XVth International Conference on Calorimetry in High Energy Physics (CALOR2012) IOP Publishing
Journal of Physics: Conference Series **404** (2012) 012045 doi:10.1088/1742-6596/404/1/012045

Pileup measurement and mitigation techniques in CMS

A Perloff, on behalf of the CMS collaboration

Graduate Student, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA

E-mail: `aperloff@physics.tamu.edu`

Abstract. When trying to reconstruct an event from a hard-scatter *pp* collision in CMS, it is of the utmost importance to correctly measure the energy from jets. The jet energy corrections (JEC) correct, on average, the energy of the reconstructed jets back to the energy of the final-state particles that initiated the jets. This effort is hindered by additional energy in the jets coming from other soft *pp* collisions. The additional energy is termed *pileup* or *offset* and comes from everything except the primary vertex (PV) and its underlying event (UE). In this paper, we describe how this pileup energy is measured and parametrized as well as the techniques used to remove this extra energy from the reconstructed jets.

1. Introduction

During the reconstruction of a *proton-proton* (*pp*) collision (event), jets are often reconstructed with a p_T that differs from that of the final-state particles that make up the jet. The jet energy corrections (JEC) correct the reconstructed jet energy back to the true energy of the final-state particles. All analyses at CMS which make use of jets must use the JEC. Therefore, the CMS collaboration has developed a factorized approach to these JEC which is composed of multiple levels representing corrections for various physics or detector effects and allows enough flexibility in the corrections to be applicable to many types of analyses [1–3].

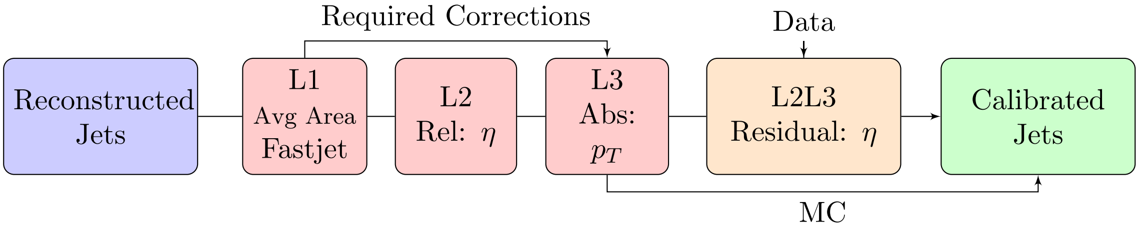

Figure 1. A graphical depiction of the jet energy correction levels required for all CMS analyses.

One of the physics effects that the JEC seek to correct for is the additional energy attributed to the jets which comes from *pp* interactions other than the hard-scatter event at the primary vertex (PV). The additional energy from these events is often called *pileup* or *offset* and will be the focus of this paper.

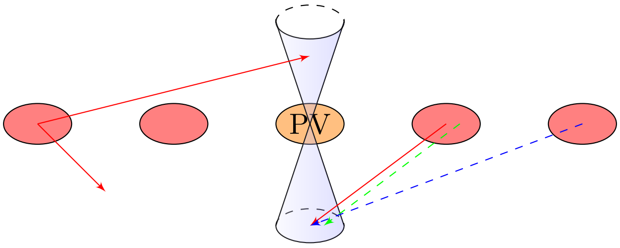

Figure 2. A basic graphic showing two jets coming from a PV with additional *pp* interactions in red. The arrows coming from the red *pp* interactions are particles which add additional energy to the jets during reconstruction. The different colors represent different types of particles (ex: photons, charged hadrons, etc.)

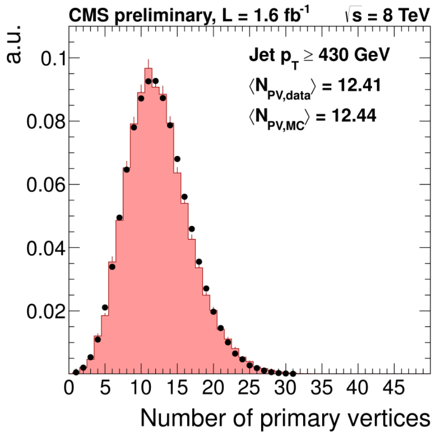

Figure 3. Distribution of the number of primary vertices (per event) in data and MC in 2012.

There are three major classifications of pileup based upon the time at which the additional energy enters the calorimeter system. In-time (IT) pileup refers to energy from *pp* collisions in the current bunch-crossing (BX) other than that at the hard scatter PV. This is the largest source of pileup energy. In addition, there is early out-of-time (EOOT) pileup, which refers to energy left in the calorimeters from previous BXs, and late out-of-time (LOOT) pileup, which refers to energy from later BXs that is integrated with the current event's energy.

2. Pileup Measurement

While pileup itself cannot be directly measured, it can be correlated to various other directly measurable quantities. As pileup comes from additional *pp* interactions, the number of PV (N_{PV}) is directly correlated to the amount of pileup; the greater the N_{PV}, the more pileup energy is added to the jets. The total offset for the event is equal to the average energy from a non-hard-scatter PV times the N_{PV}. The ability to correlate pileup with N_{PV} is predicated on the CMS detector having a high reconstruction efficiency for PVs. In creating the corrections we use the cuts $N_{dof} > 4$ and $IsFake = false$ to make sure we are measuring only good PV. Figure 3 shows the good agreement between the N_{PV} in data and in MC [4].

Another quantity which is directly correlated with the amount of pileup energy is the median energy per area (ρ), which is calculated using the Fastjet algorithm [5]. In order to make this correlation, pileup must be approximated as a homogeneous noise which can be subtracted off.

3. Pileup Corrections

The JEC corrections at CMS can be separated into two categories, those done at the hardware level and those done in software. Detector level corrections can be timing changes, threshold level modifications, or any other change to the hardware that might reduce the actual amount of pileup energy within the jets during reconstruction. Software level corrections are performed after reconstruction and scale jet p_T up or down based on the *measured* amount of pileup.

3.1. Detector Level

Detector level mitigation techniques have been extremely effective in reducing the amount of out-of-time (OOT) pileup. The barrel and endcap hadronic calorimeter system (HBHE) used four time-slices (TS=25 ns) as its integration time during 2011. The measured LOOT pileup contributed about 20% of the pileup energy with EOOT pileup contributing another 2-4%. In

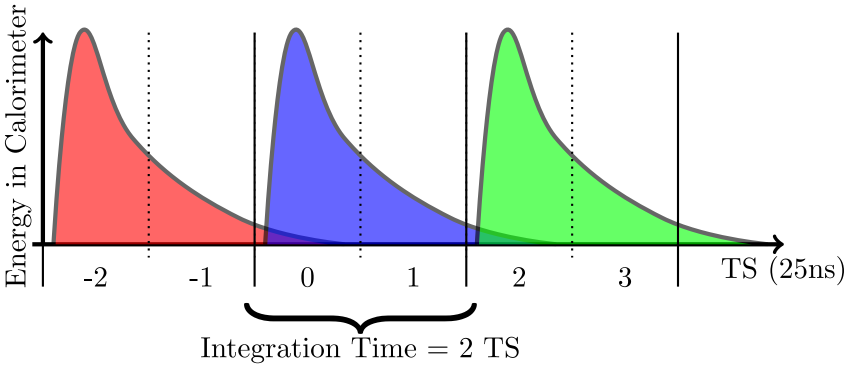

Figure 4. A graphic showing, on the most basic level, the energy deposited in the barrel and endcap hadronic calorimeters (HBHE). The central curve is the current event with the left and right curves being the events before and after, respectively.

2012, the integration time in the HBHE system was changed to two TS, which reduced the LOOT pileup to almost zero and also reduced the EOOT pileup. This change is depicted in figure 4. With four TS starting from TS 0, the energy that is integrated will contain all of the energy from the current event as well as a significant portion of the energy from the next event. The two TS scheme includes little to none of the LOOT pileup. This is a significant improvement as it not only reduces one of the contributors to the pileup to nearly zero, but it also means that the rest of the pileup energy is more easily correlated with the IT N_{PV}.

3.2. Software Level

Software level pileup mitigation techniques are essential to most analyses. The purpose of these offset correction techniques is to measure and remove the extra energy inside of the jets that is not associated with the high-p_T vertex and its UE [1–4; 6].

3.2.1. Average Offset The average offset (AO) method seeks to determine the average amount of energy added to an event due to low-p_T scatters (pileup). The correction is calculated using three different samples; a Zero-Bias data dataset, a Zero-Bias MC sample that is made by overlaying a neutrino gun sample with Minimum Bias events, and a QCD MC sample. In each of the three samples, the offset ($p_{T,offset}$) is measured on an event-by-event basis by finding the p_T deposited in a jet with a certain cone size (usually 0.5 or 0.7 at CMS), centered at a specific $\eta - \phi$ location. The ϕ direction of the *jet* is chosen at random and a scan of η, in steps of 0.1 within $|\eta| < 5.0$, is performed for each event. For calorimeter (Calo) jets, the energy in the jet is the sum of the energy inside the calorimeter towers whose $\eta - \phi$ location is inside the radius of the jet cone. For particle flow (PF) jets, the energy is that of the sum of the PF candidates inside the jet cone [6]. This procedure is called the *Random Cone Algorithm.*

These results are then binned in 1-D histograms, one for each value in $|\eta|$, with N_{PV} on the x-axis and the $< p_{T,offset} >$ on the y-axis. Figure 5 shows a sample plot of the AO versus N_{PV} for PF jets in $2.0 < |\eta| < 2.1$. The offset, shown in figure 5, is modeled by:

$$Offset(N_{PV}) = [0] + [1](N_{PV}) + [2](N_{PV})^2 \tag{1}$$

As expected, the value at the y-intercept of the line ($N_{PV} = 0$) is the p_T contribution of the OOT pileup. Whereas the first and second order terms in N_{PV} are dominated by IT pileup, with a small contribution from OOT pileup.

The correction to the raw, uncorrected jet p_T (p_T^{RAW}) is then derived to be a multiplicative scale factor which, when applied to the p_T^{RAW}, removes the added pileup energy. The scale factor for the AO method is:

$$C_{AverageOffset}(N_{PV}, p_T^{RAW}, \eta) = 1 - \frac{< Offset(N_{PV}, \eta) >}{p_T^{RAW}} \tag{2}$$

Once applied, the scale factor should correct the p_T^{RAW} of the jet back to the true (generator) p_T:

$$p_T^{COR} = p_T^{GEN} = C_{AverageOffset}(N_{PV}, p_T^{RAW}, \eta) \cdot p_T^{RAW} \tag{3}$$

3.2.2. Hybrid Jet Area When using the AO method, one assumes that every jet contains the same amount of additional energy from pileup. Therefore, the best that the AO corrections can do is to correction on an event-by-event basis. The jet area (JA) method takes into account variations in the energy from pileup on an event-by-event basis and on a jet-by-jet basis by using a new parametrization [5; 7]. The hope is to improve the jet energy resolution and gain adaptability for future pileup conditions.

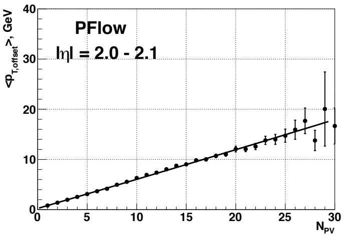

Figure 5. The average PF jet offset as a function of N_{PV} within $2.0 < |\eta| < 2.1$.

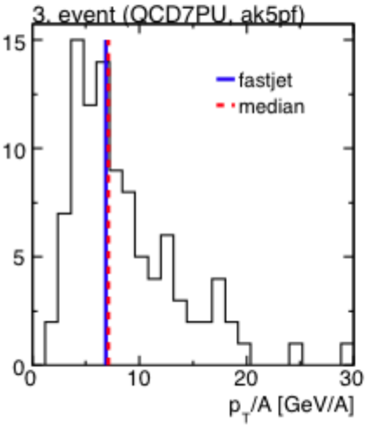

Figure 6. Distribution of p_T/A for all of the jets in a single event.

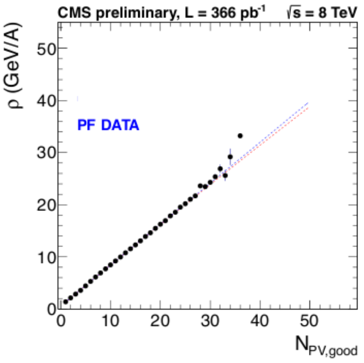

Figure 7. Distribution of ρ versus N_{PV} in data.

The new parametrization was developed based on the Fastjet algorithm. First the area (A_j) is calculated for each jet in the event. The next quantity needed is the median energy density of an event, ρ. This quantity is defined as the median of the distribution of p_{Tj}/A_j, where j is the jet index. The ρ computation is uses the k_T jet clustering algorithm (R=0.6) because of its well-defined energy clustering and is insensitive to jets from a hard scatter as seen in figure 6. The sparsely populated values around $p_T/A = 30$ are the hard jets in the event (higher p_T per unit area), whereas the large clump of jets lower in the spectrum represent the jets from pileup.

While the JA method is more granular and can correct on a jet-by-jet basis, the Fastjet algorithm used to parametrize such corrections does not know that the response of the detector is η dependent. In the AO corrections, this was taken into account by binning the corrections in η. Therefore, a method was developed that combines the η parametrization benefit of the AO method with the jet-by-jet granularity and parametrization of the JA method. This new method was called the hybrid jet area (HJA) method.

We start by parametrizing ρ in terms of N_{PV}, as is done in figure 7. Once this relationship is obtained, the formula is inverted to obtain $N_{PV}(\rho)$. This relationship is then inserted into the function for the offset derived above.

$$Offset(N_{PV}(\rho)) = [0] + [1](N_{PV}(\rho)) + [2](N_{PV}(\rho))^2 \tag{4}$$

This function has all of the same characteristics, irrespective of the parametrization, as the equation derived by the AO method. However, the HJA scale factor

$$C_{Fastjet-based}(\rho, A, p_T^{RAW}, \eta) = 1 - A\frac{<\frac{Offset(\rho),\eta}{A}>}{p_T^{RAW}} \tag{5}$$

uses the JA to correct on a jet-by-jet basis. As seen in figure 8, both of these methods measure the pileup energy almost equivalently, on average [4].

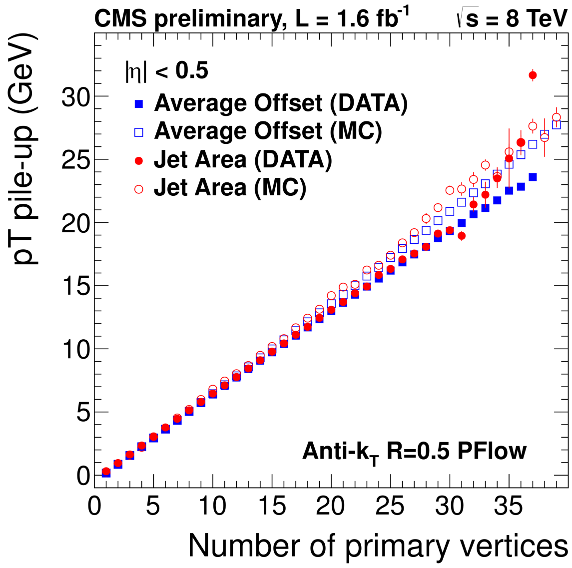

Figure 8. Comparison of the pileup measured in PF jets using AO method and the HJA method.

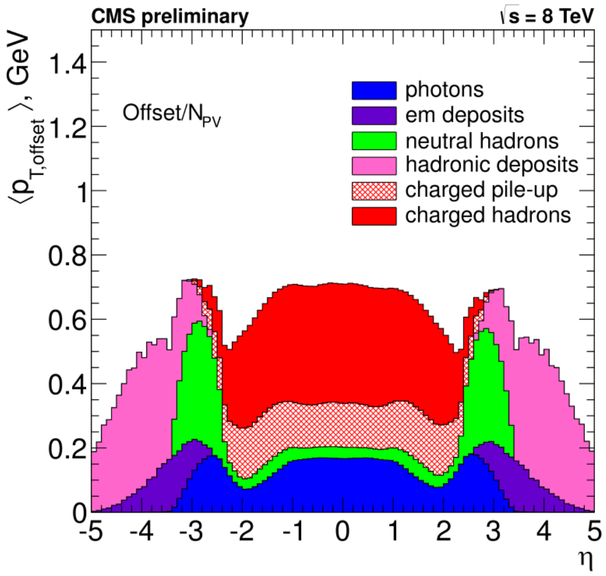

Figure 9. The particle flow composition of an $N_{PV} = 1$ event in MC.

3.2.3. Charged Hadron Subtraction Charged hadron subtraction (CHS) is an additional pileup subtraction tool that can be used in conjunction with either of the methods above, but which requires the use of particle plow jets. CHS uses the CMS detector's excellent tracking capabilities to identify and remove jet constituents (charged hadrons) which are known to have originated from pileup vertices. The algorithm does not remove unassociated tracks, as they may have originated from the high-p_T vertex.

The benefit to applying this additional algorithm to the jet corrections is that it lowers the amount of pileup energy that the AO or HJA methods must compensate for, thus lowering the multiplicative scale factors. Furthermore, CHS has the additional advanatage of being a particle-by-particle pileup subtractions technique, while at best the other algorithms work on an event-by-event or jet-by-jet basis.

3.3. Closure

In order to determine the efficacy of the corrections described above, We study the response of the detector to the jets, defined as $R =< p_T^{COR}/p_T^{GEN} >$. If the pileup corrections were ideal, we would expect the response for the jets to be identical at all N_{PV}, meaning that there was no response dependence upon the pileup conditions inside the detector. In figure 10 the left plot shows the response as a function of the GenJet p_T for PF jets. The plot clearly shows that the response for low p_T jets is not equal to 1 (R=1 being a perfect response). In addition, the

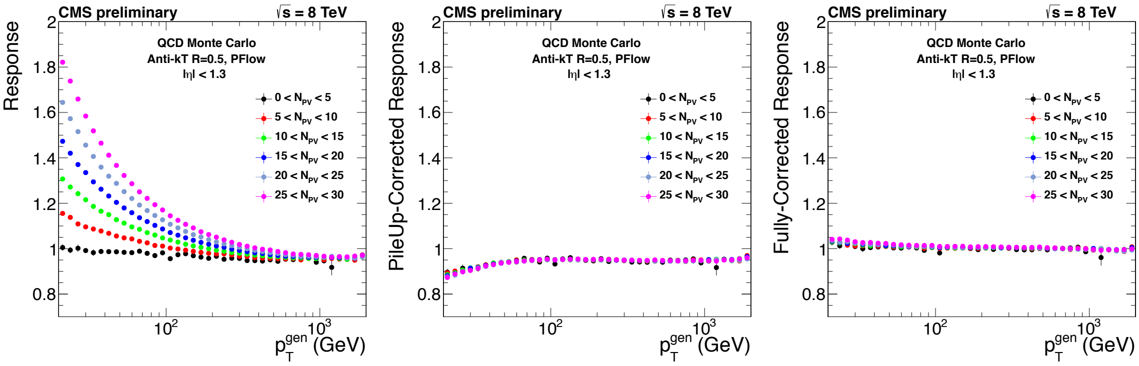

Figure 10. Response versus GenJet p_T for PF jets (left) **before** correcting for pileup, (middle) **after** correcting for pileup, and (right) after the required MC correction chain. The same plots can be made for the Calo jets with similar results.

response varies significantly for differing values of N_{PV}. However, the right plot show the same thing after the pileup corrections have been applied. One can see not only a better response at low p_T, but that there is no N_{PV} (pileup) dependence. Because all of the pileup energy has been removed from the jets, they also have a significantly lower response at low p_T. Any non-closure still present is due to various other effects, such as the detector's η and p_T dependencies, which are corrected for using additional steps in the factorized jet correction chain [3; 4].

4. Conclusion

All of the various pileup removal techniques discussed in this paper have one goal, remove the dependence on pileup energy from the jet response. The first level of mitigation against pileup comes during the detector level readout and reconstruction efforts (ex: time-slices in the HBHE system). These have been shown to provide significant improvements in reducing OOT pileup.

The second level of pileup mitigation was at the software level and came in several of varieties. The AO method is the recommended correction for Calo jets as there is empirical evidence for slightly higher resolution in Calo jets when using this method. The Fastjet-based, HJA method is used for most analyses in CMS and is the recommended correction for PF jets. The key abilities of this method were its adaptability and per event/jet granularity. CHS is an additional tool which is very promising when used in conjunction with the other pileup mitigation techniques.

Once applied, there is good closure over most of the jet p_T spectrum (in MC). Any remaining non-closure in the jet response is well understood and is removed at later stages of the factorized JEC chain [6].

References

[1] The CMS Collaboration 2010 CMS PAS JME-10-003

[2] The CMS Collaboration 2010 CMS PAS JME-10-010

[3] The CMS collaboration 2011 *J. of Inst.* **6** P11002 URL `http://stacks.iop.org/1748-0221/6/i=11/a=P11002`

[4] The CMS Collaboration 2012 CMS DP-2012/012

[5] Cacciari M and Salam G P 2008 *Phys. Lett.* B **659** 119 – 126 ISSN 0370-2693 URL `http://www.sciencedirect.com/science/article/pii/S0370269307011094`

[6] Kirschenmann H, on behalf of the CMS collaboration 2012 *JPCS*

[7] Cacciari M, Salam G P and Soyez G 2008 *JHEP* **2008** 005 URL `http://stacks.iop.org/1126-6708/2008/i=04/a=005`