

## The LHCb trigger system: performance and outlook

---

**Simone Stracka**\* †

*Scuola Normale Superiore and INFN Pisa*

*E-mail:* [simone.stracka@cern.ch](mailto:simone.stracka@cern.ch)

The LHCb experiment is a spectrometer dedicated to the study of heavy flavor at the LHC. The rate of proton-proton collisions at the LHC is 15 MHz, of which only 5 kHz can be written to storage for offline analysis. The trigger system plays a key role in selecting signal events and rejecting background, and is comprised of a hardware level (L0), reducing the rate to the maximum at which the detector can be fully read out, and a High Level Trigger (HLT) -implemented in software and deployed on a farm of roughly 25 000 parallel processing cores- responsible for reducing the rate to the 5 kHz which can be processed offline. The LHCb trigger system allowed LHCb to run at twice its design luminosity in 2012, and performed beyond the nominal design in terms of signal yields. The design and performance of the selection algorithms are discussed in the context of the 2012 data taking, and planned improvements for RunII are presented.

*Technology and Instrumentation in Particle Physics 2014,  
2-6 June, 2014  
Amsterdam, the Netherlands*

---

\*Speaker.

†on behalf of the LHCb Collaboration.

## 1. Introduction

The LHCb experiment is a single-arm forward spectrometer for precision measurements in the heavy flavor sector at the Large Hadron Collider, with an acceptance extending from 2 to 5 in pseudorapidity. The detector [1] consists of a high-precision tracking system, two ring-imaging Cherenkov detectors for particle identification purposes, hadronic and electromagnetic calorimeters for measuring the energies of photons and neutral hadrons, and a muon system consisting of alternating layers of iron and multiwire proportional chambers. The tracking capabilities rely on a vertex locator (VELO) surrounding the  $pp$  interaction region and two tracking systems, located upstream (TT) and downstream (T) of a 4Tm dipole magnet and instrumented with silicon-strip detectors and straw tubes.

In 2012, LHCb collected proton-proton collision data at 8TeV and at an instantaneous luminosity of  $4 \times 10^{32} \text{cm}^{-2} \text{s}^{-1}$ . In these conditions, beauty and charm hadrons are produced in acceptance at rates exceeding 30kHz and 600kHz respectively (these rates correspond to the production cross-sections at 7TeV,  $\sigma_{H_b} = 75.3 \pm 14.1 \mu\text{b}$  and  $\sigma_{H_c} = 1419 \pm 134 \mu\text{b}$  [2, 3]), while light quark events represent the dominant contribution to the 15 MHz rate.

The LHCb trigger system must fulfill two requirements: first, it is responsible for reducing the event rate to the maximum rate at which the full detector response can be digitized and read out (1.1 MHz); second, it must be highly selective and maximize the physics potential of the experiment within the maximum allowed output bandwidth for writing the events to permanent storage (corresponding to 5 kHz in 2012). These tasks are performed by two trigger levels: the first level, called L0, is performed in hardware and provides a read-out decision in  $\approx 4 \mu\text{s}$ , after which data are processed in the second layer, a fully-software High-Level Trigger (HLT).

## 2. L0-Trigger Overview

The strategy for data reduction adopted by the LHCb trigger system is common to many flavor experiments at hadron colliders, and is based on the identification of muon tracks and of displaced high transverse-momentum tracks, which allow to exploit the key signatures of heavy flavors: long lifetime and relatively high mass of  $b$  and  $c$  hadrons; muon signatures, associated to several rare or semileptonic decays of interest, further benefit from lower multiplicities in the muon stations.

While muons are readily identified at hardware level, for hadrons the use of tracking information is necessary to reduce the background rate to the desired levels. Since tracking is computationally demanding and the tracking devices are anyway subject to the 1.1 MHz maximum read-out rate limitation, the track-based selection is postponed to the software trigger, and the rate is preliminarily regulated in hardware based on calorimetry information [4].

The L0 muon trigger is based on track segments pointing towards the interaction point and reconstructed in the muon stations, which allow to determine muon tracks momenta with a resolution of 25% compared to the offline reconstruction. The selection is then based on the two highest  $p_T$  track segments (a single muon  $p_T$  above 1.76 GeV or  $p_T$  product above  $(1.6 \text{ GeV})^2$  for any pair of candidates are required), resulting in an efficiency for  $B$  decays with muons in the final state of 90% at 400 kHz, as evaluated on a  $B^+ \rightarrow J/\psi K^+$  control sample (here, and in what follows, the

efficiency is defined as the ratio of the yields of offline-reconstructed candidates, obtained with and without the selection under study [5, 6]).

The remaining two-thirds of the L0 output rate are dedicated to calorimeter-based selections. The main contribution (490kHz) is devoted to heavy quarks decaying to hadronic final states and is based on the presence of a large transverse energy deposit in the hadronic calorimeter ( $E_T \gtrsim 3.7\text{ GeV}$ ). This selection is moderately effective on  $B$ -meson decays but performs quite poorly on charm samples. For electrons and photons, contributing 150kHz to the L0 rate, a transverse-energy deposit above 3 GeV in the electromagnetic calorimeter is required, and the preshower and scintillator pad detectors are used to discriminate between the two particle hypotheses. An upper cut on the charged particle multiplicity in the event is also applied.

### 3. Deferral

The L0-accepted data is transferred to the Event Filter Farm devoted to the High Level Trigger processing. The farm consists of roughly 1000 nodes (a typical node is comprised of two Intel X5650 six-core processors at 2.67 GHz with hyper-threading, resulting in 24 logical processing cores) equipped with 1 – 2 TB disk space (for a total capacity of 1 PB). In 2012, only 80% of the data were processed in real time in the High Level Trigger, while approximately 20% of the data were buffered to the local disk-space of the nodes for processing during the gaps between the LHC machine fills: thus, the online system relies on the fact that the LHC is delivering collisions for roughly 30% of the time to optimize the usage of the online farm, reducing the idle time. The peak deferral disk space usage in 2012 did not exceed 90% of the final allocated disk space. For a short time ( $\approx 30\text{ min}$ ) during 2012 operations, all the L0-accepted events were buffered to the local disks of the farm, thus testing the robustness of the deferral system.

### 4. High Level Trigger

The High Level Trigger adopts a staged approach consisting of 2 rate-reduction steps. In the first one (HLT1) an inclusive selection mainly based on tracking information is performed. In the second one (HLT2), a full offline-like reconstruction is applied, and exclusive selections are also introduced. In this way, the input data is skimmed to fit the event output rate of 5 kHz (roughly 2 kHz are allocated for beauty events,  $\approx 2\text{ kHz}$  for charm events, and  $\approx 1\text{ kHz}$  for muon events). In 2012 HLT1 and HLT2, although logically separated entities, were run as a single process, whose total execution time was on average 35 ms per event.

The inclusive selections in HLT1 rely on the presence of one or two tracks, with impact parameter (IP) and/or transverse-momentum ( $p_T$ ) above set thresholds. HLT1 selections are also classified according to the L0 input category, apart from a single-track high-IP selection algorithm which accepts all L0-passed events. Track reconstruction is based on the approach detailed in [5], simplified with respect to the offline. In order to enhance pattern recognition speed, only the tracks originating in the VELO are considered in HLT1. After the identification of the primary vertices (originating from the  $x - y$   $pp$  interaction envelope estimated at the beginning of each fill), the 3D seeds reconstructed in the VELO undergo a first selection: e.g., for events with muon candidates passing the L0 selection, matching of the VELO seed to a segment in the muon stations is required;

for the single-track high-IP selection, seeds are selected based on the minimum impact parameter with respect to the primary vertices, and the number of hits in the VELO. The surviving seeds are then propagated through the TT and T tracking stations. In the TT-stations, the presence of hits in a search window around the trajectory resulting from a linear extrapolation of the VELO track can be required: however, because of a non-negligible magnetic field in the upstream region where the TT-stations are located, this requirement is inefficient for low-momentum tracks. In the T-stations, conversely, the extrapolation of the VELO seeds takes into account the track curvature in the magnet, and the search window is reduced by imposing minimum  $p$  and  $p_T$  requirements on the track hypotheses ( $p > 3\text{ GeV}$ ,  $p_T > 0.5\text{ GeV}$  for muon-matched seeds;  $p > 3\text{ GeV}$ ,  $p_T > 1.25\text{ GeV}$  for high-IP seeds). The pattern recognition in the T-station results in a momentum estimate which is further used to select the seeds before the Kalman fit and the final track-candidate selection (for seeds matched to a muon segment:  $p > 3\text{ GeV}$ ,  $p_T > 1\text{ GeV}$ ; for high-IP seeds:  $p > 3\text{ GeV}$ ,  $p_T > 1.6\text{ GeV}$ ).

The HLT1 selections result in a total output rate of  $\approx 80\text{ kHz}$  and are nearly 90% efficient for muon and beauty events. Since the selection is based on high-IP, high- $p_T$  tracks, the efficiency for events containing the lighter and shorter-lived charm hadrons is instead somewhat reduced, as shown in Tab. 1.

Channel	HLT1 trigger	HLT1 $\epsilon^{TOS}$ (%)
$B^+ \rightarrow J/\psi K^+$	$\mu/\mu\mu$	90/69
$B^0 \rightarrow K^+ \pi^-$	single high-IP track	86
$B^0 \rightarrow D^+ \pi^-$	single high-IP track	89
$D^0 \rightarrow K^- \pi^+$	single high-IP track	67
$D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$	single high-IP track	60

**Table 1:** HLT1 selection efficiencies ( $\epsilon^{TOS}$ ) on a set of representative channels. The offline-reconstructed candidates considered for the numerator of the efficiency are only the ones that have a correspondance with the accepted HLT1 candidate [6].

Channel	HLT2 $\epsilon^{TOS}$ (%)
$B^0 \rightarrow K^+ \pi^-$	78
$B^0 \rightarrow D^+ \pi^-$	76
$D^0 \rightarrow K^- \pi^+$	90
$D^{*+} \rightarrow D^0 \pi^+$ , $D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$	26

**Table 2:** HLT2 selection efficiencies ( $\epsilon^{TOS}$ ) on a set of representative channels. The offline-reconstructed candidates considered for the numerator of the efficiency are only the ones that have a correspondance with the accepted HLT2 candidate [6].

In contrast to HLT1, HLT2 is based on a more complete reconstruction, with tracking performances approaching those of offline reconstruction. Transverse momentum thresholds for the tracking are lowered to 300 MeV and tracks that do not originate from seeds in the VELO are also reconstructed. Particle identification information is introduced at this stage. For the selection of events containing beauty hadrons, HLT2 relies on inclusive selections based on multivariate analysis techniques trained on Monte Carlo simulated events. First, tracks are combined in two- three-

and four-track vertices. Then a 5-variable bagged decision tree (BDT) with binned discriminating variables [7] is constructed. In order to increase the efficiencies for partially reconstructed multi-body decays, a corrected mass is introduced, obtained by calculating the fraction of momentum, transverse to the flight direction, which is not accounted for by the reconstructed topology [5]. Using the BDT classifier, a rate reduction factor of nearly 50 is achieved, while keeping the efficiency close to 80% even for multibody hadronic decays (Tab. 2). Dimuons are also selected inclusively, and purity is improved by cuts on impact parameter, dimuon mass or dimuon  $p_T$ ; candidates not included in these categories are prescaled. Charm, on the contrary, demands for a wider use of exclusive selections in order to retain high efficiencies while keeping the rate under control.

## 5. Post LS1 LHCb trigger architecture

	Design	2012	2015
Instantaneous luminosity ( $\text{cm}^{-2}\text{s}^{-1}$ )	$2 \times 10^{32}$	$4 \times 10^{32}$	$> 4 \times 10^{32}$
Energy (TeV)	14	8	13
Bunch separation (ns)	25	50	25
Mean visible $pp$ interactions/crossing	0.4	1.6	$> 1$
HLT output rate to tape (kHz)	2	5	12.5

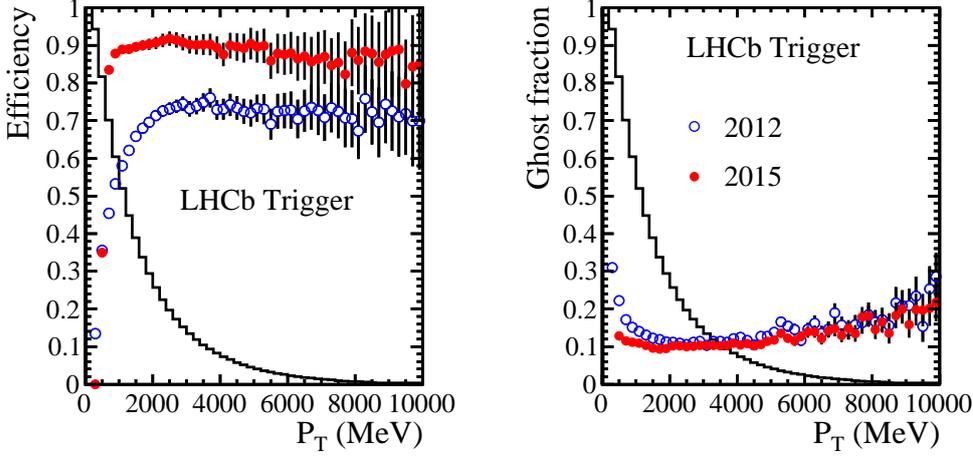
**Table 3:** LHCb design and running conditions in RunI relevant to the trigger and expected values for 2015.

Table 3 compares the running conditions in 2012 to design: the LHCb trigger was able to cope with doubled instantaneous luminosities with respect to the one initially foreseen, albeit at reduced collision energy and increased bunch separation. In 2015, the energy will be increased compared to 2012 (from 8 TeV to 13 TeV), and the peak bunch-crossing rate will reach the design value of 40 MHz. In these conditions the same instantaneous luminosity of 2012 can be achieved at a lower pile-up (1 interactions/crossing instead of 1.6); the yield will further benefit from increased cross-sections for b-hadron and c-hadron production. The requirements on the High Level Trigger will be partly relaxed by doubling the output rate for storing data to tape (12.5 kHz). On the other hand, the maximum read-out rate will be still limited to 1.1 MHz, and the new conditions will be challenging for the L0 trigger.

The computing resources will be adjusted to the 2015 conditions by doubling the Event Filter Farm CPU power and increasing the disk space for deferral to 4 PB. Changes will also be introduced in the architecture. HLT1 and HLT2 will be split in two separate physical processes, thus allowing for deferral after HLT1 rate reduction (in addition to a possible deferral after the L0 stage): this will result in a more efficient use of disk space and of the deferral mechanism, ultimately leading to an increase in the resources available for the High Level Trigger processing. Splitting the HLT stages will also allow to perform online detector alignment and calibration, which will then be available for the deferred HLT2 processing (in 2012, the calibration and fine alignment of the detector elements were obtained from an earlier data-taking period).

## 6. Revision of the HLT tracking sequence for RunII

The increased resources allocated for the High Level Trigger will be devoted to improving

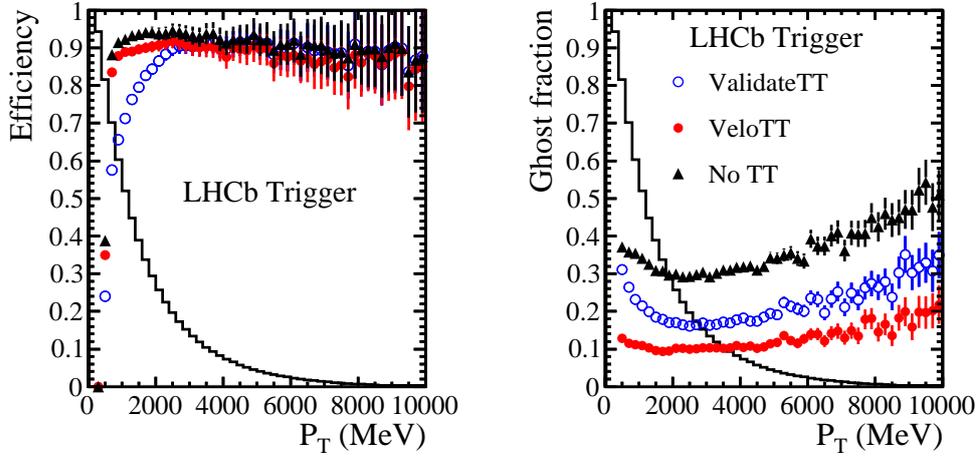


**Figure 1:** Comparison of the 2012 tracking sequence (no IP selection applied, open circles) with the 2015 tracking sequence (full circles), on Monte Carlo simulated events at 13 TeV. Left: efficiency as a function of generated particle  $p_T$ . Right: ghost rate as a function of generated  $p_T$ . The background histograms display the complementary cumulative distribution of generated  $p_T$ . Electrons are excluded from the study.

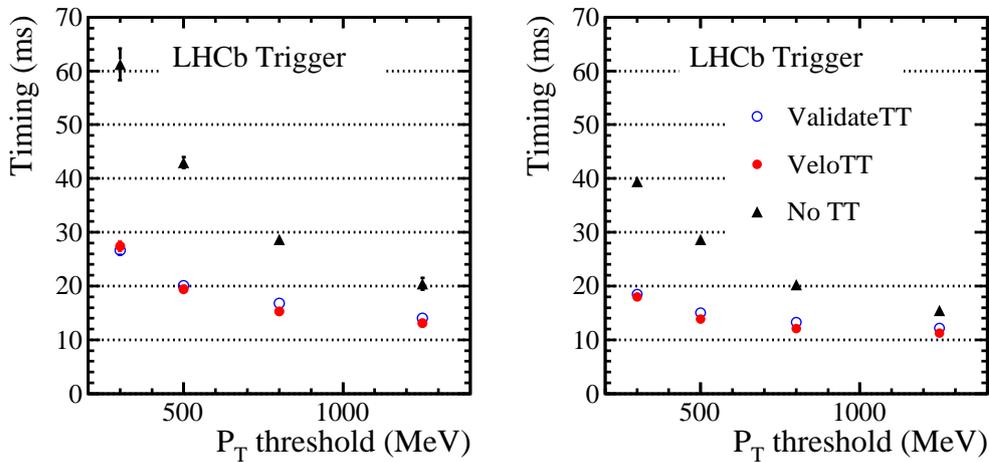
the tracking performances. Figure 1 shows a performance comparison between the 2015 track-finding sequence and the 2012 tracking, evaluated on Monte Carlo simulated events. One major improvement will consist in using the TT information for pattern recognition and momentum estimate [8] (VeloTT option in Fig. 2), instead of using TT-hits only for seed confirmation purposes (ValidateTT in Fig. 2). For a small penalty in timing, this results in higher tracking efficiencies at low momenta and in lower ghost rates (as illustrated in Fig. 2), and provides a momentum estimate also for the tracks that are not propagated through the T-stations. In turn, the lower ghost rate and the higher available CPU budget will allow the VELO multiplicity requirements on the tracks to be removed: in fact, although these cuts help in reducing ghost rate and reduce HLT timing, they are very penalizing in terms of efficiency (Fig. 1).

Monte Carlo simulation is found to severely underestimate absolute timing values, by factors comprised between 1.2 and 1.6 for VELO tracking, and between 1.6 and 1.8 for pattern recognition in the downstream tracking stations (here the lower values are associated to events not undergoing L0 selection, while the higher values correspond to L0-selected events). Therefore, the timing of each tracking sequence option has been evaluated on 2012 L0-selected data for a typical fully-loaded node of the Event Filter Farm. The values found in 2012 data, reported in Fig. 3, must then be extrapolated to 2015 conditions by correcting for the timing ratio between 2015 and 2012 Monte Carlo simulated events: for a luminosity of  $4 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$ , this ratio is evaluated to be 1.3, accounting for the increase in energy and decrease in pile-up in RunII. A comparison of the expected values with the available timing budget for the HLT1 sequence (estimated to be 51 ms after taking into account the bonus from deferral), suggests that the CPU resources allocated for the High Level Trigger allow to adopt  $p_T$  track-finding thresholds as low as 500 MeV in HLT1 for all VELO seeds.

Lowering the tracking  $p_T$  thresholds is expected to improve the HLT1 performances for charm. In addition, a more efficient tracking is a prerequisite to adopting more refined rate-reduction strate-



**Figure 2:** Tracking performances for three options in the usage of the TT-stations in the pattern recognition sequence, on Monte Carlo simulated events at 13TeV: TT-stations not considered for the pattern recognition (NoTT, triangles); validation of VELO seeds with TT-stations (ValidateTT, open circles); momentum estimate with the TT-stations (VeloTT, full circles). Left: efficiency as a function of generated particle  $p_T$ . Right: ghost rate as a function of generated particle  $p_T$ . The background histograms display the complementary cumulative distribution of generated  $p_T$ . Electrons are excluded from the study.



**Figure 3:** Timing of the pattern recognition sequence (excluding the track fit) for three options in the usage of the TT-stations in the pattern recognition sequence, on 2012 L0-selected data: TT-stations not considered for the pattern recognition (NoTT, triangles); validation of VELO seeds with TT-stations (ValidateTT, open circles); momentum estimate with the TT-stations (VeloTT, full circles). Left: no IP-selection of the VELO seeds. Right: VELO seeds required to have IP  $> 100\mu\text{m}$ . Timing values have been obtained on a fully loaded Event Filter Farm node (24 processes) and are not rescaled for the 2015/2012 correction factor.

Channel	$b\bar{b}$	$c\bar{c}$	$\mu/\mu\mu$
	$B^0 \rightarrow K^+\pi^-$	$D^0 \rightarrow K^-\pi^+$	$B^+ \rightarrow J/\psi K^+$
$\epsilon^{TOS}$ (L0) (%)	40	27	89
$\epsilon^{TOS}$ (HLT1   L0) (%)	86	67	90
$\epsilon^{TOS}$ (HLT2   (HLT1 $\wedge$ L0)) (%)	78	90	87
Total $\epsilon^{TOS}$ (%)	27	16	70

**Table 4:** Trigger selection efficiencies ( $\epsilon^{TOS}$ ) on a set of representative channels. The offline-reconstructed candidates considered for the numerator of the efficiency are only the ones that have a correspondance with the accepted trigger candidate [6].

gies in HLT1, based on the reconstruction of multiple tracks and on the topology of the reconstructed vertices.

## 7. Summary

The LHCb trigger performed successfully in 2012, doubling the instantaneous luminosity with respect to design while achieving high trigger efficiencies: modes with muon pairs in the final state have been selected with efficiencies close to 80%, while hadronic modes of beauty hadrons exhibit efficiencies larger than 60% (Tab. 4).

During RunII, starting in 2015, the trigger will face new challenges, and changes in the online infrastructure will be adopted to cope with them, such as a more extensive and efficient use of the deferral. Improvements in the tracking sequence will allow to adopt multi-track trigger selections already in the first stage of the High Level Trigger. At present, the main limitation is represented by the 1.1 MHz maximum full-detector readout rate: this limitation will be only removed with the LHCb upgrade after RunII, when a full 40MHz readout will be implemented [9, 10].

## References

- [1] A. A. Alves Jr. *et al.*, *The LHCb Detector at the LHC*, JINST **3**, S08005 (2008).
- [2] R. Aaij *et al.*, *Measurement of  $\sigma(pp \rightarrow b\bar{b}X)$  at  $\sqrt{s} = 7\text{ TeV}$  in the forward region*, Phys. Lett. **B 694**, 209 (2010).
- [3] R. Aaij *et al.*, *Prompt charm production in pp collisions at  $\sqrt{s} = 7\text{ TeV}$* , Nucl. Phys. **B 871**, 1 (2013).
- [4] R. Cornat *et al.*, *Level-0 decision unit for LHCb*, LHCb 2003-065 (2003).
- [5] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8**, P04022 (2013).
- [6] J. Albrecht *et al.*, *Performance of the LHCb High Level Trigger in 2012*, J. Phys.: Conf. Ser. **513**, 012001 (2014).
- [7] V. V. Gligorov and M. Williams, *Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree*, JINST **8**, P02013 (2013).
- [8] E. Bowen and B. Storaci, *VeloUT tracking for the LHCb Upgrade*, LHCb-PUB-2013-023 (2103).
- [9] I. Bediaga *et al.*, *Framework TDR for the LHCb Upgrade*, CERN-LHCC-2012-007 (2012).
- [10] I. Bediaga *et al.*, *LHCb Trigger and Online Upgrade Technical Design Report*, CERN-LHCC-2014-016 (2014).