

**Arbeit zur Erlangung des akademischen Grades  
Bachelor of Science**

**Selection of  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decays for  
normalisation of the  $B^0 \rightarrow D^0 \bar{D}^0$  branching  
ratio**

Ludwig Neste  
geboren in Castrop-Rauxel

2020

Lehrstuhl für Experimentelle Physik V  
Fakultät Physik  
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Bernhard Spaan  
Zweitgutachter: Prof. Dr. Kevin Kröniger  
Abgabedatum: 19. August 2020

## Abstract

In this thesis selections of the  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decay are developed. The data were taken by the LHCb detector at CERN in 2017 using proton-proton collisions at a centre-of-mass-energy  $\sqrt{s}$  of 13 TeV. Univariate selections are used to remove physical background. First, physical background coming from misidentified final state particles is removed. Afterwards, physical background coming from different decays with the same final state particles is removed. Finally, a Boosted Decision Tree classifier is trained to remove combinatorial background in a multivariate analysis. With these selections a mass peak of the  $B^0$  meson is visibly separated from background. This analysis can be used as a basis for an future analysis of  $CP$  violation in the  $B^0 \rightarrow D^0 \bar{D}^0$  decay, which will utilise the decay in this thesis as a normalisation mode.

## Kurzfassung

In dieser Arbeit werden Selektionen des  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  Zerfalls entwickelt. Die Daten wurden am LHCb Detektor am CERN bei einer Schwerpunktsenergie von 13 TeV 2017 aufgenommen. Dabei wurden Proton-Proton Kollisionen verwendet. Univariate Selektionen werden benutzt um den physikalischen Hintergrund zu entfernen. Zuerst werden dabei Hintergrundereignisse entfernt, die durch eine Falschidentifizierung von Teilchen im Endzustand entstehen. Danach werden die Hintergrundereignisse entfernt, die durch Zerfälle mit den selben Endzustandteilchen entstehen. Zuletzt wird ein Boosted-Decision-Tree-Klassifikator trainiert um in einer multivariaten Analyse den kombinatorischen Hintergrund zu entfernen. Mit den ausgewählten Ereignissen ist eine Resonanz in der Massenverteilung des  $B^0$ -Mesons zu erkennen. Diese Analyse soll eine Grundlage für eine zukünftige Analyse von  $CP$ -Verletzung im  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  Zerfall darstellen, die den Zerfall in dieser Arbeit als Normalisierungsmodus nutzen wird.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Standard Model</b>	<b>2</b>
2.1	The Particles of the Standard Model . . . . .	2
2.2	$CP$ Violation . . . . .	4
2.3	$CP$ Violation in the $B^0$ System . . . . .	5
2.4	Measuring $D^0\bar{D}^0$ . . . . .	7
<b>3</b>	<b>The LHCb Detector</b>	<b>9</b>
<b>4</b>	<b>Boosted Decision Trees</b>	<b>11</b>
<b>5</b>	<b>Physical Background Selection</b>	<b>13</b>
5.1	Preselections . . . . .	13
5.2	PID Requirements . . . . .	13
5.3	$D^*(2010)^\pm$ Veto . . . . .	16
5.4	Further Misidentification . . . . .	16
5.5	Charmless Background . . . . .	17
<b>6</b>	<b>Combinatorial Background Selection</b>	<b>19</b>
6.1	Feature Selection and Hyperparameters . . . . .	19
6.2	Training of the Boosted Decision Tree . . . . .	20
6.3	Figure of Merit Optimisation . . . . .	22
6.4	Selection Results . . . . .	23
<b>7</b>	<b>Conclusion</b>	<b>25</b>
<b>A</b>	<b>Additional Information on the Physical Background</b>	<b>26</b>
A.1	$B^0$ mass Distributions . . . . .	27
A.2	Misidentifications . . . . .	28
<b>B</b>	<b>Additional Information on the Combinatorial Background</b>	<b>30</b>
B.1	Correlation Coefficients with the $B^0$ Mass . . . . .	32
B.2	Alternative cut on BDT output . . . . .	32
B.3	Comparison to Neural Network and Random Forest . . . . .	33
	<b>Bibliography</b>	<b>35</b>
	<b>Acknowledgement</b>	<b>38</b>

# 1 Introduction

Four fundamental forces of the universe are known and three of them are described by the *Standard Model of particle physics* (SM). Only gravity cannot be quantised and is described by the general theory of relativity since 1915. The three forces described by the SM are the electromagnetic, weak and strong force. Only the weak force breaks the so called Charge-Parity-Symmetry (*CP* Symmetry) and it is the only known process, which differentiates between matter and antimatter. The current understanding of the beginning of the universe predicts that the same amount of antimatter and matter were created after the Big Bang. But, the visible matter in the universe consists mostly of matter, and almost no antimatter is present [1]. The full solution to this problem is still unknown, but *CP* violation in the SM explains a small amount of this asymmetry [2]. With precise measurement of *CP* violation, we can search for deviations from SM predictions. Such deviations could hint at new physics Beyond the Standard Model (BSM), but no such deviations have been measured with sufficient accuracy, yet.

The class of  $B^0 \rightarrow DD$  decays offers access to different parameters to measure *CP* violation, such as the CKM-angle  $\gamma$  [3].  $DD$  represents all physically possible combinations of the particles  $D^\pm, D^*, D^0, \bar{D}^0$ . These decays can also be used to indirectly measure BSM effects, for example, the SM predicts the amplitudes of different *B*-meson decays to be related by [3]

$$\mathcal{A}_{\bar{B}^0 \rightarrow D^- D^+} + \mathcal{A}_{\bar{B}^0 \rightarrow \bar{D}^0 D^0} \approx \mathcal{A}_{B^0 \rightarrow D^- D^0}.$$

Thus deviations from this relationship could show new physics. Experimentally, *CP* violation in  $B^0$  decays has been studied since the late nineties in so called *B*-factories [4, 5]. Nowadays, it is mainly studied at the Large Hadron Collider beauty (LHCb) detector at the Large Hadron Collider (LHC) in CERN and the new *B*-factory *Belle II* in Japan [6]. The high amounts of data that LHCb is producing allows to observe enough of the relatively rare  $B^0 \rightarrow DD$  events to study *CP* violation and indirectly search for BSM contributions.

The  $B^0 \rightarrow D^0 \bar{D}^0$  decay has not been studied very well, and was only studied by the LHCb collaboration using data produced in 2011 with  $\sqrt{s} = 7$  TeV. There was a hint of signal with  $2.4\sigma$  [7, p. 7] and a branching ratio of  $(1.4 \pm 0.7) \times 10^{-5}$  [8]. This thesis lays the foundation for a future analysis of LHCb Run II data that will study *CP* violation in  $B^0 \rightarrow D^0 \bar{D}^0$  decays. To increase the precision of the branching fraction of  $B^0 \rightarrow D^0 \bar{D}^0$  decays, it will be measured relative to the branching fraction of  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decays, which will be analysed in this thesis. Because of the high amount of background, multiple techniques of data analysis are used to separate signal data from background data. The strategy of this thesis is to start by applying univariate selections to remove physical background in the  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  data. Afterwards, statistical learning techniques are used to distinguish signal events from combinatorial background with a *Boosted Decision Tree* (BDT) classifier. With these selections the  $B^0$  mass peak is separated from background.

## 2 The Standard Model

The Standard Model of particle physics is a theory developed in the second half of the 20th century, describing the interactions between all known particles. These are six quarks and six leptons, together they build the elementary fermions, which make up matter. Then, there are five elementary bosons, which can be thought of as carriers of the three forces which govern the interaction of the leptons. The three forces are electromagnetism (conveyed by the photon), the weak force (conveyed by the  $Z$  and  $W^\pm$  bosons) and the strong force (conveyed by the gluon). Gravity, the last of the four known fundamental forces, is not described by the SM. To explain the mass of elementary particles, especially the mass of the  $Z$  and  $W^\pm$  bosons, another mechanism was introduced in 1964 by Peter Higgs [9] and also independently by Robert Brout and François Englert at the same time [10], the so called *Brout-Englert-Higgs mechanism*.<sup>1</sup> The Higgs mechanism predicted a new fundamental field, and as all quantum fields in quantum field theories, there is a particle associated with it: the Higgs boson. The theory we call Standard Model today consists of two parts. The first is the Glashow-Weinberg-Salam model, developed in the sixties, which describes the unification of electromagnetism with the weak force (electroweak unification) [11–13]. The second part is Quantum-Chromo-Dynamics (QCD), which was developed throughout the sixties and seventies, describing the strong force [14–18].

The SM is one of the most successful and accurate theories of physics to date. For example, the measured anomalous magnetic dipole moment of the electron  $(g - 2)/2$  is in agreement with the best SM calculation within 12 decimal places [19, 20]. Despite the huge success of the Standard Model, it cannot be considered complete for a variety of reasons. For example, the measured anomalous magnetic dipole moment of the muon disagrees with SM predictions by about  $3\sigma$  [8], it assumes massless neutrinos, which is disproved experimentally and it does not account for dark matter.

This chapter is based on refs. [21] and [22].

### 2.1 The Particles of the Standard Model

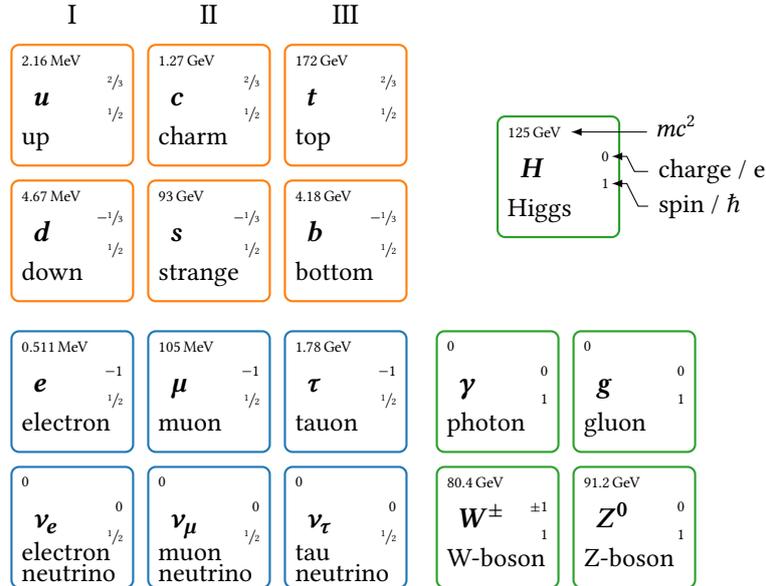
The six fundamental leptons are shown in Figure 2.1 in blue. Each one also has an anti-particle, with the same but opposite charge. They are structured in three *families*, with each family consisting of one massive lepton and its massless neutrino. All SM processes conserve the lepton family number  $L_i$ , which is defined as

$$L_i = n_i - n_{\bar{i}} + n_{\nu_i} - n_{\bar{\nu}_i}$$

---

<sup>1</sup>It is important to note that most mass in everyday life comes from binding energy, for example inside a proton. It is not coming from the fundamental particles.

with  $i = e, \mu, \tau$ . Only non-SM processes like neutrino oscillation break this conservation law. When reaching energies where all leptons can be approximated to have a very low mass, all three families are expected to behave the same, this is known as lepton flavour universality, which is tested by the LHCb experiment [23]. Leptons interact with the electromagnetic and weak force, but not with the strong force.



**Figure 2.1:** The particles of the Standard Model. The quarks are marked orange, the leptons blue and the bosons green. The Values are taken from Ref. [8] and the mass is rounded to three decimal points.

The six quarks in the Standard Model can also be categorised in three families, as seen in Figure 2.1 (orange colour). The important difference to the leptons is that the weak force can change the family type of quarks and thus, the quark family number is no conserved property. Quarks with a charge of  $\frac{2}{3}e$  are called up-type quarks, while quarks with a charge of  $-\frac{1}{3}e$  are called down-type quarks. The quarks interact with both the strong and the electroweak force. Their interaction, as described by QCD, has two important properties: the first being *confinement*, meaning that quarks cannot be observed individually, like all other SM particles. QCD introduces a new type of charge, the *colour charge*. Each (anti) quark carries an (anti) colour. The confinement can be understood by looking at the colour of QCD states. The charge can either be (anti)-red ( $r/\bar{r}$ ), (anti)-green ( $g/\bar{g}$ ) or (anti)-blue ( $b/\bar{b}$ ), and the only other strongly interacting particle, the gluon, carries both a colour and an anti-colour. In QCD, only bound states where the colour adds up to *white* (e.g.  $r + \bar{r}$  or  $r + g + b$ ) can be observed, which is another way to describe confinement. Quarks are always bound in hadrons and hadrons are further divided into mesons and baryons. The former consist of a quark and an anti-quark, which can have for example a red and an anti-red colour, resulting in an overall vanishing colour, or more generally any even number of quarks. Baryons, on the other hand contain three quarks,

acquiring their *white* colour from the combination  $r + g + b$ , which can again be generalised into any number of odd quarks. Recently, the LHCb collaboration is discovering different states of these higher number tetra- and pentaquarks [24–26].

The second important property of QCD is called *asymptotic freedom*, which means that the coupling strength between quarks decreases as the energy scale increases. This results in quarks asymptotically behaving like free particles when their energy increases.

The four different elementary bosons of the SM are listed in Figure 2.1 in green. As already mentioned, they transmit the forces described in the SM. While there is only one type of photon, the gluon comes in 8 different types, which contain a combination of two colours (one colour and one anticolour). With three different colours, one would expect there to be nine different gluons, instead of eight. But since the symmetry group of QCD is  $SU(3)$  instead of  $U(3)$ , only eight different gluons exist, with the colourless gluon being forbidden. The  $W^\pm$  and the  $Z$  bosons carry the weak force.

## 2.2 CP Violation

The *CPT* theorem states that physical processes do not change under the combined application of time reversal  $t \rightarrow -t$  ( $T$ ), charge conjugation  $q \rightarrow -q$  ( $C$ ) and spatial inversion  $\vec{x} \rightarrow -\vec{x}$  ( $P$ ). It is essential for all quantum field theories that the *CPT* theorem holds true [27]. The individual transformations were believed to hold true even separately, which was disproved by Chien-Shiung Wu in 1956 [28]. Even the combined transformation  $C \cdot P$  is no symmetry, as discovered in  $K$ -meson systems in 1964 in the Cronin-Fitch-Experiment [29].

The relevance of  $b$  physics arises from *CP* violation in  $B$  meson decays. In the Standard Model *CP* violation originates from a complex phase  $\delta$  in the *CKM matrix* [30, 31]. The CKM mechanism is based on the observation that the  $W$  and  $Z$ -boson couple with different strengths to the different quarks. To have a theory of only one underlying interaction, the quark fields in the weak eigenstates are assumed to be different quark fields than the ones in the flavour eigenstates. The linear transformation between the weak eigenstates ( $d', s', b'$ ) and the flavour eigenstates ( $d, s, b$ ) of the down type quarks <sup>2</sup> is given by the CKM matrix  $V_{\text{CKM}}$  as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \underbrace{\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}}_{V_{\text{CKM}}} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.1)$$

Where  $|V_{ij}|^2$  is the probability of an  $i$ -quark to transition into an  $j$ -quark when interacting with a  $W$ -boson. The CKM matrix is named after its inventors Cabibbo, Kobayashi and Maskawa. Its elements are generally complex, meaning there are  $3 \cdot 3 \cdot 2 = 18$  parameters to be determined. But because of its unitarity

$$\mathbf{V}\mathbf{V}^\dagger = \mathbf{1} \quad (2.2)$$

<sup>2</sup>To transform the down-type quark is only convention, the up-type quarks could be transformed as well.

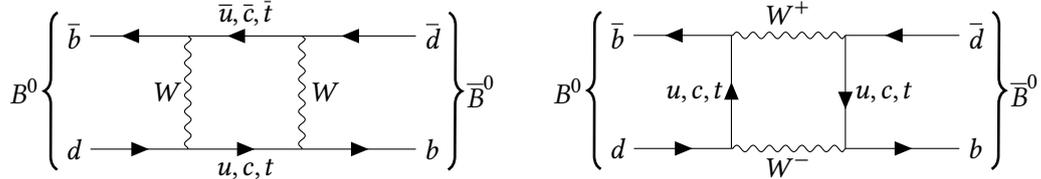
nine additional equations must be fulfilled, which leaves only nine parameters to be determined. With six quarks their fields can have five relative phases, which can be defined into the quark fields, with no physical meaning. This results in the fact that the CKM matrix can be parameterised with four parameters. This also means that it is not a rotation matrix in three dimensions, because such a matrix only has three free parameters. The additional parameter is a phase  $\delta$  that makes some of the CKM matrix elements complex, which is the cause for  $CP$  violation in the Standard Model. One common representation of the CKM matrix is the Wolfenstein expansion, which reflects the size of the different elements [8]:

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \quad (2.3)$$

With  $\lambda \approx 0.2 < 1$  [8], this representation immediately shows that quarks are most likely to transition in their own family, because of the big diagonal elements. Transitions to a neighbour family is of the order  $\lambda$  (Cabibbo suppressed) and to skip one family is even more unlikely and proportional to  $\lambda^3$  (doubly Cabibbo suppressed). With this representation  $CP$  violations means that  $|\eta| > 0$ , because  $\eta$  is proportional to  $\sin(\delta)$  and  $V_{\text{CKM}}$  becomes complex for non vanishing  $\eta$ .

### 2.3 $CP$ Violation in the $B^0$ System

The quark content of the neutral  $B$  meson is down anti-bottom ( $d\bar{b}$ ), and anti-down bottom ( $\bar{d}b$ ) for its anti-particle. Like the  $K^0$  mesons in the Cronin-Fitch experiment mentioned above,  $B^0$  can oscillate into  $\bar{B}^0$  via the box diagrams shown in Figure 2.2. The wave function of this



**Figure 2.2:** The dominant box diagrams for  $B^0$  oscillation. The diagram with the top quark is the dominant one.

system is a superposition of the  $B$  and  $\bar{B}$  states. A phenomenological Hamiltonian  $H$  describing the oscillation and decay of the mixing in the flavour eigenstates is given by

$$H = M - \frac{i}{2}\Gamma = \begin{pmatrix} m - i/2\Gamma & M_{12} - i/2\Gamma_{12} \\ M_{12}^* - i/2\Gamma_{12}^* & m - i/2\Gamma \end{pmatrix}, \quad (2.4)$$

which has to be applied to the flavour eigenstates. It resembles the Hamiltonian  $H = m - i/2\Gamma$  of a decaying particle with mass  $m$  and the decay-width  $\Gamma$ . The diagonal has to be a multiple of

the unit matrix, because that describes the evolution of the flavour eigenstates without mixing, and for the *CPT* theorem to hold true, the decay widths and masses of  $B^0$  and  $\bar{B}^0$  have to be the same. The eigenvectors of this matrix are the mass eigenstates and can be expressed as a superposition of the flavour eigenstates

$$|B_{H,L}\rangle = p|B^0\rangle \pm q|\bar{B}^0\rangle. \quad (2.5)$$

Where  $H$  stands for the heavier eigenstate (+) and  $L$  for the lighter one (-). This allows to define the mass  $m_{H,L}$  and decay width  $\Gamma_{H,L}$  of the two eigenstates as the real and (two times) the imaginary part of the eigenvalues of (2.4). The eigenvalues are

$$\lambda_{H,L} = m - i/2\Gamma \pm \sqrt{(M_{12} - i/2\Gamma_{12})(M_{12}^* - i/2\Gamma_{12}^*)}. \quad (2.6)$$

With these the relation

$$\left(\frac{q}{p}\right)^2 = \frac{M_{12}^* - i/2\Gamma_{12}^*}{M_{12} - i/2\Gamma_{12}} \quad (2.7)$$

is derived. The time evolution of the eigenstates is calculated with the Schrödinger equation

$$i\frac{d}{dt} \begin{pmatrix} |B^0\rangle \\ |\bar{B}^0\rangle \end{pmatrix} = \mathbf{H} \begin{pmatrix} |B^0\rangle \\ |\bar{B}^0\rangle \end{pmatrix} \quad (2.8)$$

with the Hamiltonian (2.4). With this the time evolution of a state which was initially  $|B^0\rangle$  is given by

$$|B^0(t)\rangle = \frac{1}{2} \left( e^{-it(m_H - i/2\Gamma_H)} + e^{-it(m_L - i/2\Gamma_L)} \right) |B^0\rangle + \frac{1}{2} \left( e^{-it(m_H - i/2\Gamma_H)} - e^{-it(m_L - i/2\Gamma_L)} \right) \frac{q}{p} |\bar{B}^0\rangle. \quad (2.9)$$

In equation (2.9) and (2.8)  $\hbar = c = 1$  is used. If  $q/p \neq 1$  it is called *CP* violation in mixing. Two further different types of *CP* violation are distinguished: *CP* violation in decay, also called direct *CP* violation, and *CP* violation in the interference of decay and mixing. Direct *CP* violation occurs if the amplitudes of  $B^0$  and  $\bar{B}^0$  to decay into a specific final state  $f$  and  $\bar{f}$  are different, which is quantified by the ratio

$$\rho_f = \frac{\langle \bar{f} | H | \bar{B}^0 \rangle}{\langle f | H | B^0 \rangle} = \frac{\bar{A}_{\bar{f}}}{A_f}. \quad (2.10)$$

The third type occurs if mixing and decay *CP* violation interfere, which is quantified by

$$\lambda_f = \frac{q \langle \bar{f} | H | \bar{B}^0 \rangle}{p \langle f | H | B^0 \rangle} = \frac{q \bar{A}_{\bar{f}}}{p A_f}. \quad (2.11)$$

This can occur, even if indirect and direct *CP* violation are not present on their own.

In case of the oscillation seen in Figure 2.2, the matrix elements are approximately proportional[32] to

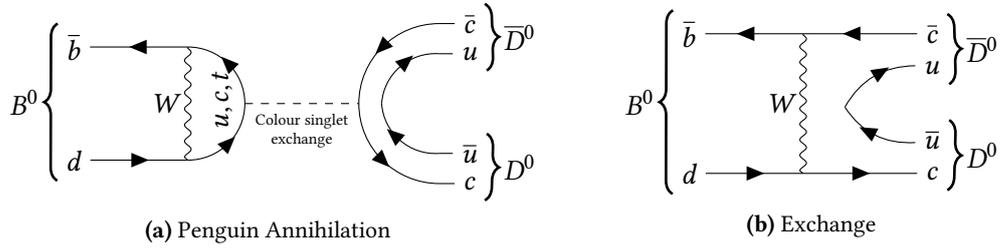
$$M_{12} \propto (V_{tb}^* V_{td})^2. \quad (2.12)$$

Furthermore,  $\Gamma_{12}$  is magnitudes smaller than  $M_{12}$  [32] and can thus be neglected in equation 2.7, which then gives

$$\frac{q}{p} \approx \sqrt{\frac{M_{12}^*}{M_{12}}} = \frac{V_{ib}^* V_{td}}{V_{ib} V_{td}^*}. \quad (2.13)$$

With this equation it is easy to see that  $q/p$  is equal to one, if the elements of the CKM matrix are not complex, which is only the case if  $\delta = 0$  or in the Wolfenstein expansion (2.3)  $\eta = 0$ .

## 2.4 Measuring $D^0 \bar{D}^0$



**Figure 2.3:** The mainly contributing Feynman diagrams for  $B^0 \rightarrow D^0 \bar{D}^0$ .

In Figure 2.3 Feynman diagrams for the  $B^0 \rightarrow D^0 \bar{D}^0$  decay are seen. Both involve the matrix elements  $V_{di}$  and  $V_{bi}$  with  $i \in \{u, c, t\}$ . Using the unitarity relation (2.2) in the third row and first column gives the equation

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0.$$

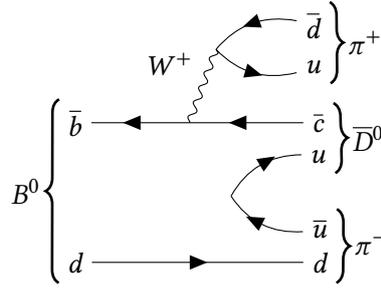
When each term in this equation is thought of as a separate complex number, it represents a triangle in the complex plane. With this unitarity triangle for the  $B^0$  meson,  $CP$  violation effects are studied. In experiment, the angles of this triangle can be determined by measuring different  $B$  decays into hadrons, which means this is a good way to find BSM effects if the angles do not add up to  $180^\circ$ .

As the title of this thesis suggests the  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decay studied in this thesis will be used to normalise the branching ratio of  $B^0 \rightarrow D^0 \bar{D}^0$ . The primarily contributing Feynman diagram for the  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decay is shown in Figure 2.4.

The  $B^0 \rightarrow D^0 \bar{D}^0$  branching ratio can be calculated with

$$\mathcal{BR}(B^0 \rightarrow D^0 \bar{D}^0) = \frac{N(B^0 \rightarrow D^0 \bar{D}^0)}{2 \mathcal{L}_{\text{int}} \sigma_{b\bar{b}} f_d \varepsilon_{D^0 \bar{D}^0}}. \quad (2.14)$$

Where  $N(B^0 \rightarrow D^0 \bar{D}^0)$  is the number of events we see in the analysis (yields),  $\sigma_{b\bar{b}}$  is the cross-section for  $b\bar{b}$  production,  $f_d$  is the probability of an  $b$  quark hadronising with a  $d$  quark into a  $B^0$  meson (fragmentation fraction) and  $\varepsilon_{D^0 \bar{D}^0}$  is the efficiency of the selection. The 2 in the



**Figure 2.4:** The leading order Feynman Diagram for  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$ . Connections in the  $u\bar{u}$  creation are not specified, production by a gluon will be dominant there.

denominator must not be used if the analysis distinguishes  $B^0$  and  $\bar{B}^0$ , which is done by flavour-tagging. In this thesis  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  and  $\bar{B}^0 \rightarrow D^0 \pi^+ \pi^-$  are not separated, which results in a factor of 2 in the denominator. The efficiency is defined as

$$\varepsilon = \frac{N_S^{\text{after selection}}}{N_S^{\text{before selection}}} \quad (2.15)$$

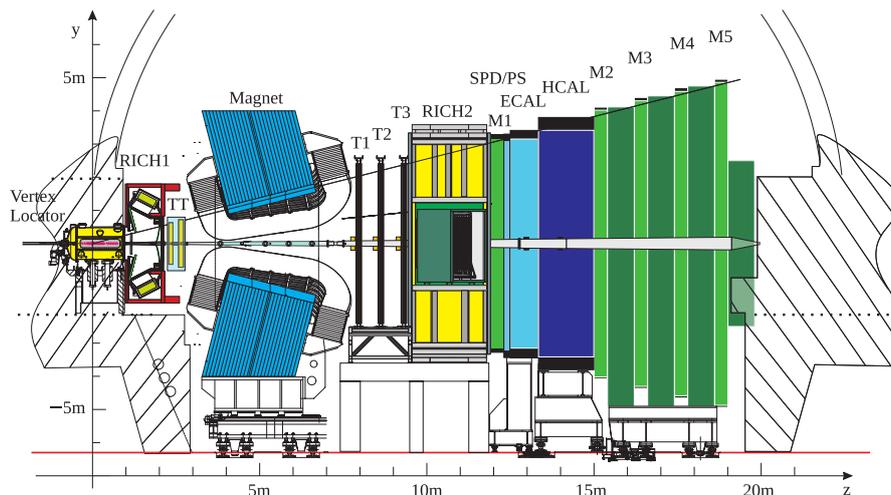
where  $N_S$  are the signal yields, so the ratio can be estimated from *Monte Carlo* (MC) simulations. The problem with equation (2.14) lies in  $\sigma_{b\bar{b}}$  and  $f_d$ , because both constants are only known with a high uncertainty [33]. For example  $\sigma_{b\bar{b}}$  is only known with an relative uncertainty of about 15 % [34]. The problem is solved by dividing equation (2.14) with the same expression for another  $B^0$  decay mode, such as  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$ :

$$\frac{\mathcal{BR}(B^0 \rightarrow D^0 \bar{D}^0)}{\mathcal{BR}(B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-)} = \frac{N(B^0 \rightarrow D^0 \bar{D}^0)}{N(B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-)} \frac{\varepsilon_{\pi^+ \pi^- \bar{D}^0}}{\varepsilon_{D^0 \bar{D}^0}} \frac{2\mathcal{L}_{\text{int}} \sigma_{b\bar{b}} f_d}{2\mathcal{L}_{\text{int}} \sigma_{b\bar{b}} f_d}. \quad (2.16)$$

The last (coloured) part of the equation cancels out and thus reduces the uncertainty of the result. The  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decay is chosen because of its similarity to the  $B^0 \rightarrow D^0 \bar{D}^0$  decay. It has the same number of final state particles and among those it decays has a relatively high branching ratio. A previous measurement of the  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  branching ratio can be used to calculate the  $B^0 \rightarrow D^0 \bar{D}^0$  branching ratio out of this result.

### 3 The LHCb Detector

The Large Hadron Collider (LHC) near Geneva, Switzerland is the worlds most powerful particle accelerator [35], i.e. no other collider in the world has reached its centre-of-mass energy ( $\sqrt{s}$ ) of 13 TeV. Proton beams are directed to collide at four points around the LHC, where detectors can observe the collision of two protons to reveal the structure of fundamental particles. There are a number of different detectors, which are able to observe different phenomena due to their different design. At the LHC  $b$ -quarks are mainly produced into the direction of the beamline. To study the effects of  $CP$  violation in  $b$ -mesons systems discussed in section 2.3 at the LHC, the Large Hadron Collider beauty (LHCb) detector, depicted in Figure 3.1, is used. It is, for this purpose, designed asymmetrical into the forward direction [36]. This is different to the other detectors at the LHC: while ATLAS, CMS and ALICE all try to cover all directions ( $4\pi$  detector) [37–39], the LHCb detector is sensitive in the pseudorapidity region of roughly  $2 < \eta < 5$ . This is equivalent to an angle between about  $0.8^\circ$  and about  $15.4^\circ$  measured to the forward direction of the beam-pipe ( $z$ -direction in Figure 3.1)[40]. The LHCb records at a luminosity of  $10^{32} \text{ cm}^{-2} \text{ s}^{-1}$  [40]. The following description of the LHCb detector shown in Figure 3.1 is based on Refs. [36, 41, 42] and [40].



**Figure 3.1:** Schematic drawing of the LHCb detector taken from Ref. [43].

The interaction region of the colliding proton beams is surrounded by the vertex locator (VELO), which allows to precisely measure track coordinates close to the interaction region. The VELO is a silicon strip detector consisting of multiple modules, each providing a measurement in polar coordinates  $r$  and  $\varphi$  of the track coordinates. The VELO can be used to reconstruct the proton-proton interaction point (Primary Vertex or PV) as well as the decay location of the daughter particles (Secondary Vertex or SV). As bottom and charm mesons typically travel a few centimeters before they decay, the VELO is crucial to measure this displacement of the

primary and secondary vertex. The other trackers of LHCb are the Tracker Turicensis (TT), located before LHCb's magnet, and T1, T2, T3, located downstream of LHCb's magnet. TT is a pure silicon tracker, while T1 to T3 are divided into inner tracker (IT) and outer tracker (OT). The ITs are also silicon trackers, but the OTs relies on gas-filled tubes for detection. LHCb's magnet is a warm magnet, i.e. it does not rely on superconductivity and must not be cooled to very low temperatures. The magnet curves the path of charged particles and the deflection is seen in the track before and after the magnet. This allows to calculate the momentum of the particle.

Aside from the tracking system there are also the two ring imaging Cherenkov detectors RICH1 and RICH2. In Run 2 both are filled with gases ( $C_4F_{10}$  and  $CF_4$  respectively). Particles entering the RICH detectors with a speed greater than the speed of light in the medium emit Cherenkov radiation. The angle of the emitted radiation depends on the refraction index and the speed of the particle. This measurement can be combined with the momentum measurement previously mentioned to calculate the mass of the particle. This allows a mass hypothesis to be assigned to the particle.

The two calorimeters can distinguish certain particles, measure their energy and also their position. In the electromagnetic calorimeter (ECAL), photons and electrons/positrons deposit all their energy by producing showers, which are detected with scintillators. Hadrons on the other hand lose only a little bit of their energy in the ECAL and deposit most into the hadronic calorimeter (HCAL), where they also produce showers. Most of the hadrons living long enough to make it to the calorimeters are pions, kaons and protons. The last components are the five muon stations M1-M5, used to identify muons.

With this many components and a bunch crossing rate of 40 MHz, it is impossible to save all recorded raw data. The process of deciding which events to save and which to throw away is called triggering. The events first enter the hardware stage (L0), where simple restrictions based on direct readouts of the detector components are made. Some detector components are more important for that than others, for example the calorimeter system is designed to make a trigger decision in just 4  $\mu$ s. This is mainly based on the transverse momentum measurement, which is improved by the first muon system, installed before the calorimeters for this reason. After the L0 trigger the amount of data is reduced to 1 MHz, and then enters the software based high level trigger (HLT). The HLT is split into two stages (HLT1 and HLT2), which both apply basic reconstruction and fits. All events are buffered into the event filter farm (EFF), which is a computing cluster that can save up 10 PB of data (about two weeks of LHCb data). As the LHCb has some smaller regular downtimes, for example when changing the magnet polarity, this buffering helps to be able to analyse more data than possible otherwise. HLT1 does a basic event reconstruction with VELO and T1-T3 data, and for example checks the high transverse momentum decision of L0. With the much decreased signal rate HLT2 is able to apply very high level reconstruction with all data available from every component, such as PID selections. The resulting data is saved at a rate of about 12.5 kHz, which means about 0.6 GB/s are saved to disk. The events are categorised additionally to allow a good selection of events for further analysis. This is called *stripping*.

## 4 Boosted Decision Trees

In this thesis background and signal events need to be separated. A classifier assigns a given input event, described by multiple different variables (features), to one of two classes: Background ( $C_B$ ) or Signal ( $C_S$ ). Such a classifier can be thought of as a function  $f_{\vec{\theta}} : \mathbb{R}^N \rightarrow \{C_B, C_S\}$ , which depends on a number of different parameters  $\vec{\theta}$ , and maps the input vector of  $N$  features of one event to the class label. The process of finding the best parameters  $\vec{\theta}$  for the problem at hand is called *learning*. Input data already assigned to a class (labelled data) is needed to be able to learn for a given task. To define what is a good classification for a given problem, a Loss function  $L(y_k, f(\vec{x}_k))$  is introduced, which measures the performance of the classification of the training event  $k$  with features  $\vec{x}_k$  and label  $y_k$ . The sum  $\sum_{k=0}^K L(y_k, f(\vec{x}_k))$  of the  $K$  learning events measures the overall performance and needs to be minimised (smaller values of  $L$  mean a better performance). Minimising the loss function is often not possible globally, so mostly greedy algorithms are applied to find a local minimum. There are many different models for such a classifier and in this thesis *Boosted Decision Trees* are used.

Instead of discrete classes a continuous output can be used. For example a two class problem can be transformed to a model with continuous output:  $f : \mathbb{R}^N \rightarrow [0, 1]$ . The assigned class of an event  $\vec{x}$  can then be predicted to be

$$C(\vec{x}) = \begin{cases} C_B & \text{if } f(\vec{x}) < c \\ C_S & \text{else} \end{cases} \quad (4.1)$$

for a cut  $c \in [0, 1]$ , which can be chosen with a Figure of Merit (see section 6.3). A decision tree  $T$  can be formally described by splitting the input space of the model into  $J$  disjunct regions  $R_j$ <sup>1</sup> and assigning a prediction  $\gamma_j$  to each region:

$$T(\vec{x}; \vec{R}, \vec{\gamma}) = \sum_{j=0}^J \gamma_j I(\vec{x} \in R_j), \quad (4.2)$$

where  $I$  is the indicator function. More intuitively they can be thought of as an iterative model, which cuts one input feature into two regions to decide to which sub-tree the event is then handed or, if there are no sub-trees left, to which class they belong. For example a simple tree of depth one can give all events which have “ $m < 5000$ ”, to the right sub-tree and all others to the left sub-tree. The right sub-tree then decides on a class by for example assigning every event with “ $E < 1000$ ” to the background class and every other event to the signal class. When tree models are built (learned) at each step the feature which splits the training data best under some criterion (for example information gain or gini-coefficient [44, p. 309]) is chosen and then the same thing is done for the sub-trees, until a given depth is reached. By remembering

---

<sup>1</sup>To be exact: we would not call all possible shape of regions a decision tree, but every decision tree has this representation.

the decision made on the feature to use in each step, a tree can also produce a metric for the importance of each feature (feature importance), for example how often a feature is used to make a cut.

A model which is only slightly better than a random prediction, is called a *weak learner*. It turns out to be a very powerful model when a number of weak learners are combined into a *committee* of models. When the combination is to average the prediction, it is called *bagging*. Boosting takes this idea one step further and starts with one weak learner and in each learning round adds another weak learner in such a way, that it increases the overall prediction [44]. This can be done in different ways, for example by reweighting falsely classified training data from the previous rounds, as done by the *AdaBoost* algorithm, which was the first boosting algorithm introduced in 1993 [44].

The boosting method used in this thesis is *gradient boosting*. It starts with a *constant* function  $f_0(\vec{x}) = \gamma_0$  as a prediction, which minimises the overall loss function  $\sum_{k=1}^K L(y_k, \gamma)$ . In each step  $m$  (*learning round*) it calculates the gradient

$$r_{mi} = \frac{\partial L(y_i, f_{m-1}(\vec{x}_i))}{\partial f_{m-1}(\vec{x}_i)}$$

with respect to the prediction of each training event of the previous round  $f_{m-1}$  of the loss function. This gradient is then fitted with a weak learner  $T_m$ . Comparable to the gradient descent method known from numerical minimisation of functions, the fitted gradient is then subtracted from the model of the previous round  $f_{m-1}$  with a weight  $\nu\gamma_m$ :

$$f_m(\vec{x}) = f_{m-1}(\vec{x}) - \nu\gamma_m T_m(\vec{x}).$$

Here  $\gamma_m$  is the step-width  $\gamma$  that minimises the loss function

$$\sum_{k=1}^K L(\vec{y}_k, f_{m-1}(\vec{x}_k) - \gamma T_m(\vec{x}_k))$$

and  $\nu \in (0, 1)$  is the learning rate. A learning rate  $\nu \neq 1$  helps the model to become more stable, but also increases the number of learning rounds needed to get an overall good learner. The iteration can be stopped at any step  $m = M$  to get the final classifier  $f_M$ . This is a general algorithm which can be applied with all kind of weak predictors. Usually for decision trees not the whole model (4.2) is trained in each step, but only the regions  $R_j$  are determined and the calculation of the constant  $\gamma_m$  is replaced by optimising each  $\gamma_i$  in (4.2) to minimise the loss function.

## 5 Physical Background Selection

This analysis does not differentiate between  $B^0$  and  $\bar{B}^0$ , so all particles and modes which are expressed in terms of  $B^0$  also include their respective charge conjugated modes and particles. In the same way  $D^0$  is also synonymous with  $\bar{D}^0$ . The  $\bar{D}^0$  in these decays is observed via the  $\bar{D}^0 \rightarrow K^\pm \pi^\mp$ ,  $\bar{D}^0 \rightarrow K^+ K^-$  and  $\bar{D}^0 \rightarrow \pi^+ \pi^-$  final states in this thesis. Among those modes, the  $K^+ \pi^-$  channel is the dominant one, with a branching ratio about ten times as high as that of the  $K^+ K^-$  channel. The analysed data was collected by LHCb in 2017, with experiment wide stripping selections applied. The stripping line can be found in Appendix A. The data includes both magnet directions.

Of course not all the stripped events contain the decays they were labelled with, they also contain high amounts of background. Two types of background events are distinguished: combinatorial background and physical background. Combinatorial background is the result of multiple random particle tracks that are reconstructed, when they do not come out of a physical decay. Physical background includes a real decay, which was falsely or only partially reconstructed. It is important to keep in mind that one physical event can appear multiple times in the dataset, as it can be labelled multiple times as different decays by the LHCb software. Physical background is removed with a number of specific cuts on variables described in this chapter, as it has a specific signature due to their real decay origin. Combinatorial background does not have this specific signature and has a flat distribution. That is why it will be removed in a multivariate analysis (MVA) in the next chapter using a BDT.

### 5.1 Preselections

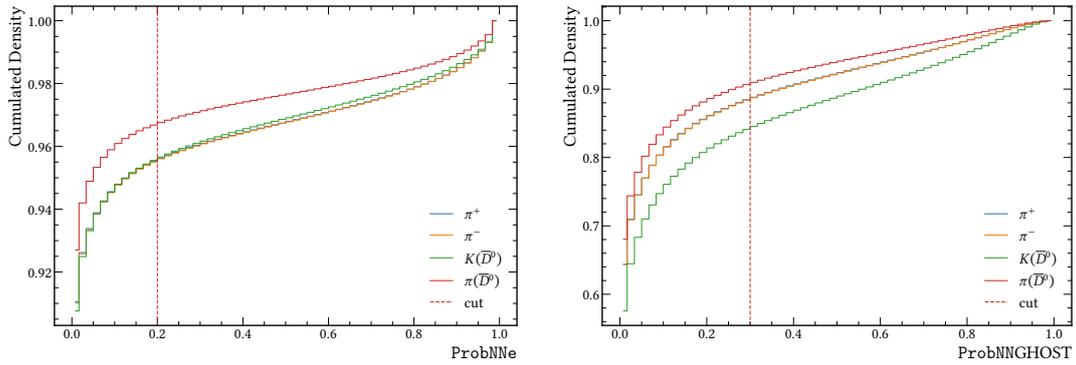
The first selection step aims to remove only the most falsely reconstructed events. It is required the decay should not produce any muons registered in the muon chambers. The vertex fits of the  $B^0$  and  $D^0$  are tested with a  $\chi^2$ -goodness of fit by the LHCb software. The  $\chi^2$  test statistic divided by the degrees of freedom of the fit, which has an expectation value of 1, are required to be less than four.

Particles are additionally reconstructed using the DecayTreeFitter (DTF), which uses extra information to constrain the reconstruction, e.g. the  $D^0$  PDG mass. It is required that the DTF fits with constrains on the PV and the  $D^0$  mass are successful. In addition, the  $B^0$  mass calculated from the DTF has to be greater than  $5000 \text{ MeV}/c^2$ .

### 5.2 PID Requirements

Kaons can be misidentified as pions and vice-versa. This means a wrong mass-hypothesis is applied to the particles, and the resulting mass distribution of the mother particle peaks

at a different location than expected. This misidentification can be suppressed with cuts on the ProbNNX variables. They are generated by the LHCb software and are a measure for the probability that the considered particle is a particle  $X$ , generated by a neural network.  $X$  can be either  $e$  for electron,  $p$  for proton,  $K$  for kaon,  $\pi$  for pion or the special one ghost. *Ghost* refers to a fake track, i.e. a track reconstructed not from a real physical particle, but multiple separate particles, or detector noise [45]. The probability of being either a ghost or electron is expected to be mostly independent of being a hadron. That is why independent cuts on these two variables are used.  $\text{ProbNN}_e < 0.2$  and  $\text{ProbNN}_{\text{ghost}} < 0.3$  is required on all four final state particles. The distribution of these variables can be seen in Figure 5.1 and with them it can be concluded, that the chosen cuts of 0.2 and 0.3 are relatively loose.

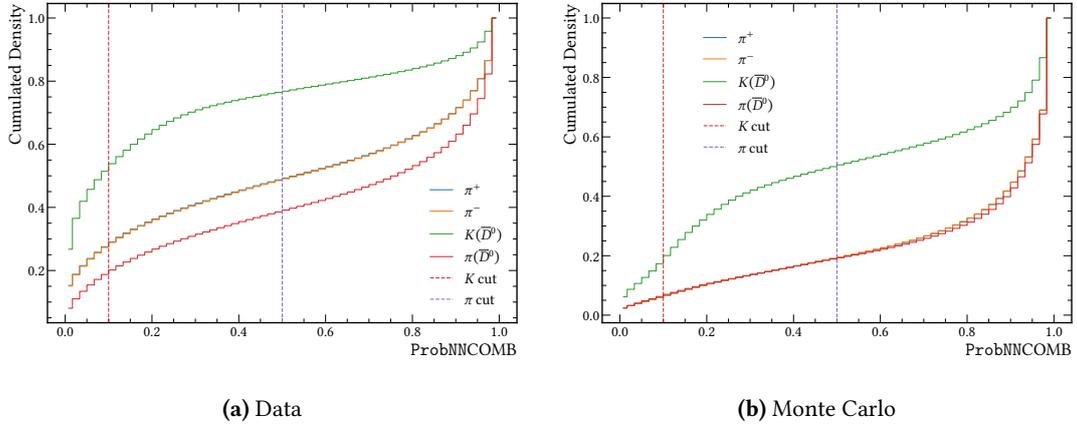


**Figure 5.1:** The distribution of  $\text{ProbNN}_e$  (left) and  $\text{ProbNN}_{\text{ghost}}$  (right) on data. The histograms are cumulated and normed to 1. One should note the different scales on the y axis.

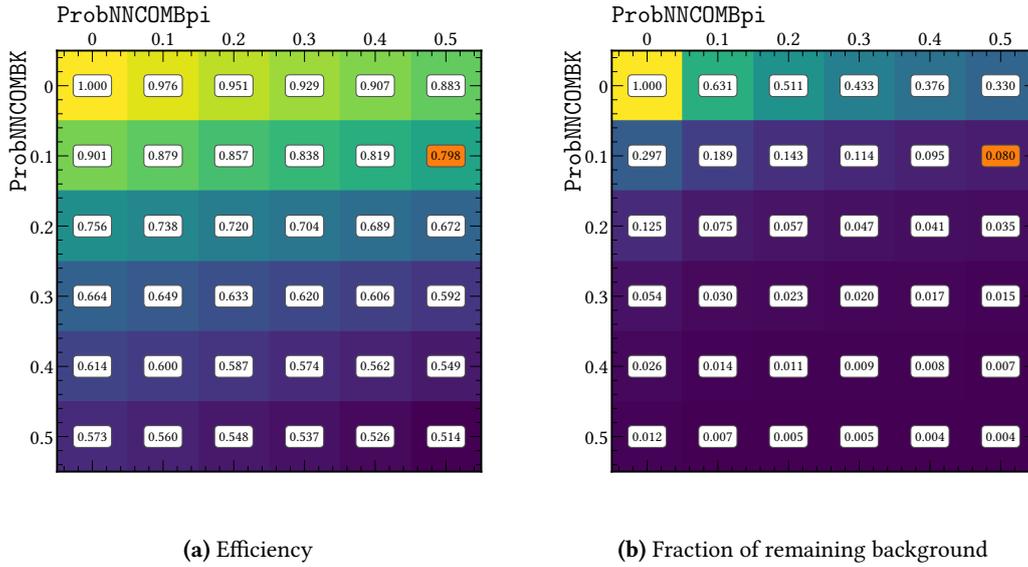
It is common to combine the ProbNNs for the hadrons to form a new variable

$$\text{ProbNNCOMB}(\pi, K) = \text{ProbNN}(\pi, K)(1 - \text{ProbNN}(K, \pi))(1 - \text{ProbNN}_p). \quad (5.1)$$

This can be thought of as a measure of how certain the assigned hadron particle type is for a final state particle. Their distribution on data and Monte Carlo samples in the  $\bar{D}^0 \rightarrow K\pi$  channel is shown in Figure 5.2. The distribution for the kaon is notably different than for the pions. For all the pions in the final state  $\text{ProbNNCOMB}_{\pi} > 0.5$  is required, where for all the kaons in the final state  $\text{ProbNNCOMB}_K > 0.1$  is used. These cuts are chosen with the help of Monte Carlo samples: As seen in Figure 5.3 they keep a reasonably good signal efficiency, while removing most of the misidentified background on Monte Carlo samples.



**Figure 5.2:** The ProbNN distribution for (a) data and (b) Monte Carlo simulations in the  $\bar{D}^0 \rightarrow K^\mp \pi^\pm$  channel. The histograms are cumulated and normed to one, so they represent the probability function.



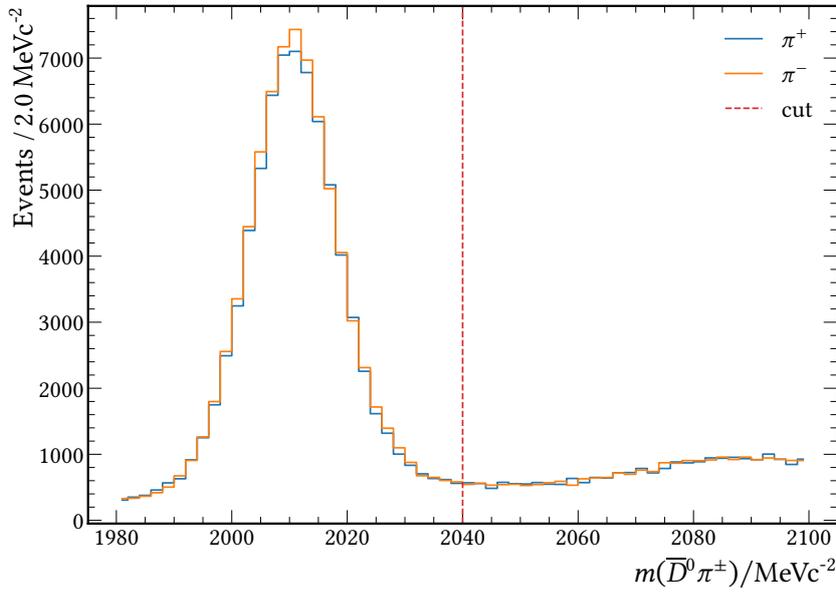
**Figure 5.3:** Effect of the PID cut on the  $\bar{D}^0 \rightarrow K\pi$  final state particles on Monte Carlo samples. On the horizontal axis the cut on the ProbNNCOMBpi variable is varied, while on the vertical axis the cut on ProbNNK is varied. It is concluded that ProbNNpi > 0.5 and ProbNNK > 0.1 remove enough background (= particle ID does not match MC truth) (92.0 %) while keeping a reasonable efficiency (79.8 %). The cells filled with orange colour are at the chosen cut.

### 5.3 $D^*(2010)^\pm$ Veto

Real decays like  $B^0 \rightarrow D^*(2010)^- \pi^+ X \rightarrow D^0 \pi^+ \pi^- X^1$  (and charge conjugate) can be mistaken for the signal decay. The value of

$$m(D^*(2010)^\pm) = (p_{\bar{D}^0}^\mu + p_{\pi^\pm}^\mu)(p_{\bar{D}^0\mu} + p_{\pi^\pm\mu}) \quad (5.2)$$

is calculated and shown in Figure 5.4, where  $p_X^\mu$  is the relativistic four-momentum of the particle X. The peaking structure at around 2010  $\text{MeV}/c^2$  indicates the presence of  $D^*(2010)^\pm \rightarrow D^0 \pi^\pm$  decays. They are removed by requiring  $m(D^0 \pi^\pm) > 2040 \text{ MeV}/c^2$ .



**Figure 5.4:** The mass distribution of the  $\bar{D}^0$  combined with  $\pi^+$  (blue) or  $\pi^-$  (orange). The peak indicates the presence of  $D^*(2010)^\pm \rightarrow D^0 \pi^\pm$  decays. Everything lower than the "cut" line is removed.

### 5.4 Further Misidentification

Even after applying the restrictions on the PID variables in section 5.2 some background due to misidentification can be present. That can be seen in the mass distribution of the combined four-momenta of particles. Where the mass in the four momentum  $p_\mu = (\sqrt{m^2 + \vec{p}^2}, -\vec{p})^T$  is changed to a new hypothesis. Different mass hypotheses for a number of particles were tested, further information is provided in Appendix A. Only in the  $\bar{D}^0 \rightarrow KK$  mode peaking

<sup>1</sup>X can be different particles, but also nothing.

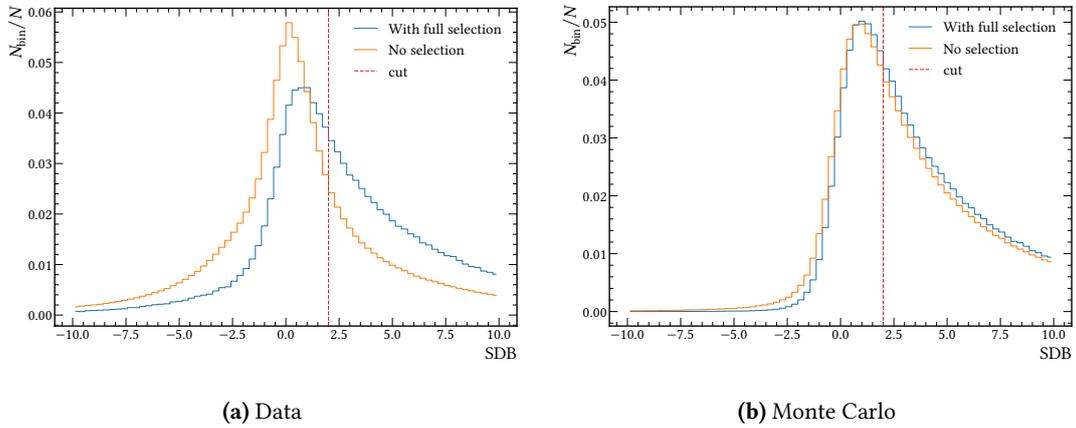
structures of physical background can be seen, but none in the other modes. The observed peaks are primarily  $\Lambda_c \rightarrow pK\pi$  decays, where a (anti) proton was misidentified as a  $(K^-) K^+$ . The mass distributions are found in the appendix section A.2. These are not further considered or removed, because the following removal of combinatorial background is performed only with the  $K\pi$  mode, as the required Monte Carlo simulations for the other modes are not yet available.

## 5.5 Charmless Background

Physical background containing only charmless mesons cannot contain any  $D^0$ -mesons. For example the decay  $B^0 \rightarrow K^+\pi^-\pi^+\pi^-$  contributes to charmless background and is mistaken as signal, because it has the same final state as the signal decay. To reduce the level of charmless background a new variable

$$SDB := \frac{(z_{D^0} - z_{B^0})}{\sqrt{\sigma_{z_{D^0}}^2 + \sigma_{z_{B^0}}^2}} \quad (5.3)$$

is introduced<sup>2</sup>. Here  $z_{D^0}$  and  $z_{B^0}$  are the  $z$ -coordinates of the end vertex of the respective particle. The coordinate system is still the same as in Figure 3.1 and because of the high forward momentum of the produced  $B^0$  meson the difference in the  $z$ -values of the vertices hold the most difference. The variable is defined in such a way that its standard deviation is one. Charmless decays do not have any  $D^0$  vertices and thus the reconstructed end vertex of the final state particles from  $D^0$  should be very near to the end vertex of the  $B^0$ .

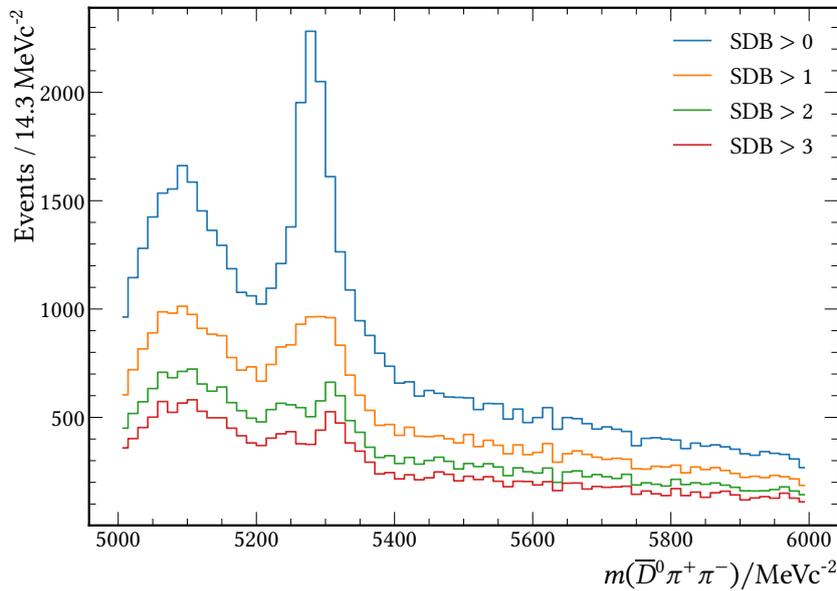


**Figure 5.5:** Distribution of the SDB variable on (a) Data and (b) Monte Carlo samples in the  $K\pi$  mode. The Distributions are normed so the sum of all bins adds up to one. "Full selection" (blue) refers to all the selections of section 5.1 to section 5.3.

<sup>2</sup>SDB is for Significance D-B

By requiring this  $SDB$  variable to be high the charmless background can be removed. The distribution of the  $SDB$  variable on data and Monte Carlo samples is shown in Figure 5.5. The distribution is symmetrical around zero before applying the selections from section 5.1 to 5.3, but not after applying the selections from the previous chapters. The distribution on Monte Carlo samples, which consist of only simulated signal events, is indeed asymmetrical for signal events.

Figure 5.6 shows the invariant  $\bar{D}^0\pi^+\pi^-$  mass distribution, considering only the  $D^0$  mass sidebands, which means that only events with  $|m(D^0) - m_{D^0}^{\text{pdg}}| > 25 \text{ MeV}/c^2$  are shown. There is a significant peak at the  $B^0$  mass which is reduced with higher  $SDB$  cuts. The cut  $SDB > 2$  is chosen as there is still a significant peak left with  $SDB > 1$  and the structure does not seem to be changed with  $SDB > 3$ .



**Figure 5.6:** The invariant  $\bar{D}^0\pi^+\pi^-$  mass distribution for the  $D^0 \rightarrow K\pi$  mode using only the  $D^0$  mass sidebands for different cuts on the  $SDB$  variable. The event are only  $K\pi$  modes in the sidebands  $|m(D^0) - m_{D^0}^{\text{pdg}}| > 25 \text{ MeV}/c^2$ .

The invariant  $\bar{D}^0\pi^+\pi^-$  mass distribution after applying all of the described selection steps, is shown in the appendix in Figure A.2. A remaining peak on the lower mass side of the  $B^0$  peak is due to  $B^0 \rightarrow D^*\pi\pi$ -decays where a pion escapes detection. Because the decay contains a real  $D^0$ , it cannot be removed with these selections, but could be considered as part of a  $B^0$  mass fit. On the upper mass side there is a small sign of  $\Lambda_b$  decays visible.

## 6 Combinatorial Background Selection

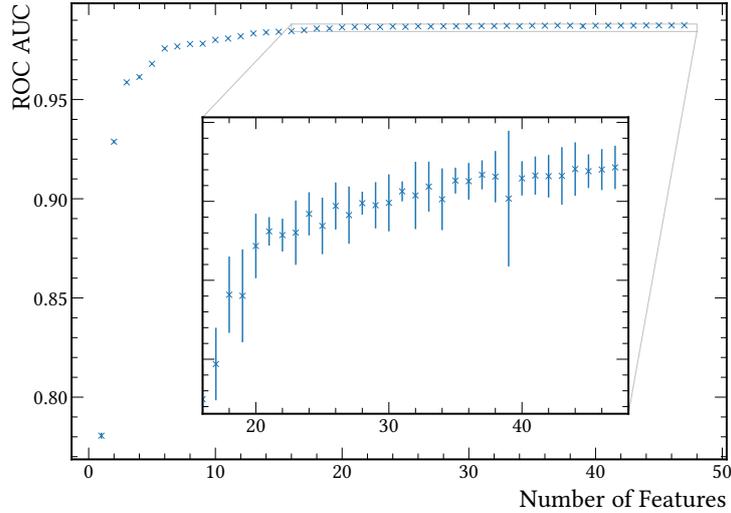
Combinatorial background will be removed by training boosted decision trees (BDTs) to label the data as signal or background. The BDT implementation of XGBoost[46] is used in the statistical learning framework `sklearn` [47]. To train the BDT labeled data is needed. For this purpose Monte Carlo simulations are used as signal data. Data from the upper  $B^0$  mass sideband (UMSB), defined as  $5400 < m(B^0) < 6000 \text{ MeV}/c^2$ , is used as combinatorial background. The problem with using the UMSB as training data is, that we later need to also classify these events again. When training data is again classified by the same classifier, it is most likely to be labeled correctly, as the classifier already “knows” the data. This is not to be confused with *overfitting*, which means that the model is so complex that it is beginning to tune the parameters to the noise in the training data, and has a decreasing score on test data with increasing model complexity (in this case number of learning rounds). For that reason five individual BDTs are trained on different events. The  $i$ th BDT is trained on every event where the row number  $n_r$  satisfies  $n_r \bmod 5 \neq i$ .<sup>1</sup> In that way it can later be used to label every event with row number  $n_r \bmod 5 = i$ . Combinatorial background can be assumed to have no peaks and be distributed exponentially over the whole region. Before training the model, the hyperparameters are chosen on a basis of some number of test learning rounds on the whole training dataset. It is noted that this means that potentially information about the truth value of the dataset can leak into the classifier which is later trained through the hyperparameters. But because only two hyperparameters are optimised with a very small grid search and a very loose feature selection is applied, it is not possible to overtrain with these few parameters on such a complex and big dataset.

### 6.1 Feature Selection and Hyperparameters

Because of the signal and background being chosen in terms of the  $B^0$  mass, we must be careful which features to include, since the model will learn to select  $m(B^0) > 5400$  if it can reconstruct the mass from the features. Thus no features are used which correlate strongly with the  $B^0$  mass. A total number of 47 features are considered to train the model, they are listed in Appendix B. An iterative backwards feature selection is used to reduce the number of features. The correlation on the finally used features with the  $B^0$  mass is considered in Appendix B. It is done by training five models on the five folded dataset, so each model is trained on  $\frac{1}{5}$  of the data and then validated on  $\frac{1}{5}$  of the data. This allows to estimate an error on some validation metric. After saving the metric the most unimportant feature is removed and the model is trained again. This is repeated until no feature is left. The result is seen in Figure 6.1. It is concluded that the use of the best 20 features is enough to give good results. These features and their corresponding average importances in the final BDTs are listed in Table 6.1.

---

<sup>1</sup>mod is the modulo operation



**Figure 6.1:** The ROC AUC score as a function of the number of features used. They are removed with the described backward feature selection.

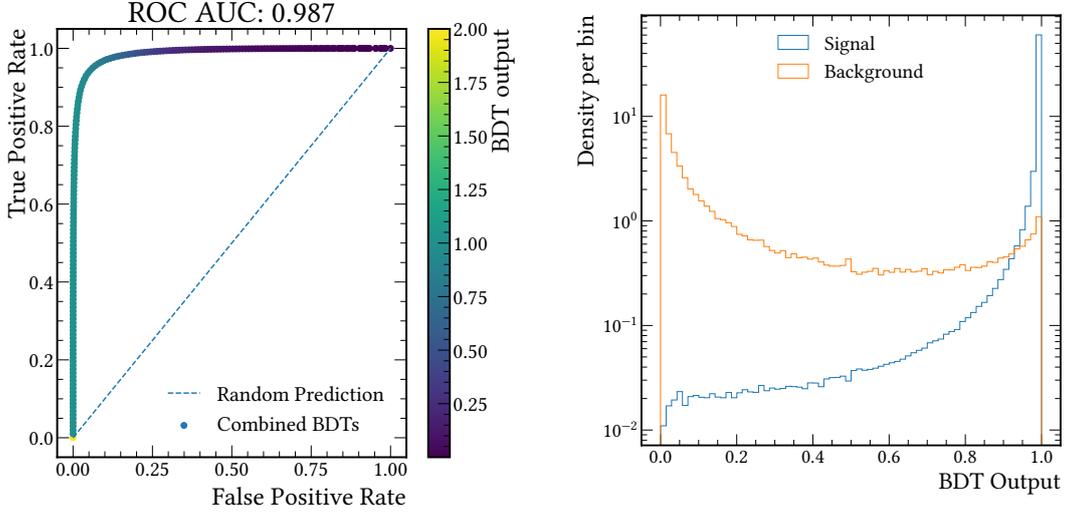
The most important hyperparameters of the model, the learning rate ( $lr$ ) and maximum depth ( $md$ ) of each tree, are optimised by testing every combination of hyperparameters in the space  $(md, lr) \in \{4, 5, 6, 7, 8\} \times \{0.1, 0.2, 0.3, 0.4\}$ . The optimal parameters are evaluated to be  $lr = 0.2$  and  $md = 8$ . Additionally the method of early stopping rounds is used to determine the maximum number of trees in the model, while training the model. In this method, training is only continued, if some test score (in our case ROC AUC) has improved in the last 10 iteration. This also helps to reduce overfitting.

## 6.2 Training of the Boosted Decision Tree

More on the details of the performance of the individual classifiers are given in Appendix B. The area under the receiver operator characteristics curve (*ROC AUC*) is used to measure the performance of the classifier. It is the parametric curve of the *false positive rate* against the *true positive rate* as a function of the cut on the classifier output  $c$ . The curve is shown in Figure 6.2. A random classification would on average have a score of  $ROC AUC = 0.5$ , while a perfect classifier would score 1 and thus split the data perfectly. Figure 6.2 shows the ROC curve for the whole dataset. When it is evaluated for each of the five classifiers individually the average score is

$$ROC AUC = 0.98723 \pm 0.00017. \quad (6.1)$$

The number of trees, as evaluated with early stopping rounds, in each model is 185, 155, 123, 214, 187, which makes an average of 172.8 trees in the classifier.



**Figure 6.2:** ROC curve for the combined BDTs on the dataset (left) and a histogram for the combined BDT output on signal events and background events. To classify between background and signal a cut on the  $x$ -axis on the right plot is chosen (see section 6.3).

**Table 6.1:** Finally used features and their average importances on the five BDTs. DOCA stands for *Distance Of Closest Approach*, IP for *Impact Parameter* and FD for *Flight Distance*.

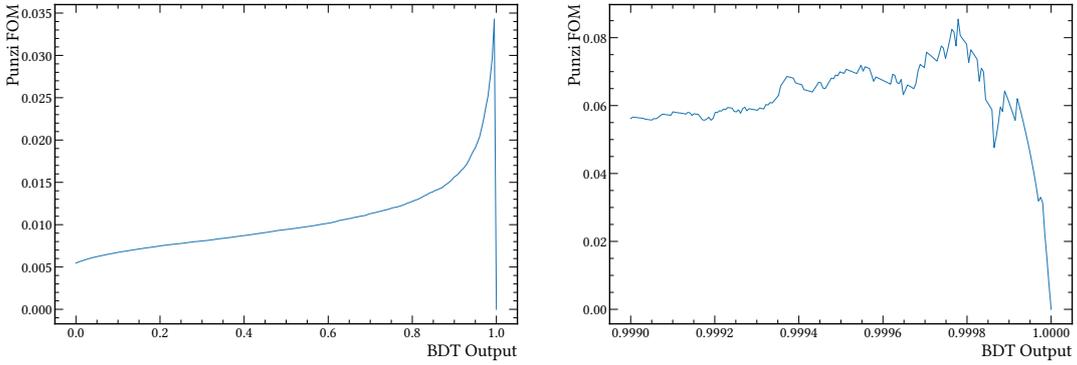
Particle	Feature	Importance	Particle	Feature	Importance
$\pi^+$	$\chi^2$ IP	$0.1981 \pm 0.0049$	$D^0$	IP	$0.0340 \pm 0.0005$
$\pi^-$	$\chi^2$ IP	$0.1965 \pm 0.0062$	$\pi^+$	$\eta$	$0.0299 \pm 0.0009$
$B^0$	$\chi^2$ DTF	$0.1070 \pm 0.0015$	$\pi^-$	$\eta$	$0.0299 \pm 0.0009$
$B^0$	$\chi^2$ FD	$0.0495 \pm 0.0014$	$B^0, \bar{D}^0$	$\cos \sphericalangle(\vec{p}_{B^0}, \vec{p}_{\bar{D}^0})$	$0.0193 \pm 0.0010$
$\pi^-$	$p_T$	$0.0464 \pm 0.0009$	$\pi^+$	IP	$0.0174 \pm 0.0003$
$B^0$	$\chi^2$ End Vertex	$0.0449 \pm 0.0010$	$\pi^-$	IP	$0.0162 \pm 0.0008$
$\pi^+$	$p_T$	$0.0431 \pm 0.0023$	$B^0$	IP	$0.0160 \pm 0.0007$
$B^0$	FD	$0.0391 \pm 0.0008$	$D^0$	$\eta$	$0.0155 \pm 0.0007$
$B^0$	DOCA to PV	$0.0371 \pm 0.0023$	$D^0$	$ \vec{p} $	$0.0148 \pm 0.0004$
$B^0$	$\chi^2$ IP	$0.0362 \pm 0.0020$	$D^0$	$\chi^2$ FD	$0.0108 \pm 0.0006$

### 6.3 Figure of Merit Optimisation

To decide where to cut the output of the BDT Punzi's figure of merit (FOM) is used [48]. It allows to select a cut  $c$  with a chosen significance level of  $N\sigma$ , here  $N = 3$ . It is defined as

$$FOM_{N\sigma}^{\text{Punzi}}(c) = \frac{\varepsilon(c)}{N/2 + \sqrt{B(c)}}. \quad (6.2)$$

The efficiency  $\varepsilon(c)$  (fraction of signal events which remain) of the cut  $c$  can be calculated from Monte Carlo data. The number of background events in the signal region of the cut  $c$  is calculated from the UMSB as follows.



**Figure 6.3:** The Punzi figure of merit as a function of the cut. The right hand side is a zoomed in version of the left hand side with a higher sampling rate.

We model the probability distribution function of the background  $f_B(m)$  to be an exponential

$$f_B(m) = Ae^{-\lambda m}. \quad (6.3)$$

The UMSB is defined as  $m_l < m < m_u$  and the signal region is defined as  $m_{sl} < m < m_{su}$ . If we divide the UMSB into half, we can express the number of background events  $N_l$  and  $N_u$  in these two regions as

$$N_l = N_B \int_{m_l}^{\frac{m_l+m_u}{2}} f_B(m) dm = \frac{N_B A}{-\lambda} (e^{-\lambda/2 m_l + m_u} - e^{-\lambda m_l}) \quad (6.4)$$

$$N_u = N_B \int_{\frac{m_l+m_u}{2}}^{m_u} f_B(m) dm = \frac{N_B A}{-\lambda} (e^{-\lambda m_u} - e^{-\lambda/2 m_l + m_u}). \quad (6.5)$$

Where  $N_B$  is the total number of background events. Introducing  $m_u - m_l = \Delta m$  gives

$$\lambda = \frac{2}{\Delta m} \ln\left(\frac{N_l}{N_u}\right) \quad (6.6)$$

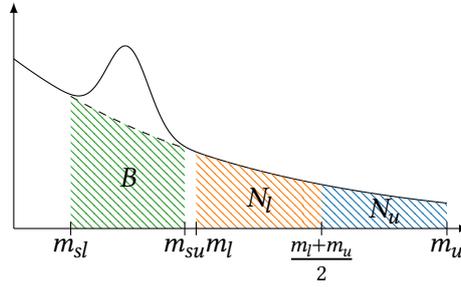
and with equation (6.4) we can get the last free parameter of equation (6.3),

$$A = \frac{-\lambda N_l}{N_B(e^{-\lambda m_u} - e^{-\lambda/2 m_l + m_u})} \quad (6.7)$$

This allows to estimate the number of background events under the signal as

$$B = N_B \int_{m_{sl}}^{m_{su}} f_B(m) dm = N_l \frac{e^{-\lambda m_{su}} - e^{-\lambda m_{sl}}}{e^{-\lambda m_u} - e^{-\lambda/2 m_l + m_u}} \quad (6.8)$$

This method of is illustrated in Figure 6.4. We define the regions as  $m_l = 5400 \text{ MeV}/c^2$ ,  $m_u = 6000 \text{ MeV}/c^2$ ,  $m_{sl} = 5225 \text{ MeV}/c^2$  and  $m_{su} = 5325 \text{ MeV}/c^2$ .



**Figure 6.4:** Illustration of the used method to estimate the number of background events in the signal region.

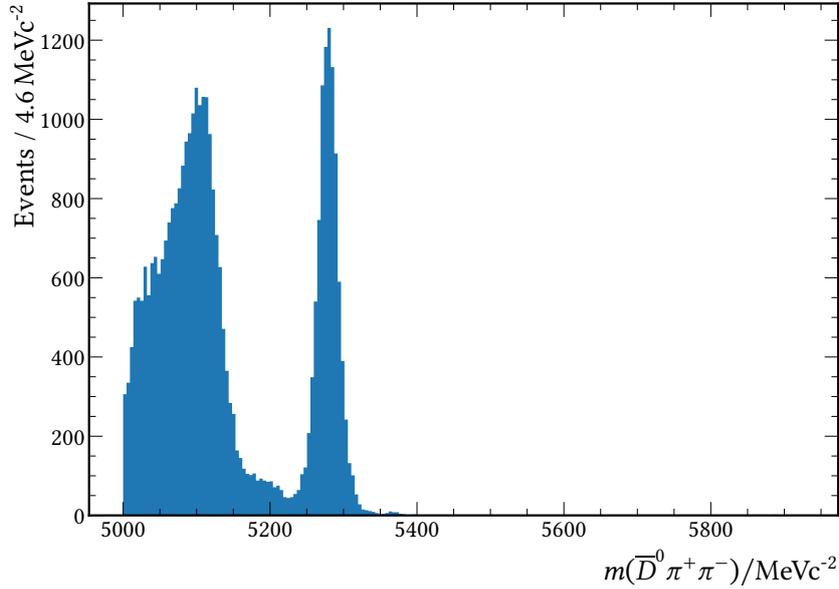
With this FOM the optimal cut is at about 0.9998, as seen in Figure 6.3. As this is quite high and results in a relative signal efficiency of about 34 % (compare Table 6.2), a smaller cut on the BDT output of 0.99 is also considered and listed in Table 6.2, which results in a relative signal efficiency of ca. 83.8 %.

## 6.4 Selection Results

The resulting mass distribution after all the selections is seen in Figure 6.5, where the cut of 0.9998 on the BDT output as found in section 6.3 is used. In Appendix B a cut 0.99 on the mass distribution is also shown for higher efficiency. The efficiencies of each selection is estimated on Monte Carlo data and listed in Table 6.2. The overall efficiency is

$$\varepsilon = 10.6 \% \quad (6.9)$$

Additionally a Random Forest classifier and a Neural Network classifier were also considered as models, but yielded very similar performances as the BDT on this dataset. Their ROC curves are seen in appendix section B.3.



**Figure 6.5:** The invariant  $\bar{D}^0\pi^+\pi^-$  mass distribution after all selections. A very clear  $B^0$  peak is visible and the peak on its lower mass side is from real  $D^*\pi\pi$  decays, which can later be considered in a fit.

**Table 6.2:** Efficiencies on Monte Carlo data for the  $\bar{D}^0 \rightarrow K\pi$  modes, using only correctly identified particles. The selections are added from top to bottom following the chapters of this thesis, with the exception of the SDB selection without the PID selection. Efficiency means the overall efficiency of all cuts, while relative efficiency means the efficiency only compared to the previous cut.

Selection Stage	MC Yield	Efficiency / %	Relative Efficiency / %
None	3 015 025	100.0	100.0
Preselection (section 5.1)	2 538 465	84.2	84.2
PID only $D^0$ daughter	2 025 114	64.2	79.8
PID $D^0$ daughter and $\pi$ from $B^0\pi$	1 558 397	51.6	77.0
PID (section 5.2)	1 474 189	48.9	94.6
Veto (section 5.3)	1 393 559	46.2	94.6
SDB (section 5.5)	925 692	30.7	66.4
SDB without PID	1 570 899	52.1	61.9
BDT selection Punzi (chapter 6)	318 164	10.6	34.4
BDT selection (Output > 0.99)	775 809	25.7	83.8

## 7 Conclusion

In this thesis a selection of the  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decay is developed. Its goal is to be used to be used as a normalisation mode for the  $B^0 \rightarrow D^0 \bar{D}^0$  decay, which will be analysed in the future in a measurement of  $CP$  violation. Because of this, the selections are designed to resemble the selections of the upcoming study.

First cuts are applied to suppress badly reconstructed events. After that restrictions on the particle identifications are made. This should remove physical background coming from misidentified particles. Other physical backgrounds have the same final state particles as the decay we want to observe, like,  $D^*(2010)^\pm$  events from the combination of a  $D^0$  meson and a  $\pi$  meson. They are removed by requiring that the invariant mass of the  $D^0$  meson and one of the  $\pi$  mesons coming out of the  $B^0$  decay is greater than  $2040 \text{ MeV}/c^2$ . Additionally, further misidentifications are studied by applying new mass hypotheses to calculate the invariant mass. We do not observe any misidentification peaks in the  $D^0 \rightarrow K\pi$  decay mode, but some are observed for the  $KK$  decay. These backgrounds in  $KK$  and possible background in  $\pi\pi$  should be investigated further in future studies. The removal of charmless background is done by requiring newly introduced SDB variable to be greater than two.

The combinatorial background is separated from signal using a Boosted Decision Tree classifier, which is trained with background data from the UMSB of  $B^0$  and signal from Monte Carlo simulations. The UMSB is split into five sets, to be able to properly classify the UMSB later again. Finally a cut on the output of the BDT is chosen using Punzi's figure of merit.

When Monte Carlo simulations of the  $KK$  and  $\pi\pi$  modes are ready, they can easily be added to this analysis using the existing framework produced in this thesis for the  $K\pi$  mode. Some analysis of backgrounds of the  $KK$  and  $\pi\pi$  modes in data was already considered in section 5.4.

The next step for the  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  normalisation would be to fit the  $B^0$  mass distribution in Figure 6.5 with an appropriate model. In the model the real  $D^* \pi\pi$  decays which are showing up in the distribution, but cannot be removed with selections, would be included. The signal yields calculated from this fit would be combined with the results of similar fits for the  $KK$  and  $\pi\pi$  modes, and together with the selection efficiencies, these will be used to normalise the  $B^0 \rightarrow D^0 \bar{D}^0$  branching ratio.

With this thesis as its foundation, we expect that the future analysis will be able to find the  $B^0 \rightarrow D^0 \bar{D}^0$  decay with more than  $3\sigma$ , if it is there. If the branching ratio is the same or larger than the branching ratio from the 2011 analysis [7], it is also expected that the  $B^0 \rightarrow D^0 \bar{D}^0$  analysis will present first results for  $CP$  violation in the decay.

# A Additional Information on the Physical Background

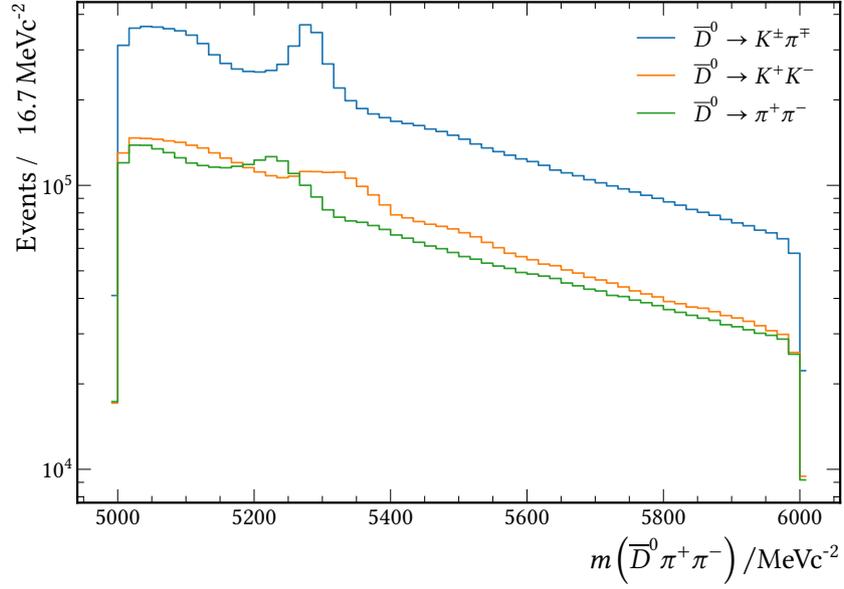
The used stripping line is listed in Table A.1. The  $B^0$  mass shape before any selection is shown in Figure A.1 and after the removal of physical background is shown in Figure A.2.

**Table A.1:** Stripping selections for the D0pipi stripping line, with D0 to 2 hadrons

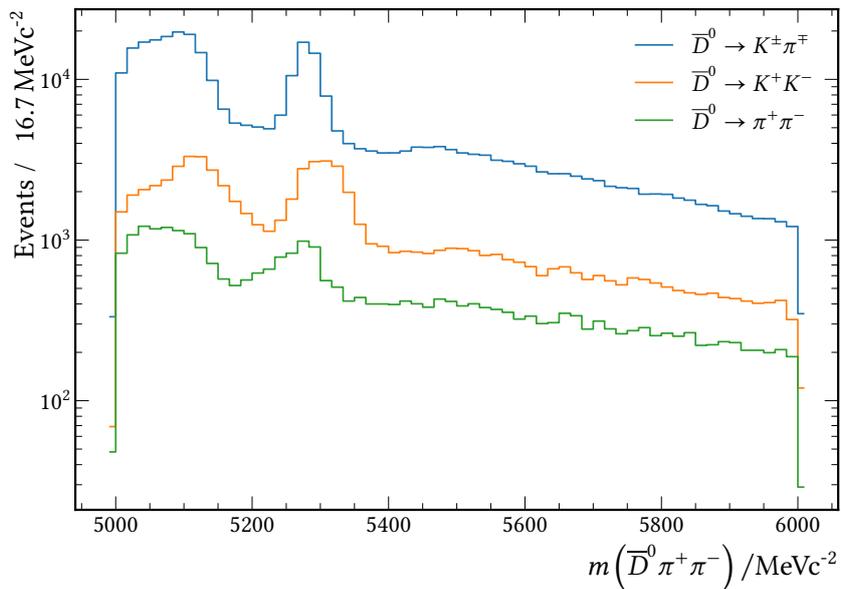
Particle or combination	Cuts
good event trigger	number of long tracks < 500 HLT2Topo or HLT2IncPhi
all pions/kaons	track $\chi^2 / \text{dof} < 4.0$ $p_T > 100 \text{ MeVc}^{-1}$ and $p > 1000 \text{ MeVc}^{-1}$ minimum $\chi_{IP}^2$ with PV > 4.0 track ghost probability < 0.4
one $h$ in $D^0 \rightarrow hh$ has	$p_T > 500 \text{ MeV}$ and $p > 5000 \text{ MeV}$
$hh$ in $D^0 \rightarrow hh$	$\sum p_T > 1800 \text{ MeV}$ $1764.84 < m(HH) < 1964.84$ for $hh$ either $\pi\pi$ , $K\pi$ , or $KK$ DOCA( $hh$ ) < 0.5mm no particles marked as proton with PIDp < -10 no particles marked as kaon with PIDK < -10 no particles marked as pion with PIDK > 20
$D^0$	vertex $\chi^2 < 10$ vertex distance $\chi^2 > 36$ DIRA > 0
pions directly from $B^0$	$p > 2000 \text{ MeV}$
one pion from $B^0$ has	$p_T > 500 \text{ MeV}$ and $p > 5000 \text{ MeV}$
$\pi\pi$ combination	$\sum p_T > 1000 \text{ MeV}$ $m(\pi\pi) < 5.2 \text{ GeV}$ DOCA( $\pi\pi$ ) < 0.5mm vertex $\chi^2$ of $\pi\pi$ combination < 16 vertex distance $\chi^2$ of $\pi\pi$ combination > 36 DIRA of $\pi\pi$ combination > 0
$D^0\pi\pi$ combination	$\sum p_T > 5000 \text{ MeV}$ $5000 < m(D^0\pi\pi) < 6000$
one hadron in $D^0\pi\pi$ has	$p > 10 \text{ GeV}$ and $p_T > 1700 \text{ MeV}$ minimum $\chi_{IP}^2$ for any PV > 16 minimum IP for any PV > 0.1mm
$B^0$	vertex $\chi^2 < 10$ $\chi_{IP}^2 < 9$ DIRA > 0.99995 lifetime relative to PV > 0.2ps

## A.1 $B^0$ mass Distributions

Comparison of Figure A.1 and Figure A.2 shows that the removal of physical background with the selections in chapter 5 already improves the  $B^0$  mass peak.



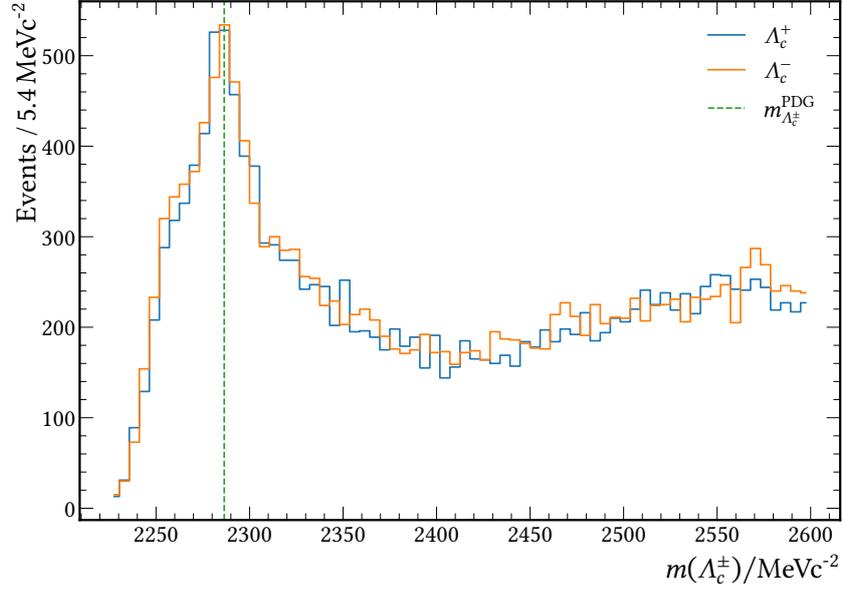
**Figure A.1:** The invariant  $\bar{D}^0 \pi^+ \pi^-$  mass distribution with only experiment wide stripping applied.



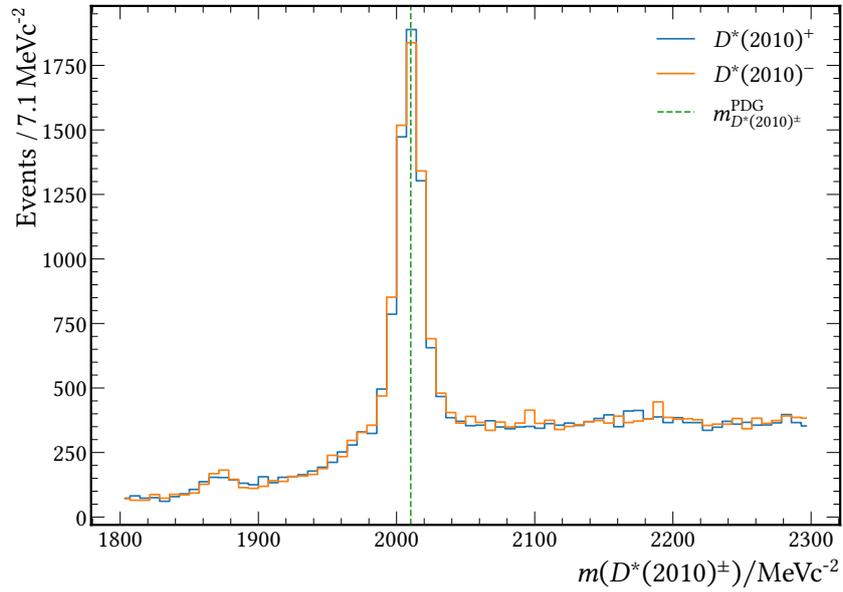
**Figure A.2:** The invariant  $\bar{D}^0 \pi^+ \pi^-$  mass distribution after all of the selection steps in chapter 5 are applied.

## A.2 Misidentifications

The considered misidentification were misidentified  $D_s^\pm$ ,  $D^\pm$  and  $\Lambda_c^\pm$  decays in the three considered modes of the  $D^0$ . After the cuts in section 5.1 to section 5.3, only two misidentification peaks in the  $D^0 \rightarrow K^+ K^-$  channel remained, they are shown in Figure A.3 and Figure A.4.



**Figure A.3:**  $\Lambda_c^\pm$  misidentification as the  $D^0 \rightarrow K^+K^-$  channel.  $m(\Lambda_c^\pm)$  means the invariant mass of  $m(p^\pm K^\mp \pi^\pm)$ , where the (anti)-proton  $p^\pm$  is misidentified as  $(K^-) K^+$ .



**Figure A.4:**  $D^*(2010)^\pm$  misidentified as the  $D^0 \rightarrow K^+K^-$  channel.  $m(D^*(2010)^\pm)$  means the invariant mass of  $m(K^\pm \pi^\mp \pi^\pm)$ , where the  $\pi^\mp$  is misidentified as  $K^\mp$ .

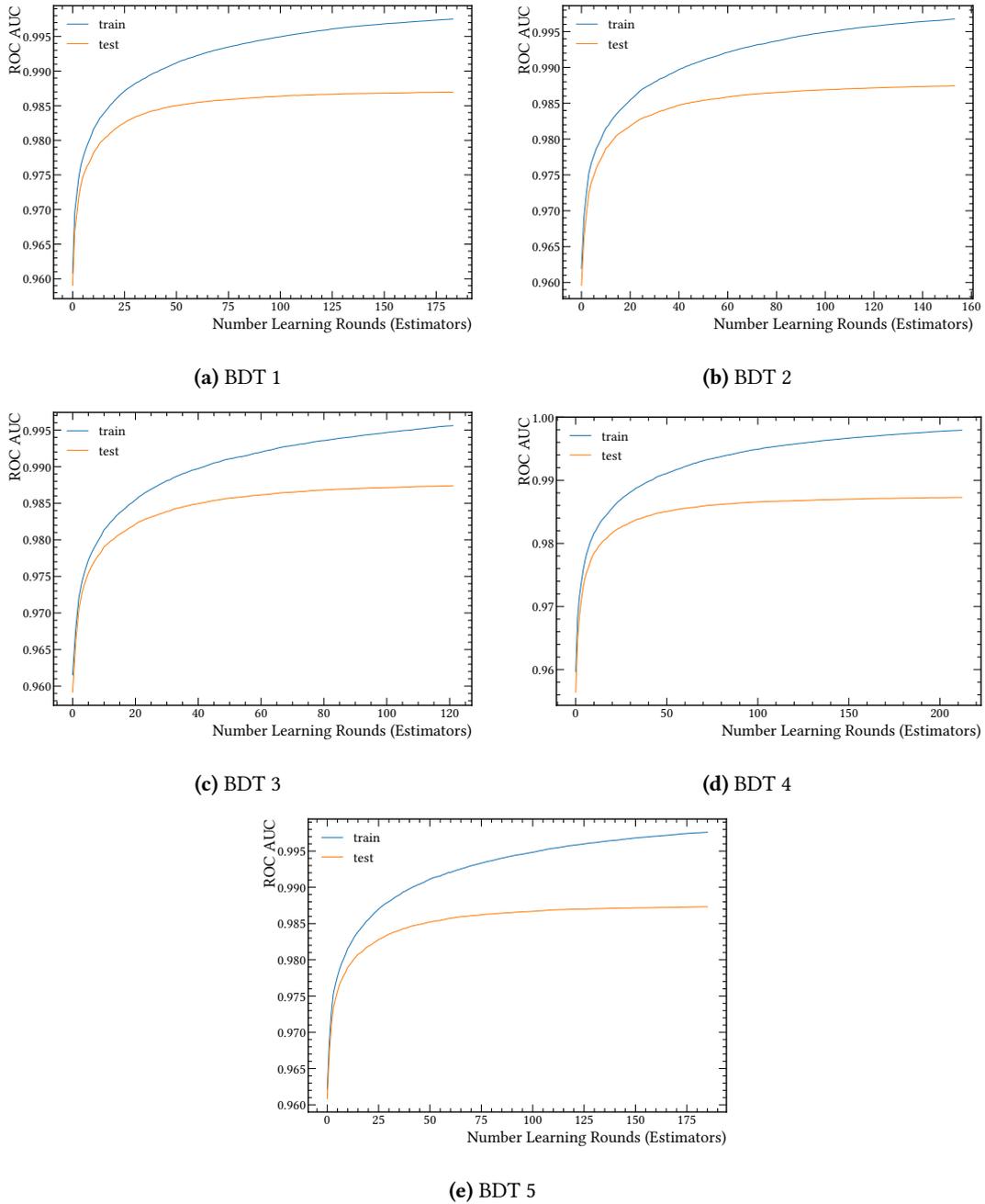
## B Additional Information on the Combinatorial Background

All 47 input variables are listed in Table B.1. To evaluate whether the five BDT models are overfitted their ROC AUC score is evaluated as a function of the number of boosting rounds on the training and test dataset, the graphs for each BDT can be seen in Figure B.1.

**Table B.1:** All considered input features for the BDT and their corresponding feature importances for a first run on the whole dataset.

Feature	Importance	Feature	Importance
pip_IPCHI2_OWNPV	0.182	K_D0_LOKI_ETA	0.007
pim_IPCHI2_OWNPV	0.151	pip_TRACK_VeloCHI2NDOF	0.007
B0_DTF_CHI2	0.081	D0_LOKI_DOCA	0.007
B0_FDCHI2_OWNPV	0.048	pi_D0_TRACK_MatchCHI2	0.007
B0_ENDVERTEX_CHI2	0.043	K_D0_IP_OWNPV	0.007
D0_IP_OWNPV	0.043	pi_D0_LOKI_ETA	0.006
pip_PT	0.037	K_D0_TRACK_VeloCHI2NDOF	0.006
pim_PT	0.036	D0_ACOS_DIRA	0.006
B0_LOKI_DOCA	0.031	pi_D0_TRACK_VeloCHI2NDOF	0.006
B0_IPCHI2_OWNPV	0.028	pip_P	0.006
B0_FD_OWNPV	0.025	pim_P	0.006
pip_LOKI_ETA	0.024	pi_D0_PT	0.005
pim_LOKI_ETA	0.023	pi_D0_IPCHI2_OWNPV	0.005
B0_LOKI_Costheta_D0	0.018	K_D0_TRACK_MatchCHI2	0.005
pim_IP_OWNPV	0.015	D0_FD_OWNPV	0.005
D0_P	0.012	pi_D0_IP_OWNPV	0.005
D0_LOKI_ETA	0.011	K_D0_P	0.005
D0_FDCHI2_OWNPV	0.011	K_D0_PT	0.005
B0_IP_OWNPV	0.010	K_D0_IPCHI2_OWNPV	0.005
pim_TRACK_MatchCHI2	0.010	D0_PT	0.004
pip_IP_OWNPV	0.009	pi_D0_P	0.004
pip_TRACK_MatchCHI2	0.008	D0_TAU_RATIO	0.004
D0_IPCHI2_OWNPV	0.008	D0_ENDVERTEX_CHI2	0.004
pim_TRACK_VeloCHI2NDOF	0.008		

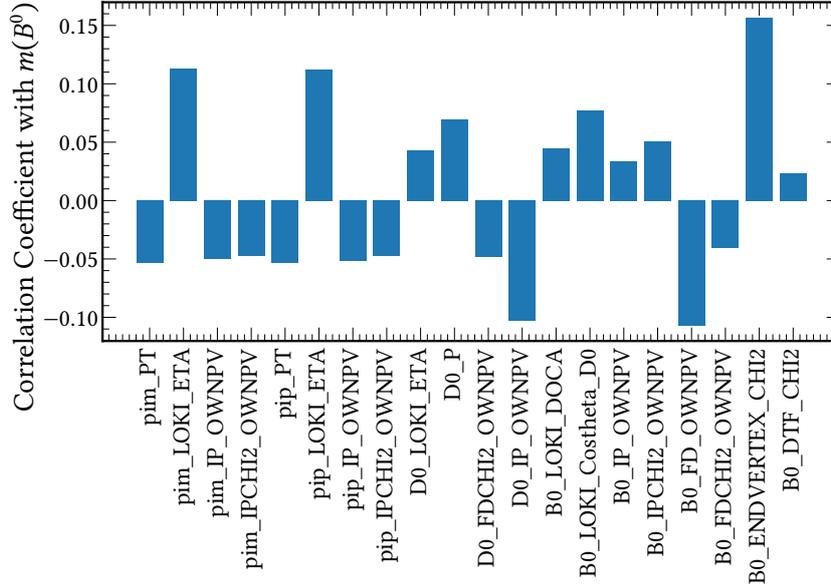
## B Additional Information on the Combinatorial Background



**Figure B.1:** The ROC AUC score of each individual BDT as a function of learning rounds. As no ROC AUC score decreases on the test dataset, it can be concluded that the models don't seem to be overtrained.

## B.1 Correlation Coefficients with the $B^0$ Mass

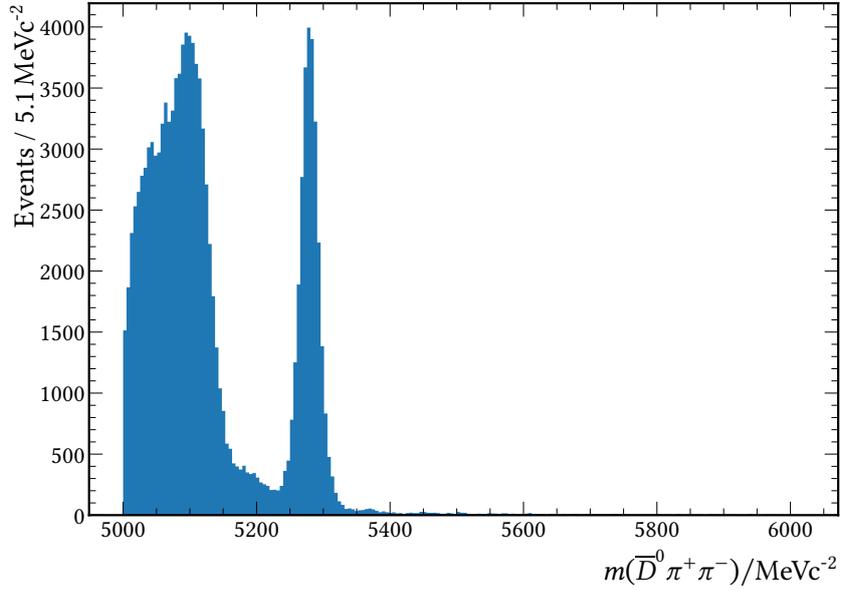
The correlation coefficient with the  $B^0$  mass has to be small, so the BDT cannot identify the definition we chose for the UMSB. This is evaluated in Figure B.2.



**Figure B.2:** Correlation coefficient of the finally used variables (see Table 6.1) with the  $B^0$  mass on the preselected dataset in the  $D^0 \rightarrow K\pi$  mode.

## B.2 Alternative cut on BDT output

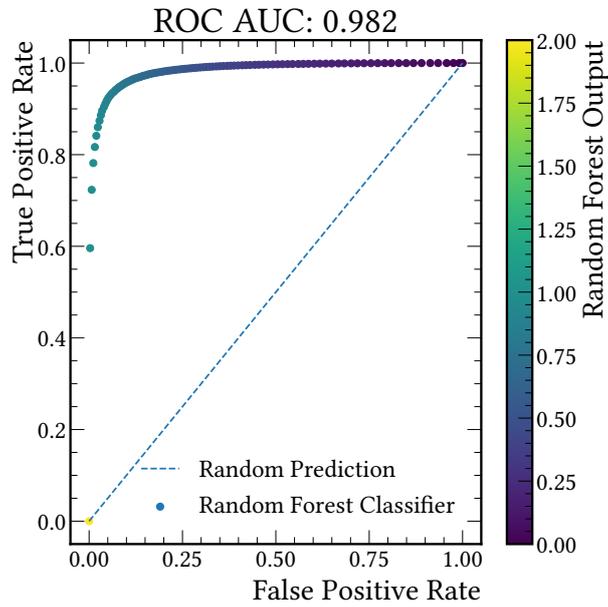
Due to the low efficiency of the optimised FOM cut on the BDT output, a cut of 0.99 is also considered and the resulting  $B^0$  mass distribution is shown in Figure B.3.



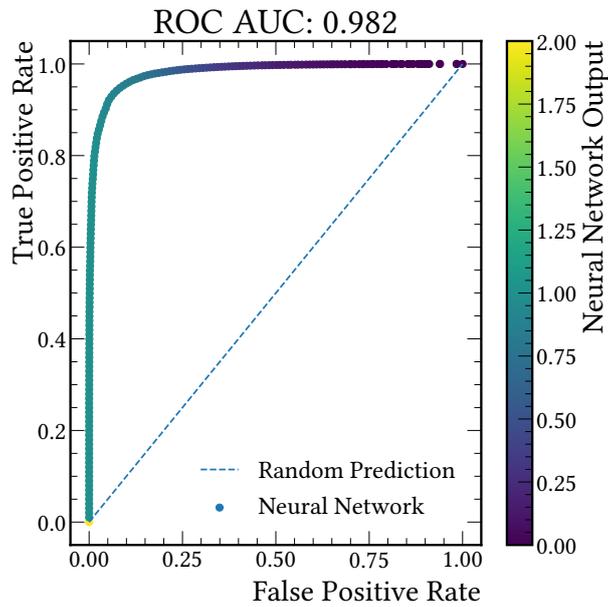
**Figure B.3:** The invariant  $\bar{D}^0\pi^+\pi^-$  mass distribution after all selections. The requirement of the BDT output is set to be greater than 0.95 here, in contrast to what is optimised in section 6.3 and shown in Figure 6.5.

### B.3 Comparison to Neural Network and Random Forest

Additionally to the Boosted Decision Tree classifier, a Neural Networks classifier and Random Forest classifier were also considered. After training them with 75% of the training data, their performance was tested on the remaining 25%. It was decided to continue using Boosted Decision Trees, as all models yielded very similar results. They were trained with all 47 variables in Table B.1. The Neural Network consists of four hidden layers of size 30, 20, 20, 20 and dropout regularisation is used. The *swish* function was used as the activation function [49]. The Random Forest consisted of 100 decision trees. The ROC curves and AUC scores of the Random Forest and Neural Network are shown in Figure B.4.



(a) Random Forest



(b) Neural Network

**Figure B.4:** The ROC curve and the AUC for the Random Forest classifier (left) and a Neural Network classifier (right) described in this section.

## Bibliography

- [1] A. G. Cohen, A. De Rujula, and S. Glashow, “A Matter - antimatter universe?”, *Astrophys. J.* **495**, 539 (1998), arXiv:astro-ph/9707087.
- [2] M. Dine and A. Kusenko, “The Origin of the matter - antimatter asymmetry”, *Rev. Mod. Phys.* **76**, 1 (2003), arXiv:hep-ph/0303065.
- [3] M. Jung and S. Schacht, “Standard Model predictions and new physics sensitivity in  $B \rightarrow DD$  decays”, *Phys. Rev. D* **91**, 034027 (2015), arXiv:1410.8396.
- [4] M. Cheng et al. (Belle Collaboration), “A Study of CP violation in B meson decays: Technical design report”, BELLE-TDR-3-95, KEK-95-1 (1995).
- [5] D. Boutigny et al. (BaBar Collaboration), “BaBar technical design report”, SLAC-R-0457, SLAC-R-457, SLAC-0457, SLAC-457, SLAC-R-95-457 (1995).
- [6] W. Altmannshofer et al. (Belle-II Collaboration), “The Belle II Physics Book”, *PTEP* **2019**, edited by E. Kou and P. Urquijo, 10.1093/ptep/ptz106 (2019), arXiv:1808.10567.
- [7] R. Aaij et al. (LHCb Collaboration), “First observations of  $\bar{B}_s^0 \rightarrow D^+ D^-$ ,  $D_s^+ D^-$  and  $D^0 \bar{D}^0$  decays”, *Phys. Rev. D* **87**, 092007 (2013), arXiv:1302.5854.
- [8] P. Zyla et al. (Particle Data Group), “Review of Particle Physics”, to be published in *Prog. Theor. Exp. Phys.* **2020**, 083C01 (2020).
- [9] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13**, 508 (1964).
- [10] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13**, 321 (1964).
- [11] S. L. Glashow, “Partial-symmetries of weak interactions”, *Nuclear Physics* **22**, 579 (1961).
- [12] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19**, 1264 (1967).
- [13] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc. C* **680519**, 367 (1968).
- [14] M. Gell-Mann, “THE EIGHTFOLD WAY: A THEORY OF STRONG INTERACTION SYMMETRY”, DOI : 10.2172/4008239 (1961).
- [15] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, “Advantages of the color octet gluon picture”, *Physics Letters B* **47**, 365 (1973).
- [16] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Non-Abelian Gauge Theories”, *Phys. Rev. Lett.* **30**, 1343 (1973).
- [17] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30**, 1346 (1973).
- [18] M. Gell-Mann, “A schematic model of baryons and mesons”, *Physics Letters* **8**, 214 (1964).

- 
- [19] D. Hanneke, S. Fogwell, and G. Gabrielse, “New Measurement of the Electron Magnetic Moment and the Fine Structure Constant”, Phys. Rev. Lett. **100**, 120801 (2008), arXiv:0801.1134.
- [20] T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio, “Tenth-Order QED Contribution to the Electron  $g-2$  and an Improved Value of the Fine Structure Constant”, Phys. Rev. Lett. **109**, 111807 (2012), arXiv:1205.5368.
- [21] M. Schellenberg, “Measurement of CP violation in the decay  $B^0 \rightarrow D^{*\pm}D^{\mp}$  with data from the LHCb experiment”, TU Dortmund, Dissertation, 10.17877/DE290R-20349 (2019).
- [22] C. Berger, “Elementarteilchenphysik”, Book, 10.1007/978-3-642-41753-5 (2014).
- [23] R. Aaij et al. (LHCb Collaboration), “Test of Lepton Flavor Universality by the measurement of the  $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$  branching fraction using three-prong  $\tau$  decays”, Phys. Rev. D **97**, 072013 (2018), arXiv:1711.02505.
- [24] R. Aaij et al. (LHCb Collaboration), “Observation of  $J/\psi p$  Resonances Consistent with Pentaquark States in  $\Lambda_b^0 \rightarrow J/\psi K^- p$  Decays”, Phys. Rev. Lett. **115**, 072001 (2015).
- [25] R. Aaij et al. (LHCb Collaboration), “Observation of  $J/\psi\phi$  structures consistent with exotic states from amplitude analysis of  $B^+ \rightarrow J/\psi\phi K^+$  decays”, Phys. Rev. Lett. **118**, 022003 (2017), arXiv:1606.07895.
- [26] R. Aaij et al. (LHCb Collaboration), “Observation of structure in the  $J/\psi$ -pair mass spectrum”, (2020), arXiv:2006.16957.
- [27] G. Luders, “On the Equivalence of Invariance under Time Reversal and under Particle-Antiparticle Conjugation for Relativistic Field Theories”, Kong. Dan. Vid. Sel. Mat. Fys. Med. **28N5**, 1 (1954).
- [28] C. Wu, E. Ambler, R. Hayward, D. Hoppes, and R. Hudson, “Experimental Test of Parity Conservation in  $\beta$  Decay”, Phys. Rev. **105**, 1413 (1957).
- [29] J. Christenson, J. Cronin, V. Fitch, and R. Turlay, “Evidence for the  $2\pi$  Decay of the  $K_2^0$  Meson”, Phys. Rev. Lett. **13**, 138 (1964).
- [30] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, Phys. Rev. Lett. **10**, 531 (1963).
- [31] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, Prog. Theor. Phys. **49**, 652 (1973).
- [32] A. Buras, W. Słominski, and H. Steger, “ $B^0 - \bar{B}^0$  mixing, CP violation and the B-meson decay”, Nuclear Physics B **245**, 369 (1984).
- [33] R. Aaij et al. (LHCb Collaboration), “Measurement of  $b$ -hadron production fractions in 7 TeV  $pp$  collisions”, Phys. Rev. D **85**, 032008 (2012), arXiv:1111.2357.
- [34] R. Aaij et al. (LHCb Collaboration), “Measurement of the  $b$ -quark production cross-section in 7 and 13 TeV  $pp$  collisions”, Phys. Rev. Lett. **118**, [Erratum: Phys.Rev.Lett. 119, 169901 (2017)], 052002 (2017), arXiv:1612.05140.

- [35] L. Evans, “*LHC Machine*”, JINST **3**, edited by L. Evans and P. Bryant, S08001 (2008).
- [36] A. A. Alves et al. (LHCb Collaboration), “*The LHCb Detector at the LHC*”, Journal of Instrumentation **3**, S08005 (2008).
- [37] G. Aad et al. (ATLAS Collaboration), “*The ATLAS Experiment at the CERN Large Hadron Collider*”, JINST **3**, S08003 (2008).
- [38] S. Chatrchyan et al. (CMS Collaboration), “*The CMS Experiment at the CERN LHC*”, JINST **3**, S08004 (2008).
- [39] K. Aamodt et al. (ALICE Collaboration), “*The ALICE experiment at the CERN LHC*”, JINST **3**, S08002 (2008).
- [40] Francesca, Dordei, “*LHCb detector and trigger performance in Run II*”, EPJ Web Conf. **164**, 01016 (2017).
- [41] A. Papanestis and C. D’Ambrosio (LHCb RICH), “*Performance of the LHCb RICH detectors during the LHC Run II*”, Nucl. Instrum. Meth. A **876**, edited by P. Krizan, S. Korpar, G. Hallewell, W. Hofmann, and E. Nappi, 221 (2017), arXiv:1703.08152.
- [42] R. Aaij et al. (LHCb Collaboration), “*LHCb detector performance*”, International Journal of Modern Physics A **30**, 1530022 (2015), eprint: <https://doi.org/10.1142/S0217751X15300227>.
- [43] R. Lindner, *LHCb layout2. LHCb schema2*, LHCb Collection., (2008-02) <http://cds.cern.ch/record/1087860>.
- [44] T. Hastie, R. Tibshirani, and J. Friedman, “*The elements of statistical learning: data mining, inference, and prediction*”, Springer Science & Business Media, Book (2009).
- [45] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, “*Fast neural-net based fake track rejection in the LHCb reconstruction*”, LHCb-PUB-2017-011 (2017).
- [46] T. Chen and C. Guestrin, “*XGBoost: A Scalable Tree Boosting System*”, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16 (2016), pp. 785–794.
- [47] F. Pedregosa et al., “*Scikit-learn: Machine Learning in Python*”, Journal of Machine Learning Research **12**, 2825 (2011), arXiv:1201.0490.
- [48] G. Punzi, “*Sensitivity of searches for new signals and its optimization*”, eConf **C030908**, edited by L. Lyons, R. Mount, and R. Reitmeyer, MODT002 (2003), arXiv:physics/0308063.
- [49] P. Ramachandran, B. Zoph, and Q. V. Le, *Searching for Activation Functions*, 2017, arXiv:1710.05941.

## Acknowledgement

First and foremost, I would like to thank my supervisor Dr. Sophie Hollitt for her immeasurable support with the creation of this thesis. Her patience with me is the reason I really learned a lot while working on this thesis, while her enthusiasm regularly contributed to the fun of this work.

I would like to thank Prof. Dr. Bernhard Spaan for offering me to write this thesis after rather sudden changes and agreeing to be the first reviewer. I also want to thank Prof. Dr. Kevin Kröninger for agreeing to be my second reviewer.

The quality of this thesis was highly improved by all the colleagues from E5a, who helped with their input in the B2OC, CPV and analysis meetings. I would especially like to thank Antje Mödden, Philipp Ibis, Quentin Führung, Vukan Jevtić and Dr. Vitalii Lisovskyi who all proofread this thesis and contributed very high-quality comments, which were very much needed!

As non LHCb people, the comments from Mirjam Bourgett and Donna Maria Mattern, who also proofread this thesis, especially helped with making things more clear, thank you!

My girlfriend is *top* at helping me clear things up, and getting a straight view of things. This thesis would look very different without her, and not for the better. I like to thank my friends Mirjam, Jonas and Jan Peter for their support and valuable advice during difficult times.

Thanks is also due to my parents, who often endured my monologues about physics and science, even if they had not asked for it. I hope I could make it up with a good cup of coffee once or twice.

Außerdem möchte ich besonders meinen Großeltern danken, die mich stets unterstützen. Der Mittagstisch am Sonntag ist eine willkommene Auszeit vom Studium und bereichert eben dadurch oft dasselbige.

# Eidesstattliche Versicherung (Affidavit)

Neste, Ludwig

205317

Name, Vorname  
(Last name, first name)

Matrikelnr.  
(Enrollment number)

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit mit dem folgenden Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present Bachelor's thesis with the following title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution.

Titel der Bachelorarbeit:  
(Title of the Bachelor's thesis):

„Selection of  $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$  decays for normalisation of the  $B^0 \rightarrow D^0 \bar{D}^0$  branching ratio“

Castrop-Rauxel, der 19.08.2020

Ort, Datum  
(Place, date)

Ludwig Neste

Unterschrift  
(Signature)

## Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG - ). Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

## Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to €50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, section 63, subsection 5 of the North Rhine-Westphalia Higher Education Act (*Hochschulgesetz*). The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:\*

Castrop-Rauxel, der 19.08.2020

Ort, Datum  
(Place, date)

Ludwig Neste

Unterschrift  
(Signature)

\*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.