.1	200	✓	✓		
HOTVR	6.2	200	✓			
N_3-BDT (CA15)	6.3	200	✓			
$m_{SD} + N_2$	6.3	200		✓	✓	✓
BEST	6.5	500	✓	✓	✓	✓
ImageTop	6.6	600	✓			
DeepAK8$^{(*)}$	6.7	200	✓	✓	✓	✓
Jet mass decorrelated algorithms						
$m_{SD} + N_2^{DDT}$	6.3	200		✓	✓	✓
double-b	6.4	300			✓	✓
ImageTop-MD	6.6	600	✓			
DeepAK8-MD$^{(*)}$	6.7	200	✓	✓	✓	✓

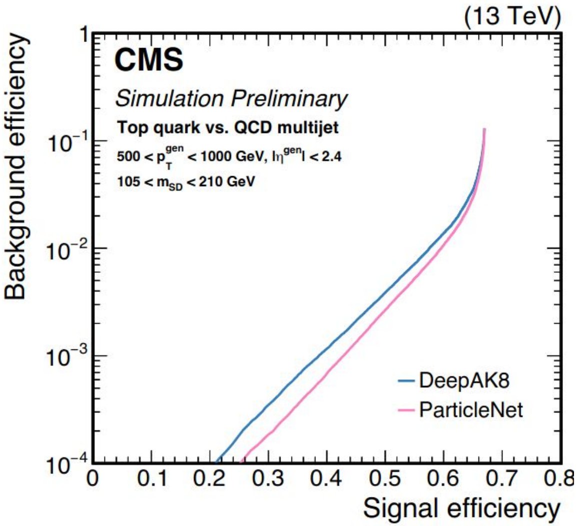
arxiv: 2004.08262

ParticleNet

Multi-classifier which profits from a Graph Neural Net and flat mass training samples.

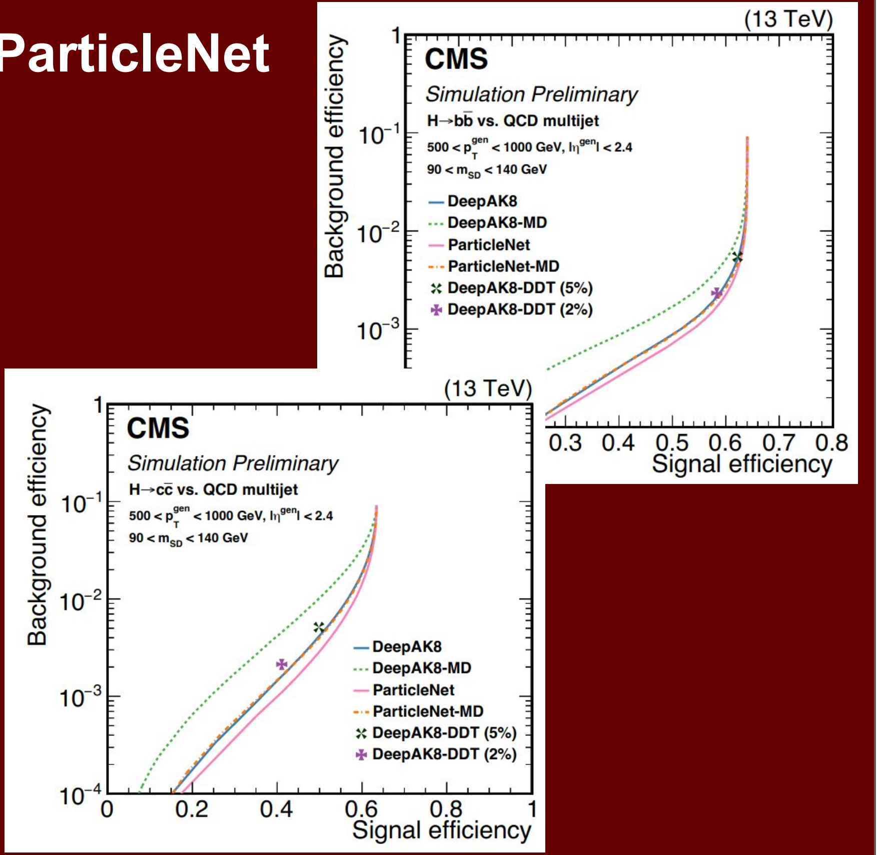




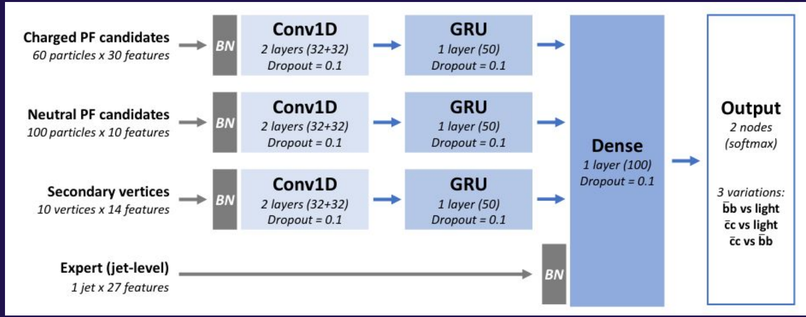


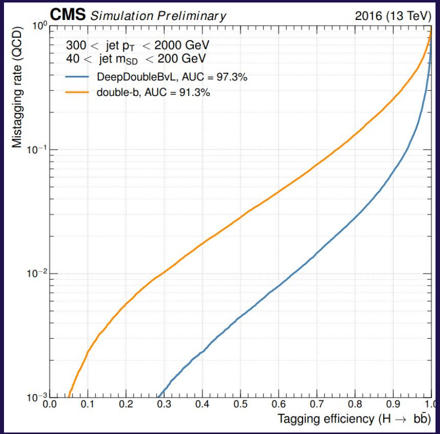
Higgs taggers (decaying to bb and cc)

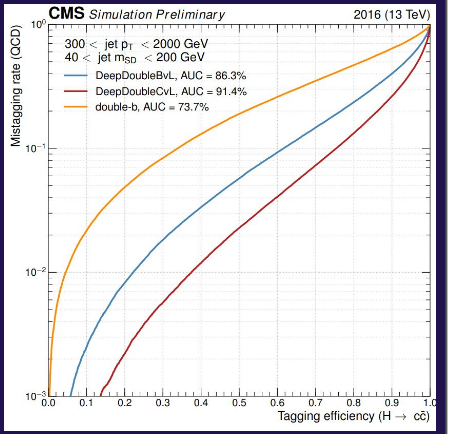
ParticleNet



DeepDoubleX

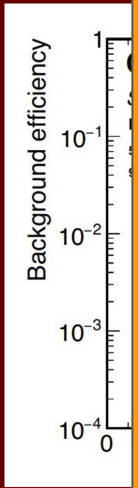


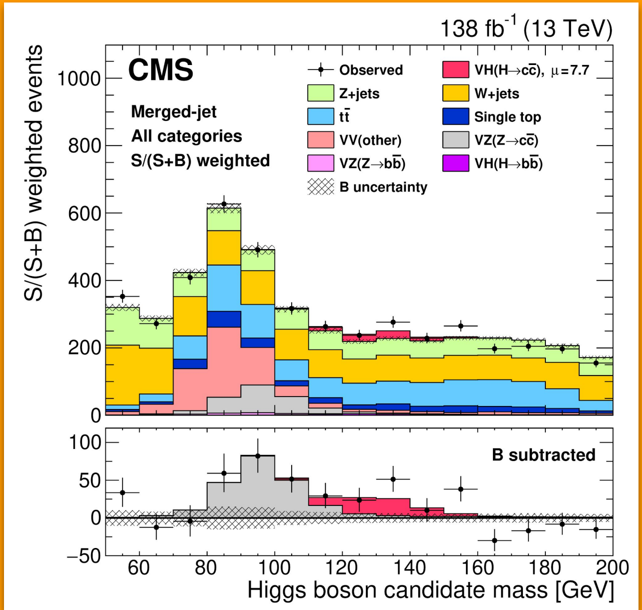




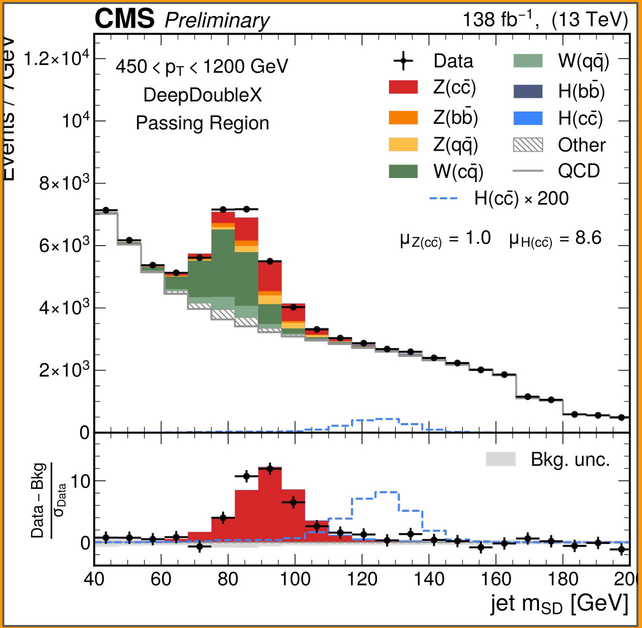
Higgs taggers (decaying to bb and cc)

Partic



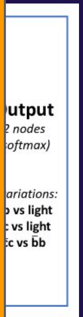


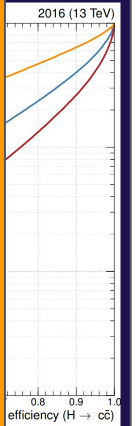
Allowed us to observed Z->cc and set strongest limits in H-> cc! (arXiv:2205.05550)



First Z->cc in this channel and the first ggH-> cc! (CMS-PAS-HIG-21-012)

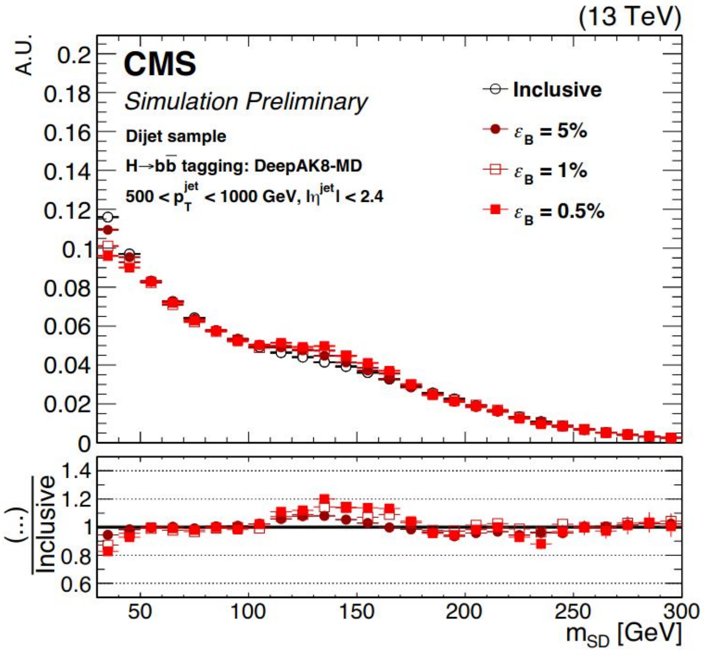
bleX

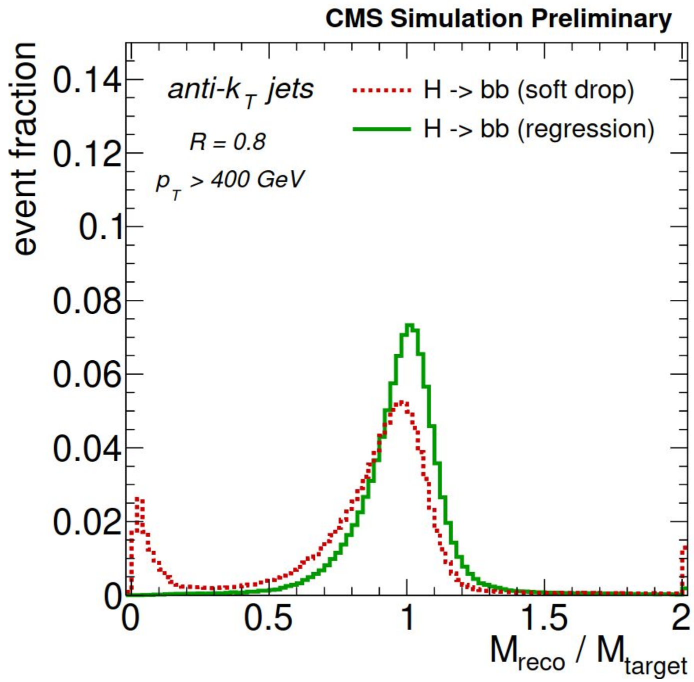




Jet mass decorrelation

Our boosted jet taggers sculpt the jet softdrop mass distribution. To mitigate this effect, our ML taggers decorrelate the jet mass from the other input variables.

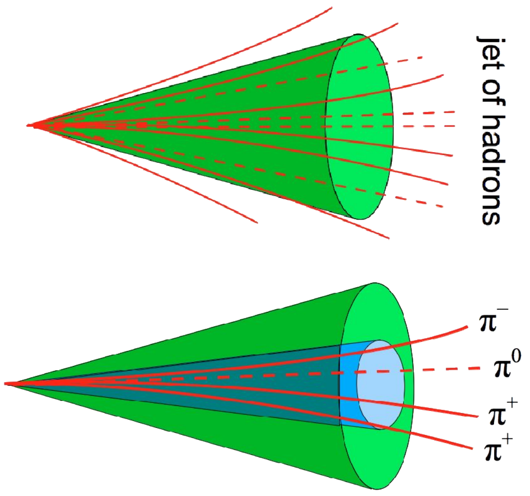


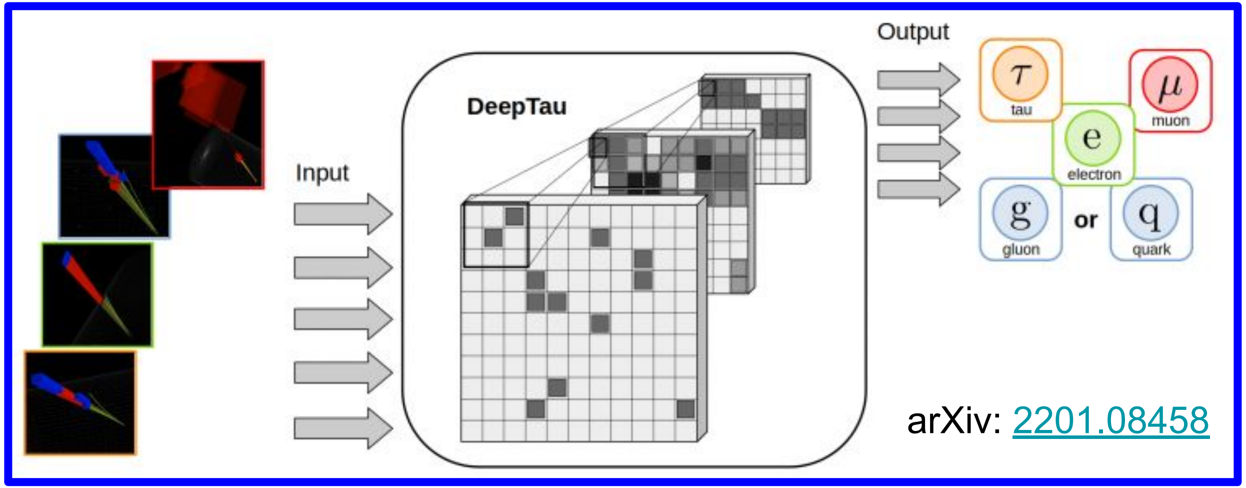


To mitigate the effects of over grooming, we developed a new mass regression algorithm.

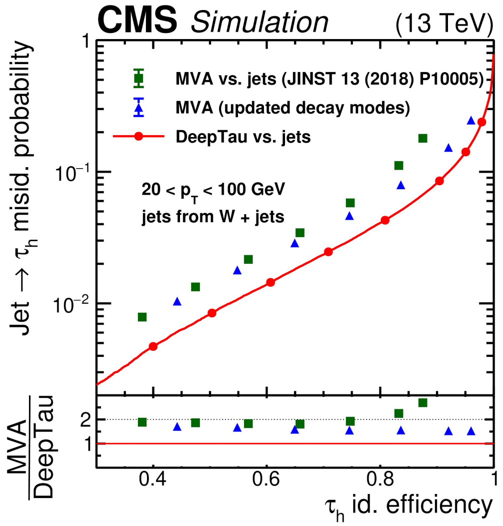
Jet mass regression

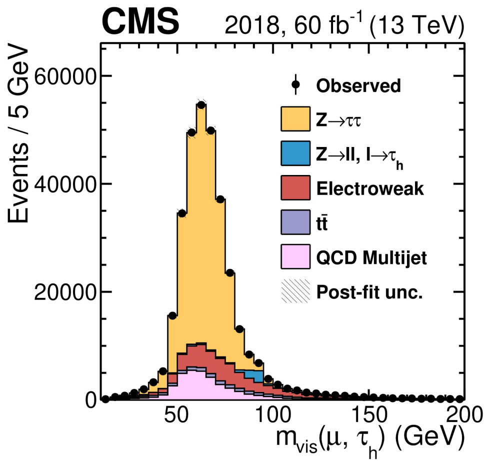
Tau jets

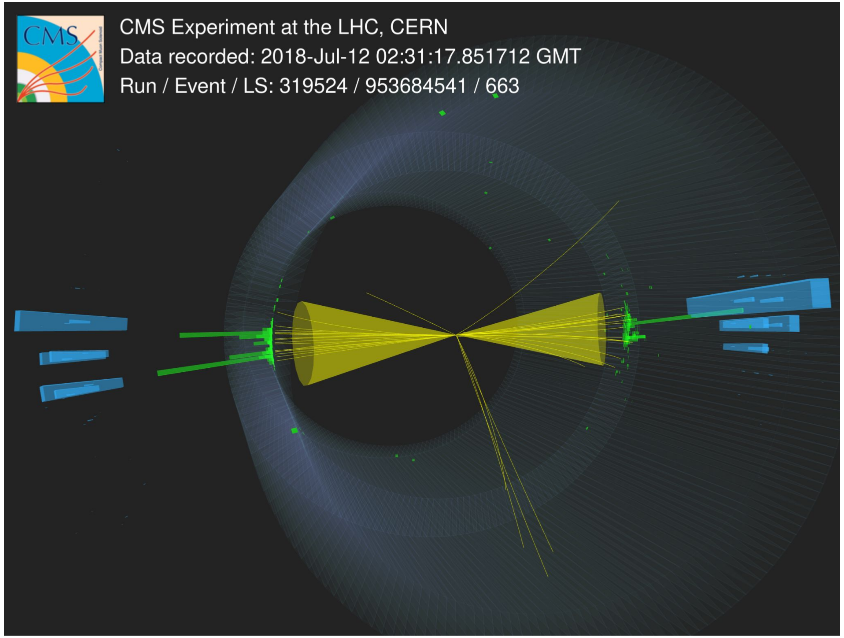




arXiv: 2201.08458

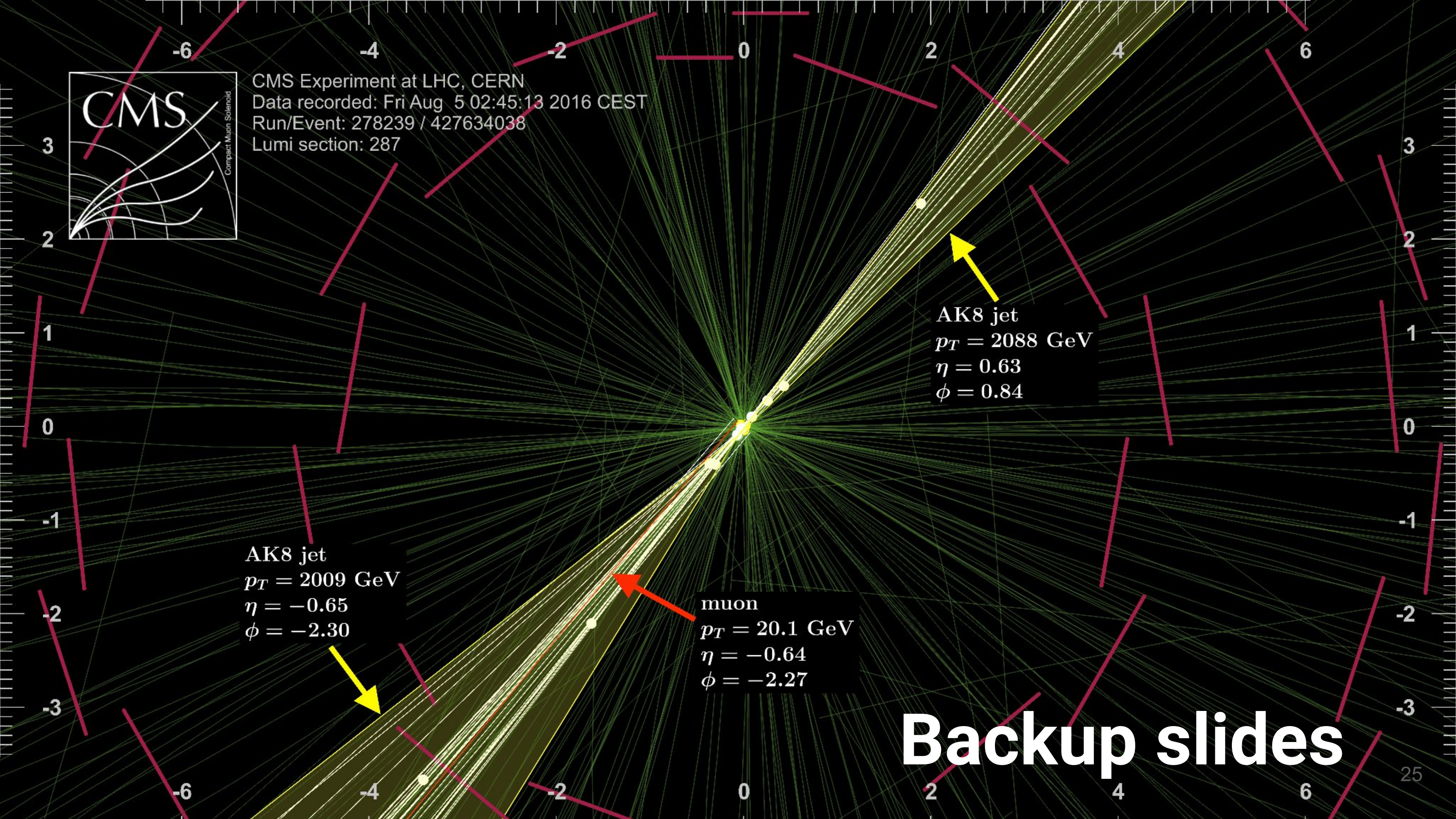






Thank you!

- This was just a quick summary of all the jet program at CMS.
- We have push the limits of jet reconstructions and algorithms for Run 2, and we are working on more improvements for Run 3.



Backup slides

PUPPI algorithm

1. Define α variable for neutral particle i:

$$\alpha_i = \log \sum_{j \neq i, \Delta R_{ij} < R_0} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \longrightarrow$$

high for hard, collinear particles
low for soft, wide angle particles

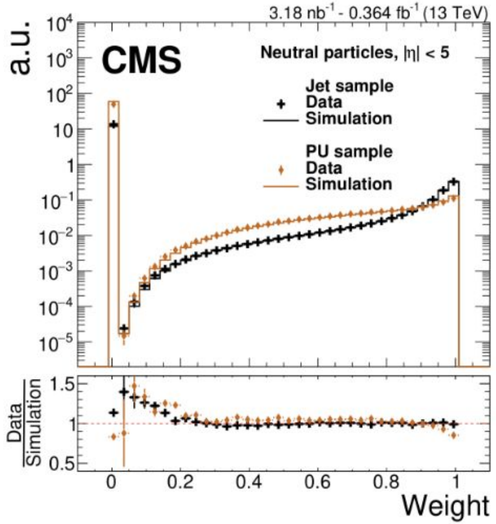
2. Suppose neutral PU same shape as charged PU
3. Calculate α distribution of charged PU, extract $\bar{\alpha}_{PU}$ and RMS_{PU}
4. For each particle i calculate:

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

5. Assign a weight w_i:

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

6. Apply w_i to particle's 4-momentum



PU Jet ID

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 $\sqrt{\sum_i p_{T_i}^2} / \sum_i p_{\mathrm{T}i}$

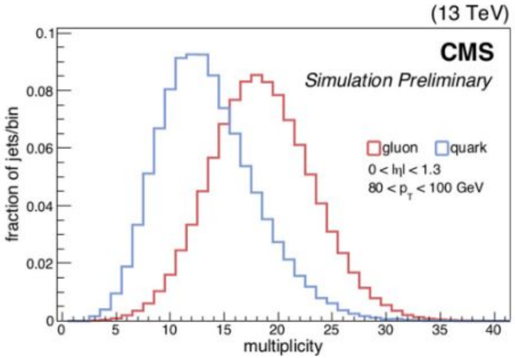
Quark Gluon Likelihood

1) Jet constituent multiplicity is higher in gluon jets than quark jets

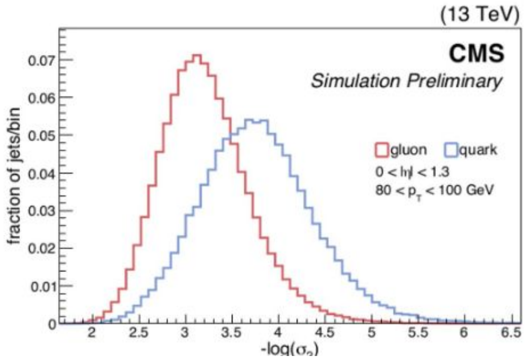
2) Quark jets are more collimated than gluon jets → the area covered by a gluon jet is larger

3) The fragmentation function of a gluon jet is softer than a quark jet → jet p_T is distributed differently among the jet constituents

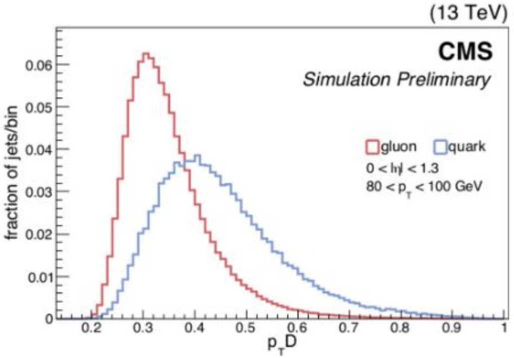
1) jet constituent multiplicity



2) minor axis of the elliptic jet

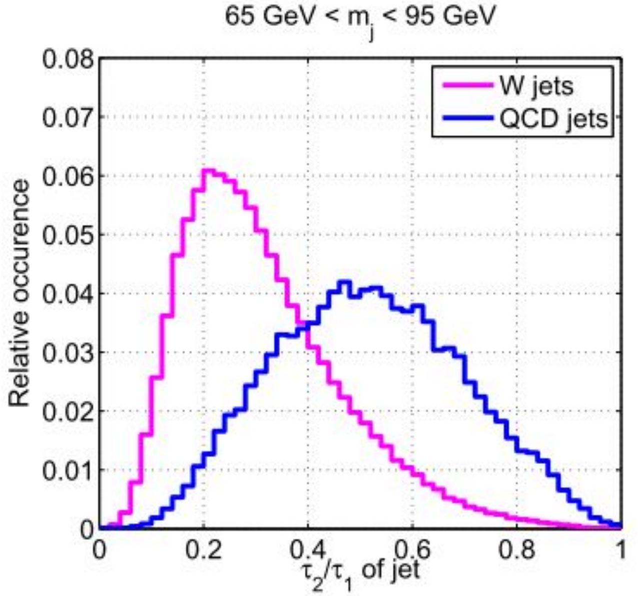


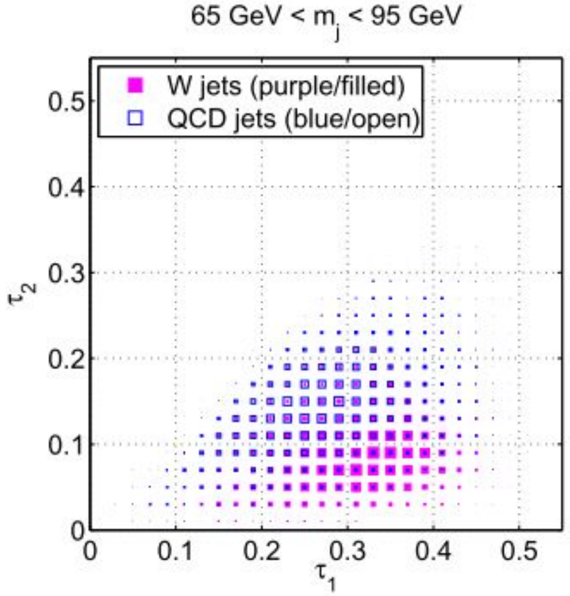
3) p_T-distribution variable



Nsubjettiness

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in \mathrm{Jet}} p_{Ti} \min\left\{\Delta R_{1i}^{\beta}, \Delta R_{2i}^{\beta}, \ldots, \Delta R_{Ni}^{\beta}\right\}.$$





DeepAK8

- Multiclassifier using AK8 jets
- 1D convolutional NN based on *ResNet*
- Mass decorrelated **DeepAK8-MD** with adversarial training technique:
  - mass predictor added to the network
  - reweighed to flat p_T and m_{SD}

Two lists of inputs:

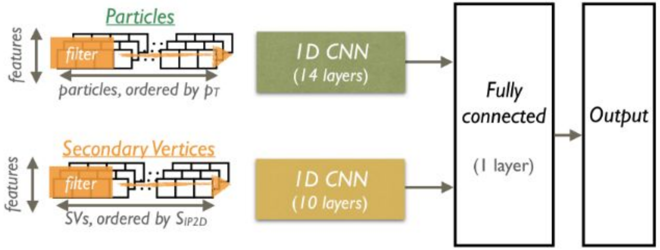
Particle list

- up to 100 PF candidates per jet
- ordered by decreasing p_T
- 42 features: p_T, change, energy, angular separation, tracking info..

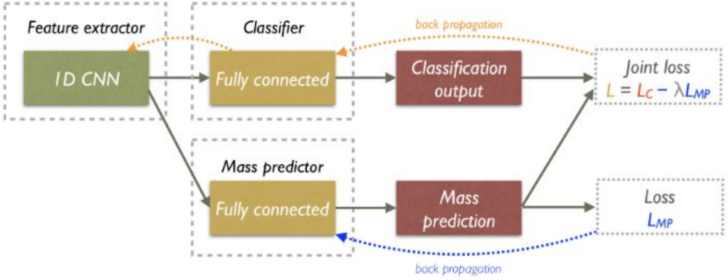
Secondary Vertex list

- up to 7 SVs
- ordered by 2D impact parameter significance
- 15 features: displacement, kinematics, quality criteria..

DeepAK8 architecture



DeepAK8-MD architecture



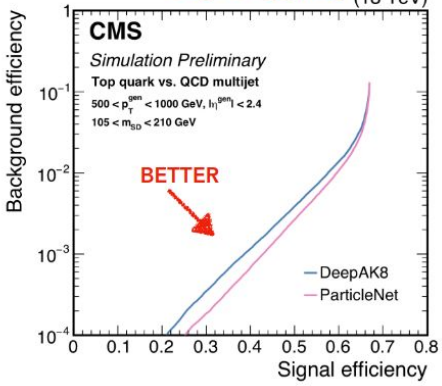
CMS: JINST 15 (2020) P06005

ParticleNet

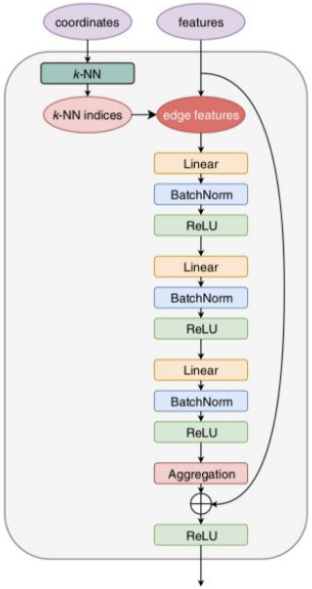
- Multiclass graph neural network
- Jet as *unordered* set of particles (**cloud**)
- Based on EdgeConv: point cloud as graph
- Inputs: PF particles and SV
- Mass decorrelated **ParticleNet-MD:**
  - training with dedicated simulation of hadronically decaying particles and QCD jets
  - flat in p_T and m_{SD}
  - same inputs and architecture as nominal network

top tagging

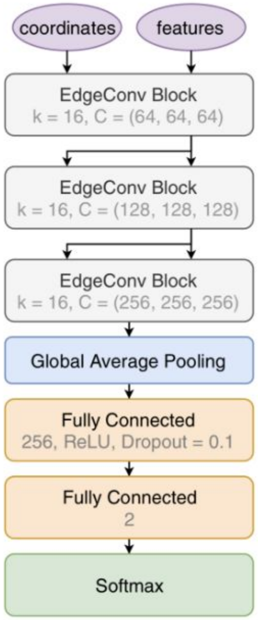
CMS-DP-2020-002



EdgeConv block



ParticleNet architecture



H. Qu and L. Gouskos: Phys.Rev.D 101 (2020) 5, 056019