CMS High Level Trigger Timing Measurements

Clint Richardson
Boston University, USA
E-mail: clint.allan.richardson@cern.ch

Abstract. The two-level trigger system employed by CMS consists of the Level 1 (L1) Trigger, which is implemented using custom-built electronics, and the High Level Trigger (HLT), a farm of commercial CPUs running a streamlined version of the offline CMS reconstruction software. The operational L1 output rate of 100 kHz, together with the number of CPUs in the HLT farm, imposes a fundamental constraint on the amount of time available for the HLT to process events. Exceeding this limit impacts the experiment’s ability to collect data efficiently. Hence, there is a critical need to characterize the performance of the HLT farm as well as the algorithms run prior to start up in order to ensure optimal data taking. Additional complications arise from the fact that the HLT farm consists of multiple generations of hardware and there can be subtleties in machine performance. We present our methods of measuring the timing performance of the CMS HLT, including the challenges of making such measurements. Results for the performance of various Intel Xeon architectures from 2009-2014 and different data taking scenarios are also presented.

1. Introduction
The CMS experiment [1] at CERN employs a non-traditional two-level triggering system for the selection of events to store for offline reconstruction. Traditional triggering systems in hadron collider experiments make use of three levels, starting with a hardware based system and then gradually introducing reconstruction software, to successively decrease the rate of input events to a level which can be stored and reconstructed offline. The CMS detector however, uses a two-level system where the first level, Level 1 (L1), is hardware based and the second level, High Level Trigger (HLT), consists of a farm of commercial CPUs that run a streamlined version of offline reconstruction software. Hence, the CMS HLT faces a larger input rate than the last tier of traditional three-level systems. The L1 trigger, comprised of custom hardware and electronics, reduces the LHC input rate of 40 MHz down to 100 kHz which is then fed as input to the HLT. Such a large input rate, together with the number of CPUs available in the HLT farm, puts a fundamental constraint on the time the HLT can spend processing events.

In 2015, the HLT will run in considerably different conditions than in 2012 and so a clear need of the CMS experiment is to derive an accurate estimate of HLT performance before main data taking occurs. The increased center-of-mass energy ($\sqrt{s}$) and instantaneous luminosity at which the LHC will operate in 2015 represent new challenges for the HLT. Further, the HLT farm will consist of three different generations of CPU during the 2015 run. Hence, in order to derive an accurate estimate of HLT performance, the effects of all these differences must be carefully studied.
2. Hardware Tested
Performance tests were carried out on four different generations of Intel CPUs: the X5650 based on the Westmere architecture, the E5-2670 based on the Sandy Bridge architecture, the E5-2650v2 based on the Ivy Bridge architecture, and three Haswell architecture based CPUs (E5-2680v3, E5-2690v3, and E5-2697v3). The first two were used in the HLT farm to take data during 2012 and will continue to be used during 2015. The Ivy Bridge based machine currently serves as a benchmark machine for CMS timing studies. The Haswell based processors were tested in order to measure expected performance gain from the newer CPUs to be used in 2015. The HLT farm will consist in large part of machines with E5-2680v3 CPUs during 2015 data taking. All machines tested had two CPU sockets. Table 1 denotes the complete specifications for all machines tested while Table 2 lists a breakdown of the configuration of the HLT filter farm in 2015.

3. Tests Performed
The effects of three main changes with respect to 2012 running were measured in order to derive an estimate of HLT performance in 2015: changes in hardware, changes in instantaneous luminosity, and changes in online reconstruction software. Further, in order to accurately compare the results of different tests, the performance of each machine had to be measured throughout a range of testing configurations. These latter tests were necessary in order to isolate the effects of Intel’s TurboBoost and HyperThreading technologies. Adding to the importance of measuring such effects is the need to extrapolate the performance of a single test job to the performance of the HLT farm as a whole, essentially the difference between running a single job on a machine and running the machine fully loaded. Here, and throughout, the word job means a single instance of HLT software.

3.1. Performance vs. CPU Occupancy
In order to understand the results of the following tests, first a detailed understanding of machine performance at different levels of occupancy was needed. When possible, machines were tested starting with a single job and increasing the number of jobs until the number running was equal to the number of virtual cores on the machine. Because all machines tested had Intel’s HyperThreading technology, the number of virtual cores is twice the number of physical cores reported in Table 1. The details of the CPU loading are as follows: each job was tasked to a specific core, each CPU on the machine was filled serially, and both CPUs were filled before HyperThreading was used. The Westmere and Sandy Bridge based machines however, do not contain enough memory to completely fill the machine with jobs. The amount of RAM per job is roughly 1.1 GB and hence only 22 and 30 jobs, respectively, could be run on those machines. For two of the Haswell based machines tested, CMS did not have access to the machines long enough to perform all tests desired, and so an abbreviated test of performance versus level of CPU occupancy was performed. All tests measuring performance versus CPU occupancy used
Table 2. Configuration of the CMS HLT filter farm in 2015

| Year Installed | 2011             | 2012             | 2015             |
|---------------|------------------|------------------|------------------|
| CPU (Architecture) | X5650 (Westmere) | E5-2670 (Sandy Bridge) | E5-2680v3 (Haswell) |
| CPUs per Motherboard | 2               | 2               | 2               |
| Cores per CPU | 6               | 8               | 12              |
| RAM           | 24 GB            | 32 GB            | 64 GB           |
| Base Freq. (Max) | 2.66 (3.06) GHz | 2.60 (3.30) GHz | 2.50 (3.30) GHz |
| Number of Cores in Farm | 3456            | 4096            | 8640            |

2012 proton-proton collisions data for input. The data were collected requiring events to pass any L1 trigger algorithm and had an instantaneous luminosity that corresponded to an average of 30 secondary collisions (pileup).

Figure 1 shows the results of the full test, left, and the abbreviated test, right. Qualitatively similar behavior is seen across all CPU generations and one can clearly see the effects of both TurboBoost and HyperThreading. While the Sandy Bridge and Ivy Bridge machines performed similarly, they bring a performance improvement of roughly 20% compared to the Westmere machine. Further, the Haswell machines, which all perform similarly to one another, bring roughly another 20% performance improvement.

Figure 1. Average processing time for various CPU configuration scenarios. Timing was measured using 2012 8 TeV data consisting of events collected by only requiring a Level 1 Trigger Accept. The tests were performed using the 2015 CMS HLT reconstruction software over data which had an average of 30 pileup collisions. The meaning of bin labels for the right plot is as follow: 1job - 1 job on the CPU; 1job NUMA - one job running on each NUMA node; NUMA - running a single CPU with one of its NUMA nodes filled; CPU - running the machine with one of its CPUs fully loaded; Full - running both CPUs on the machine fully loaded; 2 jobs HT - two jobs running on the same core using HyperThreading; NUMA HT - the same as NUMA but doubling the jobs and using HyperThreading; CPU HT - the same as CPU but doubling the number of jobs and using HyperThreading; Full HT - the same as Full but doubling the number of jobs and using HyperThreading. NB: The Sandy Bridge E5-2670 points only go up to 30 because the test machine did not have enough RAM available to run 32 jobs at once. Similarly, the Westmere X5650 machines only had enough RAM to run 22 jobs simultaneously and hence the test for it stops there.
3.2. Performance vs. Pileup
During 2012 proton-proton collisions CMS saw an average of between 20 and 30 pileup collisions during a standard LHC fill. However, in 2015, that number is expected to increase to 40 during the main data taking runs of the year. In order to characterize the effect that this increase in pileup will have, HLT performance was measured at seven distinct levels of pileup using 2012 proton-proton collision data. The four lower pileup points were taken during normal LHC running. The three higher pileup points were taken during special LHC fills which were run in specific conditions to generate high levels of pileup. All levels of pileup were tested using all machines to which CMS had access. HLT performance is seen to degrade linearly with the number of pileup collisions. The difference in slope between the regular fills and the special fills is due to the lack of out-of-time pileup in the special fills. Performance is qualitatively the same between all CPU generations. Similar improvements to those seen in the CPU occupancy tests manifest themselves again: The Sandy Bridge and Ivy Bridge machines bring roughly a 20% performance improvement with respect to the Westmere machine, while the Haswell machines bring another 20% improvement over the Sandy Bridge machine. Figure 2 shows these results.

![Figure 2](image.png)

Figure 2. Average processing time versus pileup for several different CPU generations. The performance was measured using 2012 8 TeV data consisting of a set of events passing any Level 1 Trigger. The machines were tested running with one CPU fully loaded without HyperThreading and using the 2015 CMS HLT reconstruction software. The difference in slope between the pileup 20 to 33 points and those between 44 and 63 is due to the fact that the higher pileup runs were taken without out-of-time pileup present.

3.3. Performance improvements from new CMS reconstruction software
During the intervening years between Run 1 and Run 2 of the LHC, the reconstruction software used by the CMS HLT underwent several changes with an eye to quickening the time it takes the HLT to reconstruct an event. In order to measure the performance improvements from these changes, a comparison was done between the exact same set of selection algorithms when running using the 2012 reconstruction software and when running using the 2015 reconstruction software. The test was performed using 2012 proton-proton collision data with events which had...
an average of 30 pileup interactions and which were required to pass any L1 algorithm. Further, the comparison was performed across a full scan of CPU occupancy using the Sandy Bridge machine. Figure 3 shows the detailed results and one can see an across the board improvement of roughly 25% when using the 2015 HLT reconstruction software compared to that used in 2012.

![Figure 3](image)

**Figure 3.** Average processing time per event as a function of CPU load. Each job is tasked to a specific processor so that HyperThreading becomes active at point 17, where both CPUs have been filled and one extra job is added. Timing was measured using 2012 8 TeV data consisting of a set of events which pass any Level 1 Trigger. The black points show performance using the HLT reconstruction software used in 2012 while the blue show the performance using the upgraded software which CMS will deploy in 2015. The HLT menu in both scenarios is the same so that the new reconstruction software is running the same selection algorithms as those used in 2012.

4. Results and expected HLT performance in 2015

Combining the information of the tests above, together with the number of cores of each generation of machine which will be present in the HLT farm in 2015 (Table 2), one can derive a timing budget for the HLT. Because of the diverse configuration of the farm, this budget is written as the maximum allowed average processing time per event for each machine, were the farm made up entirely of that type of machine. When projecting the farm to be made of only one type of machine, the relative performance is taken into account so that, for instance, each Sandy Bridge machine counts as roughly 1.2 Westmere machines. Hence, writing the timing budget for each machine in this manner accurately represents how quickly events need to be processed on each generation of machine. Further, it easily allows for understanding the results of the tests of each machine and what percentage of the timing budget the results actually consume. When scaled to the Ivy Bridge machine which CMS uses as a benchmark for timing studies, the timing budget for a single job running alone on a CPU is roughly 160 ms per event. Table 3 shows the full results.

In the most extreme conditions expected during 2015 the HLT will need to process data which has an average of 40 pileup collisions, corresponding to the highest projected level of
Table 3. Timing budget per machine generation for the CMS HLT filter farm in 2015. *The filter farm does not contain any Ivy Bridge machines, but since it is used by CMS as a benchmark machine it is listed here.

| CPU                  | Single Job | Fully Loaded (with HyperThreading) |
|----------------------|------------|-----------------------------------|
| X5650 (Westmere)     | 226 ms     | 367 ms                            |
| E5-2670 (Sandy Bridge)| 172 ms     | 310 ms                            |
| E5-2650v2 (Ivy Bridge)* | 162 ms     | 316 ms                            |
| E5-2680v3 (Haswell)  | 141 ms     | 283 ms                            |

instantaneous luminosity \((1.4 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1})\). During the year, as the LHC ramps up to this most intense scenario, it will for a significant period of time operate at a level of 20 pileup collisions (or an instantaneous luminosity of \(7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}\)). Hence CMS is preparing for both of these main data taking scenarios. In order to derive an estimate for the HLT performance in these conditions, the HLT software was run over simulated proton-proton events for both pileup scenarios and at the higher \(\sqrt{s} = 13 \text{ TeV}\) center-of-mass energy. Running a single job on the Ivy Bridge benchmark machine and using the 2015 HLT reconstruction software, the average processing time is found to be 162 (66.5) ms for the scenario with 40 (20) average pileup collisions. This result is perfectly consistent with the timing budget derived above.

Figure 4. Processing time distribution for both main instantaneous luminosity scenarios. The black line shows the performance for an instantaneous luminosity (average number of pileup collisions) of \(7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}\) (20) while the blue shows performance for \(1.4 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}\) (40). The timing was measured using 13 TeV Monte Carlo simulating proton-proton collisions which were required to pass any Level 1 Trigger and represents expected HLT performance in 2015. The CPU used for measuring this performance was the Ivy Bridge based E5-2650v2 which was configured to run only a single job for each test.
5. Conclusion
A detailed study of the performance of the CMS HLT farm, in terms of its hardware, software, and running conditions has been performed. Using the relative performance and number of the different machines in the HLT farm, a timing budget for 2015 running has been derived and is estimated to be roughly 160 ms when running a single job on the benchmark machine used by CMS for timing studies. The performance of the current set of selection algorithms employed by the HLT to select events for offline storage has been studied using Monte Carlo simulation of proton-proton collisions with the conditions expected for the 2015 data taking period. The performance of this set of selection algorithms is consistent with the timing budget even for the most challenging data taking scenario.

References
[1] Chatrchyan S et al. (CMS) 2008 JINST 3 S08004