

**Department of Physics and Astronomy**  
**University of Heidelberg**

Bachelor Thesis in Physics  
submitted by

**Maurice Morgenthaler**

born in Kehl (Germany)

**2019**



---

**First observation of the decay**  
 **$\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$  at the LHCb**  
**experiment**

This Bachelor Thesis has been carried out by Maurice Morgenthaler at the  
Physikalisches Institut in Heidelberg  
under the supervision of  
Dr. Sebastian Neubert



---

## Abstract

This bachelor thesis is dedicated to the first observation of the decay  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$  with the LHCb detector using all available data during Run 1 and Run 2 of the LHC. The decay is good reference channel for the proposed decay  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + \bar{p} + S^0(\text{udsuds})$  where  $S^0$  represents a hypothetical sexaquark which is a dark matter candidate. Using an cut-based selection resulted in a signal yield of  $152 \pm 25$  for Run 1 and  $447 \pm 46$  for Run 2. Alternatively a boosted decision tree (BDT) was trained which resulted in a Run 1 yield of  $96 \pm 13$  and Run 2 yield of  $290 \pm 20$  corresponding to a total observation significance of  $24.1 \sigma$ . The given errors are statistical. The BDT can be used to select signal events of this decay channel in the future.

## Zusammenfassung

Diese Bachelorarbeit ist der ersten Beobachtung des Zerfalls  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$  mit dem LHCb Detektor unter Benutzung aller verfügbarer Daten während Run 1 und Run 2 des LHC gewidmet. Der Zerfall ist ein guter Referenzkanal für den vorgeschlagenen Zerfall  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + \bar{p} + S^0(\text{udsuds})$  wobei  $S^0$  ein hypothetisches Sexaquark darstellt, welches ein Kandidat für Dunkle Materie ist. Die Benutzung einer schnittbasierten Selektion resultierte in  $152 \pm 25$  Signalereignisse für Run 1 und  $447 \pm 46$  für Run 2. Alternativ wurde ein Boosted Decision Tree (BDT) trainiert welcher in einer Run 1 Signalausbeute von  $96 \pm 13$  und Run 2 Signalausbeute von  $290 \pm 20$  resultierte, was einer kombinierten Observations-signifikanz von  $24.1 \sigma$  entspricht. Die angegebenen Fehler sind statistisch. Der BDT kann in der Zukunft dazu verwendet werden Signalereignisse dieses Zerfallskanals zu selektieren.



---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Basic Standard Model</b>	<b>4</b>
2.1	Particles in the SM . . . . .	4
2.1.1	Fermions . . . . .	5
2.1.2	Bosons . . . . .	5
2.2	Fundamental forces . . . . .	5
<b>3</b>	<b>Search for exotics</b>	<b>7</b>
3.1	Search for Pentaquarks . . . . .	7
3.2	Search for Hexaquarks . . . . .	8
<b>4</b>	<b>The LHCb detector</b>	<b>9</b>
4.1	Detector layout . . . . .	9
4.2	Trigger process . . . . .	12
<b>5</b>	<b>Observation of <math>\Xi_b^- \rightarrow J/\Psi + p^+ + K^- + K^-</math> with cut-based selection</b>	<b>14</b>
5.1	Data generation . . . . .	14
5.1.1	Stripping process . . . . .	14
5.1.2	Creation of nTuples . . . . .	17
5.1.3	Monte Carlo as reference data . . . . .	19
5.2	Application of offline-cuts . . . . .	19
5.3	Signal fit with MC data . . . . .	22
5.3.1	Testing different shapes . . . . .	23
5.3.2	Fixing parameters . . . . .	25
5.4	Determining $\Xi_b^-$ mass and signal yield . . . . .	25
<b>6</b>	<b>Observation of <math>\Xi_b^- \rightarrow J/\Psi + p^+ + K^- + K^-</math> with BDT selection</b>	<b>27</b>
6.1	Function of BDT's . . . . .	27
6.2	Deciding for training variables . . . . .	28
6.3	Choosing training samples . . . . .	29
6.4	Preparation of MC sample . . . . .	30
6.4.1	Correction of ProbNN with PIDCalib . . . . .	30
6.4.2	Kinematic reweighting . . . . .	31
6.4.3	Checking the distributions . . . . .	34

## CONTENTS

---

6.5	Training the BDT . . . . .	34
6.5.1	Study of the ROC and response curve . . . . .	35
6.5.2	Lessons from overtraining suppression . . . . .	38
6.6	Finding the best BDT response cut . . . . .	40
6.6.1	Preparing a pseudo significance . . . . .	41
6.6.2	Scanning the FoM . . . . .	42
6.7	Analysing the BDT created spectrum . . . . .	42
6.7.1	Changes to the fitting model . . . . .	44
6.7.2	Uncertainty of the signal yield . . . . .	45
6.7.3	Significance . . . . .	46
6.8	A look at systematic uncertainties . . . . .	47
<b>7</b>	<b>Conclusion and outlook</b>	<b>49</b>
	<b>Appendix</b>	<b>51</b>
	<b>References</b>	<b>54</b>
	<b>Acknowledgments</b>	<b>58</b>

## 1 Introduction

The basics of our current understanding of particle physics is summed up in the Standard Model (SM). Excluding gravity, it describes three of four known fundamental forces in our universe. Those so called strong, electromagnetic and weak interactions are mediated by Gauge bosons which are an example of elementary particles. As the most basic components of our universe the categories of elementary particles are also described by the Standard Model. Chapter 2 provides a further description of the Standard Model.

The Standard Model is not complete. The largest particle physics laboratory in the world is the *European Organization for Nuclear Research* (CERN). It is home to the LHC, the largest particle collider in the world, as well as seven detector experiments like the LHCb. The data used in this Bachelor thesis originates from this detector described in chapter 4.

The data from the LHCb mostly contains classical particles like baryons consisting of three quarks although theoretically it is also possible for them to consist of five or even six quarks. The first two pentaquarks  $P_c^+(4380)$  and  $P_c^+(4450)$ [1] were discovered in 2015 by the LHCb collaboration. In march 2019 the LHCb collaboration announced having observed a third pentaquark  $P_c^+(4312)$  as well as resolving the 4450 MeV peak to actually be two states namely  $P_c^+(4440)$  and  $P_c^+(4457)$  [2]. Chapter 3 describes the historical search for exotics further. Even though many hexaquarks were proposed none were found yet. This is also true for the H-Dibaryon proposed by Robert L. Jaffe in 1976 [3], nowadays called *sexaquark*. If it were to be found, it would be an dark matter candidate due to it's theorized characteristics. Consisting out of the quark structure *uuddss* it is a neutral singlet with presumably high binding energy. According to different hypotheses it's mass lies in a range of 1.0 to 2.3 GeV [4]. With a mass under  $m < 2.054$  GeV it would be only able to decay weakly resulting in a lifetime longer than the age of the universe[5]. For an even lower mass of  $m < 1.876$  GeV the weak channel would also be forbidden and the particle stable. A possible decay channel to observe this exotic is  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + \bar{p} + S^0$ .

This bachelor thesis records the observation of a similar decay channel  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ . It's Feynman diagram can be seen in figure 1.1. Previous work carried out by the author already used *RapidSim*[6]

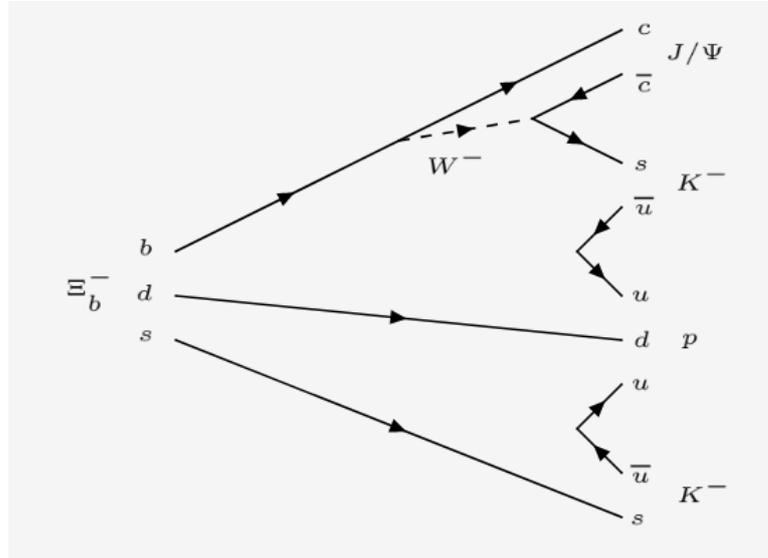


Figure 1.1: Feynman diagram of the observed decay channel

simulated data of this decay channel to test a potential model to calculate the mass of the dark matter candidate [7]. This decay channel as reference has the advantage that the two kaons of the decay can be used as a sexaquark proxy. This means that the mass models are applied to the data assuming that the proxy mass is unknown. Comparing the results with the real kaon mass gives information about the quality of the model. While the previous work used simulations a first cut-based preselection was used to search for hints of a signal peak. While an initial peak was observed, at this point no formal observation could be made. This was the case due to too tight cuts during nTuple creation as well as no available full detector simulated MC data to optimize a cut based selection on. The first part of the thesis in chapter 5 therefore describes a full cut-based selection. 5.1 talks about the complete process from the stripping to the finished new nTuples with looser cuts, as well as about fully detector generated MC data created for this analysis. In subsection 5.2 the offline selection is explained while subsection 5.3 describes the creation of a fit model with help of the MC data. Finally 5.4 is about the results of fitting the data and the signal yield coming from this method. The resulting fit gave an initial idea about the signal yield but could be still optimized in terms of background suppression.

In section 6 therefore instead of using the cut-based selection a boosted decision tree (BDT's explained in 6.1) was trained. It's training variables

and samples are described in subsection 6.2 and 6.3 respectively. For the samples the MC was used which had to be corrected beforehand which is shown in subsection 6.4. In subsection 6.5 and 6.6 the final BDT was optimized. In the final step in subsection 6.7 the BDT created dataset was scanned and fit for different signal yields to calculate the observation significance as well as an uncertainty for the signal yield. Subsection 6.8 then discusses some systematic uncertainties while in section 7 a final conclusion is given.

## 2 Basic Standard Model

In 1935 *Yukawa Hideki* published his work “On the Interaction of Elementary Particles” postulating a force field between elementary particles accompanied by a further particle [8]. This resulted in a great search for the so called “Yukawa-Particle” but a systematic search for new particles was only possible with the help of accelerators after 1953 [9]. The latest confirmed particle is the “Higgs-Boson” in 2013. The elementary particles as we currently know them are shown in figure 2.1.

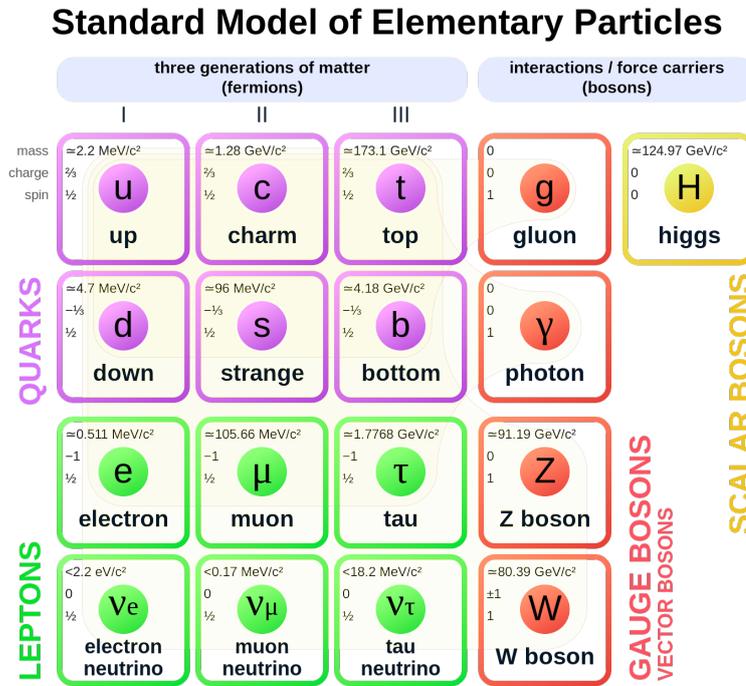


Figure 2.1: Overview of elementary particles in the Standard Model[10]

### 2.1 Particles in the SM

The particles which can be found in the standard model can be split into a multitude of different subcategories. One way the classify them is via their spin, namely into fermions and bosons.

### 2.1.1 Fermions

Fermions are elementary particles with a half-integer spin which follow Fermi-Dirac-Statistics. The Standard Model organizes them in three *generations*. Every generation consists of two quarks, a lepton and its corresponding neutrino which is also a lepton. One of the two quarks of each generation has an electrical charge of  $\frac{2}{3}$  while the other one has a charge of  $-\frac{1}{3}$ . Particles consisting of quarks are called *hadrons*. Depending on their quark content they can be further divided into *mesons* consisting of quark antiquark pair ( $q\bar{q}$ ) and baryons consisting out of three quarks ( $qqq$ ). For every quark and lepton exists an antiquark and antilepton with the same mass and inverted charge-like quantities.

### 2.1.2 Bosons

Bosons are elementary particles with integer spin following Einstein-Bose-Statistics. There are two different kinds of bosons in the Standard Model. At one hand are four Gauge bosons. They mediate the three forces described in the model. At the other hand is the Higgs boson which is a scalar boson. It is produced in the Higgs field which causes the mass of all elementary particles which aren't massless.

## 2.2 Fundamental forces

Except for gravitation the Standard Model describes all fundamental forces. The three forces which are also called interactions can be for example characterized through their range, strength or the Gauge bosons which mediate the forces. They all can be described by a field theory.

**Strong force:** As the name implies the strong interaction is about 60 times stronger than the electromagnetic and over  $10^4$  times stronger than the weak force. It is described by the field theory of *Quantum Chromo Dynamics* (QCD). The responsible Gauge boson is called gluon and interacts via three *colours* which is comparable to charges. It's a property of gluons, quarks and antiquarks hence only hadrons and quarks themselves experience this force. The QCD also states that there are no free quarks, meaning that they only exist in *confinement*, held together by the strong interaction. Hadrons need to be colour neutral, which means that a colour has to be paired with its

## 2. BASIC STANDARD MODEL

---

anticolour or all three possible colours need to be present. At a size scale of protons and neutrons the strong interaction dominating the electromagnetic interaction also explains the existence of nucleons as we know them.

**Electromagnetic force:** The electromagnetic force is the only one of these three forces which we experience in our daily live. It is described by *quantum electrodynamics*. It has an infinite range in contrast to the strong and the weak force. Photons are the Gauge bosons of this interaction and it's coupled to regular electric charges. It is the cause of many phenomenons in physics like the electron clouds around a nucleon as well as other sciences like Van der Waals forces in chemistry.

**Weak force:** This force is described by the unified theory between the weak and electromagnetic interaction *electroweak theory* (EWF). Mediated by  $W^\pm$  and  $Z^0$  bosons it is found on sub-atomic distances. Coupling on the six flavours of quarks and leptons it alone can change quark flavour and by that is the relevant mechanism for radioactive decays like  $\beta^-$  decay.

### 3 Search for exotics

In chapter 2 it was described that there are two types of hadrons. There are mesons with a quark content of  $(q\bar{q})$  and baryons with the structure  $(qqq)$ . This is only the basic configuration, following a paper of 1964 by M. Gell-Mann it would be also possible for particles with a structure of  $(qq\bar{q}\bar{q})$  and  $(qqqq\bar{q})$  to exist [11]. Collectively called exotic hadrons they are worldwide searched for.

#### 3.1 Search for Pentaquarks

In 1997 D. Diankonov et al. predicted a pentaquark with a mass of about 1530 MeV [12]. Just a few years later in the early to mid 2000s several groups claimed to have observed such a particle. Most notably would be the claim of T.Nakano *et al.* who reported the discovery of pentaquark  $\theta^+$  with a mass of 1.54 GeV/c<sup>2</sup> [13]. Those claims couldn't be supported by later experiments, though, which used order of magnitudes bigger statistics. The *Particle Data Group* has several reports on this pentaquark search with numerous papers which reject the observation of such a pentaquark[14].

The first discovery of two pentaquarks was announced in July 2015 by the LHCb collaboration [1]. Data of the decay  $\Lambda_b^0 \rightarrow J/\Psi + p^+ + K^-$  presented a resonance like peak in the mass spectrum of  $J/\Psi + p^+$ . A full amplitude analysis was conducted to see if it was possible to reproduce the data based on classic hadrons. Fitting with 14 different  $\Lambda^*$  states of a dominant decay channel  $\Lambda_b^0 \rightarrow J/\Psi + \Lambda^*(\rightarrow p^+ + K^-)$  wasn't able to fully describe the data. A resonance like  $\Lambda_b^0 \rightarrow P_c^+(\rightarrow J/\Psi + p^+) + K^-$  would decay strongly and would include a pentaquark  $P_c^+$  with a minimum quark content of  $(c\bar{c}uud)$ . As a matter of fact it was possible to fully recreate the data shape by including a  $P_c^+(4380)$  and  $P_c^+(4450)$  into the analysis and by that claim their discovery with a significance of 9  $\sigma$  respectively. On the 26<sup>th</sup> of March 2019 the LHCb collaboration announced the observation of a new pentaquark state  $P_c^+(4312)$  with 7.3  $\sigma$  significance. Being able to analyse over nine times the amount of data as in 2015 they found the  $P_c^+(4450)$  peak to be a two peak structure consisting of two pentaquarks  $P_c^+(4440)$  and  $P_c^+(4457)$  with a significance of 5.4  $\sigma$  [2].

### 3. SEARCH FOR EXOTICS

---

#### 3.2 Search for Hexaquarks

In comparison to pentaquarks there are no discoveries of hexaquarks yet. One especially interesting candidate was proposed by Robert Jaffe in 1976 [3]. He theorized a particle with a quark structure of (udsuds) with a mass of 2150MeV and called it H(Hyperon)-Dibaryon due to its content. It should be a flavour-singlet, have a spin of zero, an even parity as well as a Baryon number of  $B=2$  and a strangeness of  $S=-2$ . Furthermore its complete spatial wave function is completely symmetric with all individual wavefunctions being asymmetric [15]. Calculations of the lifetime of this hexaquark H were done for masses corresponding to likely decay channels. For the case of  $m_N + m_\Lambda < m_H < 2m_\Lambda$  where  $\Lambda$  refers to the  $\Lambda$ -baryon (uds) and N to a nucleon the hexaquark would decay weakly and have a lifetime higher than the age of the universe. For the case  $m_H < m_N + m_\Lambda$  the weak decay would also be forbidden resulting in a stable particle[15]. Those properties make the H-Dibaryon an excellent Dark Matter candidate, in case it is available in an abundance high enough to equal those calculated for dark matter. From a practical point of view its not a trivial task to detect such an exotic. Characteristics like a radius smaller than a quarter of a Neutron's and being neutral and by that not directly measurable by the LHCb detector are valid reasons why it wasn't found yet [15]. To prepare reference data for a search in the decay channel  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + \bar{p} + S^0$  this thesis examined a similar classic decay  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ .

## 4 The LHCb detector

The *Large Hadron Collider* (LHC) accelerator is the proton-proton collider of the *European Organization for Nuclear Research* (CERN) with its headquarter in Meyrin in the Swiss canton of Geneva. The LHC had a collision energy of 7 TeV in 2011, 8 TeV in 2012 and 13 TeV since 2015 making it the most powerful particle collider in the world. The LHCb detector is the detector of the LHCb experiment in CERN with its segment of the LHC located in France. The b in the name stand for the bottom-quark as it is in general used for experiments containing bottom and charm quarks. The main task of the LHCb experiment is the investigation of CP violation and by that find an explanation for the amount of asymmetry between matter and anti-matter in our known universe. It also can be used for research into different aspects. This work on a reference channel for the sexaquark search is one example of it.

In the following will be a brief description of the LHCb detector. At first the general layout of the detector will be discussed. In the process the different elements are explained in the order in which they appear in the detector going downstream, which is the the direction from the collision point to the back of the detector. A schematic of the detector can be found in figure 4.1. Afterwards there will be a short description of the trigger system used to filter events.

### 4.1 Detector layout

The detector is specialized on the measurement of decays involving hadrons consisting of heavy b- and c-quarks. It has high precision in particle identification as well as a high resolution in characteristics like the decay length. Furthermore at high energies both particles of a  $b\bar{b}$  pair tend to move in the same direction cone resulting from high production correlation. This property caused the LHCb detector to be build as a single-arm forward detector unlike the other detectors at LHC [16]. A default right-handed coordinate system for the detector was defined. Parallel to the particle beam is the z-axis while the y-axis is vertical and the x-axis horizontal respectively. Furthermore for better access most parts of the detector are divided vertically into an A- and C-side to be movable in the x-direction.

## 4. THE LHCb DETECTOR

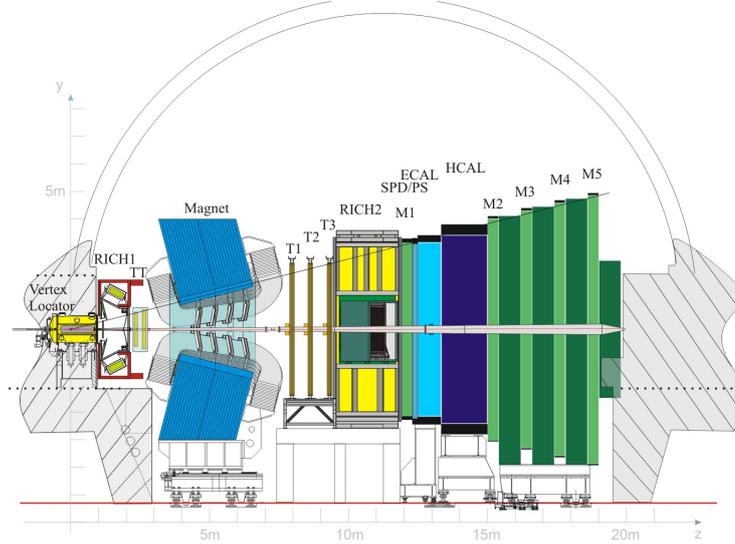


Figure 4.1: Cross-Section through the LHCb detector [17]

**Vertex Locator (VELO):** The VELO of the LHCb detector together with further trackers described below are used as the tracking system [16]. The VELO itself consists of several semi-circular silicon modules placed along the particle beam. Coming from a polar coordinate perspective they measure the R- and  $\Phi$ -Coordinate of incoming particles and by that measure the coordinates of primary vertex. For the best possible data resolution the modules are only 7mm away from the particle beam. To safely inject the p-p beam the modules are automatically retracted during this time [18].

**Ring Imaging Cherenkov (RICH) detectors:** Two RICH detectors (RICH1, RICH2) are used in the LHCb detector. They're used to identify 9charged hadrons namely pions, kaons and protons [19]. The effect used to distinguish between the potential hadrons is Cherenkov radiation. Charged particles which travel through a medium with refractive index  $n$  with a speed  $v$  faster than light in the medium,  $c_{\text{med}} = c/n$ , emit photons in a specific angle  $\theta$ . The relation is as followed:

$$\cos(\theta) = \frac{c}{nv} = \frac{c_{\text{med}}}{v} \quad (4.1)$$

To maximize the window of measurement, the RICH1 detector covers a low energy while RICH2 covers high energy window.

**Tracking Stations:** In addition to the VELO the LHCb has a Trigger Tracker (TT) behind RICH1 and three tracking stations T1-T3 in front of RICH2. The tracking stations have inner trackers (IT) and outer trackers (OT). They can precisely detect hits of charged particles to reproduce particle tracks together with vertex coordinate data of the VELO system. While the TT and IT use regular silicon modules the OT uses drift-tube gas modules [16]

**Magnet:** The magnet used in the LHCb detector is a warm dipole magnet. At an integrated field of 4 Tm it is constructed to produce a homogeneous magnetic field  $\vec{B}$  [20]. Due to the Lorentz force,  $\vec{F}_{\text{Lorentz}}$

$$\vec{F}_{\text{Lorentz}} = q \cdot \vec{v} \times \vec{B} \quad (4.2)$$

with  $q$  being the the charge and  $\vec{v}$  the velocity of the particle, it becomes apparent that oppositely charged particles will be deflected in opposing directions and that it's possible to calculate the momentum  $p$  of decay products with the flight trajectory curvature  $r$ .

$$p = qrB \quad (4.3)$$

**Calorimeters:** The calorimeter system of the LHCb detector consists of four parts stacked behind each other. First particles hit the Scintillating Pad Detector (SPD) which checks if the particle is charged or not. The following Pre-Shower Detector (PS) is there to identify if in case of a charged hit it is an electron or for the neutral case a photon. Finally there is an electromagnetic calorimeter (ECAL) as well as a hadronic calorimeter (HCAL) which are used to measure the energy of photons and electrons or hadrons respectively [21]. The calorimeters use the same principle to measure this quantity. Layers of plastic cause scintillation UV light which is captured by photo multipliers. The amount of light is proportional to the energy which allows it's quantification.

**Muon System:** The last subdetector is the Muon system which is especially important as bottom decays tend to create muons. Composed out of five rectangular stations M1 - M5 it is used to identify muons and calculate their momentum as well as transverse momentum  $p_T$ . The system is at the

## 4. THE LHCb DETECTOR

---

back of the detector and iron layers are placed between the stations. This helps with identification as only muons are able to reach the end of the detector and penetrate the layers [22]. The system information is used for the L0 trigger described below. The stations themselves are built from several gas chambers which give binary information if a muon crossed them. This allows to trace their track [23] [24].

### 4.2 Trigger process

While the previous section described how the LHCb detector is able to detect events, it would be impossible to use every occurring event for bandwidth reasons. The LHCb detector has a two level trigger system to filter the events.

**Hardware Trigger (L0):** The first trigger level of the LHCb detector reduces the rate of measured interactions of 40MHz to 1MHz. This is done by using information from the calorimeters for a hadronic L0 and the muon system for a muon L0 trigger [25]. The hadronic trigger uses information of the calorimeter 2x2 clusters about the deposited transverse energy  $E_{\text{trans}}$

$$E_{\text{trans}} = \sum_{i=1}^4 E_i \sin \theta_i \quad (4.4)$$

with  $\sin \theta$  corresponding to the angle between the line connecting proton-proton interaction and cell center and the z-axis. Additional information of the SPD and PS is used to identify the particle and if it has a transversal energy above a certain threshold it is accepted. For the muon L0 the muon system measures the transverse momentum  $p_T$ . Events are accepted in either one of two different ways. The first criteria to be accepted is to identify the muon with the highest  $p_T$  value, if it's above the *high  $p_T$  muon* threshold it is accepted. The second criteria is for the product of  $p_T$  of the muons with the highest and second highest value to exceed the corresponding threshold known as *Dimuon* [26]. A list of the threshold can be found in table 4.1.

**High Level Trigger (HLT):** The HLT also called software trigger is used after the hardware trigger for further processing. It's composed of two stages HLT1 and HLT2. The first stage HLT1 uses partial event reconstruction to

L0 trigger		E <sub>T</sub> or p <sub>T</sub> threshold				
		2011	2012	2015	2016	2017
Hadron	[GeV]	3.50	3.70	3.60	3.70	3.46
Photon	[GeV]	2.50	3.00	2.70	2.78	2.47
Electron	[GeV]	2.50	3.00	2.70	2.40	2.11
Muon	[GeV]	1.48	1.76	2.80	1.80	1.35
high p <sub>T</sub> Muon	[GeV]	—	—	6.00	6.00	6.00
Dimuon	[GeV <sup>2</sup> ]	1.296	1.60	1.69	2.25	1.69

*Table 4.1: Summary of the different L0 thresholds [27] [28]*

further filter incoming events, reducing the data frequency to 110kHz and to store the selected ones in a 10 PB buffer. This is enough storage to collect the data of 2 weeks before the second stage has to be activated. The buffer data is used to calibrate and align the detector in real time to further improve the reconstruction. Due to its large capacity the buffer can be used in regular runs or to delay to input for HLT2 in case the alignment experiences problems. Data given to the HLT2 is processed to fully reconstruct the events and further selections are applied. The data output is saved in a storage with a frequency of about 12.5 kHz [26].

## 5 Observation of $\Xi_b^- \rightarrow J/\Psi + p^+ + K^- + K^-$ with cut-based selection

The first method which was used to search for the decay channel in question was the application of variable cuts to LHCb data. Before it is possible to do offline work with data it has to pass a process named stripping. In the process after passing the trigger and reconstruction candidates for the desired channel  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$  have to be selected centrally. This is controlled by the application *DaVinci* and is based on loose predefined cuts which are defined in so called stripping lines [29]. Further initial cuts are applied when the LHCb data is used to construct nTuples to be used offline. To finally find a peak in the dataset one had to apply offline cuts to reduce background.

### 5.1 Data generation

The way the data was produced and refined is depicted. All cuts starting with the initial stripping and ending with the offline cuts are discussed.

#### 5.1.1 Stripping process

Standard selections are applied by the stripping line. The stripping line *FullDSTDiMuonJpsi2MuMuDetachedLine* was used which corresponds to the decay products of the decay channel. It's input is coming from the line *StdLooseJpsi2MuMu* which in turn gets it's input from *StdAllLooseMuons*. Furthermore *StdLooseProtons* and *StdLooseKaons* are needed. For the analysis samples from all available years are used. Stripping line versions change with every year. An overview can be found on the *Stripping Project* homepage [30]. A flowchart for the stripping process can be found in figure 5.2. An overview over the cuts which are applied by this stripping step is shown in table 5.1. The meaning and usage is explained in the following for some especially important ones. A complete list can be found in the appendix. For a better understanding the equations used in the overview have their actual name in the stripping added in brackets. If a variable doesn't have a name in brackets it's already the actual name.  $\chi^2$  in general describes the result of a  $\chi^2$ -test done on the corresponding variable.

#	stripping step	applied to	cut
0	StdAllLoseMuons	$\mu$	IsMuon = True
1		$\mu$	$\chi_{\text{TR}}^2 < 5$
2		$\mu^+ + \mu^-$	$M_x - M_{\text{J}/\Psi, \text{PDG}} < 100$ [MeV]
3	StdLooseJpsi2MuMu	$\mu^+ + \mu^-$	$\chi_{\text{DOCA}}^2 < 30$
4		J/ $\Psi$	$\chi_{\text{DV}}^2 < 25$
5	FullDSTDiMuonJpsi- 2MuMuDetachedLine	$\mu$	min PID( $\mu$ ) > 0
6		$\mu$	min $p_{\text{T}} > 500$ [MeV]
7		J/ $\Psi$	BPVDLS > 3    BPVDLS < -3
8		J/ $\Psi$	$2996.916 < M < 3196.916$ [MeV]
9		J/ $\Psi$	$\chi_{\text{DV,DOF}}^2 < 20$
10	only Run 2 (v34)	$\mu$	$p < -8000$ [MeV]
11		$\mu$	$\chi_{\text{TR,DOF}}^2 < 5$
12		J/ $\Psi$	$p_{\text{T}} > -1000$ [MeV]
13	StdLooseKaons	K	$p_{\text{T}} > 250$ [MeV]
14		K	min $\chi_{\text{IP}}^2(\text{PV}) > 4$
15		K	$\Delta \log(\mathcal{L})(\text{K} - \pi) > -5$
16	StdLooseProtons	p	$p_{\text{T}} > 250$ [MeV]
17		p	min $\chi_{\text{IP}}^2(\text{PV}) > 4$
18		p	$\Delta \log(\mathcal{L})(\text{p} - \pi) > -5$

Table 5.1: Cuts done by the stripping. For the names of the cuts as given in the stripping line check the text. For cut 2 and 3 instead of the mother particle the decay products are listed to indicate that the cut is a combination cut and therefore applied before the vertex fit. Cuts 10 through 12 are parts of FullDSTDiMuonJpsi2MuMuDetachedLine

$\chi_{\text{TR}}^2$  (**MaxChi2Cut**): Cuts away all tracks above a certain  $\chi^2$ -value. This ensures that no unreasonable tracks are chosen as candidates.

$M_x - M_{\text{J}/\Psi, \text{PDG}}$  (**ADAMASS(J/psi(1S))**): This variable is the absolute difference between the measured mass of a particle “x” and its value documented by the *Particle Data Group*. Defining an upper boundary on this difference is an easy way to discard events with unreasonable masses.

$\chi_{\text{DOCA}}^2$  (**ADOCACHI2CUT**): A cut on the maximum  $\chi^2$  for the “Distance Of Closest Approach”. This selects muon pairs which were reasonably close at one point in time to be the decay product of the same J/ $\Psi$ . Pairs of particles which are too far apart are discarded.

## 5. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH CUT-BASED SELECTION

---

$\chi_{DV}^2$  (**VFASPF(VCHI2)**): Limits the maximum  $\chi^2$  of vertex fits performed with the characteristics of the accepted particle pairs. This ensures the quality of the reconstructed decay vertex positions of the mother particle.

**min PID( $\mu$ )** (**MINTREE('mu+'==ABSID,PIDmu)**): A loose cut on the particle identification variable of a muon. Effectively it can be written as the difference between the logarithmic likelihood that the particle is a muon and the logarithmic likelihood that it's a pion  $\log(\mathcal{L}_\mu/\mathcal{L}_\pi)$ .  $\mathcal{L}$  labels the likelihood function. It discards every muon candidate where it's more probable that it's a pion instead of a muon.

**BPVDLS**: Restricts candidates to have a minimum “Decay Length Significance” with respect to the primary vertex. The decay length significance is the decay length over it's error. Therefore this checks the quality of the reconstruction by discarding events with uncertain decay length.

$\chi_{DV,DOF}^2$  (**VFASPF(VCHI2PDOF)**): Similar cut to  $\chi_{DV}^2$ . It also evaluates the  $\chi^2$  of the decay vertex but also divides it by the degrees of freedom. By that it is more forgiving if the decay vertex has many degrees of freedom otherwise it's a strict cut.

$\chi_{TR,DOF}^2$  (**MAXTREE('mu+'==ABSID,TRCHI2DOF)**): Similar to the  $\chi^2$  track cut in *StdAllLoseMuons*. It checks again the reconstructed tracks of the muons, this time with respect to the degrees of freedom.

**min  $\chi_{IP}^2(PV)$**  (**MIPCHI2DV(PRIMARY)**): This variable corresponds to the minimum  $\chi^2$  of the particles *Impact Parameter* to a given vertice. The impact parameter is the perpendicular and by that shortest distance of a particles track to a decay vertex. In this case the primary vertex is specified. This ensures the reasonable assignments to a primary vertex.

$\Delta \log(\mathcal{L})(K - \pi)$  (**CombDLL(x - pi)**): This variable stands for the difference in the *Delta Log Likelihood* of a particle “x” and a pion. It is the equivalent to “min PID( $\mu$ )” for kaons and protons. Like before the combined information about the candidates are used to discard those which are more likely to actually be pions.

### 5.1.2 Creation of nTuples

At the end of the stripping process we obtain three different selections,  $J/\Psi$  decaying into two muons, loose kaons and loose protons. Those selections are needed to build the whole decay tree  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ , which is also done by the LHCb’s analysis program *DaVinci*. Taking the particle selections and the desired decay channel as input it builds nTuples, data tuples of the CERN ROOT software [31]. *DaVinci* is able to apply further cuts to get a more pure output. *DaVinci* cuts used in this analysis are loose and remove obvious outliers. They are loose as nTuples of this decay channel created for earlier work had too few statistics after applied off-line cuts [7]. As data for Run 1 and Run 2 just slightly differ in certain variables due to upgrades and increased collision energy it was safe to create the nTuples for both with the same *DaVinci* script.

#	applied to	cut
0	$J/\Psi$	$3040 > M > 3150$ [MeV]
1	$J/\Psi + p^+ + K^- + K^-$	$p_T > 5000$ [MeV]
2	$J/\Psi + p^+ + K^- + K^-$	$5000 > M > 7000$ [MeV]
3	$\Xi_b^-$	$\chi_{\text{DV,DOF}}^2 < 10$
4	$\Xi_b^-$	$\chi_{\text{IP,BPV}}^2 < 18$
5	$\Xi_b^-$	$\text{BPVDIRA} > 0.999$

Table 5.2: Cuts applied by *DaVinci* during nTuple creation. A sum of decay products indicates a combination cut done before the vertex fit

Table 5.2 shows the cuts as applied by the *DaVinci* software. Following is a description of new used variables.

$\chi_{\text{IP,BPV}}^2$  (**BPVIPCHI2**): Similar cut to “ $\min \chi_{\text{IP}}^2(\text{PV})$ ”. In this cut the  $\chi^2$  of the impact parameter to the *best* primary vertex.

**BPVDIRA**: A particle’s spatial trajectory is sometimes not exactly equal to the combined momenta vector of the daughter particles. The *DIRection Angle* describes the cosine between those vectors. A visualisation can be found in figure 5.1

## 5. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH CUT-BASED SELECTION

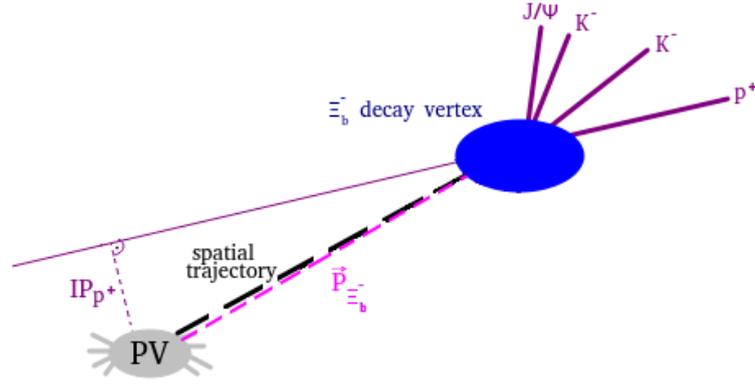


Figure 5.1: A visual demonstration of the impact parameter,  $IP$  and the direction angle cosine,  $DIRA$  on the decay  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ . The trajectory of the proton (thick line) is projected backwards (thin line) to better show from which point the perpendicular distance is calculated. The dashed black and pink lines represent the  $\Xi_b^-$ 's spatial trajectory and momentum vector respectively. The cosine of their angle is the  $DIRA$ .

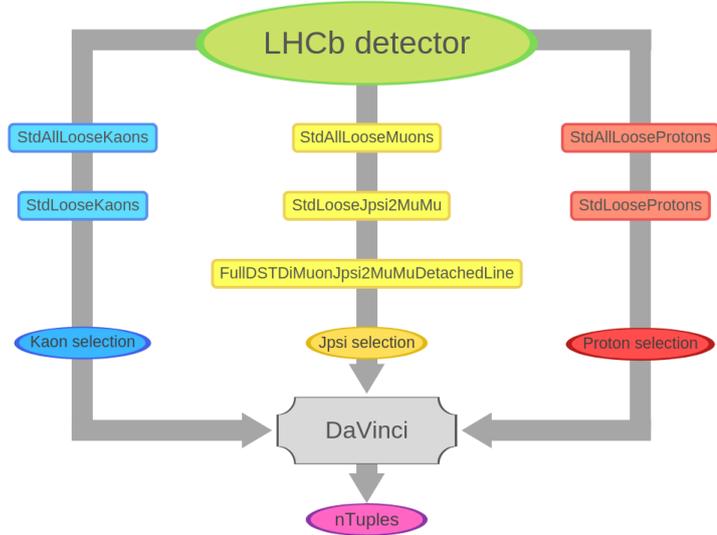


Figure 5.2: A flowchart showing the passed stages to generate the selections for  $DaVinci$ , which built the  $nTuples$  for this work

### 5.1.3 Monte Carlo as reference data

Using real data as the only dataset would bring several difficulties to the analysis as a reference sample is useful in multiple steps like fitting a model or checking the efficiencies of different cuts. Therefore *Monte Carlo* data was used as such a reference sample. In this case the simulated data is fully detector simulated and sent through the complete process of stripping and nTuple creation. Like that it is possible to circumnavigate the problem of detector specific artifacts in the real data. Such intrinsic behaviour like finite resolution is known for the simulation data and can be taken into account. As Monte Carlo data also includes background of different sources it is useful to cut on the data set. The variables *Xib BKGCAT* and *Jpsi BKGCAT* show into which category the candidates fall, for example partially reconstructed background [32]. Cutting the background candidates away resulted in a pure signal sample which could be used as reference for the desired signal shape.

## 5.2 Application of offline-cuts

After the stripping process the resulting nTuples are ready to use. Relevant for the observation of the desired decay is a clear peak in the mass spectrum of the mother particle. In this case this is the  $\Xi_b^-$  with a mass of  $M_{\Xi_b^-} = 5794.5 \pm 1.4 \text{ MeV}$  [33]. As at first no peak was to be seen a multitude of offline cuts had to be applied as shown in table 5.3. New cuts are again explained. Furthermore in this work, when mass is mentioned actually what is meant is the mass value coming from the *Decay Tree Fitter* (DTF).

**Context to the DTF:** The DTF is a tool used to create more accurate decay trees than it would be possible with the classical approach. The standard procedure fits a decay tree starting from the final decay products and working it's way back to the *primary vertex*, the reconstructed location of the initial particle collision. This mimics the way the detector reconstructs a decay. The disadvantage of this method is that it's not possible to propagate information of a vertex to it's descendants. The DTF in comparison looks at a decay chain as a whole. Applying further restrictions to vertices like mass constraints and fitting it using a Kalman filter allows for more accurate distributions. For this reason the mass output of the the DTF was chosen over the standard fit mass to work with. The restrictions which were

## 5. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH CUT-BASED SELECTION

---

applied for this decay channel were that the mother particle  $\Xi_b^-$  originates from the primary vertex as well as a mass constraint on the  $J/\Psi$ .

#	applied to	cut (Run 1, Run 2)
0	$K^-$	ProbNNk > 0.1
1	$K^-$	$\chi_{IP}^2$ (PV) > 6, > 10
2	$\mu^+$	ProbNNmu > 0.1
3	$\mu^-$	ProbNNmu > 0.1
4	$p^+$	ProbNNp effi > 0.1
5	decay products	$\sum_{x_i} p_T(x_i) > 4000, > 5000$
6	$J/\Psi$	$3050 < M < 3150$ [MeV]
7	$J/\Psi$	$\chi_{DY}^2 > 150, > 200$
8	$J/\Psi$	$\chi_{FD}^2$ (OV) < 5
9	$\Xi_b^-$	$\chi_{FD}^2$ (DV) < 30
10	$\Xi_b^-$	$\chi_{FD}^2$ (PV) > 250, > 300
11	$\Xi_b^-$	$\chi_{IP}^2$ (PV) < 11, < 12
12	$\Xi_b^-$	DIRA > 0.9999
13	$\Xi_b^-$	$\eta < 4.9$

Table 5.3: Overview of cuts applied offline on the produced  $nTuples$ . Differences in cuts between Run 1 and Run 2 data are marked by different colors. Red signals the cut value on Run 1 and green on Run 2 data

$\chi_{FD}^2$  (PV) (FDCHI2 OWNPV): As for all  $\chi^2$  cuts a cut on the quality of the variable. Here it's the flight distance to the primary vertex of the decay.

$\chi_{FD}^2$  (OV) (FDCHI2 ORIVX): Similar to the other cuts on  $\chi_{FD}^2$  in regard of quality control of a flight distance. The difference here is that  $J/\Psi$  as only intermediate particle has a decay vertex, which was cut in the stripping as well as a vertex of origin. Therefore this variable can be cut to control the flight distance to all three vertices.

$\chi_{DV}^2$  (ENDVERTEX CHI2):  $\chi^2$ -test of the reconstructed decay vertex location.

**ProbNNx:** ProbNNx variables are based on information from the several particle identification detectors. Evaluating information from them they are

a probability that a candidate is a particle of the type  $x$ . They are strong variables for the training of classifiers. They're hard to simulate due to their dependency on many inputs such as occupancy. The term "ProbNNp effi" shown in the offline cut list refers to the combined ProbNN term ProbNNp(1-ProbNNk).

Looking at table 5.3 the cuts can be roughly split into four categories. First of all *Particle Identification* (PID) are applied to the decay products of the decay channel. As already mentioned they combine several sources of information to calculate a probability of a candidate to be a certain particle. Therefore they are used to filter events with misidentified particles. In the case of the proton the cut was expanded to be more sensitive for kaons which were falsely categorized as protons. Second, the  $\chi^2$  of the flight distance FD in comparison to the primary vertex or the decay vertex were cut. Thus indirectly the lifetime was used to reject background.

Furthermore the impact parameter was once more used to check if the kaons are from the primary vertex. It was also used on the mother particle.

Finally some general cuts were applied. The DIRA variable, cosine of the direction angle, and the pseudorapidity  $\eta$  are cut in a way which stems from the LHCb geometry. It is known that the geometrical acceptance of the detector is  $\eta \in [2, 5]$  therefore an upper cut was made. Also the  $J/\Psi$  mass was further restricted and a minimum on the combined transverse momentum of the decay products was introduced.

To avoid bias, the cut values weren't chosen based on the investigated data. Instead the MC data and an upper sideband sample were used. Applying the cuts on the MC signal events and the background events gave a good idea how they would perform on real data. The percentage of rejected events sorted by runs can be found in table 5.4. As one can see while the cuts reject most of the background one also loses about 40% of the signal events. Keeping this in mind the resulting peaks over background are shown in figure 5.3. Applying a cut on the invariant mass of  $\Xi_b^-$  ( $5750 < M_{\Xi_b^-} < 5850$  [MeV]) shows the peak of interest clearly. Furthermore it leaves enough sideband to later fit the background on. The resulting datasets for Run 1 and Run 2 were used for the following procedure.

## 5. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH CUT-BASED SELECTION

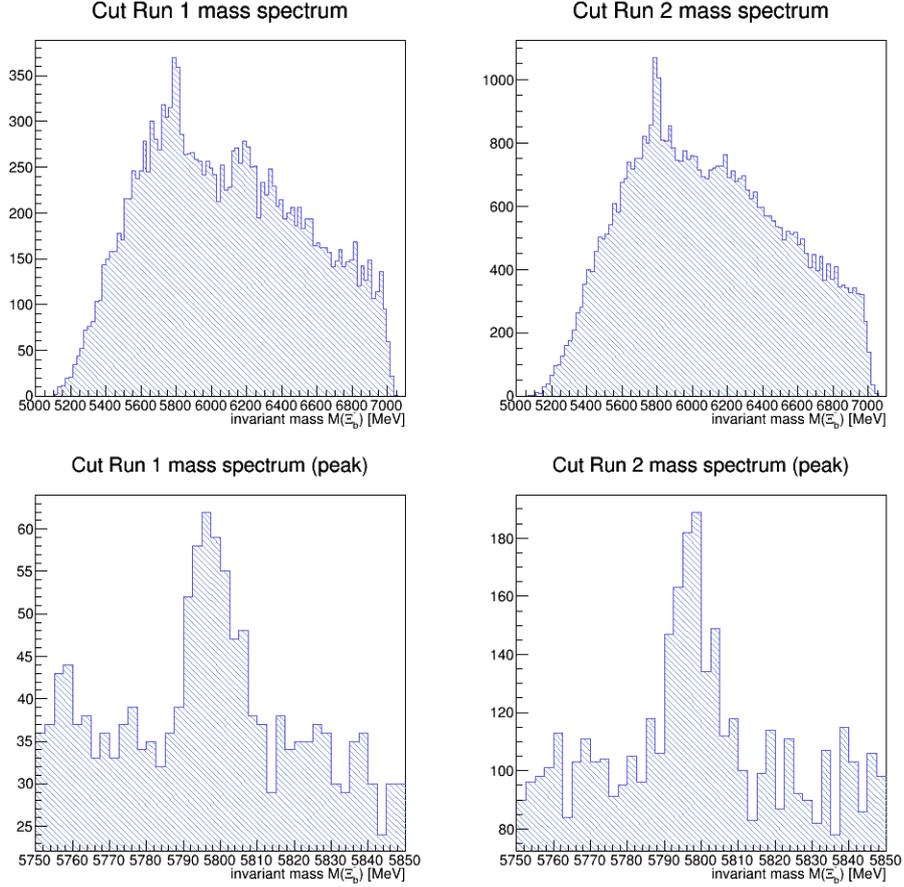


Figure 5.3: The first row shows real datasets of Run 1 and Run 2 with applied offline cuts. Applying the further mass cut on  $\Xi_b^-$  allowed to have a clear peak over background which is shown in the second row

### 5.3 Signal fit with MC data

To gain a signal yield for this data sample, calculate the significance of the signal and claim an observation of the decay channel  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ , the data must be fit. For this a tool named *beef*[34] based on *RooFit*[35] was utilized. To be precise a probability density function composing of a background term and a signal term is defined to fit the total amount of background and signal. It is necessary to say that with this method it is not possible to categorize a single event but merely to create a statistical statement about the yields.

Cut Nr.	event rejection [%]			
	Run 1		Run 2	
	Sig	Bkg	Sig	Bkg
0	6.22	77.06	1.95	46.40
1	3.09	25.97	11.12	59.59
2	0.44	11.25	0.04	5.83
3	0.41	10.92	0.04	5.88
4	17.75	88.40	9.76	64.82
5	5.98	35.36	10.05	62.32
6	0.42	4.18	0.28	4.07
7	3.49	65.26	5.09	67.80
8	3.29	19.44	2.07	14.80
9	5.20	73.30	3.1	63.83
10	4.34	74.70	5.57	74.37
11	2.26	15.04	0.52	12.42
12	1.57	67.31	0.72	58.70
13	0.18	1.02	0.17	1.66
All	40.41	99.85	35.38	99.40

Table 5.4: Event rejection of the single cuts in percent. The cut numbers correspond to table 5.3. They were calculated by applying them to MC data and an upper sideband. Cut “All” is the rejection of the combined cuts.

### 5.3.1 Testing different shapes

The MC data can be used to create the model for the signal term of the fit model. As the MC is a clean signal sample there doesn't have to be a background model for it. Four different common signal shapes were fit to it to decide on the best agreement:

1. Gaussian
2. Double Gaussian
3. Crystal-Ball
4. Breit-Wigner

The double Gaussian function defined as the sum of single Gaussians proved to be the best shape for this particular signal. Therefore it was chosen as the signal component  $f_{DG}$  for the model  $f$  as in the following:

The mean  $\mu$  and width  $\sigma$  are fit parameters.

$$\text{Gauss}(M; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{1}{2} \left(\frac{M - \mu}{\sigma}\right)^2\right) \quad (5.1)$$

## 5. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH CUT-BASED SELECTION

---

The two Gaussian terms in the double Gaussian share the same mean and width. The width of the second Gaussian is modified by a scaling factor  $r_\sigma$  which is also a fit parameter. A second scaling factor  $r_G$  modifies the size of the first Gaussian with respect to the second. A normalization factor isn't explicitly needed as RooFit normalizes the model pdf on it's own.

$$f_{DG}(M; r_G) = r_G \cdot \text{Gauss}_a(M; \mu, \sigma) + \text{Gauss}_b(M; \mu, r_\sigma \cdot \sigma) \quad (5.2)$$

For the background a common exponential background was used. Judging from the data after the cuts it will be almost flat. Nonetheless an exponential is used for more flexibility. It contains the decay constant  $\tau$ .

$$f_{Bkg}(M; \tau) = \exp(\tau \cdot M) \quad (5.3)$$

The complete fit model was created by adding the signal and background term with two additional factors  $N_{Sig}$  and  $N_{Bkg}$  representing the signal and background yield respectively. RooFit fits them considering the total amount of events given.

$$f(M; \dots) = N_{Sig} \cdot f_{DG}(M; \dots) + N_{Bkg} \cdot f_{Bkg}(M; \dots) \quad (5.4)$$

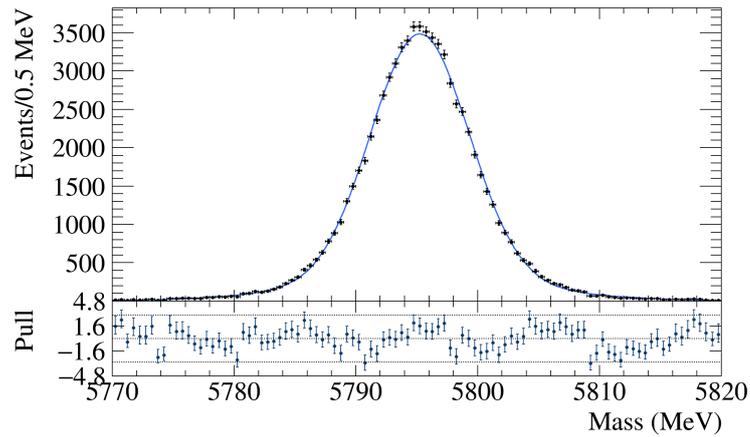


Figure 5.4: Fit of the signal term to the 2011 MC data. The blue line represents the fit PDF. Here the double Gaussian shape is used. Due to it's almost perfect agreement it is used in the fit of real data.

### 5.3.2 Fixing parameters

The fit to MC data helps with choosing the ranges for the parameters of the actual fit. Furthermore it was possible to fix two parameters from the MC. While the width may vary between the different runs the scaling factors  $r_G$  and  $r_\sigma$  should stay the same. Therefore they were fixed to the values which were gained from the 2011 MC:

$$r_G = 0.837$$

$$r_\sigma = 2.35$$

### 5.4 Determining $\Xi_b^-$ mass and signal yield

Following the preparations the fit to real data was performed. As they may slightly differ from each other the datasets of the two runs weren't combined into a single one. Instead an unbinned simultaneous fit was done. This means that both datasets were simultaneously fit with the same model. So called shared parameters are fit in a way they describe both sets in an optimal way, while the rest gives different values per sample.

The yields  $N_{\text{Sig}}$  and  $N_{\text{Bkg}}$  are separated parameters. The mean of the Gaussian was also selected as a separated parameter. This parameter corresponds to the mass of the mother particle  $\Xi_b^-$  and comparing the returned value of both runs gives an idea about the agreement of the two fits.

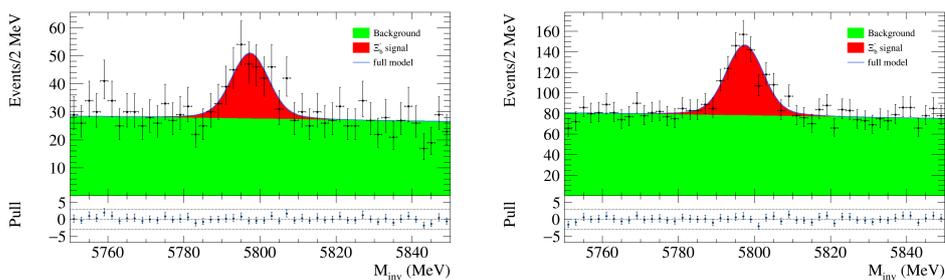


Figure 5.5: Model fit to the Run 1 (left) and Run 2 (right) datasets. The exponential background is shown in green while the red shape corresponds to the double-Gaussian signal shape. The binning is purely visual as the fit was done unbinned.

The results can be seen in figure 5.5 and in table 5.5. It is apparent that the mass of the particle  $\Xi_b^-$  is coherent between both runs.

5. OBSERVATION OF  $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$  WITH CUT-BASED SELECTION

---

	$M_{\Xi_b^-}$ [MeV]	$N_{\text{Sig}}$	$N_{\text{Bkg}}$
Run 1	$5797.31 \pm 0.97$	$152 \pm 25$	$1381 \pm 43$
Run 2	$5797.36 \pm 0.53$	$447 \pm 46$	$3911 \pm 74$

Table 5.5: Results of the fit to the datasets of Run 1 and Run 2. The mass of the mother particle  $\Xi_b^-$  is to be precise the mean of the double Gaussian.

This concludes the cut-based selection of the datasets of Run 1 and Run 2. It's apparent that while a convincing peak is shown over an exponential or even almost flat background it can be improved. For calculating an observation significance it is best to maximize the the ratio between amount of signal and background events. For this purpose it is useful to apply a different method which performs better in rejection background events while keeping the yield almost completely untouched. One way to do this was applied in the next section.

## 6 Observation of $\Xi_b^- \rightarrow J/\Psi + p^+ + K^- + K^-$ with BDT selection

While it was possible to gain good signal yields in the previous section there is room for improvement. Looking at the expected loss of signal due to the data cuts in table 5.4 about 40% is lost. As alternative to the cut-based selection, *Boosted Decision Trees* (BDT) were used. They allow to reject more background while keeping the signal yield relatively high. Even an increase wouldn't be impossible in comparison to the high loss in the first method. Rejecting enough background once more disclosed a signal peak in the data which was fit with the same model as before.

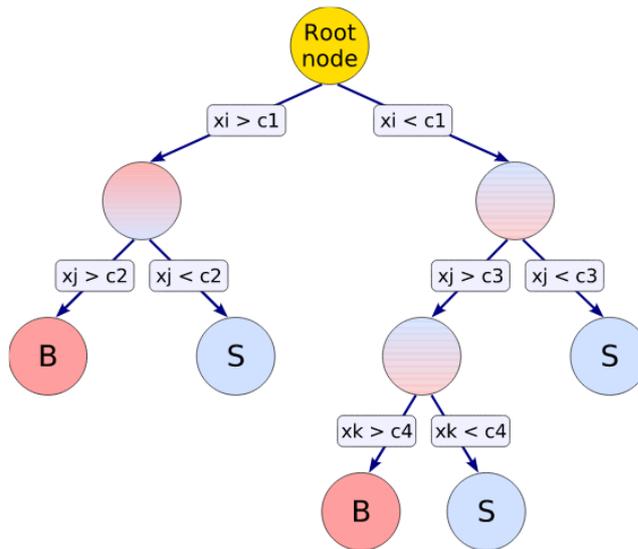


Figure 6.1: Example of a decision tree taken from the TMVA user guide [36]

### 6.1 Function of BDT's

In this analysis *gradient boosted decision trees* were used as a means of *multivariate analysis* (MVA), where a multitude of correlated variables are analysed [36]. As the name suggests a decision tree is a straightforward method where training samples for signal and background are sent through iterations of binary variable decisions which splits them into subsamples. The splitting criterion is always chosen in a way that the best separation between

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

background and signal is given. The tree grows till it reaches a stopping condition. These final subsamples are contained in *leaf nodes* and labeled as signal or background depending on which is more prevalent. A stopping condition would be for example reaching a certain *depth* or the amount of events in a subsample falls below a certain threshold. A problem of decision trees is that for large ones the risk of learning specific fluctuations (*overtraining*) is high. A good way to increase the stability of decision trees is boosting them. This means that a whole *forest* of small trees is grown from a training subsample. The trees are weighted and summed up so that the resulting final model  $F(\mathbf{x})$  where  $\mathbf{x}$  are the input variables, and the real classification  $y$  of the event are as close as possible. The deviation between true and modeled value is measured with a *loss-function*  $L(F,y)$  which is different depending on the chosen boosting method. In this analysis as already mentioned the *gradient boosting* was used. It uses a loss function which is relatively robust against fluctuations.

$$L(F, y) = \log \left( 1 + \exp^{-2F(\mathbf{x})y} \right) \quad (6.1)$$

The name comes from the fact that the minimization has to be done by scanning for the highest gradient.

### 6.2 Deciding for training variables

For the training of the BDT one must decide which variables to train on. In this case, as already a cut-based selection exists, it is possible to take the same variables as before. This was done for the majority of them which can be seen in table 6.1. In comparison to the first BDT three variables were omitted. First of all the mass of  $J/\Psi$  isn't used as training sample as this would force the mass spectrum directly in a desired shape. In the final BDT *atan ProbNNp effi* was used for a preselection and therefore it's quality as training variable shrunk. The other two were removed because of redundancy. A new variable *DIRA effi* describes the term  $(1-DIRA)*DIRAError$  which is expected to be a good training variable. Using the natural logarithm of variables results in a more homogeneous, wider distribution, making it easier for the BDT to find optimal cuts.

#	variable	particle	first BDT	opt. BDT	comment
0	atan ProbNNk	$K^-$	•	•	
1	$\log \chi_{IP}^2$ (PV)	$K^-$	•	•	
2	atan ProbNNmu	$\mu^+$	•	•	only Run 2
3	atan ProbNNmu	$\mu^-$	•	•	only Run 2
4	atan ProbNNp effi	$p^+$	•		
5	$\sum_{x_i} p_T(x_i)$	dec. prod.	•	•	
6	$\log \chi_{FD}^2$ (PV)	$J/\Psi$	•	•	
7	$\log \chi_{FD}^2$ (OV)	$J/\Psi$	•		
8	$\log \chi_{FD}^2$ (PV)	$E_b^-$	•	•	
9	$\log \chi_{IP}^2$ (PV)	$E_b^-$	•	•	
10	DIRA	$E_b^-$	•		
11	$\eta$	$E_b^-$	•	•	
12	log DIRA effi	$E_b^-$	•	•	new
13	$\log \chi_{DV}^2$	$E_b^-$	•	•	

Table 6.1: Overview over the training variables used for training. For a comparison the first and the optimized final BDT are shown. Used variables are marked with a •. Variables which weren't used in the last BDT were either applied as a preselection or removed for redundancy.

### 6.3 Choosing training samples

To train a BDT a signal sample as well as a background sample are needed. For the background sample the uncut real data can be used. Therefore there is an abundance of training events. To make sure that there is only a neglectable amount of signal events in the training sample, a sideband is utilized. For this training the upper side band events in range  $M_{E_b^-} \in [5820, 5900]$  were used. The reason why the upper sideband was used instead of both bands is that partially reconstructed events are found in a mass range below the desired peak. Those signal like events aren't desired in the training sample.

The signal sample should be as pure as possible to avoid training on wrong values. The investigated dataset can't be used as this could cause a bias. Instead the MC data can be used. To get the best possible training sample the MC data variables have to be checked to test their agreement with real signal data. This is described in the following section.

Both samples in general underwent cuts on nonphysical values, to be precise, PID variables which didn't lie in the interval  $\text{ProbNNx} \in [0, 1]$  were rejected.

## 6.4 Preparation of MC sample

The MC data has to be checked for agreement with real data to be used for the BDT training. Most important here are kinematic variables and the PID variables which were chosen as training variables. A description of the correction treatment for the MC is given.

### 6.4.1 Correction of ProbNN with PIDCalib

PID variables are often used in data analysis. Therefore they have to be simulated in the MC data which isn't a trivial task. As it combines the response of many different parts of the LHCb detector like the muon station and the RICH detectors those have to be simulated carefully. There are several sensitive parameters involved like the alignment of the detector, the occupancy or kinematics of the tracks. Due to that reason the so called *PIDCalib* software package was developed to correct PID variable distributions of MC datasets based on internal real data calibration samples [37]. There are currently two ways to correct PID variables with PIDCalib.

The classical approach is PIDGen. It's a resampling technique which means that the ProbNNx variables are completely generated anew based on calibration samples in three kinematic variables, namely momentum  $p$ , pseudorapidity  $\eta$  and number of tracks  $N_{tr}$ . The problem of this approach is that the PID variables depend on other variables, too and the correlations between the PID variables is lost making it impossible to use more than one in the training of the BDT [38]. As more than one of those variables were used this approach was unusable.

A newer approach is using PIDCorr. It is used for analyses which use more than one PID variable for the training of MVA. Using the same variables and calibration samples as before this method doesn't independently generate new values. Instead it aims to correct the PDF of the PID variables while preserving the correlations between them [39]. While this method has still the disadvantage of using only a few variables for correction and small calibration samples the results are acceptable [40].

While PIDCorr needs the three kinematic variables mentioned above for the procedure it doesn't need them to be corrected beforehand. Instead it checks if they're correct on it's own without outputting reshaped versions of them [40]. The only output it generates is a new tree with the corrected PID

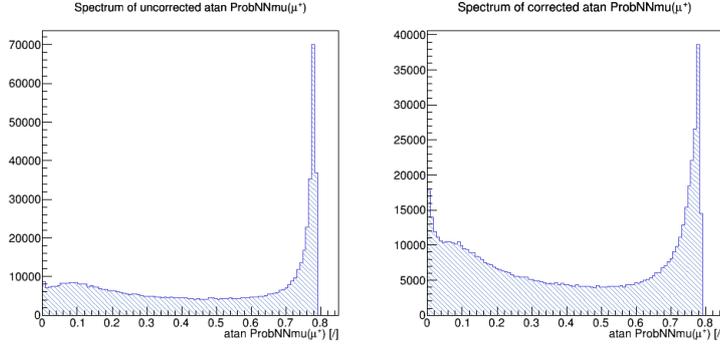


Figure 6.2: Comparison of the PID variable  $\text{atan ProbNN}\mu(\mu^+)$  before and after the correction with *PIDCorr*

variables. To get the training variables as described in section 6.2 *PIDCorr* was used to get the proper base *ProbNNx*. As there were no calibration samples for Run 2 *ProbNNmu* at the time of this work, this variable couldn't be reshaped and was not used as a training variable for the Run 2 BDT.

#### 6.4.2 Kinematic reweighting

As the *PIDCorr* method doesn't correct the shape of kinematic variables it has to be done manually. The most important kinematic variables are the pseudorapidity  $\eta(\Xi_b^-)$ , transverse momentum  $p_T(\Xi_b^-)$  and the number of tracks  $N_{\text{tr}}$ . Especially the number of tracks differs from the actual shape which has to be correct for the MVA.

The so called kinematic reweighting is performed by applying weights to every single event, giving the resulting histogram the desired shape. To do the weighting the shape has to be compared to signal events coming from real data. As already mentioned the fit done to the real datasets are not able to distinguish on a event-to-event basis what is signal and what's background. Although the shape can be drawn out of the fit with help of a background subtraction called *sPlot*.

*sPlots* are a statistical tool which is able to reproduce the shape contribution of different event sources, like signal and background, which are merged in a single sample. Using maximum likelihood fits *sweights* are calculated which, if applied to the dataset, are able to unfold the contributions of the sources [41]. In the case of wanting to subtract background it's only needed to distinguish between signal and non-signal.

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

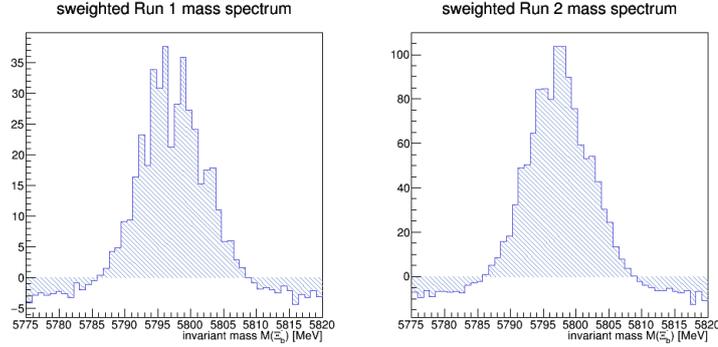


Figure 6.3: *sweighted dataset of the cut-based selection. One can see that the exponential background was suppressed and the double Gaussian signal remains.*

Beef, which was used to create the initial fits to the cut-based selection of events is able to create sweights to separate signal from background. Obvious background of the set was removed beforehand. This was achieved by a cut removing all events outside of  $M_{\Xi_b^-} \in [5775, 5820]$  [MeV]. Figure 6.3 shows the background subtracted signal shape of the cut-based distribution. As one can see the distribution is negative for events near the borders of the cut. This comes from negative sweights which are needed to compensate for the background. They are only present in this mass spectrum as it was used to fit the two sources and by that create the sweights. Applying the sweights to the real dataset removed the background events. Looking at the variables other than mass, for some bins which were close to zero the weight overshoot giving the bin a minuscule negative content which would be unphysical. In such cases the bin was set to zero to not create problems later on.

With this reweighted signal data sample it was possible to reweight the MC data to fit their PDF. Both MC and actual data was binned into  $n$  different bins where  $n$  didn't change between the different variables. As the real data is low on statistics and therefore has some strong statistical fluctuations  $n$  was chosen in a way that would allow to minimize the amount of fluctuations but at the same time be as big as possible. This is important as a too coarse binning would lead to a rough reweighting. As a second way to compensate for the low statistics a ROOT method was applied to smooth the histogram of the cut-based selection dataset.

Binning the data which was to be compared in  $n$  bins allowed for binwise weights. For the three variables which were corrected the ratio of bincontent between the datasets served as a weight for this specific variable  $x$ :

$$w_{x,i} = \frac{N_{\text{data},x,i}}{N_{\text{MC},x,i}}, \quad x \in (p_T(\Xi_b^-), \eta(\Xi_b^-), n_{\text{tr}}) \quad (6.2)$$

Here the index  $i$  labels the number of a bin. The MC data used in this formula under the index MC isn't uncut like for fixing the fit model parameters. The cuts of the cut-based selection which were used to unveil the signal peak used for the weights may influence variable shapes. Using the uncut MC data would may result in weights which skew the data into undesired PDF's. Therefore the MC data used to generate the kinematic weights has the cuts from the cut-based selection applied to them.

To preserve correlations the three weights must be combined into one single one called  $w_{\text{comb},j}$  on event-to-event basis. Combining the weights results in not perfect copies of the shape which makes it important to reweight only a few variables as more variables would give less accurate shapes when looking at isolated variables.

$$w_{\text{comb},j} = w_{\text{eff}} \cdot \prod_x w_{x,j} \quad (6.3)$$

$w_{\text{eff}}$  is a normalization factor of actual weights. This normalization is important for the calculation of uncertainties when calculating a branching ratio. It makes sure that the statistical power of the weighted sample is lower than before the weighting [42]. The normalization factor is defined as:

$$w_{\text{eff}} = \frac{\sum_{j=1}^N w_{\text{comb},j}}{\sum_{j=1}^N w_{\text{comb},j}^2} \quad (6.4)$$

$w_{\text{comb},j}$  are the event-wise weights while  $N$  is the total amount of events. Weighting all MC events with the final weights  $w_{\text{comb},i}$  corrects them to the right distribution which is controlled in the following. On a side note in this analysis the normalization is not directly needed. The BDT needs for a correct training only the correct shape, which is turned into a normalized pdf. Therefore this normalization is applied for the completeness of the weights for possible calculations in future projects.

### 6.4.3 Checking the distributions

As mentioned before for the kinematic reweighting the cuts applied to the signal channel and the MC data have to be identical. This resulted in a problem. For the training of the BDT all cuts applied to a sample, be it signal or background sample, must be applied to the other one as well. This means the cuts on the MC data which was corrected to serve as the signal sample have to be applied to the background sample as well. While the statistics of the MC data is still plenty the sideband statistics get heavily reduced by the cuts as this is their initial purpose. This increases the risk of lower quality BDT's. The best possible options would be to have weights applied to the uncut MC data so the full statistics of the upper band can be used. This would be possible if the difference between the weighted shapes before and after the cuts on the MC would be sufficiently small. Three variables were used for the kinematic reweighting so it's enough to check them for a comparison. Figure 6.4 shows this comparison. The blue set is at the edges slightly wider distributed causing a lower maximum in comparison. Taking this into account it's reasonable to allow the usage of the kinematic weights on uncut MC as a BDT signal sample.

As a final test on the sample the combined weights should accurately describe the signal shape of the real data. Figure 6.5 and 6.6 show in the left column the uncut, unweighted MC data in red and the background subtracted cut-based signal in blue. The right column features the MC data after applied combined weights. The weights are able to sufficiently recreate the PDF, even the extreme difference in the number of tracks.

## 6.5 Training the BDT

Using the corrected samples allowed to train a BDT with the training variables shown before in 6.1 and the used options in 6.2. Figure 6.7 shows at the left side the distribution of signal and background events against BDT response for training and testing samples. Potential overtraining could be observed in such a plot. The figure also shows the *Receiver Operating Characteristic* curve which plots the rejected background against the signal efficiency.

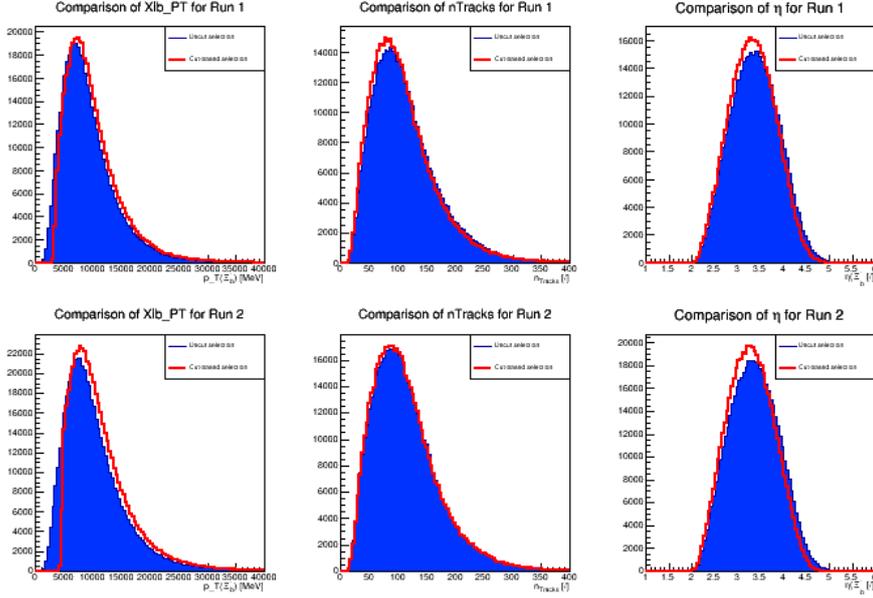


Figure 6.4: Comparison of the weighted cut (red) and uncut (blue) MC dataset. The blue distribution is the uncut and the red the cut-based selection. The upper row shows Run 1 data and the lower one shows Run 2.

### 6.5.1 Study of the ROC and response curve

Looking at the events distribution curve first of all one sees that in general there is a strong distinction between the classified signal and background events. This is also shown in the ROC curve where only for extremely high signal efficiencies the background rejection rate starts to get noticeably worse. This is first and foremost a characteristic of gradient boosted decision trees. One can also see that for almost a perfect signal efficiency still a background rejection of over 80 % is given. While this is generally desirable it can also indicate that the BDT used the strongest variables to apply cuts which only reject obvious background. Such a behaviour could result in events which are misclassified as the BDT didn't train harder to resolve differences in other variables. An indicator that this might've been the case is that for high BDT responses the background sample shows a peak indicating background events classified as signal. Two approaches were applied to suppress this. One idea for what those events are was that some signal was left in the training sample. Therefore in comparison to the first

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

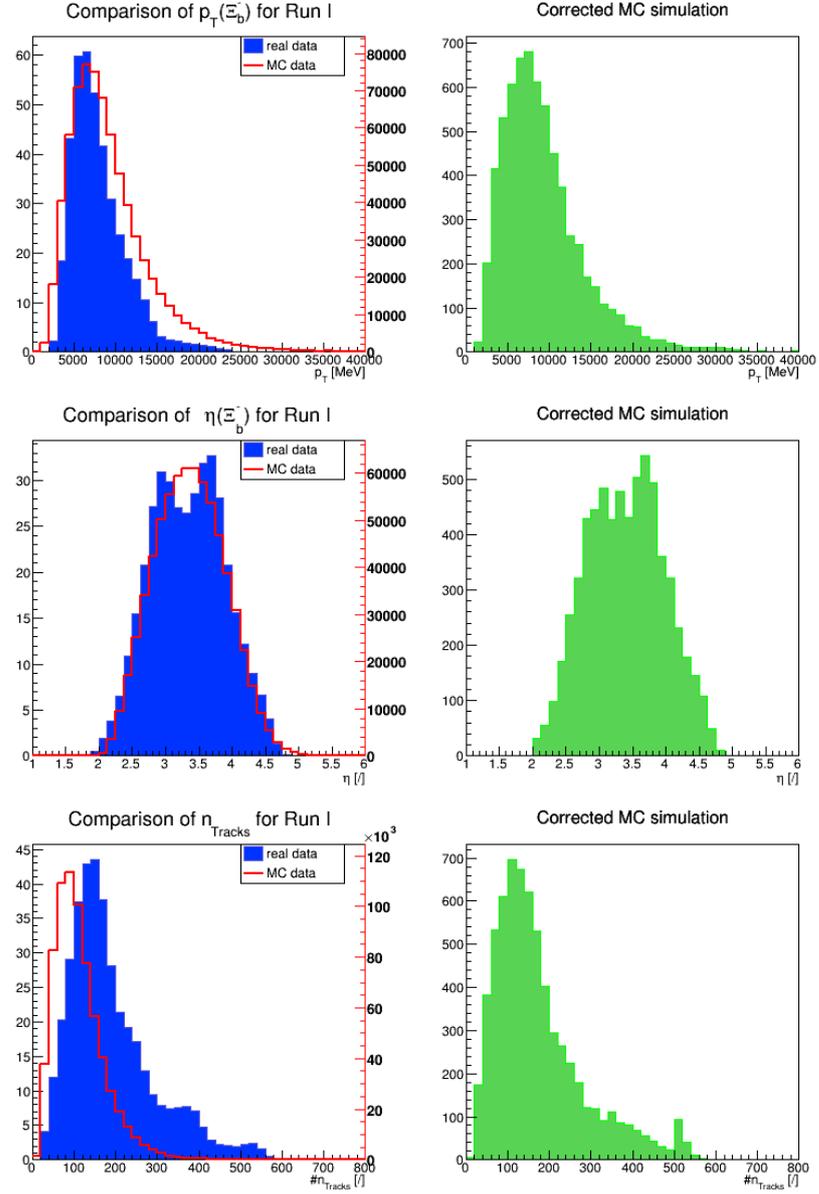


Figure 6.5: Comparison between the signal sample before and after applying kinematic weights. On the left the MC dataset before weights (red) and the real data sample (blue). The green distribution shows the MC data with applied weights. The data is coming from Run 1.

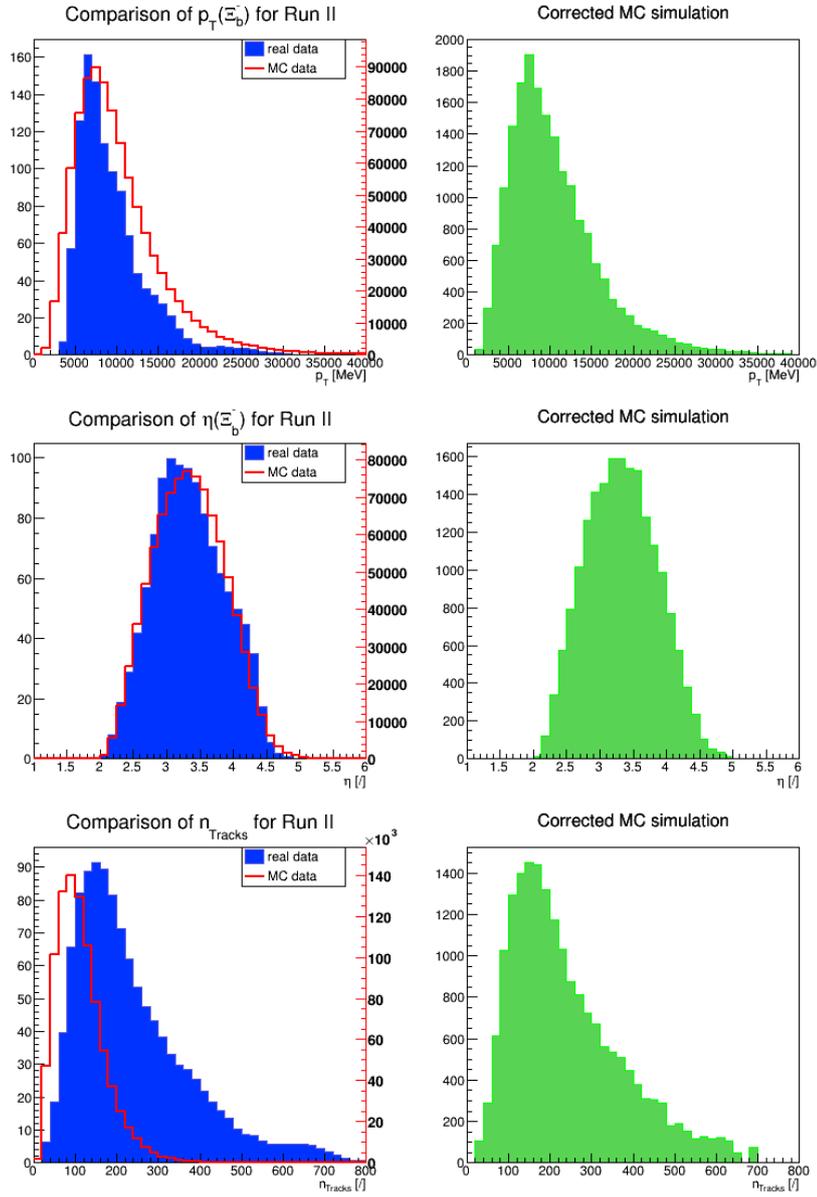


Figure 6.6: Comparison between the signal sample before and after applying kinematic weights. On the left the MC dataset before weights (red) and the real data sample (blue). The green distribution shows the MC data with applied weights. The data is coming from Run 2.

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

BDT the upper band background sample range was moved from [5820, 5900] MeV to [5900, 6100] MeV. While already for the first range the signal events should've been negligible, in this range there should be no events at all. A second theory was that the peak was caused by background events with similar characteristics to signal. The approach applied to suppress this was a preselection of events before the BDT training. As preselection variable a cut  $\text{atanProbNNpeffi}(p^+) > 0.2$  was applied. Applying the strongest classifier as preselection cut instead of using it as training variable forces the BDT to use less strong variables for the classification and by that train harder. Both applied methods didn't manage to suppress the peak. A complete study of this structure wouldn't achieve a sufficient increase in BDT quality to justify the consumed time. Nonetheless both methods are featured in the final BDT as they have secondary purposes. As already seen in the cut-based selection the amount of background in comparison to the signal yield is vast. Even with a ROC curve as good as in this BDT the needed minimal BDT response to get a significant signal peak would be very high. By cutting obvious background with a preselection the response doesn't have to be as high to get a good signal to background ratio. While this reduced the background training sample, this was counteracted by the increased side-bandrange. A harsher preselection was also tested but no noticeable change in the ROC curve or response plot indicated a reasonable quality increase. Instead it caused the samples to lose too much statistics again so that only the single preselection cut was kept.

### 6.5.2 Lessons from overtraining suppression

Overtraining is a common occurrence training a multivariate classifier. As seen in figure 6.8 overtraining causes a decrease in rejection power when applied to the testing sample [43]. This is shown in this plot as the blue distribution lays over the training events shown as markers with errorbars. Those in comparison are lower than expected. The cause for such a behaviour is that the sample has too few statistics to train for the algorithm parameters. Options causing this phenomenon were therefore identified and optimized while creating the BDT.

Looking at the options in the BDT training (6.2) a number of them were tuned to create a better BDT with less overtraining, first of all *NTrees*, the number of trees which are grown from the data. As mentioned boosted deci-

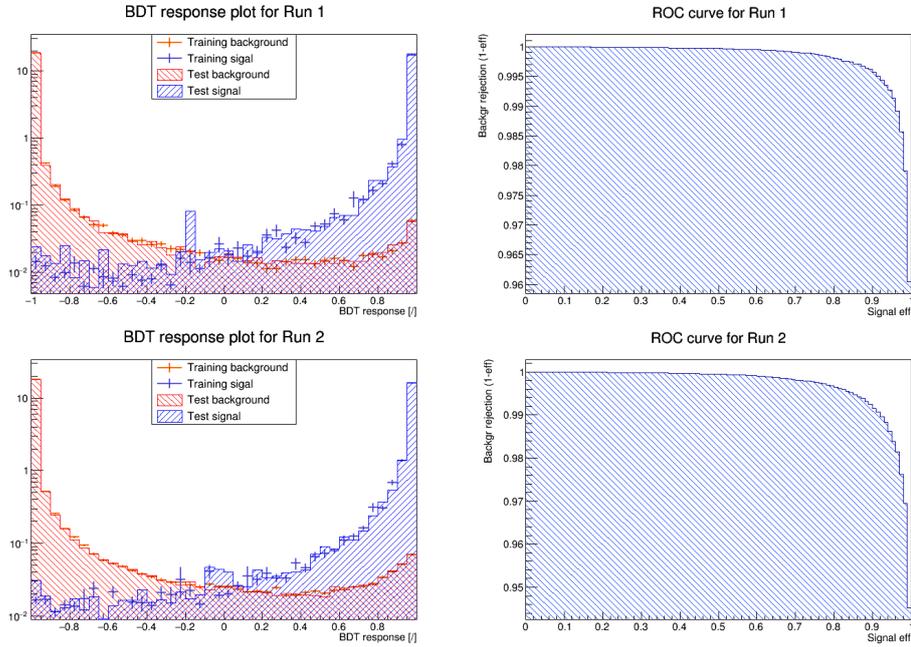


Figure 6.7: The response plot of the best BDT and it's ROC curve for Run 1 and Run 2. In the response plot the distribution shows the training events (errorbars) versus the test events (filled shape). Red is for the background sample and blue for the signal.

Option	First BDT	Final BDT
Preselection	none	p atan ProbNNp effi > 0.2
NTrees	1600	1000
MaxDepth	3	2
MinNodeSize	5.0 %	2.0 %
nCuts	80	80
BoostType	Grad	Grad
UseBaggedBoost	True	True
BaggedSampleFraction	0.6	0.8
Shrinkage	0.1	0.05
NodePurityLimit	0.6	0.5
Train/test Bkg events	600000	150988
Train/test Sig events	600000	473357

Table 6.2: Settings for the final BDT in comparison to the settings of the overtrained first. If an option isn't listed it is set to the default value. The amount of training/test events is the total event count. They are evenly split between training and testing. A list with every possible option can be found here [36]

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

sion trees need a forest of trees which are weighted to get a final classification model. It's not directly possible to create overtraining with a high number of trees as after a certain amount the classifier will reach a saturation. By checking the information gain per tree this saturation point was found and use to decrease unnecessary computation time. Two options closely related are the *MaxDepth* and the *MinNodeSize*. Both are stopping criterion and therefore vital to prevent a tree too specific to a data sample. Either when the maximal allowed depth, namely the number of steps from the first node to a leaf, or the minimal amount of events in a node is reached the splitting is stopped. By setting a low minimal node size and depth the growth will get limited by the latter. Lowering those variables increases the need for more trees but as already seen the number of trees was decreased. This could be done as due to the preselection the amount of available training statistics also decreased and the initial number of trees was a fair amount too high which offsets the change. Changing those two variables were effective in not only suppressing overtraining but also reducing the training time. A third variable linked to the amount of trees is the *shrinkage*. A smaller shrinkage means that in the boosting a smaller weight is applied per weak classifier. This decreases the influence of trees which learned statistical fluctuations but again increases the amount of needed trees. *nCuts* is the amount of datapoints used to calculate the splitting criterion. Changing it barely causes a difference. Finally so called bagging was used. This means that random subsamples are used to train every tree in the forest. Allowing to sample the same event several times effectively makes it less probable to sample statistical fluctuations. This causes a good suppression of overtraining. This was enhanced by increasing the amount of randomly sampled events (*BaggedSampleFraction*).

### 6.6 Finding the best BDT response cut

The aim of the thesis was first and foremost the first observation of the decay channel  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ . Therefore it was necessary to find the best possible BDT response in terms of maximising the observation significance. As the BDT will also reject signal events this means that it was desirable to find a BDT cut which maximises the background rejection with minimal signal loss.

The process of choosing the BDT response cut should be done blind which

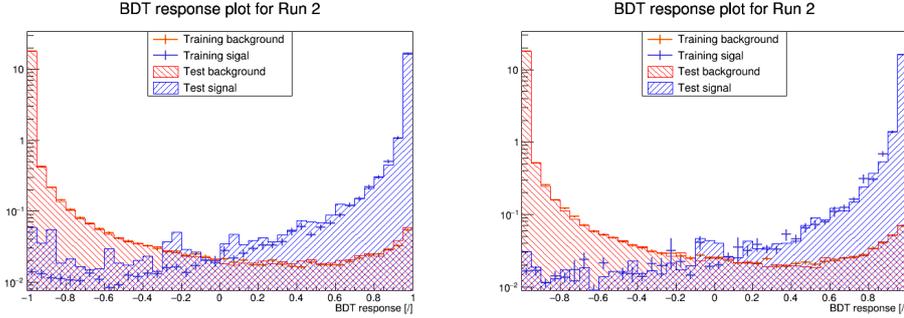


Figure 6.8: Comparison of a BDT which wasn't optimized in terms of overtraining and one that was. They are shown left and right respectively and both are for Run 2 data. The suppression of overtraining can clearly be seen.

means that it is chosen based on data independent from the dataset to which the BDT will be applied. Doing the optimization on the to be investigated dataset can cause a bias which influences the significance. Using a blind approach, meaning using an independent reference sample as for example the training samples, bypasses this problem.

### 6.6.1 Preparing a pseudo significance

For doing a scan of the BDT response while trying to maximize the significance one can use a pseudo significance called *Figure of Merit* (FoM). A FoM isn't equal to a significance but it behaves in a sufficiently similar way to a significance in terms of maximization with respect to a variable. There are several FoM which can be used each with their own advantages and disadvantages but in this analysis the *Punzi figure of merit* is an appropriate choice [44].

$$\text{FoM}(t) = \frac{\epsilon(t)}{a/2 + \sqrt{B(t)}} \tag{6.5}$$

The figure of merit is dependent on a cut  $t$  which in this case will be the BDT response cut.  $\epsilon$  denotes the signal efficiency for a given cut. The parameter  $a$  is the amount of significance in sigma one wants to simulate which is  $5\sigma$  needed for the claim of an observation.  $B$  is the background yield after applied cut  $t$ . The reason why this FoM was chosen is that it doesn't need a prior assumption of the signal yield in the sample you

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

optimize the response cut for. In comparison, for other FoM, signal and background yields  $S$  and  $B$  of the comparison samples have to be scaled to an amount mimicking the real yields to get proper feedback. Such a signal assumption could for example be created by analysing a reference channel or through theoretical studies. Another possibility is applying an unoptimized preselection to the data so that you can see a first peak to get a conservative yield assumption. Specifically for this analysis all of those weren't viable methods, motivating the Punzi FoM. For the background scaling it was used that the comparison sample is the upper band of the investigated sample and after the preselection the amount of events scales in this range linearly with the mass. A linear fit to the upper band was used to get a blind estimate of the actual amount of background needed for the FoM.

### 6.6.2 Scanning the FoM

As a independent, comparison sample the training samples of the BDT were chosen. After the preparations described above the FoM was step wise calculated for different BDT response value cuts. A gradient boosted decision tree generally has it's optimal cut in higher areas than for other methods and the preselection still left a lot of background. Therefore the scan is performed in a high range of  $BDT_{\text{response}} \in [0.800, 1.000)$  with a stepsize of 0.001. The maximum can be used as the best cut. The scan for both Run samples can be seen in figure 6.9. In Run 1 as all other points describe a descension after 0.994 except for one, this outlier wasn't chosen as optimal value even though it technically was the maximum. Therefore the optimal cuts used were:

for Run 1: 0.994

for Run 2: 0.992

The resulting datasets after applying those cuts were used to calculate an observation significance in the following.

### 6.7 Analysing the BDT created spectrum

In this final part of the analysis the BDT and it's optimized response cut were applied to the datasets of Run 1 and Run 2. The resulting dataset was fit simultaneously with effectively the same fit model as in section 5.4.

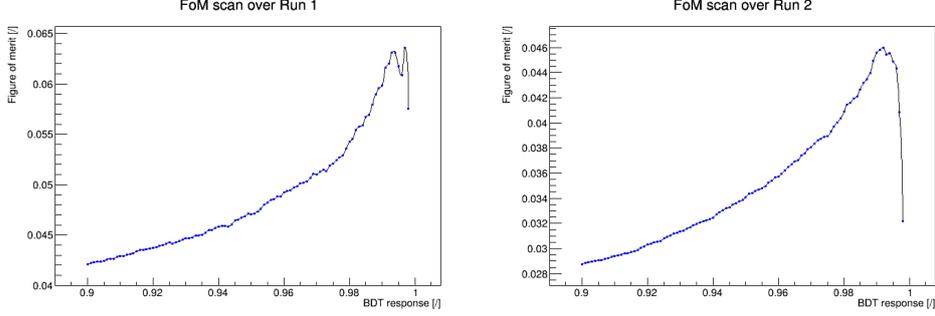


Figure 6.9: Scan of the FoM in dependency on the BDT response for both Runs.

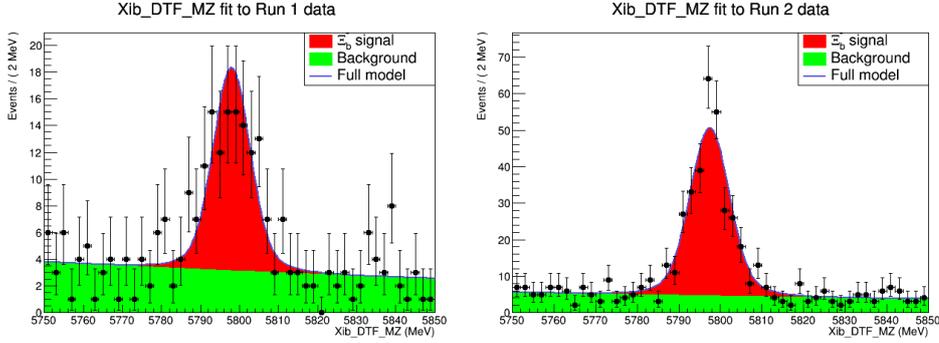


Figure 6.10: Best fits of the likelihood scan for Run 1 and Run 2

Figure 6.10 shows the resulting fits for both Runs. You can see at first sight that in comparison to the fits of the cut-based selection (5.5) the background was, as expected, greatly reduced but in exchange the signal yield also seems to have decreased. The yield values resulting from the fit can be found in table 6.3

Further relevant values which were to be extracted from the fits were a statistical calculated uncertainty of the signal yield and of course the significance

	$M_{\Xi_b^-}$ [MeV]	$N_{\text{Sig}}$	$N_{\text{Bkg}}$
Run 1	$5797.96 \pm 0.81$	$96 \pm 13$	$158 \pm 15$
Run 2	$5797.41 \pm 0.37$	$290 \pm 20$	$233 \pm 19$

Table 6.3: Results of the fit to the datasets of Run 1 and Run 2 after applying the BDT to the dataset.

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

of the signal. Both were calculated by a likelihood scan. In a statistical sense the likelihood  $\mathcal{L}(\Theta|x)$  is a measure of how likely a parameter or set of parameters  $\Theta$  is given a dataset of values  $x$ . A high likelihood here means that the given parameters are highly coherent to the data, giving a good description.

### 6.7.1 Changes to the fitting model

Doing a likelihood scan or more precisely a *profile likelihood* scan means that for each iteration the scanned variable is fixed to a value and the others parameters are fit to the data in a way that the resulting likelihood is maximized. As before the fit is done simultaneously. A total significance including both Runs is to be calculated. Therefore the scan has to be done over the total signal yield which is why the fit model had to be rewritten:

$$\text{Run1} : f_1(M; \dots) = N_{\text{Sig,tot}} \cdot r_{\text{Sig}} \cdot f_{\text{DG}}(M; \dots) + N_{\text{Bkg}} \cdot f_{\text{Bkg}}(M; \dots) \quad (6.6)$$

$$\text{Run2} : f_2(M; \dots) = N_{\text{Sig,tot}} \cdot (1 - r_{\text{Sig}}) \cdot f_{\text{DG}}(M; \dots) + N_{\text{Bkg}} \cdot f_{\text{Bkg}}(M; \dots) \quad (6.7)$$

Hereby  $N_{\text{Sig,tot}}$  is the total yield and  $r_{\text{Sig}}$  the yield scaling. Both are shared between the Runs. The rest of the parameters are treated the same as before. The best fit plots in figure 6.10 and the corresponding values in table 6.3 were also created with this updated model. The values for the yields separated by Run were calculated by using the yield scaling and applying error propagation. Doing the likelihood scan with this model resulted in figure 6.11 and 6.12. They come from the same scan and 6.11 was zoomed in. Every black dot represents one scan point. For convenience reasons the negative doubled logarithm of the likelihoods with minimum shifted to zero was drawn. The shift to zero was done by subtracting the likelihood of the best fit. The best fit was created by letting the signal yield float instead of fixing it. This subtracting corresponds to a delta log likelihood  $\Delta \log(\mathcal{L}(N_{\text{Sig}}|x)) = \log(\mathcal{L}(N_{\text{Sig}}|x)) - \log(\mathcal{L}(\widehat{N}_{\text{Sig}}|x))$  where  $\mathcal{L}(\widehat{N}_{\text{Sig}}|x)$  is the likelihood for the best fit. All in all this means that the negative delta logarithmic likelihood  $-2 \Delta \log(\mathcal{L}(N_{\text{Sig}}|x))$  was plotted.

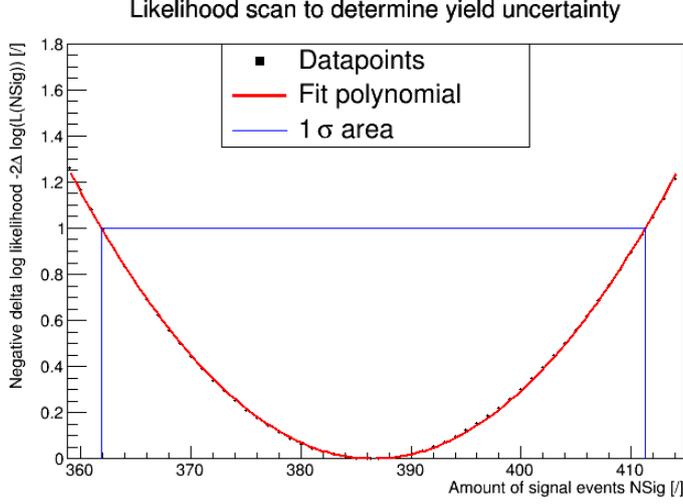


Figure 6.11: Scan of the negative delta log likelihood with dependency on the total signal yield  $NSig$ . It's the same scan as 6.12 zoomed to the area where the negative delta log likelihood falls under 1. The signal yields where this happens are the 68% uncertainty values on the best fit signal yield.

### 6.7.2 Uncertainty of the signal yield

A *confidence interval* with a chosen *confidence level*  $1 - \alpha$  can be constructed. A  $1 - \alpha$  confidence level means that for repeated measurements and calculation of the confidence interval it will cover the true value with  $1 - \alpha$  probability.

In the case of the likelihood scan a 68.3% confidence interval can be determined with the help of *Wilk's theorem* which states that for large samples  $-2\Delta \log(\mathcal{L}(\Theta|x))$  follows a d-dimensional  $\chi^2$ -distribution [45]. d is the difference in free parameters between the likelihoods in the delta log likelihood. Effectively this means that the confidence interval can be constructed by checking for which  $NSig$

$$-2\Delta \log(\mathcal{L}(NSig|x)) \leq \chi_{1-\alpha,d}^2 \quad (6.8)$$

where  $\chi_{1-\alpha,d}^2$  is the  $1 - \alpha$  quantile of the d-dimensional  $\chi^2$ -distribution. For calculating the yield uncertainty there is just one free parameter difference and  $1 - \alpha = 0.683$  which corresponds to an interval containing every  $NSig$  where equation 6.8 falls under 1.

As the likelihood scan is drawn in a way that it shows  $-2\Delta \log(\mathcal{L}(NSig|x))$

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

	fit uncertainty	scan interval
total signal yield $N_{\text{Sig}}$	$386 \pm 25$	$386_{-25}^{+24}$

Table 6.4: Comparison of uncertainty calculated by the fitter and by using Wilk's theorem.

this was easily determined. Figure 6.11 shows that first a parabola function was fit to the data as a red line. This fit then could be evaluated to find where the yield falls under the threshold. The blue lines encompass the interval where this holds true. Table 6.4 shows the comparison between the uncertainty given by the fit and the interval determined by the scan.

As one can easily see the uncertainties are coherent.

### 6.7.3 Significance

Finally one has to calculate the observation significance of the signal yield. Wilk's theorem can also be applied to construct such a significance. Choosing a likelihood ratio as a *test statistic* which is a variable which combines relevant information about a dataset the significance  $Z$  is defined as:

$$Z \approx \sqrt{-2 \log \left( \frac{\mathcal{L}(0|x)}{\mathcal{L}(\widehat{N}_{\text{Sig}}|x)} \right)} \quad (6.9)$$

The term in the natural logarithm is the likelihood ratio for the hypothesis “No Signal” against “Signal” with an optimized signal yield  $\widehat{N}_{\text{Sig}}$ . The Neyman-Pearson lemma states that for no nuisance parameters the likelihood ratio has the best performance [46]. In this case there are still nuisance parameters for example in the form of a not perfectly known background yield, nonetheless it is one of the best options for this case. The way the likelihood is drawn one had just to look up the ordinate value for the scan point  $N_{\text{Sig}} = 0$ . The square root of this value is the yield significance. The negative delta logarithmic likelihood  $N_{\text{Sig}} = 0$  is 580.177 which corresponds to a significance of  $24.1\sigma$ . This surpasses the needed  $5\sigma$  significance to claim an observation by far and by that confirms the first observation of  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ .

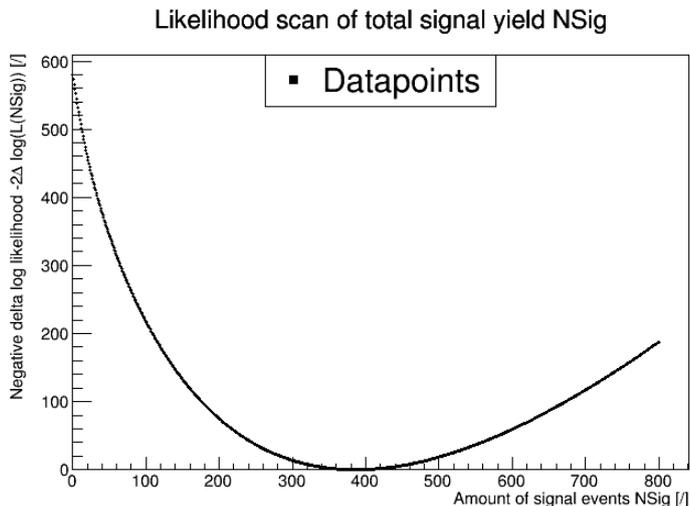


Figure 6.12: Scan of the negative delta log likelihood with dependency on the the total signal yield  $NSig$ . Every datapoint represents one fit with a certain signal yield. The square root of the value at  $NSig = 0$  corresponds to the yield significance.

### 6.8 A look at systematic uncertainties

With the calculation of the yield significance the observation of  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$  can be formally claimed. It is to be noted that all error values given in this analysis are of statistic nature. This means that through systematic uncertainties the yield significance might be lower. Looking at the extremely high significance of the observation this is of no concern. Nonetheless it is helpful to be aware of the different systematic uncertainties in this analysis . Therefore this subsection will give an overview about some systematic uncertainties which are present. Most of the main systematic uncertainties are coming from the fit model as it plays a major role in calculating the yield and significance. One of them is introduced by the general fit model shapes. The signal shape was chosen by fitting the 2011 MC data. This includes the underlying double-Gaussian shape as well as the fixed parameters for both Runs. In principle the MC could change between the years with it's resolution as only one example. This can introduce a large uncertainty. In comparison to that the exponential background was chosen based on the qualitative look of the sidebands. It would be possible to analyze the background for contributions from other sources which aren't covered by the chosen exponential function. A second

## 6. OBSERVATION OF $\Xi_B^- \rightarrow J/\Psi + P^+ + K^- + K^-$ WITH BDT SELECTION

---

systematic uncertainty related to the fit model is which parameters were chosen to be shared in the simultaneous fit. The width of the double-Gaussian was shared between the Run datasets. This assumed that the detectable width of the signal didn't drastically change between the Runs. Looking at the fit on Run 1 the width could've been wider. This introduces a systematic uncertainty on the signal yield. Another major uncertainty stems from the kinematic reweighting of the MC data to use it as BDT training sample. To calculate the weights the data had to be binned. The bin width was chosen in a way that no statistical fluctuations from the reference sample gets propagated to the MC data. As the bin width influences the quality of the distribution it introduces an uncertainty. It could be minimized by using adaptive binning, bins with variable width, to maximize the available statistics. This summarises the major sources of uncertainty. Given enough time a full analysis of the systematic uncertainties can be produced.

## 7 Conclusion and outlook

The aim of this thesis was the first observation of the decay channel  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ . This channel could prove as a useful reference for a sexaquark search. The complete dataset from Run 1 and Run 2 were first utilized in a cut-based selection. Cuts which were tested on fully detector simulated MC data were applied to the dataset. The efficiencies of the cuts in terms of signal efficiency and background efficiency are given in table 7.1:

Run	Signal efficiency [%]	Background efficiency [%]
1	59.59	3.83
2	64.62	11.8

Table 7.1: Cut efficiencies for signal and background in Run 1 and Run 2.

For the fit model a double-Gaussian signal over exponential background was chosen based on testing different models on MC data. Using a simultaneous fit of the Run 1 and Run 2 datasets gave as yields for signal and background as seen in the overview table 7.2.

Selection	Run	$M_{\Xi_b^-}$ [MeV]	$N_{\text{Sig}}$	$N_{\text{Bkg}}$
Cut-based	1	$5797.31 \pm 0.97$	$152 \pm 25$	$1381 \pm 43$
	2	$5797.36 \pm 0.53$	$447 \pm 46$	$3911 \pm 74$
BDT-based	1	$5797.96 \pm 0.81$	$96 \pm 13$	$158 \pm 15$
	2	$5797.41 \pm 0.37$	$290 \pm 20$	$233 \pm 19$

Table 7.2: Overview over the yields for both Runs as seen in the two used methods. The mass of the motherparticle is also shown.

As there was still a high amount of background an alternative approach for the observation was used. This took the form of a gradient boosted decision tree. The BDT was trained on MC data as signal sample and the upper sideband of the dataset as background sample. The MC was corrected beforehand by using the tool PIDCorr to correct the PID variables and weights for the kinematic variables. Using a Punzi figure of merit the BDT response cut was optimized blindly. The resulting best response cuts were

## 7. CONCLUSION AND OUTLOOK

---

for Run 1: 0.994

for Run 2: 0.992

Applying the response cut and fitting the resulting reduced dataset gave the yields which are also shown in table 7.2. A likelihood scan was performed to calculate the  $1\sigma$  confidence interval of the total signal yield as well as the signal significance. The total yield was measured as  $386_{-25}^{+24}$ . The signal significance was determined as  $24.1\sigma$ .

This concludes the first observation of the channel  $\Xi_b^- \rightarrow J/\Psi(\rightarrow \mu^+ + \mu^-) + p^+ + K^- + K^-$ . It is clear that this channel exists and that it can be used as a reference channel to test models for future sexaquark searches. The BDT can be used to extract the signal events for this decay channel for future datasets.

Apart from that it is possible to further improve the quality of the BDT for future work. The quality can be improved by investigating the minor signal like background which was shown in the BDT overtraining plot. Furthermore a branching ratio could be measured. The masses of the motherparticle  $\Xi_b^-$  look promising to be competitive to the current values on PDGlive. To get a true mass value the datasamples have to be momentum scaled and an analysis of the systematic uncertainties has to be conducted. This was just a first step in fully understanding this decay channel. Nonetheless the results are promising and the future will show what further analyzes will bring.

## Appendix

### Overview over all cut names

**IsMuon:** Boolean statement if the selected particle was detected by the muon system of the LHCb. It therefore selects muons in general.

$\chi_{\text{TR}}^2$  (**MaxChi2Cut**): Cuts away all tracks above a certain  $\chi^2$ -value. This ensures that no unreasonable tracks are chosen as candidates.

$M_x - M_{J/\Psi, \text{PDG}}$  (**ADAMASS(J/psi(1S))**): This variable is the absolute difference between the measured mass of a particle “x” and its value documented by the *Particle Data Group*. Defining an upper boundary on this difference is an easy way to discard events with unreasonable masses.

$\chi_{\text{DOCA}}^2$  (**ADOCACHI2CUT**): A cut on the maximum  $\chi^2$  for the “Distance Of Closest Approach”. This selects muon pairs which were reasonably close at one point in time to be the decay product of the same  $J/\Psi$ . Pairs of particles which are too far apart are discarded.

$\chi_{\text{DV}}^2$  (**VFASPF(VCHI2)**): Limits the maximum  $\chi^2$  of vertex fits performed with the characteristics of the accepted particle pairs. This ensures the quality of the reconstructed decay vertex positions of the mother particle.

**min PID( $\mu$ )** (**MINTREE('mu+'==ABSID,PIDmu)**): A loose cut on the particle identification variable of a muon. Effectively it can be written as the difference between the logarithmic likelihood that the particle is a muon and the logarithmic likelihood that it's a pion  $\log(\mathcal{L}_\mu/\mathcal{L}_\pi)$ .  $\mathcal{L}$  labels the likelihood function. It discards every muon candidate were it's more probable that it's a pion instead of a muon.

**min  $p_T$**  (**MINTREE('mu+'==ABSID,PT)**): This cut checks the decay tree for muons and rejects if the momentum transverse to the collider beam is below a certain threshold.

**BPVDLS:** Restricts candidates to have a minimum “Decay Length Significance” with respect to the primary vertex. The decay length significance is the decay length over it’s error. Therefore this checks the quality of the reconstruction by discarding events with uncertain decay length.

**M (MM):** The “Measured Mass” or *invariant mass* of a particle. It is calculated by the energy-momentum relation

$$M = \sqrt{E^2 - \vec{p}^2}$$

Hereby the light speed  $c$  is set to 1. This cut around  $J/\Psi$  cuts away obvious background while the  $\Xi_b^-$  mass cut is a loose boundary.

$\chi_{\text{DV,DOF}}^2$  (**VFASPF(VCHI2PDOF)**): Similar cut to  $\chi_{\text{DV}}^2$ . It also evaluates the  $\chi^2$  of the decay vertex but also divides it by the degrees of freedom. By that it is more forgiving if the decay vertex has many degrees of freedom otherwise it’s a strict cut.

$\chi_{\text{TR,DOF}}^2$  (**MAXTREE(‘mu+’==ABSID,TRCHI2DOF)**): Similar to the  $\chi^2$  track cut in *StdAllLoseMuons*. It checks again the reconstructed tracks of the muons, this time with respect to the degrees of freedom.

$p_T$ : The momentum of a particle transverse to the z-direction of the LHCb detector . As mentioned in section 4 this direction corresponds to the beam direction.

$\min \chi_{\text{IP}}^2(\text{PV})$  (**MIPCHI2DV(PRIMARY)**): This variable corresponds to the minimum  $\chi^2$  of the particles *Impact Parameter* to a given vertice. The impact parameter is the perpendicular and by that shortest distance of a particles track to a decay vertex. In this case the primary vertex is specified. This ensures the reasonable assignments to a primary vertex.

$\Delta \log(\mathcal{L})(K - \pi)$  (**CombDLL(x - pi)**): This variable stands for the difference in the *Delta Log Likelihood* of a particle “x” and a pion. It is the equivalent to “min PID( $\mu$ )” for kaons and protons. Like before the combined information about the candidates are used to discard those which are more likely to actually be pions.

$\chi_{\text{IP,BPV}}^2$  (**BPVIPCHI2**): Similar cut to “min  $\chi_{\text{IP}}^2(\text{PV})$ ”. In this cut the  $\chi^2$  of the impact parameter to the *best* primary vertex.

**BPVDIRA**: A particle’s spatial trajectory is sometimes not exactly equal to the combined momenta vector of the daughter particles. The *DIRection Angle* describes the cosine between those vectors.

## References

- [1] **LHCb** Collaboration, R. Aaij *et al.*, “Observation of  $J/\psi p$  Resonances Consistent with Pentaquark States in  $\Lambda_b^0 \rightarrow J/\psi K^- p$  Decays,” *Phys. Rev. Lett.* **115** (2015) 072001, [arXiv:1507.03414 \[hep-ex\]](#).
- [2] **LHCb** Collaboration, R. Aaij *et al.*, “Observation of a narrow pentaquark state,  $P_c(4312)^+$ , and of two-peak structure of the  $P_c(4450)^+$ ,” [arXiv:1904.03947 \[hep-ex\]](#).
- [3] R. L. Jaffe, “Perhaps a Stable Dihyperon,” *Phys. Rev. Lett.* **38** (1977) 195–198. [Erratum: *Phys. Rev. Lett.* 38,617(1977)].
- [4] T. Sakai, K. Shimizu, and K. Yazaki, “ $H$  dibaryon,” *Prog. Theor. Phys. Suppl.* **137** (2000) 121–145, [arXiv:nucl-th/9912063 \[nucl-th\]](#).
- [5] G. R. Farrar and G. Zaharijas, “Nuclear and nucleon transitions of the  $H$  dibaryon,” *Phys. Rev.* **D70** (2004) 014008, [arXiv:hep-ph/0308137 \[hep-ph\]](#).
- [6] G. A. Cowan, D. C. Craik, and M. D. Needham, “*RapidSim: an application for the fast simulation of heavy-quark hadron decays*,” *Comput. Phys. Commun.* **214** (2017) 239–246, [arXiv:1612.07489 \[hep-ex\]](#).
- [7] M. Morgenthaler, “*Studies of Hexaquark missing mass approximation*,” Jan, 2019. not published.
- [8] H. Yukawa, “On the Interaction of Elementary Particles I,” *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48–57. [*Prog. Theor. Phys. Suppl.* 1,1(1935)].
- [9] P. W. Demtröder, *Kern-, Teilchen- und Astrophysik Experimentalphysik IV*. Springer Verlag, 5th editio ed., 2017.
- [10] “Webpage: Standard model of elementary particles.” [https://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg). Accessed: 2019-01-09.
- [11] M. Gell-Mann, “A Schematic Model of Baryons and Mesons,” *Phys. Lett.* **8** (1964) 214–215.
- [12] D. Diakonov, V. Petrov, and M. V. Polyakov, “Exotic anti-decuplet of baryons: Prediction from chiral solitons,” *Z. Phys.* **A359** (1997) 305–314, [arXiv:hep-ph/9703373 \[hep-ph\]](#).

- 
- [13] **LEPS** Collaboration, T. Nakano *et al.*, “Evidence for a narrow  $S = +1$  baryon resonance in photoproduction from the neutron,” *Phys. Rev. Lett.* **91** (2003) 012002, [arXiv:hep-ex/0301020](#) [hep-ex].
- [14] C. G. Wohl, “Pentaquarks,” *Phys. Lett.* **B667** (2008) 1124–1125.
- [15] G. R. Farrar, “Stable Sexaquark,” [arXiv:1708.08951](#) [hep-ph].
- [16] **LHCb** Collaboration, R. Aaij *et al.*, “LHCb Detector Performance,” *Int. J. Mod. Phys.* **A30** no. 07, (2015) 1530022, [arXiv:1412.6352](#) [hep-ex].
- [17] “Webpage: LHCb detector along the bending plane.”  
<https://commons.wikimedia.org/wiki/File:Lhcbview.jpg>.  
Accessed: 2019-01-11.
- [18] **LHCb VELO Group** Collaboration, T. Latham, “Performance of the LHCb Vertex Locator,” *Phys. Procedia* **37** (2012) 830–837.
- [19] **LHCb RICH Group** Collaboration, M. Adinolfi *et al.*, “Performance of the LHCb RICH detector at the LHC,” *Eur. Phys. J.* **C73** (2013) 2431, [arXiv:1211.6759](#) [physics.ins-det].
- [20] “Webpage: LHCb Magnet - Technical Design Report.”  
<http://lhcb-magnet.web.cern.ch/lhcb-magnet/TDR/html/lhcb-magnet.html>. Accessed: 2019-01-13.
- [21] **LHCb** Collaboration, P. Perret, “First Years of Running for the LHCb Calorimeter System,” *PoS TIPP2014* (2014) 030, [arXiv:1407.4289](#) [physics.ins-det].
- [22] F. Soomro and P. Campana, “The LHCb experiment,” *Scholarpedia* **11** no. 7, (2016) 32452. revision #153343.
- [23] “Webpage: Large Hadron Collider beauty experiment - Muon system.”  
<https://lhcb-public.web.cern.ch/lhcb-public/en/detector/Muon2-en.html>. Accessed: 2019-01-13.
- [24] A. A. Alves, Jr. *et al.*, “Performance of the LHCb muon system,” *JINST* **8** (2013) P02022, [arXiv:1211.1346](#) [physics.ins-det].
- [25] T. Head, “The LHCb trigger system,” *JINST* **9** (2014) C09015.
- [26] R. Aaij *et al.*, “Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC,” [arXiv:1812.10790](#) [hep-ex].
- [27] **LHCb HLT project** Collaboration, J. Albrecht *et al.*, “Performance of the LHCb High Level Trigger in 2012,” *J. Phys. Conf. Ser.* **513** (2014) 012001, [arXiv:1310.8544](#) [hep-ex].

## REFERENCES

---

- [28] R. Aaij *et al.*, “Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC,” [arXiv:1812.10790 \[hep-ex\]](#).
- [29] “Webpage: DaVinci Physics Analysis Application.” <https://twiki.cern.ch/twiki/bin/view/LHCb/DaVinci>. Accessed: 2019-03-21.
- [30] “Webpage: The Stripping Project - Stripping versions.” <http://lhcbdoc.web.cern.ch/lhcbdoc/stripping/>. Accessed: 2019-04-11.
- [31] “Webpage: ROOT Data Analysis Framework - User’s Guide.” <https://root.cern.ch/root/html/doc/guides/users-guide/ROOTUsersGuide.html>. Accessed: 2019-04-15.
- [32] “Webpage: TupleToolMCBackgroundInfo.” <https://twiki.cern.ch/twiki/bin/view/LHCb/TupleToolMCBackgroundInfo>. Accessed: 2019-03-23.
- [33] **Particle Data Group** Collaboration, M. Tanabashi *et al.*, “Review of Particle Physics,” *Phys. Rev.* **D98** no. 3, (2018) 030001.
- [34] “Webpage: beef - A generic B-fitter based on roofit.” <https://gitlab.cern.ch/sneubert/beef>. Accessed: 2019-03-29.
- [35] W. Verkerke, “RooFit,” *EPJ Web of Conferences* **4** (04, 2010) .
- [36] A. Hoecker *et al.*, “TMVA - Toolkit for Multivariate Data Analysis,” [arXiv:physics/0703039 \[physics.data-an\]](#).
- [37] L. Anderlini *et al.*, “The PIDCalib package,” Tech. Rep. LHCb-PUB-2016-021. CERN-LHCb-PUB-2016-021, CERN, Geneva, Jul, 2016. <https://cds.cern.ch/record/2202412>.
- [38] A. Poluektov, “Correction of simulated particle identification response in LHCb using kernel density estimation,” Tech. Rep. LHCb-INT-2017-007. CERN-LHCb-INT-2017-007, CERN, Geneva, May, 2017. [https://svnweb.cern.ch/cern/wsvn/lhcbdocs/Notes/INT/2017/007/drafts/LHCb-INT-2017-007\\_v0r0.pdf](https://svnweb.cern.ch/cern/wsvn/lhcbdocs/Notes/INT/2017/007/drafts/LHCb-INT-2017-007_v0r0.pdf).
- [39] A. Poluektov, “Correction of simulated PID response with variable transformation.” <https://indico.cern.ch/event/572934/contributions/2318786/attachments/1357558/2052812/pidcorr.pdf>. Accessed: 2019-04-14.
- [40] “Webpage: Unbinned PID resampling using kernel density estimation.” <https://twiki.cern.ch/twiki/bin/view/LHCb/MeerkatPIDResampling>. Accessed: 2019-01-25.

- 
- [41] M. Pivk and F. R. Le Diberder, “*SPlot: A Statistical tool to unfold data distributions*,” *Nucl. Instrum. Meth.* **A555** (2005) 356–369, [arXiv:physics/0402083](#) [[physics.data-an](#)].
- [42] S. Stahl, *Measurement of CP asymmetry in muon-tagged  $D^0 \rightarrow K^-K^+$  and  $D^0 \rightarrow \pi^- \pi^+$  decays at LHCb*. PhD thesis, Heidelberg U., 2014-06-02.
- [43] A. Hocker *et al.*, “*TMVA - Toolkit for Multivariate Data Analysis*,” [arXiv:physics/0703039](#) [[physics.data-an](#)].
- [44] G. Punzi, “*Sensitivity of searches for new signals and its optimization*,” *eConf* **C030908** (2003) MODT002, [arXiv:physics/0308063](#) [[physics](#)]. [,79(2003)].
- [45] S. Neubert, “*Statistical Methods in Particle Physics - Lecture Notes*.” <https://uebungen.physik.uni-heidelberg.de/vorlesung/20182/smipp>. Accessed: 2019-03-13.
- [46] J. Neyman and E. Pearson, “*On the Problem of the Most Efficient Tests of Statistical Hypotheses*,” *Philosophical Transactions of the Royal Society, A* **231** (01, 1933) 289–337.

## Acknowledgments

First of all, I want to thank Sebastian Neubert for allowing me to do my Bachelor thesis in his group and sharing his knowledge with me even though he did the difficult split between his family and his work. Furthermore I want to thank Marian Stahl for always having time to help me fix problems when the framework was acting strange again, Nicola Skidmore for volunteering to read through my written work to check for grammatical monstrosities, Alessio Piucci for lifting my spirits with the cake he brought for no reason whatsoever as well as the complete working group in general for the insightful weekly discussions. Thank you Nils Hoyer, Bernd Mumme and Nicola Ackermann for the great atmosphere in our working group, speaking words of encouragement to each other when we were collectively sighing about new problems in our projects. I'm grateful for Finja Reichardt, who's always able to help me forget my stress, supporting me in every way she could when I worked into the night again. I appreciate every single one of "das Lerngruppe", walking with me through the everyday life of the university since the first semester. Finally, I of course want to thank my family for supporting me in every decision I make and for standing by my side unconditionally.

## **Erklärung**

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 15.04.2019,