

LHCb Run 2 trigger performance

Barbara Sciascia^{*†}

Laboratori Nazionali di Frascati, INFN

E-mail: barbara.sciascia@cern.ch

During the first long shutdown of the LHC (2013-2014, LS1), the LHCb detector remained essentially unchanged, while the trigger system has been completely revised. Upgrades to the LHCb computing infrastructure have allowed for high quality decay information to be calculated by the software trigger, making separate offline event reconstruction unnecessary. Reaching the ultimate precision of the LHCb experiment in real time as the data arrive has the power to transform the experimental approach to processing large quantities of data.

*16th International Conference on B-Physics at Frontier Machines
2-6 May 2016
Marseille, France*

^{*}Speaker.

[†]on behalf of the LHCb Collaboration.

LHCb is a dedicated heavy flavour physics experiment at the Large Hadron Collider (LHC). The experiment [1] is designed for precision measurements of CP violation and rare decays of beauty and charm hadrons. LHCb published more than 320 papers using mainly Run 1 (2010-2012) data. During Long Shutdown 1 (LS1) the LHCb detector remained essentially unchanged, while major upgrades are foreseen for subsequent long shutdowns [2][3]. What has been completely revised is the software part of the high-level trigger (HLT). This paper describes the characteristics of the new HLT and its performance during Run 2.

In 2015, for the beginning of Run 2, the LHC operation team commissioned the accelerator for a higher target energy, moving from 4 TeV/beam achieved in 2012 to 6.5 TeV/beam. At the centre-of-mass energy of 13 TeV the LHCb experiment collected a total of 320 pb^{-1} in 2015 and of 840 pb^{-1} at present in 2016. Even though the LHC accelerator increased the total intensity of the beams throughout the whole year, the LHCb experiment collected data at a fixed value of μ , defined as the average number of visible pp interactions per bunch crossing. This was possible owing to a mechanism called *luminosity levelling* [4], where the beams are separated in the vertical plane in order to keep the target value of luminosity and pileup. This allows the detector to record data in a stable way, collecting a very homogeneous dataset, thus easing the online and offline processing of events and allowing for consistent selections of physics candidates.

The LHCb detector is a single-arm forward spectrometer that covers the pseudorapidity range $2 < \eta < 5$. The detector includes a high-precision tracking system consisting of a silicon-strip vertex detector (VELO) [6] surrounding the pp interaction region, a large area silicon-strip detector (TT) located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors (IT) and straw drift tubes (OT) [7] placed downstream. The combined tracking system has a momentum resolution $\delta p/p$ that varies from 0.4% at 5 GeV/c to 0.6% at 100 GeV/c and an impact parameter resolution of $20 \mu\text{m}$ for tracks with high transverse momentum. Charged hadrons are identified using two ring-imaging Cherenkov (RICH) detectors [8]. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic calorimeter and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers [9].

1. Trigger system

The online event selection is performed by a trigger [10] which consists of two levels. At the first stage, Level-0 (L0), which is implemented in hardware and based on information from the calorimeter and muon systems, the event rate is reduced to below 1 MHz. At this rate the whole detector can be read out. Typical values (2012) for the thresholds in L0 are $p_T > 1.76 \text{ GeV}$ for the single muon selection, while for the di-muon selection the product of the p_T has to be $> 1.6 \text{ GeV}^2$; $E_T > 3.7 \text{ GeV}$ and $E_T > 3 \text{ GeV}$ are the thresholds to select hadrons, and photons and electrons, respectively. Low multiplicity events can also be triggered and very complex events are filtered out. The efficiencies of the L0 triggers as measured in 2012 data are shown in Fig. 1.

The trigger hardware level will be removed during the next long shutdown (LS2), and the detector will be read out at 40 MHz. This requires a complete redesign of the DAQ and trigger which will also recover efficiency at low p_T [11].

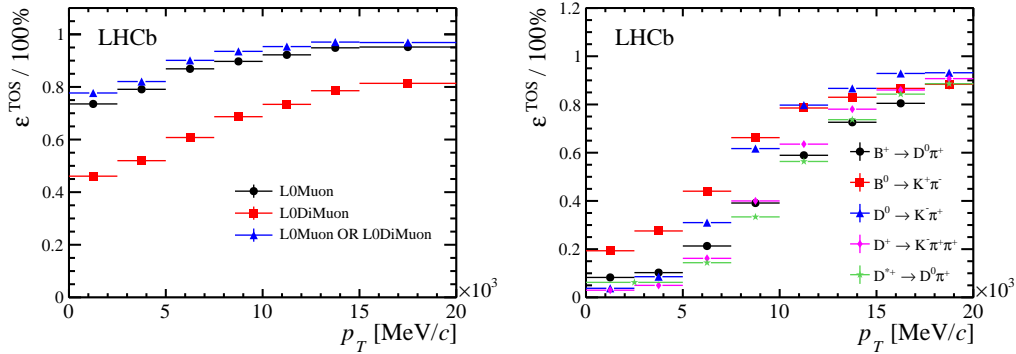


Figure 1: Left: L0 muon trigger efficiency for selected $B^+ \rightarrow J/\psi(\mu\mu)K^+$ candidates as a function of B p_T . Right: L0 hadron trigger efficiency for different selected beauty and charm decay modes as a function of B or D p_T .

The second stage, HLT, consists of software algorithms implemented in the same framework as the software used for the offline reconstruction. The HLT runs on an Event Filter Farm (EFF) of $\sim 52,000$ logical CPU cores ($\sim 29,000$ in Run 1) situated in the cavern close to the detector, outside the radiation area. Events selected by the HLT are sent to the storage system for later offline analysis.

The HLT is in turn divided in two levels (see Fig. 2, left). In the first level, HLT1, a partial event reconstruction is performed. HLT1 reconstructs charged particle trajectories using information from the VELO and tracking stations. If at least one track is found that satisfies strict quality and transverse momentum criteria, then the event is passed to the second level of the software trigger (di-muon combinations may also trigger the first software level). HLT1 thresholds can be tuned for an optimal output rate (in 2012 the rate was ~ 150 kHz).

In the second level, HLT2, the complete event is reconstructed. To satisfy the time constraints, the HLT event reconstruction in Run 1 was simpler than that used offline. Also the HLT was unable to use the latest alignment and calibration constants that were calculated at a later stage and used in the subsequent offline reconstruction.

During Run 1, LHC delivered stable colliding beams for about 30% of the time. HLT operation was synchronous with the delivery of the particle collisions when the front-end boards send to the local disk of each HLT node data. The latter have to be quickly processed by the HLT1 to reduce the amount of data sent to the slower HLT2 level. To profit from the inactivity time of the farm between an LHC fill and another, in 2012 $\sim 20\%$ of the L0-passed events were buffered to local disks and the HLT selection applied later to these events (see Fig. 2, centre). The result was a sizable improvement of $\sim 25\%$ of the effective CPU power. Nevertheless for large periods without beam, the HLT farm was still idle.

During Run 2, HLT1 and HLT2 became two independent asynchronous processes running on the same node. The implementation of this separation required a substantial change of the flow of event data in the experiment. All the events passing the HLT1 selection are buffered in the local disk and later processed by HLT2 (see Fig. 2, right). The allocated resources allow for a maximal execution time of ~ 35 ms/event in HLT1 and of ~ 650 ms/event in HLT2. Since events not accepted

by HLT1 are lost for later offline physics analysis, while in collision the HLT1 processes run with high priority. At the same time HLT2 runs on each node sending the finally accepted events to long-term storage, however during data taking HLT2 has a lower priority to not negatively interfere with the HLT1 processing.

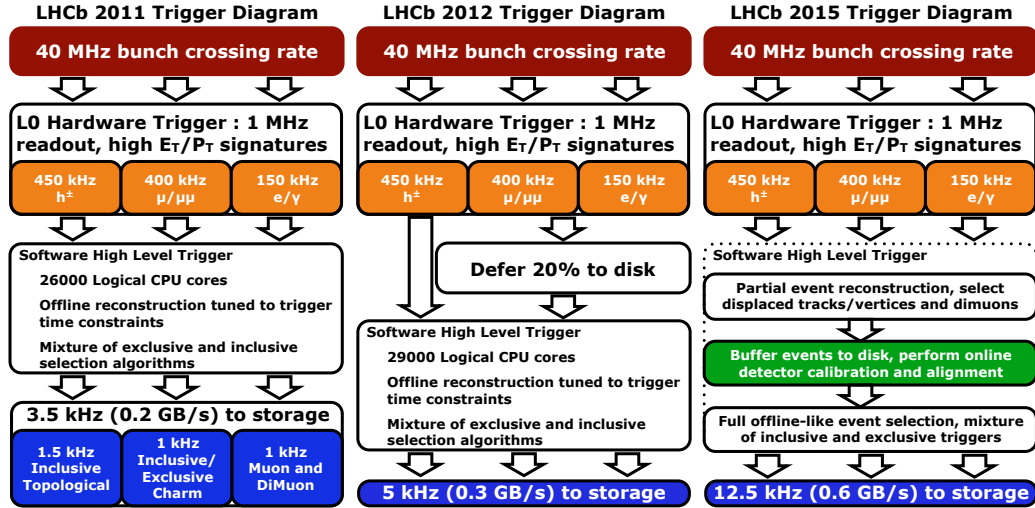


Figure 2: From left to right: HLT schemes in Run 1/2011 (no events deferred), Run 1/2012 (20% of L0 accepted events deferred for a later processing), and Run 2 (all HLT1 accepted events deferred). (See text for details.)

The separation of the two HLT levels allows to implement the high-quality alignment and calibration before running the HLT2 on the events selected by HLT1. Using the optimal calibration constants, this allows to have offline-quality reconstruction already in the trigger.

As an added benefit to the additional computing resources and the optimized code, in HLT1 tracks are reconstructed down to $p_T=500$ MeV. The main features of HLT1 are single and two track multivariate algorithms (MVA) [5] which provide inclusive charm and beauty selection with improved performance with respect to Run 1 (e.g. the 2-body charm trigger is a factor ~ 2 more efficient), inclusive single and di-muon triggers, and exclusive lines for the “lifetime unbiased” beauty and charm selection (selections which do not cut on quantities correlated with the decay time of the signal particle).

HLT2 performs the full event reconstruction starting from the information (vertices and tracks) calculated in HLT1. It reconstructs all tracks, while in Run 1 a $p_T > 300$ MeV cut was initially applied, and HLT2 builds the full particle identification information for long tracks. Reconstruction is followed by physics selections, including MVA lines based on 2-, 3-, and 4-body detached vertices to inclusively select beauty and charm decays, exclusive beauty decays such as $B \rightarrow \phi\phi$ and $B \rightarrow \gamma\gamma$, charmed baryons, and electroweak bosons.

A further improvement on the global HLT quality comes from the management of the storage disks. During the first year of Run 2, the storage of the HLT farm was 5 PiB. As a figure of merit, this allows for 160 hours of data taking with an HLT1 output rate of ~ 150 kHz (60 kiB/event). For redundancy, in 2015 the whole storage was mirrored. From the 2012 experience of disk failure, the

risk of data loss due to unrecoverable errors has been evaluated to be $\sim 0.1\%$. The per mille loss expected without mirroring the HLT disks has to be compared with a potential $\sim 15\%$ increase of triggered charm physics or with an improved quality of the online reconstruction. Starting from 2016 data taking, the HLT disks are unmirrored and the whole 10 PiB is used in the HLT running. The data loss rate is monitored and agrees with expectations.

2. Alignment and calibration

The increased computing resources in the EFF allows the deferral of all HLT1-accepted events and the alignment and calibration tasks to supply high-quality information to HLT2. This removes the need for further reprocessing. In order to align and calibrate the detector, dedicated samples from HLT1 are taken as input. The automatic calibration and alignment is performed in the trigger farm within a few minutes. Dedicated samples from HLT1 are taken as input for each system, for instance $D^0 \rightarrow K^+\pi^-$ candidates for the Tracker and $J/\psi \rightarrow \mu^+\mu^-$ candidates for the Muon system. These calibrations are done before HLT2 has processed the data, allowing some physics results to be performed directly on a dedicated trigger output (see Section 3). The calibrations are also used by HLT1 for subsequent processing. The alignment and calibration tasks are performed at regular intervals which can be as frequent as each time a new LHC proton beam is formed or less frequent depending on the calibrations being calculated. The calibration tasks are performed in a few minutes using the nodes from the EFF. The resulting alignment or calibration parameters are updated if they differ significantly from the values stored in the database at a given time. Details of the real-time alignment procedure are provided in [12] and are summarised below.

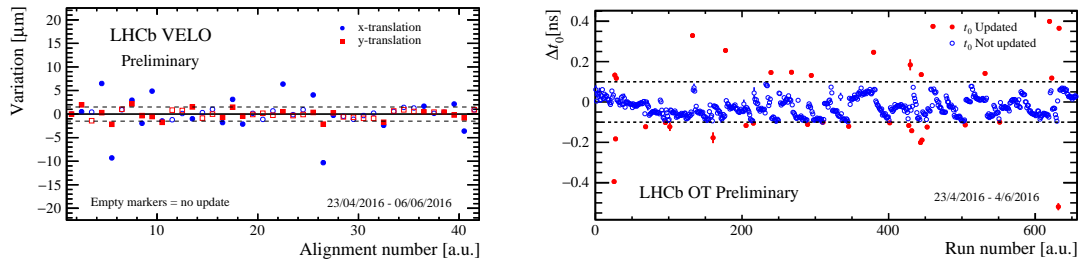


Figure 3: Left: Stability of the VELO halves alignment. Right: OT global t_0 stability. In both cases the differences were calculated in 2016 data with the latest calibrations with respect to the previous values used.

Misalignment of the VELO and tracking stations has a direct impact on the momentum resolution of charged particles. The alignment is achieved through minimisation of the residuals of a Kalman fit [13] to a set of well reconstructed tracks from HLT1. The alignment of the VELO is performed first, followed by the corresponding alignment of the tracking stations. This order is chosen due to the nature of the VELO, which is centred around the collision point from a retracted position after each beam injection

The outer tracker uses a drift (gas) tube design [7]. The measurement of the drift time in the straw tubes is susceptible to differences between the true collision time and the LHCb clock. Such a difference causes the measured drift time to be different from the estimated time arising from the

distance of the wire to the track. The residuals of a sample of well reconstructed tracks are used to provide a global drift time offset, t_0 , for the tubes. Examples of the stability of the alignment and calibration tracking constants for the VELO and the OT are shown in Fig. 3.

Misalignment of the RICH detectors causes the nominally circular rings on the detection plane to become distorted which means the distance from the projected position of the track to the Cherenkov ring varies as a function of the azimuthal angle. This distortion requires an individual correction for each mirror which is calculated using a set of well reconstructed tracks from HLT1. The refractive indices of the gas radiators vary as a function of temperature and pressure, which change over time. From a set of well reconstructed tracks originating from a particle of known mass, e.g. from $D^* \rightarrow D(K\pi)\pi$ events, the expected Cherenkov angle can be calculated using accurate momentum measurements provided by the tracking stations. The distribution of expected versus measured Cherenkov angles provides the basis for the refractive index calibration of the radiator.

The complete calibration of the full detector is a complex enterprise. The achievement of automating and providing accurate calibrations within a few minutes is a substantial achievement without which analysis-quality reconstruction in HLT2 would be impossible.

3. Turbo Stream

The quasi real-time alignment and calibration available between HLT1 and HLT2 provides online the same offline-quality reconstruction and allows to achieve a more effective selection and a higher signal purity of relevant decay channels. To exploit this possibility, a new dedicated trigger output has been defined, commonly referred to as Turbo Stream [14]. This approach allows physics candidates to be available from the Turbo Stream with optimal calibration and alignment, ready for physics analyses only a few hours after having been recorded. The events directed to the Turbo Stream consist of the Turbo-flagged physics lines. This stream does not require further event reconstruction, since data are ready for user analysis directly once the trigger information is restored through the Tesla [15] application designed to process the information calculated by the trigger, with the resulting output used to directly perform physics measurements. Since all sub-detector information may be discarded, a further advantage of the Turbo Stream is that the event size is O(1 KiB), an order of magnitude smaller than data following the traditional full stream.

The first full test of the functionalities of the Turbo stream has been performed using the first proton-proton collision data of the LHC Run 2. These series of early measurements included the cross-sections for quarkonia, beauty and charm productions.

Measurement of the J/ψ production cross-section in early Run 2 data pioneered the usage of the Turbo Stream technique for selection of the decay candidates [16]. The number of the prompt J/ψ and the J/ψ -from-b candidates is estimated from a simultaneous maximum likelihood fit of the unbinned distributions of the invariant mass and the pseudo-decay time, t_z ¹, chosen as the best approximation of the lifetime. The mass fit is used to separate inclusive J/ψ candidates from combinatorial background; the t_z fit allows to estimate contributions from peaking backgrounds, prompt J/ψ and the J/ψ -from-b separately. Fig. 4 shows the fit to the invariant mass and t_z

¹ defined as $((d_z \times M_{J/\psi})/p_z)$, with d_z the distance between the pp collision and the J/ψ decay point.

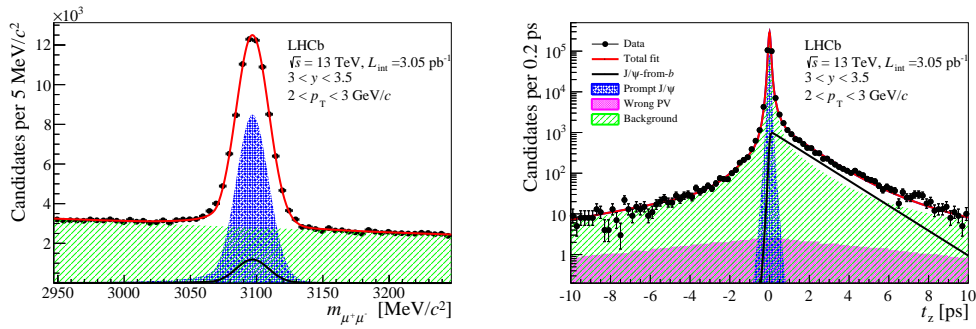


Figure 4: Invariant mass (left) and pseudo-decay time (right) distributions for a specific transverse momentum (p_T) and rapidity (y) bin, with fit results superimposed. The solid (red) line is the total fit function, the shaded (green) area corresponds to the background component. The prompt J/ψ contribution is shown by the cross-hatched area (blue), J/ψ -from- b by a solid (black) line and the tail contribution due to the association of J/ψ with the wrong primary vertex is shown by the full filled (magenta) area. The tail contribution is not visible in the invariant mass plot.

distributions available directly from the Turbo Stream after Tesla processing. The preliminary results of the J/ψ were presented one week after the the data taking of the analysed sample.

4. Conclusions

The LHCb trigger in Run 2 has been largely improved with respect the Run 1 configuration. LHCb is the first experiment in high energy physics implementing a fully automatic tracking system alignment, PID calibration and track reconstruction in the online system. This has been achieved by a major effort in tracking improvements and development of the new automatic procedure during LS1. The work includes the reorganization of the software trigger to allow buffering after HLT1 and a global software optimization to accommodate the reconstruction within the HLT time budget. The full offline-quality reconstruction avoids the need to reconstruct the data offline, allowing for a more effective use of computing resources. The new tools, developed to run directly on a dedicated trigger output, save storage resources and allow physics results to be achieved a few days after data taking, representing a working model for future experiments.

References

- [1] Alves Jr A. A. et al. (LHCb Collaboration), The LHCb detector at the LHC, 2008 JINST 3 S08005.
- [2] S. Akar, The LHCb Upgrade, J. Phys. Conf. Ser. 556, 2014.
- [3] F. Teubert, these proceedings.
- [4] Jacobsson R. et al., The LHCb Online Luminosity Monitoring and Control, in proceedings of ICALEPCS2013, 2013.
- [5] A. Hoecker et al, TMVA: Toolkit for Multivariate Data Analysis, PoS ACAT (2007) 040, arXiv:physics/0703039..

- [6] Aaij R. et al. (LHCb Collaboration), Performance of the LHCb Vertex Locator, 2014 JINST 9 P09007.
- [7] Arink R. et al., Performance of the LHCb Outer Tracker, 2014 JINST 9 P01002.
- [8] Adinolfi M. et al., Performance of the LHCb RICH detector at the LHC, 2013 Eur.Phys.J.C73 2431.
- [9] Alves Jr A. A. et al., Performance of the LHCb muon system, 2013 JINST 8 P02022.
- [10] Aaij R. et al., The LHCb trigger and its performance in 2011, 2013 JINST 8 P04022.
- [11] Albrecht J. et al., The upgrade of the LHCb trigger system, Proceedings, Workshop on Intelligent Trackers (WIT2014).
- [12] G. Dujany and B. Storaci, Real-time alignment and calibration of the LHCb Detector in Run II, J. Phys. Conf. Ser. 664 (2015), no. 8 082010.
- [13] R. Fruhwirth, Application of Kalman filtering to track and vertex fitting, Nucl. Instrum. Meth. A262 (1987) 444.
- [14] Benson S. et al., The LHCb Turbo Stream, in proceedings of CHEP2015, 2015.
- [15] R. Aaij et al., Tesla: An application for real-time data analysis in High Energy Physics, Computer Physics Communications (2016) doi:10.1016/j.cpc.2016.07.022.
- [16] R. Aaij et al., Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV, JHEP10(2015)172.