Performance of the CMS High Level Trigger

Andrea Perrotta
Istituto Nazionale di Fisica Nucleare, sez. Bologna, I-40127 Bologna
E-mail: andrea.perrotta@bo.infn.it

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 increases in center-of-mass energy and luminosity will raise the event rate to a level challenging for the HLT algorithms. The increase in the number of interactions per bunch crossing, on average 25 in 2012, and expected to be around 40 in Run II, will be an additional complication. We present here the expected performance of the main triggers that will be used during the 2015 data taking campaign, paying particular attention to the new approaches that have been developed to cope with the challenges of the new run. This includes improvements in HLT electron and photon reconstruction as well as better performing muon triggers. We will also present the performance of the improved tracking and vertexing algorithms, discussing their impact on the b-tagging performance as well as on the jet and missing energy reconstruction.

1. The High Level Trigger of CMS

The role of the trigger in a High Energy Physics experiment is to reduce the rate of recorded collision to a level which is manageable by the following Data Acquisition (DAQ) and Reconstruction steps. At LHC the proton beams are organized in bunches. Those bunches were interleaved by 50 ns during the Run1 period (i.e. the first data taking period at LHC, that ended with the Heavy Ion run of February 2013), while they will eventually be spaced by 25 ns at Run2 (i.e. the data taking period which is going to start in June 2015). This implies that when LHC will move to 25 ns running at Run2 the collision rate will be of the order of 40 MHz (in reality a bit less, because not all bunches in the machine are filled with protons). The maximum acceptable rate for data acquisition and storage is of the order of 1 kHz, and the trigger must be designed to reduce the rate to that level, by accepting the largest possible cross-section of the interesting physics events from the collisions and rejecting efficiently the non interesting ones.

The design chosen for the trigger of the CMS experiment [1] is a two-level system. The Level 1 Trigger (L1) is based on FPGA and custom ASIC technology and uses information from the calorimeters and muon spectrometers of the experiment in order to accept or reject an event; it reduces the event rate down to approximately 100 kHz, acceptable by the readout electronics. The High Level Trigger (HLT) is implemented in software running on a farm of commercial computers which includes about 16,000 CPU cores, and reduces the L1 output rate to the sustainable level for storage and physics analysis of about 1 kHz. The HLT software consists of a streamlined version of the offline reconstruction algorithms; it exploits the same
sophisticated software used for offline reconstruction and analysis, optimized in order to comply with the strict time requirements of the online selection.

The HLT menu in CMS has a modular structure, which is graphically depicted in figure 1. The menu is subdivided in paths (there are up to now more than 400 different HLT paths prepared for the data taking at Run2). Each path is a sequence of reconstruction and filtering modules, and it reproduces the offline selection for a given physics object (photons, electrons, muons, jets, missing momenta, b-tagged jets, etc.), for combinations of them, or even for more sophisticated pre-selections used in the complicate physics analyses. The modules within a path, either object producers or filters, are arranged in blocks of increased complexity, so that faster algorithms are run first and their products are filtered: if a filter fails, the rest of the path is skipped. There are other important features that differentiate the algorithms used at HLT to the ones used for the offline reconstructions, all meant to reduce the CPU time consumption at HLT: amongst them, one can recall here the regionality (detector read-out and reconstruction are restricted to narrow regions around the L1 or higher-level candidates), and the simplified tracking.

![Figure 1. Schematic representation of a HLT menu in CMS and of the HLT paths in it. The final trigger decision is the logical OR of the decisions of the single paths.](image)

2. The challenge of the LHC Run2
LHC at Run2 will provide collision with a higher centre of mass energy, a bigger instantaneous luminosity and it will change its bunch structure with respect to LHC run1. All that is going to be very demanding for the trigger, and the HLT of CMS (as well as L1) was redesigned and updated during the recent LHC shutdown in order to comply with it. The centre of mass energy of the p-p collisions will increase from the previous maximum of 8 TeV to 13 TeV at Run2. This will originate approximately a factor two increase in cross section for all typical processes, and such an increase will be even larger for multiple objects triggers because of the combinatorial. The peak luminosity reached $7 \cdot 10^{33}$ cm$^{-2}$s$^{-1}$ at Run1, and it will reach up to $1.4 \cdot 10^{34}$ cm$^{-2}$s$^{-1}$ at Run2, and this factor two will also directly reflect in the rates. Finally, having proton bunches spaced by 25 ns instead of 50 ns, while allowing large luminosities with correspondingly lower in time pile-up, still will make more important the effect of the out of time pile-up. Let remind here that with pile-up (PU) events we intend those collisions amongst beam protons that superimpose with the p-p collision which originate the “truly interesting” event. Those PU events can be “in time” if they come from the very same bunch crossing (BX) of the main collision, or out “of time” if they show up in the nearby crossings. Signals from
out of time PU can still be present in slow detectors, as the calorimeters, and pollute the ones
produced by the interesting event.

In summary, during normal running conditions at the LHC Run2 there will be roughly a
factor four increase in rate with respect to Run1, an average number of in time PU will raise
from 25 to 40, and the out of time PU will have a much larger impact. The challenge for
the trigger system will be that it must give out a similar rate of selected events to the data
acquisition system as during Run1, but it must be extracted from a much harsher environment.
The planned budget for Run2 is in fact similar to what was at Run1: 1.35 kHz peak HLT rate
for 1 kHz average offline rate.

In general, there are three handles that one can switch in order to have a sustainable output
rate also at Run2:

• to raise the thresholds for accepting physics objects;
• to buy additional CPU’s and/or storage elements, or to improve otherwise the performance
  of the hardware;
• to improve the performance of the software algorithms used by HLT.

Of course, one should better avoid raising unnecessarily the acceptance thresholds in the filters,
as it would negatively impact the physics capabilities of the experiment.

The technical improvements and updates applied to the HLT in CMS are described in another
paper in these proceedings [2]. This paper will concentrate instead on the improvements prepared
for the software algorithms and for the design of the HLT paths to be used at Run2 with respect
to the ones of Run1.

3. The Trigger of CMS at Run2

For the year 2015 three main trigger scenarios are foreseen, and for all them a dedicated trigger
menu is being finalized:

• an initial period in which LHC will run with 50 ns BX, attaining $5 \cdot 10^{33}$ cm$^{-2}$s$^{-1}$ peak
  luminosity (which corresponds to an average PU of 30);
• an intermediate period during which LHC will move to the 25 ns BX structure, and will
  reach $7 \cdot 10^{33}$ cm$^{-2}$s$^{-1}$ peak luminosity (which corresponds to an average PU of 20);
• towards the end of the year, with the same 25 ns BX structure a $1.4 \cdot 10^{34}$ cm$^{-2}$s$^{-1}$ peak
  luminosity will be reached (which corresponds to an average PU of 40).

In addition, other menus are also going to be prepared for low PU runs (PU = 0.01, 0.4), Heavy
Ion runs and for commissioning.

In the menus dedicated to all those scenarios object producer and filter algorithms have
been improved with respect to the ones used at Run1, as well as the design of the HLT paths themselves.

3.1. L1 updates for the run in 2015

It is not the scope of this presentation to describe the improvements in the (hardware) L1 trigger
of CMS for the Run2. However, L1 filters are an intrinsic preselection step of the final HLT
selection, and L1 objects may also be used as seeds for the subsequent HLT algorithms and
paths. Let therefore at least list here the most relevant updates to L1 that CMS will profit of
in its menus of 2015:

• implemented pile-up subtraction for jets, energy sums, e/gamma isolation;
• dedicated tau trigger candidates (looking at the energy content of particularly shaped 4x4
  or 4x8 calo towers);
• additional muon chambers in the endcaps;
• increased granularity of the muon readout in the endcaps.

Full L1 upgrade will be commissioned during 2015, and made then available for 2016 runs.

3.2. HLT at Run2
There are several improvements in the HLT algorithms and menu prepared for Run2. Overall, the most effective one is the move of basically all HLT paths in CMS to Particle Flow based methods and event reconstruction. Particle Flow (PF) [3] is a reconstruction technique widely used in CMS analyses. It makes use of the full detector information to describe the global collision event, by identifying particles individually and clustering them into more complex objects. Its use in HLT was already started during Run1, for tau leptons, jets and missing energy reconstructions. Moving basically all paths to PF at Run2 improves the energy resolution of trigger objects and makes the online event reconstruction and selection much closer to the corresponding ones performed in the offline reconstruction and analyses. It provides in addition more efficient methods for PU mitigation. PU in fact affects the performance of many triggers: the additional tracks and calorimetric deposits tend to increase the rate of jet related triggers, as well as reduce the efficiency for the isolation algorithms. In order to mitigate these effects, minimum pT thresholds and vertex constraints are applied to tracks and other objects; moreover, the average energy density is measured in each event and subtracted from jet areas and lepton isolation cones [4].

Tracking certainly represents the most important ingredient for applying PF techniques to the whole event reconstruction. Tracking at HLT [5] uses the same kind of iterative procedure as in the offline reconstruction. Such a procedure iteratively removes hits once associated with tracks, reducing in that way the combinatorial in the subsequent steps. It starts with prompt, high pT tracks, then it looks for displaced ones, and only after for less frequent track topologies. At HLT only the first iterations are considered, while displaced tracks (seeded by strip triplets) are only used in a subset of trigger paths where displaced tracks are needed. Figure 2 shows the efficiency for reconstructing tracks of the four iterations considered in the HLT tracking, as function of their distance from the position of the primary vertex. The challenge at 25 ns BX comes from the increased strip and pixel occupancy due to hits from out of time PU. However, clusters from out of time PU have low collected charge, because they are measured in correspondence of a different BX: by cutting on the charge of clusters one suppresses fakes, as shown in figure 3. There are several other updates and optimizations included in the tracking algorithm at HLT, all meant to reduce the CPU time consumption, and improve efficiency and purity of the reconstructed objects: amongst them one can recall here the regional tracking (track are only searched for in a region of interest defined by the track seed), constraining the passage through the primary vertex for pointing tracks, and the use of a simplified parametrized magnetic field instead of its full detailed map. All those optimizations together lead to a CPU time reduction of about a factor four with respect to the previous tracking algorithm, for an average PU of 40 which is the one expected in the last period of run of 2015.

b-Tagging is used at HLT to identify jets originated by the fragmentation of heavy flavour quarks in the event [5]. It heavily relies on tracking, and using tracks reconstructed with the improved iterative tracking at Run2 will already improve its performance. Other updates to the algorithm for the new run include the use of deterministic annealing for the reconstruction of the primary vertex, improved secondary vertex algorithm and usage of a secondary vertex finder. Figure 4 compares the performance on a simulated tt sample of the old algorithm with the one of the new algorithm prepared for Run2, for the three previously defined scenarios. Performance in the three scenarios are quite comparable amongst themselves, while it is evident that for a given b-jet efficiency the rejection power against light quark jets improves significantly when switching to the Run2 algorithm.
Figure 2. Track reconstruction efficiency as function of the distance in the transverse plane from the primary vertex for the four tracking iterations used at HLT in Run2.

Figure 3. Effect of cluster charge cut, represented by the vertical line in the figure. It is meant to remove spurious clusters from hits off track.

Figure 4. Comparison of the efficiencies of the old (CSV) and new (CSVv2+IVF) b-tagging algorithms in CMS HLT in the three scenarios expected for the LHC running in 2015: the efficiency for selecting b-jets is displayed as function of the efficiency for selecting jets originated by light quarks.

Electron identification and reconstruction [6] will also profit of the introduction of the improved iterative tracking at Run2 and of the wide usage of PF techniques at HLT. Figure 5 shows the improved energy resolution of the electron superclusters when reconstructed with PF with respect to the resolution of the calorimetric superclusters used at Run1, as function of the transverse energy of the electrons. Even for the isolation a “PF cluster based” algorithm will be used to compute the calorimetric content within the isolation cone: it does not fully corresponds to the finally identified electrons, photons or hadrons as resulted from the complete application of the PF to the whole event, but rather to the calorimetric clusters that will be used later for the full PF reconstruction, already reconstructed and PU subtracted with the PF methods. Fig 6 shows the electron isolation efficiency measured as function of the number of reconstructed vertexes in the event (and therefore as function of the number of superimposed PU events) in the electromagnetic and hadronic calorimeters of CMS, with and without applying the pile-up subtraction.
Figure 5. Resolution on the measured energy of electron superclusters reconstructed with the Run1 technique and with the one prepared for Run2.

Figure 6. Isolation efficiency when cutting on the energy content in the electromagnetic and hadronic calorimeters with the Run2 “PF cluster” method, and with and without PU subtraction, as function of the number of reconstructed vertexes in the event.

Also jets will certainly take profit of the widespread use of PF techniques at the HLT. Let however report here only about a new development meant to select jets with sub-structures within them, as expected from boosted hadronic topologies characteristic of several new physics scenarios. The new algorithm will first reconstruct “fat” jets in the event with a coarse grained jet reconstruction algorithm (AK8), and then the invariant mass of trimmed jets within it is computed. In figure 7 one can appreciate the improved shape of the turn-on curve for the scalar sum of the leading and subleading jet $p_T$’s as obtained with the new HLT path when compared with the one obtained by other HLT paths which select simple jets, or even by a path which looks at the total sum of the transverse momenta in the event.

Figure 7. Turn-on curves for the selection of simulated events with boosted jets at 13 TeV: the performance of the newly invented HLT path meant to select “trimmed jets” is compared with the one of another HLT path with a simpler jet algorithm, and with the one of a path selecting on the overall scalar sum of the jet transverse momenta in the event.

Muon reconstruction at HLT [7] starts with track fitting in the outer muon spectrometer, which builds the so called level-2 (L2) muons. The level-3 (L3) muon reconstruction is the muon HLT step that builds full muon tracks (including tracker information) starting from those L2 muons. A sequence of three algorithms is tried in cascade from the fastest to the most...
CPU consuming one and stops as soon as a L3 track is found. These algorithms are named “OIState” (seeds to reconstruct tracker tracks are built on the basis of L2 muons propagated to tracker layers), “OIHit” (seeds to reconstruct tracker tracks are built on the basis of L2 muons propagated to tracker with the addition of information from one tracker hit) and “IOHit” (L2 muons are used to build regions where to look for hits but seeds to reconstruct tracker tracks are built using pixel/tracker hit information). Quality cuts can be applied during the L3 reconstruction (which is one of the improvements prepared for the Run2 version of the algorithm) to ensure good quality/purity of muons firing the HLT, and can be tuned according to the muon signature that has to be detected. Additionally, muon isolation can be required with the same Particle Flow based technique already explained for the electron isolation, and by subtracting the average energy from PU. Figure 8 compares the L3 muon efficiency as function of the number of reconstructed vertexes in the event between the Run1 version of the algorithm and the improved one for Run2, where the new algorithm is demonstrated to be more insensitive to the PU conditions. Another large improvement prepared in the muon sector is in the triggering on highly displaced muons: the adoption of a mean-timer technique in the barrel muon chambers of CMS and the use of cosmic-like seeds for the L2 muons largely improve the $p_T$ resolution of the reconstructed muons and this result in a significantly larger efficiency for displaced muons at Run2, as it is demonstrated in figure 9 where the efficiency for this kind of muons is plotted as function of the muon $p_T$.

4. Conclusions
The High Level Trigger of CMS is in good shape and ready for data taking at 13 TeV during the Run2 of LHC. Improvements in the software, and in the producers and filtering techniques will allow an efficient data taking for all the relevant physics searches and analyses foreseen at the LHC Run2. Thanks to it, the CMS collaboration does not expect to be forced to increase too much the selection thresholds of the filtering modules of its trigger, in spite of working in a much harsher environment than at Run1. With such an improved HLT, CMS will not close its door to possible new physics knocking at it during Run2: it will then be a task for the analysis teams in CMS to verify whether there is actually anything out there.

References
[1] CMS Collaboration, “The CMS experiment at the CERN LHC”, JINST 3 (2008) S08004.
[2] A. Bocci, “The CMS High Level Trigger”, these proceedings.
[3] CMS Collaboration, “Commissioning of the Particle Flow Event Reconstruction with the First LHC Collisions Recorded in the CMS Detector”, CMS-PAS-PFT-10-001 (2010).
[4] M. Cacciari and G.P. Salam, “Pileup subtraction using jet areas”, Phys. Lett. B 659 (2008), 119.
[5] M. Tosi, “Performance of Tracking, b-tagging and Jet/MET reco at the CMS High Level Trigger”, these proceedings.
[6] A. Aizuddin Bin Anuar, “Electrons and photons at High Level Trigger in CMS for Run II”, these proceedings.
[7] B. Radburn-Smith, “Performance of muon-based triggers at the CMS High Level Trigger”, these proceedings.