Jet substructure in VBS measurements by CMS and ATLAS experiments

Nurfikri Norjoharuddeen

nurfikri.bin.norjoharuddeen@cern.ch

On behalf of CMS & ATLAS collaborations

COMETA 1st General Meeting

University of Izmir Bakırçay, İzmir, Türkiye

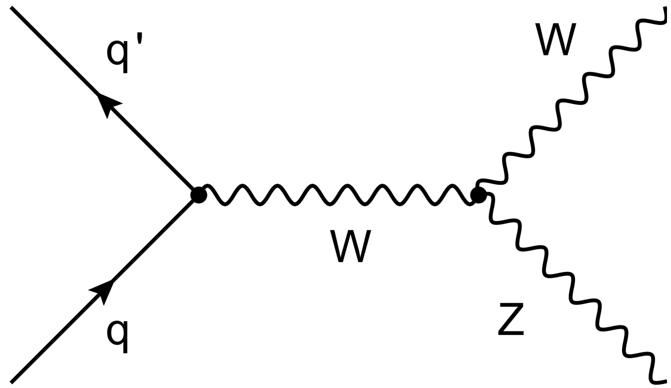
28 February 2024

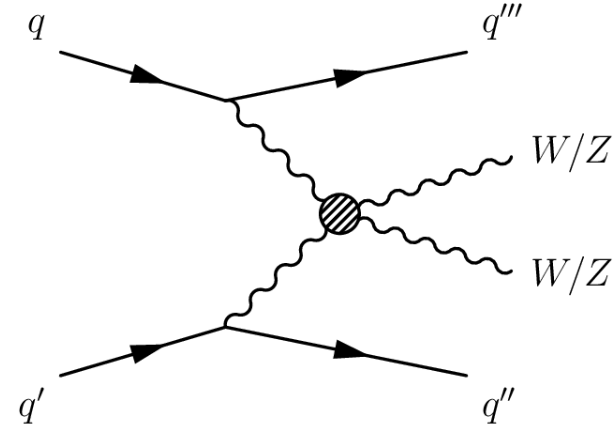
Outline

- Why jet substructure for Vector Boson Scattering (VBS) measurements?
- Focus on the jet substructure techniques developed by CMS and ATLAS.
- Highlight CMS & ATLAS VBS Run 2 measurements using jet substructure techniques.
- Recent state-of-the-art techniques by CMS & ATLAS.

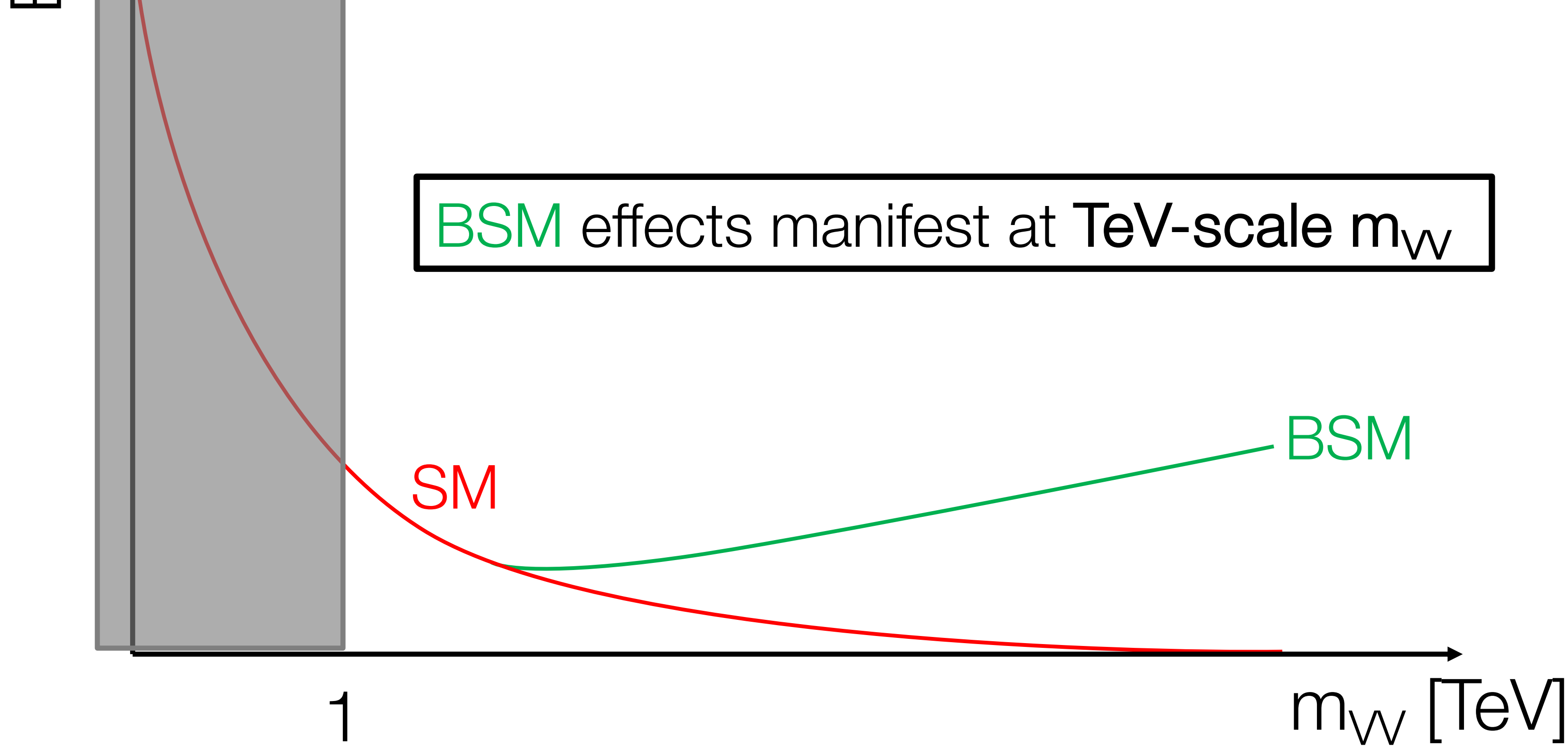
Why jet substructure for VBS analyses?

High-mass m_{VV} measurement





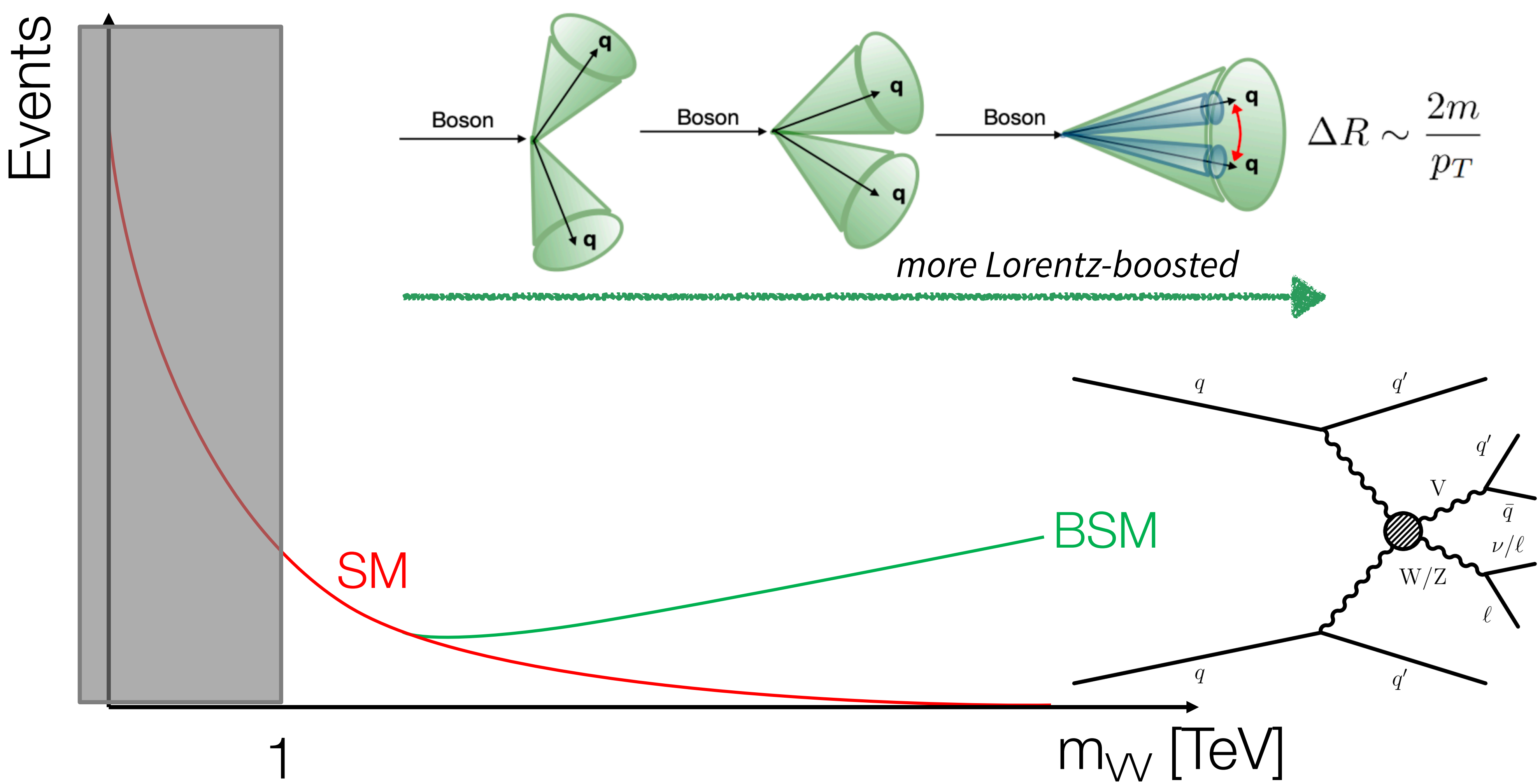
BSM effects manifest at TeV-scale m_{VV}



High-mass m_{VV} measurement

TeV-scale m_{VV}

"boosted" boson decays -> collimated decay products



Additional challenges in VBS analyses

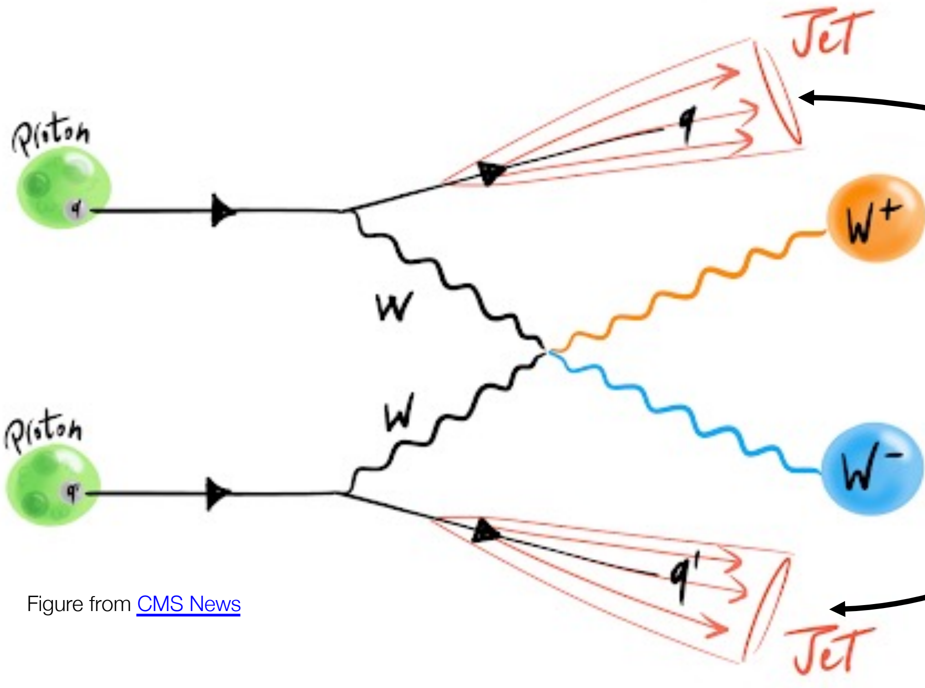


Figure from [CMS News]

2 Forward Jets

Pile-up Jet Suppression

Quark vs Gluon Discrimination

Jet substructure techniques

Jet Algorithms at the LHC

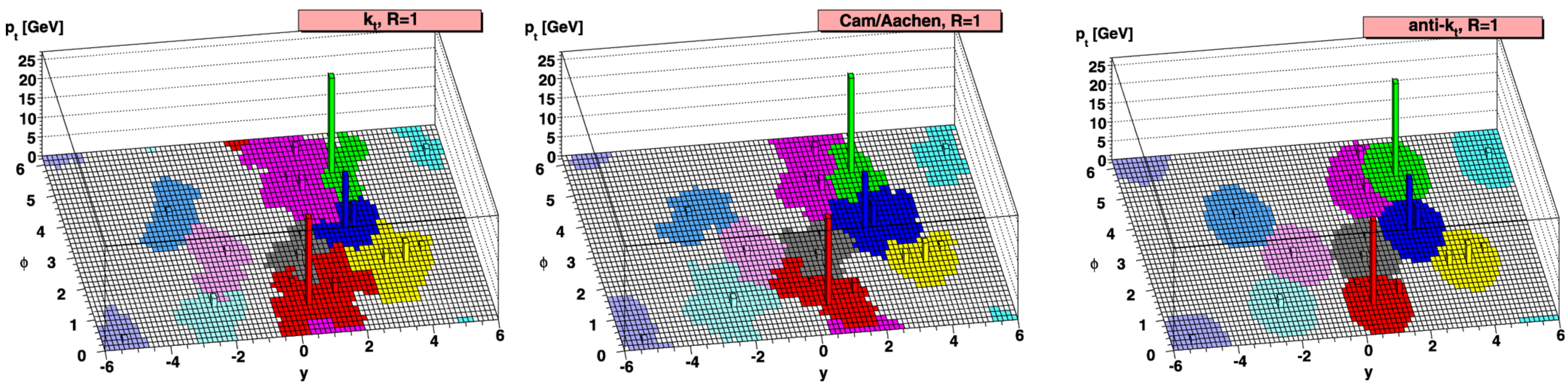
EPJ C 67, 637–686 (2010)

Sequential clustering algorithms with distance parameter R

1) k_T
2) Cambridge-Aachen (CA)
3) anti-k_T (AK)

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = p_{ti}^{2p},$$



"Standard" jet algorithms by experiments

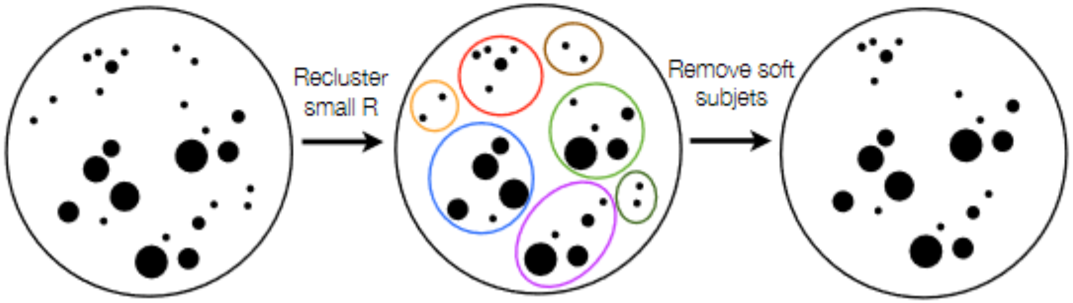
- Small-R jets: anti-k_T R = **0.4 [AK4]**
- Large-R jets: anti-k_T R = 0.8 **[AK8]** for CMS, 1.0 **[AK10]** for ATLAS

"Non-standard" jets used also e.g Variable-R jets
JINST 15 (2020) P06005
ATL-PHYS-PUB-2017-010

JHEP 1002 (2010) 084
Eur. Phys. J. C 76 (2016) 154

ATLAS

Trimming



1. Recluster with k_t algorithm with R = 0.2.
2. Remove subjets with p_T fraction < 5%.

Figures by J. Dolen

CMS

Soft Drop

1. Recluster with CA algorithm.
2. Reverse clustering history.
3. Check criterion:

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

$z_{cut} = 0.1$

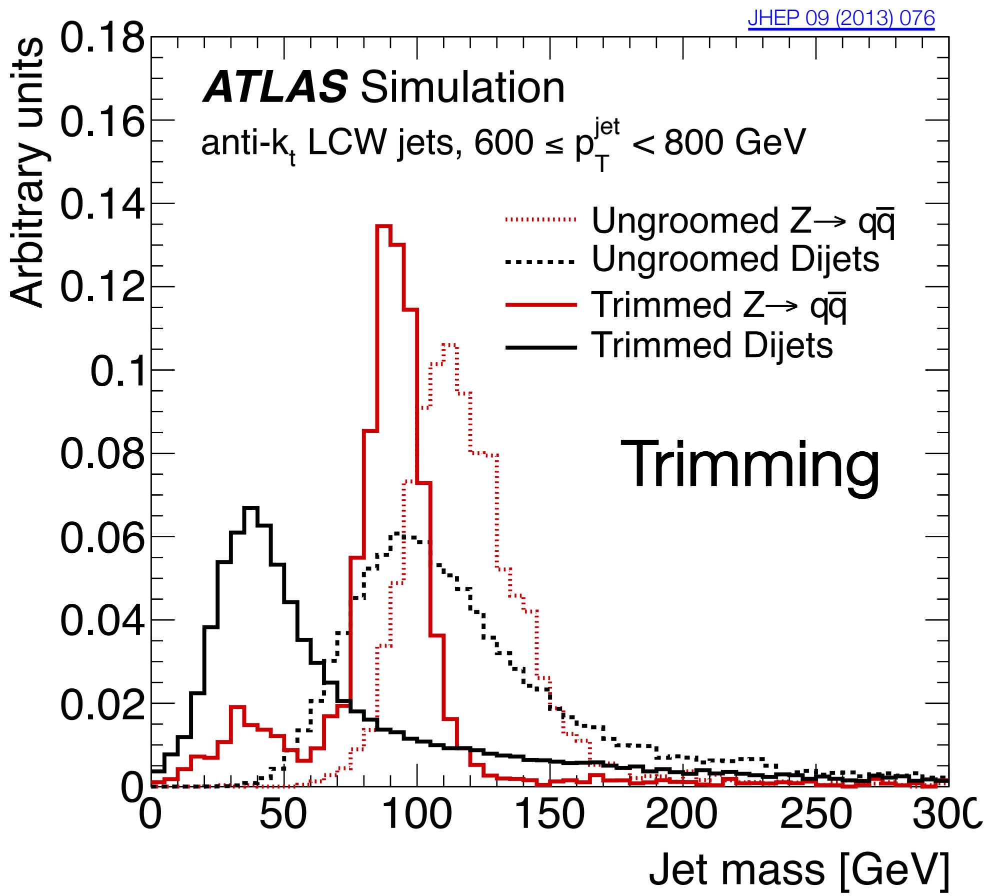
$\beta = 0$

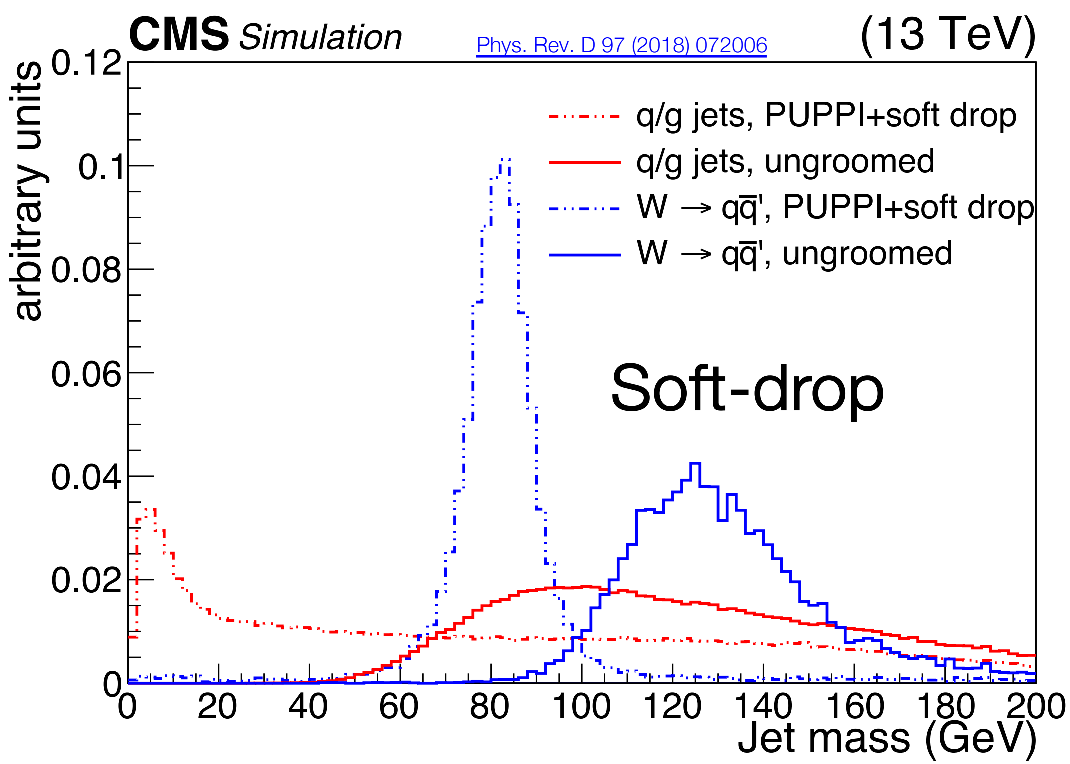
4. Pass: two subjets are final.
   Fail: remove sub-leading subjet & repeat (1).

JHEP 1405 (2014) 146
JINST 15 (2020) P06005

(Large-R) Jet Grooming

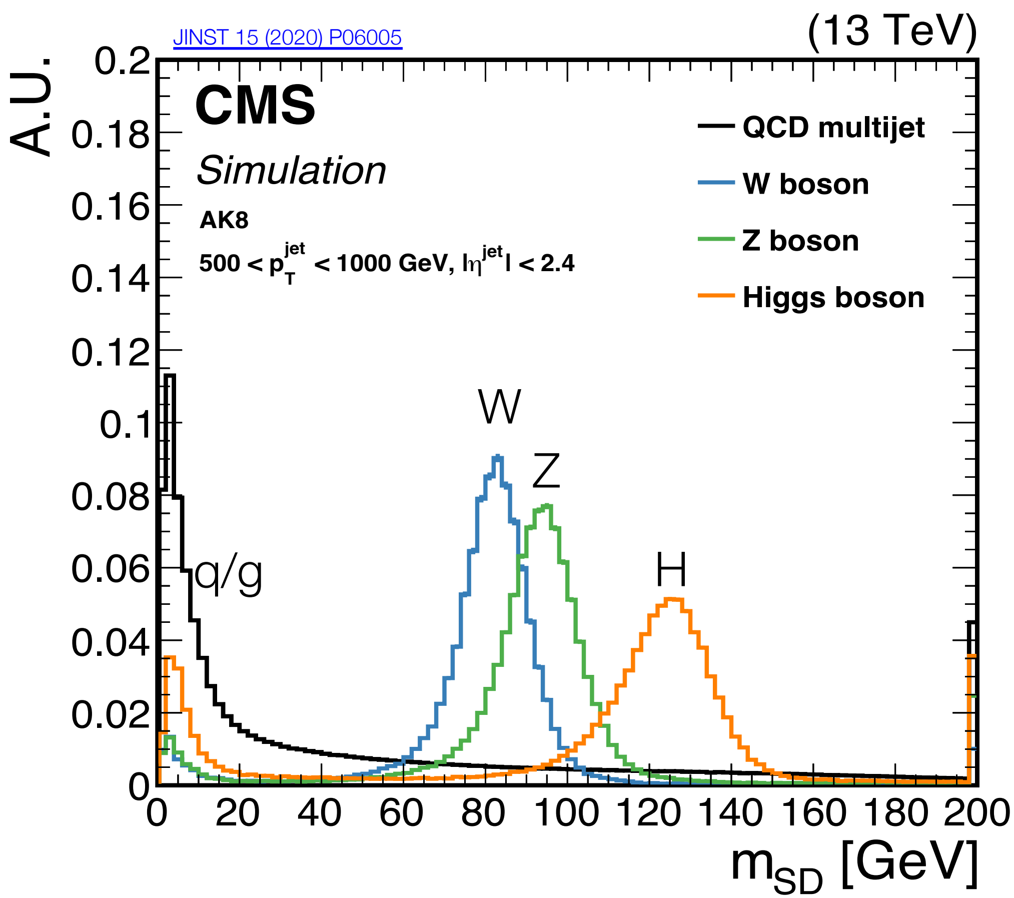
Jet grooming (+ Pileup Mitigation) “cleans” up large-R jets





Simplest observable

Large-R jet mass



CMS

Soft-Drop Mass

ATLAS

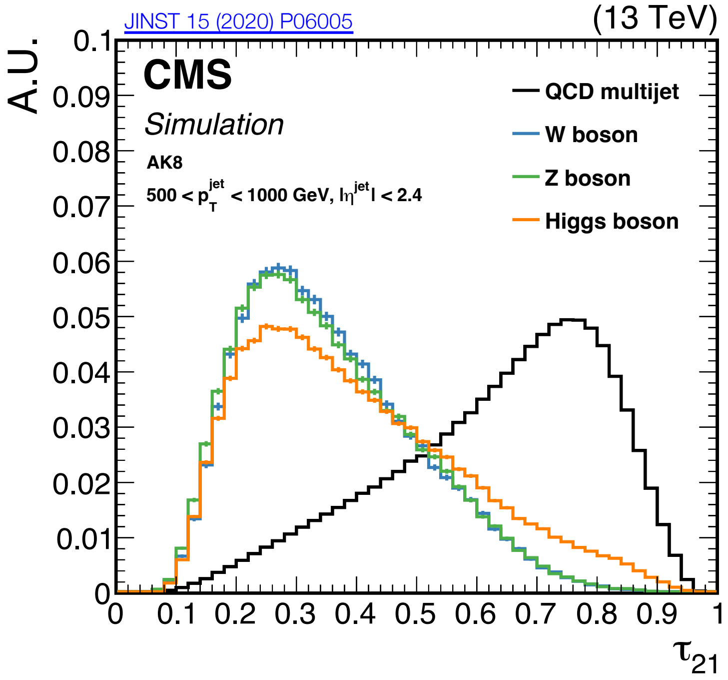
"Combined" Mass

Combination of **calorimeter-based** & **track-assisted** mass

ATLAS-CONF-2016-035
Eur. Phys. J. C 79 (2019) 375

Jet Tagging With Substructure Moments

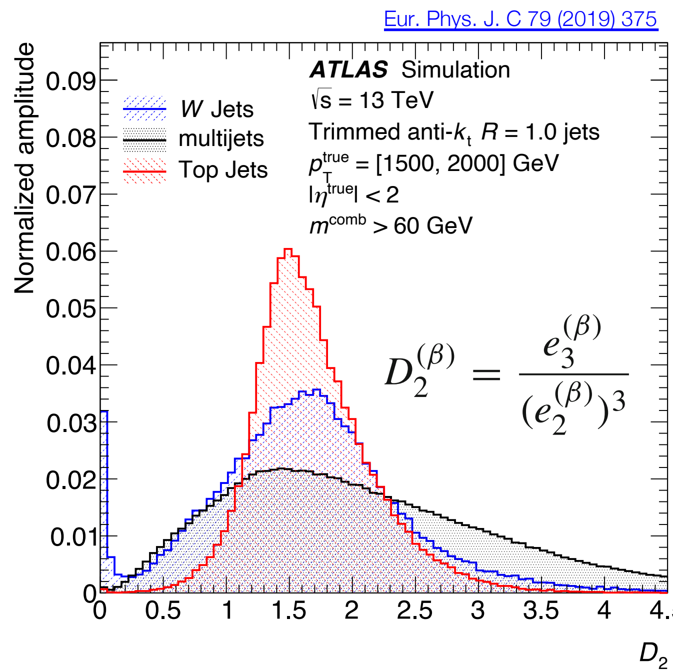
Measures of energy distributions inside a jet



N-subjettiness

- Subjet axes approach (exclusive k_T algorithm)

$$\tau_N = \frac{1}{d_0} \sum_k p_T^k \min(\Delta R_{1,k}, \ldots, \Delta R_{N,k}) \qquad d_0 = \sum_k p_T^k R_0$$



Energy Correlation Functions (ECFs)

- Axes-free approach
  - ➢ reduces dependence of the jet pt

$$E_{CF2}(\beta) = \sum_{i<j\in J} p_{T_i} p_{T_j} (\Delta R_{ij})^{\beta} \qquad e_2^{(\beta)} = \frac{E_{CF2}(\beta)}{E_{CF1}(\beta)^2}$$

$$E_{CF3}(\beta) = \sum_{i<j<k\in J} p_{T_i} p_{T_j} p_{T_k} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^{\beta} \qquad e_3^{(\beta)} = \frac{E_{CF3}(\beta)}{E_{CF1}(\beta)^3}$$

JINST 15 (2020) P09018

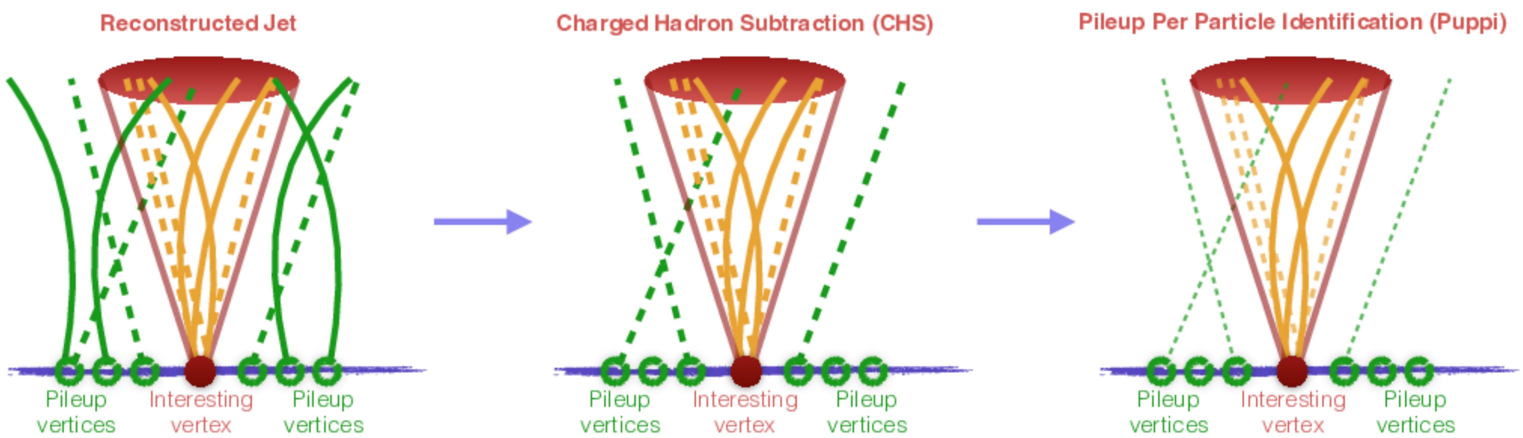
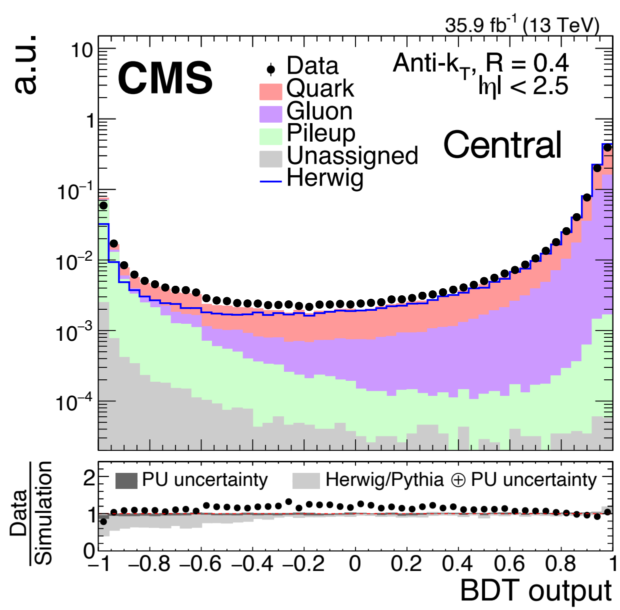
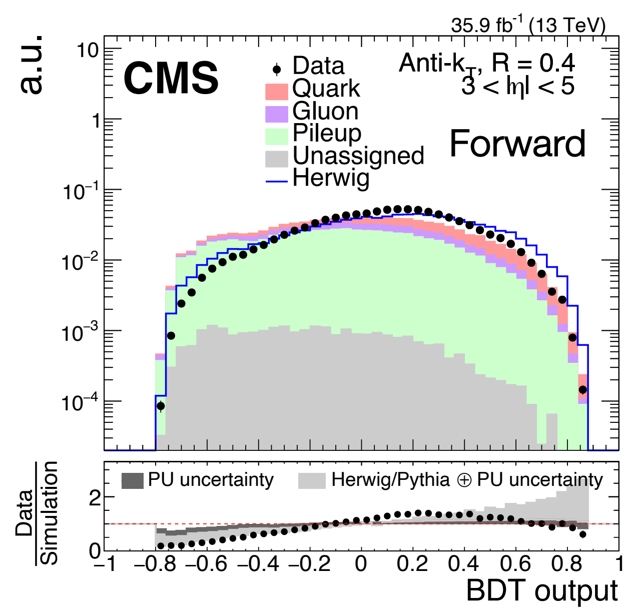


Figure by Andrea Malara

Constituent-level pileup mitigation

Pileup Jet Identification with BDT

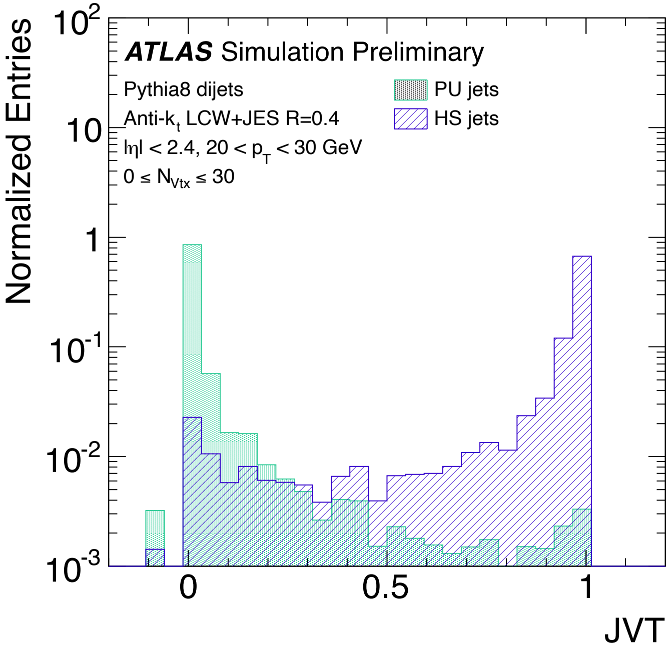




Pileup Mitigation (ATLAS)

Eur. Phys. J. C 76 (2016) 581

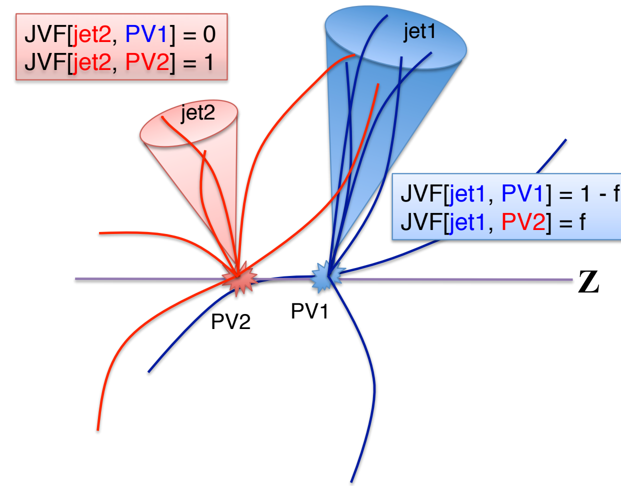
ATLAS-CONF-2014-018

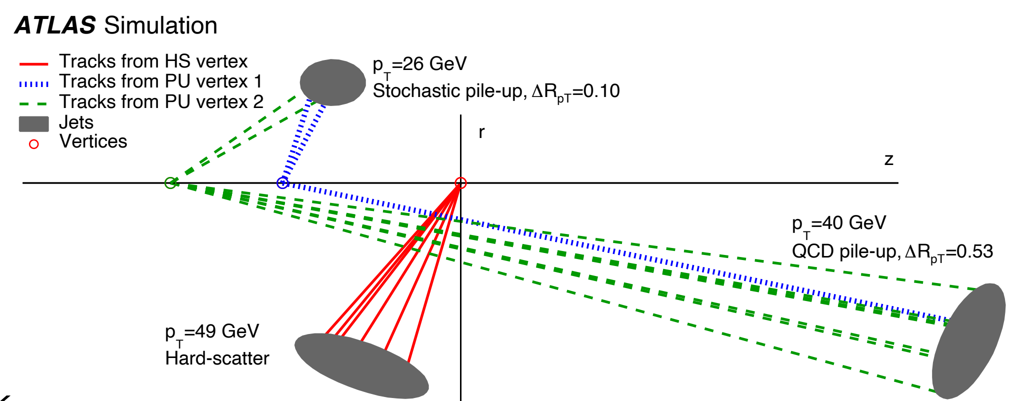


Jet Vertex Tagger (JVT)

Likelihood discriminant built from:

- ➢ corrected Jet Vertex Fraction (corrJVF)
- ➢ R_{pT}: ratio of track p_T sum to jet p_T





Forward Jet Vertex Tagger (fJVT)

Use MET calculated at each PU vertex to check for forward jets that correct imbalance

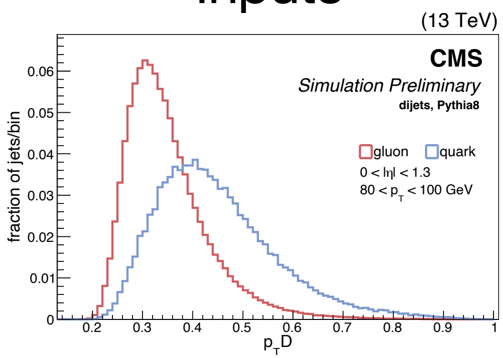
Eur. Phys. J. C 77 (2017) 580

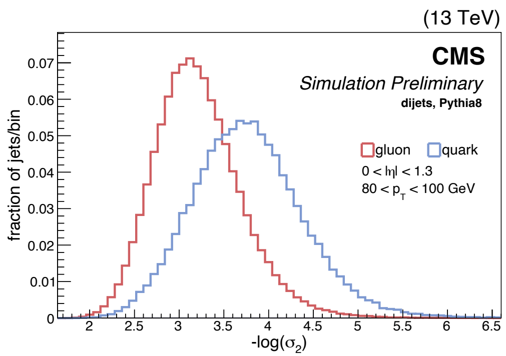
Quark-Gluon Jet Tagging (CMS)

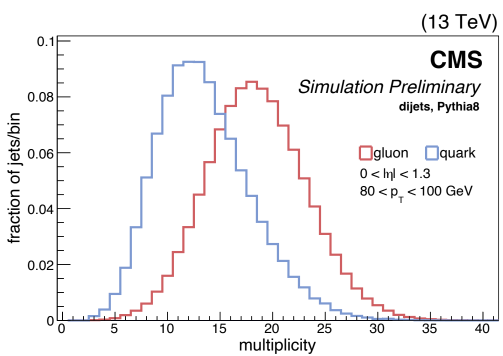
CMS-PAS-JME-16-003

Quark-Gluon Likelihood (QGL)

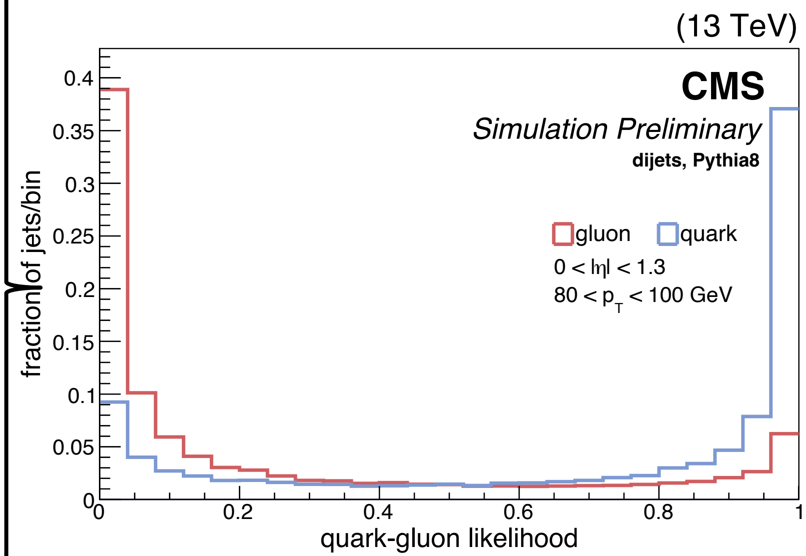
Inputs



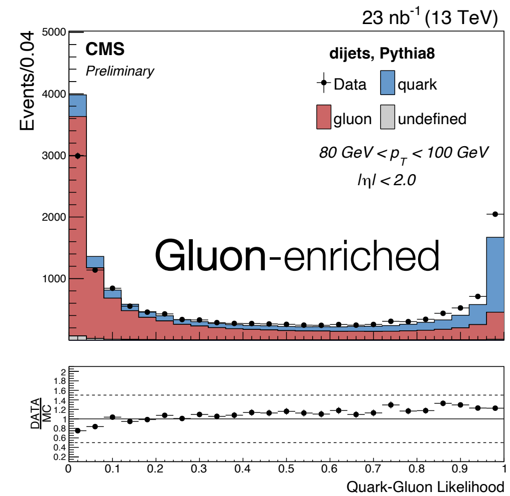


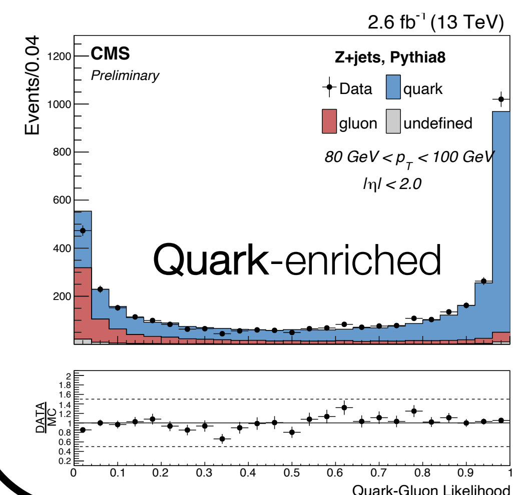


QGL discriminant



Validation with Data

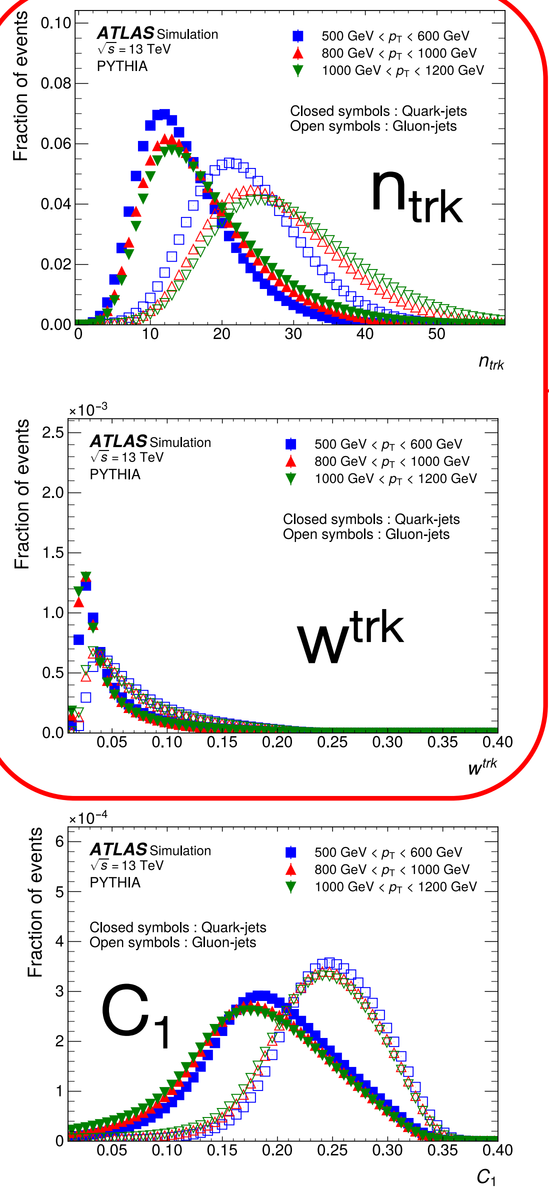




Quark-Gluon Jet Tagging (ATLAS)

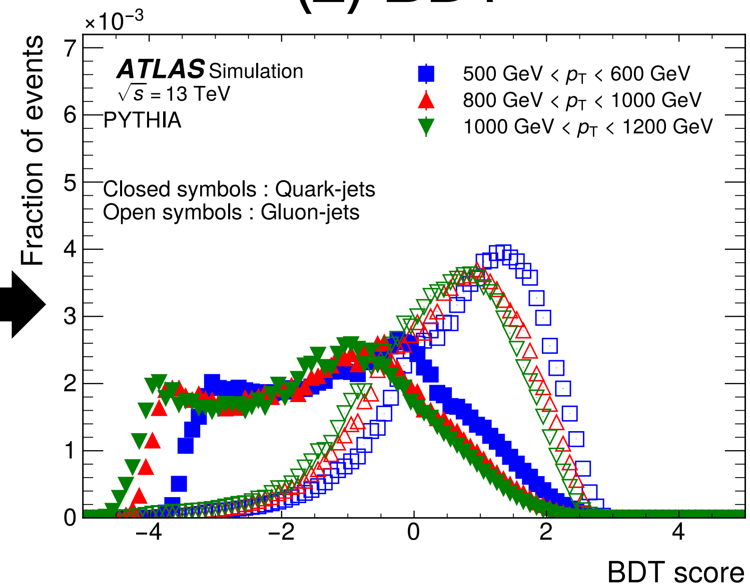
Chinese Phys. C 48 (2024) 023001

Inputs to BDT

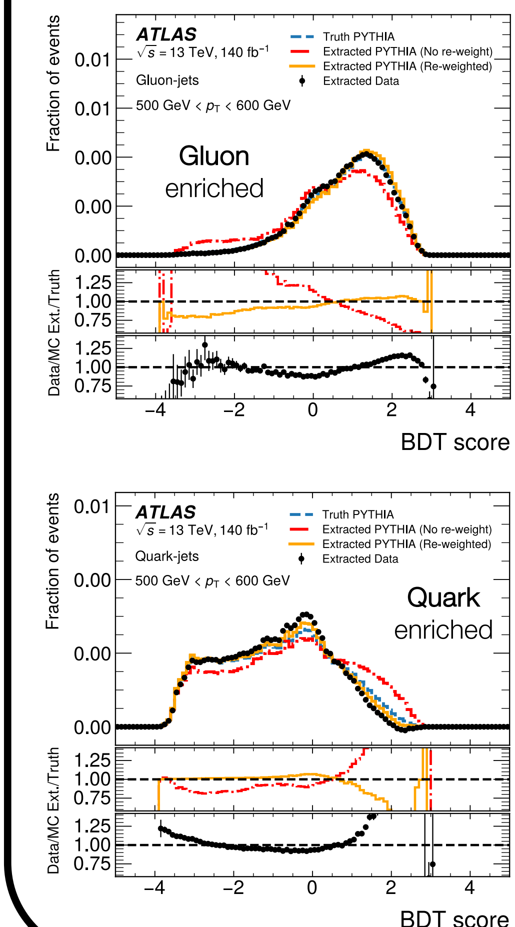


(1) Cut-based (n_{trk}, w^{trk})

(2) BDT



Validation with Data



Run 2 VBS measurements with jet substructure

CMS: VBS WW/WZ -> lvqq

Phys. Lett. B 834 (2022) 137438

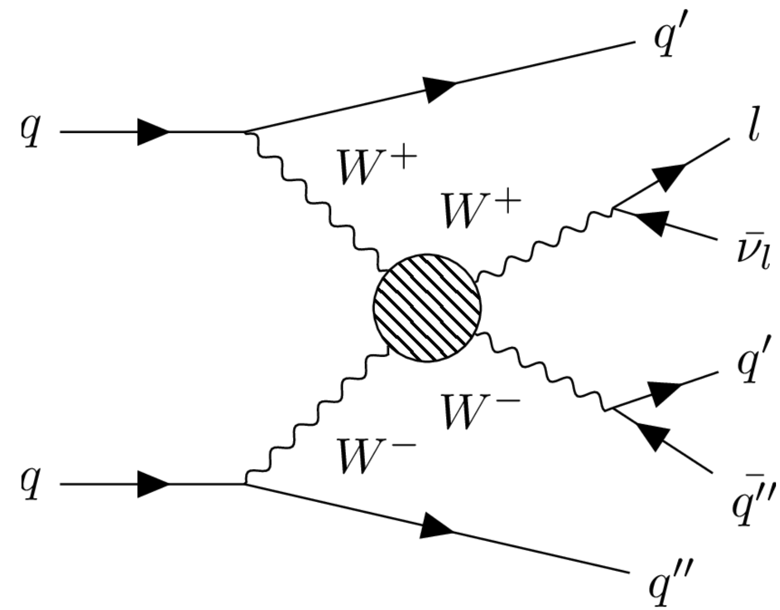
- Full Run-2 dataset (L = 138 fb^{-1}) collected with single-lepton triggers.
- W->lν: Single e/μ + MET.

Boosted selection

- **V-jet candidate:** 1 **AK8** Puppi jet, τ_{21}, $40 < m_{SD} < 250$ GeV.
- **VBS tag jets:** 2 **AK4** CHS jets with highest invariant mass.

Resolved selection

- **VBS tag jets:** 2 AK4 jets with highest invariant mass
- **V-jet candidate:** 2 AK4 jets with invariant mass closest to 85 GeV.

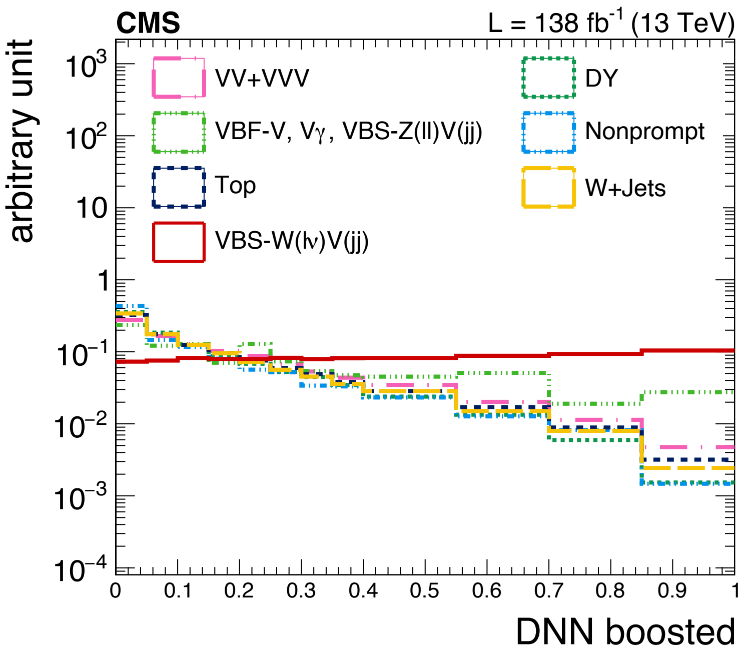


- Pileup Jet ID applied for AK4 CHS jets.

Phys. Lett. B 834 (2022) 137438

Signal extraction strategy

Train a DNN from event-level and jet-level observables and fit the DNN distribution.

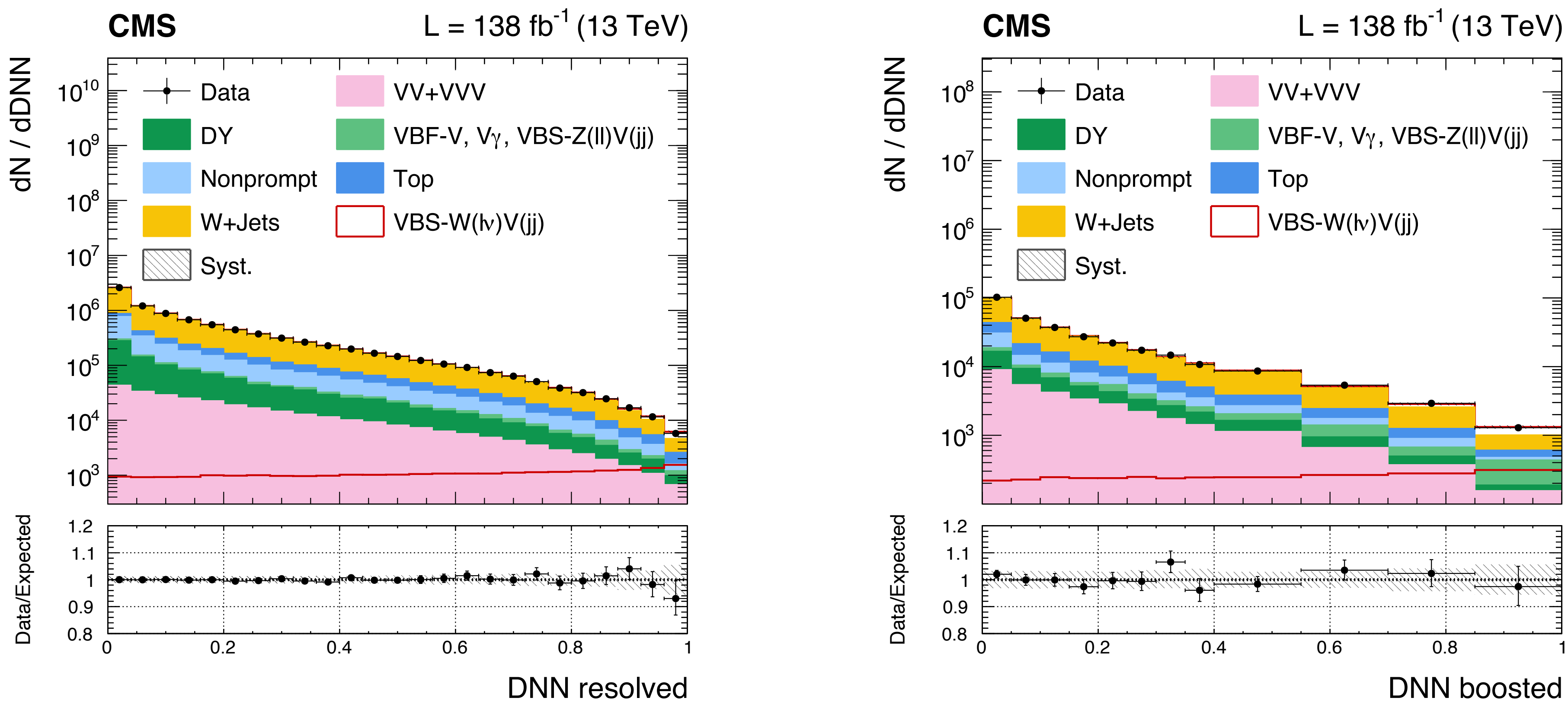


Variable	Resolved	Boosted	SHAP ranking Resolved	SHAP ranking Boosted
Lepton pseudorapidity	✓	✓	13	12
Lepton transverse momentum	✓	✓	16	10
Zeppenfeld variable for the lepton	✓	✓	2	2
Number of jets with $p_T > 30\,\text{GeV}$	✓	✓	7	3
Leading VBS tag jet p_T	-	✓	-	11
Trailing VBS tag jet p_T	✓	✓	7	6
Pseudorapidity interval $\Delta\eta_{jj}^{VBS}$ between tag jets	✓	✓	4	4
Quark/gluon discriminator of leading VBS tag jet	✓	✓	9	7
Azimuthal angle distance between VBS tag jets	✓	-	10	-
Invariant mass of the VBS tag jets pair	✓	✓	1	1
p_T of the leading V_{had} jet	✓	-	14	-
p_T of the trailing V_{had} jet	✓	-	12	-
Pseudorapidity difference between V_{had} jets	✓	-	8	-
Quark/gluon discriminator of the leading V_{had} jet	✓	-	3	-
Quark/gluon discriminator of the trailing V_{had} jet	✓	-	5	-
p_T of the AK8 V_{had} jet candidate	-	✓	-	8
Invariant mass of V_{had}	✓	✓	11	5
Zeppenfeld variable for V_{had}	-	✓	-	9
Centrality	-	✓	15	13

CMS: VBS WW/WZ -> lvqq

Phys. Lett. B 834 (2022) 137438

DNN output distribution in signal regions

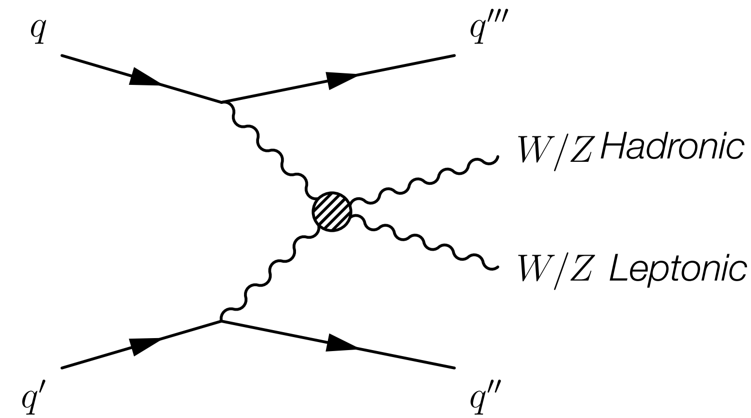


4.4σ (5.1σ) observed (expected)
significance of VBS EWK WV signal

ATLAS: VBS VV -> vvqq / lvqq / llqq

Phys. Rev. D 100, 032007 (2019)

- 2016 dataset with single-lepton / MET triggers.



Three lepton channels simultaneously:

- Leptonic W: Single e / μ.
- Leptonic Z: ee /μμ
- Z->νν: MET

Boosted selection

- **V-jet candidate:** 1 **AK10** jet, D_2, p_T-dependent m_{comb} cut.
- **VBS-tag jets:** 2 **AK4** jets with highest invariant mass.
- Split to High-Purity and Low-Purity signal regions based on (D_2, m_{comb}) cut.

Resolved selection

- **V-jet candidate**: 2 **AK4** jets with invariant mass closest to W/Z mass GeV.
- **VBS-tag jets:** 2 **AK4** jets with highest invariant mass

ATLAS: VBS VV -> vvqq / lvqq / llqq

Phys. Rev. D 100, 032007 (2019)

Signal extraction strategy

Train a BDT from event-level and jet-level observables and fit the BDT distribution.

Variable	0-lepton	1-lepton	2-lepton
m_{jj}^{tag}	✓	...	✓
$\Delta\eta_{jj}^{\mathrm{tag}}$	...	...	✓
$p_{\mathrm{T}}^{\mathrm{tag},j_2}$	✓	✓	✓
m_J	✓	...	...
$D_2^{(\beta=1)}$	✓	...	✓
$E_{\mathrm{T}}^{\mathrm{miss}}$	✓	...	...
$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, J)$	✓	...	...
η_ℓ	...	✓	...
$n_{j,\mathrm{track}}$	✓	...	...
ζ_V	...	✓	✓
m_{VV}	...	...	✓
p_{T}^{VV}	...	...	✓
m_{VVjj}	...	✓	...
p_{T}^{VVjj}	...	...	✓
w^{tag,j_1}	✓	...	...
w^{tag,j_2}	✓	...	...

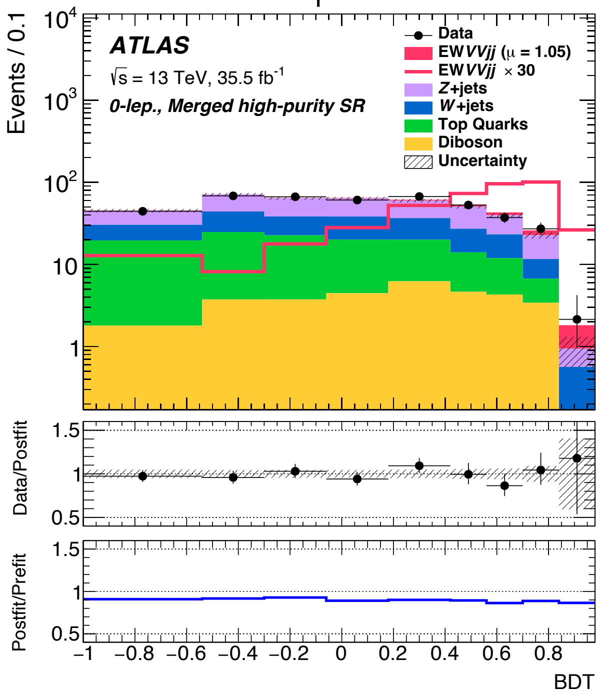
Variable	0-lepton	1-lepton	2-lepton
m_{jj}^{tag}	✓	...	✓
$\Delta\eta_{jj}^{\mathrm{tag}}$	...	...	✓
$p_{\mathrm{T}}^{\mathrm{tag},j_1}$	✓	✓	...
$p_{\mathrm{T}}^{\mathrm{tag},j_2}$	✓	✓	✓
$\Delta\eta_{jj}$	✓	✓	✓
$p_{\mathrm{T}}^{j_1}$	✓	...	...
$p_{\mathrm{T}}^{j_2}$	✓	✓	✓
w^{j_1}	✓	✓	✓
w^{j_2}	✓	✓	✓
$n_{\mathrm{tracks}}^{j_1}$	...	✓	✓
$n_{\mathrm{tracks}}^{j_2}$	...	✓	✓
w^{tag,j_1}	✓	✓	✓
w^{tag,j_2}	✓	✓	✓
$n_{\mathrm{tracks}}^{\mathrm{tag},j_1}$	...	✓	✓
$n_{\mathrm{tracks}}^{\mathrm{tag},j_2}$	...	✓	✓
$n_{j,\mathrm{track}}$	✓	...	✓
$n_{j,\mathrm{extr}}$	✓	...	...
$E_{\mathrm{T}}^{\mathrm{miss}}$	✓	...	...
η_ℓ	...	✓	...
$\Delta R(\ell,\nu)$	...	✓	...
ζ_V	...	✓	✓
m_{VV}	...	...	✓
m_{VVjj}	...	✓	...

ATLAS: VBS VV -> vvqq / lvqq / llqq

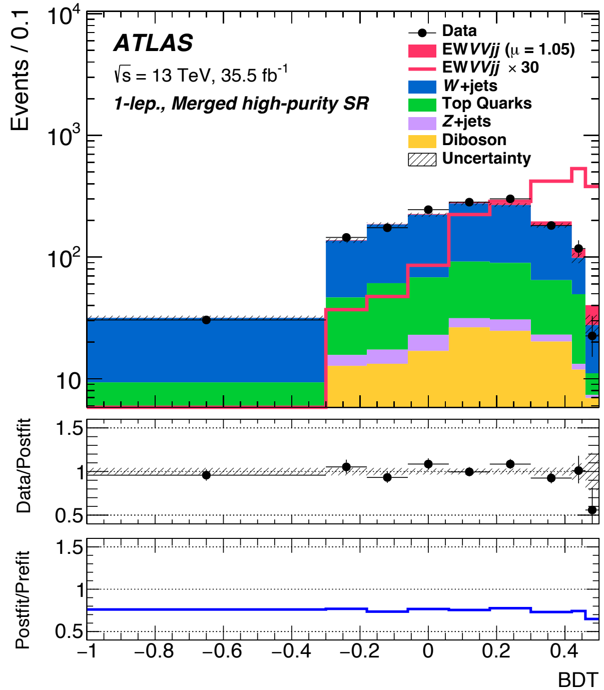
Phys. Rev. D 100, 032007 (2019)

BDT output distribution in Boosted category signal regions

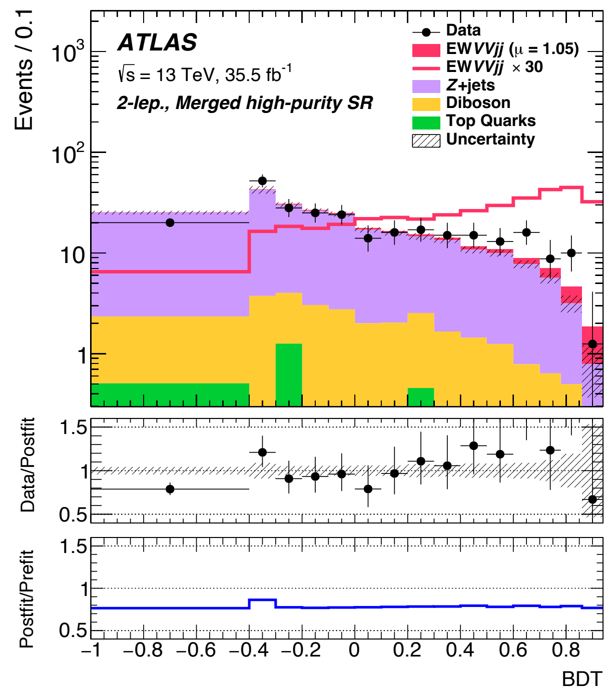
0-lepton



1-lepton



2-lepton

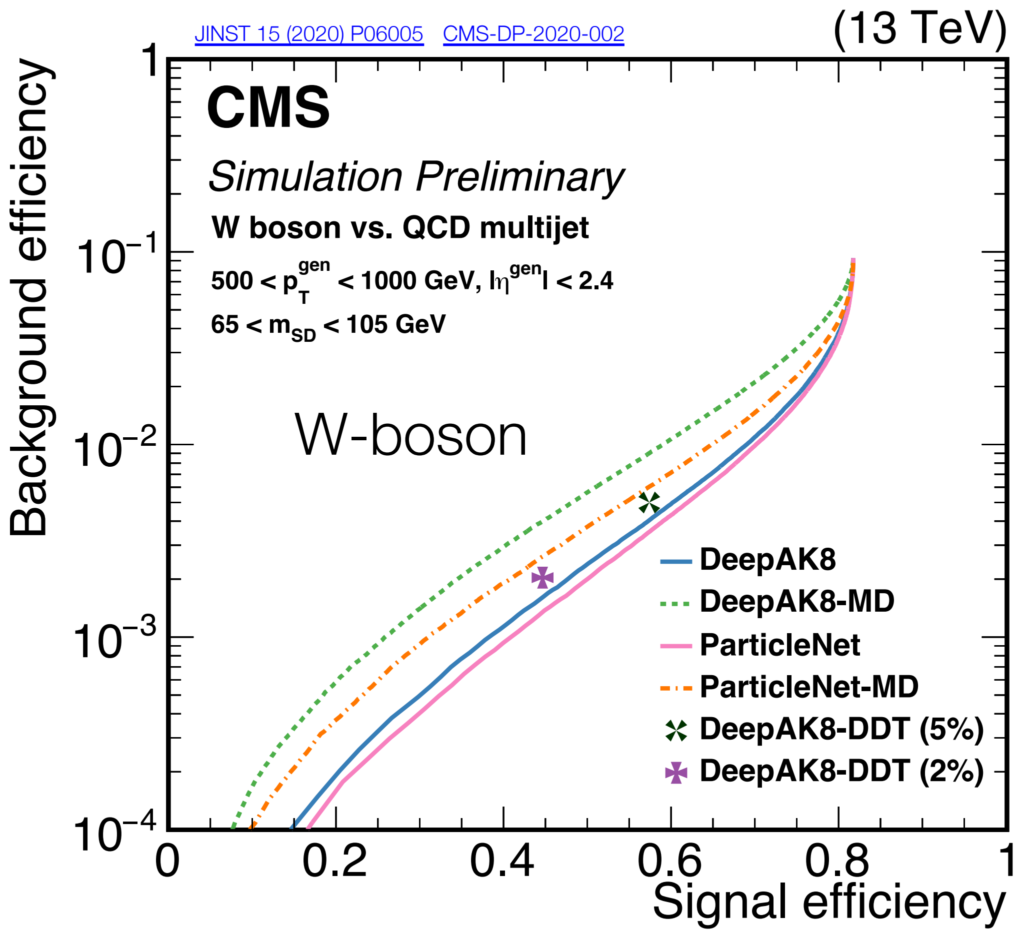


2.7σ (2.5σ) observed (expected)
significance of VBS EWK VV signal

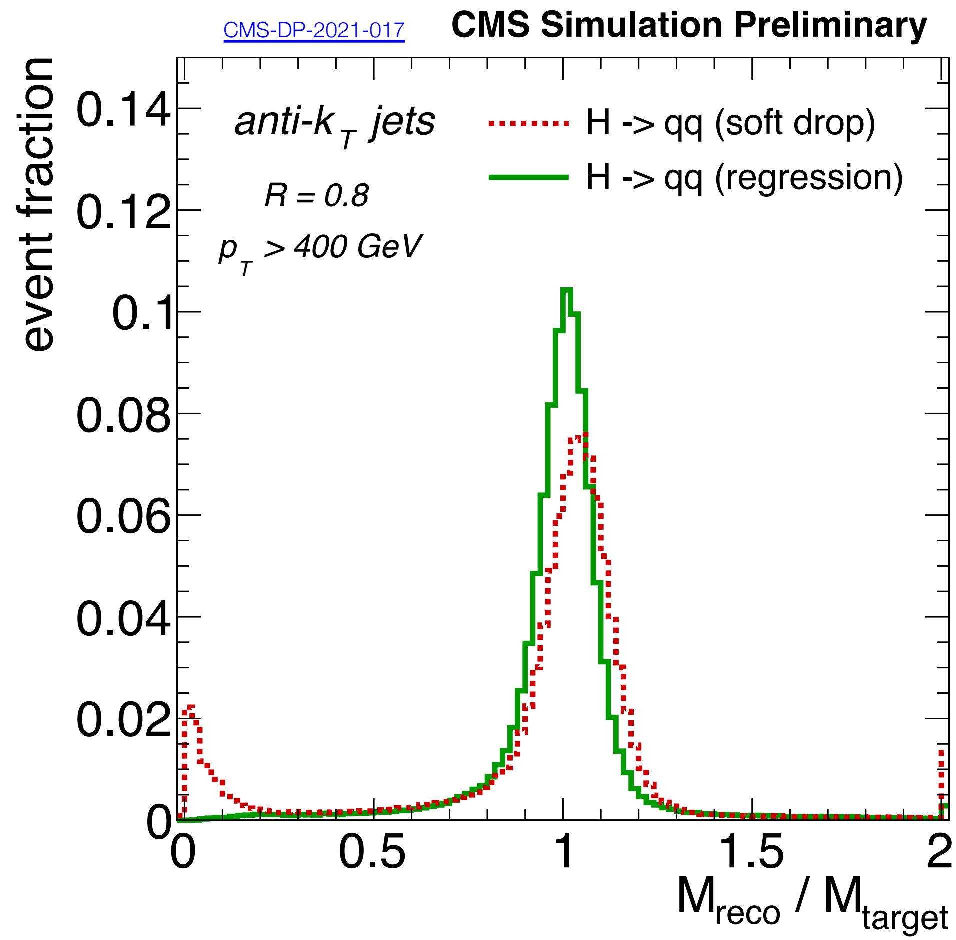
The future of jet tagging & substructure

State-of-the-art (CMS)

Constituent-based jet taggers



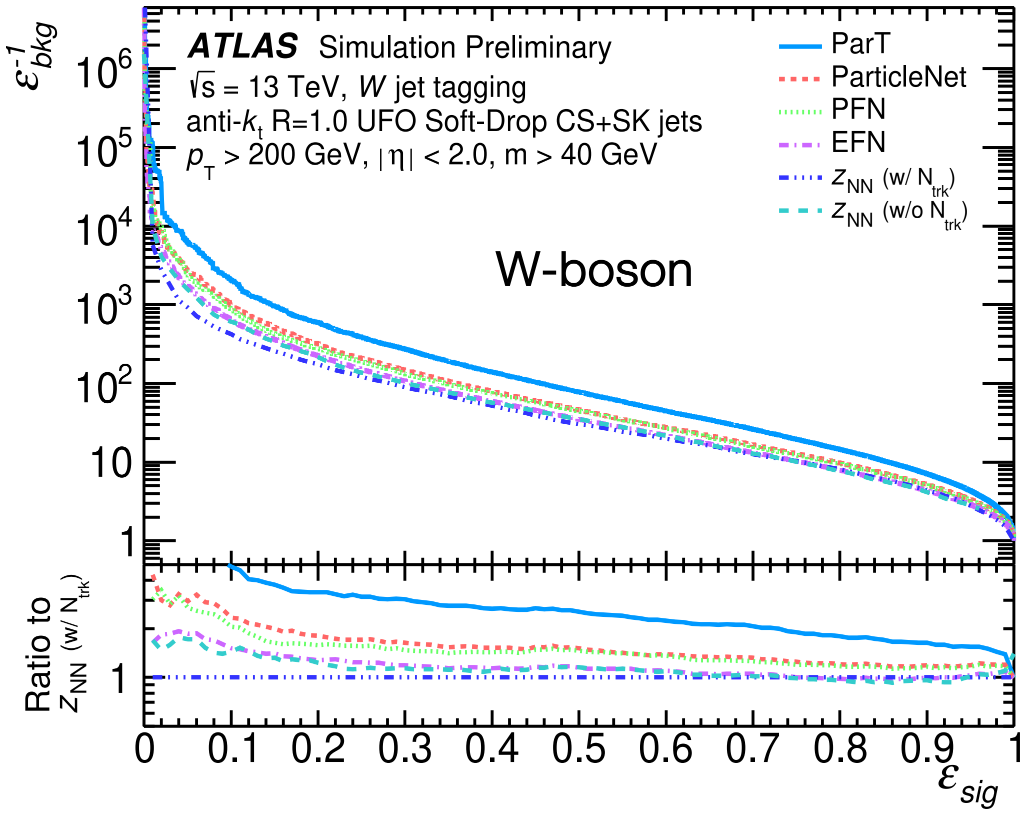
Jet mass regression with jet constituents



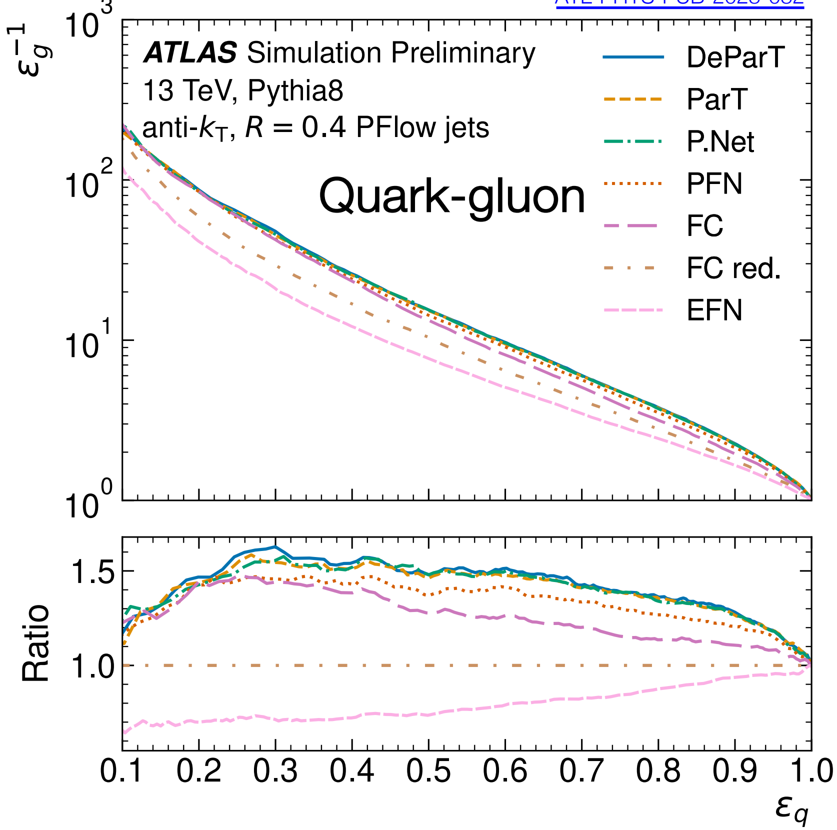
State-of-the-art (ATLAS)

- New jet input: Unified Flow Objects (**UFO**) Eur. Phys. J. C 81 (2021) 334
  - ➢ Particle Flow (**PF**) + Track Calo-Clusters (**TCC**).
- New baseline large-R jet: **CS+SK UFO Soft-Drop** Anti-k_T R = 1.0 jets
  - ➢ Constituent-Subtraction + Soft-Killer for pileup mitigation.
- Constituent-based jet taggers:

ATL-PHYS-PUB-2023-020



ATL-PHYS-PUB-2023-032



Summary

- High-mass m_{VV} regime in VBS measurements is a crucial phase space to discover or constrain BSM physics.

- In this regime, jet substructure techniques enhances the ability to identify boosted hadronically decaying V-bosons.
  - Together with quark vs gluon and pileup jet discrimination for VBS event tagging.

- CMS & ATLAS have utilized jet substructure in Run 2 VBS measurements.

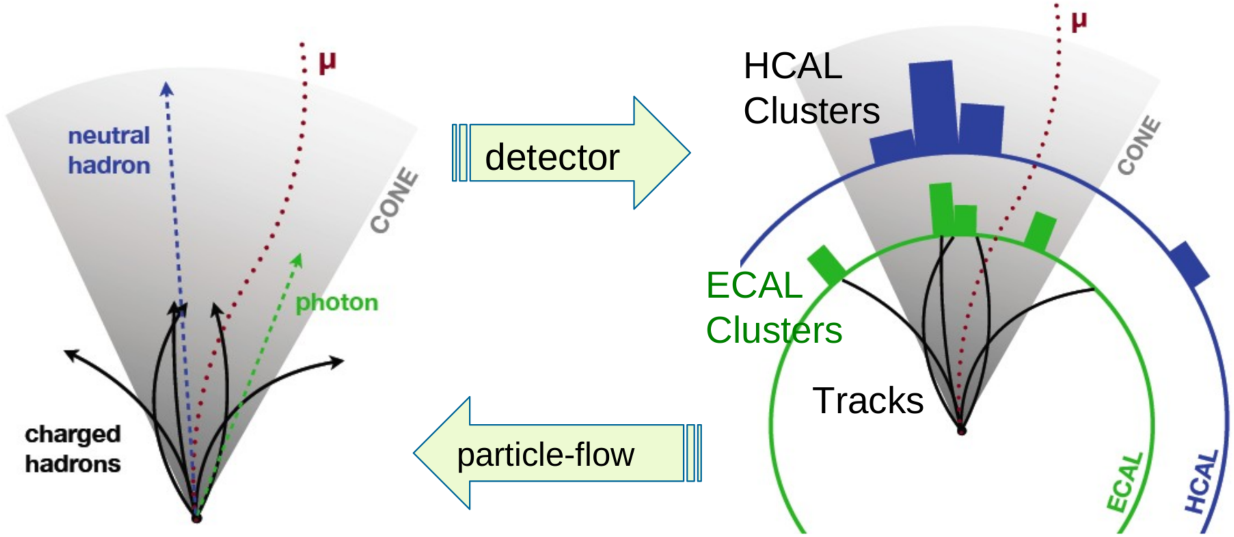
- Latest state-of-the-art substructure techniques to be used for Run 2/3 CMS and ATLAS analyses.

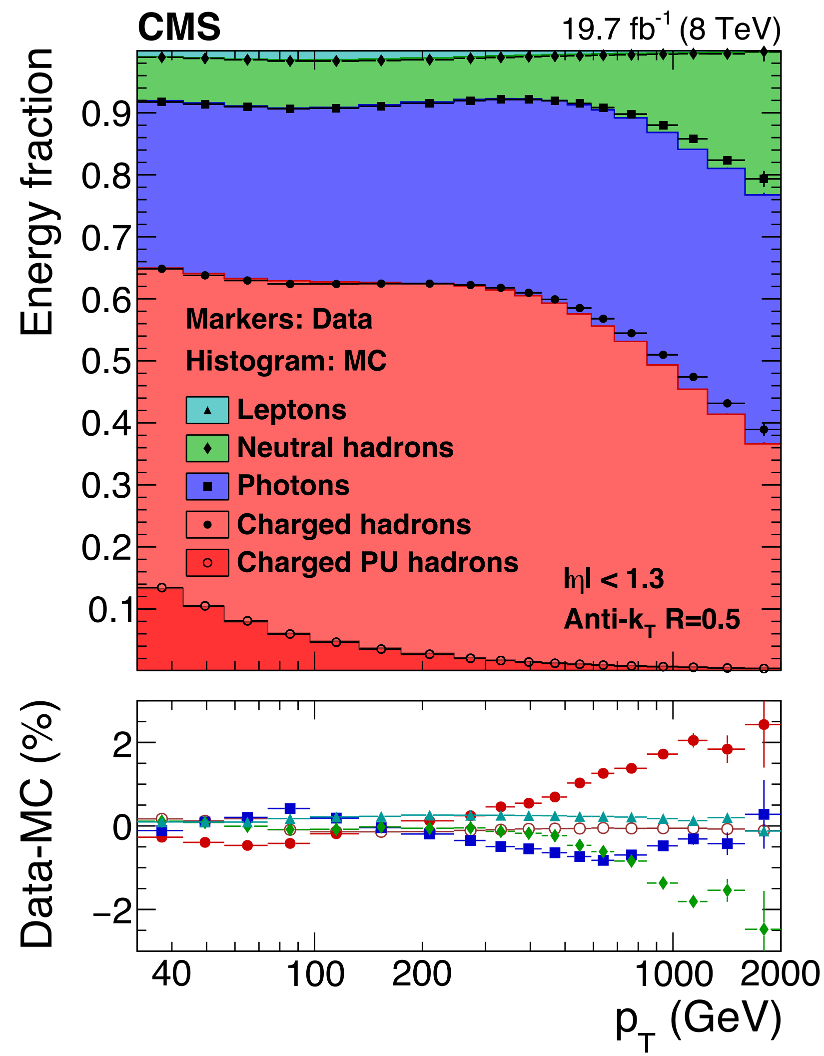
EXTRA SLIDES

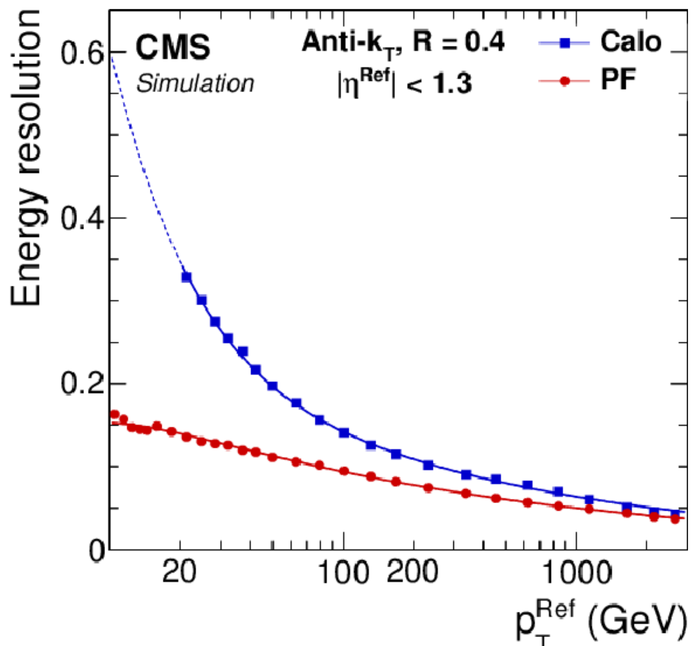
Jet Reconstruction: Input for CMS

JINST 12 (2017) P10003

Particle Flow (PF) Algorithm







Evolution throughout Run-2

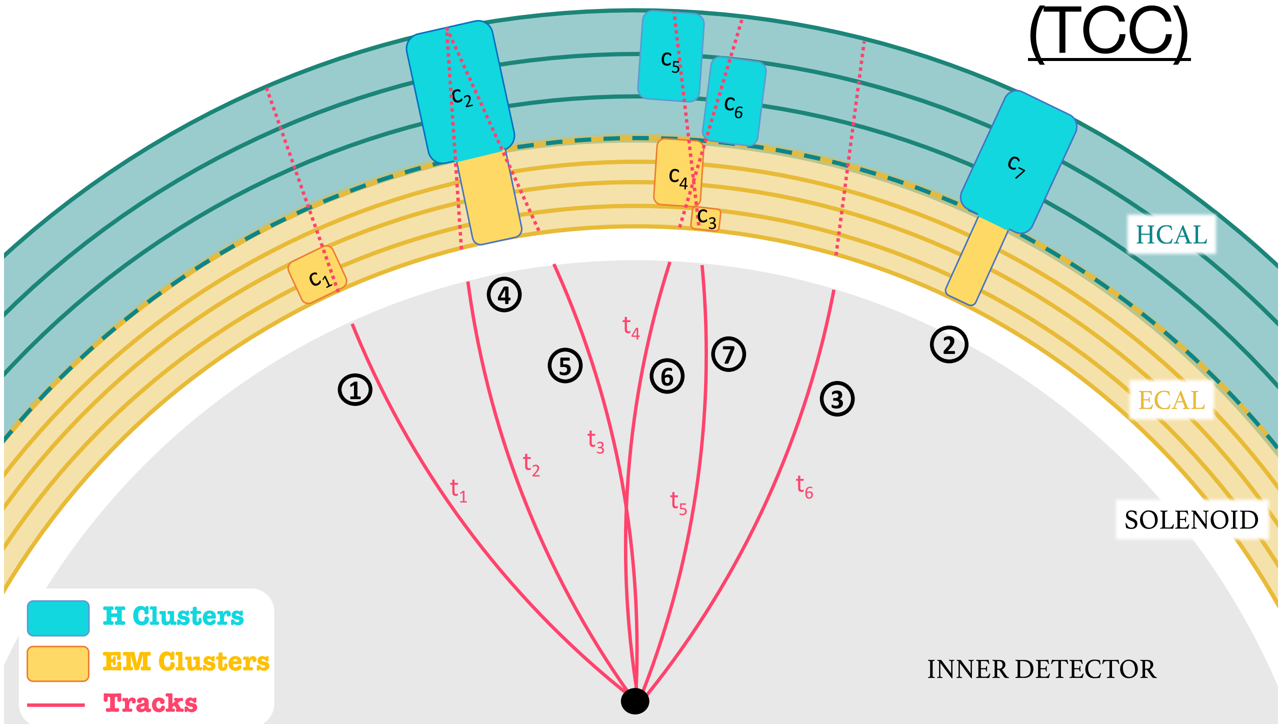
Calorimeter (Topo)clusters

Particle Flow

Track-Calo Clusters (TCC)

Unified Flow Object (UFO)

Eur. Phys. J. C 81 (2021) 334

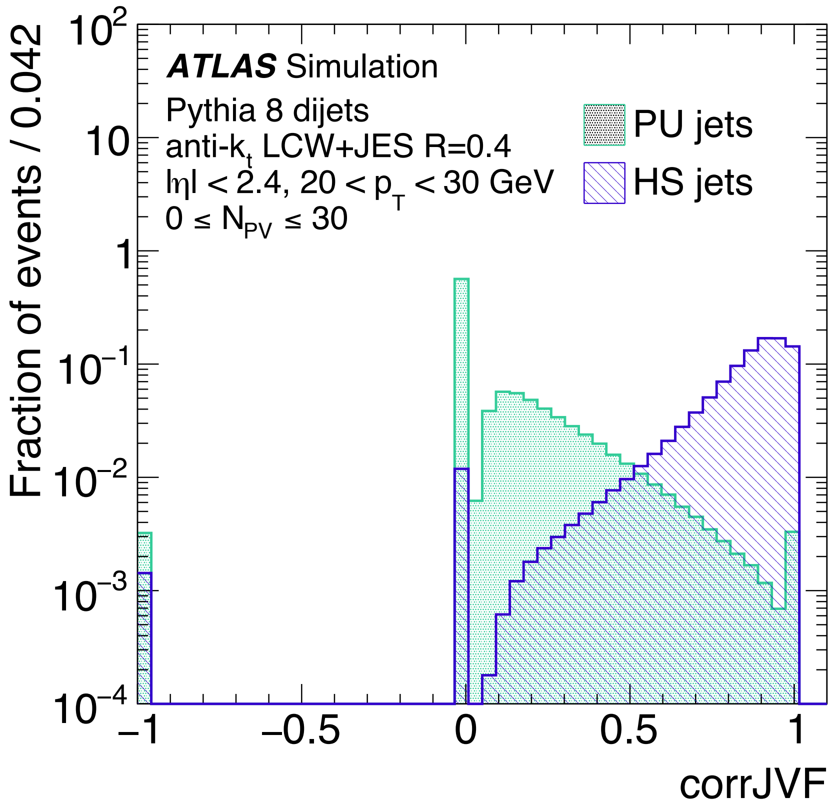


ATL-PHYS-PUB-2017-015

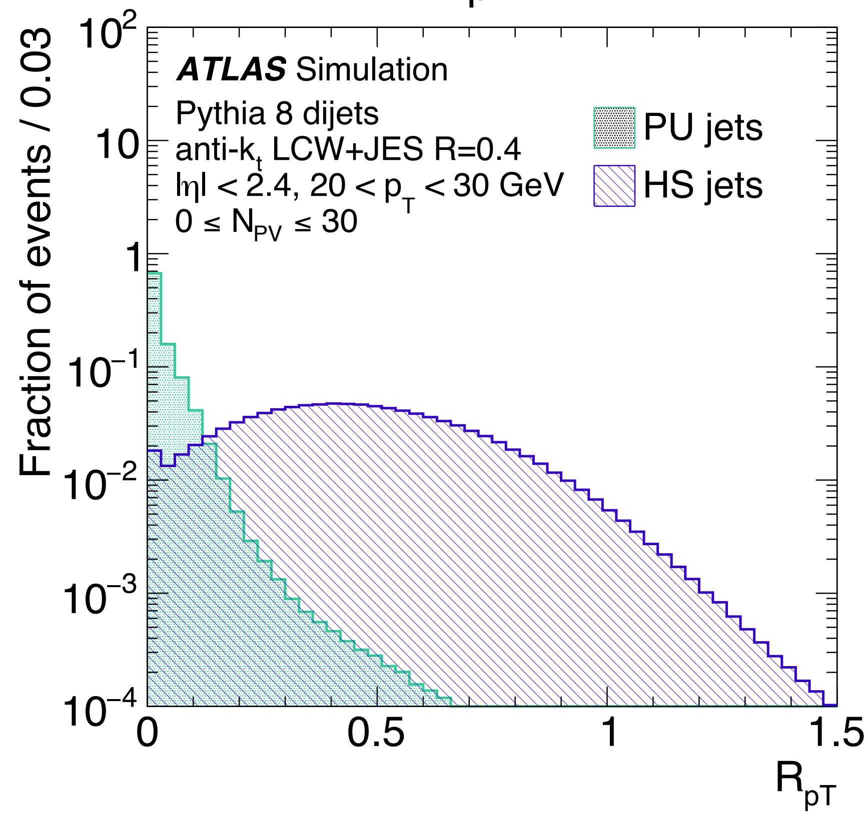
Pileup Jet Id (ATLAS): Inputs for JVT

Eur. Phys. J. C 76 (2016) 581

corrJVF



R_{pT}



$$\mathrm{corrJVF} = \frac{\sum_m p_{\mathrm{T},m}^{\mathrm{track}}(\mathrm{PV}_0)}{\sum_l p_{\mathrm{T},l}^{\mathrm{track}}(\mathrm{PV}_0) + \frac{\sum_{n\geq 1}\sum_l p_{\mathrm{T},l}^{\mathrm{track}}(\mathrm{PV}_n)}{(k \cdot n_{\mathrm{track}}^{\mathrm{PU}})}}$$

$$R_{\mathrm{pT}} = \frac{\sum_k p_{\mathrm{T},k}^{\mathrm{track}}(\mathrm{PV}_0)}{p_{\mathrm{T}}^{\mathrm{jet}}}$$

JINST 15 (2020) P09018

BDT Input Variables

Input variable	Definition
β	Fraction of p_{T} of charged particles associated with the LV, defined as $\sum_{i\in\mathrm{LV}} p_{\mathrm{T},i} / \sum_i p_{\mathrm{T},i}$ where i iterates over all charged PF particles in the jet
N_{vertices}	Number of vertices in the event
$\langle \Delta R^2 \rangle$	Square distance from the jet axis scaled by p_{T}^2 average of jet constituents: $\sum_i \Delta R^2 p_{\mathrm{T},i}^2 / \sum_i p_{\mathrm{T},i}^2$
$f_{\mathrm{ring}X}$, $X = 1, 2, 3$, 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
$p_{\mathrm{T}}^{\mathrm{lead}} / p_{\mathrm{T}}^{\mathrm{jet}}$	p_{T} fraction carried by the leading PF candidate
$p_{\mathrm{T}}^{\mathrm{l.\,ch.}} / p_{\mathrm{T}}^{\mathrm{jet}}$	p_{T} 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
$p_{\mathrm{T}}^{\mathrm{D}}$	Jet fragmentation distribution, defined as $\sqrt{\sum_i p_{\mathrm{T},i}^2} / \sum_i p_{\mathrm{T},i}$

CHS(+Pileup ID) vs Puppi Performance in simulation