Jet substructure measurements in CMS experiment

Suman Chatterjee

for the CMS Collaboration

HEPHY Vienna

26/07/2021

EPS-HEP Conference 2021

~~Hamburg, Germany~~ Online

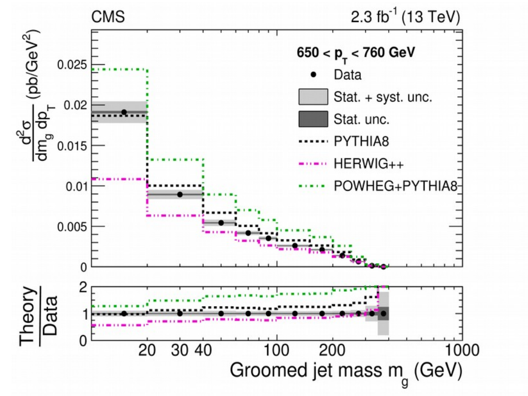
Email: suman.chatterjee@cern.ch

Why jet substructure measurements?

Why jet substructure measurements?

Improved understanding of QCD

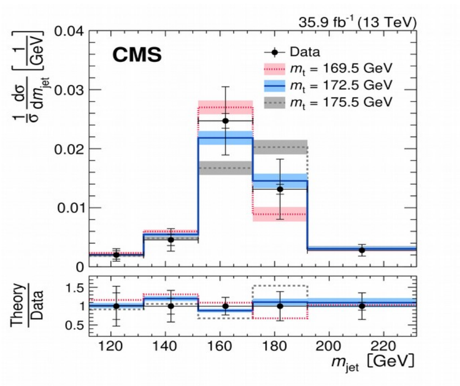
Resummation region enhanced by jet grooming

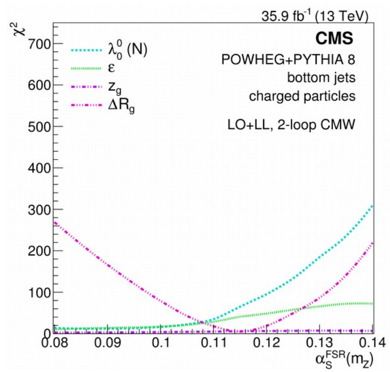


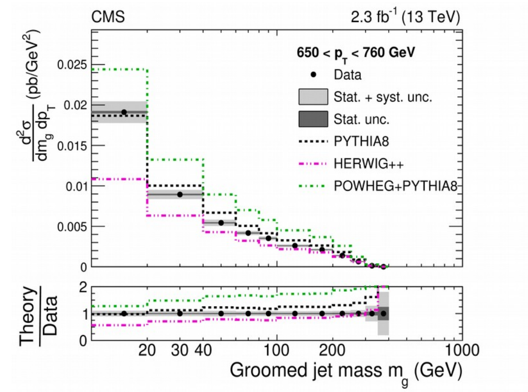
Why jet substructure measurements?

Improved understanding of QCD

Resummation region enhanced by jet grooming



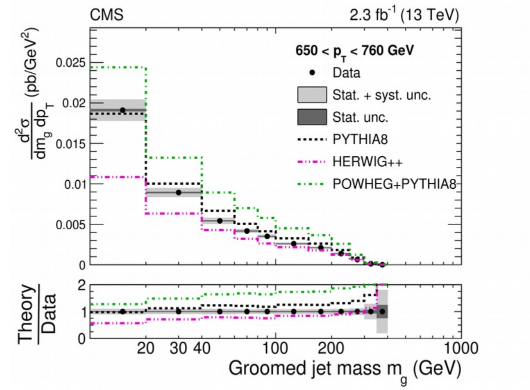




Precision tests of SM parameters (and probing BSM effects)

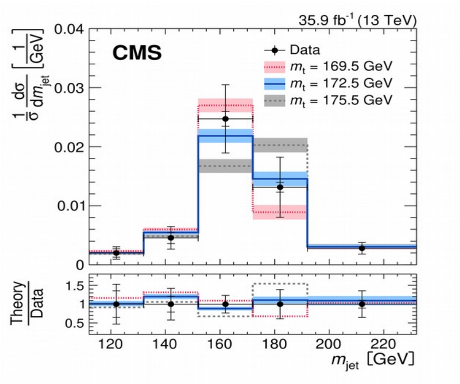
α_s running, top quark mass

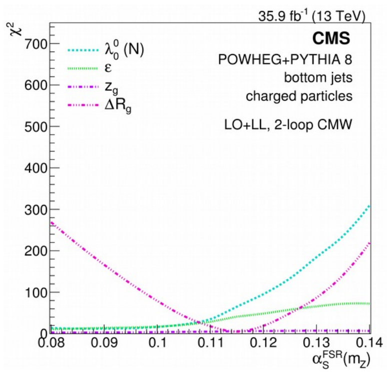
Why jet substructure measurements?



Improved understanding of QCD

Resummation region enhanced by jet grooming



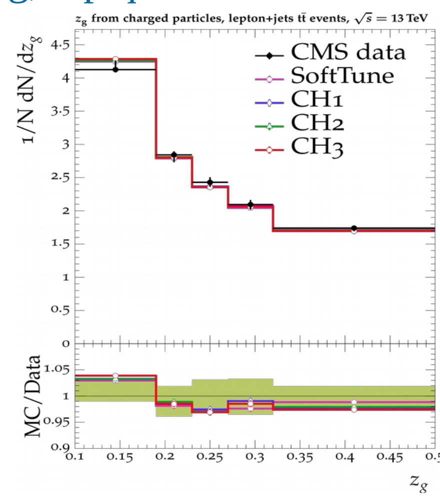


Precision tests of SM parameters (and probing BSM effects)

α_S running, top quark mass

Tuning general purpose MC generators

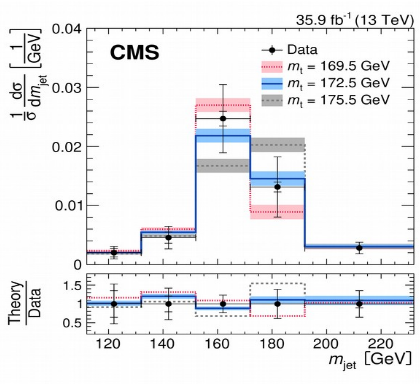
Understanding parton shower evolution
Improving underlying event description
...
Improves modeling of particle taggers → Helps measurements & searches

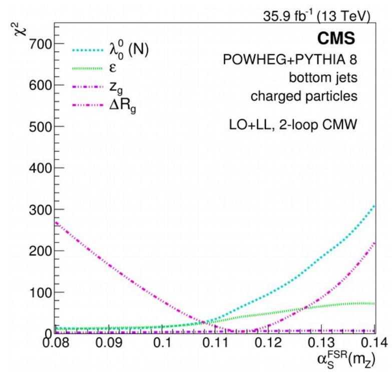


Why jet substructure measurements?

Improved understanding of QCD

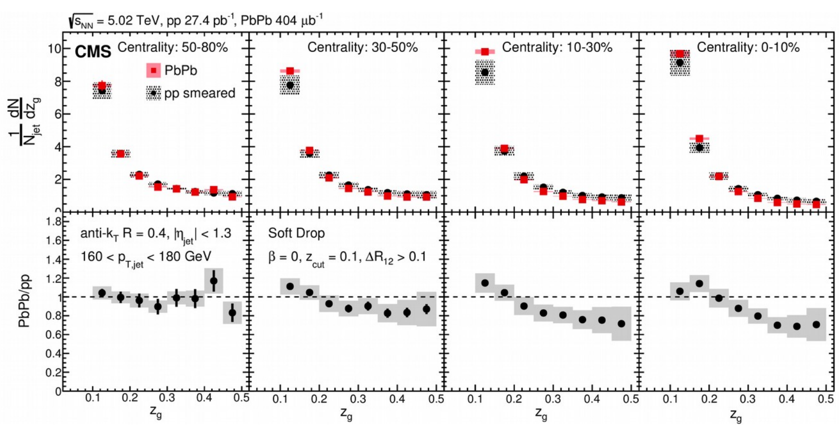
Resummation region enhanced by jet grooming

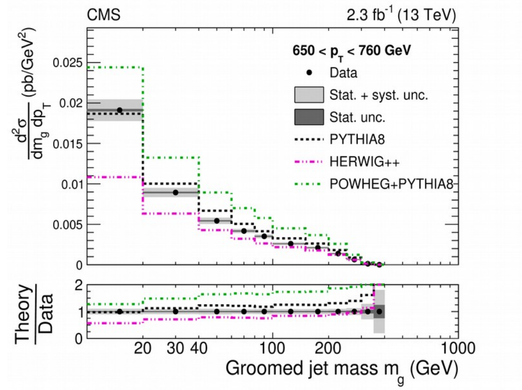




Tuning general purpose MC generators

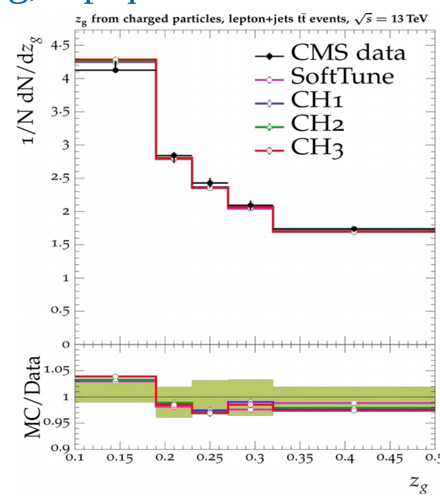
Understanding parton shower evolution
Improving underlying event description
...
Improves modeling of particle taggers → Helps measurements & searches





Precision tests of SM parameters (and probing BSM effects)

α_s running, top quark mass

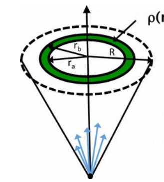


Sensing emergent phenomena

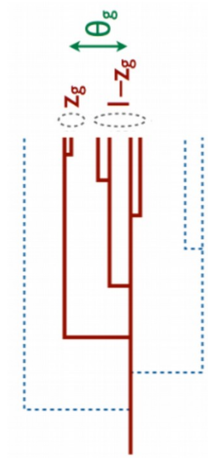
QGP
Dead cone effects
...

Reference	√s	Collision	Events	Jet sample and kinematic region	Observables
1204.3170	7 TeV	pp	Inclusive jets	q/g-jets (AK7), $20 < p_T < 1000$ GeV q/g-jets (AK5), $50 < p_T < 1000$ GeV	Jet shapes, charged hadron multiplicity, width
1205.5872	2.76 TeV	pp/PbPb	Dijet	q/g-jets (AK3), $40 < p_T < 320$ GeV	Jet fragmentation function
1310.0878	2.76 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK3), $100 < p_T < 300$ GeV	Jet fragmentation function
1406.0932	2.76 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK3), $p_T > 100$ GeV	Jet fragmentation function
1310.0878	2.76 TeV	pp/PbPb	Dijet	q/g-jets (AK3), $p_T > 100$ GeV	Jet shapes
1809.08602	2.76 TeV	pp/PbPb	Dijet	q-jets (AK3), $p_T > 30$ GeV	Jet shapes
HIN-19-003	2.76 TeV	pp/PbPb	Dijet	q/g-jets (AK4), $p_T > 50$ GeV	Jet shapes
QCD-10-041	7 TeV	pp	Dijet	q/g-jets (KT6), $97 < p_T < 1032$ GeV	Subjet multiplicities
1706.05868	8 TeV	pp	Inclusive jets	q/g-jets (AK5), $400 < p_T < 1500$ GeV	Jet charge
2004.00602	5.02 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK4), $p_T > 120$ GeV	Jet charge
1703.06330	8 TeV	pp	$t\bar{t}$	top-jets (CA12), $p_T > 400$ GeV	Jet mass
1303.4811	8 TeV	pp	Dijet + W/Z jet	q/g-jets (AK7), $220 < p_T < 1500$ GeV q-jets (AK7, CA8/12), $125 < p_T < 450$ GeV	Nominal + groomed (trimming, pruning, filtering) jet mass
1805.05145	5.02 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK4), $140 < p_T < 300$ GeV	Soft-drop jet mass
1807.05974	13 TeV	pp	Dijet	q/g-jets (AK8), $200 < p_T < 3000$ GeV	Nominal + soft-drop jet mass
1911.03800	13 TeV	pp	$t\bar{t}$	top-jets (XC12), $p_T > 400$ GeV	Xcone groomed jet mass
1708.09429	5.02 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK4), $140 < p_T < 500$ GeV	Soft-drop splitting function
1808.07340	13 TeV	pp	$t\bar{t}$	q-jets (AK4), $p_T > 30$ GeV g-jets (AK4), $p_T > 30$ GeV b-jets (AK4), $p_T > 30$ GeV	Jet substructure and soft-drop observables
SMP-20-010	13 TeV	pp	Dijet, Z+jets	q/g-jets (AK4), $50 < p_T < 4000$ GeV q-jets (AK4), $50 < p_T < 4000$ GeV	Jet angularities

$$z = \frac{p_{\parallel}^{\text{track}}}{p^{\text{jet}}}, \quad \xi = \ln\frac{1}{z},$$



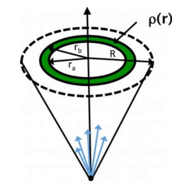
List of jet substructure measurements in CMS



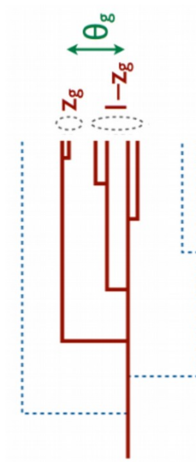
Reference	√s	Collision	Events	Jet sample and kinematic region	Observables
1204.3170	7 TeV	pp	Inclusive jets	q/g-jets (AK7), 20 < p_T < 1000 GeV q/g-jets (AK5), 50 < p_T < 1000 GeV	Jet shapes, charged hadron multiplicity, width
1205.5872	2.76 TeV	pp/PbPb	Dijet	q/g-jets (AK3), 40 < p_T < 320 GeV	Jet fragmentation function
1310.0878	2.76 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK3), 100 < p_T < 300 GeV	Jet fragmentation function
1406.0932	2.76 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK3), p_T > 100 GeV	Jet fragmentation function
1310.0878	2.76 TeV	pp/PbPb	Dijet	q/g-jets (AK3), p_T > 100 GeV	Jet shapes
1809.08602	2.76 TeV	pp/PbPb	Dijet	q-jets (AK3), p_T > 30 GeV	Jet shapes
HIN-19-003	2.76 TeV	pp/PbPb	Dijet	q/g-jets (AK4), p_T > 50 GeV	Jet shapes
QCD-10-041	7 TeV	pp	Dijet	q/g-jets (KT6), 97 < p_T < 1032 GeV	Subjet multiplicities
1706.05868	8 TeV	pp	Inclusive jets	q/g-jets (AK5), 400 < p_T < 1500 GeV	Jet charge
2004.00602	5.02 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK4), p_T > 120 GeV	Jet charge
1703.06330	8 TeV	pp	$t\bar{t}$	top-jets (CA12), p_T > 400 GeV	Jet mass
1303.4811	8 TeV	pp	Dijet + W/Z jet	q/g-jets (AK7), 220 < p_T < 1500 GeV q-jets (AK7, CA8/12), 125 < p_T < 450 GeV	Nominal + groomed (trimming, pruning, filtering) jet mass
1805.05145	5.02 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK4), 140 < p_T < 300 GeV	Soft-drop jet mass
1807.05974	13 TeV	pp	Dijet	q/g-jets (AK8), 200 < p_T < 3000 GeV	Nominal + soft-drop jet mass
1911.03800	13 TeV	pp	$t\bar{t}$	top-jets (XC12), p_T > 400 GeV	Xcone groomed jet mass
1708.09429	5.02 TeV	pp/PbPb	Inclusive jets	q/g-jets (AK4), 140 < p_T < 500 GeV	Soft-drop splitting function
1808.07340	13 TeV	pp	$t\bar{t}$	q-jets (AK4), p_T > 30 GeV g-jets (AK4), p_T > 30 GeV b-jets (AK4), p_T > 30 GeV	Jet substructure and soft-drop observables
SMP-20-010	13 TeV	pp	Dijet, Z+jets	q/g-jets (AK4), 50 < p_T < 4000 GeV q-jets (AK4), 50 < p_T < 4000 GeV	Jet angularities

Highlight of today's talk

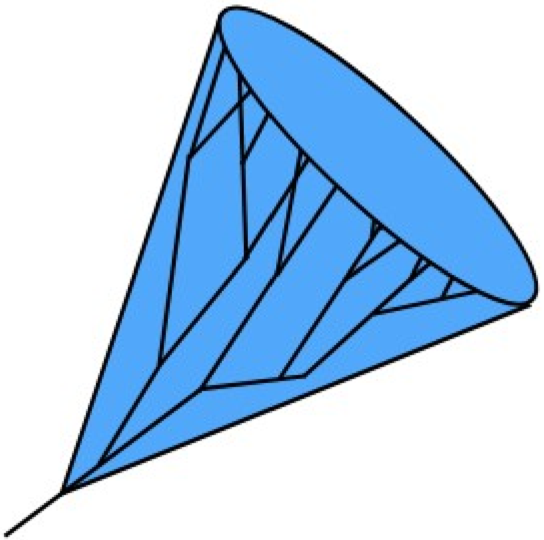
$$z = \frac{p_{\parallel}^{\text{track}}}{p^{\text{jet}}}, \quad \xi = \ln\frac{1}{z},$$



List of jet substructure measurements in CMS



Jet substructure measurements in CMS



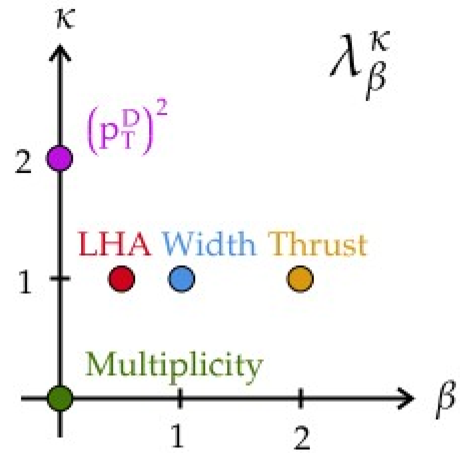
<u>Observables</u>

Generalized angularities

$$\lambda_\beta^\kappa = \sum_{i \in jet} (z_i)^\kappa (\theta_i)^\beta$$

$$z_i = p_{T\,i} / \sum p_{T\,j} \qquad \theta_i = \Delta R_{i,\hat{n}} / R_{jet}$$

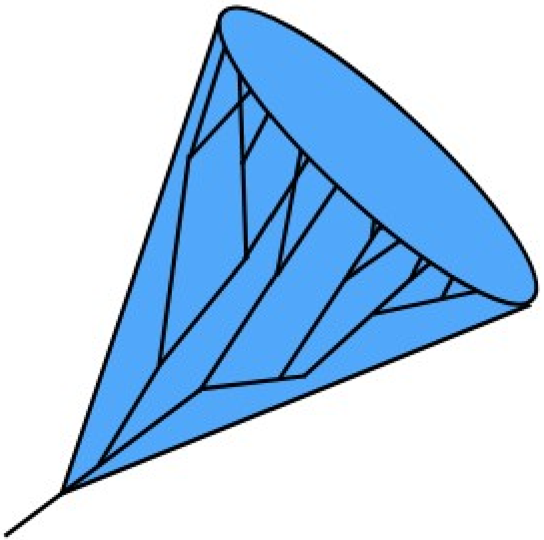
n: jet axis



β<=1: Winner-takes-all (WTA) axis

β >1: Anti-k_T axis

Jet substructure measurements in CMS



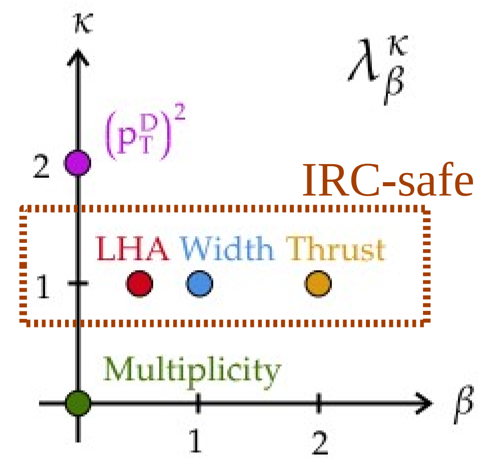
Observables

Generalized angularities

$$\lambda_\beta^\kappa = \sum_{i \in jet} (z_i)^\kappa (\theta_i)^\beta$$

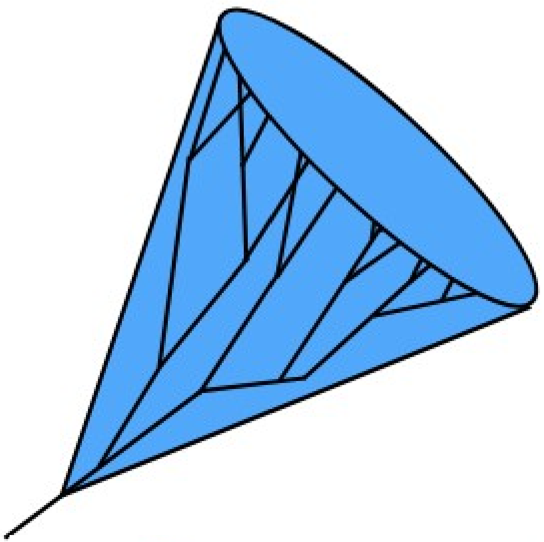
$$z_i = p_{T\,i} / \sum p_{T\,j} \qquad \theta_i = \Delta R_{i,\hat{n}} / R_{jet}$$

n: jet axis



β<=1: Winner-takes-all (WTA) axis
β >1: Anti-k_T axis

Jet substructure measurements in CMS



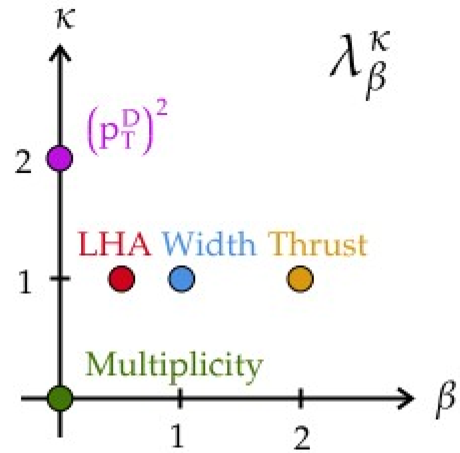
Observables

Generalized angularities

$$\lambda_\beta^\kappa = \sum_{i \in jet} (z_i)^\kappa (\theta_i)^\beta$$

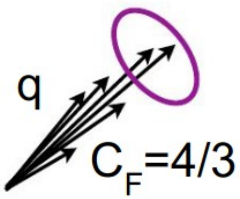
$$z_i = p_{T\,i} / \sum p_{T\,j} \qquad \theta_i = \Delta R_{i,\hat{n}} / R_{jet}$$

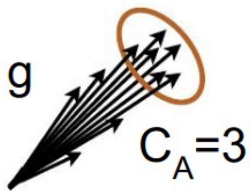
n: jet axis



β<=1: Winner-takes-all (WTA) axis

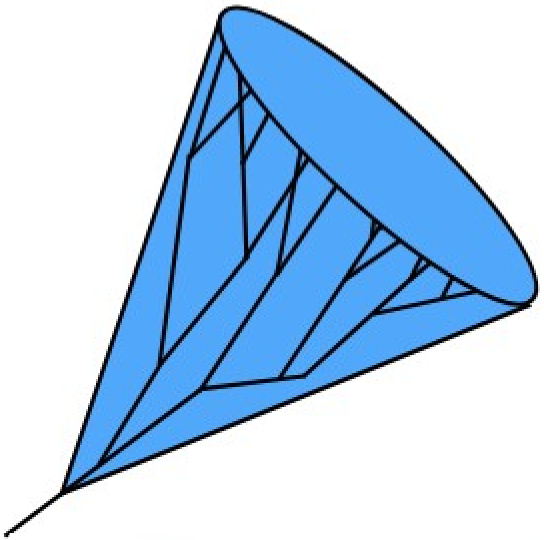
β >1: Anti-k_T axis

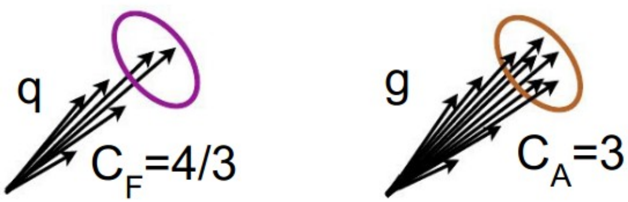




← Sensitive to differences between quark & gluon jets

Jet substructure measurements in CMS





Observables

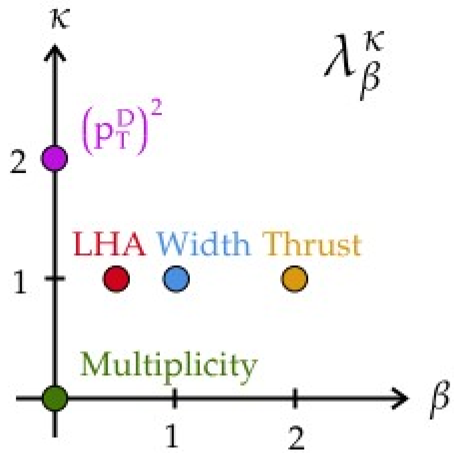
Generalized angularities

$$\lambda_\beta^\kappa = \sum_{i \in jet} (z_i)^\kappa (\theta_i)^\beta$$

$$z_i = p_{T\,i} / \sum p_{T\,j} \qquad \theta_i = \Delta R_{i,\hat{n}} / R_{jet}$$

n: jet axis

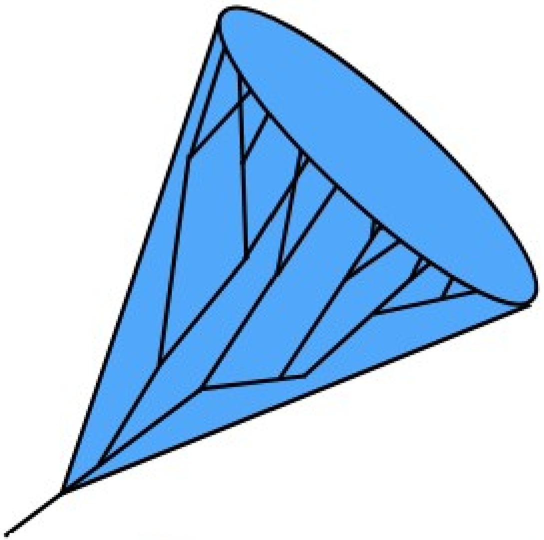
β<=1: Winner-takes-all (WTA) axis
β >1: Anti-k_T axis



Measurements: Five λ^β_κ variables in quark- and gluon-enriched jet samples (results reported in particle-level)

Study motivated by JHEP 1707 (2017) 091 (Gras et al.)

Jet substructure measurements in CMS



Observables

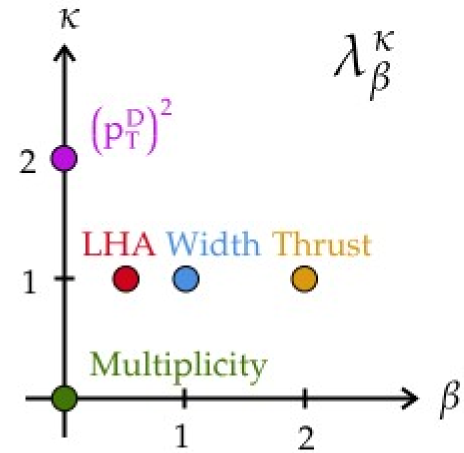
Generalized angularities

$$\lambda^{\kappa}_{\beta} = \sum_{i \in jet} (z_i)^{\kappa} (\theta_i)^{\beta}$$

$$z_i = p_{T\,i} / \sum p_{T\,j} \qquad \theta_i = \Delta R_{i,\hat{n}} / R_{jet}$$

n: jet axis

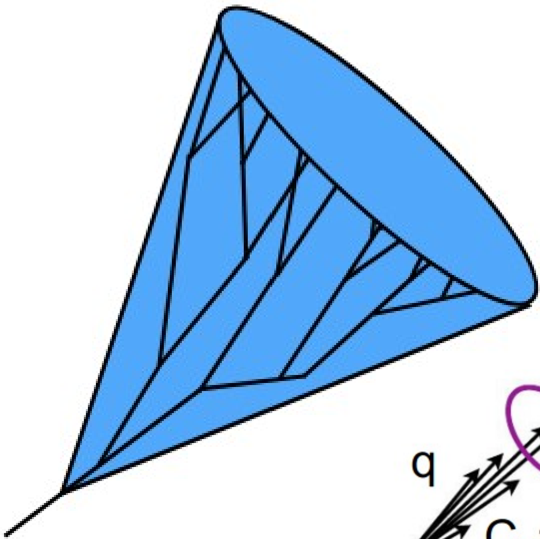
β<=1: Winner-takes-all (WTA) axis
β >1: Anti-k_T axis



Measurements: Five λ^{β}_{κ} variables in quark- and gluon-enriched jet samples (results reported in particle-level)

Purpose: Detailed understanding of jet composition & its modeling (not designing a quark-gluon tagger)

Jet substructure measurements in CMS

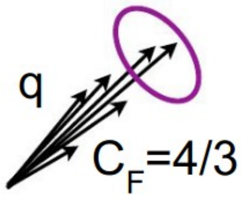


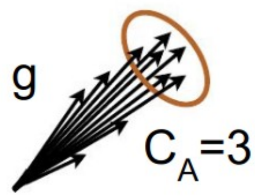
Observables

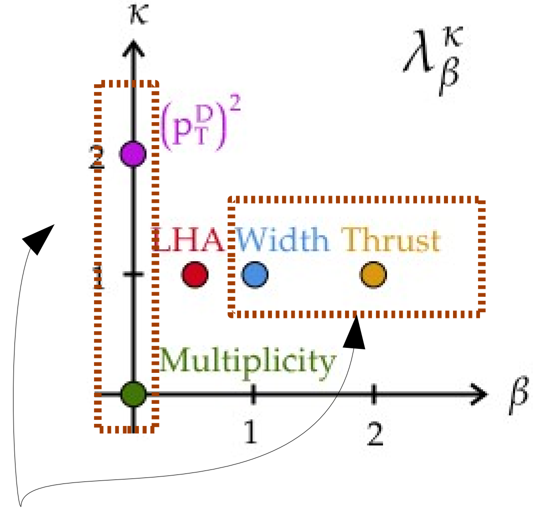
Generalized angularities

$$\lambda_\beta^\kappa = \sum_{i \in jet} (z_i)^\kappa \, (\theta_i)^\beta$$

$$z_i = p_{T\,i} / \sum p_{T\,j} \qquad \theta_i = \Delta R_{i,\hat{n}} / R_{jet}$$

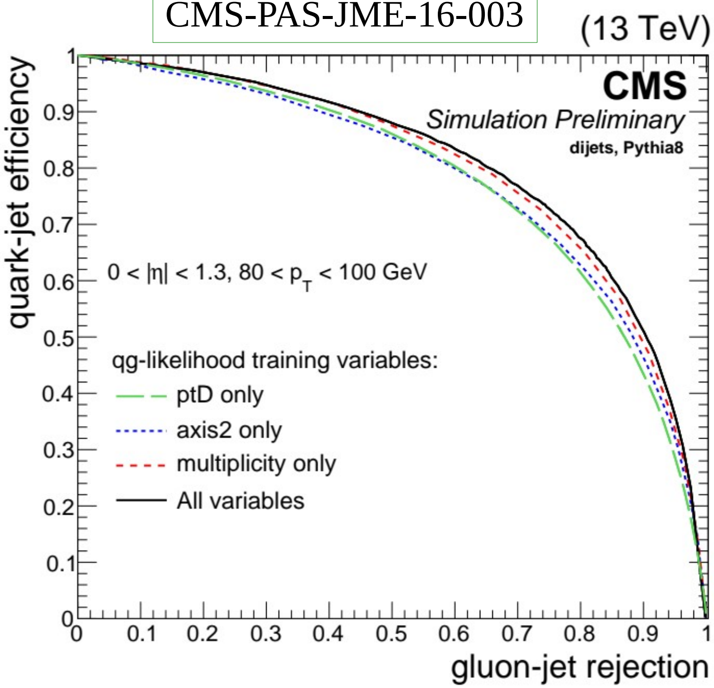




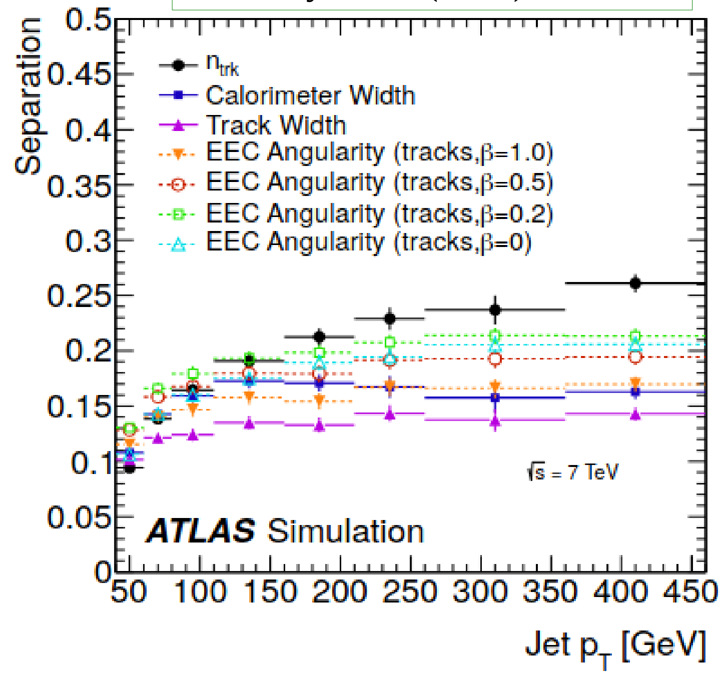


Used by experiments for quark-gluon tagging

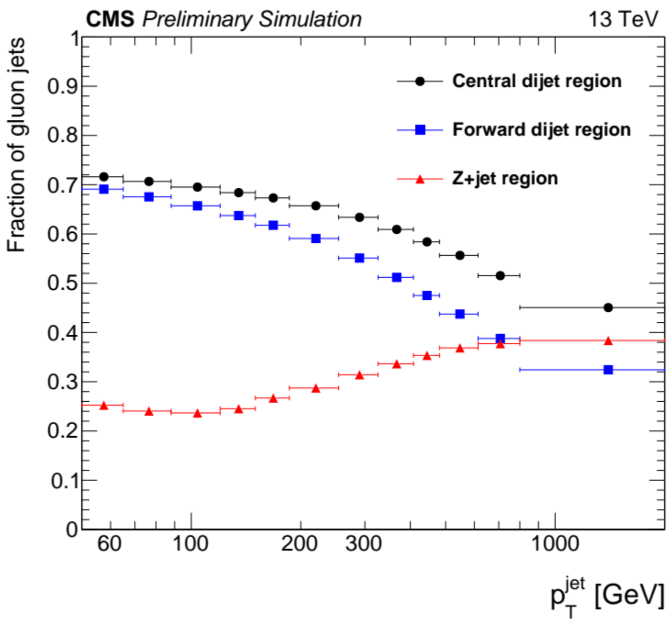
CMS-PAS-JME-16-003



Eur. Phys. J. C (2014) 74: 3023



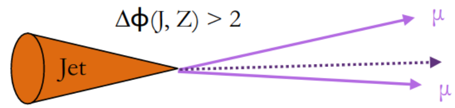
Quark- and gluon-enriched event samples



Quark-enriched jets from **Z(μμ) + jets** region:

1+ jets, 2+ muons
Leading jet must pass p_T, $|y|$ cuts,
not overlap with Z muons

$\Delta\phi(J, Z) > 2$



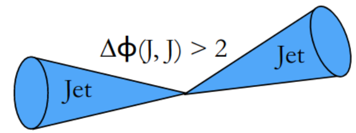
$|p_T(J) - p_T(Z)| / \sum p_T < 0.3$

$|m_{\mu\mu} - 90| < 20$ GeV
$p_T^{\mu\mu} > 30$ GeV

Gluon-enriched jets from **dijet** region:

2+ jets
Leading & subleading jets must pass p_T, $|y|$ cuts

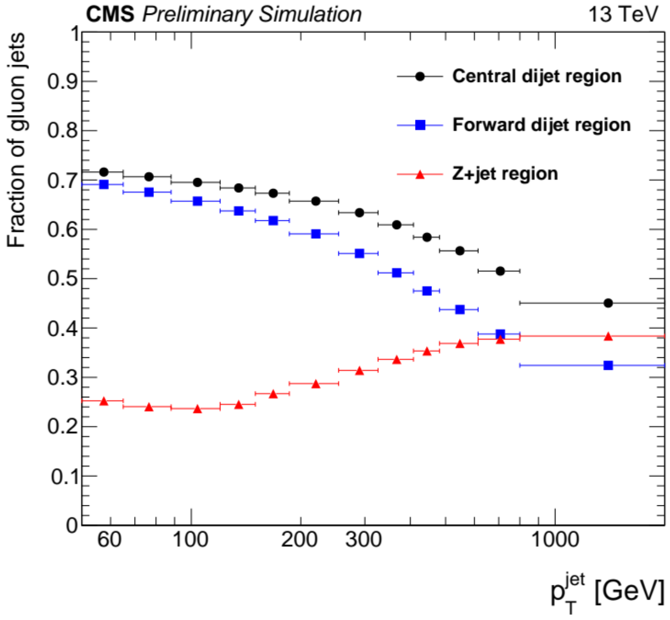
$\Delta\phi(J, J) > 2$



$\Delta p_T / \sum p_T < 0.3$
(avoids cutting on 3rd jet explicitly)

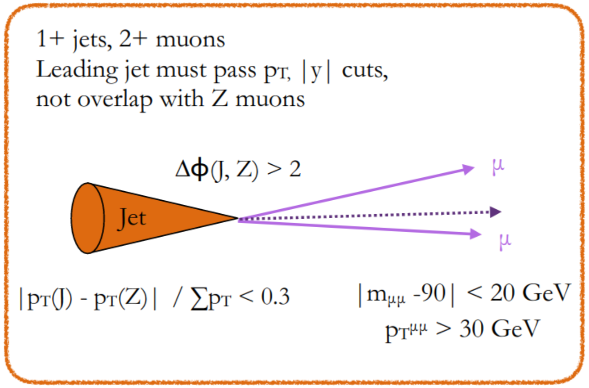
CMS-PAS-SMP-20-010

Quark- and gluon-enriched event samples

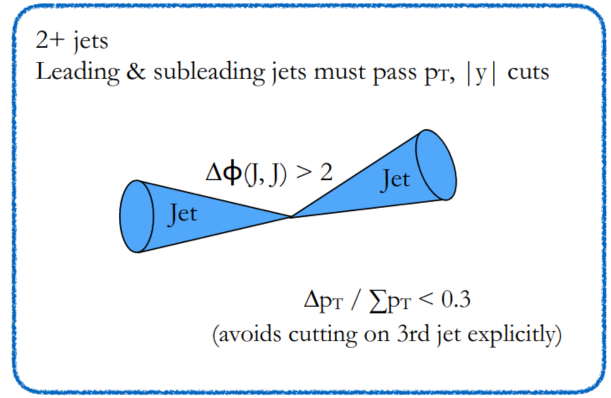


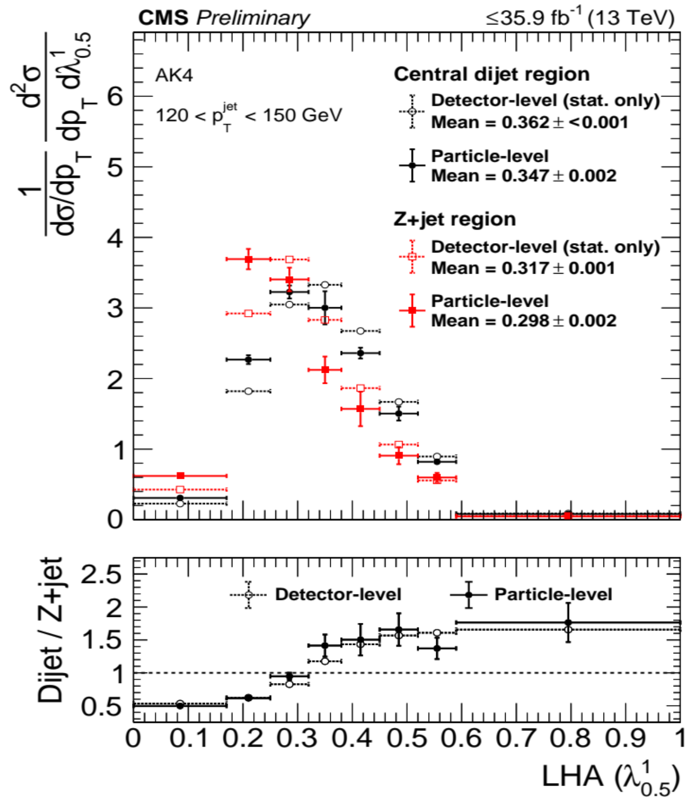
CMS-PAS-SMP-20-010

Quark-enriched jets from **Z(μμ) + jets** region:



Gluon-enriched jets from **dijet** region:





Small impact of detector response on ratio

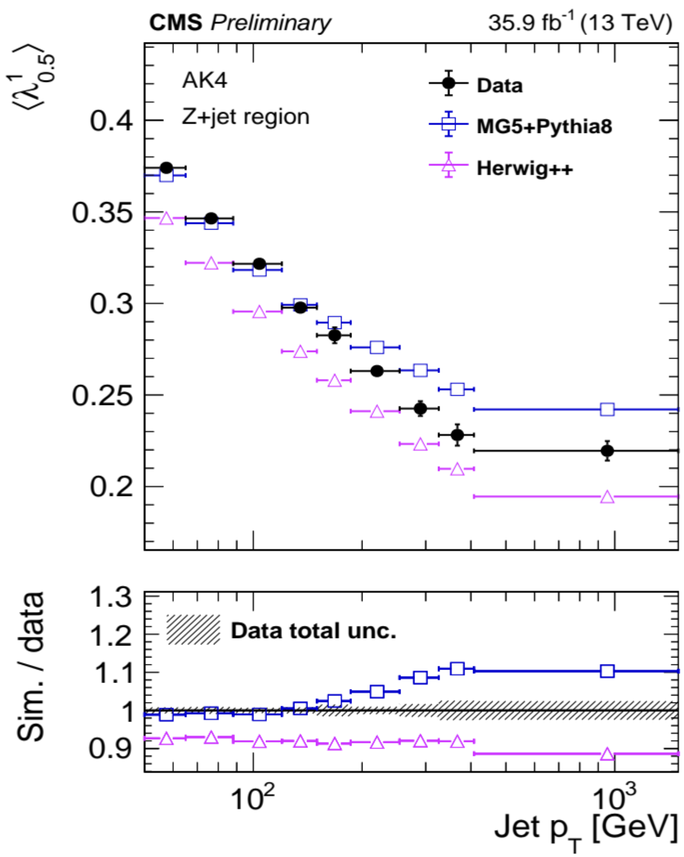
Statistical + systematic uncertainty ~ 5%

Radiation more spread out in gluon jet
→ Larger value of LHA

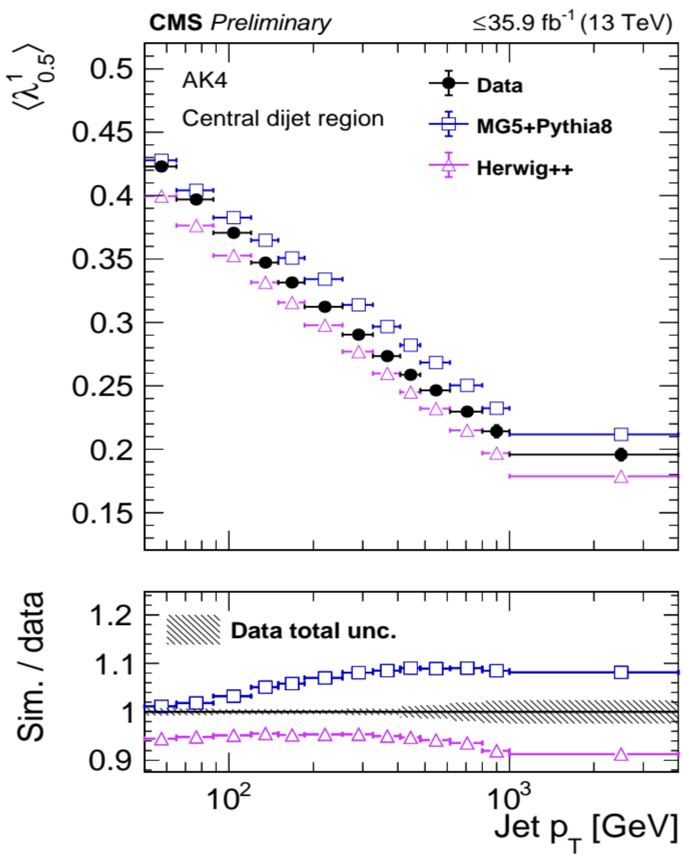
Differential distributions at particle-level

CMS-PAS-SMP-20-010

Quark-enriched sample



Gluon-enriched sample



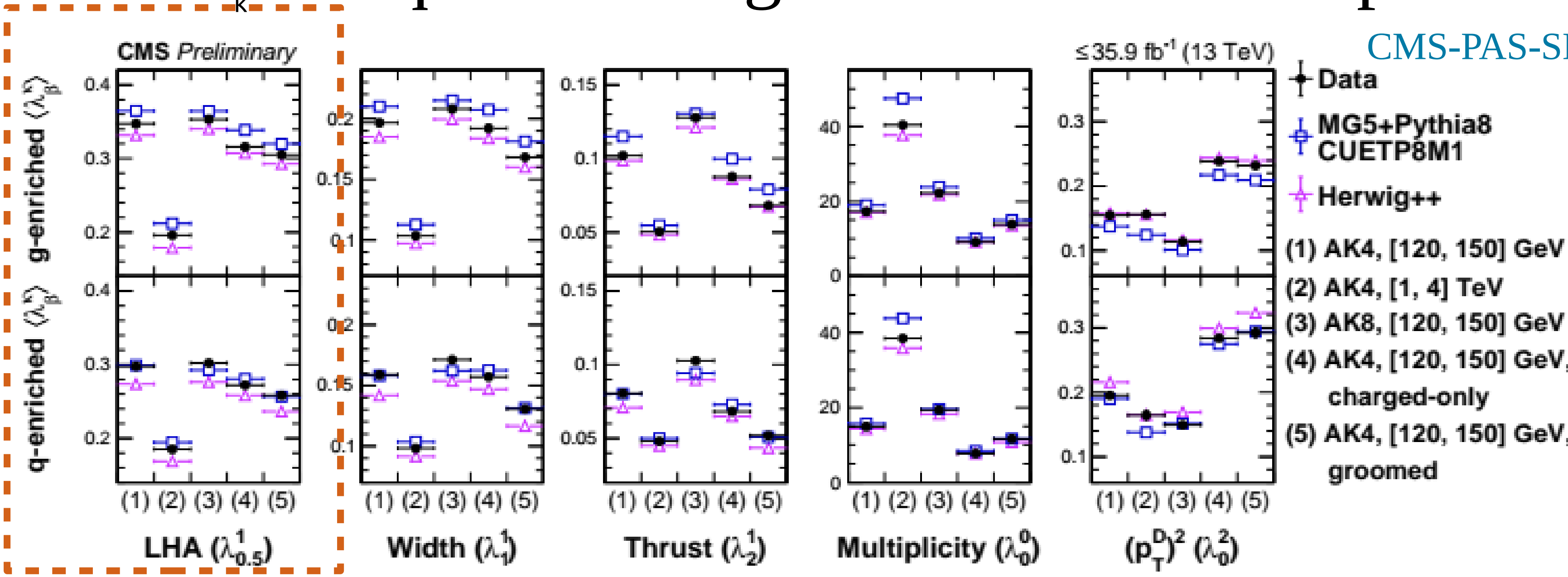
Good modeling of data by PYTHIA parton shower in quark-enriched samples

PYTHIA & HERWIG++ on opposite sides of data in gluon-enriched samples

HERWIG++ describes the p_T-dependence well but underestimates normalization

$<\lambda^{\beta}_{\kappa}>$ in quark- and gluon-enriched samples

CMS-PAS-SMP-20-010

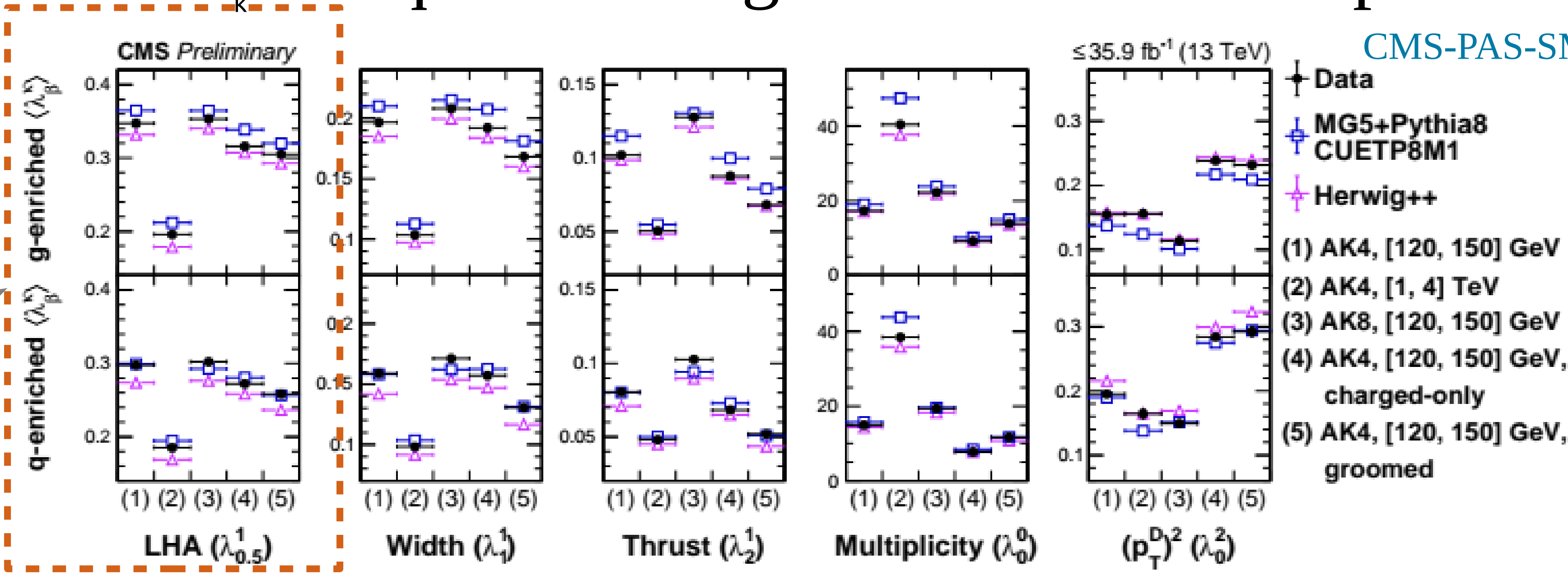


Measurements in

Event sample	Inputs from jet	Jet size	Grooming	Energy
Dijet (central)	All particles in jet	AK4	No grooming	In ranges of (ungrommed) jet p_T (50 GeV → 1 TeV)
Dijet (forward)	Charged particles only	AK8	Soft drop ($z_{cut} = 0.1$, $\beta = 0$)	
Z+ jet				

$<\lambda^{\beta}_{\kappa}>$ in quark- and gluon-enriched samples

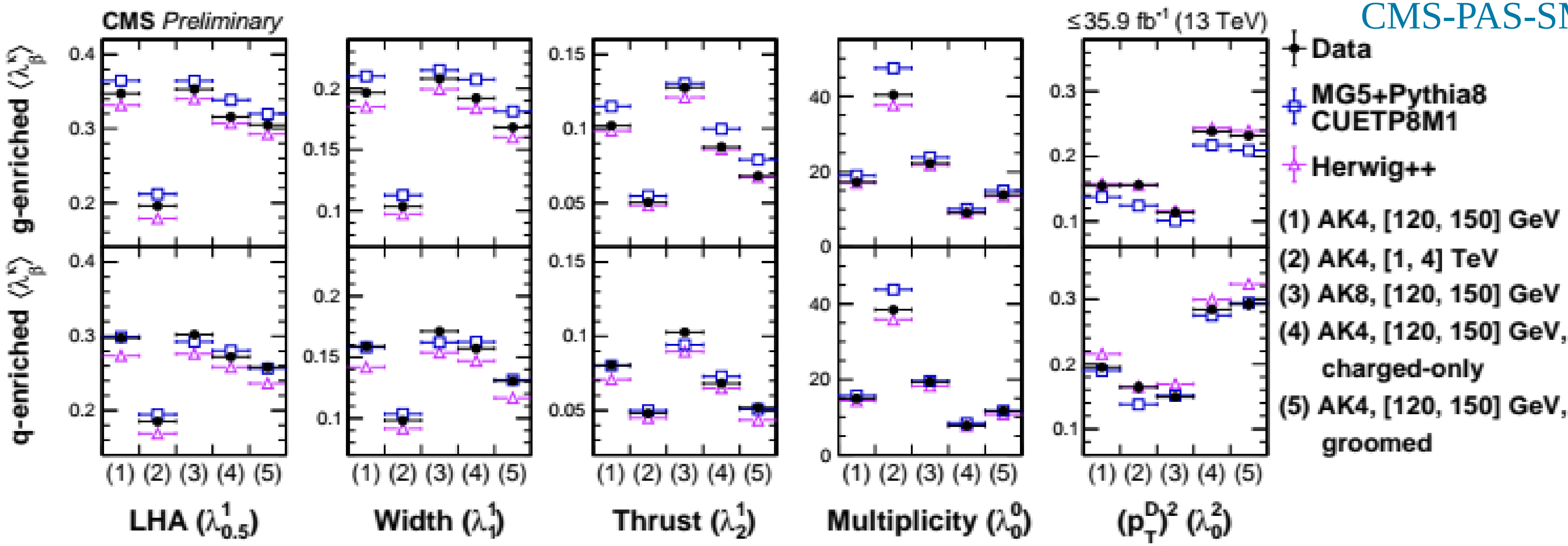
CMS-PAS-SMP-20-010

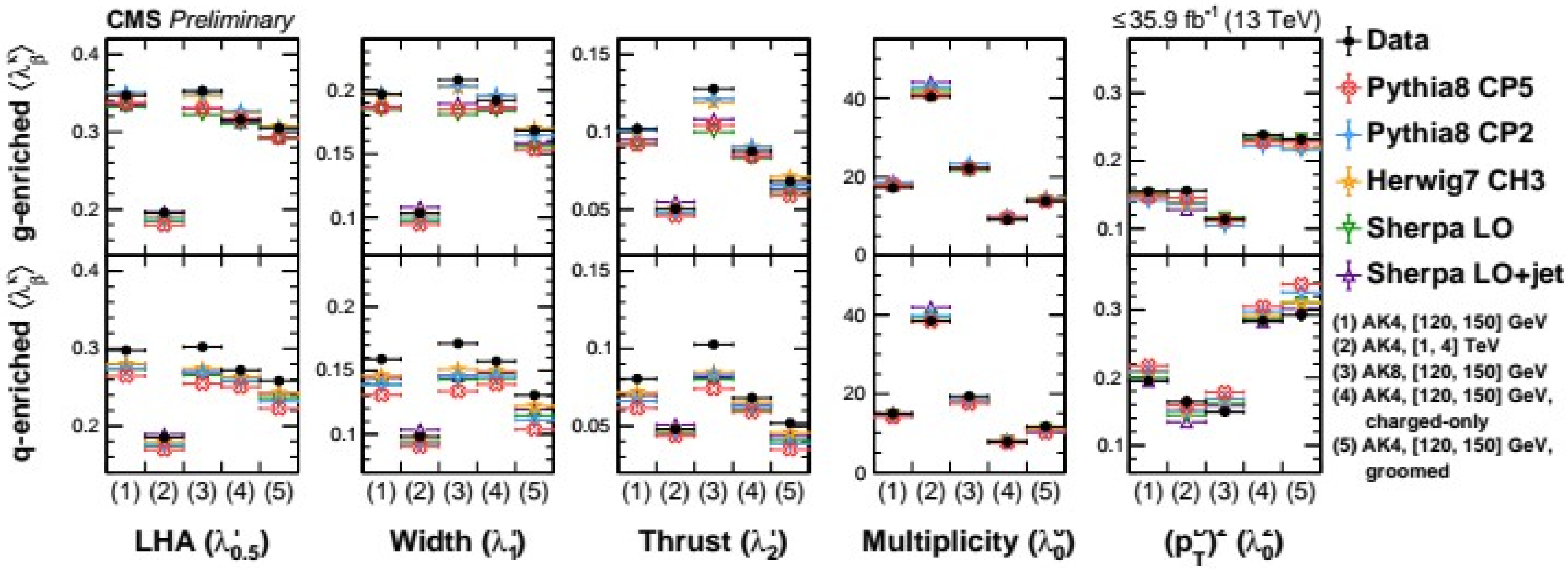


Event sample	Jet composition
Dijet (central) →	g-enriched
Dijet (forward) →	q-enriched at [1, 4] TeV
Z+ jet →	q-enriched at [120, 150] GeV

$<\lambda^{\beta}_{\kappa}>$ in quark- and gluon-enriched samples

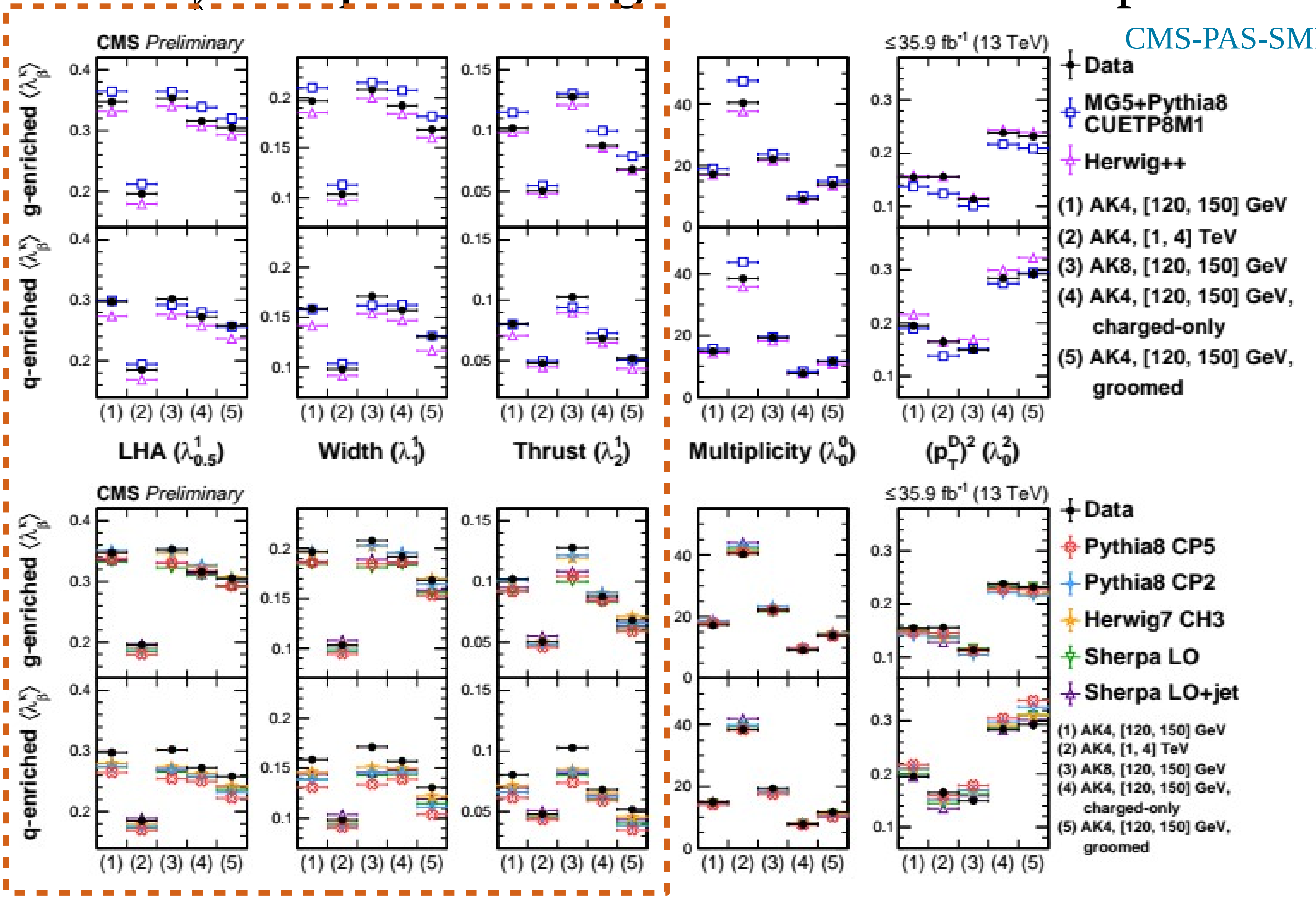
CMS-PAS-SMP-20-010





<λ^{β}_{κ}> in quark- and gluon-enriched samples

CMS-PAS-SMP-20-010

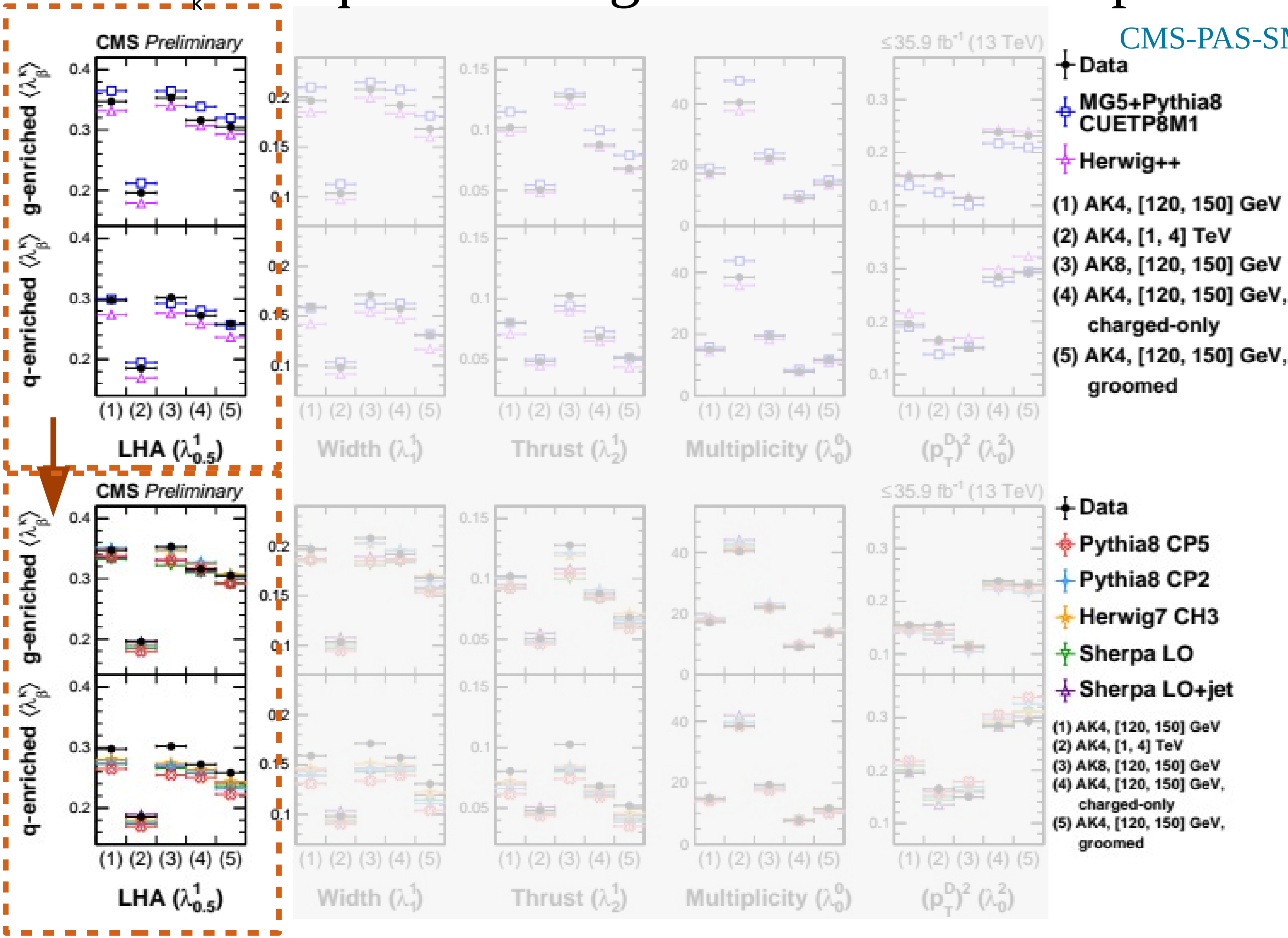


- IRC-safe variables not necessarily better modeled

=> Mismodeling by generators not only in non-perturbative region

$\langle\lambda^{\beta}_{\kappa}\rangle$ in quark- and gluon-enriched samples

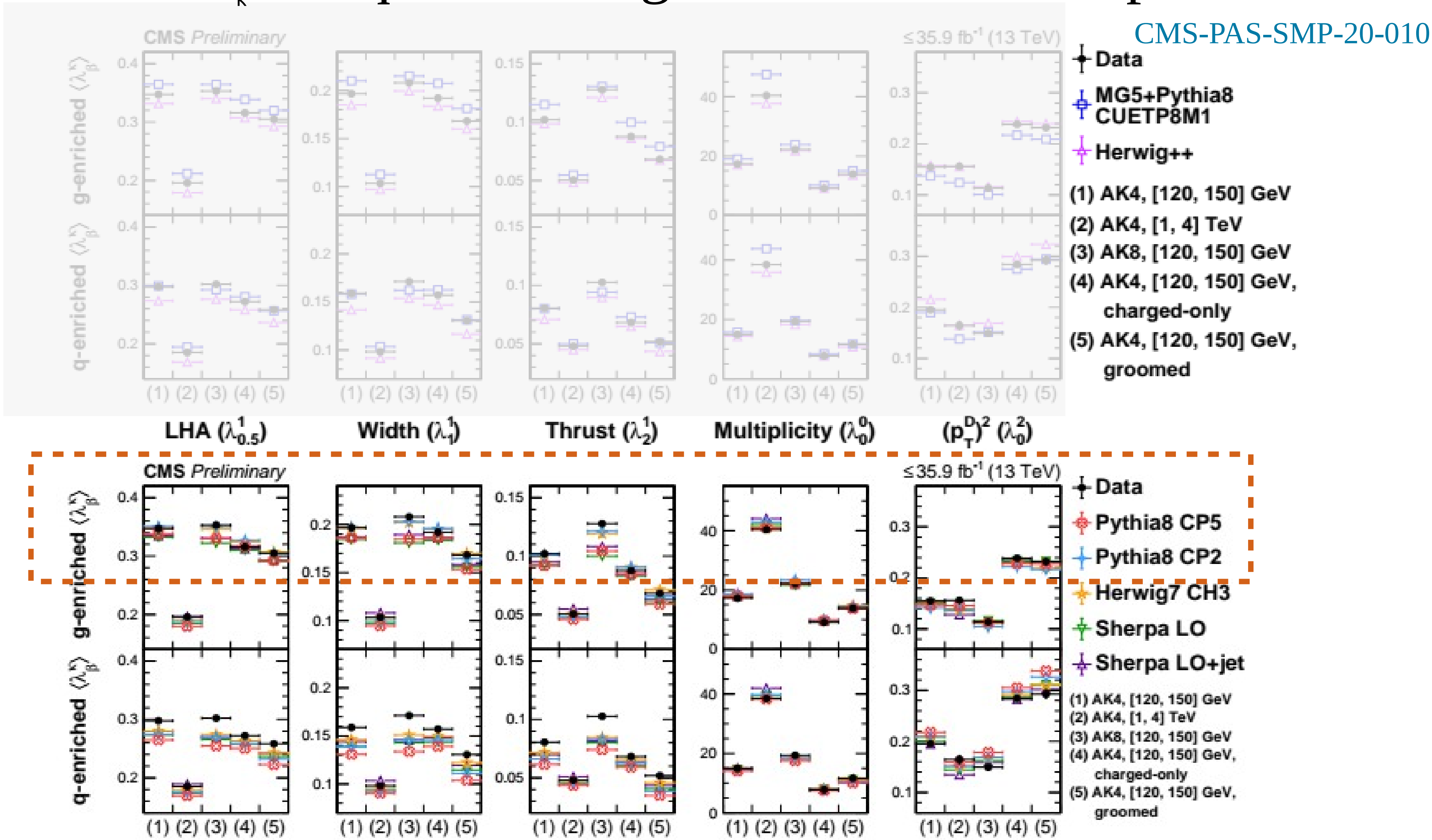
CMS-PAS-SMP-20-010



- Better modeling of gluon-enriched samples in modern generators + UE tunes

$<\lambda^{\beta}_{\kappa}>$ in quark- and gluon-enriched samples

CMS-PAS-SMP-20-010

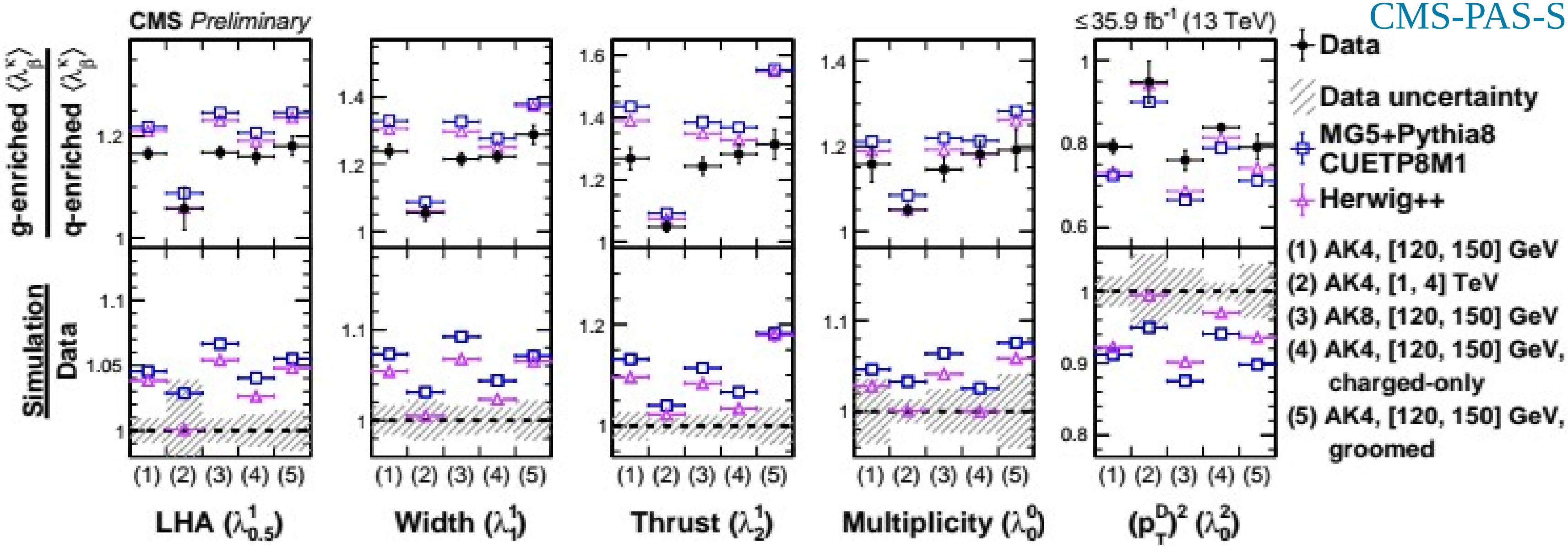


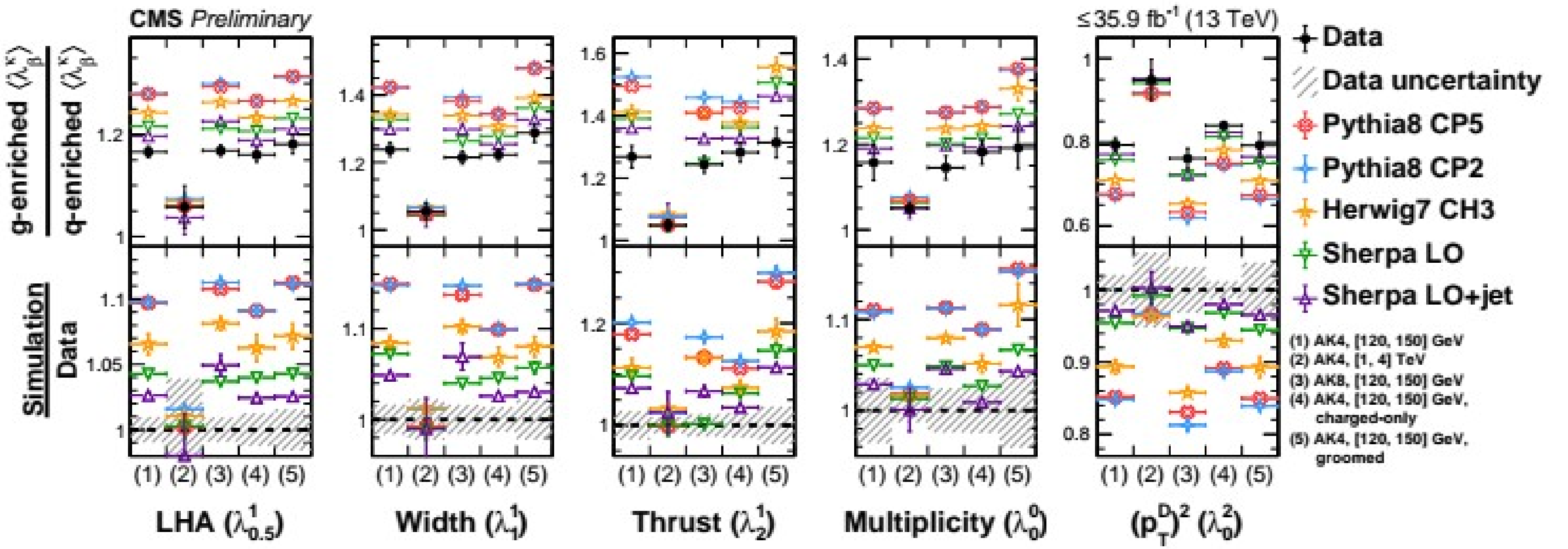
- CP2 tune describes gluon-enriched samples better than CP5 tune

(larger α_S & smaller color reconnection range in CP2)

Quark-gluon discrimination power in λ^{β}_{κ}

CMS-PAS-SMP-20-010



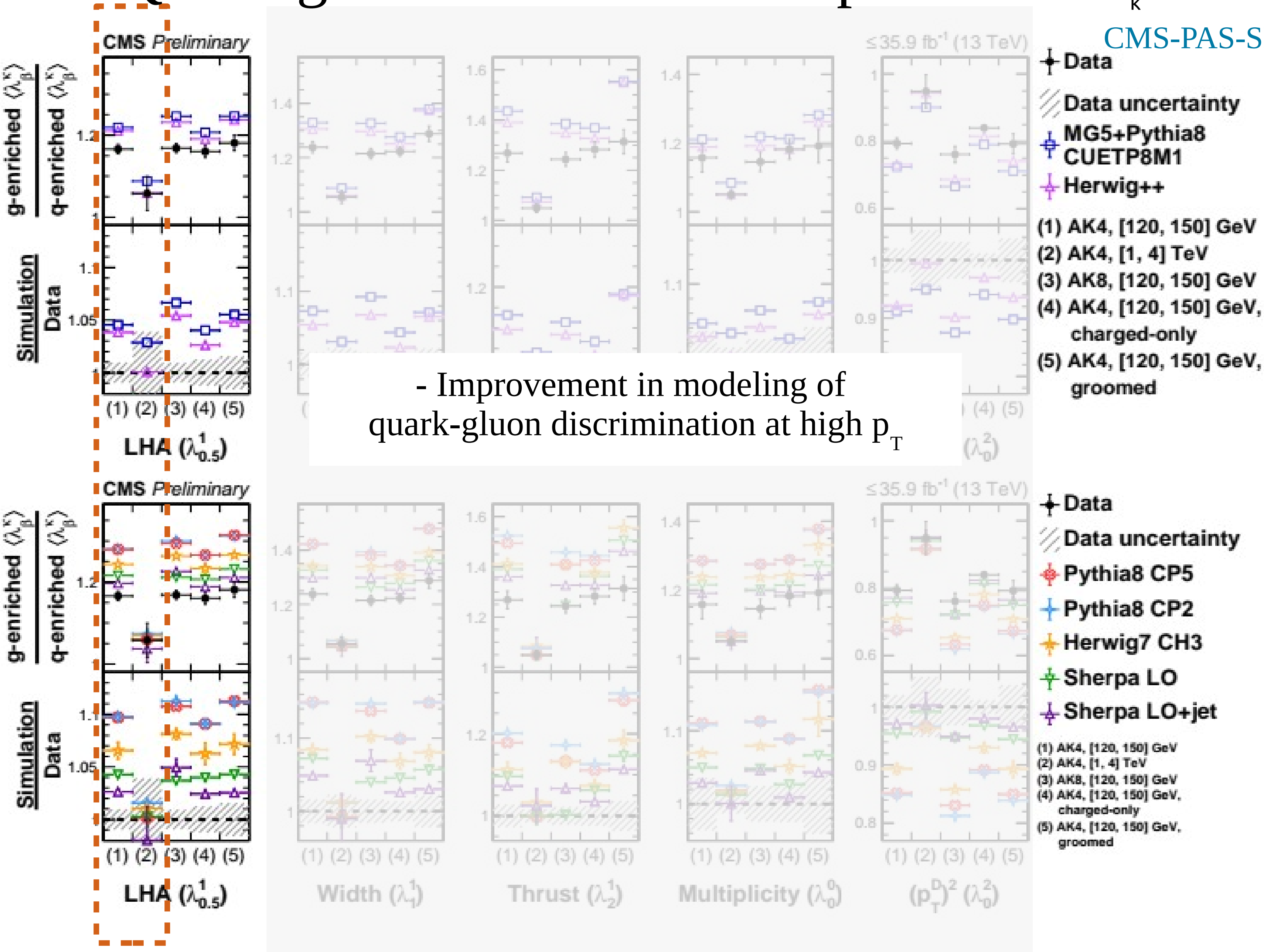


- Overestimation of quark-gluon discrimination by all the generators (except for λ^{2}_{0})

- Sherpa LO+jet simulation is the closest to data

CMS-PAS-SMP-20-010

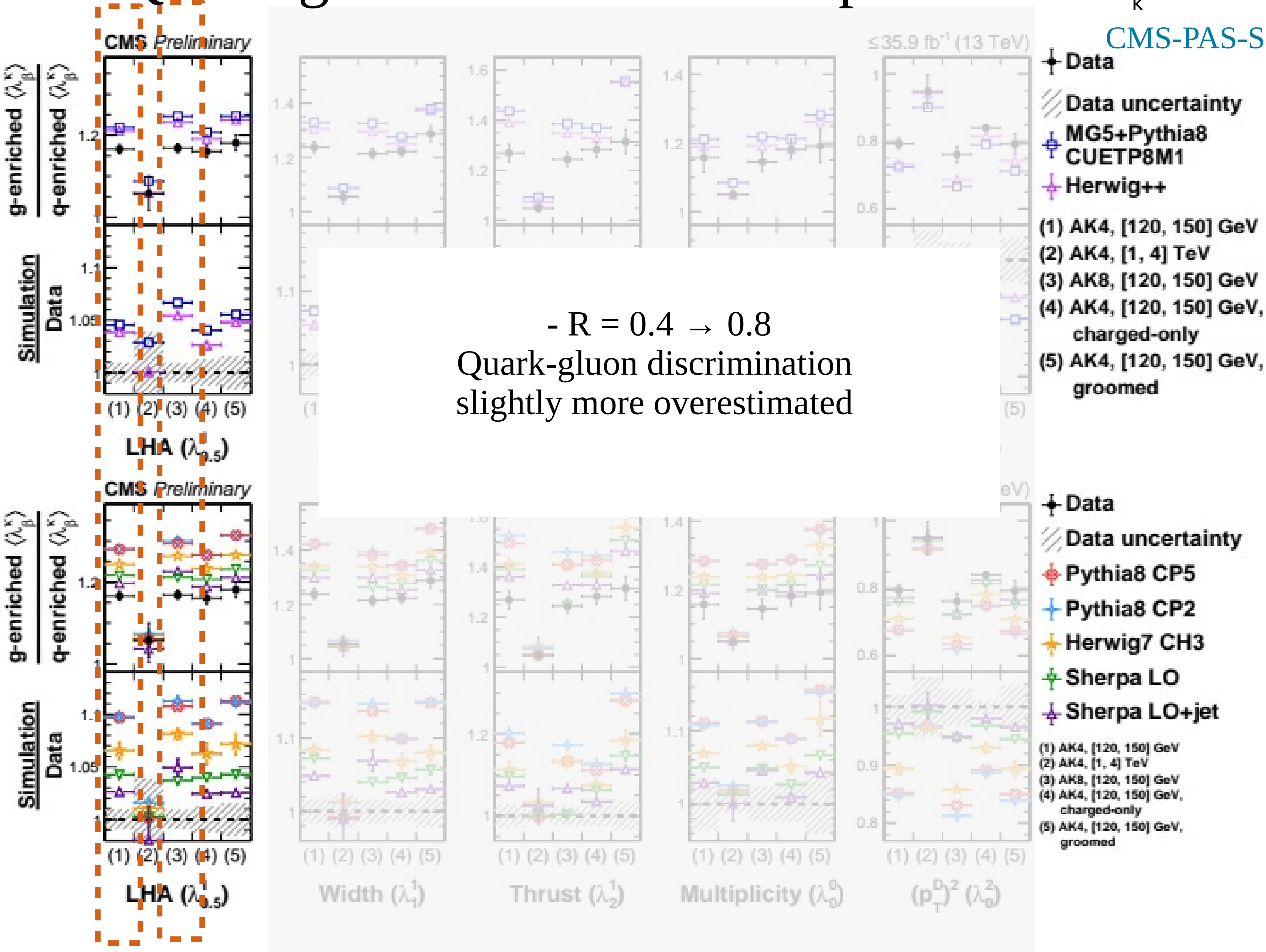
Quark-gluon discrimination power in λ^{β}_{κ}



- Improvement in modeling of quark-gluon discrimination at high p_T

Quark-gluon discrimination power in λ^{β}_{κ}

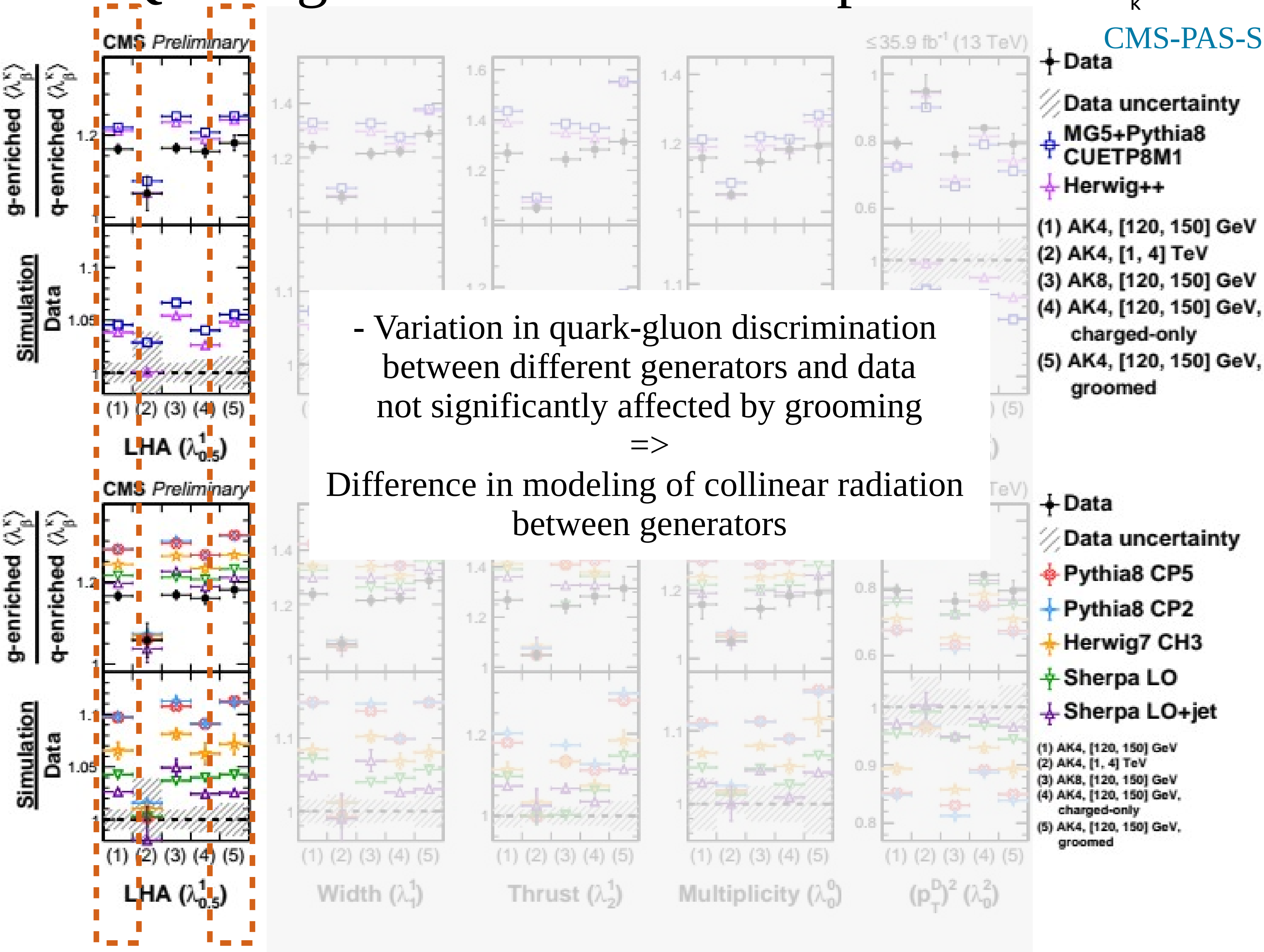
CMS-PAS-SMP-20-010



- R = 0.4 → 0.8
Quark-gluon discrimination slightly more overestimated

Quark-gluon discrimination power in λ^{β}_{κ}

CMS-PAS-SMP-20-010



- Variation in quark-gluon discrimination between different generators and data not significantly affected by grooming

=>

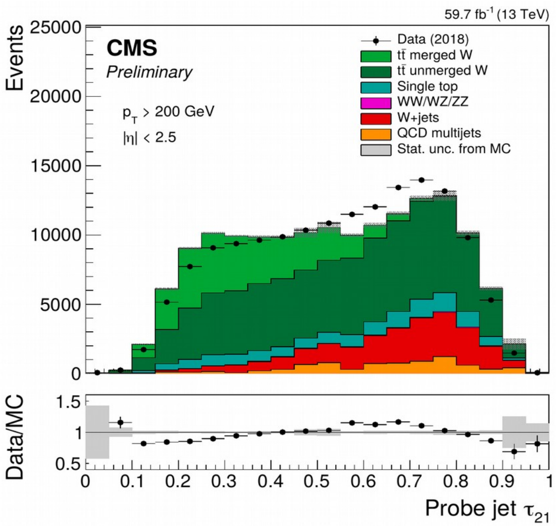
Difference in modeling of collinear radiation between generators

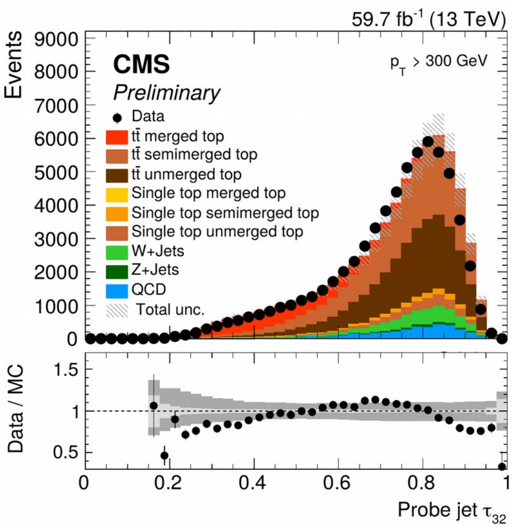
W & top jet identification with n-subjettiness ratio

$\tau_{n(n-1)} = \tau_n/\tau_{n-1}$ measures consistency of jet with n-prong structure

$\tau_{21} \rightarrow$ W tagging
$\tau_{32} \rightarrow$ top tagging

CMS-DP-20-025





- Measurements in muon + jets events (dominated by semileptonic $t\bar{t}$)

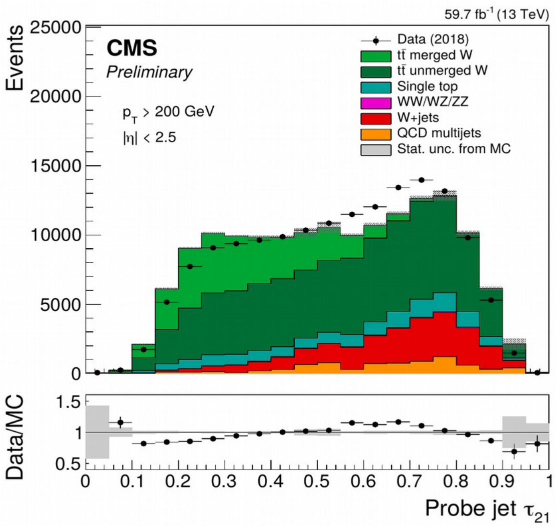
- Room for improvement of description by simulation

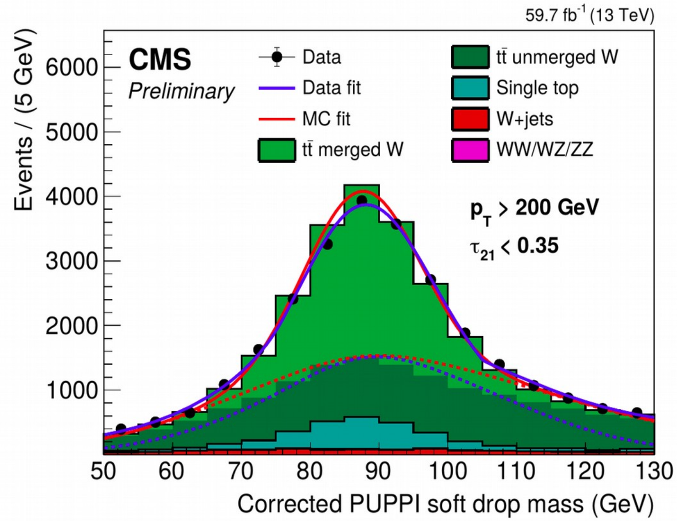
W & top jet identification with n-subjettiness ratio

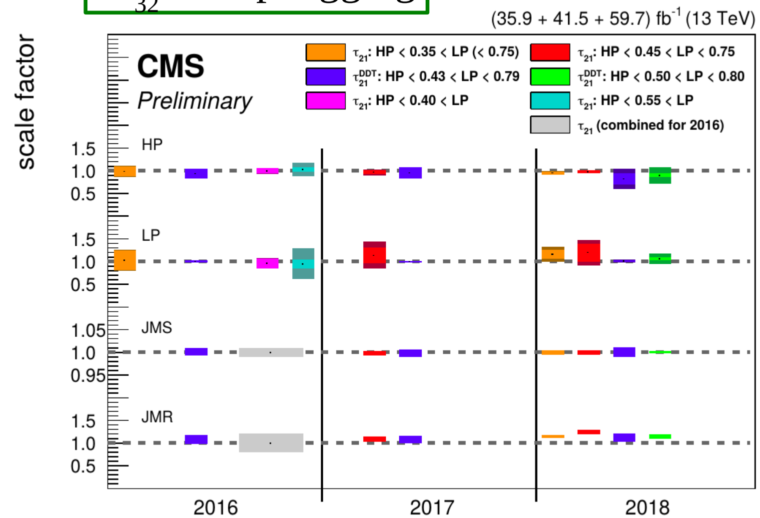
$\tau_{n(n-1)} = \tau_n / \tau_{n-1}$ measures consistency of jet with n-prong structure

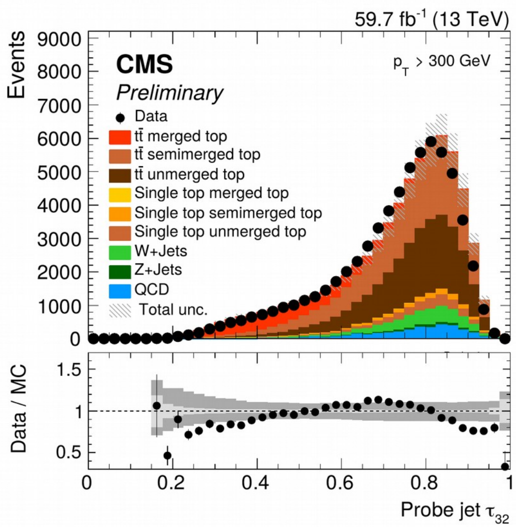
$\tau_{21} \rightarrow$ W tagging
$\tau_{32} \rightarrow$ top tagging

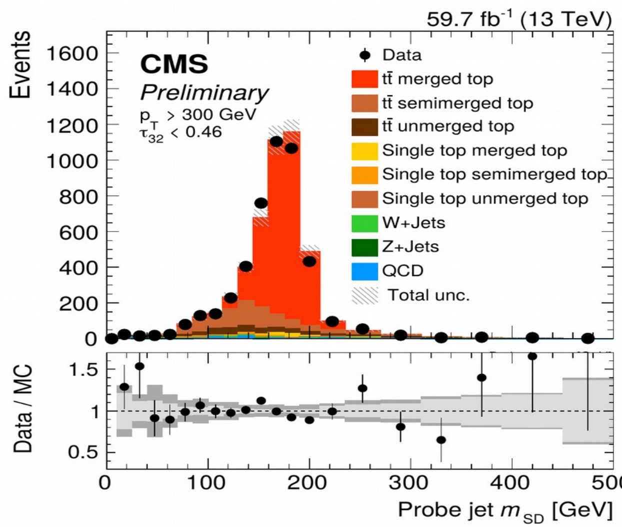
CMS-DP-20-025

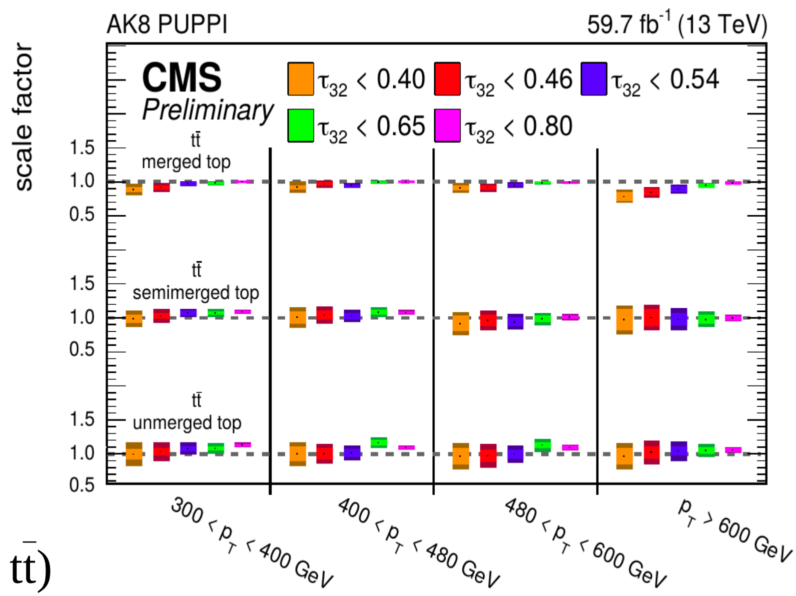










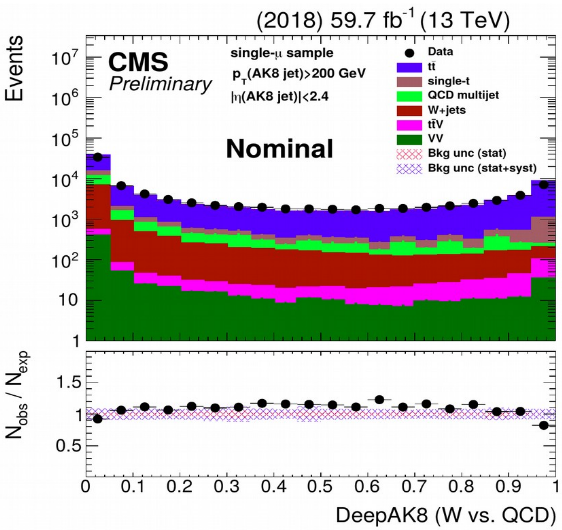


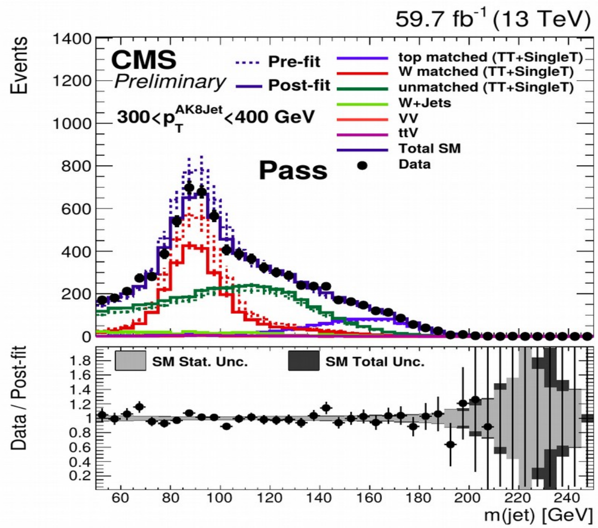
- Measurements in muon + jets events (dominated by semileptonic $t\bar{t}$)
- Room for improvement of description by simulation
- Tagging efficiencies (measured using tag-and-probe method) are similar in data & simulation

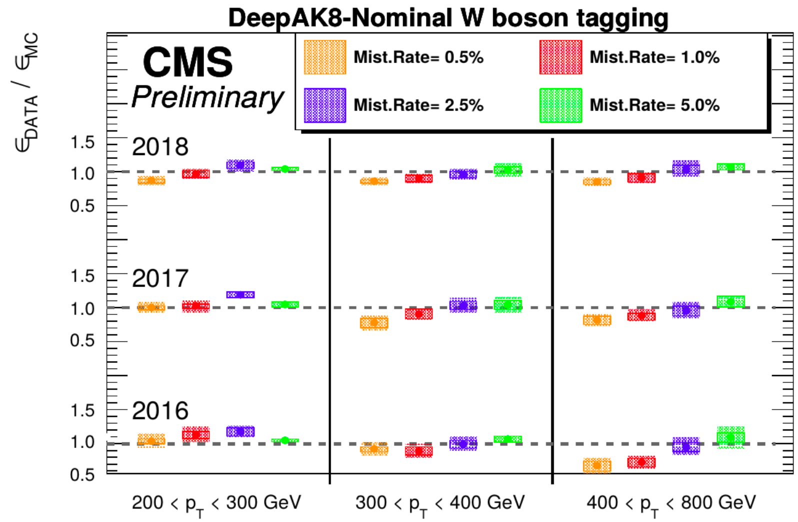
W & top jet identification with DeepAK8 tagger

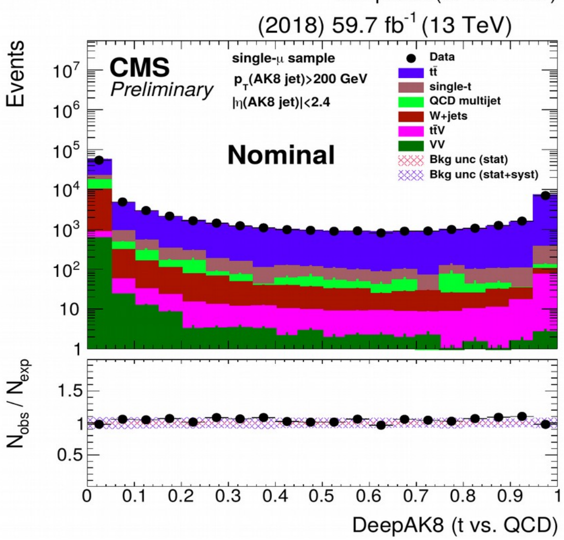
DNN-based tagger using (almost) all the particles & secondary vertices in AK8 jet

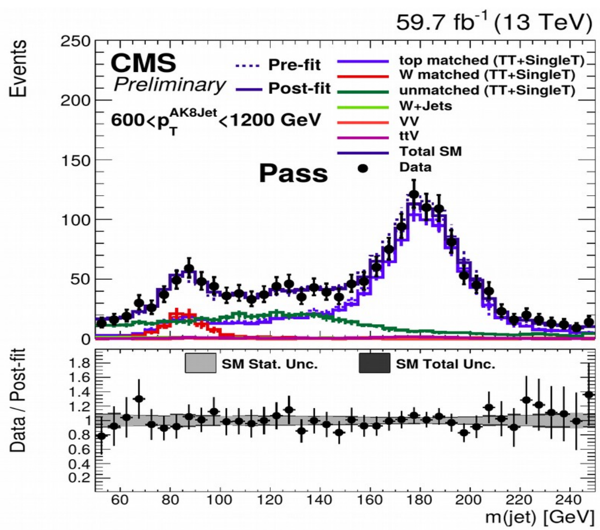
CMS-DP-20-025

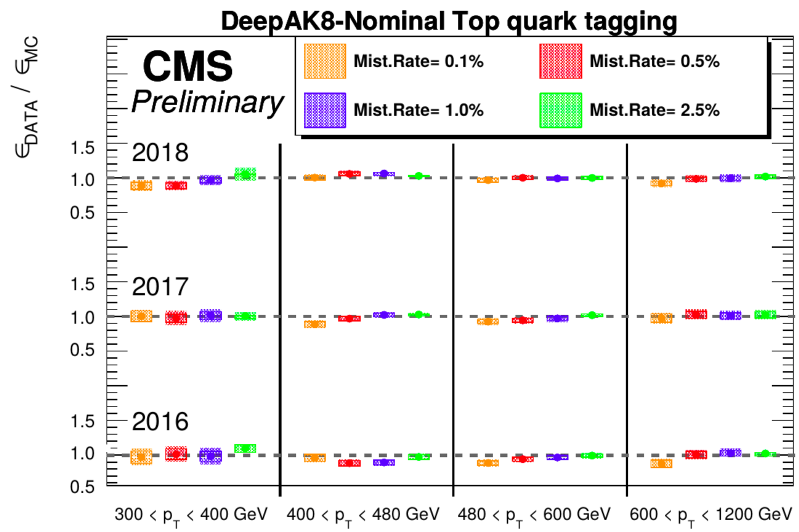












- Measurements in muon + jets events (dominated by semileptonic $t\bar{t}$)
- Moderately good description by simulation

See applications in Anna Benecke's talk

Details of DeepAK8 tagger in JINST 15 (2020) P06005

- Tagging efficiencies (measured using tag-and-probe method) are similar in data & simulation

Summary and Outlook

- A detailed study of jet substructure observables is performed

Useful for

Better modeling of

Parton shower

Non-perturbative effects (hadronization, UE, ...)

pQCD calculations

Fixed order + resummation

Handle to reduce modeling uncertainty of particle identifiers exploiting jet substructure → Improve measurements & searches

- Comparison with analytic calculation in progress

  [JHEP 07 (2021) 076]

- Calibration measurement performed for heavy particle taggers

- More measurements are coming soon!

Summary and Outlook

- A detailed study of jet substructure observables is performed

Useful for

Better modeling of

Parton shower

Non-perturbative effects (hadronization, UE, ...)

pQCD calculations

Fixed order + resummation

Handle to reduce modeling uncertainty of particle identifiers exploiting jet substructure → Improve measurements & searches

- Comparison with analytic calculation in progress

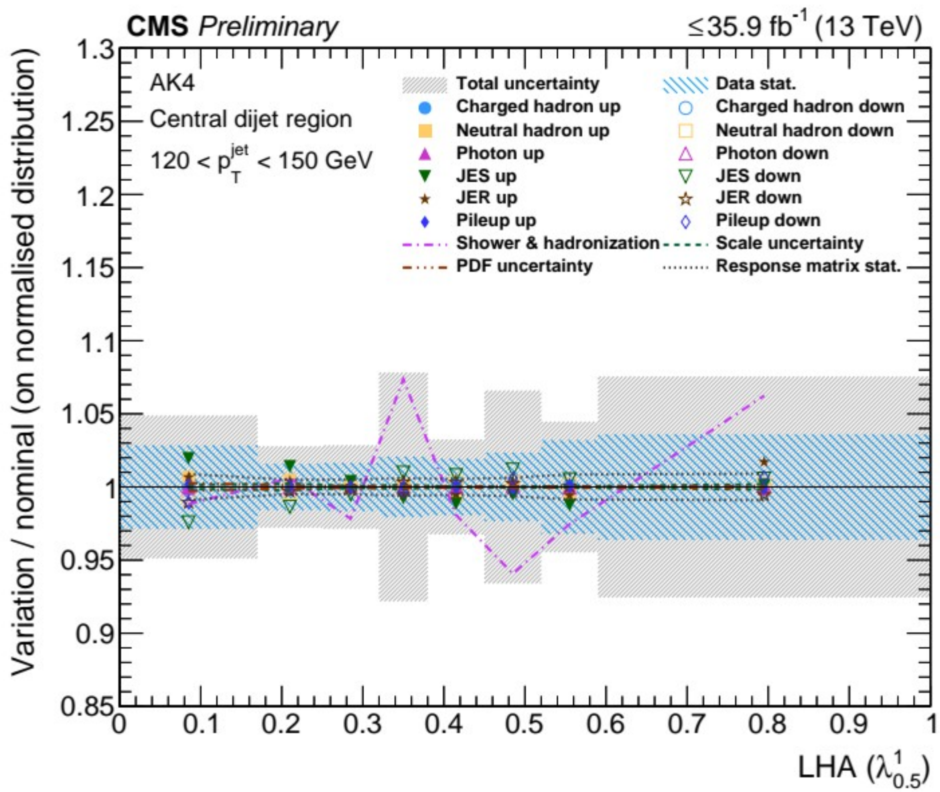
  [JHEP 07 (2021) 076]

- Calibration measurement performed for heavy particle taggers

- More measurements are coming soon!

Extra material

Uncertainties in LHA



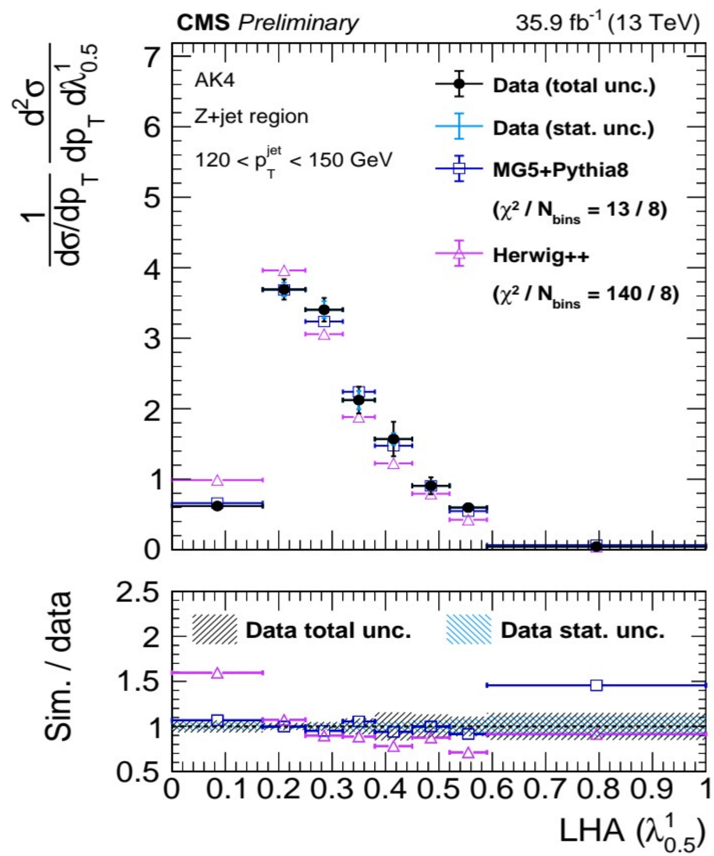
Showering+hadronization uncertainty is the dominant among all the systematic uncertainty sources

Differences of the unfolded distributions derived using response matrices constructed from MG5+Pythia8 and Herwig++

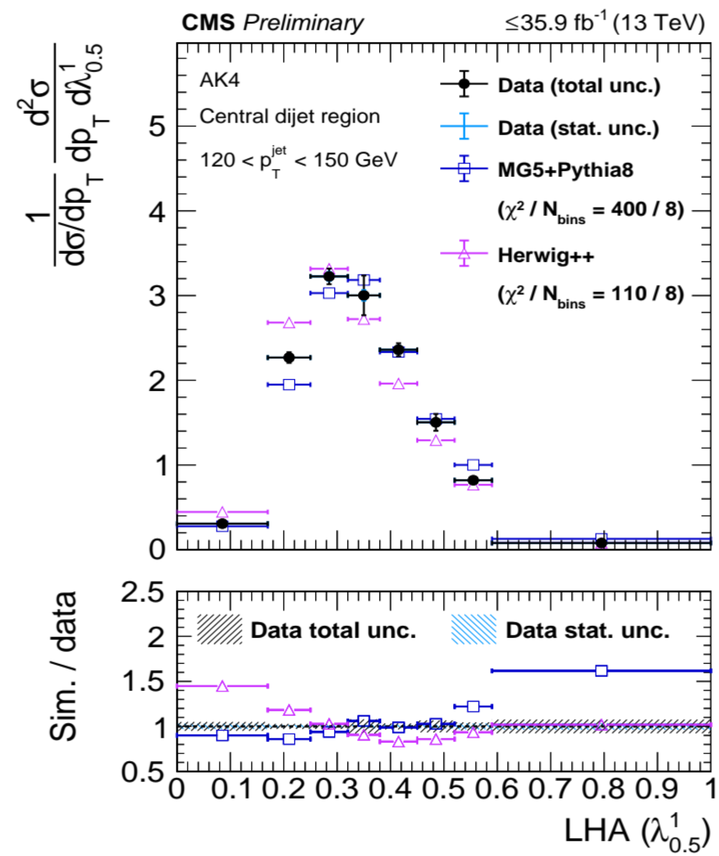
Differential distributions at particle-level

CMS-PAS-SMP-20-010

Quark-enriched sample



Gluon-enriched sample



Good modeling of data by PYTHIA parton shower in quark-enriched samples

PYTHIA & HERWIG++ on opposite sides of data in gluon-enriched samples

Top tagging efficiency

CMS-DP-20-025

Tagger	Working point	Signal efficiency [%]	Background efficiency [%]
AK8 PUPPI	$\tau_{32} < 0.40$	17	0.2
	$\tau_{32} < 0.46$	26	0.5
	$\tau_{32} < 0.54$	37	1.7
	$\tau_{32} < 0.65$	49	5.1
	$\tau_{32} < 0.80$	62	15.9
AK8 PUPPI + subjet btag	$\tau_{32} < 0.40$	16	0.1
	$\tau_{32} < 0.46$	23	0.3
	$\tau_{32} < 0.54$	33	0.6
	$\tau_{32} < 0.65$	43	1.8
	$\tau_{32} < 0.80$	53	5.3
HOTVR PUPPI	$\tau_{32} < 0.56$	37	2.6
DeepAK8	Mistag rate 0.1%	37	0.1
	Mistag rate 0.5%	52	0.5
	Mistag rate 1.0%	59	1.0
	Mistag rate 2.5%	67	2.5
DeepAK8 MD	Mistag rate 0.1%	28	0.1
	Mistag rate 0.5%	48	0.5
	Mistag rate 1.0%	57	1.0
	Mistag rate 2.5%	66	2.5

Table 1: The top tagging efficiencies including the respective mass windows are estimated from simulation in 2018 before the template fit for the AK8 PUPPI, HOTVR and DeepAK8 taggers. The efficiencies for signal are measured in $t\bar{t}$ events where the angular distance between the generated top and the probe jet is $\Delta R < 0.6$. Background efficiencies are measured using QCD multijet events. Only events are considered where the probe jet fulfills $480 < p_T < 600$ GeV. The exact values of efficiencies strongly depend on the selection applied. Therefore, the purpose of the presented numbers is to give a rough estimate and comparison between algorithms.

W tagging efficiency

CMS-DP-20-025

Tagger	Working point	Signal efficiency [%]	Background efficiency [%]
AK8 PUPPI	$\tau_{21} < 0.35$	53	2.0
	$\tau_{21} < 0.45$	69	4.7
	$\tau_{21}^{\mathrm{DDT}} < 0.43$	12	0.1
	$\tau_{21}^{\mathrm{DDT}} < 0.50$	32	0.6
DeepAK8	Mistag rate 0.5%	51	0.5
	Mistag rate 1.0%	62	1.0
	Mistag rate 2.5%	74	2.5
	Mistag rate 5.0%	79	5.0
DeepAK8 MD	Mistag rate 0.5%	38	0.5
	Mistag rate 1.0%	50	1.0
	Mistag rate 2.5%	66	2.5
	Mistag rate 5.0%	76	5.0

Table 2: The W tagging efficiencies including the respective mass windows are estimated from simulation in 2018 for the working points of the AK8 PUPPI and DeepAK8 taggers. The efficiencies for signal are measured in $\mathrm{Z}' \to \mathrm{WW}$ events where the angular distance between generated W and the probe jet is $\Delta R < 0.6$. Background efficiencies are measured using QCD multijet events. Only events are considered where the probe jet fulfills $300 < p_{\mathrm{T}} < 500$ GeV. The exact values of efficiencies strongly depend on the selection applied. Therefore, the purpose of the presented numbers is to give a rough estimate and comparison between algorithms.