Introduction
N-subjettiness
W-jet mass
Summary & Conclusions
Back-up

Recent jet (substructure) measurements at CMS

ESP-HEP 2025 Marseille

Patrick L.S. Connor

on behalf of the CMS Collaboration

Organisation européenne pour la recherche nucléaire

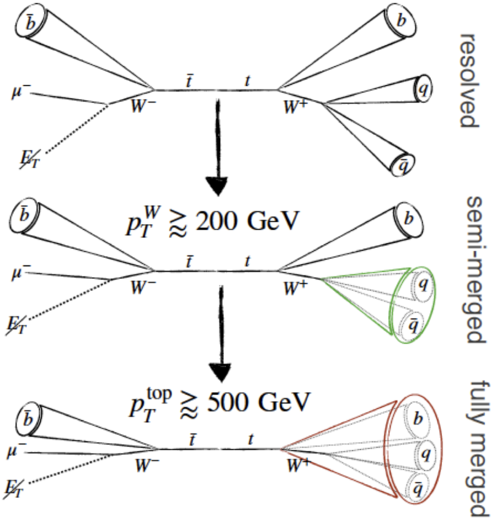
8 July 2025

Introduction

Motivation
Reminder

Introduction

Motivation

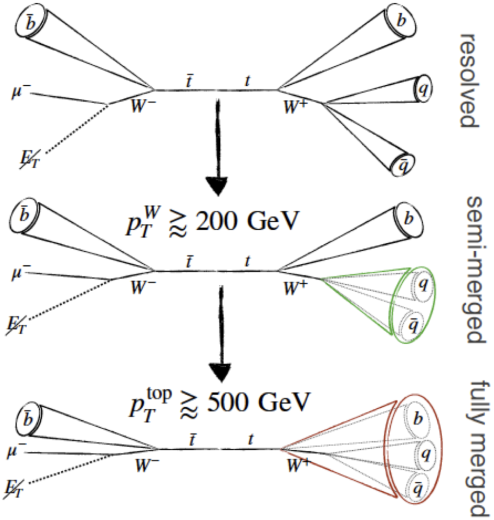


What does the substructure of boosted jets tell us?

- 1-pronged: pure QCD jet
- 2-pronged: boosted, hadronically decaying vector boson
- 3-pronged: boosted, hadronically decaying top quark

Introduction

Motivation



What does the substructure of boosted jets tell us?

1-pronged	pure QCD jet
2-pronged	boosted, hadronically decaying vector boson
3-pronged	boosted, hadronically decaying top quark

Introduction

Reminder

N-subjettiness [1, 2]

$\tau_N^{(\beta)}$ tests the **compatibility** of N axes in a jet with **fully-merged N-prong** decays:

$$\tau_N^{(\beta)} \propto \frac{1}{p_\mathrm{T}} \sum_{i \in \text{jet}} p_\mathrm{T}^{(i)} \min \left\{ \Delta R_{1i}^\beta, \ldots, \Delta R_{Ni}^\beta \right\}$$

- β regulates how much soft radiations should contribute (IRC for $\beta \geq 0$).
- The N axes need be defined to calculate $\tau_N^{(\beta)}$, e.g. by running the k_T jet clustering algorithm in exclusive mode.

Introduction
Motivation
Reminder
N-subjettiness
W-jet mass
Summary & Conclusions
Back-up

Introduction

Reminder

Grooming techniques

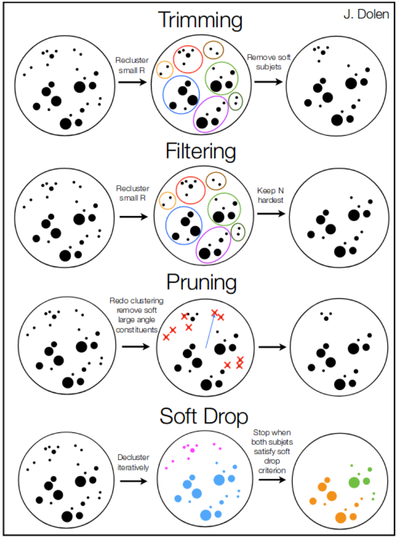
1. Jets acquire mass from splittings in PS.
2. Apply grooming to better discriminate pure QCD jets, boosted V jets or boosted top jets.
3. CMS typically uses the soft-drop (SD) algorithm to **remove soft and wide-angle radiations** [3] with $\beta = 0$ and $z_{\text{cut}} = 0.1$.

Taggers

N-subjettiness ratio $\tau_{NM} = \tau_N / \tau_M$ (at CMS often with $\beta = 1$) to separate N- and M-pronged jets [2].

energy correlator function (ECF) ratios where ECF are also function sensitive to the number of prongs, but w/o the need to define axes [4].

ParticleNet GNN approach [5].



Introduction

Reminder

Grooming techniques

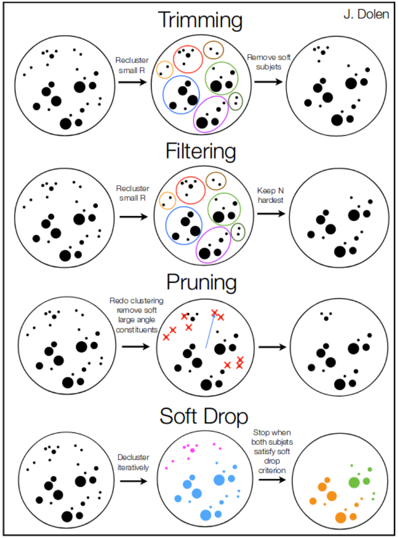
1. Jets acquire mass from splittings in PS.
2. Apply grooming to better discriminate pure QCD jets, boosted V jets or boosted top jets.
3. CMS typically uses the soft-drop (SD) algorithm to **remove soft and wide-angle radiations** [3] with $\beta = 0$ and $z_{\text{cut}} = 0.1$.

Taggers

N-subjettiness ratio $\tau_{NM} = \tau_N / \tau_M$ (at CMS often with $\beta = 1$) to separate N- and M-pronged jets [2].

energy correlator function (ECF) ratios where ECF are also function sensitive to the number of prongs, but w/o the need to define axes [4].

ParticleNet GNN approach [5].



N-subjettiness in various topologies

Goal
Methodology
Unfolding
Results

N-subjettiness in various topologies

Observables

- We want to measure the following set of observables in jets obtained from pure QCD jets and boosted hadronic W/top-enriched jets:

$$\left\{\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \ldots, \tau_{M-2}^{(0.5)}, \tau_{M-2}^{(1)}, \tau_{M-2}^{(2)}, \tau_{M-1}^{(1)}, \tau_{M-1}^{(2)}\right\}$$

$\longrightarrow$ The $3M-4$ observables completely determine the kinematics of M resolved emissions.

- We choose $M=6$ and include also $\beta = 0.25, 1.5$.

$\longrightarrow$ Overcomplete (OC) basis to mitigate the finite resolution.

- For each data set, calculate statistical correlations among all observables.

$\longrightarrow$ Repeat (nearly) the same analysis procedure for 3 different topologies.

$\longrightarrow$ **3×25 observables in total!**

Application

- Test state-of-the-art models.
- Input for tuning of MC generators.
- Input for training of ML-based discriminants.

N-subjettiness in various topologies

Observables

- We want to measure the following set of observables in jets obtained from pure QCD jets and boosted hadronic W/top-enriched jets:

$$\left\{\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \ldots, \tau_{M-2}^{(0.5)}, \tau_{M-2}^{(1)}, \tau_{M-2}^{(2)}, \tau_{M-1}^{(1)}, \tau_{M-1}^{(2)}\right\}$$

$\longrightarrow$ The $3M-4$ observables completely determine the kinematics of M resolved emissions.

- We choose $M=6$ and include also $\beta = 0.25, 1.5$.

$\longrightarrow$ Overcomplete (OC) basis to mitigate the finite resolution.

- For each data set, calculate statistical correlations among all observables.

$\longrightarrow$ Repeat (nearly) the same analysis procedure for 3 different topologies.

$\longrightarrow$ **3×25 observables in total!**

Application

- Test state-of-the-art models.
- Input for tuning of MC generators.
- Input for training of ML-based discriminants.

N-subjettiness in various topologies

Generalities

- Full Run-2 data set ($135 - 138\ \text{fb}^{-1}$ at $\sqrt{s} = 13\ \text{TeV}$).
- Using AK8 (a.k.a. fat) jets with $|y| < 1.7$ to ensure full reconstruction in tracker acceptance and $p_\text{T} > 200\ \text{GeV}$.

Dijet topology

- Follow **similar strategy as in dijet cross section measurements** [6].
- Require back-to-back, isolated jets.

Boosted topologies

- Both rely on **semileptonic $\text{t}\bar{\text{t}}$ events**.
  $\longrightarrow$ Require exactly one isolated muon ($p_\text{T} > 55\ \text{GeV}$) and no additional leptons, MET $> 30\ \text{GeV}$, one b-tagged jet, ...
- Then define **orthogonal, enriched regions**.
  $\longrightarrow$ Purity above 80%.

N-subjettiness in various topologies

Generalities

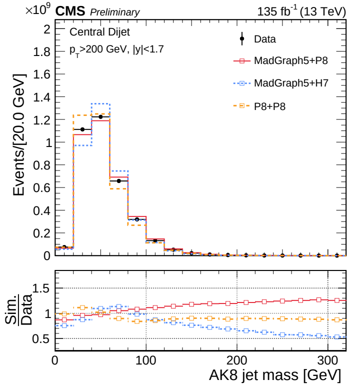
- Full Run-2 data set ($135 - 138\ \mathrm{fb}^{-1}$ at $\sqrt{s} = 13\ \mathrm{TeV}$).
- Using AK8 (a.k.a. fat) jets with $|y| < 1.7$ to ensure full reconstruction in tracker acceptance and $p_\mathrm{T} > 200\ \mathrm{GeV}$.

Dijet topology

- Follow **similar strategy as in dijet cross section measurements** [6].
- Require back-to-back, isolated jets.

Boosted topologies

- Both rely on **semileptonic $t\bar{t}$ events**.
  $\longrightarrow$ Require exactly one isolated muon ($p_\mathrm{T} > 55\ \mathrm{GeV}$) and no additional leptons, MET $> 30\ \mathrm{GeV}$, one b-tagged jet, ...
- Then define **orthogonal, enriched regions**.
  $\longrightarrow$ Purity above 80%.



N-subjettiness in various topologies

Generalities

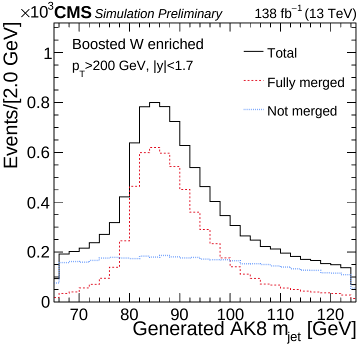
- Full Run-2 data set ($135 - 138\ \mathrm{fb}^{-1}$ at $\sqrt{s} = 13\ \mathrm{TeV}$).
- Using AK8 (a.k.a. fat) jets with $|y| < 1.7$ to ensure full reconstruction in tracker acceptance and $p_\mathrm{T} > 200\ \mathrm{GeV}$.

Dijet topology

- Follow **similar strategy as in dijet cross section measurements** [6].
- Require back-to-back, isolated jets.

Boosted topologies

- Both rely on **semileptonic $t\bar{t}$ events**.
  $\longrightarrow$ Require exactly one isolated muon ($p_\mathrm{T} > 55\ \mathrm{GeV}$) and no additional leptons, MET $> 30\ \mathrm{GeV}$, one b-tagged jet, ...
- Then define **orthogonal, enriched regions**.
  $\longrightarrow$ Purity above 80%.



N-subjettiness in various topologies

Generalities

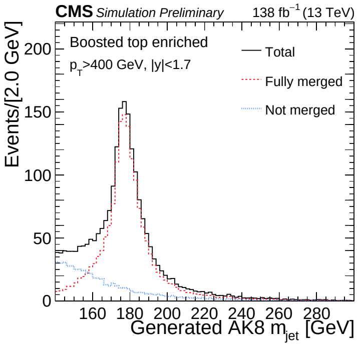
- Full Run-2 data set ($135 - 138\ \text{fb}^{-1}$ at $\sqrt{s} = 13\ \text{TeV}$).
- Using AK8 (a.k.a. fat) jets with $|y| < 1.7$ to ensure full reconstruction in tracker acceptance and $p_\text{T} > 200\ \text{GeV}$.

Dijet topology

- Follow **similar strategy as in dijet cross section measurements** [6].
- Require back-to-back, isolated jets.

Boosted topologies

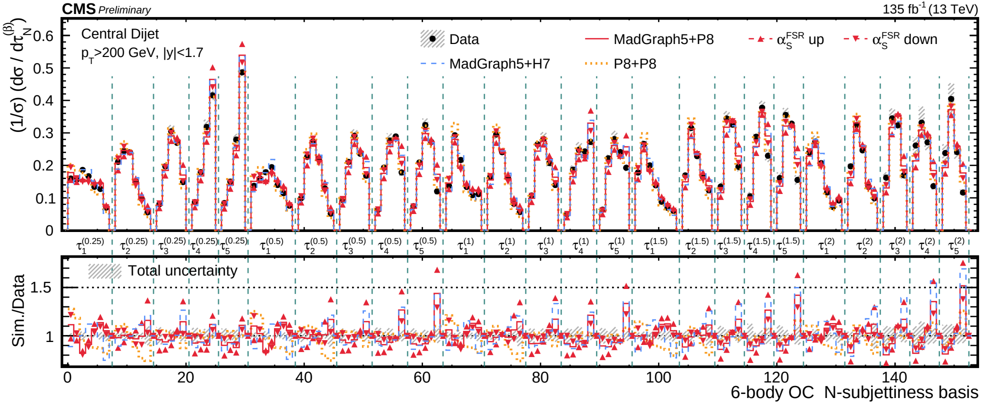
- Both rely on **semileptonic $t\bar{t}$ events**.
  $\longrightarrow$ Require exactly one isolated muon ($p_\text{T} > 55\ \text{GeV}$) and no additional leptons, MET $> 30\ \text{GeV}$, one b-tagged jet, ...
- Then define **orthogonal, enriched regions**.
  $\longrightarrow$ Purity above 80%.



Introduction
N-subjettiness
Goal
Methodology
Unfolding
Results
W-jet mass
Summary & Conclusions
Back-up

N-subjettiness in various topologies

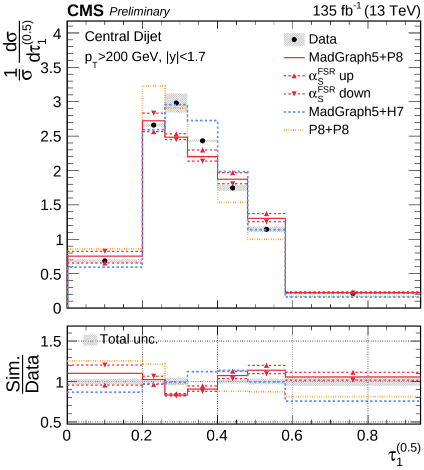
Unfolding

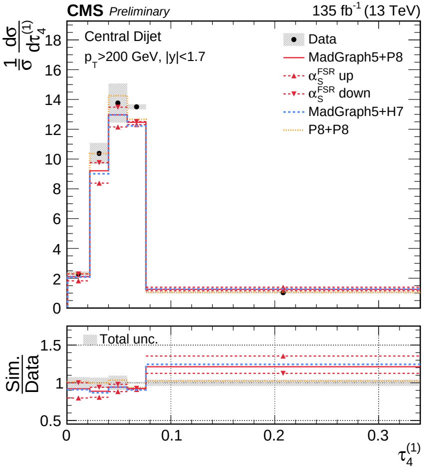


Procedure

- Using TUNFOLD package without regularisation (i.e. equivalent to pseudo-inversion).
- For each category, **unfold** all 25 observables **simultaneously**.
- Unfolding is repeated for each systematic variation independently.

N-subjettiness in various topologies



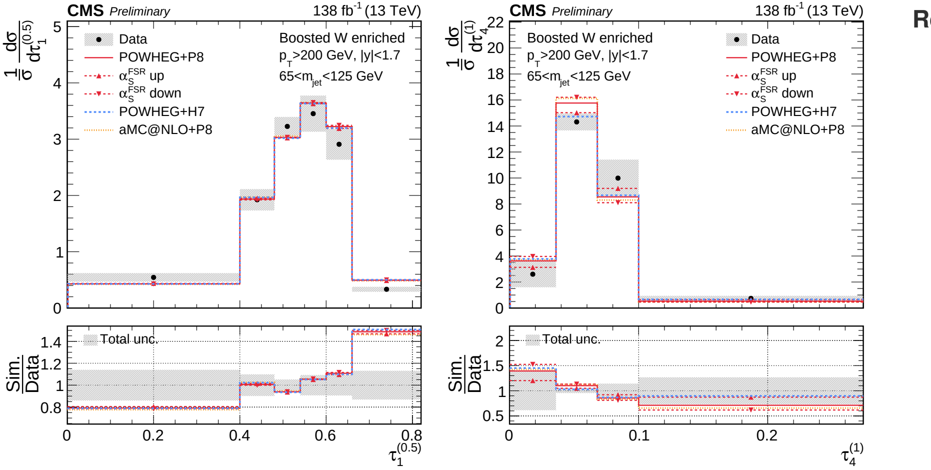


Observations

- In dijet selection, predictions generally **envelope** the data and show disagreements of 10-20%.
- In boosted W- and top-enriched selections, predictions demonstrate **similar shape differences** to the data.

N-subjettiness in various topologies

Results

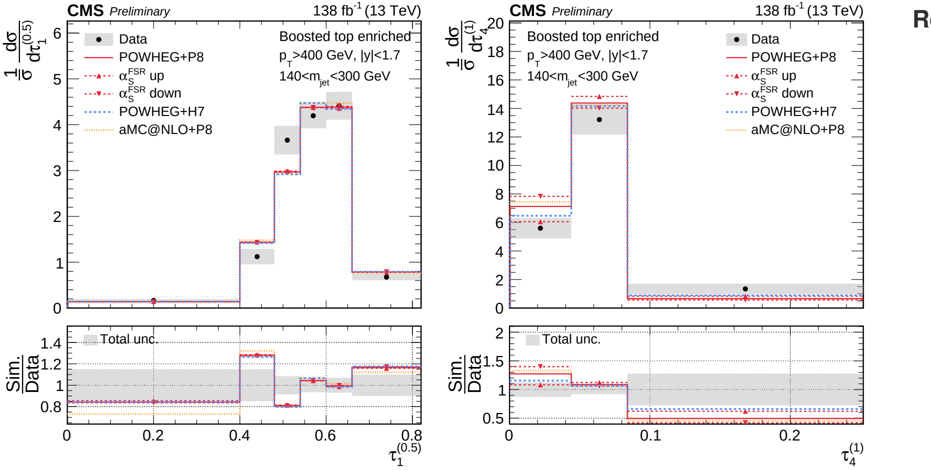


Observations

- In dijet selection, predictions generally **envelope** the data and show disagreements of 10-20%.
- In boosted W- and top-enriched selections, predictions demonstrate **similar shape differences** to the data.

N-subjettiness in various topologies

Results



Observations

- In dijet selection, predictions generally **envelope** the data and show disagreements of 10-20%.
- In boosted W- and top-enriched selections, predictions demonstrate **similar shape differences** to the data.

Boosted W-jet groomed mass

Goal
Methodology
Results

Introduction
N-subjettiness
W-jet mass
Goal
Methodology
Results
Summary & Conclusions
Back-up

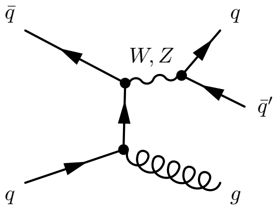
Boosted W-jet groomed mass

Measurements of groomed mass

quark/gluon jets	top quark jets	**W jets**
SMP-16-010 [7]	TOP-21-012 [8]	SMP-24-012

Motivation

- Such jets are subject to many direct BSM searches and SM measurements.
- Often limited by imperfect modelling of the jet substructure.



This analysis

Measurement of **groomed jet mass of fully hadronic $\mathrm{W}(\mathrm{qq}') + $ jets events** in bins of p_{T}:

- Standard candle jet substructure observable to help constrain MC models.
- First determination of W mass in hadronic decay at a hadron collider.

Introduction
N-subjettiness
W-jet mass
Goal
Methodology
Results
Summary & Conclusions
Back-up

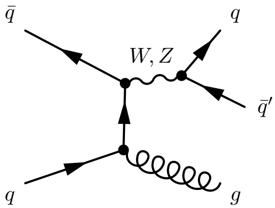
Boosted W-jet groomed mass

Measurements of groomed mass

quark/gluon jets	top quark jets	**W jets**
SMP-16-010 [7]	TOP-21-012 [8]	SMP-24-012

Motivation

- Such jets are subject to many direct BSM searches and SM measurements.
- Often limited by imperfect modelling of the jet substructure.



This analysis

Measurement of **groomed jet mass of fully hadronic $\mathrm{W}(\mathrm{qq}') + $ jets events** in bins of p_T:

- Standard candle jet substructure observable to help constrain MC models.
- First determination of W mass in hadronic decay at a hadron collider.

Introduction
N-subjettiness
W-jet mass
Goal
Methodology
Results
Summary & Conclusions
Back-up

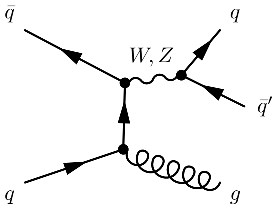
Boosted W-jet groomed mass

Measurements of groomed mass

quark/gluon jets	top quark jets	**W jets**
SMP-16-010 [7]	TOP-21-012 [8]	SMP-24-012

Motivation

- Such jets are subject to many direct BSM searches and SM measurements.
- Often limited by imperfect modelling of the jet substructure.



This analysis

Measurement of **groomed jet mass of fully hadronic $\mathrm{W}(\mathrm{qq}') + $ jets events** in bins of p_T:

- Standard candle jet substructure observable to help constrain MC models.
- First determination of W mass in hadronic decay at a hadron collider.

Introduction
N-subjettiness
W-jet mass
Goal
Methodology
Results
Summary & Conclusions
Back-up

Boosted W-jet groomed mass

Strategy

1. Apply **standard jet reconstruction and calibration** procedures at CMS to a data set obtained from single fat jet triggers with $p_T^{\text{HLT}} > 450, 500$ GeV ($p_T^{\text{offline}} > 575, 650$ GeV).
2. **Groom** jets with soft-drop (SD) algorithm (cf. introduction).
3. **Classify** events using two different mass-decorrelated jet substructure taggers [9].
4. Use a data-driven approach to **control the large QCD background**.
5. **Unfold** using a maximum-likelihood approach.

Classification

	$N_2^{1,\text{DDT}}$ $\varepsilon_{\text{W jets}}$	$\varepsilon_{\text{backg.}}$
pass	18.54%	4.14%
fail	81.46%	95.86%

Introduction
N-subjettiness
W-jet mass
Goal
Methodology
Results
Summary & Conclusions
Back-up

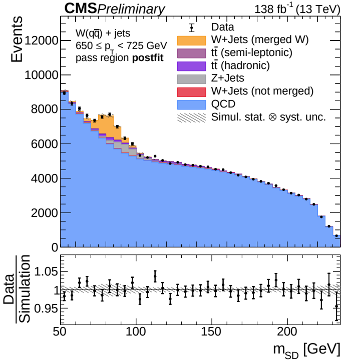
Boosted W-jet groomed mass

Strategy

1. Apply **standard jet reconstruction and calibration** procedures at CMS to a data set obtained from single fat jet triggers with $p_T^{\mathrm{HLT}} > 450, 500$ GeV ($p_T^{\mathrm{offline}} > 575, 650$ GeV).
2. **Groom** jets with soft-drop (SD) algorithm (cf. introduction).
3. **Classify** events using two different mass-decorrelated jet substructure taggers [9].
4. Use a data-driven approach to **control the large QCD background**.
5. **Unfold** using a maximum-likelihood approach.

Classification

	$N_2^{1,\mathrm{DDT}}$ $\varepsilon_{\mathrm{W\ jets}}$	$\varepsilon_{\mathrm{backg.}}$
pass	18.54%	4.14%
fail	81.46%	95.86%



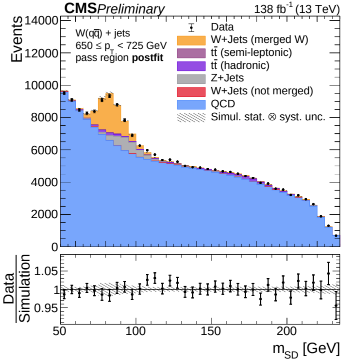
Boosted W-jet groomed mass

Strategy

1. Apply **standard jet reconstruction and calibration** procedures at CMS to a data set obtained from single fat jet triggers with $p_T^{HLT} > 450, 500$ GeV ($p_T^{offline} > 575, 650$ GeV).
2. **Groom** jets with soft-drop (SD) algorithm (cf. introduction).
3. **Classify** events using two different mass-decorrelated jet substructure taggers [9].
4. Use a data-driven approach to **control the large QCD background**.
5. **Unfold** using a maximum-likelihood approach.

Classification

	$P^{PN,DDT}_{W vs. QCD}$ $\varepsilon_{W\ jets}$	$\varepsilon_{backg.}$
pass	38.20%	4.20%
fail	61.80%	95.80%



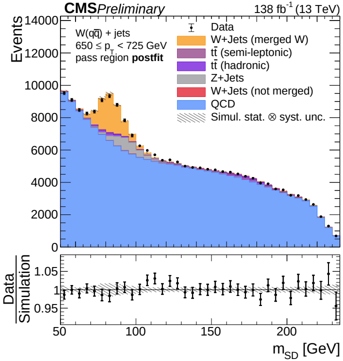
Boosted W-jet groomed mass

Strategy

1. Apply **standard jet reconstruction and calibration** procedures at CMS to a data set obtained from single fat jet triggers with $p_T^{\text{HLT}} > 450, 500$ GeV ($p_T^{\text{offline}} > 575, 650$ GeV).
2. **Groom** jets with soft-drop (SD) algorithm (cf. introduction).
3. **Classify** events using two different mass-decorrelated jet substructure taggers [9].
4. Use a data-driven approach to **control the large QCD background**.
5. **Unfold** using a maximum-likelihood approach.

Classification

	$P^{\text{PN,DDT}}_{\text{W vs. QCD}}$	
	$\varepsilon_{\text{W jets}}$	$\varepsilon_{\text{backg.}}$
pass	38.20%	4.20%
fail	61.80%	95.80%

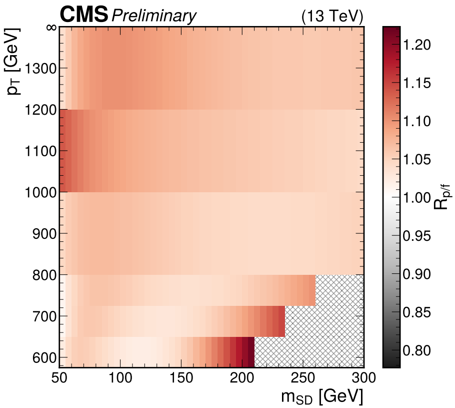


Boosted W-jet groomed mass

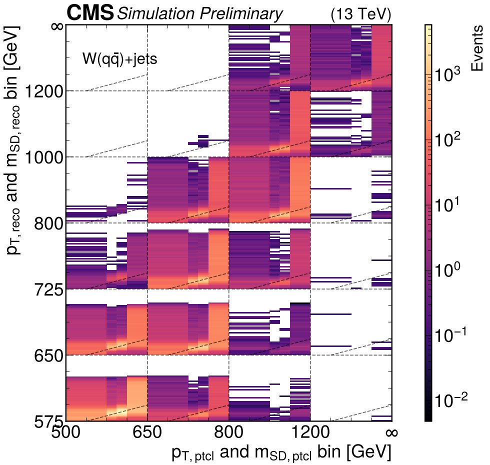
Background

- Shape of m_{SD} for the multijet background is assumed to be the same in the pass and fail regions [10, 11].
  $\longrightarrow$ Ideally, it should only be a single factor.
- However, tagging efficiency varies with p_{T} & mass-decorrelation scheme is not perfect.

  $\longrightarrow$ **2D transfer function**



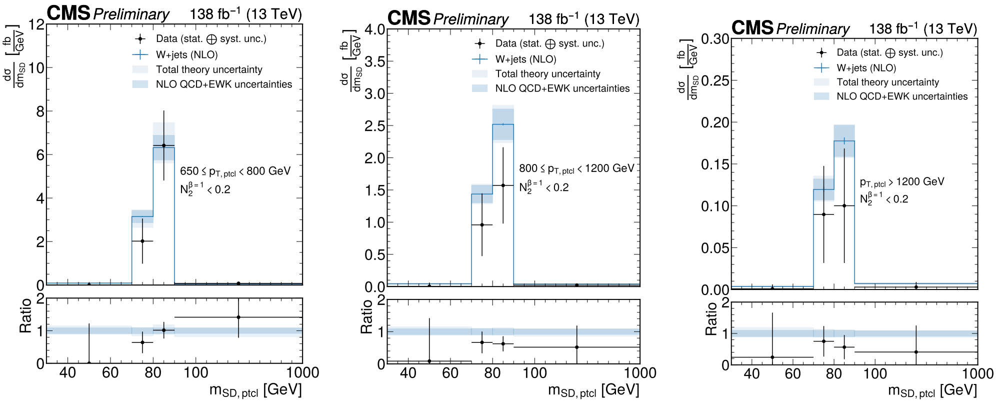
Boosted W-jet groomed mass



Unfolding

- Nominal unfolding using $P^{\mathrm{PN,DDT}}_{\mathrm{W vs.QCD}}$ in the background estimation,
- An explicit matching to a particle-level $W(qq')$ is required, including $N_2^1 < 0.2$ in the particle-level definition.
- Unfolding is implemented as a likelihood function in COMBINE [12], treating all systematic uncertainties with nuisance parameters.
- SVD regularisation is applied (minimising the global correlation coefficient).

Boosted W-jet groomed mass

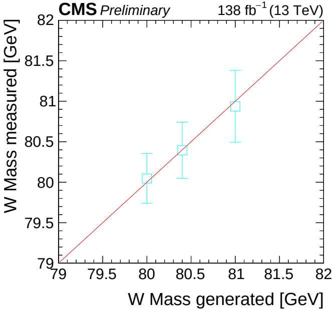


Results

- The background stays the dominant limitation, especially in the tails.
- Both PYTHIA & MADGRAPH+PYTHIA describe the peak reasonably well.

Introduction
N-subjettiness
W-jet mass
Goal
Methodology
Results
Summary & Conclusions
Back-up

Boosted W-jet groomed mass



Extracting the W boson mass

- We generate several PYTHIA 8 samples with different W mass hypotheses: $79.0, 80.0, \mathbf{80.385}, 81.0, 82.0$ GeV
- We exclude the two outermost mass bins, which suffer from large uncertainties from the hadronization model.
- We have checked that the fit closes using MADGRAPH5_AMC@NLO+PYTHIA/HERWIG as pseudodata.
- The final template fit is carried out in $70 < m_{\text{SD}} < 90$:

$$m_{\text{W}} = 80.77 \pm 0.57 \text{ GeV}$$

Summary & Conclusions

Summary & Conclusions

- CMS has just released **two new results** on jet substructure using full Run-2.
- A **better understanding of jet substructure** provides further insights on QCD and helps improve our searches for new physics.
- The first result consists in a measurement of an OC basis of **N-subjettiness** in pure QCD jets, boosted W-jets, and boosted top-jet events.
- The second result extends a series of differential measurements of **jet groomed mass**, here for boosted W-jets, and includes a measurement of the W mass in the hadronic channel.

Merci de votre attention!

Introduction
N-subjettiness
W-jet mass
Summary & Conclusions
Back-up

Summary & Conclusions

- CMS has just released **two new results** on jet substructure using full Run-2.
- A **better understanding of jet substructure** provides further insights on QCD and helps improve our searches for new physics.
- The first result consists in a measurement of an OC basis of **N-subjettiness** in pure QCD jets, boosted W-jets, and boosted top-jet events.
- The second result extends a series of differential measurements of **jet groomed mass**, here for boosted W-jets, and includes a measurement of the W mass in the hadronic channel.

Merci de votre attention!

Back-up

Acronyms I

AK anti k_T algorithm. 11–14

BSM beyond the SM. 20–22

CMS Compact Muon Solenoid. 1, 6, 7, 23–26, 32, 33

ECF energy correlator function. 6, 7

GNN graph neural network. 6, 7

IRC infrared and collinear safe. 5

MC Monte Carlo. 9, 10, 20–22

MET missing transverse energy. 11–14

ML machine learning. 9, 10

OC overcomplete. 9, 10, 32, 33

PS parton shower. 6, 7

QCD quantum chromodynamics. 3, 4, 6, 7, 9, 10, 23–26, 32, 33

SD soft-drop. 6, 7, 23–27, 30

SM standard model. 20–22

SVD singular value decomposition. 28

References I

[1] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: **JHEP** 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[2] Jesse Thaler and Ken Van Tilburg. "Maximizing Boosted Top Identification by Minimizing N-subjettiness". In: **JHEP** 02 (2012), p. 093. DOI: 10.1007/JHEP02(2012)093. arXiv: 1108.2701 [hep-ph].

[3] Andrew J. Larkoski et al. "Soft Drop". In: **JHEP** 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[4] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. "Energy Correlation Functions for Jet Substructure". In: **JHEP** 06 (2013), p. 108. DOI: 10.1007/JHEP06(2013)108. arXiv: 1305.0007 [hep-ph].

[5] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". In: **Phys. Rev. D** 101.5 (2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019. arXiv: 1902.08570 [hep-ph].

[6] Aram Hayrapetyan et al. "Measurement of multidifferential cross sections for dijet production in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$". In: **Eur. Phys. J. C** 85.1 (2025), p. 72. DOI: 10.1140/epjc/s10052-024-13606-8. arXiv: 2312.16669 [hep-ex].

References II

[7] Albert M. Sirunyan et al. "Measurements of the differential jet cross section as a function of the jet mass in dijet events from proton-proton collisions at $\sqrt{s} = 13$ TeV". In: **JHEP** 11 (2018), p. 113. DOI: 10.1007/JHEP11(2018)113. arXiv: 1807.05974 [hep-ex].

[8] Armen Tumasyan et al. "Measurement of the differential $t\bar{t}$ production cross section as a function of the jet mass and extraction of the top quark mass in hadronic decays of boosted top quarks". In: **Eur. Phys. J. C** 83.7 (2023), p. 560. DOI: 10.1140/epjc/s10052-023-11587-8. arXiv: 2211.01456 [hep-ex].

[9] James Dolen et al. "Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure". In: **JHEP** 05 (2016), p. 156. DOI: 10.1007/JHEP05(2016)156. arXiv: 1603.00027 [hep-ph].

[10] Albert M Sirunyan et al. "Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: **Phys. Rev. D** 100.11 (2019), p. 112007. DOI: 10.1103/PhysRevD.100.112007. arXiv: 1909.04114 [hep-ex].

[11] Armen Tumasyan et al. "Search for Higgs Boson and Observation of Z Boson through their Decay into a Charm Quark-Antiquark Pair in Boosted Topologies in Proton-Proton Collisions at $\sqrt{s} = 13\ \mathrm{TeV}$". In: **Phys. Rev. Lett.** 131 (2023), p. 041801. DOI: 10.1103/PhysRevLett.131.041801. arXiv: 2211.14181 [hep-ex].

References III

[12] Aram Hayrapetyan et al. "The CMS Statistical Analysis and Combination Tool: Combine". In: **Comput. Softw. Big Sci.** 8.1 (2024), p. 19. DOI: 10.1007/s41781-024-00121-4. arXiv: 2404.06614 [physics.data-an].

Acronyms
References
Visiting card

Patrick L.S. Connor
patrick.connor@cern.ch
CERN, 40/A1-016