Feasibility studies of the decay $B^0 \to K^{0*}(892)\tau^+\tau^-$ at LHCb

Alexander Dätwyler

supervisors
Prof. Nicola Serra, Dr. Rafael Coutinho

July 20, 2016

Abstract

In this thesis, sensitivity studies of the deday mode $B^0 \to K^*(892)\tau^+\tau^-$, with both taus reconstructed from $\pi^+\pi^+\pi^-\nu_\tau$, at LHCb are performed. This analysis is based on data samples corresponding to an integrated luminosity of $1\,\text{fb}^{-1}$ for background, respectively 1.5 million monte carlo simulated events for signal, which are used as reference to extrapolate for expected luminosities for Run2 and LHCb upgrade. The results of this study indicate that the first observation of this mode at LHCb in a near future is unlikely with the current setup.

Contents

1 Introduction

1.1 Standard Model

The Standard Model (SM) is the theory that describes the interaction between elementary particles. These interactions are the strong, the weak and the electromagnetic force. At high energies the weak and electromagnetic force are combined to the electroweak force. The SM also categorizes the subatomic particles. There are the following categories: fermions and bosons.

All the elementary particles can be found in Fig. 1. The fermions consist of quarks and leptons. Quarks are the up-,down-,charm-,strange-,top- and bottom-quark and their corresponding anti-quark. The top- and bottom-quark are also called truth- and beauty-quark. Leptons are the electron e^- the muon μ^-, the tau τ^-, all the corresponding neutrinos and the corresponding anti-leptons.

In the SM, apart from the mass and lifetime, charged leptons are identical. For example the transition $Z^0 \to l^+l^-$ (where $l^\pm$ is any charged leptons) is, in the current theory, the same for any lepton. This means that the chance that a Z^0 decays into a electron positron pair is the same as the chance to decay to $\mu^+\mu^-$ or $\tau^+\tau^-$. This is called Lepton Flavour Universality (LFU).

1.2 Theoretical and experimental motivations

Flavour Changing Neutral Current (FCNC) processes are among the main goals of particle physics today. These are transitions in which a quark changes its flavour without emitting a charged boson ($W^\pm$), which is only possible via loop-decay. As-yet undiscovered particles may contribute in these loop-decays with comparable amplitudes, many observables can deviate from their expected SM values. In particular, $b \to s\, l^+l^-$ transitions are highly sensitive to new degrees of freedom and this can be exploited in indirect searches for New Physics (NP) effects. [1] One example is the $B^0 \to K^{*0}l^+l^-$ mode, which dominant decay amplitudes are shown in Fig. 2. Some hints for NP effects in $b \to s\, l^+l^-$ transitions have been seen measured value of the ratios of branching fractions $R_K = \frac{B(B^+ \to K^+\mu^+\mu^-)}{B(B^+ \to K^+e^+e^-)}$. The theoretical value according to the SM should be $1 + O(10^{-3})$ [11], but the measured value deviates from that with 2.6σ [4], see Fig. 3. LHCb performed this analysis with $3\,\text{fb}^{-1}$ with a centre of mass energy of 7 and $8\,\text{TeV}$ in a
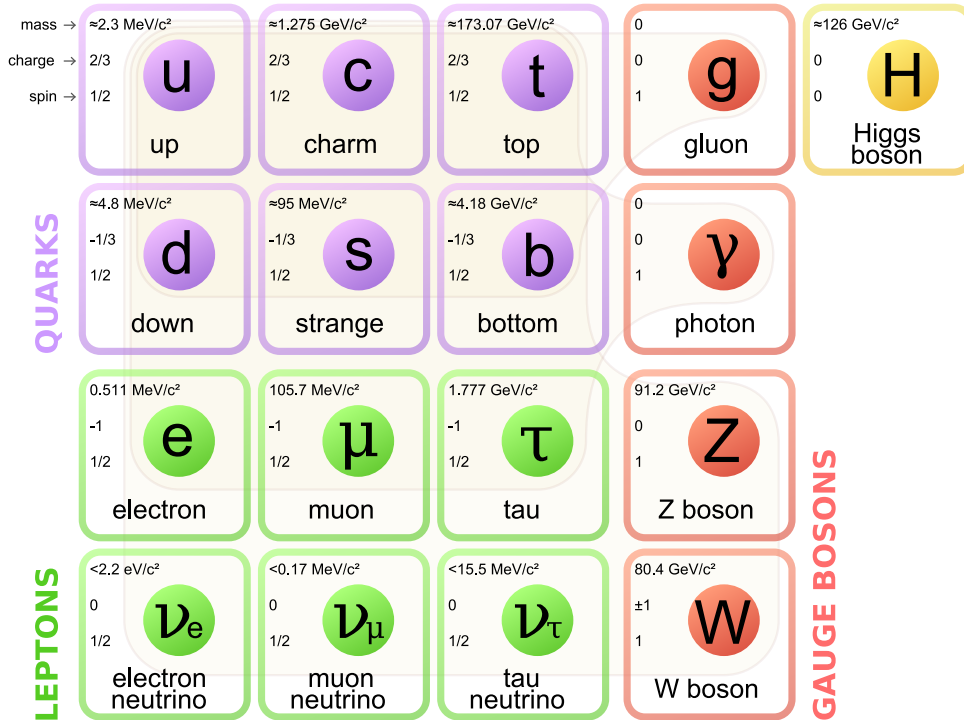

Figure 1: The standard model. [17]

---

[1]Throughout this thesis the inclusion of conjugate is implied, unless otherwise explicitly stated.
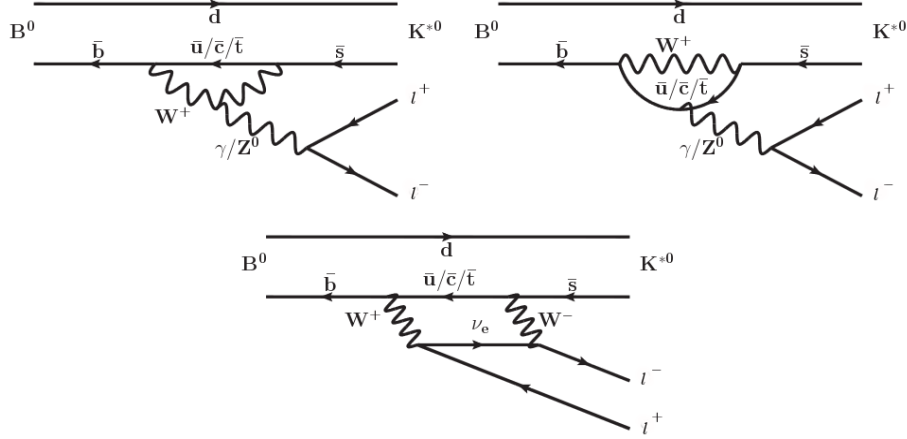
Figure 2: Three main modes for the $B^0 \to K^{*0}l^+l^-$ decays. The $l^{\pm}$ stands for any of the three leptons $e^{\pm}$, $\mu^{\pm}$ or $\tau^{\pm}$. The two Feynmann-diagrams on the top are loop decays, the one on the bottom is a box decay [3].

specific region as seen in the plot. This can be calculated by $R_K = \frac{\int_{q^2_{min}}^{q^2_{max}} \frac{dB(B^0 \to K^+\mu^+\mu^-)}{dq^2} dq^2}{\int_{q^2_{min}}^{q^2_{max}} \frac{dB(B^0 \to K^+e^+e^-)}{dq^2} dq^2}$ [4], where

q stands for the di-lepton invariant mass squared. Moreover, in the ratios $R_{D^{\pm}} = \frac{B(B^0 \to D^{\pm}\tau\bar{\nu}_{\tau}^{\pm})}{B(B^0 \to D^{\pm}\mu\bar{\nu}_{\mu}^{\pm})}$ and $R_{D^{*\pm}} = \frac{B(B^0 \to D^{*\pm}\tau\bar{\nu}_{\tau}^{\pm})}{B(B^0 \to D^{*\pm}\mu\bar{\nu}_{\mu}^{\pm})}$ the difference between calculations from SM and the measured values of $R_{D^{\pm}}$ and $R_{D^{*\pm}}$ deviates about 3.9 σ [5] Both measurements indicate the possibilities Lepton Flavour Universality (LFU) breaking.

An interesting complementary result is given by the angular analysis of the $B^0 \to K^0\mu^+\mu^-$ decay, performed by the LHCb using 3 fb^{-1}. In a given observable, called P_5', some deviation from the SM has been seen with and deviation of 3.9 σ [10], see Fig. 4 The deviation can be explained by problems in the current understanding of Quantum Chromo Dynamics (QCD), *e.g.* not properly estimating contributions from charm loops ($B^0 \to J/\psi K^{*0}$), *e.g.* a vector like contribution could have unknown contribution with a tree level Z' with mass of a few TeV. It has been suggested that this anomaly could be related to the observed hits of LFU breacking. Therefore, the extension of these studies to other final states is of great interest.

In this thesis, the feasibility to measure the decay $B^0 \to K^{0*}(892)\tau^+\tau^-$ at the LHCb is investigated using a monte carlo simulation of the decay and real data from the LHCb as background.
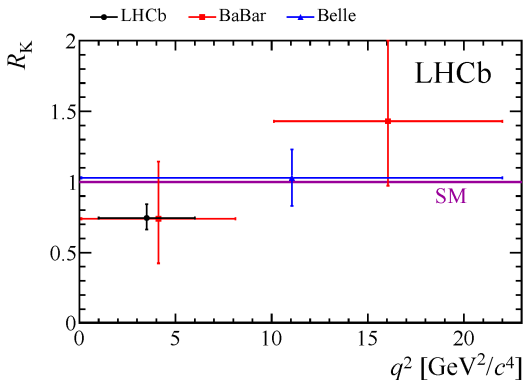

Figure 3: Difference of the ratio of branching fractions R_K between measurement and standard model prediction [4].
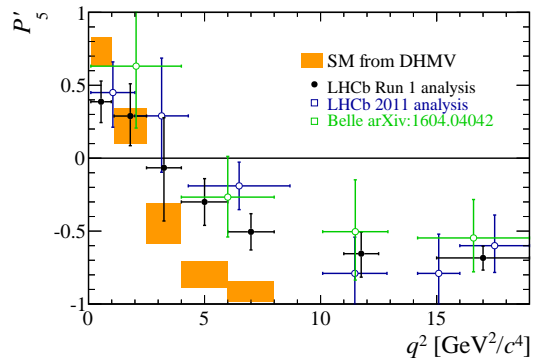

Figure 4: The difference in the P_5' between measurement and prediction by the standard model [10].
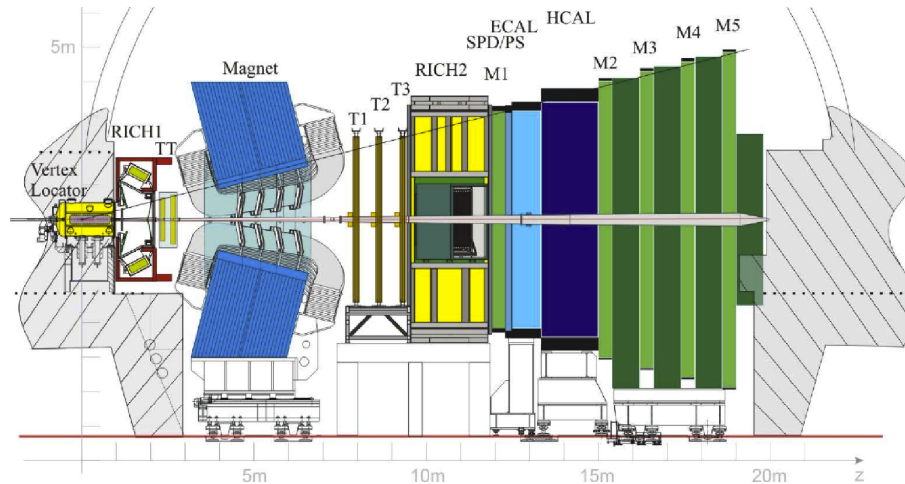
Figure 5: Schematic view of the LHCb spectrometer. Layout in the right-handed coordinate system with the z-axis long the beam, the y-axis along the vertical and the x-axis in the horizontal [9].

1.3 The LHCb detector

This analysis is performed within the LHCb, Large Hadron Collider beauty, framework. It is 21 meters long, 13 meters wide, 10 meters high and weights about 5600 tons. The LHCb is a single arm forward spectrometer designed to perform high-precision measurements of hadrons containing c- and b-quarks. The main purpose of the experiment is to search for CP-violation and rare decay processes. The LHCb detector is one of the four collision points of the Large Hadron Collider (LHC) at CERN in Geneva, Switzerland. The LHC is a circular hadron accelerator of a total length of 26.7 km which accelerates either protons or lead-ions. In the year 2011, the protons in the beam had an energy of 3.5 TeV and in 2012 it had an energy of 4 TeV. Currently (year 2016) the beam has an energy of 6.5 TeV, so the center of mass energy sums up to 13 TeV. Apart from the LHCb, the LHC has three other major experiments, CMS, ATLAS and ALICE. These detectors have a toroidal structure, which measure the particles in all directions. The purpose of CMS (Compact Muon Solenoid) and ATLAS (A Toroidal Lhc ApparatuS) are general purpose experiments with physics programmes ranging from the conformation and measurements of the observables of the standart model(*e.g.* Higgs boson) to search from for NP at the TeV scale. ALICE (A Large Ion Collider Experiment) is specialised on the research of dense quark-gluon-plasma, the state of matter produced shortly after the big bang.

The Vertex Locator (VELO) [2] is the closest sub-detector to the collision point and is composed of two retractable halves. These can be moved away from the beam to prevent beam-induced damage to the sensors. In the event, the beam is focused, the sensors are brought up to 7 mm to the beam. Each VELO half is composed of 21 silicone modules that measure the distance and angular placement of the outgoing particle. It has to be near the collision point to guarantee a high precision. This high precision is required for the precise track coordination to reconstruct the primary and secondary vertx of an event. Charged particles traverse the silicon and generate electron-hole pairs. These electrons are sensed using application specific-electronics [8]. The VELO has a signal to noise ratio of 20 and a best hit resolution of 4 μm [2].

Positioned either side of magnet, Ring Imaging Cherenkov detectors (RICH1 and RICH2) are designd tocharged particle using Cherenkov radiation. It is filled with a medium, so that light in this medium travels slower than usual. If a particles travels through that material faster than the local speed of light it emits a light-cone, similar to a sonic boom, know as Cherenkov-radiation. The light then gets reflected to photo-diodes, where it gets detected. From the opening-angle of the light-cone and the known refractive index of the medium, the velocity of the particle can be measured with the formula $\cos(\theta) = \frac{1}{n\beta}$, where θ is the angle of the cone, n is the refracting index of the medium and β is the the velocity of the particle measured in c. RICH1 measures particles with a velocity from 1 GeV to 60 GeV and RICH2 up to 100 GeV [8].

Behind Rich1 is the Trigger Tracker (TT). Among the main attributes of the TT are the reconstruction of the trajectory of the particles, in particular low momentum tracks and the reconstruction of particles that decay outside the VELO acceptance. The TT consists of several layers of silicon micro-strip detectors. Those strips have a different orientation in each layer, therefore the TT measures the trajectory of the particle up to a precision of 0.05 mm. The silicon micro-strips work the same as the silicon modules of the

---

[2] All the abbreviations in the text are the same as in Fig. 5

3

VELO [8].

Following the TT station is the dipole magnet, which bends the tracks of charged particles. From the bending radius, the momentum of the particles can be calculated. To calculate the momentum, the measurements from the TT and from the Tracking chambers are needed. The magnet is made out of two coils inside an iron yoke. The coils weight 27 tonnes without the yoke. The yoke weights about 1450 tonnes. The magnet inverses its polarity from time to time, to cancel out detection asymmetries. The polarity gets inverted roughly twice per month, the total time of the polarities is more or less the same [8].

After the magnet, there are the three Tracking chambers T1, T2 and T3. T1, T2 and T3 which measure the trajectory and the momentum of the particles. The trackers are split up in two parts, the inner and the outer tracker. The inner tracker is made out of strips of silicon, which work in the same way as the modulus in the VELO and the TT, measuring the tracks near the beam. The outer tracker is made out of tubes filled with gas with a wire in the middle with a high electrostatic potential. If a particle with a higher energy than the ionisation-energy of the gas enters the tube, the gas gets ionized and the free electrons travels to the wire in the middle and can be measured as current of the wire. The tubes have a very fast counting rate, so two different proton-proton collisions are distinguishable [8].

The Electromagnetic Calorimeter (ECAL) is positioned after the RICH2. It measures the energy and position of photons and electrons. In order to provide additional information for the detection of neutral particles, two scintillator layers (PS and SPD) are placed upstream of the ECAL. After that, the particles go through a layer of 12 mm lead, which releases a shower of particles. Behind the lead there is another layer of scintillators. The ECAL is composed of alternating layers of 3 mm thick scintillators and 2 mm thick lead. In the lead, the particles release a shower of photons and electrons, which are then measured in the scintillators. With the amount of particles released in the lead, the energy of the particle is determined. Scintillators are blocks of transparent plastic, which absorbs the energy of photons and electrons. The energy then gets emitted as light, which is detected by photo multipliers [8]. With the data from the photo multipliers, the energy of the particle can be recalculate with a precision of 2 % [9].

Downstream of the ECAL is placed the Hadronic Calorimeter (HCAL) . The HCAL works very similar to the ECAL, but instead of lead, the layers are made out of 16 mm iron. Between the layers of iron, a layer of scintillators measures the energy of the particles with a precision of 3 % [9]. The HCAL measures the energy of the hadrons, such as protons, pions and kaons [8].

Muon chambers are positioned at the downstream region of the spectrometer and consist of five stations (M1 to M5). These stations are the last part because the muons travel through the whole detector. With the information of M1 to M5 and the trackers, the track of the muons can be reconstructed. M1 is positioned in front of the ECAL in order to avoid multiple scattering processes originating in hadronic cascades and therefore improve the track-reconstruction. Between M2 to M5, there are layers of lead (80 cm thick) to shield remaining hadrons. The spectrometer measures the track of the μ with drift-tubes and multi-wired proportional counters. A drift-tube is a tube with a gas inside and in the middle of the tube there is a wire with a high electric potential. If a μ flies through a tube, it reacts with the gas and the gas will release electrons, which then drift towards the wire. From the time the electron needs to get to the wire, the distance between the track and the wire is calculated. with the combined data of all drift pipes, the muon tracks can be reconstructed. The multi-wired proportional counters work the same way as the tubes in the outer tracker in the Tracking Chambers, with the difference that in the counters is a grid of wires that collects the free electrons [8].

To reduce the data-output of 40 MHz of the LHCb to a computable level, uninteresting events have to be filtered out before they get stored. This is the task of the trigger-system. The LHCb trigger consists of two levels: Level 0 (L0) and High Level Trigger (HLT). The L0 trigger is implemented in hardware and makes decisions based on information from the calorimeter and muon systems in order to reduce the event rate to below 1 MHz, at which point the whole detector can be read out. The HLT is a software application running on the event filter farm (EFF). A fraction of L0 accepted events are stored to a disk for processing by the HLT during the time, when the LHC is not running, optimising the use of available EFF resources. EFF performs a simplified reconstruction of the tracks in the VELO. HLT1 limits the number of VELO tracks that are passed through the forward tracking algorithm that searches for matching hits in the tracking stations. VELO tracks must have a significant impact parameter with respect to all primary vertexes, or be matched to muon chamber hits by a fast muon identification algorithm. HLT1 reduces the event rate to about 80 kHz, which is sufficiently low to allow the forward tracking of all VELO tracks in the High level Trigger 2 (HLT2). The events that pass the HLT are saved and later processed with a more accurate alignment and calibration of the sub-detectors and reconstruction [9].

2 Analysis methodology

The sensitivity studies presented in this analysis are essentially conceptual in nature. In other words, the strategy of this analysis is to investigate the limitations and prospects for the first observation of the decay $B^0 \to K^{*0}\tau^+\tau^-$ at LHCb. This section is organized as follows. The topology of the final state has a distinctive signature in the detector. The features and implications for the reconstruction are discussed in Section 2.1 and Section 2.2. Finally, the selection algorithms, both online and offline, are discussed in Section 2.3.

2.1 $B^0 \to K^{0*}(892)\tau^+\tau^-$ topology

The main goal of this thesis is to investigate the feasibility of $B^0 \to K^{0*}(892)\tau^+\tau^-$ decays at LHCb. The B^0 meson is created in a proton-proton collision inside the VELO. It flies a short distance, until it decays still in the VELO to the two taus and the $K^{0*}(982)$. The $K^{0*}(982)$. resonance decays strongly to a $K\pi$ pair, which is reconstructed using information of the full tracking system (thereafter reffered to as long tracks). From the tracks of the $K^{\pm}$ and $\pi^{\mp}$, the secondary vertex of the B^0 is reconstructed. The tau can decay to several final states, however, in this study only the decay chain $\tau^{\pm} \to \pi^{\pm}\pi^{\pm}\pi^{\mp}\nu_{\tau^{\pm}}$ is examined. This transition is investigated due to the high branching ratio $(9.31 \pm 0.06)\%$ [13] and the possibility to reconstruct the tau momentum analytically. Notice that the ν_{τ} is not measured by the detector. Details on the B mass reconstraction using this approach is given in Section 2.2. The signature of the whole decay is schematically shown in Fig. 6

2.2 Reconstruction of the B^0-mass

To reconstruct the mass of the B^0 meson, the determiation of the τ momentum is required. In order to perform a full reconstruction of the tau lepton, two assumptions are considered: The mass of the tau is known and corresponds to its nominal value; neutrinos are massless. Note that additional neutral particles would invalidate this approach, since the combined mass is unknown. The following solution can be examined:

$$m_\nu = 0 \tag{2.1}$$

$$0 = (p_\tau - p_{3\pi})^2 \tag{2.2}$$

$$m_\tau^2 = E_\tau^2 - \vec{p_\tau}^2 \tag{2.3}$$

$$\vec{p_\tau} = |\vec{p_\tau}| \cdot \vec{u} \tag{2.4}$$

where m_ν is the mass of the neutrino, p_τ is the 4-momentum of the τ, $p_{3\pi}$ is the sum of the 4-momentum of the three pions, m_τ is the rest mass of the τ, E_τ is the total energy of the τ, $\vec{p_\tau}$ is the 3-momentum
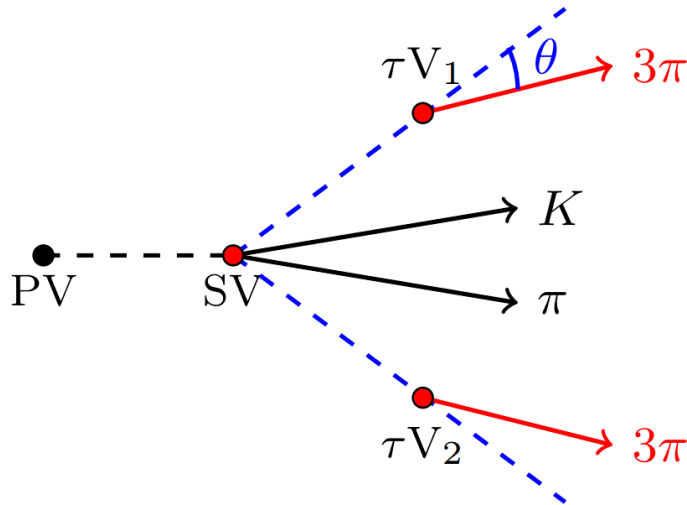

Figure 6: Topology of the decay. All the dotted lines represent particles that do not leave hits in the detector. The vertices get reconstructed by the tracks of the pions and Kaon [16].

Figure 7: Explenation of the variables in Eq. (2.5) and the two possibilities. 3π means the sum of the three pions.[12]

the τ and $\vec{u}$ is the direction of $\vec{p_\tau}$. Equations 2.1-2.3 ensure the correct mass for the ν_τ and τ, whereas Eq. (2.4) ensures the correct direction of the $\vec{p_\tau}$. The momentum of the $\tau^\pm$ is calculated as follows [12]:

$$|\vec{p_\tau}| = \frac{\left(m_\tau^2 + m_{3\pi}^2\right)|\vec{p}_{3\pi}|\cos^2(\theta) \pm E_{3\pi}\sqrt{\left(m_\tau^2 - m_{3\pi}^2\right)^2 - 4m_\tau^2|\vec{p}_{3\pi}|^2\sin^2(\theta)}}{2\left(E_{3\pi}^2 - |\vec{p}_{3\pi}|^2\cos^2(\theta)\right)}, \tag{2.5}$$

where $m_\tau = 1777\,\text{MeV}$ [13] and $m_{3\pi}$ respectively are the mass of the $\tau^\pm$ and the sum of the masses of the three pions, $E_{3\pi}$ is the total energy of the sum of the pions and θ is the angle between the momentum of the $\tau^\pm$ and the pions as seen in Fig. 7. The mass of the pion is given by $m_{3\pi} = \sqrt{\left(\Sigma_{i=1}^3 E_{\pi i}\right)^2 - \left(\Sigma_{i=1}^3 \vec{p}_{\pi i}\right)^2}$. If $\left(m_\tau^2 - m_{3\pi}^2\right)^2 - 4m_\tau^2|\vec{p}3\pi|^2\sin^2(\theta) < 0$, then $|\vec{p_\tau}|$ would have an imaginary part. In this case, the real part of the solution is taken.

Finally, the momentum of the $K^{0*}(892)$ is given by the sum of the momenta of the π and the K particle. To reconstruct the B^0 mass, the total energy each particle can be calculated by its momentum and mass

$$E_{particle} = \sqrt{m_{particle}^2 + \left|\vec{p}_{particle}^2\right|}, \tag{2.6}$$

where $E_{particle}$ is the energy of the particle, $m_{particle}$ is the rest mass and $\vec{p}_{particle}$ is the momentum of the particle. With all that information, the B^0 mass can finally be reconstructed as

$$m_B = \sqrt{E_{tot}^2 - p_{tot}^2} = \sqrt{\left(E_{\tau^+} + E_{\tau^-} + E_{K^{0*}}\right)^2 - \left(|\vec{p}_{\tau^+}| + |\vec{p}_{\tau^-}| + |\vec{p}_{K^{0*}}|\right)^2}, \tag{2.7}$$

where m_B is the mass of the B^0 meson, E_{tot} is the sum of the energies from Eq. (2.6) and p_{tot} is the sum of the momenta of all the outgoing particles.

As can be noticed, for each τ-momentum there are two solutions, due to the sign in Eq. (2.5). These solutions are schematically represented in Fig. 7. All four different combinations are analytically valid solutions for the mass of the B^0 meson. In order to examine a single mass hypothesis, the following criteria is considered

(i) The absolute value of both of the momentum of the τ's have to be positive;

(ii) if more than one solution: one of the remaining possibilities is chosen randomly;

(iii) if there is no solution, the event gets removed.

The first criteria is applied in order to preserve a validate solution for Eq. (2.7). The investigation of alternative approaches which could explore some related variable or improvements on the analyticity of the tau momentum is a natural extension of these studies. For the purpose of this analysis, the methodology provides a general overview of the current sensitivity.

2.3 Event selection

2.3.1 Dataset

This analysis is performed using the *pp* collision data, corresponding to an integrated luminosity of approximately $50\,\text{pb}^{-1}$ collected at a center-of-mass energy of $8\,\text{TeV}$ recorded by the LHCb experiment. These events are pre-filtered by the `Stripping B2KstTauTau_TauTau`, included in the `Bhadron.mdst`

stream. The stripping conditions can be seen in Table 1. Candidates are reconstructed with the `Reco14` version of the reconstruction software. This dataset is hereafter referred as Opposite Sign line (OS-line) The second dataset is pre-filtered in the same way as the OS-line, with addition of `Same Sign` name. In this dataset, events are collected, by reconstructing both taus with have the same charge. As a consequence of charge conservation, these events are compound of background-only. The second dataset is called Same-Sign line (SS-line). Monte Carlo (MC) simulated events are used to study the signal contribution using the `Sim08_c` production version, and similar conditions as it was used in the data taking. Samples are generated with 2012 conditions, *i.e.* a pileup $\mu = 2.5$ and `CONDDDB = Sim08-20130503-vc-mu100`, `DDDB = Sim08-2010503` database tags. The number of events generated corresponds to 1514911 and consists of the magnet polarity in "Down" configuration.

Variable	Acccepted value
B mass	[2000,10000]MeV
B P_T	>3000
B vertex χ^2	<100
B flight distance	<40
$K^*(892)$ mass	[700,1100]MeV
$K^*(892)$ P_T	>1000
$K^*(892)$ vertex χ^2	<15
$K^*(892)$ flight distance	>3000

Table 1: Stripping conditions

2.3.2 Preselection

In order to further reduce the size of the data sample, high signal-efficiency cuts are applied on the top of the stripping. Candidates on the opposite direction of the detector are removed by requiring that the cosine of the angle between the the reconstructed momentum of the B meson and its direction of flight to be positive. Particle identification requirements are applied on all charged pion tracks. A mild criteria on the logarithm of the likelihood ratio between kaon and pion hypothesis is placed to discriminate essentially combinatorial backgrounds. It is worth to mention that the simulated samples are pre-matches to the true-signal information. These cuts will be referred as sanity cuts later on.

Another important consideration is the association of a trigger object with a signal track. An event is classified as TOS if the signal nder study triggers the event , whereas TIS categorises the trigger object not associated to the signal. Using these categories, criteria's utilised in this analysis is given as follows

- Level 0: (`L0Hadron TOS` or `L0Global TIS`)
- High level trigger 1, `HLT1TrackAllL0Decision_TOS`, will be called HLT1 bellow
- High level trigger 2, `HLT2Topo4BodyBBDTDecision_TOS`, will be called HLT2 bellow [3].

To find out if an event gets triggered, the different trigger have to be combined as follows:

$$T = (\texttt{L0Hadron TOS} \vee \texttt{L0Global TIS}) \wedge HLT1 \wedge HLT2, \quad (2.8)$$

where T is either true or false. The total trigger efficiency is then calculated by simply dividing the numbers of triggered events by the total amount of signal-events:

$$\epsilon_{Trigger} = \frac{N_{T=true}}{N_{total}} \quad (2.9)$$

2.3.3 Variables

Further separation of signal from background is achieved with a multivariate method (MVA). In the scope of supervised machine learning, the classifiers investigated have the purpose to provide classification to apparent undistinguishable samples. Several algorithms can be used as classifier, each one of them needs a set of variables to train. For the purpose of this thesis, a set of criteria was used in the choice of the variables. First, the distribution of the background and the signal should differ from each other as much as possible (see Fig. 14). Second, the correlation between the variables should be as small as possible

---

[3]Additional triggr lines including 2 and 3 body decays have been examined and should increase the efficiency level in a percent level.

(see Fig. 8). Finally, the amount of variables should be presentably small, in order to reduce potential systematic uncertainties related to discrepancies between MC and data. Variables can be combined, to reduce the number of variables, or altered by mathematical functions, so the difference of the distributions between background and signal is enhanced. In this thesis, the following variables are used:

- χ^2 of the impact parameter of the primary vertex of the B-meson (`B_IPCHI2_OWNPV`)

- $\log(1 - \cos(\text{Dira angle}))$(`log_cos_1min_DIRA_OWNPV`)

- χ^2 of the decay position (endvertex) of the $K^{0*}(892)$ (`Kst_ENDVERTEX_CHI2`)

- Distance in the XY-plane of the detector between primary- and endvertex of the $K^{0*}(892)$ (`Kst_BPVVDRHO`)

- Isolation value of the $K^{0*}(892)$ (`Kst_BDTiso3`), *i.e* number of extra tracks (excluding tracks that have been used to form candidates) that an form a vertex.

- Logarithm of the sum of minimal impact parameter divided by 2 of the $\pi^{\pm}$ and $K^{\mp}$ with respect to $K^{0*}(892)$ (`log_K_pi_MINIP_div2`)

- Isolation value of one of the $\tau^{\pm}$ (`tauM_BDTiso3`), same definition as `Kst_BDTiso3`.

- Cosine of the angle between the momentum of the $\tau^{\pm}$ and the momentum of the three π (`tauM_cos3pi`)

- Isolation value of the other $\tau^{\pm}$ (`tauP_BDTiso3`), same definition as `Kst_BDTiso3`.

- Same as `tauM_cos3pi`, but for the opposite sign τ (`tauP_cos3pi`)

2.3.4 MVA training

Nine different classifier have been studied in the framework of scikit-learn package [14]. These classifier have to be trained, afterwards they calculate the probabilities of events of being a signal event or a background event, or they directly classify the events as signal or background. The stability and accuracy of the classifier is verified by considering a 10% fraction of the original signal and background samples. In this subset, the processing times of the algorithms are significantly reduced, though keeping enough data [4] [5] to compare the results of the different methods. Two third of those 10% are used as a training sample for the learning process and the rest as test sample. Each method is trained and tested with the same dataset. In the first run, the baseline parameters implemented of the methods in the framework are used. To compare the classifiers, the precision of the direct classification for all events was measured (total precision), whereas the precision of the signal events is unified in the test sample (signal precision). The signal precision is used to verify if there is any issue with the classifier. For example, if twice as many signal-events are found as there was signal in the test-sample, then it is likely that there is a bias in the classifier.

The best three classifier which had almost the same total and signal precision are the Gradient Boosted tree Classifier (GBC), the Random Forest Classifier (RFC) and the Boosted Decision Tree with the AdaBoost algorithm classifier (BDT).

2.3.5 Gradient tree Boosting Classifier (GBC)

The gradient tree boosting classifier uses decision trees of a fixed size. A decision tree is a set of simple decisions to classify data (*e.g.* is the value higher or lower than a threshold). It has some advantages to other classifier. For example, the data for the decision tree does not need much preparation such as normalisation. The principle of a decision tree is shown in Fig. 9.

Gradient Tree Boosting or Gradient Boosted Regression Trees (GBRT) is a generalization of boosting to arbitrary differentiable loss functions. GBRT is an accurate and effective off-the-shelf procedure that can be used for both regression and classification problems. Gradient Tree Boosting models are used in a variety of areas including Web search ranking and ecology. The Advantages of the GBC are its predictive power, its natural handling of data of mixed type and its robustness to outliers in output space. The disadvantage of the GBC is the scalability, due to the sequential nature of boosting it can hardly be parallelized [14].

---

[4]On the data for the training the cuts described in Section 2.3.2 and the trigger have been applied

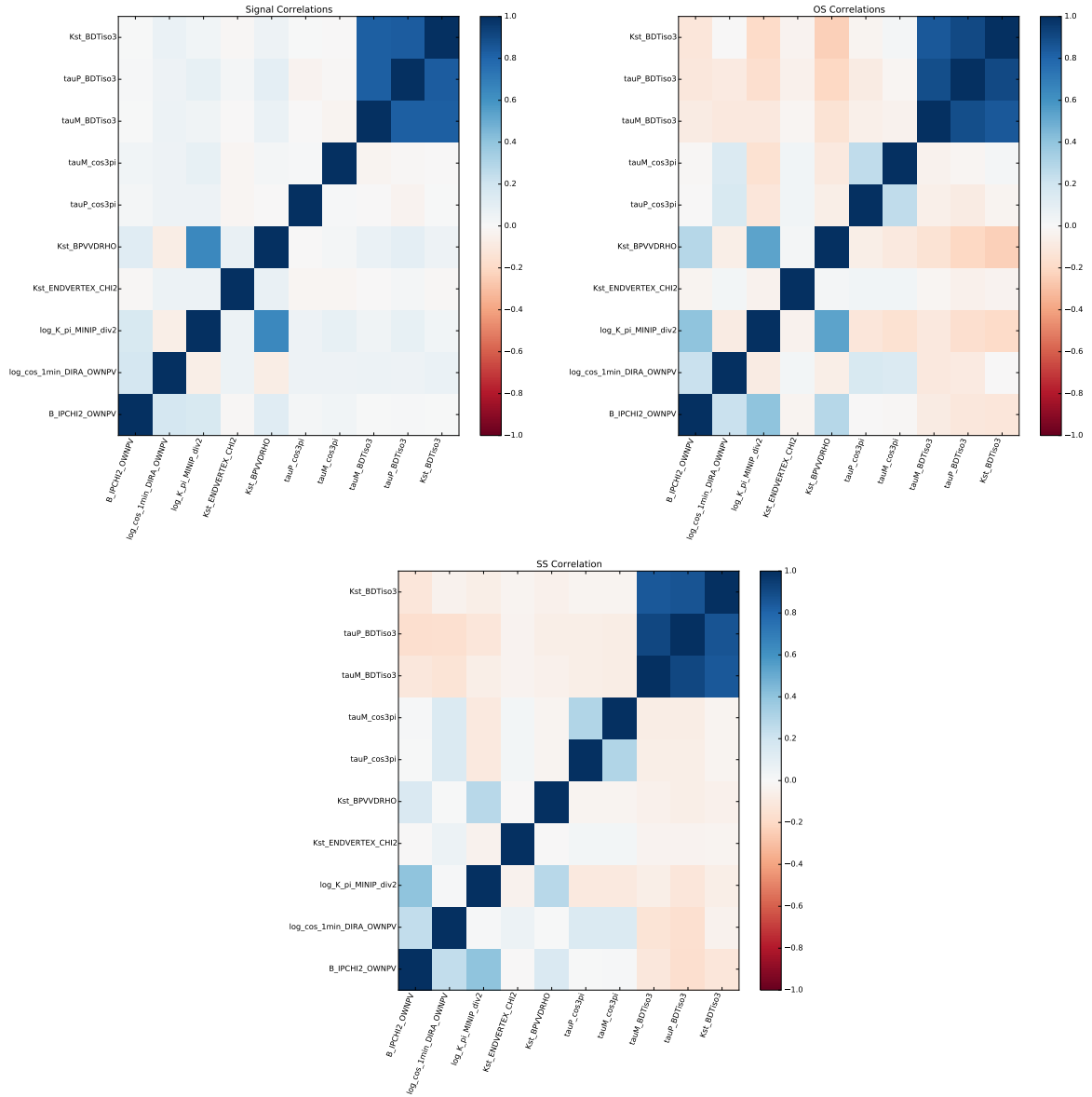[5]For the training, only data of the signal- and SS-line were used.

Figure 8: Correlations between the variables used in the machine learning training for simulated $B^0 \to K^{*0}\tau^+\tau^-$ decays (top left) and background from the OS-line (top right) and SS-line (bottom).

2.3.6 Random Forest Classifier (RFC)

In random forests, each decision tree in the ensemble is built from a sample drawn with replacement (*i.e.* a bootstrap sample) from the training set. In addition, when splitting a node during the construction of the tree, the split that is chosen is no longer the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. As a result of this randomness, the bias of the forest usually slightly increases (with respect to the bias of a single non-random tree) but, due to averaging, its variance also decreases, usually more than compensating for the increase in bias, hence yielding an overall better mode [14].

2.3.7 Adaptive Boosted Classifier with decision tree (BDT)

The core principle of the BDT is to fit a sequence of weak learners (decision trees that are slightly better than a random selection) on repeatedly modified versions of the data. The predictions from all of them are then combined through a weighted majority vote to produce the final prediction. The data modifications, in each so-called boosting iteration, consist of applying weights to each of the training samples. Initially, those weights are all set equally, so that the first step simply trains a weak learner on the original data. For each successive iteration, the sample weights are individually modified and the learning algorithm is re-applied to the reweighted data. At a given step, those training examples that were incorrectly predicted
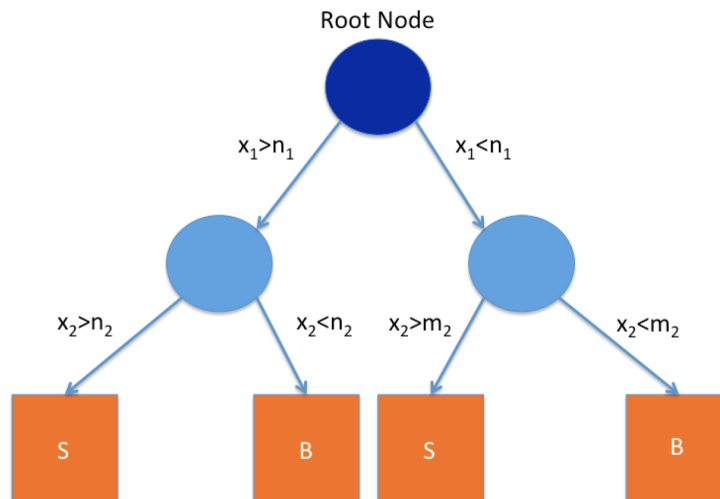
9

Figure 9: The principle of a decision tree. For a given variable x_i, a set of decisions are found that give the best separation between background and signal [6].

by the boosted model at the previous step have their weights increased, whereas the weights are decreased for those that were predicted correctly. As iterations proceed, examples that are difficult to predict receive ever-increasing influence. Each subsequent weak learner is thereby forced to concentrate on the examples that are missed by the previous ones in the sequence [14].

2.3.8 Classifier optimisation and cross validation

The total precision of the three classifier is maximized by optimising the parameters of the classifiers. Those parameters are optimized by scanning over the different possible combinations of the parameters, referred to as grid-search. The grid search is done with the same amount of data used before to determine the best classifiers. Again the total and signal precision were used to compare and determine the best parameters. Because there is still no significant difference between the three methods after optimizing the parameters, all three of them are used for classifying of the the whole data and calculate the probabilities for all events.

To get the total efficiency of the machine learning, one has to do a cross-validation. This is done as following

(i) split the data into 10 sets;

(ii) train a classifier with all but one sets;

(iii) calculate the probabilities for the events of the remaining set;

(iv) repeat step (ii) and (iii) for all sets;

(v) stack the results of the tests back together;

In the end, the training is performed with the with the whole data and applied to the OS-line.

2.3.9 Figure-of-Merits (FoM)

The results of the cross validation are used for the calculation of the figure of merits. In particular, the number of events expected for both signal and background contributions can be estimated for each working point of the classification probability. The choice of the optimal cut on the algorithm response can be driven by several FoM, *e.g.* [6]:

$$\text{FoM}_1 = \frac{S}{\sqrt{B+S}}, \tag{2.10}$$

$$\text{FoM}_2 = \frac{\epsilon}{a/2 + \sqrt{B}}, \tag{2.11}$$

---

[6]On the data to calculate the figure of merits, also the reconstruction of Section 2.2 cut has been applied.

where S is the yield of the signal, B is the yield of the background, $a = 5$ is a constant and ϵ is the efficiency of the machine learning of the signal after all the other cuts were made. The signal yield is determined as

$$S = \int \mathcal{L} \cdot BR(B^0 \to K^{*0}\tau^+\tau^-) \cdot BR(\tau^- \to \pi^+\pi^-\pi^-\nu_\tau)^2 \cdot \sigma_{b\bar{b}} \cdot f(B^0) \cdot 2 \cdot \epsilon \qquad (2.12)$$

where $\int \mathcal{L}dt$ is the luminosity, $BR(B^0 \to K^{*0}\tau^+\tau^-) = 8.14 \cdot 10^{-8} \pm 12.5\%$ [7][7] and $BR(\tau^- \to \pi^+\pi^-\pi^-\nu_\tau) = (0.093 \pm 0.0006)$ [13] the branching ratios, $\sigma_{b\bar{b}} = 284 \pm 53\mu b$[1] the cross section and $f(B^0) = 0.402 \pm 0.007$ [13] is the hadronisation factor.

Notice that $\texttt{FoM}_1$ is the standart choice of the maximal significance of the signal with respect to background. However, in the search for previously unobserved modes or in the event there are no reliable predictions, $\texttt{FoM}_2$ provides an unbiased estimator, simply by choosing the value a that quantifies the target level of significance in units of standard deviations. Both approaches are evaluated in this thesis.

The number of background events in Eq. (2.10) and Eq. (2.11) is estimated in a different manner. First, calculate the region in the reconstructed B^0-mass of one σ of the signal after the probability cut. Then, apply the probability cut on the background sample which lies in the same region as the one σ region. Finally, the number of Background events is mutiplied by the scaling factor n, defined as follows

$$n = \frac{N_{OS}}{N_{SS}}\frac{L_{OS}}{L_{SS}}. \qquad (2.13)$$

N_{OS} is the amount of events in the opposite sign line after all cuts for the events which have a measured B^0-mass above $5.3GeV$ [8], N_{SS} is the same but for the same sign line, $L_{OS} = 47.39 \pm 0.55pb^{-1}$ the integrated luminosity of the opposite sign line and $L_{SS} = 47.29 \pm 0.55pb^{-1}$ the integrated luminosity of the same sign line. n is there to ensure that the signal gets compared with an accurate amount of background with the correct luminosity. There is the possibility that the factor $\frac{N_{OS}}{N_{SS}}$ is different in the mass region of the signal. For the purpose of this thesis, it was assumed that this is a good approximation.

---

[7]Numerical predictions provided by Danny van Dyk

[8]This is the B^0-mass that got reconstructed without any assumptions of the missing energy of the neutrinos.

3 Results

3.1 Efficiencies

3.1.1 Preselection efficiency

All the particle of the decay have to travel through the detector, otherwise the event does not get registered. The geometric efficiency describes therefore the percentage of events in which all the particles stay inside the detector. In Table 2, the geometric efficiency is gathered with all the others the efficiencies described in Section 2.3.1.

	Efficiency	error
geometric	2.28%	±0.08%
stripping	8.69%	±0.01%
sanity cuts	2.16%	±0.04%
total preselection	0.43%	$\pm 8 \cdot 10^{-5}\%$

Table 2: Efficiencies of the preselection cuts described in Section 2.3.1 and the geometric efficiency of the detector.

3.1.2 Trigger efficiency

In order to have a more accurate perception of the efficiencies of test criteria, each individual decision is shown in Table 3. Notice that the final $\epsilon_{trigger}$ is given by combining the intermediate steps into an overall efficiency. The L0 criterion used in this analysis is designed to select at least one hadronic trigger object associated to a high traverse-momentum track from the signal or an object that has no associated overlap with the signal. The obained efficiency is consisten with similar final states [15], which is explained by the inclusive hadronic approach applied. Similary, the indicated high level trigger efficiency ae in agreement with other hadronic multi-body decay.

Trigger	Efficiency	error
L0	49.2%	±1.6%
Hlt1	88.2%	±3.4%
Hlt2	63.9%	±2.9%
All trigger	27.8%	±1.1%

Table 3: The efficiencies of the triggers calculated using equation (2.9).
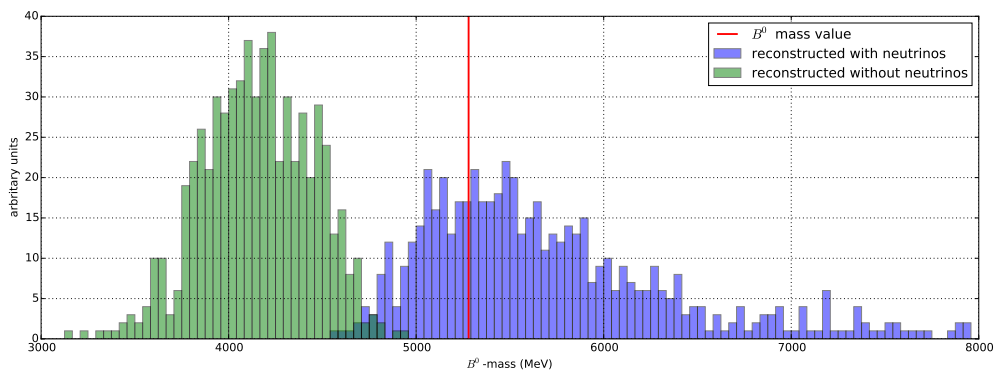

Figure 10: Selected B-masses after the selection, as described in Section 2.3.2. "Reconstructed with neutrinos" refers to reconstructed method described in Section 2.2 and "reconstructed without neutrinos" refers to the initial method to reconstruct masses. The red line indicates the nominal B^0-mass value [13].
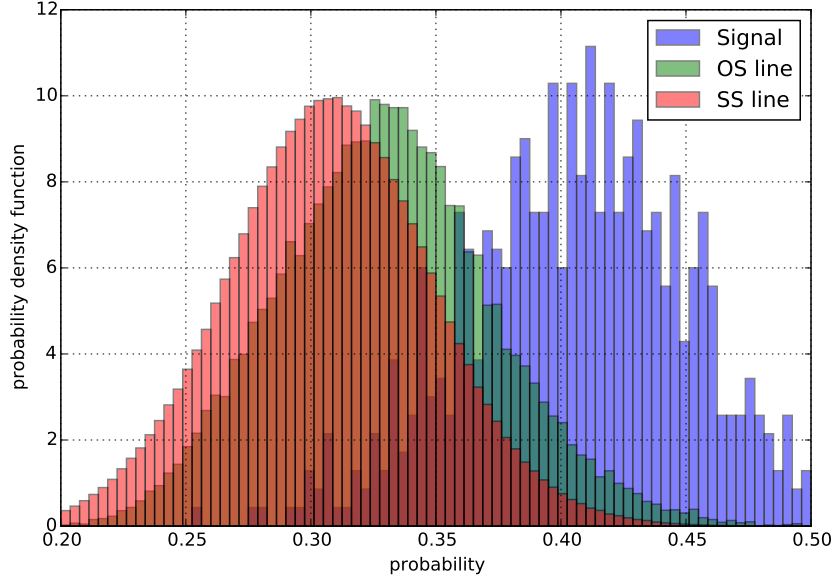
Figure 11: Binned probability density function of the BDT trained with the the full SS-line. x-axis is the probability, which the classifier calculated.

3.1.3 B-mass reconstruction

The B meson candidate invariant mass distribution before the multivariant selection is shown in Fig. 10. For completeness the four possible hypotheses are shown in Fig. 17 in the Appendix A.3. There is a clear displacement of the mean value of the distribution towards the nominal B mass. This is explained by the rescale of the τ momentum, which somehow corrects for the energy lost from the undetected neutrinos. The B^0-mass is reconstructible for $(79.3 \pm 04.3)\%$ of the events. Interestingly, the term bellow the square-root in Eq. (2.5) was never negative for the data events. The τ-momentum was negative due to the cosine in Eq. (2.5), which can be negative. This can also be seen in Fig. 14 in the diagram `tauM_cos3pi` and `tauP_cos3pi`.

3.1.4 Multivariate selection

Although the preliminary studies described in Section 2.3.4 are useful in the commissioning of the framework, the sample size for the SS line is rather limited. In particular, there is no event in the SS line with associated probability in the high-region of the signal, as shown in Fig. 15. Therefore, the cross-validation is re-evaluated with a larger SS-line sample, with a corresponding luminosity of $(980.8 \pm 0.8)pb^{-1}$. The distribution of the probabilities of the BDT can be seen in Fig. 11, whereas for the GBC and RFC methods are shown in Fig. 16. Since the probabilities of the GBC and RFC are significantly clustered, the figure-of-merits are examined only with the probabilities of the BDT.

The Receiver Operating Characteristic curve (ROC-curve) of the BDT of the SS-line and OS-line can be seen in Fig. 18. The optimal values for `FoM`$_1$ and `FoM`$_2$ are 0.414 and 0357, respectively. The figure of Merits can be seen in Fig. 12 and Fig. 13. The efficiencies for the multivariate analysis are gathered in Table 4.

	efficiency	error
`FoM`$_1$	43.8%	$\pm 3.17\%$
`FoM`$_2$	86.9%	$\pm 5.09\%$

Table 4: The efficiencies after the cuts calculated in the figure of merits

3.2 $B^0 \to K^{*0}(892)\tau^+\tau^-$ sensitivity

In Table 5 and Table 6 are gathered all the efficiencies of the signal mode. The most significant cuts are coming from the detector, due to its forward structure, and the sanity cuts. Note that the trigger is responsible for a reduction of $(62.2 \pm 1.1)\%$ of the total number of events. For $(79.3 \pm 4.2)\%$ of the events it is possible to reconstruct the B^0-mass. The cuts of the machine learning removes 50% and 15% for $\texttt{FoM}_1$ and $\texttt{FoM}_2$, respectively. This corresponds to an efficiency of $(4.1 \cdot 10^{-4} \pm 3.3 \cdot 10^{-7})$ % and $8.18 \cdot 10^{-4} \pm 6.64 \cdot 10^{-6}$ % for the machine learning cut for $\texttt{FoM}_1$ and $\texttt{FoM}_2$. For an integrated luminosity of $(0.0473 \pm 0.00055)\text{fb}^{-1}$, this results in a yield of (0.004 ± 0.001) events, respectively (0.007 ± 0.002).

The estimated yield for Run1 at the LHCb experiment is found to be 0.0020 ± 0.0005 and 0.0040 ± 0.0010 events for $\texttt{FoM}_1$ and $\texttt{FoM}_2$, respectively. These predictions can be extended to higher luminosity scenario: Run2 with a luminosity of $10\,\text{fb}^{-1}$ and the future upgrade of the LHCb with an integrated luminosity of $300\,\text{fb}^{-1}$. The predictions are shown in Table 7. Notice that even in the upgrade setup the possibility for the $B^0 \to K^{0*}(892)\tau^+\tau^-$ is limited to the low efficiency of the detector. In other words, with the current apparatus, it is unlikely that the first observation of this mode can be performed in LHCb.

	Efficiency	error
geometric	2.28%	±0.08%
stripping	8.69%	±0.01%
sanity cuts	2.16%	±0.04%
L0	49.2%	±1.6%
Hlt1	88.2%	±3.4%
Hlt2	63.9%	±2.9%
trigger total	27.8%	±1.1%
reconstruction	79.3%	±4.2%

Table 5: Efficiencies of all the cuts before the offline event selection.

Figure of merits	$\texttt{FoM}_1$		$\texttt{FoM}_2$	
	efficiency	error	efficiency	error
machine learning cut	43.8%	±3.2%	86.9%	±5.1%
total efficiency	$5.29 \cdot 10^{-4}\%$	$\pm 3.3 \cdot 10^{-7}\%$	$8.18 \cdot 10^{-4}\%$	$\pm 6.7 \cdot 10^{-7}\%$

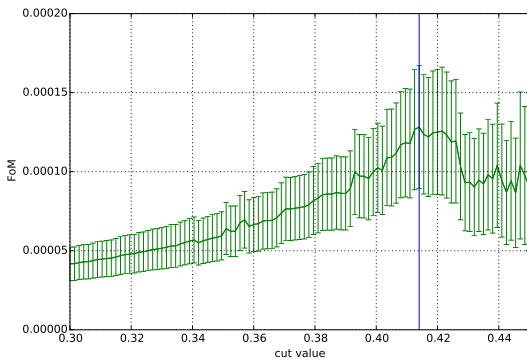Table 6: Efficiencies of the offline event selection and total efficiency.

Figure 12: Figure of Merits as calculated in equation (2.10). The maximum is at the cut of 0.414, represented by the blue line.
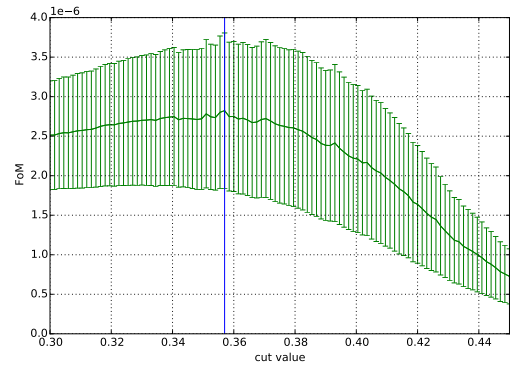
Figure 13: Figure of Merits as calculated in equation (2.11). The maximum is at the cut of 0.357, represented by the blue line.

	Luminosity	yield	error on yield
FoM$_1$	(0.047 ± 0.001)fb^{-1}	$3.1 \cdot 10^{-5}$	$\pm 0.8 \cdot 10^{-5}$
	Run1 (3fb^{-1})	0.0020	± 0.0005
	Run2 (10fb^{-1})	0.0066	± 0.0017
	Upgrade (300fb^{-1})	0.20	± 0.05
FoM$_2$	(0.047 ± 0.001)fb^{-1}	$6.2 \cdot 10^{-5}$	$\pm 1.6 \cdot 10^{-5}$
	Run1 (3fb^{-1})	0.0040	± 0.0010
	Run2 (10fb^{-1})	0.013	± 0.003
	Upgrade (300fb^{-1})	0.35	± 0.10

Table 7: Estimated number of events for $B^0 \to K^{*0}\tau^+\tau^-$ decays for different integrated luminosities, separated for FoM$_1$ and FoM$_2$.

4 Conclusion

In this thesis, the feasibility of the decay $B^0 \to K^{0*}(892)\tau^+\tau^-$ at LHCb was investigated. To calculate the yield, the efficiencies of the stripping, triggers, reconstruction and offline selection were estimated using a combination of monte carlo simulated signal and background from pp-collision. The reconstruction uses the transition $\tau^- \to \pi^-\pi^-\pi^+\nu_\tau$, for which the B mass is reconstructed analytically. The offline selection uses machine learning, for which two figure of Merits were maximised.

At the current state of the LHCb experiment, it is unlikely that the decay $B^0 \to K^{0*}(892)\tau^+\tau^-$ can be observed in a near future. Therefore improvements on the candidate selection are mandatory, in particular related to the reconstruction. It is worth mentioning that more realistic conditions (*e.g.* inclusion of specific background and detailed efficiency estimators) should be provided and properly considered in the measurement of this mode.

Another improvement that can be done for this mode is the inclusion of other final states. For instance, the transition $\tau^- \to \nu_\tau\mu^-\overline{\nu}_\mu$ which has a higher branching ratio of $(17.41 \pm 0.04)\%$ [13]. The disadvantage of this mode is the presence of to neutrinos. Since both neutrinos cannot be detected, the secondary vertex of the tau is almost impossible to reconstruct with a good precision, and therefore, the momentum of the tau is not analytically reconstructed. Notice that for this final state it is expected a higher efficiency, which is essentially driven by the presence of a muon in the final state. Finally, the combination of a tau from the hadronic chain with a tau $\to \nu_\tau\mu^-\overline{\nu}_\mu$ is a very interesting scenario. Although this is still a two-fold ambiguity , the sensitivity for this mode is boosted due to higher efficiency from the muon, therefore, an extension of the studies with these final states is rather appealing.

A   Appendix

A.1   Acknowledgements

I Want to thank to the department of physics of the University of Zürich, in particular professor Nicola Serra and his research-group, to make this thesis possible and give me an interesting look into the research at LHCb.

Next thanks goes to my college Jonas Eschle, who help me a lot with problem in Linux and programming in general.

At last, a very special thanks goes to Doctor Rafael Silva Coutinho, who basically helped with everything and had the patience to answer all the questions I had.

A.2   Error calculation

A.2.1   Error on trigger efficiency

$$e_\epsilon = \frac{1}{\sqrt{N_{T=True}}} \tag{A.1}$$

A.2.2   Error on Figure of Merits

The relative error on the efficiency is:

$$r_\epsilon = \frac{1}{\sqrt{S_c}} \tag{A.2}$$

with S_c the number of events after applying the cut.

The relative error on the Yield is:

$$r_S = \sqrt{r_L^2 + r_c^2 + r_h^2 + r_B^2 + r_\epsilon^2} \tag{A.3}$$

with the relative errors of the luminosity L , cross section c, hadronisation-factor h, branching-ratio B and the efficiency ϵ.

Error on Eq. (2.10)

$$e_{\texttt{FoM1}} = \sqrt{\left(\frac{1}{\sqrt{S+B}} - \frac{S}{2(S+B)^{3/2}}\right)^2 (S \cdot r_S)^2 + \frac{S^2}{4(S+B)^3} \cdot B} \tag{A.4}$$

Error on Eq. (2.11)

$$e_{\texttt{FoM2}} = \sqrt{\left(\frac{\epsilon \cdot r_\epsilon}{5/2 + \sqrt{B}}\right)^2 + \left(\frac{\epsilon^2}{(5/2 + \sqrt{B})^4}\right)} \tag{A.5}$$
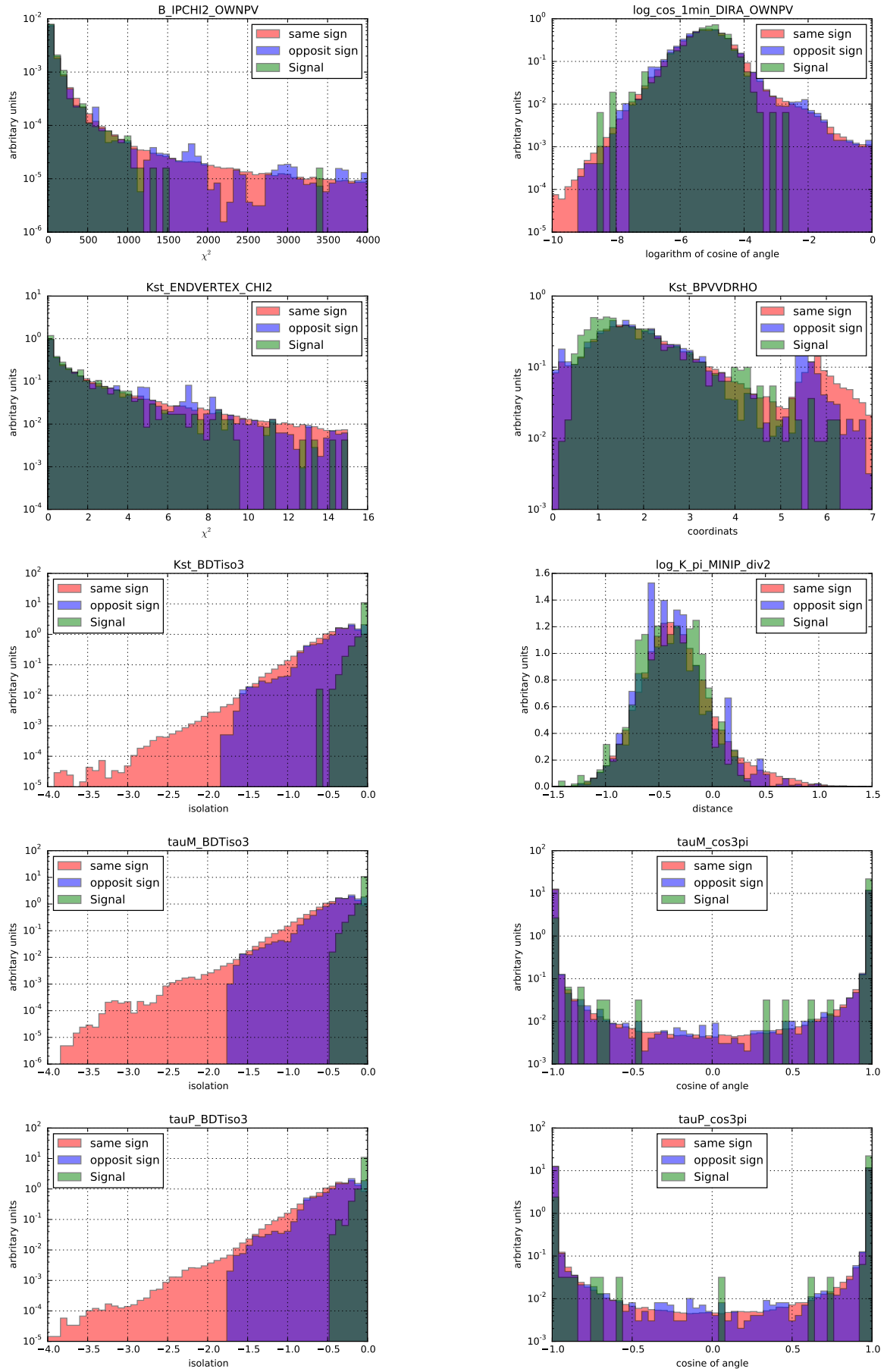
A.3   Complementary figures

Figure 14: Signal and background distribution of all the variables used in machine-learning. All histograms are normalized. The SS-line was calculated with the sample with corresponding luminosity of $980.8 \pm 0.8 pb^{-1}$.
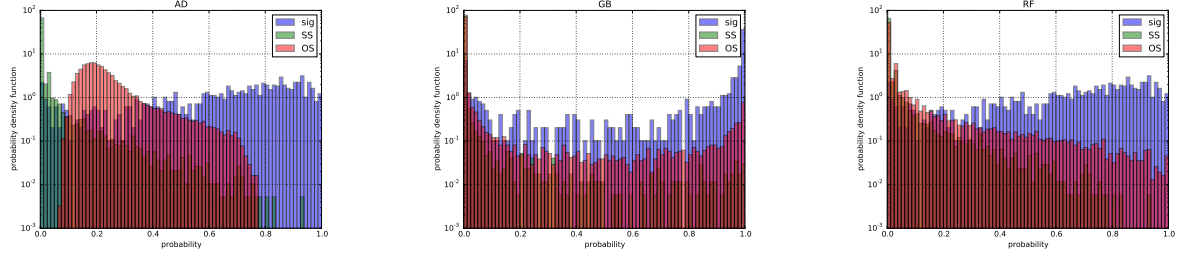
Figure 15: Binned probability density function of the different classifier trained with the smaller SS-line. From right to left: BDT, GBC and RFC. x-axis is the probability, which the classifier calculated. No points of the SS-line are in the regime of the signal line
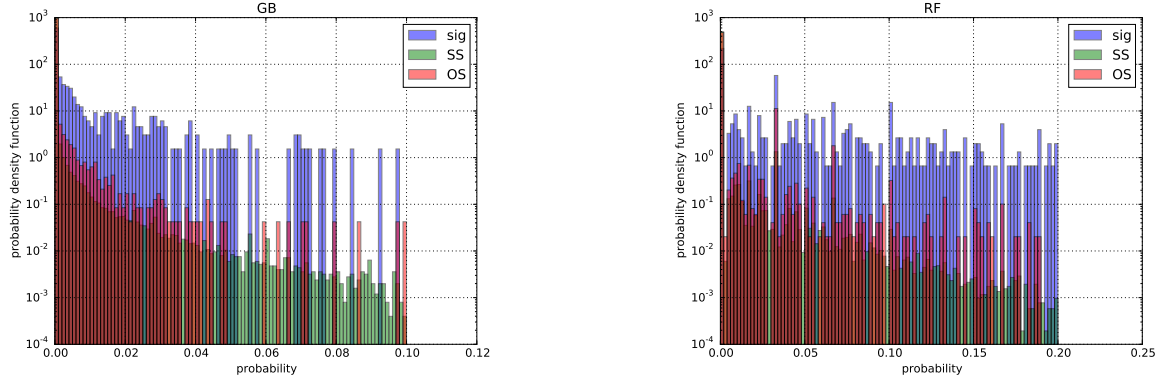

Figure 16: Binned probability density function of the different classifier trained with the the full SS-line. left is the GBC and right is the RFC. x-axis is the probability, which the classifier calculated. It's straightforward to see that the probabilities of the GBC and RFC are extremely clustered
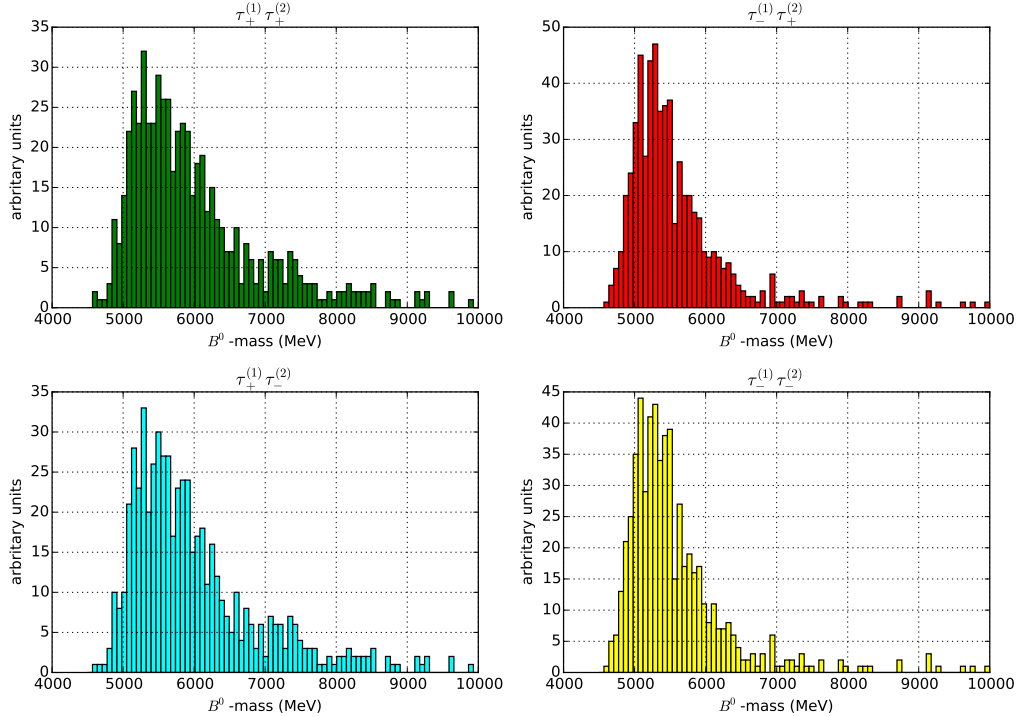

Figure 17: The four different reconstructed B masses as calculated with Eq. (2.7). The sign in the lower index of the τ refers to the sign used in the Eq. (2.5) for the calculation of the momentum.
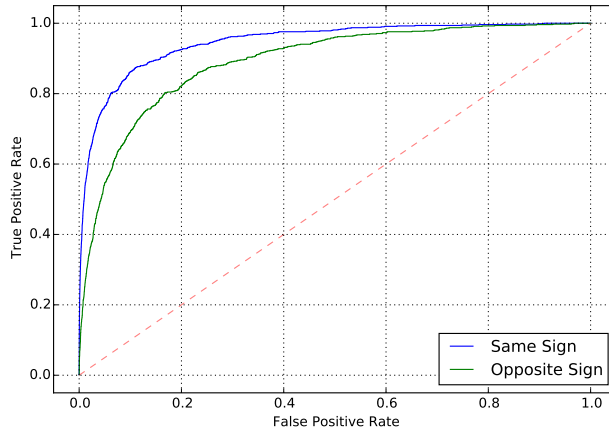
Figure 18: The Receiver Operating Characteristic curve (ROC-curve) of the BDT for SS and OS-line compared to signal. The ROC-curve shows the chance for an signal event to be classiefied as a signal (true positive event) in dependency of an background event getting classiefied as signal (false positive event).

References

[1] R. Aaij et al. Measurement of $\sigma(pp \to b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region. *Phys. Lett.*, B694:209–216, 2010.

[2] R Aaij et al. Performance of the lhcb vertex locator. *Journal of Instrumentation*, 9(09):P09007, 2014.

[3] Roel Aaij et al. Measurement of the ratio $\mathcal{B} = B^0 \to K^{*0}\mu^+\mu^-/B^0 \to K^{*0}e^+e^-$. [in preparation].

[4] Roel Aaij et al. Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 113:151601, 2014.

[5] Roel Aaij et al. Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$. *Phys. Rev. Lett.*, 115(11):111803, 2015. [Addendum: Phys. Rev. Lett.115,no.15,159901(2015)].

[6] Sean Benson. *Searching for CP Violation in the $B^0{}_s \to \phi\,\phi$ Decay at LHCb*. PhD thesis, PhD thesis, University of Edinburgh, United Kingdom, 201 4, CERN-THESIS-2014-087, 2014.

[7] Christoph Bobeth, Gudrun Hiller, and Danny van Dyk. The benefits of decays at low recoil. *Journal of High Energy Physics*, 2010(7):1–31, 2010.

[8] Cern. http://www.lhc-facts.ch/index.php?page=lhcb, 2011.

[9] LHCb Collaboration et al. Lhcb detector performance. *International Journal of Modern Physics A*, 2015.

[10] Sebastien Descotes-Genon, Joaquim Matias, and Javier Virto. Understanding the $B \to K^*\mu^+\mu^-$ Anomaly. *Phys. Rev.*, D88:074002, 2013.

[11] Gudrun Hiller and Frank Kruger. More model independent analysis of $b \to s$ processes. *Phys. Rev.*, D69:074020, 2004.

[12] Anne Keune. *Reconstruction of the Tau Lepton and the Study of $B^0 \to D\,\tau\,\nu\tau$ at LHCb*. PhD thesis, école polytechnique fédérale de Lausanne, 2012.

[13] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[15] Rafael Silva Coutinho. *Studies of charmless three-body b-hadron decays at LHCb*. PhD thesis, University of Warwick , CDS CERN-THESIS-2015-117, 2015.

[16] Andrey Tayduganov. [Intern communication, very rare decays meeting, 20 Jan 2016].

[17] Wikipedia. Standard model — Wikipedia, the free encyclopedia, 2004. [Online; accessed 03.05.2016].