Challenges of Particle Flow reconstruction in the CMS High-Granularity Calorimeter at the High-Luminosity LHC

Frank Chlebana, on behalf of the CMS Collaboration.
Fermi National Accelerator Laboratory, P.O. Box 500, MS 205, Batavia, IL 60510-5011
E-mail: chlebana@fnal.gov

Abstract. The challenges of the High-Luminosity LHC (HL-LHC) are driven by the large number of overlapping proton-proton collisions (pileup) in each bunch-crossing and the extreme radiation dose to detectors at high pseudorapidity. To overcome this challenge CMS is developing an endcap electromagnetic+hadronic sampling calorimeter employing silicon sensors in the electromagnetic and front hadronic sections, comprising over 6 million channels, and highly-segmented plastic scintillators in the rear part of the hadronic section. This High-Granularity Calorimeter (HGCAL) will be the first of its kind used in a colliding beam experiment. Clustering deposits of energy over many cells and layers is a complex and challenging computational task, particularly in the high-pileup environment of HL-LHC. Baseline detector performance results are presented for electromagnetic and hadronic objects, and studies demonstrating the advantages of fine longitudinal and transverse segmentation are explored.

1. Introduction
The Large Hadron Collider (LHC) is expected to be upgraded to the High Luminosity LHC (HL-LHC) starting in 2024, resulting in an average of 140-200 interactions per crossing (pileup), and is projected to deliver an integrated luminosity of 3000 fb$^{-1}$ over a ten year running period. Not all subdetectors of the Compact Muon Solinoid (CMS) will be able to survive in the high radiation environment and will have to be replaced. In order to achieve the full benefit of the HL-LHC, the physics objects should be reconstructed with high efficiency, low fake rate, and high resolution in the higher luminosity environment of the HL-LHC. CMS has selected a highly granular sampling calorimeter (HGCal) to replace the current endcap calorimeter. The HGCal consists of an electromagnetic section (EE) of 28 sampling layers of silicon, followed by the front hadronic section (FH) consisting of 12 sampling layers of silicon, and finally the back hadronic section (BH) consisting of 12 sampling layers of scintillator. The design uses silicon wafers of three different types. Wafers with an active thickness of 300 $\mu$m and 200 $\mu$m have 128 channels with a cell size of about 1 cm$^2$. Wafers with a 100 $\mu$m active thickness have 256 channels with a cell size of about 0.5 cm$^2$. Wafers with smaller cells are used at higher pseudorapidity providing finer segmentation and better radiation tolerance. More details can be found in Reference [1].

Particle flow (PFlow) reconstruction [2] was used in CMS with great success for the analyses in Run 1. The reconstruction attempts to measure and identify each particle originating from the collision and makes optimum use of the tracker, calorimeter, and muon detectors. Detector
features that help improve the performance of PFlow reconstruction include: efficient track reconstruction with high purity, a strong magnetic field to separate charged and neutral particles, and high calorimeter transverse granularity to separate overlapping showers. Significant improvements to the tracking system are planned for the phase 2 upgrades which will reduce the overall tracker mass, improve the track reconstruction efficiency, and extend the tracking coverage. Further details on the phase 2 upgrades can be found in the CMS technical proposal [3].

2. Simulation and Clustering
Performance studies for the technical proposal [3] were done using a GEANT4 [4] based simulation of the absorber structure as well as a detailed full detector simulation of the HGCal using square cells and the CMS PFlow clustering algorithm. Recently the simulation has been updated to more closely reflect the geometry described in the technical proposal. Changes include using hexagonal cells for EE and FH sections, updating the absorber structure, and improvements to the electronics simulation including the digitization and readout of time information. The active layers are simulated as a plane of hexagonal sensors and uninstrumented areas necessary for mechanical assembly are not currently simulated. A finer segmentation for the BH section was implemented with the ability to gang readout into different transverse sizes allowing studies on the effects of segmentation. The reconstruction of clusters and the energy sharing between adjacent clusters is a crucial input to the PFlow algorithm and the performance of several clustering algorithms have been investigated. The CMS PFlow algorithms that have been used for the much coarser granularity Run 1 detector, and were adapted for the technical proposal studies, use layer by layer clustering [2] that is seeded by the local maximum energy. The Pandora software development kit [5] provides a flexible framework to test different clustering algorithms. A forward projective cone algorithm from Pandora was studied as well as the ARBOR clustering approach which models showers as following the topology of a tree [6]. The clustering algorithms were originally developed for low occupancy environments and when used in the high pileup environment of the HL-LHC the execution times were found to be slow. Clustering algorithms employ nearest neighbor searches and require efficient identification of the outermost points of the clusters. Modern computing techniques were used to improve both the CPU and memory usage. Utilization of a k-d tree provided several orders of magnitude in speedup. Techniques adapted from graph theory such as the use of quick unions allow efficient association of hits that form clusters. Data structures were optimized for more efficient memory management and the code was rewritten to be thread safe. These software improvements reduced execution times from several hours to several minutes per event. Performance of the clustering algorithms was good, however, when examining the hit patterns using an event display it could be seen that the topological based clusters did not efficiently resolve closely adjacent energy deposits. Experience gained from these studies is being used to modify the clustering algorithms to better separate multiple narrow shower cores.

3. Reconstruction and Performance
The electron resolution is shown in Figure 1 for silicon sensors having different active layer thicknesses and for different regions of pseudorapidity as obtained from the GEANT4 simulation with no pileup. The $e/\gamma$ resolution is $20-24\% / \sqrt{E}$. In the presence of pileup, electron clusters will contain subclusters originating from pileup which can be identified and removed. Low $p_T$ neutral photons can be resolved using the transverse granularity, late interacting charged hadrons can be rejected using the longitudinal segmentation, measurement of the shower profile helps reduce pileup contamination, and the lower density design of the tracker results in more efficient tracking. The reconstructed energy of the electron supercluster is shown as a function of the number of vertices in Figure 2 before and after pileup subtraction was applied. The dependence of the energy measurement on the pileup is clearly removed.
Figure 1. The electron resolution for sensors with different active thickness and for different pseudorapidity regions.

Figure 2. Energy of the electron super cluster before and after applying pileup subtraction.

The electron reconstruction efficiency for the phase 1 detector is degraded as the electromagnetic crystals are damaged from radiation and with increasing pileup as is shown in Figure 3. The performance of the HGCAL for 140 pileup events is comparable with the aged phase 1 endcap performance for 50 pileup and the increased coverage of the phase 2 tracker extends the acceptance of electrons. Pure electromagnetic showers have a higher energy density than hadronic showers. The high transverse and depth granularity allows us to measure the cluster density and a correction can be applied to the energy of the hadronic cluster to help compensate for electromagnetic fluctuations and improve the jet energy resolution. The results shown in Figure 4 were obtained by applying the method developed by the CALICE collaboration [7] and improves the jet resolution by 10%. Development is ongoing to further exploit the benefits that the high granularity provide and to improve the reconstruction performance.

Figure 3. The electron reconstruction efficiency for the current endcap calorimeter and the HGCAL with different pileup events.

Figure 4. The jet energy resolution, before and after software compensation is applied.

4. Timing
A collision can occur anywhere in the overlapping region as the colliding bunches pass through each other and can occur at the same \( z \) location but with different times. The actual beam spot distribution in time and space depends on the LHC operating conditions. Precision timing
can be used to mitigate pileup by helping to associate energy deposits with the originating vertex. This is of particular importance for clusters originating from neutral particles such as photons. The HGCal provides precise timing from the silicon. A study was done to estimate the time resolution of electromagnetic clusters for different assumptions of the time resolution for the individual cells. The results are shown in Figure 5 where a cell time resolution of 50 ps for deposits with 60 fC results in a cluster resolution of 20 ps for 10 GeV photons. A time resolution of 20 ps in silicon has been achieved in test beams [8].

![Figure 5. The photon cluster time resolution obtained for different assumptions of the cell time resolution and as a function of the photon energy.](image)

5. Summary
The preliminary studies for the CMS phase 2 technical proposal indicate that the HGCal will be able to maintain the necessary physics performance in the high pileup environment of the HL-LHC. The energy clustering in the calorimeter is an essential input to particle flow reconstruction and we investigated several existing algorithms that were developed for imaging calorimeters of the ILC. Modern computational techniques were used to dramatically improve clustering algorithm execution times. The clustering performance was good, however we did see that nearby clusters were not fully resolved in very dense environments and are exploring methods of energy weighting the hits to help with the separation of clusters. The simulation of the geometry has been updated to better reflect the description in the technical proposal, and we have changed from using square cells to hexagonal cells which required a major update of the geometry description and reworking the clustering algorithms. The HGCal design is still evolving and we will need to incorporate the details of uninstrumented areas necessary for the mechanical support. We have only started to exploit the full potential that the high granularity provides for pileup mitigation and improved event interpretation.

References
[1] C. Ochando, “HGCAL: A High-Granularity Calorimeter for the Endcaps of CMS at HL-LHC”, these proceedings.
[2] F. Beaudette, “The CMS Particle Flow Algorithm”, arXiv:1401.8155.
[3] D. Contardo et al., “Technical Proposal for the Phase-II Upgrade of the CMS Detector”, CERN-LHCC-2015-010.
[4] S. Agostinelli et al., “GEANT4 - a simulation toolkit”, Nuclear Instruments and Methods in Physics Research A 506 (2003) 250-303.
[5] M. A. Thomson, “Particle Flow Calorimetry and the PandoraPFA Algorithm”, Nucl. Instrum. Meth. A611: 25-40, 2009.
[6] M. Ruan, “Arbor, a new approach of the Particle Flow Algorithm”, arXiv:1403.4784.
[7] C. Adloff et al., “Hadronic energy resolution of a highly granular scintillator-steel hadron calorimeter using software compensation techniques”, JINST 7 P09017 (2012), arXiv:1207.4210.
[8] N. Akchurin, “Fast-timing capabilities of silicon sensors for the CMS High-Granularity Calorimeter at the High-Luminosity LHC”, these proceedings.