RAPID 2018 Vertexing Challenge

November 10, 2018

1 Introduction and objective

A central challenge when reconstructing events at the LHC is identifying and locating the vertices at which interesting particle interactions happened. These come in two main categories : "primary" vertices where the LHC protons collided, and "secondary" vertices where long-lived particles produced in these proton-proton collisions decayed. Finding all the vertices in an event, and correctly classifying the primary and secondary vertices, is important because it tells us where the different particles originated from. Accurately measuring the positions of these vertices allows for an accurate measurement of the lifetimes of the different particles, which allows us to identify different types of particles and decay processes, as well as being an important quantity in many physics analyses. Finally all this vertex reconstruction must fit within the resource budget of real-time data analysis, since many of the real-time classifiers which decide which events should be saved to permanent storage for further analyses require a good quality secondary vertex.

The objective of this challenge will be to compare the performance of different vertex reconstruction algorithms, and especially to understand the tradeoffs between the accuracy of their reconstruction and their resource cost when used in a real-time context. The LHCb detector, illustrated in Fig. 1 will be used as the test case. The RAMP platform will be used to measure and compare the performance based on predefined criteria, and to find areas in which the different algorithms complement each other. We now describe the geometry of the LHCb detector which is relevant to this challenge, the data which will be provided for the challenge, the scoring criteria, and finally give the performance of LHCb's traditional vertex finding algorithms as an idea of the baseline for this challenge.

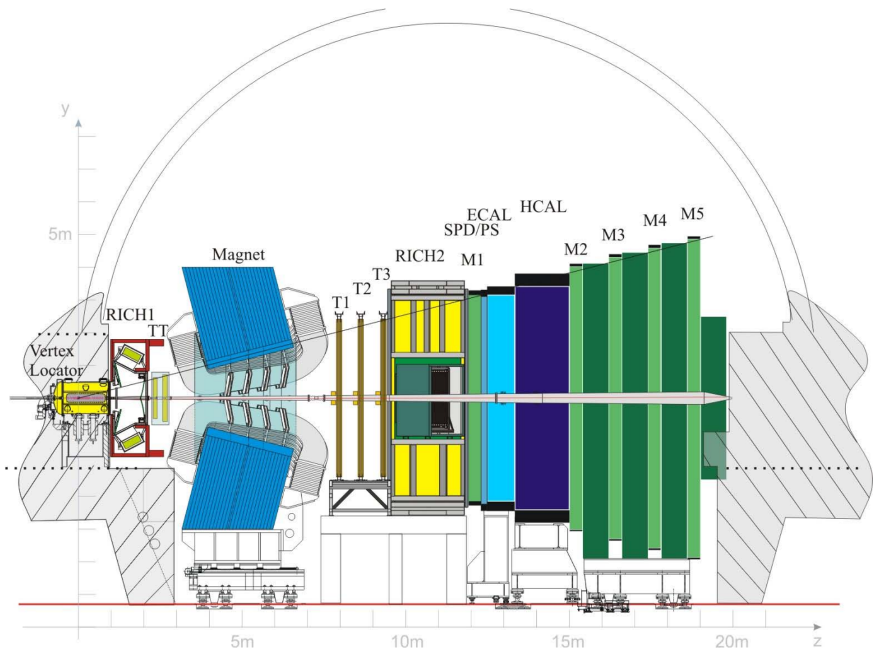

Figure 1: The LHCb detector. The "forward" and "backward" directions respectively refer to particles travelling from the VELO into, or out of, the LHCb acceptance.

2 The LHCb detector geometry

LHCb is a forward spectrometer optimized for the study of particles between around 2 GeV and 200 GeV of energy. It has a 4 Tm dipole magnet which bends charged particles in the horizontal plane, and whose polarity can be reversed to reduce charge asymmetries in the detector reconstruction. LHCb has an outer acceptance of 300 mrad in the bending and 250 mrad in the non-bending plane. The inner acceptance is around 10 mrad for the tracking system, corresponding to a pseudorapidity acceptance of $2 < \eta < 5$. The system for finding charged particle trajectories (tracking) consists of a silicon-strip vertex detector (VELO) with 55 micron square pixels, placed outside the magnetic field so that the tracks are straight lines, which surrounds the LHC interaction region and measures the location of the pp collisions (primary vertices). These positions are then used to calculate the distance of closest approach between tracks and primary vertices, known as their impact parameter. Tracks with small impact parameters are identified as direct products of a pp collision, while tracks with large impact parameters are identified as decay products of particles such as strange, charmed, or beauty hadrons whose lifetimes and boosts were sufficiently large to measurably displace their decay points from the primary vertex. In addition, there are three stations of trackers (T1-T3, collectively the T-stations) after the dipole magnet which use scintillating fibres, and a tracking station placed just before the magnet (UT) which uses silicon-strip sensors. The UT plays a crucial role in correctly matching track segments in the VELO to segments in the T-stations, and extending LHCb's acceptance for long-lived light particles such as K^0_S mesons and Λ baryons. LHCb uses a number of track finding algorithms, which are optimized for tracks originating from different points in the detector acceptance and provide a measure of redundancy. A detailed summary of their performance can be found in a dedicated technical paper [1]. The most important category for the purposes of this challenge is VELO tracks, which are charged particle trajectories reconstructed in the vertex detector, whose geometry is illustrated in Fig. 2.

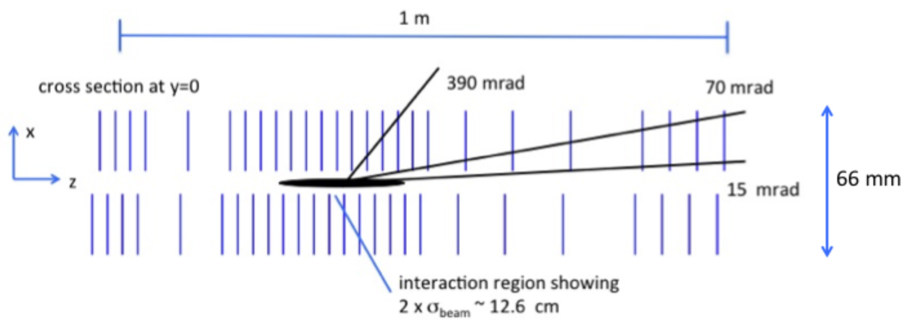

Figure 2: The LHCb VELO detector layout, reproduced from [].

As can be seen, the vertex detector surrounds the interaction region, and the LHCb coordinate system is set up so that primary proton-proton interactions (and hence primary vertices) occur around $\mathbf{z} = 0$. The VELO modules are more finely spaced in this region to give maximal coverage and reduce the distance between the primary vertex and the first measured point on a track. There is no magnetic field in the VELO so all tracks are straight lines. Tracks are reconstructed whether they travel forward towards the rest of the LHCb detector or backwards away from it. The backwards tracks are very important in measuring the position of primary vertices, but only the forward tracks are relevant for measuring the position of secondary vertices.

3 Input dataset and format

The dataset for this challenge will be based on simulated 13 TeV minimum bias events in LHCb upgrade conditions, which means an average of around six proton-proton collisions per event. Both the raw hits in the vertex detector and the reconstructed charged particle trajectories (tracks) will be provided, and teams are welcome to use any mixture of this information in order to identify and estimate the positions of vertices in the event.

The format in which the data is extracted and how it is provided in the challenge is described in this notebook. To summarize, the (x,y,z) coordinates of the hits in the Velo are available, as well as the reconstructed Velo track states at the point closest to the beam line and the corresponding covariance matrices.

For more details, please see the notebook above and the data extraction repository, with which the data is produced.

4 Scoring criteria

The following are the basic performance criteria which could be used to evaluate the performance of a primary vertex finding algorithm

1. The efficiency to correctly reconstruct and identify a primary vertex. The denominator will be the total number of true primary vertices present in the test sample which could have been reconstructed. The criterion that a true primary vertex could have been reconstructed is that it produced at least two (**3, 4, 5? This needs to be optimized**) particles each of which leaves at least three hits in the VELO. The numerator will be the number of reconstructed primary vertices which are associated to true primary vertices. The association criterion will be that the reconstructed primary vertex is closer than a certain minimal distance to a true primary vertex. While not perfect, in particular for overlapping primary vertices, this criterion is much simpler to implement than criteria based on matching hits on particles produced in the primary vertex.

2. The fraction of fake reconstructed primary vertices which are not matched to any true vertex.

3. The **z** position resolution for the primary vertices.

These performance criteria can be combined into many measures of quality, and in particular the "physics" performance criteria matter in different ways for different analyses. In order to simplify things, for this first challenge only the efficiency (true positive rate) ϵ_{rec}, fake (false positive) rate f, and the **x,y,z** resolutions σ_i will be used to score the algorithm, using the formula

$$\text{score} = \frac{\epsilon_{rec} \cdot (1 - f)}{\sigma_x \cdot \sigma_y \cdot \sigma_z} \tag{1}$$

5 Baseline/traditional algorithm performance

References

[1] Roel Aaij et al. Measurement of the track reconstruction efficiency at LHCb. *JINST*, 10(02):P02007, 2015.