The CMS Particle Flow Algorithm

Milos Dordevic[1,2]on behalf of the CMS Collaboration [*]

[1]Vinca Institute of Nuclear Sciences, PO Box 522, 11001 Belgrade, Serbia

[2]University of Belgrade, 1 Studentski trg, 11000 Belgrade, Serbia

Abstract. The event reconstruction at the Compact Muon Solenoid (CMS) experiment at the CERN Large Hadron Collider (LHC) is predominantly based on the Particle Flow algorithm. This algorithm for a global event description uses the information from all subdetector systems, unlike the previous, traditional approaches that were focused on the localized information in each subdetector. These traditional methods use the raw information (tracks, hits), while the Particle Flow algorithm completely reconstructs the event by identifying and reconstructing the comprehensive list of final-state particles (photons, electrons, muons, charged and neutral hadrons), resulting in superior reconstruction of jets, missing transverse energy, tau leptons, electrons and muons. This approach also allows for efficient identification and mitigation of the pileup effect. The concept and performance of the Particle Flow algorithm, together with the prospects for its development in the context of the upgraded CMS detector, are presented in this overview.

1 Introduction

The traditional event reconstruction at hadron colliders implied that jets consist of hadrons and photons, having their energy measured exclusively by the calorimeters, without identification of individual particles within jets and without using the information from tracker and muon detectors. The missing transverse momentum (p_T^{miss}) was reconstructed following the same approach. The reconstruction of isolated photons and electrons was performed mostly using the electromagnetic calorimeter, while the tagging of jets from hadronic tau decays and b quark hadronization was based mostly on the information from the tracking detector.

The Particle Flow (PF) algorithm [1] is based on a concept of global event reconstruction as it performs a correlation of the basic elements (i.e. tracks and clusters) obtained from all subdetector systems, in order to identify all particles in the event and measure their properties. An example is shown in Fig. 1 and Fig. 2 where a jet with transverse momentum of 65 GeV, simulated in the CMS detector [2], is made of only five particles: two charged hadrons (a π^+ and a π^-), two photons (from the decay of a π^0) and one neutral hadron (a K_L^0). Charged hadrons are identified by linking the track to one or more calorimeter clusters, while photons and neutral hadrons are in general identified by calorimeter clusters that are not linked to the track. Electrons are identified by linking the track and the cluster in the electromagnetic calorimeter, without any signal left in the hadronic calorimeter, while muons are identified as a track in the tracker detector that is linked with the corresponding track in the muon stations.

---

[*]e-mail: milos.dordevic@cern.ch

The PF algorithm was first derived and applied at the ALEPH experiment of the LEP collider [3], and today it is driving the design of future experiments. A key component is a fine spatial granularity of detector layers. From the list of reconstructed PF particles it is possible to obtain superior physics objects with higher efficiency, resolution and reduced misidentification rate. At the CMS experiment, the PF algorithm was first studied using MC events in 2009, then it was commissioned with data and applied in physics analyses in 2010, followed by the implementation in the High Level Trigger (HLT) system and heavy ion collision analyses in 2011. The CMS detector is proven to be very suitable for the usage of the PF algorithm, having a high magnetic field to separate neutral from charged hadrons, fine granularity of the tracker detector for jets with transverse momenta up to 1 TeV, a highly segmented electromagnetic calorimeter (ECAL) to determine the energy fractions in jets, followed by a hermetic hadronic calorimeter (HCAL), as well as an excellent system to identify muon tracks.
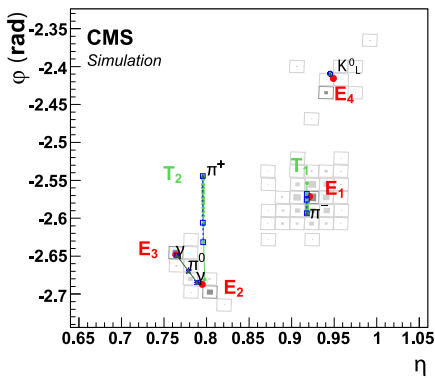
Figure 1. Event display of a jet made of five particles only in the (η,ϕ) view on the ECAL surface, presenting charged particle tracks, clusters and track-cluster links corresponding to particles in the jet [1].
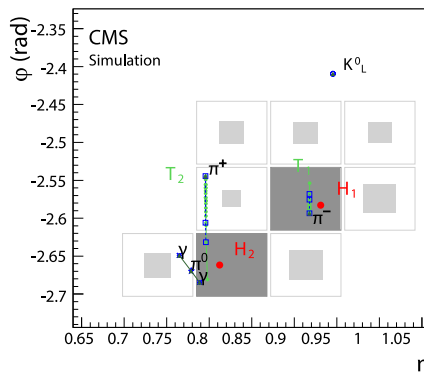
Figure 2. Event display of a jet made of five particles only in the (η,ϕ) view on the HCAL surface, presenting charged particle tracks, clusters and track-cluster links corresponding to particles in the jet [1].

2 Reconstruction of the particle-flow elements

The reconstruction of the trajectories of charged particles in the inner tracker represents the initial step of the PF algorithm. Charged-particle track reconstruction was originally aimed at measuring the momenta of high-energy isolated muons, identifying isolated hadronic tau decays and at tagging b quark jets. With the traditional approach, even if the misidentification rate was at the order of percent, the efficiency of charged pion reconstruction amounted to only 70-80% for tracks with transverse momenta above 1 GeV, compared to 99% for isolated muons. Each charged hadron missed by the tracking algorithm would be detected (if at all) as a neutral hadron. The tracking inefficiency could be reduced by accepting the tracks with smaller transverse momentum and with fewer hits, but this comes at the expense of an exponential increase of the combinatorial rate of misreconstructed tracks. The solution to increase the efficiency while keeping the misidentification rate at a similar level was the iterative tracking algorithm. At each step (iteration) qualitative criteria are added to the tracks, selected hits

are then masked, keeping the misidentification rate at a low level, while progressively adding to the overall efficiency. A total of 10 iterations is used, each targeting a particular track type having a different origin, in order to achieve the highest efficiency. The comparison between the traditional (black), prompt (green) and the full iterative tracking algorithm (red) efficiencies and misidentification rates is presented in Fig. 3 and Fig. 4.

3 Particle identification and reconstruction

A given particle, in general, gives rise to several PF elements in the various CMS subdetectors. Therefore, the reconstruction of a particle proceeds with a link algorithm that connects the PF elements from different subdetectors. The link algorithm is limited in terms of the CPU time and is thus restricted to the nearest neighbours (tracks, clusters). The distance between two linked elements is quantifying the quality of the link. The link algorithm then produces a PF block of elements associated either by a direct link or by an indirect link through common elements. Examples are a link between a track and a cluster in the calorimeter, a link for bremsstrahlung photons which is a tangent to the track extrapolated towards ECAL, cluster to cluster links, a link to the secondary vertices or links between the tracks in the tracker and tracks in muon stations. In each PF block, the muon candidates are reconstructed first, and the corresponding PF elements are removed from the PF block. The electron reconstruction follows, with the aim of collecting also the energy of bremsstrahlung photons. Energetic and isolated photons, converted or unconverted, are identified in the same step. The remaining elements in the block are then subject to the cross-identification of charged hadrons, neutral hadrons and photons. Secondary particles from nuclear interactions are reconstructed as well.
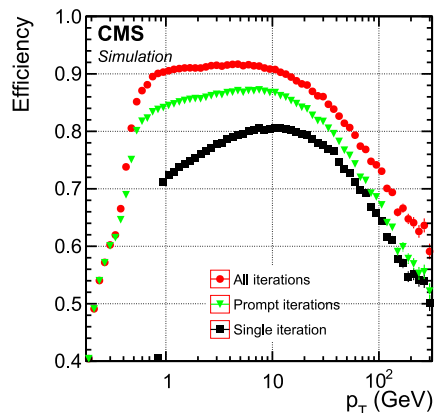
Figure 3. Efficiency of the global combinatorial track finder (black squares); and of the iterative tracking method (green triangles: prompt iteations based on seeds with at least one hit in the pixel detector; red circles: all iterations) [1].
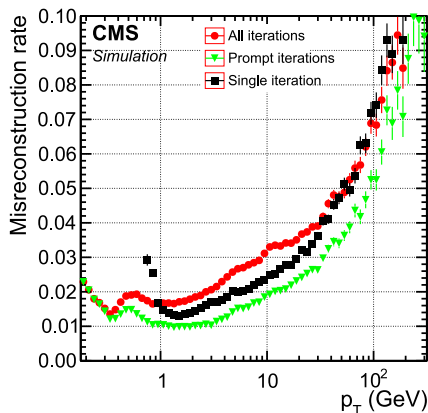
Figure 4. Misreconstruction rate of the global combinatorial track finder (black squares); and of the iterative tracking method (green triangles: prompt iteations based on seeds with at least one pixel hit; red circles: all iterations) [1].

4 Performance in simulation

Jets reconstructed by clustering the PF candidates have an energy response, energy resolution and angular resolution superior to the calorimeter-based jets, which results mostly from

a more precise and accurate measurement of the jet charged-hadron momentum in the PF algorithm. The performance improvement in the reconstruction of missing transverse energy, brought by the PF algorithm, is quantified by comparing $\vec{p}_{T,PF}^{miss}$ and $\vec{p}_{T,Calo}^{miss}$ to the reference $\vec{p}_{T,Ref}^{miss}$ calculated with all stable particles from the event generator, excluding neutrinos. The p_T^{miss} resolution was studied for events in which the p_T^{miss} has been calibrated to unity and therefore the $p_{T,Ref}^{miss}$ was required to be larger than 70 GeV. The relative p_T^{miss} resolution and the $\vec{p}_T^{miss}$ angular resolution in each bin of $\vec{p}_{T,Ref}^{miss}$ are shown in Fig. 5 and Fig. 6.
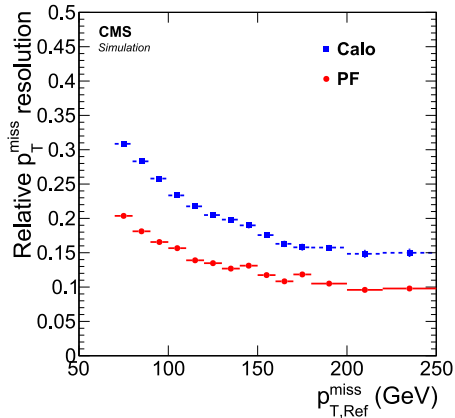
Figure 5. Relative p_T^{miss} resolution as a function of $p_{T,Ref}^{miss}$ [1].
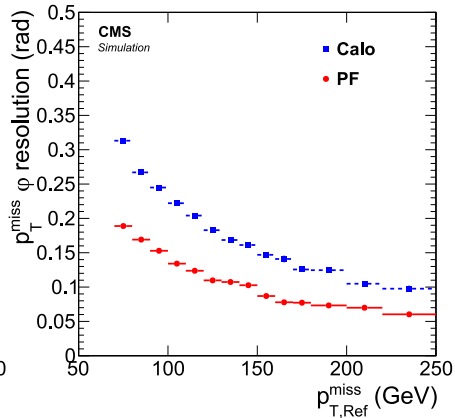
**Figure 6.** Resolution on the $\vec{p}_T^{miss}$ direction as a function of $p_{T,Ref}^{miss}$ [1].

The online reconstruction of a single event at the HLT has to be done one hundred times faster than offline, to cope with the incoming event rate. Therefore, the reconstruction was simplified at the HLT. Tracking is reduced to three iterations, dropping time-consuming reconstruction of tracks with low p_T or arising from nuclear interactions. The electron identification and reconstruction are not integrated in the online PF algorithm. The ability of HLT PF reconstruction to reproduce the offline results is tested with jets and hadronically decaying taus (τ_h) and shown in Fig. 7 and Fig. 8. The consistent use of PF jets at the HLT leads to a sharper jet triggering efficiency curve than with calorimeter jets. The sharp rise of the triggering efficiency curve demonstrates the excellent agreement between the τ_h p_T reconstructed online and offline.

## 5 Performance in data and pileup mitigation

The PF algorithm is able to resolve the particles arising from the τ decay and to reconstruct the surrounding particles to determine its isolation. The particles are used as an input to the hadrons-plus-strips (HPS) algorithm [4] to reconstruct and identify PF τ_h candidates. Before the advent of PF reconstruction, τ_h candidates were reconstructed as collimated and isolated calorimetric jets, called Calo τ_h. The HPS (PF) algorithm is able to assign the correct decay mode to the reconstructed and identified τ_h, where the generated decay mode is typically found for about 90% of the τ_h. The τ_h momentum is found to be reconstructed with a much better accuracy and precision than with the calorimeters. The ability and the accuracy of the HPS (PF) algorithm to identify and assign properly the τ decay mode and efficiently reconstruct the τ_h invariant mass, denoted as m_{τ_h}, are demonstrated in Fig. 9 and Fig. 10.
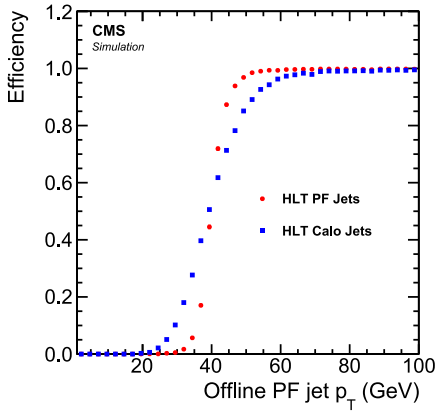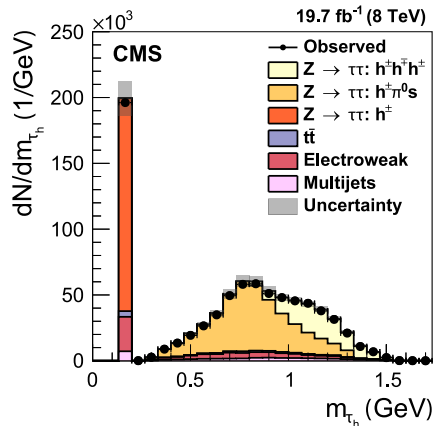
Figure 7. Probability to find at HLT a jet with $p_T > 40$ GeV matching the jet reconstructed offline, as a function of the offline jet p_T [1].
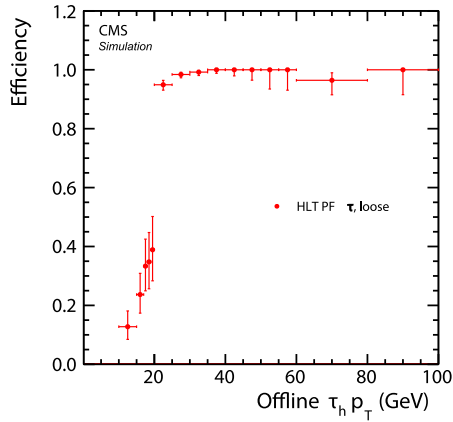
Figure 8. Probability to find a τ_h with $p_T > 20$ GeV at HLT matching the τ_h reconstructed and identified in the offline [1].
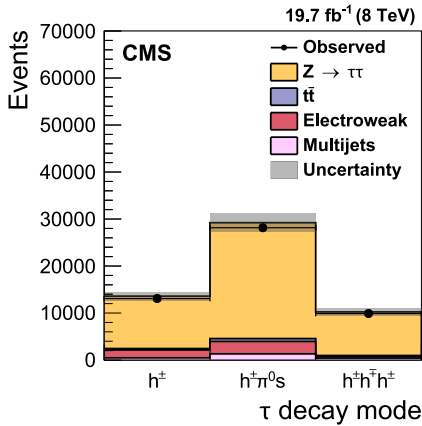

Figure 9. Distribution of the reconstructed τ decay mode in $Z/\gamma^* \to \tau\tau$ events selected in data compared to MC expectation [1].
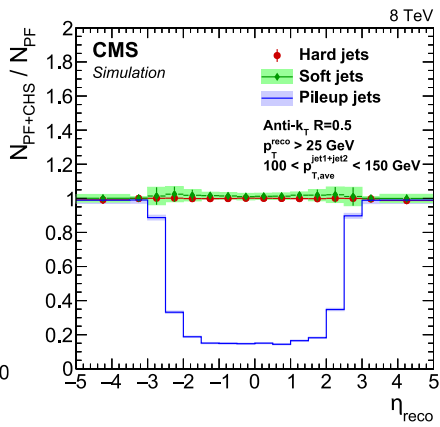
Figure 10. Distribution of the reconstructed τ_h mass in $Z/\gamma^* \to \tau\tau$ events selected in data compared to MC expectation [1].

The particles produced in pileup interactions give rise to additional charged hadrons, photons and neutral hadrons. As a consequence, reconstructed particles from pileup affect jets, E_T^{miss}, the isolation of leptons and the identification of hadronic τ decays. The charged hadrons that are unambiguously associated with the pileup vertices are removed from the list of PF particles used to form physics objects with an algorithm called charged hadron subtractions (CHS). The impact of pileup on the jet energy resolution is presented in Fig. 11, where the application of CHS improves the jet energy resolution for lower-p_T jets and the improvement is larger for higher numbers of pileup interactions. The ratio of the numbers of PF+CHS and PF jets with $p_T > 25$ GeV is shown in Fig. 12 as a function of pseudorapidity. The CHS is able to reduce the number of pileup jets by ∼85% within the tracker acceptance.
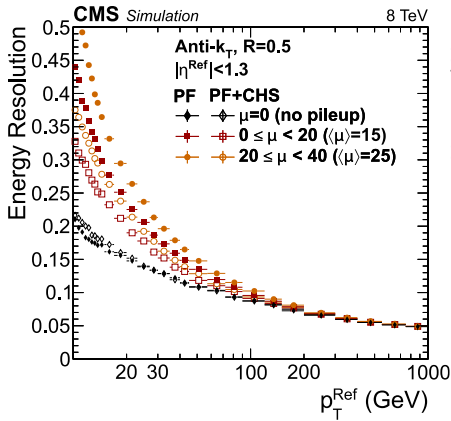
Figure 11. Jet p_T resolution PF+CHS jets (open markers) and PF jets (full markers) under three different pileup conditions, as a function of p_T^{Ref} [1].

Figure 12. Ratio of PF jet multiplicity with and without application of CHS, for hard jets, pileup jets and soft jets, as a function of the reconstructed jet η [1].

The jet energy contribution from different types of particles are measured with the tag-and-probe technique [5] in back-to-back dijet events. In the Fig. 13 and Fig. 14 the PF jet composition, corresponding to photons, leptons, neutral hadrons, charged hadrons and a fraction of the latter that is removed by the CHS algorithm, is presented as a function of jet p_T and η, showing a good agreement between the data and the MC prediction.
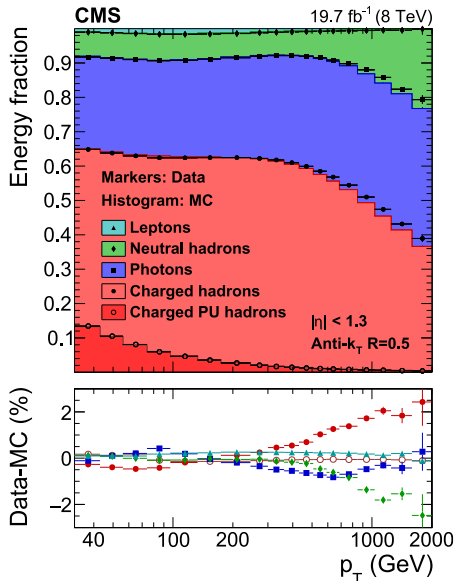
Figure 13. Jet energy composition in observed and simulated events as a function of p_T [1].
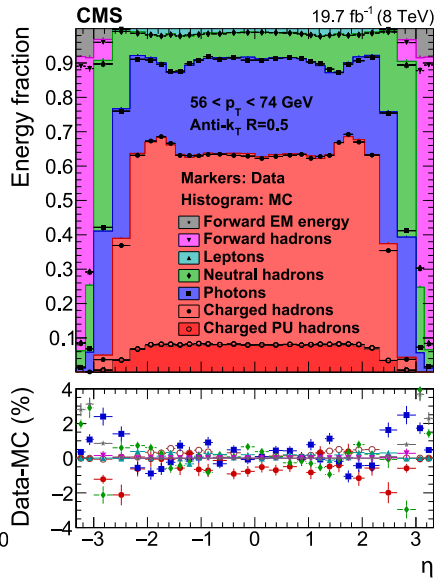
Figure 14. Jet energy composition in observed and simulated events as a function of η [1].

6 Summary and outlook

The CMS detector is found to feature properties well suited for PF reconstruction which motivated the first implementation of the PF algorithm in a hadron collider experiment. The technical challenges posed by the complexity of proton-proton collisions and the amount of material in the tracker were overcome with the development of new, high-performance reconstruction algorithms in different subdetectors. The PF reconstruction computing time was kept under control both for offline data processing and for triggering the data acquisition. As a result, the performance of all physics objects (efficiency, purity, energy and angular resolution etc.) was improved, to the benefit of many physics analyses. Excellent agreement was obtained between the data and simulation, thereby validating the use of PF reconstruction in real data-taking conditions, with an average pileup of 20 events and up to 35 concurrent pileup interactions.

The future CMS detector upgrade is planned to provide optimal conditions for PF performance. The HL-LHC will include a lighter and extended tracker (integrated into the Level-1 trigger) and a high-granularity endcap calorimeter, enhancing the PF capabilities for online and offline reconstruction [6]. These detector evolutions, accompanied by the necessary PF software development, should help to respond to the new challenges posed by the 200 pileup interactions per bunch crossing, foreseen at the LHC by the end of the next decade.

References

[1] The CMS Collaboration, JINST 12 P10003 (2017).
[2] The CMS Collaboration, JINST 3 S08004 (2008).
[3] The ALEPH Collaboration, The DELPHI Collaboration, The L3 Collaboration, The OPAL Collaboration, The LEP Electroweak Working Group, *Physics Reports*, **532**, 4, (2013).
[4] The CMS Collaboration, JINST 7 P01001 (2012).
[5] The CMS Collaboration, JHEP01 080 (2011).
[6] The CMS Collaboration, CERN-LHCC-2015-019 (2015).