PROCEEDINGS OF SCIENCE

# Jet performance and pileup mitigation in Run 3 in CMS

**Anna Benecke**[a,*] **for the CMS Collaboration**

[a] *UCLouvain, Louvain-La-Neuve, Belgium*

*E-mail:* anna.benecke@cern.ch

With the start of Run 3 data taking the standard pileup mitigation technique for the CMS Collaboration is the Pileup Per Particle Identification (PUPPI) algorithm. Recently, the track-vertex association in PUPPI was optimised in order to recover an inefficiency for hadronically decaying $\tau$ leptons at low transverse momentum ($p_T$). The jet energy scale and resolution are sensitive to many different subdetector systems, making continuous monitoring crucial. It is shown that Run 3 promptly reconstructed data have an excellent jet energy resolution performance with respect to Run 2. Furthermore, a flavor-aware $p_T$-regression was derived and calibrated for the first time in the CMS Collaboration, which is expected to improve the jet energy resolution by up to 18%. Finally, two new algorithms have been developed to identify jets originating from the decay products of heavy particles. The first is a boosted decision tree that identifies hadronically decaying top quarks within variable-size cone jets over a wide $p_T$ range. The second is a neural network that distinguishes jets initiated by positively charged W bosons from those initiated by negatively charged W bosons.

*42nd International Conference on High Energy Physics (ICHEP2024)*
*18-24 July 2024*
*Prague, Czech Republic*

*Speaker





PoS(ICHEP2024)893

## 1. Introduction

For Run 3, the CMS Collaboration [1, 2] aimed to unify the flavor definition of small-cone jets, including the identification of hadronically decaying $\tau$ leptons ($\tau_h$). However, the current pileup mitigation technique demonstrated inefficiencies in reconstructing $\tau_h$. To eliminate such inefficiency, the pileup mitigation technique was optimized [3], as described in Sec. 2.

In addition, the CMS Collaboration is constantly monitoring the jet energy scale and resolution of the incoming Run 3 data. Measurements of jet energy scale and resolution are sensitive to various detector effects, making them crucial tools for monitoring the performance of the calorimeters. In Run 3, thanks to a continuous calibration of the calorimeters during data taking, the promptly reconstructed data performs comparably to the legacy reconstructed data from Run 2 [4]. The jet energy calibration used in Run 3 is presented in Sec. 3, together with a description of the newly developed flavor-aware $p_T$-regression [5], including its calibration and expected performance improvements.

In Sec. 4, two new taggers are presented. Firstly, a boosted-decision-tree (BDT) is trained on top of Heavy Object Tagger with Variable R (HOTVR) [6] jets to improve the identification of jets initiated by hadronically decaying top quarks over a wide $p_T$ range [7]. Secondly, a charge tagger based on the ParticleNet [8] architecture was developed, making it possible to discriminate jets originated by hadronically decaying $W^+$, $W^-$, and $Z$ bosons [9].



## 2. Optimisation of pileup mitigation techniques for hadronic $\tau$ lepton identification

Since Run 3, the Pileup Per Particle Identification (PUPPI) algorithm [10, 11] is the default pileup mitigation technique for small-cone jets. However, the PUPPI algorithm was not optimised for the reconstruction of hadronic $\tau$ leptons, which can exhibit displaced charged tracks that are less

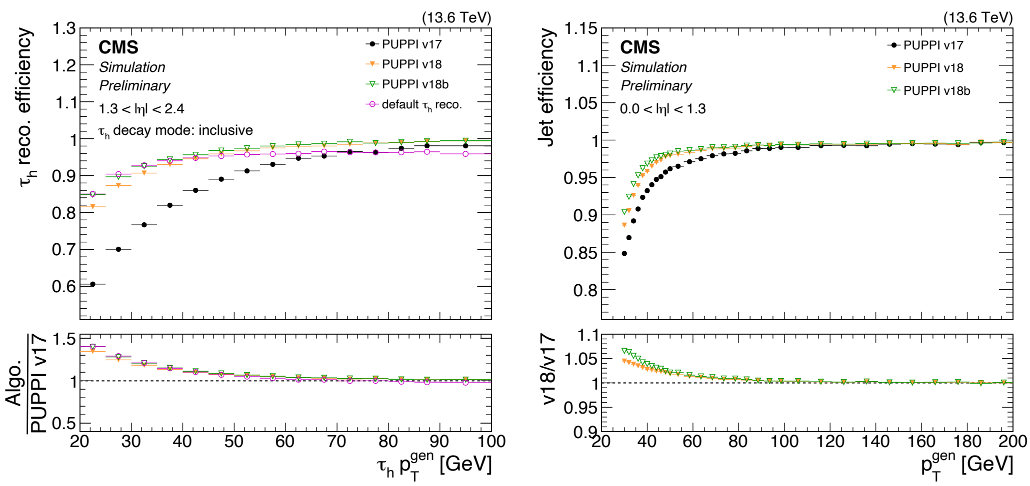


**Figure 1:** (left) Reconstruction efficiency as a function of the generator-level $\tau_h$ $p_T$ of hadronically decaying $\tau$ for the old PUPPI v17, the new PUPPI v18(b) tune. It can be compared to the default $\tau_h$ reconstruction. (right) Jet reconstruction efficiency for the old PUPPI v17 and the new PUPPI v18(b) tune as a function of generator $p_T$. Taken from Ref. [3].

favored in the previous PUPPI versions. Fig. 1 (left) shows a lower reconstruction efficiency of the old PUPPI tune v17 (closed circles) for $\tau_h$ at low $p_T$ with respect to the default $\tau_h$ reconstruction (open circles). Charged particles associated to the $\tau_h$ are displaced from the leading vertex (LV) and therefore often not included in the LV fit, especially when the charged particle has a low momentum. In PUPPI v17, charged particles with $p_T > 20$ GeV are kept even if they are not used in the LV fit. Charged particles with $p_T < 20$ GeV and $|\eta| < 2.4$ are evaluated based on their probability of originating from pileup. Below $p_T$ of 20 GeV and $|\eta| > 2.4$ charged particles are kept only if they satisfy a requirement on the absolute value of their distance along the z-axis $d_z$ to be less than 0.3 cm. In the new PUPPI v18, charged particles that are closest in $d_z$ to the LV or within 700 $\mu m$ from the jet axis originating at the LV are retained if their $p_T$ exceeds 4 GeV. This adjustment leads to significant improvements in the reconstruction of $\tau_h$ at low $p_T$ (closed triangles in Fig. 1 left). At the same time, it is important not to worsen the performance of the reconstruction of jets originating from quarks and gluons. Fig. 1 (right) shows an improvement in the jet reconstruction efficiency, especially at low $p_T$. However, the jet purity is slightly worse by less than 2% at low $p_T$ [3].

## 3. Jet energy scale and resolution in Run 3

The jet energy scale and resolution are sensitive to various detector effects in the tracking and calorimeter systems. Therefore, studying the performance of the jet energy calibration is an important tool for monitoring detector conditions, particularly with the increased pileup and upgraded detector components. In general, jet energy fractions are very stable in Run 3 as can be seen in Ref. [13].

However, during the data taking in 2023, 27 modules in the Barrel Pixel Layers 3 & 4 (representing around 1.5% of the total modules in pixel detector) became inoperable due to an issue in distributing the LHC clock signals. They cover a sector spanning approximately 0.4 radians (around 23 degrees) in azimuthal angle $\phi$ at negative pseudorapidity. This results in a loss of jet energy of up to 6% as can be seen in Fig. 2 (left, orange cross).

To account for this effect, the CMS Collaboration has introduced $\phi$-dependent simulation-based corrections, both offline and online. Data-driven corrections are applied on top of the simulation-based corrections to mitigate residual differences between data and simulation. The time stability of these corrections can be seen in Ref. [4]. As shown in Fig. 2 (right), the jet energy resolution performance from the promptly reconstructed Run 3 data collected in 2023 is better than the jet energy resolution performance from the Run 2 legacy reconstructed data (with improved calibrations applied after data-taking) in the central region $|\eta| < 2.5$, used by most of the analyses in the CMS Collaboration.

In addition, the CMS Collaboration is constantly working on the improvement of the determination of the jet energy scale and resolution. Fig. 3 shows the jet energy resolution after the calibration of the newly developed $p_T$-regression with the ParticleNet architecture. Both the visible and the invisible (including neutrinos) regressions with respect to generator-level jets, excluding and including neutrinos, respectively, are shown. After applying the corrections, a negligible difference between the reconstructed jet response and the simulated jet response is found. In addition, a difference of less than a few percent is observed within the tracker acceptance between data and simulation when using the residual in-situ correction derived from the non-regressed jets [5]. This

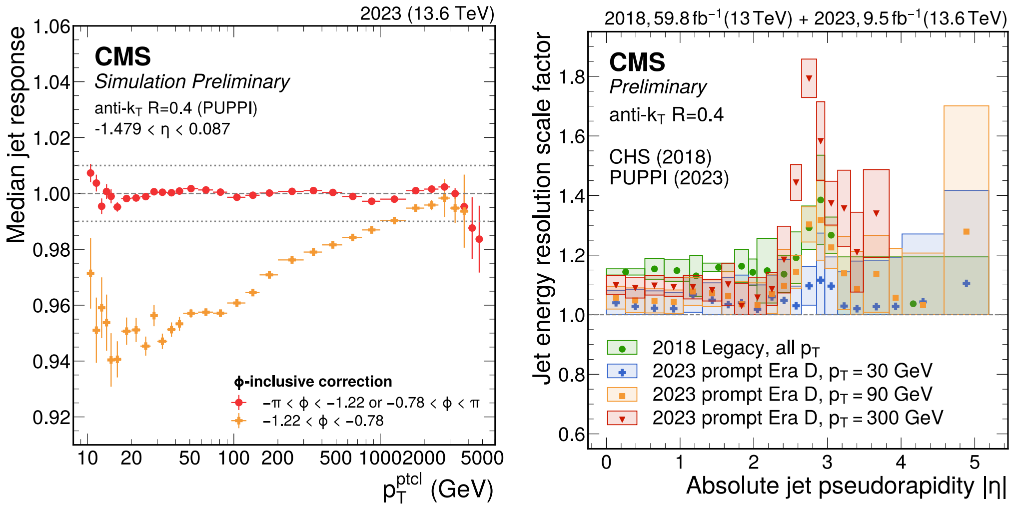


**Figure 2:** (left) Median jet response as a function of the particle-level $p_T$ for jets within the issue region (orange) and outside (red). (right) Jet energy resolution correction factor as a function of $|\eta|$ for 2018 legacy inclusive in $p_T$ and three $p_T$-bins for 2023 prompt data taking. Taken from Ref. [4].

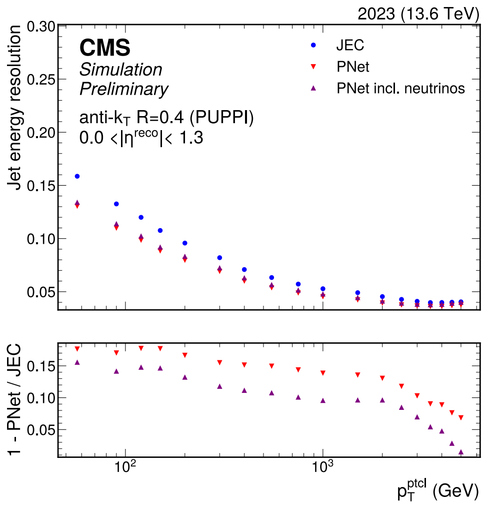


**Figure 3:** (left) Jet energy resolution as a function of the particle-level $p_T$ for the standard jet energy reconstruction (JEC, blue) and the $p_T$-regression including (purple) and excluding (red) neutrinos. Taken from Ref. [5].

shows that no additional artifacts are introduced when using machine learning based $p_T$-regression. A significant improvement in the resolution is expected over the whole $p_T$ range.

## 4. New developments in the identification of large-cone jets

Reconstructing and identifying jets that originated from top quarks is an important task for the CMS physics program. However, the jet radius can vary significantly depending on the $p_T$ of the top quark. The fixed-cone jet reconstruction algorithms typically used by the CMS Collaboration

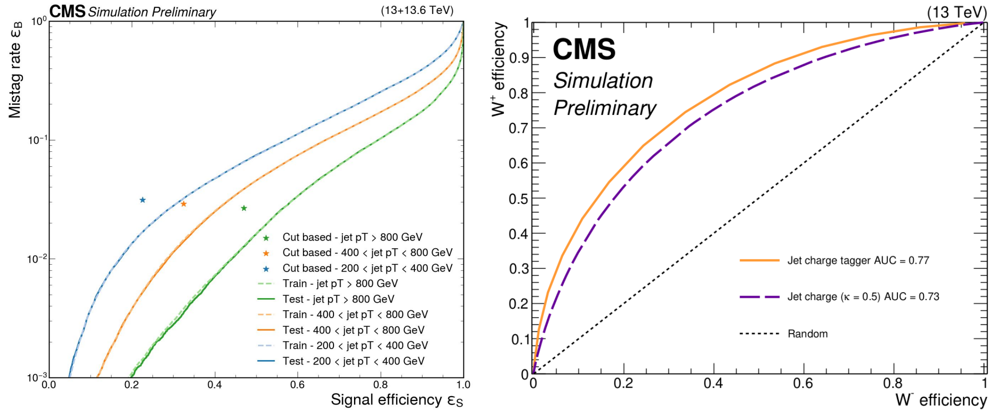


**Figure 4:** (left) ROC curves in three different $p_T$-bins for the HOTVR algorithm with the BDT to identify jets initiated by top quarks. The three stars indicate a working point for the old, cut-based approach. Taken from Ref. [7]. (right) ROC curves comparing the performance to distinguish $W^+$ from $W^-$ for the jet charge (dashed) and the new jet charge tagger (solid line). Taken from Ref. [9].

are therefore not optimal when reconstructing jets initiated by top quarks over a wide $p_T$ range. The CMS Collaboration explored the use of the HOTVR algorithm [6] which is a variable-R jet clustering algorithm. A BDT has been developed to identify the hadronic decays of top quarks within HOTVR jets. In Fig. 4 (left), the Receiver-Operating Characteristic (ROC) curve of the newly trained BDT on top of HOTVR jets can be seen. A significant improvement with respect to the cut-based approach, defined in Ref. [6] is shown. This is especially useful when considering multi-scale challenges like the ones offered by four-top-quark final states [7].

In addition, Fig. 4 (right) presents the ROC curve of a newly developed charge tagger based on the ParticleNet [8] architecture, compared to the ROC curve of a more standard jet charge algorithm. A clear improvement in the performance of distinguishing jets initiated by $W^+$ and $W^-$ bosons can be observed with the jet charge tagger. This tagger can be very helpful when studying vector boson scattering with hadronically decaying W bosons in the final state. The tagger score has a very good data-to-simulation agreement, as detailed in Ref. [9].

## 5. Summary

A summary of the latest improvements and developments in pileup mitigation, jet energy scale and resolution, and jet tagging was presented. The track-vertex association of PUPPI was optimised such that hadronically decaying $\tau$ leptons are reconstructed more effectively [3]. This allows for a unified flavor identification of small-cone jets [12].

Simulation-based, $\phi$-dependent corrections are now available to recover the energy lost due to an issue with the CMS pixel detector. Corrections to mitigate the residual discrepancy between simulation and data are derived from Run 3 promptly reconstructed data. Their performance, especially the jet energy resolution performance, is as good or better than in Run 2 re-reconstructed data [4]. A first look into the full calibration of the $p_T$-regression was shown. A small non-closure of less than a few percent in the tracker acceptance is observed without dedicated residual corrections. A jet energy resolution improvement of up to 18% is expected [5].

A BDT was developed on top of HOTVR jets in order to improve the identification of jets originating from top quarks. A significant improvement is found with respect to the cut-based approach [7]. The ParticleNet architecture was used to develop a charged tagger which is able to differentiate jets originated by $W^+$, $W^-$, and $Z$ bosons. This new method shows an improvement over the standard jet charge reconstruction [9].

## References


[1] CMS Collaboration, *The CMS experiment at the CERN LHC*, DOI:10.1088/1748-0221/3/08/S08004, JINST 3, Aug. 2008

[2] CMS Collaboration, *Development of the CMS detector for the CERN LHC Run 3*, arXiv:2309.05466, JINST 19, May. 2024

[3] CMS Collaboration, *Optimizing the pileup per particle identification algorithm in the context of $\tau_h$ lepton identification in Run3*, CMS-DP-2024-043, Jul. 2024

[4] CMS Collaboration, *Jet Energy Scale and Resolution Measurements Using Run3 Data Collected by CMS in 2022 and 2023 at 13.6 TeV*, CMS-DP-2024-039, Jun. 2024

[5] CMS Collaboration, *Jet energy scale and resolution of jets with ParticleNet $p_T$ regression using Run3 data collected by the CMS experiment in 2022 and 2023 at 13.6 TeV*, CMS-DP-2024-064, Jul. 2024

[6] Lapsien, Tobias and Kogler, Roman and Haller, Johannes, *A new tagger for hadronically decaying heavy particles at the LHC*, arXiv:1606.04961, Eur. Phys. J. C 76, Nov. 2016

[7] CMS Collaboration, *Hadronic top quark tagging with variable-sized jets for the CMS experiment*, CMS-DP-2024-038, Jun. 2024

[8] Qu, Huilin and Gouskos, Loukas, *ParticleNet: Jet Tagging via Particle Clouds*, arXiv:1902.08570, Phys. Rev. D 101, Mar. 2020

[9] CMS Collaboration, *A novel approach for discriminating hadronically decaying $W^+$, $W^-$, and $Z$ bosons in the CMS experiment*, CMS-DP-2024-044, Jul. 2024

[10] Bertolini, Daniele and Harris, Philip and Low, Matthew and Tran, Nhan, *Pileup Per Particle Identification*, arXiv:1407.6013, JHEP 10, Oct. 2014

[11] CMS Collaboration, *Pileup mitigation at CMS in 13 TeV data* , arXiv:2003.00503, JINST 15, Sep. 2020

[12] CMS Collaboration, *A unified approach for jet tagging in Run 3 at $\sqrt{s}$ =13.6 TeV in CMS*, CMS-DP-2024-066, Jul. 2024

[13] CMS Collaboration, *Determination of jet identification criteria with proton-proton collision at 13.6 TeV data collected with the CMS detector at the CERN LHC*, CMS-DP-2024-028, May. 2024