

Real-time monitoring of LHCb interaction region with a fast trackless methodology

Giulio Cordova^{1,3,*}, Elena Graverini^{2,3,4}, Federico Lazzari^{2,3}, Michael J. Morello^{1,3}, Daniele Passaro^{1,3}, and Giovanni Punzi^{2,3}

¹Scuola Normale Superiore, Pisa, Italy

²Università di Pisa, Italy

³INFN Sezione di Pisa, Italy

⁴École Polytechnique Fédérale, Lausanne, Switzerland

Abstract. The increasing computing power and bandwidth of FPGAs opens new possibilities in the field of real-time processing of high-energy physics data. The LHCb experiment has implemented a cluster-finder FPGA architecture aimed at reconstructing hits in its innermost silicon-pixel detector on-the-fly during readout. In addition to accelerating the event reconstruction procedure by providing it with higher-level primitives, this system enables further opportunities. LHCb triggerless readout architecture makes these reconstructed hit positions available for every collision, amounting to a flow of 10^{11} hits per second, that can be used for further analysis. In this work, we have implemented a set of programmable counters, counting the hit rate at many locations in the detector volume simultaneously. We use these data to continuously track the motion of the beams overlap region and the relative position of the detector elements, with precisions of $O(\mu\text{m})$ and time granularity of $O(\text{ms})$. We show that this can be achieved by simple linear combination of data, that can be executed in real time with minimal computational effort. This novel approach allows a fast and precise determination of the beamline position without the need to reconstruct more complex quantities like tracks and vertices. We report results obtained with pp collision data collected in 2024 at LHCb.

1 Introduction

The LHCb experiment at the Large Hadron Collider (LHC) is a forward spectrometer designed to investigate CP violation and rare heavy-hadron decays [1].

A significant upgrade of the LHCb detector was prepared in view of the LHC Run 3, capable of supporting an event rate of 30 MHz by means of improved granularity and renewed readout electronics [2]. The innermost tracking detector, used to reconstruct particle collision vertices and named VERtEx LOcator (VELO), has been upgraded from silicon-strip to silicon-pixel sensors [3]. This detector consists of 26 layers transverse to the beam axis; each layer consists of two separable modules that can move horizontally in the transverse plane. These modules can be retracted up to 25 mm from the beam axis for protection during beam injection and brought as close as 5.1 mm to the beam for optimal data collection. Each module is made of four hybrid silicon sensors with $55\ \mu\text{m} \times 55\ \mu\text{m}$ pixels.

*e-mail: giulio.cordova@cern.ch

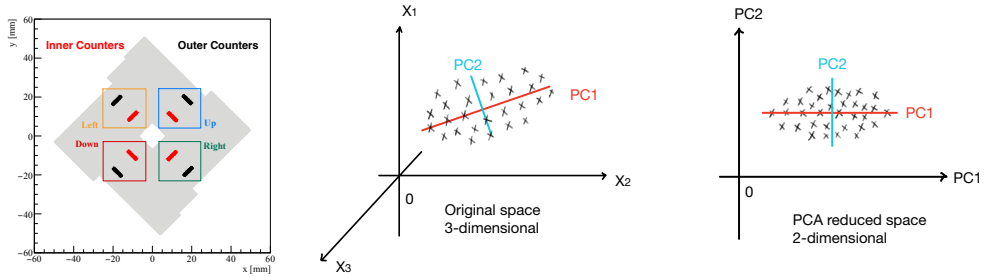


Figure 1. (Left) Position of the cluster counting regions over the VELO sensors. The colour-code highlight the position of the counters. (Right) Schematic of PCA: a 3D dataset is transformed into a 2D space using the first two principal components PC1 and PC2, ordered by variance.

A new triggerless readout architecture is one of the most significant upgrades implemented by the LHCb collaboration, making all detector data read out at every collision event [4, 5]. In each bunch crossing, sub-detector data are time-synchronized and first grouped into event fragments before being assembled for full event reconstruction. The processing pipeline involves two software-based trigger stages (HLT1 and HLT2), where high-level physics objects such as particle tracks and primary vertices are reconstructed in real-time for immediate or offline analysis.

A notable enhancement is the introduction of a two-dimensional clustering algorithm embedded on the readout boards of the VELO [6], which use Intel Arria10 GX 1150 FPGAs [7]. This allows particle hits to be reconstructed at the earliest stage of the data acquisition process. The availability of a rate of $\sim 10^{11}$ hits/s in the readout FPGAs enables potentially interesting measurements to be performed in real-time. In principle, the analysis of high-statistics hit samples could serve various purposes, including detector diagnostics, alignment, luminosity, and the determination of the luminous region position. In this work, we demonstrate reconstruction of the position of the interaction region, as well as the relative positioning of VELO detector elements relative to it. These measurements are obtained based solely on the real-time flow of reconstructed clusters, without relying on particle tracks.

2 Real-time cluster counters on the VELO

The clustering algorithm, which is run in parallel for each VELO half-module, produces a list of x, y coordinates of cluster centroids. The latter is transmitted to the servers where partial event data are gathered together and sent off for processing by the HLT1 routines.

Two geometrical regions are defined on top of each sensor of the VELO, at different radial distances from the beam axis. Clusters recorded within these geometrical regions are accumulated directly in firmware to provide counters that are then used for luminosity and luminous region position analysis [8]. This results in a total of 208 ($2 \times 4 \times 26$) individual counters. The left-hand side of Figure 1 sketches one VELO station, with the position of the counting regions highlighted in red and black for the inner and outer regions, respectively. A naming scheme is introduced to distinguish the four counting regions (Up, Down, Left and Right) within each VELO sensor. Each region is located on a different sensor of the VELO and comprises 20×110 pixels. Cluster counting uses only 0.62% of the logic elements, 2.4% of the M20K memory blocks and 2.8% of the available digital signal processors.

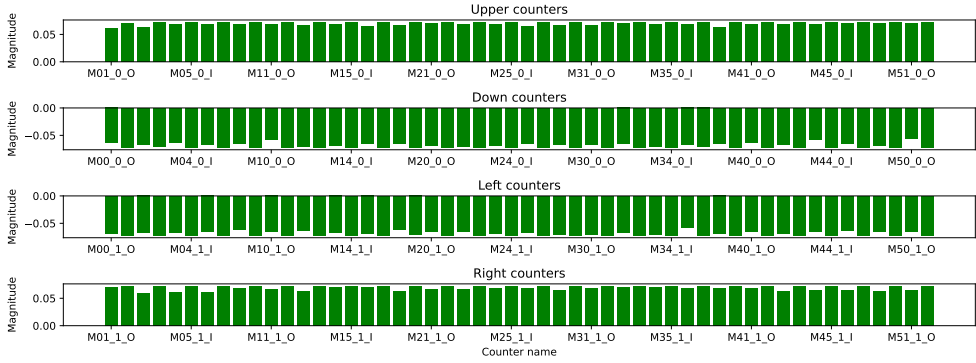


Figure 2. Magnitude of each component of the eigenvectors $\mathbf{w}^{\mathbf{x}}_{(1)}$. The elements relate to the geometrical position of the cluster counters on the VELO in this way: the first panel shows the weights w_U applied to the upper counters c_U , the second panel shows the weights w_L applied to the lower counters c_L , while the third and the fourth shows the weights w_L and w_R applied the left-hand side counters c_L and right-hand side counters c_R , respectively. The different magnitude between the weights of outer and inner counters (even/odd columns), associated with the different acceptance, is evident. This plot also shows that the quantity $\mathbf{t}^{\mathbf{x}}_1$ can be written in a form that highlights the weighted asymmetry up/down and right/left that is computed at each VELO layer i : $\mathbf{t}^{\mathbf{x}}_1 = \sum_i [c_U^i \cdot |w_U^i| - c_D^i \cdot |w_D^i| + c_R^i \cdot |w_R^i| - c_L^i \cdot |w_L^i|]$.

3 Method

Information from all available cluster counters is combined in the experiment slow control software to produce estimators of the position of the interaction region. Several approaches can be considered for this purpose; the method presented in this paper adopts a linear combination of the counters, where the coefficients are obtained from Monte Carlo (MC) simulated data by means of a Principal Component Analysis (PCA).

Principal Component Analysis

Principal Component Analysis (PCA) [9] is a statistical technique for dimensionality reduction, feature extraction, and data compression. It transforms a dataset via an orthogonal linear transformation that maximizes variance along successive Principal Components (PCs). The right-hand panel of Figure 1 provides an intuitive visualization.

Let \mathbf{C} be an $n \times p$ data matrix, where the average of each column is 0. PCA finds a set of l orthonormal weight vectors $\mathbf{w}_{(k)}$, mapping each row $\mathbf{c}_{(i)}$ to ordered PC scores:

$$\mathbf{t}_{\mathbf{k}(i)} = \mathbf{c}_{(i)} \cdot \mathbf{w}_{(k)} \quad \text{for } i = 1, \dots, n \quad k = 1, \dots, l. \quad (1)$$

The first PC represents the direction on which the projected dataset has the greatest variance, leading to the optimization problem:

$$\mathbf{w}_{(1)} = \arg \max_{\|\mathbf{w}\|=1} \left\{ \sum_{i=1}^n (\mathbf{c}_{(i)} \cdot \mathbf{w})^2 \right\} = \arg \max_{\|\mathbf{w}\|=1} \{ \|\mathbf{C}\mathbf{w}\|^2 \} = \arg \max \left\{ \frac{\mathbf{w}^T \mathbf{C}^T \mathbf{C} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \right\}. \quad (2)$$

This maximization corresponds to the largest eigenvalue of $\mathbf{C}^T \mathbf{C}$, with $\mathbf{w}_{(1)}$ as its associated eigenvector [10]. For subsequent components, a residual matrix $\hat{\mathbf{C}}_k$ is defined by removing

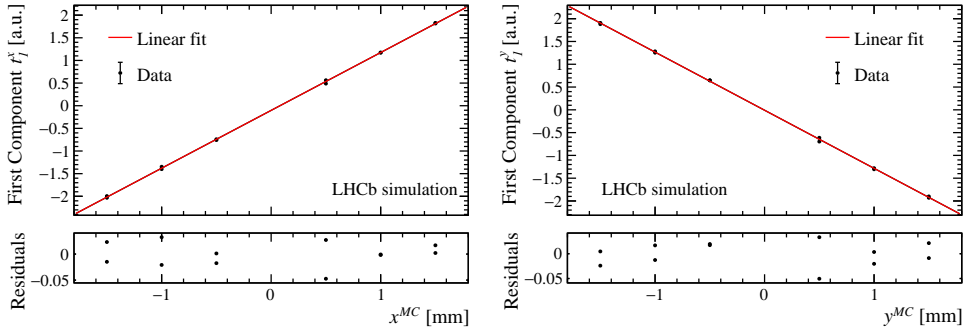


Figure 3. Linearity of the first component calculated with the PCA with respect to luminous region position shifts along the x (left) and y directions (right).

the contributions of the first $k - 1$ components. The problem reduces to an eigenvalue decomposition of $\hat{\mathbf{C}}_k^T \hat{\mathbf{C}}_k$, ensuring orthogonality. The PCA technique, ultimately, diagonalizes the sample covariance matrix:

$$\mathbf{S} = \frac{1}{n-1} \mathbf{C}^T \mathbf{C}, \quad (3)$$

where the eigenvectors of \mathbf{S} are ordered by eigenvalue magnitude, with the largest eigenvalue corresponding to the first PC. The percentage of variance that each PC represents, corresponds to the normalized eigenvalues.

Parameter estimation in Monte Carlo simulation

We construct two estimators \hat{x} and \hat{y} , to track the two transverse components of interaction region position, using a set of l scores $\{t^j\}_{k=1\dots l}$, with $j = x, y$:

$$\hat{x} = \hat{x}(t_1^x, \dots, t_l^x), \quad \hat{y} = \hat{y}(t_1^y, \dots, t_l^y) \quad (4)$$

Each of the t_k scores is described by equation (1), hence let us define $\mathbf{c}_{(i)}$ and $\mathbf{w}_{(k)}$. The $\mathbf{c}_{(i)}$ vector is a p -dimensional vector¹ containing the mean cluster count per event, normalized to the sum of each of the p implemented counters. The $\mathbf{w}_{(k)}$ is the vector of weights to calculate the k -th component with the PCA algorithm. Training MC datasets \mathbf{X}_{train} or \mathbf{Y}_{train} are used to estimate the weights. The MC datasets are generated using the LHCb simulation workflow, where the beam spot is shifted with respect to its nominal position. As described in Section 3, the covariance matrices of these datasets are calculated and diagonalized in order to obtain k p -dimensional weight vectors $\mathbf{w}_{(k)}$, that represent the new basis of the space computed by the PCA. In our setup, more than 90% of the variance of \mathbf{X}_{train} and \mathbf{Y}_{train} is explained by the first PC, while the others give marginal contributions. This behaviour is expected because, in the MC training dataset, the only varying feature (i.e., the main source of variance) is the position of the luminous region. As a consequence, the first PC captures most of the variance and is expected to be directly proportional to the luminous region position, while the remaining components primarily describe statistical fluctuations in the dataset. Therefore, we will only use the first PC for each Cartesian component. A visualisation of the components of the 208-dimensional vector $\mathbf{w}_{(1)}^x$ is depicted in Figure 2.

¹where $p = 208$, i.e. the vector has one entry per counter.

In order to check the relation between \mathbf{t}_1^j –the scores of the first PC– and the luminous region position, these scores are calculated using rates $\mathbf{c}_{(i)}$ from test MC datasets. The $\mathbf{w}_{(1)}^j$ weights are estimated from the training datasets. Two different datasets are used in order to prevent bias. Being the test datasets completely independent from the training datasets, the validity of this method is assessed. In fact, when operating on real collision data, the scores \mathbf{t}_1^j are obtained in real time from the cluster counters $\mathbf{c}_{(i)}$, and the vectors $\mathbf{w}_{(1)}^j$ are pre-calculated using the train MC. The relation between the first PC and the luminous region position can be assessed in the test datasets as depicted on the left and right-hand side of Figure 3 for the x , and y components, respectively.

The approximation of linearity is only valid for x and y , because in this case the shifts of the luminous region are small with respect to the detector size. For the the z direction –where the observed range is comparable to the size of the detector– a cubic relation with \mathbf{t}_1^z was observed. Determination of the z component is however of lesser importance due to wide spread of the interaction region in this coordinate, and will not be given further attention in the present work.

Calibration on real data

Equation (4) can be rewritten explicitly to depend only on the first PC \mathbf{t}_1^j :

$$\hat{x} = \hat{x}(t_1^x) = \alpha_x t_1^x + \beta_x; \quad \hat{y} = \hat{y}(t_1^y) = \alpha_y t_1^y + \beta_y. \quad (5)$$

The coefficients α_j and β_j (with $j = x, y$) are obtained from a fit to calibration data, minimising the residuals of the cluster-counter based position estimators \hat{x} , \hat{y} with respect to the position readings provided by the VELO Closing Monitoring Tasks (VCMT). These positions are obtained from distributions of PVs calculated using VELO tracks sampled every few milliseconds.

The length-scale calibration (LSC) phase of a short van der Meer (vdM) [11] scan performed on April 6th, 2024 is used as calibration dataset. During a LSC, the beams are shifted head-to-head in order to calibrate the size of the vdM displacement steps. For the purpose of this study, the LSC is useful because the interaction region is shifted and the estimator presented in this paper can be calibrated and tested. The left and right panels of Figure 4 show the calibration lines of the x and y estimators, respectively.

On real collision data, we have to account for the possibility of missing counters or outliers, which could affect the computation of t_1^j , since this requires exactly p components. Let us define $m^r(\tau)$ as the median value of the counters $c_i^r(\tau)$ at a specific time τ , where the superscript r denotes either inner or outer counters. If a counter value $c_i^r(\tau)$ falls outside the range $m^r(\tau) \pm 50\%$, it is marked as an outlier and replaced with the median value $m^r(\tau)$ when computing t_1^j . This procedure ensures the robustness of the method under experimental conditions where some counters may be missing or affected by detector or data acquisition glitches.

4 Results on pp collision data

The left side of Figure 5 compares the estimated x position versus the input from the VCMT, for a period of 30 minutes following the calibration. An identity line is overlaid to the scatter plot; the associated residuals are shown on the right-hand side of Figure 5. The same plots for the y position are shown in Figure 6. The estimated statistical resolution of $4 \mu\text{m}$ is achieved accumulating counters every 90 ms. In nominal proton-proton running conditions, the same amount of counts used to compute these estimates can be achieved in just 1 ms, due to the different number of colliding bunches and different pile-up.

In Figure 7 we show the time evolution of the luminous region x and y positions as estimated by the counters presented in this work, as well as by the VCMT.

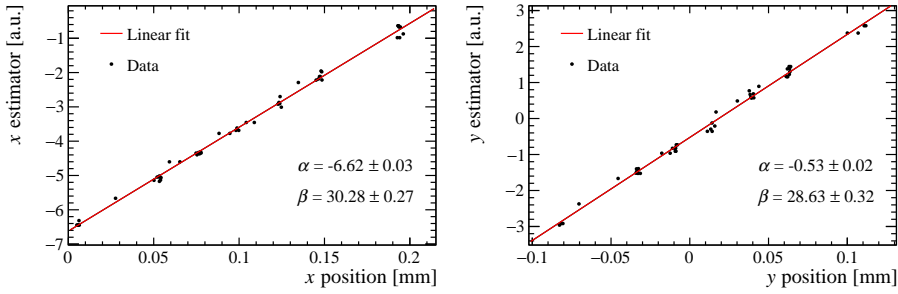


Figure 4. Calibration line for the estimator of the interaction region \hat{x} position (Left) and \hat{y} (Right). The position shifts due to the LSC steps performed during the short vdM scan of Fill 9475 of April, 6th 2024. The y -axis represents the score t_j^i for $j = x, y$, while the x -axis shows the position measured by the VCMT. A linear fit is performed to estimate α and β of Equation (1).

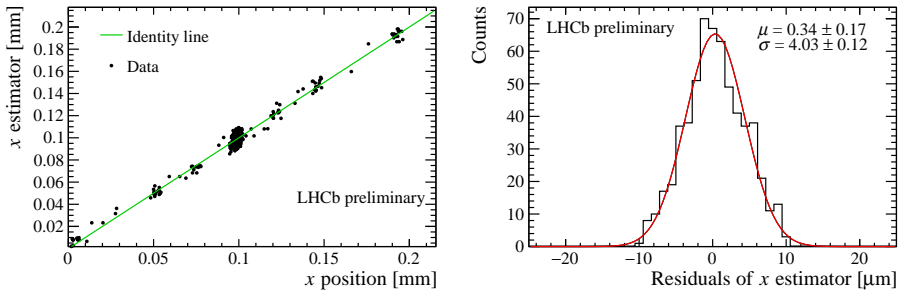


Figure 5. On the left-hand panel, a comparison of the calibrated x estimator [y -axis] with the position measured by the VCMT [x -axis] is shown. The plotted data correspond to a 30-minute period following the first LSC of Fill 9475, of April 6th, 2024. The green line represents the identity line. On the right-hand panel, the residuals are shown, fitted with a Gaussian distribution. The mean and standard deviation of the Gaussian indicate the bias and resolution of the estimator on a test data sample that was not used for calibration.

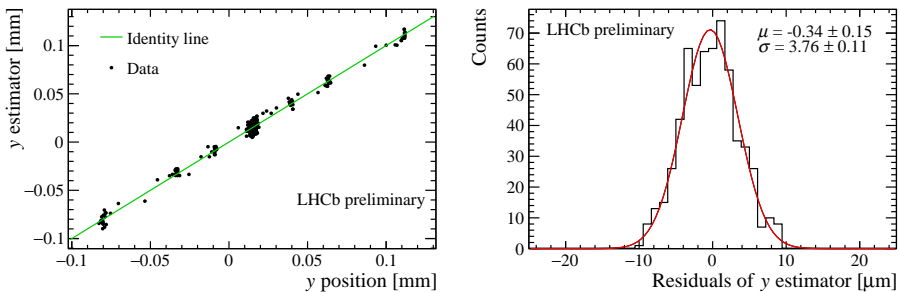


Figure 6. Analogous of Figure 5 for the y estimator.

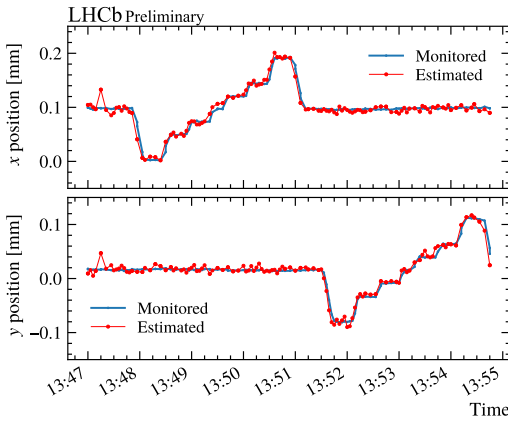


Figure 7. Position of the x and y estimators (red) as a function of time, overlaid to the transverse beamline position provided by the VCMT (blue). The time period corresponds to a second LSC scan performed during Fill 9475 [April, 6th 2024], 30 minutes after the one used for calibration.

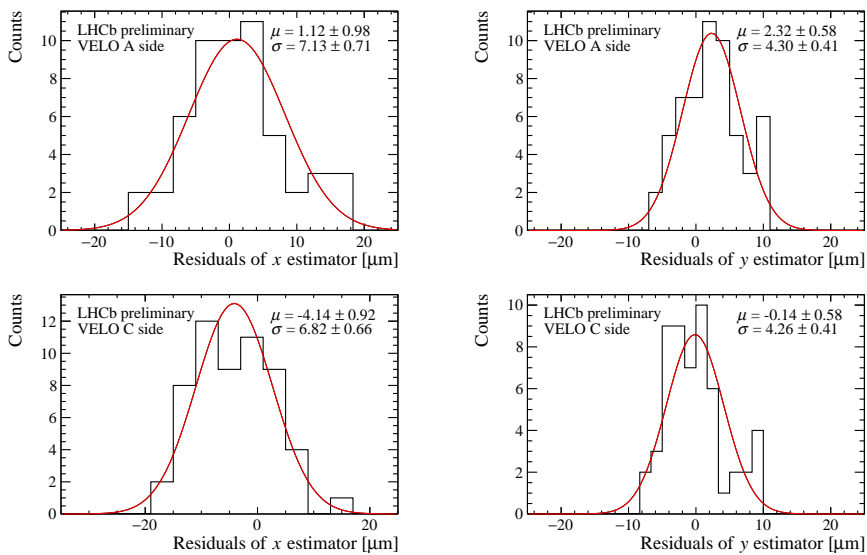


Figure 8. Residuals of estimators of the position of the VELO halves with respect to the luminous region position. (Top Left) position x of VELO A. (Top Right) position y of VELO A. (Bottom Left) position x of VELO C. (Bottom Right) position y of VELO C.

Results obtained using only one half of the VELO

Alternatively, the position of the interaction region can also be estimated using only one side of the VELO. The procedure is the same as in Equation (1), with the difference that $\mathbf{c}_{(i)}$ refers only to the counters placed on one half of the VELO, and the weights \mathbf{w}_1^i are correspondingly recalculated on MC. The calibration of the VELO A-side and C-side position estimators follows the same procedure described above.

These estimators provide insight about the relative position of either VELO half with respect to the interaction region, or with respect to one another. The VELO-half position estimates are compared with the VCMT positions, throughout a period of a few minutes following the LSC scan used for calibration. Figure 8 shows the bias and resolution of these estimates for the x , y directions and for both sides of the VELO. Both y estimators have a

statistical resolution of $4\ \mu\text{m}$, while the x estimators have a statistical resolution of $7\ \mu\text{m}$. The reason for the different resolution is still under study.

5 Conclusions and future prospects

This work focused on an application of the implementation of on-the-fly hit counters in specific VELO regions that provides real-time information embedded in the readout at the 30 MHz collision rate. By combining these counters, we achieve a real-time trackless monitor of the luminous region position with an extrapolated statistical precision of $4\ \mu\text{m}$ using an amount of data achievable every 1 ms in nominal pp running conditions. It is also possible to restrict the measurement to cluster counters from one half of the VELO. This way, we achieve a resolution of $7\ \mu\text{m}$ on the position of the luminous region; this measurement can however be interpreted also as a measurement of the VELO half position with respect to the luminous region, which could in principle prove useful for detector monitoring.

These results can be further improved by implementing a measurement of the z component of the luminous region position, as well as a measurement of the luminous region shape and inclination in the horizontal and vertical planes. This study highlights the potential of real-time FPGA-based data processing, showing that it can provide reconstruction-level quantities at a very early stage in the data acquisition chain.

References

- [1] LHCb Collaboration, The LHCb Detector at the LHC, JINST **3**, S08005 (2008). [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005)
- [2] LHCb Collaboration, The LHCb Upgrade I, JINST **19**, P05065 (2024), 2305.10515. [10.1088/1748-0221/19/05/P05065](https://doi.org/10.1088/1748-0221/19/05/P05065)
- [3] LHCb Collaboration, LHCb VELO Upgrade Technical Design Report, Tech. rep., CERN (2013), <https://cds.cern.ch/record/1624070>
- [4] LHCb Collaboration, Computing Model of the Upgrade LHCb experiment, Tech. rep., CERN (2018), <https://cds.cern.ch/record/2319756>
- [5] LHCb Collaboration, Upgrade Software and Computing, Tech. rep., CERN (2018), <https://cds.cern.ch/record/2310827>
- [6] G. Bassi, L. Giambastiani, K. Hennessy, F. Lazzari, M.J. Morello, T. Pajero, A.F. Prieto, G. Punzi, A FPGA-Based Architecture for Real-Time Cluster Finding in the LHCb Silicon Pixel Detector, IEEE Transactions on Nuclear Science **70**, 1189–1201 (2023). [10.1109/tns.2023.3273600](https://doi.org/10.1109/tns.2023.3273600)
- [7] Intel® Arria® 10 Device Datasheet, Intel Corporation (2022), available at www.intel.com/content/www/us/en/docs/programmable/683771
- [8] LHCb Collaboration, Calibration of online luminosity measurement using VELO RetinaCluster counters at LHCb (2024), <https://cds.cern.ch/record/2904669>
- [9] K. Pearson, Liii. on lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **2**, 559 (1901), <https://doi.org/10.1080/14786440109462720>. [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720)
- [10] R.A. Horn, C.R. Johnson, Matrix Analysis, 2nd edn. (Cambridge University Press, Cambridge; New York, 2013), ISBN 9780521839402, <https://books.google.it/books?id=07sgAwAAQBAJ>
- [11] V. Balagura, Van der Meer scan luminosity measurement and beam–beam correction, Eur. Phys. J. C **81**, 26 (2021), 2012.07752. [10.1140/epjc/s10052-021-08837-y](https://doi.org/10.1140/epjc/s10052-021-08837-y)