Optimization of the LHCb track reconstruction

Barbara Storaci on behalf of the LHCb collaboration
University of Zurich, Physik-Institut, Winterthurerstrasse 190, Zurich, Switzerland
E-mail: Barbara.Storaci@cern.ch

Abstract.
The LHCb track reconstruction uses sophisticated pattern recognition algorithms to reconstruct trajectories of charged particles. Their main feature is the use of a Hough-transform like approach to connect track segments from different sub-detectors, allowing for having no tracking stations in the magnet of LHCb. While yielding a high efficiency, the track reconstruction is a major contributor to the overall timing budget of the software trigger of LHCb, and will continue to be so in the light of the higher track multiplicity expected from Run II of the LHC. In view of this fact, key parts of the pattern recognition have been revised and redesigned. In this document the main features which were studied are presented. A staged approach strategy for the track reconstruction in the software trigger was investigated: it allows unifying complementary sets of tracks coming from the different stages of the high level trigger, resulting in a more flexible trigger strategy and a better overlap between online and offline reconstructed tracks. Furthermore the use of parallelism was investigated, using SIMD instructions for time-critical parts of the software.

1. Introduction
The LHCb detector, discussed in detail in Ref. [1], is a single arm forward spectrometer that covers a pseudo-rapidity range of $2 < \eta < 5$, with the primary purpose of performing precision measurements in the search for New Physics in CP violation and rare decays of beauty and charm hadrons. Essential requirements to perform such a physics program are: excellent tracking (momentum, impact parameter and primary vertex resolution), precise decay time resolution and exceptional particle identification. During the data taking period between 2010 and 2012, called Run I, the LHCb experiment achieved extraordinary performance while facing operation conditions above the design values [2]. Even more challenging conditions are expected for the next run, called Run II, as shown in Table 1. To be ready to achieve the desirable performance in this new scenario, detailed studies were carried out to optimize the reconstruction chain, as presented in this proceeding. Run II represents also an interesting test bench for new concepts devised for the upgrade of the LHCb detector, as for example the fact that, in the upgrade scenario, a full software trigger is chosen [3]. This choice will allow to do a complete physics analysis based directly on data produced by the High Level Trigger (HLT). More precisely, in the “turbo stream” the trigger will write out a compact summary of “physics” objects containing all information necessary for analyses, and this will allow an increased output rate boosting all the physics, like charm physics, which are limited by trigger output rate constraints [4]. Such a strategy imposed to have the same reconstruction chain offline and online, with strong requirements on the reconstruction time and on the performance online. Using Run II as proof of concept, the optimization studies tried to improve the time of the reconstruction chain in...
order to meet the stringent time requirement present at HLT1 level, as well as pushing the
online performance to the excellence obtained normally offline. An other important ingredient,
as presented in [5], is the possibility to run a real-time detector alignment and calibration which
reaches equivalent performance in the online and offline reconstruction.

### Table 1.
Comparison of some of the key elements in the conditions faced in Run I and expected
in Run II at LHCb. The time budget for the HLT steps were estimated in simulation to safely
use only a quarter of the disk buffer available, see Figure 3.

| Conditions       | Run I           | Run II          |
|------------------|-----------------|-----------------|
| $\sqrt{s}$      | 7 – 8 TeV       | 13 TeV          |
| Bunch spacing    | 50 ns           | 25 ns           |
| Pile-up          | 1.7             | 1.3 – 2.4       |
| Output rate HLT1 | 80kHz           | 150kHz          |
| Output rate HLT2 | 5kHz            | 12.5 kHz        |
| Time budget HLT1 | 20ms/evt        | 35ms/evt        |
| Time budget HLT2 | 150ms/evt       | 350ms/evt       |

2. Track reconstruction
The geometry of LHCb allows the definition of several different track types depending on the
sub-detectors in which the tracks have measurements. The LHCb tracking system consists of
a silicon-strip vertex detector (VELO) surrounding the pp interaction region [6], a large-area
silicon-strip detector (TT) located upstream of a dipole magnet with a bending power of about
4 Tm, and three stations (T-stations) of silicon-strip detectors (IT) and straw drift tubes (OT)
[7] placed downstream of the magnet. A schematic diagram of the LHCb tracking system along
with each of the track types is shown in Figure 1. The following definitions are used:

- **Long tracks**: traverse full tracking system from VELO to the T-stations. As they have the
  most accurate momentum estimate, they are often the most useful for physics analysis.
- **VELO tracks**: have hits in both the R- and $\Phi$-sensors of the VELO but are not matched to
  hits in other sub-detectors. They can be at large polar angles or backwards and are used
  for primary vertex reconstruction.
- **Upstream tracks**: have hits in the VELO and TT only. Often low momentum particles
  that are bent out of acceptance by the magnetic field. They can be used also as input for
  reconstructing long tracks.
- **T tracks**: only reconstructed in the T-stations. Can originate from very long-lived particles
  or material interactions.
- **Downstream tracks**: have hits in the TT and T-stations. Allow the reconstruction of charged
daughters of long-lived particles ($K^0_s$, $\Lambda$) with a decay vertex displaced from the interaction
point.

In order to reconstruct the different track types, several tracking algorithms are employed.
There are two stand-alone algorithms, VELO tracking and T-seeding, while the other algorithms
use input from these two algorithms in order to perform a further track reconstruction, see
Figure 2. It should be noted that two complementary algorithms, the forward tracking and the
track matching, reconstruct long tracks, therefore a loss of efficiency in one algorithm can be
compensated by the other, guaranteeing a full final efficiency.
3. Optimization for Run II

To reduce the execution time of the reconstruction chain, two main optimizations were performed: a new sequence of algorithms is proposed, and vectorization in the form of the single instruction, multiple data (SIMD) paradigm is applied in some algorithms. In the reconstruction chain used in Run I, track segments reconstructed in the VELO were passed to the algorithm in charge of finding matching hits in the tracking stations downstream of the LHCb dipole magnet (forward tracking). The time budget for the HLT steps in Run II were estimated in simulation to safely use only a quarter of the disk buffer available\(^1\), see Figure 3, and a time constraint around 13 ms is set as a reasonable benchmark for the track reconstruction in HLT1. With the Run I reconstruction chain it was not possible to achieve such a goal. It was proposed to reduce the execution time by executing an intermediate algorithm which extended the VELO-track segments to the TT stations upstream of the magnet (VELO-TT algorithm). These two algorithms are presented in Section 3.1 and 3.2 respectively and the effect of this modification on the total execution time is discussed in Section 3.3. More details on the others tracking algorithms can be found in [6, 8, 9]. Some details about the effect of the introduction of vectorization in some algorithms is described in Section 3.4.

3.1. Forward tracking

The Forward tracking algorithm [10] is used to find long tracks, see Figure 4. A Hough transform is utilised to associate hits in the T-stations to each VELO-track. The VELO-track is linearly extrapolated to the T-stations and a search window is opened in each x layer. The VELO-track direction and knowledge of the B-field are used to project each selected hit to the z position of a reference plane. Hits from the same particle are expected to be projected to the same x position while random hits should be uniformly distributed. The resulting clusters are fitted and outliers

---

\(^1\) A total disk buffer space of 4000 TB is available.
are removed using a $\chi^2$ criterium. An additional cluster search is used to add stereo hits that are consistent with the $x$-$z$ track. This 3D track is then fitted, outliers are removed and the best track candidate is chosen based on its $\chi^2$/dof.

Figure 4. Schematic description of the forward algorithm.

3.2. VELO-TT algorithm
The VELO-TT tracking algorithm is used to find upstream tracks, see Figure 5. Each VELO-track is linearly extrapolated to the TT. Hits within a window around the extrapolated track are selected. Track candidates are searched for by first forming doublets (two hits in the first TT station but in different layers), and then extending those doublets to the opposite station and searching for compatible hits to form triplets or quadruplets. If no quadruplets are found, the process is repeated in the reverse direction. Each track candidate is fitted and the track parameters, as the $q/p$, are estimated with a $\chi^2$ minimisation. Due to the fringe B-field between the VELO and the TT a momentum estimate of $\delta p/p \sim 15\%$ is possible. The best track candidate is chosen based on the number of TT layers containing measurements and the $\chi^2$ of the fit. Due to overlap between the sensors of the TT, a particle can leave multiple hits in a single layer. These overlap hits are searched for and added to the track candidate.

Figure 5. Schematic description of the VELO-TT algorithm.

3.3. Effect of the introduction of the VELO-TT algorithm
Thanks to the fringe field of the magnet in the region of the TT, the charge sign of the particle can be determined and its momentum measured with a resolution of about 15%. The availability of this extra information has two advantages. Firstly, a selection on the VELO-track segments to be passed to the tracking stations downstream of the magnet can be performed, requiring a minimum momentum or transverse momentum. Secondly, the charge can be used to reduce search windows in the downstream tracking stations. By running the forward algorithm only on VELO-tracks that are upgraded as upstream-tracks and that have a transverse momentum bigger than 400 MeV/$c$, the forward tracking has to process only half of the tracks, with a
reduction on the execution time as well as a reduction in the fake track rate. On the other hand, the geometrical acceptance of the TT detector is such that around 3% of efficiency loss is introduced. Part of these tracks can be retrieved later by the matching algorithm, or by running the forward tracking in the HLT2 step, where more time is available, on the unused VELO-tracks.

A reconstruction chain 3 times faster than the one used in Run I, with the possibility of running an event in 32 ms in HLT1 was obtained. More notably this new sequence allowed to have tracks with a transverse momentum ($p_T$) bigger than 500 MeV/c already at the first step of the high level trigger with no need, for example, a selection on the impact parameter as was required during Run 1\(^2\). Also a reduction of four times of the the fake track rate was possible, achieving at HLT1 an absolute value of 6%, with a single track efficiency of 87% for tracks with a momentum bigger than 3 GeV/c and a transverse momentum larger than 500 MeV/c.

The second step of the high level trigger has a time budget of 350 ms. This allowed to start from HLT1 tracks and run a second iteration of the forward tracking with looser requirements ($p > 0.5$ GeV/c, and $p_T > 80$ MeV/c) on unused VELO-tracks and unused hits in all the tracking systems. Afterwards all the other algorithms (seeding, downstream, and matching) are run on all the hits. An efficiency greater than 90% with only a 12% fake track rate was achieved for a single long track from B daughter spending only a quarter of the total HLT2 budget in the pattern recognition.

3.4. Vectorization

The use of parallelism was investigated for time-consuming parts of the software. A complete redesign of the pattern recognition and tracking algorithms was not possible for the Run II time scale. On the other hand the identification of a number of bottlenecks in the existing codes where it was possible to take advantage of the SIMD paradigm, allowed to achieve a speed up of order of 30% in several algorithms, such as fast track fits, Kalman filter fits and the evaluation of the magnetic field map, see Figure 6.

Figure 6. Graphics on how the execution time is distributed inside the Kalman filter code before and after the vectorization is introduced in the most time-critical parts.

4. Conclusions

A new data taking period is approaching and huge effort has been made to optimize the reconstruction chain to allow an innovative scheme of data processing where some analyses can

---

\(^2\) Note that in Run I at HLT1 level a requirement on the impact parameter: $\text{IP} > 100\mu m$ and $p_T > 1.5$ GeV/c, as well as VELO hits $> 9$, missing VELO hits $< 3$ were applied.
be done directly from the output produced by the high level trigger, as devised for the upgrade scenario of LHCb. To be ready to test this novel approach, detailed studies are performed to minimize the time for track reconstruction while maintaining the performance at the same level expected from an offline process. A new algorithm, VELO-TT, was introduced as input for the main algorithm used for building tracks that are traversing the full tracking system in LHCb. The introduction of this algorithm in the reconstruction chain allowed a speed up of three times, with a reduction in the fake track rate, allowing to run the HLT1 step within 32 ms, i.e. below the expected time budget. Notably tracks with a transverse momentum bigger than 500 MeV/c and a momentum bigger than 3 GeV/c are now available already at HLT1 level. At the same time is was shown that the same performance as offline are possible after the HLT2 step remaining also in this case in the time budged of 350 ms. Further improvements of order of 30% in the execution time of time-critical parts in the tracking reconstruction are achieved by the introduction of vectorization.

References
[1] Alves Jr A A et al. (LHCb collaboration) 2008 JINST 3 S08005
[2] Aaij R et al. (LHCb collaboration) 2015 Int. J. Mod. Phys. A30 1530022
[3] Albrecht J et al. 2014 JINST 9 C10026
[4] Benson S The LHCb turbo stream in this proceeding
[5] Dujany G and Storaci B Real-time alignment and calibration of the LHCb Detector in Run II in this proceeding
[6] Aaij R et al. 2014 JINST 9 P09007
[7] Arink R et al. 2014 JINST 9 P01002 (Preprint 1311.3893)
[8] Callot O 2011 Tech. Rep. CERN-LHCb-PUB-2011-001 CERN Geneva
[9] Callot O and Schiller M 2008 Tech. Rep. CERN-LHCb-2008-042 CERN Geneva
[10] Callot O et al. 2007 Tech. Rep. CERN-LHCb-2007-015