Measuring Matter Antimatter Asymmetries at the Large Hadron Collider

M. Gersabeck[1], C. Parkes[1].

[1] *The University of Manchester, Manchester, United Kingdom*

Abstract

This project aims to measure a fundamental difference between the behaviour of matter and antimatter through the analysis of data collected by the LHCb experiment at the Large Hadron Collider. The three-body decays $B^{\pm} \to h^{\pm}h^{+}h^{-}$, where $h^{\pm}$ is a $\pi^{\pm}$or $K^{\pm}$ are studied. The inclusive matter antimatter asymmetry is calculated, and larger asymmetries are searched for in localized regions of the phase-space.

Last updated: March 18, 2022

1 Aims & Objectives of Laboratory Project

Knowledge of special relativity and basic statistical techniques, as taught in the first year courses, are assumed. A detailed knowledge of particle physics is not required for the project.

This project assumes basic familiarity with `Python` syntax.

1.1 Aims

1. To utilise and understand data analysis techniques used in modern particle physics experiments.
2. To understand how matter antimatter differences can be observed.
3. To understand the design of a particle physics detector.

1.2 Objectives

1. To write a `Python` program to select and analyse three-body decays of charged B mesons.
2. To use the data to observe intermediate resonance states in the decays.
3. To use the data to observe matter antimatter differences.

1.3 Intended Learning Outcomes

By the end of this lab experiment you should be able to

1. Use detector output to calculate derived quantities and separate signal from background candidates
2. Produce high-quality plots to display particle physics quantities
3. Fit models to data distributions to determine model parameters
4. Evaluate statistical and systematic uncertainties and judge the significance of results based on their uncertainties

2 Introduction

The Large Hadron Collider [1] (LHC) is a particle accelerator at CERN, the Eurpoean Organisation for Nuclear Research, located near Geneva. In the data analysed in this project, the LHC collided protons with protons at energies of 3.5 TeV, giving a centre-of-mass energy of 7 TeV. The LHC beams consist of around 2000 bunches of protons, with each bunch containing around 10^{11} protons. For the data sample used here, the bunches from the two beams were crossed in the LHCb detector at a rate of 15 MHz. Each time the beams cross (called 'events') there is a high probability of one or more collisions, that is inelastic interactions between the constituents of the protons.

The first physics runs of the LHC started in 2010 and operations continued until early 2013. The second running period lasted from 2015 until 2018. The LHC, with various upgrades, is expected to continue until the late 2030s. The main aims of the LHC experiments are to test the Standard Model (SM) of particle physics and to search for new phenomena beyond the SM.

LHCb [2] is one of the four large experiments at the LHC. It is an experiment dedicated to searching for new physics from beyond the Standard Model through the study of matter antimatter differences (*CP* Violation) and rare decays of hadrons containing b and c quarks. The design of the detector is described in Sect. 4. In this project you will analyse data recorded by the experiment to attempt to observe *CP* Violation.

The processes being observed are rare, so statistical analysis is used to sieve information from large quantities of data. You have been given a subset of data collected by the LHCb experiment which has been preselected (*i.e.* filtered) to contain events of primary interest for this analysis.

3 CP Violation

Why does the universe contain much more matter than antimatter? This is one of the major outstanding questions of fundamental physics. The big-bang is expected to have produced equal amounts of matter and antimatter. The universe today is made primarily of matter. A fundamental difference between the behaviour of matter and antimatter (*CP* Violation) is one of the requirements, known as the Sakharov conditions, to explain how this can have arisen [3].

CP-symmetry is the product of two symmetries: C for charge conjugation that transforms a particle into its antiparticle, and P for parity that reflects the spatial co-ordinates. A violation of this symmetry was discovered in 1964 [4] in weak decays of the kaons for which the 1980 Nobel Prize was awarded. A mechanism to include *CP* Violation in the SM (through the CKM matrix [5]) was suggested in 1973 and verified through tests of *CP* Violation in weak decays of the B mesons at the BaBar and Belle experiments from 1999-2010 [6], with the 2008 Nobel Prize being awarded for this work. *CP* violation in charm decays were discovered by the LHCb experiment in 2019 [7]. Further information on *CP* Violation can be found in Refs. [8–10].

The LHC experiments have discovered the last particle of the SM, the Higgs boson [11, 12], and are engaged in searching for new physics from beyond the SM. The SM of particle-physics includes *CP* Violation, but the magnitude of the effect is far too small to explain the matter antimatter imbalance in the universe. Consequently additional sources of *CP* Violation are required. New particles from beyond the SM will in general add additional sources of *CP* Violation. Hence measurements of *CP* Violation provide an important method of searching for new phenomena from beyond the SM. These new physics particles could potentially also provide explanations for other major unanswered questions such as the nature of dark-matter.

Experimental studies of *CP* Violation can be conducted by comparing the decays of a particle and antiparticle state. The LHCb experiment is the experiment at the LHC dedicated to such studies and has produced a number of major results. Regularly updated highlights from its physics programme are described on the web pages at [13, 14]. The analysis undertaken here is based on the largest *CP* Violation effects that have been observed [15, 16].

4 Detector, software and data sample

The LHCb detector [2, 17] is a single-arm forward spectrometer and is shown in Fig. 1 as operated during the acquisition of the data analysed here. A right-handed co-ordinate system is used with the z axis along the beam into the detecor and the y axis vertical. The negative (positive) z side of the detector is known as upstream (downstream). The detector is designed for the study of particles containing b (or c) quarks. At LHC energies these particles are produced with a strong forward boost along the directions of the proton beams, and the decay products of those with a polar angle of 15–300 mrad can be observed in the detector. The hadron containing the b quark decays after travelling a few mm to a cm in the detector, so is reconstructed from its decay products as shown in the example in Fig. 2.

The detector includes a high-precision tracking system which detects charged particles. A silicon-strip vertex detector, known as the vertex locator (VELO) [19], surrounds the pp interaction region and measures at least three (typically ten) points on each track. This allows the position of the pp interaction point (primary vertex) to be reconstructed and the position of the decay of the hadron containing the b quark (secondary vertex) to be measured. The charged particles passing through the silicon create electron-hole pairs and these are drifted in an electric field across the silicon, and a signal obtained. A large-area silicon-strip detector, known as TT, has four detection layers and is located upstream of the dipole magnet whose field is oriented vertically. Charged particles are bent by the magnetic force, and the dipole magnet has a bending power of about 4 Tm. The direction of the magnetic field is regularly reversed, denoted as 'Magnet Up' and 'Magnet Down'. This allows systematic uncertainties from oppositely charged particles bending into different regions of the detector, with potentially different detection efficiencies, to be controlled. Three stations of silicon-strip detectors (each with four detection layers) and

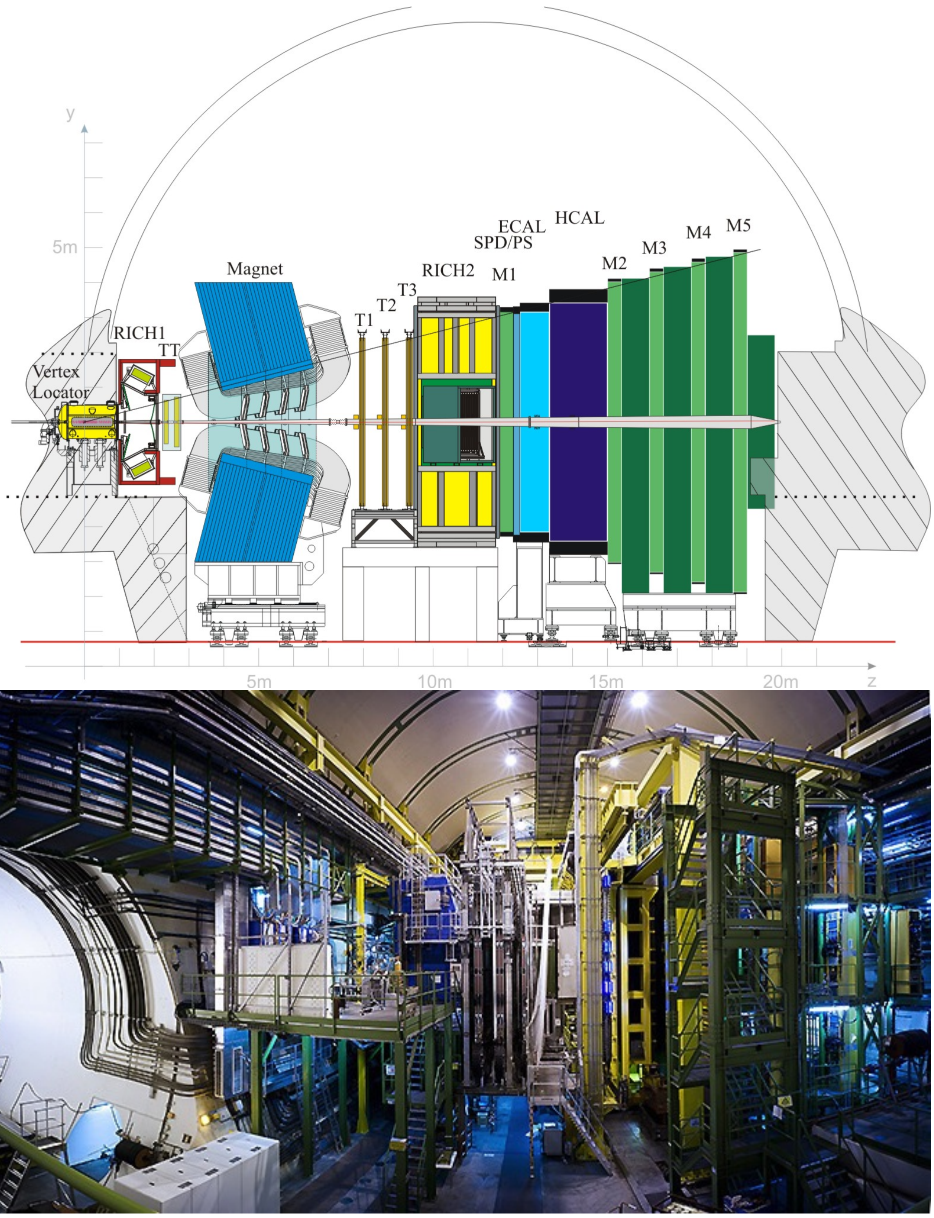

Figure 1: (top) Layout of the LHCb experiment. The beams travel along the horizontal pipe in the centre of the detector and collide in the Vertex Locator at the left hand side of the diagram. The subdetector components are described in the text. (bottom) Photograph of the LHCb cavern. The image has been mirror reflected to be in the same orientation as the layout diagram.

straw drift tubes [20] are placed downstream of the magnet and are known as T1, T2, T3. The straw drift tubes are filled with a gas (Ar, CO_2 and O_2) that is ionised by the charged particles passing through, and this charge is drifted under an electric field to be measured.

The measured points are used to reconstruct the track of charged particles. The

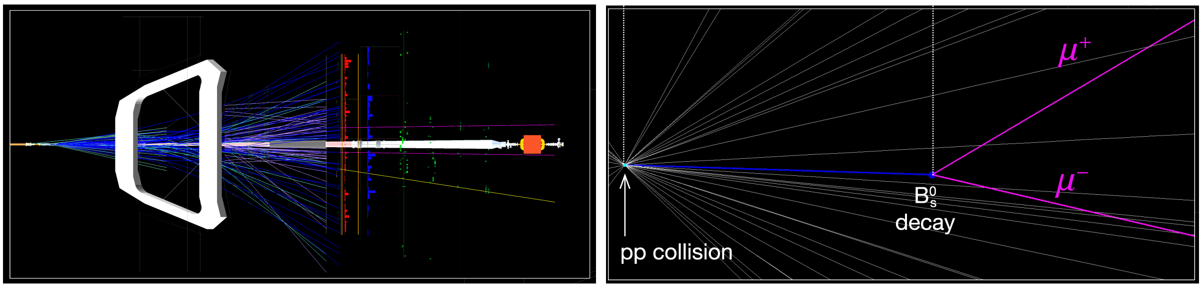

Figure 2: Event display images showing tracks reconstructed in a candidate $B^0_s \to \mu^+\mu^-$ event. The first evidence for this decay [18] was one of the major physics results from the LHC run 1 as it places significant constraints on new physics in Supersymmetry models. (left) All reconstructed tracks in LHCb. (right) A zoom on the VELO region, showing the reconstructed primary vertex and secondary decay vertex of the B^0_s meson separated by 14 mm. The reconstructed muon tracks are highlighted in magenta.

momentum of the charged particle can be obtained from the bend of the track. The combined tracking system provides a momentum measurement with a relative uncertainty that varies from 0.5% at low momentum, p, to 1% at 200 GeV/c, and an impact parameter (see section 5.3) measurement with a resolution of 20 μm for charged particles with large transverse momentum, p_{T}.

Different types of charged hadrons (pions, kaons, and protons) are distinguished using information from two ring-imaging Cherenkov detectors [21]. Charged particles travel through volumes of gases (CF_4, C_4F_{10}) or ultra-light solids (aerogel) faster than the local phase velocity of light in those materials, and radiate photons. This is known as the Cherenkov effect. These photons are measured and the angle at which the light is given off is directly related to the velocity of the particle. By combining the velocity information with the momentum information, the particle mass can be deduced and used to determine the type of the charged hadron.

Photon, electron and hadron candidates are identified by a calorimeter system consisting of scintillating-pad (SPD) and preshower (PS) detectors, an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). The calorimeters are based on alternating layers of iron and scintillator. The particles interact and are absorbed in the iron and the amount of scintillation light detected allows the reconstruction of the energy of the paticles. Photons and electrons are primarily absorbed in the electromagnetic calorimeter. Hadrons are primarily absorbed in the hadron calorimeter. Charged hadrons, which are the particles analysed in this experiment, have a relative energy resolution in the HCAL ranging from 70% at 1 GeV to 10% at 200 GeV [22]. Only neutrinos, which remain undetected, and muons typically pass through the calorimeters. Muons are identified by a system (M1–5) composed of alternating layers of iron and multiwire proportional chambers [23].

Around 11 MHz of bunch-bunch crossings contain one or more inelastic pp interactions in the data sample studied here. The size of the event data produced by the LHCb detector means that this could only be written to disc at a rate of 3 kHz. The system used to determine which events to store is known as the trigger [24]. It consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which applies a full event reconstruction. The software stage is run in a large farm of multiprocessor CPUs, which run 26,110 copies of the program. The trigger aims to retain as high efficiency as possible for the decays channels that will be studied while rejecting events of low interest.

The data is then processed at CERN and by computers in an worldwide computing system known as the GRID. The data undergoes preselection to write out events of interest for paticular physics processes. Individual groups of analysts then write programs, using the LHCb software framework, to perform further selection and reduce the data size storing only the variables of primary interest. The data that you are provided with in the experiment is of this form.

Simulations, in which the physics process under study or background processes are also generated for comparison with the data. These may also include detailed simulations of the response of the detector, so that the simulation data is directly comparable with the data. Simplified simulation data, not including detector repsonse, is provided here.

The data sample provided has been selected by the trigger and offline selection from approximately 10^{14} beam crossings that occurred at LHCb in 2011. The sample provided includes only the most important reconstructed variables needed to perform this analysis in order to reduce the data file size and decrease the processing time for your analysis. The samples contain 3.4 million (5.1 million) events corresponding to an integrated luminosity of 434 pb^{-1} (584 pb^{-1}) with the magnet in the "up" ("down") polarity.

5 Experimental Strategy

The analysis performed here compares the rates for the decays $B^+ \to h^+h^+h^-$ and its antiparticle equivalent $B^- \to h^-h^-h^+$, where $h^\pm$ is a $\pi^\pm$ or $K^\pm$. The B^+ mesons are composed of a $\overline{b}$ quark and u quark. The π^+ meson is composed of $\overline{d}$ and u quarks. The K^+ meson is composed of u and $\overline{s}$ quarks. The decays studied proceed through the weak force (via $W^\pm$ bosons), where the CP symmetry is known to be violated.[1]

You will be assigned by your demonstrator to perform the analysis of one of the three decays:

- $B^+ \to \pi^+\pi^+\pi^-$
- $B^+ \to K^+K^+K^-$
- $B^+ \to K^+\pi^+\pi^-$.

A simple quark flow diagram can also be written down for the $B^+ \to \pi^+K^+K^-$ decay channel but experimentally the decay has somewhat higher background and is therefore not included in this experiment.

[1]The observation of CP Violation requires interference between two decay routes to the same final state (this is beyond the scope of this project and is discussed in the Frontiers in Particle Physics 2 course [9]).

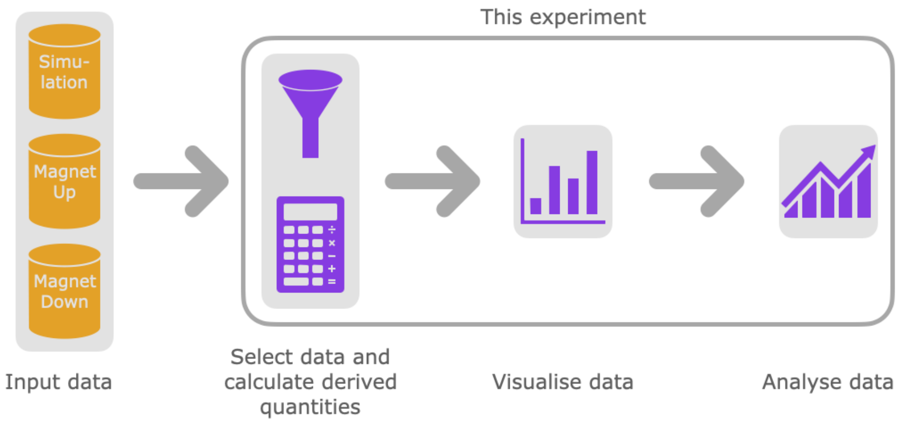

Figure 3: Schematic depiction of the data analysis workflow of this experiment. The input data shown on the left is provided.

5.1 Getting Started

A general overview of the data analysis workflow of this experiment is given in Fig. 3. You are given three input data files, which are in the `.root` file format that is commonly used for storing particle physics data. These have been produced with the CERN `ROOT` software, but you do not need to use this yourself. The `Python` module `uproot` takes care of reading in these data. The data stored in the `.root` files can be visualised as big tables with one row per recorded collision event and the columns containing the values of characteristic variables for the three measured final state particles. The variables available are given in Appendix A.

Two samples of data collected by the LHCb experiment are provided with different orientations of the magnetic field. Data in the same format from a simple simulation of the decay channel $B^+ \to K^+K^+K^-$ are also provided. This simulated data only has the momentum and charge variables filled, all other variables should not be used.

An example analysis code is provided in the form of a `Jupyter` notebook. This contains three cells of code, each corresponding to one of the three stages in Fig. 3.

5.1.1 Setting up

There are different ways of running this lab and the best way for you depends on your system and how comfortable you are with installing new software.

Web based (default)
The default mode for this lab is built on the University's web-based `JupyterHub`, which runs `Jupyter` notebook servers on the Cloud.

To get started, launch the notebook server by going to the following URL in your favourite browser:

Figure 4: Screenshot after launching the `Jupyter` server.

https://jupyter.cs.man.ac.uk/hub/user-redirect/git-pull?repo=https%3A%2F%2Fgithub.com%2Fgersabec%2FUoM_MatterAntimatterLab&urlpath=tree%2FUoM_MatterAntimatterLab%2F&branch=master

The first time you do this will create your personal account on the service, which is based on your University login. Once the server is launched (see Fig. 4), execute the example code by clicking on `Analysis_example.ipynb`. This will open a new tab in which the notebook runs.

The notebook automatically saves, but you can equally save and download the file to your computer as a backup. Finally, you will need to download your output files, which will be your plots. To do this, go to the first tab that has the list of files as shown in Fig. 4 and tick the box next to the file you want to download (you need to do this one file at a time). At this point a `Download` button will appear above the file list with which you can download whichever of the files you wish to download. Please don't use this to download the large data files but rather follow the instructions given below for the local installation.

Local installation

This is recommended for students who are familiar with installing `Python packages`.

You can download the example code from the following `GitHub` repository:

https://github.com/gersabec/UoM_MatterAntimatterLab

On a `Linux/OS X` system you can execute the `postBuild` file to download the data. On `Windows` you just need to go to the URLs contained in this file. Move the data to the same directory as the code.

The next step is to ensure that you have the `Python` packages installed that are listed as dependencies in the file `environment.yml`.

In a command line terminal launch the server by executing the following command:

```
jupyter notebook
```

Then proceed as above by clicking on `Analysis_example.ipynb`. This will open a

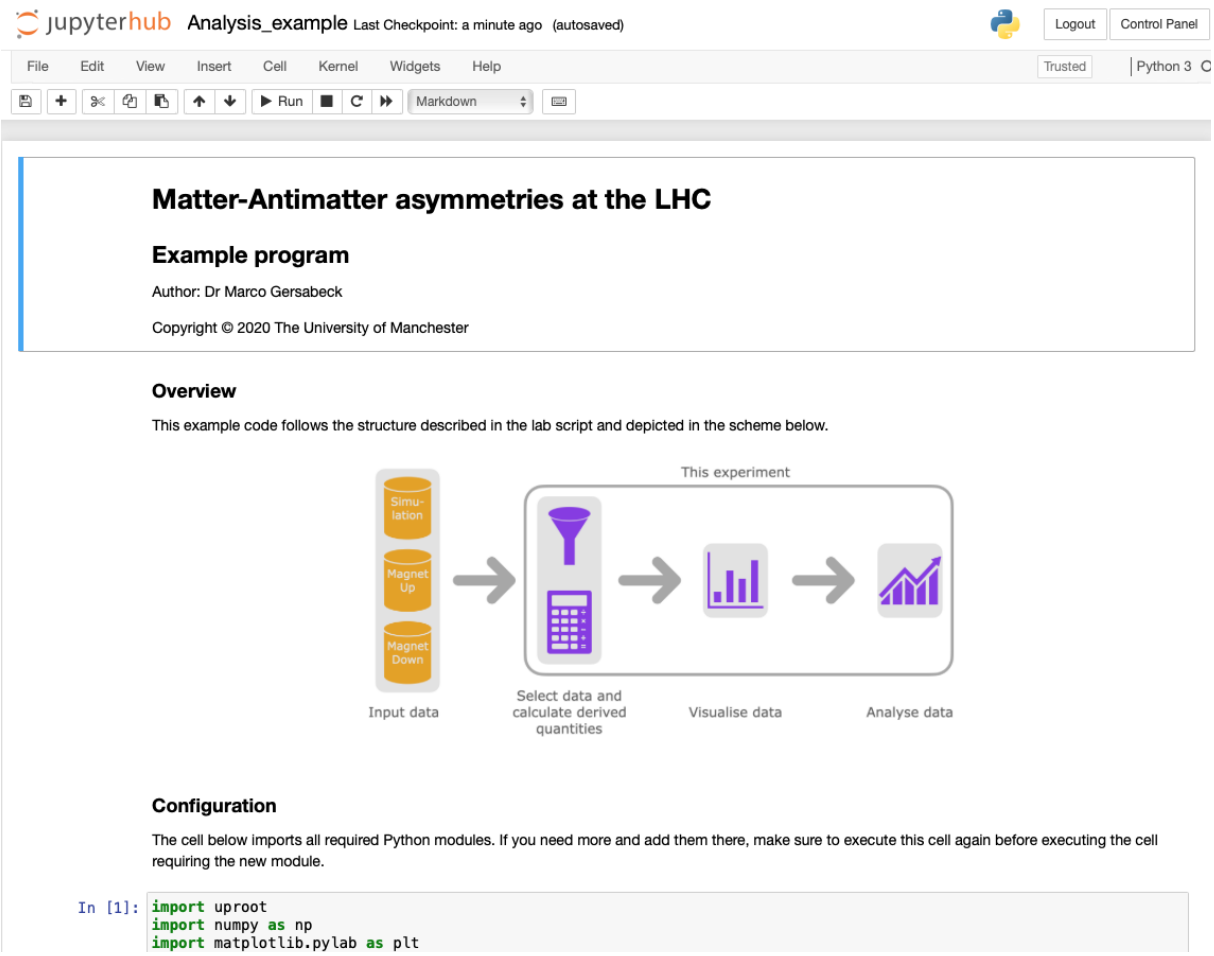

Figure 5: Screenshot after launching the `Jupyter` notebook.

new tab in which the notebook runs.

Aim: Run the example program. Display the example histograms that are produced from the data and run the analysis.

Hints:

You can execute each cell in the notebook by pressing `Ctrl+Enter` *or* `Shift+Enter`*, where the latter will then move the focus to the next cell. Alternatively you can use the menu and select* `Cell` → `Run Cells`*.*

Aim: Learn how to book, fill, save and plot histograms. Produce a plot of the probability that the final state particles are pions or kaons. You can find the list of variables in the appendix. Note that these variables are not filled in simulation, so run over e.g. 50000 events of collision data. Please discuss the plot with your demonstrator.

5.2 Invariant Mass Reconstruction

5.2.1 Invariant mass

Aim:

Produce a plot of the invariant mass of the B meson using the simulation data file. Assume that all final state particles are kaons.

Hints: The invariant mass can be calculated from special relativity using energy and momentum conservation of the decay process. The energies of the final state particles should be calculated from the measured momenta and known masses [25]. Derive the required formulas on paper first, and then code them. The formula will not be entirely trivial and it will be helpful to use intermediate variables in the calculation.

The simulation data should be used for these plots. You should obtain a sharp peak at the known B meson mass with a width much less than $1\,\mathrm{MeV}$. *The plots for the experimental data are more complex to understand as they contain a mixture of pions and kaons and background events. The plots for the experimental data will be analysed later.*

Please discuss your plot with your demonstrator.

5.2.2 Momentum resolution

The invariant mass reconstruction relies on the measurement of the particle momenta. The direction of the momentum vector is determined predominantly by the vertex detector. The magnitude of the momentum vector is determined by the combination of the dipole magnet and the tracking stations (see Fig. 6).

Aim: (i) Show that, for a dipole magnetic field in y direction and an initial momentum in z direction, the change in the x component of the momentum in units of GeV/c is given by

$$\frac{\Delta p_x}{q} = 0.3 \int B_y dx, \tag{1}$$

with q the particle charge in units of e. You may assume that $p_z \gg p_x$.

(ii) The bending power of the LHCb dipole magnet is $\int B_y dx = 4$ Tm. The momentum is measured by the deflection of the particle in the x direction. Show that the relative momentum resolution is given by

$$\frac{\sigma(p_z)}{p_z} = \frac{\sigma(\Delta x)}{\Delta x}, \tag{2}$$

with $\sigma(i)$ the absolute measurement uncertainty for i.

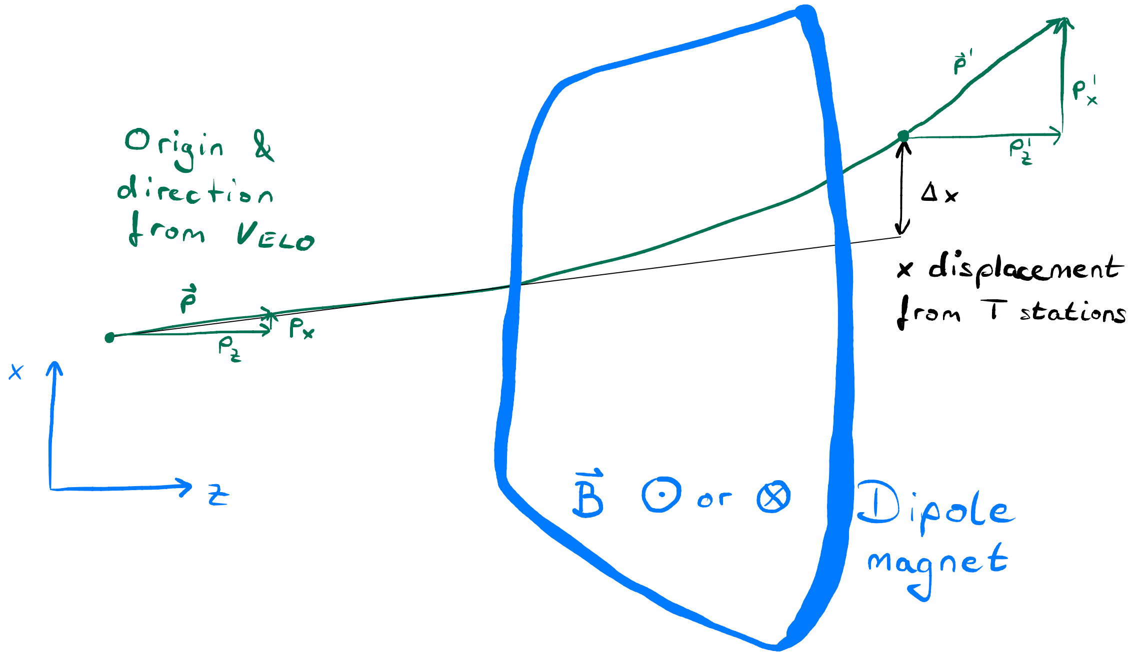

Figure 6: Schematic of momentum measurement.

Hint: Relate p_z to Δx with Δp_x.

(iii) Assuming a position resolution in the x direction of $\sigma(x) = 1$ mm after the dipole magnet and a negligible uncertainty of the vertex detector, determine the relative momentum resolution for a particle of total momentum of 10 GeV/c and of 100 GeV/c. Compare this to the calorimeter energy resolution given in Sec. 4.

Please discuss your explanation with your demonstrator.

5.3 Selection

The data sample you are given has pre-selection cuts applied to select a sample of events with relatively low background. The most important pre-selection cuts that have been applied are listed in table 1, and explained here:

- *Momentum.* The three final state tracks should all have significant momentum (p) and momentum in the plane transverse to the beam direction (p_{T}) as they originate from a B decay, whereas tracks from background processes will typically have low momenta.

Variable	Selection Cut
Track Transverse Momentum (p_{T})	$> 0.1\,\mathrm{GeV}/c$
Sum of p_{T} of Tracks	$> 4.5\,\mathrm{GeV}/c$
Track Momentum (p)	$> 1.5\,\mathrm{GeV}/c$
$B^{\pm}$ candidate mass assuming all tracks are $K^{\pm}$ (M_{KKK})	$5.05 < M_{KKK} < 6.30\,\mathrm{GeV}/c^2$
Track Impact Parameter (IP) χ^2	> 1
Sum of IP χ^2 of Tracks	> 500
$B^{\pm}$ candidate vertex fit χ^2	< 12

Table 1: The most important pre-selection cuts applied to the data sample (see text for details).

- $B^{\pm}$ Mass. The mass of the candidate $B^{\pm}$ meson can be computed from the measured momenta of the final state particles (under an assumption as to whether they are kaons or pions) and events with masses close to the known mass of this particle [25] are selected.

- *Impact Parameters.* The $B^{\pm}$ meson has a mean lifetime τ of 1.6×10^{-12} s [25], and will travel a distance $\gamma c\tau$ before decaying, where γ is the Lorentz factor of special relativity. This distance is typically a few mm in LHCb. The distance between its production point where the proton beams collide, known as the primary vertex, and the decay vertex into the three final state tracks is key to identifying a clean sample of events. The closest distance of approach of the final state tracks to the primary vertex is known as the impact parameter (IP). These tracks will thus have a significant IP. The IPχ^2 is the IP divided by its estimated uncertainty all squared. Hence a track from a background process that originates at the primary vertex will typically have a low IPχ^2.

The sample has not had any identification of the final state particles applied and hence contains all combinations of charged pions and kaons. The event selection is common between all the $B^{\pm} \to h^{+}h^{+}h^{-}$ channels as the kinematics of all the channels will be very similar. Using the information from the particle identification systems [21] of the detector, particularly the RICH detectors to separate pions and kaons, the decay channel that is to be studied can be separated out. This is a statistical process, there will still be a significant mis-identification rate of the h particle as pion/kaon in order to retain a high efficiency for selecting the signal.

Aim:
(i) Select the sample of events for the channel that you will study using particle identification information.
(ii) Show that a clear mass peak is observed in the data for the $B^{\pm}$ once the correct mass hypotheses for the final state particles are assumed. Identify that there are regions in

the mass plot dominated by signal events and regions dominated by background (known as side-bands).

(iii) Discuss why the signal peak appears to have approximately the shape of a Gaussian distribution.

Please discuss the selection, mass plot and Gaussian shape with your demonstrator.

Hints:

You do not need to repeat the cuts listed in Table 1 as that pre-selection has already been applied to the data set and only need to work with pion/kaon identification variables. In addition, you should require that none of the particles has been identified as a muon (consult the list of variables in the appendix to find out how to do this).

Students studying the decay $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}$ will need to determine which final state particle is most likely to be the kaon in each event in order to impose the correct mass hypothesis for the final state particles. Make use of the charge of the particles when doing this.

Students studying the decays $B^{\pm} \to \pi^{\pm}\pi^{+}\pi^{-}$ or $B^{\pm} \to K^{\pm}K^{+}K^{-}$ should consider applying different particle identification requirements on the three particles to exploit the fact that the values of the particle identification variables even for true pions or kaons are random samples of a distribution.

5.4 Two-body Resonances

The decays can proceed either directly to the three-body final state or via an intermediate particle. For example, $B^{+} \to h_1^{+}h_2^{+}h_3^{-}$, could proceed through the decay $B^{+} \to h_1^{+}R^{0}$, where R^{0} is a neutral particle resonance which can decay $R^{0} \to h_2^{+}h_3^{-}$.

Aim: Plot the two neutral two-body invariant mass distributions and identify the two-body resonances in the data for your decay channel. Please discuss your findings with your demonstrator.

Hints: The mass, width and other properties of all known particles can be obtained from [25].

You can achieve a cleaner picture by including in the plots only events for which the three-body invariant mass lies within the $B^{\pm}$ peak observed in the previous section.

For the channels with all the same type of final state particles ($B^{\pm} \to \pi^{+}\pi^{+}\pi^{-}$ and $B^{\pm} \to K^{+}K^{+}K^{-}$), there are two identical pairings ($\pi^{+}\pi^{-}$ or $K^{+}K^{-}$). It is conventional to distinguish these by plotting the higher and lower mass pairs of each event.

The simulation contains no resonances so you should not observe these peaks in that sample.

The analysis performed here is to study the CP Violation in the charmless B meson decays to pions and kaons. However, the most frequent decays of the B mesons occur through the b quark decaying into a c quark. These events can be discarded by removing the decays that may have proceeded via a D^0 meson (which contains the c quark).

There is a small probability of a muon track being mis-identifed as a charged pion track in the detector. Consequently the $B^\pm \to J/\psi K^\pm$, with $J/\psi \to \mu^+\mu^-$ can be a background process that fakes the signal if both muons are misidentified as pions.

Aim: Reject events from D^0 meson decays and (if relevant) from $B^\pm \to J/\psi K^\pm$ from the event sample you are studying. Do not reject any other resonance or part of the data set. Please discuss your particle rejection with your demonstrator.

Hints: The D^0 and possibly the J/ψ should have appeared in the two-body invariant mass distributions you studied before. Select suitable mass ranges around these peaks and veto any events falling inside these ranges.
Consult Ref. [25] to determine which D^0 decay final state involving two charged hadrons is the most likely. If you are analysing a different final state you may see a slightly broader distribution next to the expected D^0 peak, which comes from the dominant decay mode and involves $K \longleftrightarrow \pi$ mis-identification. This should be rejected as well and can be done either by rejecting the broad distribution or, better, by plotting the same events under the mass hypothesis of the dominant final state and rejecting a narrow range around the nominal D^0 mass there.
The J/ψ may have been sufficiently suppressed by the muon veto applied in the selection, so if it is not visible you're find and can move on.

5.5 Global CP Asymmetry

The *CP* asymmetry A is defined as:

$$A = \frac{N^- - N^+}{N^- + N^+},$$

where N^+ is the number of $B^+ \to h^+h^+h^-$ events observed in the channel analysed and N^- the number for the corresponding B^- decay. The value of $N^\pm$ can be estimated by fitting the $B^\pm$ invariant mass distribution to extract the signal event contribution.

Other effects can fake a *CP* asymmetry. The LHC is a proton-proton collider and hence the initial state is not matter antimatter symmetric. While the primary production mechanisms will produce $b\bar{b}$ quark-pairs there may be a small production asymmetry, it is estimated that this could be approximately 1%. The LHCb detector could be more efficient for detecting the B^+ or B^- final states as the positive and negative particles will be bent by the magnet in different directions in the detector[2].

[2]Another mechanism is that final state particles of one charge can have a higher probability of

Aim: (i) Fit the 3-body invariant mass distribution with an appropriate model and determine $N^{\pm}$ from the number of events contributing to the peak around the nominal B mass for the $B^{\pm}$ sample, respectively. Calculate the global CP asymmetry. Have you found evidence for CP violation?

(ii) Show that the statistical uncertainty on the asymmetry is given by $\sigma_A = \sqrt{\frac{1-A^2}{N^- + N^+}}$.

(iii) Discuss the limitations of the formula in (ii) as this only considers signal events.

Please discuss the fit model you plan to use with your demonstrator prior to implementing it. Please discuss your results and the formula for the uncertainty with your demonstrator.

Hints: Correctly reconstructed signal decays will peak around the $B^{\pm}$ mass. Background events from random combinations of tracks that have passed the event selection cuts will have a much smoother variation in mass. Peaking backgrounds from four-body decays of $B^{\pm}$ mesons where only three tracks have been observed may also be present.

An example how to fit a function to data points is provided. You should decide on appropriate functions to use to describe the line shape and write the appropriate code. As a minimum you will need a Gaussian function and an exponential (provided in the example). You should give the program reasonable starting values for the free parameters to allow the fit to converge.

The asymmetry will be a binomially distributed random variable, and its statistical uncertainty must be calculated appropriately, see for example Ref. [26]. The asymmetry can be rewritten in terms of efficiencies $\epsilon_1 = \frac{N^-}{N^-+N^+}, \epsilon_2 = \frac{N^+}{N^-+N^+}$. The uncertainty on these efficiencies can then be calculated. The uncertainty on the asymmetry is calculated most easily if the one of the two efficiencies is eliminated from the equation you just derived.

5.6 Dalitz Plots

The Dalitz Plot [27] is an important technique for analysing three-body decays. The kinematics of a three-body decay can be fully described using only two variables. The axes for a standard Dalitz Plot are the squares of the invariant masses of two pairs of two-body masses. Consequently, the intermediate resonances will be visible as bands on the Dalitz Plot. Hence in the decay $B^+ \to h_1^+ h_2^+ h_3^-$, $M^2_{h_1^+ h_3^-}$ and $M^2_{h_2^+ h_3^-}$ are a potential choice of variables. The kinematic constraints of a Dalitz Plot are shown in Fig. 7. The main properties of a Dalitz Plot are:

- Decays that proceed simply according to phase space, not proceeding via resonances, uniformly populate the Dalitz Plot.

interacting with the detector than those of the other. This is a consequence of the detector being made of matter rather than anti-matter.

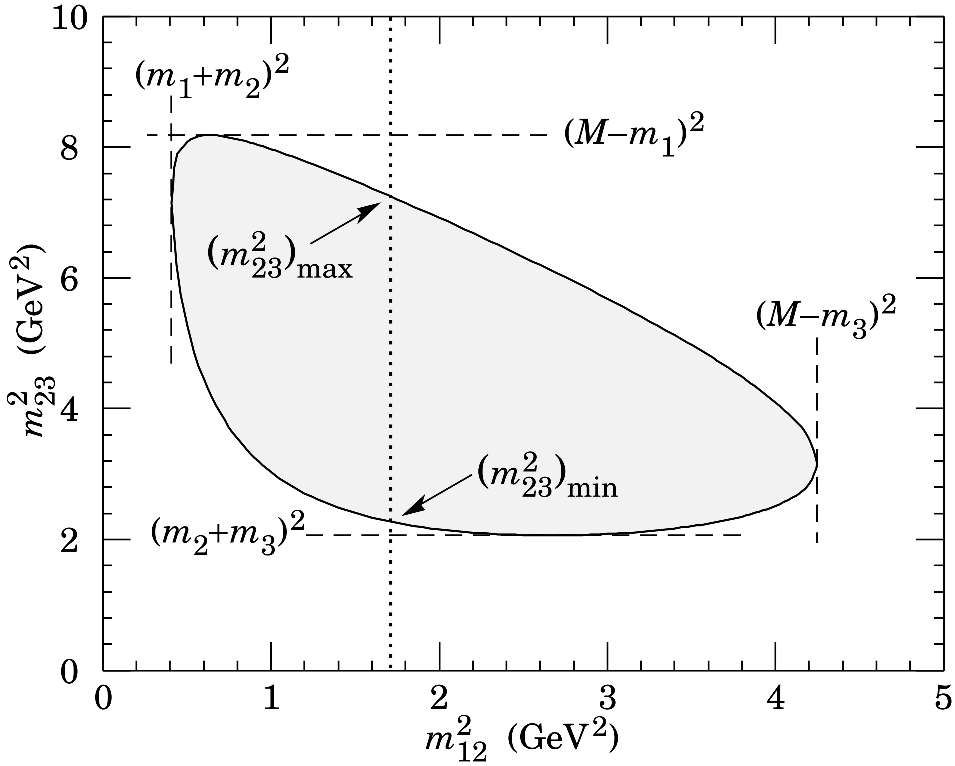

Figure 7: Dalitz Plot for an example three body final state. Four-momentum conservation restricts events to the shaded region. Reproduced from Ref. [25], where additional information on the technique is available.

- Resonances appear as bands of events in the Dalitz Plot.
- The position and size of the band relate to the mass and width of the resonance.
- The spin of the resonance governs the distribution of events along the band. Spin 0 (scalar) resonances will produce uniformly populated bands. Spin 1 (vector) resonances will show a minimum in the band.
- The exact pattern of events in the Dalitz Plot is determined by interference between the various contributing states.

Aim:

(i) Produce a Dalitz Plot for the data sample you are analysing.

(ii) Produce a Dalitz Plot which describes only the signal events in the data sample by subtracting the background.

(iii) Revisit your identification of two-body resonances and observe how they are distributed on the Dalitz Plot.

Please discuss your approach to subtracting the background component with your demonstrator as well as the resulting plots and observations.

Hints: You may wish to produce the Dalitz Plots with several alternative binnings using 2D histograms. Fine bins may allow you to observe the resonances more easily. However, for the asymmetry analysis below, large bins (say $2.5\,GeV^2/c^4 \times 2.5\,GeV^2/c^4$) will be needed to keep the uncertainties on the number of entries in each bin small. If preferred, you can generate a histogram with variable bin widths by passing a list of bin boundaries when calling `hist2d`.

As in Sect. 5.4 it is conventional to use the higher and lower mass h^+h^- pairs in each event for the case of final states with all pions or all kaons.

Discuss with your demonstrator your ideas for performing the background subtraction and specifically how to treat cases where this subtraction leads to negative entries.

5.7 Local CP Asymmetry

In Sect. 5.5 the global *CP* asymmetry was calculated. However, *CP* Violation may arise from interference between decays through different resonances, and hence the magnitude and sign of the *CP* Violation may vary across the Dalitz Plot. Equation 5.5 can also be applied to calculate the *CP* asymmetry in local regions of the plot. Consider the statistical and systematic uncertainty on your observed *CP* Violation. The statistical significance (asymmetry divided by error on asymmetry) of the *CP* Violation in these regions can also be determined, to identify if you have obtained strong evidence for *CP* Violation. Results exceeding 3 sigma are usually labelled in particle physics as "evidence" and those exceeding 5 sigma an "observation" or "discovery".

Aim:

(i) Produce plots to show *CP* Violation across the Dalitz Plot.

(ii) Identify any regions in which you find significant evidence for *CP* Violation. Determine the magnitude and significance of the *CP* Violation.

(iii) You should produce three Dalitz plots: the asymmetry, the error on the asymmetry, and the significance.

Please discuss your results with your demonstrator.

Hints: You need to calculate the error on the asymmetry as well. You can use the formula derived for the error on the global asymmetry or an alternative as you will have discussed. In order to inspect regions of significant local asymmetries further, it can be instructive to plot two-body or three-body invariant mass distributions only for events falling into the regions you want to investigate.

5.8 Systematic Uncertainties & Further Work

A *CP* asymmetry effect could be faked by either a production or detection effect (as discussed in Sect. 5.5). Other sources of systematic uncertainty may be related to your analysis technique.

5.8.1 Detector asymmetries

The detector is not perfectly symmetric, either due to built-in elements like staggered detector components or due to localised malfunctioning components. The final states under study contain a different number of positively and negatively charged particles. Any localised left-right asymmetry of the detector will be reversed when then polarity of the magnetic field is swapped. To cancel such detector-induced asymmetries results from both magnet polarities are typically averaged.
Aim: Calculate the global asymmetries separately for each magnet polarity and determine the average to cancel detector asymmetries. Should you use a weighted average or an arithmetic average? Should you assign a corresponding systematic uncertainty, if so how would you evaluate this? Please discuss your approach with your demonstrator.

5.8.2 Further work

Aim: Consider the sources of systematic uncertainty on your measurement, and where possible estimate their contribution to your *CP* asymmetries.

Consider any additional analyses that could be performed with this data sample. Examples for possible extensions are:

- If you have not yet done so, perform a statistical study to find the optimal set of particle identification criteria.
- If you have not yet done so, you may want to compare different fit models for signal and/or background and compare these in a statistically meaningful way.
- If you have not yet done so, you could implement variable bin widths for your Dalitz plot as described in Sec. 5.6.
- If you have identified a region of large local asymmetries, you could perform a fit to the three-body invariant mass exclusively for events in that region to obtain a more precise estimate of the asymmetry in the region.
- You could also perform a statistical analysis of the local asymmetry significance values. In the CP symmetric case, these should be distributed according to a normal distribution, so you could test the compatibility with that hypothesis. The method originates in Ref. [28] and allows for a correction of effects of a global asymmetry, which is best explained in Ref. [29].

Please discuss your ideas and outcomes with your demonstrator.

6 Additional Information

This section is meant to give further information concerning the report and presentation for the interview. In general, try to give your work a professional look and be consistent in the layout. This section contains hints, but cannot be exhaustive, and should be undertstood as a guidance rather than a strict prescription. You should consult the marking rubric to be aware of the goals for excellent reports and interviews, not all of which are touched upon here.

6.1 Report

The report should have a clear structure including

- Abstract: Collating in few sentences a description of the topic, the methodology, and the main result.
- Introduction: This should cover a connection of the topic in question to the wider field, a brief overview of the theoretical background, a brief review of existing results, and an overview of the report.
- Methodology: The methodology and apparatus should be described in sufficient detail to allow readers to retrace the steps taken in the analysis. It will not be possibile to be exhaustive to achieve full repeatability, but the report should clearly state the main approach and highlight any major decisions taken in sufficient detail.
- Results: Where results consist of a single or few numbers, they should be presented as numerical results with appropriate uncertainties. Where appropriate, tables and figures should be used to illustrate the results. Systematic uncertainties should be discussed in detail, covering what effect they assess and how their size was determined.
- Conclusions: This should summarise the main results (no need to repeat all where they are too many) and draw conclusions on how they relate to the bigger picture discussed in the introduction.

6.2 Figures and tables

All figures and tables should have clear captions explaining what is shown. These don't need to repeat every detail that may be discussed in the text, but should contain sufficient information for someone who is in principle familiar with the topic to understand the figure or table without consulting the text. Figures and tables should have the same overall layout throughout the report, e.g. the same line layout for tables.

Figures should

- Have axis labels including units. For histograms, the vertical axis should have a label such as "entries per 1 MeV", where 1 MeV is the bin width.
- Not have titles as these are redundant to the axis labels and captions.
- Sufficiently big labels, thick lines and symbols, to be visible when printed in the report or projected in a presentation.
- Use high contrast colours and styles that also ensure readability for colourblind readers.
- Have a legend describing the different components where there is more than one.
- Avoid including statistics boxes where they contain unnecessary information.

6.3 Presentation

The presentation should contain a set of slides, which should

- Have a clear structure with each slide conveying one clear message. An outline slide is not required for a short presentation.
- Cover all aspects of the experiment, i.e. include all parts from introduction to conclusions, just like the report.
- Not be overloaded with text. The text on slides should just be keywords to make things understandable, but the audience should listen to what you say and not find all of that spelled out on the slide.
- Have slide numbers. These are an important aid for someone wanting to ask a question later.

The delivery of the presentation should end on the conclusions slide, which should contain the main results and stay on as a reminder to the audience of what the main outcomes were.

Acknowledgements

We are grateful to Vava Gligorov for his advice and preparation of the initial ntuples. We wish to thank Irina Nasteva for her guidance on the stripping line and offline selection cuts. The experiment was initially run in a `C++` version and we thank José Gutiérrez for his help in preparing this. Many thanks go to Benjamin Rise and Nathan Doorenbos for their contributions to the development of the `Python` version and to Fedor Bezrukov for his suggestions and improvements when demonstrating the experiment.

Appendices

A Data Variables

The variables available for each event in the `.root` data files are described below. The data is stored in a `TTree` structure. The $B^{\pm}$ candiate is reconstructed from three charged tracks which are labelled H{1,2,3} in the following. Not all variables are filled for the simulated data. Not all variables are required for the project, but additional variables have been included as they may be useful for extensions of the project.

Variable	**Description**
`H1_PX`	Reconstructed momentum component of particle in X direction [MeV/c
`H1_PY`	Reconstructed momentum component of particle in Y direction [MeV/c
`H1_PZ`	Reconstructed momentum component of particle in Z direction [MeV/c
	The momentum is reconstructed from the curvature of the path
	of the track in the magnetic field.
`H1_ProbK`	Likehood of particle being Kaon [range 0 to 1].
`H1_ProbPi`	Likehood of particle being Pion [range 0 to 1].
	The particle likelihood is obtained from combining information from
	the RICH and tracking detectors.
`H1_Charge`	Particle charge (+1 or -1)
	Obtained from direction of curvature of path of the track
	in the magnetic field.
`H1_isMuon`	Identification of track as a muon, obtained from muon chamber hits
	(0 is false, 1 is true)
`B_FlightDistance`	The distance travelled by the B candidate before decaying. Obtained fr
	the distance from the primary vertex to the
	vertex made by three charged tracks [mm].
`B_VertexChi2`	χ^2 of the quality of the vertex made by the three charged tracks.
`H1_IPChi2`	Impact Parameter χ^2.

References

[1] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[2] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[3] A. Sakharov, *Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32.

[4] J. Christenson, J. Cronin, V. Fitch, and R. Turlay, *Evidence for the 2 π Decay of the k_2^0 Meson*, Phys. Rev. Lett. **13** (1964) 138.

[5] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[6] BaBar and Belle collaborations, A. Bevan *et al.*, *The Physics of the B Factories*, Eur. Phys. J. **C74** (2014), no. 11 3026, `arXiv:1406.6311`.

[7] LHCb collaboration, R. Aaij *et al.*, *Observation of CP Violation in Charm Decays*, Phys. Rev. Lett. **122** (2019), no. 21 211803, `arXiv:1903.0872`.

[8] B. R. Martin and G. Shaw, *Particle physics*, Wiley-Blackwell, 2008.

[9] M. Gersabeck, *Frontiers in Particle Physics 2, University of Manchester*, `http://www.hep.manchester.ac.uk/u/gersabec/FPP2_Part1_LectureNotes.pdf`, 2020.

[10] M. S. Sozzi, *Discrete symmetries and CP violation: From experiment to theory*, .

[11] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, `arXiv:1207.7214`.

[12] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30, `arXiv:1207.7235`.

[13] LHCb collaboration, *LHCb Public Page*, `http://lhcb-public.web.cern.ch/lhcb-public/`, 2014. [Online; accessed 27-May-2014].

[14] LHCb collaboration, *LHCb-UK Public Page*, `www.lhcb.ac.uk/`, 2014. [Online; accessed 27-May-2014].

[15] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation in the phase space of $B^\pm \to K^+K^-\pi^\pm$ and $B^\pm \to \pi^+\pi^-\pi^\pm$ decays*, Phys. Rev. Lett. **112** (2014) 011801, `arXiv:1310.4740`.

[16] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation in the phase space of $B^\pm \to K^\pm\pi^+\pi^-$ and $B^\pm \to K^\pm K^+K^-$ decays*, Phys. Rev. Lett. **111** (2013) 101801, `arXiv:1306.1246`.

[17] LHCb collaboration, R. Aaij *et al.*, *LHCb Detector Performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, `arXiv:1412.6352`.

[18] LHCb collaboration, R. Aaij *et al.*, *First Evidence for the Decay $B_s^0 \to \mu^+\mu^-$*, Phys. Rev. Lett. **110** (2013) 021801, `arXiv:1211.2674`.

[19] LHCb VELO Group, R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) 09007, arXiv:1405.7808.

[20] LHCb Outer Tracker group, R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) 01002, arXiv:1311.3893.

[21] LHCb collaboration, M. Adinolfi *et al.*, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. **C73** (2013) 2431, arXiv:1211.6759.

[22] LHCb collaboration, I. Machikhiliyan, *Current status and performance of the LHCb electromagnetic and hadron calorimeters*, J. Phys. Conf. Ser. **293** (2011) 012052.

[23] F. Archilli *et al.*, *Performance of the Muon Identification at LHCb*, JINST **8** (2013) P10020, arXiv:1306.0249.

[24] R. Aaij *et al.*, *The LHCb Trigger and its Performance in 2011*, JINST **8** (2013) P04022, arXiv:1211.3055.

[25] Particle Data Group, P. Zyla *et al.*, *Review of Particle Physics*, PTEP **2020** (2020), no. 8 083C01, [Online access at https://pdglive.lbl.gov].

[26] R. J. Barlow, *Statistics: A guide to the Use of Statistical Methods in the Physical Sciences*, Wiley-Blackwell, 1989.

[27] R. Dalitz, *On the analysis of tau-meson data and the nature of the tau-meson*, Phil. Mag. **44** (1953) 1068.

[28] I. Bediaga *et al.*, *On a CP anisotropy measurement in the Dalitz plot*, Phys. Rev. D **80** (2009) 096006, arXiv:0905.4233.

[29] LHCb collaboration, R. Aaij *et al.*, *Search for CP violation in $D^+ \to K^-K^+\pi^+$ decays*, Phys. Rev. **D84** (2011) 112008, arXiv:1110.3970.