Electrons and photons at High Level Trigger in CMS for Run II

Afiq A. Anuar on behalf of the CMS Collaboration
National Center for Particle Physics, University of Malaya, 50603 Kuala Lumpur, MY
E-mail: afiq.anuar@cern.ch

Abstract. The CMS experiment has been designed with a 2-level trigger system. The first level is implemented using custom-designed electronics. The second level is the so-called High Level Trigger (HLT), a streamlined version of the CMS offline reconstruction software running on a computer farm. For Run II of the Large Hadron Collider, the increase in center-of-mass energy and luminosity will raise the event rate to a level challenging for the HLT algorithms. New approaches have been studied to keep the HLT output rate manageable while maintaining thresholds low enough to cover physics analyses. The strategy mainly relies on porting online the ingredients that have been successfully applied in the offline reconstruction, thus allowing to move HLT selection closer to offline cuts. Improvements in HLT electron and photon definitions will be presented, focusing in particular on: updated clustering algorithm and the energy calibration procedure, new Particle-Flow-based isolation approach and pileup mitigation techniques, and the electron-dedicated track fitting algorithm based on Gaussian Sum Filter.

1. Introduction
In the CMS experiment [1], reconstruction of electrons and photons (eγ) at the HLT [2] is done in a sequence of steps. Events passing a pre-selection, implemented using custom-designed electronics in the so-called Level 1 (L1) seeding step, are processed through a streamlined version of the CMS offline reconstruction algorithm. The algorithm starts by clustering the hot crystals inside the electromagnetic calorimeter (ECAL) to collect the full energy deposit of the candidates. eγ candidates are then selected with some identification criteria involving ECAL cluster shape and isolation observables, which are defined as the sum of transverse energy (E_T) deposit around the candidate. For electrons only, the presence of a well reconstructed track matched to the supercluster is also required. Finally, a selection on the Tracker isolation is applied on the eγ candidates. The reconstruction chain is summarized in Figure 1.

This sequence of cuts, designed to optimize the timing of the HLT selection, has been used in Run I and remains unchanged in Run II. However, due to the increased energy and luminosity, overall improvements in the algorithm are necessary to ensure that the HLT output rate is kept manageable while maintaining thresholds low enough to cover physics analyses. This is achieved by porting online, with some simplifying approximations to stay within HLT resource limitations, the ingredients that have been successfully applied in the Run I offline reconstruction [3], thus allowing to move HLT selection closer to offline cuts. Improvements in HLT eγ definitions will be presented in this paper, focusing in particular on: updated clustering algorithm and the energy calibration procedure, new Particle-Flow-based [4] isolation approach and pileup mitigation techniques, and the electron-dedicated track fitting algorithm based on Gaussian Sum Filter.
techniques, and the electron-dedicated track fitting algorithm based on Gaussian Sum Filter (GSF) \cite{5}.

2. L1 seeding
In this step, custom-designed electronics perform a fast pre-selection on the events using input from the calorimeter system, reducing the event rate from the nominal 40 MHz to about 100 kHz for further processing at the HLT step \cite{6}. In Run II, on top of filtering on the transverse energy deposit in the ECAL as done in Run I, pseudorapidity and coarse isolation of the candidates are filtered on in the $e\gamma$ trigger paths, in order to keep the threshold at a reasonable level despite the overwhelming rate and without the use of a prescale.

3. ECAL clustering and identification

3.1. Run II clustering algorithm
Although electron energy is spread very narrowly when its bremsstrahlung energy loss is small, depositing as much as 97\% of its energy within a single 5×5 crystal array \cite{7}, within CMS this mode of energy loss is a significant one due to the dense intervening Tracker system, and up to 90\% of the electron energy could be lost before it reaches the ECAL. It is therefore essential to collect also the energy of the radiated photons in order to have an accurate measurement of the electron’s energy. In Run I, this was done by using what was called the stand-alone approach, which is divided into two clustering algorithms, depending on the pseudorapidity of the candidate. The ‘hybrid’ algorithm takes advantage of the fact that the crystals in barrel ECAL (EB, $|\eta| < 1.4791$) are arranged in $\eta \times \phi$ geometry and that the shower is more spread out in $\phi$ due to the influence of the magnetic field. In the endcap ECAL (EE, $|\eta| > 1.4791$), due to a different arrangement of the crystals, the ‘multi 5×5’ algorithm was used, which collects the energy deposit within clusters of 5×5 crystals around a hot crystal seed which are then grouped together into a supercluster if their total energy exceeds a certain threshold.

In Run II, an alternative approach is used that is part of the Particle Flow (PF) reconstruction algorithm \cite{4}. In this approach, called ‘mustache’ clustering, clusters are reconstructed by grouping together all crystals contiguous to a seed crystal if their energy deposit is two standard deviations above the electronic noise. The requirement of a crystal to be taken as a seed is that its energy must be above these thresholds; $E_{\text{seed}} > 230$ MeV in the EB and $E_{\text{seed}} > 600$ MeV or $E_{T, \text{seed}} > 150$ MeV in the EE. This is the approach of choice in Run II as it provides significant improvements to energy resolution, as shown in Figure 2.

3.2. Multifit pulse amplitude reconstruction algorithm
In Run II, the LHC will be operating at a bunch spacing of 25 ns, half that of Run I. This shorter spacing between bunches means a significant increase of out of time pile-up (OOT PU), contributions from additional collisions not within the current bunch crossing. In ECAL this problem is mitigated by the use of the Multifit algorithm; essentially a fit of a template, obtained
from a simulation of ECAL pulses in a PU-less environment, to multiple superimposed ECAL pulses fulfilling a $\chi^2$ criterion. This leads to a significant suppression of OOT PU while also improving the quality of the in time pulses, particularly for the low energy and high PU regime.

3.3. Crystal transparency correction

Due to the harsh radiation environment the ECAL crystals are subjected to, they lose their transparency as the data accumulates; affecting the energy measurement. It is therefore necessary to apply corrections to counter this effect. In Run I, this was applied only in the EE where the effect was bigger, with a single correction factor to account for the average of a group of crystals in rings of $\phi$. Nevertheless, the correction had a large impact as could be seen in Figure 3.

In Run II, the harsher conditions demand the correction to be applied in both regions. Also, as

Figure 2: Comparison of energy resolution of the two clustering algorithms in (a) $\eta$ and (b) $E_T$ after the energy correction procedure.

Figure 3: Impact of crystal transparency correction in Run I, applied only in the endcap region.
the crystals are affected differently depending on their position within the ECAL, the correction factor will be derived for each individual crystals and applied weekly.

3.4. \( e\gamma \) identification in the calorimeter system
After the supercluster creation as described above, identification observables can be defined to discriminate the \( e\gamma \) candidates from the other particles depositing energy in the ECAL, such as hadronic particles inside jets; exploiting the fact that \( e\gamma \) energy deposit are typically very narrow. The observables that are used in the HLT are energy-log weighted cluster shape, \( \sigma_{\text{inj}} \), used in \( e\gamma \) paths and the ratio between 3x3 crystal array and supercluster energy deposit, R9, used to discriminate between converted (to electron pairs) and unconverted photons. Figure 4 shows the performance of these observables.

![Figure 4: (a) Signal and background efficiencies of the \( \sigma_{\text{inj}} \) observable in the EE region. The blue star denotes the efficiency of the filter used in the single electron trigger. (b) Signal efficiency of the R9 observable versus \( p_T \) for the diphoton trigger. Also shown is the efficiency of isolation and calorimeter identification used to recover the signal portion failing the R9 selection.](image)

Following the ECAL identification step, the \( e\gamma \) candidates are subjected to another selection filter based on the input from the hadronic calorimeter (HCAL). This is because electrons and photons, being electromagnetic particles, are primarily stopped in the ECAL, thus it is expected that their energy deposit in the HCAL to be small.

4. Calorimeter and track isolation
Isolation is a handle to discriminate between the prompt and background \( e\gamma \) candidates, capitalizing on the fact that prompt candidates are typically not aligned along other final state objects in the event. At the HLT, relative isolation is the quantity most commonly filtered on; it is the ratio of the isolation sum and the \( E_T \) of the candidate.

In Run I, the isolation algorithm with the best performance was the PF isolation algorithm, which made use of the information provided by the entire detector combined in such a way as to reconstruct the individual particle showers. As the strategy for managing the rate in Run II is to bring the HLT reconstruction algorithms as close to offline as possible, the PF
isolation algorithm is an obvious choice. However, this could not be done in a direct way as the PF isolation algorithm requires a complete set of inputs to be computed, one of which is the information from the Tracker system, which in turn requires track reconstruction that is too time-intensive to be feasible at the HLT. In order to get around this obstacle, the PF isolation is broken down into three parts, one for each subdetector; ECAL, HCAL and Tracker. The ECAL and HCAL isolation take as their input the PF clusters of the respective calorimeter system while the Tracker isolation is moved to the end of the filtering steps following the track reconstruction, when the event rate has been reduced to a point that regional track reconstruction is feasible within HLT timing constraints. The performance of the isolation algorithm to be used in Run II is shown in Figure 5 in comparison with the isolation algorithm used in Run I.

As isolation is a measure of activity around the candidate of interest, PU naturally affects its performance significantly and therefore has to be accounted for. At HLT, this is done by applying a correction factor which is derived from the fact that both isolation and average energy density for every event vary linearly with PU. By applying this correction factor the HLT manages to maintain a fairly robust performance of the isolation selection criteria against PU, even in the harsh conditions of Run II, as shown in Figure 5.

Figure 5: (a) Performance of the Run I and Run II isolation algorithms in the barrel and endcap region (b) Efficiency of the ECAL and HCAL isolation in barrel and endcap, with and without PU correction factor applied.

5. Track reconstruction and identification

5.1. Pixel matching

While the track reconstruction at HLT is done only within a region compatible with the supercluster, this region could still contain many tracks, only one of which could be that of the electron. In order to ensure only the most relevant tracks are reconstructed, hits in the pixel detector forming the seed for track reconstruction are matched with the supercluster; in $\phi$ windows for the first layer of pixel detector and in windows of $r$, $z$ and $\phi$ for the second layer, as schematically shown in Figure 6. In the event where multiple seeds exist and fulfill the matching criteria, the one with the smallest distance (as defined using the observables used in the matching) is chosen as the best candidate.
5.2. GSF track reconstruction algorithm

While it is possible to reconstruct electron tracks using the standard Kalman Filtering (KF) technique as with other charged particles in the CMS detector [8], the large energy loss through bremsstrahlung renders this procedure inadequate, due to the large bending in $\phi$ the electrons experience. For this reason, electron tracks are reconstructed with the GSF algorithm, a non-linear generalization of the KF that utilizes multiple weighted Gaussian distributions to model the energy loss. The algorithm provides significant improvements to both momentum and angular resolution compared to the KF algorithm, as can be seen in Figure 7.

5.3. Track identification observables

As the track reconstruction results in another fresh set of parameters associated with the candidate, more identification observables can be defined to aid in discriminating the true,
prompt electrons from the collection of candidates. In the electron triggers three observables are used; the difference between inverse of supercluster energy and track momentum, $1/E - 1/P$, difference between supercluster and track $\eta$, $\Delta \eta$ and $\Delta \phi$. These observables exploit the fact that electron candidates are measured using two independent subdetectors, ECAL and Tracker; they are therefore measures of compatibility between the two, in terms of energy response and angular positioning respectively. The distributions and performances of these observables are shown in Figure 8.

6. Summary
The electron and photon reconstruction algorithm at the HLT in the CMS experiment has been discussed, focusing on the strategies adopted in preparation for the harsher data-taking conditions in Run II. Individual trigger paths are created to suit the needs of physics analyses and their performance has been estimated. For the leading (unprescaled path with the lowest threshold) single electron trigger path designed for top physics ($p_T^e < 32$ GeV), the expected signal efficiency is 90% at a rate of 157 Hz, while for the double photon path designed for Higgs physics ($p_T^\gamma < 34, 18$ GeV), the expected signal efficiency is 100% at a rate of 41 Hz. The performance has been evaluated assuming an instantaneous luminosity of $1.4 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and average 40 PU, the harshest condition in the 2015 Run II data-taking period, in keeping with the primary purpose of the trigger development.
Acknowledgments
The author wishes to thank Stephanie Beauceron, Javier Fernandez, Matteo Sani and the rest of the Trigger Studies Group for their helpful input in preparing this presentation.

References
[1] Chatrchyan S et al. 2008 Journal of Instrumentation 3 S08004 URL http://stacks.iop.org/1748-0221/3/i=08/a=S08004
[2] Chatrchyan S et al. (CMS Collaboration) 2012 CMS Physics Analysis Summary TRG-12-001
[3] Chatrchyan S et al. (CMS Collaboration) 2015 (Preprint 1502.02701)
[4] Chatrchyan S et al. (CMS Collaboration) 2009 CMS Physics Analysis Summary PFT-09-001
[5] Adam W, Fruhwirth R, Strandlie A and Todorov T 2005 Journal of Physics G: Nuclear and Particle Physics 31 N9 URL http://stacks.iop.org/0954-3899/31/i=9/a=N01
[6] Chatrchyan S et al. (CMS Collaboration) 2000 The TriDAS project; CMS Technical Design Report, Volume 1: The Trigger Systems (CERN)
[7] Adzic P 2007 Journal of Instrumentation 2 P04004
[8] Chatrchyan S et al. (CMS Collaboration) 2014 Journal of Instrumentation 9 P10009 (Preprint 1405.6569)