UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Use of jet-charge algorithms in the study of $B_s \to J/\psi\, \phi$ decay at CMS experiment

Relatore

Prof. Paolo Ronchese

Laureando

Jakub Skowronski

Anno Accademico 2017/2018

Contents

Chapter 1

Introduction

Physics of B-mesons in the TeV scale offers an intriguing opportunity to search for New Physics and do precision tests of the Standard Model. The weak decays of mesons can violate charge-conjugation parity (CP) symmetry stating that physics should remain the same if a particle is interchanged with its antiparticle and the mirror inversion is taken of the coordinate system. In particular, B_s^0 meson provides a rich source of opportunities to probe the consistency of the Standard Model. The measurement of the weak phase ϕ_s, the observable connected to CP violation in the B_s^0 - $\bar{B}_s^0$ systems, and the decay width difference $\Delta\Gamma_s$ between the light and heavy B_s^0 mass eigenstates are among the most interesting ones [5]. The measurement of those quantities requires the knowledge of the B_s^0 flavor at production time, but, due to the phenomenon of oscillations, the flavor of B_s^0 can change before the meson decay; moreover the flavor can be impossible to reconstruct from the final state in decays as $B_s^0 \to J/\psi\phi$.

This thesis presents the development and optimization of an algorithm used to determine, with the CMS experiment, the flavour of neutral B_s^0 mesons at production time. The same work has been done, using data collected in 2012 at a collision energy $\sqrt{s} = 8$TeV. We try now to apply the same algorithm to data collected in 2016 and 2017 at a collision energy $\sqrt{s} = 13$TeV. The flavour tagging algorithm developed exploits jets induced by b mesons produced in pp collisions at the LHC. Due to *beauty* flavor conservation in the hard scattering process, $b-\bar{b}$ pairs are created and two different hadronization processes are started. There still exist processes that create only one b (or $\bar{b}$) quark, but their cross-section are much smaller, hence they are ignored. The two b-hadrons are called "Same-Side" and "Opposite-Side", and they are represented with the symbols B_{ss} and B_{os}. The B_s^0-meson which the flavor is to be inferred is associated to the Same Side. The jet-charge algorithm is used to estimate the charge of the B_{os}, as explained in Chapter 3, and thus, considering charge and flavor conservation, to infer the flavour of B_{ss}. Four simulated samples (two from 2016, two from 2017) samples of $B_s^0 \to J/\psi\phi$ and $B^\pm \to J/\psi K^\pm$ decays are exploited to develop and test the algorithm.

The tagging performances are further measured and calibrated on two samples of self-

tagging $B^+ \to J/\psi\ K+$ decays collected by the CMS experiment during 2016 LHC run, corresponding to 8.74 fb^{-1} luminosity, and during 2017 LHC run, corresponding to 13.45 fb^{-1} luminosity. In the reconstruction of $B^{\pm}$ decay the charge of the $B^{\pm}$ itself determines the flavor; the comparison with the estimation from the OS jet-charge allows the measurement of the misassignment probability.

Chapter 2

The CMS Experiment

The CMS experiment [1] is a typical detector for high-energy collissions based on the concept of cylindrical detection layers, nested around the beam axis. Starting from the beam interaction region, particles first enter a tracker, in which charged-particle trajectories (tracks) and origins (vertices) are reconstructed from signals (hits) in the sensitive layers. The tracker is immersed in a magnetic field that bends the trajectories and allows the electric charges and momenta of charged particles to be measured. Electrons and photons are then absorbed in an electromagnetic calorimeter (ECAL). The corresponding electromagnetic showers are detected as clusters of energy recorded in neighbouring cells, from which the energy and direction of the particles can be determined. Charged and neutral hadrons may initiate a hadronic shower in the ECAL as well, which is subsequently fully absorbed in the hadron calorimeter (HCAL). The corresponding clusters are used to estimate their energies and directions. Muons and neutrinos traverse the calorimeters with little or no interactions. While neutrinos escape undetected, muons produce hits in additional tracking layers called muon detectors, located outside the calorimeters. This simplified view is graphically summarized in Fig. 2.1, which displays a sketch of a transverse slice of the CMS detector.

The CMS coordinate system is oriented such that the x axis points to the centre of the LHC ring, the y axis points vertically upward, and the z axis is in the direction of the counterclockwise proton beam, when looking at the LHC from above. The origin is centred at the nominal collision point inside the experiment. The azimuthal angle ϕ (expressed in rad) is measured from the x axis in the (x, y) plane, and the radial coordinate in this plane is denoted r. The polar angle θ is defined in the (r, z) plane with respect to the z axis and the pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. The component of the momentum transverse to the z axis is denoted p_T.

The central feature of the CMS design is a large superconducting solenoid magnet. It delivers an axial and uniform magnetic field of 3.8T over a length of 12.5m and a free-bore radius of 3.15m. This radius is large enough to accommodate the tracker and both the ECAL and HCAL, thereby minimizing the amount of material in front of the

calorimeters. This feature is an advantage for PF reconstruction, as it eliminates the energy losses before the calorimeters caused by particles showering in the coil material and facilitates the link between tracks and calorimeter clusters. At normal incidence, the bending power of 4.9T·m to the inner surface of the calorimeter system provides strong separation between charged and neutral particle energy deposits.

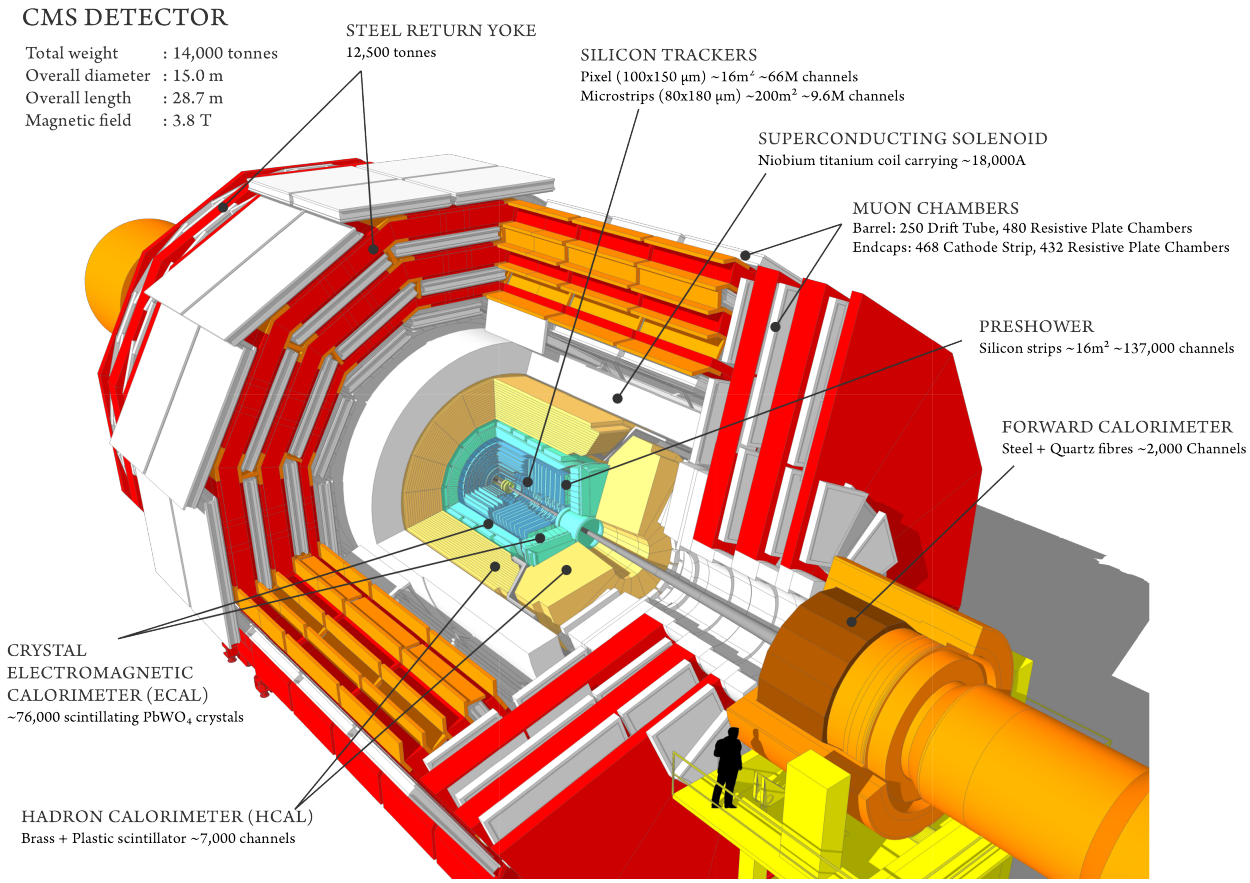

Figure 2.1: The CMS Detector

2.1 The CMS Tracker

The CMS tracker [3] is an all-silicon detector with a sensitive area of over 200m^2. The sensors are arranged in concentric cylinders around the interaction region of the LHC beams and are situated in a 3.8T magnetic field. The purpose of the detector is to provide high precision measurement points in three dimensions along the curved trajectories of charged particles up to pseudorapidities $|\eta| < 2.5$. The charged particle tracks are used to reconstruct the positions of the primary interaction and secondary decay vertices. The tracker allows for rapid and precise measurements with temporal and spatial resolutions that fulfill the challenges posed by the high luminosity LHC collisions, which occur at a frequency of 40MHz. The high particle flow induces radiation damage, which also presents a challenge for the operation and data-reconstruction in the inner layers of the tracker.

The tracker is composed of two sub-detectors with independent cooling, powering, and read-out schemes. The inner sub-detector, the pixel detector, has a surface area of 1.1m^2. It is segmented into 66 million n^+ pixels of size 100μm by 150μm implanted into n-type bulk and p-type back side. The pixel detector contains 15840 read-out chips (ROC). The ROCs are arranged into modules which transmit data via 1312 read-out links. The sub-detector surrounding the pixels, the strip detector, is segmented into 9.6 million p^+ strips which are implanted into n-type bulk in the inner layers or disks and n-type back side. The detector has 10 tracking layers in the barrel region that span radii from 25cm to 110cm and along the z axis up to 120cm: 4 layers in the inner barrel (TIB) and 6 in the outer barrel (TOB). It also has 12 disks in the endcap region with radii up to 110cm and in z up to 280cm: 3 inner disks (TID) inside and 9 endcap disks (TEC) outside the TOB. In the pixel detector, the pixel coordinates are also transmitted. For the strips, all data-processing happens in off-detector electronics.

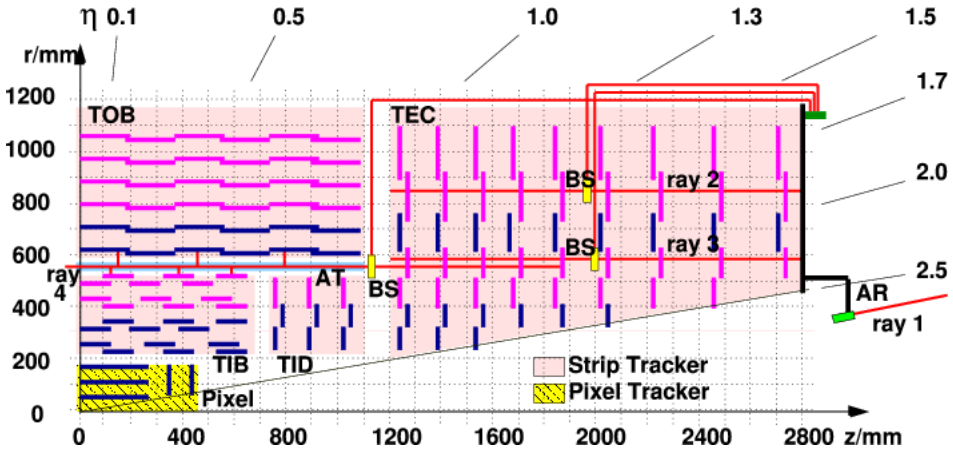

Figure 2.2: Scheme of the CMS tracker

2.2 The CMS Calorimeter

The CMS calorimeter is composed of an inner electromagnetic calorimeter (ECAL) surrounded by a hadronic calorimeter (HCAL) enclosed within a solenoid operating at 3.8T. The ECAL consists of a barrel region (EB) covering the pseudorapidity range of $1.5 < |\eta|$ and an endcap section (EE) covering $1.4 < |\eta| < 3.0$. The HCAL barrel (HB) covers the region $\eta < 1.3$ and consists of 36 azimuthal wedges assembled into two half-barrels (HB+ and HB-). The Hadronic endcap calorimeter (HE) covers the pseudo rapidity range $1.3 < |\eta| < 3$. The hadronic forward calorimeter (HF) extends the coverage $3.0 < |\eta| < 5.0$. Additional scintillators (HO) are located outside of the solenoid and act as "tail catchers" effectively increasing the thickness of the calorimeter in the central pseudo rapidity region. Figure 2.3 is an elevation view of the CMS detector showing the HCAL components with lines of constant pseudo rapidity overlayed.

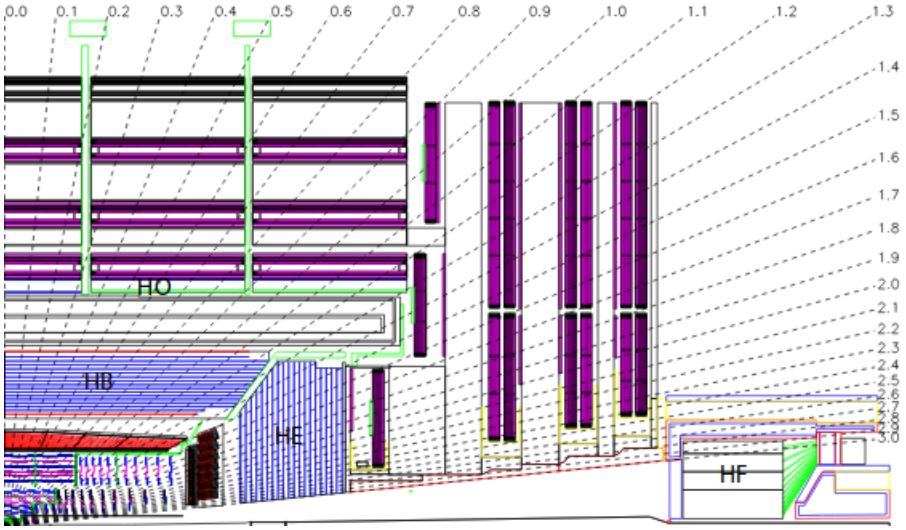

Figure 2.3: A scheme of the CMS Calorimeter

s

2.3 Trigger and data acquisition

The trigger system [2] is composed by an L1 hardware trigger and an HLT array of commercially available computers running high-level physics algorithms. The L1 trigger is a hardware system with a fixed latency. Within $4\mu s$ of a collision, the system must decide if an event should be tentatively accepted or rejected using information from the calorimeter and muon detectors.

The trigger primitives (TP) from electromagnetic and hadron calorimeters (ECAL and

HCAL) and from the muon detectors (drift tubes (DT), cathode strip chambers (CSC) and resistive-plate chambers (RPC)) are processed in several steps before the combined event information is evaluated in the global trigger (GT) and a decision is made whether to accept the event or not. For every LHC bunch crossing, the GT decides to reject or accept a physics event for subsequent evaluation by the HLT. This decision is based on trigger objects from the L1 muon and calorimeter systems, which contain information about transverse energy E_T or transverse momentum p_T, location (pseudorapidity and azimuthal angle), and quality. Similarly, special trigger signals delivered by various subsystems are also used to either trigger or veto the trigger decision in a standalone way ("technical triggers") or to be combined with other trigger signals into logical expressions ("external conditions"). These technical triggers (up to 64) are also used for monitoring and calibration purposes of the various CMS sub-detectors including L1 trigger system itself.

Tracking and vertex finding is very important for reconstruction at the HLT. A robust and efficient tracking algorithm can help the reconstruction of particles in many ways, such as improving the momentum resolution of muons, tracking-based isolation, and b-jet tagging. Since track reconstruction is a CPU-intensive task, many strategies have been developed to balance the need for tracks with the increase in CPU time.

2.4 CMSSW and Event Data Model

The overall collection of software, referred to as CMSSW [7], is built around a Framework, an Event Data Model (EDM), and Services needed by the simulation, calibration and alignment, and reconstruction modules that process event data so that physicists can perform analysis. The primary goal of the Framework and EDM is to facilitate the development and deployment of reconstruction and analysis software.

The CMSSW event processing model consists of one executable, called *cmsRun*, and many plug-in modules which are managed by the Framework. All the code needed in the event processing (calibration, reconstruction algorithms, etc.) is contained in the modules. The same executable is used for both detector and Monte Carlo data.

The CMS Event Data Model (EDM) is centered around the concept of an *Event*. An *Event* is a C++ object container for all RAW and reconstructed data related to a particular collision. During processing, data are passed from one module to the next via the *Event*, and are accessed only through the *Event*. All objects in the *Event* may be individually or collectively stored in ROOT [8] files, and are thus directly browsable in ROOT. This allows tests to be run on individual modules in isolation.

Chapter 3

Neutral B Flavour Tagging

The aim of flavour tagging algorithms is to determine whether, at production time, a physical state is produced as a particle or as an antiparticle. While this issue is trivial for self-tagging charged particles like B^+, it is far more complicated for neutral states such as B^0 which undergo flavour mixing and can decay to nontaggable final states. Flavour tagging algorithms are developed to infer the flavour of neutral B-mesons at production time by associating to each B^0 a tag decision (d) usually defined as $d = +1$ for B^0 and $d = -1$ for $\bar{B}^0$. If the tagger is not able to provide a decision, the latter is set to $d = 0$. Tagging algorithms are characterized by their effective tagging efficiency, or tagging power, described by the following relation:

$$P_{tag} = \varepsilon D^2 = \varepsilon(1 - 2\omega)^2 \tag{3.1}$$

where ε is defined as the tag efficiency, ω is the mistag (or wrong-tag) fraction, and D as the dilution.

The tagging efficiency of a tagger represents the fraction of events to which the algorithm is able to assign a tag decision, i.e. the fraction of tagged events. Given U the number of events for which the tagging algorithm is not able to determine the initial flavour, and defined R and W the number of events for which the decision taken by the algorithm is right and wrong respectively, the tagging efficiency can be represented by the following ratio:

$$\varepsilon = \frac{R + W}{R + W + U} \tag{3.2}$$

The mistag is the fraction of tagged events with an incorrectly assigned tag decision, and it can be represented as the probability of having assigned a wrong tag decision to the event. The mistag can be calculated as follows:

$$\omega = \frac{W}{R+W} \tag{3.3}$$

The dilution term D is simply related to the mistag, and defined as $D = 1 - 2\omega$.

Two main families of tagging algorithms can be defined depending on the tagging strategy used in order to infer the flavour of the initial state for a B-meson: Opposite Side (OS) and Same Side (SS) taggers. A Same Side tagger can be used to infer the flavour of the B_{ss} at production time by studying the sign of the charged particles produced nearby. Also Opposite Side taggers exploit nearby particles to infer the flavour of the B_{os}. In fact, when the $b - \bar{b}$ pair is created from the pp collision, it usually hadronizes creating B-hadrons, as visualized in Fig. 3.1. They create two different jets which tend to have a charge one slighlty negative, the other slightly positive, as the generators of these jet were charged positively ($\bar{b}$) or negatively (b). Thus if the B_{ss} is a B^+ or B_s^0, so containing a $\bar{b}$, we expect the B_{os} to be a B-hadron containging a b. Likewise, if the B_{ss} is a B^- or a $\bar{B}_s^0$, so containing a b, the other one is expected to contain a $\bar{b}$. This kind of charge and flavor correlation between the two B-hadrons is used in the jet-charge algorithm constructed below.

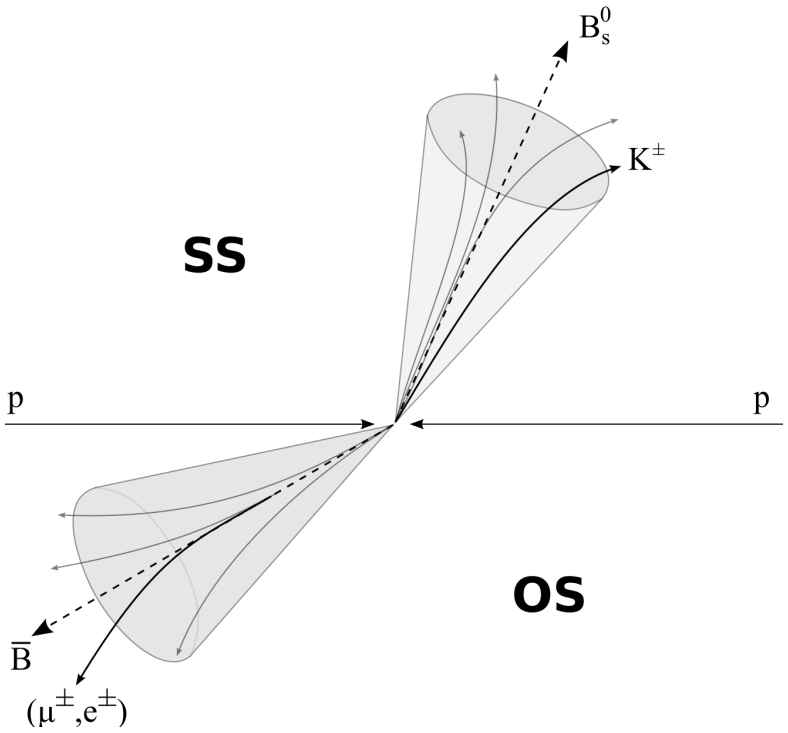

Figure 3.1: Scheme of B_s^0 and $\bar{B}$ production from pp collision

Chapter 4

Jet-Charge Algorithm

Among the possible algorithms that can be used to infer the initial flavour of a B-meson, the so called "jet-charge algorithms", exploits the mean charge of the particles in a jet opposite to the B-meson itself. It is defined as:

$$Q_{jet} = \frac{\sum_i^n q_i (p_T)_i^k}{\sum_i^n (p_T)_i^k} \tag{4.1}$$

where n is the number of particles in the jet, q_i and p_T are respectively the charge and the transverse momentum of a single particle, and k is generic weight on the momentum, as the algorithm is weight-dependant. By constrction, $-1 \leq Q_{jet} \leq +1$.

In particular, jet charge is used to distinguish jets initiated by b quarks from those initiated by $\bar{b}$ quarks, and thus to infer the flavour of a B-meson.

4.1 Data Analysis

The data analysis procedure has been tested on following data sets:

- data from 2017 LHC run.
- data from 2016 LHC run.
- two 2017 samples of simulated data containing a $B_s^0 \to J/\psi\, \phi$, the first one (5368595 events), and $B^{\pm}$, $\to J/\psi\ K^{\pm}$ channel, the second one (549776 events).
- another two samples of simulated data from 2016, for $B^{\pm}$ (1000000 events) and B_s^0 (1000000 events) respectively.

The simulated events were generated by PYTHIA8 [9], decayed by EVTGEN [10], and simulated by GEANT4 [11]. Both Monte Carlo and CMS data are stored as ROOT

objects, which are managed by dedicated software illustrated above. A C++ code implemented in the CMSSW framework is used in order to access the data, stored in form of ROOT Ntuples, analyze them and create histograms of desired quantities.

4.1.1 Same-Side

The first step is the B_{ss} selection from CMS data. $B^{\pm}$ candidates are formed by combining a reconstructed J/ψ, , reconstructed in its turn combining two opposite charge muons, and a track with $p_T > 2$GeV/c under the kaon mass assumption. The two muons and the track are fitted to a common vertex while constraining the dimuon invariant mass to the word average J/ψ mass. B_s^0 candidates are formed by combining a reconstructed J/ψ and reconstructed ϕ candidates, reconstructed in their turn combining two opposite charge tracks assumed to be kaons. The two muons and the two kaons tracks are fitted to a common vertex in the same way as for $B^{\pm}$, applying a constraint on the dimuon invariant mass. The B_s^0 and $B^{\pm}$ candidates having a reconstructed invariant mass between 5.0 and 5.5 GeV/c^2 and having a χ^2 vertex fit probability greater than 2% have been selected. All the selection criteria are summariezed in Table 4.1.

Variable	$B^{\pm} \to J/\psi K^{\pm}$	$B_s^0 \to J/\psi\phi$
P_{χ^2}	> 0.02	> 0.02
M(J/ψ)	2.80 - 3.40 GeV/c^2	2.80 - 3.40 GeV/c^2
M(ϕ)	-	1.005 - 1.035 GeV/c^2
M(B)	5.0 - 5.5 GeV/c^2	5.0 - 5.5 GeV/c^2
$p_{T\mu}$	> 2.0 GeV/c	> 2.0 Gev/c
$p_{TK^{\pm}}$	> 0.7 Gev/c	> 0.7 Gev/c

Table 4.1: B-mesons vertex reconstruction conditions

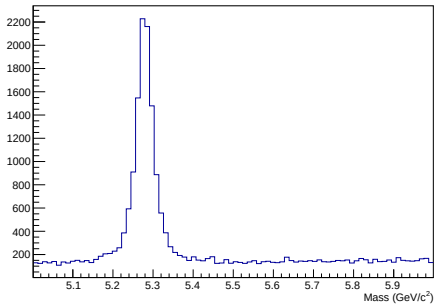

Figure 4.1: $B^{\pm}$ mass distribution from vertex reconstruction

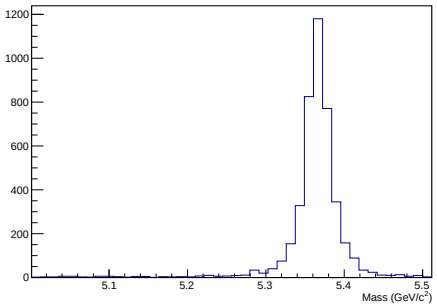

Figure 4.2: B_s^0 mass distribution from vertex reconstruction

In real data the performances of the jet-charge algorithm can be studied only in the self-tagging channel $B^{\pm} \to J/\psi K^{\pm}$, so only this one is used. In the simulated samples

instead, simulated particles can be associated to reconstructed ones on the basis of their distance in the (η, ϕ) plane:

$$\Delta R^2 = (\Delta\eta)^2 + (\Delta\phi)^2 < 2 \cdot 10^{-5} \tag{4.2}$$

where Δ represents the distance between the corrisponding momentum of simulated and a reconstructed objects. If two or more particles are found, the one which minimizes ΔR is chosen.

In this way it is then possible to reconstruct the $B^{\pm}$ or B_s^0 as described above and assign the flavour looking at the associated simulated particles, so in the simulated sample the study can be done for both decays.

Once the B_{ss}-meson has been found, the jet linked to B_{ss} is searched for, using all of the tracks coming from the selected vertex: the one that is linked to the majority of tracks is chosen. In the following the reconstructed meson and the jet it belongs to will be labelled B_{ss}^{reco} and Jet_{ss}^{reco} respectively.

In the simulated samples, it is also possible to find the B-mesons looking at the simulated particles and searching for a $B^{\pm}$ or $B_s^0/\bar{B}_s^0$ decaying to J/ψ and a $K^{\pm}$ or ϕ respectively. The particles coming from the decay are then linked to reconstructed ones using the criterion of minimal ΔR distance (formula 4.2).

Finally, the jet that is linked to the majority of previously found particles is chosen as the one corresponding to the B-meson decay. In the following, this simulated meson and the jet it belongs to will be labelled as B_{ss}^{sim} and Jet_{ss}^{sim}.

4.1.2 Other-Side

If the B_{ss} is found, the B_{os} jet is searched for, if not, the event is ignored. Jets fulfilling a set of quality criteria, summarized in Tab. 4.3, are searched for, excluding the Same-Side jet; the jet having the highest probability of being originated from a b quark is then selected. The probability is estimated on the basis of an overall compatibilty of the particles in the jet to come from a single production point or from the decay of a flying particle, as a B-hadron, forming a "secondary vertex". This is summarized in an estimator called "Combined Secondary Vertex" probability CSV_{jet}. The CSV distributions for b-jets and for jets not originated from b are shown in Fig. 4.3 and 4.4. In the following the jet selected according to this criteria will be labelled Jet_{os}^{reco}.

In simulated data, an Opposite Side jet can be looked for also by searching for another B-hadron, taking care that it is not in the same decay chain of the B_{ss}^{sim}. This second B-hadron will be called B_{os}^{sim} in the following. Likewise in the B_{ss}^{sim} case, daughter particles are linked to reconstructed ones as described in par. 4.1.1.. Then the jet that is linked to the majority of previously found particles is chosen as the one corresponding

$\text{Jet}_{os} \neq \text{Jet}_{ss}$
$\text{p}_T > 20$ GeV
$\lvert\eta\rvert < 2.7$
$\text{CSV}_{jet} > 0.8$

Table 4.2: Jet_{os}^{sel} criteria

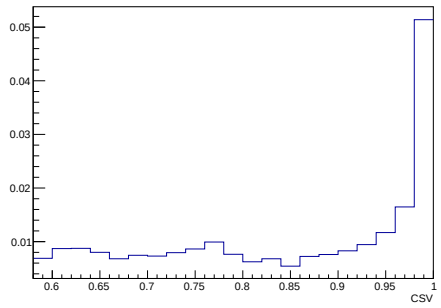

Figure 4.3: *CSV* distribution for *b*-jets

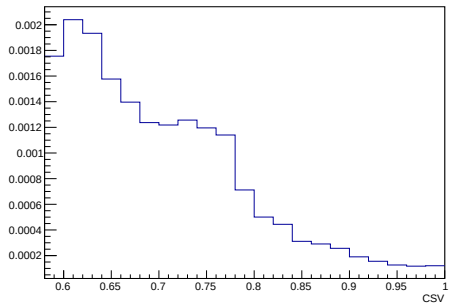

Figure 4.4: *CSV* distribution for other jets

to the B-hadron decay. In the following the jet selected in this way will be labelled Jet_{os}^{sim}.

4.1.3 Jet-Charge

Jet-charges for the found B_{os} jets are calculated using the definition at 4.1 using different k-weights with $0 \leq k \leq 2$. In the calculation of Q_{jet}, particles in the jet are required to pass the following selection:

- $|\eta| < 2.5$
- $\text{p}_T > 0.2$
- $q \neq 0$
- $|d\bar{z}_{B_{ss}} - dz_i| < 4\sigma_{d\bar{z}_{B_{ss}}}$, where dz is the impact parameter in z.

The charge of the B_{ss}-meson can be then compared with the one inferred from the Jet_{os}, as described in the following, to test the tagging power of the algorithm. In doing this, B_s^0 mesons are associated to B^+ and $\bar{B}_s^0$ to B^-. The B_{ss}^0 flavor can not be determined from the reconstructed decay products, and the jet-charge algorithm performances , in case of real data, can be studied only for the $B_{ss}^{\pm}$-mesons, while both $B^{\pm}$ and $B_s^0/\bar{B}_s^0$ can be used in the simulation. The results were not computed for the $B_{ss}^{\pm,\ reco} - B_{os}^{reco}$ pair and for the $B_{s,\ ss}^{0,\ reco} - B_{os}^{reco}$ in the 2017 simulated samples as the number of reconstructed B_{ss}-mesons is very small and not sufficient to obtain significant results.

As the algorithm depends on k-weight, the first step is to select the one that maximizes the algorithm discriminating capacity. In order to evaluate that capacity, the distributions of Q_{jet} for positive B_{ss} (P(Q_{jet})) and negative B_{ss} (N(Q_{jet})) are built, both normalized to 1. The Right Assignment Rate ($f_{R(x)}$) and Wrong Assignment Rate ($f_{W(x)}$) are then defined as the integrals between -1 and x of P(Q_{jet}) and N(Q_{jet}) respectively, with $-1 < x < +1$. The integral of the plot of $f_{R(x)}$ against $1 - f_{W(x)}$, called ROC (Receiver Operating Characteristic), gives an estimation of the discriminating capability of the variable Q_{jet}: values near 1 corresponds to a very good discrimination, while a value at 0.5 corresponds to a random assignment or no discrimination. Values of ROC integral are computed and illustrated in the Appendix A. In the following, the k-weight value that maximize ROC integral for each sample is used. Two jet-charge plots, blue for a positive B_{ss}, red for a negative one, are shown, for each every $B_{ss} - B_{os}$ pair in Fig. 4.5 - 4.16. While in simulated data the Q_{jet} distributions for positive and negative B_{ss} are separated, although partly overlapped, the distributions for data are almost completely overlapped, leading to a very small discriminating capability.

The tagging efficency, ε_{tag}, and the mistag fraction, ω_{os}, are then computed:

$$\varepsilon_{tag} = \frac{N_{sel}}{N_{B_{ss}}} \tag{4.3}$$

$$\omega_{os} = \frac{N_{wrong}}{N_{wrong} + N_{right}} \tag{4.4}$$

where N_{sel} is the number of times a jet is selected for B_{os}, and $N_{B_{ss}}$ is the number of found B_{ss}. N_{right} counts how many times for a positive (negative) B_{ss} a $Q_{jet} < 0$ ($Q_{jet} > 0$) is obtained. N_{wrong} instead how many times for a positive (negative) B_{ss} a $Q_{jet} > 0$ ($Q_{jet} < 0$) is obtained.

In the end, the tagging power P_{tag} is computed for every Same-Side and Other-Side pair as defined in (4.4).

The performances of the algorithm is much less than expected, especially in real data; a tagging power much less than 10^{-3} is obtained, while a value near 10^{-2} was expected.

Particles	ε_{tag}	ω_{os}	P_{tag}
$B^{\pm}$ CMS data	0.177(1)	0.484(2)	0.000 18(7)
$B_{ss}^{\pm,\ sim} - B_{os}^{reco}$	0.0536(7)	0.438(2)	0.000 82(5)
$B_{s,\ ss}^{0,\ sim} - B_{os}^{reco}$	0.0812(9)	0.406(2)	0.0028(1)

Table 4.3: Tagging power values for 2017 data. The number with brackets is the error.

Particles	ε_{tag}	ω_{os}	P_{tag}
$B^{\pm}$ CMS data	0.0541(7)	0.489(2)	0.000 02(1)
$B_{ss}^{\pm,\ sim} - B_{os}^{reco}$	0.0346(6)	0.422(2)	0.000 84(5)
$B_{ss}^{\pm,\ reco} - B_{os}^{reco}$	0.0394(6)	0.431(2)	0.000 74(4)
$B_{s,\ ss}^{0,\ sim} - B_{os}^{reco}$	0.0404(6)	0.409(2)	0.001 33(6)
$B_{s,\ ss}^{0,\ reco} - B_{os}^{reco}$	0.0148(4)	0.403(2)	0.000 55(3)

Table 4.4: Tagging power values for 2016 data. The number with brackets is the error.

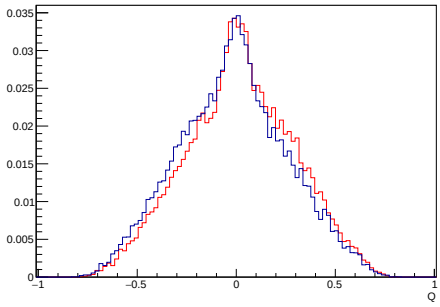

Figure 4.5: OS Jet Charge for $B^{\pm}$ from 2017 CMS data

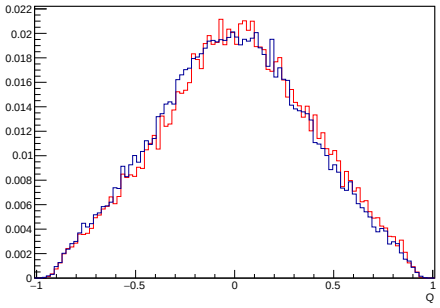

Figure 4.6: OS Jet Charge for $B^{\pm}$ from 2016 CMS data

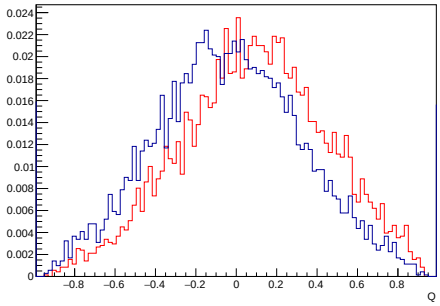

Figure 4.7: OS Jet Charge for $B_{ss}^{\pm,\ sim}$ – B_{os}^{sim} from 2017 simulated data

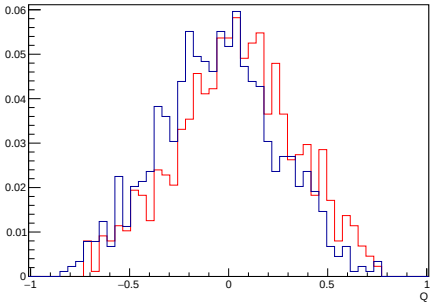

Figure 4.8: OS Jet Charge for $B_{ss}^{\pm,\ sim}$ – B_{os}^{reco} from 2017 simulated data

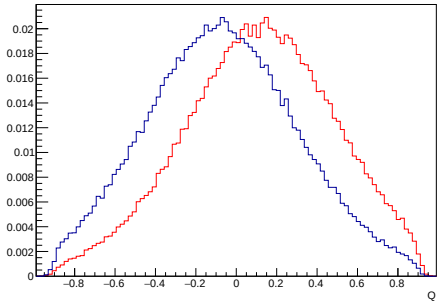

Figure 4.9: OS Jet Charge for $B_{s,\ ss}^{0,\ sim}$ – B_{os}^{sim} from 2017 simulated data

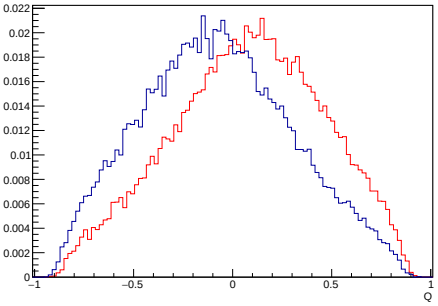

Figure 4.10: OS Jet Charge for $B_{s,\ ss}^{0,\ sim}$ – B_{os}^{reco} from 2017 simulated data

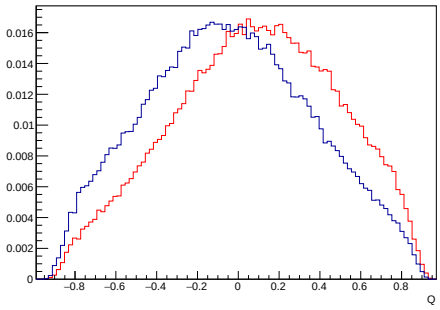

Figure 4.11: OS Jet Charge for $B_{ss}^{\pm,\ sim}$ – B_{os}^{sim} from 2016 simulated data

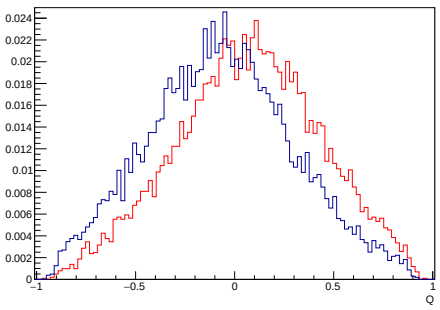

Figure 4.12: OS Jet Charge for $B_{ss}^{\pm,\ sim}$ – B_{os}^{reco} from 2016 simulated data

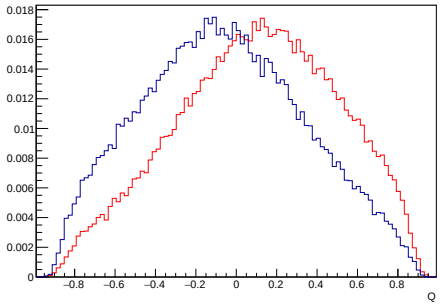

Figure 4.13: OS Jet Charge for $B_{s,\ ss}^{0,\ sim}$ – B_{os}^{sim} from 2016 simulated data

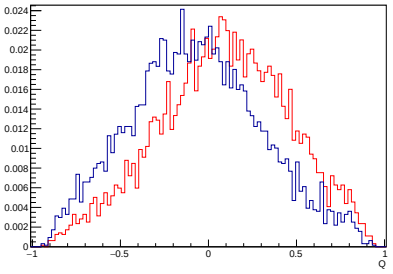

Figure 4.14: OS Jet Charge for $B_{s,\ ss}^{0,\ sim}$ – B_{os}^{reco} from 2016 simulated data

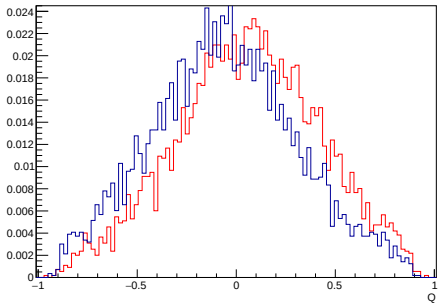

Figure 4.15: OS Jet Charge for $B_{ss}^{\pm,\ reco}$ – B_{os}^{reco} from 2016 simulated data

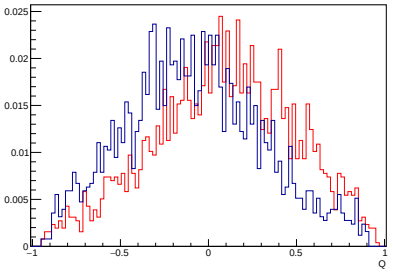

Figure 4.16: OS Jet Charge for $B_{s,\ ss}^{0,\ reco}$ – B_{os}^{reco} from 2016 simulated data

Chapter 5

Conclusions

The jet-charge algorithm for the flavor determination of Opposite-Side B-hadrons at production time has been presented. Even though it was calibrated on simulated samples with promising results, the tagging power for the CMS data both from 2016 and 2017 LHC run is much lower than expected, as illustrated in Table 4.3 and 4.4. This is probably due to a higher energy of the pp collisions than before, at $\sqrt{s} = 13$ TeV instead of 8 Tev.

The aim of this work was to implement the jet-charge algorithm through C++ code that can be implemented in CMS Software framework, but further work is necessary in order to gain significant performance improvement and to make it usable for analysis aiming to make high-precision measurements of $B_s^0 - \bar{B}_s^0$ mixing and CP asymmetries in B_s^0 decays.

Appendices

Appendix A

ROC Values

k-weight	$B^{\pm}_{\text{CMS data}}$	$B^{\pm}_{\text{sim - sim}}$	$B^{\pm}_{\text{sim - reco}}$	$B^{0}_{s,\text{ sim - sim}}$	$B^{0}_{s,\text{ sim - reco}}$
0	0.537	0.549	0.561	0.570	0.570
0.1	0.538	0.557	0.573	0.583	0.586
0.2	0.538	0.565	0.583	0.595	0.595
0.3	0.537	0.580	0.586	0.607	0.603
0.4	0.536	0.586	0.587	0.618	0.609
0.5	0.535	0.591	0.586	0.627	0.613
0.6	0.533	0.596	0.585	0.634	0.615
0.7	0.531	0.599	0.585	0.639	0.617
0.8	0.529	0.601	0.583	0.643	0.617
0.9	0.528	0.603	0.582	0.645	0.617
1.0	0.527	0.604	0.582	0.647	0.617
1.1	0.526	0.605	0.580	0.648	0.616
1.2	0.525	0.605	0.580	0.649	0.616
1.3	0.524	0.605	0.579	0.649	0.615
1.4	0.523	0.606	0.578	0.649	0.614
1.5	0.523	0.605	0.576	0.649	0.613
1.6	0.522	0.605	0.575	0.649	0.613
1.7	0.522	0.605	0.574	0.649	0.613
1.8	0.521	0.605	0.574	0.648	0.612
1.9	0.521	0.605	0.573	0.648	0.611
2.0	0.521	0.605	0.572	0.648	0.611

Table A.1: ROC values for 2017 data

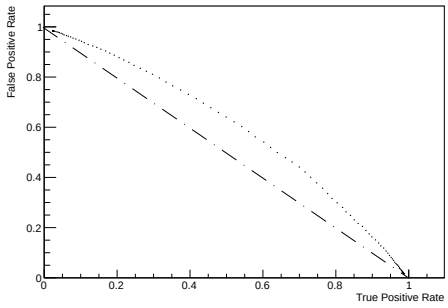

Figure A.1: $B^{\pm}_{\text{sim - sim}}$ ROC curve for 2017 data and $k = 1.1$

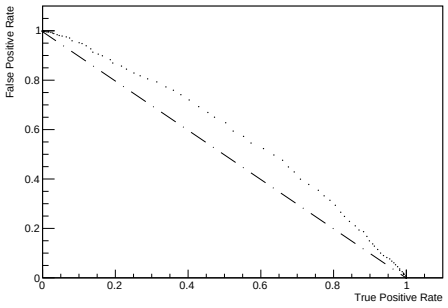

Figure A.2: $B^{\pm}_{\text{sim - reco}}$ ROC curve for 2017 data and $k = 0.7$

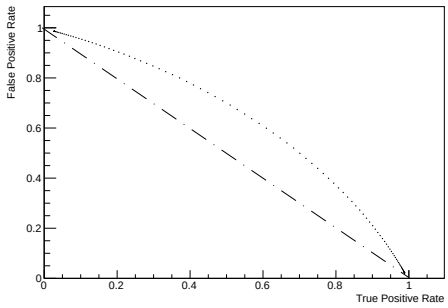

Figure A.3: $B^{0,}_{s,\text{ sim - sim}}$ ROC curve for 2017 data and $k = 1.2$

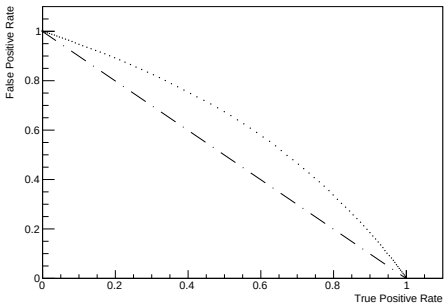

Figure A.4: $B^{0,}_{s,\text{ sim - reco}}$ ROC curve for 2017 data and $k = 0.7$

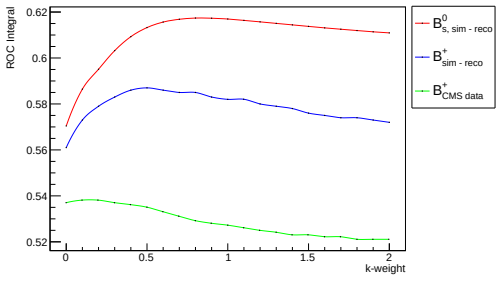

Figure A.5: k-weight dependence of ROC Integral for 2017 data

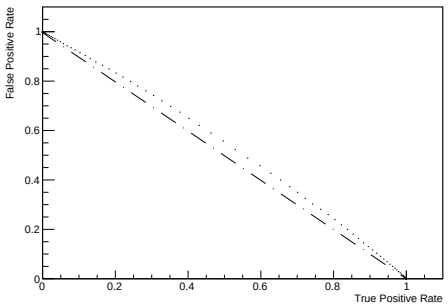

Figure A.6: $B^{\pm}_{\text{CMS data}}$ ROC curve for 2017 data and $k = 0.2$

k-weight	$B^{\pm}_{\text{CMS data}}$	$B^{\pm}_{\text{sim - sim}}$	$B^{\pm}_{\text{sim - reco}}$	$B^{0}_{s,\text{ sim - sim}}$	$B^{0}_{s,\text{ sim - reco}}$	$B^{\pm}_{\text{reco - reco}}$	$B^{0}_{s,\text{ reco - reco}}$
0	0.508	0.556	0.567	0.574	0.575	0.562	0.569
0.1	0.510	0.564	0.581	0.584	0.591	0.573	0.585
0.2	0.511	0.570	0.589	0.592	0.600	0.580	0.596
0.3	0.513	0.577	0.596	0.600	0.608	0.585	0.605
0.4	0.514	0.583	0.601	0.608	0.614	0.589	0.612
0.5	0.515	0.588	0.605	0.615	0.617	0.591	0.617
0.6	0.515	0.592	0.607	0.620	0.620	0.592	0.619
0.7	0.516	0.595	0.607	0.624	0.621	0.592	0.621
0.8	0.516	0.597	0.607	0.628	0.621	0.592	0.621
0.9	0.516	0.599	0.607	0.630	0.620	0.591	0.621
1.0	0.516	0.601	0.606	0.631	0.620	0.590	0.621
1.1	0.516	0.602	0.605	0.633	0.619	0.589	0.620
1.2	0.516	0.603	0.604	0.634	0.618	0.589	0.620
1.3	0.516	0.603	0.603	0.635	0.617	0.588	0.619
1.4	0.516	0.603	0.602	0.635	0.616	0.587	0.618
1.5	0.516	0.604	0.601	0.636	0.615	0.586	0.617
1.6	0.516	0.604	0.600	0.636	0.614	0.585	0.616
1.7	0.516	0.604	0.600	0.636	0.614	0.584	0.615
1.8	0.516	0.604	0.599	0.636	0.613	0.584	0.614
1.9	0.516	0.604	0.599	0.636	0.612	0.583	0.614
2.0	0.516	0.604	0.598	0.636	0.611	0.583	0.613

Table A.2: ROC values for 2016 data

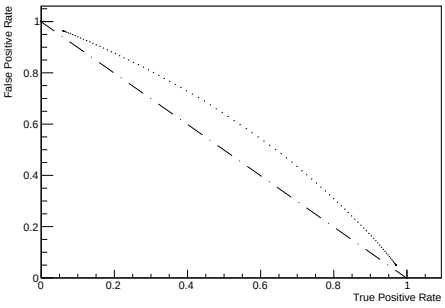

Figure A.7: $B^{\pm}_{\text{sim - sim}}$ ROC curve for 2016 data and $k = 1.2$

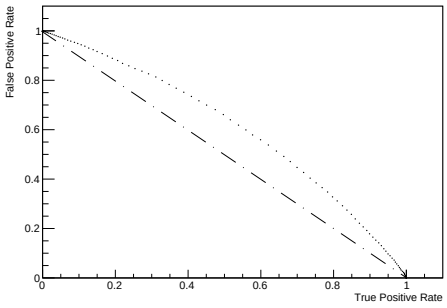

Figure A.8: $B^{\pm}_{\text{sim - reco}}$ ROC curve for 2016 data and $k = 0.7$

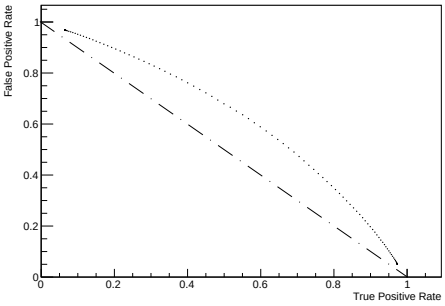

Figure A.9: $B^{0}_{s,\text{ sim - sim}}$ ROC curve for 2016 data and $k = 1.3$

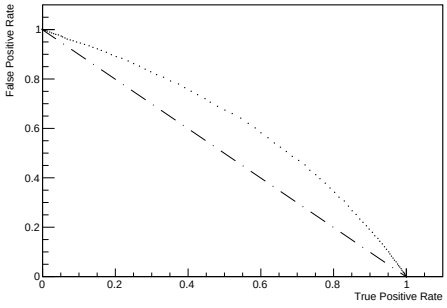

Figure A.10: $B^{0,\,reco}_{s,\text{ sim - reco}}$ ROC curve for 2016 data and $k = 0.7$

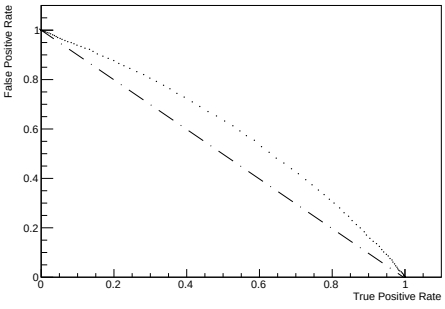

Figure A.11: $B^{\pm}_{\text{reco - reco}}$ ROC curve for 2016 data and $k = 0.7$

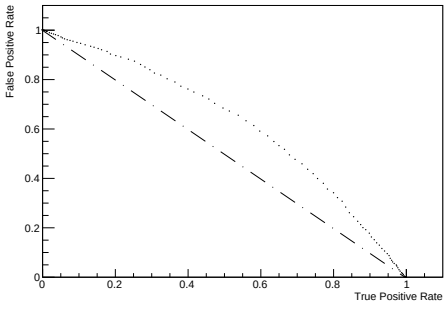

Figure A.12: $B^{0,\,reco}_{s,\text{ reco - reco}}$ ROC curve for 2016 data and $k = 0.7$

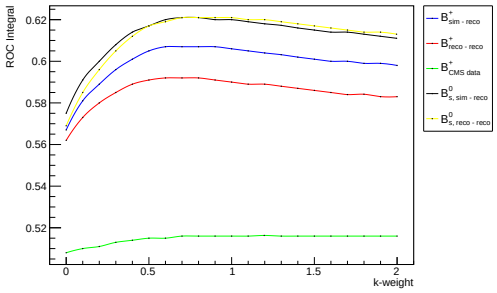

Figure A.13: k-weight dependence of ROC Integral for 2016 data

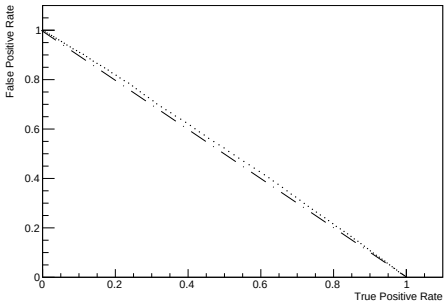

Figure A.14: $B^{\pm}_{\text{CMS data}}$ ROC curve for 2016 data and $k = 0.7$

Bibliography

[1] The CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST 03 (2008) S08004.

[2] The CMS Collaboration, *The CMS trigger system*, JINST 12 (2017) P01020, arXiv:1609.02366.

[3] The CMS Collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, JINST 09 (2014) P10009.

[4] The CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, JINST 12 (2017) P10003, arXiv:1706.04965.

[5] CMS Collaboration, *Measurement of the CP-violating weak phase ϕ_s and the decay width difference $\Delta\Gamma_s$ using the $B_s \to J/\psi\phi(1020)$ decay channel in pp collisions at $\sqrt{s}$ = 8 TeV*, PLB 757 (2016) 97, arXiv:1507.07527.

[6] C. Weiser, *A Combined Secondary Vertex Based B-Tagging Algorithm in CMS*, CMS-NOTE-2006-014.

[7] The CMS Collaboration, *The CMS Offline WorkBook* , see `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBook`.

[8] Rene Brun and Fons Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86.

[9] Torbjörn Sjöstrand, *A Brief Introduction to PYTHIA 8.1*, CompPhys Comm. 178 (2008) 852.

[10] David Lange, *The EvtGen particle decay simulation package*, NIM A 462 (2001) 152.

[11] GEANT4 Collaboration, *Geant4 — a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment; Volume 506, Issue 3, 1 July 2003, Pages 250-303.