

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

**Optimization of Flavour Tagging  
Algorithms for the LHCb Experiment**

Kevin Heinicke

30. September 2016

Lehrstuhl für Experimentelle Physik V  
Fakultät Physik  
Technische Universität Dortmund

Erstgutachter	Prof. Dr. Bernhard Spaan
Zweitgutachter	Prof. Dr. Kevin Kröninger
Betreuer	Dr. Julian Wishahi

## Abstract

Studies of  $CP$  violation can be used to test the Standard Model and might give insight into New Physics. Therefore, a wide range of  $CP$  measurements, including time-dependent decay rate measurements, are performed with the LHCb Experiment. Many of these are subject to mixing of neutral  $B$  meson states with their antiparticles. The knowledge of the initial  $B$  flavour is essential in these cases which is why several Flavour Tagging algorithms are used to deduce this information from the available event properties.

These algorithms must be adjusted to changes in the shape of the event properties, resulting from an upgrade of the LHC centre-of-mass energy to  $\sqrt{s} = 13\text{TeV}$ . To simplify this process, the Flavour Tagging software is re-implemented. The tagging power of the muon, electron and kaon tagger is measured based on  $B^+ \rightarrow J/\psi K^+$  data, which is processed within the new framework. It is found to be  $(0.782 \pm 0.018)\%$ ,  $(0.243 \pm 0.011)\%$  and  $(0.649 \pm 0.020)\%$  for Run 1 data, respectively, and  $(0.67 \pm 0.05)\%$ ,  $(0.134 \pm 0.019)\%$  and  $(0.54 \pm 0.04)\%$  for Run 2 data.

Furthermore a new inclusive strategy for the muon tagger is implemented that improves its tagging power to  $\varepsilon_{\text{eff}} = (1.09 \pm 0.04)\%$  on Run 1 data and  $\varepsilon_{\text{eff}} = (0.83 \pm 0.06)\%$  on Run 2 data.

## Kurzfassung

Untersuchungen von  $CP$ -Verletzung stellen einen wichtigen Test des Standardmodells dar und können Hinweise auf Neue Physik liefern. Daher werden verschiedene Messungen dieser Art mit dem LHCb-Experiment durchgeführt. Unter anderem sind dies zeitaufgelöste Messungen von Zerfallsraten die oft der Mischung neutraler  $B$ -Mesonen unterliegen. Hierbei ist die Information des initialen  $B$ -Flavours essentiell, welche von mehreren Flavour Tagging Algorithmen aus den Eigenschaften der Teilchenkollisionen ermittelt wird.

Diese Algorithmen müssen an Änderungen in den Verteilungen dieser Eigenschaften angepasst werden, die aus der im Zuge des ersten LHC Upgrades erhöhten Schwerpunktennergie von  $\sqrt{s} = 13\text{TeV}$  hervorgehen. Um dieses Unterfangen zu vereinfachen, wird die Flavour Tagging Software re-implementiert. Die Tagging Power des Myon-, Elektron- und Kaon Taggers wird auf  $B^+ \rightarrow J/\psi K^+$  Daten bestimmt, die in dem neuen Framework prozessiert werden. Sie betragen jeweils  $(0.782 \pm 0.018)\%$ ,  $(0.243 \pm 0.011)\%$  und  $(0.649 \pm 0.020)\%$  für Run 1 Daten und  $(0.67 \pm 0.05)\%$ ,  $(0.134 \pm 0.019)\%$  und  $(0.54 \pm 0.04)\%$  für Run 2 Daten.

Darüber hinaus wird ein neuer, inklusiver Ansatz für den Myon Tagger implementiert, der dessen Tagging Power auf bis zu  $\varepsilon_{\text{eff}} = (1.09 \pm 0.04)\%$  für Run 1 Daten und  $\varepsilon_{\text{eff}} = (0.83 \pm 0.06)\%$  für Run 2 Daten verbessert.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Theoretical Framework</b>	<b>3</b>
2.1 <i>CP</i> Violation in the Standard Model of Particle Physics . . . . .	3
2.1.1 Basics of the Standard Model and the CKM Triangle . . .	3
2.1.2 Mixing of <i>B</i> Mesons . . . . .	6
2.1.3 <i>CP</i> Violation . . . . .	8
2.2 Machine Learning and Data Analysis in Particle Physics . . . .	10
2.2.1 Decision Trees and Boosting . . . . .	10
2.2.2 Calibration of Probabilities . . . . .	12
2.2.3 Model Fitting and $\chi^2$ Weights . . . . .	12
<b>3 The LHCb Experiment at the Large Hadron Collider</b>	<b>15</b>
3.1 The LHC . . . . .	15
3.2 The LHCb Experiment . . . . .	16
3.2.1 Tracking System . . . . .	17
3.2.2 Particle Identification System . . . . .	21
3.2.3 Trigger System . . . . .	23
3.2.4 Software Stack . . . . .	24
<b>4 Flavour Tagging of <i>B</i> Mesons</b>	<b>27</b>
4.1 Variables and Definitions of Flavour Tagging . . . . .	27
4.2 Flavour Tagging Algorithms . . . . .	31
4.2.1 Opposite Side Single Track Taggers . . . . .	32
4.2.2 Vertex Charge Tagger . . . . .	35
4.2.3 Charm Tagger . . . . .	35
4.2.4 Same Side Proton and Pion Tagger . . . . .	36
4.2.5 Same Side Kaon Tagger . . . . .	36
<b>5 Reimplementation of the LHCb Flavour Tagging Software</b>	<b>37</b>
5.1 Structure of LHCb Data Analysis Algorithms . . . . .	38
5.2 Re-factoring of the Flavour Tagging Software Package . . . . .	39
<b>6 Optimization Studies of Single Track Taggers for LHCb Run 2</b>	<b>41</b>
6.1 Dataset properties and Run Period Differences . . . . .	41
6.2 Muon Tagger with Logistic Calibration . . . . .	46
6.3 Electron and Kaon Taggers with Logistic Calibration . . . . .	52
6.4 Studies of a BDT-based, Inclusive Muon Tagger . . . . .	55
<b>7 Conclusion and Outlook</b>	<b>61</b>
<b>Bibliography</b>	<b>63</b>



# 1 Introduction

Today's efforts in the field of particle physics, to understand and describe the fundamental laws of matter (and antimatter) are unmatched. The world's biggest machine, the Large Hadron Collider with its four major experiments provides experimental data to thousands of researchers to test, verify, and extend the Standard Model of particle physics, which precisely describes fundamental particles and their interactions. Despite the huge success of this theory, many questions remain unanswered. For example, it does not explain the origin of dark matter and dark energy, which seem to be the major energy content of our universe; it does not provide an explanation of the imbalance of matter and antimatter in the observable universe; gravitation is neglected in the Standard Model – it is no Grand Unified Theory and no such theory has been experimentally verified so far.

In addition to studies of rare decays of  $b$  and  $c$  hadrons, the LHCb Experiment is optimized for the measurement of  $CP$  violation, a phenomenon allowing tests of the Standard Model and providing a way to discover New Physics. Many of these measurements are performed on data of neutral  $B$  meson decays, which tend to oscillate between their particle and antiparticle states, such as  $B^0$  and  $\bar{B}^0$ . In the LHCb detector, these particles only travel few millimetres before they decay into energetically favourable, lighter particles, which can effectively be measured with the detector's components. Since  $CP$  violation might affect the decay rates of the  $B$  mesons and their initial quark content cannot be deduced from the final state, the information whether the meson contained a  $b$  quark or a  $\bar{b}$  quark at production needs to be known. The aggregation of this flavour information is called Flavour Tagging.

The Flavour Tagging software bundles several algorithms, which deduce the tag information from the event information of secondary particles that are produced in the hadronization process of the signal decay or the non-signal  $b$  quark. Each algorithm provides a tag prediction and an estimation for the probability of that tag to be wrong, called mistag estimate. These mistag estimates are computed with multivariate analysis tools (MVAs) such as Boosted Decision Trees or Neural Networks. Furthermore MVA outputs usually need to be calibrated to provide probabilistic estimations. Therefore, to use the Flavour Tagging algorithms for an analysis, its MVAs need to be trained with data that is kinematically similar to the signal decay and for which the true tags are known.

This training and optimization process is especially required after an upgrade of the centre-of-mass energy of the LHC. The MVA and calibration properties, determined on Run 1 data cannot be applied for Run 2 data

and all algorithms need to be re-trained for the new physics conditions. To simplify this process and make use of newly developed external MVA packages, the Flavour Tagging software is re-implemented.

The differing performance of Flavour Tagging before and after the upgrade will be shown for the example of the Single Track taggers, while the data for this comparison is processed in the re-implemented software. Furthermore, a study to re-optimize the classic approach of the Muon tagger is performed. While the changed physics conditions seem to generally decrease the performance of Flavour Tagging in the case of the muon tagger, a new, inclusive approach for this tagging algorithm is presented.

In this thesis, first the theoretical framework will be introduced, describing relevant parts of the Standard Model of particle physics as well as data analysis strategies used for further studies. Afterwards, the experimental foundation is explained, introducing the Large Hadron Collider and the LHCb Experiment. The different Flavour Tagging algorithms and their characteristics are described in the next chapter. The following chapter briefly outlines the design of the re-implemented Flavour Tagging software within the LHCb software stack. An optimization study is presented afterwards, based on the re-implemented software, and comparing the performance of single track taggers for Run 1 and Run 2 data. Furthermore, two approaches to improve the muon tagger algorithm are described. A conclusion of the results, together with an outlook for further studies is given in the last chapter.

## 2 Theoretical Framework

Modern particle physics uses powerful accelerators and detectors to gather huge amounts of data of fundamental processes of nature. The data is used to verify or falsify detailed theoretical predictions and give hints towards the creation of new theoretical models.

In the following sections, the parts of the most advanced and widely proofed theoretical model for particle physics that are relevant for this thesis will be introduced. Afterwards, several machine learning algorithms that are widely used within the particle physics community as well as within this thesis will be explained.

### 2.1 *CP* Violation in the Standard Model of Particle Physics

The Standard Model of particle physics (SM) is today's theory of choice to describe all known particles and their interactions. It evolved in the 1960's and has been experimentally verified multiple times since then. The last particle that was predicted by the SM and has been discovered by the LHC experiments ATLAS and CMS in 2012 is the Higgs Boson [1, 2].

Despite its success, some open questions cannot be answered by this theory. For example, it does not give any explanation on the origin of dark matter or dark energy, although these seem to contribute the major fraction of energy within the known universe. More and more precise measurements of the SM predictions therefore might lead to some inconsistencies, which would be indicators for New Physics [3].

The basic principles of the SM will be shown in the following sections. Furthermore the phenomenon of *CP* violation will be introduced, which is intensively studied by the LHCb-collaboration and requires a Flavour Tagging software discussed in this thesis.

#### 2.1.1 Basics of the Standard Model and the CKM Triangle

The Standard Model describes all fundamental particles in a relativistic quantum field theory, following the principles of local gauge invariance and symmetries. Particles are represented in terms of field operators  $\phi(x)$  and their dynamics is described by a Lagrange function  $\mathcal{L}(\phi(x), \partial_\mu \phi(x))$ . By requiring Poincaré invariance, the SM is symmetric under continuous spatial translations, rotations and Lorentz transformations. Additionally, the SM contains three discrete symmetries: *P*, the inversion of all spatial coordinates, *C*, which inverts all additive quantum numbers, and *T*, the

inversion of time. Due to the Lorentz invariance of the SM, the combination *CPT* is required to be conserved [4].

Fundamental particles can generally be grouped into fermions, with a spin of  $\frac{1}{2}$  or bosons, with an integer spin. The fermions can be arranged in three families of two particle doublets. Each family contains a doublet of an up-type quark  $u$ ,  $c$  or  $t$  with its down-type quark  $d$ ,  $s$  or  $b$  as well as a doublet of a charged lepton  $e^-$ ,  $\mu^-$  or  $\tau^-$  and its neutral partner  $\nu_e$ ,  $\nu_\mu$  or  $\nu_\tau$  respectively:

$$\begin{array}{l} \text{quarks} \\ \text{leptons} \end{array} \quad \begin{array}{ccc} \begin{pmatrix} u \\ d \end{pmatrix} & \begin{pmatrix} c \\ s \end{pmatrix} & \begin{pmatrix} t \\ b \end{pmatrix} \\ \begin{pmatrix} e^- \\ \nu_e \end{pmatrix} & \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} & \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix} \end{array}$$

The up-type quarks carry a  $+2/3$ -charge while the down-type quarks have a charge of  $-1/3$ . Furthermore they carry a colour charge of the values ‘red’, ‘green’ or ‘blue’. The quark-type is also referred to as the quarks flavour. The masses show a hierarchy of  $m_u < m_d < m_s < m_c < m_b < m_t$  ranging over several orders of magnitude from the lowest mass  $m_u = 2.3 \text{ MeV}/c^2$  to the highest mass  $m_t = 173 \text{ GeV}/c^2$  [5]. The leptons  $e^-$ ,  $\mu^-$  and  $\tau^-$  carry a charge of  $-1$  and their neutrino-partners  $\nu_e$ ,  $\nu_\mu$  and  $\nu_\tau$  are uncharged. While the charged lepton masses range from the lowest mass  $m_e = 511 \text{ keV}/c^2$  to the highest mass  $m_\tau = 1.8 \text{ GeV}/c^2$ , the neutrino masses are very light and have not been measured yet [5]. The overall hierarchy of these fermions can not be explained within the SM since it does not predict any of these values.

Additionally to these, every particle has an antimatter partner with the exact same properties except for all additive quantum numbers being inverted, such as charge, colour charge or lepton family number. The charge-conjugation *C* transforms matter- into antimatter states and vice versa. Antiparticles are usually marked with a bar, for example  $\bar{\nu}_e$  is the antiparticle of  $\nu_e$ .

Matter that is naturally occurring in everyday life is usually made from the first two doublets containing the lightest particles: the up- and down-quark as well as the electron and electron-neutrino. A hydrogen atom for example has a core of a single proton  $p$  consisting of three quarks  $|p\rangle = |uud\rangle$  and a single electron  $e^-$ . In general, multiple quarks combine to a hadron, such that the charge of the resulting particle is integer and the colour charge is ‘white’ either by combining all three colour or anticolour charges or by combining a colour charge with its own anticolour. Two-quark states are called mesons, three-quark states baryons. The LHCb-collaboration recently also discovered the first pentaquark state, consisting of five quarks and has found evidence for four quark states [6, 7, 8].

All interactions between quarks and leptons are carried out by the force-mediating particles of the three fundamental forces: The electromagnetic force couples to the charge of particles via exchange of massless spin-1 photons  $\gamma$ .

The strong force is described in the  $SU(3)$  colour symmetry group by Quantum Chromodynamics [9]. It couples to the colour charge of quarks via exchange of massless spin-1 gluons  $g$ . These gluons carry colour charge themselves and thus are capable of self-coupling. The strong force usually acts on subatomic ranges of up to  $10^{-15}$  m. An effect that is especially relevant for hadron-physics like the LHC-physics is ‘colour confinement’: If a pair of quarks is separated, the creation of a new quark-anti quark pair is at some point energetically favourable and the separated quarks will combine with these spontaneously created quarks. As a consequence single quarks cannot be observed, instead a ‘jet’ of hadrons is seen, which is why this process is also called ‘hadronization’.

The weak force couples to the left-handed doublets of quarks and leptons via exchange of neutral  $Z$  or charged  $W^\pm$  bosons. Both particles are massive vector-bosons with  $m_Z = 91.2 \text{ GeV}/c^2$  and  $m_{W^\pm} = 80.4 \text{ GeV}/c^2$ . Quantum numbers are conserved in the exchange of  $Z$  bosons, which is why these provide the only mechanism for elastic scattering of neutrinos. In contrast, exchanges of  $W^\pm$  bosons change the flavour number of quarks. The coupling of the weak force cannot be described in terms of the mass-eigenstates of quarks and leptons and thus these states need to be transformed into eigenstates of the weak interaction. Neutrinos are assumed massless in the SM, thus the mass-eigenstates of the leptons can trivially be written as eigenstates of the weak force like  $\ell' = \ell$ , where  $\ell$  is any lepton <sup>1</sup>. However, quarks cannot be assumed massless and therefore are transformed to weak force eigenstates via the Cabibbo-Kobayashi-Maskawa matrix  $V_{\text{CKM}}$ . By convention, the transformation is applied to the down-type quarks  $d$ ,  $s$  and  $b$ :

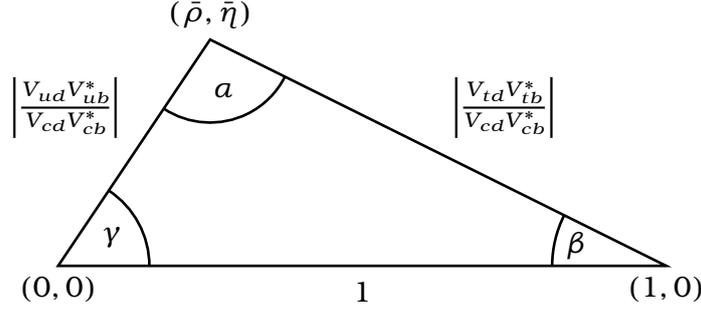
$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (2.1)$$

By construction, the CKM-Matrix is complex, unitary and defined by four parameters. The squared norm of one element of the matrix,  $|V_{ij}|^2$  is proportional to the probability of a flavour-transition  $i \leftrightarrow j$ . The CKM-matrix can be written in the Wolfenstein parametrization [11] in terms of three parameters  $A$ ,  $\rho$ ,  $\lambda$  and a complex phase  $\eta$ :

$$V_{\text{CKM}} \approx \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix}, \quad (2.2)$$

$$\text{with } \lambda = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \quad A\lambda^2 = \lambda \left| \frac{V_{cb}}{V_{us}} \right|, \quad V_{ub}^* = A\lambda^3(\rho + i\eta).$$

<sup>1</sup> Experimental evidence for neutrino oscillation has been seen and requires massive neutrinos, which can only be described with extensions to the SM, e.g. the Seesaw mechanism [10]. The neutrino mass eigenstates are transformed into eigenstates of the weak interaction through the PMNS matrix, which is the leptonic analogue to the quarks’ CKM matrix.



**Figure 2.1:** The most commonly used CKM triangle. It is determined by the CKM matrix elements. Its angles  $\alpha$ ,  $\beta$  and  $\gamma$ , the point  $(\bar{\rho}, \bar{\eta})$  and the lengths of its two non-fixed sides can be measured experimentally. The area of the triangle is determined by the complex phase of the CKM matrix and is a measure for the  $CP$  violation in the SM.

The unitarity of the matrix implies  $\sum_i V_{ij} V_{ik}^* = \delta_{jk}$  and  $\sum_j V_{ij} V_{kj}^* = \delta_{ik}$  which can be represented as triangles in the complex plane, if  $j \neq k$  or  $i \neq k$ , respectively. The most commonly used triangle follows from the relation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \quad (2.3)$$

Hereby, the angles of the CKM triangle are defined as

$$\alpha = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \beta = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \gamma = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right). \quad (2.4)$$

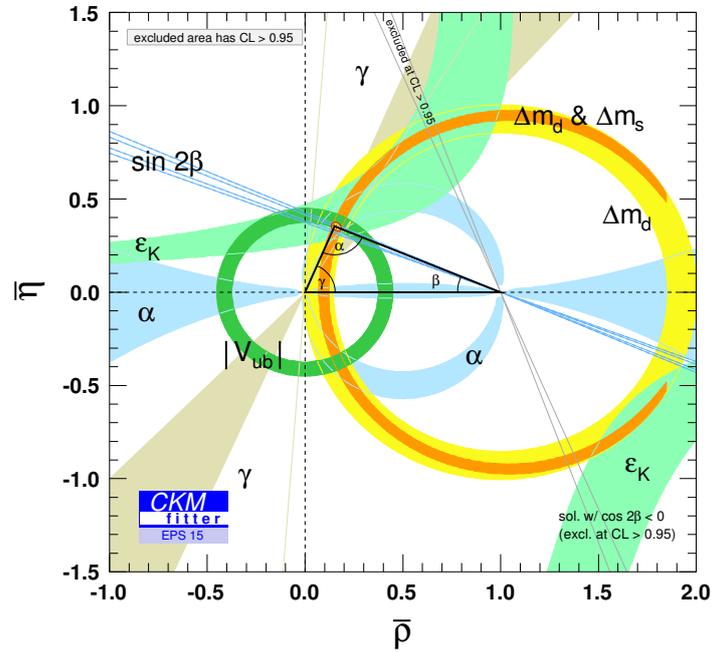
All angles as well as the non-fixed sides and the point  $(\rho, \eta)$  can be measured experimentally. The parameters have been determined to be

$$\begin{aligned} \lambda &= 0.22537 \pm 0.00061, & A &= 0.814^{+0.023}_{-0.024} \\ \bar{\rho} &= 0.117 \pm 0.021, & \bar{\eta} &= 0.353 \pm 0.013, \end{aligned}$$

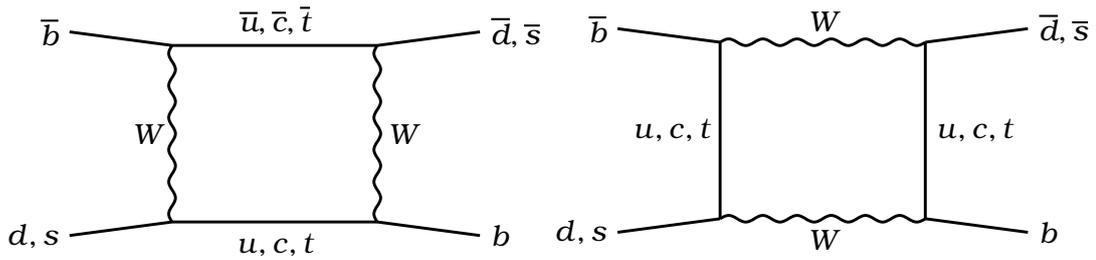
and the CKM triangle with its parameters is shown in Figure 2.1 [5]. The CKM triangle is therefore overdetermined, which allows for precise measurements of and cross-checks between its parameters. A deviation from the SM expectation would be an indicator for New Physics. As shown in Figure 2.2, the constraints on the CKM parameters have already been measured with high precision, and no inconsistencies with the SM have been found. Nonetheless, small tensions of up to  $2.2\sigma$  require even more precise measurements [12].

### 2.1.2 Mixing of $B$ Mesons

The previously introduced CKM matrix allows for weak flavour-changing neutral currents of pairs of  $W$  particles in the SM. As a result, neutral mesons, such as the  $|B^0\rangle = |\bar{b}d\rangle$  or  $|B_s^0\rangle = |\bar{b}s\rangle$  meson can oscillate between particle and antiparticle states as shown in Figure 2.3. This has first been observed in the system of neutral  $K^0$  mesons by Cronin and Fitch in 1964 [13]. To simplify the notation of the following equations,  $P^0$  and  $\bar{P}^0$  will be



**Figure 2.2:** Current experimental constraints on the different CKM triangle parameters. The coloured areas indicate the 68 % confidence levels resulting from the average of all available measurements [12].



**Figure 2.3:** Mixing of  $B^0$  or  $B_s^0$  with  $\bar{B}^0$  or  $\bar{B}_s^0$  mesons. Via the exchange of  $W^\pm$  bosons, the flavour of the initial mesons can change. This process is dominated by top quarks due to their high mass. A possible contribution of yet unknown heavier particles instead of the top quarks could indicate New Physics. Mixing allows for different types of  $CP$  violation, which are described in section 2.1.3.

used for  $B_s^0$  and  $\bar{B}_s^0$  or for  $B^0$  and  $\bar{B}^0$  mesons, respectively.

The decay time wise evolution of a system of neutral  $B$  mesons can be described by the Schrödinger equation

$$i\frac{d}{dt}\begin{pmatrix} P^0 \\ \bar{P}^0 \end{pmatrix} = \mathbf{H}\begin{pmatrix} P^0 \\ \bar{P}^0 \end{pmatrix} = (\mathbf{M} - i\mathbf{\Gamma})\begin{pmatrix} P^0 \\ \bar{P}^0 \end{pmatrix}, \quad (2.5)$$

with the Hamilton operator  $\mathbf{H}$  and the hermitian matrices  $\mathbf{M}$  and  $\mathbf{\Gamma}$ . As a consequence of the  $CPT$  theorem, the diagonal elements of  $\mathbf{M}$  and  $\mathbf{\Gamma}$  fulfil  $M_{11} = M_{22}$  and  $\Gamma_{11} = \Gamma_{22}$ . The eigenvalues  $\mu$  of the Hamiltonian then are  $\mu_{1,2} = M_{1,2} - i/2\Gamma_{1,2}$ , with the masses  $m_{1,2}$  and decay widths  $\Gamma_{1,2}$  of the mass eigenstates of this system. In case of  $B$  mesons these are referred to as  $m_H$  and  $\Gamma_H$  or  $m_L$  and  $\Gamma_L$  for the high and low mass and decay widths. Furthermore, a mass difference  $\Delta m = m_H - m_L$  and a decay width difference  $\Delta\Gamma = \Gamma_H - \Gamma_L$  can be defined. The mass eigenstates can then be written as a combination of the flavour eigenstates:

$$|B_L\rangle = p|P^0\rangle + q|\bar{P}^0\rangle, \quad |B_H\rangle = p|P^0\rangle - q|\bar{P}^0\rangle, \quad (2.6)$$

where the complex factors  $p$  and  $q$  obey the normalization  $|p|^2 + |q|^2 = 1$ . Furthermore, the decay amplitudes  $A$  of a  $B$  meson into a final state  $f$  follow to be

$$\begin{aligned} A_f &= \langle f|H|P^0\rangle, & A_{\bar{f}} &= \langle \bar{f}|H|P^0\rangle, \\ \bar{A}_f &= \langle f|H|\bar{P}^0\rangle, & \bar{A}_{\bar{f}} &= \langle \bar{f}|H|\bar{P}^0\rangle, \end{aligned} \quad (2.7)$$

where  $H$  depicts the Hamilton operator. Mixing processes might involve heavier particles than the known top quark and might therefore give access to New Physics. The importance of mixing for the phenomena of  $CP$  violation will be pointed out in the next section.

### 2.1.3 $CP$ Violation

While the SM is symmetric under  $CPT$  transformation and no violation of this has been observed [14, 15], the transformations  $C$ ,  $P$ ,  $T$  and the combination  $CP$  are violated [16, 17, 18, 19]. This effect is theoretically described by the CKM parameter  $\eta$ , such that the size of the CKM triangle measures the strength of  $CP$  violation in the SM [20]. Given Equations (2.6) and (2.7), three types of  $CP$  violation can be defined:

- Direct  $CP$  violation – also called  $CP$  violation in the decay – leads to different decay amplitudes

$$|A_f| \neq |\bar{A}_{\bar{f}}| \quad \text{or} \quad |A_{\bar{f}}| \neq |\bar{A}_f|.$$

This is the only type of  $CP$  violation that can be observed in charged  $B$  meson decays and has first been observed by the LHCb collaboration in the  $B_s^0 \rightarrow K^- \pi^+$  and  $B^0 \rightarrow K^+ \pi^-$  decay channels [21].

- Indirect  $CP$  violation – also called  $CP$  violation in the mixing – leads to different transition rates from particle into antiparticle state and vice versa. This is the case if

$$\left| \frac{p}{q} \right| \neq 1.$$

The parameter  $\alpha_{\text{sl}}$ , defined as

$$\alpha_{\text{sl}} = \frac{|p/q|^2 - |q/p|^2}{|p/q|^2 + |q/p|^2}$$

has been measured by the LHCb collaboration in the  $B^0$  and  $B_s^0$  system, to be  $\alpha_{\text{sl}}^d = -0.02 \pm 0.19 \pm 0.30\%$  and  $\alpha_{\text{sl}}^s = -0.39 \pm 0.26 \pm 0.20\%$ , while the latter value corresponds to a  $1.5\sigma$  deviation from the SM expectation of  $\alpha_{\text{sl}} = 0$  [22, 23].

- Even if none of the above effects occur independently, there can be  $CP$  violation in the interference of decay,  $P^0 \rightarrow f$ , and mixing,  $P^0 \rightarrow \bar{P}^0 \rightarrow f$ , when

$$\text{Im } \lambda_f \neq 1, \quad \text{with } \lambda_f = \eta_f \frac{q \bar{A}_f}{p A_f},$$

where  $\eta_f = \pm 1$  is the  $CP$  eigenvalue of the final state.

All types of  $CP$  violation are usually constrained by measurements of the asymmetry  $\mathcal{A}_f$  of the decay rates  $\Gamma(P^0/\bar{P}^0 \rightarrow f/\bar{f})$ . In the case of  $CP$  violation in interference of decay and mixing, the decay-time-dependent decay rates of signal candidates  $P^0$  or  $\bar{P}^0$  into a common final state  $f$ ,  $\Gamma(P^0/\bar{P}^0(t) \rightarrow f)$  are studied. The asymmetry of these rates is defined as

$$\mathcal{A}_f(t) = \frac{\Gamma(B^0(t) \rightarrow f) - \Gamma(\bar{B}^0(t) \rightarrow f)}{\Gamma(B^0(t) \rightarrow f) + \Gamma(\bar{B}^0(t) \rightarrow f)}. \quad (2.8)$$

An example of this kind of analyses is the latest LHCb  $CP$  measurement with  $B^0 \rightarrow D^+ D^-$  decays [19]. In this specific case, where no  $CP$  violation in mixing occurs and the decay widths difference  $\Delta\Gamma$  of the mass eigenstates is approximately zero, the  $CP$  observables  $S$  and  $C$ , are defined as

$$S_f = \frac{2 \text{Im } \lambda_f}{1 + |\lambda_f|^2} \quad \text{and} \quad C_f = \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2}, \quad (2.9)$$

and linked to the CKM angle  $\beta$  via  $S/\sqrt{1 - C^2} = -\sin(2\beta + \delta)$  with a phase shift  $\delta$ , representing higher order contributions. Assuming no direct  $CP$  violation, the SM predicts the parameters  $C$  and  $\delta$  to vanish, such that  $S = -\sin(2\beta)$ . The asymmetry can be written as

$$\mathcal{A}_f(t) = S_f \sin(\Delta m t) - C_f \cos(\Delta m t). \quad (2.10)$$

As shown here for a specific decay channel, the initial flavour of the  $B$  mesons needs to be known for studies of  $CP$  violation. In general, this is the case for any time-dependent analysis of  $CP$  violation in the interference of decay and mixing with neutral  $B$  mesons. Flavour Tagging is therefore an essential ingredient for a variety of analyses.

## 2.2 Machine Learning and Data Analysis in Particle Physics

To test, verify or falsify theories, such as the models described above, the experiment's data needs to be statistically analysed. Since processes, which are described by the SM and tested with experiments such as LHCb, superimpose each other, the relevant information for a given analysis needs to be extracted. An example would be the selection of a specific decay channel such as  $B^+ \rightarrow J/\psi K^+$  or single particles, e.g. a muon out of that decay chain. This classification problem can be solved with multivariate analysis tools, which are introduced in the following sections.

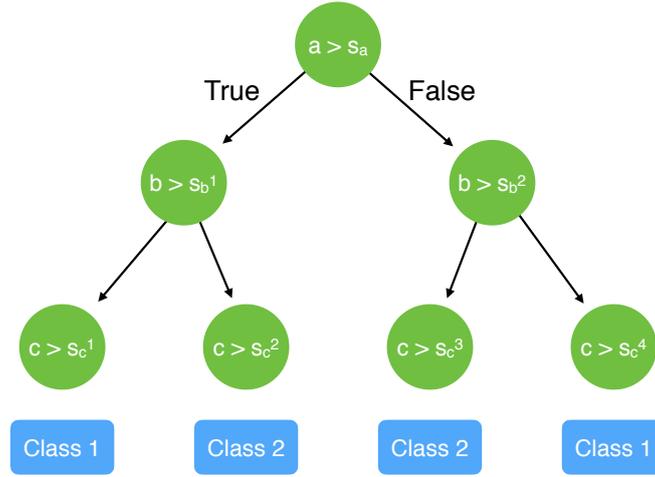
Each of these approaches adjusts a mathematical model such that the class, e.g. 'signal' or 'background' of a given vector of numbers that correspond to e.g. properties of particles is predicted. Therefore data of the same shape, but with the known class information  $y$  is passed to the model and its parameters  $\theta$  are changed until an objective function  $\text{Obj}(\theta)$ , describing how well the model predicts the data, has reached a minimum. This process is called training and the objective function is often chosen to be the mean squared error

$$\text{Obj}(\theta) = L(\theta) = \sum_i (y_i - \hat{y}_i)^2, \quad (2.11)$$

where  $\hat{y}_i$  is the predicted and  $y_i$  the true class of the  $i$ -th data sample. The general schema is often referred to as 'Machine Learning' and has become especially popular in the field of image pattern recognition in the last decade due to improvements of parallel computing performance. Variables that are used for training of Machine Learning algorithms will be called 'features' of that algorithm in the following sections. The predicted class will be referred to as the 'target' of the training.

### 2.2.1 Decision Trees and Boosting

One intuitive Machine Learning algorithm is based on ensembles of decision trees. Starting from one feature, the training data is split into two sub samples, based on a cut on that feature that minimizes the objective function. Afterwards, another feature is taken into account within both subsets and so on. A so formed decision tree is shown in figure Figure 2.4 with a depth of three cuts. Using an arbitrary number of cuts would lead to a single decision tree which could potentially minimize the objective function to zero, but would also represent the whole training dataset and therefore



**Figure 2.4:** Sketch of a three level decision tree, classifying data consisting of features  $\vec{x}_i = (a_i, b_i, c_i, \dots)^T$  into the two classes  $y_i = (\text{Class 1} \vee \text{Class 2})$  by applying different rectangular cuts  $s$ . Machine Learning algorithms such as Boosting or Random Forests are ensembles of Decision Trees and can be used to predict the class of a data sample, once trained.

has no predictive abilities for unknown data. This is called over-fitting and can be tested by applying a trained model to unknown data and comparing their performances. To avoid over-fitting, multiple, shallow decision trees can be combined. If their output is averaged and their feature sets are chosen randomly, the model is called a ‘Random Forest’.

Instead of randomly choosing and combining Decision Trees, a number of  $M$  possible trees  $\mathcal{F}$  of a given depth can be combined with a set of weights  $\theta_m$ , such that the overall model to predict a given event  $i$  can be represented as a weighted combination of these trees:

$$\hat{y}_i = \sum_m^M \theta_m f_m(x_i) \quad \text{with } f_m \in \mathcal{F}. \quad (2.12)$$

This approach is called ‘boosting’. The weights are chosen to minimize the objective function.

The boosting method, used in this thesis, takes the complexity of the model into account with an additional term within the objective function:

$$\text{Obj}(\theta) = L(\theta) + \Omega(\theta), \quad (2.13)$$

where  $\Omega$  is often chosen to be the L2 norm  $\|w_m\|_2 = \sqrt{\sum w_m^2}$  of the decision tree weights  $w_m$ .

## 2.2.2 Calibration of Probabilities

Although, MVAs can predict whether its more likely for a set of features to be of one class or another, after they have been trained, their prediction can not necessarily be interpreted as a proper probability for that class. Often, a simple logistic transformation of the form

$$p_{\text{raw}} \rightarrow \frac{1}{1 - e^{p_{\text{raw}}}} \quad (2.14)$$

is applied to the MVA output value  $p_{\text{raw}}$ , which is shown as a red curve in Figure 2.5. This transformation only ensures that all values are within the range  $[0, 1]$  as expected for a probability estimation. For proper probability values, the true number  $N_c^k$  of a predicted class  $c$  within a certain range  $k$  of prediction values should be comparable to the mean predicted value within this range:

$$\frac{N_1^k}{N^k} \approx \frac{\sum_i^{N^k} p_{i,\text{raw}}^k}{N^k}. \quad (2.15)$$

If this does not hold due to non-probabilistic MVA output values, these can be calibrated by applying a logistic binary regression. Here, a function of the form (2.14) with additional parameters is used:

$$S(p_{\text{raw}}) = \frac{1}{1 - \exp\left[\sum_j^k \beta_j p_{\text{raw}}^j\right]}. \quad (2.16)$$

The parameter  $\beta_j$  can be determined by a maximum likelihood fit, which is described in the following section. The polynomial sum of the BDT value helps to reflect non-linear relations of the BDT values and the true probabilities. An example with  $k = 4$  non-zero parameters is given with the blue curve in Figure 2.5.

## 2.2.3 Model Fitting and $\mathcal{L}$ Weights

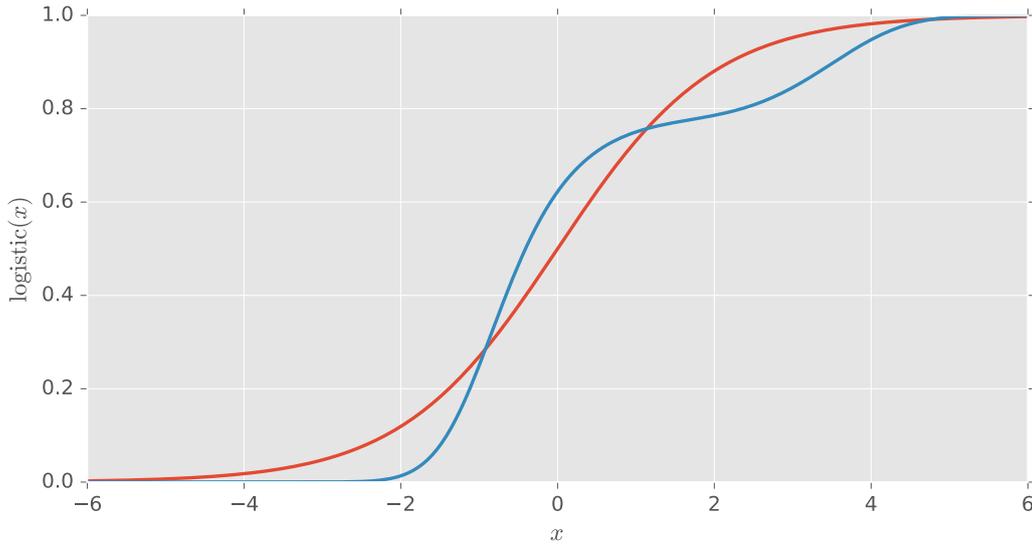
To measure parameters of a specific model inside a given dataset, usually the parameters of that model are adjusted such that the model best fits to the data. Therefore, often the likelihood  $\mathcal{L}$  of the model  $\mathcal{A}(x, p)$  with a set of parameters  $p$ , given the data  $x$ , is maximized:

$$p_{\text{best}} = \operatorname{argmax}_p [\mathcal{L}(\mathcal{A}|x)], \quad (2.17)$$

with the likelihood being the product of every data point's likelihood

$$\mathcal{L}(\mathcal{A}|x) = \prod_i^N \mathcal{L}(\mathcal{A}|x_i). \quad (2.18)$$

The  $\mathcal{L}$ Plot method [24] uses a maximum likelihood fit to determine the contributions of different components to a one-dimensional variable distribution. The distribution of this discriminating variable needs to be well-modelled



**Figure 2.5:** Logistic transformations of the parameter  $x$ . The red curve shows an example for the transformation, given in Equation 2.14, while the blue curve shows Equation 2.16 with parameters  $\beta_0 = 0.5$ ,  $\beta_1 = 1$ ,  $\beta_2 = -0.5$ , and  $\beta_3 = 0.1$ .

within all components that should be determined, which is the case for e.g. the invariant mass of a signal candidate particle. By fitting the model of this variable, weights can be determined that indicate the contribution of each component to that distribution. Assuming that the discriminating variable is uncorrelated with the other variables, the weights can then be applied to those which unfolds each component within the distributions.



## 3 The LHCb Experiment at the Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest particle accelerator. It serves its four major experiments ALICE, ATLAS, CMS, and LHCb with colliding protons of very high energy. The accelerator is part of the European Organization for Nuclear Research (CERN) in Geneva. CERN was founded by twelve European countries in 1954 and its first synchrocyclotron accelerator started operation in 1964. Currently, CERN has 22 member states and eleven associated members and observers [25]. The laboratory currently employs around 3200 employees and has approximately 13 000 collaborator scientists from its member states [26].

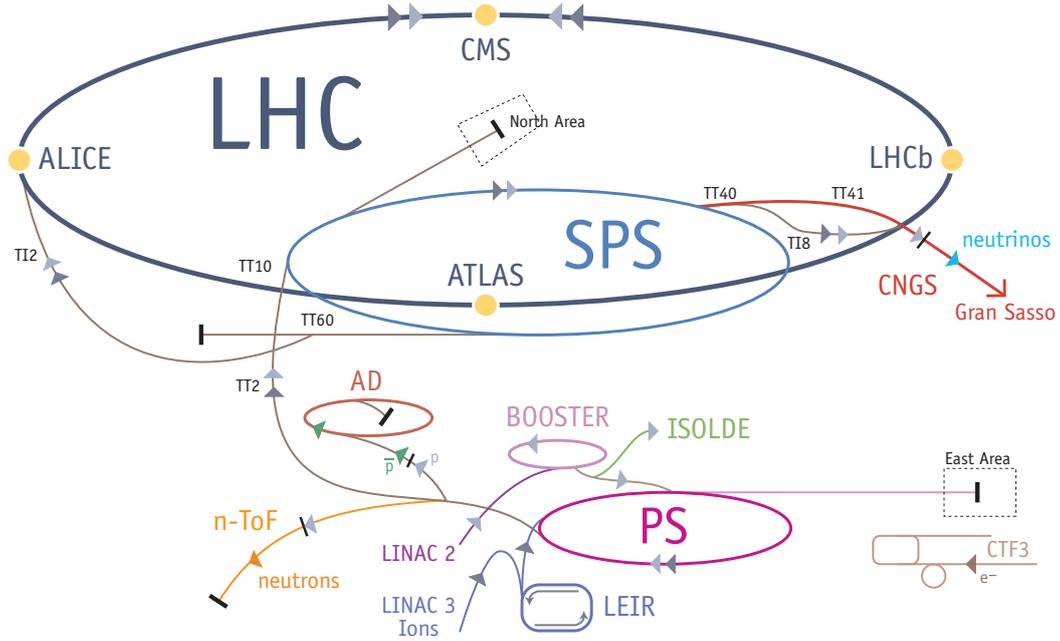
The LHC accelerator and the LHCb experiment will be introduced in more detail in the following sections.

### 3.1 The LHC

The Large Hadron Collider is the successor accelerator of the Large Electron Positron Collider (LEP) whose structures were built between 1983 and 1989 [27]. The LHC hardware has been installed in the 27 km-circumference tunnel, 100 m below ground level, during the early 2000s. It first started operation in 2008 [28].

Under nominal operation conditions the protons are fed into the LHC ring by a chain of pre-accelerators at a beam energy of 450 GeV. The accelerator chain consists of the Linear Accelerator 2 (LINAC2) and several synchrotrons, increasing in size and energy (Proton Synchrotron Booster, PSB; Proton Synchrotron, PS; Super Proton Synchrotron, SPS), which are depicted in Figure 3.1. The protons are accelerated opposite, in two separate beam pipes. A total number of 16 radiofrequency cavities increase their centre-of-mass energy up to  $\sqrt{s} = 14$  TeV. The proton-beams are kept on a circular trajectory with the help of 1232 superconducting dipole magnets, producing magnetic fields of up to 8.3 T with a 11.85 kA current. The beam shape and stability is controlled with over 8000 additional magnets. Completely filled up, both beams will contain 2808 bunches of approximately  $10^{11}$  protons each. The distance between two bunches is about 7 m, which results in a temporal bunch spacing of 25 ns [29, 30].

The beam tubes are merged in four points of the LHC and the beams' usual diameters of several millimetres are magnetically focused down to 16  $\mu\text{m}$  and directed onto an intersecting trajectory [31].



**Figure 3.1:** Schematic view of the Large Hadron Collider (LHC) and its pre-accelerator chain, consisting of Linear Accelerator 2 (LINAC2), Proton Synchrotron Booster (PSB), Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) [30].

This leads to four interaction points, at which the detectors of the major experiments mentioned above are positioned. The instantaneous luminosity

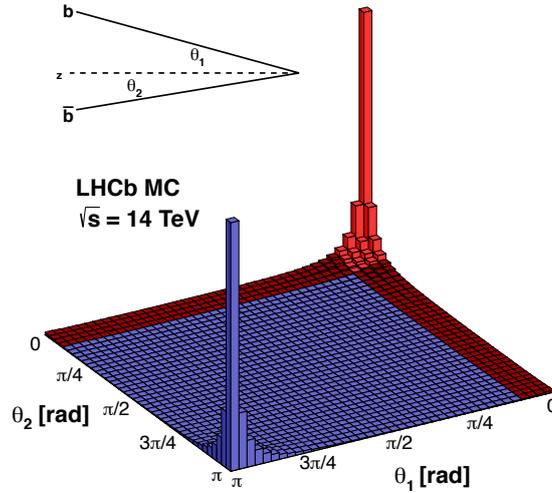
$$\mathcal{L} = \frac{1}{\sigma} \frac{dR}{dt} \quad (3.1)$$

can reach values up to  $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  at these points, where  $\sigma$  is the cross-section of some event type and  $dR/dt$  is its rate.

Currently, the LHC is operating in its second run period (Run 2), which started mid 2015 and is planned to last until 2018. After the first season (Run 1), the beam energy has been increased from 4 TeV to 6.5 TeV per beam, corresponding to a centre-of-mass energy of  $\sqrt{s} = 13 \text{ TeV}$ , which has never been reached before [32]. This increase affects the kinematics of particle interactions such that data analysis algorithms, which were developed and optimized on Run 1 data, will need to be adjusted for Run 2. This will be covered in case of the Flavour Tagging software in chapter 6.

## 3.2 The LHCb Experiment

The LHCb-detector is specifically designed to study  $b$ - and  $c$ -mesons. Hence it is built as a single-arm forward spectrometer, covering an angular acceptance from 10 mrad to 300 mrad. The relatively light mesons will typically be strongly boosted along the beam axis. This results in roughly  $\frac{1}{4}$

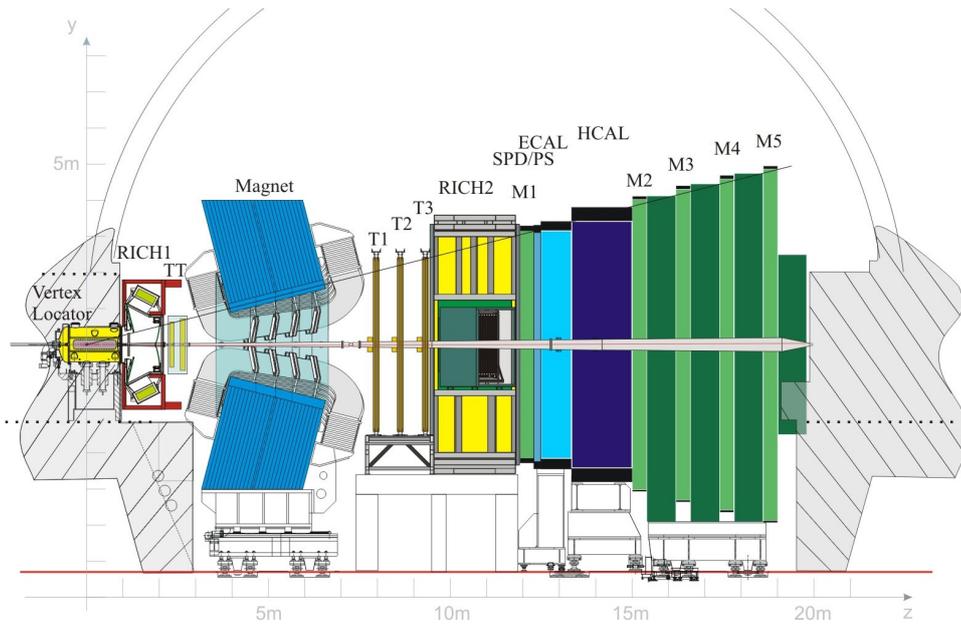


**Figure 3.2:** Monte Carlo simulation of the angular distribution of  $b$  and  $\bar{b}$  quarks produced in  $p\bar{p}$ -interactions at a centre-of-mass energy of  $\sqrt{s} = 14 \text{ TeV}$  [34]. The red highlighted area shows the detector's acceptance.

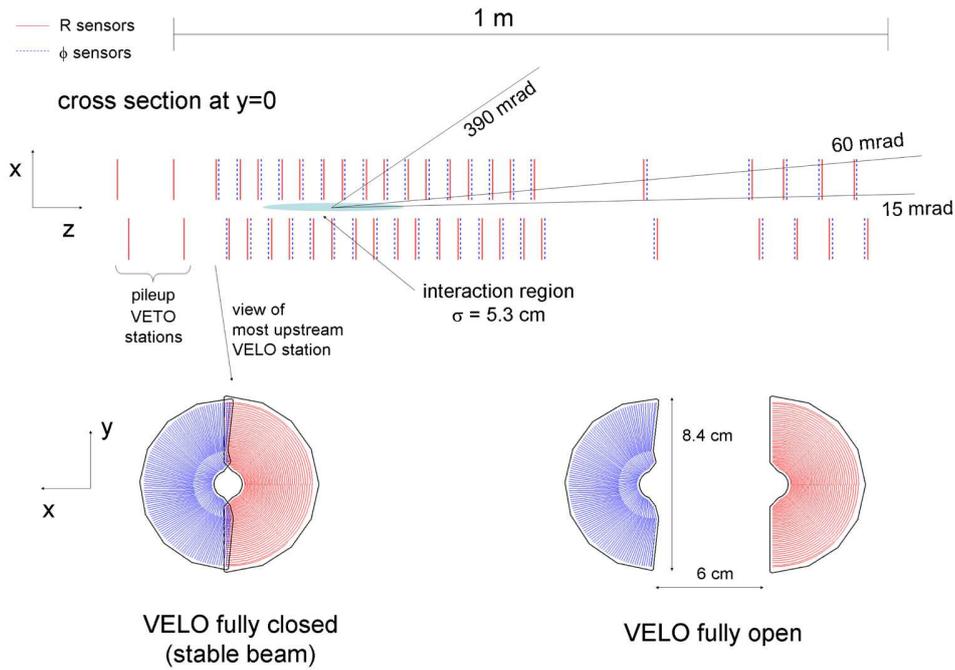
of these particles propagating into the detector's acceptance. The angular distribution of  $b$  and  $\bar{b}$  quarks at the interaction point is shown in Figure 3.2. A speciality of the LHCb detector is the constant value of average  $p\bar{p}$  interactions per bunch crossing, defined as  $\mu$ . This technique is called 'lumi levelling' and provides a very homogeneous data signature. The detectors components are explained in more detail in the following sections. If not stated otherwise, the information is referring to Ref. [33]. A sketch of the detector is shown in Figure 3.3. The components can be categorized into a tracking system and a particle identification system, which will be explained in the following sections. Additionally, a 1500 t heavy, normal conducting dipole magnet is installed downstream of the interaction point. It provides an integrated magnetic field of 4 Tm while its polarity can be changed.

### 3.2.1 Tracking System

The tracking system is used to reconstruct particle trajectories within the LHCb detector. It is divided into a silicon based detector, mounted very close to the interaction point, and two drift tube stations which are positioned downstream to the interaction point. Both detector components will be described in the following sections.



**Figure 3.3:** Side-view of the LHCb-detector [35] with its components. Starting at the interaction point at  $z = 0$ , these are the Vertex Locator (VELO), the first Cherenkov detector (RICH1), the first part of the tracking stations (Tracker Turicensis, TT), the dipole magnet, the second part of the tracking stations (T1-T3), the second Cherenkov detector (RICH2), a single muon chamber (M1), the electromagnetic and hadronic calorimeter system (ECAL and HCAL), including the PS and SPD and the final set of muon chambers (M2-M5). Each component is described in more detail in sections 3.2.1 to 3.2.2.



**Figure 3.4:** A sketch of the VELO. A cross section, showing the VELO module arrangement in the  $x$ - $z$ -plane is displayed at the top. Possible particle trajectories, defined by the design requirements are drawn as thin black lines. The bottom of the image shows the VELO in beam direction, in the closed state on the left and in the opened state on the right. Here, the  $\phi$ - (blue) and  $r$ - (red) silicon stripes can be seen [33].

## Vertex Locator

The Vertex Locator (VELO) is specifically designed to locate primary and secondary vertices with high resolution. The VELO consists of two series of silicon modules, lined up along to the beam axis. Both parts of the detector are retractable and can be moved up to 5 mm towards the beam axis, while sensitive material will be only 8 mm away.

Every module of the VELO contains two silicon sensors for measuring the  $r$ - and  $\phi$ -coordinates of charged particle trajectories. In total there are 42 modules which overlap in a closed state, arranged as shown in Figure 3.4. The module's layout is essentially defined by the prerequisites for the VELO. First, a particle, produced in the interaction region, needs to hit at least three modules to be reconstructed. This needs to be satisfied in the whole polar detector acceptance. Combining this with the minimum distance to the beam axis defines the minimal proportions of about 65 cm of the detector in the  $z$ -direction, as well as a maximal distance of 5 cm between two modules next to the interaction region. Furthermore the overlap of the two VELO sides provides full azimuthal coverage. In addition to a very close distance to the beam, a low material budget is needed for the high spatial resolution. Therefore, the VELO is completely operating in vacuum that is separated from the beam vacuum by a foil.

Due to the close distance from the beam, the VELO is susceptible to malfunctions of the LHC and might suffer irreparable damage if the beam directly hits its modules. Therefore a Beam Condition Monitoring system (BCM) is installed [36], consisting of two arrangements of eight artificial diamond sensors, located 2.1 m upstream and 2.8 m downstream of the interaction point. By measuring radiation resulting from interactions of the beam with the VELO foil, the BCM is able to detect beam misalignments before the sensitive VELO material is damaged. An exceed of the particle flux above certain thresholds will trigger a ‘beam dump request’ resulting in empty beams within 90  $\mu$ s after the signal has reached the beam dump system to protect the VELO hardware.

Given this design, the VELO reaches a spatial resolution of  $\sigma_{x,y} \approx 10 \mu\text{m}$  in the  $x$ - $y$ -plane and  $\sigma_z \approx 40 \mu\text{m}$  in the  $z$ -axes, the proper decay time resolution is  $\sigma_\tau \approx 40 \text{fs}$  [37].

## Tracking Stations

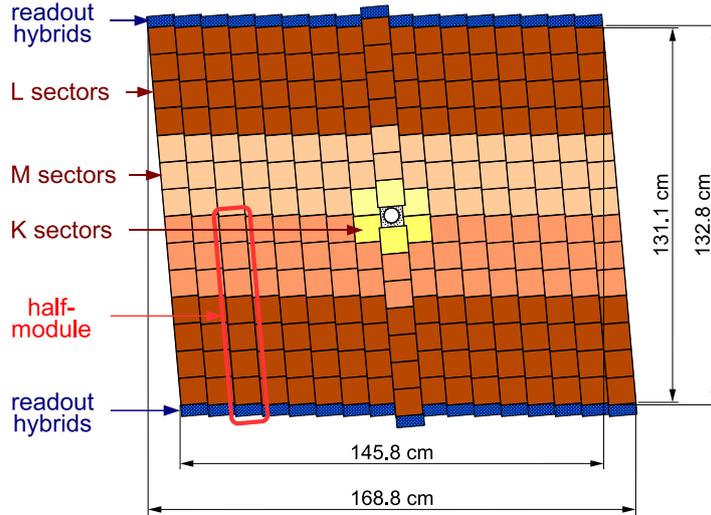
The Tracker Turicensis (TT) and the tracking stations T1 to T3 form the second part of the tracking system. Together with vertex and track information from the VELO these detectors allow for full track reconstruction inside the LHCb acceptance by measuring the  $x$ - and  $y$ -positions of charged particle tracks.

The track information combined with the magnetic field allows for charge identification due to the different direction of the Lorentz force for particles of different charges inside the magnet.

The TT consists of four layers of silicon micro strip sensors and is positioned upstream to the magnet, directly after the RICH1 detector. The tracking stations T1 to T3 are installed downstream to the magnet just before the RICH2 detector. Each of the stations contains an inner part, based on the same silicon sensors as the TT, positioned next to the beam pipe (inner tracker, IT). Both, IT and TT provide a spatial resolution of 50  $\mu\text{m}$ .

The outer parts of the tracking stations T1 to T3 form the outer tracker (OT) and are based on drift tubes. These are gas-filled tubes with a central wire. An electric voltage is applied between this wire and the tube wall such that a charged particle will ionize the gas while travelling through the tube and induce a charge avalanche which can be detected as a current in the nearest wire. The OT provides a spatial resolution of 200  $\mu\text{m}$ .

TT and the OT stations consist of multiple layers of single modules, as shown in Figure 3.5, where a half module is marked with a red rectangle. While two of these modules are aligned vertically, two are tilted around the beam axis with an angle of about  $\pm 5^\circ$  each. This set-up provides a high resolution in the  $x$ - $y$ -plane which is necessary due to the magnet’s magnetic field orientation and the result that charged particles will preferably spread into this plane.



**Figure 3.5:** Sketch of the frontal view of a single layer of the Tracker Turicensis (TT). The central beam pipe is indicated with a white circle. The layer is tilted by  $5^\circ$ . Each sector contains several silicon microstrip sensors with combined read-out [33].

### 3.2.2 Particle Identification System

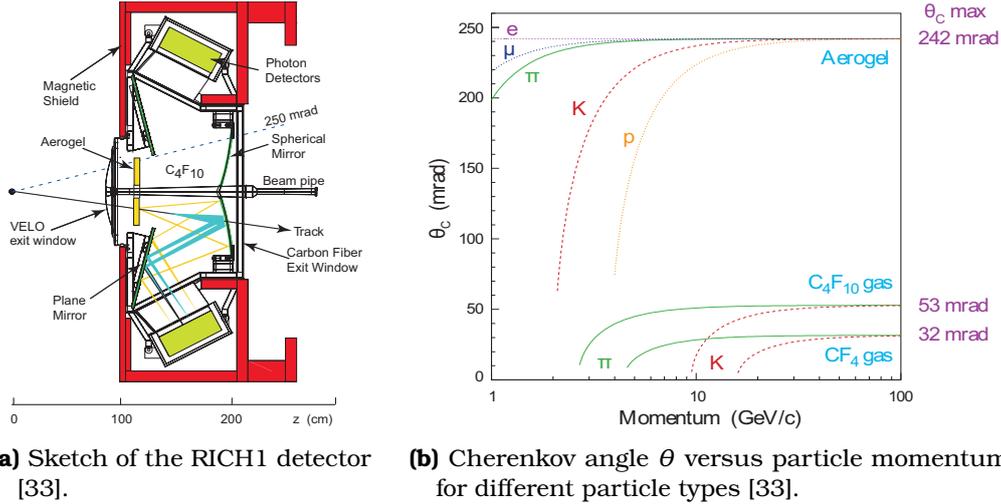
The Particle Identification System (PID) bundles information from Cherenkov detectors, calorimeters and muon chambers to assign particle hypotheses to particle tracks. Its components will be described in the following sections.

#### Ring Imaging Cherenkov Detectors

The Ring Imaging Cherenkov (RICH) detectors provide particle type information by measuring Cherenkov angles. If a particle travels through some material with a higher velocity than the speed of light within this material, it will emit a cone of Cherenkov light. The Cherenkov angle  $\theta$  in which the light is emitted allows for measurement of the velocity via

$$\cos \theta = \frac{1}{n\beta} \quad (3.2)$$

with the refractive index  $n$  of the medium and the ratio  $\beta$  of the particles velocity and the speed of light. Given the different masses of pions, protons, kaons and leptons that will reach the RICH detectors, these particles can be identified by their Cherenkov angles. A sketch of the RICH1 detector is given in Figure 3.6a and the distribution of the Cherenkov angle for different particles and momenta is shown in Figure 3.6b. Since the momentum range of the particles varies from low momenta at large polar angles to very high momenta at small polar angles, two detectors are installed. The first detector RICH1, positioned upstream to the LHCb magnet, is optimized for the lower momentum range from 1 GeV/c to 60 GeV/c and used Aerogel and  $C_4F_{10}$  gas radiators, while it only uses  $C_4F_{10}$  during Run 2. The second



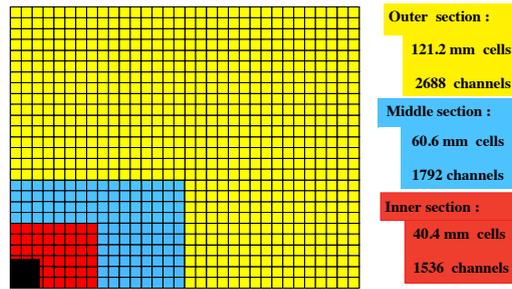
**Figure 3.6:** The PID system contains two Cherenkov detectors of which one is shown in a. As shown in b and described in the text, the Cherenkov angle  $\theta$  can be used to identify different types of particles.

detector RICH2 is installed downstream of the magnet and covers a momentum range of approximately 15 GeV/c to 100 GeV/c and beyond. It is operating with CF<sub>4</sub> gas as its radiator. Both detectors use a set of plane and convex mirrors, symmetrically mounted on both sides of the beam pipe, to guide the Cherenkov photons to photo detectors and allow a software-based reconstruction of  $\theta$ .

### Calorimeter System

The LHCb calorimeter system consists of a combination of an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). It provides identification power for electrons, photons and hadrons and measures the momentum and  $x$ - $y$ -position of particles. Although the calorimeters provide some spatial information for the tracking system, their main purpose is to identify particle types. Both calorimeters use sequences of scintillators and lead, connected to photo-multipliers via wavelength shifting fibres. The ECAL thickness is about 25 radiation lengths to measure the full showers of high energy photons while the HCAL thickness is about 5.6 radiation lengths, which is limited by the localities.

To suppress background from  $\pi^0$  and  $\pi^\pm$  particles for the  $e^\pm$ -identification, a Scintillating Pad Detector and a Preshower detector are mounted upstream of the ECAL. Furthermore, both calorimeters' segmentation is finer next to the beam pipe and coarser in the outer region, since the particle flow varies over two magnitudes in this range. A sketch of the ECAL is shown in Figure 3.7.



**Figure 3.7:** Sketch of the layout of one quadrant of the electronic calorimeter (ECAL). The segmentation is smaller closer to the beam pipe (black) to better resolve single particles in the high particle flux in this area. [33].

## Muon Chambers

The muon chambers M1 to M5 provide essential information to many LHCb analyses by identifying muons and measuring their position. Similar to the calorimeters, the muon chambers add some spatial information for the tracking system, but primarily serve to identify muons. Especially one of the Flavour Tagging algorithms, which will be covered in more detail in section 4.2.1 profits from the clean signature of these particles.

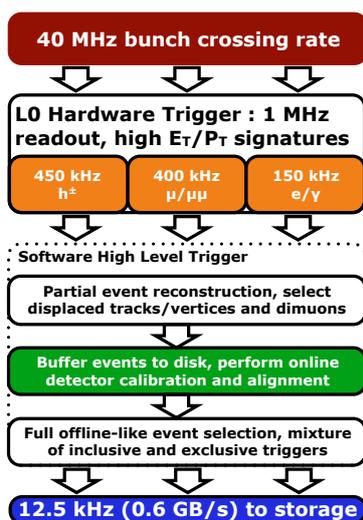
The first station M1 is located upstream to the calorimeters and provides the highest spatial resolution. Chambers M2 to M5 are placed downstream to the calorimeters, with iron blocks serving as absorbers between each station.

Due to the high radiation load next to the beam pipe in the M1 chamber, the detector uses Gas Electron Multiplier detectors in this area. The outer region of M1 as well as the other muon stations are based on Multi Wire Proportional chambers. In both detector technologies gas is ionized by a particle and the signal is amplified in a charge avalanche by an external high voltage.

### 3.2.3 Trigger System

While the LHCb detector system provides very high spatial, momentum and mass resolution it must deal with the 40 MHz collision rate of the LHC. As of 2015, the size of a single event of the LHCb detector is 50 kB which would lead to a raw data rate of the detector of 2 TB/s. Since it is not yet technically possible to store such data rates over a long period of time, a system of successive triggers is implemented to reduce the recorded rate down to 12.5 kHz corresponding to a data rate of 600 MB/s which is written to disk and available for off-line analyses [38].

The first trigger stage (L0) is implemented directly on hardware and has a fixed 1 MHz output rate. This trigger level selects events with high transverse momenta and energies based on information from the calorimeters and the muon chambers M2 to M5.



**Figure 3.8:** A diagram of the LHCb trigger system as of 2015. The initial bunch crossing rate of 40 MHz is reduced to an output rate of 12.5 kHz by applying a set of on-line selection criteria. The trigger system is divided into three stages one of which is implemented in hardware (L0) and two implemented in software (HLT1 and HLT2) [38].

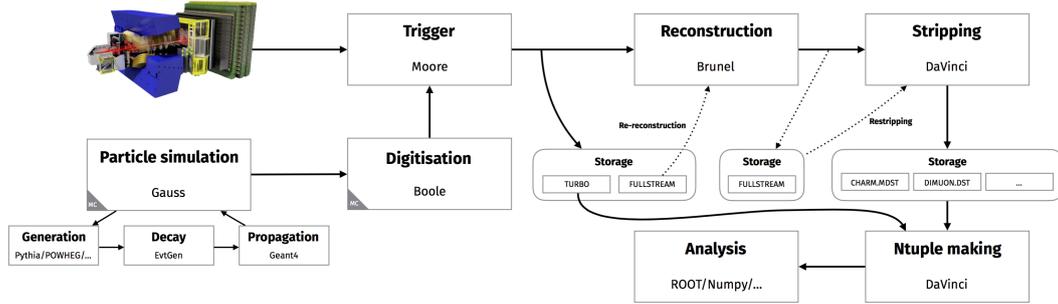
The output of the level-0 trigger is further processed by a set of high level triggers (HLT1 and HLT2). These software based triggers apply first partial track reconstructions (HLT1) with an output rate of approximately 150 kHz followed by a full event reconstructions with the final output rate of 12.5 kHz (HLT2). A diagram of the trigger system is shown in Figure 3.8.

While the detector has not been changed in between Run 1 and Run 2 (see section 3.1), the software of the trigger system has been revisited and the trigger hardware was upgraded. Instead of chaining all reconstruction steps of the HLT, the HLT1 output is buffered onto 5 TB (2015) or 10 TB (2016) of disk space. The HLT2 reads the event data from this buffer which not only allows a higher utilization of the computing resources even after data taking, but also allows for continuous, asynchronous alignment of the trigger parameters. The computing resources have therefore been increased to a total number of 56 000 available CPU cores [38]. The trend of more flexible trigger implementations, providing higher data rates is especially important, when looking towards further increases of the LHC luminosity during Run 3 or the High-Luminosity LHC.

### 3.2.4 Software Stack

The LHCb software stack contains several software packages to turn the raw detector output into manageable files which can be used for analysis. It is based on the High Energy Physics Framework GAUDI [39] and uses the C++ data processing framework ROOT [40].

The raw detector output is parsed to raw data containing full reconstructed events in the MOORE framework, which is also running in the High Level



**Figure 3.9:** Diagram of the LHCb software and the data flow in the analysis pipeline [51].

Raw data, either from the detector, or from a chain of simulation software is passed to the Trigger software. From there on, the data is stored to disk the first time and directly accessible for analyses via the ‘Turbo Stream’. Furthermore, the data is used to reconstruct whole decay chains after which another copy of the data is stored. In the so-called ‘Stripping’, further selection and reconstruction criteria are computed and applied for the specific output streams. The final datasets which are used for the different analysis and in which the Flavour Tagging algorithms are applied are produced in the DAVINCI-based ‘NTuple making’.

Trigger (section 3.2.3) [41]. The raw output can also be simulated within the GAUSS package [42, 43] by generating single particles with e.g. PYTHIA [44, 45], simulating their decays with the EVTGEN package [46] and propagating the resulting particles through a digital model of the detector with GEANT4 [47]. A copy of the raw data is kept on disk. The data is further processed to combine track and particle ID information into so-called proto particles in the BRUNEL framework [48] and written to disk again. Specific selections for different physics modes are applied during the so-called stripping, which is implemented in the DAVINCI framework [49]. The stripped data is then written into a Data Summary Tape (DST) file containing the full information of all particles in the event as well as  $\mu$ DST files only containing specific signal information of the active stripping configuration. These files are available for the final tuple production, which is also running in the DAVINCI framework and in which all reconstruction and analysis algorithms for the specific analyses are applied. The Flavour Tagging software is running on the full DST or  $\mu$ DST files at this stage. Usually the final tuple is stored as ROOT-files and can be studied with a variety of data analysis tools. The structure of the LHCb software stack is depicted in Figure 3.9.

A novelty of the Run 2 period is the so-called turbo stream, in which a part of the raw data is directly written to DST files using the TESLA framework [50].



## 4 Flavour Tagging of $B$ Mesons

In the following chapter, the general strategy that is currently used to compute tag information in the tagging algorithms will be outlined. Therefore all available Flavour Tagging algorithms will be described while focusing on single track taggers and in particular on the muon tagger. Furthermore all variables that are common for Flavour Tagging will be introduced, while this section follows Ref. [52].

### 4.1 Variables and Definitions of Flavour Tagging

The knowledge of the initial flavour of neutral  $B$  mesons is essential for many LHCb analyses (see section 2.1.3). Since neutral  $B$  mesons can oscillate between particle and antiparticle states and thus between the two possible  $b$  flavours, it is not possible and not necessarily useful to deduce their production-flavour from their decay products. Nevertheless, a variety of other particles is produced during the signal candidate's production and hadronization process. As already shown in Figure 3.2, these predominantly include pairs of  $b$  quarks within the LHCb acceptance. Due to the conservation of additive quantum numbers, the flavours of these particles are linked to the initial flavour of the signal particle. This information is exploited with LHCb's Flavour Tagging algorithms, which assign a tag to the signal candidate. The selection of the non-signal particles that provide tagging information within a given event therefore is the main task of these algorithms. These particles will be referred to as 'tagging particles' from now on. Since a single event usually contains a large number of non-signal particles, dominated by kaons and pions, the selection of the correct particle constitutes the major difficulty.

As described in section 2.1.3,  $CP$  analyses usually determine the decay time-dependent asymmetry  $\mathcal{A}_f(t)$  of the decay rates  $\Gamma$ . The decay rates itself are experimentally determined by counting the number  $N_{B^0}(t)$  of signal candidates at decay time  $t$ , which defines the observed asymmetry to be

$$\mathcal{A}_{\text{obs}}(t) = \frac{N_{\bar{B}^0}(t) - N_{B^0}(t)}{N_{\bar{B}^0}(t) + N_{B^0}(t)}. \quad (4.1)$$

Due to experimental imperfections, this observed asymmetry does not necessarily reflect the true asymmetry and this fact is usually parametrized with a dilution term  $D$ . Neglecting intrinsic asymmetries, the observed asymmetry, up to  $\mathcal{O}(\mathcal{A}(t))$ , can then be written as

$$\mathcal{A}_{\text{obs}}(t) \approx D\mathcal{A}(t). \quad (4.2)$$

The uncertainty  $\sigma_{\mathcal{A}}$  of this value can be calculated using Gaussian error propagation. Assuming a small dilution and neglecting the dilution's uncertainty leads to:

$$\sigma_{\mathcal{A}}^2 \approx \frac{1}{D^2} \frac{1}{N_{\bar{B}^0} + N_{B^0}}, \quad (4.3)$$

with the total number  $N_{\bar{B}^0} + N_{B^0}$  of tagged events, which can also be parametrized by the total number  $N = N_{\bar{B}^0} + N_{B^0} + N_u$  of events, including the number  $N$  of untagged events, and an efficiency  $\varepsilon$  of the whole algorithm, such that

$$\sigma_{\mathcal{A}} \approx \frac{1}{\sqrt{\varepsilon D^2 N}}. \quad (4.4)$$

Apart from the statistical uncertainty  $\sqrt{N}$ , this term also contains the efficiency of the algorithm and reflects the imperfection of the experimental set-up.

The figure of merit, usually used to optimize the performance of Flavour Tagging algorithms therefore combines these additional variables. The tagging efficiency  $\varepsilon_{\text{tag}}$  hereby is parametrized in terms of the number  $N_r$  of correctly tagged events and the number  $N_w$  of wrongly tagged events:

$$\varepsilon_{\text{tag}} = \frac{N_{\bar{B}^0} + N_{B^0}}{N} = \frac{N_w + N_r}{N_w + N_r + N_u}. \quad (4.5)$$

The average dilution  $\bar{D}$  can be parametrized with the fraction of wrongly tagged events within all tagged events, which is referred to as the average mistag rate:

$$\bar{D} = \frac{N_r - N_w}{N_r + N_w} = 1 - 2\bar{\omega}, \quad (4.6)$$

$$\text{with } \bar{\omega} = \frac{N_w}{N_w + N_r}. \quad (4.7)$$

Following Equation 4.4, the combination of these variables provides a general figure of merit for all tagging algorithms and is called average tagging power

$$\bar{\varepsilon}_{\text{eff}} = \varepsilon_{\text{tag}} (1 - 2\bar{\omega})^2 = \varepsilon_{\text{tag}} \bar{D}^2, \quad (4.8)$$

The average tagging power of a given algorithm indicates that a high tagging efficiency  $\varepsilon_{\text{tag}}$ , resulting in a large number of events for which a tag is assigned, does not necessarily correspond with a better performance. Instead, the additionally tagged events might lower the average dilution which can be seen as a worse quality of the corresponding tagging particles.

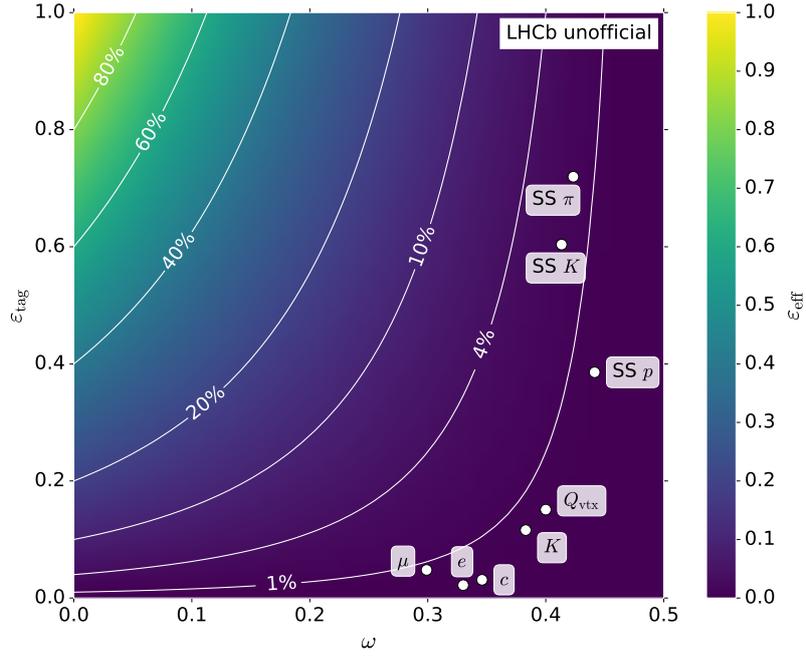
In addition to the average tagging power and mistag fraction, the knowledge of the mistag probability  $\omega_i$  for a given event  $i$  allows for the definition of a mean per-event tagging power via

$$\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}} \langle D^2 \rangle = \frac{1}{N_w + N_r + N_u} \sum_{i=1}^{N_w + N_r} (1 - 2\omega_i)^2. \quad (4.9)$$

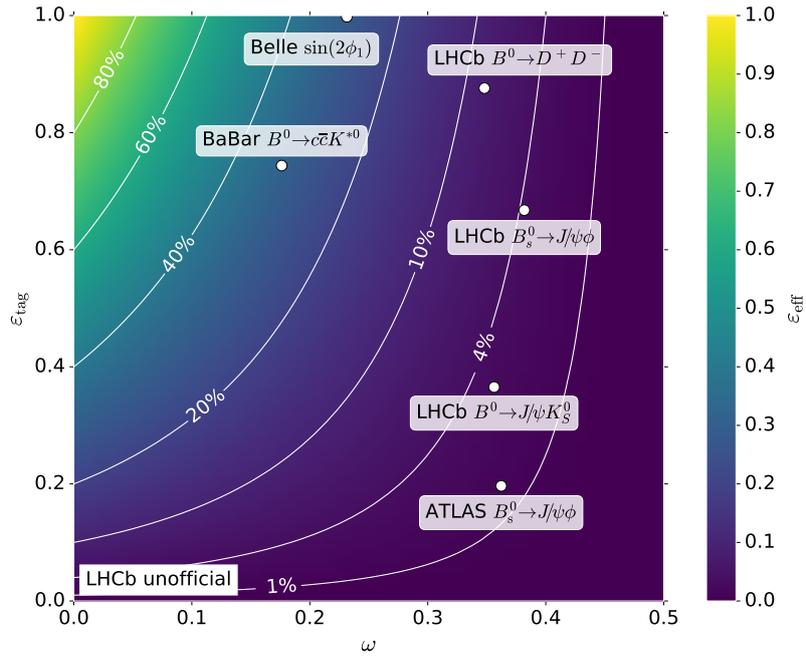
This definition has the advantage that events of low quality, which is corresponding to a mistag probability  $\omega_i \approx 0.5$ , are given less weight in the overall per event tagging power. A higher tagging efficiency can therefore only increase the per event tagging power, assuming that the mistag estimation is correct. An estimation of the mistag probability  $\omega_i$  can be determined with multivariate analysis tools such as Boosted Decision Trees (BDTs) (see section 2.2) or Artificial Neural Networks (ANNs). Analyses that use this per-event mistag information can therefore profit from a generally higher average per-event tagging power. The calculation of the tag decisions and the information used to predict the per event mistag will be further explained in section 4.2.

A distribution of the tagging power for different  $(\varepsilon_{\text{tag}}, \overline{\omega})$ -pairs is shown in Figure 4.1, where the specific tagging power values of the Run 1 tagging algorithms are marked in a, and the combined tagging power values of recent flavour tagged LHCb analyses as well as flavour tagged analysis from other experiments are marked in b.

As described previously, the true tag of the signal candidates and with it the numbers  $N_w$ ,  $N_r$  and  $N_u$  cannot be determined on the same data to which the flavour tagging algorithms are applied. Instead, these can either be determined on simulated data or on self-tagged decay channels, where the decay product is clearly linked to the initial flavour of the signal candidate. Usually the decays  $B^+ \rightarrow J/\psi K^+$  or  $B^0 \rightarrow J/\psi K^{*0} (\rightarrow K^+ \pi^-)$ , for which a Feynman graph is shown in Figure 4.2, are used in the latter case. If not stated otherwise, decay descriptors such as  $B^+ \rightarrow J/\psi K^+$  will always also refer to the charge-conjugated decays ( $B^- \rightarrow J/\psi K^-$  in this case) in the following sections and the remaining chapters. The charge of the  $K^+$  and  $K^+ \pi^-$  final states, which can be detected in the LHCb detector is clearly connected to the signal candidates initial flavour. In the first case this is also the tag at production time, since charge is conserved and no  $B^+ \leftrightarrow B^-$  oscillation occurs. To suppress incorrect tags due to mixing in the second case, a decay-time cut of usually  $\tau_{B^0} \leq 3$  ps is applied. All optimization studies that will be presented in chapter 6 are performed on  $B^+ \rightarrow J/\psi K^+$  data samples.

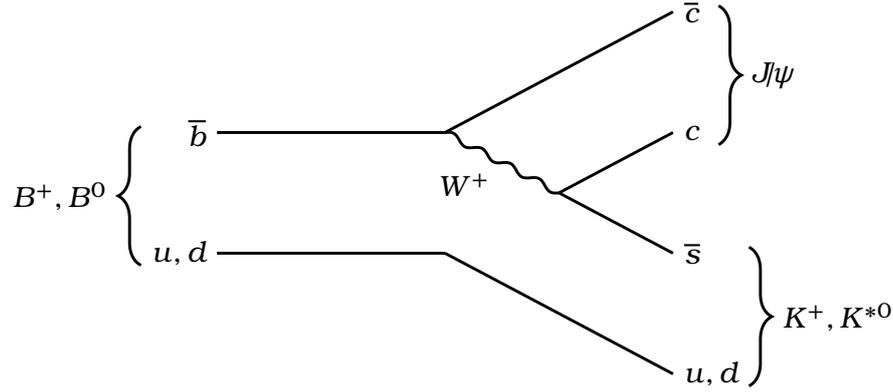


(a) Performance of LHCb Flavour Tagging algorithms.



(b) Combined performance of different flavour tagged analyses.

**Figure 4.1:** Tagging power  $\varepsilon_{\text{eff}}$ , measuring the performance of Flavour Tagging algorithms. The  $x$ - and  $y$ - axes show the mistag fraction  $\omega$  and the tagging efficiency  $\varepsilon_{\text{tag}}$ , respectively. The colour maps the corresponding tagging power. The white lines indicate contours of constant tagging power. The tagging power values of the LHCb Flavour Tagging algorithm are highlighted in Figure 4.1a [53, 54, 55]. The combined tagging power of recent flavour tagged analyses of the LHCb collaboration, the ATLAS experiment and the  $e^+e^-$  collider experiments BaBar and Belle are highlighted in Figure 4.1b [19, 56, 57, 58, 59, 60].



**Figure 4.2:** Feynman graph of the self-tagging decay channels  $B^+ \rightarrow J/\psi K^+$  and  $B^0 \rightarrow J/\psi K^{*0}$ . The charge of the final state kaons can clearly be identified with the tag of the signal candidate. In the case of neutral  $B^0$  mesons, a decay-time cut is applied to reduce the fraction of mixed signal candidates.

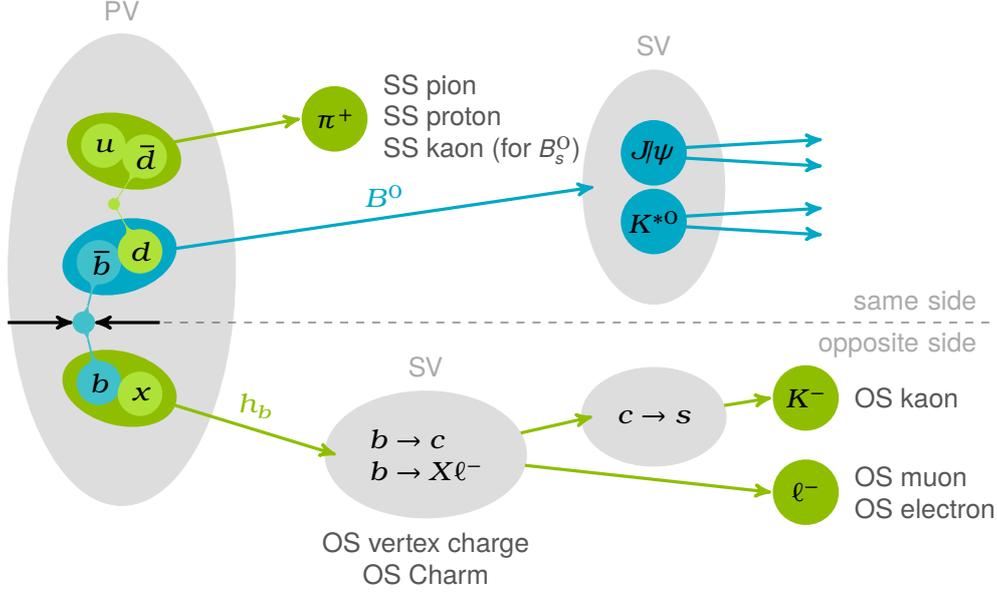
## 4.2 Flavour Tagging Algorithms

The LHCb Flavour Tagging software bundles a set of different algorithms which each determine a tag decision  $d_i$  and a mistag estimate  $\omega_i$  based on the information of a set of particles within a given event  $i$ . Each algorithm is designed to use different types of particles and particle tracks. Therefore currently eight such algorithms are implemented, and will be explained in the following sections. Since the optimization studies that are presented in chapter 6 are primarily based on the muon tagger, the corresponding section 4.2.1 will be more detailed than the remaining ones.

The  $b$  quarks are dominantly produced in  $b\bar{b}$  pairs, with a relatively low angular distance  $\Delta\phi$  (see Figure 3.2). Thus the Flavour Tagging algorithms are commonly divided into such that use particles originating from non-signal  $b$  decays – called Opposite-Side (OS) taggers – and such that use particles, connected to the signal hadronization – called Same-Side (SS) taggers. A sketch of a  $B$  meson decay with its non-signal  $b$  partner and all decay products that are used by the different Flavour Tagging algorithms, as well as primary and secondary vertices (PV and SV) in that event is shown in Figure 4.3.

While the here presented study is focused on the evaluation of single tagging algorithms, a  $CP$  analysis usually combines the outputs of all available algorithms to receive a maximum tagging power. The combined probability  $P(b)$  or  $P(\bar{b})$  of a given event, to contain a  $b$  or  $\bar{b}$  quark is hereby defined as

$$P(b) = \frac{p(b)}{p(b) + p(\bar{b})} \quad \text{and} \quad P(\bar{b}) = 1 - P(b), \quad (4.10)$$



**Figure 4.3:** Schematic diagram of the different Flavour Tagging algorithms. The signal decay chain is highlighted in blue and determines the same side (SS). The other  $b$  hadron, determining the opposite side (OS) is flavour-connected to the signal candidate and can therefore also be used to deduce tag information. The different Flavour Tagging algorithms are listed next to the particles on which they rely to identify the tag and the mistag estimate. The algorithms are explained in more detail in section 4.2.

with the products  $p(b)$  and  $p(\bar{b})$  of each algorithm:

$$p(b) = \prod_k \left( \frac{1 + d_k}{2} - d_k (1 - \eta_k) \right), \quad (4.11)$$

$$p(\bar{b}) = \prod_k \left( \frac{1 - d_k}{2} + d_k (1 - \eta_k) \right), \quad (4.12)$$

where  $d_k$  is the tag decision and  $\eta_k$  the calibrated mistag estimation of the  $k$ -th algorithm. In general, correlations of taggers on the same side or opposite side need to be taken into account to correctly estimate the mistag probability. This will be discussed further in section 6.4.

#### 4.2.1 Opposite Side Single Track Taggers

The general design of the kaon, muon, and electron taggers is identical. Each tagger applies a set of rectangular selections to kinematic and PID observables of the particles within a given event to select a single, best tagging particle. These cuts aim to select particles that originate from OS  $b \rightarrow X\mu^-$  or  $b \rightarrow Xe^-$  transitions in case of the muon or electron tagger, and  $b \rightarrow c \rightarrow s$  transition for the kaon tagger. The particle  $X$  can be any hadron in this case. As shown in Figure 4.3, the initial OS  $b$  was produced with a  $\bar{b}$  partner, which forms the signal decay. Thus, the signal candidate's flavour can be determined by measuring the  $\mu^-$ ,  $e^-$  or  $K^+$  charge. Since the OS  $b$

quarks might hadronize into neutral  $B$  mesons, and are therefore subject to  $B$  oscillation, the OS taggers have a certain chance of being wrong although selecting the correct tagging particle.

The selection criteria for the OS single track taggers include different subsets of a variety of event information. These are the transverse and full momentum  $p$  and  $p_T$  of the tagging particle, and the transverse momentum of the signal candidate,  $p_{T,B}$ . Furthermore, the minimum angular distance  $\min(\Delta\phi)$  between the tagging particle and all other tracks in the event, the significance of the track per degrees of freedom  $\chi_{\text{track}}^2/\text{ndf}$  are taken into account. The selections also accounts for the absolute significance  $|\chi^2|$  of the impact parameter IP with respect to the primary vertex PV or all other vertices that occurred in the given event and called pile-up vertices PU. The fraction  $E_{\text{cal}}/p$  of the calorimeter energy  $E_{\text{cal}}$  and the momentum  $p$  and the VELO charge  $Q_{\text{VELO}}$ , which estimates the number of tracks within the VELO that contribute to a given track, are also used as selection criteria. It is checked, whether muon-PID information is available and whether the particle is a daughter of the signal candidate. Finally, the probability  $P_{\text{NN}}(X)$  of a given particle hypothesis  $X$ , which is based on a Neural Net and the PID information, the track's ghost probability  $P_{\text{GHOST}}$ , the number  $N_{\text{tracks}}$  of tracks in the event, the number  $N_{\text{PV}}$  of primary vertices in the event, and whether the particle was in the acceptance of the HCAL are used for selection criteria. The specific selection values for each single track tagger are listed in Table 4.1.

After these selections are applied, from the remaining particles the one with the maximum transverse momentum  $p_T$  is chosen to be the tagging particle. The charge of this particle is further used to determine the tag  $d_i$  of the signal candidate. For all OS single track taggers, the tag is  $d_i = +1(-1)$  for a negatively (positively) charged tagging particle (see Figure 4.3).

Some of the features listed above, which are also marked in Table 4.1 will afterwards be passed to a BDT to determine a raw estimation  $p_{\text{raw}}$  for the tag  $d_i$  to be correct. This is first transformed into a raw mistag estimation  $\eta$  via

$$\eta_i = \begin{cases} 1 - p_{\text{raw}} & \text{if } p_{\text{raw}} \geq 0.5 \\ p_{\text{raw}} & \text{if } p_{\text{raw}} < 0.5 \end{cases}, \quad (4.13)$$

while the tag  $d_i$  is inverted if  $p_{\text{raw}} < 0.5$ . Finally, the raw mistag estimation  $\eta_i$  is calibrated by fitting an function  $\bar{w}(\eta)$  to the average mistag rates  $\bar{w}$  in bins of the raw mistag estimation. Hereby, a linear function has usually been used in previous analyses [53, 61, 19]. These transformations and calibrations will be described in more detail in chapter 6, where a per-event fit of a polynomial logistic regression will be used instead of a binned fit of a linear function. The performance values of the single track taggers are listed in Table 4.2 for Run 1 data and were measured in the  $B^+ \rightarrow J/\psi K^+$  decay channel.

**Table 4.1:** All features that are used for selection of tagging particles (if a selection expression is listed) and as BDT input (if the corresponding row is marked with ‘•’). The variables are defined within the text of section 4.2.1.

Variable	Selections								
	$\mu^-$		$K^+$		$e^-$				
$p_T/\text{GeV}$	>	1.1	•	>	0.7	•	>	1.1	•
$p/\text{GeV}$	-		•	>	2.0	•	-		•
$p_{T,B}/\text{GeV}$	-		•	-		•	-		•
$\min(\Delta\phi)$	>	0.005		>	0.005		>	0.005	
$\chi^2_{\text{track}}/\text{ndf}$	<	3.0	•	<	3.0	•	<	3.0	•
$ \chi_{\text{IP(PV)}} $	-		•	<	4.0	•	>	3.5	•
$E_{\text{cal}}/p$	-			-			>	0.85	•
	-			-			<	2.0	
$ \chi_{\text{IP(PU)}} $	>	3.0	•	>	6.0	•	>	4.0	•
$\mathcal{Q}_{\text{VELO}}$	-			-			>	0.0	
	-			-			<	1.4	
$ \text{IP(PV)} $	-		•	<	1.6		-		
PID( $\mu$ )		True			False			False	
IsDaughter		False			False			False	
$P_{\text{NN}}(\mu)$	>	0.35	•	<	0.8		-		
$P_{\text{NN}}(\pi)$	<	0.8		<	0.8	•	<	0.8	
$P_{\text{NN}}(e)$	<	0.8		<	0.8		>	0.1	
$P_{\text{NN}}(K)$	<	0.8		>	0.25	•	>	0.8	
$P_{\text{NN}}(p)$	<	0.8		<	0.8	•	<	0.8	
$P_{\text{NN}}(K) - P_{\text{NN}}(p)$	-			>	0.0		-		
$P_{\text{NN}}(K) - P_{\text{NN}}(\pi)$	-			>	-0.6		-		
$P_{\text{NN}}(e) - P_{\text{NN}}(\pi)$	-			-			>	-0.8	
$P_{\text{GHOST}}(\text{track})$	<	0.4	•	<	0.35	•	<	0.4	•
$N_{\text{tracks}}$	-		•	-		•	-		•
$N_{\text{PV}}$	-			-		•	-		
$\text{Acc}_{\text{HCAL}}$	-			-				True	

**Table 4.2:** The performance of the single track taggers divided into their selection efficiency  $\varepsilon_{\text{tag}}$  and the per event dilution  $\langle D^2 \rangle$  and combined to the tagging power  $\varepsilon_{\text{eff}} = \varepsilon_{\text{tag}} \langle D^2 \rangle$  [53]. The numbers have been measured on  $B^+ \rightarrow J\psi K^+$  data and are therefore comparable to the performance values presented in the upcoming sections.

Tagger	$\varepsilon_{\text{tag}}/\%$	$\langle D^2 \rangle/\%$	$\varepsilon_{\text{eff}}/\%$
$\mu^-$	$4.8 \pm 0.1$	$16.2 \pm 1.1$	$0.77 \pm 0.07$
$e^-$	$2.2 \pm 0.1$	$11.3 \pm 1.5$	$0.25 \pm 0.04$
$K^+$	$11.6 \pm 0.1$	$5.5 \pm 0.5$	$0.63 \pm 0.06$

## 4.2.2 Vertex Charge Tagger

Instead of deferring tag information from single charged particles like the single track taggers, the vertex charge tagger uses the total charge of a OS secondary vertex, that originates from the PV. Therefore, an inclusive secondary vertex on the opposite side is reconstructed and the weighted sum of the charges of all tracks that originate from this vertex is calculated. The vertex is formed by combining two tracks and evaluating the probability of that vertex originating from a  $b$  hadron by a BDT. The best combination among all tracks that pass a set of selection criteria including kinematic and track quality cuts is used as a seed vertex. Other tracks that are compatible with originating from this seed vertex, but which do not originate from the primary vertex, are added to the seed vertex to form the final inclusive vertex. The tag decision is based on

$$\mathcal{Q}_{\text{vtx}} = \frac{\sum_i^{N_{\text{tracks}}} \mathcal{Q}_i p_{T,i}^\kappa}{\sum_i^{N_{\text{tracks}}} p_{T,i}^\kappa} \quad (4.14)$$

of the track charges  $\mathcal{Q}_i$ , their transverse momenta  $p_{T,i}$  and an exponent  $\kappa = 0.4$  that is chosen to optimize the average tagging power  $\bar{\varepsilon}_{\text{eff}}$  of this tagger. The mistag estimate  $\eta_i$  is calculated by a Neural Network that was trained on simulated data with the averages of the kinematic properties of the tracks, as well as the vertex charge  $\mathcal{Q}_{\text{vtx}}$  as input variables. The average tagging power on  $B^+ \rightarrow J\psi K^+$  Run 1 data is  $\bar{\varepsilon}_{\text{eff}} = (0.60 \pm 0.06)\%$  data [53].

## 4.2.3 Charm Tagger

The charm tagger exploits the tagging potential of OS charm hadrons originating from a non-signal  $b \rightarrow c$  transition. Therefore seven decay modes of charm hadrons are considered and candidates for these hadrons are reconstructed. These have to meet several selection criteria such as mass constrains of the associated particle hypothesis, a reconstructed track that points back to the PV and a good separation from any other PV. Furthermore a BDT is trained on simulated data to separate charm candidates from background particles. Afterwards, the candidate with the best BDT value is chosen to be the tagging particle and the BDT output is calibrated to represent

the mistag estimate  $\omega_i$ . The tagging power of the charm tagger, calculated on Run 1  $B^+ \rightarrow J/\psi K^+$  data is  $\varepsilon_{\text{eff}} = (0.30 \pm 0.01(\text{stat}) \pm 0.01(\text{sys}))\%$  [61].

#### 4.2.4 Same Side Proton and Pion Tagger

The SS- $p$  and SS- $\pi$  taggers determine a tag decision  $d_i$  by selecting protons and pions that originate from excited  $b$  hadron states such as  $B^{**}$ ,  $\Lambda_b^{**}$  or  $\Sigma_b^{**}$ , decaying into the signal candidate  $B$  meson and additional hadrons. A BDT is trained to identify the correctly correlated tagging particles and the particle with maximum BDT value is chosen to determine the signal tag. The mistag estimate  $\omega_i$  is calculated by first applying a polynomial transformation  $\text{BDT} \rightarrow \eta_i$  to the BDT output and then calibrating the  $\eta_i$  values with a linear function  $\omega_i(\eta_i)$ . In contrast to the other taggers, the SS- $\pi$  and SS- $p$  taggers have only been trained on  $B^0 \rightarrow D^+ \pi^-$  data, in which  $B^0$  oscillation occurs. The fraction of oscillated mesons is therefore reduced by applying a lifetime cut. The tagging power of these algorithms is measured on  $B^0 \rightarrow D^+ \pi^-$  Run 1 data to be  $\varepsilon_{\text{eff}} = (1.69 \pm 0.10)\%$  for the SS- $\pi$  and  $\varepsilon_{\text{eff}} = (0.53 \pm 0.05)\%$  for the SS- $p$  tagger [54].

#### 4.2.5 Same Side Kaon Tagger

The SS- $K$  tagger uses a set of two Neural Networks to select kaons which are produced in the signal candidate's hadronization process as tagging particle candidates. The first network is trained on simulated  $B_s^0 \rightarrow D_s^- \pi^+$  events to select proper kaon tracks. The second network uses up to three best output values from the first network and other event information to predict the probability that the given signal candidate was produced with a tag  $d = +1$ . This value is finally transformed into a mistag estimate by a linear calibration function and the tag decision is set correspondingly. The performance of the SS- $K$  tagger is measured on  $B_s^0 \rightarrow D_s^- \pi^+$  Run 1 data to be  $\varepsilon_{\text{eff}} = (1.80 \pm 0.19(\text{stat}) \pm 0.18(\text{sys}))\%$  [55].

## 5 Reimplementation of the LHCb Flavour Tagging Software

The Flavour Tagging software contains several key features that will be described in the following section. While running during the tuple production for all flavour tagged analyses, it loads all  $e^-$ ,  $\mu^-$ ,  $K^+$ ,  $p$  and  $\pi$  particles that are associated with a given event  $i$  into memory after applying a general set of pre-selections to these particles. Furthermore, the charm tagger (see section 4.2.3) requires to reconstruct several charm hadron candidates and applies a different set of selections, which is unique for each charm particle hypothesis, before reading these particles into memory. This first set of selections will be further referred to as S1 and S2<sup>1</sup>. The adjustment of these selections can be used to control the overall computing time consumed by the Flavour Tagging package, while this might affect the tagging efficiency or the mistag probability and with that the tagging power of the algorithms.

Afterwards the different tagging algorithms are executed, using the event information, which includes the selected tagging particle candidates, a given signal candidate and an associated vertex. As depicted in Figure 5.1, every algorithm  $k$ , described in chapter 4, will apply a different set of selections S3 and use different kinds of multivariate analysis algorithms to provide its tag decision  $d_i^k$  and its raw mistag estimate  $\eta_i^k$ . Furthermore, the raw mistag estimates need to be calibrated to provide a correct mistag probability  $\omega_i^k$ . Finally, the tag decisions and mistag estimates and/or calibrated mistag probabilities  $\omega_i^k$  of every algorithm and optionally a combined tag decision  $d_i^{\text{comb}}$  and mistag probability  $\omega_i^{\text{comb}}$  must either be forwarded to other algorithms for further processing or stored to disk.

In addition to the output of each tagging algorithm, the software is required to provide the flexibility to apply different selection criteria and store the complete vectors of tagging particle candidates instead of the single tag decisions and mistag estimates. This additional information is essential for the training and optimization process of the algorithms and their internally used MVA tools.

Conclusively, the main functionalities of the Flavour Tagging framework are to

- provide the algorithms with event information, consisting of signal candidates, tagging particle candidates and vertices,

---

<sup>1</sup> The selections S1 are applied within the `FilterDesktop` methods, provided by the `LoKi` software package [62], while S2 is computed within the Flavour Tagging software and refers to relative information, based on combinations of particle and vertex information.

- apply selections to the events, based on the event information,
- call different MVA tools with a subset of the event information,
- and apply a calibration of the MVA output.

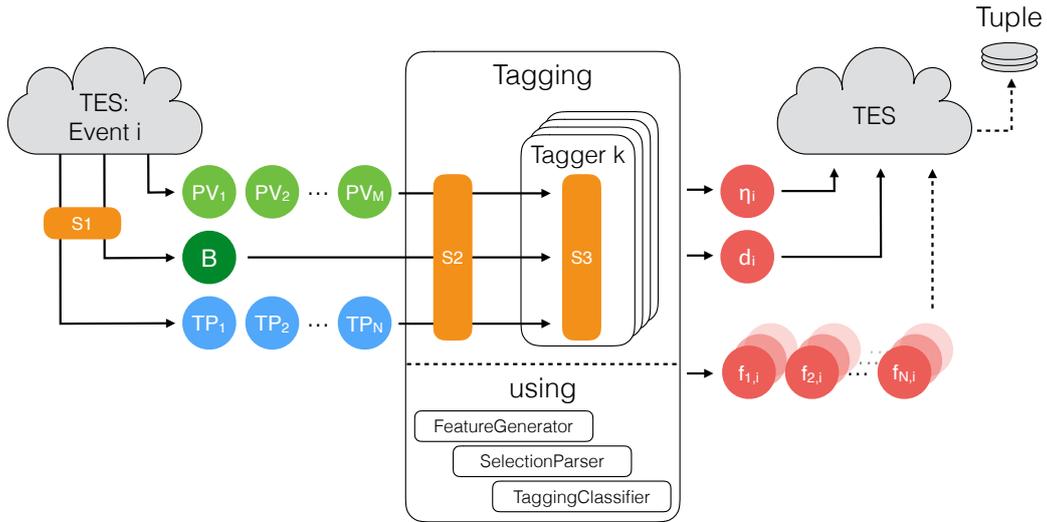
The classic Flavour Tagging software, used for Run 1 analyses, has been developed since more than a decade ago [63]. From a current point of view it lacks configurability and provides a deprecated code base which is hardly maintainable or re-usable. Facing the optimization campaign for Run 2, the software has partly been re-factored and re-implemented during the course of this thesis. While trying to use available LHCb software functionality where possible, the overall dependency of the LHCb software has been reduced to simplify future developments of independent lightweight sub modules. The tools and design patterns that are provided within the `DAVINCI` framework and used within the Flavour Tagging software will be presented in the next section. Afterwards the re-designed structure of the software will be shown in more detail.

## 5.1 Structure of LHCb Data Analysis Algorithms

The LHCb software is designed in terms of projects, which contain packages that provide specific functionalities and are independently developed. The Flavour Tagging software is provided within the `Phys/FlavourTagging` package. The packages are built with the `CMAKE` build tool and can include specific versions of other packages or external libraries of the LHCb software stack in a modular way. The core functionalities which are applied to the analysed data in the tuple production stage (see section 3.2.4) are implemented as `DaVinciAlgorithms`. These can be configured via `PYTHON` files by providing an implementation of the `GaudiTool` interface.

This design allows for modular definitions of analysis tool chains to prepare data tuples for specific analyses. A tool chain needs to specify a list of `DST` input files and a list of tools which internally call the corresponding algorithms with the given input data. Each algorithm is then responsible for loading the data into memory which is provided within `C++` objects containing e.g. particle or vertex properties from the so-called transient event store (TES). The results of the algorithms are finally returned to the TES or other algorithms or are written to disk.

To reduce the processing time of these algorithms, multiple instances of the software usually run on different nodes of a distributed computing grid which provide a homogeneous software stack. The distribution, configuration and management of the applications with their input and output files for the different server nodes is controlled within the `GANGA` software kit [64].



**Figure 5.1:** Structure of the LHCb Flavour Tagging software: For a given event, a signal candidate, together with a set of vertices (PV) and a set of tagging particles (TP) are read from the TES and passed to Flavour Tagging package. It further executes the different tagging algorithms  $i$  (see chapter 4) and returns the tag decisions  $d_i$  together with the mistag estimate  $\eta_i$  to the output file. Optionally, the matrix of all features  $f$  and all tagging particles can be written to the output file. Several selections (S1-S3) are applied. The re-implemented package provides general classes for feature generation, particle selection and MVA execution, which are described in the text of this section.

## 5.2 Re-factoring of the Flavour Tagging Software Package

The Flavour Tagging software has been re-factored with three main aspects in mind that were not covered in the classic implementation: The definition and configuration of selection cuts should be simple and well-structured, the latest features of the LHCb software should be used wherever necessary, while the general dependency on the LHCb software stack should be reduced, the usage of external multivariate analysis tools should be made possible and configurable.

After reviewing the logical structure of the Flavour Tagging process which is depicted in Figure 5.1, three major classes are provided, that implement simple interfaces for the main recurring tasks of the different tagging algorithms.

The generation of vectors of feature values for a specific event is implemented within a FeatureGenerator class. Its state is defined by a combination of a signal candidate particle, a set of tagging particles together with one specific tagging particle, an associated primary vertex and a list of additional  $pp$  interaction vertices, called pile-up vertices. Different sets of features can be requested while the functions to calculate these features can either be manually defined within the FeatureGenerator or can be in-

stantiated with LoKi functors which are widely used throughout the LHCb software. This centralized definition of features helps to use a consistent naming throughout the software package and improves reproducibility. The FeatureGenerator is able to store feature values of multiple tagging particle candidates such that a single event, together with the features of all tagging particles can be accessed and stored within an output file.

The selection of tagging particles is implemented within a SelectionParser class. It uses feature vectors of the FeatureGenerator and can easily be configured to apply different sets of selections to these features.

The execution of different multivariate analysis tools with a given set of features is implemented in the TaggingClassifier class. It provides a simple interface to pass a feature vector to an MVA tool and receive and return its output. Currently, this has been implemented for several BDTs and ANNs that are based on TMVA as well as for XGBoost BDTs [65, 66].

Each tagging algorithm can use these core classes to produce its raw BDT output value  $\eta_i$  while profiting from the configurability for adding new features and adjusting selection criteria. The specific behaviour of each algorithm can be configured within corresponding Python files, which bundle the main properties of that algorithm. Furthermore, default configurations of the Flavour Tagging software can be provided by combining several tagging algorithm configuration files.

The results of the re-implemented software have been compared with the corresponding output of the classic implementation for the OS single track taggers as well as the charm tagger on a bit-wise basis. The  $SS-p$ ,  $SS-\pi$ ,  $SS-K$  and Vertex Charge taggers are being transferred to the re-implemented framework at the time of this thesis.

Due to the fact that the Espresso Performance Monitor (EPM) [67] provides functionality to calibrate and combine tagging algorithms, no specific classes for these functions have been implemented. Instead, its core library can be used to apply and configure different calibration functions. Furthermore, it provides a comprehensive set of tools to measure the performance of the tagging algorithms for a given analysis.

# 6 Optimization Studies of Single Track Taggers for LHCb Run 2

As described in section 3.1, the LHC's centre-of-mass energy has been increased after the long shut down 2014/2015 from  $\sqrt{s} = 7/8\text{TeV}$  to  $\sqrt{s} = 13\text{TeV}$ . As a result, the kinematic properties of the particles within an event, which are used by the Flavour Tagging algorithms, have changed. Studies how these changes affect performance of these algorithms and how they might be adjusted and improved will be presented in this section while focusing on OS single track taggers and especially the muon tagger. These are performed, using  $B^+ \rightarrow J\psi K^+$  data and the newly implemented software, described in chapter 5.

## 6.1 Dataset properties and Run Period Differences

The optimization studies have been performed on LHCb Run 1 data from 2011 and 2012, recorded at a centre-of-mass energy of  $\sqrt{s} = 7\text{TeV}$  and  $\sqrt{s} = 8\text{TeV}$ , respectively, corresponding to an integrated luminosity of  $\mathcal{L}_{\text{int}} \approx 3\text{fb}^{-1}$  and Run 2 data from 2015 and 2016 (as of late August) at  $\sqrt{s} = 13\text{TeV}$ , corresponding to  $\mathcal{L}_{\text{int}} \approx 1.5\text{fb}^{-1}$ .

The  $B^+ \rightarrow J\psi K^+$  stripping line has been used to select the corresponding signal decay candidates. During the stripping, the events are required to fulfil the selection criteria, which will be described now and are also listed in Table 6.1. Stripping version 21 is used for Run 1 data, versions 23r1 and 26 are used for Run 2 2015 and 2016 data, respectively. All stripping configurations state that each event must contain at least one primary vertex. Each kaon is required to have PID information available and a track fit, fulfilling  $\chi^2_{\text{track}}/\text{ndf} < 5$ . The  $J\psi$  candidate, reconstructed in the  $\mu^+\mu^-$  mode, needs to have a mass that is compatible with the known  $J\psi$  mass,  $m_{J\psi, \text{PDG}} = 3096.916\text{MeV}/c^2$  [5], within  $\pm 150\text{MeV}/c^2$ . Its vertex fit has to fulfil  $\chi^2_{\text{VTX}} < 16$ , while PID information needs to be available for its daughter muons. They also need to have a transverse momentum of  $p_{\text{T}, \mu} > 500\text{MeV}/c$ , and their distance of closest approach (DOCA) must lead to  $\chi^2_{\text{DOCA}} < 20$ . The combined  $J\psi K^+$ -mass has to fulfil  $5150\text{MeV}/c^2 < m_{J\psi K^+} < 5450\text{MeV}/c^2$ , while the vertex fit quality is required to be  $\chi^2_{\text{VTX}}/\text{ndf} < 10$ . The kaon decaying from the signal candidate is furthermore required to have a transverse momentum of  $p_{\text{T}, K^+} > 500\text{MeV}/c$ .

After stripping,  $s$ Weights for the signal and background component are calculated based on the signal candidate  $B^+$  mass distribution (see sec-

tion 2.2.3). To further reduce the number of non-signal events for the maximum likelihood fit, additional selections are applied to the data. The track fit for the muons is required to provide  $\chi_{\text{track}}^2/\text{ndf} < 4$ . The  $J/\psi$  candidate mass range is tightened to  $m_{J/\psi K^+} \in m_{J/\psi K^+, \text{PDG}} \pm 60 \text{ MeV}/c^2$ . The  $K^+$  meson daughter from the signal candidate  $B^+$  needs to have a transverse momentum of  $p_{T, K^+} > 1 \text{ GeV}/c$  and a total momentum of  $p_{K^+} > 10 \text{ GeV}/c$ . After the decay tree has been fitted, the decay vertex of the reconstructed  $B^+$  meson has to fulfil  $\chi_{\text{VTX}}^2/\text{ndf} < 5$  and the reconstructed impact parameter to the signal candidate PV must lead to  $\chi_{\text{PV}}^2 < 25$ . The reconstructed  $B^+$  mass is required to have a difference smaller  $\pm 1 \text{ GeV}/c^2$ , with respect to the known mass  $m_{B^+, \text{PDG}} = 5279 \text{ MeV}/c^2$  [5]. To ensure that the reconstructed track provides a clean signature, either the event is required to only have one reconstructed PV or the  $\chi_{\text{IP}}^2$  including the next best PV is required to be  $\chi_{\text{IP}}^2(\text{next}) > 50$ . The decay time  $t$  of the  $B^+$  candidate needs to be in the range  $0.3 \text{ ps} < t < 14 \text{ ps}$  and its reconstruction uncertainty estimate less than  $0.12 \text{ ps}$ . Furthermore, several high level trigger decisions ‘on signal’ (TOS) are required, including a  $J/\psi$  decision of the HLT2 muon trigger, a HLT2 dimuon high mass decision without a HLT1 decision or a detached  $J/\psi$  HLT2 decision and a decision of the HLT1 muon track or a general L0 track decision.

After these selection steps, a maximum likelihood fit of the signal mass distribution, described by two Gaussian distributions with a shared mean value and the background mass distribution, described by an exponential function is applied. The corresponding plots of the mass distributions are shown in Figure 6.1. If not stated otherwise, the resulting  $s$ Weights  $w_i$  are used like regular weights for all following calculations and during BDT training to effectively use a pure signal data sample. The tagging power, defined in Equation 4.9 therefore becomes

$$\varepsilon_{\text{eff}} = \frac{\sum_{i=1}^N w_i (1 - 2\omega_i)^2}{\sum_{i=1}^N w_i}, \quad (6.1)$$

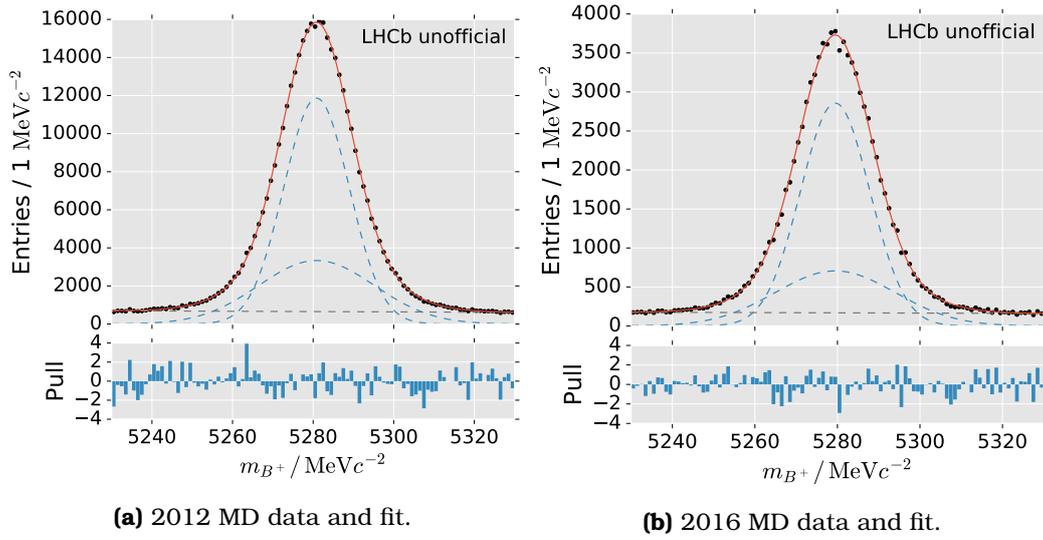
where  $\omega_i = 0.5$  for untagged events, leading to a dilution of  $D = 0$ , such that these events do not contribute to the tagging power. Furthermore, the whole 2011 and 2012 as well as 2015 and 2016 datasets for both LHCb magnet polarities are only separated by the Run periods.

The differences between Run 1 and Run 2 data can exemplary be seen in the distribution of some variables that are used by multiple tagging algorithms. Normed histograms of the number of tracks  $N_{\text{tracks}}$  per event and the probability estimation for the muon particle hypothesis  $P_{\text{NN}}(\mu)$  of the tagging particle after applying the muon tagger selections are shown in Figure 6.2 to illustrate these differences.

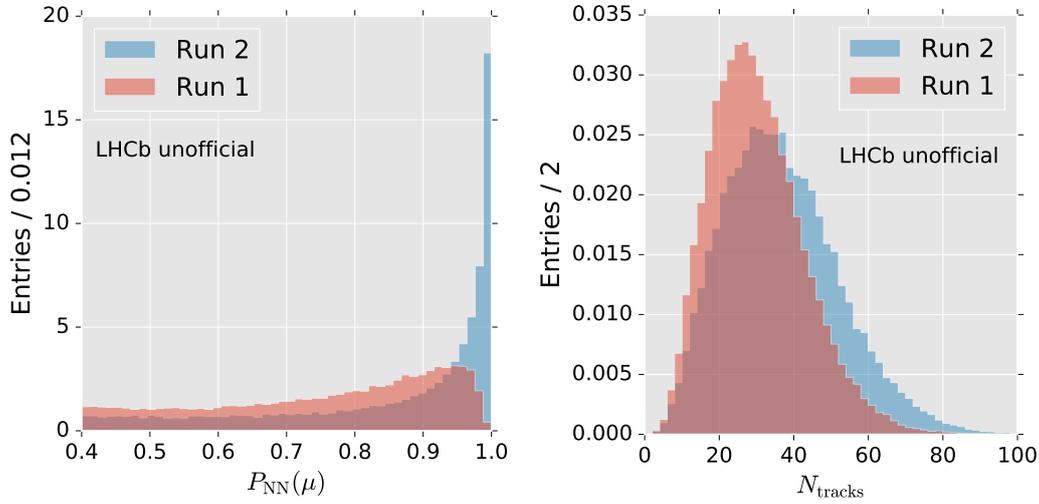
To quantize the difference between these data samples, a BDT is trained to predict the Run period on random, equally sized subsets of both datasets. If the difference between these samples is small, the area under the Receiver Operator Characteristics (ROC AUC) score is expected to be around 50%.

**Table 6.1:** Selections applied to the  $B^+ \rightarrow J/\psi K^+$  data sample during stripping and before  $s$ Weight calculation. Multiple kinematic features as well as PID information are taken into account. These are mainly  $\chi^2_{\text{track}}/\text{ndf}$  values of the track fit,  $\chi^2_{\text{VTX}}$  and  $\chi^2_{\text{DOCA}}$  values of the vertex and the DOCA fit, as well as momentum and mass cuts.

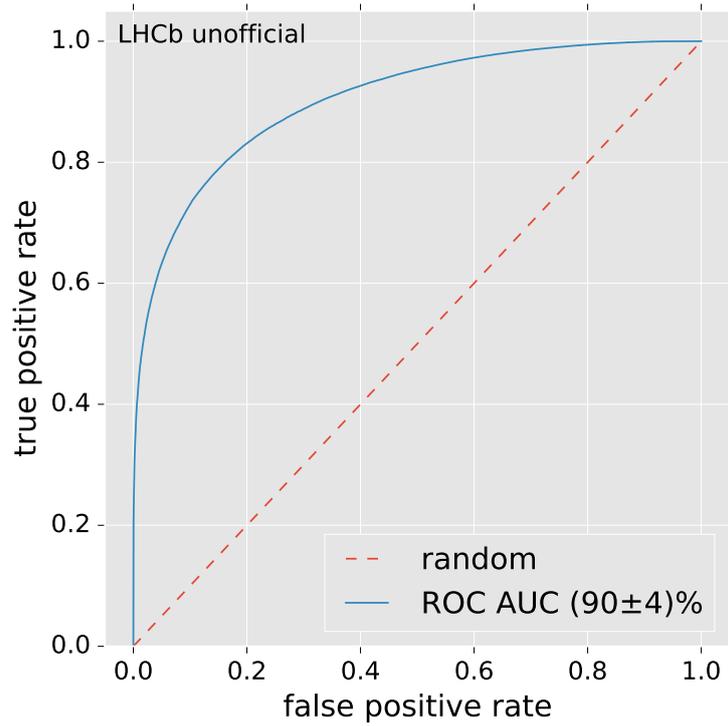
Selection Phase	Candidates	Variable	Selection Criteria	Unit
Stripping $B^+ \rightarrow J/\psi K^+$	$B^+$	$m_{J/\psi K^+}$	$\in [5150, 5450]$	$\text{MeV}/c^2$
		$\chi^2_{\text{VTX}}/\text{ndf}$	$< 10$	
	$J/\psi$	$m_{J/\psi}$	$\in [m_{J/\psi, \text{PDG}} \pm 150]$	$\text{MeV}/c^2$
		$\chi^2_{\text{VTX}}$	$< 16$	
$K^+$	PID( $K^+$ )	$> 0$	$\text{MeV}/c$	
	$\chi^2_{\text{track}}/\text{ndf}$	$< 5$		
$\mu^+ \mu^-$	$\mu^+ \mu^-$	$p_{\text{T}, K^+}$	$> 500$	$\text{MeV}/c$
		PID( $\mu$ )	$> 0$	
		$p_{\text{T}, \mu}$	$> 500$	
$s$ Plot Selection $B^+ \rightarrow J/\psi K^+$	$B^+$	$m_{J/\psi K^+}$	$\in [m_{J/\psi K^+, \text{PDG}} \pm 60]$	$\text{MeV}/c^2$
		$\chi^2_{\text{track}}/\text{ndf}$	$< 5$	
		$\chi^2_{\text{PV}}$	$< 25$	
		$\chi^2_{\text{next PV}}$	$> 50$	
		$\tau_{B^+}$	$\in [0.3, 14]$	
	$J/\psi$	$m_{J/\psi}$	$\in [m_{J/\psi, \text{PDG}} \pm 1]$	$\text{GeV}/c^2$
	$K^+$	$p_{\text{T}, K^+}$	$> 1$	$\text{GeV}/c$
$p_{, K^+}$		$> 10$	$\text{GeV}/c$	
$\mu^+ \mu^-$	$\mu^+ \mu^-$	$\chi^2_{\text{track}}/\text{ndf}$	$< 4$	



**Figure 6.1:** Exemplary mass distributions of the combined  $J/\psi K^+$  states, used to determine  $s$  Weights. The plots show subsets for the given year with 'magnet down' (MD) polarity of the LHCb magnet. The solid line represents a fit of a combination of two Gaussian distributions and one exponential function to describe the signal and the background, respectively. The bottom part of the plot shows the distribution of the pulls  $p_i$ , defined as  $p_i = \Delta m / \sigma_m$ , where  $\Delta m$  is the deviation of the fit from the corresponding data point at bin  $i$  and  $\sigma_m$  is the statistical uncertainty on that bin.



**Figure 6.2:** Distributions of the PID variable  $P_{NN}(\mu)$  and the number  $N_{tracks}$  of tracks per event, both variables are used to select tagging particle candidates and train the BDT for mistag estimation. These variables exemplary show the difference of Run 1 and Run 2 data. The differences make a re-optimization of the Flavour Tagging algorithms necessary for Run 2 analyses.



**Figure 6.3:** The ROC curve of a BDT that has been trained on all muon tagger selection variables, listed in Table 4.1, to classify the Run period. If the data differences between Run 1 and Run 2 data is small, a ROC AUC score of 50% is expected. The presented score has been calculated with a 30-fold cross validation.

All selection variables for the muon tagger (Table 4.1) are used as input variables. The ROC curve for the training dataset and the remaining data of a single split is shown in Figure 6.3, the average ROC score on a 30-fold cross-validation is  $\text{ROC}_{\text{cv}} = (90 \pm 4) \%$ . As a consequence of these differences, it is not expected that the performance of Flavour Tagging algorithms remains unchanged between both data taking periods. Measurements of the performance of the classic single track taggers, using polynomial logistic calibration will be shown in the next sections.

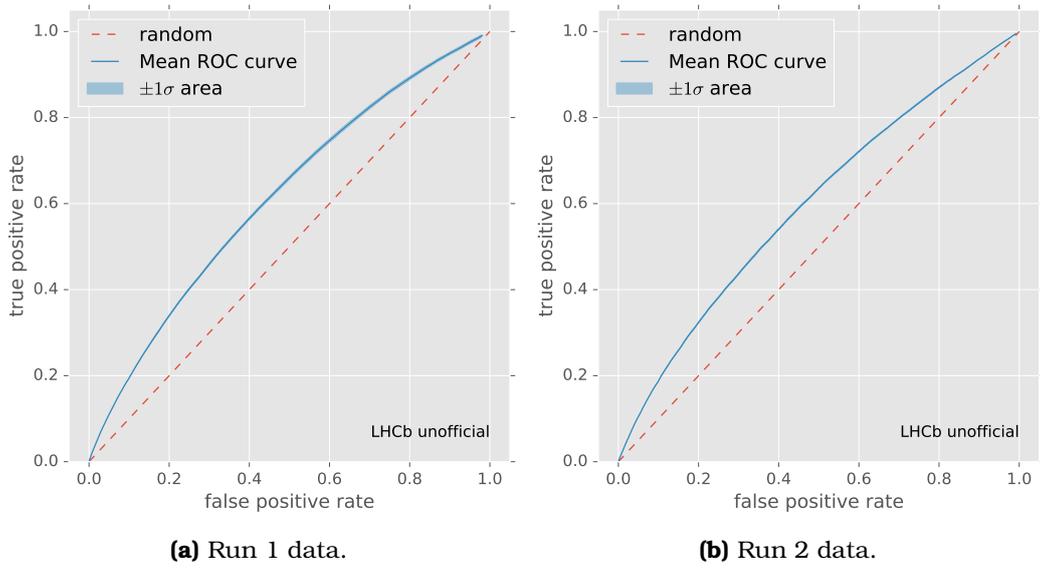
## 6.2 Muon Tagger with Logistic Calibration

The muon tagging algorithm introduced in section 4.2.1 is now applied to Run 1 and Run 2 data.

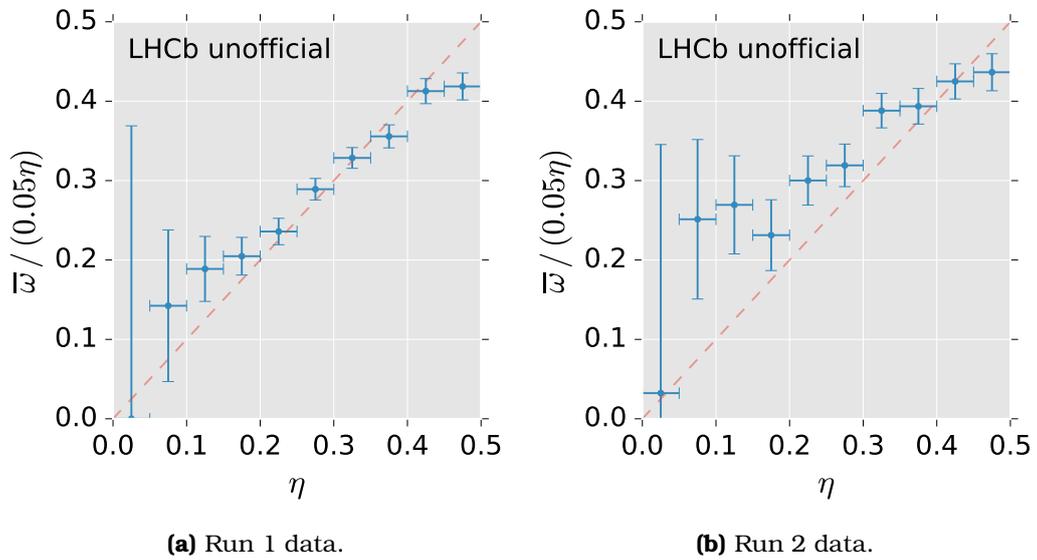
To verify the compatibility of new optimization strategies with classic implementations, the muon tagger has been implemented inside an IPYTHON based development environment, using JUPYTER NOTEBOOKS [68]. While the classic tagging uses ROOT's TMVA framework for multivariate analyses, XGBoost is used instead as it provides a simple interface, can be transferred into to re-factored Flavour Tagging software, and is highly parallelizable.

After applying the classic selections for the muon tagger on Run 1 data and selecting the particle with maximum  $p_T$  for every remaining event, the tagging efficiency is found to be  $\varepsilon_{\text{tag}} = (5.533 \pm 0.023) \%$ , the average mistag rate is  $\bar{w} = (31.31 \pm 0.23) \%$  and the average tagging power follows to be  $\bar{\varepsilon}_{\text{eff}} = (0.773 \pm 0.019) \%$ . While the selection efficiency and mistag rate are slightly above previous values, the average tagging power is compatible with values from Ref. [53].

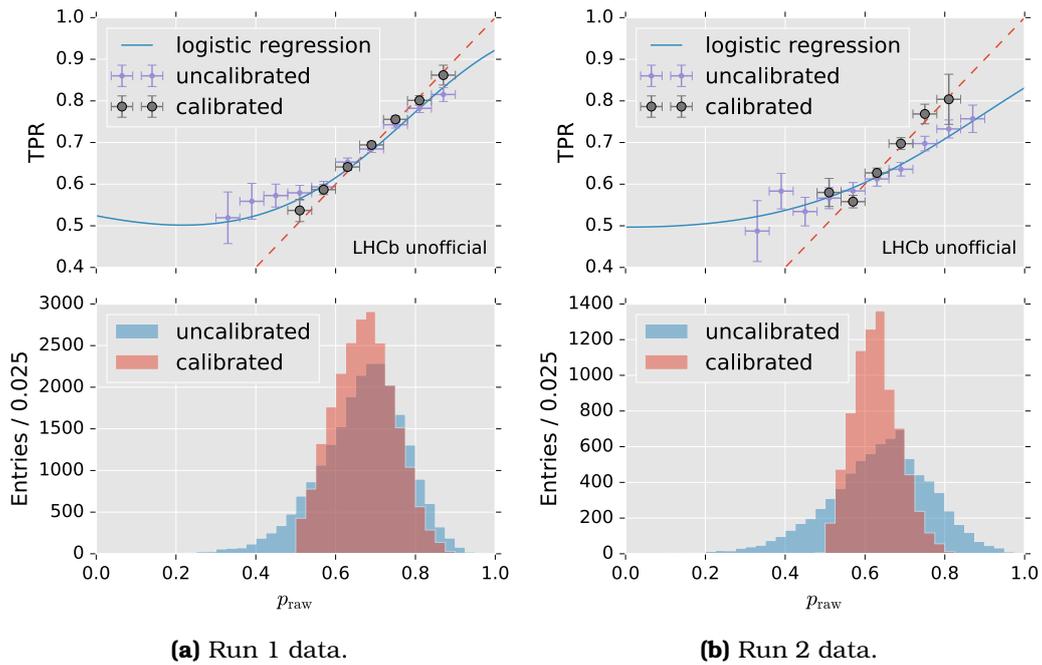
Afterwards, the data is randomly divided into equally sized training, calibration and testing sets. An XGBoost BDT is trained on the training subset to predict whether the charge of the tagging particle correctly predicts the initial tag of the signal candidate. The depth of the trees is limited to  $B_d = 4$ , the number of estimators is set to  $B_N = 400$  and the learning rate is set to  $B_l = 0.01$ . The corresponding area under ROC curve for Run 1 and Run 2 exemplary calculated on each test dataset are 60.88% and 61.54%, respectively, and the average curves for 30 bootstrap iterations are shown in Figure 6.4. The BDT output  $p_{\text{raw}}$  can be translated into the raw per event mistag estimate  $\eta_i$  via Equation 4.13. The binned distribution of the predicted  $\eta_i$  values is compared to the true mistag ratios for the corresponding bins in Figure 6.5. Assuming a probabilistic BDT output, a diagonal distribution is expected. The deviation of the data points from the diagonal therefore indicates an non-probabilistic BDT output. To be able to use the BDT estimations for calculation of the tagging power, a polynomial logistic calibration with the power  $k = 3$  (see section 2.2.2) is applied to the raw output  $p_{\text{raw}}$ . The distribution of the true positive rate in bins of the BDT output  $p_{\text{raw}}$ , together with its distribution, before and after the calibration is shown in Figure 6.6. After calibration, the per event tagging power can be



**Figure 6.4:** ROC curves of the calibrated mistag estimations, calculated with 30 bootstrap iterations, for Run 1 and Run 2 data.



**Figure 6.5:** Distribution of true mistag rates  $\bar{w}$  per predicted mistag estimate  $\eta_i$  bin. The red dotted line indicates the expected distribution, if the BDT output is a correct probability. The prediction underestimates the mistag rates at low values and slightly overestimates the values at high values.



**Figure 6.6:** Distribution of true positive rates (TPR) per BDT output  $p_{\text{raw}}$  bin (top figures) and the distributions of the BDT output (bottom figures) before and after applying a polynomial logistic regression. The purple markers show the raw data points, while the grey circles show the calibrated data. The calibration function is drawn as the blue line. Perfectly calibrated data is expected to follow the red dotted line.

**Table 6.2:** Calibrated performance values of the re-implemented classic muon tagger for both data taking periods. While the selection efficiency of the classic selections is higher for Run 2 data, the dilution  $\langle D^2 \rangle$  is lower than on Run 1 data, which indicates an overall worse quality of the tagging particles in terms of the muon tagger and leads to a decreased tagging power.

Period	$\varepsilon_{\text{tag}}/\%$	$\langle D^2 \rangle/\%$	$\varepsilon_{\text{eff}}/\%$
Run 1	$5.533 \pm 0.023$	$14.14 \pm 0.32$	$0.782 \pm 0.018$
Run 2	$7.52 \pm 0.05$	$9.0 \pm 0.6$	$0.67 \pm 0.05$

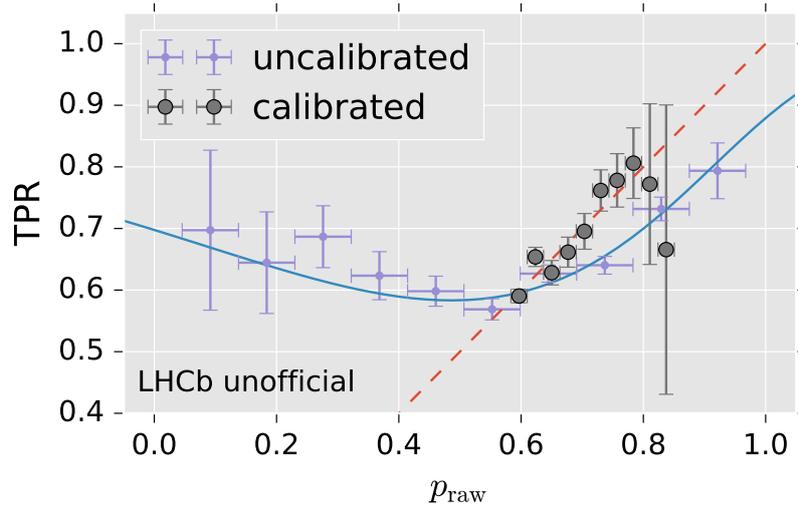
calculated based on the now probabilistic BDT estimate. The corresponding values are listed in Table 6.2. It can be seen that the tagging efficiency rises for Run 2 data while the quality of the tagging particle decreases which results in an overall worse tagging power for Run 2 data.

The three-fold cross validation is repeated 10 times to compute an estimate of the uncertainties for the average per event dilution  $\langle D^2 \rangle$  and the tagging power  $\varepsilon_{\text{eff}}$ . The results in Table 6.2 show the average performance values of all  $3 \cdot 10 = 30$  evaluation steps, while the error is the standard deviation of these values. As a consequence, the error estimates not only incorporate statistic but also systematic effects. Hence a detailed evaluation of the calibration parameters is not necessary and skipped in this study. A large error would furthermore indicate a badly designed algorithm for this specific problem. The relatively small uncertainties, which are obtained here, therefore indicate a robust tagging algorithm while the increased statistics of the whole Run 1 dataset reduce their absolute values with respect to the previous ones.

The BDT, trained on Run 1 data is furthermore applied to Run 2 data, to test whether a re-training could be omitted. The corresponding distribution of the raw BDT output is shown in Figure 6.7. One can see that the BDT output needs to be further calibrated, while the difference of the true positive rates is especially large at low  $p_{\text{raw}}$  values. A polynomial logistic regression with  $k = 4$  is applied to the raw BDT values and the resulting tagging power is  $\varepsilon_{\text{eff}} = (0.41 \pm 0.03)\%$ . Hereby, no bootstrapping is performed and the uncertainties are obtained by scaling the bootstrapped uncertainty of the full Run 2 algorithm. Conclusively, the Run 1 BDT cannot be applied to Run 2 data without a significant drop of the corresponding tagging power.

### Randomized Grid Search for Cut optimization

To study the optimization potential of this approach, a grid search for several selection parameters is performed. For every cut parameter combination, the data is split as described previously, a BDT is trained to predict whether the tag is correct, the BDT output is calibrated with a polynomial logistic regression, and the tagging power is evaluated. The BDT training, calibration and evaluation steps are again performed on different, equally



**Figure 6.7:** Distribution of the true positive rates in bins of the raw (violet) and calibrated (grey dots) BDT output. The BDT is trained on Run 1 data and applied to Run 2 data. The blue line is the corresponding polynomial logistic calibration function.

sized data subsets. To use the whole dataset for each step, the subsets are rotated after every iteration such that overall, three BDTs are trained, three calibrations are performed and each subset is used for testing. Due to the size of the grid, which is spanned by the cut ranges that are listed in Table 6.3, a random subset of  $N_{\text{gp}} = 300$  grid points is chosen. The cut parameter sets, which lead to the five best mean tagging-power scores within the corresponding three-fold values are listed in Table 6.4, while no bootstrapping is performed and the error estimations are omitted. It can be seen that no cut parameter combination is found that increases the tagging power on Run 1 or Run 2 data. It is generally noticeable, that different combinations of cut parameters lead to comparable tagging power values.

**Table 6.3:** Cut parameter ranges used for a randomized grid search of the cut-based muon tagger algorithm. The parameter grid contains a given number of values, listed in the ‘Samplesize’ column for every variable. In total, 300 parameter combinations are randomly chosen. The classic muon tagger algorithm is afterwards applied to Run 1 and Run 2 data with the given parameter set and the tagging power is calculated.

Variable	Range	Samplesize
$p_T/\text{GeV}$	[0, 3]	16
$p/\text{GeV}$	[1, 5]	11
$\min(\Delta\phi)$	[0, 0.5]	6
$P_{\text{GHOST}}(\text{track})$	[0, 0.6]	6
$ \chi_{\text{IP(PU)}} $	[1, 4]	16
$\chi_{\text{track}}^2/\text{ndf}$	[2, 5]	6
$P_{\text{NN}}(\mu)$	[0, 0.6]	6

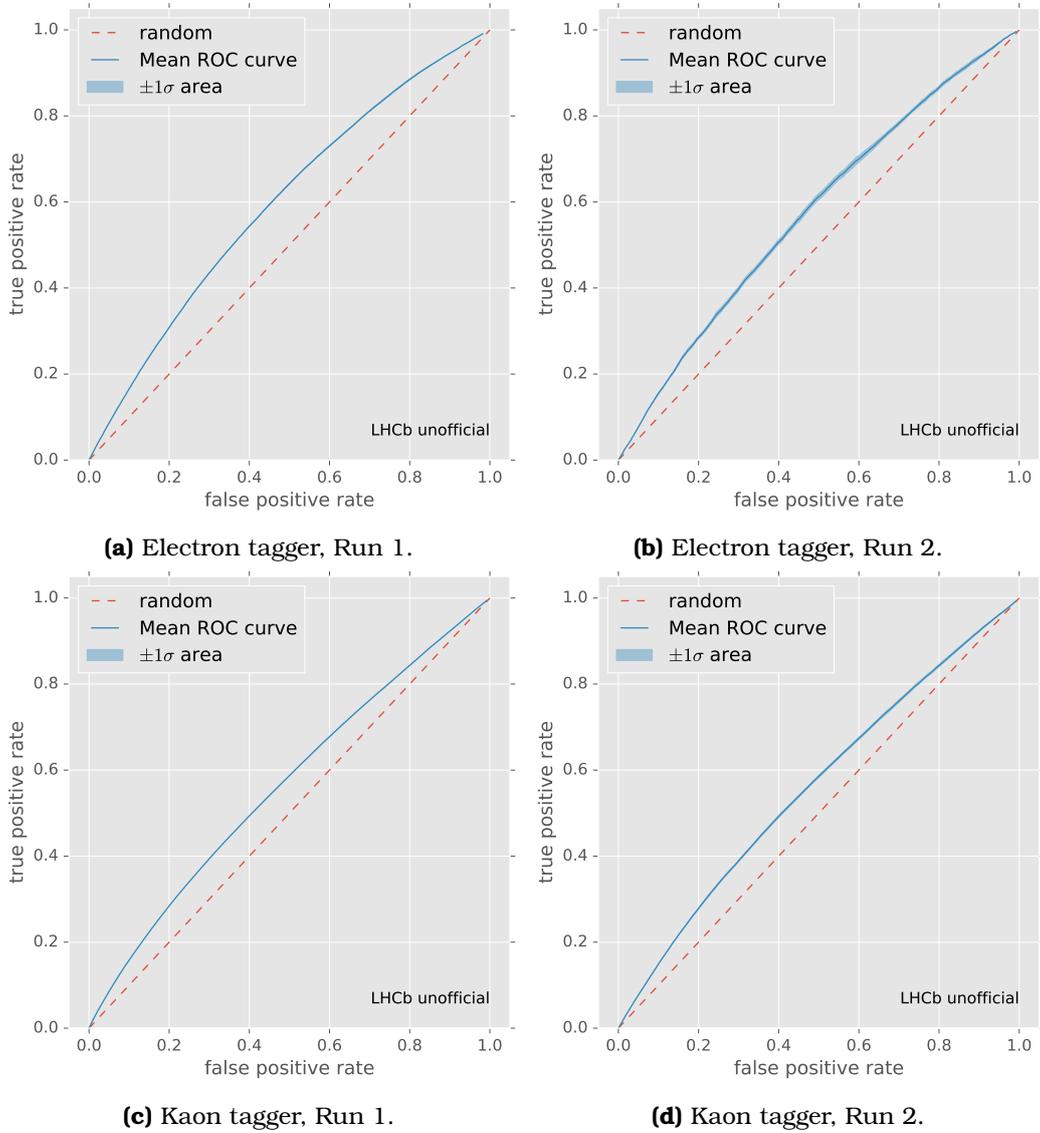
**Table 6.4:** Tagging power score for different cut parameter combinations. The best 20 combinations are presented. While the single cut parameters vary widely, the best scores are compatible with the classic cut parameter set. Errors on the tagging power score are omitted.

Period	Variable	Values
Run 1	$p_T/\text{GeV}/c >$	0.8 1.0 1.0 1.0 0.2
	$p/\text{GeV}/c >$	4.2 3.2 2.8 4.4 2.0
	$\min(\Delta\phi) >$	0.0 0.0 0.0 0.0 0.0
	$P_{\text{GHOST}}(\text{track}) <$	0.36 0.6 0.36 0.36 0.48
	$ \chi_{\text{IP(PU)}}  >$	1.4 3.0 1.6 3.6 1.4
	$\chi_{\text{track}}^2/\text{ndf} <$	4.4 2.6 4.4 4.4 5.0
	$P_{\text{NN}}(\mu) <$	0.36 0.48 0.12 0.48 0.36
	$\varepsilon_{\text{eff}}/\%$	0.72 0.72 0.71 0.71 0.71
Run 2	$p_T/\text{GeV}/c >$	0.2 0.8 1.0 0.6 1.0
	$p/\text{GeV}/c >$	1.2 3.8 3.0 2.2 1.4
	$\min(\Delta\phi) >$	0.0 0.0 0.0 0.0 0.0
	$P_{\text{GHOST}}(\text{track}) <$	0.48 0.24 0.48 0.6 0.36
	$ \chi_{\text{IP(PU)}}  >$	2.0 3.4 2.0 2.8 3.4
	$\chi_{\text{track}}^2/\text{ndf} <$	3.2 3.8 3.8 5.0 2.6
	$P_{\text{NN}}(\mu) <$	0.6 0.36 0.12 0.6 0.36
	$\varepsilon_{\text{eff}}/\%$	0.64 0.64 0.63 0.63 0.62

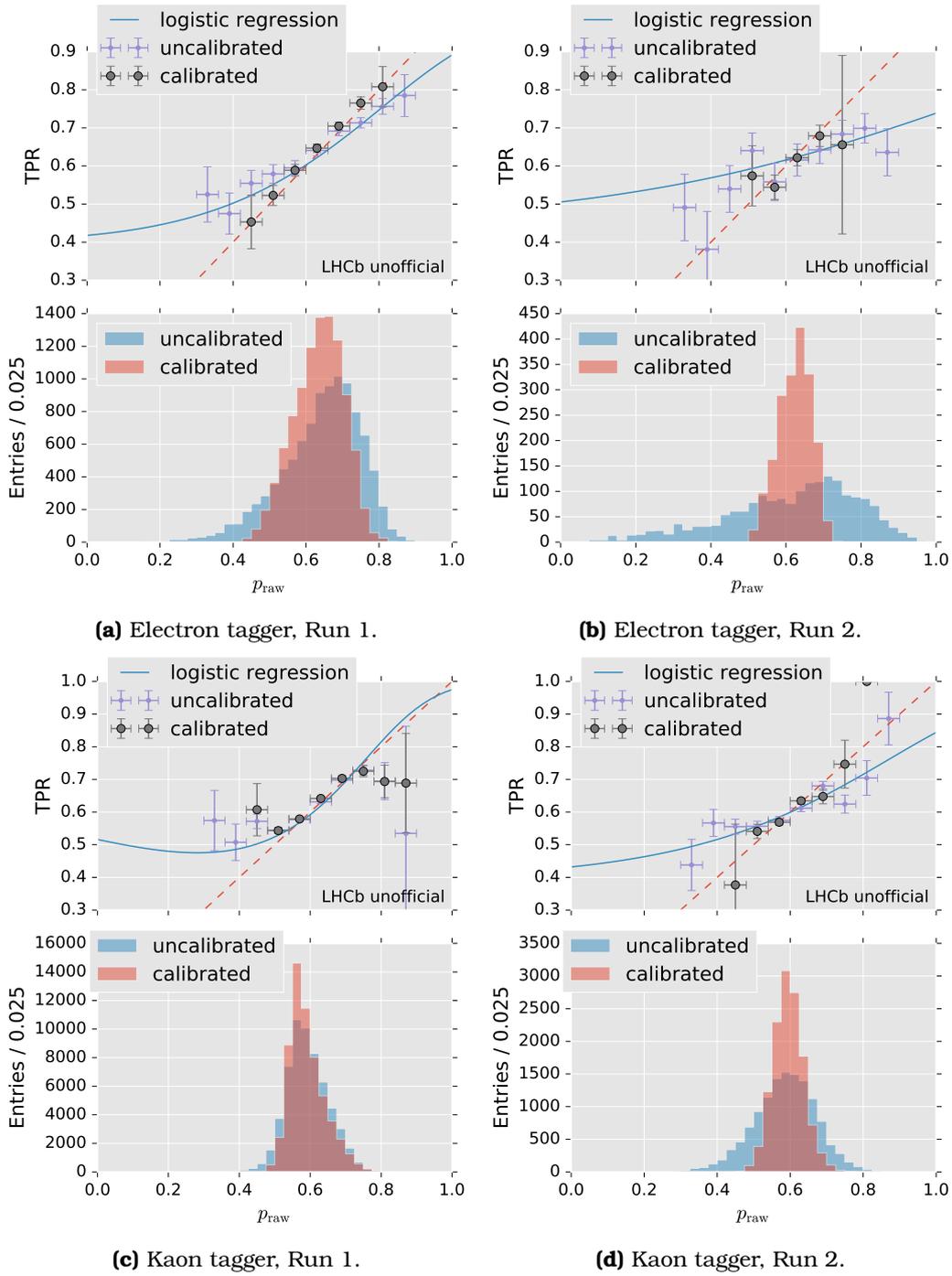
### 6.3 Electron and Kaon Taggers with Logistic Calibration

Analogous to the previous section, the performance of the electron and kaon tagger is determined for Run 1 and Run 2 data. For both taggers, the set of selections of the cut based Run 1 algorithms, which is listed in Table 4.1 is applied. While the tagging efficiency for the electron tagger is compatible with previous values, the tagging efficiency for the kaon tagger exceeds previous values. After selecting the tagging particle with maximum transverse momentum within each remaining event, a random three-fold split is applied to the data. A BDT is trained on one part of the data in both cases, while the training target is whether the tagging particle's charge is correctly linked with the signal particle's flavour. The electron tagger's BDT is trained with the same depth of trees, number of estimators and learning rate as the muon tagger, namely  $B_d = 4$ ,  $B_N = 400$  and  $B_l = 0.01$ , while the kaon tagger's BDT is configured with  $B_N = 500$  estimators and a depth of  $B_d = 3$ . The uncertainties on the dilution  $\langle D^2 \rangle$  and tagging efficiency  $\varepsilon_{\text{eff}}$  are obtained as in the previous section. The three-fold split is performed and evaluated 10 times while the order of events is randomized before each iteration. Therefore, the uncertainties combine systematic and numeric effects. The ROC curves for these BDTs for one bootstrapping iteration are shown in Figure 6.8.

The resulting BDT output for both taggers is calibrated with a polynomial logistic regression of the power  $k = 4$ . The distributions of the true positive rate in bins of the predicted correct tag probability  $p_{\text{raw}}$  are shown in Figure 6.9. While the polynomial logistic calibration is able to transform the BDT output well for the Run 1 electron tagger, the Run 2 electron and the kaon BDT data are not described very well. Despite this, the calibrated distributions for the Run 2 electron and kaon taggers are compatible with the expected distributions. In case of the Run 1 kaon tagger, the calibrated BDT output and with that the mistag estimation  $\eta_i$  overestimates the true probability. The resulting performance values for tagging efficiency, dilution and tagging power are listed in Table 6.5. Both single track taggers additionally show the same tendency towards a decreasing tagging power on Run 2 data, which has been also observed in the muon tagger.



**Figure 6.8:** Exemplary ROC curves for the electron and kaon tagger split for Run 1 and Run 2 data. Each plot displays the ROC curves on the test and calibration datasets, which are independent from the BDT training dataset. The dashed line indicates the curve of a random selection.



**Figure 6.9:** Exemplary distributions of true positive rates (TPR) per BDT output  $p_{\text{raw}}$  bin (top figures) and the distributions of the BDT output (bottom figures) before and after applying a polynomial logistic regression for the electron and kaon taggers. The calibration function is drawn as the blue line. Perfectly calibrated data is expected to follow the red dotted line.

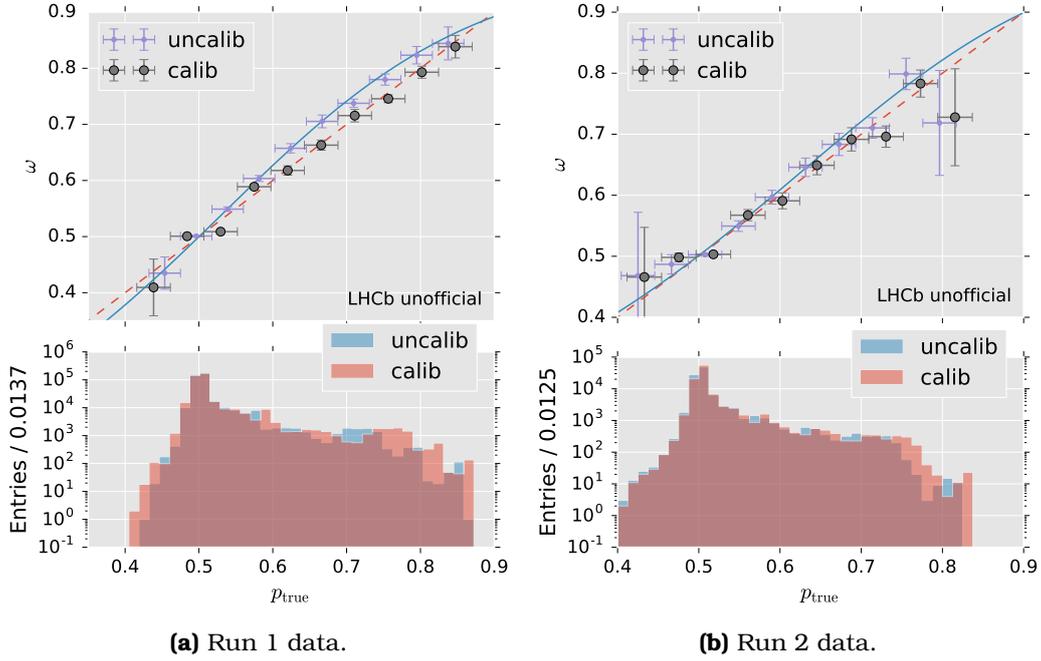
**Table 6.5:** Calibrated performance values of the re-implemented classic electron and kaon taggers for both data taking periods. While the tagging power values for Run 1 data are compatible to former values, the trend of a generally worse tagging power on Run 2 data can also be seen here. The kaon tagger, marked with ‘(\*)’, lists the result for a modified set of selection criteria, which is explained in the text.

Tagger	Period	$\varepsilon_{\text{tag}}/\%$	$\langle D^2 \rangle/\%$	$\varepsilon_{\text{eff}}/\%$
$e$	Run 1	$2.432 \pm 0.015$	$7.0 \pm 1.0$	$0.243 \pm 0.011$
	Run 2	$1.903 \pm 0.026$	$9.0 \pm 0.6$	$0.134 \pm 0.019$
$K$	Run 1	$14.04 \pm 0.04$	$4.62 \pm 0.14$	$0.649 \pm 0.020$
	Run 2	$12.45 \pm 0.07$	$4.38 \pm 0.31$	$0.54 \pm 0.04$
$K (*)$	Run 1	$16.46 \pm 0.04$	$2.62 \pm 0.08$	$0.431 \pm 0.013$
	Run 2	$14.63 \pm 0.08$	$2.28 \pm 0.20$	$0.333 \pm 0.029$

## 6.4 Studies of a BDT-based, Inclusive Muon Tagger

To investigate a possible improvement of the muon tagger, adopting the strategies that are used by e.g. the charm tagger and some SS taggers, all selection cuts are dropped. An XGBoost BDT is trained on one third of the data with the same target as in the classic implementation, predicting whether a tag decision based on the tagging particle charge is correct. In contrast to the classic strategy, the events contain all tagging particles after the pre-selections S1 and S2. Afterwards only tagging particles with maximum BDT value in each event are kept and a polynomial logistic regression is applied to receive a probabilistic BDT value. The raw BDT output  $p_{\text{raw}}$ , the calibrated output  $\eta_i$ , and the calibration function are shown in Figure 6.10. Finally, the tagging power is computed after different cuts on the mistag estimate  $\eta_i(p_{\text{raw}})$  are applied, ranging from  $\eta_i \leq 0$  to  $\eta_i \leq 0.5$ . The  $\varepsilon_{\text{eff}}-\eta$  curves for Run 1 and Run 2 is shown in Figure 6.11. Although the correlation is not linear, the tagging power increases monotonically. This behaviour is expected due to the definition of the per event dilution  $D_i$ , which is approximately zero for events with low quality while the tagging efficiency is 100% for  $\eta \geq 0.5$ . The maximum tagging power reached with this approach is listed in Table 6.6 and exceeds the classic tagging power (see Table 4.2) for both data periods, while the Run 1 performance is still better than the Run 2 performance.

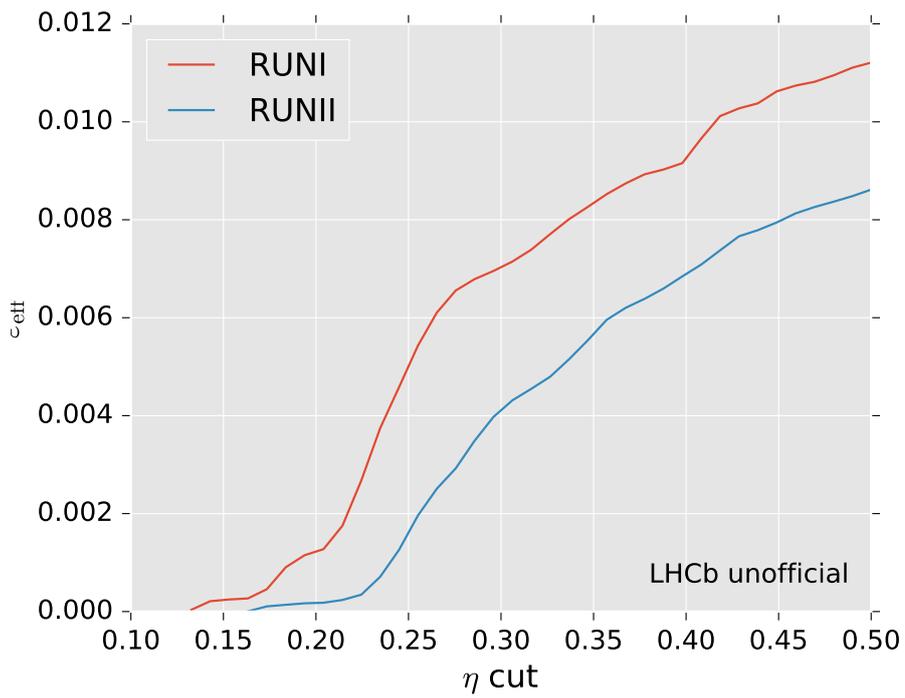
To check whether the selected particles are valid muons, the fraction  $f_{\text{same}}$  of events for which the same tagging particle was chosen in the classic and the inclusive algorithms is computed. No selection for the BDT parameter is applied, since this provides the best tagging power. Furthermore only the subset of events for which both algorithms provide a tag decision is taken into account. The fraction of  $N_{\text{same}}$  events in which both algorithms use the same tagging particle and the number  $N_{\text{sel}}$  of classically selected events is



**Figure 6.10:** Distribution of true positive rates (TPR) per BDT output  $p_{\text{raw}}$  bin (top figures) and the distributions of the BDT output (bottom figures) before and after applying a polynomial logistic regression. In contrast to the plots shown in Figure 6.6, no selections are applied to the tagging particles before BDT training. The purple markers show the raw data points, while the grey circles show the calibrated data. The calibration function is drawn as the blue line. Perfectly calibrated data is expected to follow the red dotted line. Note that the view is zoomed in and the histogram is scaled logarithmically.

**Table 6.6:** Maximum performance values of the BDT based muon tagger for both data taking periods. Since effectively no cuts are applied to the events, the tagging efficiency is 100% and the dilution equals the tagging power. The error estimations arise from a three-fold cross-validation and does not take the sample size into account.

Period	$\varepsilon_{\text{tag}}/\%$	$\langle D^2 \rangle/\%$	$\varepsilon_{\text{eff}}/\%$
Run 1	100	$1.11 \pm 0.04$	$1.11 \pm 0.04$
Run 2	100	$0.83 \pm 0.06$	$0.83 \pm 0.06$



**Figure 6.11:** Distribution of tagging power of a BDT based muon tagger for Run 1 and Run 2 data against different cut values on the predicted mistag estimate  $\eta_i$ . It can be seen that the overall performance of this algorithm is better on Run 1 data. The best tagging power is reached for the loosest  $\eta$  cut, which outperforms the classic Run 1 tagging power, even for Run 2 data.

found to be

$$f_{\text{same}}^{\mu} = \frac{N_{\text{same}}^{\mu}}{N_{\text{sel}}^{\mu}} = \begin{cases} (98.6 \pm 0.5) \% & \text{Run 1} \\ (94.0 \pm 0.9) \% & \text{Run 2} \end{cases} . \quad (6.2)$$

This indicates that almost all tagging particles, which are selected in the classic muon tagger algorithm, are also selected within the BDT-based approach.

Furthermore, the correlation with the electron and kaon tagger are tested to study whether a later combination of these taggers might overestimate the tagging power by using the same tagging particle multiple times. By calculating the fraction of events with same tagging particles within the classically selected events, defined in Equation 6.2, it can be seen that the correlation with the kaon tagger is small and will be neglected:

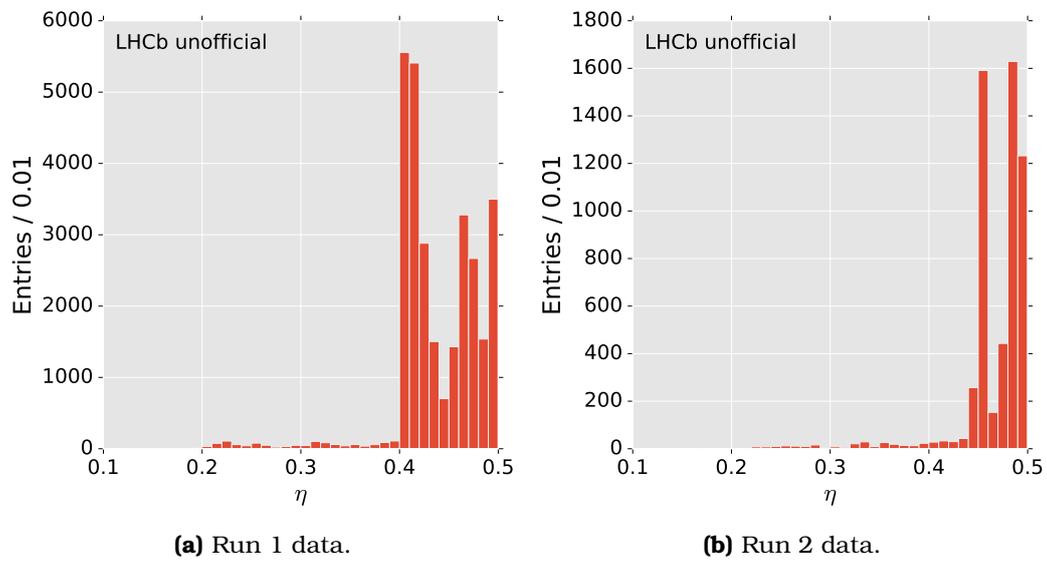
$$f_{\text{same}}^K = \frac{N_{\text{same}}^{\mu}}{N_{\text{sel}}^K} = \begin{cases} (0.65 \pm 0.06) \% & \text{Run 1} \\ (0.2 \pm 0.5) \% & \text{Run 2} \end{cases} . \quad (6.3)$$

A larger overlap is found with the electron tagger:

$$f_{\text{same}}^e = \frac{N_{\text{same}}^{\mu}}{N_{\text{sel}}^e} = \begin{cases} (74 \pm 6) \% & \text{Run 1} \\ (70.5 \pm 1.6) \% & \text{Run 2} \end{cases} . \quad (6.4)$$

Nevertheless, the contribution of the corresponding events to the tagging power of the BDT-based muon tagger is small due to the low muon-specific quality of these events. The distribution of the calibrated  $\eta_i(p_{i,\text{raw}})$  values is shown in Figure 6.12. Therefore, the events in which both, electron and BDT-based muon tagger use the same tagging particles are ignored for the BDT-based muon tagger. As a consequence, the resulting tagging power is not correlated with the electron tagger, but still improves the classic tagging power for Run 1 and Run 2 data (see Table 6.2):

$$\varepsilon_{\text{eff}} = \begin{cases} (1.09 \pm 0.04) \% & \text{Run 1} \\ (0.83 \pm 0.06) \% & \text{Run 2} \end{cases} . \quad (6.5)$$



**Figure 6.12:** Distribution of calibrated BDT values  $\eta_i$  of tagging particles which are also used by the electron tagger for Run 1 and Run 2 data. Since the contribution of these events to the BDT-based muon tagger performance is negligible, they are ignored in the new approach.



## 7 Conclusion and Outlook

In the course of this thesis, it has been shown that a re-optimization campaign for the Flavour Tagging algorithms is required after the LHC upgrade. The implementation of these algorithms within a re-implemented Flavour Tagging software has been simplified in. The updated and newly developed algorithms can therefore interchangeably be used for upcoming analyses with minimal configuration effort. Moreover the training and testing data, used for development and optimization of the Flavour Tagging algorithms is ensured to originate from the same internal functionality as the production tuples used for physics analyses.

This software has been used to perform an exemplary optimization process for the muon tagger and calculate the performance of all OS single track taggers for Run 1 and Run 2 data. Hereby, the data differences after the LHC centre-of-mass energy upgrade were found to be significant, such that BDT was well able to determine Run 1 from Run 2 data. The influence of these differences were tested for the cut-based single track taggers. Their performance was found to be  $\varepsilon_{\text{eff}}^{\mu} = (0.782 \pm 0.018) \%$ ,  $\varepsilon_{\text{eff}}^e = (0.243 \pm 0.011) \%$  and  $\varepsilon_{\text{eff}}^K = (0.649 \pm 0.020) \%$  for the muon, electron and kaon tagger, respectively. While these values were compatible with the previous values for Run 1 data, the performance dropped to  $\varepsilon_{\text{eff}}^{\mu} = (0.67 \pm 0.05) \%$ ,  $\varepsilon_{\text{eff}}^e = (0.134 \pm 0.019) \%$  and  $\varepsilon_{\text{eff}}^K = (0.54 \pm 0.04) \%$  for Run 2 data.

To find improvements of the classic strategy, 300 random selection parameter combinations have been applied for the muon tagger. In each step, its performance was measured with the per event tagging power. This led to several parameter combinations with a slightly lower tagging power, which has decreased by approximately 8% on Run 1 and Run 2 data. It is worth mentioning that the parameter combinations, resulting from the grid search were showing some fluctuations although their tagging power was comparable.

After testing the classic muon tagger strategy, a new, BDT-based, inclusive approach for that tagger has been implemented. Instead of applying cut-based selection criteria, a BDT is used to choose a ‘best’ tagging particle in every event. It has been shown that this algorithm selects most of the tagging particles in the former muon tagger, while having a negligible correlation with the kaon tagger. Ignoring tagging particles, which are used by the electron tagger, eliminates the correlation with this tagger while reducing the tagging power by only a small fraction. Therefore, the tagging power of the muon tagging algorithm could be significantly increased with this new approach from  $\varepsilon_{\text{eff}} = (0.782 \pm 0.018) \%$  to  $\varepsilon_{\text{eff}} = (1.09 \pm 0.04) \%$  on Run

1 data and  $\varepsilon_{\text{eff}} = (0.67 \pm 0.05) \%$  to  $\varepsilon_{\text{eff}} = (0.83 \pm 0.06) \%$  on Run 2 data. Since the correlations with the vertex charge and the charm tagger has not been investigated, further studies should be performed before using this approach for upcoming analyses.

Furthermore, all presented studies have only been performed on  $B^+ \rightarrow J/\psi K^+$  data. This could be expanded to use  $B^0 \rightarrow J/\psi K^{*0}$  or corresponding simulated data samples, as well. The differences in the MVA variable distributions, which are especially large in the  $P_{\text{NN}}$  variables and originate from changes in the PID software in Run 2, could be calibrated to reconstruct the previous shapes. The grid search was limited by memory resources and used only one third of the Run 1 data and one half of the Run 2 data. This problem could be circumvented by implementing less memory-consuming approaches for the training and calibration phase of the optimization campaign. More multivariate analysis tools could be tested, especially ones that provide Deep Neural Networks that have not been covered in this thesis.

Conclusively, a re-optimization campaign can be efficiently performed within the re-implemented Flavour Tagging software, while new approaches for the tagging algorithms promise to restore or even exceed the Run 1 performance of Flavour Tagging for Run 2 analyses.

# Bibliography

- [1] G. Aad et al. (The ATLAS collaboration). ‘Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC’. In: *Physics Letters B* 716.1 (2012), pp. 1–29. doi: 10.1016/j.physletb.2012.08.020.
- [2] S. Chatrchyan et al. (The CMS collaboration). ‘Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC’. In: *Physics Letters B* 716.1 (2012), pp. 30–61. doi: 10.1016/j.physletb.2012.08.021.
- [3] S. F. Novaes. ‘Standard Model: An Introduction’. In: *Particles and fields. Proceedings, 10th Jorge Andre Swieca Summer School*. 1999, pp. 5–102. arXiv: hep-ph/0001283 [hep-ph].
- [4] O. W. Greenberg. ‘CPT Violation Implies Violation of Lorentz Invariance’. In: *Phys. Rev. Lett.* 89 (Nov. 2002), p. 231602. doi: 10.1103/PhysRevLett.89.231602.
- [5] K. A. Olive et al. (The Particle Data Group collaboration). ‘Review of Particle Physics’. In: *Chin. Phys.* C38 (2014), p. 090001. doi: 10.1088/1674-1137/38/9/090001.
- [6] R. Aaij et al. (The LHCb collaboration). ‘Observation of  $J/\psi p$  Resonances Consistent with Pentaquark States in  $\Lambda_b^0 \rightarrow J/\psi K^- p$  Decays’. In: *Phys. Rev. Lett.* 115 (Aug. 2015), p. 072001. doi: 10.1103/PhysRevLett.115.072001.
- [7] R. Aaij et al. (The LHCb collaboration). ‘Observation of  $J/\psi \phi$  structures consistent with exotic states from amplitude analysis of  $B^+ \rightarrow J/\psi \phi K^+$  decays’. Tech. rep. LHCb-PAPER-2016-018. June 2016. arXiv: 1606.07895.
- [8] R. Aaij et al. (The LHCb collaboration). ‘Amplitude analysis of  $B^+ \rightarrow J/\psi \phi K^+$  decays’. Tech. rep. LHCb-PAPER-2016-019. June 2016. arXiv: 1606.07898.
- [9] F. Halzen and A. Martin. ‘Quarks and leptons: an introductory course in modern particle physics’. Wiley, 1984. isbn: 0471887412.
- [10] M. Gonzalez-Garcia and M. Maltoni. ‘Phenomenology with massive neutrinos’. In: *Physics Reports* 460.1–3 (2008), pp. 1–129. doi: 10.1016/j.physrep.2007.12.004.
- [11] L. Wolfenstein. ‘Parametrization of the Kobayashi-Maskawa Matrix’. In: *Phys. Rev. Lett.* 51 (Nov. 1983), pp. 1945–1947. doi: 10.1103/PhysRevLett.51.1945.

- [12] J. Charles et al. (The CKMfitter collaboration). ‘Current status of the standard model CKM fit and constraints on  $F = 2$  new physics’. In: *Phys. Rev. D* 91 (Apr. 2015), p. 073007. doi: 10.1103/PhysRevD.91.073007.
- [13] J. H. Christenson et al. ‘Evidence for the  $2\pi$  Decay of the  $K_2^0$  Meson’. In: *Phys. Rev. Lett.* 13 (July 1964), pp. 138–140. doi: 10.1103/PhysRevLett.13.138.
- [14] E. Abouzaid et al. ‘Precise measurements of direct  $CP$  violation,  $CPT$  symmetry, and other parameters in the neutral kaon system’. In: *Phys. Rev. D* 83 (May 2011), p. 092001. doi: 10.1103/PhysRevD.83.092001.
- [15] R. Aaij et al. (The LHCb collaboration). ‘Search for violations of Lorentz invariance and  $CPT$  symmetry in  $B_{(s)}^0$  mixing’. In: *Phys. Rev. Lett.* 116 (Mar. 2016), 241601. 18 p. doi: 10.1103/PhysRevLett.116.241601.
- [16] J. McDonough et al. ‘New searches for the  $C$ -noninvariant decay  $\pi^0 \rightarrow 3\gamma$  and the rare decay  $\pi^0 \rightarrow 4\gamma$ ’. In: *Phys. Rev. D* 38 (Oct. 1988), pp. 2121–2128. doi: 10.1103/PhysRevD.38.2121.
- [17] C. S. Wu et al. ‘Experimental Test of Parity Conservation in Beta Decay’. In: *Phys. Rev.* 105 (Feb. 1957), pp. 1413–1415. doi: 10.1103/PhysRev.105.1413.
- [18] J. P. Lees et al. (The BABAR collaboration). ‘Observation of Time-Reversal Violation in the  $B^0$  Meson System’. In: *Phys. Rev. Lett.* 109 (Nov. 2012), p. 211801. doi: 10.1103/PhysRevLett.109.211801.
- [19] R. Aaij et al. (The LHCb collaboration). ‘Measurement of  $CP$  violation in  $B^0 \rightarrow D^+ D^-$  decays’. In: (Aug. 2016). arXiv: 1608.06620.
- [20] C. Jarlskog. ‘Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal  $CP$  Nonconservation’. In: *Phys. Rev. Lett.* 55 (Sept. 1985), pp. 1039–1042. doi: 10.1103/PhysRevLett.55.1039.
- [21] R. Aaij et al. (The LHCb collaboration). ‘First Observation of  $CP$  Violation in the Decays of  $B_s^0$  Mesons’. In: *Phys. Rev. Lett.* 110 (May 2013), p. 221601. doi: 10.1103/PhysRevLett.110.221601.
- [22] R. Aaij et al. (The LHCb collaboration). ‘Measurement of the Semileptonic  $CP$  Asymmetry in  $B^0$ - $\bar{B}^0$  Mixing’. In: *Phys. Rev. Lett.* 114 (Jan. 2015), p. 041601. doi: 10.1103/PhysRevLett.114.041601.
- [23] R. Aaij et al. (The LHCb collaboration). ‘Measurement of the  $CP$  Asymmetry in  $B_s^0$ - $\bar{B}_s^0$  Mixing’. In: *Phys. Rev. Lett.* 117 (Aug. 2016), p. 061803. doi: 10.1103/PhysRevLett.117.061803.
- [24] M. Pivk and F. L. Diberder. ‘ $\sqrt{s}$ Plot: A statistical tool to unfold data distributions’. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1–2 (2005), pp. 356–369. doi: 10.1016/j.nima.2005.08.106.

- [25] C. I. Relations. ‘Member States’. 2016. URL: <http://international-relations.web.cern.ch/International-Relations/ms/> (visited on Aug. 25, 2016).
- [26] ‘CERN Annual Personnel Statistics 2015’. CERN-HR-STAFF-STAT-2015. 2015. URL: <https://cds.cern.ch/record/2154389>.
- [27] C. O’Luanaigh. ‘The Large Electron-Positron Collider’. July 2012. URL: <https://cds.cern.ch/record/1997351> (visited on Aug. 25, 2016).
- [28] C. O’Luanaigh. ‘The Large Hadron Collider’. Jan. 2014. URL: <https://cds.cern.ch/record/1998498> (visited on Aug. 25, 2016).
- [29] L. Evans and P. Bryant. ‘LHC Machine’. In: *Journal of Instrumentation* 3.08 (2008), S08001. doi: 10.1088/1748-0221/3/08/S08001.
- [30] C. Lefevre. ‘LHC: the guide (English version). Guide du LHC (version anglaise)’. Feb. 2009. URL: <https://cds.cern.ch/record/1165534>.
- [31] M. Benedikt et al. ‘LHC Design Report’. Geneva: CERN, 2004. doi: 10.1088/1748-0221/11/05/P05010.
- [32] P. L. Rocca and F. Riggi. ‘The upgrade programme of the major experiments at the Large Hadron Collider’. In: *Journal of Physics: Conference Series* 515.1 (2014), p. 012012. doi: 10.1088/1742-6596/515/1/012012.
- [33] A. Augusto Alves Jr et al. (The LHCb collaboration). ‘The LHCb Detector at the LHC’. In: *Journal of Instrumentation* 3.08 (2008), S08005. doi: 10.1088/1748-0221/3/08/S08005.
- [34] C. Elsasser. ‘ $\bar{b}b$  production angle plots.  $\sqrt{s} = 14$  TeV’. URL: [http://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb\\_ProductionAngles.html](http://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_ProductionAngles.html) (visited on Aug. 26, 2016).
- [35] R. Lindner. ‘LHCb layout 2’. Feb. 2008. URL: <https://cds.cern.ch/record/1087860>.
- [36] C. Ilgner et al. ‘The Beam Conditions Monitor of the LHCb Experiment’. Tech. rep. Jan. 2010. arXiv: 1001.2487.
- [37] M. Kucharczyk, P. Morawski, and M. Witek. ‘Primary Vertex Reconstruction at LHCb’. Tech. rep. LHCb-PUB-2014-044. Geneva: CERN, Sept. 2014. URL: <https://cds.cern.ch/record/1756296>.
- [38] B. Sciascia. ‘LHCb Run 2 Trigger Performance’. In: *The 16th International Conference on B-Physics at Frontier Machines*. May 2016. URL: <https://cds.cern.ch/record/2208038>.
- [39] G. Barrand et al. ‘CHEP2000 GAUDI — A software architecture and framework for building HEP data processing applications’. In: *Computer Physics Communications* 140.1 (2001), pp. 45–55. doi: 10.1016/S0010-4655(01)00254-5.
- [40] I. Antcheva et al. ‘ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization’. In: *Computer Physics Communications* 180.12 (2009), pp. 2499–2512. doi: 10.1016/j.cpc.2009.08.005.

- [41] ‘The MOORE project’. 2007. URL: <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/> (visited on Aug. 28, 2016).
- [42] M. Clemencic et al. ‘The LHCb Simulation Application, Gauss: Design, Evolution and Experience’. In: *Journal of Physics: Conference Series* 331.3 (2011), p. 032023. DOI: 10.1088/1742-6596/331/3/032023.
- [43] I. Belyaev et al. ‘Handling of the generation of primary events in Gauss, the LHCb simulation framework’. In: *Journal of Physics: Conference Series* 331.3 (2011), p. 032047. DOI: 10.1088/1742-6596/331/3/032047.
- [44] T. Sjöstrand et al. ‘High-energy-physics event generation with Pythia 6.1’. In: *Computer Physics Communications* 135.2 (2001), pp. 238–259. DOI: [http://dx.doi.org/10.1016/S0010-4655\(00\)00236-8](http://dx.doi.org/10.1016/S0010-4655(00)00236-8).
- [45] T. Sjöstrand, S. Mrenna, and P. Skands. ‘PYTHIA 6.4 physics and manual’. In: *Journal of High Energy Physics* 2006.05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026.
- [46] D. J. Lange. ‘The EvtGen particle decay simulation package’. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1–2 (2001), pp. 152–155. DOI: [http://dx.doi.org/10.1016/S0168-9002\(01\)00089-4](http://dx.doi.org/10.1016/S0168-9002(01)00089-4).
- [47] S. Agostinelli et al. ‘Geant4—a simulation toolkit’. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. DOI: [http://dx.doi.org/10.1016/S0168-9002\(03\)01368-8](http://dx.doi.org/10.1016/S0168-9002(03)01368-8).
- [48] ‘The BRUNEL project’. 2001. URL: <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/> (visited on Aug. 28, 2016).
- [49] ‘The DAVINCI project’. 2005. URL: <http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/> (visited on Aug. 28, 2016).
- [50] S. Benson et al. ‘The LHCb Turbo Stream’. In: *J. Phys.: Conf. Ser.* 664.LHCb-PROC-2015-013. CERN-LHCb-PROC-2015-013. 8 (May 2015), 082004. 8 p. DOI: 10.1088/1742-6596/664/8/082004.
- [51] L. Starterkit. ‘Changes to the data flow in Run 2’. 2015. URL: <https://lhcb.github.io/first-analysis-steps/21-run-2-data-flow.html> (visited on Aug. 28, 2016).
- [52] J. Wishahi. ‘Measurement of CP Violation in  $B^0 \rightarrow J/\psi K_s^0$  Decays with the LHCb Experiment’. PhD thesis. 2013.
- [53] R. Aaij et al. ‘Opposite-side flavour tagging of B mesons at the LHCb experiment’. In: *The European Physical Journal C* 72.6 (2012), pp. 1–16. DOI: 10.1140/epjc/s10052-012-2022-1.
- [54] M. Calvi et al. ‘New algorithms to tag the flavour of  $B^0$  mesons using pions and protons’. Version 2.2. July 2016. URL: [https://twiki.cern.ch/twiki/pub/LHCbPhysics/SSpAndSSpiBDT/LHCb-ANA-2015-061-v2\\_2.pdf](https://twiki.cern.ch/twiki/pub/LHCbPhysics/SSpAndSSpiBDT/LHCb-ANA-2015-061-v2_2.pdf) (visited on Sept. 28, 2016).

- [55] R. Aaij et al. (The LHCb collaboration). ‘A new algorithm for identifying the flavour of  $B_s^0$  mesons at LHCb’. In: *Journal of Instrumentation* 11.05 (2016), P05010. DOI: 10.1088/1748-0221/11/05/P05010.
- [56] R. Aaij et al. (The LHCb collaboration). ‘Precision Measurement of CP Violation in  $B_s^0 \rightarrow J/\psi K^+ K^-$  Decays’. In: *Phys. Rev. Lett.* 114 (Jan. 2015), p. 041801. DOI: 10.1103/PhysRevLett.114.041801.
- [57] R. Aaij et al. (The LHCb collaboration). ‘Measurement of CP Violation in  $B^0 \rightarrow J/\psi K_S^0$  Decays’. In: *Phys. Rev. Lett.* 115 (July 2015), p. 031601. DOI: 10.1103/PhysRevLett.115.031601.
- [58] G. Aad et al. (The ATLAS collaboration). ‘Measurement of the CP-violating phase  $\phi_S$  and the  $B_s^0$  meson decay width difference with  $B_s^0 \rightarrow J/\psi \phi$  decays in ATLAS’. In: *Journal of High Energy Physics* 2016.8 (2016), pp. 1–45. DOI: 10.1007/JHEP08(2016)147.
- [59] B. Aubert et al. (The BaBar collaboration). ‘Measurement of time-dependent CP asymmetry in  $B^0 \rightarrow c\bar{c}K^{(*)0}$  decays’. In: *Phys. Rev. D* 79 (Apr. 2009), p. 072009. DOI: 10.1103/PhysRevD.79.072009.
- [60] H. Kakuno et al. (The Belle collaboration). ‘Neutral flavor tagging for the measurement of mixing-induced CP violation at Belle’. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 533.3 (2004), pp. 516–531. DOI: 10.1016/j.nima.2004.06.159.
- [61] R. Aaij et al. (The LHCb collaboration). ‘B flavour tagging using charm decays at the LHCb experiment’. In: *Journal of Instrumentation* 10.10 (2015), P10005. DOI: 10.1088/1748-0221/10/10/P10005.
- [62] ‘LoKi – Smart & Friendly C++ Physics Analysis Toolkit’. URL: <http://lhcb-comp.web.cern.ch/lhcb-comp/Analysis/Loki/>.
- [63] M. Calvi, O. Dormond, and M. Musy. ‘LHCb flavour tagging performance’. Tech. rep. LHCb-2003-115. Oct. 2003. URL: <http://cds.cern.ch/record/691595>.
- [64] K. Harrison et al. ‘GANGA: a user-Grid interface for Atlas and LHCb’. In: *cs.SE/0306085* (June 2003), 9 p. URL: <https://cds.cern.ch/record/622197>.
- [65] A. Hoecker et al. ‘TMVA: Toolkit for Multivariate Data Analysis’. In: *PoS ACAT (2007)*, p. 040. arXiv: physics/0703039.
- [66] T. Chen and C. Guestrin. ‘XGBoost: A Scalable Tree Boosting System’. In: *CoRR* (2016). arXiv: 1603.02754.
- [67] J. Wimberly. ‘EspressoPerformanceMonitor’. 2014. URL: <https://gitlab.cern.ch/lhcb-ft/EspressoPerformanceMonitor> (visited on Sept. 22, 2016).
- [68] F. Pérez and B. E. Granger. ‘IPython: a System for Interactive Scientific Computing’. In: *Computing in Science and Engineering* 9.3 (May 2007), pp. 21–29. DOI: 10.1109/MCSE.2007.53. URL: <http://ipython.org>.



# Danksagung

An dieser Stelle bedanke ich mich herzlich bei Prof. Dr. Bernhard Spaan für die Möglichkeit, die Masterarbeit in seiner Arbeitsgruppe der Experimentellen Physik 5a schreiben zu können. Außerdem möchte ich mich bei Herrn Prof. Dr. Kevin Kröninger für sein Interesse und die Zeit bedanken, diese Arbeit zu lesen und zu beurteilen.

Dem gesamten Lehrstuhl gilt mein Dank für die Unterstützung bei allen Belangen und für die durchweg positive Stimmung, die auch in stressigen Zeiten stets für eine gute Atmosphäre gesorgt hat. Britta danke ich für die Hilfe bei allen angenehmen und unangenehmen organisatorischen Angelegenheiten. Timon möchte ich für die gute und inspirierende Zusammenarbeit in den frühen Tagen der Masterarbeit danken. Bei allen Masterstudenten und vor allem bei Nico und Thomas möchte ich mich für die offenen Ohren, schönen Tafelmalereien und vielen getragenen Litern Milch bedanken.

Besonders möchte ich mich bei Julian für seine Betreuung, Unterstützung und den ehrlichen und offenen Umgang, sowie das Vertrauen in meine Arbeit bedanken. Außerdem bedanke ich mich bei der Flavour Tagging Gruppe, vor allem bei Alex und Vanessa, die sich viel Zeit für Diskussionen und Fragen genommen haben und bei Jack für das reichliche Feedback während meiner Arbeit.

Schließlich danke ich meiner Familie, meinen Freunden und meiner geliebten Freundin Kristina dafür, dass sie mir so oft geholfen haben, die richtigen Entscheidungen zu treffen und mich motiviert haben, vollen Einsatz zu zeigen, ohne die wichtigsten Dinge aus den Augen zu verlieren.



## Eidesstattliche Versicherung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Masterarbeit mit dem Titel „Optimization of Flavour Tagging Algorithms for the LHCb Experiment“ selbständig 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.

---

Ort, Datum

Unterschrift

## 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.

---

Ort, Datum

Unterschrift