Reconstruction of jets and missing transverse momentum at the CMS experiment: Run 2 and perspective for Run 3

Andrea Malara[a,*]

[a]*Institut für Experimentalphysik, Universität Hamburg,*
Luruper Chaussee 149, Hamburg, Germany

E-mail: andrea.malara@cern.ch

The performance of reconstruction and calibration of jets and missing transverse momentum in the CMS Collaboration, together with the latest tools for pileup mitigation, is presented. The results shown are relative to the data collected from 2016 to 2018 (Run 2) in proton-proton collisions at a centre-of-mass energy of 13 TeV. An outlook to the new techniques foreseen for Run 3 is discussed.

[*]on behalf of the CMS Collaboration

1. Introduction

In the hadronic environment at the LHC, many quarks and gluons are produced, which, due to QCD confinement, create a collimated spray of hadrons, which appear as a cluster of energy deposited in a localised area of the detector, called a jet. Precise calibration of both the energy scale and resolution of jets plays a crucial role across the whole physics programme at the CMS experiment. Similarly, an accurate estimation of the missing transverse momentum (p_T^{miss}) is of crucial importance, for example, in standard model measurements involving the invisible decay products of the top quark, τ lepton and the W, Z and Higgs bosons, as well as beyond the standard model searches targeting new weakly interacting neutral particles.

Moreover, the presence of multiple collisions in the same bunch-crossing (pileup) represents a challenge to the reconstruction and calibration procedure. Several techniques can be used to limit the effect of additional particles. An overview of the respective methods and their impact on the performance is presented in the following.

2. Jet Reconstruction

The event reconstruction in CMS [1] uses a similar scheme for both the online and offline reconstruction, with some difference in the amount of information used and level of corrections applied to achieve fast performance during data-taking. Starting from a local reconstruction, the information from each sub-detector is combined and successively used by the Particle Flow (PF) [2] algorithm to identify the particles produced in each event. Jets are usually clustered, starting from PF candidates combined with pileup mitigation techniques. It is possible to reconstruct several types of jets, based on different algorithms. Furthermore, we define particle-level jets as clustered from all stable ($c\tau > 1$ cm) and visible particles (excluding neutrinos) in simulated events. Finally, it is possible to calibrate the energy of the jets and propagate these corrections onto missing transverse momentum (p_T^{miss}) and apply additional pileup rejection.

3. Pileup mitigation

The instantaneous luminosities reached during data-taking imply multiple proton-proton (pp) collisions to occur in the same bunch crossing. Additional particles coming from the secondary interactions, known as pileup (PU), can deteriorate the measurement since they may be clustered in the reconstructed jets. Hence, the identification of interesting collisions has become an ever-growing challenge at the LHC. During Run 2, the CMS detector collected data with up to 60 interactions per bunch crossing, with an average pileup of 30 interactions (figure 3.1). Moreover, the expected average pileup for Run 3 is even higher, making PU rejection an even more challenging task.

The CMS Collaboration uses a variety of techniques for PU mitigation. One example is the Charged Hadron Subtraction (CHS) algorithm, which has been the standard method in Run 2. It uses the information from the tracker to remove the charged particles that are associated with a pileup vertex from the jet clustering procedure. Due to its limited coverage in η, outside the tracker no information on the charge of a particle is available; consequently, dedicated jet energy corrections are applied to account for the impact of charged PU outside the tracker coverage, and of neutral PU everywhere.
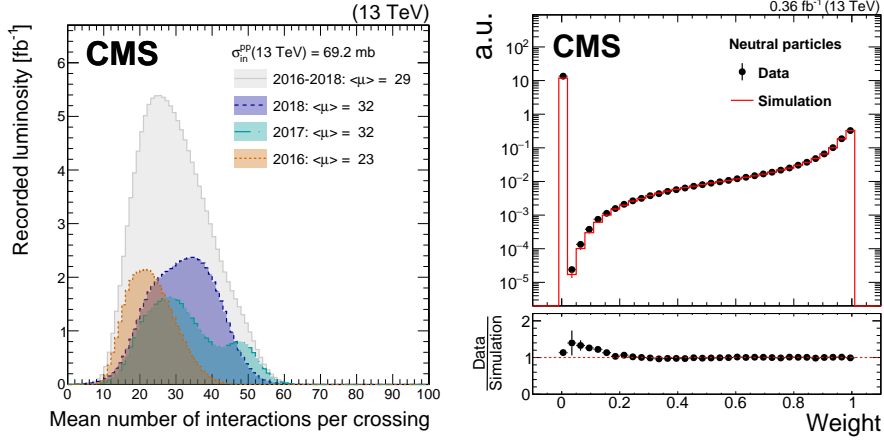
Figure 3.1: Left: Distribution of the mean number of inelastic interactions per crossing (pileup) in data for pp collisions [3]. Right: Data-to-simulation comparison of the PUPPI weight distribution for neutral particles [3].

This approach is limited since the additional corrections act on the four-momentum and not on the jet shape or substructure. To overcome this limitation, an alternative technique for PU mitigation, pileup per particle identification (PUPPI), is introduced [3]. It calculates, event by event, a probability that each particle originates from the leading primary vertex and scales the energy of these particles based on that probability (figure 3.1). As a consequence, objects clustered from hadrons, such as jets, p_T^{miss}, and lepton isolation, are expected to be less susceptible to PU when PUPPI is used. In figure 3.2, a schematic representation of CHS and PUPPI is shown.
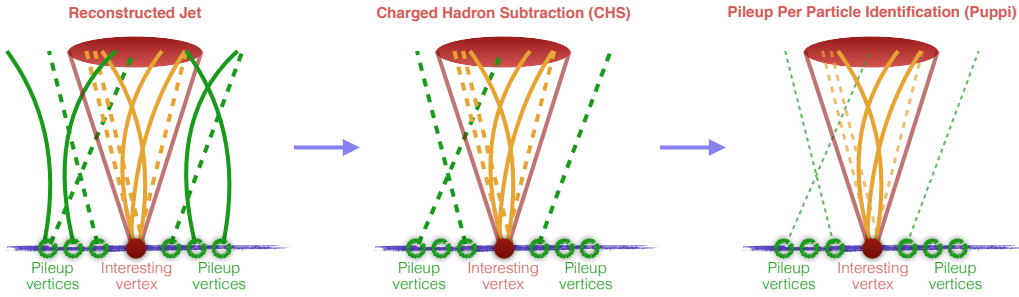

Figure 3.2: Sketch of PU suppression techniques. Solid (dashed) lines refer to charged (neutral) PF candidates. The weights applied by the PUPPI algorithm are represented by thin lines.

These techniques are complementary, as highlighted in figure 3.3. Inside the tracker acceptance, PUPPI has a good performance in both efficiency and purity, defined as the fraction of reconstruction-level jets with $p_T \geq 30$ GeV that match within $\Delta R \leq 0.4$ with a particle-level jet with $p_T \geq 20$ GeV. In contrast, for CHS, even though the efficiency is nearly close to 100%, the purity is significantly reduced at high pileup. To improve the purity, but at the cost of a reduction in efficiency, one can apply the pileup jet ID, a boosted decision tree based technique to identify low-p_T jets coming from PU [4]. At high values of $|\eta|$, the purity drops more rapidly in all the cases, and, even though PUPPI performs better than CHS only, the usage of pileup ID on top of

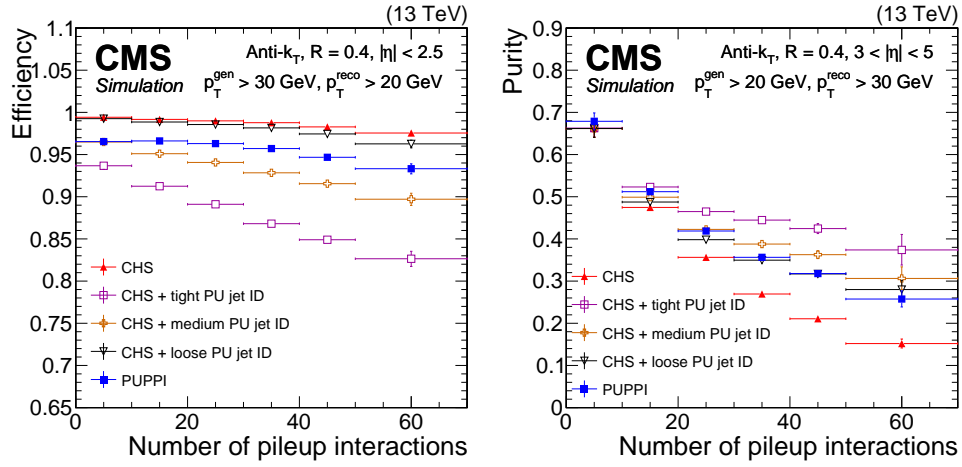CHS improves the performances compared to PUPPI.

Figure 3.3: Efficiency and purity as a function of the number of interactions in different η bins for PUPPI, CHS and CHS+PU jet ID at different working points [3].

4. Jet energy calibration

For jet energy calibration, CMS uses a factorised approach. The first step consists of PU subtraction. It is a simulation-based correction that aims to remove the average energy offset coming from PU, monitored for each type of PF candidate (see figure 4.1). The calibration procedure heavily relies on the second step, which corresponds to the calibration of the jet energy response (see figure 4.1). Here, the primary goal is to account for detector effects that introduce a discrepancy in the measured energy between the particle-level jets and the reconstructed ones.
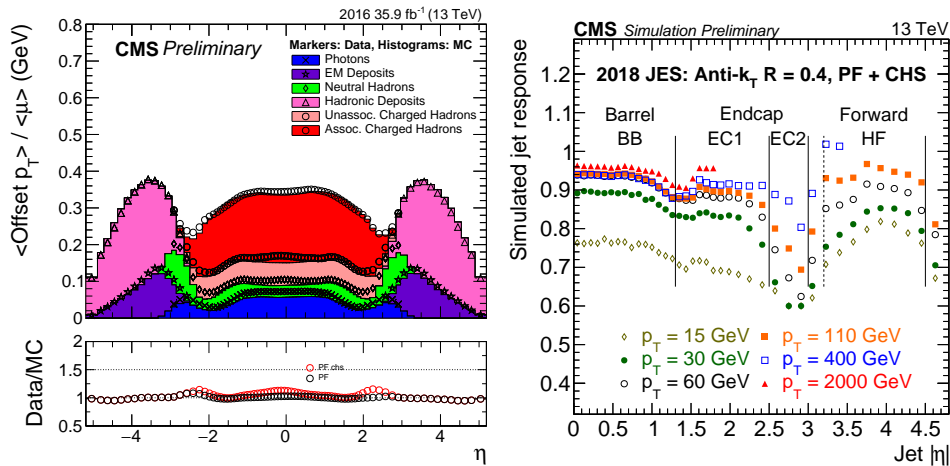

Figure 4.1: Left: Data-to-simulation comparison for average offset per pileup interaction, calculated for each type of PF candidate. Right: Jet response corrections derived as a function of $|\eta|$ and p_T. Changes in performance at high $|\eta|$ and low p_T are caused by detector acceptance. Taken from [5].

Residual differences between data and simulation are then corrected in two steps: an $|\eta|$-dependant correction, to correct the different response of each sub-detector with respect to the central, better-calibrated part of the detector; a p_T-dependant correction, mitigating the absolute scale difference in the central region of the detector.

Several techniques involving precisely calibrated reference objects ($Z \to \mu\mu$, $Z \to ee$ and γ) are used in a global fit to reach a precision at a percent level, as shown in figure 4.2. After the jet energy scale has been corrected, the jet energy resolution (JER) in simulation needs to be adjusted to match the jet resolution in data. Therefore, scale factors (SFs) are applied in simulation to broaden the detector response distribution (figure 4.2). In the $\eta \in [2.5, 3]$, the imperfect calibration of overlapping sub-detectors leads to larger SFs.
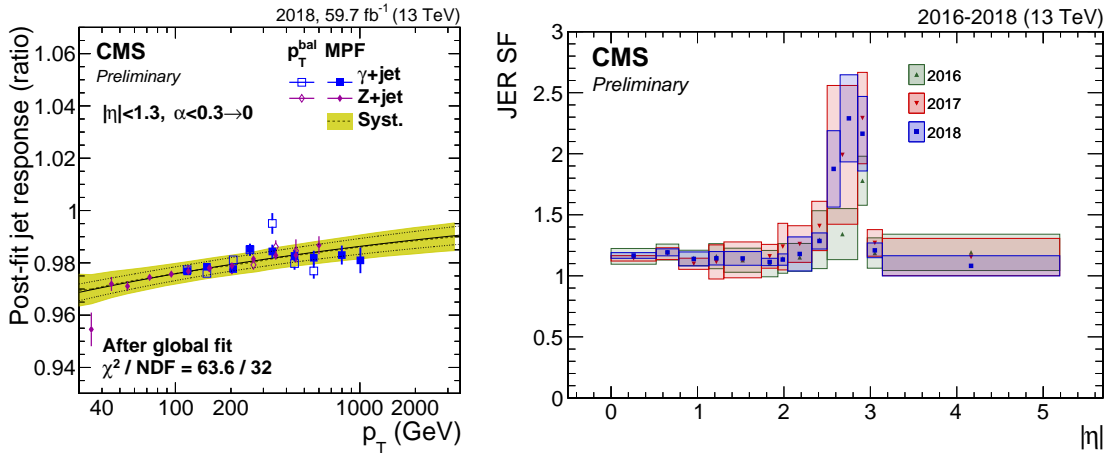

Figure 4.2: Left: Data-to-simulation p_T-dependant residual corrections for the jet response. Right: $|\eta|$-dependant data-to-simulation jet energy resolution scale factors. Taken from [5].

5. Missing transverse momentum

Missing transverse momentum arises from invisible particles, either from standard model neutrinos or new physics, not interacting with the detector. It is estimated from the momentum conservation in the transverse plane using the visible part of the event, which, if miscalibrated, can lead to an inaccurate estimation of p_T^{miss} [6]. Furthermore, anomalous high-p_T^{miss} events can appear from detector effects. Figure 5.1 shows the effectiveness of several mechanisms adopted to suppress these spurious events in data. The p_T^{miss} response and resolution can be assessed in events either with an isolated photon or with a Z boson decaying to a pair of electrons or muons, balancing the momentum of the vector boson to the hadronic recoil system ($\vec{u_T}$), where no genuine p_T^{miss} is expected. The performance of the PUPPI algorithm on p_T^{miss} is presented in figure 5.1. A direct improvement in the resolution of transverse mass in W+jets events, where genuine p_T^{miss} is expected, is observed [6]. Also in this case, PUPPI shows better stability against PU, and the gain at the expected PU for Run 3 is significant.
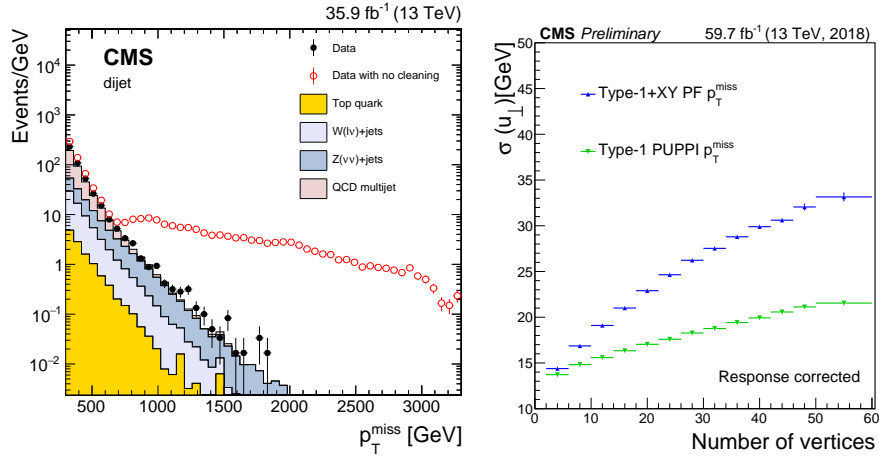
Figure 5.1: Left: p_T^{miss} distribution with the event filtering algorithms applied on a dijet selection [6]. Right: The resolution of the hadronic recoil in γ+jets events as a function of the number of vertices for different PU suppression algorithms. Results for Run 2 are shown in [7].

6. Summary

Several techniques for jet and p_T^{miss} reconstruction and calibration adopted by the CMS Collaboration during Run 2 are presented. The improved performance of pileup mitigation algorithms developed in this period will allow coping with the harsher conditions in the upcoming Run 3, during which PUPPI will be used as the default algorithm for pileup suppression.

References

[1] CMS Collaboration, *The CMS experiment at the CERN LHC*, *Journal of Instrumentation* **3** (2008) S08004 https://doi.org/10.1088

[2] CMS Collaboration, *Particle-flow reconstruction and global event description with the cms detector*, *Journal of Instrumentation* **12** (2017) P10003–P10003 http://dx.doi.org/10.1088/1748-0221/12/10/P10003.

[3] CMS Collaboration, *Pileup mitigation at CMS in 13 TeV data*, *JINST* **15** (2020) P09018 [`arxiv:2003.00503`].

[4] CMS Collaboration, *Performance of the pile up jet identification in CMS for Run 2*, https://cds.cern.ch/record/2715906.

[5] CMS Collaboration, *Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018*, https://cds.cern.ch/record/2715872.

[6] CMS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$ = 13 TeV using the CMS detector*, *JINST* **14** (2019) P07004 [`1903.06078`].

[7] CMS Collaboration, *Performance of missing transverse momentum reconstruction in events containing a photon and jets collected by CMS during proton-proton collisions at $\sqrt{s}$ = 13 TeV in 2018*, https://cds.cern.ch/record/2723010.