UNIVERSIDAD NACIONAL DE COLOMBIA

Improvement on the Track Reconstruction Algorithms at the LHCb Experiment

Jeysson Alexander Riaño Pascagaza

Universidad Nacional de Colombia
Science Faculty, Physics Department
Bogotá, Colombia
2019

Improvement on the Track Reconstruction Algorithms at the LHCb Experiment

Jeysson Alexander Riaño Pascagaza

Thesis submitted to get the degree of :
Master of Sciences-Physics

Advisor:
Ph.D. Diego Alejandro Milanés Carreño

Research field:
Experimental particle physics
Collaboration Group:
Grupo de particulas Fenyx

Universidad Nacional de Colombia
Science Faculty, Physics Department
Bogotá, Colombia
2019

Everyone knows that debugging is twice as hard as writing a program in the first place. So if you are as clever as you can be when you write it, how will you ever debug it?

Brian Kernighan

Abstract

In this MSc. Thesis work, the performance of the ghost categorization algorithms at the LHCb experiment were studied, using the $B_s^0 \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$, $B^0 \rightarrow J/\psi(\mu^+\mu^-)$ $K_s^0(\pi^+\pi^-)$, $B^0 \rightarrow \mu^+\mu^-K^*(K^+\pi^-)$ and $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$ simulated data samples for the years 2011 and 2012. Since, ghost tracks which correspond to reconstructed tracks that were not simulated, affect directly the error of the experimental measurements. Several studies were made in order to have better reconstruction algorithms. The first simulated data samples were used to check the differences between all the variables involved in the reconstruction process. An adaptive binning was done with the information of the χ^2_{ndof} and the transverse momentum ($\mathrm{P_T}$), the pseudorapidity (η) and the number of tracks per event. A selection criterion in the χ^2_{ndof} variable discards ghost tracks, however, some signal is also lost. Furthermore, two kaon tracks, coming from the same ϕ meson, were analyzed to discriminate the differences between the meson containing one or two ghost tracks. If one of the kaons is a ghost, a peak in the mass distribution was observed, meanwhile, a flat distribution was obtained for two ghost kaons. The compatibility of this peaking distribution with the real signal was tested using a multivariated selection which gave us as a result that only a few ghost events are compatible with the signal. Finally, the D^{*+} sample for the year 2012 was used to test the modification of the linking fraction between the reconstructed and the generated particle. Decreasing this fraction results in a reduction of ghost tracks and therefor an increase in the signal yields, although the background increases as well. Two scenarios are discussed maximizing the signal gain and minimizing the background increase, and a proposal to improve the linking is presented.

Keywords: ghost track, signal, background, migration

Resumen

En este trabajo de tesis de maestría el desempeño de los algoritmos de categorización de trazas fantasma fue estudiado, empleando para ello muestras de datos simulados de las desintegraciones de los mesones: $B_s^0 \rightarrow J/\psi(\mu^+\mu^-)\phi(K^+K^-)$, $B^0 \rightarrow J/\psi(\mu^+\mu^-)$ $K_s^0(\pi^+\pi^-)$, $B^0 \rightarrow \mu^+\mu^-K^*(K^+\pi^-)$ y $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$, para los años 2011 y 2012. Ya que las trazas fantasma son aquellas trazas reconstruidas las cuales no fueron generadas y estas afectan directamente el error de las medidas experimentales. Varios estudios fueron realizados para implementar mejoras en los algoritmos de reconstrucción. Las primeras muestras de datos simulados fueron empleadas en la búsqueda de diferencias entre las distribuciones de todas las variables cinemáticas involucradas en el proceso de reconstrucción. Un estudio de paso adaptativo fue llevado a cabo con la información del χ^2_{ndof} y el momento transversal (P_T), la pseudorapidez (η) y el número de trazas por evento. Un criterio de selección en la variable χ^2_{ndof} fue empleado para descartar trazas fantasmas sin embargo éste también descarta eventos de señal. Además, dos trazas de kaones, proveninetes del mismo meson $\phi \rightarrow K^+K^-$, fueron analizadas para observar las diferencias entre el meson que contiene una o dos trazas fantasma. Un pico en la distribución de masa fue observado cuando tan solo uno de los kaones es fantasma, en caso contrario se obtiene una distribución plana. La compatibilidad de los eventos en el pico de la distribución eventos fue comparada con la distribución de eventos de señal arrojando como resultado que sólo unos pocos eventos fantasmas son compatibles con la señal. Finalmente la muestra de datos para el D^{*+} del año 2012 fue empleada para testear una modificación en la fracción de relación entre las partículas reconstruidas con las partículas generadas. Imponiendo una condición menos severa sobre esta fracción la cantidad de trazas fantasma reconstruidas se reduce y adicionalmente la cantidad de trazas bien reconstruidas aumenta. Sin embargo, hay también un aumento de la cantidad de fondo reconstruidos. En este trabajo dos escenarios son discutidos, minimizando la cantidad de fondo y maximizando la cantidad de señal reconstruidos para los cuales la propuesta de ganar eventos de señal es cumplida.

Palabras clave: traza fantasma, señal, fondo

Content

List of Figures

List of Tables

1. Introduction

The tracks of charged particles coming from heavy hadron decays in a proton-proton collision (pp) at the Large Hadron Collider (LHC) experiment allow us to correctly identify the identity of the heavy hadron that generates them. This information should be verified by Monte Carlo (MC) simulations. The Large Hadron Collider beauty (LHCb) experiment [1] has a specialized software to perform this task. Those algorithms show us the differences between generated and reconstructed tracks of each charged particle originated from hadrons containing heavy quarks.

The LHCb detector is a one arm spectrometer which covers the pseudo-rapidity range between $2 < \eta < 5$. This was designed specifically to identify particles with heavy quarks (charm and bottom). The election of the detector shape is justified, since at high energies the b and $\bar{b}$ hadron decays are predominantly produced forward and backward[1] [2] of the beam axis. The detector is composed by tracking and PID (particle identification) systems. The tracking system is subdivided in (VErtex LOcator) VELO, (Tracker Turicensis) TT, a dipole superconductor magnet and three tracking stations T1, T2 and T3. On the other hand, the PID system is subdivided into two ring imaging Cherenkov detectors, RICH1 and RICH2, an electromagnetic calorimeter (ECAL), and an hadronic calorimeter (HCAL). Moreover, there are five muon stations located at the end of the detector. The detector is well described in Chapter 3. This design minimizes the amount of material used to build the detector. Although it allows high track reconstruction efficiencies, even for low momentum tracks, the limited number of tracking stations over the long distance generate a non negligible amount of mis-reconstructed tracks [3]. The performance of the detector is affected by multiple scattering which decreases efficiency and momentum resolution. Hence, checking the right performance of the LHCb detector becomes a crucial task. It is important to inspect the algorithms thoroughly in order to ve-

Background Category	Code
Undefined	-1
Signal	**0**
QuasiSignal	10
FullyRecoPhysBkg	20
Reflection	30
PartRecoPhysBkg	40
LowMassBkg	50
Ghost	**60**
Clone	63
Hierarchy	66
FromPV	70
AllFromSamePV	80
FromDifferentPV	100
bbar	110
ccbar	120
uds	130
LastGlobal	1000

Table 1-1.: Background Categories of particles at LHCb

---

[1] The LHCb detector covers the forward region

rify the confidence of the reconstructed data. MC simulations help us to determine differences between generated and reconstructed tracks. Among the algorithms LHCb has designed to cope with this labor, one of the most important packages is the MC background info which reports a background category for each reconstructed particle [4]. All background categories are reported in the Table. **1-1**, they are described in more detail in Chapter. 5.

The analysis in the LHCb experiment are carried out based on charged particles that come from the primary interaction vertex. Therefore, it is necessary to define some terminology used within this document.

Prompt particle: Prompt particles are defined as all particle coming from the primary interaction vertex or from a decay chain in which the sum of mean lifetimes does not exceed 10 ps.

Charged particles: All quasi-stable particles carrying electric charge are defined as charged particles [3] such as electrons, muons, charged pions, charged kaons, protons and their anti-particles.

Truth Matched tracks: Truth Matched or signal tracks are all successfully reconstructed particle which shares at least 70 % of its interactions with the detector (hits) with a simulated particle. These kind of tracks should satisfy several conditions [4, 5]:

1. All MC-truth particles matched to final state daughters of candidate have the same mother.
2. All MC-truth particles originating from this mother are matched to final state daughters of the candidate.
3. All MC-truth particles matched to final state particles of the candidate are correctly identify.
4. The common MC-truth mother has the same PID[2] as the candidate.

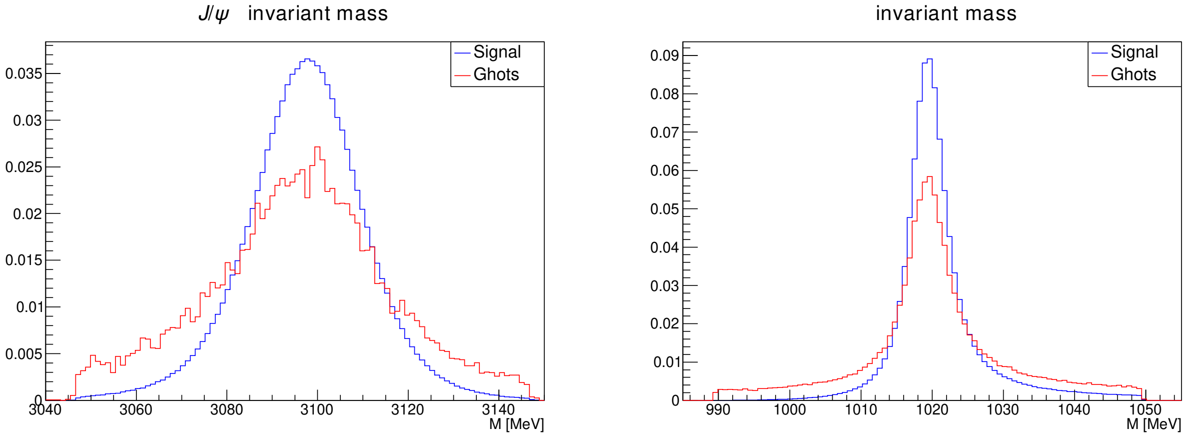
Ghost tracks: Ghost tracks are the most important kind of tracks in the scope of this work. If a track cannot be matched to a simulated particle, it is identified as ghost track. Ghost tracks become more relevant depending on the data sample, the analysis , etc. These can be up to 30 % out of the total sample, in some cases.

---

[2] PID is a system in which each particle has a code assigned, Ex. $e^- \rightarrow 11$, this is done to ease the interaction among the event generation, reconstruction and analysis software. [6]

Clone tracks: If two tacks share more than 70 % of their hits, the worst reconstructed one is classified as a clone track. Since they are matched to the same simulated particle.

The charged particles traversing the LHCb tracking system can be reconstructed if the algorithms successfully store a minimum of the hits [3]. There are several track types at LHCb: VELO-, T-, Long-, Upstream- and Downstream-track; these are fully described in Chapter. 4.4. For each track type, a different pattern recognition algorithm was created to reconstruct as much tracks as possible. The description of the reconstruction algorithms is given in Chapter. 5.



(a)

(b)

Figure 1-1.: J/ψ (a) and ϕ (b) reconstructed invariant mass distributions for signal and ghost candidates

In the track reconstruction, ghost rate affects experimental quantities that depend on the MC samples such as efficiencies and resolution. LHCb experiment has designed several ways and algorithms to cope with the problem. However these are not perfect. There are some failures discussed in Chapter 6. Within these flaws, one that calls the attention is the behavior shown in Figure **1-1** which is not expected. For ghost tracks there is a peak on the mass of the particle, which is unexpected since the distribution must be flat as these events do not correspond to real tracks. This suggest problems in the reconstruction algorithms. Taking into account the cost of production of MC samples (timing, storage and computing) and the fact that the events categorized as ghost events cannot be used in posterior analysis, modifications to the reconstruction algorithms need to be studied in order to find a better categorization that allows not only the migration from ghost to signal candidates, but also the recovery of the truth information of the generated particle, this is the main goal of this work. To do this, the LHCb software is used to reconstruct a data sample from MC simulated events and identify the most relevant variables in the reconstruction chain and study the behaviour of these variables to get an optimum configuration. Once the variables

were identified, the main idea is to implement some modifications in the reconstruction algorithms involved in the ghost recognition.
Within this thesis the natural system of units was used, $\hbar = c = 1$.

2. The Standard Model

The Standard Model (SM) of particle physics is our best approximation to the behavior of fundamental particles and their interactions. It describes the phenomenology of matter and radiation at microscopic level [3]. The SM combines three out of four fundamental interactions of nature: weak, electromagnetic and strong. The weak force describes the decays of leptons and quarks, this force is responsible for radioactivity phenomena such as the beta decay of the neutron. The electromagnetic force is responsible for the interaction between electrically charged particles. For instance, atoms, molecules and solid state bodies are held together by this force [3]. The strong force is responsible for binding protons and neutrons to nuclei. Although this is a short distance interaction, it is the strongest force. Finally the fourth force, gravity is not described within the SM framework. The SM describes the matter throughout quantum fields that represent particles called fermions that carry a half-integer spin. They are divided into quarks and leptons, where both are arranged in three generations of doublets. The six different quark flavors are grouped in up-type (up, charm and top) and down-type quarks (down, strange and bottom) [3].

Fermions (Spin 1/2)

	Quarks			**Leptons**		
Generation	Flavour	Electric Charge	Mass	Flavour	Electric Charge	Mass
1^{st}	u	$+2/3e$	$2{,}3^{+0{,}7}_{-0{,}5}$ MeV	ν_e	0	< 2 eV
	d	$-1/3e$	$4{,}8^{+0{,}5}_{-0{,}3}$ MeV	e	$-1e$	511,0 keV
2^{nd}	c	$+2/3e$	$1{,}28 \pm 0{,}03$ GeV	ν_μ	0	< 2 eV
	s	$-1/3e$	95 ± 5 GeV	μ	$-1e$	105,7 MeV
3^{rd}	t	$+2/3e$	$173{,}2 \pm 1{,}2$ GeV	ν_τ	0	< 2 eV
	b	$-1/3e$	$4{,}18 \pm 0{,}03$ GeV	τ	$-1e$	$1776{,}8 \pm 0{,}2$ MeV

Table 2-1.: Properties of fermions in the SM. Quarks and leptons are grouped in three generations [6].

All quarks can participate in all fundamental interactions due to the fact that they carry a color charge (strong interaction), flavor charge (weak interaction) and an electric charge (electromagnetic interaction). Depending on its type electric charge can be a positive charge of 2/3 of the elementary charge (e) or a negative charge $1/3e$. Moreover, the quark masses have a hierarchy from the first generation which is the lightest to the third generation which is the heaviest.

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}.$$

There are also three generations of leptons that appear in three different flavors: electron, muon, and tau. Each doublet consist of a charged particle and a tinny neutral particle called neutrino:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}.$$

All leptons are weakly interacting due to the weak charge they carry. However, only the electrically charged leptons can also interact electromagnetically [3]. A summary of all fermions is shown in Table **2-1**. Each fermion, f, has its corresponding anti-fermion $\bar{f}$ which carries the opposite quantum numbers but the same mass.

The SM lagrangian, which describes the dynamics of particles, is invariant under local gauge transformations of the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. This group has 12 generators that are associated to gauge bosons (particles with integer spin). Gauge bosons are responsible for each fundamental interaction.

Bosons

Spin 1	Mass	Interaction	**Spin 0**	Mass	couples to
γ	0	electromagnetic	H^0	125,7 GeV	mass
Z^0	$91{,}188 \pm 0{,}002$ GeV	weak			
$W^\pm$	$80{,}385 \pm 0{,}015$ GeV				
$g(8)$	0	strong			

Table 2-2.: Properties of bosons in the SM [6].

The strong interaction is described by Quantum Chromo-Dynamics (QCD) which is based on the symmetry group $SU(3)_C$. This group has 8 generators corresponding to eight massless particles called gluons, g. Color charge is the quantity conserved in QCD. There are three different color charges: red, green, blue (and their corresponding anti-colors). Due to the fact that only quarks and gluons carry color charge, quark-gluon and self-interactions (gluon-gluon) are allowed. The QCD coupling constant, α_s, is not really a constant, it depends on the energy scale. The dependency on the energy of α_s is shown in Figure. **2-1**.

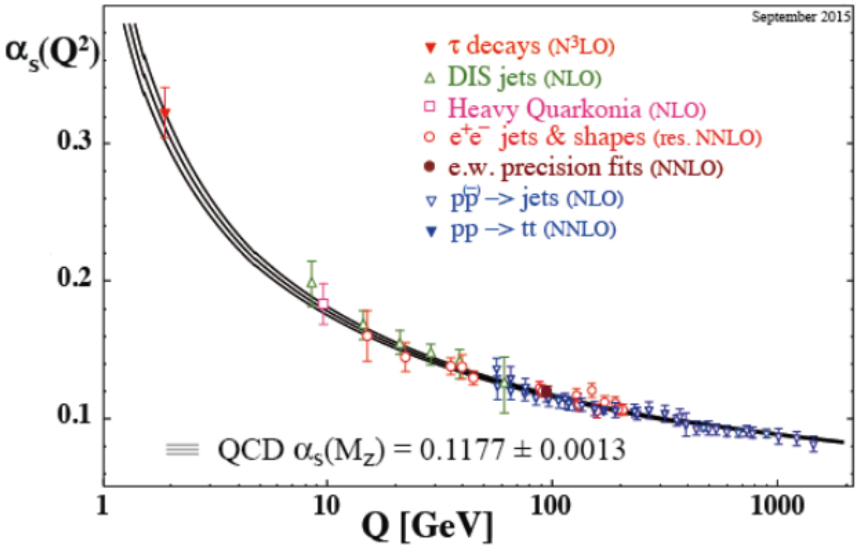


Figure 2-1.: Summary of measurements of α_s as a function of the energy scale Q [7].

This dependency leads to two important effects. At low energies, an effect called confinement is produced due to the large value of α_s. This phenomenon describes that color charges cannot be isolated, and therefore, they cannot be directly observed in normal conditions. Towards large energies, α_s decreases and leads to asymptotic freedom of the colored particles. As a result, quarks are quasi-free while they keep only small distances compared to the hadron radius [3]. Moreover, confinement implies that only color-neutral objects can exist. These objects are called hadrons which are divided into some groups depending on the number of quarks that composed them: barions made of three quarks, Mesons made of a pair quark-antiquark and there are predictions of other states of four and five quarks called Tetra-quark and Penta-quark respectively [8, 9].
Glashow proposed, in 1961, the symmetry group $SU(2)_L \otimes U(1)_Y$ to unify weak and electromagnetic interactions in a single theory. This predicts four massless bosons a massless triplet W_μ and a massless singlet B_μ. The triplet is associated to the $SU(2)_L$ group of the weak interaction in which the conserved quantum number is the weak isospin T. And the singlet is associated to the group of the electromagnetic interaction in which the conserved quantum number is the hypercharge Y. The electric charge Q is defined as $Q = Y - T_3$, where the T_3 is the third component of the weak isospin. The W_μ bosons can only couple to left-handed particles (or right-handed anti-particle) with weak isospin. Thus, the theory is called a chiral theory and the symmetry group has an index of L for left-handed. As all particles carry hypercharge, the B_μ boson can couple to all particles [3].

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{em} \tag{2-1}$$

The symmetry described in eq. 2-1 is spontaneously broken introducing a complex scalar field

(Φ). As a result, the theory predicts the existence of four physical vector boson eigenstates, $W^{\pm}$, Z^0 and γ, as a linear combination of the gauge bosons. The mediators of charged currents are the W^+ and W^-, ($M_{W^{\pm}} \approx 80$ GeV), Z^0, W^-, ($M_{Z^0} \approx 90$ GeV) shown in the Table **2-2**, is the one that mediates neutral currents and the photon, γ, the electromagnetic mediator that couples to the electric charge of the particle. Since the photon is massless the electromagnetic interaction has an infinite range.

The fundamental particles described in the gauge symmetry have to be massless, otherwise, this violates the local gauge invariance. It means massive particles such as bosons $W^{\pm}$ and Z^0 are not allowed. The Higgs mechanism solves the problem by introducing a scalar doublet. A combination of these doublets gives us the Higgs potential depicted in Figure **2-2** which has a non-zero ground state and spontaneously breaks the electroweak symmetry giving the masses to the particles. This predicts the existence of a spin-zero particle called the Higgs boson.

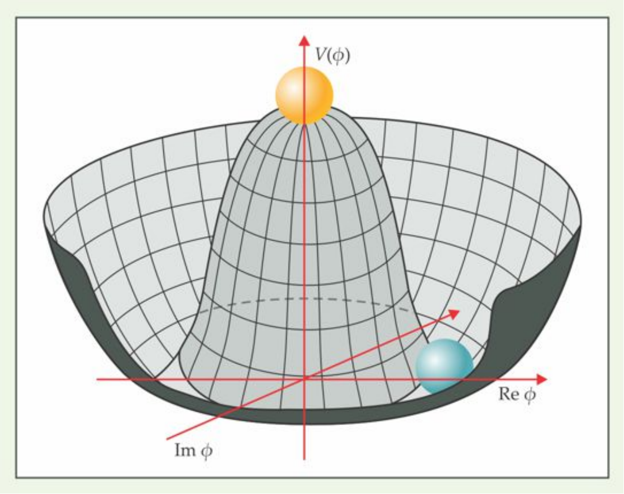


Figure 2-2.: Higgs potential [10].

The combined results from ATLAS and CMS collaborations, in July 2012, announced the discovery of a new particle compatible with the SM Higgs boson [11, 12]. The discovery completes the whole particle content in the SM.

2.1. CKM mechanism

The SM cannot explain some crucial aspects, for instance, Why is there such a large number of parameters for quark masses and the Cabibbo–Kobayashi–Maskawa (CKM) mixing matrix, spanning such a wide range of values? Why are the electroweak and strong interactions treated separately? Why is antimatter absent from the observed universe, even though the amount of CP violation in the SM is too small to produce the observed matter-antimatter asymmetry? [13]. Therefore, the CKM is a key element in describing flavor dynamics. This is able to describe a large range of phenomena in the quark sector such as CP-violation and flavor changing currents.

After the spontaneous symmetry breaking (SSB), the lagrangian for the Yukawa couplings of the Higgs boson to the quark fields is

$$\mathcal{L} = -(M_d)_{ij}\overline{D'_{Li}}D'_{Rj} - (M_u)_{ij}\overline{U'_{Li}}U'_{Rj}, \tag{2-2}$$

where i and j are family indices, with $U' = (u', c', t')$ and $D' = (d', s', b')$, and L and R indicate the components with left- and right-handed chiralities, respectively [13] and the matrices M_u and M_d are related to the Yukawa coupling matrices as $M_q = vY^q/\sqrt{2}$ where v is the vacuum expectation value of the Higgs field. Those matrices should be diagonalised. The Lagrangian can be rewritten in terms of the mass eigenstates letting the neutral gauge bosons unaffected, nevertheless, the structure of charged-current interactions between quarks and $W^{\pm}$ is modified.

$$\mathcal{L} = -\frac{g}{\sqrt{2}}\overline{U_i}\gamma^{\mu}\frac{1-\gamma^5}{2}(V_{CKM})_{ij}D_j W^{\pm}_{\mu} + hc., \tag{2-3}$$

where g is the electroweak coupling constant and $V_{CKM} = V^{\dagger}_{uL}V_{dL}$ is the unitary CKM matrix that can be written as

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{2-4}$$

At tree level, the CKM matrix is responsible for the flavor-changing transitions inside and between the quark generations in the charged sector nonetheless, there are no flavor-changing transitions in the neutral sector. The CKM matrix elements represent the couplings between up-type and down-type quarks. Notwithstanding, the conventions used to defined this matrix were quite arbitrary. For instance, some parameters can be redefined to reduce the number of real parameters needed such as the relative phases among the left-handed quark fields. Another important fact, based on experimental evidence, is the suppression of the transitions, it means the transitions between the same generation are characterized by V_{CKM} elements of $\mathcal{O}(1)$, those between the second and the third generation have a factor $\mathcal{O}(10^{-1})$ and finally the transitions between the first and the third generations are suppressed by a factor $\mathcal{O}(10^{-2})$. The hierarchy can be expressed by the four-phase convention-independent quantities [13].

The parameterization of the CKM matrix proposed by Wolfenstein is one of the most used.

$$\lambda = \frac{|V_{us}|^2}{\sqrt{(|V_{ud}|^2 + |V_{us}|^2}}), \tag{2-5}$$

$$A\lambda^2 = \lambda\left|\frac{V_{cb}}{V_{us}}\right|, \tag{2-6}$$

$$A\lambda^3(\rho + i\eta) = V^{*}_{ub}, \tag{2-7}$$

however, a useful parametrization of the CKM matrix is used to let all its parameters in terms of powers of λ which corresponds to $\sin\theta_C \approx 0{,}22$, exploiting the unitarity of V_{CKM} to highlight its hierarchical structure. This expansion yields the following parametrization up to $\mathcal{O}(\lambda^5)$.

$$V_{CKM} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2-\frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho-i\eta) \\ -\lambda+\frac{1}{2}A^2\lambda^5[1-2(\rho+i\eta)] & 1-\frac{1}{2}\lambda^2-\frac{1}{8}\lambda^4(1+4A^2) & A\lambda^2 \\ A\lambda^3[1-(\rho+i\eta)] & -A\lambda^2+\frac{1}{2}A\lambda^4[1-2(\rho+i\eta)] & 1-\frac{1}{2}A^2\lambda^4 \end{pmatrix} \tag{2-8}$$

2.2. CP violation

The charge conjugation parity symmetry (CP) violation in the Standard Model of particle physics, plays an important role not only in attempts to explain the dominance of the matter over antimatter in the present universe but also, in the study of the weak interactions in particle physics. The presence of strong and weak phases in the CKM matrix is the source of the CP violation, which is allowed if and only if η is different from zero. It is also important in the searches of new physics beyond the SM.

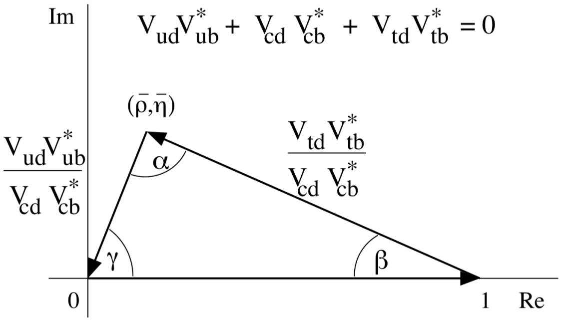


Figure 2-3.: The unitarity triangle shown in the complex plane [14]

The SM predicts CP violation in transitions that involve top and bottom quarks, specifically in the V_{td} and V_{ub} components of the CKM matrix. This fact reveals the importance of experiments such as b-factories. Although CP violation can be studied in the top quark, experiments using the b quark are preferable due to the available energy. Due to the quark masses, mesons that contain bottom quarks are easier to produce than top quarks. CP violation can be studied in charm mesons also. However, these transitions are highly suppressed since the term that contains CP violation has a factor λ^5 which is too difficult to measure in the experiments.

Figure 2-3 shows one of the six unitarity triangles predicted by the SM. The area of the triangles is directly related to the quantity of CP violation presented in a certain decay channel. All six unitarity triangles have an equal surface area, denoted as $J = A^2\lambda^6\eta$. It is

important to mention that the CP violation involves the measurement of the γ angle in the triangle. If that angle was zero there was no CP violation at all.
The CP violation effects can be classified in various types such as CP violation in decay, CP violation in mixing, and CP violation in interference between a decay with and without mixing [14]. A brief explanation of these effects will be given as follows[1].
The CP violation in decay occurs when the decay rate of a B and an anti-B to a final state f and its CP-conjugated state $\overline{f}$ differ.

$$\Gamma(P^0 \to f) \neq \Gamma(\overline{P^0} \to \overline{f}), \tag{2-9}$$

when the CP-asymmetry can be measured as

$$A_{CP} = \frac{\Gamma(B \to f) - \Gamma(\overline{B} \to \overline{f})}{\Gamma(B \to f) + \Gamma(\overline{B} \to \overline{f})} < 0. \tag{2-10}$$

This type of CP violation can occur in neutral and charged mesons for instance in the $B^0 \to K^+\pi^-$ decay where the asymmetry is big enough to be measured at the current experiments. The CP violation in mixing implies that the oscillation from a meson to an anti-meson differs from the oscillation from an anti-meson to a meson as follows

$$\mathbf{Prob}(P^0 \to \overline{P^0}) \neq \mathbf{Prob}(\overline{P^0} \to P^0). \tag{2-11}$$

Semileptonic decays of the B^0 and $\overline{B^0}$ are important for CP violation searches of this type. If the B pair is coherently produced, for instance, $\Upsilon \to B^0\overline{B^0}$, the B^0 meson decay to a positively charged lepton and vice versa. Therefore the two leptons with the same charge in the final state imply that one of the mesons oscillated. This allows us to compare the asymmetry in the number of two positive and two negative leptons.

$$A_{CP} = \frac{N_{++} - N_{--}}{N_{++} + N_{--}} = \frac{|p/q|^2 - |q/p|^2}{|p/q|^2 + |q/p|^2}, \tag{2-12}$$

where the violation is presented if $|q/p| \neq 1$. However in the B_s and B^0 systems the ratio is approximately 1 within the experimental accuracy. So, this is observed in the K-system.
The CP violation in interference between a decay with and without mixing occurs when there are oscillations involved. In this case, the CP violation is measured in decays that have a common final state for the B and $\overline{B}$ masons. If the condition

$$\Gamma(P^0_{(\to\overline{P^0})} \to f) \neq \Gamma(\overline{P^0}_{(\to P^0)} \to f), \tag{2-13}$$

is satisfied the CP symmetry is violated. There are two amplitudes that contribute to the transition amplitude from the initial state $|B^0\rangle$ to the final state f, as a direct consequence

---

[1] See ref [14] for further information.

of $f = \overline{f}$. The two amplitudes are $A(B^0 \to f)$ and $A(B^0 \to \overline{B^0} \to f)$. If $|q/p| = 1$, the asymmetry can be written as

$$A_{CP} = \frac{\Gamma(P^0(t) \to f) - \Gamma(\overline{P^0}(t) \to f)}{\Gamma(P^0(t) \to f) + \Gamma(\overline{P^0}(t) \to f)} = \frac{C_f \cos(\Delta m t) - S_f \sin(\Delta m t)}{\cosh(\Delta\Gamma t/2) + D_f \sinh(\Delta\Gamma t/2)}. \tag{2-14}$$

If the transition is dominated by one amplitude, the previous expression is simplified since one of the amplitudes C_f or S_f could be zero or negligible.

Finally, an alternative classification of the CP violation effects is direct and indirect where the direct type appears in the CP violation in decay and, the indirect type of CP violation that involves mixing in any way [14].

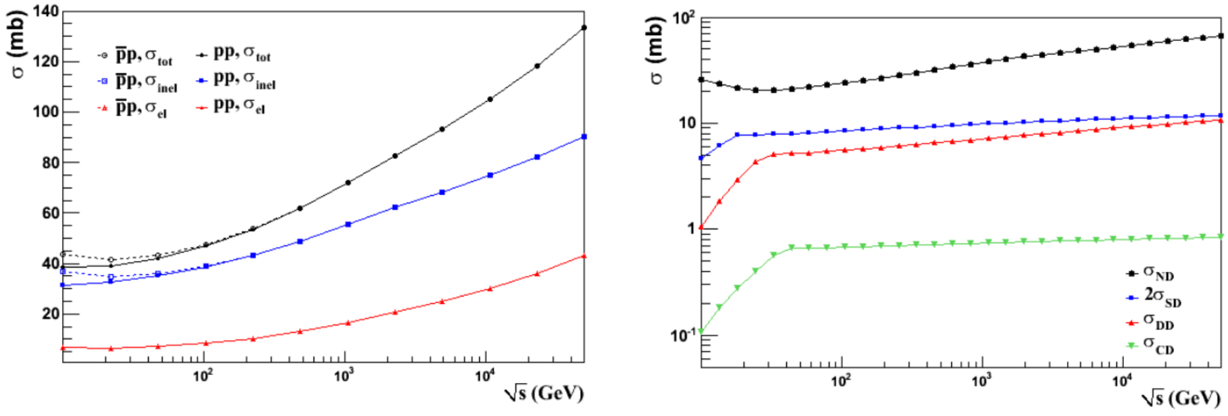
2.3. Particle production in pp collisions at LHC

High energy protons collide at a centre-of-mass energy of to $\sqrt{s} = 13$ TeV[2]. The pp interactions become extremely complex at these energies. Depending on the momentum transferred the collision between the two protons can be studied as the scattering of two compound objects (low momentum transferred), or inelastic collisions (high momentum transferred) which have enough energy to break up the proton allowing its constituents (partons) to participate in the collision. A proton is compound by three valence quarks (uud), sea quarks and gluons. The total momentum of the proton is shared among all its partons. The momentum is described by parton density functions (PDFs). This distribution shows the probability density of finding a certain parton with a momentum fraction of x of the proton. At the LHC operation energies, the interaction is not only sensitive to the valence quarks but also the gluons and sea quarks creating a pair quark/antiquark from the vacuum. As a result, the partons in the two protons can interact via quark-quark, quark-gluon and gluon-gluon which are mediated by the strong interaction (QCD). There are hard and soft hadronic processes. The hard processes involve one or two partons that have a large momentum fraction of the proton. Hence, there is a large momentum transferred in the interaction at the called hard energy which is typically > 1 GeV^2. Due to the nature of α_s depicted in Figure **2-1**, where the larger the energy is, the smaller the α_s becomes. Therefore, hard QCD processes can be calculated using perturbation theory.

This behavior allows to accurately predict production rates and event properties. Heavy bosons, e.g. $W^{\pm}$ or H^0 bosons, leptons, e.g. τ , or heavy quarks, such as bottom or top quarks are produced in the hard QCD collisions due to the available energy. The probability to produce heavy particles depends on several parameters within the most important are; the mass of the produced particles, the available energy and the particle species that participate in the collision.

---

[2] Energy for Run II which comprises the years 2015-2019, for Run I the energy was $\sqrt{s} = 8$ TeV.



(a) (b)

Figure 2-4.: pp cross section as a function of the energy $\sqrt{s}$. Total, elastic and total-inelastic *pp* and p-$\bar{p}$ cross sections (a), and Diffractive (SD, DD, CD) and non-diffractive (ND) cross sections (b) [15].

The inelastic pp cross-section at the LHC energy of $\sqrt{s} = 13$ TeV has been measured by the LHCb, ATLAS and CMS experiments to be $75{,}4 \pm 3{,}0 \pm 4{,}5$ mb [16], $78{,}1 \pm 2{,}9$ mb [17] and $71{,}3 \pm 0{,}5 \pm 2{,}1 \pm 2{,}7$ mb [18] respectively. The cross-section for different processes can differ by several orders of magnitude as shown in Figure. **2-4**b. The main characteristic of soft processes is the transferred momentum typically of a few hundred MeV. The perturbation theory can not be applied at this regime due to the α_s value as shown in Figure. **2-1**. Soft processes also include diffractive dissociation in hadron-hadron collisions.

The total pp cross-section is composed of different contributions; the elastic scattering (EL), the single diffractive (SD), the double diffractive (DD), the central diffractive (CD) and the non-diffractive (ND) which will be briefly explained afterward.

$$\sigma_{Tot} = \sigma_{EL} + \sigma_{SD} + \sigma_{DD} + \sigma_{CD} + \sigma_{ND}, \tag{2-15}$$

where the first term is the elastic scattering and the other terms correspond to the inelastic processes. In the inelastic collision, both protons come and out unscattered ($p + p \rightarrow p + p$). The single diffractive occurs when one of the protons remains intact and the other breaks up to create particles ($p + p \rightarrow p + X$). It is important that no quantum numbers are exchanged between the protons and thus the quantum numbers of the produced final-state particles are the same as the ones of the original proton. Besides, The double diffractive occurs when both protons break up to create particles ($p + p \rightarrow X + X$). Once again, no exchange of quantum numbers between both incident particles is allowed. If none of the protons survive the collision and quantum numbers are exchanged the process is called non-diffractive ($p + p \rightarrow X$). Additionally, there are central diffractive processes which create neutral color objects such as J/ψ or H^0 and preserve the quantum numbers of the initial protons, $(p+p \rightarrow p+X+p)$. The identification of diffractive processes can be a hellish task from the experimental point of view. Therefore, most analyses of inelastic interactions are treated as inclusive inelastic interactions. The elastic interactions are not considered due to the fact that LHCb was not designed to measure them.

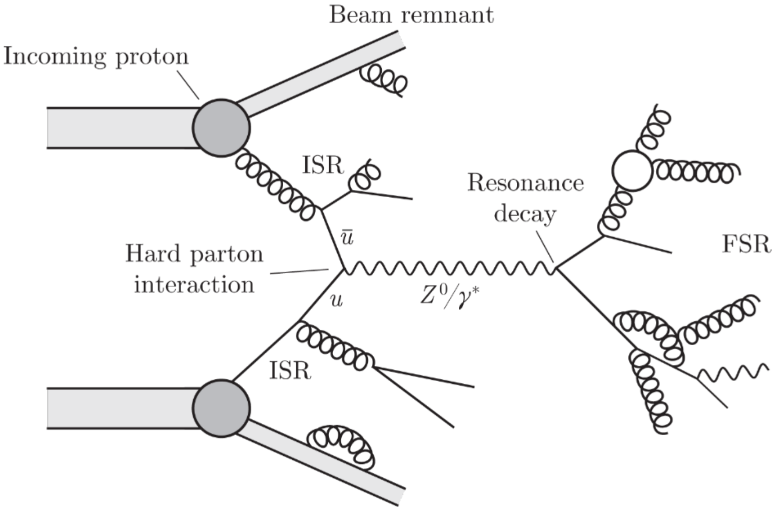


Figure 2-5.: Illustration of an inelastic pp collision at the LHC [3].

A typical inelastic pp collision at high energies is illustrated in Figure. **2-5**. The partons of one of the incoming protons interact with the partons of the other proton via the strong interaction. The beam remnant is the part of the proton which usually survives the collision and it is slightly deflected from its original path. The heavy resonances such as the Z^0 boson in the example can be produced by hard parton interactions. There is final-state radiation where produced partons further emit bremsstrahlung or gluon-strahlung. In addition, there are soft interactions between the incoming partons which also produce quarks or gluons resulting in multiple-particle production of light hadrons. The partons produced by any of these processes must all form color-neutral mesons or baryons. This procedure is called harmonization [3].

2.4. Production of b hadrons at LHC

The leading-order (LO) mechanisms for the $b\bar{b}$ production are quark-antiquark annihilation and gluon fusion. The examples of these kinds of processes are given in the Figure. **2-6** a) and b). For a centre of mass energies at the TeV scale, next to leading order (NLO) contributions, such as flavor excitations and gluon splitting, become more and more important. At the energies achieved by LHC $\sqrt{s} = 13$ TeV, the dominant processes are flavor excitation mediated. Figure. **2-6** c) and d) show examples of them. The mass of the b quark, $\approx 4{,}2$ GeV, is too small compared with the available energy in a pp collision at the LHC operation scales. Therefore, there is necessary that just one of the partons involved in the collision carries large energy to produce a $b\bar{b}$ pair. This is especially important for the experiment thus, it results in a boost of the created $b\bar{b}$ pair. The boost is directed along the flight direction of one of the colliding protons.

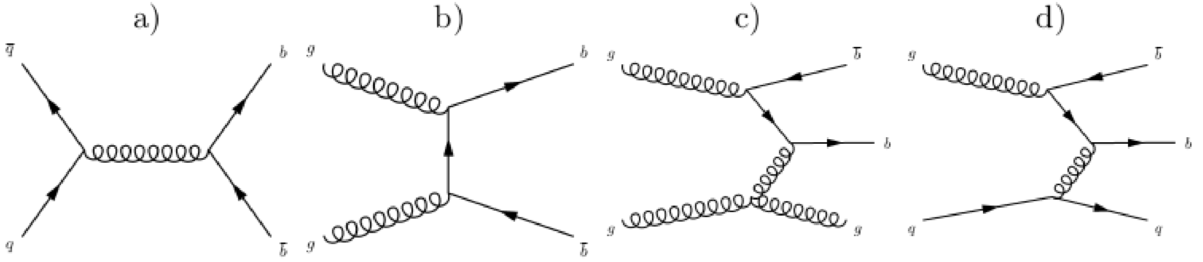


Figure 2-6.: Feynman diagrams for the production of $b\bar{b}$ pairs in a pp collision at the LHC. Examples are given for leading-order: $q\bar{q}$ annihilation (a) and gluon fusion (b); next to leading order: flavor excitation (c-d) [3].

LHCb has measured the cross-section for producing b quarks in the reaction $pp \to b\bar{b}X$ for 7 TeV and 13 TeV collisions at the LHC as a function of the pseudorapidity in the range $2 < \eta < 5$ which corresponds with the acceptance of the LHCb experiment. The cross-section values in the covered η range are $72{,}0 \pm 0{,}3 \pm 6{,}8$ μb and $144 \pm 1 \pm 21$ μb, where the quoted uncertainties are statistical and systematic, respectively. Which means that $N = \int \sigma d\mathcal{L}$. Bosons have been produced within the LHCb acceptance.

3. The LHCb Experiment at LHC

The Large Hadron Collider (LHC) is the largest and most powerful particle accelerator in the world. It is located at CERN, the European Organization for Nuclear Research [1], in Geneva, Switzerland. Two superconducting rings are installed in a 26,7 km tunnel. This tunnel has eight straight sections and eight arcs and lies between 45 m and 170 m below the surface on a plane inclined at 1,4 % sloping towards the Léman lake [19]. The LHC is designed as a high-luminosity pp collider, which has the main purpose to study physics beyond the SM. However, it can also be operated to collide heavy ions. It consists of two beamlines surrounded by powerful superconducting magnets that keep charged particles on their orbit. The beams are brought to collide at four intersection points, where the main LHC experiments are located [3].

CERN has a total of seven experiments associated with the LHC, each of the four collision points accommodates one of the four big experiments ATLAS, CMS, ALICE, and LHCb. The main purposes of ATLAS (A Toroidal LHC ApparatuS) [20] and CMS (Compact Muon Solenoid) [21] are the direct searches of new physics. Their goals are to discover new particles coming from pp collisions, such as candidates for dark matter, supersymmetry and the most important one achieved in July 2012, the Higgs boson discovery[11] [12], which was the final brick that completes the Standard Model.

ALICE (A Large Ion Collider Experiment) [22] is a heavy-ion detector focused on the QCD sector. The quark-gluon-plasma is a new phase of matter that can be achieved by strongly interacting particles at extreme energies and temperatures in Pb-Pb collisions. In order to vary energy density and interaction volume proton-nuclei collisions are also studied.

The LHCb experiment (Large Hadron Collider beauty) [2] was designed to look for indirect evidence of new physics in CP violation and rare decays of beauty and charm hadrons [2].

3.1. Physics Motivations at LHCb

The results obtained in heavy flavor physics, B factories, and Tevatron, are consistent with the CKM mechanism. However, weak interactions of the SM only, explain a small part of the difference between matter and antimatter observed in the universe. Many models of physics beyond the standard model have been proposed to add new CP violation sources. Effects such as CP-violating phases, rare decays branching fractions, and new decays that

---

[1] The abbreviation originates from the French name, Conseil Européen pour la Recherche Nucléaire

are forbidden by the SM, might be seen in heavy flavor physics. Therefore, rare decays and CP violation should be studied in B_d, B_s and D with high statistics and different decay modes.

A large $b\bar{b}$ production cross-section of $\sim 500\ \mu b$ at 14 TeV is expected, which becomes LHCb a huge source of B mesons. The experiment is also capable of producing large quantities of B_c and b-barions such as Λ_b. LHCb reduces the luminosity on purpose for two reasons: To limit the number of pp interactions per bunch, and to reduce the occupancy in the detector as a consequence the radiation damage is limited. Each data taking year collects up to 10^{12} b-$\bar{b}$ pairs at the luminosity of 2×10^{32} cm^{-2} s^{-1} [2]. This is tunned by changing the beam focus and the interaction point independently of the other detectors and interaction points.

LHCb has design the trigger system to cope with the harsh hadronic environment and exploit the huge amount of data. This should be sensitive to different final states and an excellent vertex, and momentum resolution is crucial for proper time-related measurements like B_s-$\bar{B}_s$ meson systems. It is also required an excellent m_0 resolution in order to reduce the combinatorial background.

3.2. The LHCb Detector

LHCb is a single arm spectrometer with a forward angular coverage from approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane. The choice of the detector geometry is justified by the fact that at high energies both the b and $\bar{b}$ hadrons are predominantly produced in the same forward or backward cone [2] as a consequence of asymmetric parton collisions. The topology of b and $\bar{b}$ motivates the key aspects of the LHCb layout [3]. The z axis is aligned with the beam direction and y axis is pointing upwards.

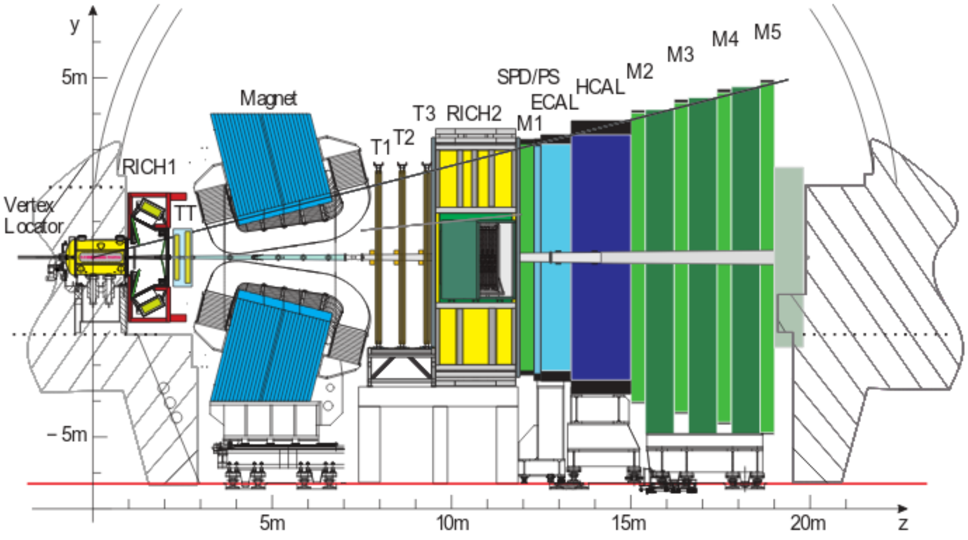


Figure 3-1.: Schematic overview of the LHCb detector in the y-z plane [2].

The histogram shows the angles between b and $\bar{b}$ quarks. The z-axis produced at the LHC. The red area indicates the coverage by the LHCb detector in the forward region.

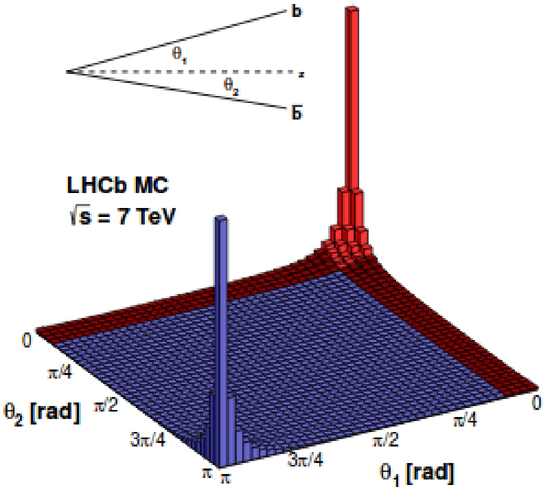


Figure 3-2.: The histogram shows the distribution of the angle between the beam axis and $b\bar{b}$ production in pp collisions [3].

The LHCb is composed by several subdetectors among them, The VELO is the first component, the second component is the Tracker Turicensis (TT) located just in front of the dipole magnet and the main tracking stations T1, T2, and T3 which consist of two independent tracking subsystems. The nearest region to the beam pipe, of the main tracking stations, is covered by the Inner Tracker (IT) and the outer region is covered by the Outer Tracker (OT). This group of sub-detectors is responsible for tracking reconstruction. The group of subdetectors in charge particle identification (PID) consists of: Two Ring Imaging Cherenkov Detectors, RICH1, and RICH2, which allow the distinction between the different hadron species. RICH1 is located behind the VELO. It deals with low momentum particles. RICH2 is located behind the T stations. It deals with high momentum particles. RICH2 is followed by the calorimetry system. The Electromagnetic (ECAL) and the Hadronic Calorimeter (HCAL) which give us the information about the energy of the particles by absorption. They are assisted by a Scintillation Pad Detector (SPD) and a Pre-Sower detector (PS), both are primarily used for particle identification and Finally, the muon system consists of five stations, M1 placed in front of the calorimeters and M2-M5 placed behind.

3.3. The tracking system

The tracking system was designed to detect the signatures of the charged particles traversing the detector[3]. The hits in each subdetector left by charged particles are used to reconstruct the corresponding track of the particle. Momentum information is calculated taking into account the deflection caused by the magnetic field. The VELO and the TT are located in front of the magnet and the T stations behind it. The combination of the information gives the momentum and the trajectory of each charged particle.

3.3.1. VErtex LOcator (VELO)

The VELO is the subdetector located closest to the primary interaction region. It provides the first measurement of the trajectories of charged particles. The VELO was designed to measure the position of the primary interaction vertex (PV) and distinguish it from any displaced vertex, secondary interactions, and pile-up within the same event. The VELO surrounds the interaction region with the highest particle flux therefore highly radiation tolerant silicon sensors are adopted. Two different types of sensors were designed (r and ϕ). These sensors employ half-discs shape made of silicon microstrips. r-type sensors use silicon micro-strips arranged in concentric circles, each strip is subdivided in 45° sections in order to reduce occupancy. Strip pitch decreases linearly from 102 μm to 38 μm.
ϕ-type sensors have an approximately radial strip design grouped in two regions of different pitch sizes. This prevents the sensors from too large occupancies and pitch sizes at the edge regions[3].

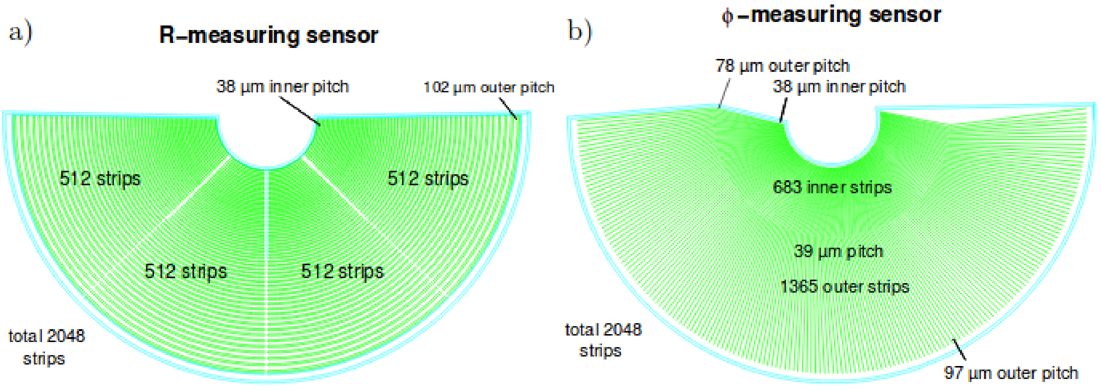


Figure 3-3.: VELO r and ϕ sensor layout [3].

The VELO has 21 stations where each has a left and right detector module of the beam axis. The modules comprise one r- and one ϕ-type sensor. The modules are slightly displaced in order to allow a small overlap between the sensors. A schematic overview of the VELO layout is shown in Figure. 3-4. Around the nominal interaction region, the modules are more closely spaced than downstream. For vetoing pile-up events, two additional stations using r-type sensors are placed upstream of the interaction point.

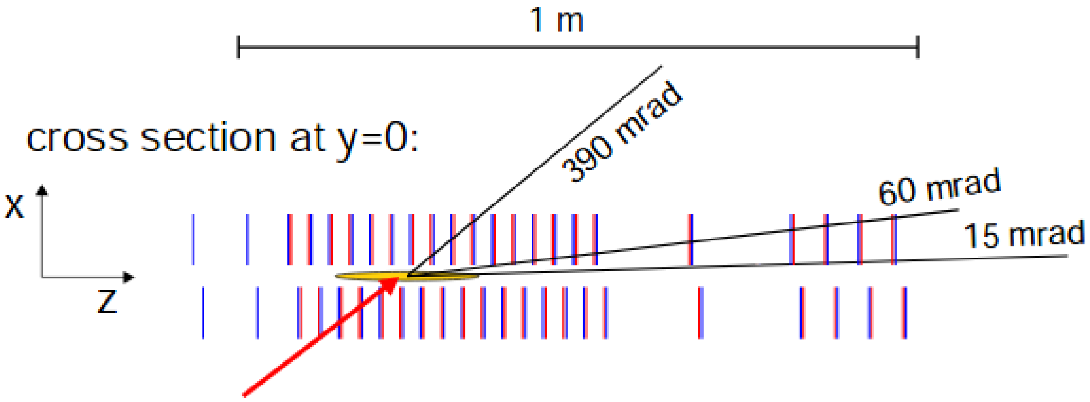


Figure 3-4.: VELO panels layout [23].

The beam pipe is removed within the VELO to reduce the material responsible for multiple interactions of particles. All the modules are mounted in a vacuum vessel.

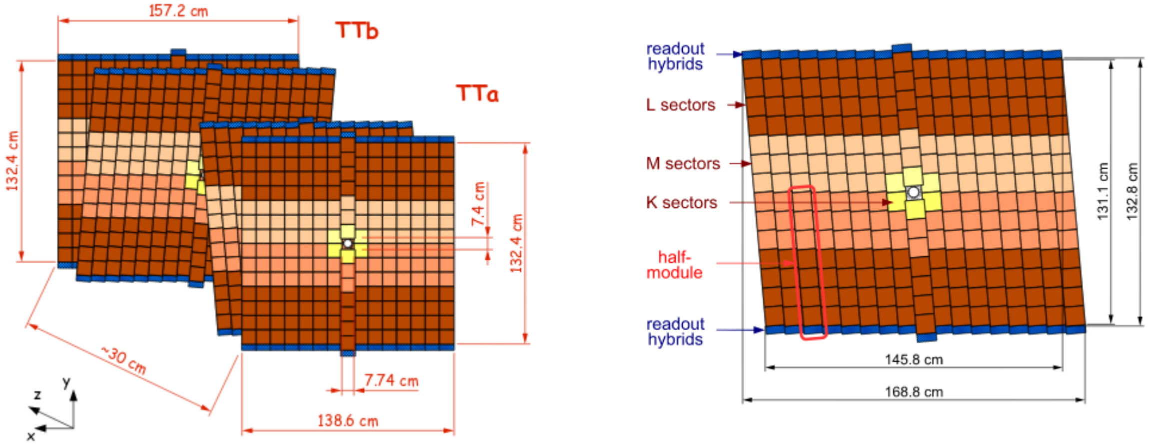
A thin RF foil, made of an aluminum alloy, is used to separate the beam and the sensor discs. This is important for two reasons. The ultra-high vacuum of the LHC beam is protected from the out-gassing of the VELO sensors, and further, the sensors are shielded from electromagnetic effects induced by the LHC beam. The RF foil accounts for around 40 % of the material budget of the VELO [3]. The two halves of the VELO are movable along the x-axis. This is important to protect the VELO sensors from severe damage due to the fact that during injection proton- or ion-beams are less stable in position. Therefore VELO halves are kept 3 cm away from the beam during this phase. Once stable beams phase is achieved VELO is automatically closed up to 8 mm from the beam.

3.3.2. Tracker Turicencis (TT)

The Tracker Turicensis (TT)[2] is placed behind RICH1 just in front of the dipole magnet. The detector has 2×2 detection layers of silicon microstrips sensors. The strips have a pitch size of 183 μm and a maximum length of 38 cm. There are two kinds of layers, so-called x-layers, and stereo-layers (u,v). One layer of each type is arranged in a station, TTa has a (x,u) and TTb has (v,x) configuration, this is especially important to aid the track reconstruction algorithms. The strips in x-layers are aligned vertically, while the strips in stereo layers are tilted by either $-5°$ (u) or $5°$ (v) with respect to the vertical axis [3]. This allows parameters to be measured with higher resolution in the bending plane, adding the y component information. TTa and TTb are separated approximately 30 cm along the beam axis.

---

[2] The TT was formerly known as the Trigger Tracker.



(a) TT station configuration

(b) Single TT layer

Figure 3-5.: TT stations layout [2]

The layout of one of the detection layers is illustrated in Figure. 3-5b. It consists of a row of seven silicon sensors organized in two or three readout sectors.

The readout hybrids for all readout sectors are mounted at one end of the module [2]. The region above and below the LHC beam pipe is covered by a one-half module each. Rows of seven (for the first two layers) or eight (for the last two layers) are mounted to both sides of the beam pipe. Two half modules assembled end to end compose a full module that covers the full height of LHCb acceptance. Adjacent detection modules are staggered 1 cm in z direction to allow a few millimeters to overlap in order to avoid gaps in the detection modules and facilitate the relative alignment. The main advantage of this design is that front-end hybrids, cooling infrastructure and module supports are located outside the active area of the detector, outside the LHCb acceptance.

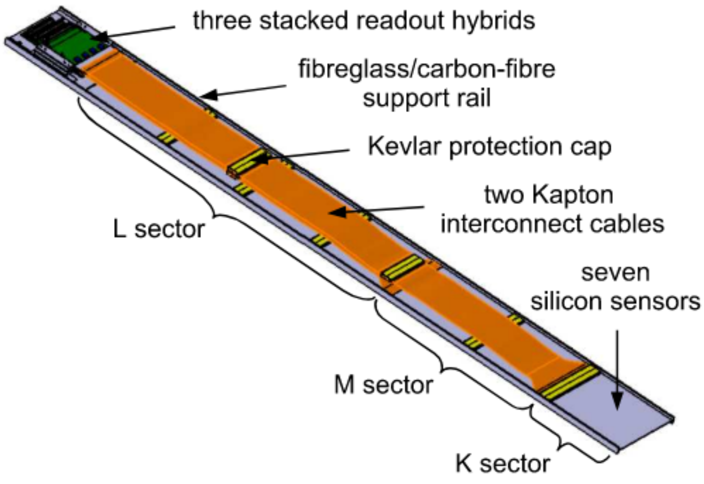


Figure 3-6.: TT Detector module [2].

TT Detector Modules: The layout of the half 4-2-1 type detector module is illustrated in Figure: 3-6. It has seven silicon sensors and two or three readout hybrids at the end. The half modules closest to the beam pipe are mounted following a 4-2-1 type module. The other modules are 4-3 type. Both cases the first sector has four sensors in which the strips are bounded together forming the L sector. The L sector is connected to the first readout hybrid. For 4-3 type the M sector is composed of three sensors bounded together connected to the second readout hybrids via a 39 cm long kapton flex cable. The M sector in a 4-2-1 type half module has two sensors. Finally, the K sector which covers the highest particle flux region consists of only one silicon sensor. Both sectors M and K are connected by 39 cm and 58 cm long kapton cables to the readout hybrids respectively. The silicon sensors are 500 μm thick. They are 9,64 cm wide and 9,44 cm long and carry 512 readout strips with strip pitch of 183 μm. The kapton interconnect cables carry 512 signal strips and two pairs of bias voltage and ground strips on a 100 μm thick kapton substrate. The strips consist of 7 μm thick copper with a 1 μm thick gold plating. They are 15 μm wide and have a pitch of 112 μm. A short pitch-adapter section in which the strip pitch widens to 180 μm permits to directly wire-bound the strips on the cable to the silicon sensor strips. A copper mesh backplane provides a solid ground connection and shielding against pick-up noise. The small strip width was required to keep the strip capacitance of the cable small but led to an unacceptably low production yield for fault free cables of the required length. The 39 cm long cables for the M sector, therefore, had to be assembled from two shorter pieces, and the 58 cm long cables for the K sector were assembled for three pieces. The pieces were joined together end-to-end by gluing them onto a thin strip of fiberglass reinforced epoxy. An electrically conductive adhesive was used to provide the electrical connection between the copper mesh backplanes of the two cable pieces [2].

3.3.3. The Magnet

A dipole magnet is used in the LHCb experiment to measure the momentum of the charged particles. The magnet consists of two trapezoidal coils bent at 45° on the two transverse sides, arranged inside an iron yoke of window-frame configuration. The magnet gap is wedge-shaped in both vertical and horizontal planes, following the detector acceptance. In order to provide space for the frames of the tracking chambers positioned inside the magnet, the planes of the pole faces lie 100 mm outside the ±250 mrad vertical acceptance and the shims on the sides of the pole faces 100 mm outside the ±300 mrad outside the horizontal acceptance. The horizontal upstream and downstream parts of the coils are mounted such that their clamps and supports do not penetrate into the clearance cone defined above for the frames of the tracking chambers [24]. the coils are identical, conical saddle shape and placed mirror-symmetrically to each other in the magnet yoke. Each coil consists of five triplets of pancakes made of pure $\mathrm{Al} - 99{,}7$ hollow conductor with a specific ohmic resistance below 28 Ω m at 20° C. This arrangement produces an integrated magnetic field of 4 T. Cast aluminum clamps are used to hold together the triplets, support, and center the coils concerning the measure mechanical axis of the iron poles with a few millimeters tolerance. Only the extremity of the coil, where the electrical and hydraulic terminations are located, is kept fixed. the other pancakes are free to slide upon their supports to ease the electromagnetic and thermal stress on the conductor. An alignment of the yoke is made just after rolling the magnet into its nominal position, to follow the 3,6 mrad slope of the LHC and its beam. The magnet is aligned to its nominal position with a precision of 2 mm while the alignment measurements in the position can be up to 0,2 mm.

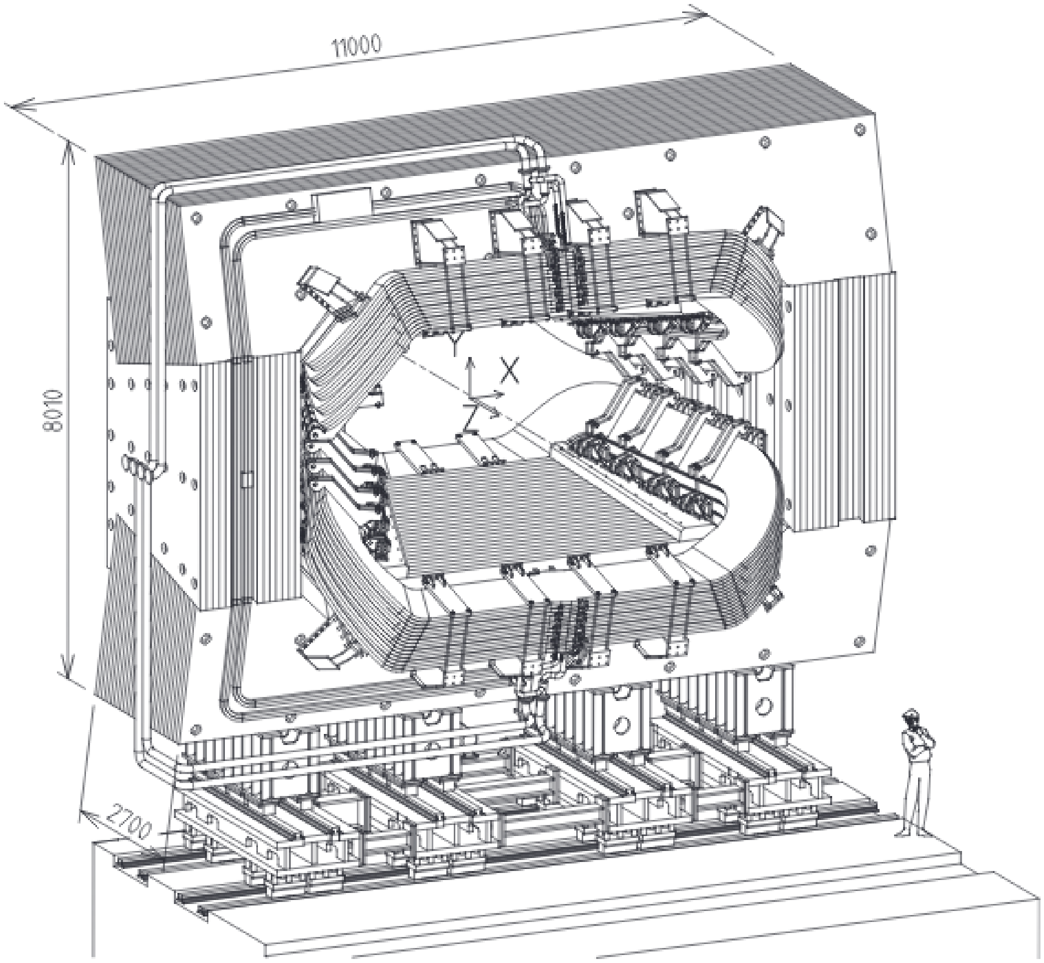
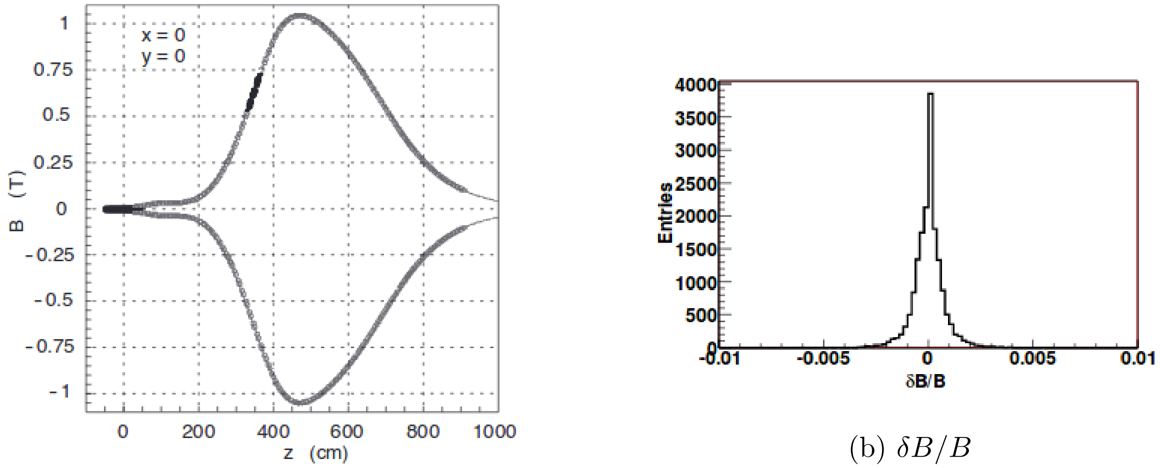


Figure 3-7.: Perspective view of the LHCb dipole magnet [2].

In order to achieve the required momentum resolution for charged particles, the $\int Bdl$ needs to be measured with relatively high precision of the order of (10^{-4}) and the position of the B-field peak with a precision of few millimeters. A quasi-automatically measuring system was designed which allows controlling the scanning along the longitudinal axis of the dipole [3]. It consists of an array of 60 sensor cards covering an 80 mm × 80 mm grid. Each sensor card contains three Hall probes mounted orthogonally on a cube whose main goal is to measure the three components of the magnetic field inside the tracking volume of the detector for the two magnetic polarities.



(b) $\delta B/B$

(a) Magnetic Field

Figure 3-8.: Magnetic field, Figure **3-8**a shows the magnetic field B_y for Up and Down configurations and Figure. **3-8**b shows the relative difference of the measurements of B using different Hall probes at the same position in the magnet. The resolution is completely dominated by the precision of the calibration of the Hall probes [2]

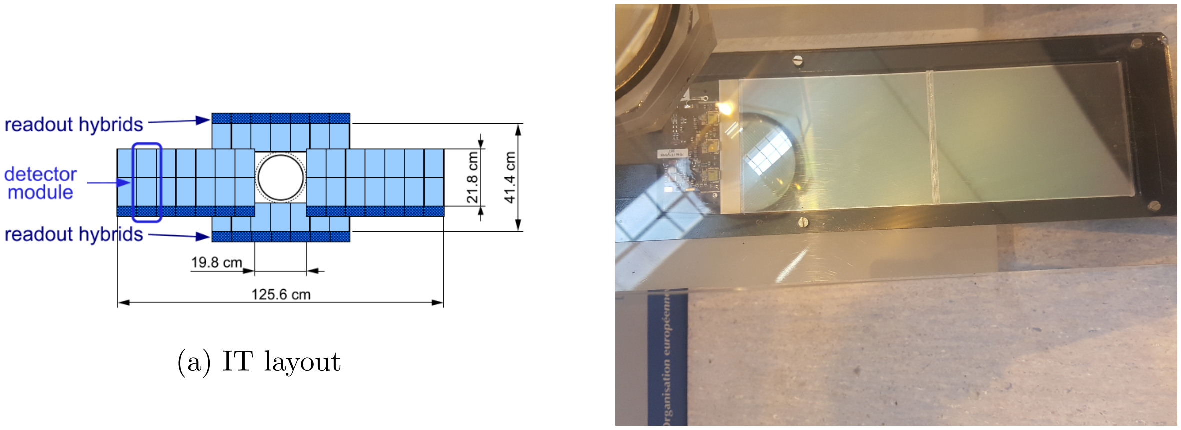
It is important to control the systematic effects of the detector by changing periodically the direction of the magnetic field for the measurement of the CP asymmetries.

3.3.4. Inner Tracker

The Inner Tracker (IT) is another microstrip detector located downstream of the magnet. It compounds the inner part of the T stations. It covers the region with the highest particle flux. The IT consists of three stations, where each station is built of four individual detector boxes. Similar to the TT, each box comprises four detection layers in a (x,u, v, x)-configuration. Each detection layer consists of seven detector modules. Adjacent modules in a detection layer are staggered by 4 mm in z and overlap by 3 mm in x to avoid acceptance gaps and facilitate the relative alignment of the modules. Detector modules in the boxes above and below the beam pipe (top and bottom boxes) consist of a single silicon sensor and a readout hybrid. Detector modules in the boxes to the left and right of the beam pipe (side boxes) consist of two silicon sensors and a readout hybrid [2].

IT detector modules: The modules are arrangements of either one of two silicon sensors that are connected to the front-end readout hybrids via a pitch adapter. There are two types of *silicon sensors* installed in the IT detector. The sensors for the two-sensor module are 7,6 cm wide and 11 cm long, carry 384 readout strips with a strip pitch of 198 μm and

410 μm thick. The sensors for the one-sensor module are 320 μm thick. These thickness were chosen to minimize the amount of material budget on the detector and ensure a significantly high signal-to-noise ratio for each module.



(a) IT layout

(b) IT silicon sensor

Figure 3-9.: IT modules [2].

The IT readout hybrids is highly similar in design and routing to the TT hybrids, and the pitch adapter is similar to the used for the M and K hybrids in TT. the two differences are: the kapton flex circuit for IT has three front-end chips and it incorporates an 89 mm readout tail with straight traces, at the end of which a 60-pin board-to-board connector is mounted [2].

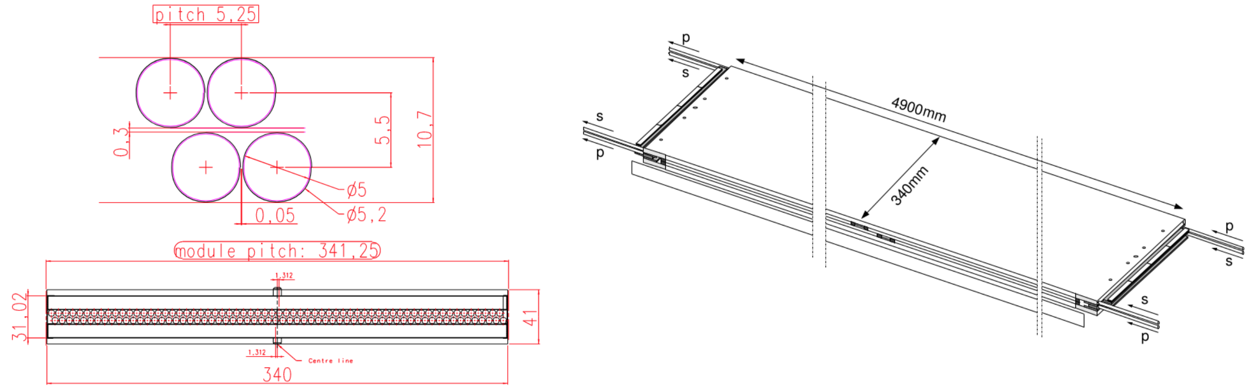
3.3.5. Outer Tracker

The farthest region from the beam pipe of the T stations is covered by the Outer Tracker (OT). It collects the information of the large area outside the IT acceptance where the particle flux is lower. The OT has a size of around 6 × 5 m width and height. OT is not made of silicon sensors due to their high cost, it is a drift-time gas detector based on arrays of straw-tubes instead. The straw-tubes are cylindrical with a diameter of 4,9 mm and are filled with a gas mixture of $Ar/CO_2/O_2(70\,\%, 28{,}5\,\%, 1{,}5\,\%)$ [3]. this configuration provides a spatial resolution of around $200\mu m$ and the required fast drift to reduce the spill-over[3] during the operation at 50 ns bunch spacing rate. In order to collect the charge of the ionized gas, there is an anode wire with a size of 25 μm located in the center of each tube. The straw-tubes were arranged in detector modules. Each detection station consists of four layers arranged in a x-u-v-x configuration, the modules containing x-layers are aligned vertically whereas those with u and v are tilted ±5°. The total active area consists of 5971 × 4850 mm^2. A cross-section of the modules is shown in Figure. **3-10**. The OT acceptance was determined by the

---

[3] Spill-over is a residual signal in the substation after 50 ns

requirement that occupancies should not exceed 10 % at a luminosity of 2×10^{32} cm^{-2} s^{-1}. The OT covers the acceptance of 300 mrad in the bending plane and 250 mrad in the non-bending plane.



(a)

(b)

Figure 3-10.: Cross section of the straw tube module [25] (a) and OT straw tube layer [2] (b).

Each station is split into two halves which are retractable on both sides of the beamline. Each half consists of two independent movable units of two half layers called C-frames. The OT C-frames are mounted in a stained steel structure (OT bridge) equipped with rails that allow the independent movement of all twelve C-frames [3]. The modules are composed of two staggered layers (monolayers) of 64 drift tubes each. There are two types of modules F-type and S-Type. The F-type modules are split longitudinally in the middle into two sections composed of individual straw tubes. Both sections are read out by the outer end. The two monolayers are split in a different position to avoid regions without sensitivity in the middle of the module. F-modules have an active length of 4850 mm and contain a total of 256 straw tubes. The S type modules are located below and above the beam pipe. These modules have about half the length of F-type modules. They contain 128 single straw tubes and are read out from the outer module end. A layer half is compound by 7 long and 4 short modules. The complete OOT active area is covered by 168 long and 96 short modules which comprises about 55000 single straw-tube channels. Errors in the mechanical alignment of the drift tubes can significantly degrade the track reconstruction. A single-cell resolution of about 200 μm requires that the drift tubes be aligned within an accuracy of 100 μm (1 mm) in the x (z) coordinate. Therefore, care was taken during each step of the detector construction and installation to minimize alignment errors. All detector C-frames which hold the modules were built with stringent requirements on the mechanical tolerances. During installation, the positions of all modules were surveyed and the C-frames positions have been adjusted accordingly [2].

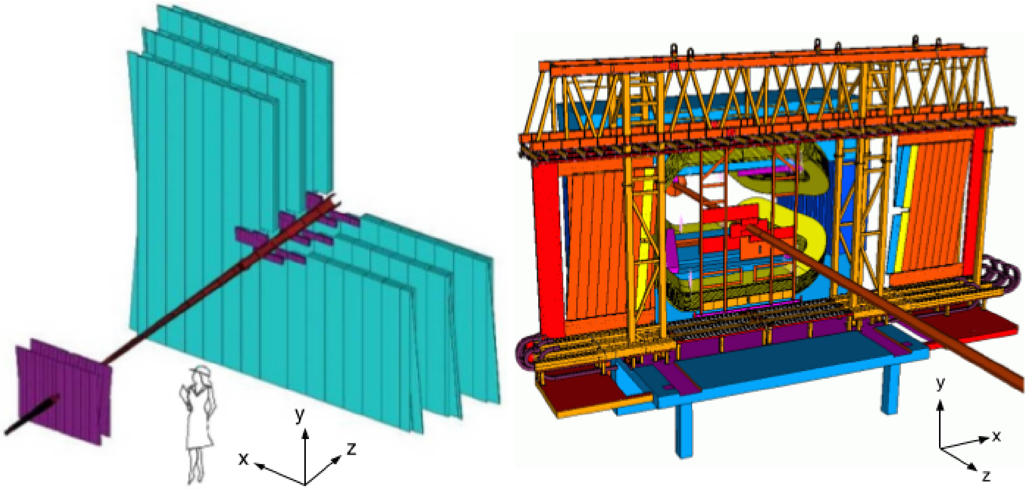


Figure 3-11.: OT straw-tubes arrangement in layers and stations (left) and overview of the bridge carrying C-frames (right). The C-frames on both sides of the beam pipe are retracted [2].

3.4. Particle Identification

Particle identification (PID) is a crucial attribute for all flavor physics experiments. The hadron identification, in particular, the separation between pions and kaons, is carried into effect by the Ring Imaging Cherenkov (RICH) system. Hadron PID plays a key role in most LHCb analysis. For instance, allowing the isolation of different two-body charmless b-hadron decays, the first observation of CP-violation in the B_s^0 system [26] or the measurement of the angle γ in the unitary triangle through the study of the $B^{\pm} \rightarrow DK^{\pm}$ decays [27]. The calorimeter system leads the photon and electron identification. It has a particular importance for radiative Penguin processes for example the $B^0 \rightarrow K^{*0}\gamma$ and $B_s^0 \rightarrow \phi\gamma$ decays [28]. Many heavy flavor decays of interest have muons in the final state. Since the LHCb muon system is the final substation which leads the discovery of the highly suppressed and important mode $B_s^0 \rightarrow \mu^+\mu^-$ [29]. PID is mandatory in flavor tagging for time-dependent CP-violation studies. Finally, the information provided by the calorimeter and muon systems are critical to the LHCb first trigger level.

3.4.1. Ring Imaging Cherenkov detectors (RICH)

The RICH system is fundamental for the goals of the LHCb experiment to separate pions from kaons in selected $b-$ and c-hadron decays over the momentum range 1,5 – 100 GeV. Due to the overlapping peaks in the invariant mass distributions for two-body hadronic B^0

and B_s^0 decays, which final states are $\pi^+\pi^-$, $K^+\pi^-$ and K^+K^-. PID is required to cleanly separate those final states. On the other hand, multibody final states such as $B_s^0 \to \phi\phi$ where $\phi \to K^+K^-$ would have severe combinatorial background without the $\pi - K$ separation. It is also important for particle-antiparticle tagging which relies on the identification of b and $\bar{b}$ hadrons. Finally, it facilitates the study of heavy-flavor baryonic states by providing an efficient proton identification. At large polar angles, the momentum spectrum is softer while at small polar angles the momentum spectrum is harder. Hence, two RICH detectors were designed to cover the full momentum range. RICH1 is located upstream of the magnet and it covers the low momentum range, $\sim 1 - 60$ GeV using silica aerogel and C_4F_{10} as radiators [30], and the angular acceptance of $25 - 300$ mrad. While RICH2 is located downstream of the magnet. it covers the high momentum range, $\sim$ 15 GeV up to 100 GeV using CF_4 radiators, and the acceptance of $15 - 120$ mrad. The tracks traversing RICH2 have higher momentum so they tend to be produced at lower angles [30].

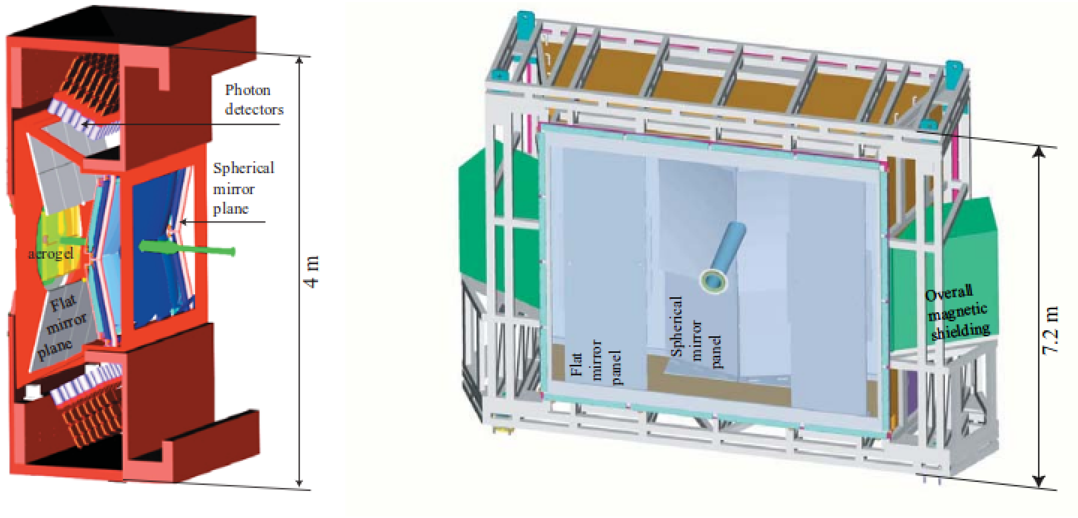


Figure 3-12.: Schematic layout of the RICH detectors RICH1 (left) and RICH2 (right) [30].

3.4.2. Calorimeter System

The calorimeter system can be divided into four components, two large electromagnetic (ECAL) and hadronic (HCAL) calorimeters and the two assisting detection layers, the Scintillating Pad Detector (SPD) and the Pre-Shower Detector (PS or PRS). The calorimeter system performs several functions; it selects the transverse energy from electron, hadron and photon candidates for the first trigger level (L0) and identifies them depending on the signatures in the calorimeter components. It also helps the reconstruction of π^0 and prompt photons with good accuracy which is important in flavor tagging and for B-meson decay studies. The set of constraints resulting from those functionalities defines the general structure and the characteristics of the calorimeter system and its electronics. Hadrons deposit the energy in the HCAL, while electrons and photons deposit the entire energy in the ECAL. It is possible to discriminate between electrons and photons using the combined information provided by two scintillating layers and a lead plate.

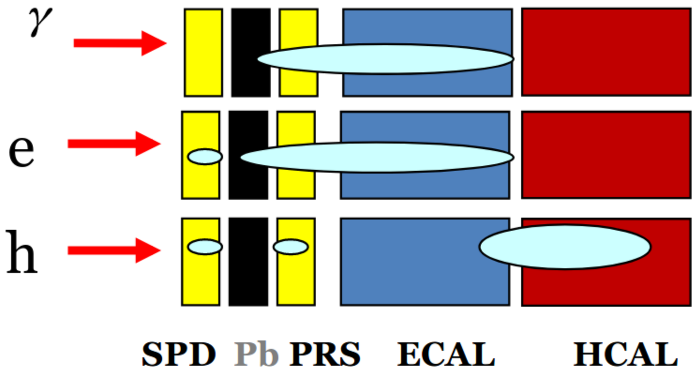


Figure 3-13.: Schematic view of the different signatures of photons, electrons and hadrons in the LHCb calorimeter system [31]

One of the most demanding tasks is the electron identification. 99 % of the inelastic pp interactions should be rejected while an enrichment factor of at least 15 in b events is kept, this goal is achieved by selecting electrons with high transverse energy E_T. A longitudinal segmentation of the electromagnetic shower direction rejects the majority of the background for charged pions. The lead thickness was chosen to improve trigger performance while maximizing the energy resolution. The addition of an SPD, in front of the PRS, is justified by the fact that it selects charged particles that reject most of the π^0's background at high E_T. Optimal energy resolution requires the showers coming from high energy electrons should be fully contained therefore ECAL thickness was chosen to be 25 radiation lengths. HCAL resolution does not impose a stringent hadronic shower containment hence 5.6 interaction lengths thickness was chosen due to space limitations. PRS/SPD, ECAL, and HCAL have variable lateral segmentation since the hit densities. PRS/SPD and ECAL are subdivided into three zones while HCAL has two zones with larger cell sizes justified by the dimension of hadronic showers.

Scintillating Pad/ Pre-Shower Detectors The first component of the calorimeter system comprises the layers of the SPD and PRS detectors which are located downstream the first muon station. Each detector consists of a 15 mm lead converter, that is sandwiched between two rectangular scintillator pads of high granularity with a total of 12032 detection channels [2]. The sensitive area of the detector is 7,6 × 6,2 m wide and high. SPD is slightly smaller than PRS due to projectivity requirements. Particle densities varies by two magnitude orders comparing the central and peripheral regions, therefore the detectors adopt a subdivision into inner middle and outer sections. The inner section contains 3072 cells arranged in 4×4 cm^2, 3584 cells in 6×6 cm^2 for the middle section and 5376 cells in 12×12 cm^2 for the outer section [2].

The cells are packed in $\sim 48 \times 48$ cm^2 detector units grouped in super-modules. The super-modules are $\approx$ 96 cm width and $\approx$ 7,7 m height arranged in detector modules from 2 rows and 13 columns. The diameter of the WLS fiber groove is a few millimeters smaller than the tile size. The square structure of a pad is cut from a 15 mm thick scintillator plate, and the scintillator surface is polished to reach the required optical quality. In order to maximize the light collection efficiency, WLS fibers are coiled and placed into a ring groove that is milled in the body of the cell. The rectangular cross-section of the groove is 4,1 mm deep and 1,1 mm wide. The groove contains 3.5 loops of WLS fiber. The number of loops was chosen to achieve an overall optimization of the light collection efficiency [32] and the signal formation [33]. Two additional grooves are milled in the scintillator allowing both ends of the WLS fiber to exit the plate.

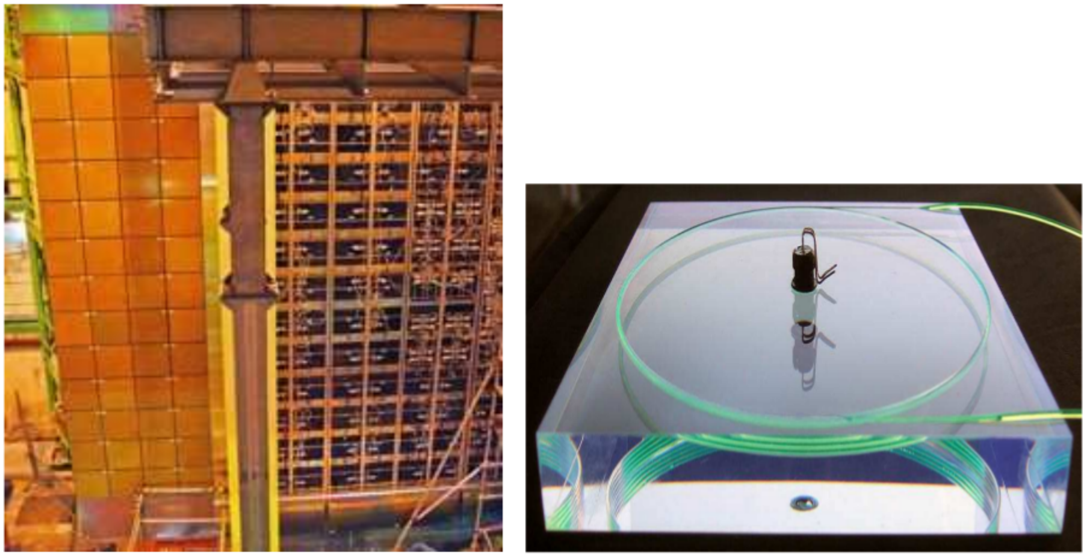


Figure 3-14.: Front view of one half of the SPD/PS installed in the LHCb experimental hall (left). Individual scintillator pad with the WLS fibre layout and the LED housing in the middle (right) [2].

The Electromagnetic Calorimeter The electromagnetic calorimeter is located next to the PRS. The ECAL harness the *shashlik* calorimeter technology. A sampling scintillator/lead structure readout by plastic WLS fibers has been chosen for the ECAL not only at LHCb but also at some other experiments[4]. The decision to use shashlik technology was made taking into account its modest energy resolution, the fast time response, an acceptable radiation resistance and the reliability and performance in its technology which was proved by other experiments. The design energy resolution of $\sigma_E/E = 10\,\%/\sqrt{E} \oplus 1\,\%$ (E in GeV) results

---

[4] For instance: HERA-B and PHENIX

in a B mass resolution of 65 MeV for the $B \to K^{*}\gamma$ penguin with a high-E_T photon and of 75 MeV for $B \to \rho\pi$ with the π^0 mass resolution of ~ 8 MeV2 [2].
The ECAL is placed 12.5m downstream of the interaction point. Its outer acceptance match those of the tracking system $\theta_x < 300$ mrad and $\theta_y < 250$ mrad while the inner acceptance is limited, to the region $\theta_{xy} > 25$ mrad around the beampipe, by the substantial radiation dose. As the hit density varies with the distance from the beampipe, the calorimeter is subdivided into inner, middle and outer regions. A module is built from alternating layers of 2 mm thick lead, 120 μm thick, white, reflecting TYVEK [5] paper and 4 mm thick scintillator tiles. The light from the scintillator tiles is absorbed, re-emitted and transported by WLS fibers which traversing the entire module. The fibers are looped to make them traverse modules twice in order to improve the light collection efficiency. For the inner modules about 3100 photoelectrons per GeV of deposited energy, 3500 for the middle and 2600 for the outer modules. The differences are due to the differing WLS fiber density in the inner, middle and outer modules and to the cell size differences which influence the probability of photon absorption in the scintillating tiles.

Figure 3-15.: Downstream view of the ECAL installed (but not completely closed) with the exception of some detector elements above the beam line. Outer, middle and inner type ECAL modules (right) [2].

The Hadronic Calorimeter The hadronic calorimeter (HCAL) is placed between the ECAL and the muon system (M2-M5) at a distance from the interaction point of 13,33 m with dimensions of 8,4 m in height, 6,8 m in width and 1,65 m in depth. The HCAL is also a sampling device made of iron (as absorber) and scintillating pads (as active material). The special characteristic of this sampling structure is the orientation of the scintillating tiles which run parallel to the beam axis. The tiles are interspersed with 1 cm of iron in the lateral direction, while in longitudinal direction the length of the tiles corresponds to the hadron interaction length in steel (λ_I).

---

[5] TYVEK of type 1025D used, a product of E.I. du Pont de Nemours and Company.

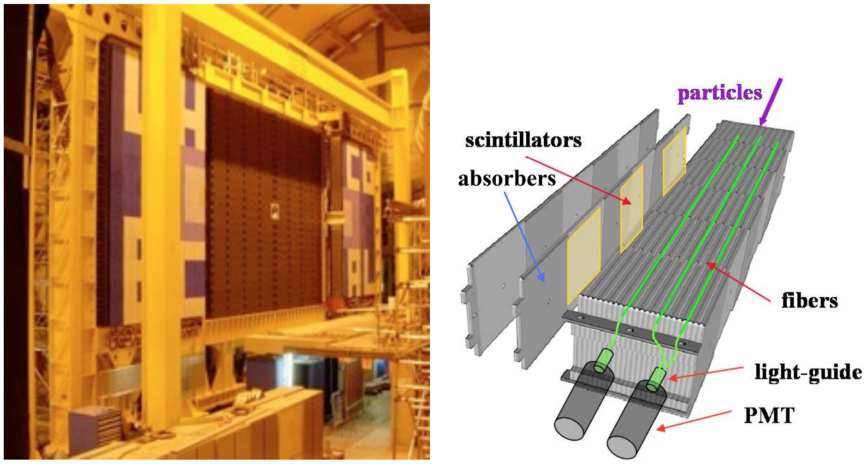


Figure 3-16.: View from upstream of the HCAL detector installed behind the two retracted ECAL halves in the LHCb cavern (left). A schematic of the internal cell structure (right). The exploded view of two scintillator-absorber layers illustrates the elementary periodic structure of a HCAL module [2].

The light is collected by WLS fibers traversing the detector towards the backside where photomultipliers are tubes (PMT) are housed. The HCAL is segmented transversely into square cells of size 131,3 mm (inner section) and 262,6 mm (outer section) [34, 35]. The calorimeter is divided into two different zones which dimensions are ±2101 mm and ±4202 mm in x and ±1838 mm and ±3414 mm in y for the inner and outer sections, respectively.

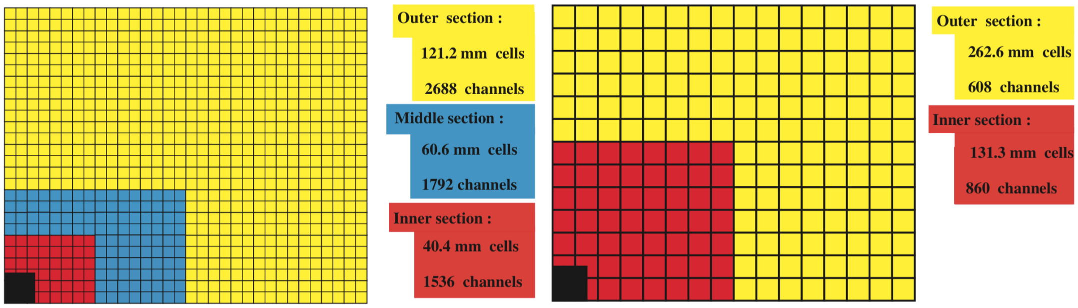


Figure 3-17.: Lateral segmentation of the SPD/PS and ECAL (left) and the HCAL (right). One quarter of the detector front face is shown. In the left figure the cell dimensions are given for the ECAL. [2].

The assembled HCAL is shown in Figure **3-16**(left). The absorber structure, shown in figure Figure **3-16**(right), is made from laminated steel plates of only six different dimensions that are glued together. Identical periods of 20 mm thickness are repeated 216 times in the module. One period consists of two 6 mm thick master plates with a length of 1283 mm and a height of 260 mm that are glued in two layers to several 4 mm thick spacers of 256,5 mm in height and variable length. The space is filled with 3 mm scintillator [2]. The comparison between the granularity of the calorimeter system, The cell sizes adopted for the SPD, PS and ECAL and HCAL and the region sizes are shown in Fig **3-17**. The position of the beampipe is indicated by the black square.

3.4.3. Muon System

Muon triggering and offline muon identification are fundamental requirements of the LHCb experiment [36, 37]. Muons are present in the final states of many CP-sensitive B decays, in particular $B_d^0 \to J/\psi(\mu^+\mu^-)K_s^0$ and $B_s^0 \to J/\psi(\mu^+\mu^-)\phi$ which are called the golden decays. Besides, The tag of initial state flavor of B neutral mesons is provided by muons coming from semi-leptonic b decays. Additionally, the study of rare B decays such as the flavor-changing neutral current decay, $B_s^0 \to \mu\mu$, may reveal new physics beyond the Standard Model [38].

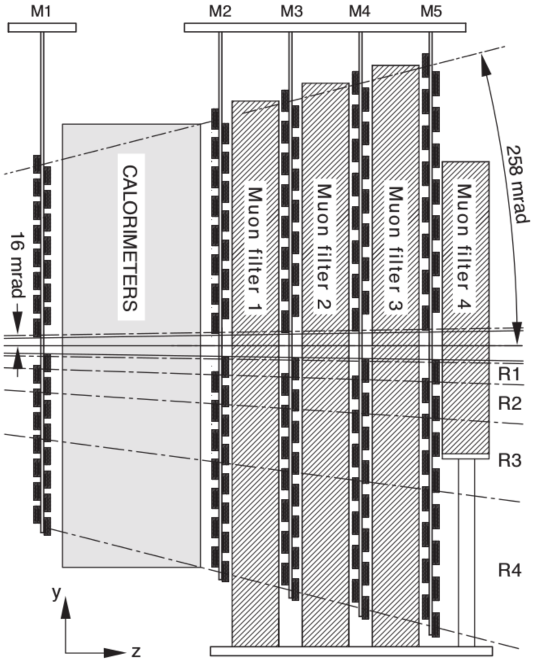


Figure 3-18.: Side view of the muon system [2].

The muon system is a compound of five stations (M1-M5) as sown in Fig **3-18**. The first station is located upstream of the calorimeters, its purpose is to improve the transverse momentum resolution of muons in the trigger due to the increment of the multiple scattering after muons traverse the calorimeters. The stations M2-M5 are located downstream of the calorimeters. These stations select penetrating muons and are interleaved with iron plates of 80 cm thick which act as absorbers. The full system covers a total area of 435 m^2 where

1380 chambers are placed. The inner and outer angular acceptances are 20(16) mrad and 308(258) mrad in the bending (non-bending) plane respectively. The minimum momentum of a muon to cross the five stations is $\sim$ 6 GeV since the total absorber thickness, including calorimeters, is $\sim$ 20 interaction lengths [2].

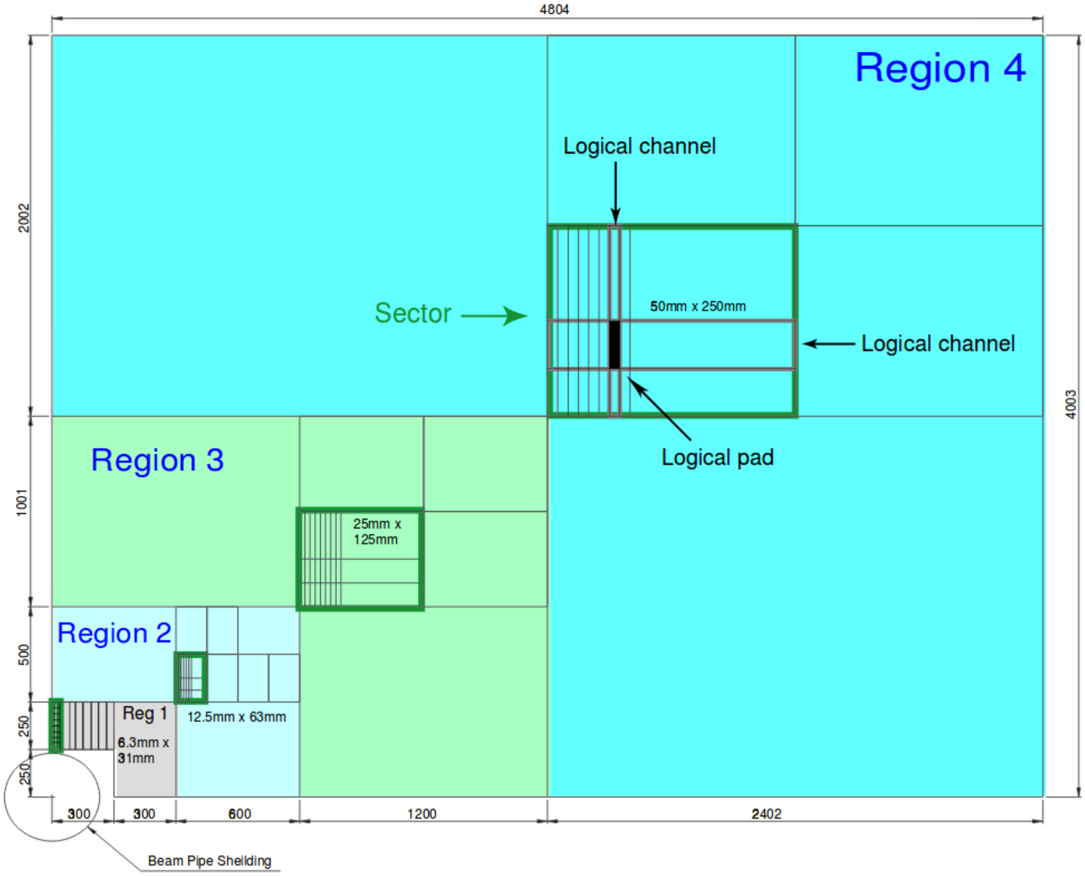


Figure 3-19.: Front view of one quadrant of a muon station showing the four regions of different granularity and their dimensions. [36].

The information from the muon stations is obtained by partitioning the muon stations into rectangular logical pads whose dimensions define the resolution in x and y coordinates. M1-M3 stations have a high spatial resolution in the bending plane. These are used to define track direction and to calculate P_T of the muon candidate with a resolution of 20 %. M4 and M5 stations have limited spatial resolution. They were built to identify penetrating particles. The operation, in absence of one muon station (M1, M4 or M5) or with missing chamber parts (worse resolution and performance), is allowed. The layout of the station M2 is shown in Figure **3-19**. Each muon station is divided into four regions with different logical pad dimensions due to the granularity requirements. This geometry ensures, particle flux and occupancy, to be roughly the same over the four regions for a given station. Region and pad sizes scale by a factor two from one region to the next. The x dimensions for the logical pads in the M1-M3 stations are determined to achieve the require muon P_T resolution for the L0 trigger. The pad y dimensions are determined by the required background rejection on triggers that do not point to the interaction region [2]. The logical pad x/y ratios are; 2,5 for the M1 stations, 5 for M2 and M3, and 1,25 for M4 and M5. The total number of logical pads in the muon system is 55296

3.5. Trigger System

The LHCb experiment operates at an average luminosity of 2×10^{32} cm^{-2} s^{-1}, much lower than the maximum LHC designed luminosity. This is intentionally reduced in order to reduce the radiation damage to the components. Moreover, to limit the number of interactions per bunch to help the trigger and reconstruction algorithms. One of the main limitations of the current experiment is that the collision rate must be reduced to the readout rate of 1,1 MHz within a fixed latency [39]. This rate corresponds to the maximum that can be written to storage. The trigger system was designed to select only potentially interesting events for future physics analysis. The crucial constraint for the trigger is to find an algorithm to reject or retain the events within a very short time window. LHCb trigger system was made with a hardware and a software component to keep flexibility and achieve the fast decision. At the mentioned luminosity the bunch crossing is expected to contain about 100 kHz of $b\bar{b}$-pairs. Nevertheless, only about 15 % of these events will include at least one B meson with all of its decay products within the acceptance of the detector [2]. Due to their large mass, B meson decays often produce particles with large transverse momentum (P_T) and energy (E_T) respectively. Therefore, the **Level-0 (L0)** trigger attempts to reconstruct whether the highest E_T hadron, electron and photon clusters in the calorimeters or the two highest P_T muons in the muon chambers. Besides, the pile-up system in the VELO estimates the number of primary pp interactions per bunch crossing. The calorimeters calculate the total observed energy and estimate the number of tracks based on the number of hits in the SPD [2]. The output rate of the L0 consists of 450 kHz assigned to hadrons, followed by 400 kHz for muons and 150 kHz for electrons/photons [3]. The second trigger stage is the software **High Level Trigger (HLT)** which processes the events selected by L0. the HLT consists of two components, HLT1, and HLT2. The HLT1 was designed to confirm the decision taken by L0 by performing a partial event reconstruction around the events selected by the hardware trigger. furthermore, some track trigger configurations (trigger lines) are added to look for topological signatures of B decays, such as high momentum tracks with large impact parameters. The output of the HLT1 is around $50 - 80$ kHz. This allows the required event rate to the HLT2 for the full event reconstruction close to the offline quality reconstructions. The HLT2 comprises a variety of algorithms written on $C++$ which reduce the trigger rate to 5kHz. These events are sent to store to be used in posterior physics analysis. See Ref. [2, 40, 39] for more details.

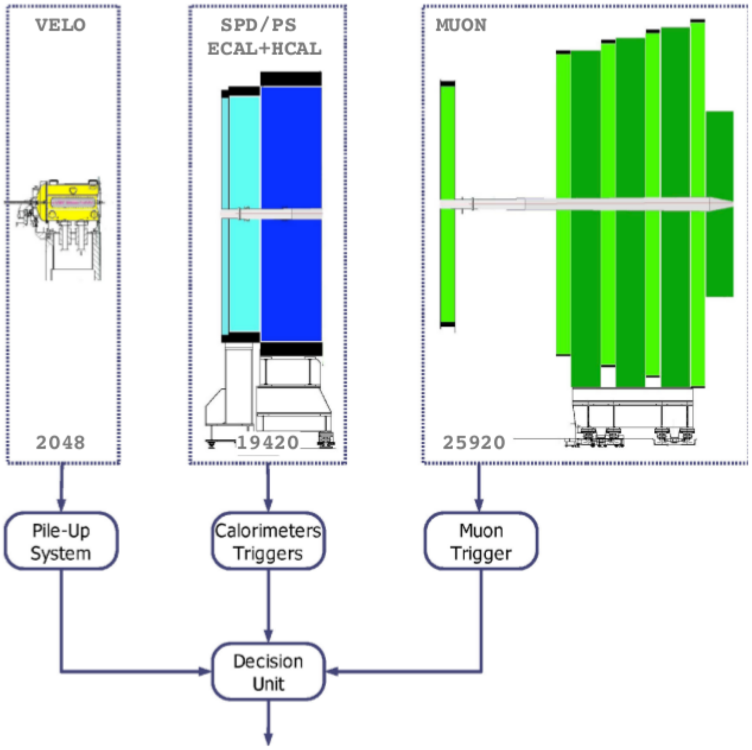
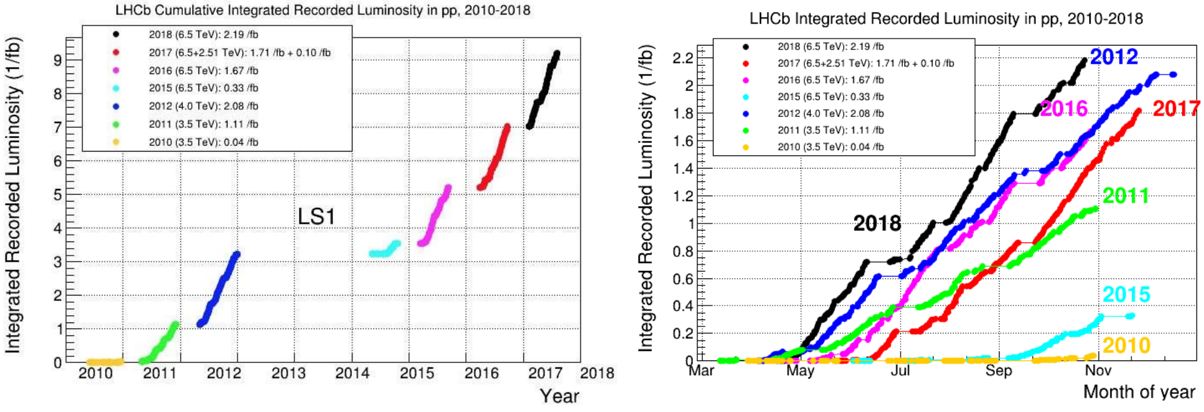


Figure 3-20.: Overview of the Level-0 trigger. Every 25 ns the pile-up system receives 2048 channels from the pile-up detector, the Level-0 calorimeters 19420 channels from the scintillating pad detector, PRS, electromagnetic and hadronic calorimeters while the Level-0 muon handles 25920 logical channels from the muon detector [2].

3.6. Data Taking

Since the LHC has started to delivered pp collisions at a center-of-mass energy of $\sqrt{s} = 900$ GeV in 2009 [3]. The LHCb detector has collected a huge amount of b-$\bar{b}$ pairs with several detector configurations and energies. The main data taking periods, RunI and RunII, of proton-proton collisions through the years 2010-2018, are summarized in this section. RunI and RunII correspond to the period between the years 2010-2012 and 2015-2018 respectively. During the RunI, the first pp collisions at the center of mass energy of $\sqrt{s} = 7$ TeV were delivered and recorded in 2010. The ideal conditions were achieved in this period in order to particle multiplicity measurements. The LHC beams were filled with less than 10 proton bunches per beam. The instantaneous luminosity was around $\mathcal{L}_{inst} = 2 \cdot 10^{28}$ cm^{-2} s^{-1} with the number of interactions per bunch crossing of less than 0.1. Therefore, There is no pile-up during this period. By the end of the year, the number of bunches was increased up to ≈ 350. It also increases the luminosity up to $\mathcal{L}_{inst} = 1{,}2 \times \cdot 10^{32}$ cm^{-2} s^{-1}. The entire data collected during 2010 was $\mathcal{L} = 0{,}04$ fb^{-1}. The conditions during 2011 were stable, LHC worked at the instantaneous luminosity $\mathcal{L}_{inst} = 3{,}8 \times \cdot 10^{32}$ cm^{-2} s^{-1} with an average number of pp interactions of 1.5 and 1380 fill bunches. The total size of the recorded data was $\mathcal{L} = 1{,}11$ fb^{-1}. In 2012, the center of mass-energy was increased up to $\sqrt{s} = 8$ TeV with an average interaction rate of 1.7. The other conditions were comparable to those of 2011. The total recorded luminosity was $\mathcal{L} = 2{,}08$ fb^{-1}.



(a) (b)

Figure 3-21.: Summary of the recorded luminosity at the LHCb detector during the pp runs in the years 2010 - 2018 [1].

During the RunII, LHCb delivered and recorded luminosity in 2018, the last year of Run 2, were $\mathcal{L} = 2{,}461$ fb^{-1} and $\mathcal{L} = 2{,}185$ fb^{-1} respectively. The corresponding numbers for 2015 are $\mathcal{L} = 0{,}371$ fb^{-1} and $\mathcal{L} = 0{,}328$ fb^{-1}; for 2016 $\mathcal{L} = 1{,}906$ fb^{-1} and $\mathcal{L} = 1{,}665$ fb^{-1}, for 2017 $\mathcal{L} = 1{,}987$ fb^{-1} and $\mathcal{L} = 1{,}817$ fb^{-1} [1].

Conditions	**RunI**	**RunII**
$\sqrt{s}$	7-8 TeV	13 Tev
Bunch spacing	50 ns	25 ns
Pile-up	1,7	1,3-2,4
Output rate HLT1	80 kHz	150 kHz
Output rate HLT2	5 kHz	12,5 kHz
Time budget HLT1	20 ms/evt	35 ms/evt
Time budget HLT2	150 ms/evt	350 ms/evt

Table 3-1.: Comparison of some of the key elements in the conditions faced in Run I and expected in Run II at LHCb [41].

4. Event generation and track reconstruction software at LHCb

The performance of the LHCb detector is constantly studied through detailed Monte Carlo (MC) simulations, which produces raw samples, as identical as possible to real data. The pp collisions are simulated with the PYTHIA program [42]. The generated tracks traversing the whole detector and its surrounding material are generated with the GEANT4 package [43], which simulates the geometry, material composition of the LHCb detector, detector response, resolution, noise, etc.
The software programs used for the performance studies execute the following tasks [44]:

- Generation of the event (GAUSS).
- Tracking of the particles through the detector and digitization (BOOLE).
- Simulation of the trigger decision (MOORE).
- Reconstruction of the event including track finding and particle identification (BRUNEL).
- Offline selection of a specific b-hadron final state (DaVinci).

LHCb has developed several packages to perform each of the previous tasks individually, processing the events as if they were from real data without employing any information from the MC truth. The MC truth information is only used at the end of the reconstruction chain to test the real performance of the LHCb detector. In this chapter, we discuss the characteristics of the MC samples, the software employed to generate them and the whole reconstruction chain.

4.1. Monte Carlo event generators

Monte Carlo event generators are essential for all particle physics experiments and the analysis carried out. They are used not only to make predictions and preparations for future experiments but also to test the performance of the current ones. Unfortunately, MC tools in high energy physics are often used as "black boxes"whose outcome is treated as data. As an example of the importance of MC, the majority of the Higgs discovery plots rely very

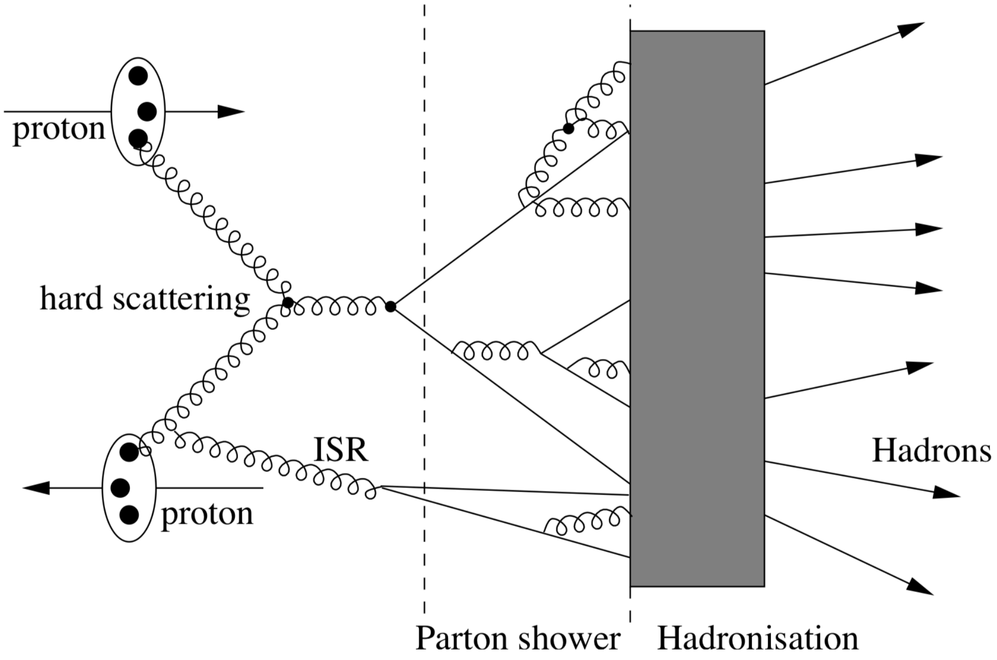


Figure 4-1.: Schematic picture of a pp collision [49].

strongly on MC predictions, to set limits on Higgses in certain parameter space regions as well as to discover them [45]. This motivates and justifies the use of MC generators for future physics discoveries and predictions. PYTHIA generator is usually used to simulate the extremely complex pp collision as shown in figure **4-1**, where single simulations steps are introduced.

4.1.1. The Pythia generator

PYTHIA is a general-purpose event generator to simulate the physics processes in and right after pp, $p\bar{p}$ and ee collisions. Its use has been extended in all the high energy physics community over the years. PYTHIA combines the two interactions, hard and soft scattering processes. hard scattering is calculated using perturbative QCD, while it makes a phenomenological description for soft interactions. The transition between the regimes is tuned by the momentum transferred ($P_{T_{min}}$). There are two different implementations of the PYTHIA generator. PYTHIA 6.4 [46, 42] and PYTHIA 8.2 [47, 48] which are written in FORTRAN and C++ respectively. The recommended version for LHC studies is PYTHIA 8.1 [3] although PYTHIA 6.4 has more implemented features. These simulate the required event types such as minimum-bias, $b\bar{b}$ inclusive, signal events among others.
The full collision process can be divided into several steps:

- Two partons coming from the protons in each beam hardly interact with each other.

This process is calculated to leading order in perturbative QCD.

- In the parton shower model, next to leading order perturbative processes are approximated instead of calculating them directly. This is done by adding initial and final state showers to the hard parton scattering which modifies the leading order processes.
- The multiple interaction model in PYTHIA is added to describe semi-hard processes that may occur among the other partons in the two interacting protons.
- The outgoing quarks are colored. Therefore they cannot leave the interaction region freely. They must fragment into colorless hadrons due to confinement effects. PYTHIA uses the Lund string fragmentation model [50].
- Finally hadronization processes are calculated in which all unstable particles decay until only stable particles are left.

Three processes contribute to the $b\bar{b}$ production cross-section, pair creation, flavor excitation, and gluon fusion. A brief description of these processes is depicted in Figure **4-2**.

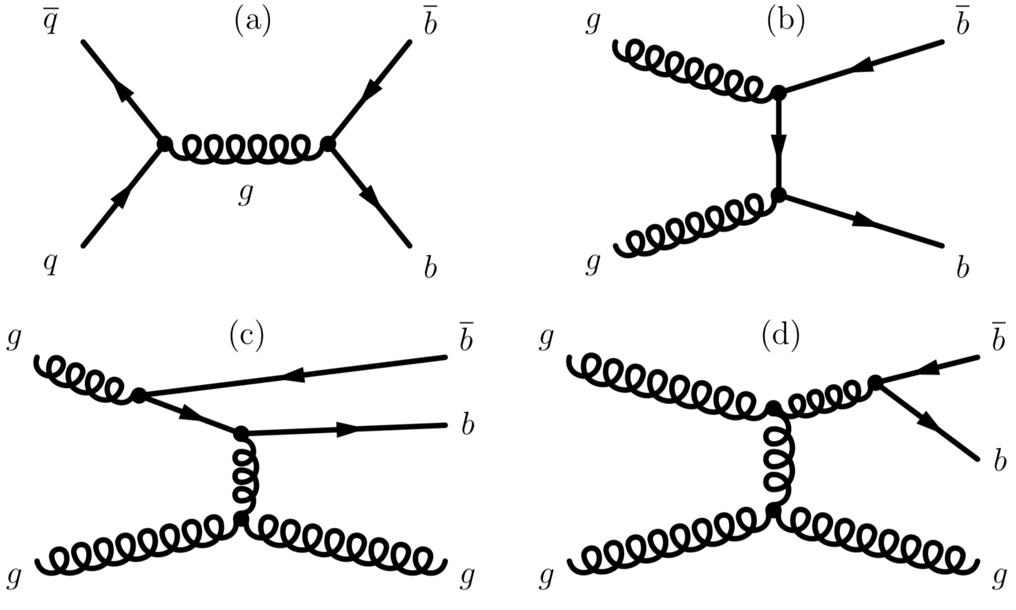


Figure 4-2.: Examples of Feynman diagrams for B production. The two leading- order diagrams are pair creation through quark-antiquark annihilation (a) and gluon fusion (b). The next-to-leading order diagrams show examples of flavour excitation (c) and gluon splitting (d) [49].

Pair creation It is a leading order interaction in which hard 2 → 2 processes are calculated as quark-antiquark annihilation ($q\bar{q} \to b\bar{b}$) and gluon fusion ($gg \to b\bar{b}$). The dominant contribution, between these two, is gluon fusion.

Flavour excitation A virtual b quark from one of the protons is put on the mass shell by the scattering with a parton from the other proton ($bq \to bq$) and ($bg \to bg$). Since the b quark is not a valence quark, and thus comes from the sea, there must be a $\bar{b}$ quark from a previous branching $g \to b\bar{b}$. Likewise, the charge-conjugated process, with the b and b interchanged, occurs with the same probability.

Gluon splitting Instead of in the hard scattering, the $b\bar{b}$ pair is created in a $g \to b\bar{b}$ branching in the initial or final state shower.

Production process		Contribution (%)
Pair creation	($q\bar{q}$ annihilation)	$0{,}20 \pm 0{,}04$
	(gluon fusion)	$14{,}4 \pm 0{,}4$
Flavour excitation		$60{,}3 \pm 0{,}5$
Gluon splitting		$25{,}0 \pm 0{,}4$

Table 4-1.: Contribution of the different B production processes in Pythia with at least one B in the LHCb acceptance [49].

The classification of these processes is given by the number of b and $\bar{b}$ involved in the leading order hard interaction [49], 2, 1, 0, respectively. Pair creation contributions are calculated with leading order perturbation theory, while flavor excitation and gluon splitting are calculated with a phenomenological description that generates large uncertainties. It is well known by PYTHIA simulations of protons at 14 TeV that flavor excitation and gluon splitting are non-negligible contributions to the total $b\bar{b}$ cross-section. It is important to include the appropriate kinematic distributions of the outgoing particles due to each production mechanism. After the pp collision most of the B mesons are produced in the forward or backward cone as shown in figure **3-2**. The boost is due to the fact that $b\bar{b}$ pair are produced in collisions in which a parton x from one proton carries a large fraction of the total momentum and another parton, from the other proton, carries a small fraction. The difference in momentum increases directly proportional to the center of mass energy. About one third of the produced $b\bar{b}$ pairs fall in the $2 < \eta < 5$ region.

The multiple parton interaction processes are caused by the other partons which produce soft collisions. These contribute to the particle multiplicity. The default parton interaction model[1] employed by PYTHIA assumes varying impact parameters of the colliding protons and a Gaussian matter distribution inside the proton which results in a particle multiplicity governed by the parameter $P_{T_{min}}$. The momentum transferred determines how hard the interaction is which is expressed in terms of the transverse momentum. The harder the process is, the more of the available energy is absorbed. The remaining energy is available for the other partons. The number of parton interactions can be calculated by taking a parton

---

[1] The different models are explained in detail in Ref. [51]

out and adjusting the leftover energy. The parton cross-section, σ_{part}, diverges when $P_T \to 0$ due to the fact that PYTHIA assumes the partons to move freely, instead of bounding them within the proton as it occurs in nature. Therefore it needs to be regularized with a cut-off setting a minimum transverse momentum. The cut-off is physically interpreted as the effective, inverse color-screening distance [49]. Finally, the outgoing quarks have to hadronize to mesons and baryons.

4.2. The LHCb Software Framework

Due to the complexity of the processes in high energy physics, the simulation and reconstruction task has been divided into several steps. LHCb has designed a specific package to perform each phase. The whole software is based on the GAUDI framework [52, 53] which provides the necessary interfaces and services for building high energy physics experiment frameworks in the domain of event data processing applications [54]. The physics analysis carried out by the LHCb experiment. The MC simulations are crucial to build and test different analysis strategies. They are also important to test the performance of the detector (e.g. efficiencies). A brief description of the main software packages is given as detailed below.

4.2.1. Generation and simulation phase - GAUSS Software

The GAUSS framework [55] is the platform to interface to various specialized packages available in the particle physics community for the generation and simulation phases to cope with those processes. GAUSS uses two different packages to generate the events, PYTHIA [47] and EvtGen [56]. PYTHIA generates the pp collision taking into account the interactions and decay particles. It stops when a ”stable”particle for PYTHIA is reached in the decay three (a meson or a hadron), while EvtGen takes care of the decay of the B mesons and other resonances relevant in high energy physics. The generations implemented in GAUSS are; forced fragmentation, repeated hadronization, and plain Pythia [57]. In the forced fragmentation, the b quark in the event is forced to hadronize with the correct light quark to give the correct the b hadron type of the signal decay mode. Repeated hadronization is used when the event contains a b quark, It is hadronized several times until the correct b hadron type is found. Finally, the plain Pythia generates and rejects full events until the correct b hadron type is found. After the generation sequence, the generated particles are transformed into G4Particles to be processed by GEANT4 [43] in a simulation sequence. The conversion algorithm ignores certain particles e.g. partons or particles that are unknown for Geant4 like B^{**}. If a particle is converted, the whole decay chain is converted independently of the status code or the type of the particle. A particle is converted if it satisfies the following conditions [57]:

- The status of the particle is 1, 2, 888, 889 [2], and,
- the particle has no production vertex, or,
- the particle is a hadron, a lepton, a nucleus or a photon and has only one mother particle which is not a hadron, neither a lepton, a nucleus nor a photon, or the status of the mother particle is 3, or,
- the particle is a hadron, a lepton, a nucleus or a photon and has several mother particles or no mother particle but a production vertex (which is the case for particle guns).

GAUSS contains several packages to simulate the LHCb detector response to the simulated particles traversing it, Gauss Tracker, Gauss Calo, and Gauss RICH. The **Gauss Tracker** package contains a sensitive implementation for all LHCb tracker devices. These implementations generate hits where the following information is stored: the energy deposition, the entry and exit points, the time of flight, and the ID if the track that created a given hit (important to set the appropriate links between hits and MC particles). Furthermore, VELO sensitive detector provides the sensor number of the hit, while the Muon sensitive detector provides Chamber ID and Gap ID. The **Gauss Calo** package contains the sensitive implementation for the LHCb calorimeter system which produce hits containing the following information: the energy deposited by charged particles during a Geant4 step, the time when the energy was deposited, the Cell ID where the energy was deposited and the link to the track that generated the hit. The **Gauss RICH** stores the following information: PDG code of the charged track, photon energy at the time of its production, Cherenkov angle θ and ϕ, PDG mass of the charged track, three momentum of the charged track and, Pre-step and Post-step point locations of the charged track. The whole information from the generation and simulations steps are stored in SIM files.
The list of event types available for production and their documentation can be found in the DecFiles web pages [58].

4.2.2. Digitization phase - BOOLE Software

The final stage of the LHCb detector simulation is performed by the BOOLE [59] digitization software. BOOLE applies the detector response to the hits previously generated in the sensitive detectors by the Geant4 in the GAUSS stage. Moreover, hits coming from spillover events and LHCb background are added at this stage. The digitization step includes not only the simulation of the detector response but also of the readout electronics as well as the L0 hardware performance. The output is digitized data in a DIGI file format that mimics the real data coming from the detector.

---

[2] The list of the status codes and their description can be found in Ref. [57].

4.2.3. Trigger emulation phase - MOORE Software

The MOORE software emulates the trigger response for simulated data using identical configurations as during data taking. MOORE will either run in the online trigger farm processing online data from the LHCb DAQ system or offline starting from real data or from the output of the detector digitization application (BOOLE) [60]. The LHCb trigger is configured via a unique key the *Trigger Configuration Key* or TCK which defines the software version, the specific sequence of algorithms and the selection requirements applied to each trigger line [61]. A TCK is identified by a 32-bit value. A given TCK configuration can only be run with a specific MOORE version. Each TCK contains HLT Tracking, PV finding, and Calorimeter reconstruction sequences. The HLT Tracking is divided into HLT1 and HLT2 tracking sequences. A brief description of the configurations is described below.

HLT1 Tracking

1. VELO tracks are built using the VELO pattern recognition [23]. This is going to be described in the next chapter.
2. If a minimum momentum of 6 GeV is detected in the muon stations, An algorithm is run to match the VELO tracks and the muon station measurements. The condition in the x and y windows are 200 mm from the predicted track.
3. A validation with the TT information may be performed on the forward tracks. The forward track algorithm will be discussed in the next chapter.
4. VELO tracks are upgraded to forward tracks if they fulfill the conditions: The number of hits of the track needs to be above 12 or for OT-only tracks, the number of OT hits needs to be above 14. This is just an initial condition to confirm the L0 predictions.
5. Track is fitted using a Kalman filter based fitter.
6. Optionally, a long track may be upgraded to a muon candidate; this step requires decoding the muon stations.

Besides, the VELO-only track fit method essentially performs a straight line fit.

HLT2 Tracking The tracking sequence in HLT2 is closer to the offline version than the HLT1 one.

1. VELO tracking algorithm is performed with the HLT2 requirements which are more restrictive than HLT1 ones.
2. Forward tracking is performed and VELO tracks are upgraded to forward tracks with the full forward algorithm which performs additional steps to prepare the hits and

perform ghost killing. Two configurations can be used in the forward tracking procedure first loop and second loop with requirements; in both cases, it is required again that either (a) the number of hits of the track is above 12 or (b) for OT-only tracks, the number of OT hits is above 14. Second loop reconstruction is typically executed after some pre-selection process that lowers the rate and it was designed to recover lower momentum particles that were not picked up in the first instance due to timing limitations.

3. Seeding[3] on unused hits in the T-stations, a matching of the VELO and seed tracks and a clone killing are performed. This allows to obtain long tracks. Additionally, the seed tracks can be used to build downstream tracks.

4. A Kalman fit is performed.

5. Particles are made from the Kalman-fitted tracks. In the case of muons, the muon ID information which is determined when needed.

PV reconstruction Primary vertex finding and reconstruction are performed from the VELO tracks by making use of the forward and backward tracks.

Calorimeter reconstruction The calorimeter is used for the reconstruction of neutral electromagnetic particles, γ and π^0. All neutral clusters are associated to photons that are reconstructed using the SPD information and their corresponding transverse energy (E_T) deposition on the ECAL. neutral pions typically decay into photons which can be resolved if the $E_T < 200$ GeV. Otherwise, the π^0 is called merged and it cannot be resolved. Merged pions are not distinguished at the trigger level. However, these are built from a single calorimeter deposition with a mass window of 60 MeV around their nominal mass, while resolved pions are build from two calorimeter depositions of $E_T > 200$ MeV with a mass window of 30 MeV around their nominal mass.

MOORE writes the information on a DIGI file which emulates the raw data that has been either simulated or triggered by the LHCb detector. This file contains the full detector response which needs to be reconstructed and prepared before to perform physics analysis. The lowest level trigger, L0, can be seen as part of the digitization job, as it is fully implemented in (programmable) electronics. On the other hand, the higher level triggers (HLT) can be seen as part of the reconstruction job, as they are executed on a dedicated processor farm, running the same software that is used in MC data [49].

---

[3] Seeding is the algorithm in charged of the reconstruction using only the seed (IT+OT) hits.

4.2.4. Offline reconstruction phase - BRUNEL Software

The raw data, or the trigger output, coming from the detector data taking, or the MOORE output, respectively are reconstructed by using the BRUNEL [62] software. Its main tasks are the reconstruction of tracks and the identification of particles. BRUNEL includes algorithms to reconstruct the properties and of all particles that are visible to the detector such as three-momentum, invariant mass, trajectories, etc. The output of BRUNEL is a Data Summary Tape (DST) file. The algorithms involved in the reconstruction program use calibration and alignment constants to make corrections in the response of the detector and its electronic components. Calibration, alignment data, and detector conditions are stored in a distributed database.

4.2.5. Analysis phase - DAVINCI Software

The DAVINCI [63] software is an event selection algorithm that comprises several physics analysis packages. Firstly, particles are selected to form a group of standard particles that are constructed from tracks and particle identification objects. Secondly, the primary vertices are calculated and fitted. Finally, specific b-hadron decays of interest are selected by making the appropriate particle combinations. The output of DAVINCI is typically smaller data samples on ROOT [64] file format which only contains a selection of events and particles used for the individual physics analysis.

4.3. Data flow

This section describes the dataflow of the LHCb computing model for all stages in the processing of the real and simulated LHCb events [65] as is shown in figure **4-3**. The role of several Tier centers as well as the computing model at LHCb. There are several stages in the processing of event data. Each phase is followed by others in a specific sequence, but some of them may be repeated a determined number of times. Although raw data and simulated data are reconstructed by the same algorithms, they follow different paths in the data processing farms. **Raw data** is produced via the Event Filter farm of the online system. The trigger system collects events of interest by processing data coming from the LHCb subsystems and applying HLT algorithms to make calibration corrections during the reconstruction of the physical properties of particles. The data not selected by the trigger to be stored cannot be recovered anymore. On the other hand, the **simulated data** are produced by a detailed MC model that contains not only the same information that a raw data file but also the extra truth information. the truth information contains the simulated hits on the detector which is used to record the physics history of a specific event and its relationships between hits and particles. rawMC data is only stored if events pass the trigger simulation.

The raw data, whether real or simulated, must then be reconstructed in order to provide physical quantities [65]: calorimeter clusters to measure the energy of electromagnetic and hadronic showers, hits in the tracking stations to determine position and momentum of the tracks. Furthermore, the particle identification information needs to be reconstructed (electron, photon, neutral pion, hadron separation, and muons). during the reconstruction the creation of smaller data files that contain only the necessary information for a specific analysis may be possible. these files are called reduced data summary tapes (rDST). The reconstruction step may be repeated in order to update the data samples with the latest improvements in the algorithms. There is a data stripping phase that performs a pre-selection in order to select events streams for further analysis. The rDST information (track, energy clusters and PID) is analyzed to calculate properties such as momentum four-vector, location of primary and secondary vertices, etc. Furthermore, the pre-selection algorithms applied to identify the candidates for composite particles of each particular decay channel of interest. The events that pass the selection criteria are fully re-reconstructed recreating the full information associated with them. These are written on a full DST file which also contains the raw data in order to have detailed information for each event. Raw data information is required to perform some analysis. Finally, the DST output of the stripping phase is processed to perform a physics analysis of interest running some algorithms to reconstruct the b-hadron decay channel under study.

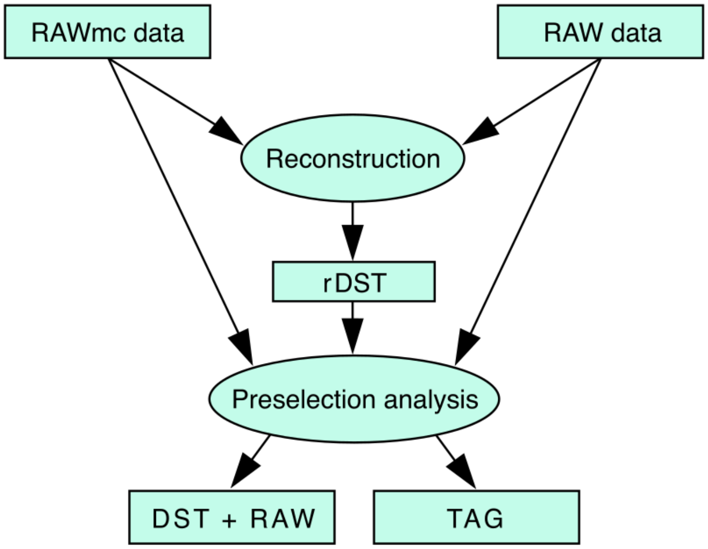


Figure 4-3.: The LHCb computing logical dataflow model [2].

4.4. Track Types at LHCb

The charged particle tracks traversing the LHCb tracking system can be reconstructed if the algorithms successfully store a minimum of the hits [3]. There are several track types at LHCb depending on their trajectories. Figure 4-4 shows the types of tracks that can be reconstructed by the LHCb experiment.

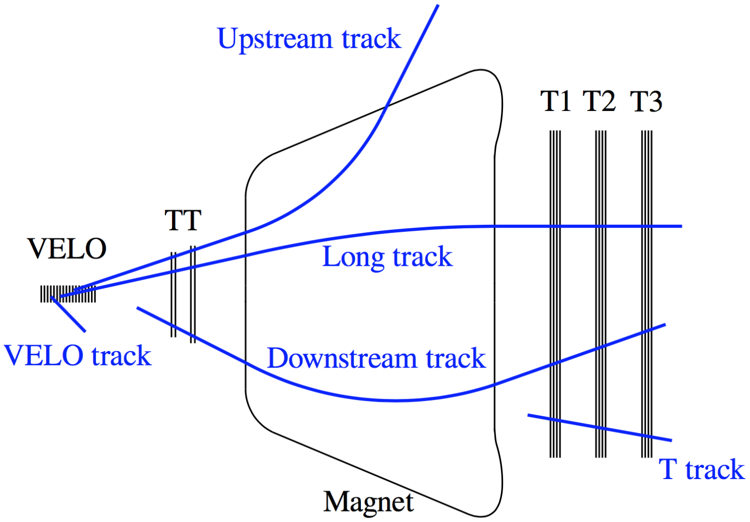


Figure 4-4.: Track types at LHCb [3]

VELO Track: These kinds of tracks are reconstructed exclusively by the VELO detector and are not used by matching algorithms. They do not have momentum information since the magnetic field in the VELO region is negligible. VELO tracks are important to measure the primary interaction vertices. They have typically large polar angles, therefore, they correspond to particles flying in backward direction in most of the cases.

Upstream Track: The upstream tracks traversing VELO and TT stations. They have low momentum ($>$ 2 GeV). Therefore, they suffer large deviations by the magnetic field.

Long Tracks: These kinds of tracks are the most important at LHCb since they traversing the whole tracking system. They have hits in all (Velo, TT and T stations) detection substations. The momentum information of these tracks is the most accurate among all track types. There are two algorithms to deal with those tracks, forward tracking, and track matching [3].

Downstream Track: These tracks traverse TT and T stations however they do not have VELO information. The momentum is worse reconstructed due to the missing hits in VELO. These are important in the reconstruction of K_s^0 and λ particles witch decay typically outside VELO acceptance.

T Track: T tracks are reconstructed exclusively by T station. They come from secondary interactions or interactions with the material of the detector.

5. Track Reconstruction at LHCb

The reconstruction of the tracks traversing the LHCb detector is a challenging task due to the large number of tracks contained in a single event (~ 100). The hits in the subdetectors (VELO, TT, IT and OT) need to be combined to find the real trajectory of the particle that generates them. The LHCb collaboration has pattern recognition software packages to determine not only as many tracks of each type as possible but also to accurately determine their track parameters. The track of a charged particles crossing a magnetic field region can be described by five parameters, its position (x,y), its direction ($dx/dx, dy/dz$) and its momentum at a given position z. These parameters can precisely determine the path of the particle through the detector if we neglect the material effects. There are several individual algorithms in the track finding procedure, VELO seeding [23], forward tracking [66], T seeding [67], track matching [49], upstream [68] and downstream [69] tracking. These algorithms work in a cooperative environment in which VELO and T seeds, are used to extrapolate long tracks first, next downstream and upstream tracks are cratered, and finally, the unused seeds are defined to be VELO or T track types. The hits used by the forward algorithm are not taken into account by any posterior tracking algorithms in order to avoid the possible creation of tracks that share a large fraction of hits called clone tracks.

5.1. VELO seeding

The most important LHCb tracking subdetector is the VELO. This gives us the first measurement of the position and direction of the produced particles. The VELO tracks also serve as seeds for posterior algorithms. Therefore, the reconstruction in the VELO is crucial for the goals of the LHCb experiment. The fast VELO [23] is the pattern recognition algorithm designed to perform this task. The tracks in the VELO can be extrapolated as straight lines since the magnetic field is negligible in this region. The algorithm has several steps to reconstruct these tracks. Firstly, the track is reconstructed in the r-z projection using only the information from r-sensors. Secondly, the information of ϕ-sensors are added to form three dimensional tracks and, finally, some algorithms are run in order to find the best fit for the tracks and eliminate possible duplicated ones.

r-z reconstruction As the main interest of the experiment is the B-decays which come from displaced vertices that are not located on the beam axis. Similarly, and with larger effect, tracks from K_s^0 and Λ are not at all coming from the beam axis and thus are not straight lines in the r-z projection [23]. The tracks crossing at least four r-sensors are desirable. However, the experiment admits tracks crossing just three r-sensors although these have a greater ghost rate. Tracks with four hits are searched first and then the tracks with three hits are looked at using only the unused hits to improve the speediness of the algorithm. The searches are done for the forward going tracks and then for the backward going tracks. The algorithm starts looking at the quadruplets from the last four sensors where the tracks are more separated. The quadruplet found in the last part of the detector is extended to the lower z region as much as possible. It allows some missing sensors in the track reconstruction, four out of five are enough. If the clusters are in the nearest region to the vertical boundary between two VELO halves, The algorithms check the clusters on the opposite side since they can be part of the same track. Clusters cannot be reused for other quadruplet searches to avoid clone creation. When the four sensors cannot contain quadruplets coming from the luminous region, the code stops. It performs the same search in the backward direction. however, the stop criterion changes, it stops when there are no tracks coming from the luminous region. Finally, shorter tracks (triplets) are searched first in backward and then forward direction although three hits are not a strong enough constrain to fit a straight line. If two triplets share a hit, only the best one is kept.

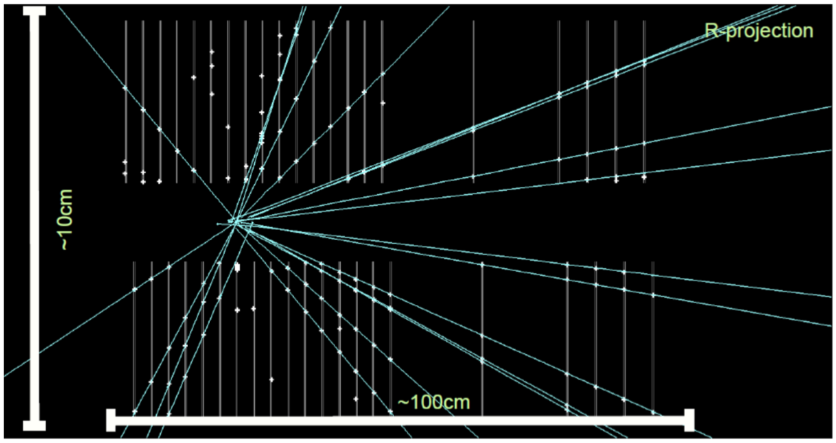


Figure 5-1.: The r-z-projection of reconstructed tracks in the VELO [3].

ϕ hits addition Taking the combined r-z track and the ϕ-sensors information, the code finds the ϕ-clusters that can make a straight line. The r-z tracks are processed by length, the longest tracks first to build more constrained tracks and reduce the ghost ones. The first step is to define the first and the last ϕ-sensor to use. To do this, the code checks the radius of the track in the sensor reference frame. The hits in the selected sensor range are collected, selected and sorted per station. The selection criterion at this point is the sine of the angle to the bisector that satisfies the condition $\sin\phi < 0{,}410$ or, $\sin\phi < 0{,}040$ for tracks in the overlap zone. Furthermore, the projection of the ϕ-strip's center and the bisector should have a minimum value. Finally, the algorithm looks for the intersection of a circle and a line that has two solutions however only one of them falls in the right r zone. The calculations are done in cartesian coordinates and, the resolution is done in x and y coordinates for zones near to vertical and horizontal respectively.

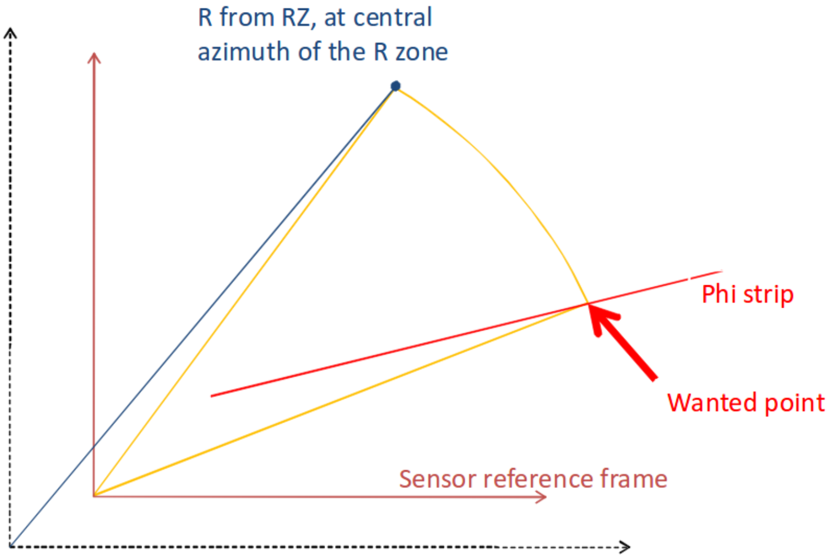


Figure 5-2.: Intersection point computation in the VELO sensors [23].

Final selection Some tests are carried out to select the best tracks to be stored. Each r-z track can have several candidates. If two candidates share more than 70 % of their ϕ hits, the shortest track, or the one with worst quality, is discarded. Tracks with three ϕ-hits that are already used are also discarded. If the track has only three r and three ϕ hits, the track is refitted eliminating hits with $\chi^2_{\text{per hit}} > 12$ [23]. the track is stored if at least six hits are left.

Last track search (Non-pointing tracks) This algorithm is added to the previous since some good tracks cannot be reconstructed by the procedure described above. the non-pointing tracks search is done for forward tracks exclusively and takes only the unused hits. The code tries to reconstruct the tracks since ϕ stations first. The algorithm looks for three ϕ- and three r-hits in consecutive stations. The seed sensor is taken from the last eight stations.

5.2. Forward Tracking

The forward tracking algorithm combines the best VELO seeds found in the previous step and the T stations measurements to find the particle trajectory. The algorithm searches for potential measurements in the x planes of the T stations, applying loose cuts based on the information from the VELO seed [66]. The next step is to search for clusters that belong to VELO seeds. The clusters are fitted by applying a third polynomial fit which selects the best hits based on the χ^2 criterion

$$\chi^2 = \frac{\Sigma\left[a(x_0+t_xz)+b(y_0+t_yz)+c\right]^2}{\sigma^2}, \tag{5-1}$$

where x and y are the position of the hit and, t_x and t_y are their corresponding slopes. Then the tracks are passed for a stereo hits preselection if they satisfy a minimum number of hits or a maximum χ^2_{dof}. The x position can be derived by combining the y information on the VELO seed and the u and v measurements on the T stations. The x deviations are stored in order to discriminate them by making a parabolic fit on the x hits and a straight line fit of the y information on the stereo hits. Once again the hits with large χ^2 are removed. Finally, the best track is selected using a track quality criterion which comprises the momentum, χ^2_{dof}, compatibility of y VELO seed and the T station and the number of hits, variables. The pattern recognition steps are described in detail in the following.

5.2.1. Selection of potential x hits

For each x plane in the T station, a y search window needs to be defined. However, as y information cannot be provided by the x planes, the restriction to the search window in y is limited to the upper and lower half for the OT detection stations and the four boxes for the IT. The VELO seed extrapolation, as a straight line until the central z position (z_0) to the x plane, is defined as the center distance (y_c) in the y coordinate.

$$y_c = y_0\big|_{\text{VELO seed}} + \left.\frac{dy}{dz}\right|_{\text{VELO seed}} \times z_0, \tag{5-2}$$

and the search window in the y coordinate is defined as Δy

$$\Delta y = y_{CT} + 50\text{mm} \times \left.\frac{dy}{dz}\right|_{\text{VELO seed}}, \tag{5-3}$$

here the maximum compatible tolerance in the y coordinate is defined as $y_{CT} = 10$mm. Each hit outside this window is discarded. A correction in the x position of the hits is made by using the y information coming from the VELO seed. All parameters, the intersection point in the y and z position of the VELO seeds and the position of the measurement plane, should be recomputed. The corrected x and z positions of the hit are named (x_{meas},z_{meas}),

$$y = \frac{y_0\big|_{\text{VELO seed}} + \frac{dy}{dz}\big|_{\text{VELO seed}} \times z_0}{1 - \frac{dy}{dz}\big|_{\text{VELO seed}} \frac{dz}{dy}\big|_{\text{plane}}}, \tag{5-4}$$

$$z_{meas} = z_0 + \left.\frac{dz}{dy}\right|_{\text{plane}} \times y, \tag{5-5}$$

$$x_{meas} = x_{meas} + \left.\frac{dx}{dy}\right|_{\text{plane}} \times y. \tag{5-6}$$

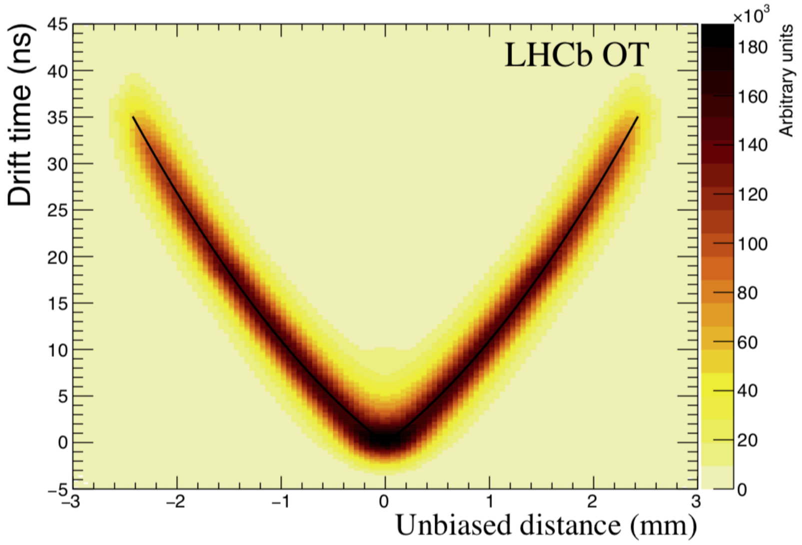


Figure 5-3.: The drift time versus the unbiased distance distribution with the overlaid TR-relation [70].

The drift distance r should be taken into account in case hits traverse the OT stations. This distance is extracted from the drift time measurement. The drift-time-distance relation (TR-relation) is calibrated from tracks that successfully pass the momentum ($P > 10$ GeV) and $\chi^2_{dof} < 2$ requirements which are parametrized with the second order polynomial.

$$d_t(r) = \left(21{,}3\frac{|r|}{R} + 14{,}4\frac{r^2}{R^2}\right) ns, \tag{5-7}$$

where the distance R is the inner radius of the straw tube, and the drift distance r should satisfy the condition $-0{,}3 < r < 2{,}5$ for the hit to be kept otherwise the hit is discarded. In the next step, the VELO seed is extrapolated, as a straight line, to a reference plane located at $z_{ref} = 8520$ mm, which modifies the search window around the extrapolated x position. The standard pattern recognition algorithm starts from VELO seeds only and the TT hits information is added later to improve track quality. The deviation of the particle's trajectories from a straight line is performed by the magnetic field, corresponds to a window in which the particle has a minimum momentum of $P_{mim} = 1$ GeV and a minimum transverse momentum of $P_{T_{min}} = 80$ MeV. The extrapolation needs to be slightly corrected if the TT hits information was used to calculate the momentum of the VELO seed.

The propagation of charged particles, through a magnetic field region with a defined field map, is theoretically well calculated. However, a fast approximation is needed for pattern recognition to find the intersection of the track and the reference plane at a fixed position. A cubical parametrization has carried out to describe the trajectory of a charged particle traversing the magnetic field region.

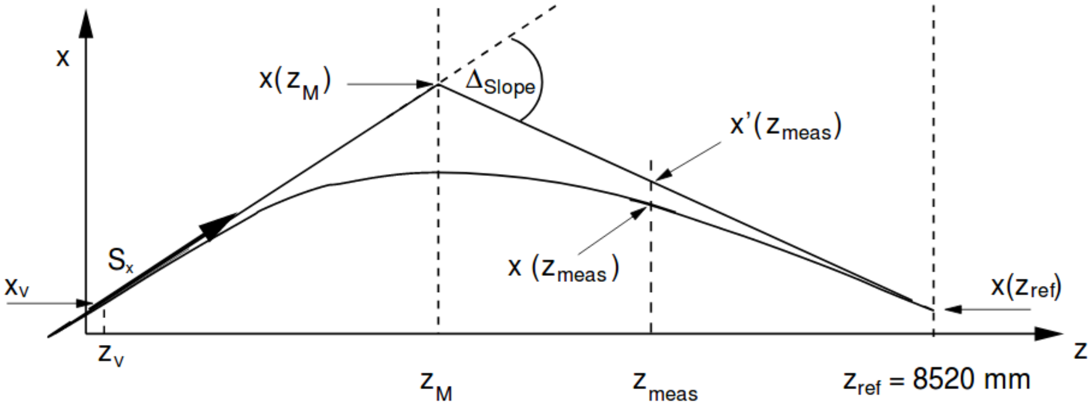


Figure 5-4.: Track parameterization used in the forward pattern recognition strategy [66].

$$x(z) = x_{ref} + B_x(z - z_{ref}) + C_x(z - z_{ref})^2 + D_x(z - z_{ref})^3, \tag{5-8}$$

$$y(z) = A_y + B_y(z - z_{ref}). \tag{5-9}$$

B_x is the tangent of the track at z_{ref}. Δ_{slope} is the angle between the x_{slope} of the VELO seed S_x and B_x and z_M is the position of the focal plane of the magnet [66]. The position of the x hits should be re-corrected because in previous calculations, due to the fact that the magnetic field has a small impact on the y_{slope}. A field component perpendicular to y causes a mild change in dy/dz which results in a correction δy of the y position of the track at the z_{meas} distance. The detailed calculations can be found at Ref. [66].

5.2.2. Search for hits

The search for the clusters of the hits and sorted a list with them is the next step. The maximum distance between the first and the last hit is calculated as:

$$\Delta x = maxSpreadX + |x_{\text{first hit}} - x_{ext}| \times maxSpreadSlopeX, \tag{5-10}$$

where $maxSpreadX = 0{,}6$ and $maxSpreadSlope = 0{,}011$ are taken as definitions. The clusters that are further away from the straight-line extrapolation usually belong to low momentum tracks. If the first hit is on the OT station, the Δx should be enlarged by 1,5mm to take into account the drift time and drift distance measurements. The algorithm requires at least five hits found within the window $[x_{\text{hit i}}, \ldots, x_{\text{hit i}} + \Delta x]$. The hit $i+1$ is taken as a potential starting point of a new cluster, in case hits do not satisfy the previous condition. This procedure is repeated until a successful cluster candidate $[x_{\text{hit j}}, \ldots, x_{\text{hit k}}]$ is found. Before a measurement was added to the cluster, several conditions should be satisfied (the starting hit of the overlapping cluster should be chosen from the hits in at least five different planes, $x_{start} - x_{end} < \Delta x$ and one hit in the range $j+1, \ldots, k+1-5$). Moreover, clusters are merged if the position of the starting hit j is within the Δx to the end position of the next accepted cluster. Otherwise, a new cluster is created. The algorithm stops when all clusters on the list are checked.

5.2.3. Selection of Potential Stereo Hits.

The x fit parametrization is extrapolated to each u and v planes within the y compatible tolerance. The u and v information is combined with the known track parameters. $A_y - B_y \times z_{ref}$ and B_y of the x candidate are used instead of the VELO seed parameters. The u and v are transformed into x measurements where dy/dx is the tilt of the stereo angle. Once again the drift distance has to be between $-0{,}3\text{mm} < r < 2{,}5\text{mm}$. If the difference between x_{ext} and x_{meas} is within

$$\Delta x = maxSpreadY + maxSpreadSlopeY \times \left(\frac{q}{P}\right)^2, \tag{5-11}$$

where $maxSpreadY = 1{,}5\text{mm}$ and $maxSpreadSlopeY = 70\text{mm}$, the hits are added to a list. Δx is enlarged by 1,5mm if the first hit corresponds to an OT measurement. Finally, the scan for clusters is performed in the same way as the scan for x ones. The minimum number of stereo planes required is four and the maximum size of the cluster is 3mm for IT hits as starting hit and 4,5 mm for OT hits. In case several good stereo clusters are found the one with the largest number of different planes is taken [66].

A fit is performed to reduce the number of wrong hits and increase the track quality. In the first step, the x and stereo hits are updated to the current track parametrization. The x projection is fitted, the x position of the stereo hits is also included in the fit. A straight line fit for the y component of the stereo hit. Therefore, a new fit in y space has to be performed due to the new parametrization. Next, the hit with the largest χ^2 contribution is removed if it is larger than $\chi^2_{max} > 20$ as well as the hits with a χ^2 larger than 1000. The process is repeated in case of hits in more than nine planes, x and stereo, are left. Once all stereo hits have $\chi^2 < 40$, the χ^2 contributions from the x hits are tested. The geometrical compatibility is tested once. Finally, the tracks that survive all filters are sent to the final track selection algorithm.

5.2.4. Final track selection.

There are several steps to select the best track candidates. The first step is to test the change in the y position of the VELO seed and the track candidate in the T stations in order to discard tracks with Δy larger than:

$$\Delta y_{tol} = \Delta y_{max} + \frac{1}{P^2} \times \Delta y_{max\ Slope}, \tag{5-12}$$

where $\Delta y_{max} = 30$ mm and $\Delta y_{max\ Slope} = 300$ mm. A minimum number of hits is required in the T stations, IT hits count two times due to the fact that there are two times more detection channels in the OT than in the IT, $n_{Hits} = 2 \times n_{IT\ Hits} + n_{OT\ Hits}$, where n_{Hits} correspond to the hits on the T stations only. For tracks that mainly traverse the IT region the minimum number of hits is 14, on the other hand, the tracks that traverse the OT region the minimum number of hits is 16. A quality variable Q is defined which combines ΔY, χ^2_{ndof} and momentum. The smaller Q value the best track

$$Q = \frac{5 \times \Delta y}{\Delta y_{tol}} + \frac{\chi^2_{ndof}}{10} + \frac{10}{P}. \tag{5-13}$$

The track with the lowest Q value (Q_{min}) is retained, as well as tracks with Q between $[Q_{min}, Q_{min} + 1]$, all others are discarded. Finally, the track has to be a minimum of 22 hits to be considered as a final track candidate. The tracks that survive these cuts are stored.

5.2.5. Adding TT hits

The forward tracks reconstructed by the algorithms can further have TT-hits. The TT-information is not essential and does not affect the track reconstruction itself. However, the quality of the tracks can be improved by adding the extra information of the TT [3].

The VELO seed is extrapolated as a straight line to the z position in the TT station. The consistency of the extrapolated y position is checked by the tolerance Tol_{TT} that has to be within

$$Tol_{TT} = TT_{tol} + \frac{TT_{tol\ slope}}{P}, \tag{5-14}$$

where $TT_{tol} = 2$ mm and $TT_{tol\ slope} = 20$ mm. Taking into account the track parameters, the position of all hits are updated to make corrections in potential slopes at the measurement planes. The magnetic field is described by a kink between the VELO seed extrapolation and the extrapolation as a straight line of the TT stations at the focal plane $z_{TT\ field} = 1650$ mm. The next step is to project the difference between the extrapolated and measured x distance ($x_{ext} - x_{meas}$) on a plane between the two TT stations at $z_{TT\ proj} = 2500$ mm if it is lower than Tol_{TT}. After the correction is done by the effect of the magnetic field, the only difference between the VELO seed and the true trajectory is a change of the slope of the seed due to multiple scattering. TT1 deviations are expected to be smaller than the ones in TT2 and their sizes are expected to have a linear dependence in z. Then, a search for the cluster of hits is performed in the projection plane. The algorithm requires at least hits in three different planes within the Δx acceptance defined as:

$$\Delta x = 2\ \text{mm} + 0{,}5 \times |x_{first\ hit}|, \tag{5-15}$$

the $x_{first\ hit}$ is the projected position of the starting hit of the cluster. A straight line and a constant fit are performed in x and y respectively.

In order to be successfully reconstructed a particle should satisfy the following conditions:

- The momentum of the particle at its production vertex has to be greater than 1 GeV
- the particle has at least three reconstructed clusters in r-sensors and three in ϕ-sensors.
- The particle has at least one reconstructed x and stereo hit in each T station.
- The particle does not interact hadronically before the end of the T stations.
- The particle is not an electron.

The Monte Carlo association criterion for forward tracking says that at least 70 % of the VELO hits and 70 % of the T hits should be associated to the same particle.

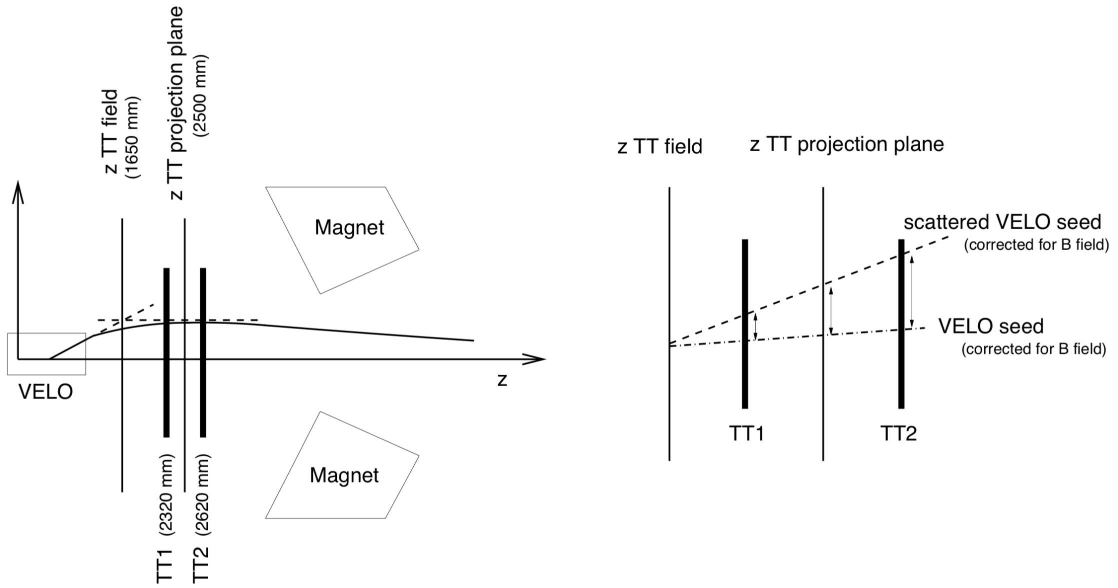


Figure 5-5.: Deviation from straight line extrapolation of VELO seed and real path of particle through the detector due to magnetic field (left). Deviation from VELO seed and scattered seed (right)[66].

5.3. T track seeding

The track seeding is a stand-alone algorithm that searches for tracks in the T stations (T1-T3). This is a challenging task due to several reasons. First, the magnetic field in the T station region cannot be negligible therefore the trajectories of the charged particles cannot be assumed as straight lines in the (x,z) projection. Besides, most of the tracks found here correspond to secondary interactions which are not interesting for the LHCb purposes. These events can have an occupancy of the detector of 40 % or more in some zones. The algorithm starts searching for straight-line candidates in the x-z plane, considering only the hits in the x layers. Then the effect of the magnetic field is taken into account adding the hits of the middle layer and parametrizing the trajectory of the track candidate as a parabola. In the final step, the stereo hits are added to confirm the track candidate and, a likelihood calculation is applied to improve the track quality and to reduce the wrong hits.

Before starting with the search for x track candidates, there are some steps to follow in order to prepare the data taken by the T stations in order to remove some hot-spots point which may cause many combinatorial problems. First, the curling tracks should be removed. If more than six straws are hit, the hits are discarded. Additionally, the hits in the modules where the occupancy is greater than 40 % are removed. The inner tracker has the same problems however, they have less impact on track reconstruction due to the higher granularity. These cuts remove up to 10 % of the total data. Nevertheless, the speediness of the algorithms is increased up to 50 % while maintaining the efficiency unaffected. The algorithm is not executed, if the total number of hits surpasses 10000.

5.3.1. Projection search

The search is performed using the five logical sectors of the T station. The OT is divided into upper and lower halves meanwhile the IT region is divided into four boxes, left, right, upper and lower. Tracks traversing the detector swept from side to side in the x plane due to the magnetic field effect. As it is vertical tracks rarely cross from upper to lower part of the OT or upper to lower box of the IT. Therefore the search is made separately for each sector. The IT boxes are overlapped to avoid acceptance gaps as well as the IT boxes with the OT halves. The tracks traversing these regions are important for alignment purposes. Since these tracks do not have enough hits in the single sector, they are left for the stub search algorithm. Here the process for pattern recognition for the OT will be described. The process for the IT recognition is closely similar however it is slightly simplified by the fact that there is no left-right ambiguity. The drift time in the OT straws measures a radius from the wire, which when coupled to an assumed direction of the track leads to an ambiguity as to whether the track passed on the left or right side of the wire. This effect is not present in the silicon IT [67], The detailed algorithm for the IT pattern recognition is described in Ref. [71].

x hit search

Two pairs of hits that belong to the first and the last T stations are selected to define a straight line in the x-z plane. The pair is kept for studies if it satisfies the conditions: the slope of the line s_x is smaller than $|s_x| < 0{,}8$, and $|x/3125\text{ mm} - s_x| < 0{,}7$ (the correlation between slope and x observed for long tracks). An additional cut in the P_T kick angle $\Delta\theta$ is applied.

$$|\Delta\theta| < 0{,}7. \qquad \text{where,} \tag{5-16}$$

$$\Delta\theta = \tan^{-1}(s_x) - \tan^{-1}\left(x_1 - \frac{s_x(z_1 - z_0)}{z_0}\right), \tag{5-17}$$

the coordinates of the first hit are (x_1,z_1), and $z_0 = 5{,}3$ m (approximately the center of the magnet). All hits in the ±10 mm window are selected for further analysis. furthermore, there have to be at least seven hits, including the two that formed the straight line, out of ten which are the expected number of hits on the OT x layers. T2 station should contain at least one of the selected hits within the window. Drift time has not been taken into account so far, because of the search window is of the order of the drift cell size.

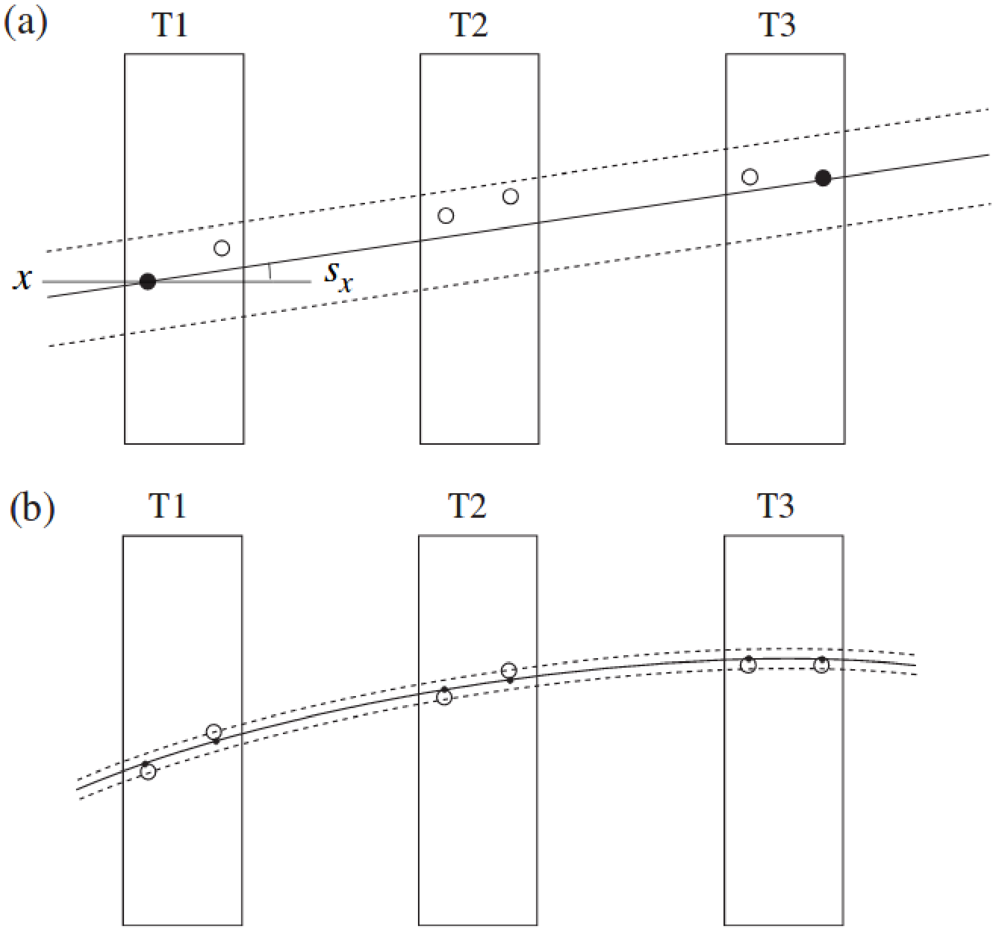


Figure 5-6.: Search for track candidates in the x projection: (a) opening a window around a straight line extrapolation between two hits in T1 and T3; (b) opening a window around a parabolic trajectory after adding a hit in T2. [67].

A hit in the T2 station is chosen among the hits selected in the previous step. A parabola is defined with those three hits. Then the left/right (L/R) ambiguity is checked for each of the three hits. Around the parabola, a 1 mm search window is defined and the hits that fall in that region are selected for the following steps. The algorithm selects the ambiguity which maximizing the number of selected hits. This approach also reduces the number of L/R combinations to be tested, by neglecting the effect on the parabolic parameterization in other stations when the ambiguity of a hit is changed [67]. A least-squares fit is made to selected hits in the parabola, the chosen ambiguities are taking into account. An iteration of the fit along the hits is made allowing the L/R ambiguity to change the sign in order to move the point as close as possible to the fitted line. If a contribution larger than 3σ remains after the optimization of the L/R ambiguity, the most outlying point is removed. Finally, some cuts are applied to select the best track candidate, at least seven hits should be kept, the x coordinate satisfy $|x_0| < 4{,}5$ m, the slope has to be smaller than $|S_{x_0}| < 0{,}8$, and the curvature term t_x should comply the condition $|5{,}3 \times 10^4 t_x - \Delta\theta| < 0{,}8$. Each accepted track candidate is stored for the stereo hit search and its hits are tagged as used to prevent their use of any subsequent search.

Add stereo hits

The search for stereo hits starts by selecting the hits that are compatible with the x candidate. Each selected hit is converted into a y measurement using the information of the x-z trajectory. The algorithm checks whether the y coordinate is within the physical boundaries of the hit channel (i.e. within the length of the wire, for an OT hit). The tilt of the T station layer is taken into account during the conversion of the stereo hits into y coordinates. The process is similar to the one used to select the x hits without the parabolic trajectory parametrization. This time a straight-line trajectory is searched among a selected pair of hits in the first and the last station. applying a cut in the slope s_y, $|s_y - y_1/z_1 < 0{,}1|$ where the $(x_1, z-1)$ is the position of the first selected hit. A 100 mm window around the straight line that connects the two hits. All hits that fall in these regions are selected for further studies. A minimum of eight hits out of ten, including the two initial hits, are required. Next, the L/R ambiguity for the two initial hits is tested and retained the one which maximizes the number of selected hits within a 10 mm window. Then the L/R ambiguities of all hits are tested to find the best straight line fit for the selected hits. The fit is iterated similar to the parabolic fit, however, this time the cut is loosened up to 4σ due to additional uncertainties introduced when converting stereo hits to y measurements. Finally, at least five stereo hits, and fifteen in total should be retained to be considered as a track candidate. Only the stereo candidate with the largest number of hits is kept. If a candidate fails the requirements, it is not deleted to avoid the need to reorder the vector that holds them, but instead, a flag is set within the seed candidate object to indicate that it should no longer be used [67].

Likelihood determination method

A Likelihood [72] function is calculated for each of the candidates that survive the x and stereo cuts. It improves the previous method used as a discriminant which consisted of a simple sum of the χ^2 from parabolic and linear fits however the tracks with more hits are penalized. The likelihood method is based on whether the track candidate is expected to give a hit on the T layers. OT inefficiency increases toward the wall of the straw tubes due to insensitive gaps between two cells of the order of 0,25 mm.It is calculated for each tracking layer where the track candidate does not have any hit. Although the fitted trajectory passes through the region. Larger likelihood values represent less efficient regions. The probability of χ^2 is taken which is the combination of a product of the likelihood values and a term that discriminates the goodness of the fit. Finally, a binomial counting term is added to compare the number of measured hits with the number of expected ones. The expected average efficiency is also taking into account. The natural logarithm of the combined likelihood value is then used as a discriminant to select good quality tracks candidates.

Some corrections to the parabolic fit need to be carried out before using the tool. The z component of the x hits is adjusted according to the y coordinate taken from the linear fit. After the refitting, the candidate is already in the LHCb coordinate system. Likelihood contribution is determined by sampling a normalized probability density function which parameter is the distance to the nearest wire r with value $\mathcal{L} = 0{,}261 + \exp(5{,}1r - 11{,}87)$ [67]. Fluctuations of the probability of χ^2 degrade the track selection performance of the likelihood method. Therefore a scaling factor is added to de-weighting that contribution. The log-likelihood can be described as

$$\ln \mathcal{L} = \ln \frac{N!}{n!(N-n)!} + n \ln \epsilon + (N-n) \ln (1-\epsilon), \tag{5-18}$$

where ϵ is the average effective efficiency which is 0,90 for the OT and 0,99 for the IT. The difference is due to the effect of missing hits in a high occupancy environment.

The seed candidates are sorted according to their log-likelihood values and the one with the highest (log-)likelihood is selected. Next, the second highest value is tested whether it shares more than three hits with the first. If so the track is discarded, otherwise, it is selected, and so on. The algorithm continues until there are no remaining seeds or the log-likelihood falls below a cut which can be tuned depending on quality requirements.

5.3.2. Stub search

The stub search is made with the hits in the IT which are not assigned to any T seed previously. They required to form a straight line segment made by hits in the four layers within a T station. Ta pair of hits in the x layers are chosen to form a straight line segment in the x-z projection with some cuts applied in the slope: $|s_x| < 0{,}5$ and $|x_1/3125\text{ mm} - s_x| < 0{,}3$. This parametrization is then used to convert stereo hits into y measurements in that box.

This ensures that the resulting coordinates are within the physical boundaries of the strips. a pair of hits in the stereo layers are selected to form a straight line in the y-z projection where the cuts,$|s_y| < 0{,}25$ and $|y_2 - y_1 - (z - 2 - z_1)y_1/z_1| < 3$ mm, are applied. The clone suppression is not necessary due to the fact that many hits were previously removed by the projection search and the IT occupancy which is lower than the presented in the OT.

The next step is to link together the stubs. to do this, two stubs, which belong to the first and second T stations respectively, are selected and a projection to the middle point between the two stations is made. If the two projections successfully fulfill the cuts: $|\Delta x| < 0{,}3$ mm, $|\Delta y| < 0{,}3$ mm and $|\Delta s_x| < 30$ mrad, the two stubs are linked together. A similar process is made to link the third stub. The pair or the triplet of stubs have to be fitted to form the seed candidate (parabolic fit in x-z and linear in y-z). The stubs that cannot be linked to any others are checked to determine whether they come from tracks that cross from IT to OT. Extending the stubs to the OT and check if there are hits consistent with the parabolic trajectory within the window of the extrapolated stub. Next, a linear trajectory search is made using the stereo hits of the same sector consistent with a line joining the interaction point to the stub y coordinate.

5.4. Match Tracking

The VELO seeds and the T seeds are linked together by the track matching algorithm to find long tracks. This algorithm uses only the hits which have not been used for the forward tracking algorithm. There are two approaches to match VELO and T seeds.

The first approach combines the VELO and T seed by extrapolating them towards the plane in the center of the magnet. The magnetic field is considered homogeneous enough to put forward, all the intersections of the extrapolated seeds occur in a focal plane at an almost fixed z position. This method is robust, fast and efficient. It also shows the feasibility to make the reconstruction of tracks without using the information on the TT stations. The second approach matches the VELO and T seeds at a plane located just downstream the last VELO station. The first step is to estimate the momentum of the T seeds which is used to extrapolate the T seed towards the matching plane. The VELO seed is also extrapolated as a straight line to the same plane. To select the best match between VELO and T seeds, a χ^2 criterion is applied. A search for the TT hits is performed in each correctly matched track. If the hits found are accepted, they are added to the final track. The second approach leads in efficiency to the first one owing it takes into account the full magnetic field shape. This is the default algorithm which is particularly special for low momentum particles. It is discussed in more detail below.

5.4.1. Momentum estimation

The momentum can be estimated by the p-kick method which consists of an instant kick on the momentum vector at the center of the magnet. the momentum kick, $\Delta\vec{p}$, depends on the integrated magnetic field along the trajectory left by each particle.

$$\Delta\vec{p} = q \int d\vec{l} \times \vec{B}. \tag{5-19}$$

The principal component, Δp_x, provides the highest precision on the momentum. This can be written as:

$$\Delta p_x = p_{x,f} - p_{x,i} = p \left(\frac{t_{x,f}}{\sqrt{1 + t_{x,f}^2 + t_{y,f}^2}} - \frac{t_{x,i}}{\sqrt{1 + t_{x,i}^2 + t_{y,i}^2}} \right) = q \int \left| d\vec{l} \times \vec{B} \right|_x , \tag{5-20}$$

where $t_{x,f}$ and $t_{y,f}$ are the slopes of the T seed which are evaluated at the T3 station. Their values are known after the parabolic fit. The slopes $t_{x,i}$ and $t_{y,i}$, and the magnetic field have to be estimated. the charge of the particle, q is determined from the sign of curvature and the field polarization.

In the LHCb experiment, the total integrated magnetic field along the z-axis is 4,2Tm. $z_{magnet} = 5150$ mm defines the plane at the center of the magnet where the integrated field has half the total value. This is the initial focal plane for the projection of all T seeds in which the two projection of the seeds intersect each other as shown in Figure 5-7. The path is extrapolated from the T seed up to the z_{magnet} plane, at this point, the path bends towards the nominal interaction point. This gives the first estimation of the slopes before the magnet. The integrated field is calculated along this path which defines a new plane, in the center of the magnetic field, located at $z = z_c$. The integrated magnetic field, as well as the slopes, are re-calculated for the new plane. A slight systematic effect is observed in momentum resolution which has a dependency of the $t_{x,f}$ slope.

$$\left\langle \frac{\delta p}{p} \right\rangle = -0{,}0092 - 0{,}112 t_{x,f}^2 . \tag{5-21}$$

After making the correction, the final momentum has a resolution of $\frac{\delta p}{p} = (0{,}6964 \pm 0{,}005)\,\%$.

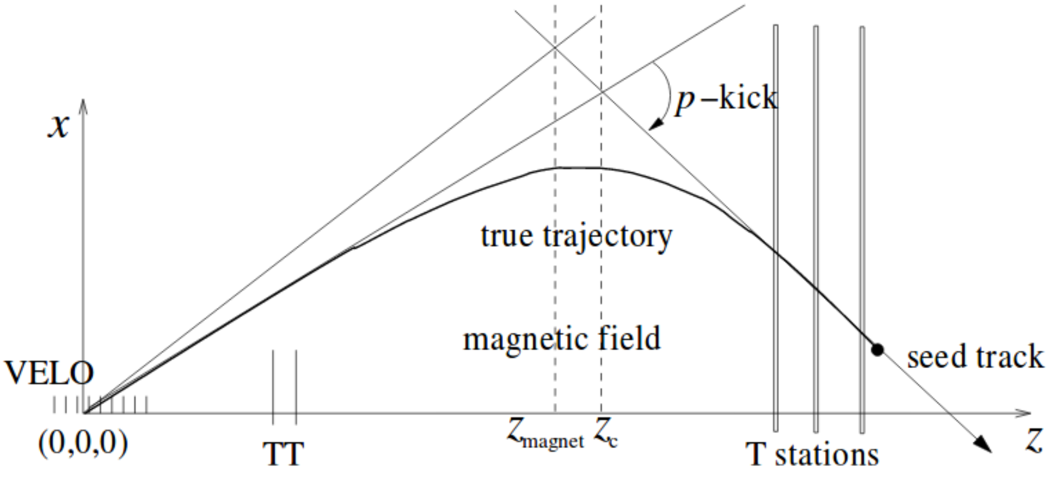


Figure 5-7.: Schematic drawing of the p-kick method (not to scale) [49].

5.4.2. Matching selection criteria

T seeds are fitted with the Kalman filter [49, 73] after applying the p-kick method. It takes into account the multiple scattering and energy loss effects. The VELO seed is also fitted, however, the correction for these effects are not carried out due to the fact that the momentum information of the VELO seeds cannot be determined. The good precision required is achieved by the track parameter fits which leads to an optimal matching efficiency. The momentum calculated from the p-kick method is used to extrapolate the trajectory of the particle (T seed) through the magnetic field region towards the $z_{match} = 752$ mm plane which is located just behind the last VELO station where it is matched with the VELO seed. χ^2_{match} is calculated as follows:

$$\chi^2_{match} = (\vec{x}_{VELO} - \vec{x}_T)(C_{VELO} - C_T)^{-1}(\vec{x}_{VELO} - \vec{x}_T), \tag{5-22}$$

where $\vec{x}_{VELO}$ and $\vec{x}_T$ are the track parameters of the VELO and T seeds at z_{match}, and C_{VELO} and C_T are the corresponding covariance matrices [49]. The momentum cannot be calculated for the VELO seed, therefore, their components are not evaluated in the eq. 5-22. The combination of seeds that falls below $\chi^2_{match} < 500$ is selected as track candidates. This cut reduces the number of wrong combinations. If a VELO or a T seed is selected more than once, the combination with the lowest χ^2_{match} is kept.
The effects of multiple scattering are despised due to the fact that these effects are roughly the same for all T seeds. If those effects are taken into account the same efficiency is obtained. However, the χ^2_{match}, due to underestimating errors, and the processing speed is penalized. The matching efficiency is calculated as:

$$\epsilon_{match} = \frac{N_{correct}}{N_{MC}}, \tag{5-23}$$

where $N_{correct}$ is the number of correct matches found and N_{MC} is the number of true match seed according to the MC information. Additionally, the fraction of wrong matching is defined as:

$$w_{match} = \frac{N_{tot} - N_{correct}}{N_{tot}} = \frac{N_{wrong}}{N_{tot}}, \tag{5-24}$$

where N_{wrong} is the number of wrong matches and N_{tot} is the total number of matches found. As it is well known the seed multiplicity affects not only the efficiency but also the wrong matches. The average seed multiplicity is $N_{VELO+T} = 142$.

5.4.3. Adding TT hits

The final step is to add the TT hits to the matched track candidate. The clusters of the hits of neighboring strips on the silicon layers have a signal-noise ratio above a threshold. The VELO seed is extrapolated to the TT region using the previously calculated momentum information from the T seed. Deviation due to multiple scattering in RICH1 is taken into account. These are approximately equal for all TT hits that belong to the same track. The algorithm searches for groups of TT hits that have approximately the same distance from the predicted trajectory.

Groups of maximum four hits are allowed it means one hit per TT layer. Hits that have a distance below $d = 10$ mm are considered for the searches. Hits in the same station cannot differ in more than $d = 1$ mm, and 2 mm if they belong to different stations. I case two hits or more hits are compatible with the same group, a new group is created. The minimum number of hits required for a group is three. On the other hand, if a matched track has more than one group compatible, the one with the lowest quality q is selected. q is defined as

$$q^2 = \bar{d}^2 + w^2_{spread} s^2_d, \tag{5-25}$$

where $\bar{d}$ is the mean distance, w_{spread} is a weight factor which is tuned to maximize the TT hit efficiency, and s_d is the RMS spread of the distances. The addition of the TT hits does not affect the track reconstruction itself. However, It has an impact on momentum resolution and the track fit.

5.5. Downstream Tracking

This part of the algorithm tries to reconstruct the tracks traversing only the TT and T station. This extrapolates T seeds and matches them with the TT hits. These tracks do not have VELO information due to the fact that they belong to final states of long live particles such as K^0_s and Λ.

The algorithm takes advantage of the fact that TT and T stations are located at both sides of the magnetic field region. This supposes TT and T seed intersect each other in the center of the magnet. The algorithm extrapolates T seed to the center of the magnet and searches for compatible measurements which are compatible with a straight line from this point towards the TT station. As the tracks are not too far from the interaction region, some constraints, such as maximum transverse momentum $P_{T\ max} = 209$ MeV, maximum momentum $P_{max} = 2$ GeV and the maximum angle between the K_s^0 and the decay pions is 100mrad, are applied to select the best track candidates. There is also a window to search for TT measurements which correspond to 12 cm, roughly 1200 strips. Many candidates fall in that region. Therefore the size of the window scales with $1/P$ and it is tuned to maximize the efficiency and reduce the ghost rate as much as possible. The same constraints are applied to Λ, Σm, Ω and Ξ since they have lower maximum transverse momentum.

5.5.1. Search for x measurements

Before the candidates' selection, some algorithms are run in order to carry out a parametrization, a first filtered and a pre-selection. The position at the center of the magnet is not a simple plane due to its trapezoidal aperture. Therefore, Monte Carlo simulations are used to make a parametrization of that position z_M in terms of known variables such as the slopes of the T seed, t_x, t_y, at $z = 9450$ mm, and the x coordinate of the T seed, x_{seed}, at this position.

$$z_M = 5368{,}54 - 2155{,}88 \times t_y^2 + 597{,}27 \times t_x^2 - 0{,}00001455 \times x_{seed}^2, \qquad (5\text{-}26)$$

where the values of the parameters are a result of the fit obtained from the Monte Carlo. The position of the center of the magnet is shifted as shown in Tab. 5-1.

Parameter	t_y	t_x	$x = 2$ m
Shift	-80 mm	$+53$ mm	$+57$ mm

Table 5-1.: Corrections of the z_M distance due to different parametrization values.

These corrections are essentially a property of the field map and do not need to be measured with real data. However, it may be possible to check them with clean events [69].

Although the momentum parametrization does not affect pattern recognition, its value is taken by the Kalman fit to calculate the final momentum with a resolution of 1 %. This uses t_x, t_y and Δx_{slope} assuming a nominal fiel map.

$$P = \frac{1190{,}86 + 605{,}67 \times t_x^2 + 2656{,}55 \times t_y^2}{\Delta x_{slope}}. \qquad (5\text{-}27)$$

The error of the x and y measurements act as the first filter.

$$\Delta x = \sqrt{\Delta x_{seed}^2 + (z - z_M)^2 \times \Delta T_{x\ seed}^2 + (\Delta z \times \Delta x_{slope})^2}, \qquad (5\text{-}28)$$

where the first two terms are the extrapolation of the T seed track error and the last term is the error of the z. The tolerance on $\delta P/P$ is quite weak and its default value is 0,7. The search starts with a pre-selection of the hits on the TT stations within the maximum tolerance in the x coordinate, $x_{tol} = 100$ mm × GeV. This value significantly reduces the ghost rate and keeps the efficiency almost unaffected. The hits that successfully pass the x tolerance are selected for the y search which retains the hits within $y_{tol} < 15$ mm window. Next, the algorithm searches for compatible x measurements with the straight-line extrapolation. the x measurements are matched and added to a temporary vector if they are close enough. If the hits are in the same layer the tolerance is 0,01 mm or in other layers varies from 2,5 mm to 3,5 mm. Finally, the track is fitted in case two or more measurements are found. The Δx and the maximum distance allowed are used to determine if the measurement is kept or rejected.

5.5.2. Adding stereo hits

The stereo measurements pre-selected are compared with the extrapolation of the T seed. If the distance is smaller than 3,0 mm, they are kept. The minimum measurements required are three out of four, otherwise, the track is rejected. These are sorted according to its z coordinate. Next, a fit over the measurements is performed. the y coordinate of each stereo measurement is computed. the y hits are checked if their y differ from the predicted track for more than three times the maximum distance, they are removed. Taking the x information in the stereo layers, the x projection is refitted with the new information. If a hit has a distance larger than the maximum distance, it is rejected.

5.5.3. Final track selection

The first criterion is the number of measurements (minimum three in all stations). If the new candidate has fewer measurements than the “current value”, the new candidate is ignored. The “current value” is the number of measurements of the best candidate (maximum four). If a candidate has more ghost hits, it is rejected. If the new candidate has at least three measurements, the candidate with the lowest χ^2_{ndof} is kept [69].

5.6. Upstream tracking

The reconstruction of short tracks is based on the principle that these kinds of tracks are low momentum tracks. They do not reach the T stations or are not reconstructed by the long track algorithms due to several reasons. Low momentum tracks are useful for global particle identification and could serve for flavor tagging especially if particles have kaons as final states. They can further employ to select some B decay channels. The momentum range is between 0,5-3 GeV. In the trigger system, this pattern recognition is useful to suppress the minimum bias events coming from the L0. This is achieved by combining the information of the high transverse momentum with the high impact parameter. This time the search momentum window is above 3 GeV. There are two reconstruction algorithms, the on-line [1] [74] and Off-line [75]. The same pattern recognition is used for both, the only difference is the tuning.

5.6.1. Pattern recognition

The upstream algorithm takes as an input values the VELO seeds and a collection of TT hits. The algorithm uses two windows to search for TT hits. The first is the **search window**, w_{search}, which is the one that opens around the straight line extrapolation of the VELO seed to the z distance in the middle of the TT stations (z_{TT}). This is used to collect the possible hits that belong to the track. The size of the w_{search} depends on the minimum momentum of the track. The hits falling into this region The combinations with the track seed segments are performed by calculating the Δx which corresponds to the distance between the x coordinate of the hit and the x coordinate of the straight line extrapolation of the VELO seed. Re-scaling the Δx, as Δx_{norm}, at the z_{TT} position. The correct hits should have the same Δx_{norm} value for all TT layers. however, In a real situation, a spread of these distances around some central value is expected [68]. The tolerance window, w_{tol}, sets the maximum allowed difference between the Δx_{norm} values belonging to different layers. The absolute value $\Delta x_{norm,i} - \Delta x_{norm,j}$ (i and j denote TT layers, $i \neq j$), cannot be used for the whole momentum range due to the multiple scattering. Therefore a relative value is defined as

$$S_{rel} = \frac{\Delta x_{norm,i} - \Delta x_{norm,j}}{\Delta x_{norm,i}}, \tag{5-29}$$

which has a weakly dependence on the momentum of the track. So the tolerance window can be expressed as

$$w_{tol} = w^0_{tol} + S_{rel} \times \Delta x_{norm}, \tag{5-30}$$

where the w^0_{tol} is the constant reflecting the spatial resolution within the limit of infinite momentum and S_{rel} embraces the effects of multiple scattering increased by low momenta. Finally, hits are sorted in ascending order according to their Δx_{norm} values. If other hits are found within the tolerance window, they are added to the combination. The best combination is selected to form the track.

---

[1] The matching algorithm described in this reference is the used for Run1 data

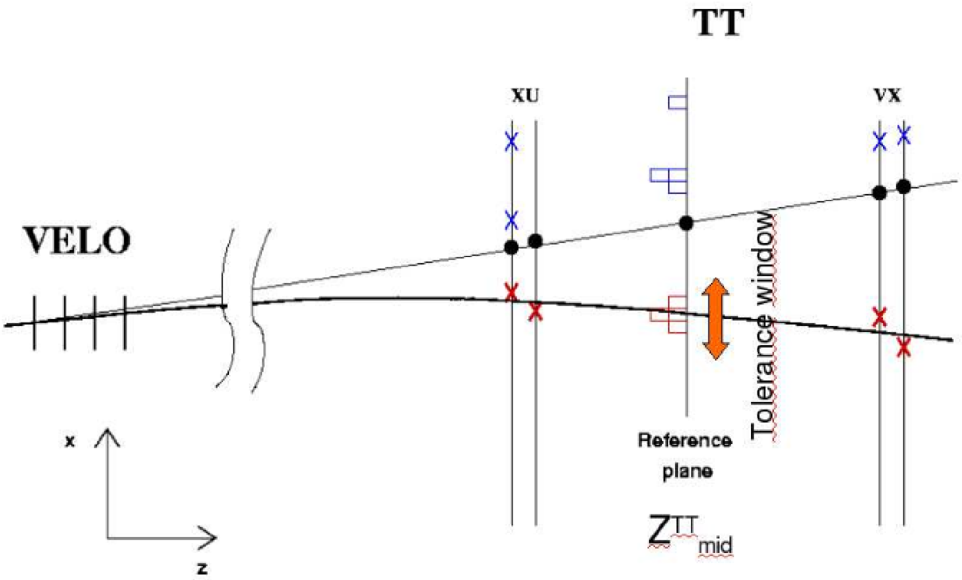


Figure 5-8.: Sketch of the upstream pattern recognition [68].

5.6.2. VELO-TT fit

Due to execution time limitations and the fact that Kalman fit could not be used in on-line reconstruction (HLT environment), a simplified VELO-TT fit has been implemented. This implementation starts with a dedicated least square fit to determine the track momentum, where a pseudo χ^2 is used to select the best combinations of TT hits to make the track. The contributions of multiple scattering are not taken into account. Therefore the χ^2 is not statistically correct.

The algorithm works under the assumption that track segments in VELO and TT can be extrapolated by straight lines which converge at one point. As the magnetic field is negligible in the VELO region, the trajectories there are well described by straight lines. However, the condition in TT stations is limited to track which momentum is above 3 GeV. The low momentum region is duly considered in the off-line reconstruction by the Kalman fit. The two linear track segments intersect each other at the point (x_{bdl}, z_{bdl}) in the x-z projection, where x_{bdl} is the coordinate at the middle of the effective magnet. In the y-$z$projection, there is no deflection, therefore, VELO and TT slopes are assumed to be the same. The total pseudo χ^2 is calculated as

$$\chi^2_{tot} = \chi^2_{VELO} + \chi^2_{TT}, \qquad \text{where} \tag{5-31}$$

$$\chi^2_{VELO} = \left(\frac{\Delta t_x^{VELO}}{\sigma_{t_x^{VELO}}}\right) = f_1(x_{bdl})^2 \qquad \text{and} \tag{5-32}$$

$$\chi^2_{TT} = \sum_{i=1}^{n}\left(\frac{\Delta x_i}{\sigma_{x_i}}\right)^2 = f_2(x_{bdl}, t_x^{TT}). \tag{5-33}$$

The first and the second components come from the uncertainty on the track slope measured in the VELO and the χ^2 of the straight line in TT respectively. The product of the inverse momentum and the charge of the track can be determined in the x-z projection by the expression:

$$\frac{Q}{P_{xz}} = \frac{3{,}3356 \times (\sin\theta_{TT} - \sin\theta_{VELO})}{\int Bdl}, \tag{5-34}$$

where θ_{TT} and θ_{velo} indicate angles of the track in the XY projection in TT and VELO, respectively. The total momentum can be found using the slope of the track in the y-z projection as

$$P = P_{xy} \times \sqrt{1 + (S_y^{TT})^2}. \tag{5-35}$$

The complete tuning procedure for on-line and off-line reconstruction can be found in Ref. [68].

5.7. Performance of the track reconstruction algorithms

The pattern recognition algorithms discussed in previous sections work in a collaborative environment in order to reconstruct as many tracks as possible. Figure **5-9** shows a scheme of the tracking algorithm collaboration. The main goal is to maximize the efficiency of each tack while minimizing the ghost rate. Separate efficiencies are defined for each pattern recognition algorithm. The tracks which are found by the track reconstruction pattern algorithms are successfully reconstructed if at least 70 % of the assigned hits belong to the same particle.

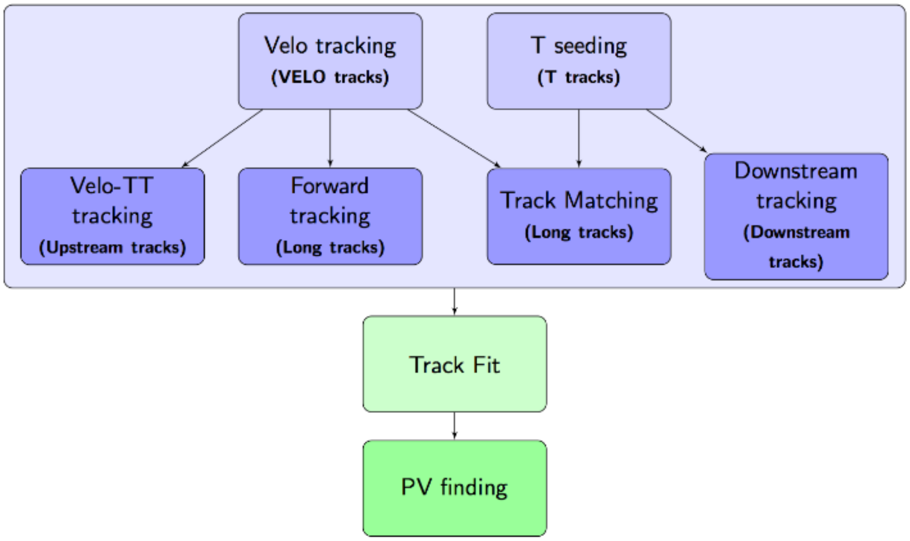


Figure 5-9.: Schematic diagram of the tracking algorithms in LHCb [41].

The tracks reconstruction efficiency can be calculated as

$$\epsilon_{track} = \frac{N_{correct}}{N_{MC}}, \tag{5-36}$$

where the $N_{correct}$ is the number of reconstructed tracks and N_{MC} is the total number of contractible particles. The efficiency error is calculated according to the binomial distribution

$$\sigma_{\epsilon} = \frac{\sqrt{N_{correct}(1-\epsilon_{track})}}{N_{MC}}. \tag{5-37}$$

The ghost rate and its corresponding error are calculated by

$$g_{track} = \frac{N_{ghost}}{N_{tot}}, \tag{5-38}$$

$$\sigma_g = \frac{\sqrt{N_{ghost}(1-g_{track})}}{N_{tot}}, \tag{5-39}$$

where N_{ghost} is the number of reconstructed tracks that cannot be associated with an MC particle and N_{tot} is the total number of reconstructed tracks. The average number of reconstructed tracks in a single $b\bar{b}$ event is 106. This number corresponds to 34 VELO tracks, 33 long tracks, 19 T tracks, 6 upstream tracks, and 14 downstream tracks.

5.7.1. Efficiency and ghost rate

The efficiency and ghost rate for each track type is described as follows.

VELO tracks: The remaining VELO seeds that cannot be matched to any long or upstream track are called VELO tracks which are useful for primary vertex reconstruction. They have an average of 9,3 hits. the efficiency is about 50 % and the ghost rate is approximately 20 %.

Long tracks: A typical long track has 35,6 measurements on average which correspond to 12,7 on VELO, 3,0 in TT, 2,4 in IT and 17,5 on OT. These are average numbers. If the track only traverses the OT region, the average number of measurements in T stations is 21,3. On the other hand, if the track only traverses the IT region the average number of measurements is 11,8. 75 % of the long tracks pass through the OT region, 12 % pass through the IT region and the rest traverse both regions. The average efficiency for long tracks in $b\bar{b}$ events is $(91{,}34 \pm 0{,}05)$ % of which $(3{,}35 \pm 0{,}03)$ % corresponds to matched tracks. There is an increment of the efficiency as a function of the momentum. For $P > 5$ GeV the efficiency reaches the value of $(94{,}31 \pm 0{,}08)$ %. The decay products of B mesons use to have high momentum which leads to an efficiency of $(93{,}1 \pm 0{,}1)$ %. Another crucial quantity to take into account is the ghost rate which is $(11{,}53 \pm 0{,}06)$ % and for $P > 5$ GeV, it is $(7{,}65 \pm 0{,}05)$ %. The multiplicity is directly correlated with efficiency and ghost rate. The higher the event multiplicity, the more difficult it becomes to assign the correct hits to the track candidates. The relative hit multiplicity, n_{rel}, is defined as

$$n_{rel} = \frac{1}{4}\left(\frac{n_{VELO}}{\langle n_{VELO}\rangle} + \frac{n_{TT}}{\langle n_{TT}\rangle} + \frac{n_{IT}}{\langle n_{IT}\rangle} + \frac{n_{OT}}{\langle n_{OT}\rangle}\right) \tag{5-40}$$

where n_{xx} is the number of hits in a given subdetector. The efficiency and the ghost rate can be written in terms of the n_{rel} as

$$\epsilon_{track} = (91{,}2 - 2{,}4 n_{rel})\,\%, \tag{5-41}$$

$$g_{track} = (1{,}5 + 8{,}0 n_{rel})\,\%. \tag{5-42}$$

The efficiency slowly decreases with an increasing multiplicity, while the ghost rate rises more steeply. Nevertheless, the reconstruction of long tracks proves to be robust against higher hit multiplicities [49].

T tracks: The T seeds that were not used to reconstruct match or downstream tracks are defined as T tracks. The have an average of 16,8 hits, 12,7 in OT and 4,1 in IT. The efficiency is about 55 %, with its corresponding ghost rate of 28 %. The inefficiency is due to secondary interactions which have typically low momentum. For $P > 5$ GeV, the efficiency reaches to 72 %.

Upstream tracks: The upstream tracks have an average of 11,3 hits in the VELO and 3,9 hits in the TT stations. In the region $P > 1$ GeV the efficiency is about 75 % and the ghost rate 15 %. The searches below that momentum region lead to high ghost rates and in consequence low efficiency.

Downstream tracks: The downstream tracks have an average of 3,2 hits in the TT stations. They are mainly used to reconstruct pions coming from K_s^0 mesons. Many of these pions decay outside the VELO acceptance, only a quarter of the pions are reconstructible as long tracks. The efficiency to find both pions from such decay is 54 %. The single track efficiency is about 74 % with a ghost rate of about 26 %. For $P > 5$ GeV the efficiency reaches 82,8 %.

5.7.2. Background Category tool and association algorithm

The background category tool [4, 5] was designed to classify composite particles in simulated events as signal or any other background category. It does not classify tracks (final states) or photons, it associates a background code to each composite particle. Based on the properties of final state particles, composite particles fall in any of the categories. The algorithm works by rejecting the hypothesis in order of priority which the first priority is the ghost background. The Tab. **5-2** contains various background categories and conditions to classify composite particles. When the association algorithm fails to match a final state particle (track) with any MC object, The tracks are classified as a ghost. All possible tracks at LHCb correspond to quasi-stable particles which are: electron, photon, muon, pion, kaon, Proton or neutrino. There are two conditions to make the link between the track and the MC object taking into account the nature of the track. For **charged tracks**, the algorithm requires at least 70 % of hits in common between the reconstructed track and the MC object. For **neutral clusters** the MC particle should contribute at least 50 % to the reconstructed cluster energy. The background category association tree is shown in figure **5-10**.

Priority	Background name and code	Description
1	Ghost (60)	One or more of the final state tracks/clusters is not matched to any MC truth object (track/neutral)
2	Clone (63)	One or more of the final state tracks/clusters are matched to the same MC truth object (track/neutral)
3	Hierarchy (66)	One of the final state particles is combined with its own true parent
4	Pileup (100)	The MC truth objects matched to the final state tracks come from two or more different primary vertices
5	Combinatorics (70/80)	At least one of the MC truth objects matched to the final state particle comes from a primary vertex. If all of the MC truth objects come from the same primary vertex, this is 80, else 70.

6	bbar (110)	At least one of the MC truth objects matched to the final state particle comes from the decay of a beauty hadron.
7	ccbar (120)	At least one of the MC truth objects matched to the final state particle comes from the decay of a charm hadron.
8	light quark (130)	This is a combinatorial background and none of the above conditions are satisfied
9	Low mass (50)	All of the final state tracks/clusters are associated with an MC truth object with the correct particle ID (no mis-identification), but the parent decay has not been fully reconstructed. Also, the true common parent must have a mass at most 100 MeV above the mass of what we are trying to reconstruct.
10	Partially reconstructed (40)	The parent decay has not been fully reconstructed and the low mass condition is not satisfied.
11	Reflection (30)	One or more of the final state tracks/clusters are associated with an MC truth object with a different particle ID.
12	Fully reconstructed (20)	The decay is correctly and fully reconstructed but the particle ID of the parent particle is mistaken.
13	Quasi-Signal (10)	The decay is correctly and fully reconstructed but an intermediate resonance or particle is misidentified.
14	Signal (0)	None of the previous conditions have been satisfied.

Table 5-2.: Description of the background categories at the LHCb experiment [5].

5.7.3. Ghost tracks

The amount of ghost tracks in the LHCb detector is not a negligible quantity as it was discussed in previous sections. Although VELO has a low ghost rate, T stations are more sensitive to mis-reconstruction due to the smaller number of detection layers and the hits induced by secondary particles and detector noise. This entails an increment of the mismatched tracks.

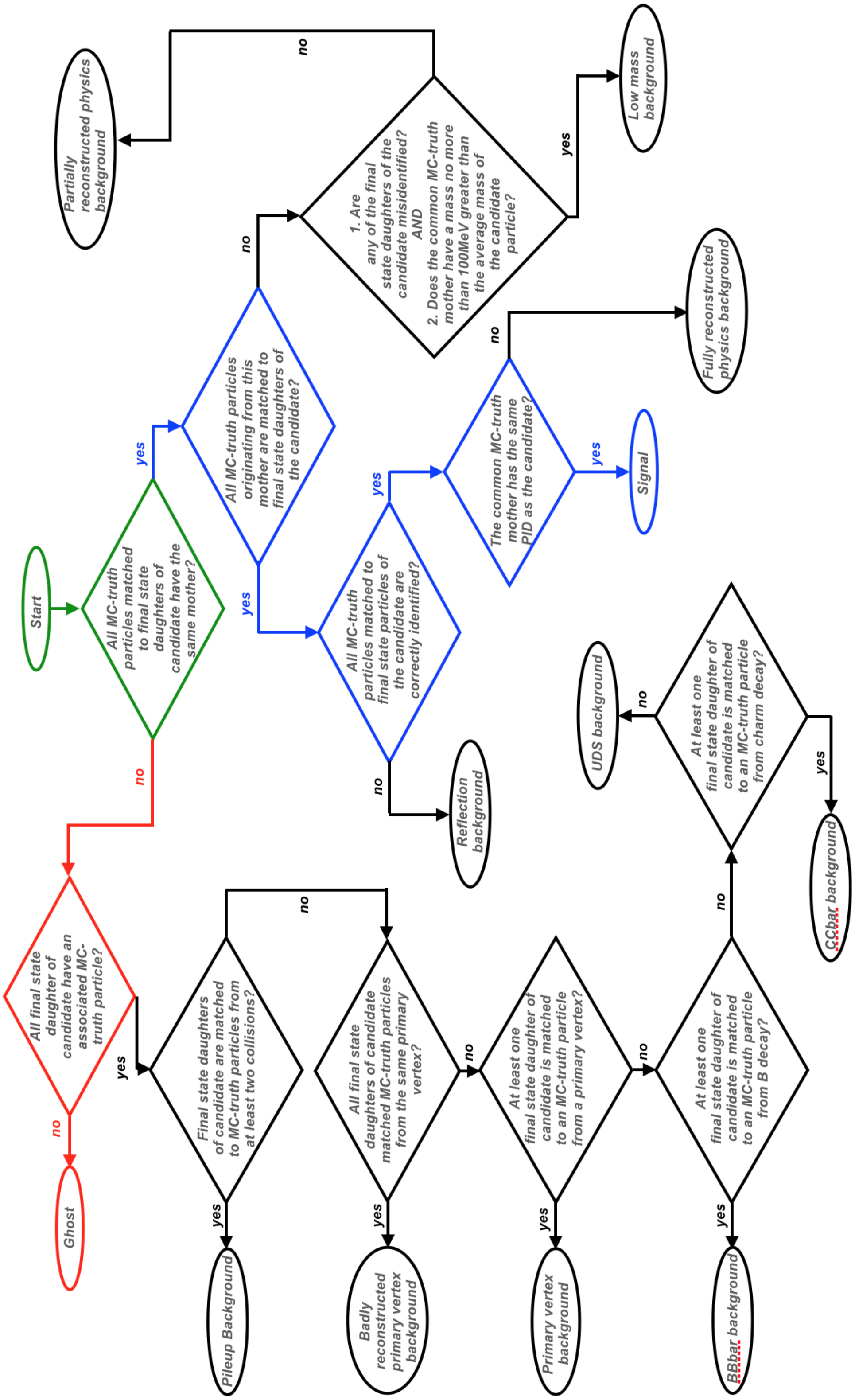
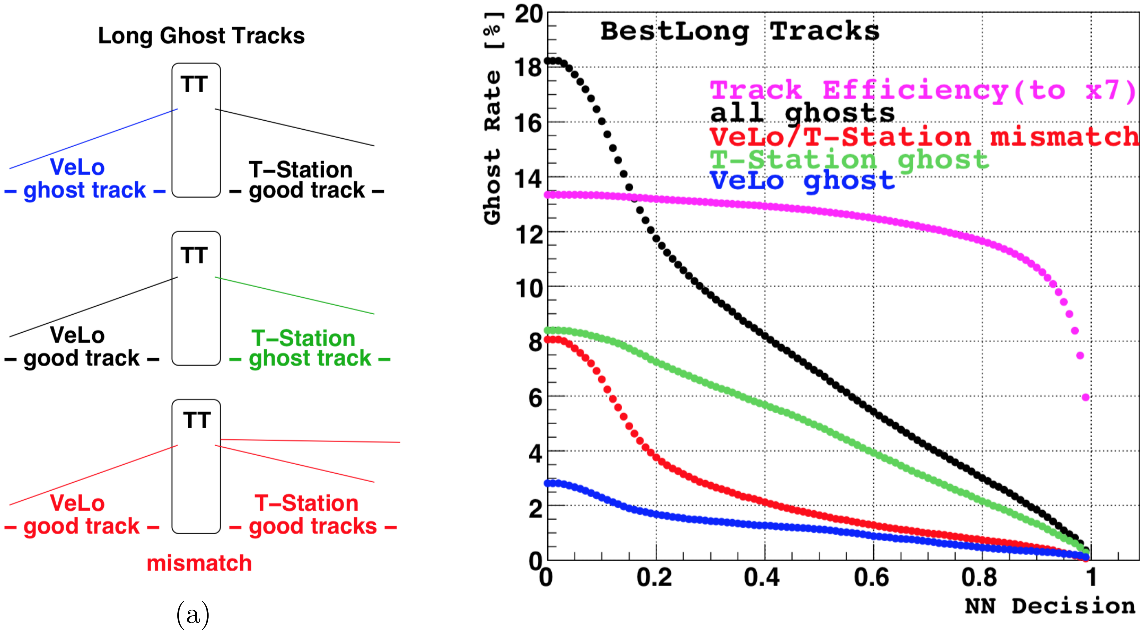


Figure 5-10.: Background category algorithm. Adapted from [5].



(a)

(b)

Figure 5-11.: Classification of ghost sources for long tracks (a) and Ghost types Rate evolution as a function of the NN decision. The corresponding Track Efficiency, down-scale factor 7 (b) [76].

All reconstructed tracks are grouped into several groups. Within this thesis, the most important groups are signal and ghost. The other categories are treated as background no matter which source they come from. There are several sources of ghost tracks as shown in Figure **5-11**a. The most important in the creation of long ghost tracks are: **hadronic interactions** of particles with the detector material and the beam pipe. If a good VELO track is extrapolated to a T track induced by hadronic interactions, it results in a ghost track. The second source is **mis-matching tracks** in which a good track segment is matched to a ghost track or two true tracks from different particles are matched. Assigning the wrong hits in the TT stations also produce ghost tracks. The LHCb experiment has developed several ways to deal with the number of ghost tracks during the years. The first attempts were to use the χ^2 and the likelihood methods as discriminant [77]. After that, a neural network (NN) algorithm was implemented in the offline reconstruction [76]. Finally, the algorithm was improved becoming it fast enough to include it in the on-line reconstruction [78].

6. Results

Figure 6-1 shows the ghost candidates introduced during the reconstruction chain by the algorithms, where the distribution for the ϕ particle peaks at 1020 MeV which is the nominal particles mass. It is important to remind that a ghost candidate is a particle that does not correspond to any generated particle in a MC sample. Hence, the expected distribution should be flat since the ghost particles do not have any physical meaning. Therefore that peaking distribution could contain some tracks where their information has been lost during the complicated reconstruction process. Our main objective is centered on discovering where the information could be accidentally erased by the algorithms or if those particles are real ghost candidates.

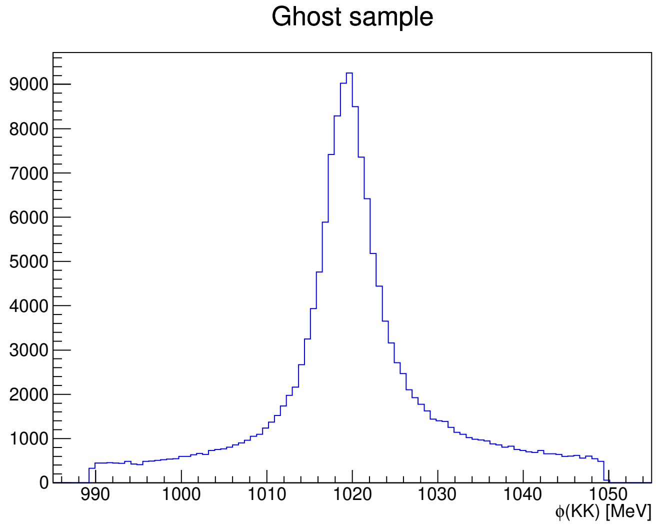


Figure 6-1.: ϕ reconstructed mass distribution for the ghost candidates.

In this chapter, the discussion will be centered on solving our main problem which is the peaking distribution just in the mass of particles for ghost candidates. Our hypothesis says those events are signal events that were miss-categorized at some point in the reconstruction chain. It means, a lack of information is causing the peak, therefore the reconstruction algorithms need to be debugged, to find where the algorithm is failing, and improved if it is required.

In the development of this work, several data analyses were carried out in order to reveal where the problem was. In the first tests, the distributions of the kinematic variables were studied to find the differences. In the second part, changes in the reconstruction algorithm chain were tested to make an improvement and increase the signal reconstructed while reducing the ghost candidates. A loose selection was applied for all data samples, $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$, $B^0 \to J/\psi(\mu^+\mu^-)K_s^0(\pi^+\pi^-)$ and $D^{*+} \to D^0(K^-\pi^+)\pi^+$. Furthermore, an additional data sample, $B^0 \to \mu^+\mu^-K^*(K^+\pi^-)$, was used to confirm the results with muons, kaons and, pions. The selected data samples correspond to Run I data.

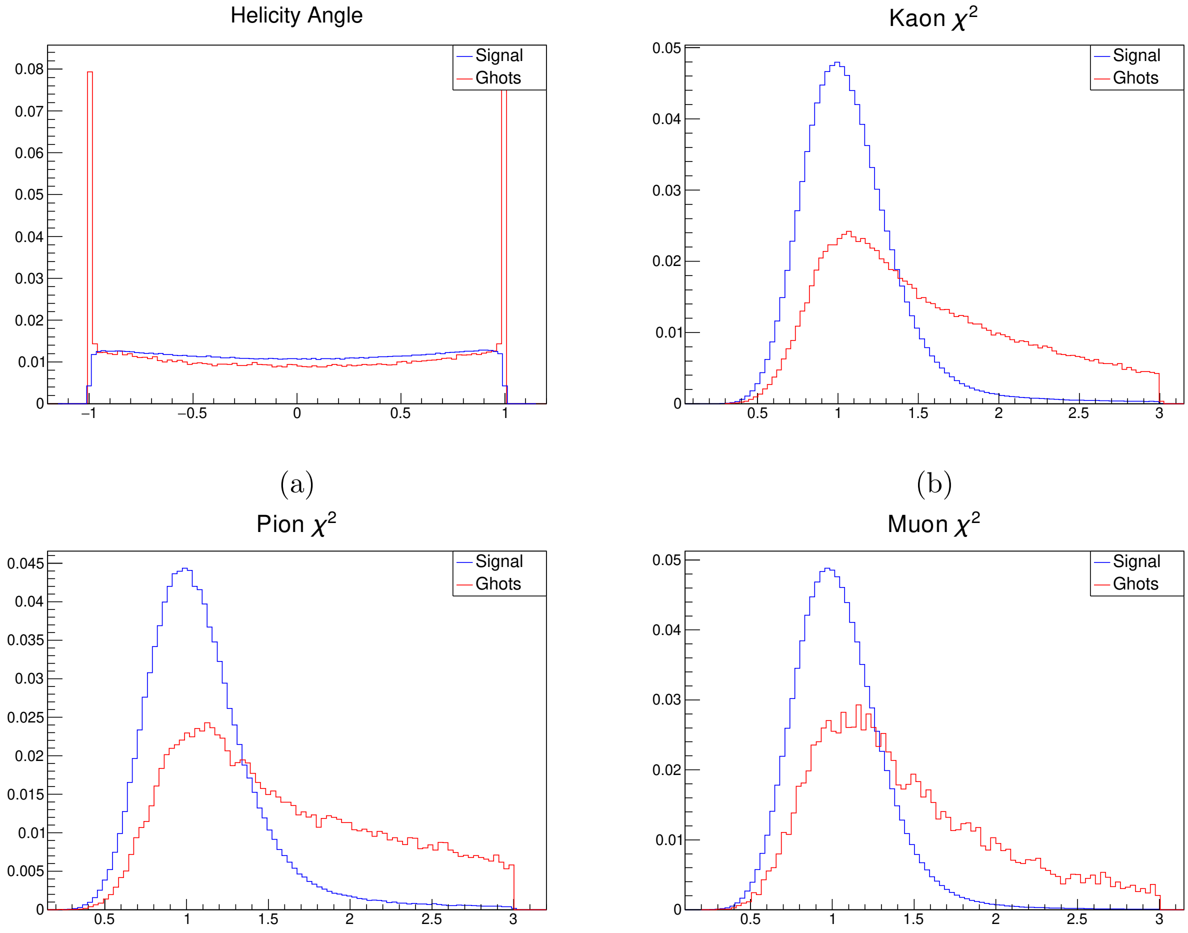
Data sample	cuts	year
$B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$	$\lvert\Delta M_\phi\rvert \leq 30$ MeV $\lvert\Delta M_{J/\psi}\rvert \leq 50$ MeV $\lvert\Delta M_{B_s^0}\rvert \leq 200$ MeV $\frac{\chi^2_{\text{v}}}{\text{v}_{ndof}} \leq 10$	2011 and 2012
$B^0 \to \mu^+\mu^-K^*(K^+\pi^-)$	$\lvert\Delta M_{K^*}\rvert \leq 60$ MeV $\lvert\Delta M_{B^0}\rvert \leq 200$ MeV $\frac{\chi^2_{\text{v}}}{\text{v}_{ndof}} \leq 10$	2012
$B^0 \to J/\psi(\mu^+\mu^-)K_s^0(\pi^+\pi^-)$	$\lvert\Delta M_{K_s^0}\rvert \leq 30$ MeV $\lvert\Delta M_{J/\psi}\rvert \leq 50$ MeV $\lvert\Delta M_{B^0}\rvert \leq 300$ MeV $\frac{\chi^2_{\text{v}}}{\text{v}_{ndof}} \leq 10$	2011 and 2012
$D^{*+} \to D^0(K^-\pi^+)\pi^+$	$\lvert\Delta M_{D^0}\rvert \leq 100$ MeV $\lvert\Delta M_{D^{*+}}\rvert \leq 100$ MeV $\frac{\chi^2_{\text{v}}}{\text{v}_{ndof}} \leq 10$	2012

Table 6-1.: Event selection criteria applied on tracks and particles for the selected MC data samples.

Where the $\frac{\chi^2}{V_{ndof}}$ is the vertex χ^2 per degree of freedom which represents the vertex quality. It was set to be a maximum of 10. The mass window, ΔM, is the maximum deviation allowed of the particles' nominal mass, and the standard loose algorithm was used to make the first selection of all particles.

6.1. Checking variables

All kinematic variables were checked in the search for differences between the signal and ghost candidate distributions. The background categories 0 and 60 were used to make the selection of the particles in which 0 means real signal and 60 represents the incorrectly reconstructed particles.

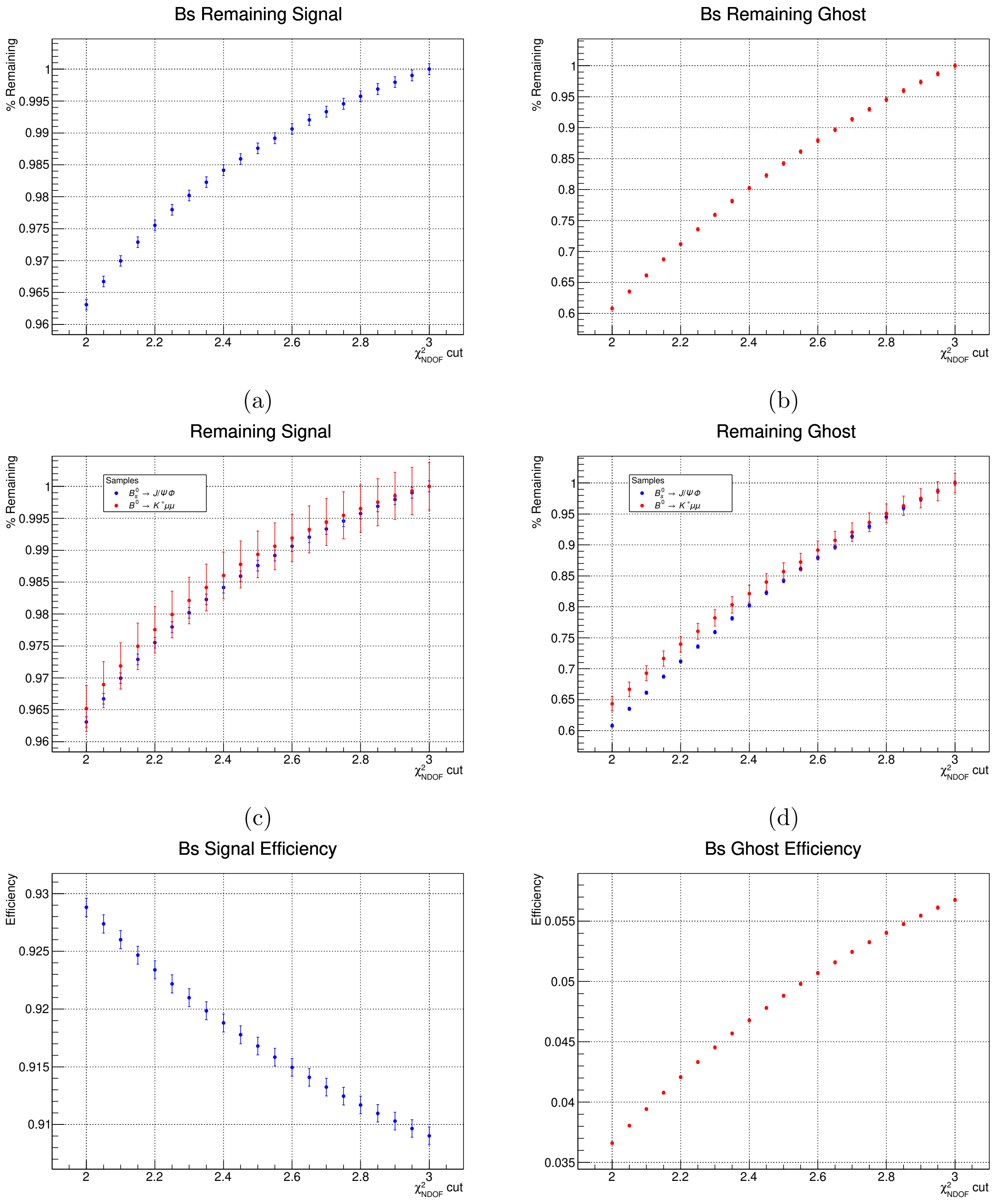


(a) (b)

(c) (d)

Figure 6-2.: Kinematic variables which presented large differences.

The distributions for the ϕ child, Figure **6-2**a, showed peaks in the forward and backward region for ghost candidates. However posterior analysis demonstrated that the peaks in the helicity angle plot do not correspond to the events in the peak of the mass distribution plot so these are in fact real ghost tracks. Another important difference is the χ^2_{ndof} for kaons **6-2**b, pions **6-2**c and muons **6-2**d. This variable has a default cut $\chi^2_{ndof} < 3$. Nevertheless based on the evidence, it possibly needs to be redefined to reduce the number of ghost candidates.



(a) (b)

(c) (d)

(e) (f)

Figure 6-3.: Remaining reconstructed signal and ghost depending on the χ^2_{ndof} cut (a),(b),(c) and (d), and reconstruction efficiency (e) and (f).

The results showed in Figure **6-3**, in which a significant reduction of ghost candidates with a small signal loss depending on the χ^2_{ndof} criterion is observed, encourage the posterior analysis to isolate regions that maximize the ghost reduction and minimize the signal loss. For instance, if $\chi^2_{ndof} = 2$, the ghost reduction is up to 40 % and the signal lost is about 5 %. Furthermore the signal efficiency increase $\sim$ 0,4 %. The majority of the ghost tracks come from hadron final states. Although the effect of muons is much smaller, it cannot be neglected.

6.2. Adaptive binning study

An adaptive binning map for ghost and signal candidates was made looking for important zones were signal and ghost candidates accumulate. These studies let on average the same number of events in each bin. To do this, three variables were chosen; the number of tracks per event ($nTracks$), the pseudo-rapidity (η), and the transverse momentum (P_T). They were compared in pairs of variables to identify any possible region in which the distributions between signal and ghost candidates differ.

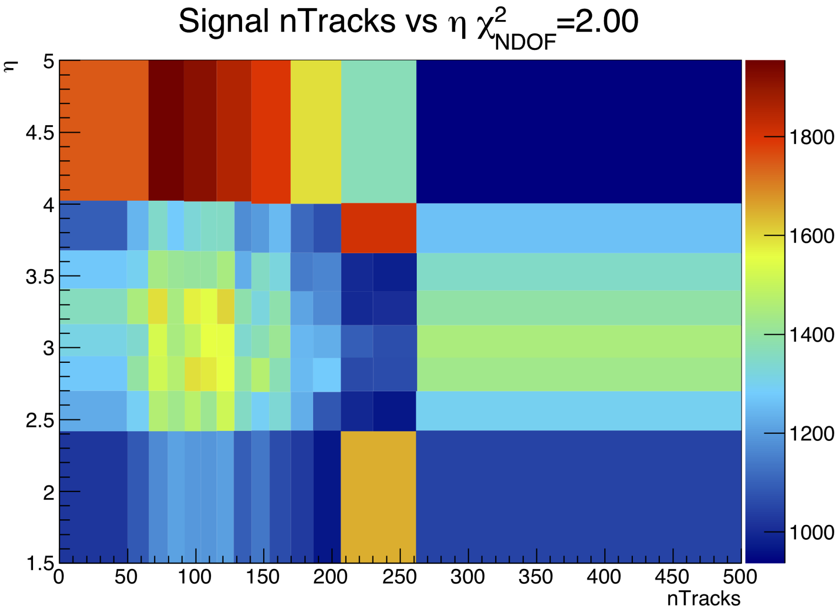
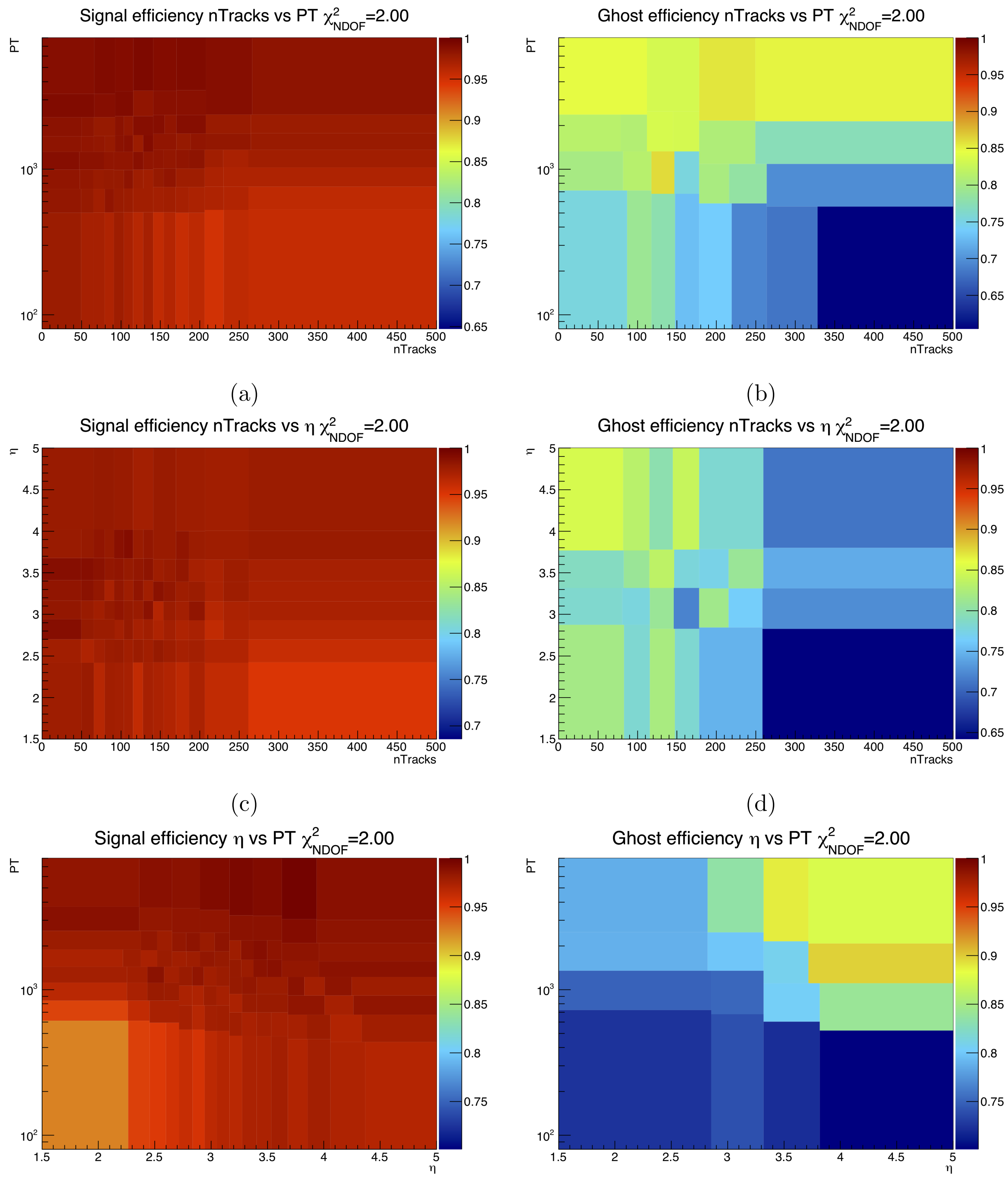


Figure 6-4.: Bin map η vs $nTracks$.

The algorithm was designed to divide each region into four if the number of events in that zone surpasses a maximum. This maximum number can be tuned according to the number of total events and the desired granularity. Figure 6-4 shows the output of the algorithm in which the number of events per bin is between 400 and 1000. It is important to note that the algorithm should avoid empty bins. Next, the efficiency bin-to-bin is calculated for each χ^2_{ndof} cut in an effort to find the optimum.


Signal efficiency nTracks vs PT χ^2_{NDOF}=2.00

Ghost efficiency nTracks vs PT χ^2_{NDOF}=2.00

Signal efficiency nTracks vs η χ^2_{NDOF}=2.00

Ghost efficiency nTracks vs η χ^2_{NDOF}=2.00

Signal efficiency η vs PT χ^2_{NDOF}=2.00

Ghost efficiency η vs PT χ^2_{NDOF}=2.00

(a) (b)

(c) (d)

(e) (f)

Figure 6-5.: Efficiency plots comparing signal and ghost candidates in a bin map.

We observed ghost candidates accumulate in the zone of high transverse momentum and high pseudo-rapidity (η). Meanwhile, there is no clear region where truth match events accumulate as shown in Figures 6-5e and 6-5f. Most of the ghost candidates accumulate at a high P_T. However, truth match[1] events accumulate in the same zone as shown in Figures 6-5a and 6-5b. Same for Figures 6-5c and 6-5d ghost and truth matched events accumulate in the same zones. Therefore we cannot assign different statistical weights to the zones due to the fact that we expected ghost and truth matched candidates to accumulate in different regions. Moreover cutting any of these variables lead to signal loss which is not the main objective.

6.3. Single and double ghost study

There are two sources of ghost candidates called single and double ghost. In the $\phi \to K^+K^-$ decay, the single ghost is given if one of the kaons is truth matched and the other is a ghost one, and double ghost, booth kaons are ghost ones. Based on the truth information available from the MC samples, we can distinguish which track is a ghost. In this case, the truth three momentum information is used.

Conditions	Category
$P_{x,true} \neq 0$ $P_{y,true} \neq 0$ $P_{z,true} \neq 0$	Truth matched
$P_{x,true} = P_{y,true} = P_{z,true} = 0$	Ghost

Table 6-2.: Categorization for truth matched and ghost tracks. $P_{x,true}$ refers to the truth information momentum in the x coordinate, and so on.

6.3.1. Ghost probability

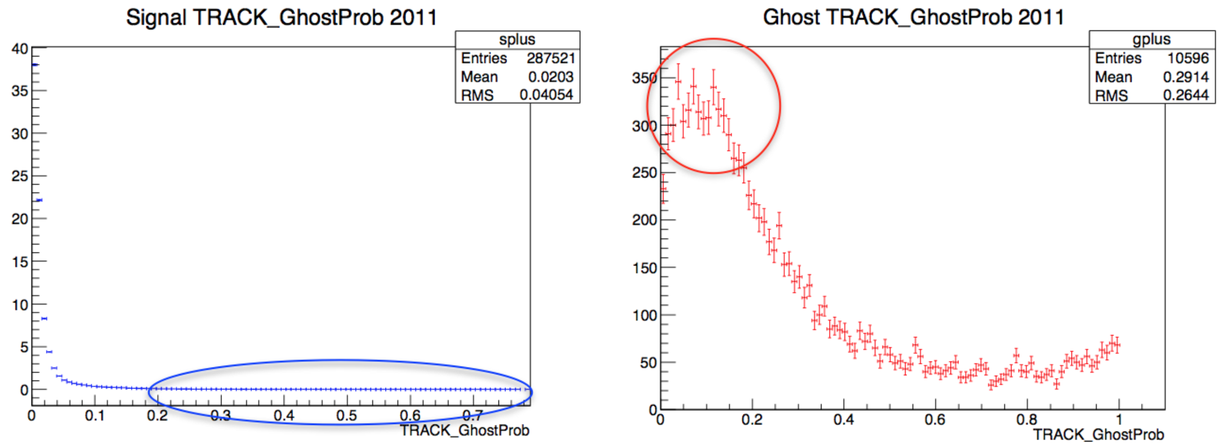
The algorithm called Ghost Probability (GP) was designed to discriminate the truth matched and the ghost candidates. However, as it has a probabilistic description it fails for some tracks. The algorithm cannot distinguish between ghosts and truth matched at low probabilities. About 5 % of the truth matched tracks have high GP and $\sim 50\,\%$ of the Ghost tracks fall in the low probability region. The behavior is repetitive for all data samples, the $B^0 \to J/\psi K_s^0$ is shown in Figure 6-6. GP $< 0{,}15$ is a cut that is usually used in reconstruction or selection to minimize the number of ghost candidates. However this cut discards truth matched events. It is an inevitable consequence.

Table 6-3 shows the number of events discarded for the decays $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$ and $B^0 \to J/\psi(\mu^+\mu^-)K_s^0(\pi^+\pi^-)$. It is easy to note that for a $GP < 0{,}15$, the number of events discarded by the algorithm is almost the same for ghost and signal candidates. The

---

[1] Signal events will be referred as truth matched events

algorithm discards worse reconstructed tracks due to the number of hits in the detector, although some of these tracks are true ones.



(a)

(b)

Figure 6-6.: Ghost probability distributions for truth matched (a) and ghost (b) candidates.

Sample	Year	Particle	Signal	Ghost	Double Ghost
$B_s^0 \to J/\psi(\mu^+\mu^-)\phi(K^+K^-)$	2011	K^+	$1{,}31 \pm 0{,}01\,\%$ 36249	$50{,}29 \pm 0{,}33\,\%$ 33453	5463 $0{,}18 \pm 0{,}01\,\%$
		K^-	$1{,}32 \pm 0{,}01\,\%$ 36348	$50{,}31 \pm 0{,}31\,\%$ 36992	
	2012	K^+	$1{,}50 \pm 0{,}02\,\%$ 42680	$53{,}77 \pm 0{,}32\,\%$ 42359	7383 $0{,}24 \pm 0,01\,\%$
		K^-	$1{,}51 \pm 0{,}02\,\%$ 42988	$53{,}76 \pm 0{,}32\,\%$ 46830	
$B^0 \to J/\psi(\mu^+\mu^-)K_s^0(\pi^+\pi^-)$	2011	π^+	$1{,}84 \pm 0{,}02\,\%$ 5300	$60{,}20 \pm 1{,}06\,\%$ 6379	1323 $0{,}38 \pm 0{,}01\,\%$
		π^-	$1{,}87 \pm 0{,}02\,\%$ 5377	$62{,}64 \pm 1{,}11\,\%$ 6902	
	2012	π^+	$2{,}10 \pm 0{,}03\,\%$ 5300	$66{,}10 \pm 1{,}07\,\%$ 6379	3474 $0{,}64 \pm 0{,}01\,\%$
		π^-	$2{,}13 \pm 0{,}03\,\%$ 9477	$67{,}38 \pm 1{,}08\,\%$ 15219	

Table 6-3.: Signal and ghost changes after applying the cut GP $< 0{,}15$. The numbers for signal and ghost show the amount of events discarded and its percentage relative and for double ghost the percentage is compared with the total data sample.

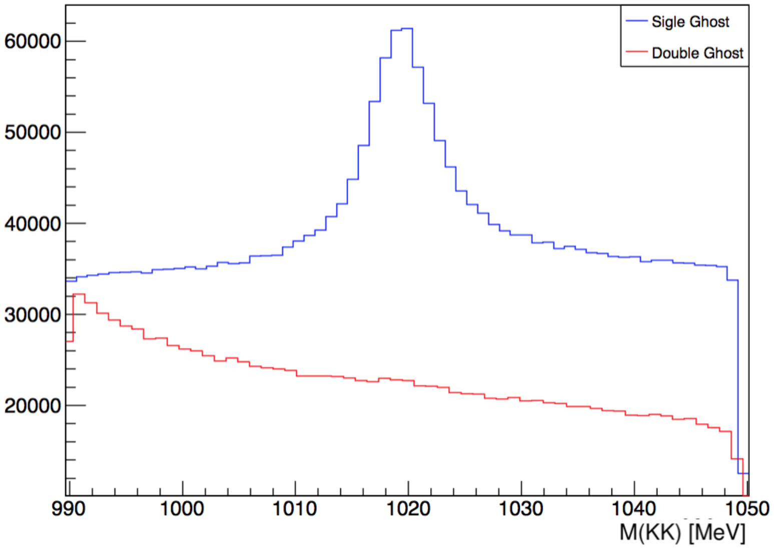
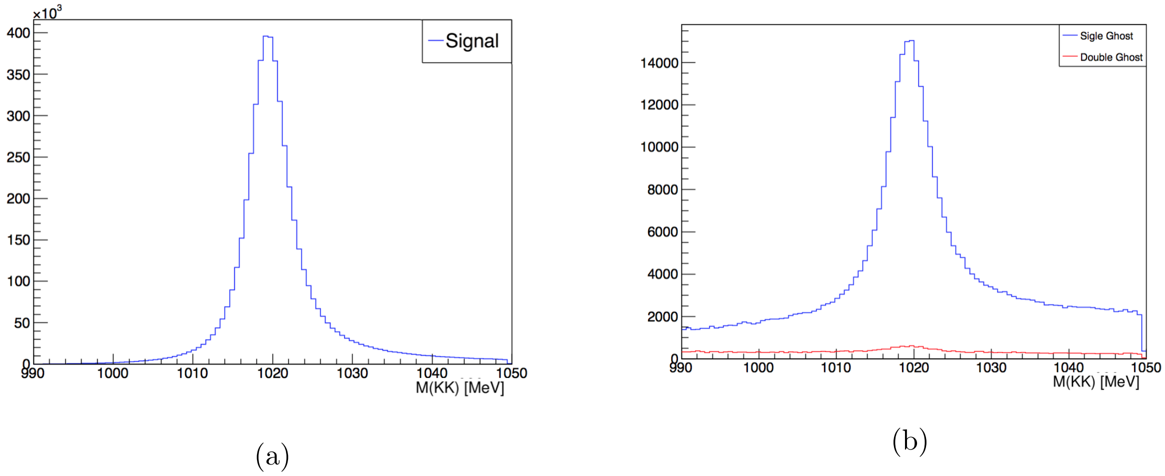


Figure 6-7.: $\phi(KK)$ distribution for single and double ghost candidates

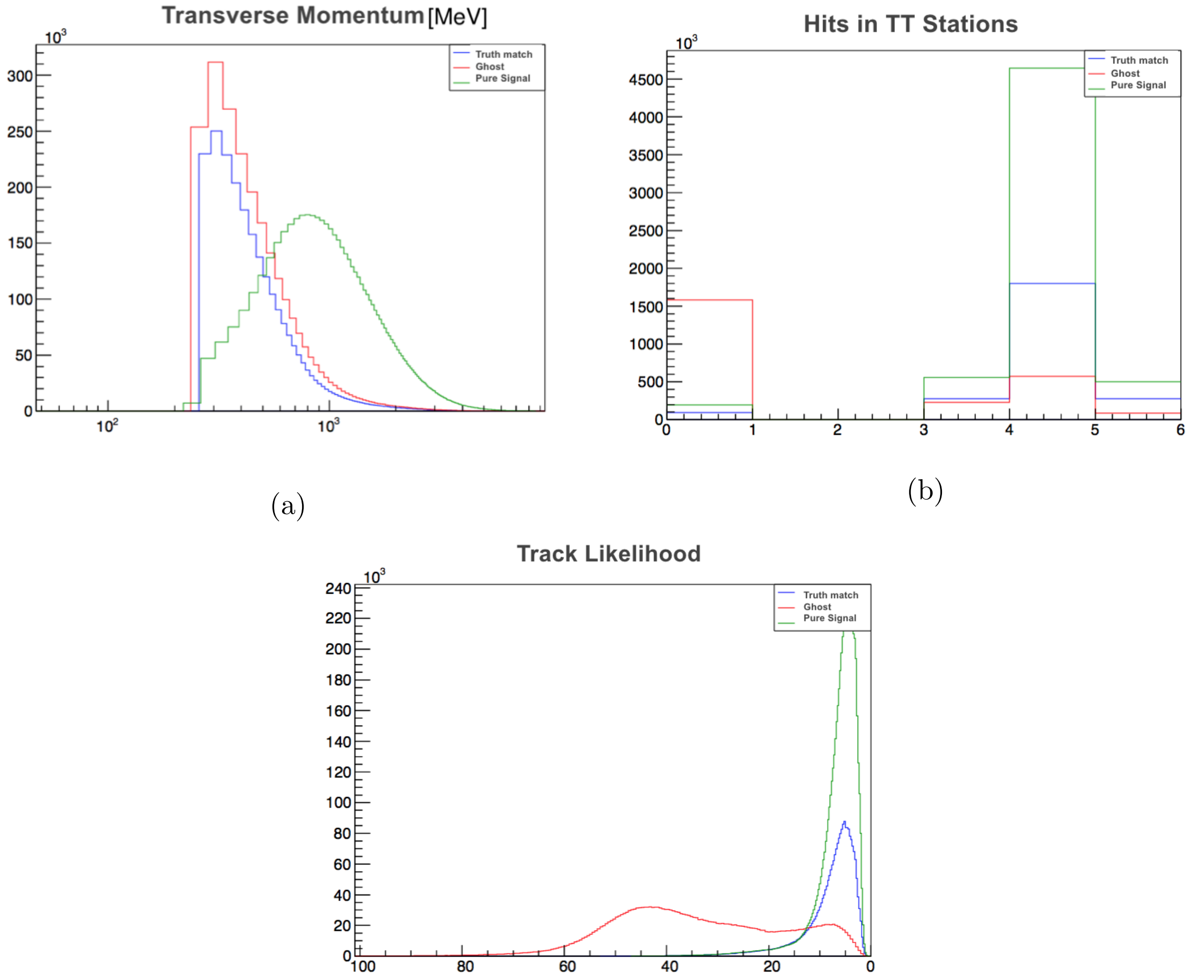
6.3.2. $\phi(KK)$ study

This analysis was performed with the B_s^0, 2012, data sample. However, only the $\phi \to K^+K^-$ was reconstructed. This data sample contains $\sim$ 11 million events with $\sim 40\,\%$ of ghost candidates.



(a) (b)

Figure 6-8.: Signal and ghost distributions after applying the cuts in P_T, nTThits and Likelihood.



(a)

(b)

(c)

Figure 6-9.: Track variables which presented the largest differences between the ghost candidate and the signal one, where pure signal refers to the candidate coming from a ϕ in which both are truth matched. Truth matched and ghost refers to the truth matched and ghost reconstructed coming from a ϕ ghost particle.

Double ghost events are the most interesting since they are the real ghost candidates. However, our analyses were centered in the single ghost events because double ghost events distribution does not show any peak. Nonetheless, the single ghost events do show a signal-like peaking distribution. So the next analysis was performed only using the single ghost events which are the ones that could be recovered as signal events. The ghost and the signal candidate are plotted for each ϕ. The major differences were found in the transverse momentum of the track, P_T **6-9**a, the number of hits in the TT stations **6-9**b and the log-likelihood of the track **6-9**c. Based on the information shown in the Figure **6-9**, the possibility of applying the cuts, nTThits$\neq$ 0, $P_T > 500$ MeV and Likelihood> -35, on these variables

to reduce the background in the single ghost distribution. The events outside these regions are discarded. The cuts applied to reduce 90 % of the double ghost candidates and reduce the background in the single ghost distribution while keeping as many events in the peak as possible.
In order to make a comparison, the distributions showed in Figure **6-8** were studied. The same selection criteria were applied to both distributions to determine their compatibility. However, there exist plenty of events in the background for the ghost distribution which difficult the selection of the wrong categorized information. This result tells us that the two distributions are not compatibles so these ghost candidates could be in fact real ghosts. However, it is to soon to discard them as a wrong categorized signal.

6.4. BDT study

To confirm the hypothesis in the previous section the Boosted Decision Tree (BDT) [79] analysis was performed to isolate the peaking ghost candidates. This analysis was done with the complete B_s^0 decay chain in which the ϕ candidates were selected. The events in the peak were called signal ghost and the events far from the peak were the true ghost as shown in Figure **6-10**. The selection for signal and ghost samples were made by using the truth information once.

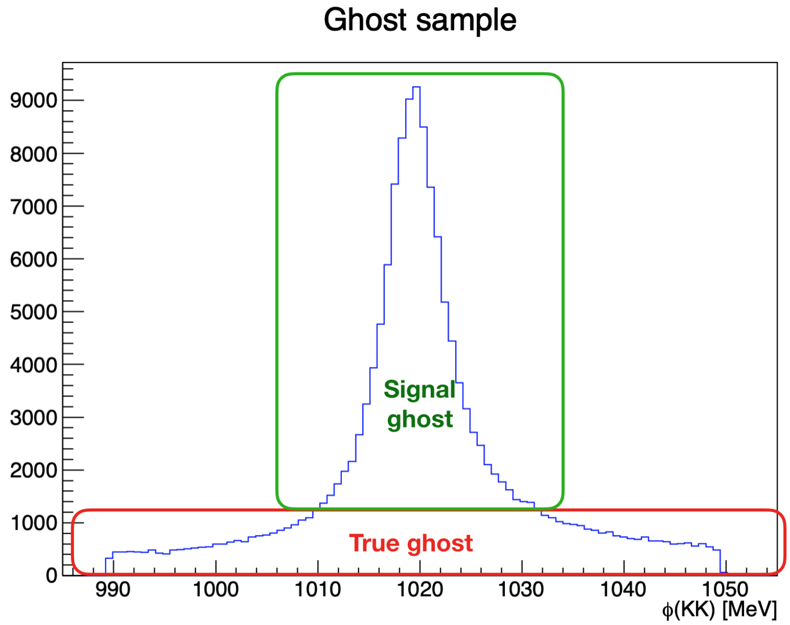


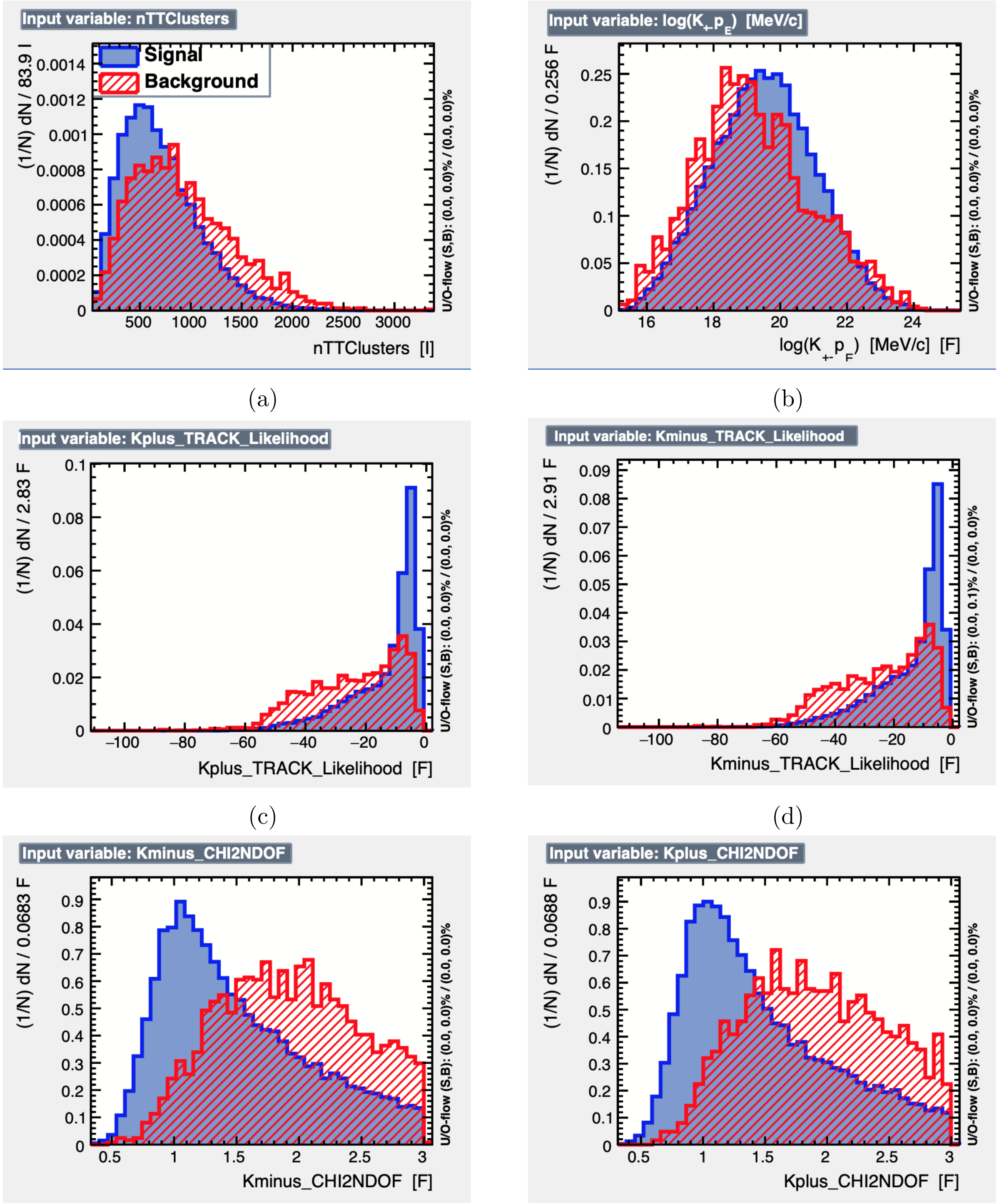
Figure 6-10.: Expected re-categorization from the BDT study.

The variables, Likelihood, χ^2_{ndof}, Key[2], the logarithm of the ghost probability, the number

---

[2] The Key variable is an identification number assigned to all generated particles

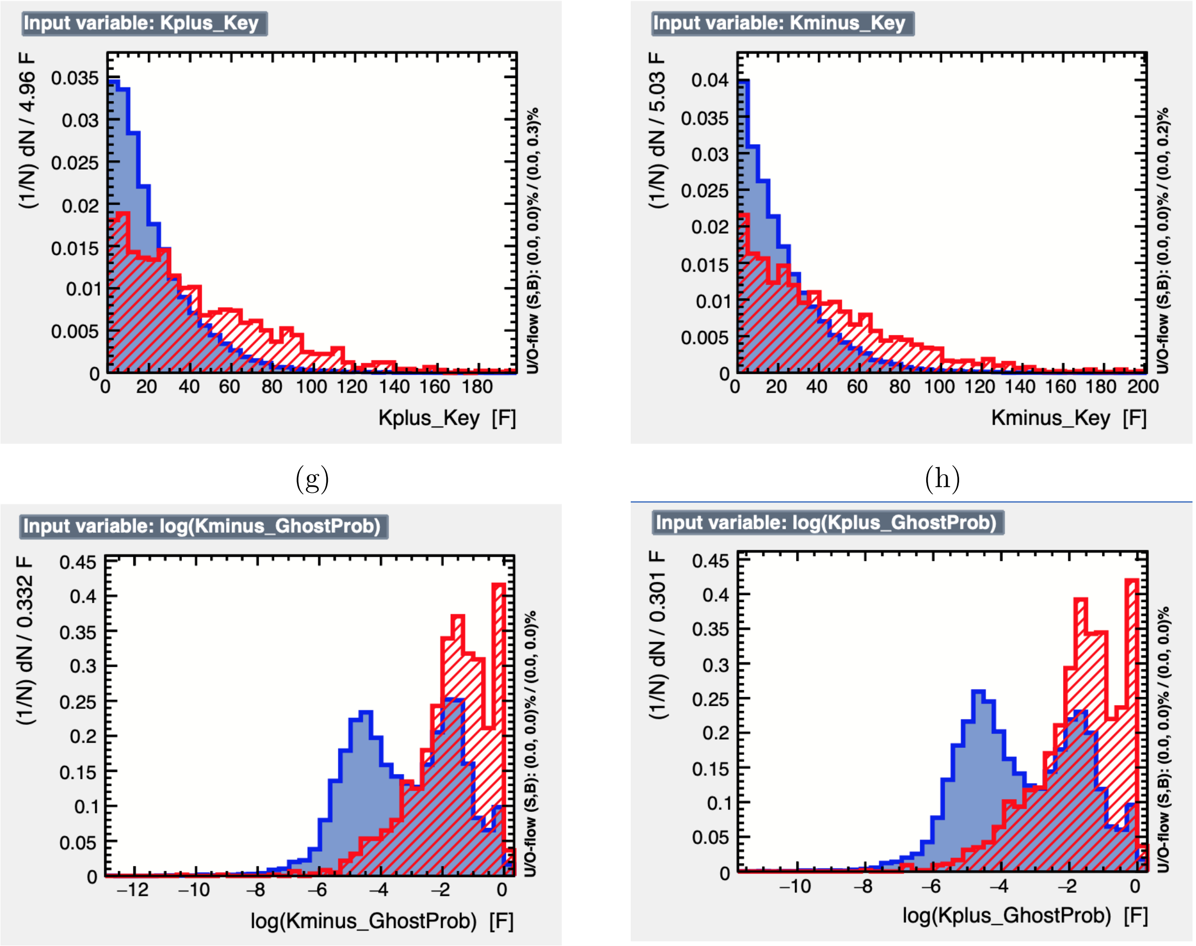
of clusters in the TT stations and $\log P_E$, were the ones highly correlated therefore their large discrimination power. The distributions for signal and background are shown in Figure 6-11. It is important to notice that for this case signal refers to the signal-ghost events and background refers to the true ghost events.



(a) (b)

(c) (d)

(e) (f)

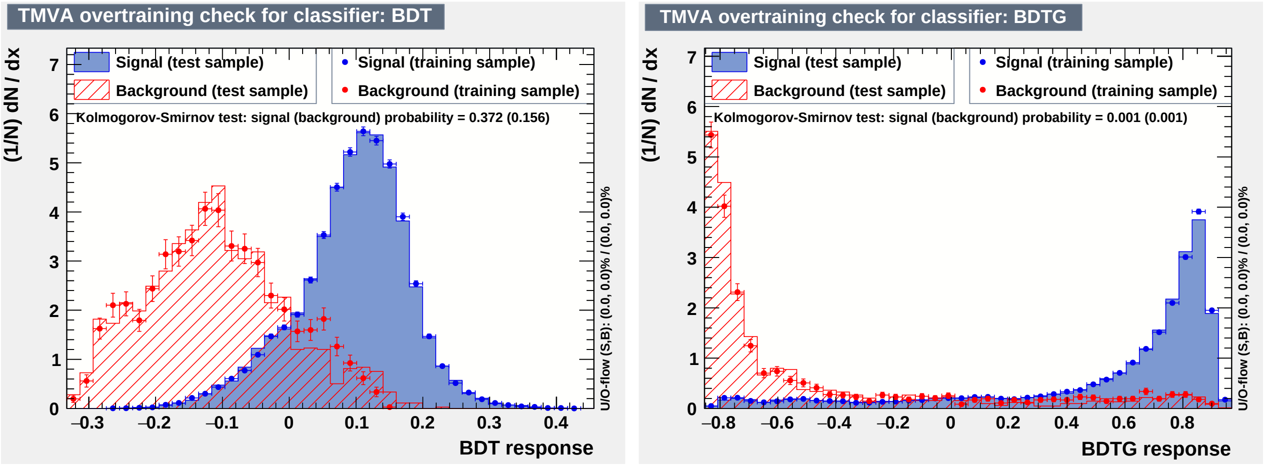


(g) (h)

(i) (j)

Figure 6-11.: Input variables for the BDT algorithm

The BDT 6-12a and the Gaussian BDTG 6-12b were calculated taking the input variables information which shows a high discrimination power, the BDTG has an excellent behavior to discriminate between signal and ghost. Figure 6-13 shows the significance, the purity and the product of the purity and the significance in order to determine the best criteria to apply to the BDTG to make the posterior analysis. However, after fitting the true signal and the signal ghost distributions for the BDTG=0 cut as shown in figures 6-14a and 6-14b respectively, the width of the distributions, $\sigma_{\text{true signal}} = 2{,}13 \pm 0{,}02$ and $\sigma_{\text{signal ghost}} = 2{,}33 \pm 0{,}20$, reveal a worst reconstructed signal ghost distribution which is expected due to the fact that these events came from tracks that contain less hits in the detector.

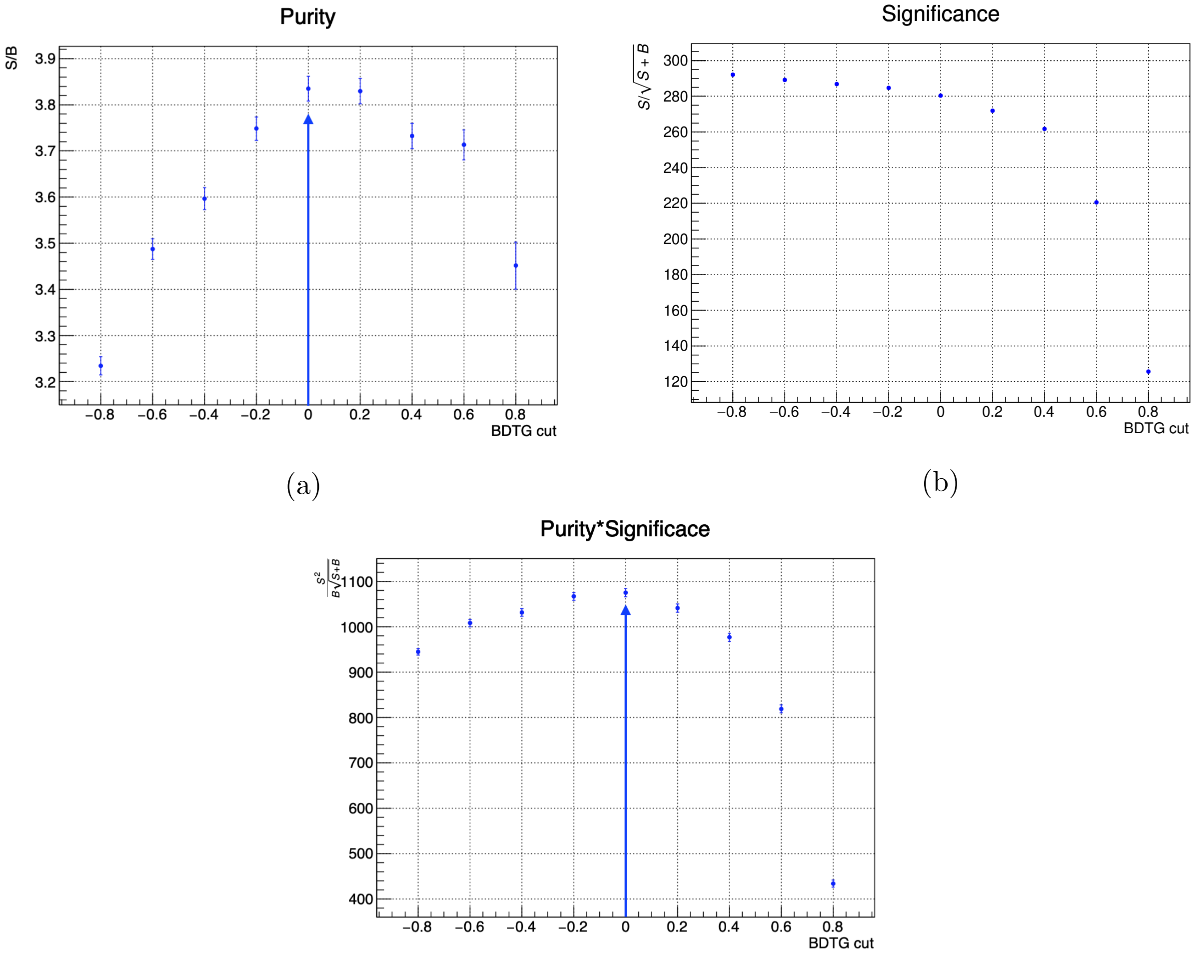


(a)

(b)

Figure 6-12.: The response for the BDT (a), BDTG (b)

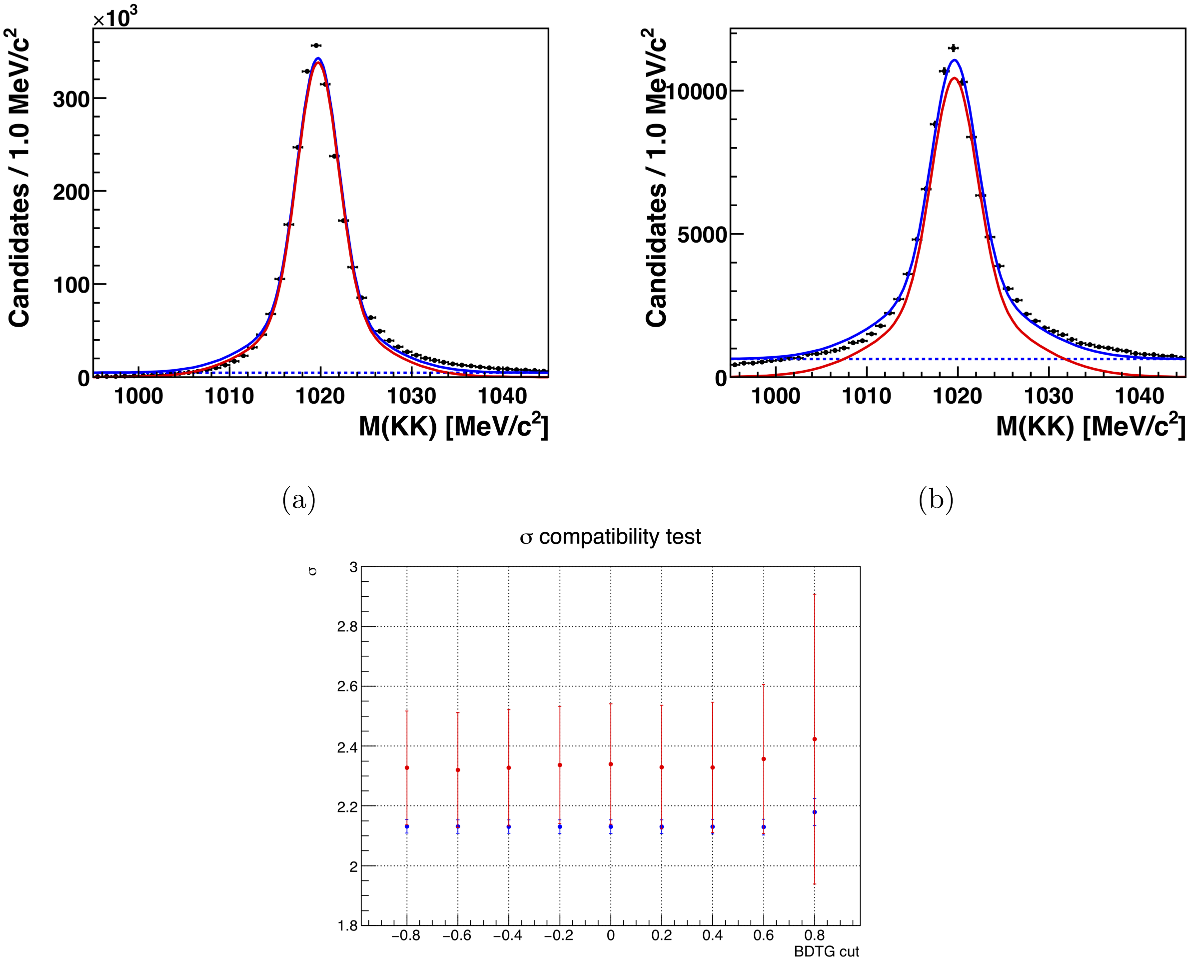
Although the fit of the distributions shows the compatibility of the data as shown in figure 6-14c, There are good hints to think that signal ghost data are in fact wrongly categorized by the reconstruction algorithms. However, more studies are needed to confirm the hypothesis that some of the events could be miss reconstructed.



(a) (b)

(c)

Figure 6-13.: The purity, the significance, and the purity times significance were calculated for several cuts in the BDTG to determine the most optimal value to apply.



(a) (b)

(c)

Figure 6-14.: The fit for the signal (a) and the signal ghost (b) candidates, and the comparison between the σ of the signal ghost (red) and real signal (blue) distributions depending on the BDTG cut (c).

6.5. Linking Particle to MC-Particle study

The algorithm in charge of linking the reconstructed particle with its corresponding MC particle is the Track Associator. As was discussed in the previous section, a charged track should share at least 70 % of hits with its corresponding generated track. The m_fractionOK is the variable within the algorithm that sets the default value. we applied a condition, reducing the 70 % condition of the hits shared between a reconstructed track and its MC track. In this case, the algorithms store more MCLinks[3] therefore the background information changes, although the total reconstructed particle content remains unaffected. The events which are thrown away are important, probably these are the ones that are in the peak of the distribution. The D^{*+} data sample was used for the following tests. This data sample contains a huge amount of background due to the fact that the $D^{*+} \to D^0(K^-\pi^+)\pi^+$ comes from the primary vertex so the soft π^+ is too complex to reconstruct. Therefore most of the analysis were carried out within the mass region $\Delta m = M_{D^{*+}-M_{D^0}}$. Finally, To select the signal and ghost candidates, the Background Category information was used.

6.5.1. Brief description of the algorithm

The algorithm decides which of the reconstructed particles is associated with an MC particle. It depends on the number of hits found in each subdetector, VELO, TT, Seed[4] and muon[5]. If the particle fulfills the conditions in all subdetectors, the link information is stored. Several conditions are depending on the subdetector.

$$\text{ratio} = \frac{\#\text{ associated hits}}{\text{total }\#\text{ hits}} \geq 0{,}70, \text{for VELO, Seeding and T stations.} \tag{6-1}$$

$$\#\text{ associated TT hits} > (\text{ total }\#\text{ TT hits - 2 }) \text{ or,} \tag{6-2}$$

$$(\text{total }\#\text{ VELO hits} > 2) \text{ and } (\text{total }\#\text{ IT+OT hits} > 2), \text{ for TT stations.}$$

where the eq. 6-1 is the decision to be fulfilled by the VELO, seed and muon stations and eq. 6-2 is the requirement for TT stations. The difference is due to the number of TT layers, as there are only four, the 70 % condition cannot be applied in the reconstruction. Moreover, some signal tracks have no hits in these stations. In these cases, the track should be correctly reconstructed by VELO and Seed stations.

6.5.2. m_fractionOK= 0,50 analysis

The first step is to prepare the data samples for posterior analysis. We need to reprocess the data sample with BRUNEL with the default configuration in order to make sure that the

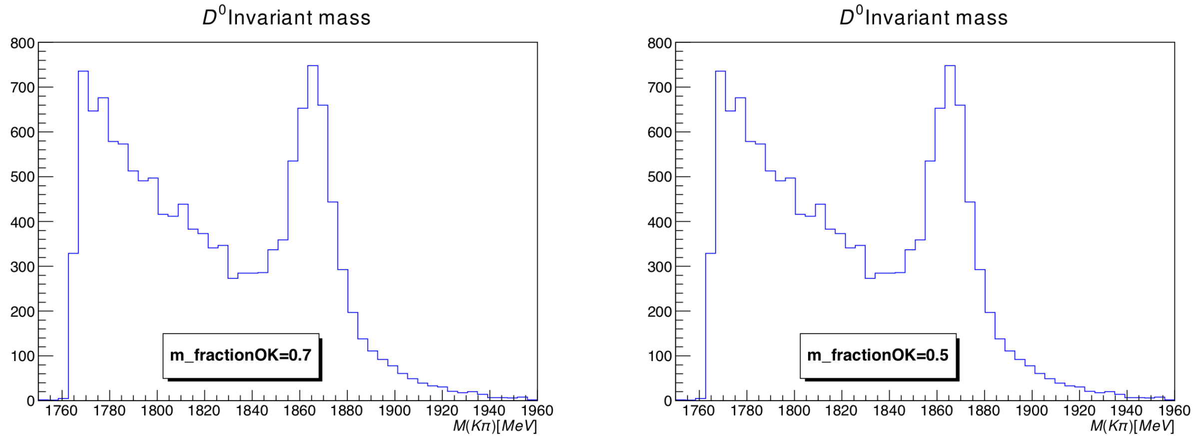
---

[3] The MCLink is the ratio between the reconstructed hits found and the generated ones in the whole detector.

[4] Seed information combines the IT and OT measurements

[5] Muon information is calculated if required

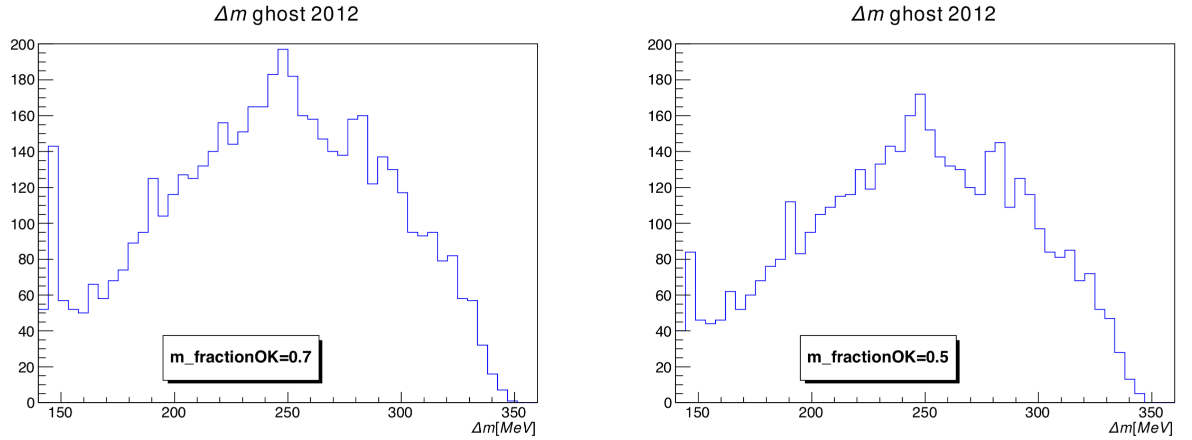
conditions are exactly the same and the unique change is in the algorithm requested due to the small improvements of the algorithms over the years.



(a) (b)

Figure 6-15.: D^0 mass distribution for 0.7 (a) and 0.5 (b) in the linking fraction.

The output of the reprocessed data sample was unaffected by the modification in the algorithm. The shape of the full MC sample does not show any change. However signal (BKGCAT== 0) and Ghost (BKGCAT== 60) shows a change. The change was applied to VELO, T and muon stations. The fraction for the TT stations was let as default.

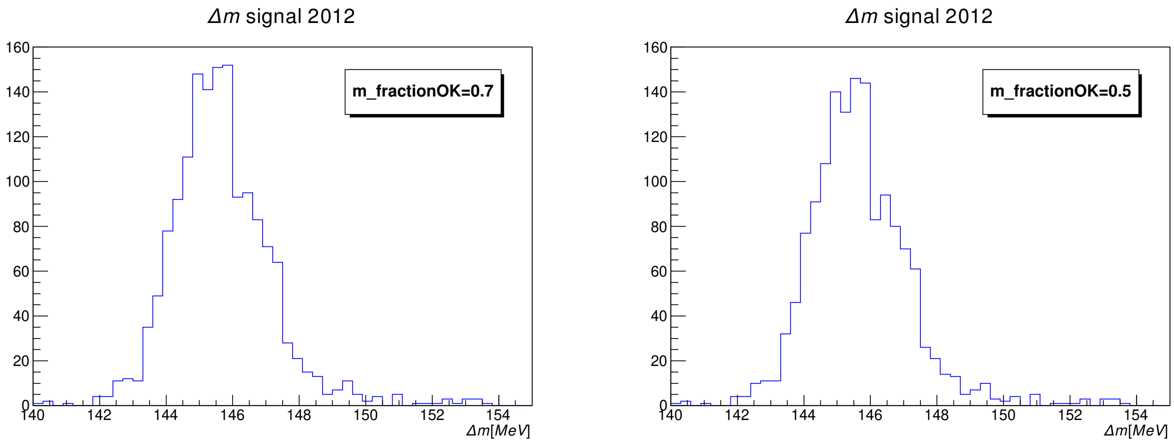


(a) (b)

Figure 6-16.: Δm ghost distribution for 0.7 (a) and 0.5 (b) in the linking fraction.

In Figure **6-16**, the evident reduction of the ghost candidates is due to the fact that the algorithm store more MCLinks that were rejected by the old one. In fact, the new tracks accepted could be generated tracks in which few hits were reconstructed. We observed $\sim 12\,\%$

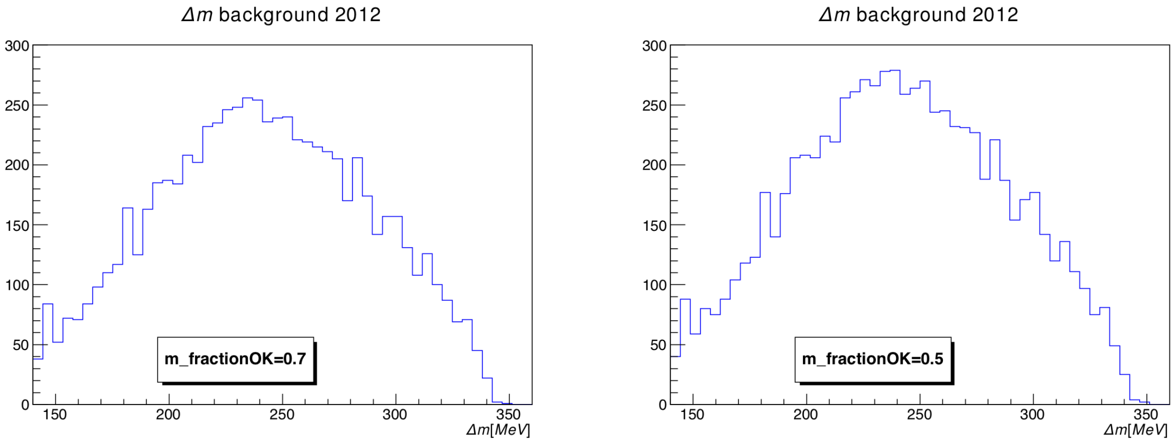
ghost candidates reduction in the whole mass region. Nevertheless $\sim 32\,\%$ ghost candidates reduction in the peak of the distribution ($\Delta m < 160$ MeV).



(a) (b)

Figure 6-17.: Δm signal distribution for 0.7 (a) and 0.5 (b) in the linking fraction.

The Figure **6-17** shows that in fact there is an increment in the signal candidates. $\sim 4\,\%$ signal gain in the Δm signal region. Although the majority of the events that migrate to signal fall in the $\Delta m < 160$ MeV region, It is important to determine the quality of the new tracks accepted by the algorithm.



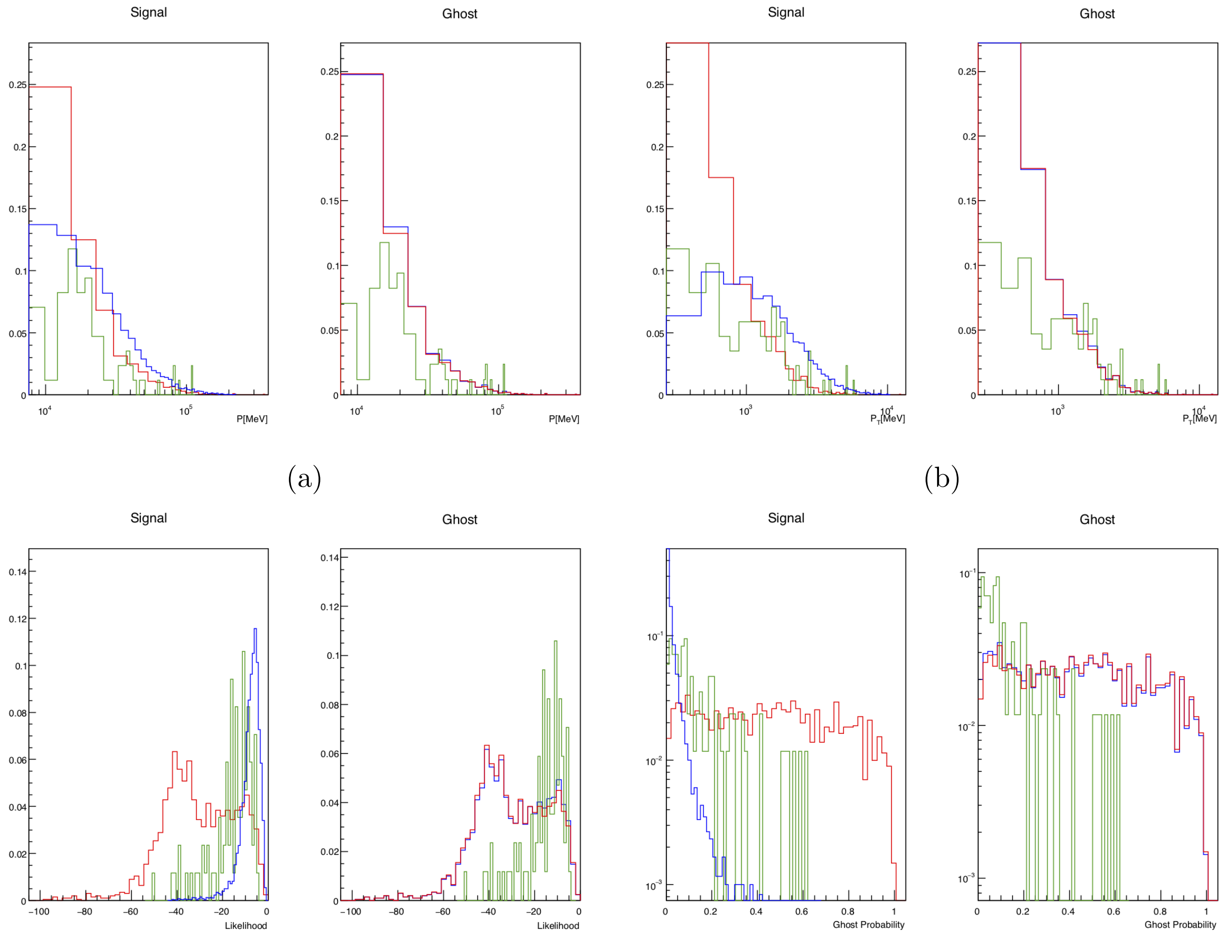
(a) (b)

Figure 6-18.: Δm background distribution for 0.7 (a) and 0.5 (b) in the linking fraction.

The majority of the new track candidates accepted were categorized as background. The background increases $\sim 8\,\%$ in the whole region. These reconstructed tracks have not enough quality to be categorized as signal. Background events were selected by using all background categories except for 0 and 60, (non Signal and non Ghost).

6.5.3. Migrated events

The events which migrate from ghost to signal candidates were isolated in order to compare their distributions with the pure signal and ghost samples in the mass region. The truth information was used to discriminate between ghost and signal candidates, specifically the true ID information was used which is equivalent to the true momentum information.



(a)

(b)

(c)

(d)

Figure 6-19.: Normalized distributions comparing signal and ghost candidates. Original sample (blue), events that migrate from ghost to background (red) and the events that migrate from ghost to signal (green).

All tracking variables show that the events which migrate, from ghost to signal, and the pure signal distributions have similar behavior. The majority of the new signal events fall in an acceptable region. Figure 6-19 shows the most relevant tracking variables.

	m_fractionOK=0.7	m_fractionOK=0.5	Migrated particles
Signal	$61{,}0 \pm 2\,\%$	$63{,}8 \pm 2\,\%$	$+2{,}85 \pm 2{,}22\,\%$
Ghost	$19{,}1 \pm 0{,}9\,\%$	$14{,}9 \pm 0{,}8\,\%$	$-4{,}25 \pm 2{,}01\,\%$
Background	$19{,}9 \pm 0{,}9\,\%$	$21{,}3 \pm 1\,\%$	$+1{,}40 \pm 2\,\%$

Table 6-4.: Signal, ghost and background ratios for 0,7 and 0,5 in the linking fraction for the $\Delta m < 170$ MeV region.

There is no reason for assigning the same linking fraction cut for the whole detector, Therefore a study varying those fractions individually for each subdetector was carried out to determine the best configuration.

6.5.4. Varying the m_fractionOK

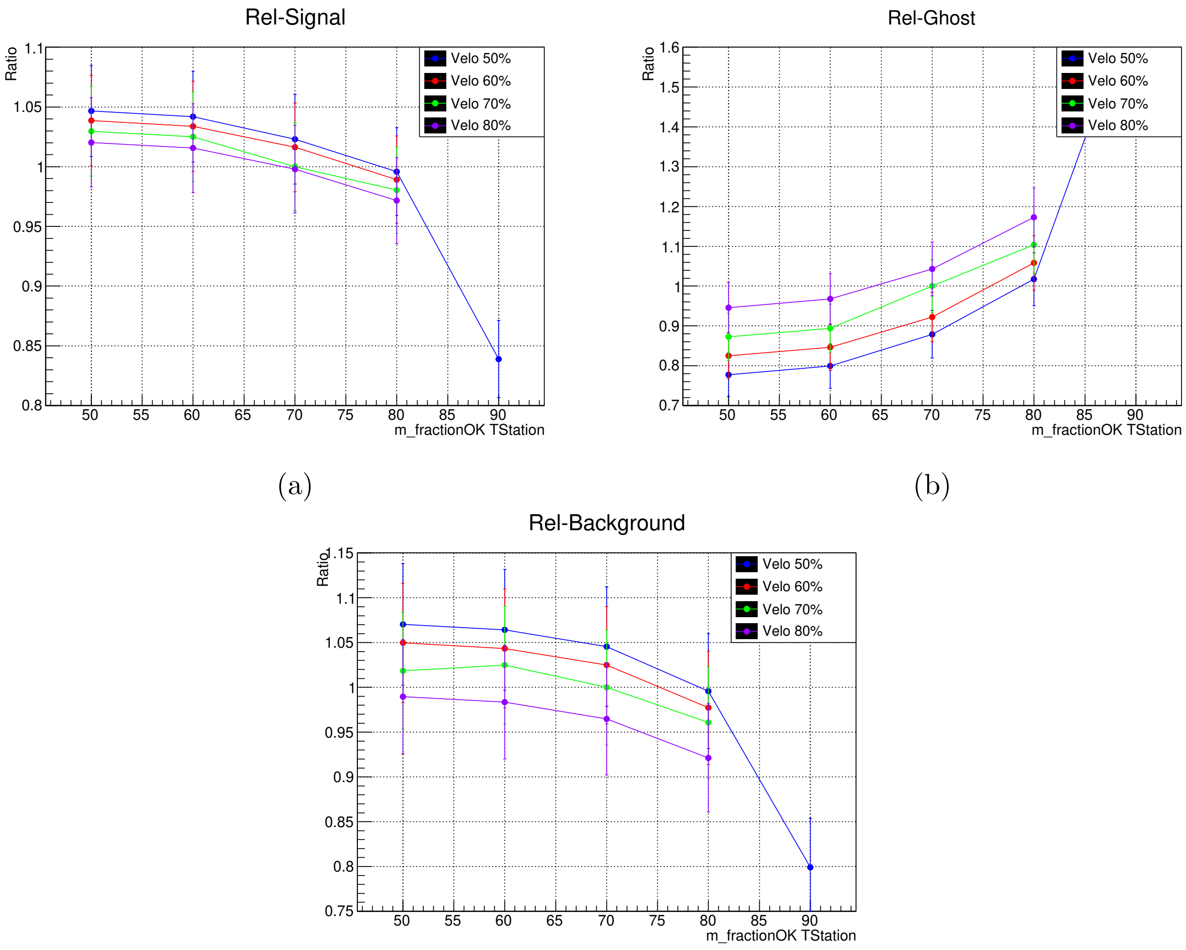
To determine the best configuration for VELO and T stations, a variation from 50 % to 80 % was made for the m_fractionOK for both subdetectors independently, m_fractionOKVelo and m_fractionOKSeed. Additionally, the 90 % configuration for the T stations was tested to determine, if it is worth. All calculations shown below were within the mass region $\Delta m < 170$ MeV. Furthermore, to select the particles, the background information for the D^{*+} was used.

The first quantities checked were the relative ratio, $ratio_R$, for signal, ghost, and background, for all possible combinations of configurations.

$$\text{Ratio}_R = \frac{\text{Events X-Y}}{\text{Events 70-70}}, \tag{6-3}$$

where Events X-Y is the number of events in a specific configuration, X means the fraction of the VELO detector and Y the one in the T stations and Events 70-70 represents the default configuration.

The values above 1 in Figure 6-20 mean that the events are migrating to that variable and below the events are throwing away. Moreover, the more restrictive the condition is, the more ghost candidates are identified. Therefore the 90 % condition was only tested for T stations, the ghost rate increases too much in that case. From Figures. 6-20a and 6-20b the best region for the migration from ghost to signal is between 50 % and 70 % however the background increases a bit too. Figure 6-20c shows an unexpected behavior for the background candidates at 80 % in the VELO configuration, all of them show a reduction of the candidates. This could be a clue to increase the VELO fraction to 80 % in order to have as good tracks as possible, taking into account that all pattern recognition algorithms start from VELO seeds.



(a)

(b)

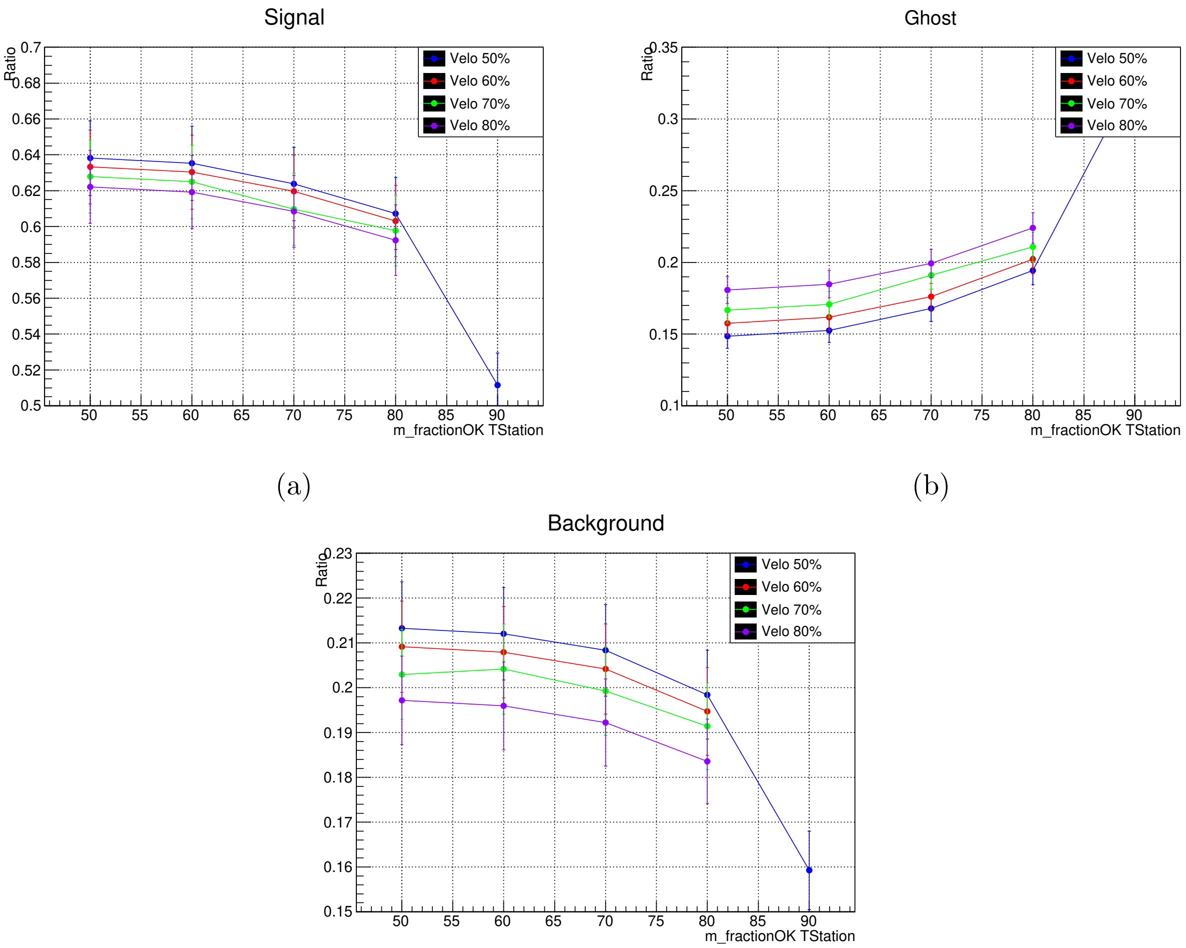
(c)

Figure 6-20.: Relative ratios for signal (a), ghost (b), and background (c).

Although the relative ratios were useful to extract some relevant information, the absolute ratios were calculated to compare and confirm the results from relative ratios.

$$\text{ratio}_{Abs} = \frac{\text{Events X-Y}}{\text{Total Sample}}. \tag{6-4}$$

The purity of the data sample is $\sim 61\,\%$ purity and the ghost rate $\sim 19\,\%$ in the mass region for the original configuration. Figure 6-21 shows the behavior of the signal, ghost, and background depending on the m_fractionOK cut. The 70 % cut is possibly the worst cut due to the fact that Figure 6-21a and 6-21b show a slightly decrement and increment respectively which are out of the tendency compared with the other configurations. Figure 6-21c show an increment of the background out of the tendency for the configuration 70-60. This not only confirms that the best cut is among the 50-70 but also discards the default configuration.



(a) (b)

(c)

Figure 6-21.: Ratios for signal (a), ghost (b), and background (c).

Another important variable to check is the ratio signal background due to the fact that the desirable cut is the one which maximizes signal and minimizes the background

$$\text{ratio}_{S/B} = \frac{\text{Signal X-Y}}{\text{Background X-Y}}. \tag{6-5}$$

The larger ratios signal vs background were found for VELO 80 %. However taking into account the information of **6-21**, it discards the configurations above the 70 % for the T stations. There is no clear configuration for VELO. Furthermore, the 70-70 configuration has a lower ratio than some other ones which confirms the hypothesis that this is not the best. Therefore the plot from the migrated events was checked to clarify the best configuration.

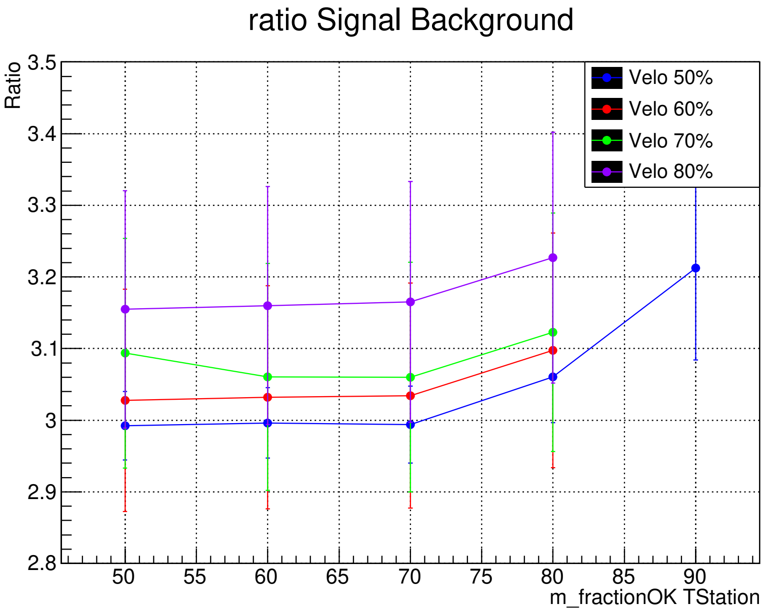


Figure 6-22.: Ratio Signal/Background

6.5.5. Best m_fractionOK cut

The comparison between the migrated events from or to signal and background were studied to determine the best cut.

$$\frac{\Delta\text{Signal}}{\Delta\text{Background}} = \frac{\text{Signal X-Y} - \text{Signal 70-70}}{\text{Background X-Y} - \text{Background 70-70}} \tag{6-6}$$

Figure 6-23 shows, there are two possible scenarios to generate the best configuration. The first one is the 80-50 configuration and the second one in the 70-50. It is clear that the best cut for the T stations is 50 % this could be due to the L-R ambiguity and for the VELO configuration, the discussion is between 70 % and 80 %.

	70-50		80-50	
	ratio_{Abs}	Δ	ratio_{Abs}	Δ
Signal	62,9 %	$+1{,}82 \pm 2{,}26\,\%$	62,2 %	$+1{,}24 \pm 2{,}25\,\%$
Ghost	16,7 %	$-2{,}44 \pm 1{,}21\,\%$	18,1 %	$-1{,}03 \pm 1{,}23\,\%$
Background	20,3 %	$+0{,}37 \pm 1{,}28\,\%$	19,7 %	$-0{,}21 \pm 1{,}27\,\%$

Table 6-5.: Comparison between the two better configurations for the total ratio and the migrated particles.

The configuration 80-50 shows a result quite fishy, there is a small migration from background to signal candidates and the ghost reduction is just about 1 % as shown in Table

6-5. This behavior is due to the conditions applied, the reduction on the ratio for the TT stations generates an increment of the background coming from ghost particles. However, the condition of the VELO in 80 % generates the decrement of the background because some tracks do not achieve the requirements to be background and these are considered as ghost. Therefore the configuration we propose for the best signal gain and the best ghost rejection is letting the VELO as default and set the T stations to 50 %. This ensures that the majority of the ghost migrated to signal candidates.

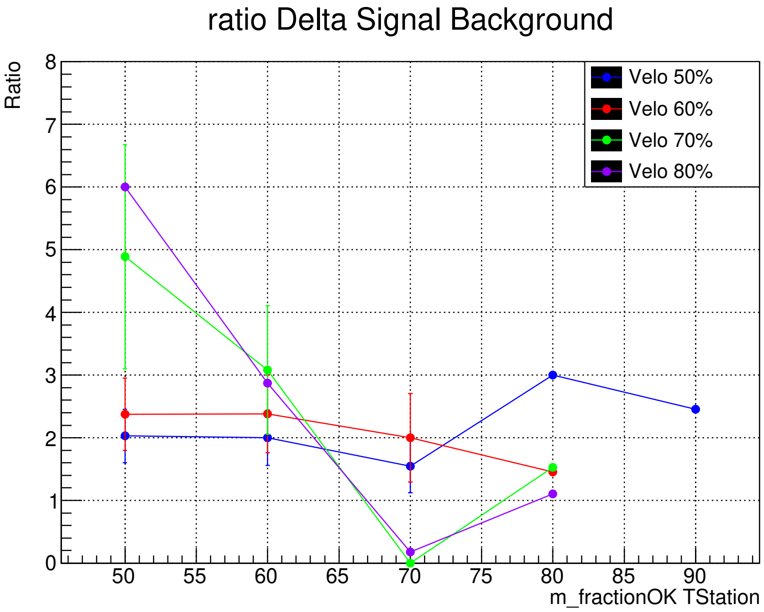


Figure 6-23.: ΔSignal/ΔBackground.

6.5.6. Setting TT station requirement to 50 %

Although the reasons why setting 50 % in the TT stations does not have too much sense, a test was made with this configuration, and 70 % for the rest of the tracking stations. Data showed this configuration obtained similar behavior that the 60 %-60 % configuration. The signal gain was 2,06 % with a ghost reduction of 2,93 % and a slight increment of the background of 0,87 %.

7. Conclusions

The tracking studies presented in this MSc. thesis on Monte Carlo simulations to improve the performance of the LHCb reconstruction algorithms by using Run I data samples (2011 and 2012) was to understand how the algorithms work and then to propose modifications in the reconstruction chain not only to be applied to Run I data but also to Run II and Run III. To successfully achieve the objective, we proposed five studies, first checking all the variables, second an adaptive binning, third a single and double ghost, fourth a BDT and, fifth a Linking Particle to MC-Particle.

In the first study, all kinematic variables were checked to look for differences between the signal and ghost distributions. Employing the background information to make the separation, large differences are presented in the helicity angle for kaons and the χ^2_{ndof} for kaons, muons, coming from the B^0_s decay, and pions coming from the B^0 decay. The difference in the helicity angle does not correspond to the peaking events in the mass distributions, therefore, we are sure these are in fact real ghost events. The differences presented in the χ^2_{ndof} could be a hint to redefine the cut which is less than three by default.

In the second study, An adaptive binning were proposed to determine if a tight χ^2_{ndof} cut, it would be plausible in the region compound by the η, PT and nTracks. The main objective was to assign different statistical weights to the regions to have a better separation between signal and ghost candidates. However, this study revealed that both signal and ghosts accumulated in the same areas, therefore, it was not possible to separate them without losing signal events.

In the third study, the ϕ ghost sample was separated into single and double ghost in order to study the performance of the ghost probability algorithm. Although this is a powerful tool to discard ghost tracks, it wrongly categorizes some tracks. Moreover, the peak in the mass of the ϕ candidate is built from one ghost kaon and one truth matched kaon. The double ghost distribution is flat as expected. The distributions of all track variables were checked to determine the differences between the ghost and the truth matched track. The larger differences were found in the transverse momentum, the number of hits in the TT stations and the likelihood. Cuts in these variables were propounded, $PT > 500$ MeV, zero hits in the TT stations, and Likelihood> -35. There are some problems with the cuts, some channels would be suppressed such as $D^0\pi$ by applying the PT cut, and the TT cut could remove some truth matched tracks. Nevertheless, the real signal and the ghost distributions were fitted after cutting in these variables to determine if they are compatible. Comparing $\sigma_s = 2{,}1123(0{,}003)$ and $\sigma_g = 2{,}486(0{,}051)$ show the distributions are not compatible however the possibility to have information lost at the reconstruction level is not discarded yet.

Furthermore, tracks that have zero TT hits should be treated differently.
In the fourth study, to confirm the previous results, a BDTG analysis was carried out which the pretend to separate the real ghost from the peaking candidates. by fitting the signal ghost and true signal distributions, some BDTG cuts show compatibility. It confirms that there are some events wrong categorized. However, we cannot recover them as signal events since those are worse reconstructed.
In the fifth study, the track associator algorithm was studied in detail. The 70 % linking fraction between the hits of the generated and reconstructed track was changed to 50 % in order to determine the differences in the reconstruction of the D^{*+} data sample. This change shows that in fact there are some signal events misidentified as ghost ones. This configuration reduces the ghost events by 4,25 % while increasing the signal candidates by 2,84 %. All the same, the background is also increased by 1,40 % out of the total sample. The majority of the migrated events from ghost to signal are in the Δm region and they behave as signal candidates as distributions show. This confirms they are in fact signal candidates. This happens because the algorithm stores more MCLinks, it means the algorithm does not flag tracks as ghost tracks which not only re-categorizes the particles but also recovers the truth information. 50 % is not probably the best cut, therefore, a variation from 50 % to 70 % was studied which shows the best configuration is 70 % for VELO and 50 % for the T stations. This configuration loosens the restriction only in the T stations due to the L-R ambiguity. The presence of this ambiguity and the way that the algorithm was built leads in the introduction of more hits in the T stations which do not correspond to the generated ones. The selected configuration increases signal and background events by $1{,}82 \pm 2{,}25\,\%$ and $0{,}37 \pm 1{,}2\,\%$ respectively and reduces the ghost events by $2{,}4 \pm 1{,}2\,\%$ out of the total sample. It is possible to recover some signal from the ghost candidates, in this channel the error in the signal is too big. However, using cleaner and bigger data samples the error should decrease.
Finally, Although the bug in the algorithm was that the linking fraction should not be the same for all the subdetectors, the track association is only the first step in ghost recognition. Therefore more studies are needed to confirm the configuration. Studies including pions and muons in the final states within B decay channels, checking the RunII data samples and the upgrade also.

A. Kalman formalism

The concept of the Kalman formalism is to consider a track as a collection of measurements and track states which establishes the right scenario for the Kalman filter as a good method to fit tracks. This filter adds the measurements one-by-one to the fit, while the local track state at the current node is updated. This method is mathematically equivalent to the least-squares fit. However, the procedure is faster than least-squares due to the fact that it avoids the slow computational processes for large matrices which is crucial at CERN experiments. the method minimizes the χ^2 of the measurements on the track.

The usage of the Kalman filter has several advantages. Among them, its use in pattern recognition where the algorithm decides if a measurement is added based on the measurement's contribution to the χ^2 without refitting the whole track. Furthermore, the track state and full trajectory are known. This information is especially important for matching tracks with RICH rings, calorimeter clusters, and muon candidates, and the noise produced by multiple scattering is naturally added in this formalism. The Kalman technique can be applied to any linear dynamic system where random disturbances are present in the evolution of the system as well as in the measurements [49].

The Kalman procedure can be divided into three different steps, prediction, filter and smoother.

A.1. Prediction

The prediction of the track state at a given node (k^{th}) is made based on the track state at the previous node ($k-1$) using the propagation relation:

$$\vec{x}_k^{k-1} = f_k(\vec{x}_{k-1}), \tag{A-1}$$

$$C_k^{k-1} = F_k C_{k-1} F_k^T + Q_k. \tag{A-2}$$

The $\vec{x}_k^{k-1}$ term refers to the predicted state vector, $\vec{x}_k^k \equiv \vec{x}_k$ refers to a filtered state kk vector, and $\vec{x}_k^n$ refers to a smoothed state vector. It means the superscript denotes the amount of information used in the estimation. The Q_k is the process noise term which is added to the predicted covariance matrix C_{k-1}.

In order to make the first prediction, the initial track state, $\vec{x}_0$, needs to be estimated as required for the track fit. this prediction usually comes from an initial fit in the track finding algorithms in case of the absence of subdetector estimation. The initial covariance matrix,

C_0, must have increased values for the errors on the diagonal. This not only suppresses any bias from the prediction but also avoids double counting of measurements. Finally the distance between the measurement, mk, and the state vector in the measurement plane is calculated. This difference is called the residual

$$r_k^{k-1} = m_k - h_k(\vec{x}_k^{k-1}), \tag{A-3}$$

$$R_{k-1} = V_k + H_k C_{k-1} H_k^T. \tag{A-4}$$

The expected variance, R_{k-1}, has a contribution from the covariance matrix of the k track state, C_{k-1}, and the measurement variance, V_k. The predicted contribution of this k measurement to the total χ^2 can be calculated as

$$(\chi^2_+)_k^{k-1} = r_{k-1}(R_k^{k-1})^{-1} r_{k-1}. \tag{A-5}$$

The main goal of the Kalman fit is to find the optimal track states that minimize the χ^2_+

A.2. Filter

In this step, the prediction is updated with information of the measurement in this node k using the filtered state vector and its covariance matrix. The prediction and filter steps are repeated until all measurements are added. The track state after each filter step is the best estimate of the trajectory based on the measurements incorporated so far [49].

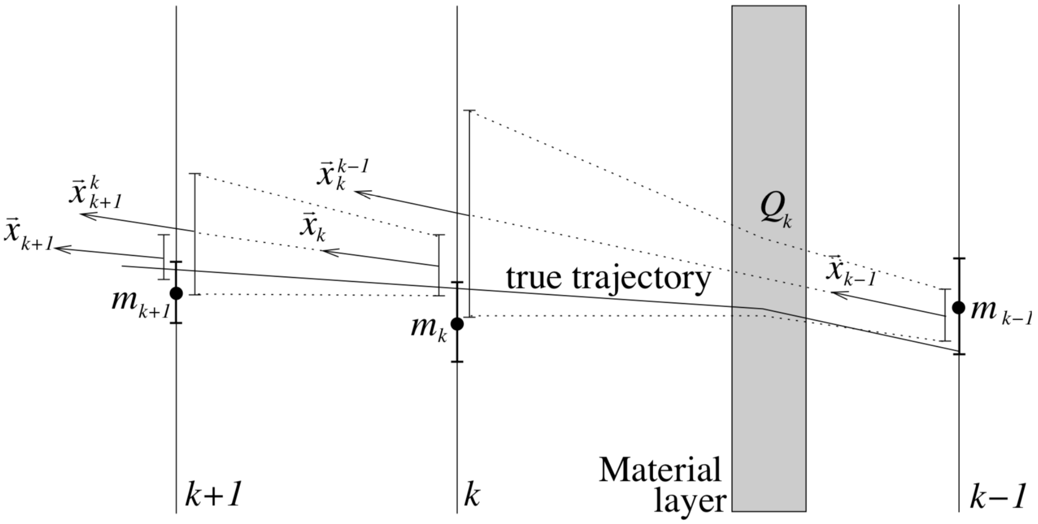


Figure A-1.: Schematic picture of the Kalman filter, showing evolution of a track state from node $k-1$ to k to $k+1$ [49].

In the gain matrix formalism, the equations are

$$\vec{x}_k = \vec{x}_k^{k-1} + K_k r_k^{k-1}, \tag{A-6}$$
$$C_k = (\mathbb{1} - K_k H_k) C_k^{k-1} \tag{A-7}$$

where K_k is a 5×1 gain matrix calculated as

$$K_k = C_k^{k-1} H_k^T (V_k + H_k C_k^{k-1} H_k^T)^{-1} = C_k^{k-1} H_k^T (R_k^{k-1})^{-1}. \tag{A-8}$$

The filtered values for the residual and its covariance matrix are calculated as follows

$$r_k = m_k - h_k(\vec{x}_k) = (1 - H_k K_k) r_k^{k-1}, \tag{A-9}$$
$$R_k = (1 - H_k K_k) V_k = V_k - H_k C_k H_k^T. \tag{A-10}$$

The contribution to the total filtered χ^2 of this measurement is

$$(\chi^2_+)_k = r_k R_k^{-1} r_k. \tag{A-11}$$

The scattering of a particle traversing a material layer between node k and $k-1$ is shown in Fig **A-1**. The track fit increases the predicted error on the state vector, $\vec{x}_k^{k-1}$, with $Q-k$. The measurement in m_k pulls the state to the true trajectory and the filtered track state, $\vec{x}_k$, is calculated as a result. the best estimation of the track state is obtained with the filtered track state at the least node ($k = n$) which includes the information from all nodes. The other nodes ($k < n$) are updated in the smoother step.

A.3. Smoother

After all, measurements are added to the track, the track state at the previous nodes is updated using the recursive method, known as the Rauch-Tung-Striebel smoother, resulting in a more smooth track. the smoothed state vector and its covariance matrix are given by

$$\vec{x}_k^n = \vec{x}_k + A_k(\vec{x}_{k+1}^n - \vec{x}_{k+1}^k), \tag{A-12}$$
$$C_k^n = C_k + A_k(C_{k+1}^n - C_{k+1}^k) A_k^T \tag{A-13}$$

where the 5×5 smoother gain matrix equals

$$A_k = C_k F_{k+1}^T (C_{k+1}^k)^{-1} \tag{A-14}$$

The smoothed residual and the error on the residual are

$$r_k^n = m_k - h_k(\vec{x}_{k+1}^n) \tag{A-15}$$
$$R_k^n = V_k - H_k C_{k+1}^n H_k^T \tag{A-16}$$

Finally, the total χ^2 of the track fit is obtained as a sum of the smoothed χ^2_+ increments

$$(\chi^2_+)_n^k = r_k^n (R_k^n)^{-1} r_k^n \tag{A-17}$$

A.4. Resolution and pull

Two important parameters for the performance of the fit are the resolution and pull distributions. The fitted track states need to be compared with the Monte Carlo true information in order to determine the performance of the fit. The error in the parameter x is calculated as

$$\delta x = x_{rec} - x_{true}, \tag{A-18}$$

where x_{rec} is the reconstructed or fitted, value, x_{true} is the Monte Carlo value and the resolution which is the Gaussian width of the δx distribution, and the pull is defined as

$$\Delta x = \frac{x_{rec} - x_{true}}{\sigma_x} \tag{A-19}$$

where σ_x is the estimated error, obtained from the covariance matrix of the fit. The pull distribution should have zero mean and a unit width.

B. Muon tracks reconstruction

A large fraction of the B decay channels studied by the LHCb experiment have muons in their final states such as $B_s^0 \to J/\psi(\mu^+\mu^-)\phi$, $B^0 \to J/\psi(\mu^+\mu^-)K_s^0$, $B^0 \to \mu^+\mu^-K^*$ among others. Therefore the identification of muons by the experiment needs to be studied and improved. The identification of muons at LHCb is lead by the five muon detection stations interleaved by four filtering iron walls. Only muons can penetrate the calorimeters and muon filters, and leave signals on the muon chambers. Several variables were built to exploit the information collected only from the muon stations. Moreover, discrimination with respect to other particles is carried out by the combined information coming from the calorimeters and RICH detectors.

B.1. Variables

The first identification variable is a boolean decision, called IsMuon, obtained from the extrapolation of a long or a downstream track through the muon stations [80]. The extrapolation, within the muon stations acceptance, yields the expected track coordinates in the stations M2-M5. The search for hits is made within an elliptic, momentum dependent, field of interest (FoI) around the extrapolated track for each station.

p [GeV/c]	Required stations	
	`IsMuon`	`IsMuonLoose`
$p < 3$	*Always false*	*Always false*
$p < 6$	M2 & M3	At least two of M2–M4
$6 < p < 10$	M2 & M3 & (M4 $\|$ M5)	At least three of M2–M5
$p > 10$	M2 & M3 & M4 & M5	At least three of M2–M5

Figure B-1.: Required stations with hits within FoI for IsMuon and IsMuonLoose with respect to track momentum [80].

The attributes IsMuon and IsMuonLoose are flagged to each track if it contains at least one hit in a minimum number of stations always including M3. The minimum momentum required for a muon to reach the M3 station without being absorbed for the material is 3 GeV. The readout of the muon detector is given by the horizontal and vertical, physical pads. A logical pad is defined as the crossing of two which the dimensions are the (x, y) pad size associated with the hit. If there is no crossing the whole dimensions of the hit physical pad are taken. this ensures low occupancies while reducing the number of readout channels. The physical and logical pads are also referred to as uncrossed hits and crossed hits respectively.

Another variable to consider is the average squared distance in units of the pad size. This is the difference between the track extrapolation into the muon stations and the closest hit for each station which separates muons and non-muons with high precision.

$$D^2 = \frac{1}{N} \sum_{i=0}^{N} \left[\left(\frac{x_{\mathrm{closest,i}} - x_{\mathrm{track,i}}}{\mathrm{Pad}_{\mathrm{x,i}}} \right)^2 + \left(\frac{y_{\mathrm{closest,i}} - y_{\mathrm{track,i}}}{\mathrm{Pad}_{\mathrm{y,i}}} \right)^2 \right] \tag{B-1}$$

D^2 distance depends on the momentum of the muon, due to multiple scattering, the traversed material, and the logical pad size. furthermore, the delta log-likelihood, DLL, depends on two probabilities. The $P(\mu)$ and $P(\mathrm{not}\mu)$ which correspond to the probability of being a muon or not respectively. The DLL is calculated as

$$\mathrm{DLL} = \log \left(\frac{P(\mu)}{P(\mathrm{not}\mu)} \right) \tag{B-2}$$

Calibration constants are needed due to the two dimensional binning. Finally, the nShared variable was created to distinguish between real tracks and potential ghost track since tracks sharing several numbers of hits are rarely two muons very close to each other.

B.2. Matching tracks to muon hits

The tag-and-probe method uses two-prong decays, where one of the decay products, the "tag", is fully reconstructed as a long track, while the other particle, the "probe", is only partially reconstructed [81]. The probe should carry enough information to reconstruct the invariant mass of the parent with a good resolution which helps to discriminate background. If a match is found between the probe and the tag the probe track is defined as efficient. There are three methods, the VELO, the T Station and the long method as shown in Figure **B-2**. These are briefly explained below.

B.2.1. VELO method

The muon track reconstruction in the VELO is performed using downstream tracks as probes. The downstream track and the long track of the same muon are matched if they share at least 50 % of their hits in the T stations. This fraction is appropriate based on simulations.

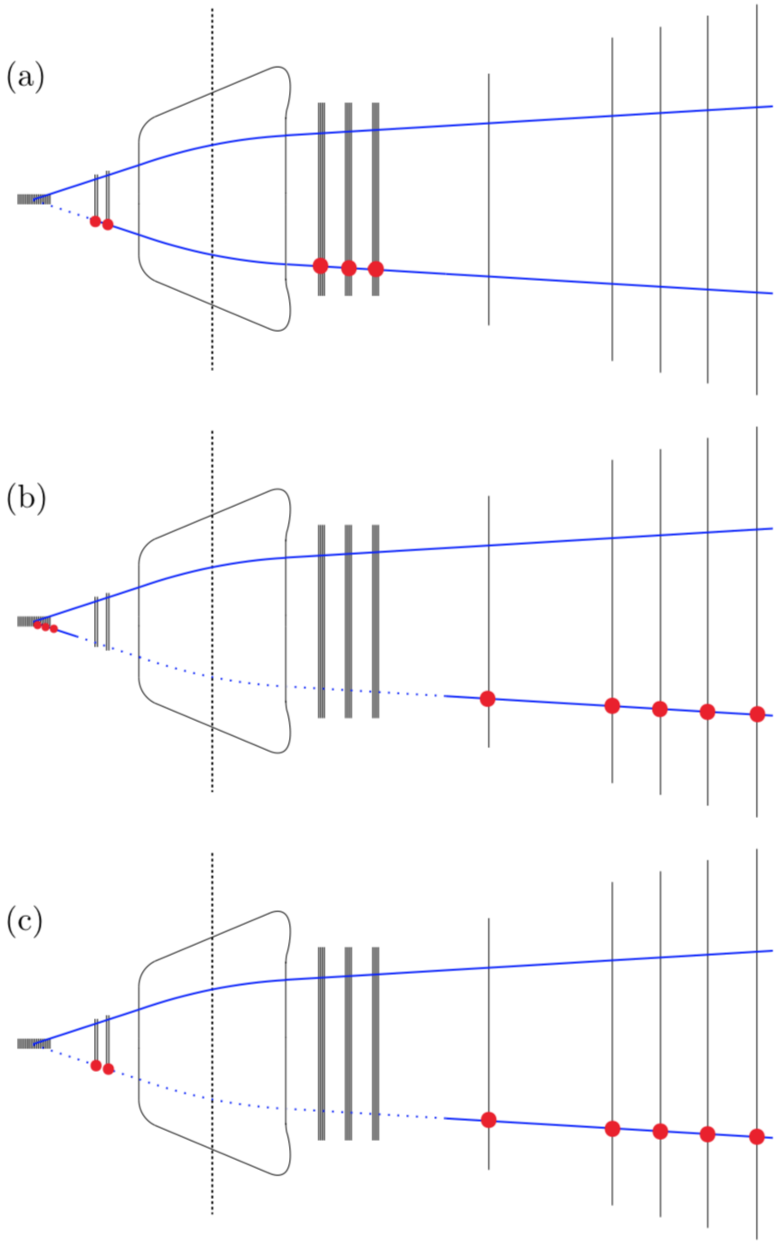


Figure B-2.: Illustration of the three tag-and-probe methods: (a) the VELO method, (b) the T-station method, and (c) the long method. The VELO (black rectangle), the two TT layers (short bold lines), the magnet coil, the three T stations (long bold lines), and the five muon stations (thin lines) are shown in all three sub-figures. The upper solid blue line indicates the tagged track, the lower line indicates the probe with red dots where hits are required and dashes where a detector is probed [81].

B.2.2. T Station method

The algorithm reconstructs muons as straight lines starting from the hits in the last muons station which are matched with VELO tracks. Two requirements need to be fulfilled in order to match the probe and the long track. The probe and the long track have to be reconstructed from the same VELO seed. And at least two hits on the probe track in the muon stations have to be compatible with the extrapolation of the track into the muon stations. This is a strong enough condition to ensure compatible trajectories of the long track and the VELO-muon probe track.

B.2.3. Long method

In the long method, the probe track has hits in the TT and muon stations. his method measures the efficiency to reconstruct long tracks because the long-track-finding algorithms do not require the presence of TT hits [81]. A TT-muon track is reconstructible as a long track if at least 70 % of the hits in the muon stations are compatible with the extrapolation of the long track. This percentage is loosened to 60 % in case the long track has TT hits. Simulations helped to optimize these fractions.

The average track reconstruction efficiency for long tracks is measured to be $(95{,}4 \pm 0{,}7)\,\%$ for 2010 data, $(97{,}78 \pm 0{,}07)\,\%$ for 2011 data and $(96{,}99 \pm 0{,}05)\,\%$ for 2012 data [81]. An upgrade of the muon reconstruction algorithms for the RunII was made in order to have a better performance. This improvement is explained in Ref. [80].

Bibliography

[1] LHCb Collaboration. LHCb - Large Hadron Collider beauty experiment. `http://lhcb-public.web.cern.ch/lhcb-public/`, 2018.

[2] LHCb Collaboration. The lhcb detector at the lhc. *Journal of Instrumentation*, 3:1–218, 2008.

[3] M. Meissner. *Measurements of particle production and particle correlations in proton-proton and proton-ion collisions with LHCb.* Phd thesis, Ruperto-Carola-University, Heidelberg, Germany, 2015. `http://inspirehep.net/record/1381400?ln=en`.

[4] Wimberley. J. Ibackgroundcategory. `https://twiki.cern.ch/twiki/bin/view/LHCb/TupleToolMCBackgroundInfo`, 2013.

[5] Gligorov. V. The backgroundcategory tool for background classification. PPTS Meeting, CERN, 2013.

[6] Olive K.A. et al. *Review of particle physics.* Chin. Phys C, Beijing, 2014.

[7] S. Bethke. World summary of α_s (2015). In *High-precision α_s measurements from LHC to FCC-ee*, pages 6–10, 2015.

[8] Marek Karliner and Jonathan L. Rosner. Discovery of the doubly charmed Ξ_{cc} baryon implies a stable $bb\overline{u}\overline{d}$ tetraquark. *Phys. Rev. Lett.*, 119:202001, Nov 2017.

[9] R. et. al. Aaij. Observation of $j/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to j/\psi K^- p$ decays. *Phys. Rev. Lett.*, 115:072001, Aug 2015.

[10] A. Nisati and G. Tonelli. The discovery of the Higgs boson at the Large Hadron Collider. *Riv. Nuovo Cim.*, 38(11):507–573, 2015.

[11] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716:1–29, 2012.

[12] CMS Collaboration. Observation of a new boson at a mass of $125 gev$ with the cms experiment at the lhc. *Physics Letters B*, 716:30–61, 2012. CMS.

[13] Sébastien Descotes-Genon and Patrick Koppenburg. The ckm parameters. *Annual Review of Nuclear and Particle Science*, 67(1):97–127, 2017.

[14] N.Tuning P.Kooijman. Lectures on cp violation. `https://www.nikhef.nl/~h71/Lectures/2015/ppII-cpviolation-29012015.pdf`, 2015.

[15] R. Ciesielski and K. Goulianos. Mbr monte carlo simulation in pythia8. arXiv:1205.1446, 2012.

[16] The LHCb collaboration. Measurement of the inelastic pp cross-section at a centre-of-mass energy of 13 tev. *Journal of High Energy Physics*, 2018(6):100, Jun 2018.

[17] ATLAS Collaboration. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the atlas detector at the lhc. *Phys. Rev. Lett.*, 117:182002, Oct 2016.

[18] Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-FSQ-15-005, CERN, Geneva, 2016.

[19] L. Evans and P. Bryant. Lhc machine. *Journal of Instrumentation*, 3:1–165, 2008.

[20] ATLAS Collaboration. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3:1–438, 2008.

[21] CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3:1–362, 2008.

[22] ALICE Collaboration. The alice experiment at the cern lhc. *Journal of Instrumentation*, 3:1–260, 2008.

[23] O Callot. FastVelo, a fast and efficient pattern recognition package for the Velo. Technical Report LHCb-PUB-2011-001. CERN-LHCb-PUB-2011-001, CERN, Geneva, Jan 2011. LHCb.

[24] F. Rohner W. Flegel (Project Coordinator), M. Losasso. Lhcb magnet technical design report. Technical report, Magnet Design Team, CERN, 2000.

[25] LHCb Collaboration. Outer tracker technical design report. Technical report, LHCb, CERN, 2001.

[26] LHCb Collaboration R. Aaij et al. First observation of CP violation in the decays of B^0_s mesons. *Phys.Rev.Lett.*, 110:221601, 2013.

[27] LHCb Collaboration R. Aaij et al. Measurement of the CKM angle γ from a combination of $B^{\pm} \to Dh^{\pm}$ analysis. *Phys.Lett.B*, 726:151, 2013.

[28] LHCb Collaboration R. Aaij et al. Measurement of the ratio of the branching fractions $B(B^0 \to K^{+0}\gamma)/B(B^0_s \to \phi\gamma)$. *Phys.Rev.Lett.*, 111:101805, 2013.

[29] LHCb Collaboration R. Aaij et al. Measurements of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for the $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment. *Phys.Rev.Lett.*, 111:101805, 2013.

[30] LHCb Collaboration. Lhcb particle identification upgrade technical design report. Technical report, LHCb, CERN, 2013.

[31] LHCb Collaboration O. Deschamps. LHCb Calorimeters - Commissioning & performance. LHCb-TALK-2010-053, 2010.

[32] S Filippov, Yu K Gavrilov, E Guschin, S V Laptev, and V E Postoev. Experimental performance of SPD/PS detector prototypes. Technical Report LHCb-2000-031, CERN, Geneva, Oct 2000.

[33] G Böhner, A Falvard, J Lecoq, P Perret, and C Trouilleau. Very front-end electronics for the LHCb preshower. Technical Report LHCb-2000-047, CERN, Geneva, Oct 2000.

[34] LHCb collaboration Yu Guz. The lhcb hadron calorimeter. *Journal of Physics: Conference Series*, 160(1):012054, 2009.

[35] L G Afanasyeva, Y Bogomolov, S Denissov, R I Dzhelyadin, A Kobelev, A K Konoplyannikov, O Kurakina, V Matveev, G D Nekipelova, V F Obraztsov, and E Tchernov. The hadron calorimeter design and construction. Technical Report LHCb-2000-045, CERN, Geneva, Oct 2000.

[36] LHCb Collaboration. LHCb muon system technical design report. Technical report, CERN, Geneva, 2001.

[37] F Archilli, W Baldini, G Bencivenni, N Bondar, W Bonivento, S Cadeddu, P Campana, A Cardini, P Ciambrone, X Cid Vidal, C Deplano, P De Simone, A Falabella, M Frosini, S Furcas, E Furfaro, M Gandelman, J A Hernando Morata, G Graziani, A Lai, G Lanfranchi, J H Lopes, O Maev, G Manca, G Martellotti, A Massafferri, D Milanes, R Oldeman, M Palutan, G Passaleva, D Pinci, E Polycarpo, R Santacesaria, E Santovetti, A Sarti, A Satta, B Schmidt, B Sciascia, F Soomro, A Sciubba, and S Vecchi. Performance of the muon identification at lhcb. *Journal of Instrumentation*, 8(10):P10020, 2013.

[38] B. De Paula, F. Marinho, and S. Amato. Analysis of the $B_s rightarrowmu^+mu^-$ Decay with the Reoptimized LHCb Detector. 2004.

[39] LHCb collaboration. LHCb Trigger and Online Upgrade Technical Design Report. Technical Report CERN-LHCC-2014-016. LHCB-TDR-016, CERN, May 2014.

[40] LHCb collaboration Antunes-Nobrega et al. *LHCb trigger system: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2003. revised version number 1 submitted on 2003-09-24 12:12:22.

[41] Barbara Storaci. Optimization of the lhcb track reconstruction. *Journal of Physics: Conference Series*, 664(7):072047, 2015.

[42] Torbjörn Sjöstrand, Patrik Edén, Christer Friberg, Leif Lönnblad, Gabriela Miu, Stephen Mrenna, and Emanuel Norrbin. High-energy-physics event generation with pythia 6.1. *Computer Physics Communications*, 135(2):238 – 259, 2001.

[43] René Brun, F Bruyant, Federico Carminati, Simone Giani, M Maire, A McPherson, G Patrick, and L Urban. *GEANT: Detector Description and Simulation Tool; Oct 1994*. CERN Program Library. CERN, Geneva, 1993. Long Writeup W5013.

[44] Antunes-Nobrega et al. LHCb reoptimized detector design and performance: Technical Design Report. Technical report, CERN, Geneva, 2003.

[45] Michael H. Seymour and Marilyn Marx. *Monte Carlo Event Generators*. Springer International Publishing, Cham, 2015.

[46] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[47] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A brief introduction to pythia 8.1. *Computer Physics Communications*, 178(11):852 – 867, 2008.

[48] Torbjörn Sjöstrand. Pythia 8 status report. arXiv:0809.0303, 2008.

[49] J Van Tilburg and M Merk. *Track simulation and reconstruction in LHCb*. PhD thesis, 2005. Presented on 01 Sep 2005.

[50] Torbjörn Sjöstrand, Patrik Edén, Christer Friberg, Leif Lönnblad, Gabriela Miu, Stephen Mrenna, and Emanuel Norrbin. High-energy-physics event generation with pythia 6.1. *Computer Physics Communications*, 135(2):238 – 259, 2001.

[51] Torbjörn Sjöstrand and Maria van Zijl. A multiple-interaction model for the event structure in hadron collisions. *Phys. Rev. D*, 36:2019–2041, Oct 1987.

[52] Marco Clemencic, Hubert Degaudenzi, Pere Mato, Sebastien Binet, Wim Lavrijsen, Charles Leggett, and Ivan Belyaev. Recent developments in the lhcb software framework gaudi. *Journal of Physics: Conference Series*, 219(4):042006, 2010.

[53] P Mato. GAUDI-Architecture design document. Technical Report LHCb-98-064, CERN, Geneva, Nov 1998.

[54] LHCb Collaboration. The gaudi project. http://gaudi.web.cern.ch/gaudi/, 2018.

[55] LHCb Collaboration. The gauss project. http://lhcbdoc.web.cern.ch/lhcbdoc/gauss/, 2018.

[56] Torbjörn Sjöstrand. Status and developments of event generators. arXiv:1608.06425, 2016.

[57] LHCb Collaboration. Gauss user guide and reference manual. http://lhcb-comp.web.cern.ch/lhcb-comp/Simulation/Gauss.pdf, 2005.

[58] LHCb Collaboration. The decfiles package. http://lhcbdoc.web.cern.ch/lhcbdoc/decfiles/, 2018.

[59] LHCb Collaboration. The boole project. http://lhcbdoc.web.cern.ch/lhcbdoc/boole/, 2018.

[60] LHCb Collaboration. The moore project. http://lhcbdoc.web.cern.ch/lhcbdoc/moore/, 2018.

[61] Albert Puig. The LHCb trigger in 2011 and 2012. Technical Report LHCb-PUB-2014-046. CERN-LHCb-PUB-2014-046, CERN, Geneva, Nov 2014.

[62] LHCb Collaboration. The brunel project. http://lhcbdoc.web.cern.ch/lhcbdoc/brunel/, 2018.

[63] LHCb Collaboration. The davinci project. http://lhcbdoc.web.cern.ch/lhcbdoc/davinci/, 2018.

[64] LHCb Collaboration. Root data analysis framework. https://root.cern.ch, 2018.

[65] Antunes-Nobrega et. al. *LHCb computing: Technical Design Report*. Technical Design Report LHCb. CERN, Geneva, 2005. Submitted on 11 May 2005.

[66] O Callot and S Hansmann-Menzemer. The Forward Tracking: Algorithm and Performance Studies. Technical Report LHCb-2007-015. CERN-LHCb-2007-015, CERN, Geneva, May 2007.

[67] R W Forty and M Needham. Standalone Track Reconstruction in the T-stations. Technical Report LHCb-2007-022. CERN-LHCb-2007-022, CERN, Geneva, Mar 2007.

[68] O Callot, M Kucharczyk, and M Witek. VELO-TT track reconstruction. Technical Report LHCb-2007-010. CERN-LHCb-2007-010, CERN, Geneva, Apr 2007.

[69] O Callot. Downstream Pattern Recognition. Technical Report LHCb-2007-026. CERN-LHCb-2007-026, CERN, Geneva, Mar 2007.

[70] P. d'Argent, L. Dufour, L. Grillo, J.A. de Vries, A. Ukleja, R. Aaij, F. Archilli, S. Bachmann, D. Berninghoff, A. Birnkraut, J. Blouw, M. De Cian, G. Ciezarek, C. Färber, M. Demmer, F. Dettori, E. Gersabeck, J. Grabowski, W.D. Hulsbergen, B. Khanji, M. Kolpin, M. Kucharczyk, B.P. Malecki, M. Merk, M. Mulder, J. Müller, V. Mueller, A. Pellegrino, M. Pikies, B. Rachwal, T. Schmelzer, B. Spaan, M. Szczekowski, J. van Tilburg, S. Tolk, N. Tuning, U. Uwer, J. Wishahi, and M. Witek. Improved performance of the lhcb outer tracker in lhc run 2. *Journal of Instrumentation*, 12(11):P11016, 2017.

[71] M Needham. Tsa: Fast and efficient reconstruction for the Inner Tracker. Technical Report LHCb-2004-075. CERN-LHCb-2004-075, CERN, Geneva, Sep 2004.

[72] Luca Lista. Practical statistics for particle physicists. *CERN Yellow Reports: School Proceedings*, 5(0):213, 2017.

[73] Rutger M. van der Eijk. *Track reconstruction in the LHCb experiment.* PhD thesis, Amsterdam U., 2002.

[74] M Witek. VELO-TT matching and momentum determination at Level-1 trigger. Technical Report LHCb-2003-060, CERN, Geneva, Aug 2003.

[75] Y Xie. Short track reconstruction with VELO and TT. Technical Report LHCb-2003-100, CERN, Geneva, Aug 2003.

[76] Adrian Perieanu. Identification of Ghost Tracks using Neural Networks. Technical Report LHCb 2007-158, CERN, Geneva, Dec 2007.

[77] M Needham. Identification of Ghost Tracks using a Likelihood Method. Technical Report LHCb-2008-026. CERN-LHCb-2008-026. LPHE-2008-004, CERN, Geneva, May 2008.

[78] Michel De Cian, Stephen Farry, Paul Seyfert, and Sascha Stahl. Fast neural-net based fake track rejection in the LHCb reconstruction. Technical Report LHCb-PUB-2017-011. CERN-LHCb-PUB-2017-011, CERN, Geneva, Mar 2017.

[79] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. Tmva - toolkit for multivariate data analysis. `https://arxiv.org/abs/physics/0703039v5`, 2007.

[80] Roel Aaij, Johannes Albrecht, Francesco Dettori, Kevin Dungs, Helder Lopes, Diego Martinez Santos, Jessica Prisciandaro, Barbara Sciascia, Vasileios Syropoulos, Sascha

Stahl, and Ricardo Vazquez Gomez. Optimization of the muon reconstruction algorithms for LHCb Run 2. Technical Report LHCb-PUB-2017-007. CERN-LHCb-PUB-2017-007, CERN, Geneva, Feb 2017.

[81] The LHCb collaboration. Measurement of the track reconstruction efficiency at lhcb. *Journal of Instrumentation*, 10(02):P02007, 2015.