Measuring Matter Antimatter Asymmetries at the Large Hadron Collider ([Open Data](#) LHCb project)

Reading simulation and real data

In order to get started and check the first code that you will be writing works correctly it is best to start by analysing simulated data rather than real data from the LHC. The real data contains not only the type of events that you wish to analyse, known as the 'signal', but also events that can fake these, known as 'background'. The real data measurements are also limited by the resolution of the detector. The simplified simulation data provided here contains only the signal events and provides the results that would be obtained for a perfect

The data contains information about 'events' that were observed in the detector. An event refers to the particles produced when an interaction took place when two proton are collided at the LHC. The data you have includes information about particles observed in the detector after each collision. If you think of the data as a table, each row of the table is the results from a different collision. The columns of the table are different quantities measured about the particles produced in the collision.

We are interested in analysing the decays of particles called B^+ or B^- mesons decaying into three other mesons called kaons (K^+ or K^-). The events you have been given are those in which this process may have occurred. The detector has been used to reconstruct tracks that may have come from the kaons. You are given the measured momenta, charge, and likelihood of the tracks being kaons. You are given information for three tracks in each event, the ones that could be the three kaons that a B^+ or B^- meson has decayed into. The following information is available about each event:

Format of Event Data

The root files you have been given contains information about many different events. For each event it contains properties that have been measured for that event that may be useful in your analysis.

Three particles have been identified in each event that are candidates for particles that were produced in the decay of the B mesons. The information for these three particles is labelled, H1_xx, H2_xx, H3_xx. Here 'xx' is the particular property of the particle. H1_PX is the X component of the momentum of particle 1. The corresponding variables for particle 2 and 3 are called H2_PX and H3_PX.

Variable	Description
H1_PX	Reconstructed momentum component of particle in X direction $[\text{MeV}/c\,]$.
H1_PY	Reconstructed momentum component of particle in Y direction $[\text{MeV}/c\,]$.
H1_PZ	Reconstructed momentum component of particle in Z direction $[\text{MeV}/c\,]$. 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 from the distance from the primary vertex to the vertex made by three charged tracks $[\text{mm}\,]$.
B_VertexChi2	χ^2 of the quality of the vertex made by the three charged tracks.
H1_IPChi2	Impact Parameter χ^2.

For the real data all variables are filled for every event. For the simulated data not all variables are filled.

Not all the variables listed in the table are required for completing this project. Additional variables have been included that were important in the selection of the data sample and might be useful for extensions of the project.

For those who have studied Physics formally some these variables will be familiar but likely some of these will not be.

Variables that you will need for this analysis project are explained further here:

- H1_PX: The measured momentum component of the first particle (H1) in the x-direction. Replace X with Y and X for the other variables. You will use the momenta when computing energies and masses.

- H1_ProbK: An estimate of how likely the first particle is to be a kaon. In this analysis you are looking to analyse the decays of B mesons into three kaons. Hence you will wish to select events where the probabilities of the three particles each being a kaon is high.

- H1_ProbPi: An estimate of how likely the first particle is to be a pion. Since in your analysis the final state particles should not be pions, you will wish to select events where the probabilities of the three particles each being a pion is low.

- H1_Charge: The electric charge (in units of the electron charge) of the first particle.

- H1_isMuon: A binary value which states if the particle is likely to be a muon. As you wish your particles to be kaons you can check that this is false (0).

The additional variables that are provided are:

- B_FlightDistance: The collision point of the protons is found from the tracks that are reconstructed that come from that point. The decay vertex of the B meson is reconstructed by looking for the point at which the three final state tracks meet. The distance between the collision point, where the B meson is created, and the decay point of the B meson, where the three final state particles are created, is the flight distance of the B meson. The B meson is a short lived particle and will typically travel a few mm to a cm before decaying.

- B_VertexChi2: You can find the B meson decay point (vertex) by finding the point at which these three final state tracks meet. In practice the measured tracks will not come from an exact point as they have measurement errors, but the best estimate of the point is found. The χ^2 (Chi2 or [chisquare](https://en.wikipedia.org/wiki/Chi-squared_distribution)) is a statistical measure measure of how compatible the three tracks are from having come from a common point. A large value (greater than 1) suggests that the tracks do not really come from a common point, and hence it is more likely that they are not from a real B meson decay.

- H1_IPChi2: The impact parameter is defined as the perpendicular distance between the path of the particle and the proton-proton collision vertex. The three final state tracks come from the B meson decay point and hence should not point back to the collision vertex. Most of the tracks in an event are produced at the collision point and thus if a track does point back to the collision vertex it is an indication that it might not come from the B meson decay. The χ^2 is a measure of how compatible this track is from coming from the collision vertex taking the uncertainties into account.

Invariant mass reconstruction

Aims:

- Plot a histogram of the momentum of one of the kaon candidates
- Calculate the energy of each of the kaon candidates
- Plot the invariant masses of the B^+ or B^- mesons

Plotting a feature:

You can plot any features of the data in a histogram. Choose any suitable binning that allows you to observed the distribution of the variable clearly. Try making a histogram for the first kaon candidate's momentum x-component (H1_PX):

Momentum is a **vector** quantity, it has x,y, and z components. Try calculating the **magnitude** of the momentum of the first kaon candidate and plotting a histogram of this, you'll need the `H1_PX`, `H1_PY` and `H1_PZ` variables.

```
# calculate a variable for the magnitude of the momentum of the first kaon
# plot a histogram of this variable
```

Energy and mass

Einstein's theory of special relativity relates Energy, mass and momentum. We have measured the momentum of the kaon candidates in the detector, and have just plotted one of the components of the momentum of the kaon, and the magnitude of the momentum. The invariant mass of the kaon is well known and you can look this up. We wish to determine the energy of the kaons.

```
# calculate the energy of the first kaon
# plot a histogram of this variable
# calculate variables for the energy of the other two kaons
```

Adding features of the B^+ meson

In this analysis we are looking for B^+ or B^- mesons (see [B meson](#)) that have decayed into the three charged [kaons](#).

```
# calculate the energy of the B meson
# calculate the momentum components of the B meson
# and the magnitude of the momentum of the B meson
# calculate the B meson invariant mass
# plot the B meson invariant mass in a histogram
```

You should have a graph that sharply peaks at the mass of the B^+ meson. The mass of the B^+ and B^- meson are the same. Check that the peak of your graph is at the [known mass](#) of the B meson. **Congratulations!**

Recall that you have made this plot for simulated data. How might you expect the plots for real data to look different? In the next section you will start to work with the real LHC data.

Histogram plotting - Take care that the range of your mass plot is set suitably that you can see the mass peak. Once you have found the peak you can set the range appropriately. You do not have to start your graph at a mass of 0.

Units - The data you are provided has energies in 'MeV' (10^6 electron volts). The mass of the B meson is often quoted in 'GeV/c^2' (10^9 electron volts).

Working with real data and applying cuts

Aims:

- Filter out data that is not from the $B^+ \rightarrow K^+K^+K^-$ channel, or the antiparticle equivalent $B^- \rightarrow K^+K^-K^-$
- Plot a histogram of B-meson mass for the real data and observe how different cuts affect the data

In the section above you have analysed the simulation data to determine the invariant mass of the B meson. Now, you can start applying the methods you have used to the real LHCb data. This data was collected by the LHCb detector at CERN during 2011, the first major year of LHC operations.

The data you are given has been filtered to select only events that are likely to have come from B^+ or B^- mesons decaying into three final state charged particles. You are interested in the case where these three final state particles are charged kaons K^+ or K^-.

Preselection

You want to apply a preselection to the three final state tracks that

- Ensures that they are not muons (i.e. `!H1_isMuon` where `!` means `not`, and similarly for `H2` and `H3`)
- Requires that they each have a low probability of being pions (e.g. `H1_ProbPi < 0.5`)
- Requires that they each have a high probability of being a kaon (e.g. `H1_ProbK > 0.5`)

You need to find a balance between making cuts that are too loose and include too many background events and too tight and reject many of your signal events.

Make histograms of the probability of a final state particle being a kaon or a pion. These will help guide you on suitable probability values at which to cut.

You can also consider more sophisticated options like 2-D plots of kaon and pion probabilities or different values of the cuts for the different final state particles.

```
# plot the probability that a final state particle is a kaon
# plot the probability that a final state particle is a pion
```

Now calculate the invariant mass of the B meson for the real data and plot a histogram of this. Compare it with the one you drew for the simulation data.

Can you explain the differences you observe?

```
# draw a histogram for the B meson mass in the real data
```

Experiment with the cuts and see the impact of harsher or more lenient cuts on the invariant mass plot. You should select a set of cuts which makes the signal most prominent with respect to the background. Once you have finalised the selection on particle identification also make cuts on the reconstructed particle mass to select the events in the B meson mass peak, removing the background events which lie at lower and higher invariant masses.

Preselection example hint

This is an example string, showing the syntax, that you could use as a preselection starting point:

```
preselection = "H1_ProbPi < 0.5 & H2_ProbPi < 0.5 & H3_ProbPi < 0.5 & H1_ProbK
> 0.5 & H2_ProbK > 0.5 & H3_ProbK > 0.5 & !H1_isMuon & !H2_isMuon & !H3_isMuon"
```

Searching for global matter anti-matter differences

You will start by reconstructing the momentum, energy and mass of the B meson using the measured momenta of the kaons. You will then look for differences in the rates at which the B$^+$ decay process and the B$^-$ decay process happen: this would be a difference in the behavior of matter and anti-matter (CP violation). You will observe CP violation yourself !.

Aims:

- Calculate the global CP asymmetry
- Work out the statistical uncertainty
- Determine if there is evidence for CP violation in this decay

In order to quantify the matter antimatter asymmetry in this process we wish to compare the B$^+$ and the B$^-$ particles. The B$^-$ is the anti-particle of the B$^+$.

How can you distinguish between events that contain B$^+$ and B$^-$ particles using `H1_Charge`, `H2_Charge` and `H3_Charge`?

`# make a variable for the charge of the B mesons`

Now count the numbers of events of each of the two types (N$^+$ and N$^-$). Also calculate the difference between these two numbers.

`# make variables for the numbers of positive and negative B mesons`

In order to calculate the asymmetry and the statistical uncertainty of the asymmetry, you can make use of the formula:

$$A = \frac{(N^- - N^+)}{(N^- + N^+)} \qquad \sigma_A = \sqrt{\frac{1 - A^2}{N^- + N^+}}$$

```
# calculate the value of the asymmetry, by using the formula above, and
then print it
# calculate the statistical significance of your result and print it
```

Congratulations! You have performed your first search for a matter anti-matter difference.

Here you have only considered the statistical uncertainty. Your measurement will also have other sources of uncertainty known as systematic uncertainties which you have not considered at this stage.
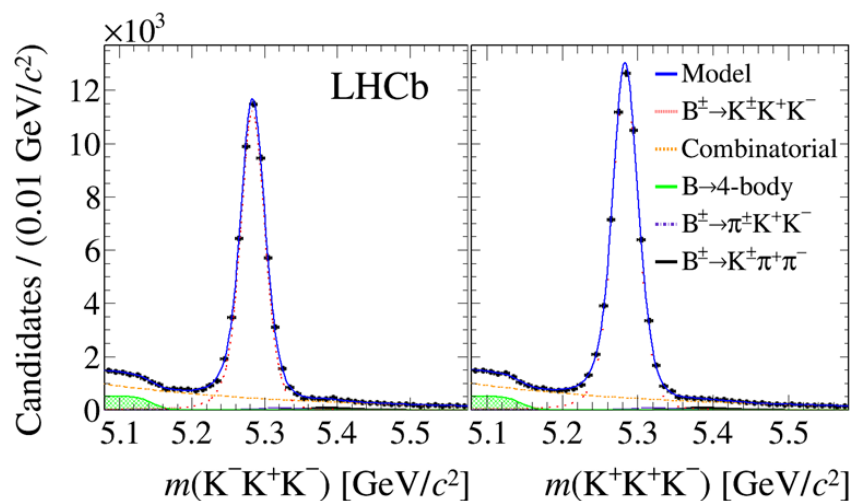
Opis warsztatu

Analizę można wykonać serwerze **lhcb1**, który znajduje się naszej Katedrze i jest widoczny poprzez sieć wydziałową, np. taurus.

Dla każdego zespołu o numerze **XX** utworzone zostały konta, przygotowany plik z danymi i przykładowy skrypt.

Zakładając, że numer zespołu to *01*, a Państwa konto na taurusie to **user,** logowanie na serwer przebiega w następujący sposób:

> **user@taurus: ssh -XY hep-lab01@lhcb1**
>
> **hep-lab01@lhcb1's password: xxxxx**

Proszę zmienić hasło do swojego konta (i zapamiętać nowe).

Do swojego **home** katalogu proszę przekopiować i wykonać plik konfiguracyjny:

> *[hep-lab01@lhcb1 ~]$ cp /datac/setenv .*
>
> *[hep-lab01@lhcb1 ~]$ source setenv*

Proszę przekopiować przykładowy program do analizy:

[hep-lab01@lhcb1 ~]$ cp /datac/storage/hep_labo/OPENDATA/Bmeson_1.cpp .

Pliki z danymi do analizy (wstępnie wybrane rzeczywiste przypadki zebrane przez LHCb i dane symulacyjne), znajdują się w repozytorium: */datac/storage/hep_labo/OPENDATA/*.

```
[hep-labMaster@lhcb1$ ls -l /datac/storage/hep_labo/OPENDATA/

-rw-r--r--. 1 hep-labMaster lab   666484974 Nov 21 15:15 B2HHH_MagnetDown.root
-rw-r--r--. 1 hep-labMaster lab   444723234 Nov 23 10:30 B2HHH_MagnetUp.root
-rw-r--r--. 1 hep-labMaster lab        3444 Nov 21   15:15 Bmeson.cpp
-rw-r--r--. 1 hep-labMaster lab     2272072 Nov 21 15:27PhaseSpaceSimulation.root
```

> ***Proszę nie kopiować pliku z danymi do swojego katalogu roboczego! Można z niego korzystać podając ścieżkę dostępu w programie do analizy!***

Projekt - ocenie podlega:

- sposób przeprowadzenia selekcji: próby różnych kryteriów, różne kombinacje ładunków, szukanie przypadków z nieprawidłową identyfikacją.
- sposób napisania programu: prosty skrypt, napisanie funkcji, klasy do selekcji, użycie TLoretnzVector, etc.
- sposób opisu: wstęp teoretyczny (krótki, ale sensowny), poprawność wykonania rysunków, porównanie z wartościami tablicowymi, wnioski.

ZACZYNAMY!

1. Zazwyczaj przy pierwszym kontakcie z plikiem z danymi, chcemy po prostu zobaczyć „co jest w środku". W tym celu uruchamiamy ROOT z plikiem z danymi.

   ```
   [hep-lab01@lhcb1 ~]$ > root /datac/storage/hep_labo/OPENDATA/B2HHH_MagnetUp.root

   root [1] TBrowser t        okno z podglądem danych
   root [4] .q                tak kończymy pracę w roocie
   ```
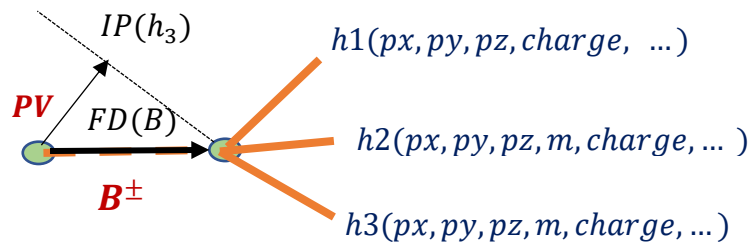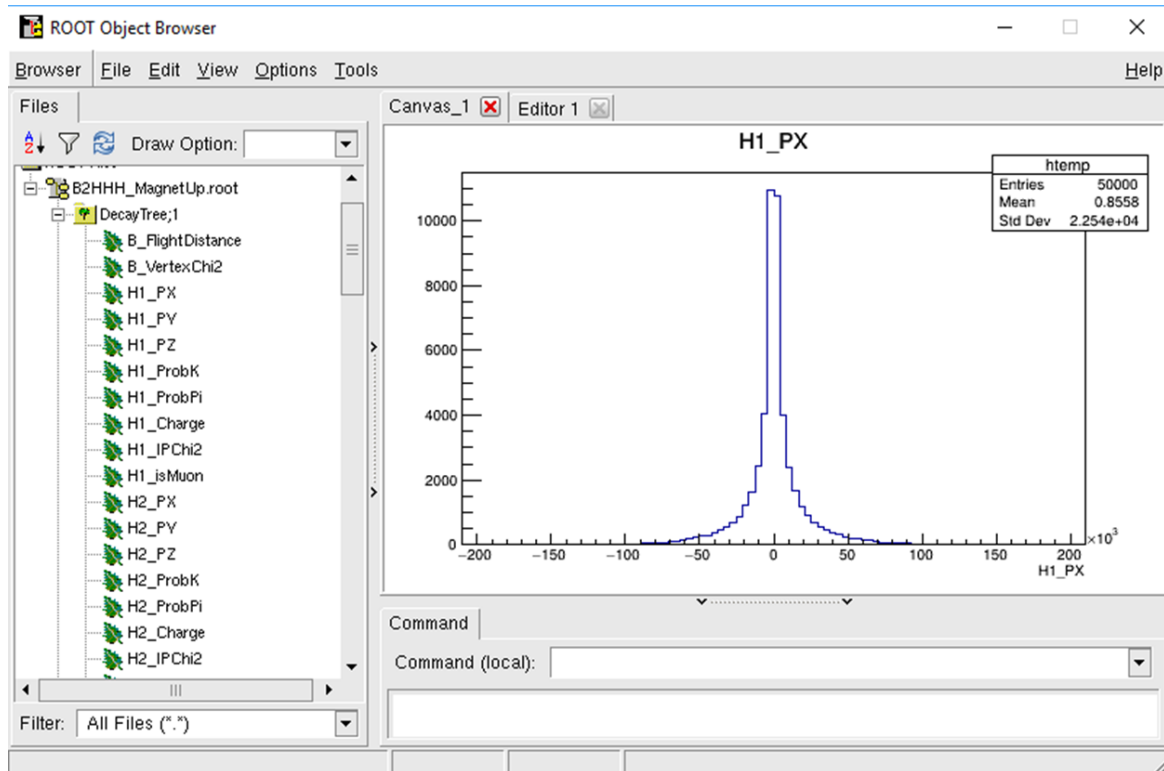

2. Analizę danych wykonujemy poprzez napisanie odpowiedniego programu. Przykładowy program z najprostszą analizą wykonamy (w roocie) poprzez:

   ```
   root [2] .x Bmeson.cpp
   ```

```
void Bmeson() {
    gROOT->Reset();
    TChain* Bchain = new TChain("DecayTree");
    Bchain->Add("B2HHH_MagnetDown.root");
//    TChain* Bchain = new TChain("PhaseSpaceTree");
//    Bchain->Add("PhaseSpaceSimulation.root");

    TTree* BDK_tree = Bchain;

    Double_t    hh12_px, hh12_py, hh12_pz, hh12_m;
    Double_t    h1_e, h2_e, h3_e;
    Double_t    hh13_px, hh13_py, hh13_pz, hh13_m;
    Double_t    hhh_px, hhh_py, hhh_pz, mass_3ka;
    Double_t    h1_px, h1_py, h1_pz,  h1_probK, h1_probPi, h1_isMuon,
h1_char, h1_IPChi2;
        (…)
    BDK_tree->SetBranchAddress("H1_PX", &h1_px);
    BDK_tree->SetBranchAddress("H1_PY", &h1_py);
    BDK_tree->SetBranchAddress("H1_PZ", &h1_pz);
        (…)

    Int_t NBINS=100;
    Double_t GeV=0.001;
    Double_t min_mass1 = 1.;
    Double_t max_mass1 = 2.5;
    Double_t min_mass2 = 1.;;
    Double_t max_mass2 = 2.5;
    Double_t min_Bmass = 5.;
    Double_t max_Bmass = 6;
    Double_t m_ka = 493.7;

   TH1D* HH12_m= new TH1D( "HH12_mass", "HH12_mass ",NBINS, min_mass1, max_mass1);
   TH1D* HH13_m= new TH1D( "HH13_mass", "HH13_mass ",NBINS, min_mass2, max_mass2);
   TH1D* mass_3ka_hist= new TH1D("mass_3ka", "mass_3ka",NBINS, min_Bmass, max_Bmass);

  Int_t Evt_tot = Bchain->GetEntries();
//    Int_t Evt_tot=10000;

    Double_t mass_h12, mass_h13;

   for(Int_t event = 0; event < Evt_tot; ++event){
      Bchain->GetEvent(event);

            hh12_px = h1_px+h2_px;
             (…)

            h1_e = sqrt(h1_px*h1_px+h1_py*h1_py+h1_pz*h1_pz+ m_ka*m_ka);

            mass_3ka = (h1_e+h2_e+h3_e)*(h1_e+h2_e+h3_e)-(hhh_px*hh12_px +
hhh_py*hhh_py + hhh_pz*hhh_pz);
            mass_3ka = sqrt(mass_3ka);
            mass_3ka_hist->Fill(mass_3ka*GeV);


    }
    TCanvas* mass_B_can=new TCanvas("mass_B_can","mass_B_can",0,0,800,600);
   mass_B_can->cd(1);
   mass_3ka_hist->Draw();


}
```

Wczytywanie pliku z danymi i jego struktury („drzewa")

Deklaracja potrzebnych zmiennych

Deklaracja histogramów

Właściwa analiza. Tutaj dla każdego przypadku można dokonać obliczeń, nałożyć kryteria, a na końcu wczytać wartość do histogramu.

Deklaracja pola na rysunek I wykonanie rysunku histogramu