



university of
groningen

faculty of mathematics
and natural sciences

van swinderen institute for
particle physics and gravity

MASTER'S THESIS

Using the $B^0 \rightarrow K^\pm \pi^\mp$ channel as normalization in the search for the charged lepton flavor violating $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay

Author:
G.S.P. WIERSEMA

Supervisor:
Dr. Ir. C.J.G. ONDERWATER

STUDENT NR. 2033178
G.S.P.WIERSEMA@STUDENT.RUG.NL

Co-Assessor:
Dr. J.G. MESSCHENDORP

MSc. QUANTUM UNIVERSE
VAN SWINDEREN INSTITUTE
UNIVERSITY OF GRONINGEN

April 30, 2017

Abstract

This master's thesis describes the use of the $B^0 \rightarrow K^\pm \pi^\mp$ channel as normalization in the search for the charged lepton flavor violating $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay. Charged lepton flavor violating decays are heavily suppressed in the Standard Model and are therefore used to probe for New Physics. Using LHC Run I data, collected by the LHCb detector in the years 2011 and 2012, this analysis determines the $B^0 \rightarrow K^\pm \pi^\mp$ channel's yield, and the efficiencies of the LHCb detector's acceptance, trigger, reconstruction&selection and particle identification. We find a value for the normalization factor α_{norm} of $(3.91 \pm 0.21) \cdot 10^{-12}$, respectively, $(1.69 \pm 0.06) \cdot 10^{-12}$ for the 2011 and 2012 samples, which are used to determine the $\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp) = \alpha_{norm} \cdot \frac{N_{sig}}{\epsilon_{sig}}$ and $\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp) = \alpha_{norm} \cdot \frac{f_d}{f_s} \cdot \frac{N_{sig}}{\epsilon_{sig}}$ branching fractions, and identify the yield determination as the largest contributor to the uncertainty on α_{norm} for both 2011 and 2012.

Keywords: *LHCb, Normalization, Charged lepton flavor violation*

Contents

1	Introduction	5
1.1	Motivation	6
1.1.1	Branching fractions	6
1.1.2	Normalization channels	8
1.1.3	Connections with previous research	9
1.2	Strategy	10
1.3	Outline	13
2	Theory	14
2.1	The Standard Model	14
2.1.1	Leptons	14
2.1.2	Quarks	15
2.1.3	Bosons	16
2.2	$B^0 \rightarrow K^\pm \pi^\mp$ decay mode	17
2.3	Hadronization fraction $\frac{f_s}{f_d}$	19
2.4	Charged lepton flavor violation	21
2.5	$B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay mode	23
3	LHCb	25
3.1	LHC	25
3.2	Detector	26
3.2.1	Tracking system	28
3.2.2	Cherenkov detectors	29
3.2.3	Calorimeters	30
3.2.4	Muon stations	31
4	Data flow	33
4.1	Trigger	33
4.1.1	Level 0 Trigger	33
4.1.2	High Level Trigger	35
4.2	Reconstruction	37
4.3	Selection	38
4.4	Particle identification	38
4.5	Samples	39
4.5.1	Data	40
4.5.2	Simulation	41

5 Yield	43
5.1 Invariant mass spectrum	43
5.2 Correcting mass hypotheses	47
5.3 Binning	49
5.4 Maximum likelihood fit	50
5.4.1 Combinatorial background shape	51
5.4.2 $B^0 \rightarrow K^\pm \pi^\mp$ peak shape	52
5.4.3 $B_s^0 \rightarrow K^\pm \pi^\mp$ peak shape	55
5.4.4 Partially reconstructed background	57
5.4.5 Resulting total fit	59
6 Efficiencies	62
6.1 Acceptance	63
6.2 Reconstruction&selection	64
6.3 Trigger	64
6.4 Particle identification	66
6.4.1 Pre and post-PID cut yield fit	66
6.4.2 Extracting the PID efficiency without knowing the signal's shape	67
6.4.3 PIDCalib	68
6.4.4 Performance histograms	68
6.4.5 Global PID weight determination	69
7 Results	73
7.1 Acceptance efficiency	73
7.2 Trigger efficiency	73
7.3 Reconstruction&Selection efficiency	75
7.4 Particle identification efficiency	75
7.5 Yield	75
7.6 Normalization factor α_{norm}	76
8 Discussion	78
8.1 Comparison of results	78
8.1.1 Comparison between 2011 and 2012	78
8.1.2 Comparison with previous $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis	80
8.2 Limitations	81
9 Summary and outlook	83
10 Acknowledgments	85

11 Appendix I: Mass defect caused by failing to include the π^0 when reconstructing $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ events	86
References	96

1 Introduction

The Standard Model describes all elementary particles and relates them to each other through their interactions; it is the best theoretical explanation currently available of the physical phenomena that are observed in experiments. Yet, it is only a *model*, because we know it not to be perfect. Several observations have been made that show the limitations of the Standard Model. For example, the Standard Model cannot explain the large amounts of *Dark Matter* and *Dark Energy* that seem to be present in the galaxy (Clowe et al., 2006). Moreover, although the Standard Model seems to describe particle interactions well in the lower energy scales (up to a couple of TeV/c^2), it is yet to be tested whether the Standard Model still holds in the higher energy ranges as well (Chakdar, 2016).

Physics not covered by the Standard Model is generally referred to as *New Physics*. New Physics models attempt to explain theoretically the New Physics phenomena that are observed in experiments and, in doing so, predict the existence of other, yet unobserved New Physics processes. Where New Physics becomes important and the Standard model breaks down, and the validity of models attempting to explain New Physics, can be tested different ways. On the one hand, one can test whether the Standard Model holds in the higher energy ranges, by comparing experimental observations of specific processes to the theoretical predictions made by the Standard Model. On the other hand, experiments can probe directly for New Physics, by searching for New Physics processes whose existence is predicted by New Physics models.

Experiments of the second category generally focus on a specific type of processes predicted by New Physics models. Namely, processes that the Standard Model predicts to be non-existent or heavily suppressed, yet one or more New Physics models predict to occur (relatively) frequently. The high frequency with which the New Physics models predict the process to occur makes the process more likely to be observed (if the New Physics models are correct) and the high suppression or non-existence according to the Standard Model makes the observation of the process more likely to yield significant evidence of New Physics (and not be a statistical fluctuation from the Standard Model's predictions).

One such process is the charged lepton flavor violating $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay, where the subscript (s) denotes that both the $B_s^0 \rightarrow e^\pm \mu^\mp$ and the $B^0 \rightarrow e^\pm \mu^\mp$ decays will be studied (the B^0 meson is also often denoted as B_d or B_d^0 , to depict its d quark content). This thesis describes the normalization process involved in the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis searching for the existence of this decay.

1.1 Motivation

Flavor changing processes of neutral leptons (i.e. neutrinos) are generally accepted to exist and can be observed from solar neutrinos, which are produced as electron neutrinos during hydrogen fusion in the sun, but arrive at earth as 65% muon and tau neutrinos and 35% electron neutrinos. This is the case because the electron neutrinos oscillate into the other neutrinos flavors; muon neutrinos and tau neutrinos. Neutrino oscillations typically take place over lengths of a 100 kilometers or more, making them unobservable under normal laboratory conditions, but its effects are clearly present in cosmic observations. The Standard Model was modified to include this effect by giving neutrinos mass, which allows them to oscillate (McDonald, Klein, & Wark, 2003). Moreover, in the quark sector, flavor changing transitions are also observed, which is described by the CKM matrix.

The existence of flavor changing processes of neutral leptons, which was initially unforeseen by the Standard Model, and the flavor changing “analogue” of quarks, as according to the CKM matrix, has given rise to investigations into the existence of flavor changing processes of charged leptons (i.e. electrons, muons and taus). One such investigation is the search for the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay.

1.1.1 Branching fractions

The search for the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay is a New Physics experiment of the “second category” described the introduction. Namely, the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay is heavily suppressed in the Standard Model, but predicted to occur relatively frequently by certain New Physics models (Kappert, 2016). Because of its heavy suppression in the Standard Model, the observation of this decay has the potential to show strong evidence for Standard Model breakdown and the existence of New Physics processes. In particle physics, the frequency with which a process occurs is typically defined as a *branching fraction*, denoted, for the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay, as $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$ (where, again, the subscript (s) denotes both the branching fractions $\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp)$ and $\mathcal{B}(B_d^0 \rightarrow e^\pm \mu^\mp)$). The branching fraction of a certain decay is defined as the fraction of the time the mother particle decays according to this specific decay mode, any time the mother particle decays (according to any decay mode). In other words, the branching fraction of a certain decay is the probability, whenever the mother particle decays, that it decays according to this specific decay.

There is always a discrepancy between the upper and the lower limit on a branching fraction determined by any experiment, which is dictated by

the analysis' accuracy and the level of statistical significance at which the researchers are able to determine upper and lower limits of the branching fraction. Moreover, to set a lower limit larger than zero on the branching of a certain decay, the decay must actually be observed in the experiment. This might seem trivial, but it is actually very relevant for the for the $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$ analysis.

Setting upper and lower limits on the branching fractions of New Physics processes, such as the $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$ decay, is important for the search for New Physics, as they can be used to reject New Physics models predicting branching fractions that fall outside of the range dictated by the upper and lower limits. Moreover, they guide theoretical and experimental physicists in the right direction of where to look for New Physics.

Given the number of $B_s^0 \rightarrow e^\pm \mu^\mp$ decays detected, e.g. by the LHCb detector, the branching fraction $\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp)$ can be calculated from

$$\mathcal{B}(B_s \rightarrow e^\pm \mu^\mp) = (\mathcal{L}^{LHCb} \cdot \sigma_{pp \rightarrow b\bar{b}} \cdot f_s \cdot 2)^{-1} \cdot \frac{N_{B_s^0 \rightarrow e^\pm \mu^\mp}}{\epsilon_{B_s^0 \rightarrow e^\pm \mu^\mp}}, \quad (1)$$

where \mathcal{L}^{LHCb} is the total integrated luminosity of the LHCb detector (which, for 2012, was equal to $4 \cdot 10^{32} \text{cm}^{-2}\text{s}^{-1}$ proton-proton collisions), $\sigma_{pp \rightarrow b\bar{b}}$ the cross-section of $b\bar{b}$ pairs production in proton-proton collisions, and f_s the probability that a b quark hadronizes into a B_s^0 meson. Moreover, $N_{B_s^0 \rightarrow e^\pm \mu^\mp}$ is the number of detected $B_s^0 \rightarrow e^\pm \mu^\mp$ decays, and $\epsilon_{B_s^0 \rightarrow e^\pm \mu^\mp}$ the efficiency with which these decays are detected. The factor 2 in equation (1) comes from the fact that the $b\bar{b}$ pair consists of a b quark and an anti b quark, which can hadronize into an anti B_s^0 meson and a B_s^0 meson, respectively. Because the neutral B mesons and their antimeson counterparts have the same weight, we cannot distinguish between them in the final fit to the mass spectrum, since the yield fit simply determines the number of $e^\pm \mu^\mp$ pairs that decayed from “something” that had exactly the B_s^0 invariant mass (in other words, the mother particle of $e^\pm \mu^\mp$ decay is identified based on its mass). Moreover, neutral B mesons oscillate into a superposition of their meson and antimeson states, making the distinction between the two states even more ambiguous. Therefore, we do not distinguish between the neutral B mesons and antimesons in the branching fraction $\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp)$, such that it is really the sum of the branching fractions of both B_s^0 mesons and anti B_s^0 mesons decaying into $e^\pm \mu^\mp$ pairs.

Equivalently, from the number of $B^0 \rightarrow e^\pm \mu^\mp$ decays detected by the LHCb detector, the branching fraction $\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp)$ can be calculated from

$$\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp) = (\mathcal{L}^{LHCb} \cdot \sigma_{pp \rightarrow b\bar{b}} \cdot f_d \cdot 2)^{-1} \cdot \frac{N_{B^0 \rightarrow e^\pm \mu^\mp}}{\epsilon_{B^0 \rightarrow e^\pm \mu^\mp}}, \quad (2)$$

where f_d denotes the hadronization fraction of b quarks into B^0 mesons.

Figure 1 illustrates how, starting from the proton-proton collision, the electron-muon pair is finally produced (for the $B_s^0 \rightarrow e^\pm \mu^\mp$ decay), illustrating the relevance of the factors $\sigma_{pp \rightarrow b\bar{b}}$ and f_s in equation (1). Furthermore, the dotted lines in figure 1 indicate other decays than the ones relevant for the cascade ending in the $B_s^0 \rightarrow e^\pm \mu^\mp$ channel, or decay products that are not detected (giving rise to the efficiency ϵ).

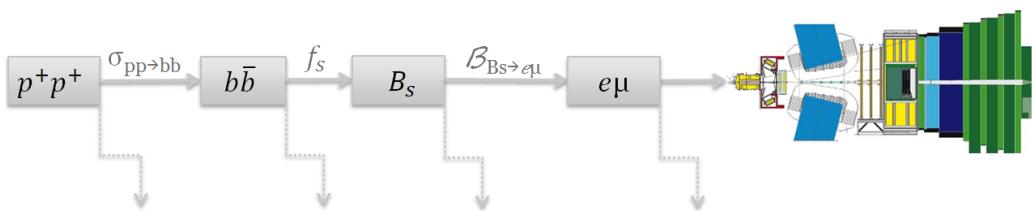


Figure 1: The cascade from proton-proton collision to the final $B_s^0 \rightarrow e^\pm \mu^\mp$ decay

Equations (1) and (2) might seem like straightforward calculations. However, the uncertainties on both the luminosity \mathcal{L}^{LHCb} and the $\sigma_{pp \rightarrow b\bar{b}}$ production cross-section are quite large, limiting the accuracy in determining the number of produced $B_{(s)}^0$ mesons. Therefore, we need an alternate approach to determine the number of $B_{(s)}^0$ mesons, which will be explained in the next section.

1.1.2 Normalization channels

To be precise, the branching fraction defined above is the *absolute* branching fraction, or, in other words, the branching fraction of a decay relative to all decay modes of the mother particle taken together (including the decay itself). Sometimes, it is more useful to use a *relative* branching fraction; a branching fraction relative to the branching fraction of another specific decay mode, which is typically a decay mode of the same mother particle. Let us call the decay mode of which we wish to determine the relative branching fraction the *signal channel*, and the decay mode against which we determine this relative branching fraction the *normalization channel* (also often referred to as the *calibration channel*). Then, the relative branching is determined from the number of signal channel decays, divided by the number of normalization channel decays.

Unless we are specifically after some *relative* branching fraction (because, for example, some New Physics model specifically predicts the relative branching fraction and not the absolute one), the concept of relative

branching fractions might seem trivial. Namely, if the absolute branching fraction of the normalization channel is known, the number of normalization channel decays can be used to determine how many of its mother particle were produced. Then, assuming that the signal and normalization channels have the same mother particle (if this is not the case, the proportion in which both mother particles are produced is also needed), the absolute branching fraction of the signal channel can be determined by dividing the number of signal channel decays by the number of produced mother particles. Hence, the importance of the direct relation between the signal and normalization channels and the relevance of the relative branching fraction might seem unclear.

However, because experiments are always limited by both statistics and systematics, it can be valuable to think of the relative branching fraction as more than just a “proportion” needed to determine the absolute branching fraction of the signal channel. Namely, a favorable choice of normalization channel has the potential of canceling systematic errors in the signal channel against those in the normalization channel to some degree, if the physics and measuring systematics are sufficiently similar.

Another consideration for the choice of normalization channel is the actual magnitude of its branching fraction and the resulting decay rate in the LHCb detector. Namely, if the decay rate is too small, too little data will be collected to reach sufficient statistical significance, whereas if the decay rate is too large, when the trigger of the LHCb detector is set up to store this decay, it will be overloaded and the detection of more rare decays will be suppressed (this will be explained further in section 4.1).

With these considerations in mind, the $B^0 \rightarrow K^\pm\pi^\mp$ decay channel, a two-body decay of a B^0 meson, was chosen as the normalization channel for the $B_{(s)}^0 \rightarrow e^\pm\mu^\mp$ analysis, because of its similar topology to the $B_s^0 \rightarrow e^\pm\mu^\mp$ and $B^0 \rightarrow e^\pm\mu^\mp$ channels (all three decay modes consist of a neutral B meson decaying into two relatively light, oppositely charged and long lived particles) and appropriate decay rate in the detector.

1.1.3 Connections with previous research

The previous analysis did not observe the $B_{(s)}^0 \rightarrow e^\pm\mu^\mp$ decay (Aaij et al., 2013). Thus, no lower limit on the $B_{(s)}^0 \rightarrow e^\pm\mu^\mp$ could be set, because no evidence was observed of the existence of the decay mode. Because the decay was not observed, the upper limit on the branching fraction (i.e. the maximum branching fraction the decay could realistically have without being observed in the analysis) that was set, was entirely dependent on the analysis’ accuracy and statistical significance level at which the researchers were able

to set the limit.

The current $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis aims to improve upon the methods used in the previous analysis to produce more accurate results. Although the current analysis is not expected to observe the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decays, the possibility does exist, as research methods have improved since the previous analysis. Moreover, because the current $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis aims to be more rigorous than its predecessor, if the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay is not observed, its existence can be rejected with a higher degree of certainty and the upper limit on the branching fraction can be lowered.

The previous analysis set the upper limit on the branching fraction of the $B_s^0 \rightarrow e^\pm \mu^\mp$ decay at $1.1(1.4) \cdot 10^{-8}$ at a confidence level of 90% (95%) and the branching fraction of the $B^0 \rightarrow e^\pm \mu^\mp$ decay at $2.8(3.7) \cdot 10^{-9}$ at a confidence level of 90% (95%) (Aaij et al., 2013).

The current analysis, using Run I data from the LHC runs in 2011 and 2012, is expected to lower the branching fraction of the $B_s^0 \rightarrow e^\pm \mu^\mp$ decay to somewhere in the 10^{-9} ranges and the branching fraction of the $B^0 \rightarrow e^\pm \mu^\mp$ decay to somewhere in the 10^{-10} ranges (within one order of magnitude from the previous results). However, the aim is to develop an analysis methodology that will straightforwardly translate to the (post-upgrade) Run II data from the LHC runs in 2015, which will soon become available. Moreover, with the data from 2015, the upper limits on the branching fractions are expected to be lowered by multiple orders of magnitude.

1.2 Strategy

The goal of the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis is to determine the branching fractions $\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp)$ and $\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp)$, and the used normalization channel is the $B^0 \rightarrow K^\pm \pi^\mp$ decay mode. Thus, to determine $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$, the signal and normalization channels' yields (i.e. the number of decays found in the fit to the mass spectrum) need to be determined. Moreover, to know how many decays actually took place, we also need to know the *efficiency* with which the decays in the yield are detected (e.g. an efficiency of 25% indicates that for every decay in the final yield fit, on average, 3 more decays took place but did not end up in the final yield fit). Additionally, we need to know the (absolute) branching fraction of the $B^0 \rightarrow K^\pm \pi^\mp$ decay, which was accurately determined by the “*B*-factories” Belle (Duh et al., 2013) and BaBar (Aubert et al., 2007).

Moreover, because the B_s^0 signal channel has a different mother particle than the B^0 signal and normalization channels, we also need to know the proportion in which B_s^0 and B^0 mesons are produced. This fraction is known

as $\frac{f_s}{f_d}$ and has been extensively studied (Koopman, 2015). The quantities discussed above can be used to calculate $\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp)$, according to the following relation:

$$\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp) = \mathcal{B}_{norm} \cdot \frac{f_d}{f_s} \cdot \frac{\frac{N_{sig}}{\epsilon_{sig}}}{\frac{N_{norm}}{\epsilon_{norm}}} = \mathcal{B}_{norm} \cdot \frac{f_d}{f_s} \cdot \frac{\epsilon_{norm}}{N_{norm}} \cdot \frac{N_{sig}}{\epsilon_{sig}}, \quad (3)$$

where, the subscripts *norm*, for *normalization*, and *sig*, for *signal*, indicate the $B^0 \rightarrow K^\pm \pi^\mp$ and $B_s^0 \rightarrow e^\pm \mu^\mp$ decays, respectively, and ϵ denotes the detection efficiency. The relation between the different quantities in equation (3) is illustrated in figure 2, where the dotted lines indicate either decay products that are “missed” in the analysis, giving rise to detection (in)efficiencies, or decays that are not part of the decay cascades ending in the $B_s^0 \rightarrow e^\pm \mu^\mp$ or $B^0 \rightarrow K^\pm \pi^\mp$ decays.

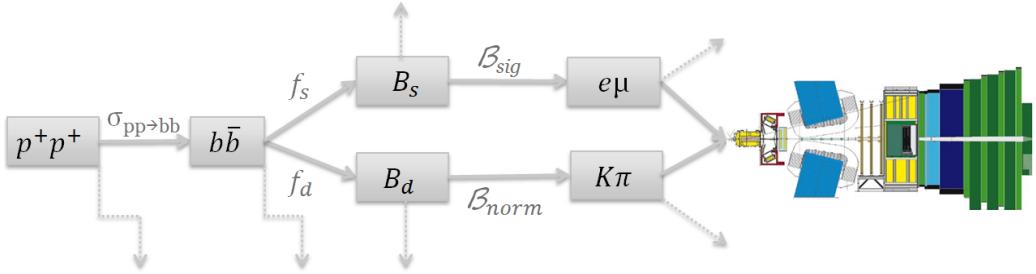


Figure 2: The relation between the signal and normalization channels

Equivalently, the branching fraction $\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp)$ can be calculated according to

$$\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp) = \mathcal{B}_{norm} \cdot \frac{\frac{N_{sig}}{\epsilon_{sig}}}{\frac{N_{norm}}{\epsilon_{norm}}} = \mathcal{B}_{norm} \cdot \frac{\epsilon_{norm}}{N_{norm}} \cdot \frac{N_{sig}}{\epsilon_{sig}}, \quad (4)$$

where, in this case, *sig* indicates the $B^0 \rightarrow e^\pm \mu^\mp$ signal channel.

It is common practice to introduce the normalization factor α , such that

$$\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp) = \alpha \cdot N_{sig}, \quad (5)$$

where

$$\alpha = \mathcal{B}_{norm} \cdot \frac{f_d}{f_s} \cdot \frac{\epsilon_{norm}}{N_{norm}} \cdot \epsilon_{sig} \quad (6)$$

for the $B_s^0 \rightarrow e^\pm \mu^\mp$ signal channel, and

$$\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp) = \alpha \cdot N_{sig}, \quad (7)$$

with

$$\alpha = \mathcal{B}_{norm} \cdot \frac{\epsilon_{norm}}{N_{norm}} \cdot \epsilon_{sig} \quad (8)$$

for the $B^0 \rightarrow e^\pm \mu^\mp$ signal channel.

Furthermore, because this thesis is concerned with the $B^0 \rightarrow K^\pm \pi^\mp$ normalization channel specifically, it is also convenient to introduce a second normalization factor, α_{norm} , defined as

$$\alpha_{norm} = \mathcal{B}_{norm} \cdot \frac{\epsilon_{norm}}{N_{norm}}, \quad (9)$$

such that

$$\mathcal{B}(B_s^0 \rightarrow e^\pm \mu^\mp) = \alpha_{norm} \cdot \frac{f_d}{f_s} \cdot \frac{N_{sig}}{\epsilon_{sig}}, \quad (10)$$

and

$$\mathcal{B}(B^0 \rightarrow e^\pm \mu^\mp) = \alpha_{norm} \cdot \frac{N_{sig}}{\epsilon_{sig}}. \quad (11)$$

Thus, the aim of the thesis is to explain how α_{norm} is determined. As can be seen from equation (9), α_{norm} consists of the $\mathcal{B}(B^0 \rightarrow K^\pm \pi^\mp)$ branching fraction, the yield of the $B^0 \rightarrow K^\pm \pi^\mp$ normalization channel and the yield's accompanying detection efficiency. Moreover, the “aggregate” efficiency ϵ is typically subdivided into different subcategories, for different phases of the LHCb data processing flow, namely the acceptance, trigger, reconstruction&selection, and particle identification efficiencies. Section 1.3 outlines the structure of this thesis and describes which sections of this thesis explain the determination methods of the different quantities that go into the calculation of α_{norm} .

1.3 Outline

This thesis will start with a review of particle physics, specifically of the Standard Model, charged lepton flavor violation, the physics of the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ and $B^0 \rightarrow K^\pm \pi^\mp$ channels, and the hadronization fraction $\frac{f_s}{f_d}$, in chapter 2. This chapter will be followed by a description of the experimental setup of the LHC and LHCb detector, in chapter 3, and explains the different (sub)detection systems of the LHCb detector in detail. Subsequently, chapter 4 describes the stepwise process of converting detector hits in the LHCb detector into manageable data formats ready for analysis. The preparation of the data and the implementation of the fit to the invariant mass spectrum is explained in chapter 5. Then, the determination of the acceptance, trigger, reconstruction&selection, and particle identification efficiencies are discussed in chapter 6. The results of this study (i.e. the found yields and efficiencies) are presented in chapter 7, which are subsequently discussed in chapter 8. Finally, chapter 9 concludes this thesis with a summary of what was discussed before and an outlook on what is to follow.

2 Theory

This section describes the theory relevant to the understanding of the physics behind, and significance of the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis. We start with a brief description of the standard model, after which we explain the physics behind the $B^0 \rightarrow K^\pm \pi^\mp$ channel and the $\frac{f_s}{f_d}$ hadronization fraction. We conclude this chapter with a discussion of charged lepton flavor violation and the physics behind the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay.

2.1 The Standard Model¹

The Standard Model describes the elementary particles and how they interact with each other. The illustration of the Standard Model in figure 3 summarizes the explanation of the Standard Model that will now follow. The Standard Model is a quantum field theory in which the elementary particles are described as excitations of the corresponding fields. The model distinguishes between fermions, which are spin- $\frac{1}{2}$ particles, and bosons, which have spin 1.

The massive fermions make up the matter around us and come in three “generations”; the first, second and third. Moreover, the fermions can be further subdivided into quarks and leptons, with each generation consisting of two quarks and two leptons. Additionally, all fermions have a corresponding anti-particle with opposite internal quantum numbers, such that quantum numbers are conserved when a fermion is created or annihilated together with its corresponding anti-particle.

2.1.1 Leptons

The Standard Model distinguishes between three different leptons with negative unitary charge (and the three corresponding anti-particles with positive unitary charge), of which the electron e is the most well-known. The electron is also the lightest of the three charged fermions, and is located in the first generation. The two remaining charged fermions are, in increasing order of generation and mass, the muon μ and the tauon τ .

For each charged lepton, there is a corresponding neutral lepton, called neutrino, which is part of the same generation as its charged counterpart. In increasing order of generation, the neutrinos are called electron neutrino ν_e , muon neutrino ν_μ , and tauon neutrino ν_τ . Thus, the lepton generation

¹Unless stated otherwise, the information in this section can be found in Morrison (2010) and Bettini (2014).

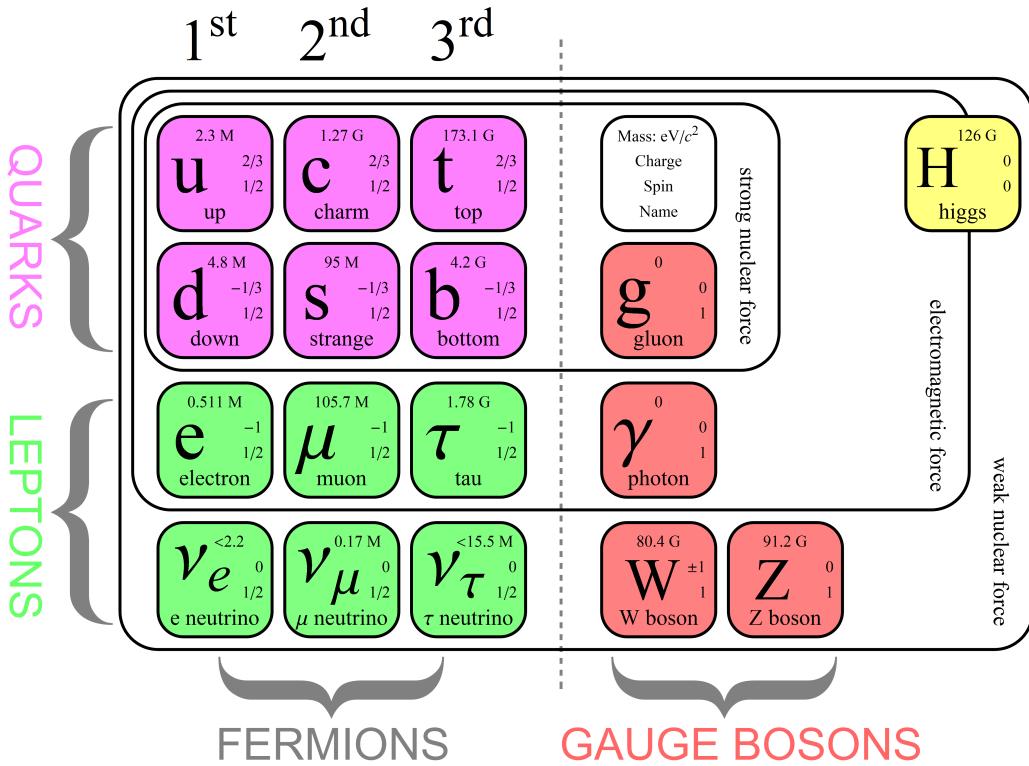


Figure 3: The Standard Model

douplets are as follows:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}. \quad (12)$$

2.1.2 Quarks

Quarks, like leptons, increase in mass with the generations. Furthermore, each generation has an up-type quark, with charge $+\frac{3}{2}$, and a down-type quark, with charge $-\frac{1}{2}$ (and, as with leptons, their anti-particles have opposite charge). In increasing order of generation, the up-type quarks are the up quark u , the charm quark c and the top quark t , and the down-type quarks are the down quark d , the strange quark s and the bottom quark b . Thus, the quark generation doublets are:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix}. \quad (13)$$

Contrary to leptons, quarks are never observed as free particles; they

hadronize into combinations of two or three quarks, with these combinations of quarks being called hadrons. The three-quark hadrons are called baryons and consist of either three quarks or three anti-quarks. The most common baryons in the universe, because of their stability, are protons and neutrons. The two-quark hadrons are called mesons and consist of a quark and an anti-quark, the most well-known mesons being the pions, which mediate the transition from a proton into a neutron, and vice versa.

2.1.3 Bosons

The bosons are the mediators of forces exerted between the fermions. The strong force is mediated by the massless gluon, the electromagnetic force by the massless photon, and the weak force by the massive, charged \mathcal{W}^\pm boson and the massive, neutral \mathcal{Z}^0 boson. The discovery at the LHC of the massive Higgs boson, which mediates the gravitational force, completed the Standard Model in 2012 (Aad et al., 2012). The bosons mediate the transitions between fermions, where transitions consisting exclusively of leptons are referred to as leptonic processes, those involving both leptons and quarks are referred to as semi-leptonic, and processes consisting of only quarks are referred to as hadronic processes.

The strong force Gluons, the mediators of the strong force, act on a charge called “color”. Therefore, the theory of the strong force is called Quantum *chromodynamics*. Quarks are the only fermions that carry color charge, and are therefore the only fermions that experience the strong force. The property of the strong force that it does not decrease with distance (after a tiny distance of about a hadron diameter), is why quarks cannot exist as free particles, but only as part of color-neutral hadrons. The strong interaction is the fastest interaction and takes place within 10^{-22} seconds.

The electromagnetic force The electromagnetic force is mediated by photons, is of intermediate strength and range (as compared to the strong and weak force), and acts on all charged fermions. Therefore, all fermions but the neutrinos are affected by the electromagnetic force. The electromagnetic force has interaction times between 10^{-20} and 10^{-14} seconds, and the theory of the electromagnetic force is called Quantum *electrodynamics*.

The weak force The weak force, mediated by the \mathcal{W}^\pm and \mathcal{Z}^0 bosons, is best understood using electro-weak theory and is the slowest of the elementary forces, with weak interactions typically taking between 10^{-14} and

10^{-8} seconds to take place. The weak force is the only force that affects all twelve fermions, and the \mathcal{W}^\pm boson has the unique property of being the only boson that can change an up-type quark into a down-type quark, and vice versa, while transferring the change in charge to another transition of opposite charge discrepancy. Furthermore, the \mathcal{W}^\pm boson is also the only boson that can change the generation of a quark, when changing an up-type quark into a down-type quark or vice versa. The coupling constants of the transitions between up-type and down-type quarks can be found from the Cabibbo-Kobayashi-Maskawa matrix,

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.9743 & 0.2253 & 0.0035 \\ 0.2252 & 0.9734 & 0.0412 \\ 0.0087 & 0.0404 & 0.9994 \end{pmatrix}, \quad (14)$$

where $|V_{\alpha\beta}|^2$ is proportional to the transition probability for an α ($= u, c, t$) quark to transition into a β ($= d, s, b$) quark or vice versa. The CKM matrix is strongly diagonal, which means that the weak force prefers to conserve the generation, when changing up-type quarks into down-type quarks and vice-versa. Decays consisting exclusively of quark generation conserving transitions are referred to as *Cabibbo allowed* (or Cabibbo favored), and decays involving a quark transition that does not conserve generation are referred to as *Cabibbo suppressed* (or double or triply Cabibbo suppressed, when consisting of two or three quark transitions that do not conserve generation, respectively).

Another unique property of the \mathcal{W}^\pm boson is that it can change a charged lepton into the corresponding neutrino from the same generation and vice-versa, and create or annihilate lepton-antilepton pairs of a charged lepton and a neutrino from the same generation. However, contrary to quarks, these transitions all have the same coupling constant, according to *Lepton Universality*. Namely, as cross-generational transitions involving charged leptons are not allowed, there is no CKM matrix for leptons (or it would be a diagonal matrix of ones). A mixing matrix similar to the CKM matrix does exist for neutrinos, called the PMNS matrix, for Pontecorvo Maki Nakagawa Sakata matrix, which describes the mixing of weak and mass eigenstates of the neutrinos.

2.2 $B^0 \rightarrow K^\pm \pi^\mp$ decay mode

The $B^0 \rightarrow K^\pm \pi^\mp$ decay is a hadronic decay, and an allowed decay within the Standard Model. The decay consists of a B^0 meson decaying into a positively

charged kaon and a negatively charged pion². The B^0 meson contains a d quark and an anti- b quark, the kaon a u quark and an anti- s quark and the pion an anti- d quark and an u quark. In the simplest decay, the Feynman tree diagram (a Feynman diagram not containing loops) shown in figure 4 (b), the anti- b quark transitions into an anti- u quark, which together with the (spectator) d quark forms a kaon. The \mathcal{W}^\pm boson that is emitted in this transition subsequently form a u , anti- d quark-pair, i.e. a pion.

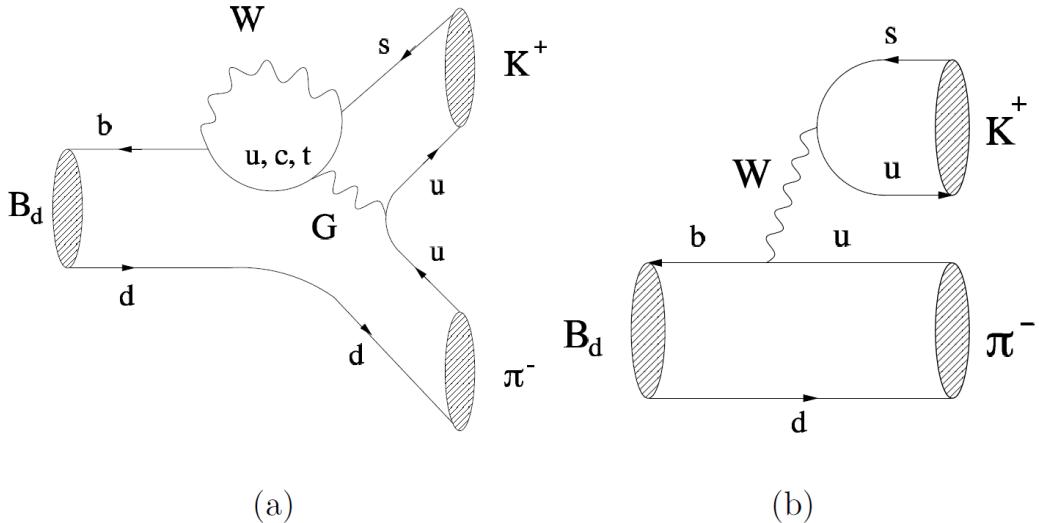


Figure 4: Feynman diagrams of the most important $B^0 \rightarrow K^\pm\pi^\mp$ decay modes (with the B^0 meson listed as B_d). Figure 4 (a) depicts the dominant penguin diagram and figure 4 (b) shows the dominant tree diagram (Buckley, 2006)

However, as can be seen from the Feynman diagram in figure 4 (b), the $B^0 \rightarrow K^\pm\pi^\mp$ decay is doubly Cabibbo suppressed, as it includes both an up-bottom and an up-strange quark transition, which both do not conserve quark generation. Therefore, the penguin diagram in figure 4 (a) actually forms the dominant contribution to the $B^0 \rightarrow K^\pm\pi^\mp$ branching fraction (Buckley, 2006).

Penguin diagrams are one-loop processes, where a quark temporarily transitions into another quark type via a \mathcal{W}^\pm boson, during which the changed quark undertakes some tree (often strong) interaction, after which the changed quark undergoes another transition via the same \mathcal{W}^\pm boson it emitted in the

²The reason we list the decay channel as $B^0 \rightarrow K^\pm\pi^\mp$, is because we do not distinguish between the B^0 meson and its anti-meson counterpart \bar{B}^0 , and the \bar{B}^0 anti-meson decays into negatively charged pion and a positively charged kaon, i.e. $\bar{B}^0 \rightarrow K^-\pi^+$.

Table 1: Estimates of $\mathcal{B}(B^0 \rightarrow K^\pm \pi^\mp)$

Value · 10^6	Year	Detector
$20.00 \pm 0.34 \pm 0.60$	2013	Belle
$19.1 \pm 0.6 \pm 0.6$	2007	BaBar
$18^{+2.3+1.2}_{-2.1-0.9}$	2003	Cleo
19.6 ± 0.5	2015	Particle Data Group

Note: This table shows the estimates, and the experiments from which they originate, on which the Particle Data Group's quoted value of the branching fraction of the $B^0 \rightarrow K^\pm \pi^\mp$ decay is based.

first place, either back into the quark type it started as, or into another, third type of quark. Because the quark first emits, and subsequently absorbs the same \mathcal{W}^\pm boson, this is referred to as a \mathcal{W}^\pm loop. Typically, such a loop suppresses the amplitude of the diagram, as compared to a regular tree diagram.

The penguin diagram of figure 4 (a) has a CKM contribution of $|V_{tb}V_{ts}|$ (singly Cabibbo suppressed), whereas the tree diagram of figure 4 (b) has a CKM contribution of $|V_{ub}V_{us}|$ (doubly), and

$$\frac{|V_{ub}V_{us}|}{|V_{tb}V_{ts}|} \approx 0.02\%. \quad (15)$$

Therefore, even though penguin diagrams are loop-suppressed, the penguin diagram of figure 4 (a) forms the dominant contribution to the $B^0 \rightarrow K^\pm \pi^\mp$ decay rate.

We use the branching fraction quoted by the Particle Data Group of $(1.96 \pm 0.05) \cdot 10^{-5}$, which is based on the measurements from Belle (Duh et al., 2013), BaBar (Aubert et al., 2007) and Cleo (Bornheim et al., 2003) listed in Table 1.

2.3 Hadronization fraction $\frac{f_s}{f_d}$

The hadronization fraction (or fragmentation fraction) $\frac{f_s}{f_d}$ is used to estimate the number of produced B_s^0 mesons, from the B^0 meson yield in the normalization channel. B mesons are produced from $b\bar{b}$ pairs, where one or both of the b quarks hadronize into a $B_{(s)}^0$ or other B meson. The production rate of a specific B meson can therefore be expressed as the product of the $b\bar{b}$ production cross section $\sigma_{b\bar{b}}$ and the probability of a b quark to hadronize into the specific B meson. These latter probabilities are written as f_q , with q denoting the quark with which the b quark is paired in the B mesons. Thus, for the B_s^0 and B^0 mesons, these are written as f_s and f_d , respectively, and their fraction $\frac{f_s}{f_d}$ is the hadronization fraction, i.e. the relative probability

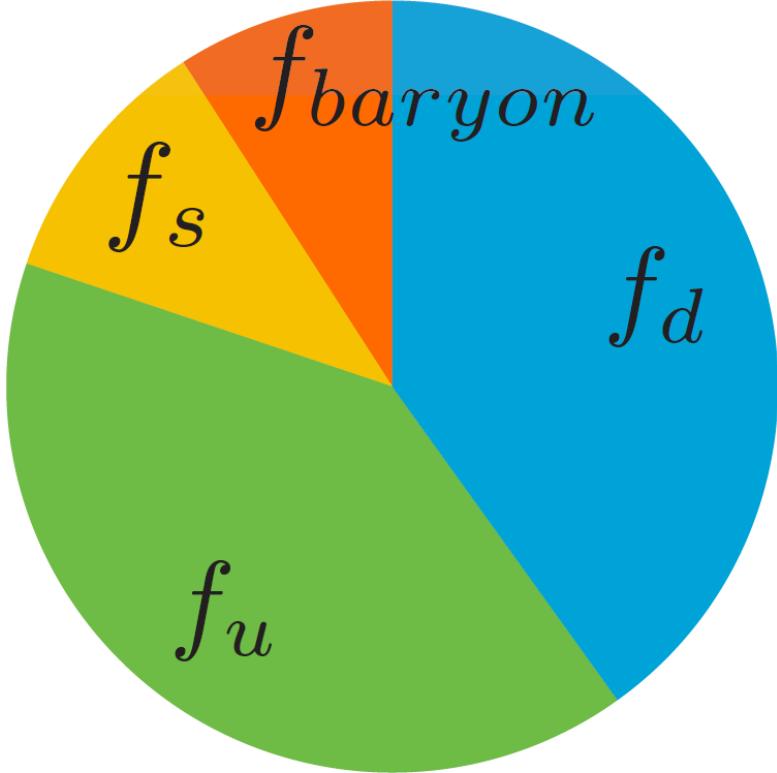


Figure 5: The relative sizes of the hadronization fractions (Tolk, 2016)

for a b quark to hadronize into a B_s^0 meson, with respect to the probability of a b quark to hadronize into a B^0 meson. Moreover, $f_s + f_d + f_u + f_c + f_{baryon}$ must be equal to unity (where f_{baryon} denotes the “hadronization” fraction of a b quark forming a baryon, and f_u the hadronization fraction of a b quark forming a B^+ meson) and f_u is often taken as equal to f_d , because of the small mass difference between the u quark and the d quark (Koopman, 2015). The relative size of the branching fractions are visualized in figure 5.

The dominant production of $b\bar{b}$ pairs is directly from pp collisions (proton-proton collisions). Figure 6 illustrates the most relevant $b\bar{b}$ pair production processes, namely the leading order gluon-gluon fusion processes and the suppressed quark-antiquark process.

The hadronization of a b quark into either a bq meson, or bqq' baryon is mediated by the strong force, where gluons act on the quarks’ color charges. This hadronization process takes place in the non-perturbative regime of

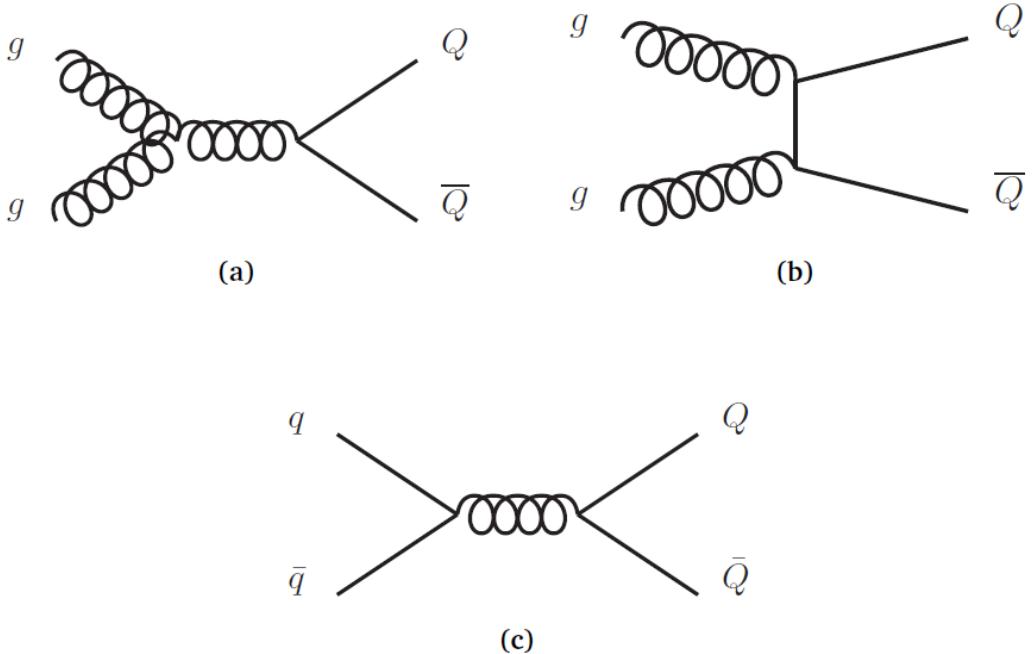


Figure 6: The leading order $b\bar{b}$ production of (a),(b) gluon-gluon fusion and (c) the suppressed quark-antiquark annihilation (Koopman, 2015)

Quantum Chromodynamics, making theoretical predictions on the hadronization probabilities f_q unreliable (Koopman, 2015). Therefore, the hadronization fractions must be determined experimentally. However, the hadronization fractions are not universal; they depend on the production process of the $b\bar{b}$ pairs and the kinematics of the decay (Zonneveld, 2016). Because the LHCb detector operates in another kinematic region than the other LHC detectors, the hadronization fractions found in the other detectors cannot simply be “copied” to the LHCb. Therefore, the hadronization fraction $\frac{f_s}{f_d}$ must be determined specifically for the LHCb, which is discussed in (Zonneveld, 2016).

2.4 Charged lepton flavor violation

In addition to energy, momentum, and charge conservation laws, other conservation laws have been derived from the observation or absence of certain decays. For quarks, there is conservation of the baryon number, which is

expressed as (Morrison, 2010)

$$B = \frac{1}{3} [N(q) - N(\bar{q})], \quad (16)$$

where $N(q)$ denotes the number of quarks, and $N(\bar{q})$ the number of anti quarks. Baryons, which consist of three quarks, thus have baryon number 1, whereas hadrons, consisting of a quark and an antiquark, have baryon number 0.

For leptons, the lepton number describes the conservation of the balance between leptons and antileptons in transitions:

$$L = N(l) - N(\bar{l}), \quad (17)$$

where $N(l)$ denotes the number of leptons and $N(\bar{l})$ the number of antileptons.

For leptons, we will refer to the generation to which they belong as their flavor. Thus, electrons and electron neutrinos have electron flavor L_e , muons and muon neutrinos have muon flavor L_μ and tauons, and tauonneutrinos have tauon flavor L_τ . The formulas for the lepton flavor numbers are therefore:

$$L_e = N(l_e) - N(\bar{l}_e), \quad (18)$$

$$L_\mu = N(l_\mu) - N(\bar{l}_\mu), \quad (19)$$

$$L_\tau = N(l_\tau) - N(\bar{l}_\tau), \quad (20)$$

where $N(l_e)$ and $N(\bar{l}_e)$ denote the number of electron and electron neutrinos, and anti-electrons and anti-electron neutrinos, respectively. Moreover, $N(l_\mu)$ and $N(\bar{l}_\mu)$ denote de number of muons and muon-neutrinos, and anti-muons and anti-muonneutrinos, respectively, and $N(l_\tau)$ and $N(\bar{l}_\tau)$ denote, respectively, the number of tauon and tauonneutrinos, and anti-tauons and anti-tauonneutrinos.

From the absence of certain decays, conservation of the lepton flavor numbers was postulated. However, neutrino oscillations have provided evidence that for neutral leptons (i.e. neutrinos) lepton flavor is not conserved. This has led physicists to believe that the conservation of the flavor of the charged leptons might be violated as well, but evidence of this has yet to be observed. Violation of the principle of conservation of the flavor of charged leptons is referred to as charged lepton flavor violation, and is the topic of the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis, to which this thesis contributes.

Within the Standard Model, the change of flavor of charged leptons is only possible via an intermediate neutrino, which oscillates into another flavor. However, this must take place over very short distances, whereas the distances over which neutrinos typically oscillate are very large. This makes the probability for this charged lepton flavor changing process to occur extremely small, with branching fractions in the order of about 10^{-40} , which is far too small to be observed. An example of such a (Standard Model allowed) neutrino oscillation mediated charged lepton flavor changing process is the charged lepton flavor violating $B_s^0 \rightarrow e^+ \mu^-$ decay illustrated in figure 7.

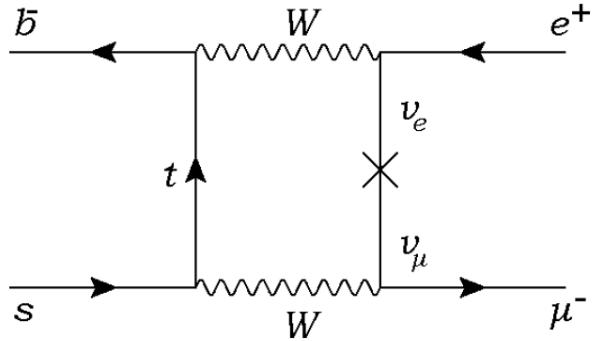


Figure 7: The neutrino oscillation mediated charged lepton flavor violating $B_s^0 \rightarrow e^+ \mu^-$ decay. Kappert (2016)

However, there exist New Physics models that propose mechanisms that can greatly increase the branching fractions of certain charged lepton flavor violating decays. Among these are Supersymmetry, Leptoquarks and Leptogenesis. For an extensive discussion of such models, we refer to Vos (2015).

2.5 $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay mode

The $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay consists of a $B_{(s)}^0$ meson, containing an s or d quark and an anti- b quark (or a b quark and anti s or d quark), decaying into a positron and a muon (or an electron and an anti-muon). Within the Standard Model, this can only occur via an intermediate electron neutrino, which “carries away” the electron flavor from the positron. This electron neutrino then needs to oscillate into an muon neutrino, and “deliver” the muon-flavor to the muon. This neutrino oscillation, which is required to make the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay possible within the standard model, heavily suppresses the branching fraction $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$. This Standard Model allowed neutrino oscillation mediated charged lepton flavor changing process is illustrated in figure 7.

However, as explained in section 2.4, certain New Physics models predict other charged lepton flavor changing mechanisms, that make it possible for the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay to occur with much larger branching fractions. One such mechanism is that of leptoquarks, which is illustrated in figure 8. For more details on how Leptoquarks, Leptogenesis, Seesaw models, Supersymmetry, and other New Physics models explain how the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ decay can take place, and how much larger the predicted $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ branching fractions are than the heavily suppressed Standard Model allowed neutrino oscillation mediated branching fraction (which is in the order of about 10^{-40}), we refer to Kappert (2016).

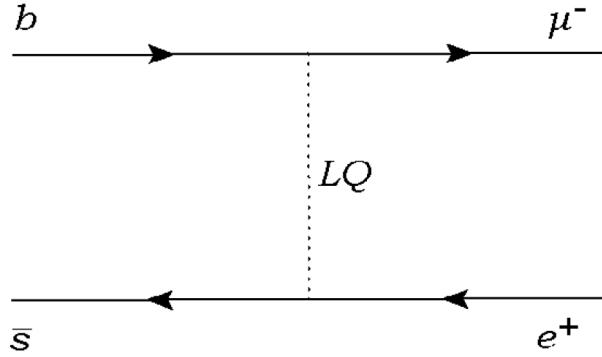


Figure 8: The leptoquark mediated charged lepton flavor violating $B_{(s)}^0 \rightarrow e\mu$ decay. Kappert (2016)

3 LHCb

The experimental setup used in this analysis was not specifically designed or constructed to perform the experiments necessary for the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis. Namely, this analysis relies on the data produced by the “Large Hadron Collider beauty” (LHCb) detector, which is used for many different b -physics analyses. Moreover, the LHCb is part of the Large Hadron Collider (LHC), which is used for an even wider range of experiments. This section starts by discussing the Large Hadron Collider, after which it focuses specifically on the LHCb detector and its sub-detectors.

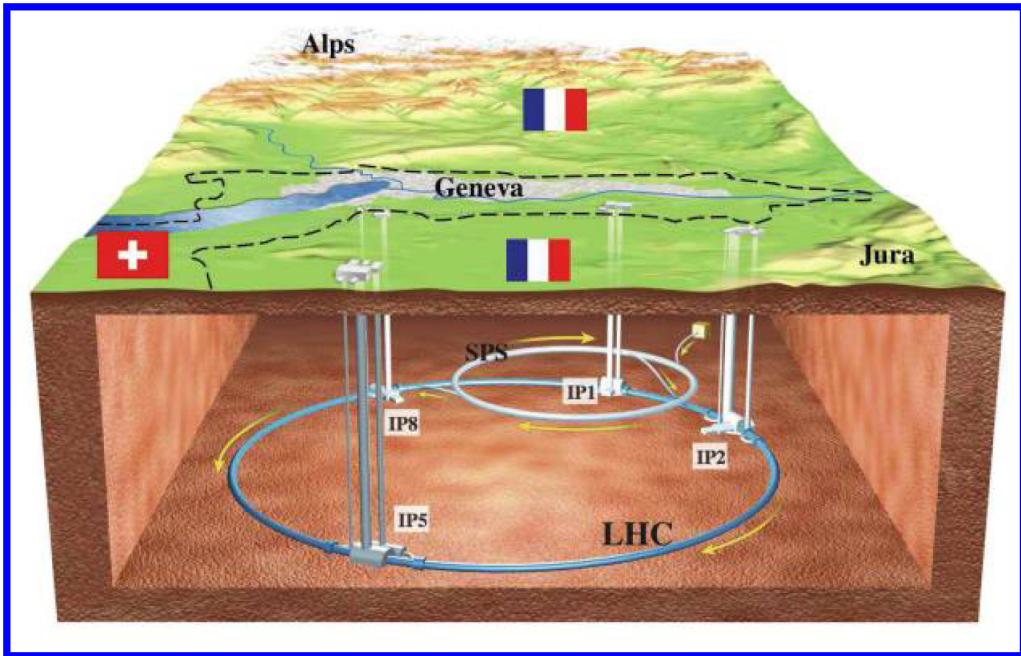


Figure 9: Location of the LHC underneath the French and Swiss territories (Myers, 2013)

3.1 LHC

The LHC is a proton-proton accelerator and collider built by the European Organization for Nuclear Research (CERN) and is located in the vicinity of Geneva, Switzerland. The LHC is constructed inside of a 27 kilometer long circular tunnel, which lies about a hundred meters underground, mostly beneath French territory, which is illustrated in figure 9. The same tunnel previously housed the Large Electron-Positron collider (LEP), but is now used to accelerate two beams of protons in opposite directions.

The disadvantage of the LHC, as compared to LEP, is that it needs to transport positively charged particles in both directions. This means that two separate magnetic fields are needed to bend the two proton beams, in opposite directions, into circular trajectories. On the other hand, LEP used a single magnetic field to constrain the electrons, in one direction, and positrons, in the other direction, to a circular motion. However, the advantage of the LHC is that the two beams do not meet, except at specific locations, where the protons are made to collide. This is an advantage that single ring accelerators, such as LEP, do not have (Myers, 2013). To bend the proton beams into a circular trajectory, the LHC uses more than a thousand dipole magnets and, to keep the beams focused, it uses almost 400 quadrupole magnets.

Before being injected into the LHC, the protons are first accelerated in a sequential process of passing through several linear and circular accelerators, in order to successively increase the protons' energy. The protons start in a linear accelerator, called LINAC2, after which they are passed through to four successive circular accelerators, which, in order, are called the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and Super Proton Synchrotron (SPS), which finally injects the protons into the LHC (Myers, 2013). The two separate proton beams are made to collide at several locations, where different types of detectors are built to observe the diverse types of collisions and decay products produced. The four main experiments are the ATLAS, CMS, LHCb, and ALICE detectors. Figure 10 depicts the different elements of the LHC that were just discussed.

3.2 Detector

The LHCb is a spectrometer, built in a cone-like shape in one direction around the LHC proton beams, with the apex of the cone coinciding with the location where the two proton beams are brought to collide. Its conic shape means it is tuned to detect decays that produce particles that have most of their momentum in the forward direction. Moreover, because the LHCb is constructed in one direction around the beam, it can only detect such decays in one of the two directions. The LHCb detector can track particles in the range from 10 mrad to 300 mrad in the horizontal plane, and in the range from 10 mrad to 250 mrad in the vertical plane (Tolk, 2016). This is illustrated in figure 11. The two proton beams are slightly defocused at the collision point to lower the luminosity (to about $4 \cdot 10^{32} \text{cm}^{-2}\text{s}^{-1}$ in 2012). The LHCb detector is about 5 meters high, 20 meters long, and is located about a 100 meters underground. The detector consists of several sub-detectors, with the VELO (VERtex LOcator), tracker turicensis (TT),

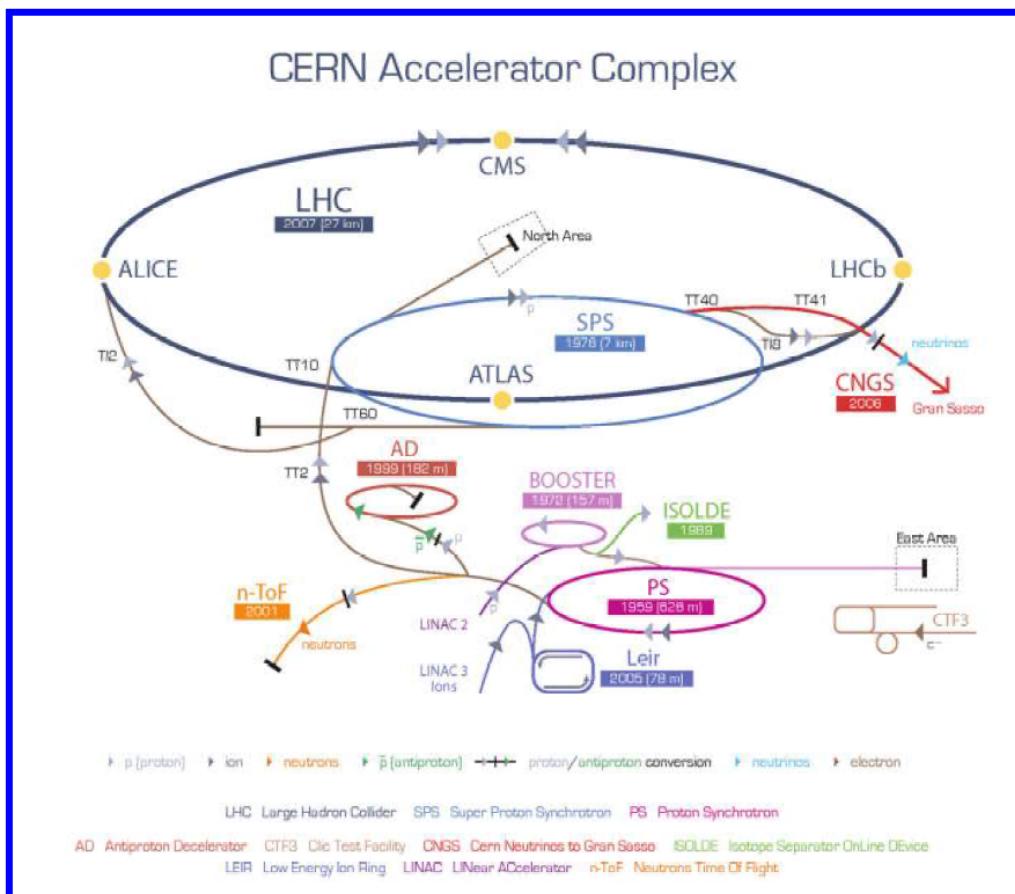


Figure 10: The different components of the Large Hadron Collider. Myers (2013)

inner tracker (IT), and outer tracker (OT) making up the tracking system, and the cherenkov detectors, calorimeters, and muon stations making up the particle identification system. The different sub-detectors are depicted in figure 12. These sub-detectors will be discussed in detail below.

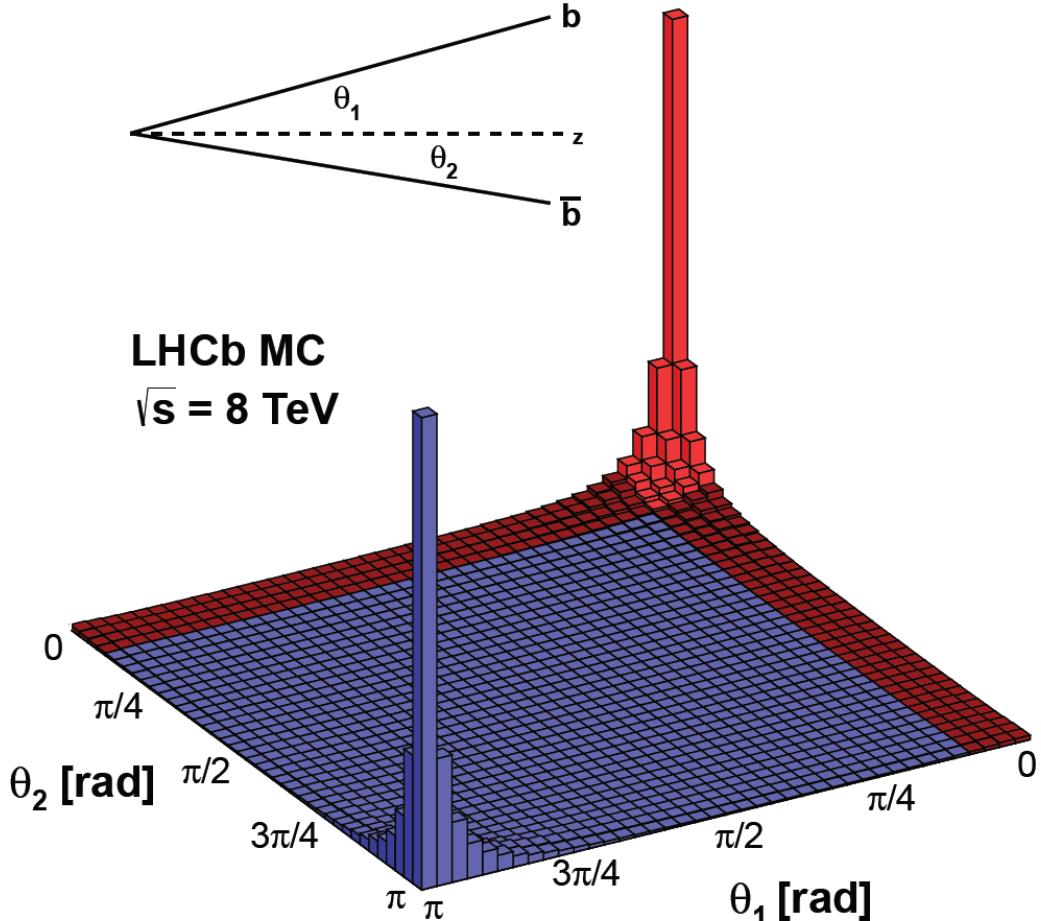


Figure 11: The LHCb detector acceptance with respect to the angular distribution of the produced $b\bar{b}$ pairs (Tolk, 2016)

3.2.1 Tracking system

The tracking system consists of the VELO, TT, IT, and OT. The VELO, TT, and IT are semiconductor trackers, consisting of narrow, doped silicon strips. These silicon modules produce small ionization currents when charged particles pass through them, which can be measured to determine where the particle passed through the detector material. Silicon trackers have a

much higher resolution, but are also more costly than trackers based on older technologies, such as gaseous straw-tube detectors like the OT. Gaseous ionization detectors consist of tubes that are filled with gas, which becomes ionized when charged particles pass through them. A wire goes down the center of the tube, which has a potential difference from walls of the tube, such that when the gas in the tube become ionized, the ionized electrons drift in one direction and the ions in the other. This induces a current that can be measured to determine that a particle passed trough the gas tube detector. The reason that gaseous straw-tubes were chosen over silicon modules for the OT is simply for cost-saving purposes; because the OT is located the farthest away from the apex of the cone-shaped LHCb detector, it has the largest lever arm (distance to the proton-proton collision point), which means it can have a coarser resolution and still measure the track angles with high accuracy. Moreover, it needs to cover the largest surface area out of the four trackers (as the cross-section of a cone is larger the farther away from its apex). The larger cross section means that both the number of detector hits per unit area (per event) is smaller and more detector material is needed to cover the entire surface area. Thus, more, but less accurate detector material is needed, which is why the cheaper gaseous-straw tubes were chosen.

The VELO is positioned in a cylindrical setup around the collision point, and has the purpose of determining the location of the proton-proton collision vertex and secondary vertices of b -hadron decays. The TT and IT are located, respectively, in front of and behind a dipole magnet that bends the trajectories of the particles depending on their charge. This allows for the distinction between charged and neutral particles, as the neutral particles' tracks remain straight in the magnet's field. Moreover, the sign of the particles' charges can be determined from the direction of deflection and, as all charged leptons and hadrons have unitary charge, the particles' momenta can be determined from the magnitude of deflection in the magnetic field, as measured by the TT and IT. The OT is the last tracker before the calorimeters and has the goal of determining the exact position of where particles enter the system of calorimeters, as well as improving momenta estimates, by contributing to the post-magnet track angles determined by the IT.

3.2.2 Cherenkov detectors

The ring imaging cherenkov (RICH) detectors, together with the calorimeters and muon stations, make up the particle identification system. The RICH detectors use of the principle of Cherenkov radiation to identify particles. Cherenkov radiation is emitted when a particle travels through a transparent medium with a higher speed than the phase velocity of light in this medium,

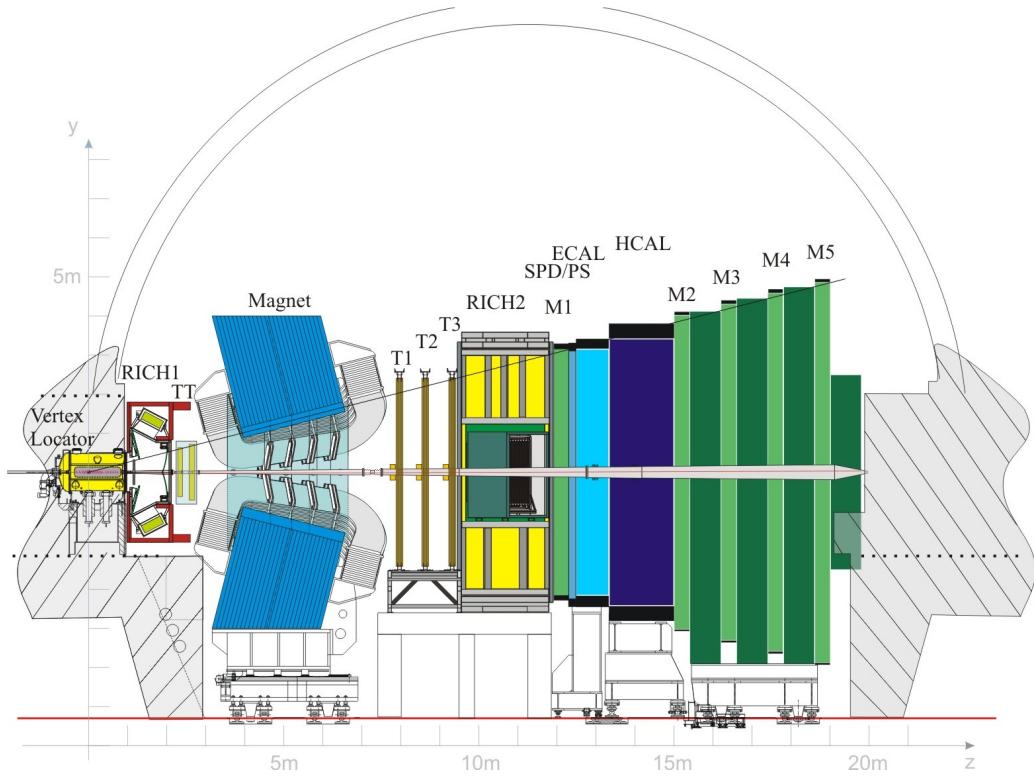


Figure 12: The different subdetectors of the LHCb detector (*LHCb Wikipedia webpage*, 2016)

i.e. if

$$v > \frac{c}{n}, \quad (21)$$

where n is the index of refraction of the traversed medium (Bettini, 2014). The Cherenkov radiation effect generates a shock wave, which is the electromagnetic wave analogue to the sonic boom shock wave produced by a super sonic jet in the atmosphere. The radiation is picked up by the detector and measured, from which, together with the information from the calorimeters, it can be used to determine a particle's identity.

3.2.3 Calorimeters

The hadronic and electromagnetic calorimeters together make up the calorimeter system. In the calorimeters, particles collide with the detector material and decay into showers of decay products, of which the deposited energy in the detector is then determined. Although there is some “cross-contamination” between the two calorimeters, electrons and photons typi-

cally decay in the electromagnetic calorimeter, whereas mesons and baryons typically manage to traverse the entirety of the electromagnetic calorimeter and decay into particle showers in the hadronic calorimeter. This is illustrated in figure 13.

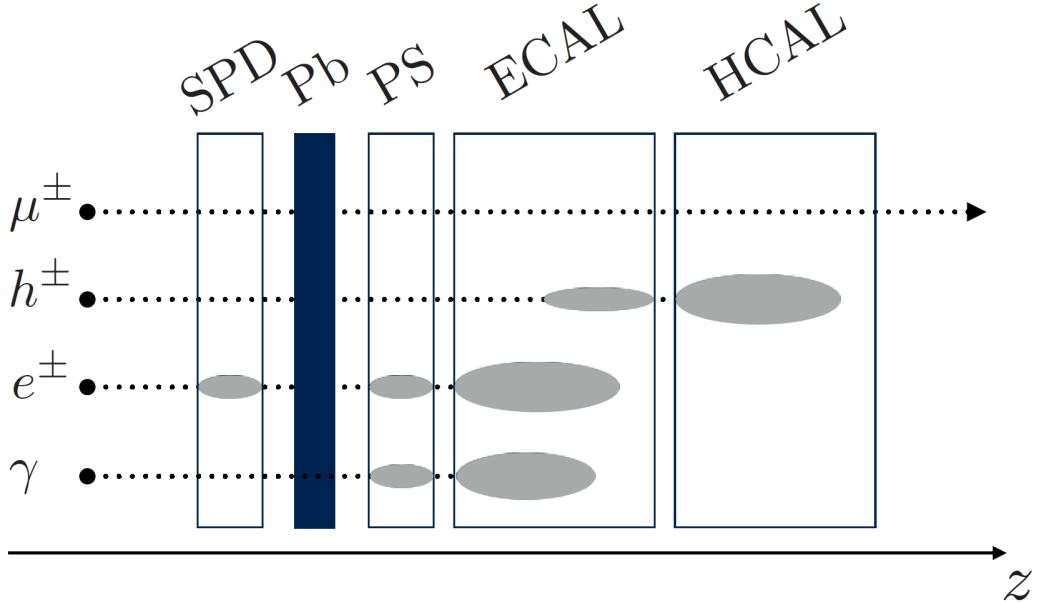


Figure 13: The energy deposits of different particles in the different calorimeter stages (The SPD measures the occupancy of events, Pb denotes a sheet of lead, and the PS determines where particles enter the calorimeter) (Tolk, 2016)

3.2.4 Muon stations

The only particles that manage to “survive” both the electromagnetic and hadronic calorimeters are the muons (and neutrinos, which none of the LHCb’s detectors can detect), which then enter the muon stations. The muon stations work on the premise that anything reaching these detectors, instead of having decayed before doing so, must be a muon (or a neutrino, but they fly right through the muon stations without leaving tracks). The muon detectors consist of thick sheets of metal, iron in the case of the LHCb detector, to select penetrating muons. The sheets of metal are alternated by sheets of detector material, the “muon stations”, which measure the muons’ trajectories. The LHCb consists of a total of five muon stations, one of which is placed in front of the calorimeters (the M1), and the other four are placed

behind the calorimeters (M2-M5). Muon stations M1 to M3 are used to distinguish between positively and negatively charged muons, and M4 and M5 are used to tag false muon candidates.

4 Data flow

Figure 14 shows how data events, starting as (real or simulated) proton-proton collisions and subsequent decays, go through the complete LHCb data processing flow and end up in the final ready-to-process data format called *Ntuple*. This process begins by an event (i.e. a decay, or collection or cascade of decays) being *triggered* and stored, after which it *reconstructed*. Then, the event passes the *selection*, after which the events are stored in Ntuples, and are ready to be analyzed.

4.1 Trigger

Every event that ends up in the data Ntuples starts with a positive trigger decision. The LHCb trigger consists of a collection of hardware and software devices, designed to eliminate as much background noise as possible, while letting as much useful data pass through for storage as possible. The trigger consists of the L0, for Level 0 trigger, the HLT1, for High Level Trigger 1, and HLT2, for High Level Trigger 2. Because the LHCb does not have the infrastructure to store all events that occur, the trigger has the purpose to reduce the data rate as much as possible, while throwing away as few useful events as possible. The L0 trigger reduces the initial interaction rate (events per second) from about 15MHz to about 1 MHz, the HLT1 trigger reduces this further to about 80 kHz, and the HLT2 then finally stores the events at a rate of about 5 kHz (a single event is about a 100 kB) (Kuindersma, 2016). The data rate reduction is illustrated in figure 15.

4.1.1 Level 0 Trigger

The L0 is referred to as the hardware trigger, which means that the trigger is embedded in the hardware of the LHCb detector. The L0 uses dedicated trigger lines to find useful events from muon, hadron, photon and/or electron detector hits. For this, it uses only the fastest subdetectors, namely the calorimeters and muon stations. The L0 consists of three triggers; the pileup trigger, which measures luminosity of the event, the muon trigger, which triggers on muon hits in the muon stations, and the calorimeter triggers, which trigger on electron or photon hits in the electromagnetic calorimeter, or hadron hits in the hadronic calorimeter (Tolk, 2016). A positive trigger decision of any of these three triggers individually is sufficient for the L0 trigger to pass the event through to the HLT1 trigger.

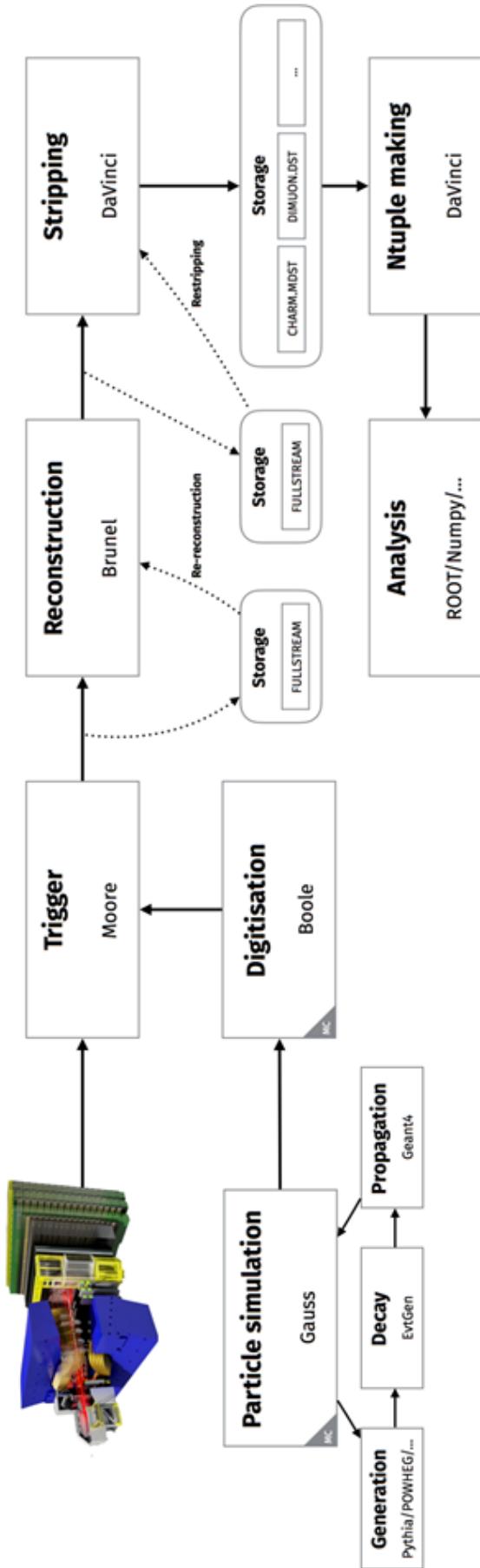


Figure 14: The LHCb dataflow

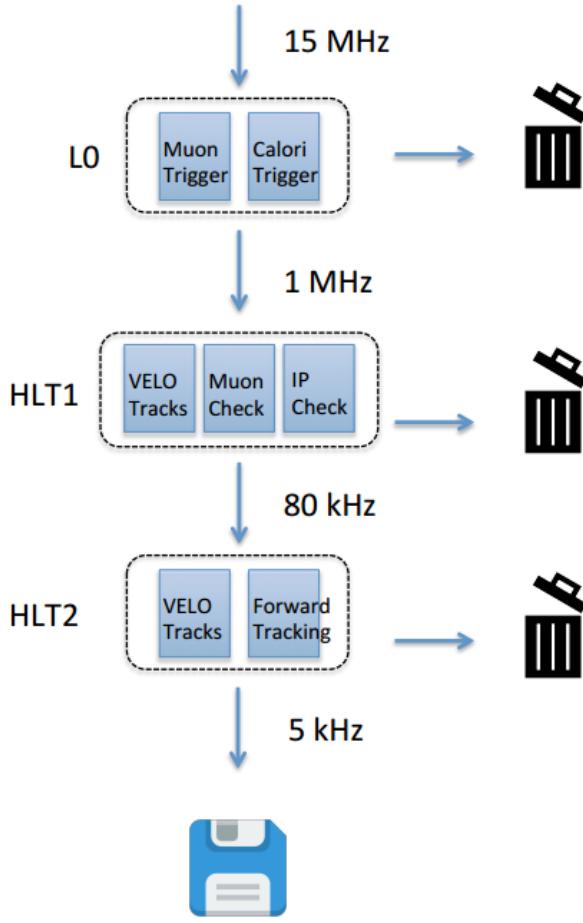


Figure 15: The sequential data rate reduction by the consecutive LHCb triggers (Kuindersma, 2016)

4.1.2 High Level Trigger

The High Level Trigger consists of the HLT1 and HLT2, which are both software-based triggers. The HLT software is called *Moore* and consists of algorithms, mainly written in C++, which run simultaneously on more than 26 thousand multiprocessor computers, which together are called the Event Filter Farm (Kuindersma, 2016). The HLT performs a reconstruction similar to the offline reconstruction, but because the offline reconstruction takes about 2 seconds per event, which is too long given the data rate the HLT needs to deal with, the HLT performs a simplified reconstruction process in two steps, namely the HLT1 and the HLT2.

HLT1 The HLT1 is the first software trigger level, and analyzes events that have been passed through to it by the L0. Because the HLT1 needs to operate very fast, it only partially reconstructs events based on the VELO tracks. Moreover, it only selects the VELO tracks that have the highest probability to originate from a signal decay. For these selected tracks, the HLT1 then matches these tracks to hits in the trackers that follow the magnet, namely the Inner Tracker (IT) and Outer Tracker (OT), in order to determine the track momentum from the deflection in the magnetic field. This matching of VELO tracks to hits in the IT and OT is called *forward tracking*.

HLT2 The second and final trigger level, the HLT2, performs a more complete reconstruction of the event, as it has a smaller event rate to process than the HLT1. Contrary to the HLT1, it performs forward tracking for all VELO tracks with high enough (transversal) momentum, and although the HLT2 is still very time constrained, it performs almost as well as the full, offline reconstruction Tolk (2016). The HLT2’s trigger lines can be divided into three categories:

- **Generic beauty trigger lines** The beauty lines are designed to trigger on partially reconstructed b -hadron decays and take up the majority of the HLT2 bandwidth.
- **Muonic trigger lines** The muon lines trigger on events that have one or two identified muons in the final state, where single muon events are only triggered if the muon has high enough momentum, and di-muon events are only selected if the reconstructed invariant mass of the two muons matches with a specific mother particle mass.
- **Charm trigger lines** Because of the high charm production rates, the charm lines only trigger on events whose final state particles have all been reconstructed. Moreover, they include only trigger lines for very specific charm decays.

Trigger categories The type (i.e. trigger line) of trigger decision is also stored as part of the event. From the trigger line that triggered the event, we can divide all decays in the event up into four different categories; Triggered On Signal (TOS), Triggered Independent from Signal (TIS), simultaneously TIS and TOS, and Triggered On Both (TOB):

- **TOS** For a specific decay, if this and only this decay was responsible for the trigger to trigger and the event to be stored, then this decay was triggered *on signal*, and is labeled TOS.

- **TIS** For a specific decay, if the event containing this decay was stored because another decay in the event was triggered, then this decay (that did not, in any way, contribute to the triggering of the event) was triggered independently from this signal, and is labeled TIS.
- **TISTOS** The two categories Triggered On Signal and Triggered Independently from Signal are not mutually exclusive; it is entirely possible for an event to contain multiple decays that all *independently from each other* match the trigger criteria. From the definitions of the categories of TIS and TOS, it can be understood that such decays are labeled both TIS and TOS, or TISTOS.
- **TOB** For the category TISTOS, it was stressed that all TISTOS decays were sufficient individually to cause the event to be triggered (i.e. the event would still have been triggered if only one of these decays was present). However, in some rare cases, it is possible that none of the decays present in an event were individually sufficient to cause the event to be triggered, but their combined detector hits were, in fact, sufficient to cause the event to be triggered. These decays that contributed together to the trigger are labeled as TOB. However, less than 0.05% of all decays are labeled TOB (Kuindersma, 2016), and will therefore not be paid further attention to.

4.2 Reconstruction

The data that pass the trigger is stored at a rate of about 500MB/s (Kuindersma, 2016). The offline reconstruction software, called *Brunel*, uses all detectors and is better calibrated than the HLT's Moore. The more elaborate reconstruction that Brunel can run, because it does not face the time constraints that Moore does, improves the reconstruction efficiency with about 1-2% per track with respect to the online reconstruction performed by the HLT2. The reconstruction tries to trace back tracks to find initial vertices and daughter particles that originate from the same decay. How well Brunel is able to reconstruct an event, is captured in certain variables that are attached to the events, such as the *impact parameter*, which describes how closely a track approaches the primary vertex from which it is deemed to originate. Brunel then passes the reconstructed event (tracks and corresponding identified particle object) to DaVinci, which subsequently combines the tracks into particles and vertices Tolk (2016).

4.3 Selection

Based on the parameters determined and stored in the reconstruction, the selection aims to further eliminate background noise while keeping as much of the signal data as possible. The selection consist of the stripping and further “Boosted Decision Tree” cuts. The stripping simply discards decays on the basis of cuts on certain variables that best describe the probability of an event being either signal or background. Subsequently, the background can be reduced further by BDT cuts, which are based on a multivariate operator called the Boosted Decision Tree (BDT). The BDT uses a number of different variables to find the best hierarchy in separating signal from background. The difference between the BDT cuts and simple stripping cuts, is that the BDT can select multiple regions in the space defined by the variables the cuts are based on, which results in an almost 20% greater reduction of the background, as compared to even tighter selection cuts with the same efficiency as the BDT cuts (Tolk, 2016). The BDT is “trained” on Monte Carlo samples, which means that best decision hierarchy/tree structure and cuts are found from simulated data that describe, as accurately as possible, the decays for which the BDT will be used. This training is done using a software package called *TMVA*, which stands for Toolkit for MultiVariate Analysis. A schematic illustration of the BDT decision tree can be seen in figure 16.

4.4 Particle identification

The particle identification (PID), using all information available, aims to determine the identity op particles as accurately as possible, according to which the corresponding masses are allocated to the particles. The particle identification uses information from the RICH detectors and calorimeters, to establish ΔLL variables (for Δ Log Likelihood), which give the difference between the logs of the likelihoods of a particle to have two different identities. For example, the $\Delta LL_{K-\pi}$ describes the likelihood that a particle is a kaon with respect to the likelihood of the particle being a pion. ΔLL variables can be graphed into two dimensional diagrams, with the loglikelihoods of the two different identities on the different axes. For a data set containing purely kaon and pions, the $\Delta LL_{K-\pi}$ vs. $\Delta LL_{\pi-K}$ plot will show a diagonal pattern, preferably with the highest concentrations in the top-left and bottom right regions. The more stringent the cut on the ΔLL variable is, the higher is the certainty that the remaining events have a certain identity, but also the lower is the efficiency of the ΔLL cut, i.e. the more signal events are discarded. Figure 17 shows the effect on a scatter plot of $B \rightarrow h^\pm h'^\mp$ events (where h is either a kaon or a pion) of a PID cut that requires that one of the two

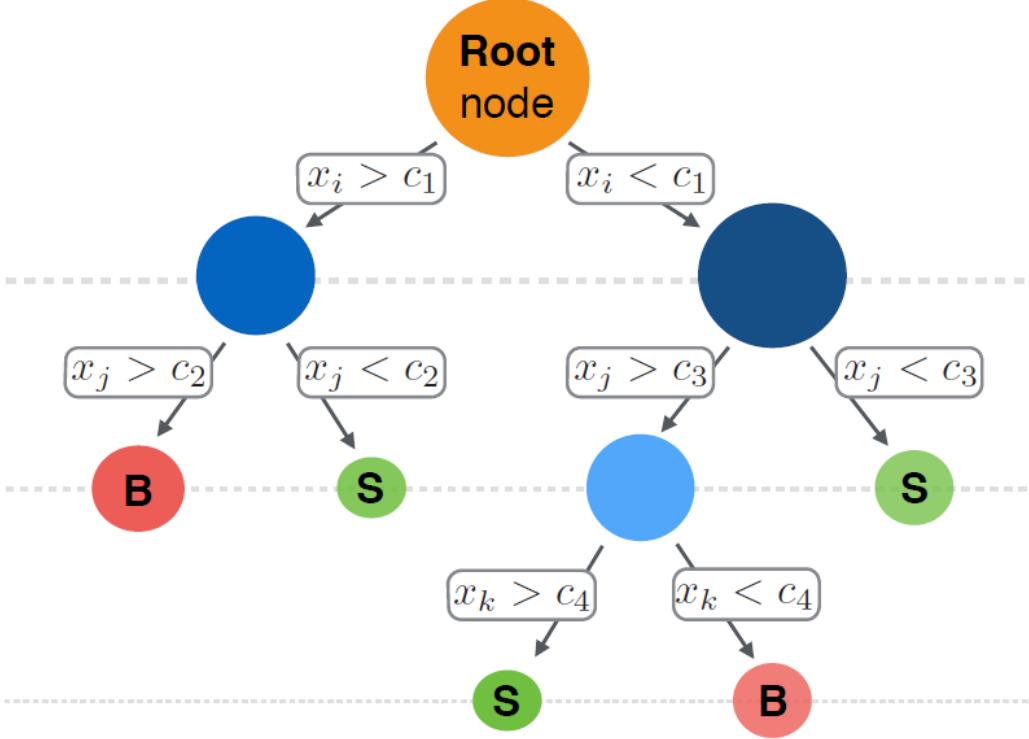


Figure 16: The BDT decision tree. For each node, the variable that best distinguishes between signal and background is used (Tolk, 2016).

particles has $\Delta LL_{K-\pi} > 0$ and the other has $\Delta LL_{K-\pi} < 0$, which is a very “loose” cut requiring that one of the particles is a kaon and the other a pion (i.e. the PID requirements cuts away decays that are likely to have either two kaons or pions in the final state). Furthermore, figure 17 shows the effect of a PID cut that requires that one of the two particles has $\Delta LL_{K-\pi} > 10$ and the other has $\Delta LL_{K-\pi} < -10$, which is a far more stringent cut requiring the final state to consist of a kaon and a pion.

4.5 Samples

Both the yield fit and the determination of the PID-efficiency use both real and simulated data. The real data will be discussed first, after which the simulated data are described.

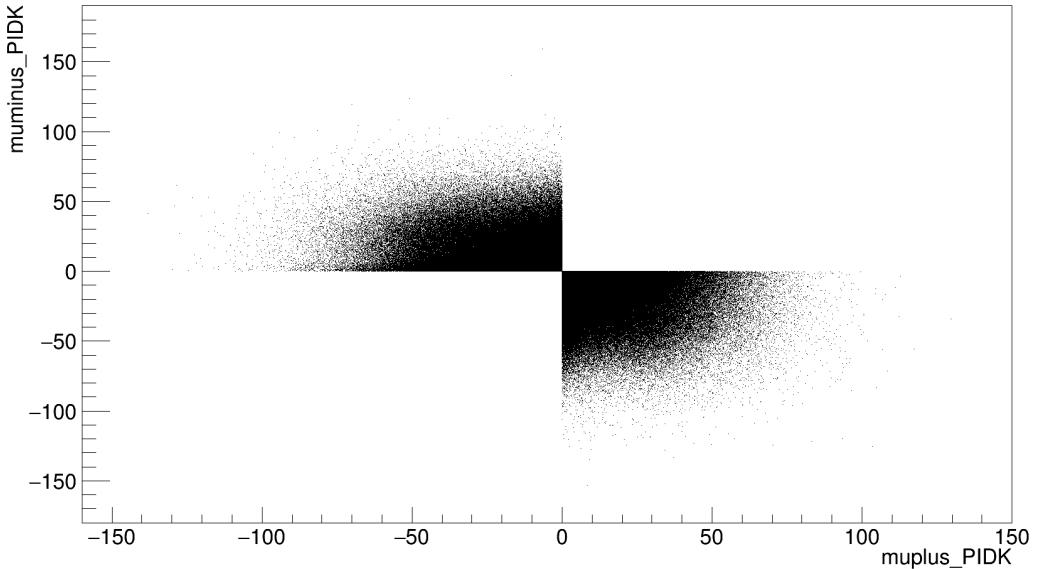


Figure 17: The effect on a scatter plot of $B \rightarrow h^\pm h'^\mp$ events (where h is a kaon or a pion) of a PID cut that requires that one of the two particles has $\Delta LL_{K-\pi} > 0$ and the other has $\Delta LL_{K-\pi} < 0$

4.5.1 Data

This section describes the data for the $B^0 \rightarrow K^\pm \pi^\mp$ normalization channel. The data, consisting of two $B \rightarrow h^\pm h'^\mp$ samples, were collected at two different center of mass energies. During 2011 the data were collected with a center of mass energy of $\sqrt{s} = 7 \text{ TeV}/c^2$. After a technical stop at the beginning of 2012, data collection continued, at a center of mass energy of $\sqrt{s} = 8 \text{ TeV}/c^2$. The data selection has been performed using stripping 21 and no further cuts were applied. The events' invariant masses are initially reconstructed with a PID requirement of $\Delta LL_K > 0$ for the kaon, and $\Delta LL_K < 0$ for the pion³. However, a ΔLL cut of $\Delta LL_K > \pm 10$ has been applied to the data that are used in the final fit, as this was found to offer the best balance between eliminating as much of the (misID) background as possible, whilst retaining as many signal sample events as possible. Moreover, the B mesons' invariant masses reconstructed from pions and kaons have been reconstructed with the pions and kaons still allocated muon masses. Therefore, the reconstructed masses need to be recalculated to compensate for this mass mis-allocation. The reason that the pions and kaons are still allocated

³Henceforth, it is implied that all PID loglikelihoods of kaons and pions are with respect to pions and kaons, respectively, so $\Delta LL_K = \Delta LL_{K-\pi}$, and $\Delta LL_\pi = \Delta LL_{\pi-K}$.

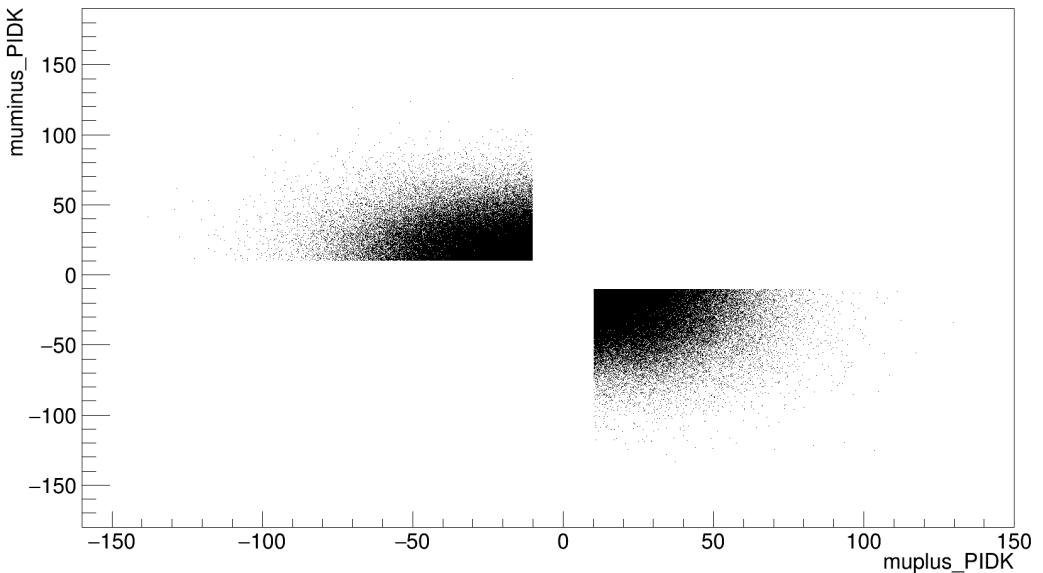


Figure 18: The effect on a scatter plot of $B \rightarrow h^\pm h^\mp$ events (where h is a kaon or a pion) of a PID cut that requires that one of the two particles has $\Delta LL_{K-\pi} > 10$ and the other has $\Delta LL_{K-\pi} < -10$

muon masses is that particle identities are only established with appreciable certainty after applying ΔLL cuts, which is done as part of this analysis and not as part of the (automated) LHCb data selection process. Furthermore, only TIS $B^0 \rightarrow K^\pm \pi^\mp$ events are used, the reason for which is explained in section 6.3.

4.5.2 Simulation

The simulated samples used in the analysis are based on Monte Carlo simulations under the conditions of 2012 data. The events from the Monte Carlo sample, like the real data events, are simulated to go through the identical sequential process from pp collision, to decay, to detector hits, to triggering, reconstruction and selection. The simulation of events occurring in the LHCb is performed with a software package called GAUSS (Tolk, 2016). GAUSS relies on several different software packages to simulate the different stages of the data flow from pp collision to final selection. PYTHIA is used to simulate the pp collision and subsequent hadronization. The output of PYTHIA, in the form of four-vectors of the simulated particles, is passed to EvtGen, which then simulates the different decays of the B mesons. Here, particles that fall outside of the LHCb detector range are discarded to speed up the

simulation, which is an important step for the *acceptance efficiency*, which will be discussed below. The detector hits of these decays are then simulated by GEANT4, which models the interaction of the different decay products with the detector materials. For radiative products, an additional software package is used, which is called PHOTOS. The final step is then the response of the LHCb detection systems to the detector hits simulated by GEANT4. This is performed by BOOLE, after which the simulated events can pass through the same trigger, reconstruction and selection process as real data, as real events have been converted to numerical data at this point as well.

5 Yield

This section describes how the number of $B^0 \rightarrow K^\pm \pi^\mp$ decays is extracted from the invariant mass spectrum, which is heavily polluted by background noise. From the data collected by the LHCb detector, after processing, spectra of the “reconstructed invariant masses” of the events are obtained. Namely, according to a certain “mass hypothesis”, which is the assumption which decay occurred and thus which identifies the detected particles had, masses can be allocated to these detected particles. Then, together with the momentum information from the trackers, the mass of the mother particle can be reconstructed. This yields the reconstructed invariant mass spectrum, in which, typically, one or more peaks can be seen that correspond(s) to the mass(es) of the mother particle(s) of the decays that occurred (and on which to the mass hypothesis/hypotheses is/are based).

5.1 Invariant mass spectrum

The invariant mass spectrum consists of events of which the invariant mass of the mother particle is reconstructed from the masses and momenta of the daughter particles. Because the identity of the mother particle cannot be known a-priori, all decays that yield (exactly) the same daughter particles as the target decay are reconstructed, which yields a “spectrum” of reconstructed invariant masses.

Although the “target” mother particle is of definite mass, it does not show up as a delta peak. Namely, because the LHCb detector is not infinitely accurate, the peak has a certain width, which is referred to as the resolution and corresponds to the LHCb detector’s inaccuracy. In addition to the finite accuracy with which the momenta and energies are measured, radiative effects, i.e. particles radiating energy in the form of photons before having their momentum measured by the trackers, also play an important role, by causing a “radiative tail” to the peak in the reconstructed invariant mass spectrum⁴. This radiative tail consists of events that have radiated photons and are therefore reconstructed with too low of an invariant mass. Namely,

⁴“Bremmstralung” is the dominant source of energy loss for electrons when passing through a dense medium, because of the electrons’ small mass (Kappert, 2016). Bremmstralung, which is German for “braking radiation”, is a radiative effect that occurs when charged particles approach the nuclei (of the atoms of the medium they are traversing) closely enough to feel these nuclei’s Coulomb fields. Consequently, the charged particle’s trajectory is bent in the Coulomb field, emitting a photon, to conserve momentum, in the process. These emitted photons cause the charged particle to lose momentum and thus decelerate, hence Bremmstralung.

as the radiated energy is not picked up by the LHCb detector and is therefore missed in the reconstruction, the momenta used in the reconstruction are “too low” (as they have decreased from radiative energy loss), resulting in a reconstructed invariant mass that is smaller than the actual mass of the decay’s mother particle.

Backgrounds In addition to the peak that corresponds to the actual mass of the target mother particle, other peaks are often also present in the reconstructed invariant mass spectrum. Such peaks are called *peaking backgrounds*; where “background” refers to the fact that they consists of other types of events than the target event.

One type of peaking background is that of another type of mother particle of similar mass to the target mother particle, that decays into exactly the same particles as the target decay. Such peaks show up to the left or the right of the target peak, depending on whether the incorrect mother particle is, respectively, lighter or heavier than the target mother particle. The occurrence of such peaks in the reconstructed mass spectrum is generally very predictable, as it is well known which particles have masses close to the mass of the target mother particle and decay into the same daughter particles (as the target decay), with significant branching fractions.

Another type of peaking background is the misidentified background. These backgrounds show up when one or more daughter particles of a decay other than the target decay are misidentified, resulting in the decay’s daughter particles’ *presumed* identities to correspond to the identities of the daughter particles of the target decay. Because one or more of the particles are actually misidentified, the event is reconstructed with either too little or too much energy, resulting in a, respectively, lower or higher reconstructed invariant mass. Depending on whether the decay’s reconstructed mass is too low, or too high, the misidentified event shows up to the left or right of the mass of its mother particle, respectively.

A third type of peaking background is the partially reconstructed background. Here, a decay that includes the daughter particles of the target decay, but also includes at least one other particle, is only partially reconstructed, missing the daughter particle(s) that is (are) not part of the target decay. Because the energy of the missed particle(s) is not included in the invariant mass reconstruction, the partially reconstructed background shows up to the left of its mother particle’s mass.

Lastly, a non-peaking background, known as the combinatorial background, is present, to some extent, in all reconstructed invariant mass spectra. The combinatorial background consist of random tracks/detector hits

that were accidentally reconstructed as the target decay. Because these are random occurrences, they can have almost any energy, and do not “prefer” any specific energies. Therefore, the combinatorial background shows up as a non-peaking background. The combinatorial background decreases with the invariant mass, as high invariant masses are less likely to be reconstructed from random tracks (the more tracks an event has, the higher the probability that random tracks are reconstructed as the target decay, but the lower the energy per track).

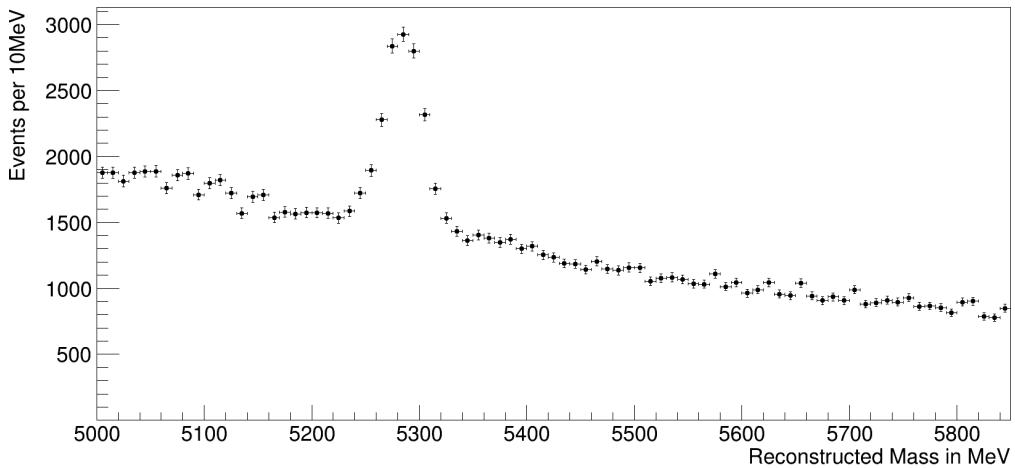


Figure 19: The invariant mass spectrum of the $B^0 \rightarrow K^\pm\pi^\mp$ channel (number of data-events per bin of 10Mev), binned in bins of $10 \text{ MeV}/c^2$ and for a PID cut of $|\Delta LL_K| > 10$

The $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum For the $B^0 \rightarrow K^\pm\pi^\mp$ decay, a typical reconstructed invariant mass spectrum, after binning in bins of 10 MeV and applying a PID cut of $|\Delta LL_K| > 10$, is shown in figure 19. As can be readily seen, the combinatorial background has a very dominant presence in the spectrum, and even where the B^0 mass peaks, the number of combinatorial background events still exceeds the number of target, or “signal” events. This is caused by the TOS requirement made on the events’ triggering category. Namely, as will be explained in the section 6.3, the events in the $B^0 \rightarrow K^\pm\pi^\mp$ analysis are required to be TOS events, which have a much higher probability to be combinatorial background events than TIS events. The high combinatorial background contribution proves to pose quite a challenge to the extraction the number of $B^0 \rightarrow K^\pm\pi^\mp$ from the reconstructed invariant mass spectrum, i.e. the yield, which will be explained in section 5.4.

Furthermore, to the right of the peak of the B^0 meson, a small peak, caused by the $B_s^0 \rightarrow K^\pm\pi^\mp$ channel, can be seen, which corresponds to the slightly higher mass of the B_s^0 meson. This peak should theoretically contain about 8 times as few events as the B^0 peak, as the hadronization fraction $\frac{f_s}{f_d}$ is about $\frac{1}{2}$ (so two times fewer B_s^0 than B^0 mesons are produced) and the $B_s^0 \rightarrow K^\pm\pi^\mp$ branching fraction is about four times as small as the $B^0 \rightarrow K^\pm\pi^\mp$ (Olive, 2016).

Depending on the PID cut applied, some “cross-contamination” between the B^0 and B_s^0 peaks shows up between the two peaks. Namely, if, of a $B_s^0 \rightarrow K^\pm K^\mp$ decay, one of the kaons is misidentified as a pion, some energy is missed because the kaon (being misidentified as a pion) is allocated too little mass. Therefore, the reconstructed mass ends up being slightly smaller than the B_s^0 meson’s mass, which causes the event to end up between the B^0 and B_s^0 peaks. Similarly, for a $B^0 \rightarrow \pi^\pm\pi^\mp$ decay, if one of the pions is misidentified as a kaon, the pion is allocated too much mass and the reconstructed invariant mass is slightly too high, which causes the event’s reconstructed invariant mass to end up between the B^0 and B_s^0 peaks as well.

The mis-ID background makes extracting the final number of $B^0 \rightarrow K^\pm\pi^\mp$ from the invariant mass spectrum quite challenging, as the two effects overlap both with each other, and have the tails underneath the B^0 and B_s^0 peaks. Fortunately, the mis-ID background can be cut away quite effectively by applying stringent cuts on the ΔLL PID variable, requiring with a higher level of certainty that the reconstructed kaon was actually a kaon, and the reconstructed pion a actually pion. No attempts were made to fit the misID background, as it was already decided to perform the yield fits with a PID cut of $|\Delta LL_K| > 10$, after which virtually no mis-ID background is left. However, the effect of the mis-ID background on the fit, when no PID cut is made, and no mass model for mis-ID background is included, can be seen from comparing figure 20 to figure 32. Namely, no PID cut was made on the events in figure 20 (note the high number of events between the B^0 and B_s^0 peaks), whereas figure 32 shows the complete fit after a PID cut of 10 was ($|\Delta LL_K| > 10$) was made on the data.

Lastly, starting about a pion mass to the left of the B^0 peak and continuing further to the left, a “structure” on top of the combinatorial background can be seen, which is assumed to consist of partially reconstructed B^0 decays, as the lightest particle typically missed in the reconstruction is the pion. This structure will be further discussed in section 5.4.

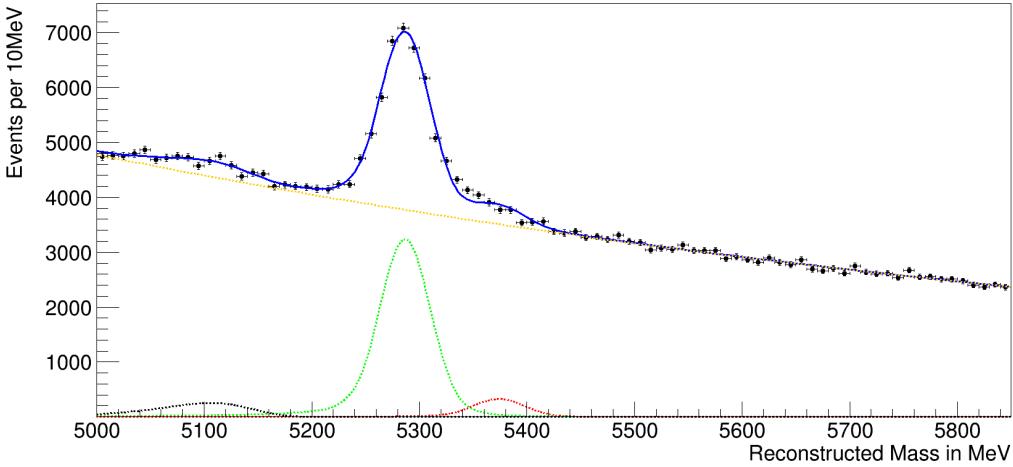


Figure 20: A typical resulting total fit (blue) to the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum without having applied a PID cut to the data (note the high number of events between the B^0 and B_s^0 peaks)

5.2 Correcting mass hypotheses

As described previously, events are reconstructed according to a certain mass hypothesis. However, the identities of the detected particles are quite ambiguous and are often expressed in “likelyhoods” of particles having a certain identity, in terms of ΔLL variables. Only after requiring stringent ΔLL values, accurate assertions as to the particles’ identities can be made. These PID cuts are applied as part of the final analysis, as it is up to the analyst to a judge what level of stringency of the PID cut provides the best balance between removing background and having a high efficiency (i.e. how many target or *signal* events are discarded because of the cut). Initially, for the construction of the ntuples, the choice of allocating particle identities is therefore quite arbitrary.

The $B \rightarrow h^\pm h^\mp$ samples used in this analysis were primarily constructed for the $B_{(s)}^0 \rightarrow \mu^\pm\mu^\mp$ analysis. As the particles’ identities are ambiguous before making stringent cuts on the ΔLL variables, in the construction of the Ntuples for the $B \rightarrow h^\pm h^\mp$ samples, masses were simply allocated based on the $B_{(s)}^0 \rightarrow \mu^\pm\mu^\mp$ mass hypothesis (i.e. on the assumption that both hadrons of the $B \rightarrow h^\pm h^\mp$ decay are muons). Only after the particles’ identities are determined with appreciable certainty, the reconstructed invariant mass is corrected according to the appropriate mass hypothesis.

The correct reconstructed invariant mass can be obtained from understanding how the invariant mass is reconstructed according to the principle

of conservation of the invariant mass. The frame-independent expression for the invariant mass of a system is given by

$$m = \sqrt{E^2 - \mathbf{p}^2}. \quad (22)$$

The invariant mass of the system of the mother B^0 particle is simply the B^0 's invariant mass. Moreover, because the invariant mass of the system is conserved in the decay, we can express the invariant mass of the mother B^0 particle in terms of the energies and momenta of the daughter kaon and pion, as

$$m_{B^0} = \sqrt{{E_{tot}}^2 - {\mathbf{p}_{tot}}^2} = \sqrt{\left(\sum E_{daughter}\right)^2 - \left(\sum \mathbf{p}_{daughter}\right)^2}. \quad (23)$$

For the $B^0 \rightarrow K^\pm \pi^\mp$ decay, the conservation of mass is expressed as

$${m_{B^0}}^2 = (E_{K^\pm} + E_{\pi^\mp})^2 - (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp})^2, \quad (24)$$

$${m_{B^0}}^2 = {E_{K^\pm}}^2 + {E_{\pi^\mp}}^2 + 2E_{K^\pm}E_{\pi^\mp} - \mathbf{p}_{K^\pm}^2 - \mathbf{p}_{\pi^\mp}^2 - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp}, \quad (25)$$

where \mathbf{p}^2 denotes the square of the absolute value of \mathbf{p} , i.e. $\mathbf{p}^2 = |\mathbf{p}|^2 = \mathbf{p} \cdot \mathbf{p}$. By using $E^2 = m^2 + \mathbf{p}^2$, we can express the energy in terms of the energies and momenta:

$${m_{B^0}}^2 = ({E_{K^\pm}}^2 - \mathbf{p}_{K^\pm}^2) + ({E_{\pi^\mp}}^2 - \mathbf{p}_{\pi^\mp}^2) + 2E_{K^\pm}E_{\pi^\mp} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp}, \quad (26)$$

$${m_{B^0}}^2 = {m_{K^\pm}}^2 + {m_{\pi^\mp}}^2 + 2\sqrt{({m_{K^\pm}}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{({m_{\pi^\mp}}^2 + \mathbf{p}_{\pi^\mp}^2)} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp}. \quad (27)$$

On the other hand, if we reconstruct the B^0 meson's mass using the muons' masses, the incorrect invariant mass, which we will denote as $m_{\mu\mu}$, will be reconstructed according to

$${m_{\mu\mu}}^2 = {m_{\mu^\pm}}^2 + {m_{\mu^\mp}}^2 + 2\sqrt{({m_{\mu^\pm}}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{({m_{\mu^\mp}}^2 + \mathbf{p}_{\pi^\mp}^2)} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp}. \quad (28)$$

We will denote the difference between the correct squared mass ${m_{B^0}}^2$, and the incorrect squared mass ${m_{\mu\mu}}^2$, as $\Delta(m^2) = {m_{\mu\mu}}^2 - {m_{B^0}}^2$, which can thus be expressed as

$$\Delta(m^2) = m_{\mu^\pm}^2 + m_{\mu^\mp}^2 + 2\sqrt{(m_{\mu^\pm}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{(m_{\mu^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)} - \\ \left(m_{K^\pm}^2 + m_{\pi^\mp}^2 + 2\sqrt{(m_{K^\pm}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{(m_{\pi^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)} \right), \quad (29)$$

where all the quantities in equation (29) can be determined from the data. Moreover, because the invariant mass is frame independent, the expression in equation (29) holds in any frame, including that of the LHCb detector. Thus, the correct invariant mass m_{B_d} can be calculated from $m_{B_d}^2 = m_{\mu\mu}^2 - \Delta(m^2)$.

5.3 Binning

Before performing the fit, the data are first binned into histograms, as binned fits take significantly less computation time. When dealing with small amounts of data, binned fits are inappropriate, as too much information is lost in the binning process. However, as we are dealing with a large number of events for the $B^0 \rightarrow K^\pm \pi^\mp$ channel, a binned fit is not only permissible, but actually preferable. The data are binned in 60 bins of $10 \text{ MeV}/c^2$, and the mass window used runs from $5200 \text{ MeV}/c^2$ to $5850 \text{ MeV}/c^2$.

The left bound on the mass window was chosen such that the “partial reconstruction” structure (to the left of the B^0 peak in the mass spectrum) is excluded from the fit. Moreover, in the construction of the ntuples, a cut was made on the minimum and maximum reconstructed mass of the $\mu^\pm \mu^\mp$ pair. This leaves a relatively wide mass window for the $B_{(s)}^0 \rightarrow \mu^\pm \mu^\mp$ analysis. However, this cut is also made on the $B \rightarrow h^\pm h'^\mp$ samples used in this analysis and when we correct the reconstructed invariant mass spectrum for the false $B_{(s)}^0 \rightarrow \mu^\pm \mu^\mp$ mass hypothesis, the mass window cuts are smeared when propagated to the corrected mass spectrum. Namely, events that are initially reconstructed with identical invariant masses under the $B_{(s)}^0 \rightarrow \mu^\pm \mu^\mp$ mass hypothesis, end up with different corrected invariant masses, depending on their momentum distributions, when adjusted for the $B^0 \rightarrow K^\pm \pi^\mp$ mass hypothesis.

Figure 21 shows the reconstructed invariant mass spectrum obtained from $\mu^\pm \mu^\mp$ pairs that are generated with random momenta and a uniform distribution of the pairs’ invariant mass, after which events with a reconstructed invariant mass outside of the range from $4875 \text{ MeV}/c^2$ to $5850 \text{ MeV}/c^2$ are discarded. Moreover, figure 22 shows the corrected invariant mass spectrum of the events in figure 21, when adjusted for the $B^0 \rightarrow K^\pm \pi^\mp$ mass hypothesis. As can be seen from figure 22, the initially “boolean” cut in figure 21 has

been smeared by transitioning to the new mass hypothesis⁵. Thus, the mass window, running from $5200 \text{ MeV}/c^2$ to $5850 \text{ MeV}/c^2$, was chosen, because figure 22 shows that the mass spectrum within this range is unaffected by the smearing of the invariant mass cut.

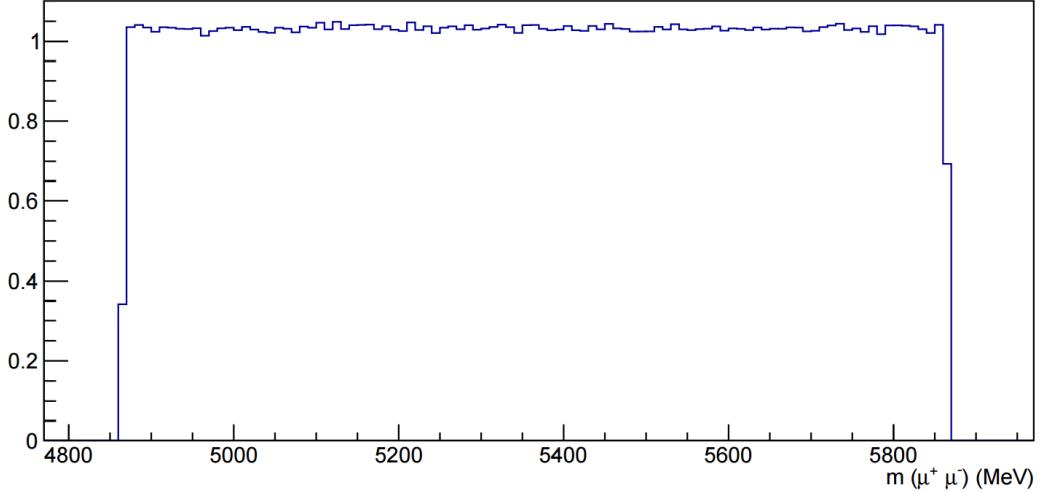


Figure 21: Initial mass window cut for the $\mu^\pm\mu^\mp$ mass hypothesis on a uniform invariant mass distribution

5.4 Maximum likelihood fit

To extract the number of $B^0 \rightarrow K^\pm\pi^\mp$ events from the reconstructed invariant mass spectrum, the background needs to be separated from the signal events. One method for doing so is called sideband subtraction, which is illustrated in figure 25. This method uses the “sidebands”, the regions

⁵Figures 23 and 24 show the corrected invariant mass spectra of the events in figure 21, when adjusted for the $B^0 \rightarrow \pi^\pm\pi^\mp$ and $B^0 \rightarrow K^\pm K^\mp$ mass hypotheses, respectively. As can be seen from figure 23, the invariant mass cut remains virtually unsmeared under the transition to the $B^0 \rightarrow \pi^\pm\pi^\mp$, whereas the smearing in figure 24 is slightly exacerbated with respect to figure 22. Thus, the smearing, of events with identical invariant mass under the $B^0 \rightarrow \mu^\pm\mu^\mp$ mass hypothesis, but with different momentum distributions, under the transition to the $B^0 \rightarrow K^\pm\pi^\mp$ mass hypothesis is almost entirely caused by the large mass difference between the kaon and the muon (the pion is less than 30% heavier than the muon, whereas the kaon is almost five times as heavy), which is unsurprising as transitioning from the $B_{(s)}^0 \rightarrow \mu^\pm\mu^\mp$ mass hypothesis to a mass hypothesis of which the daughter particles’ masses are identical to the $B_{(s)}^0 \rightarrow \mu^\pm\mu^\mp$ daughter particles’ masses (which is simply the identity operation), leaves the invariant mass cut completely unsmeared, irrespective of the momenta involved.

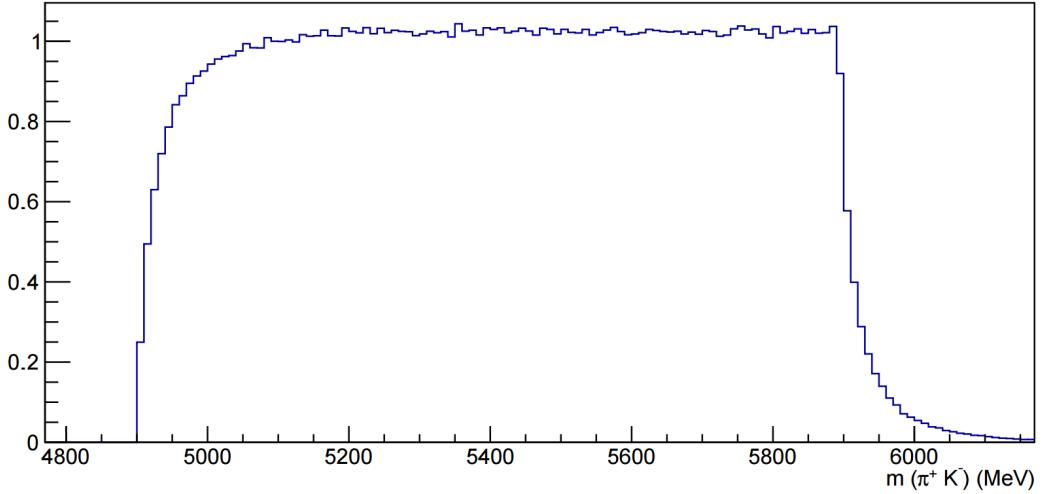


Figure 22: Resulting smeared mass window cut for transitioning from the $\mu^\pm\mu^\mp$ mass hypothesis to the $K^\pm\pi^\mp$ mass hypothesis

far enough away from the target peak for its tails to be negligible, to estimate the amount background “underneath” the main peak. This straightforward method works relatively well for eliminating the combinatorial background, but becomes unreliable when peaking backgrounds have a dominant presence in the reconstructed invariant mass spectrum. Therefore, for the $B^0 \rightarrow K^\pm\pi^\mp$ mass spectrum, a more elaborate method, called maximum likelihood fitting, is used. This method relies on finding parametrized expressions for the signal and different backgrounds, and fitting the sum of these expression to the reconstructed invariant mass spectrum. For this method, the shapes of the signal and different backgrounds need to be understood and modeled correctly. The procedure for this will be explained below. Furthermore, an example of a typical $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum, from which the signal yield needs to be extracted, is shown in figure 19.

5.4.1 Combinatorial background shape

The fit model for the combinatorial background is the most straightforward; an exponential function, which uses a single parameter, is fitted to the combinatorial background. The exponential parameter only determines the shape (slope) of the exponential function, as in the final fit, the proportion, in which the different polynomials are summed, can scale the exponential shape up or down. An example of a typical exponential fit to the combinatorial background in the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum is shown

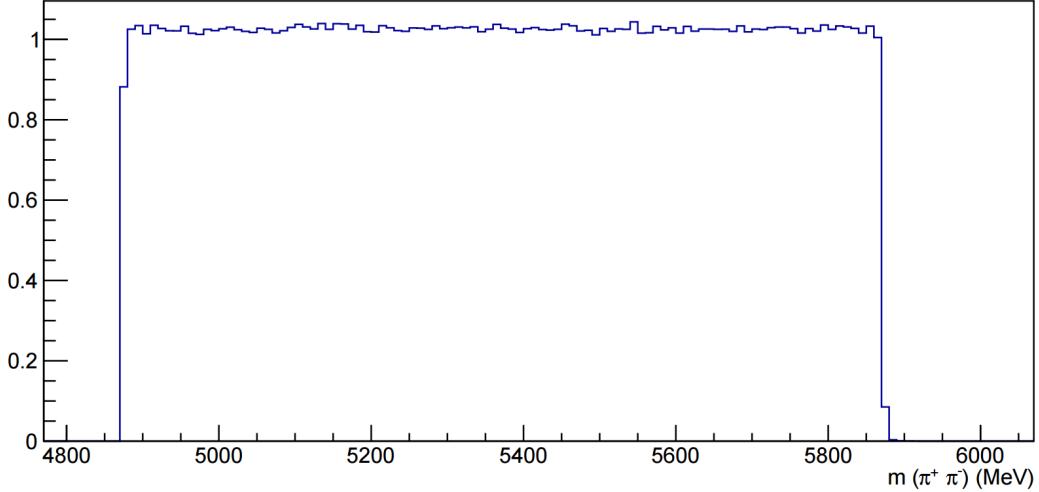


Figure 23: Resulting (almost) unsmeared mass window cut for transitioning from the $\mu^\pm\mu^\mp$ mass hypothesis to the $\pi^\pm\pi^\mp$ mass hypothesis

in yellow in figure 26.

5.4.2 $B^0 \rightarrow K^\pm\pi^\mp$ peak shape

The $B^0 \rightarrow K^\pm\pi^\mp$ peak is fitted using two Crystal ball functions, to account for non-Gaussian radiative and resolution effects in the tails of the distribution.

The Crystal ball function consist of a Gaussian core, and a power-law tail for either the left or right tail. The Crystal ball has four parameters, two to describe the mean and sigma of the Gaussian core, one to describe where the Gaussian tail transitions into the power-law tail (how many sigma's to the left or right of the mean), and one to describe the shape of the power-law tail. The Crystal ball function is defined as (*Crystal Ball Function Wikipedia webpage*, 2016)

$$f(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}, & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha, \\ A \cdot \left(B - \frac{x-\bar{x}}{\sigma}\right)^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha, \end{cases} \quad (30)$$

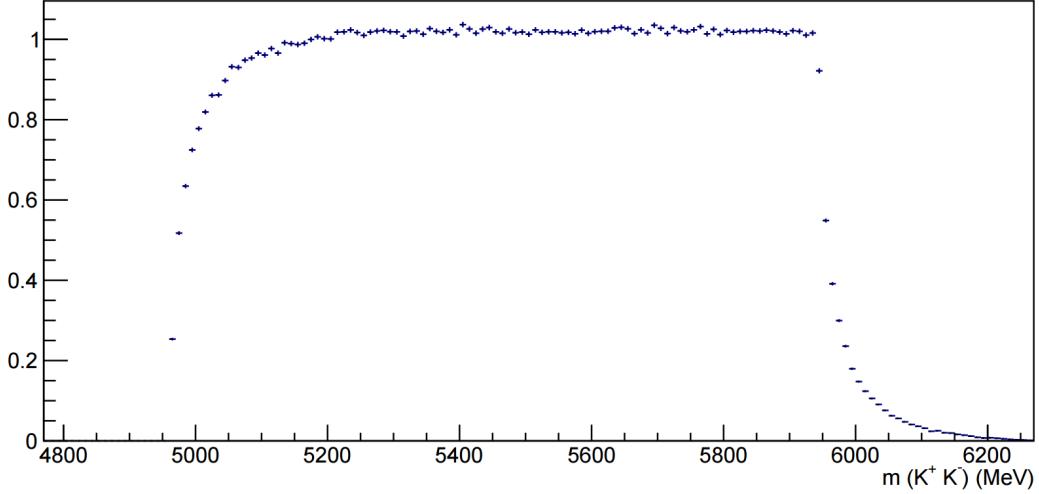


Figure 24: Resulting smeared mass window cut for transitioning from the $\mu^\pm\mu^\mp$ mass hypothesis to the $K^\pm K^\mp$ mass hypothesis

where

$$\begin{aligned}
 A &= \left(\frac{n}{|\alpha|} \right)^n \cdot e^{\left(-\frac{|\alpha|^2}{2} \right)}, \\
 B &= \frac{n}{|\alpha|} - |\alpha|, \\
 N &= \frac{1}{\sigma(C + D)}, \\
 C &= \frac{n}{|\alpha|} \cdot \frac{1}{n-1} \cdot e^{\left(-\frac{|\alpha|^2}{2} \right)}, \\
 D &= \sqrt{\frac{\pi}{2}} \left(1 + \text{erf} \left(\frac{|\alpha|}{\sqrt{2}} \right) \right),
 \end{aligned}$$

and α , n , \bar{x} and σ are the four previously described parameters. α determines where the Gaussian tail transitions into the power-law tail, n is the shape-parameter of the power-law tail, and \bar{x} and σ are, respectively, the mean and sigma of the Gaussian core. Moreover, N is a normalization factor, and ‘‘erf’’ is the error function (*Crystal Ball Function Wikipedia webpage*, 2016). Examples of different Crystal ball functions, for different α and n parameters, can be seen in figure 27.

Using two Crystal ball functions, one with the power-law tail to the left, and one with the power-law tail to the right, and using the same mean and sigma for the Gaussian core of both Crystal ball functions, a shape is

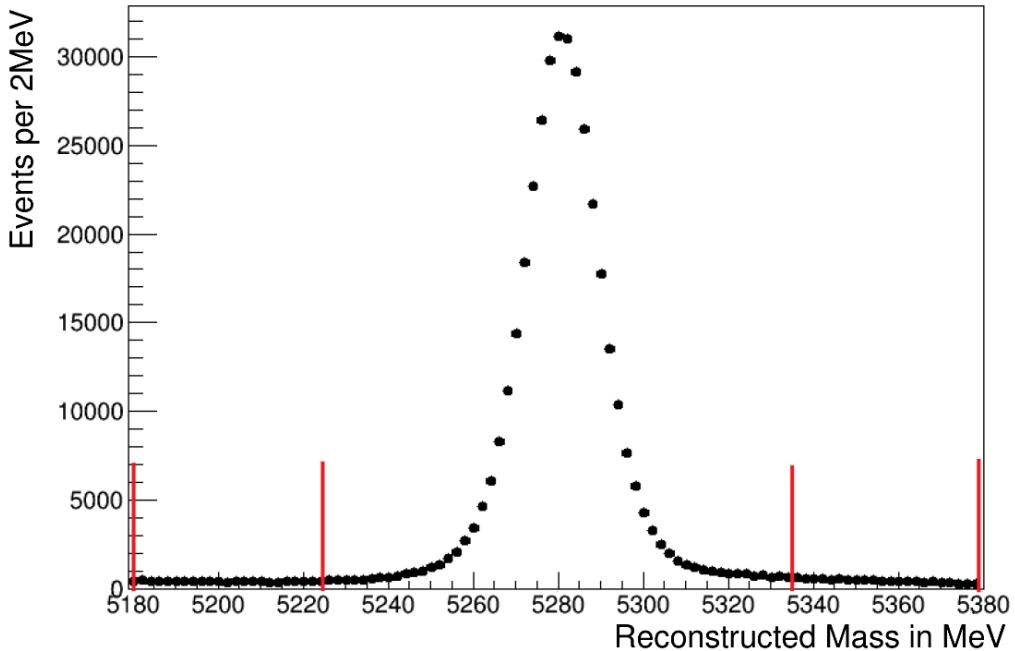


Figure 25: The sidebands subtraction method (Zonneveld, 2016)

obtained that consists of a Gaussian core and two power-law tails. However, out of the six parameters that describe the full “double Crystal ball” shape, only two can be accurately determined from fitting the shape to the data, namely the Gaussian mean and sigma parameters. The four parameters describing the two power-law tails are impossible to determine from the data, because they strongly resemble (and therefore correlate with) the large combinatorial background. Therefore, these parameters are determined from fitting the double Crystal ball shape to the Monte Carlo simulated sample shown in figure 28, which represents the shape of the signal channel well and contains a negligible amount of background. However, an important downside of this method is that ΔLL variables are not modeled well by the Monte Carlo sample. Therefore, although the double Crystal ball shape is fitted to the data after applying a ΔLL cut of 10, the power-law tail parameters of the double Crystal ball shape are determined from simulated data to which no ΔLL cut was applied. Consequently, any changes to the shape of the signal peak caused by the ΔLL cut, are not incorporated in the shape of the simulated invariant mass spectrum, causing the tail parameters, and therefore the final yield, to be biased. Nevertheless, this method was chosen to be used for the final yield fit, as no better alternatives or remedies were found.

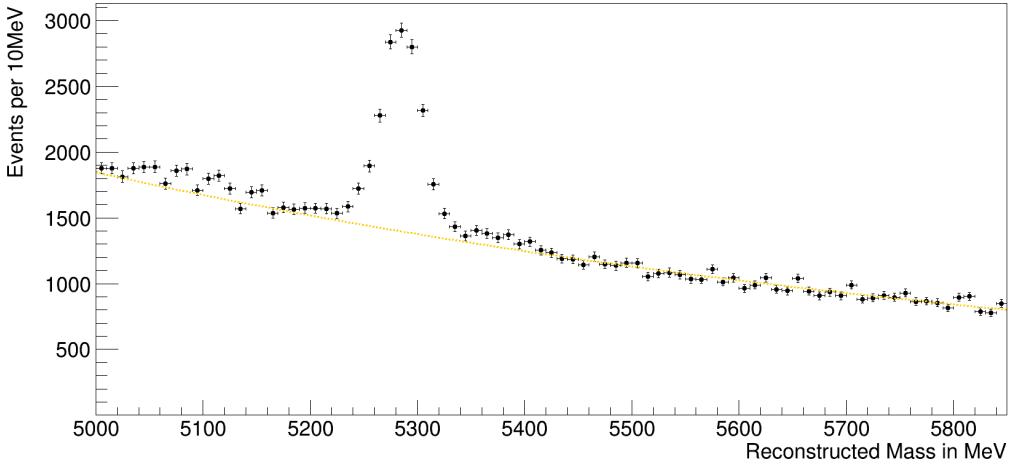


Figure 26: An exponential fit to the combinatorial background in the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum

Moreover, the tail parameters found from the Monte Carlo data were not used to fix, but only to ‘‘constrain’’ the tail parameters of the double Crystal ball shape. This means that the tail parameters, and their variances, as found from the simulated sample, are transferred to RooFit (the fitting algorithm used) as additional constraints. Consequently, if sufficient evidence is present in the data, the fitted tail parameters can deviate from the constraining values, with more statistical evidence being needed to deviate more from the constraining value. However, the final fitted values only differ within error bars from the constraining values, showing the need for external information to determine the tail parameters, as they cannot be found from the data. An example of a typical double Crystal ball fit to the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum is shown in green in figure 29.

5.4.3 $B_s^0 \rightarrow K^\pm\pi^\mp$ peak shape

The shape of the B_s^0 peak is similar to that of the B^0 peak. Therefore, the B_s^0 peak is also fitted using two Crystal ball functions with identical mean and sigma. However, fitting the B_s^0 peak is much more challenging than fitting the B^0 peak. Namely as the data contain about eight times fewer $B_s^0 \rightarrow K^\pm\pi^\mp$ events, the B_s^0 peak is much more difficult to separate from the combinatorial background, making it almost impossible to determine the fit parameters accurately from the data. However, the parameters found from the fit to the B^0 peak can be used to determine the B_s^0 peak’s parameters, in the following way: the calibration of the LHCb is slightly off, causing the mean of the B^0 mass to be slightly shifted, but this effect is the same for

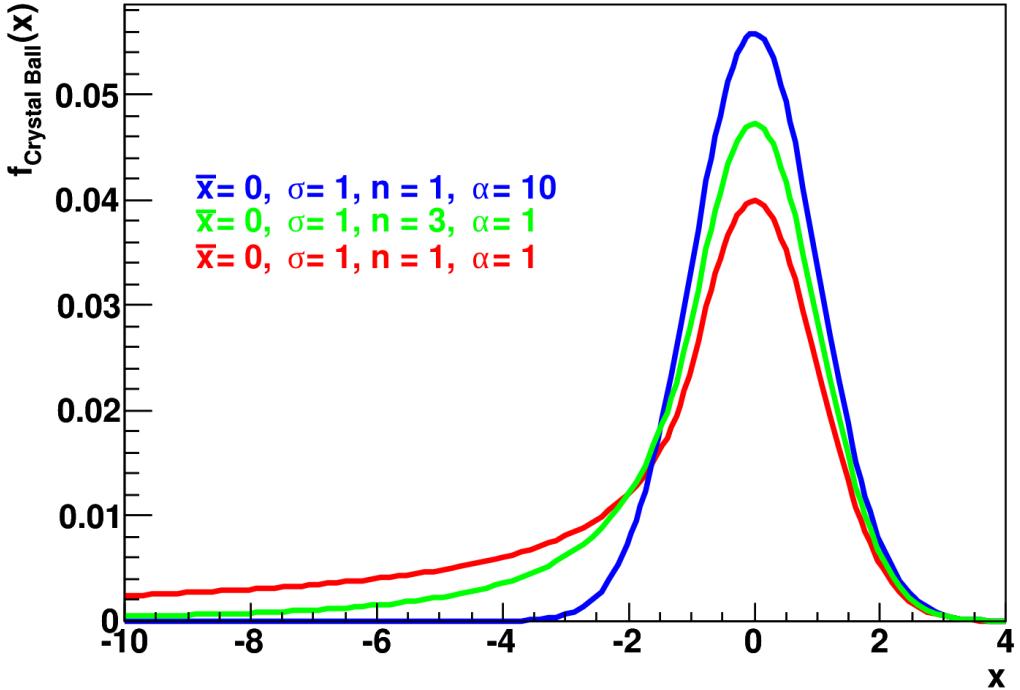


Figure 27: Examples of different Crystal ball functions as depending on the α and n parameters (*Crystal Ball Function Wikipedia webpage*, 2016)

all events. Therefore, the mean of the B_s^0 peak can be fixed to the B^0 mass mean found from the data, increased by the difference between the B^0 and B_s^0 masses, as known from the literature. This means that the B^0 and B_s^0 mass means are simultaneously fitted, with a fixed difference between the two, but as the B^0 peak is about eight times as large, in practice, the B^0 peak is very dominant in determining these means. The width of the peak, often referred to resolution, increases slightly with increasing invariant mass. This behavior is very well understood⁶, from which the resolution of the B_s^0 peak is predicted to be about 1.6 percent larger than the B^0 peak's resolution. Again, the B^0 peak's sigma and B_s^0 peak's sigma are fitted simultaneously, with the B_s^0 peak's sigma fixed to being one percent larger than the B^0 peak's resolution. However, in practice, the B^0 peak again plays the dominant role in the determination of the sigma's. The non-Gaussian radiative and resolution effects of the B_s^0 peak are similar to those of the B^0 meson, so the tail parameters of the B_s^0 peak are constrained to the same values as the the B^0 peak. An example of a typical double Crystal ball fit to the $B_s^0 \rightarrow K^\pm \pi^\mp$

⁶The dependence of the resolution on the invariant mass can be approximated by a power law plus a constant: $\approx a \cdot m^y + C$, where $y = 1.33 \pm 0.07$ (Adrover et al., 2013).

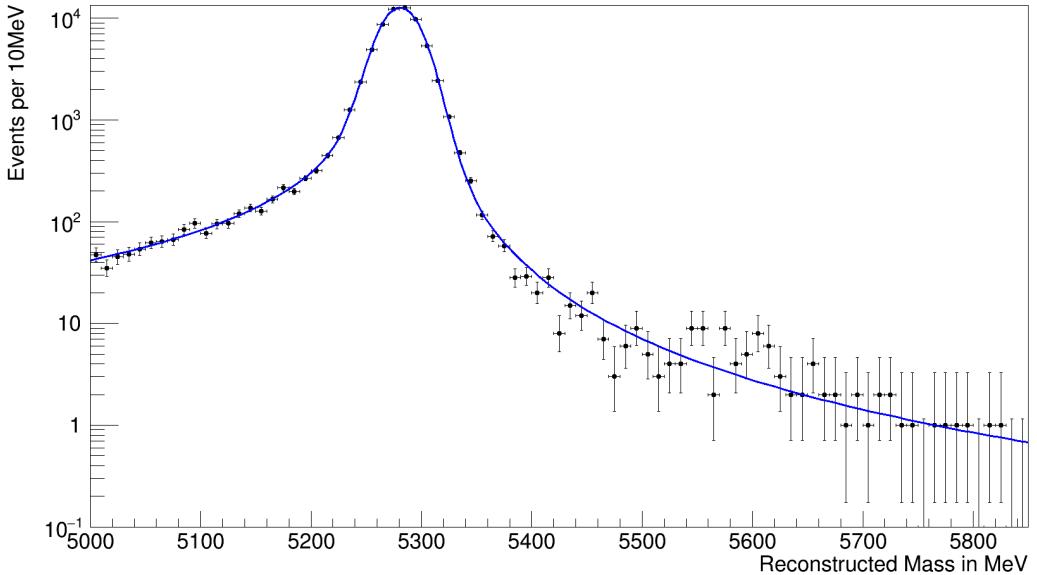


Figure 28: Double Crystal Ball fit to the Monte Carlo sample of simulated $B^0 \rightarrow K^\pm \pi^\mp$ events (note: log scale)

reconstructed invariant mass spectrum is shown in red in figure 30.

5.4.4 Partially reconstructed background

As explained above, it was finally decided to narrow the mass window, excluding the ‘‘partial reconstruction structure’’. However, this was not before significant effort was spent on understanding and fitting this peaking background. Namely, the mass window should ideally include sufficient data on each side of the B^0 peak, such that the combinatorial background can be estimated accurately on each side (i.e. such that the exponential background function can be fit accurately to the combinatorial background on each side of the B^0 peak). Then, the data on both sides of the B^0 peak contribute to the estimation of the combinatorial background underneath the B^0 peak, and the combinatorial background underneath the B^0 peak be ‘‘interpolated’’ from the fit to the combinatorial background on each side of the B^0 peak. Thus, the goal of modeling the partial reconstruction structure was to separate it from the combinatorial background, such that we could also fit the combinatorial on the left side of the B^0 peak, which would improve our estimate of the combinatorial background underneath the B^0 peak.

Two different methods for fitting partially reconstructed backgrounds that are generally used, are the Argus-Gauss convolution and RooPhysBkg shapes. They work in a conceptually similar manner, by describing the shapes

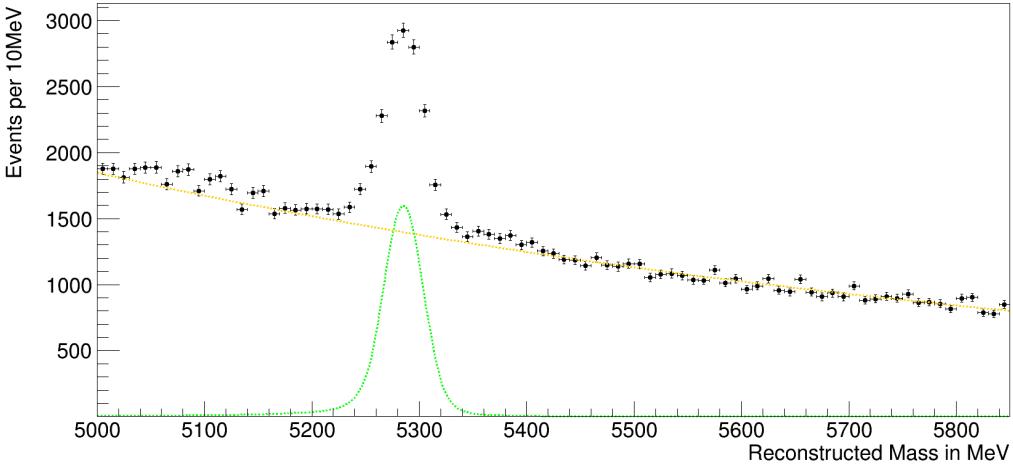


Figure 29: A double Crystal ball fit to the B^0 peak (green) and exponential fit to the combinatorial background (yellow) in the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum

as initially consisting of a non-smeared mass-peak and accompanying resolution, and then smearing it to the left (the lower-mass region), depending on the momentum and mass of the particle that was missed in reconstruction. The Argus-Gauss convolution uses a Gauss function to describe the initial distribution, and ‘‘convolutes’’ it with the Argus shape, representing the smearing caused by missing a daughter particle in the reconstruction. The RooPhysBkg shape is formulated less intuitively but functions in a similar manner. An example of a typical RooPhysBkg fit to the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum is shown in black in figure 31.

From figure 31, it can be seen that the partial reconstruction structure starts about a pion mass to the left of the B^0 peak. In Appendix I, it is explained that when failing to reconstruct the π^0 of a $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ event, the resulting minimum mass defect is indeed the π^0 mass, which matches the observations from the data. However, the expected theoretical yields do not even match the order of magnitude of the observed partial reconstruction yields (which are more than ten times larger than the predicted yields).

As it was not well-understood what caused this structure (it is likely to be caused by several different effects), its shape could not be predicted from theory, and could therefore only be estimated from the data itself, by extrapolating the combinatorial background on the right side of the B^0 peak to the left side of the peak. However, by doing so, the region to the left of the B^0 peak does not actually contribute to the estimation of the combinatorial background underneath the B^0 peak (as, on the left side of

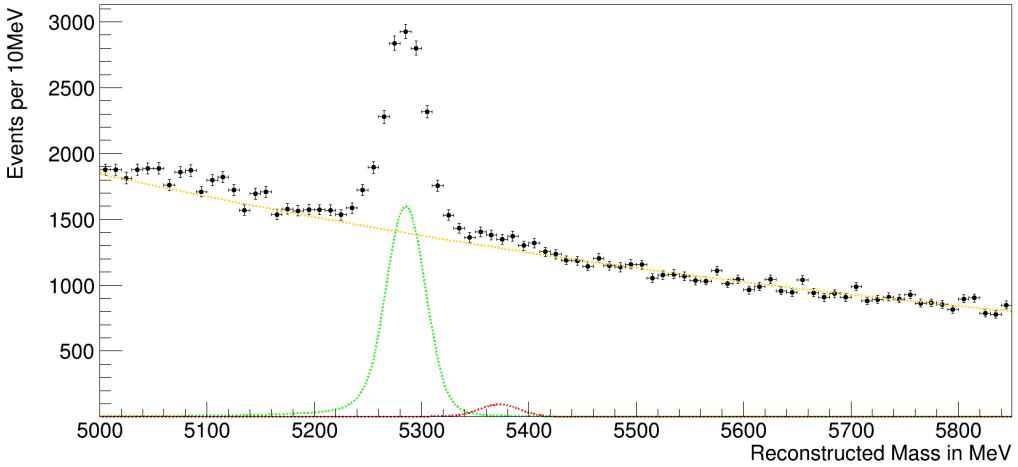


Figure 30: Two double Crystal ball fits to the B_s^0 peak (red) and B^0 peak (green) and an exponential fit to the combinatorial background (yellow) in the $B^0 \rightarrow K^\pm \pi^\mp$ reconstructed invariant mass spectrum

the B^0 peak, the combinatorial background can not be separated from the “partial reconstruction structure” without using the fit to the combinatorial background on the right side of the B^0 peak).

Thus, as the region to the left of the B^0 peak did not contribute to the estimation of the background underneath the B^0 peak, and events with a reconstructed invariant mass smaller than $5200 \text{ MeV}/c^2$ are affected by the (smeared) $\mu^\pm \mu^\mp$ mass cut, we decided to exclude the region from the fit.

5.4.5 Resulting total fit

Having determined the shapes of the combinatorial background, B^0 peak, and B_s^0 peak, the shapes can be scaled up and down in order for the sum of the shapes to match the complete reconstructed invariant mass spectrum. In RooFit, the fitting package used, these scaling parameters are calibrated in such a way that they match the yields of the corresponding shapes, such that the yields of the B^0 and B_s^0 peaks (and the combinatorial background) are directly obtained from the fit. Figures 32 and 33 show the resulting total fits from using the RooPhysBkg and Argus-Gauss convolution shape, respectively, to model the partial reconstruction structure, and figure 34 shows the final total fit, using a narrower mass window, from which the yields presented in section 7 were determined.

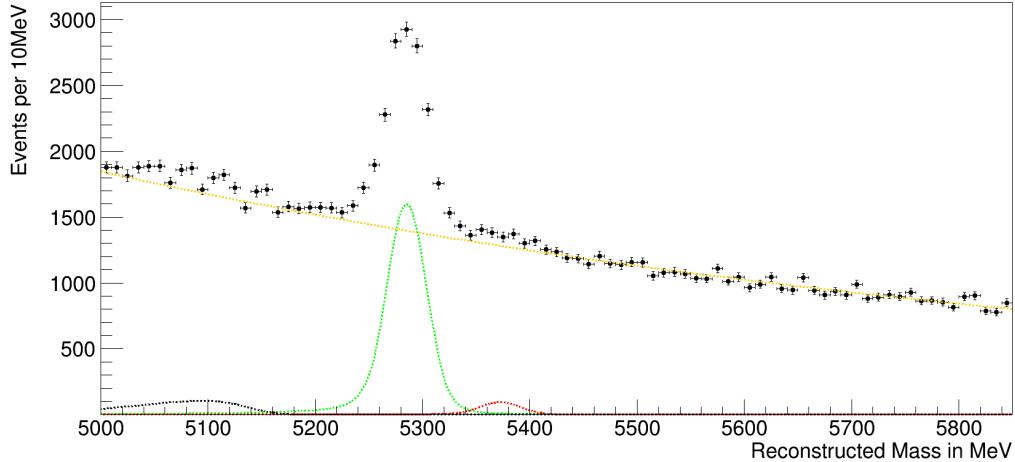


Figure 31: A RooPhysBkg fit to the partial reconstruction structure (black), two double Crystal ball fits to the B_s^0 peak (red) and B^0 peak (green) and an exponential fit to the combinatorial background (yellow) in the $B^0 \rightarrow K^\pm \pi^\mp$ reconstructed invariant mass spectrum

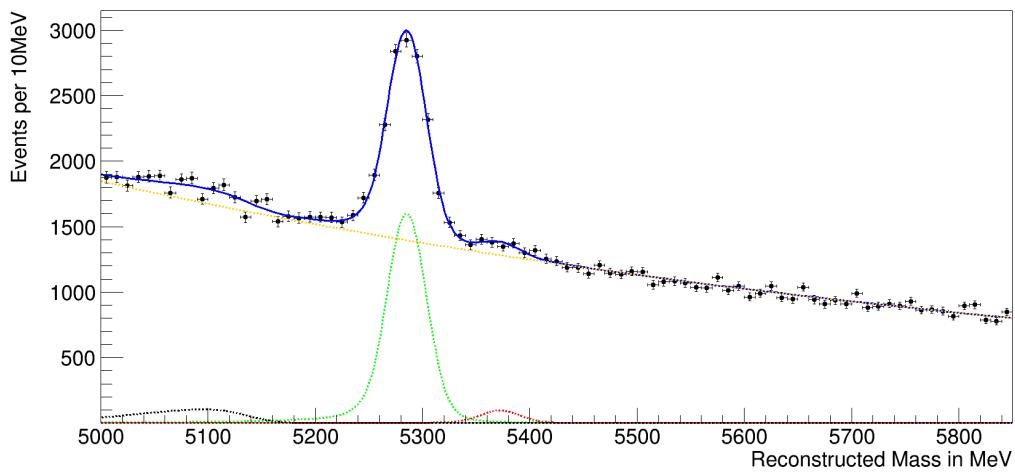


Figure 32: The resulting total fit (blue) from a RooPhysBkg fit to the partial reconstruction structure (black), two double Crystal ball fits to the B_s^0 peak (red) and B^0 peak (green) and exponential fit to the combinatorial background (yellow) in the $B^0 \rightarrow K^\pm \pi^\mp$ reconstructed invariant mass spectrum

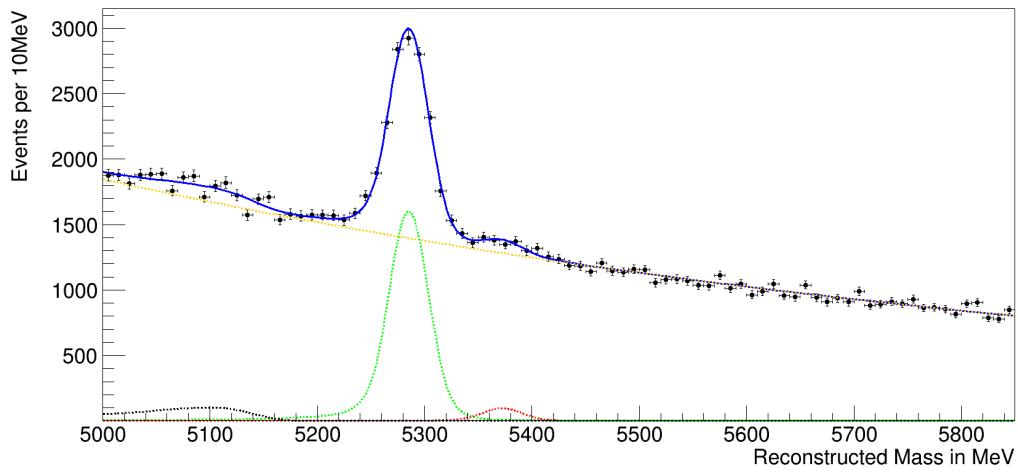


Figure 33: The resulting total fit (blue) from a Argus-Gauss convolution fit to the partial reconstruction structure (black), two double Crystal ball fits to the B_s^0 peak (red) and B^0 peak (green) and exponential fit to the combinatorial background (yellow) in the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum

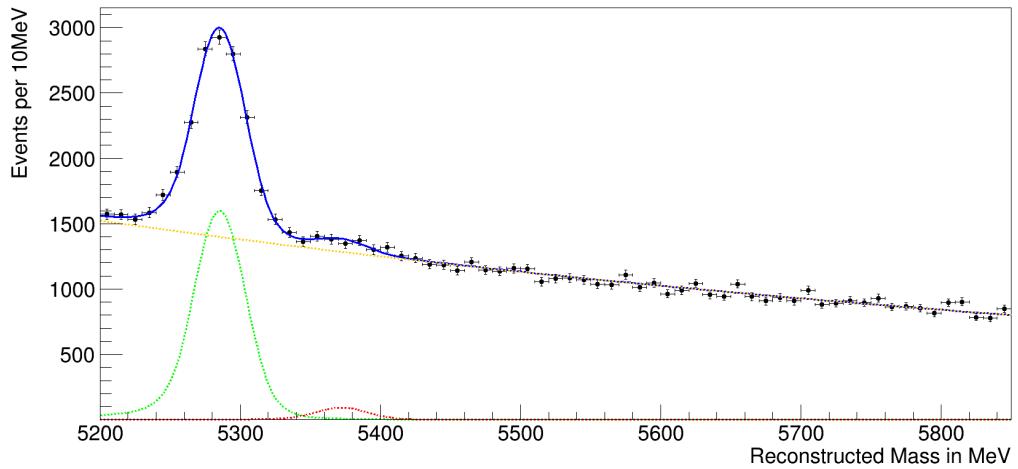


Figure 34: The resulting total fit (blue) from a narrower mass window cut to exclude the partial reconstruction structure, two double Crystal ball fits to the B_s^0 peak (red) and B^0 peak (green) and exponential fit to the combinatorial background (yellow) in the $B^0 \rightarrow K^\pm\pi^\mp$ reconstructed invariant mass spectrum

6 Efficiencies

To understand how many $B^0 \rightarrow K^\pm\pi^\mp$ decays actually occurred in the LHCb detector, the yield found, as according to the method of section 5, needs to be compensated for the fact that events are lost in all steps of the LHCb data flow outlined in section 4. This compensation factor is called the (detection) *efficiency* and estimates what fraction of the *produced* signal events the *detected* signal events represent.

The most straightforward method is to run a simulated Monte Carlo sample through the same processes the data are run through. Because it is known how many signal events were simulated, from counting how many events “survived” whatever process under consideration, the efficiency is obtained by dividing the number of “surviving” events by the total number of simulated events. However, this method only works if the relevant parameters are modeled well in the Monte Carlo sample (e.g. the parameters on which the stripping or PID cuts are made). Therefore, in some cases, *data-driven* methods are used, relying on real data, instead of simulated data, to estimate efficiencies. These data can originate both from the signal channel or (data bases on) one or more other decay channels.

Moreover, the particles flying through the detector are only converted into data (which can be substituted by simulated data to calculate the efficiencies) after their tracks have been accepted by the L0 (hardware) trigger. Therefore, any efficiencies occurring before the high level trigger need to either be obtained from data-driven methods, or, when using simulated data, not only the events themselves, but also their interactions with the detector need to be simulated. Section 4.5.2 explains the different software packages involved in simulating these interactions.

The efficiency is can be divided into four contributions, based on the order in which the efficiencies occur in the LHCb data flow and on the way the efficiency is determined⁷. The four efficiencies are called the acceptance efficiency ϵ^{Acc} (sometimes referred to as the generator level efficiency), trigger efficiency ϵ^{Trig} , reconstruction&selection efficiency ϵ^{SelRec} , and particle identification efficiency ϵ^{PID} , such that the aggregate efficiency becomes

$$\epsilon_{norm} = \epsilon_{norm}^{Acc} \cdot \epsilon_{norm}^{RecSel|Acc} \cdot \epsilon_{norm}^{Trig|RecSel} \cdot \epsilon_{norm}^{PID|RecSel}. \quad (31)$$

As can be seen from equation 31, reconstruction&selection and trigger efficiencies are formulated as conditional probabilities (and the PID efficiency

⁷Typically, the particle identification efficiency is not included in the efficiency, but is reported separately as a direct weighing factor for the yield. However, we choose to include it as an efficiency similar to the other efficiencies reported in this section.

as well), but the order might seem backwards. Namely, the trigger efficiency is determined for events that get reconstructed and selected, and the reconstruction&selection efficiency is determined for events that are within the LHCb detector acceptance, even though the events need to pass the trigger before entering the reconstruction&selection stage. However, this is for practical reasons; the trigger efficiency is the only efficiency that is determined from real data. Therefore, data used for determining the trigger efficiency have already passed the Reconstruction&Selection stage. Thus, the reconstruction&selection efficiency is determined for all events that fall within the LHCb detector acceptance.

6.1 Acceptance

The geometrical detector acceptance is given by the fraction of $B^0 \rightarrow K^\pm \pi^\mp$ events that have all decay products' tracks point inside the detector acceptance. Moreover, some particle trajectories that fall just outside of this range can still be bent into the detector by the magnet, but trajectories that point just inside the detector can still be bent out of the detector by the magnet as well (depending, in both cases, on the charge and momentum of the particle). However, as such effects are considered in the reconstruction&selection efficiency, the acceptance efficiency is slightly “optimistic”, assuming that more tracks fall within the detector acceptance than actually do, and the reconstruction&selection efficiency retroactively takes into account the loss of events that fall outside of the detector acceptance, including losses caused by magnet effects.

Theoretically, the reconstruction&selection efficiency could incorporate the acceptance efficiency as well, as the reconstruction&selection efficiency can include the entire range of events, including the 50% of events that fly in the entirely opposite direction. However, the more finely-tuned reconstruction&selection efficiency determination method is only needed for events that fall within or on the edge of the detector acceptance, whereas the faster acceptance efficiency determination method is sufficient for considering the loss of events that fall generously outside of the detector acceptance.

Thus, after simulating the proton-proton collision and subsequent decays, as explained in 4.5.2, the tracks that fall well outside of the detector acceptance are discarded, such they do not need to be run through the timely, and computationally expensive, reconstruction&selection, and the acceptance efficiency is introduced to accommodate for the loss of (discarded) events. As the acceptance efficiency is independent of any tweaks to the subsequent software stages in the LHCb data flow, the acceptance efficiency is determined once after every hardware upgrade of the detector, after which the value

does not need to be re-calibrated until the next change to the physics and/or hardware of the LHCb.

6.2 Reconstruction&selection

Like the acceptance efficiency, the reconstruction&selection efficiency determination is based on simulation. The reconstruction and selection are software-driven, well-understood processes, and the parameters of the data these methods rely on are modeled accurately in the Monte Carlo simulation, which makes using simulated data for the evaluation of the reconstruction&selection efficiency a viable method. The reconstruction&selection efficiency considers almost the entire LHCb data flow, including part of the detector acceptance, and only excludes the trigger and particle identification efficiencies. Therefore, contrary to the acceptance efficiency, the reconstruction&selection efficiency needs to be determined for every individual analysis (unless two analyses use identical reconstruction and selection processes for the same decay channel), after the reconstruction and selection process has been firmly established. One example is the stripping, which is often tweaked with the goal of achieving an even better balance between eliminating background events and retaining signal events. Another example is the actual mass window used. Namely, narrow mass windows have the potential of losing some signal events by cutting off the tails of the signal distribution. Such effects need to be incorporated into the reconstruction&selection efficiency and therefore, the efficiency can only be determined after the analysis methods have been determined. By using simulated data, the efficiency is determined from the ratio between the events that pass the reconstruction and selection criteria, and the total number of events that were simulated.

6.3 Trigger

As explained in section 4.1, the trigger system consists of the L0, the HLT1 and the HLT2. The HLT2, performing a similar reconstruction as the off-line reconstruction (explained in section 4), is, like the off-line reconstruction, software-based and well-understood, and the parameters it acts upon are modeled accurately in the Monte Carlo simulation. Therefore, the HLT2 efficiency can be estimated from using a simulated sample.

On the other hand, the L0 trigger (and, to a lesser extent, the HLT1 as well) is a hardware trigger and makes its trigger decision based on event parameters that are not modeled correctly in the simulated Monte Carlo sample. Therefore, a suitable alternative is found in data-driven efficiency

determination methods. The trigger efficiency is defined as

$$\epsilon_{norm}^{Trig|SelRec} = \frac{N_{norm}^{Trig|RecSel}}{N_{norm}^{RecSel}}, \quad (32)$$

where $N_{norm}^{Trig|RecSel}$ denotes the number of triggered events that are reconstructed and selected, which are the events in our data, whereas N_{norm}^{RecSel} denotes the total amount of events that would get reconstructed and selected, if all events would be triggered. However, N_{norm}^{SelRec} is not possible to be determined directly, as only a select number of events is triggered and, of the untriggered events, it is impossible to know if they would pass the reconstruction and selection.

Therefore, we will consider a specific subsample of the triggered events, in order to derive the trigger efficiency. We will henceforth omit the superscript $|RecSel$ and subscript $norm$ with the understanding that all of the following trigger efficiencies are efficiencies of the normalization channel and are defined on the same reconstructed and selected events. Given that we previously (in section 4.1) explained the trigger categories TIS, TOS and TISTOS, we can define

$$\epsilon_{TIS|TOS} = \frac{N_{TISTOS}}{N_{TOS}}, \quad (33)$$

i.e. the TIS efficiency for TOS triggered events, is given by the number of events that were triggered both TIS and TOS, divided by the total number of TOS events (irrespective of whether they are also triggered TIS). If the TIS trigger category is indeed triggered truly *independent* of the signal, then the TIS sample is independent of the TOS sample and thus

$$\epsilon_{TIS} = \epsilon_{TIS|TOS} = \frac{N_{TISTOS}}{N_{TOS}}, \quad (34)$$

where the quantities N_{TISTOS} and N_{TOS} can be measured from the data.

However, as we have seen in section 5, the $B^0 \rightarrow K^\pm \pi^\mp$ sample is heavily polluted by background. Therefore, if we would (naively) use the $B^0 \rightarrow K^\pm \pi^\mp$ sample, we would mainly measure the background TIS efficiency, instead of the signal TIS efficiency.

The method of s-weights allows for the recovery of the true signal distributions of certain parameters from a background-polluted sample, based on the yield fit's mass model (Pivk, 2006), but the method is imperfect and elaborate. However, since the TIS efficiency is independent of the signal, the efficiency is the same for all b decays. Contrary to the $B^0 \rightarrow K^\pm \pi^\mp$ channel, the $B^+ \rightarrow J/\psi K^+$ TIS sample has very little background, which is shown in

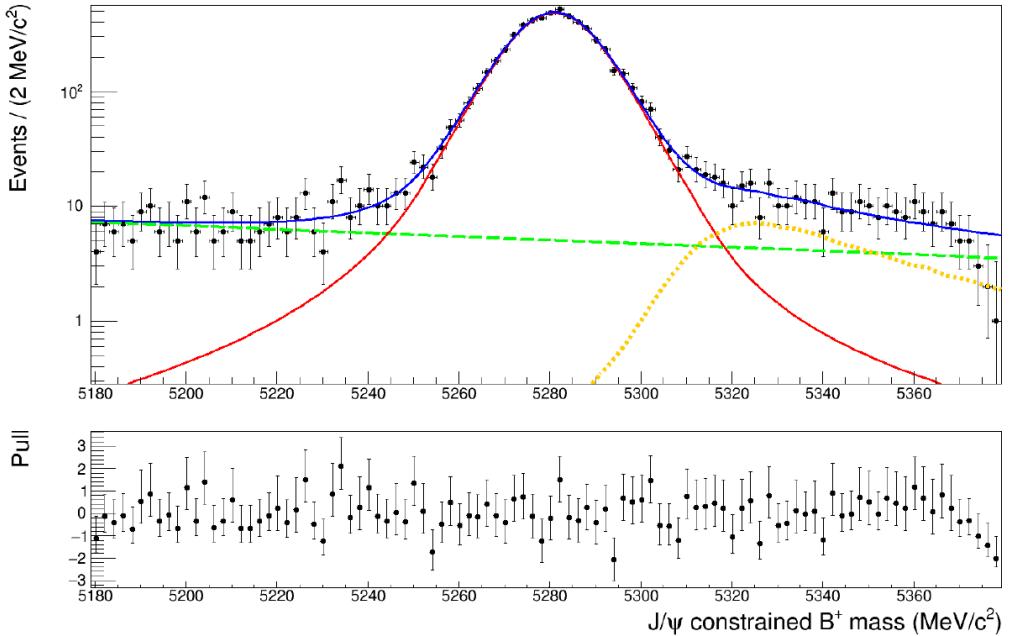


Figure 35: The reconstructed invariant mass spectrum of the $B^+ \rightarrow J/\psi K^+$ TIS sample from 2011 data (Zonneveld, 2016)

figures 35 and 36. Therefore, the TIS efficiency used for the $B^0 \rightarrow K^\pm \pi^\mp$ TIS sample is calculated from the $B^+ \rightarrow J/\psi K^+$ TIS sample according to equation (34).

6.4 Particle identification

The goal of the particle identification efficiency is to account for the yield loss from $|\Delta LL_K| > 10$ cut made on the data, in order to “recover” the signal events lost in the PID cut. The common practice of using a simulated sample to determine efficiencies cannot be used, as the Monte Carlo simulated sample does not model the ΔLL variable correctly.

6.4.1 Pre and post-PID cut yield fit

Theoretically, the yield could be determined both from fitting data with, and without having the $|\Delta LL_K| > 10$ cut applied, and the particle identification efficiency could subsequently be determined from the ratio between the two yields. However, as illustrated in figure 20, without a stringent PID cut, the mis-ID background makes the B_s^0 peak hard to be separated from the B^0 peak and therefore hard to be fit. Moreover, the mis-ID background

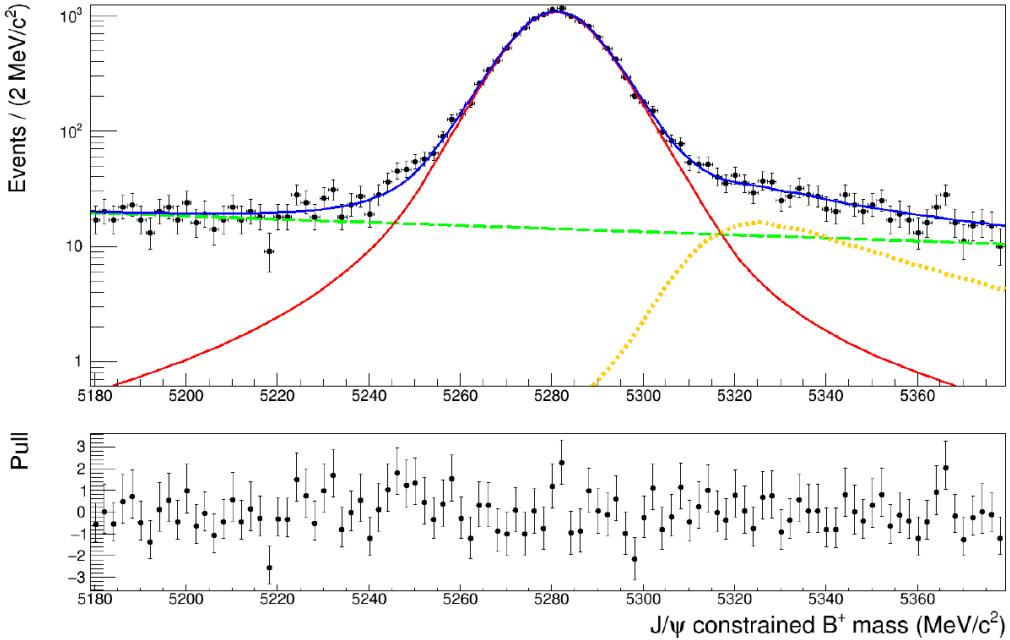


Figure 36: The reconstructed invariant mass spectrum of the $B^+ \rightarrow J/\psi K^+$ TIS sample from 2012 data(Zonneveld, 2016)

itself consists of two different overlapping effects, and the aggregate mis-ID background overlaps with the tails of both the B^0 and B_s^0 peaks, making it an exceptionally challenging shape to be fit. Therefore, the reconstructed invariant mass spectrum is difficult to fit accurately and the resulting particle identification efficiency would measure both the correct effect of signal event loss because of the PID cut, and the incorrect effect of differences in accuracy between the fits to the non-PID cut, and PID cut mass spectra, making the particle identification efficiency determined by the above method inaccurate.

6.4.2 Extracting the PID efficiency without knowing the signal's shape

Kappert (2016) describes a method of determining the particle identification efficiency without knowing the signal's shape. However, this method hinges on the assumption that the PID cut does not change the signal's shape in the mass spectrum. For the $B^0 \rightarrow K^\pm \pi^\mp$ decay's shape, the validity of this assumption is difficult to asses, as the tail parameters, because of the large combinatorial background in the data, can only be determined from the Monte Carlo simulated sample, which does not model the ΔLL variables

well. Therefore, the effect of the PID cut on the $B^0 \rightarrow K^\pm\pi^\mp$ signal shape cannot be established.

6.4.3 PIDCalib

Another alternative, and the method that was decided to be used for the $B^0 \rightarrow K^\pm\pi^\mp$ analysis, makes use of the PIDCalib package. PIDCalib relies on a data-driven method for determining PID efficiencies (Anderlini et al., 2016). Moreover, PIDCalib is a database that consists of samples of decays that are used for the calibration of cuts based on ΔLL variables. These decays are so-called “golden modes”, that can be reconstructed without the use of the RICH detectors. For the $B^0 \rightarrow K^\pm\pi^\mp$ decay, the efficiency of PID cuts on kaons and pions needs to be determined. The decay used that establishes the identities of kaons and pions without using the rich detector is a D^{*+} decay, namely

$$D^{*+} \rightarrow D^0\pi^+. \quad (35)$$

The decay can be accurately selected by stringent cuts on the identity on the of the π^+ , as the π^+ moves slowly with respect to the D^0 . and the charge of the π^+ can be measured. The D^0 subsequently decays according to

$$D^0 \rightarrow K^-\pi^+, \quad (36)$$

where the kaons and pions can be identified based on their charge alone, with 99,7% certainty (which is ratio between the $D^0 \rightarrow K^-\pi^+$ and $D^0 \rightarrow K^+\pi^-$ branching fractions (Olive, 2016)), as the pions of the decays in equations (35) and (36) must have identical charge (a pair of negatively charged pions is also possible, for the decay of the D^{*+} meson’s anti particle D^{*-} , but the pions’ charges are always identical with 99,7% certainty). Thus, from the decays in equations (35) and (36), a pure sample of pions and kaons can be obtained, based on the charges of the particles alone, and, thus, without use of information from the RICH detectors. Therefore, these pions and kaons can be used to calibrate the efficiency of the PID cuts, by applying the pion and kaon requirements cuts and calculating what percentage of true kaons and pions is lost.

6.4.4 Performance histograms

It is assumed that the quantities that determine the ΔLL value of a particular track can be reduced to a small set of independent variables. The PID-Calib samples, obtained from the corresponding calibration channels, can

be visualized as a high-dimensional space, consisting of PID efficiencies for the different combinations of parameters attached to a $D^{*+} \rightarrow D^0(K^-\pi^+)\pi^+$ decay. Following the above assumption, this high-dimensional space is reduced into a lower-dimensional “performance histogram” for a certain PID cut ($\Delta LL_K > 10$ for the kaon and $\Delta LL_K < -10$ for the pion, for this analysis), which are subsequently used to evaluate the PID efficiency for this PID cut on a certain signal sample. For the $B^0 \rightarrow K^\pm\pi^\mp$ analysis, the performance histograms were binned in three dimensions; the track momentum $|\mathbf{p}|$, the “pseudo-rapidity” η , and the track occupancy N_{tracks} . The pseudo-rapidity is defined as

$$\eta = \ln \left(\tan \left(\frac{\theta}{2} \right) \right), \quad (37)$$

which is preferred over the polar angle θ , as differences between different values of η are independent of the frame of reference. Moreover, the shape of the angular distribution, measured in terms of the pseudo-rapidity, is frame-independent (Tolk, 2016). PID efficiencies typically decrease for extreme values of $|\mathbf{p}|$ and η as that reduces the LCHb detector’s accuracy. Furthermore, the efficiency also decreases for increasing values of N_{tracks} , as busier events are measured more inaccurately as well.

6.4.5 Global PID weight determination

After constructing the performance histograms, the global PID efficiency (i.e. the efficiency for the entire signal sample) is determined by evaluating the performance histograms over a sample that is representative of the signal channel (specifically, representative of the $|\mathbf{p}|$, η and N_{tracks} distributions for the entire signal sample). The $B^0 \rightarrow K^\pm\pi^\mp$ invariant mass spectrum reconstructed from the data cannot be used for this evaluation, as the mass spectrum is dominated by the combinatorial background, and thus the PID efficiency would mainly measure the PID efficiency of combinatorial background events, instead of signal events. Therefore, another sample, that is representative of the signal sample, needs to be used.

As we have seen in the previous sections, the Monte Carlo simulated sample is typically the go-to sample, when a clean (i.e. background-free) sample, that is representative for the data signal sample, is needed. However, although the simulated sample represents the $|\mathbf{p}|$ and η distributions of the data signal sample well, the N_{tracks} distribution is not modeled accurately in the simulated sample. Namely, because the Monte Carlo simulates specifically the $B^0 \rightarrow K^\pm\pi^\mp$ decay channel, the simultaneous occurrence of other tracks is underrepresented in the simulated sample. Therefore, another sam-

ple that was previously used to represent the $B^0 \rightarrow K^\pm\pi^\mp$ sample as well, will be used to compensate for the misrepresentation of N_{tracks} in the Monte Carlo simulated sample.

Namely, N_{tracks} is a global event parameter that is independent of the specific signal channel considered (however, combinatorial background events have, on average, a higher N_{tracks} value, because when more tracks are present, the probability, of randomly combining tracks into a “fictitious” signal events, is higher, which is why we cannot use the $B^0 \rightarrow K^\pm\pi^\mp$ data sample). Therefore, the $B^+ \rightarrow J/\psi K^+$ data sample can be used to represent the N_{tracks} distribution of the $B^0 \rightarrow K^\pm\pi^\mp$ signal channel, as it has very little background. The N_{tracks} distributions of the $B^+ \rightarrow J/\psi K^+$ data sample and $B^0 \rightarrow K^\pm\pi^\mp$ simulated sample are shown in figure 37, showing that the two distribution clearly differ (i.e. the Monte Carlo sample indeed does not model N_{tracks} correctly).

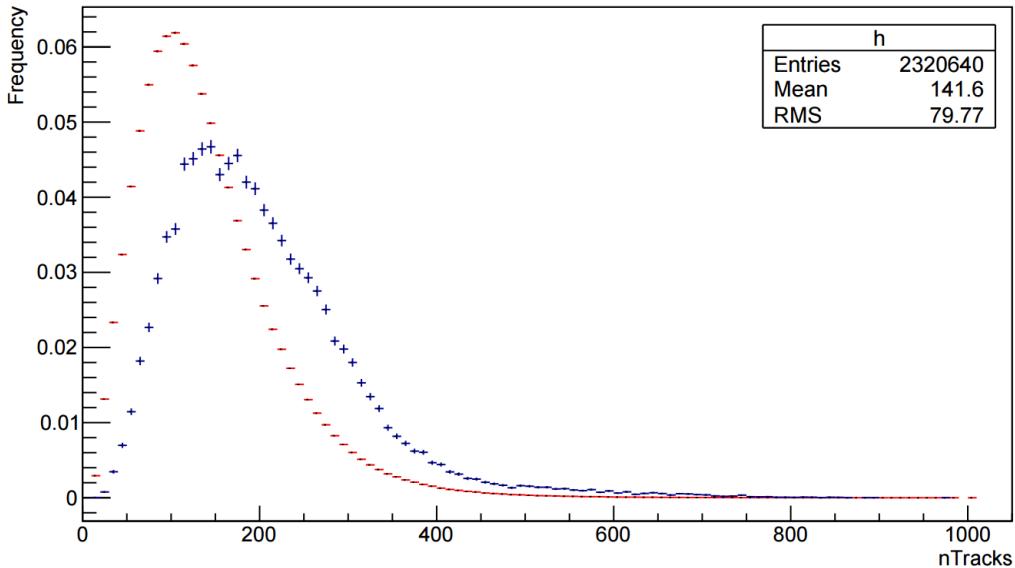


Figure 37: The n_{tracks} distributions of the data $B^+ \rightarrow J/\psi K^+$ sample (blue) and Monte Carlo simulated $B^0 \rightarrow K^\pm\pi^\mp$ sample (red)

As can be seen from figures 38 and 39, $|\mathbf{p}|$ and η are mostly uncorrelated with N_{tracks} (or we would expect a clear diagonal band to be present in figure 38 or 39). Therefore, by using the $B^+ \rightarrow J/\psi K^+$ sample, the Monte Carlo simulated sample can be reweighed, giving higher weights to events that have values of N_{tracks} that are underrepresented in the Monte Carlo sample, and lower weights to events that have values of N_{tracks} that are overrepresented in the Monte Carlo sample, such that the reweighed Monte Carlo sample has

the same N_{tracks} distribution as the $B^+ \rightarrow J/\psi K^+$ sample. Because $|\mathbf{p}|$ and η are not strongly correlated with N_{tracks} , this N_{tracks} reweighting does not significantly alter the distributions of $|\mathbf{p}|$ and η in the Monte Carlo sample. Thus, through reweighting the Monte Carlo simulated sample, a sample is obtained that accurately represents the $|\mathbf{p}|$, η and N_{tracks} distributions of the $B^0 \rightarrow K^\pm \pi^\mp$ signal sample. Subsequently, the performance histograms are evaluated over the reweighted Monte Carlo sample, yielding the global PID efficiency for the $\Delta LL_K > 10$ cut for the kaon and $\Delta LL_K < -10$ cut for the pion on the $B^0 \rightarrow K^\pm \pi^\mp$ data sample.

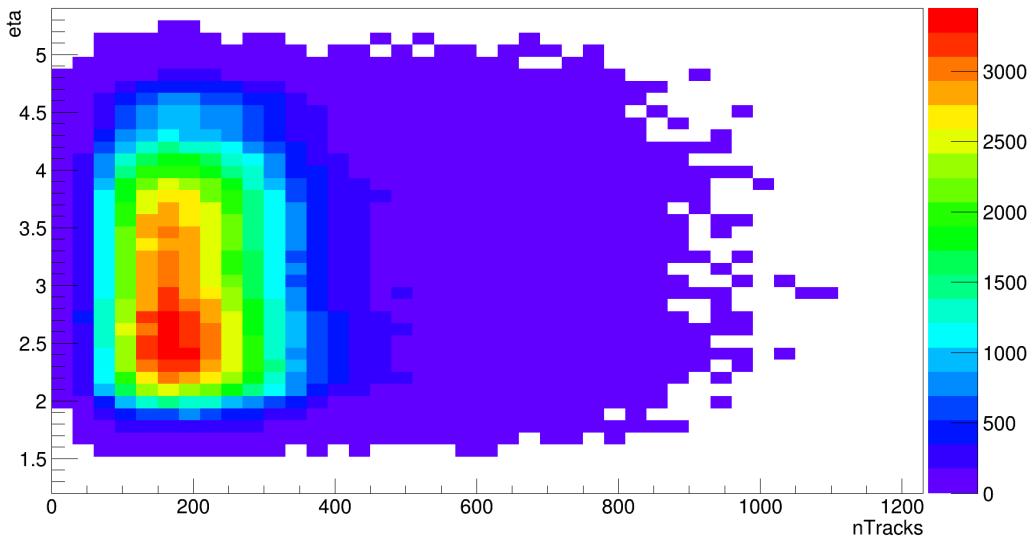


Figure 38: Plot of the correlation between η and N_{tracks}

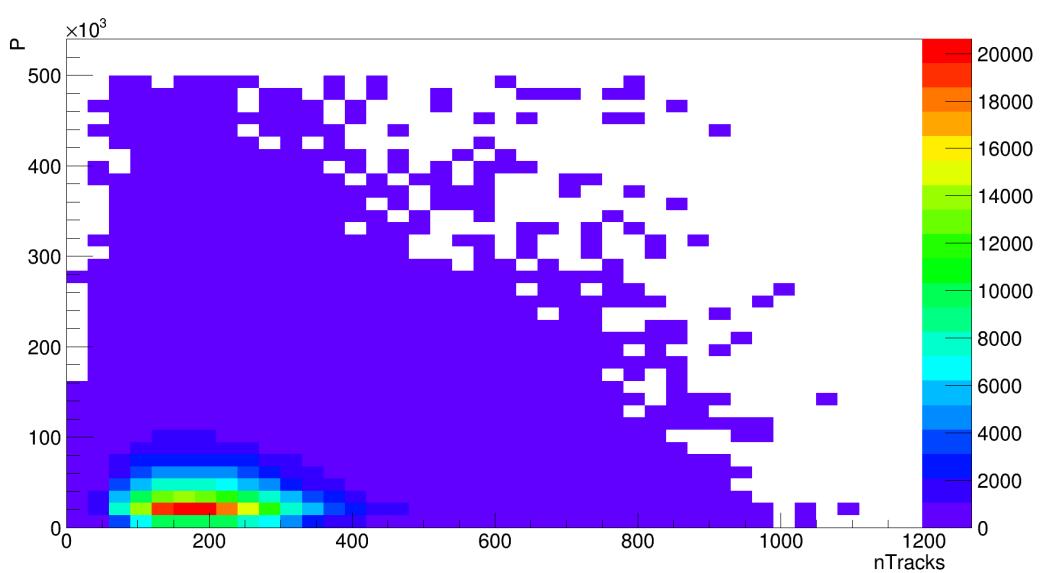


Figure 39: Plot of the correlation between $|\mathbf{p}|$ and N_{tracks}

7 Results

The determined yield and efficiencies and the resulting normalization factor α_{norm} are presented in table 3.

As can be seen from table 3, the acceptance and particle identification efficiencies are determined separately for both magnet polarizations (“MagUp” or “Magdown”), whereas the determined trigger efficiency, reconstruction&selection efficiency, and the yield are already averaged over both magnet polarizations. As we do not need to specifically determine the normalization factor for each magnet polarization, but only the average, we determine the average acceptance and particle identification efficiencies from the MagUp to MagDown luminosity ratios for 2011 and 2012 of, respectively, $\frac{\mathcal{L}_{2011}^{MagUp}}{\mathcal{L}_{2011}^{MagUp}} = 0.746$ and $\frac{\mathcal{L}_{2012}^{MagUp}}{\mathcal{L}_{2012}^{MagUp}} = 1.008$ (Adrover et al., 2013).

Furthermore, as the errors are on the efficiencies and yield are mainly statistical, calculations of uncertainties in this chapter are based on the propagation rules for normally-distributed errors (see section 8 for a discussion of systematic errors).

The following sections elaborate further on how the efficiencies, the yield, and the normalization factor α_{norm} are calculated.

7.1 Acceptance efficiency

Using the MagUp to MagDown luminosity ratios, we find that the resulting weighted average acceptance efficiency for the 2011 sample is

$$\epsilon_{2011}^{Acc} = \frac{0.746}{1 + 0.746} (17.73 \pm 0.04)\% + \frac{1}{1 + 0.746} (17.74 \pm 0.03)\% = (17.74 \pm 0.02)\%, \quad (38)$$

and for the 2012 sample is

$$\epsilon_{2012}^{Acc} = \frac{1.008}{1 + 1.008} (19.04 \pm 0.08)\% + \frac{1}{1 + 1.008} (18.92 \pm 0.07)\% = (18.98 \pm 0.05)\%, \quad (39)$$

where we have used that the weighted errors add quadratically when adding the weighted efficiencies.

7.2 Trigger efficiency

$\epsilon^{TIS,L0HLT1}$ is the combined efficiency of an event that contains a $B^0 \rightarrow K^\pm \pi^\mp$ decay to trigger both the L0 and HLT1 triggers independently from the $B^0 \rightarrow K^\pm \pi^\mp$ “signal” decay. Furthermore, $\epsilon^{TOS,HLT2}$ is the efficiency with which a $B^0 \rightarrow K^\pm \pi^\mp$ decay triggers the HLT2 trigger. As the TIS and TOS

Table 2: Efficiencies, yield, and normalization factor estimates

2011			
	MagUp	MagDown	Average
ϵ^{Acc}	(17.73 \pm 0.04)%	(17.74 \pm 0.03)%	(17.74 \pm 0.02)%
$\epsilon^{TIS,L0HLT1}$			(4.97 \pm 0.04)%
$\epsilon^{TOS,HLT2}$			(72.31 \pm 0.38)%
ϵ^{TRIG}			(3.593 \pm 0.035)%
$\epsilon^{RecSel Acc}$			(23.29 \pm 0.05)%
$\epsilon^{PID,B^0 \rightarrow K^+ \pi^-}$	(39.35 \pm 0.27)%	(39.36 \pm 0.01)%	
$\epsilon^{PID,B^0 \rightarrow K^- \pi^+}$	(39.12 \pm 0.05)%	(39.43 \pm 0.01)%	
ϵ^{PID}	(39.235 \pm 0.14)%	(39.40 \pm 0.01)%	(39.326 \pm 0.080)%
ϵ_{norm}			(0.0584 \pm 0.0006)%
N_{norm}			2925 \pm 132
B_{norm}			(19.6 \pm 0.5) $\cdot 10^{-6}$
α_{norm}			(3.91 \pm 0.21) $\cdot 10^{-12}$
2012			
	MagUp	MagDown	Average
ϵ^{Acc}	(19.04 \pm 0.08)%	(18.92 \pm 0.07)%	(18.98 \pm 0.05)%
$\epsilon^{TIS,L0HLT1}$			(5.69 \pm 0.03)%
$\epsilon^{TOS,HLT2}$			(75.49 \pm 0.10)%
ϵ^{TRIG}			(4.295 \pm 0.023)%
$\epsilon^{RecSel Acc}$			(21.81 \pm 0.01)%
$\epsilon^{PID,B^0 \rightarrow K^+ \pi^-}$	(37.35 \pm 0.24)%	(38.18 \pm 0.07)%	
$\epsilon^{PID,B^0 \rightarrow K^- \pi^+}$	(37.54 \pm 0.3)%	(38.36 \pm 0.00)%	
ϵ^{PID}	(37.45 \pm 0.19)%	(38.27 \pm 0.03)%	(37.855 \pm 0.098)%
ϵ_{norm}			(0.0673 \pm 0.0006)%
N_{norm}			7814 \pm 211
B_{norm}			(19.6 \pm 0.5) $\cdot 10^{-6}$
α_{norm}			(1.69 \pm 0.06) $\cdot 10^{-12}$

Note: Empty cells indicate undetermined values, that were not needed for the calculation of the normalization factor.

trigger probabilities are independent from each other, the total trigger efficiency is simply the product of the combined L0 and HLT1 TIS efficiency and the HLT2 TOS efficiency. Thus, the total trigger efficiency for 2011 is

$$\epsilon_{2011}^{TRIG} = (4.97 \pm 0.04)\% \cdot (72.31 \pm 0.38)\% = (3.593 \pm 0.035)\%, \quad (40)$$

and for 2012 is

$$\epsilon_{2012}^{TRIG} = (5.69 \pm 0.03)\% \cdot (75.49 \pm 0.10)\% = (4.295 \pm 0.023)\%, \quad (41)$$

where we have used that the relative errors add quadratically when multiplying the efficiencies.

7.3 Reconstruction&Selection efficiency

Under 2011 conditions, out of 775,505 simulated $B^0 \rightarrow K^\pm\pi^\mp$ events that are within the detector acceptance, 180,611 events are selected and reconstructed, from which we find a selection and reconstruction efficiency of $\epsilon_{2011}^{RecSel|Acc} = (23.29 \pm 0.05)\%$. Furthermore, under 2012 conditions, out of 10,543,584 simulated $B^0 \rightarrow K^\pm\pi^\mp$ events that are within the detector acceptance, 2,299,768 events are selected and reconstructed, and the selection and reconstruction efficiency is found to be $\epsilon_{2011}^{RecSel|Acc} = (21.81 \pm 0.01)\%$. Moreover, the (statistical) errors are calculated using the \sqrt{N} law (i.e. for $N \rightarrow \infty$, typical fluctuations are of the order \sqrt{N}).

7.4 Particle identification efficiency

The PID efficiency is determined separately not only for both magnet polarizations, but also for both decay modes $B^0 \rightarrow K^+\pi^-$ and $\bar{B}^0 \rightarrow K^-\pi^+$. However, as the $B^0 \rightarrow K^+\pi^-$ and $\bar{B}^0 \rightarrow K^-\pi^+$ decays occur at equal rate⁸ we can simply take the (unweighted) average over the efficiencies for both decays. Furthermore, the average over MagUp and MagDown is again calculated using the luminosity ratios, such that the PID efficiency for 2011 is found to equal

$$\epsilon_{2011}^{PID} = \frac{0.746}{1.746}(39.235 \pm 0.14)\% + \frac{1}{1.746}(39.40 \pm 0.01)\% = (39.326 \pm 0.080)\%, \quad (42)$$

and for 2012 is found to equal

$$\epsilon_{2012}^{PID} = \frac{1.008}{2.008}(37.45 \pm 0.19)\% + \frac{1}{2.008}(38.27 \pm 0.03)\% = (37.855 \pm 0.098)\%. \quad (43)$$

7.5 Yield

Figure 40 shows the final fits to the $B^0 \rightarrow K^\pm\pi^\mp$ mass spectrum. The $B^0 \rightarrow K^\pm\pi^\mp$ yield of the 2011 sample, which corresponds to the green peak in the top panel of figure 40, is $N_{norm}^{2011} = 2925 \pm 132$ events. Furthermore, the $B^0 \rightarrow K^\pm\pi^\mp$ yield of the 2012 sample, which corresponds to the green peak in the bottom panel of figure 40, is $N_{norm}^{2012} = 7814 \pm 211$ events.

⁸ b and \bar{b} quarks, which are produced in pairs, hadronize into, respectively, B^0 and \bar{B}^0 mesons at the same rate and the $\mathcal{B}(B^0 \rightarrow K^+\pi^-)$ and $\mathcal{B}(\bar{B}^0 \rightarrow K^-\pi^+)$ branching fractions are equal.

7.6 Normalization factor α_{norm}

The normalization factor α_{norm} is calculated according to

$$\alpha_{\text{norm}} = \epsilon_{\text{norm}} \cdot \frac{\mathcal{B}_{\text{norm}}}{N_{\text{norm}}} = \epsilon_{\text{norm}}^{\text{Acc}} \cdot \epsilon_{\text{norm}}^{\text{Trig}} \cdot \epsilon_{\text{norm}}^{\text{RecSel}} \cdot \epsilon_{\text{norm}}^{\text{PID}} \cdot \frac{\mathcal{B}_{\text{norm}}}{N_{\text{norm}}}, \quad (44)$$

where the subscript norm denotes the $B^0 \rightarrow K^\pm \pi^\mp$ normalization channel. Using the (total) detection efficiencies of $\epsilon_{\text{norm}}^{2011} = (0.0584 \pm 0.0006)\%$ and $\epsilon_{\text{norm}}^{2012} = (0.0673 \pm 0.0006)\%$, we find that the normalization factor for the 2011 sample is equal to $\alpha_{\text{norm}}^{2011} = (3.91 \pm 0.21) \cdot 10^{-12}$ and for the 2012 sample is equal to $\alpha_{\text{norm}}^{2012} = (1.69 \pm 0.06) \cdot 10^{-12}$.

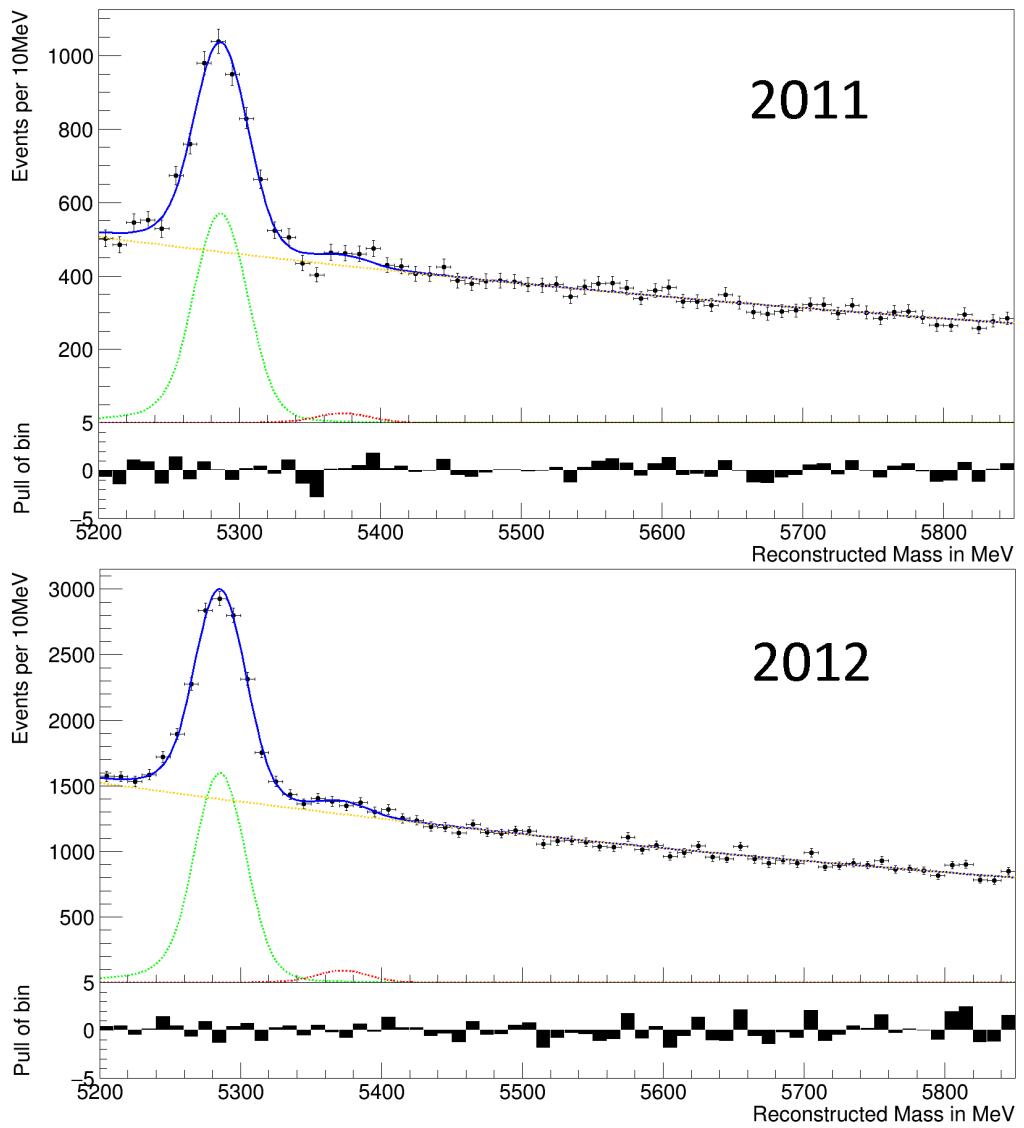


Figure 40: The top and bottom panels show the final fit to the $B^0 \rightarrow K^\pm \pi^\mp$ reconstructed invariant mass spectrum for, respectively, the 2011 and 2012 samples.

8 Discussion

In this chapter, we discuss the results presented in the previous section, specifically, comparing differences between the results for 2011 and 2012, and comparing the results of this analysis to those of the previous $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis. Furthermore, we discuss the limitations of the methodology used to determine these results. To facilitate this discussion, table 3 summarizes the estimates presented in the previous section and, specifically, lists their relative errors, as they directly determine the (relative) errors on the calculated normalization factors.

Table 3: Summary of estimates and their absolute and relative errors

Year	ϵ_{norm}^{Acc}	ϵ_{norm}^{Trig}	ϵ_{norm}^{RecSel}	ϵ_{norm}^{PID}	ϵ_{norm}	N_{norm}	\mathcal{B}_{norm}	α_{norm}
2011	0.1774	0.0359	0.2329	0.3933	$5.84 \cdot 10^{-4}$	2925	$19.6 \cdot 10^{-6}$	$3.91 \cdot 10^{-12}$
± abs.	0.0002	0.0003	0.0005	0.0008	$0.06 \cdot 10^{-4}$	132	$0.5 \cdot 10^{-6}$	$0.21 \cdot 10^{-12}$
± rel.	0.11%	0.96%	0.21%	0.20%	1.01%	4.51%	2.6%	5.28%
2012	0.1898	0.0430	0.2181	0.3786	$6.73 \cdot 10^{-4}$	7814	$19.6 \cdot 10^{-6}$	$1.69 \cdot 10^{-12}$
± abs.	0.0005	0.0002	0.0001	0.0010	$0.04 \cdot 10^{-4}$	211	$0.5 \cdot 10^{-6}$	$0.06 \cdot 10^{-12}$
± rel.	0.26%	0.54%	0.05%	0.26%	0.66%	2.70%	2.6%	3.77%

Note: ± abs. and ± rel. denote, respectively, the absolute and relative errors.

8.1 Comparison of results

For a very rare (or nonexistent) decay, such as $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$, the efficiencies should ideally be large, to increase the probability of observing the decay⁹. However, for a relatively strong decay, such as the $B^0 \rightarrow K^\pm \pi^\mp$ channel, the magnitudes of the efficiencies are far less relevant, and their errors more important, as the latter determine the accuracy with which the normalization factor α_{norm} can be determined. The *relative* errors are of particular importance, because the relative error on the normalization factor α_{norm} is given by the root of the summed squares of the efficiencies' and yield's relative errors (as the errors are mainly statistical).

8.1.1 Comparison between 2011 and 2012

Comparing our results for 2011 with those for 2012 in table 3, we find that the total detection efficiency is higher in 2012 and its relative error smaller, although this is not true for all individual efficiencies. Namely, the PID

⁹If the decay is not observed, the analysis can only determine an upper bound on the branching fraction, which is determined by the *lower* bound on the detection efficiency. Thus, both the magnitude and accuracy of the detection efficiency are important for determining the branching fraction of a very rare (or nonexistent) decay.

and reconstruction&selection efficiencies are higher for 2011, and the relative errors of the acceptance and PID efficiencies smaller.

Furthermore, table 3 shows that the yields in 2011 and 2012 have the largest relative errors of all components of the normalization factor α_{norm} , and are therefore the main determinant of the inaccuracy of α_{norm} . Nevertheless, the relative error of the yield for 2012 is smaller than for 2011, which is likely to be at least partially due to the larger yield for 2012, making the fit to the invariant mass spectrum *relatively* more accurate. Moreover, as can be seen from the pull-plots in figure 40, the pull of the 5350-5360 MeV/ c^2 bin for 2011 is rather large, showing that the 2011 fit fails to accurately follow the data in this bin (between the B^0 and B_s^0 peak), which is not the case for the 2012 fit.

Thus, improving the fit to the mass spectrum has the potential to greatly improve the accuracy of the normalization factor α_{norm} . Specifically, the aforementioned “structure” on top the combinatorial background to the left of the $B^0 \rightarrow K^\pm\pi^\mp$ peak was cut away by taking a narrower mass-window in this analysis, as its shape was not well-understood and needed to be determined from the data itself (by estimating the combinatorial background underneath the structure, which limits the contribution of this region of the mass-window to the estimation of the combinatorial background underneath the $B^0 \rightarrow K^\pm\pi^\mp$ peak). If this structure is better understood, such that its shape can be determined from theory and not from the data itself, this region in the mass spectrum could contribute to estimation of the background underneath the $B^0 \rightarrow K^\pm\pi^\mp$, and to the improvement of the accuracy of the fit overall.

The second largest contributor to the inaccuracy of the normalization factor α_{norm} , of the determined efficiencies and yield, is the trigger efficiency, specifically the combined trigger efficiency of the L0 and HLT1 triggers. As the L0 and HLT1 triggers are hardware-driven, their efficiencies are notoriously difficult to estimate (the TISTOS method was specifically developed for this). Therefore, a larger relative error was to be expected and cannot straightforwardly be improved upon. Furthermore, the accuracy of the $\mathcal{B}(B^0 \rightarrow K^\pm\pi^\mp)$ branching fraction also has an effect on the accuracy of the normalization factor α_{norm} , which is actually stronger than that of the trigger efficiency. However, improving the accuracy of the $\mathcal{B}(B^0 \rightarrow K^\pm\pi^\mp)$ branching fraction is beyond the scope of this analysis, as this would require a complete analysis of its own.

The relative errors on the PID, reconstruction&selection, and acceptance efficiencies are negligible compared to the relative error on the yield. Therefore, the accuracy of these efficiencies shows little need for improvement in future analyses. Nevertheless, the ambitious goal of the next $B_{(s)}^0 \rightarrow e^\pm\mu^\mp$

analysis, using Run II data, of lowering the upper limit on the branching fraction by multiple orders of magnitude will require all efficiencies to be determined as accurately as possible.

The combination of smaller relative errors on both the total detection efficiency and yield for 2012, results in a smaller relative error on the normalization factor α_{norm} for 2012 than for 2011, showing a modest improvement over time. Namely, the relative error decreases by almost a factor 1.5 times by moving from the 2011 data to the 2012 data.

8.1.2 Comparison with previous $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis

If we now compare our results to those of the previous $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis (Aaij et al., 2013), we find slight differences in the estimated efficiencies, which was to be expected, because of different choices made in the analysis, such as with regard to the stripping. Furthermore, we find that the relative errors on the efficiencies in the current analysis have strongly improved with respect to the previous analysis, as they have decreased by about a factor 5 on average.

The previous analysis took a different approach to determining the yield of the $B^0 \rightarrow K^\pm \pi^\mp$ channel, namely, by fitting the total $B^0 \rightarrow h^\pm h'^\mp$ yield, where h is a kaon or a pion, and using the fraction $f_{B^0 \rightarrow K^\pm \pi^\mp} = (60.5 \pm 2.7)\%$ determined by Aaij et al. (2012). Therefore, we cannot compare the PID efficiencies of the current analysis to the previous analysis, nor the yield.

Nevertheless, by calculating the PID-reweighed $B^0 \rightarrow K^\pm \pi^\mp$ yield for the current analysis, we can compare its accuracy to that of the yield of the previous analysis. The relative error on the previous analysis' yield is over 9%, whereas the relative errors on the yield for 2011 and 2012 in this analysis are smaller than 5% and 3%, respectively. This is largely due to the larger relative error on the $B^0 \rightarrow h^\pm h'^\mp$ yield in the previous analysis than the relative errors on the $B^0 \rightarrow K^\pm \pi^\mp$ yields in this analysis, as ϵ^{PID} and $f_{B^0 \rightarrow K^\pm \pi^\mp}$ have small relative errors and do not significantly decrease the accuracy of the yield. Thus, we can conclude that the use of the $B^0 \rightarrow K^\pm \pi^\mp$ yield fit significantly increases the accuracy of the yield with respect to the $B^0 \rightarrow h^\pm h'^\mp$ fit used in the previous $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis.

Furthermore, although the previous analysis did not specifically calculate α_{norm} , as the required signal channel data were available to directly determine the total normalization factor α , we can use the quoted efficiencies and yield to calculate α_{norm} for the previous analysis and compare its accuracy to that of the α_{norm} normalization factors presented in this paper. In doing so, we find that the relative errors on the normalization factors for 2011 and 2012 of the current analysis are about 2, respectively, 3 times smaller than the

relative error on α_{norm} for the previous analysis.

Having discussed the improvements in the accuracy of the normalization factor α_{norm} , it is important to reiterate that the choices made in this analysis were not aimed at specifically improving the accuracy of the normalization factor α_{norm} , but at improving the overall accuracy of the determination of the branching fraction $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$, to which α_{norm} contributes. Specifically, the selection of $B^0 \rightarrow K^\pm \pi^\mp$ events was aimed at being as similar as possible to the signal channel, to cancel out systematic errors. Therefore, it is theoretically possible that accuracy in the calibration was sacrificed to gain overall accuracy. Thus, only when the complete $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$ is ready, we can definitively determine if and how much the current analysis has improved over the previous one and how much accuracy is gained from using the data from 2012 instead of 2011.

Furthermore, although we have concluded that the yield fit is the strongest determinant of the accuracy of α_{norm} , as the accuracy of the calculated branching fraction $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$ is given by the root of the summed squares of the relative errors of the components in equation 3, how much the accuracy of α_{norm} affects the accuracy of the $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$ branching fraction depends on the accuracies of the other components. Specifically, as the electron trigger and PID efficiencies are typically difficult to determine, these are likely to have a strong effect on the accuracy of the $\mathcal{B}(B_{(s)}^0 \rightarrow e^\pm \mu^\mp)$ branching fraction.

8.2 Limitations

In the determination of the PID efficiency and the $B^0 \rightarrow K^\pm \pi^\mp$ yield, we did not estimate systematic errors. The determination of systematic errors is non-trivial, typically requiring some alternative estimation method and comparing the results, which was not possible within the scope of this research. For a detailed discussion and analysis of the systematic errors involved in using the $B^0 \rightarrow K^\pm \pi^\mp$ decay as the normalization channel, see Adrover et al. (2013), which uses the $B^0 \rightarrow K^\pm \pi^\mp$ channel as normalization for the $B_{(s)}^0 \rightarrow \mu^\pm \mu^\mp$ analysis.

Furthermore, it is important to be aware of the assumptions that were made in this analysis, especially because we are searching for beyond the Standard Model physics, but base many of our assumptions on the validity of Standard Model physics. The limitations of this research mainly pertain to how accurately the Monte Carlo simulations (on which this analysis relies to determine efficiencies and overcome challenges resulting from the large combinatorial background in the data caused by the TIS trigger requirement) model the true physics taking place in the LHCb detector. Namely, the

simulated data were used both in the evaluation of the efficiencies, and for constraining the tail parameters of the B^0 and B_s^0 peaks in the fit to the invariant mass spectrum, as the large combinatorial background made it impossible to determine the tail parameters from the data. However, as the simulation does not model ΔLL variables well (which is also why PIDCalib was used to evaluate the PID efficiencies instead), we could not determine the effect of the PID cut on the shape of the B^0 and B_s^0 peaks. Therefore, we used simulated data, without having a PID cut applied, to determine the tail parameters used in the fit to the 2011 and 2012 data, which did have a PID applied. Thus, if the PID cut changes the shape of the B^0 and B_s^0 peaks, using the tail parameters from the simulated data will introduce a systematic error in the determined yields.

Alternatively, the tail parameters could be determined from the invariant mass spectrum of the $B^0 \rightarrow K^\pm \pi^\mp$ channel having the TOS trigger requirement, which would strongly decrease the combinatorial background and make it easier to estimate the tail parameters from the data. However, the possibility that the trigger requirement affects the B^0 and B_s^0 peak shapes cannot be excluded. Nevertheless, future analyses could (and should) use such an approach to determine the tail parameters and compare the resulting yield to yield that results from using the simulated data to determine the tail parameters, and use the difference between the two yields to estimate the systematic error.

A final limitation of this analysis worth mentioning, is that by cutting off the part of the mass window to the left of the B^0 and B_s^0 peaks, the combinatorial background underneath the B^0 and B_s^0 peaks is determined by extrapolating the background on the right of the B^0 and B_s^0 peaks. If the region to the left of B^0 and B_s^0 peaks in the mass windows is better understood, such that it can be accurately modeled based on theory, the combinatorial background underneath the B^0 and B_s^0 peaks can be interpolated using the backgrounds on the left and right of B^0 and B_s^0 peaks.

Nevertheless, based on the analysis of the statistical and systematic errors involved in the analysis of the $B^0 \rightarrow K^\pm \pi^\mp$ channel for the search for the $B_s^0 \rightarrow \mu^\pm \mu^\mp$ and $B^0 \rightarrow \mu^\pm \mu^\mp$ decays by Adrover et al. (2013) (which employs a methodology that is very similar to the analysis of the $B^0 \rightarrow K^\pm \pi^\mp$ channel presented in this paper), we do not expect the systematic errors to dominate over the statistical errors.

9 Summary and outlook

Charged lepton flavor violating decays are heavily suppressed in the Standard Model, and are used to probe for New Physics. This thesis discussed using the $B^0 \rightarrow K^\pm\pi^\mp$ channel as normalization in the search for the charged lepton flavor violating $B^0 \rightarrow e^\pm\mu^\mp$ and $B_s^0 \rightarrow e^\pm\mu^\mp$ decays. The data used were collected by the LHCb detector during LHC Run I, in the years 2011 and 2012. The aim of this thesis was to determine the $B^0 \rightarrow K^\pm\pi^\mp$ channel's yield, and the acceptance, trigger, reconstruction&selection, and PID efficiencies, to determine normalization factor $\alpha_{norm} = \mathcal{B}_{norm} \cdot \frac{\epsilon_{norm}}{N_{norm}}$. The normalization factor α_{norm} is used in the search for the charged lepton flavor violating $B^0 \rightarrow e^\pm\mu^\mp$ and $B_s^0 \rightarrow e^\pm\mu^\mp$ decays, to determine the branching fractions $\mathcal{B}(B^0 \rightarrow e^\pm\mu^\mp)$ and $\mathcal{B}(B_s^0 \rightarrow e^\pm\mu^\mp)$, by using $\mathcal{B}(B^0 \rightarrow e^\pm\mu^\mp) = \alpha_{norm} \cdot \frac{N_{sig}}{\epsilon_{sig}}$ and $\mathcal{B}(B_s^0 \rightarrow e^\pm\mu^\mp) = \alpha_{norm} \cdot \frac{f_d}{f_s} \cdot \frac{N_{sig}}{\epsilon_{sig}}$.

To determine the $B^0 \rightarrow K^\pm\pi^\mp$ yield, a maximum likelihood fit was used, describing the B^0 and B_s^0 peaks using double Crystal ball functions, and using an exponential function to model the combinatorial background. A narrow mass window was used, excluding the region directly to the left of the B^0 peak, as events in this region were affected by an up-stream cut on the reconstructed invariant mass, and this region was dominated by a poorly understood peaking background, leading to the region to not contribute to the estimate of the combinatorial background underneath the B^0 peak (which is the only purpose including the region in the fit serves).

The HLT2 and reconstruction&selection efficiencies were determined using Monte Carlo simulated data, by running the simulated data through these processes and comparing the number of “surviving” events to the total number of generated events. The combined L0 and HLT1 trigger efficiency is determined using a data-driven method called the TISTOS method, as the simulated data do not model the variables on which L0 and HLT1 trigger decisions are based accurately. The TIS requirement causes the $B^0 \rightarrow K^\pm\pi^\mp$ invariant mass spectrum to be dominated by background, but as the TIS efficiency is channel-independent, another b decay, the $B^+ \rightarrow J/\psi K^+$ channel, which has negligible background, can be used to determine the TIS efficiency.

The PID efficiency is also determined using a data-driven method, because the simulated data do not model the ΔLL variables, on which the PID cuts are based, well. The PID efficiency is determined using PIDCalib, which is a database that consists of samples of decays that are used for the calibration of cuts based on ΔLL variables. These decays are so-called “golden modes”, that can be reconstructed without the use of the RICH detectors. We determined a global PID efficiency by constructing three-dimensional per-

formance histograms, binned in $|\mathbf{p}|$, η , and N_{tracks} , and evaluating these over the $|\mathbf{p}|$, η , and N_{tracks} distributions of the $B^0 \rightarrow K^\pm \pi^\mp$ samples. Because of the large presence of combinatorial background events in our $B^0 \rightarrow K^\pm \pi^\mp$ samples, we use Monte Carlo simulated data to represent the $B^0 \rightarrow K^\pm \pi^\mp$ samples' $|\mathbf{p}|$ and η distributions, and, as the simulated data do not model N_{tracks} correctly, we use the $B^+ \rightarrow J/\psi K^+$ channel's N_{tracks} distribution to reweigh the simulated data to obtain the correct N_{tracks} distribution in the simulated sample (N_{tracks} is a global event parameter, so its is channel independent, and it is uncorrelated with $|\mathbf{p}|$ and η).

We find a normalization factor $\alpha_{\text{norm}}^{2011}$ of $(3.91 \pm 0.21) \cdot 10^{-12}$ for the 2011 sample, and a normalization factor $\alpha_{\text{norm}}^{2012}$ of $(1.69 \pm 0.06) \cdot 10^{-12}$ for the 2012 sample. The largest contributor to the uncertainty on $\alpha_{\text{norm}}^{2011}$ and $\alpha_{\text{norm}}^{2012}$ is the yield, with the second and third largest being the $\mathcal{B}(B^0 \rightarrow K^\pm \pi^\mp)$ branching fraction and the trigger efficiency, respectively. The acceptance, reconstruction&selection, and PID efficiencies had a negligible contribution to the branching fraction.

Given the expectation that no $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ events will be observed, the $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis aims to lower the upper limit on the branching fraction set by the previous $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ analysis by Aaij et al. (2013), with the current analysis being expected to lower the branching fraction within one order of magnitude. However, the analysis is set-up to straightforwardly translate to the Run II data from the LHC runs in 2015, and, using these data, we aim to lower the upper limit on the $B^0 \rightarrow e^\pm \mu^\mp$ and $B_{(s)}^0 \rightarrow e^\pm \mu^\mp$ branching fractions by multiple orders of magnitude.

Because of these ambitions, it is of the utmost importance to ensure that the methodologies used in this analysis are as accurate as possible. For the $B^0 \rightarrow K^\pm \pi^\mp$ normalization channel, we believe the greatest potential lies with improving the maximum likelihood yield fit to the invariant mass spectrum. Namely, widening the mass window by including the region to the left of the B^0 peak should improve estimates of the background underneath the B^0 , *if* the partial reconstruction structure is better understood, such that its shape can be determined based on theory and not from the data, and the aforementioned upstream invariant mass cuts are circumvented. Furthermore, although this is beyond the scope of this analysis, reducing the uncertainty on the $\mathcal{B}(B^0 \rightarrow K^\pm \pi^\mp)$ branching fraction will also strongly improve the accuracy of α_{norm} . Lastly, reducing the uncertainty on the trigger efficiency would also improve the accuracy of α_{norm} , but as the estimation of the L0 and HLT1 trigger efficiencies is rather challenging, we cannot propose any straightforward solutions for reducing its uncertainty.

10 Acknowledgments

I would like to thank my supervisor Gerco Onderwater for his inspiring enthusiasm and involvement, which has been a main driver of my thorough enjoyment of this research project, and for trusting me with the responsibilities and freedom that have made this project so very educational.

Furthermore, I am grateful to Mick Mulder for all his time and effort spent on explaining and discussing the relevant intricacies of the $B_{(s)}^0 \rightarrow \mu^\pm \mu^\mp$ analysis, and to Rosa Kappert for helping me develop the programming skills necessary for this analysis.

Finally, I would like to thank Antonio Pellegrino for taking me to see the LHCb detector from up-close, which proved to be the initial spark for my interests in this field of research, and Marcel Merk for facilitating the many fruitful days I spent at Nikhef.

11 Appendix I: Mass defect caused by failing to include the π^0 when reconstructing $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ events

Summary

Failing to include the π^0 when reconstructing a $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ event results in a reconstructed invariant mass of the B^0 meson that is significantly smaller than its actual correct mass. This causes a peaking background in the $B^0 \rightarrow K^\pm\pi^\mp$ mass spectrum left of the B^0 mass peak. In this appendix, it will be shown that the minimum mass defect when missing a π^0 is achieved when the π^0 has zero momentum in the center of mass frame (i.e. the frame where the mother B^0 meson is at rest) and that when this is the case, the mass defect in the reconstruction of the B^0 meson's invariant mass is exactly the mass of the π^0 . These findings are valid for any three-body decay reconstructed as a two body decay, i.e., for any such decay, the minimum mass defect is exactly equal to the omitted daughter particle's mass and occurs when the daughter particle omitted in the reconstruction is at rest in the center of mass frame of the mother particle.

Motivation

As can be seen from figure 41, which shows the B^0 invariant mass spectrum reconstructed from $B^0 \rightarrow K^\pm\pi^\mp$ events, there are significantly more events in the $5000 - 5200$ MeV/ c^2 range than the trend of the combinatorial background fit to the higher energy ranges predicts. The dominant cause is assumed to be partial reconstruction of $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ events. Failing to include the π^0 when reconstructing a $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ event results in a reconstructed invariant mass of the B^0 meson that is significantly smaller than its actual correct mass. One would assume this mass defect to be smallest when the missed π^0 has negligible momentum, such that the invariant mass reconstruction only lacks the mass-energy of the π^0 . This “partial reconstruction spectrum” tops off at around 5140 MeV/ c^2 , such that the difference with the B^0 mass (≈ 5280 MeV/ c^2) coincides quite nicely with the energy of a π^0 with zero momentum (i.e. its mass, ≈ 140 MeV/ c^2). As will be shown below, the mass defect in the reconstruction of the B_d meson's invariant mass when missing a π^0 is at its minimum when the π^0 has zero momentum in the center of mass frame (i.e. the frame where the mother B^0 meson is at rest). Moreover, it will be shown that when this is the case, the mass defect equals exactly the mass of the π^0 .

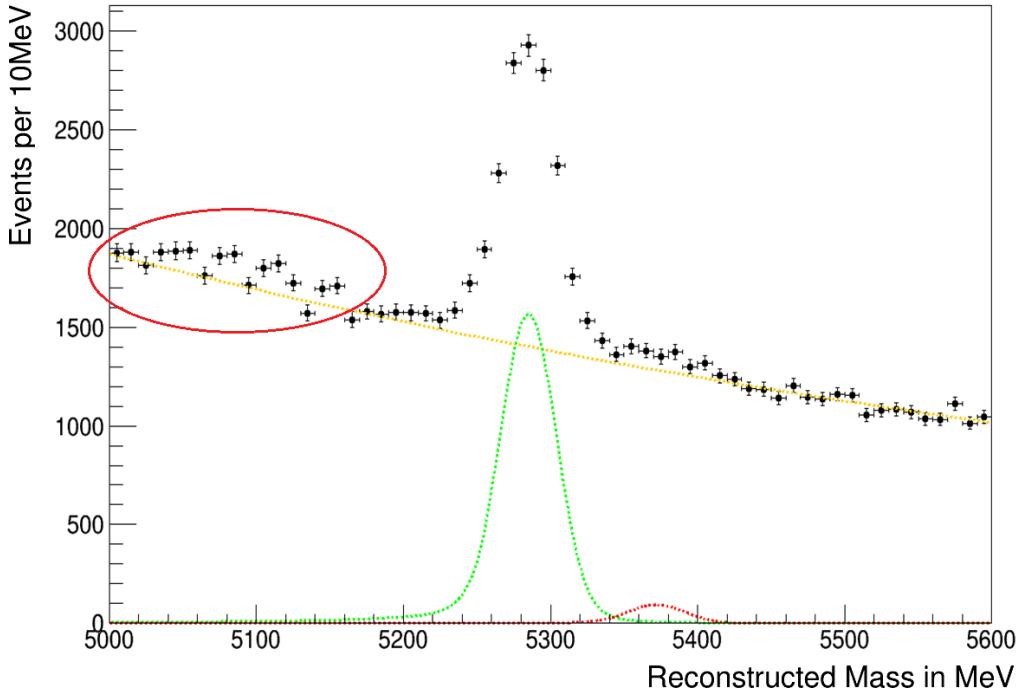


Figure 41: The invariant mass spectrum of the $B^0 \rightarrow K^\pm \pi^\mp$ channel (number of data-events per bin of $10 \text{ MeV}/c^2$). Green denotes the B^0 mass peak, red is the B_s^0 mass peak and yellow the combinatorial background. The red ellipse points out the partially reconstructed background structure.

Invariant mass reconstruction

The frame-independent expression for the invariant mass of a system is given by

$$m = \sqrt{E^2 - \mathbf{p}^2}. \quad (45)$$

When considering the decay of a single mother particle into multiple daughter particles, the invariant mass of the system of the mother particle is simply the mother particle's mass. Moreover, because the invariant mass of the system is conserved in the decay, we can express the invariant mass of the the mother particle in terms of momenta and energies of the daughter particles, as

$$m_{mother} = \sqrt{{E_{tot}}^2 - {\mathbf{p}_{tot}}^2} = \sqrt{\left(\sum E_{daughter}\right)^2 - \left(\sum \mathbf{p}_{daughter}\right)^2}. \quad (46)$$

For the $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ decay, the conservation of mass is expressed as

$$m_{B^0}^2 = (E_{K^\pm} + E_{\pi^\mp} + E_{\pi^0})^2 - (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp} + \mathbf{p}_{\pi^0})^2, \quad (47)$$

$$m_{B^0}^2 = E_{K^\pm}^2 + E_{\pi^\mp}^2 + E_{\pi^0}^2 + 2E_{K^\pm}E_{\pi^\mp} + 2E_{K^\pm}E_{\pi^0} + 2E_{\pi^\mp}E_{\pi^0} - \mathbf{p}_{K^\pm}^2 - \mathbf{p}_{\pi^\mp}^2 - \mathbf{p}_{\pi^0}^2 - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^0} - 2\mathbf{p}_{\pi^\mp} \cdot \mathbf{p}_{\pi^0}, \quad (48)$$

where \mathbf{p}^2 denotes the square of the absolute value of \mathbf{p} , i.e. $\mathbf{p}^2 = |\mathbf{p}|^2 = \mathbf{p} \cdot \mathbf{p}$. We would like to express the invariant mass in quantities determined by the detector. Therefore, using $E^2 = m^2 + \mathbf{p}^2$, we can eliminate the energy terms in equation 48. Namely,

$$m_{B^0}^2 = (E_{K^\pm}^2 - \mathbf{p}_{K^\pm}^2) + (E_{\pi^\mp}^2 - \mathbf{p}_{\pi^\mp}^2) + (E_{\pi^0}^2 - \mathbf{p}_{\pi^0}^2) + 2E_{K^\pm}E_{\pi^\mp} + 2E_{K^\pm}E_{\pi^0} + 2E_{\pi^\mp}E_{\pi^0} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^0} - 2\mathbf{p}_{\pi^\mp} \cdot \mathbf{p}_{\pi^0}, \quad (49)$$

$$m_{B^0}^2 = (m_{K^\pm}^2) + (m_{\pi^\mp}^2) + (m_{\pi^0}^2) + 2\sqrt{(m_{K^\pm}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{(m_{\pi^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)} + 2\sqrt{(m_{K^\pm}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{(m_{\pi^0}^2 + \mathbf{p}_{\pi^0}^2)} + 2\sqrt{(m_{\pi^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)}\sqrt{(m_{\pi^0}^2 + \mathbf{p}_{\pi^0}^2)} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^0} - 2\mathbf{p}_{\pi^\mp} \cdot \mathbf{p}_{\pi^0}, \quad (50)$$

and, using $\mathbf{p}_1 \cdot \mathbf{p}_2 = p_{1,x}p_{2,x} + p_{1,y}p_{2,y} + p_{1,z}p_{2,z}$, we can write equation 50 in such a way that we can express invariant mass in terms of quantities that are determined by the detector;

$$\begin{aligned} m_{B^0}^2 = & m_{K^\pm}^2 + m_{\pi^\mp}^2 + m_{\pi^0}^2 + \\ & 2\sqrt{(m_{K^\pm}^2 + p_{K^\pm,x}^2 + p_{K^\pm,y}^2 + p_{K^\pm,z}^2)}\sqrt{(m_{\pi^\mp}^2 + p_{\pi^\mp,x}^2 + p_{\pi^\mp,y}^2 + p_{\pi^\mp,z}^2)} + \\ & 2\sqrt{(m_{K^\pm}^2 + p_{K^\pm,x}^2 + p_{K^\pm,y}^2 + p_{K^\pm,z}^2)}\sqrt{(m_{\pi^0}^2 + p_{\pi^0,x}^2 + p_{\pi^0,y}^2 + p_{\pi^0,z}^2)} + \\ & 2\sqrt{(m_{\pi^\mp}^2 + p_{\pi^\mp,x}^2 + p_{\pi^\mp,y}^2 + p_{\pi^\mp,z}^2)}\sqrt{(m_{\pi^0}^2 + p_{\pi^0,x}^2 + p_{\pi^0,y}^2 + p_{\pi^0,z}^2)} - \\ & 2(p_{K^\pm,x}p_{\pi^\mp,x} + p_{K^\pm,y}p_{\pi^\mp,y} + p_{K^\pm,z}p_{\pi^\mp,z}) - 2(p_{K^\pm,x}p_{\pi^0,x} + \\ & p_{K^\pm,y}p_{\pi^0,y} + p_{K^\pm,z}p_{\pi^0,z}) - 2(p_{\pi^\mp,x}p_{\pi^0,x} + p_{\pi^\mp,y}p_{\pi^0,y} + p_{\pi^\mp,z}p_{\pi^0,z}). \end{aligned} \quad (51)$$

From equation 51, the invariant mass of the B^0 meson can be reconstructed, using the momenta in the x , y , and z directions determined by the detector and the masses according to the particle identities determined by the detector. Similarly, for the two-body decay $B^0 \rightarrow K^\pm \pi^\mp$, the invariant

mass of the mother B^0 meson can be expressed in terms of the momenta and masses of the the daughter particles, as

$$m_{B^0}^2 = (E_{K^\pm} + E_{\pi^\mp})^2 - (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp})^2, \quad (52)$$

$$m_{B^0}^2 = E_{K^\pm}^2 + E_{\pi^\mp}^2 + 2E_{K^\pm}E_{\pi^\mp} - \mathbf{p}_{K^\pm}^2 - \mathbf{p}_{\pi^\mp}^2 - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp}, \quad (53)$$

$$m_{B^0}^2 = (E_{K^\pm}^2 - \mathbf{p}_{K^\pm}^2) + (E_{\pi^\mp}^2 - \mathbf{p}_{\pi^\mp}^2) + 2E_{K^\pm}E_{\pi^\mp} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp}, \quad (54)$$

$$m_{B^0}^2 = m_{K^\pm}^2 + m_{\pi^\mp}^2 + 2\sqrt{(m_{K^\pm}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{(m_{\pi^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^\mp}, \quad (55)$$

$$\begin{aligned} m_{B^0}^2 &= m_{K^\pm}^2 + m_{\pi^\mp}^2 - 2(p_{K^\pm,x}p_{\pi^\mp,x} + p_{K^\pm,y}p_{\pi^\mp,y} + p_{K^\pm,z}p_{\pi^\mp,z}) + \\ &2\sqrt{(m_{K^\pm}^2 + p_{K^\pm,x}^2 + p_{K^\pm,y}^2 + p_{K^\pm,z}^2)}\sqrt{(m_{\pi^\mp}^2 + p_{\pi^\mp,x}^2 + p_{\pi^\mp,y}^2 + p_{\pi^\mp,z}^2)}. \end{aligned} \quad (56)$$

Mass defect for a π^0 with zero momentum

When the π^0 , for whatever reason, is not identified as part of the decay, the $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ decay is reconstructed according to 56, under the assumption that the event consisted of a $B^0 \rightarrow K^\pm \pi^\mp$ decay. When the invariant mass of a $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ decay is reconstructed according to equation 56, the difference, which we will denote as m_{defect}^2 , between the squares of the correct B^0 mass (equation 51) and of the reconstructed B^0 mass (equation 56), is

$$\begin{aligned} m_{B^0}^2 &= m_{\pi^0}^2 - 2(p_{K^\pm,x}p_{\pi^0,x} + p_{K^\pm,y}p_{\pi^0,y} + p_{K^\pm,z}p_{\pi^0,z}) - \\ &2(p_{\pi^\mp,x}p_{\pi^0,x} + p_{\pi^\mp,y}p_{\pi^0,y} + p_{\pi^\mp,z}p_{\pi^0,z}) + \\ &2\sqrt{(m_{K^\pm}^2 + p_{K^\pm,x}^2 + p_{K^\pm,y}^2 + p_{K^\pm,z}^2)}\sqrt{(m_{\pi^0}^2 + p_{\pi^0,x}^2 + p_{\pi^0,y}^2 + p_{\pi^0,z}^2)} + \\ &2\sqrt{(m_{\pi^\mp}^2 + p_{\pi^\mp,x}^2 + p_{\pi^\mp,y}^2 + p_{\pi^\mp,z}^2)}\sqrt{(m_{\pi^0}^2 + p_{\pi^0,x}^2 + p_{\pi^0,y}^2 + p_{\pi^0,z}^2)}, \end{aligned} \quad (57)$$

or, using inner product notation,

$$m_{defect}^2 = m_{\pi^0}^2 + 2\sqrt{(m_{K^\pm}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{(m_{\pi^0}^2 + \mathbf{p}_{\pi^0}^2)} + 2\sqrt{(m_{\pi^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)}\sqrt{(m_{\pi^0}^2 + \mathbf{p}_{\pi^0}^2)} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^0} - 2\mathbf{p}_{\pi^\mp} \cdot \mathbf{p}_{\pi^0}. \quad (58)$$

Please be aware that m_{defect}^2 denotes the difference between the squares of the correct and reconstructed B^0 masses, whereas m_{defect} (which will be used below) denotes the difference between the (non-squared) correct and reconstructed B_d masses. Thus, $\sqrt{m_{defect}^2} \neq m_{defect}$. This is a slight inconsistency in the notation, with the purpose of simplicity. Moreover, for a π^0 with zero momentum, in the center of mass frame, equation 58 equals

$$m_{defect}^2 = m_{\pi^0}^2 + 2m_{\pi^0}\sqrt{(m_{K^\pm}^2 + \mathbf{p}_{K^\pm}^2)} + 2m_{\pi^0}\sqrt{(m_{\pi^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)}, \quad (59)$$

$$m_{defect}^2 = m_{\pi^0}^2 + 2m_{\pi^0}E_{K^\pm} + 2m_{\pi^0}E_{\pi^\mp}, \quad (60)$$

$$m_{defect}^2 = m_{\pi^0}^2 + 2m_{\pi^0}(E_{K^\pm} + E_{\pi^\mp}). \quad (61)$$

Because energy is conserved in the decay, $E_{B^0} = E_{K^\pm} + E_{\pi^\mp} + E_{\pi^0}$, so $E_{B^0} - E_{\pi^0} = E_{K^\pm} + E_{\pi^\mp}$ and we can write equation 62 as

$$m_{defect}^2 = m_{\pi^0}^2 + 2m_{\pi^0}(E_{B^0} - E_{\pi^0}). \quad (62)$$

Moreover, because we are in the center of mass frame, the B^0 meson has no momentum, so $E_{B^0} = m_{B^0}$, and, since the π^0 also has zero momentum, we also have $E_{\pi^0} = m_{\pi^0}$. Thus,

$$m_{defect}^2 = m_{\pi^0}^2 + 2m_{\pi^0}(m_{B^0} - m_{\pi^0}), \quad (63)$$

$$m_{defect}^2 = -m_{\pi^0}^2 + 2m_{\pi^0}m_{B^0}, \quad (64)$$

With the squared masses defect of equation 64, the reconstructed mass of a $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ event when missing a π^0 with zero momentum, becomes

$$m_{rec}^2 = m_{B^0}^2 - m_{defect}^2, \quad (65)$$

$$m_{rec}^2 = m_{B^0}^2 + m_{\pi^0}^2 - 2m_{\pi^0}m_{B^0}, \quad (66)$$

$$m_{rec}^2 = (m_{B^0} - m_{\pi^0})^2, \quad (67)$$

$$m_{rec} = m_{B^0} - m_{\pi^0}, \quad (68)$$

and

$$m_{defect} = m_{B^0} - m_{rec} = m_{\pi^0}. \quad (69)$$

Thus, m_{defect} , the difference between the correct and reconstructed masses of a $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ event when missing a π^0 with zero momentum, is exactly the mass of the π^0 . Moreover, it can be shown that this is also the minimum mass defect, which will be the purpose of the next section.

Proof that the minimum mass defect equals m_{π^0}

In order to prove that m_{π^0} is the minimum mass defect, we express m_{defect}^2 using inner product notation, as according to equation 58. As it turns out, equation 58 can be greatly simplified, such that finding its minimum becomes trivial. Namely,

$$\begin{aligned} m_{defect}^2 &= m_{\pi^0}^2 + 2\sqrt{(m_{K^\pm}^2 + \mathbf{p}_{K^\pm}^2)}\sqrt{(m_{\pi^0}^2 + \mathbf{p}_{\pi^0}^2)} + \\ &\quad 2\sqrt{(m_{\pi^\mp}^2 + \mathbf{p}_{\pi^\mp}^2)}\sqrt{(m_{\pi^0}^2 + \mathbf{p}_{\pi^0}^2)} - 2\mathbf{p}_{K^\pm} \cdot \mathbf{p}_{\pi^0} - 2\mathbf{p}_{\pi^\mp} \cdot \mathbf{p}_{\pi^0}, \end{aligned} \quad (70)$$

$$m_{defect}^2 = m_{\pi^0}^2 + 2E_{K^\pm}E_{\pi^0} + 2E_{\pi^\mp}E_{\pi^0} - 2\mathbf{p}_{\pi^0} \cdot \mathbf{p}_{K^\pm} - 2\mathbf{p}_{\pi^0} \cdot \mathbf{p}_{\pi^\mp}, \quad (71)$$

$$m_{defect}^2 = m_{\pi^0}^2 + 2E_{\pi^0}(E_{K^\pm} + E_{\pi^\mp}) - 2\mathbf{p}_{\pi^0} \cdot (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp}), \quad (72)$$

$$m_{defect}^2 = m_{\pi^0}^2 + 2E_{\pi^0}(E_{B_d} - E_{\pi^0}) - 2\mathbf{p}_{\pi^0} \cdot (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp}), \quad (73)$$

$$m_{defect}^2 = m_{\pi^0}^2 - 2E_{\pi^0}^2 + 2E_{\pi^0}E_{B_d} - 2\mathbf{p}_{\pi^0} \cdot (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp}), \quad (74)$$

For the next steps, we need to specify that we are in the center of mass frame, such that the mother B^0 meson is at rest (i.e. $E_{B^0} = m_{B^0}$) and, because momentum is conserved in the decay, the total momentum of the system of daughter particles is also zero (i.e. $\sum \mathbf{p}_{daughter} = \mathbf{0}$). Therefore,

$$m_{defect}^2 = m_{\pi^0}^2 - 2(m_{\pi^0}^2 + \mathbf{p}_{\pi^0}^2) + 2E_{\pi^0}m_{B^0} - 2\mathbf{p}_{\pi^0} \cdot (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp}), \quad (75)$$

$$m_{defect}^2 = m_{\pi^0}^2 - 2m_{\pi^0}^2 + 2E_{\pi^0}m_{B^0} - 2\mathbf{p}_{\pi^0} \cdot (\mathbf{p}_{K^\pm} + \mathbf{p}_{\pi^\mp} + \mathbf{p}_{\pi^0}), \quad (76)$$

$$m_{defect}^2 = -m_{\pi^0}^2 + 2E_{\pi^0}m_{B^0} - 2\mathbf{p}_{\pi^0} \cdot (\mathbf{0}), \quad (77)$$

$$m_{defect}^2 = -m_{\pi^0}^2 + 2E_{\pi^0}m_{B^0}. \quad (78)$$

Thus, the squared masses defect m_{defect}^2 equals $-m_{\pi^0}^2 + 2E_{\pi^0}m_{B^0}$ and, because m_{π^0} and m_{B^0} are both invariant, E_{π^0} is the only term that varies. Therefore, the mass defect is at its minimum when E_{π^0} is at its minimum. Therefore, since we specified that we are in the center of mass frame of the mother B_d meson, the mass defect is at its minimum when the missed π_0 is at rest in the center of mass frame, such that $E_{\pi_0} = m_{\pi_0}$ and

$$m_{defect} = m_{B^0} - \sqrt{m_{B^0}^2 - 2E_{\pi^0}m_{B^0} + m_{\pi^0}^2} = m_{B^0} - \sqrt{(m_{B^0} - m_{\pi^0})^2} = m_{\pi^0}. \quad (79)$$

It is important to realize that we restricted our physics to the center of mass frame of the B^0 meson. Therefore, the result $m_{defect} = m_{\pi^0}$, for $\mathbf{p}_{\pi^0} = 0$ in the center of mass frame, is specific to this reference frame. However, remember that the mass defect m_{defect} is the difference between the reconstructed B^0 mass and the correct B^0 mass, when reconstructing the invariant mass from just the charged pion and kaon for a $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ decay (i.e. failing to include the neutral pion). Therefore, the reconstructed mass is simply the invariant mass of the charged pion-kaon pair, which is frame independent, and thus the mass defect $m_{defect} = m_{B^0} - \sqrt{m_{B_d}^2 - 2E_{\pi^0}m_{B_d} + m_{\pi^0}^2}$ is simply the difference between two frame independent quantities, namely invariant mass of the kaon-pion pair and the invariant mass of the B^0 meson, and thus the mass defect m_{defect} is frame independent (the invariant mass of the charged pion-kaon pair varies with the momentum of the missed neutral pion, but is Lorentz *invariant* under a change of reference frame). Thus, in the lab frame (or any other inertial reference frame), the equality

$$m_{defect} = m_{B^0} - \sqrt{m_{B^0}^2 - 2E_{\pi^0}m_{B^0} + m_{\pi^0}^2} \quad (80)$$

holds as well. Nevertheless, the requirement $\mathbf{p}_{\pi^0} = 0$ is frame dependent, on the other hand. Therefore, the minimum mass defect

$$m_{defect} = m_{\pi^0} \quad (81)$$

holds in any inertial reference frame (i.e. in any inertial reference frame, the minimum mass defect equals the mass of the neutral pion), but is achieved

if and only if the momentum of the neutral pion is zero *in the center of mass frame of the B^0 meson.*

□

Discussion

As was shown in the previous section, the squared mass defect can be expressed as

$$m_{\text{defect}}^2 = -m_{\pi^0} + 2m_{B^0}E_{\pi^0,cm}, \quad (82)$$

where the subscript *cm* of the term $E_{\pi^0,cm}$ denotes that the quantity is measured in the center of mass frame of the mother B^0 meson. Moreover, as is obvious from 82, the mass defect is at its minimum when $E_{\pi^0,cm}$ is at its minimum, i.e. π^0 is emitted at rest in the center of mass frame. Furthermore, as we have not used any features specific to the $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ decay, these results hold for any three-body decay. Namely, for any three-body decay reconstructed as a two-body decay, the following two relations hold:

$$m_{\text{defect}}^2 = -m_d + 2m_pE_{d,cm}, \quad (83)$$

and

$$m_{\text{defect}} = m_p - \sqrt{m_p + m_{\pi^0} - 2m_{B^0}E_{\pi^0,cm}}, \quad (84)$$

where the subscript *p* denotes the mother particle, the subscript *d* denotes the missed daughter particle, and the subscript *cm* specifies that the quantity is measured in the center of mass frame of the mother particle. Therfore, for any three-body decay reconstructed as a two-body decay, the minimum mass defect is exactly the mass of the daughter particle that was omitted in the reconstruction of the mother particle's invariant mass. Moreover, this minimum occurs when the omitted daughter particle is emitted at rest in the center of mass frame of the mother particle, i.e. the (post-decay) velocity of the omitted daughter particle equals the (ante-decay) velocity of the mother particle in the lab frame. This is illustrated nicely in figure 42, for which a great number of random $B^0 \rightarrow K^\pm\pi^\mp\pi^0$ and $B_s^0 \rightarrow K^\pm\pi^\mp\pi^0$ decays were simulated (weighted for their phase-space elements) and reconstructed without the π^0 . The invariant mass spectra of these partially reconstructed events are plotted together with the correct mass peaks of the B mesons. As can be seen clearly from figure 42, the differences between the correct mass peaks and right-most limits of the corresponding partial reconstruction spectra of the B^0 and the B_s^0 mesons coincide nicely with the π^0 mass ($\pm 140\text{MeV}$). These

findings are of great value for the improved understanding of the shapes and features of partially reconstructed channels showing up as backgrounds in reconstructed invariant mass spectra, such as the $B^0 \rightarrow K^\pm \pi^\mp$ mass spectrum. For example, for the $B^0 \rightarrow K^\pm \pi^\mp$ reconstruction, from the understanding that the mass defect is at least the π^0 mass, we know that for the reconstructed invariant mass spectrum of the $B^0 \rightarrow K^\pm \pi^\mp$ decay, the region between $m_{B^0} - m_{\pi^0}$ and m_{B^0} ($\approx 5140 - 5280$ MeV/ c^2) is uncontaminated from partially reconstructed $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ events.

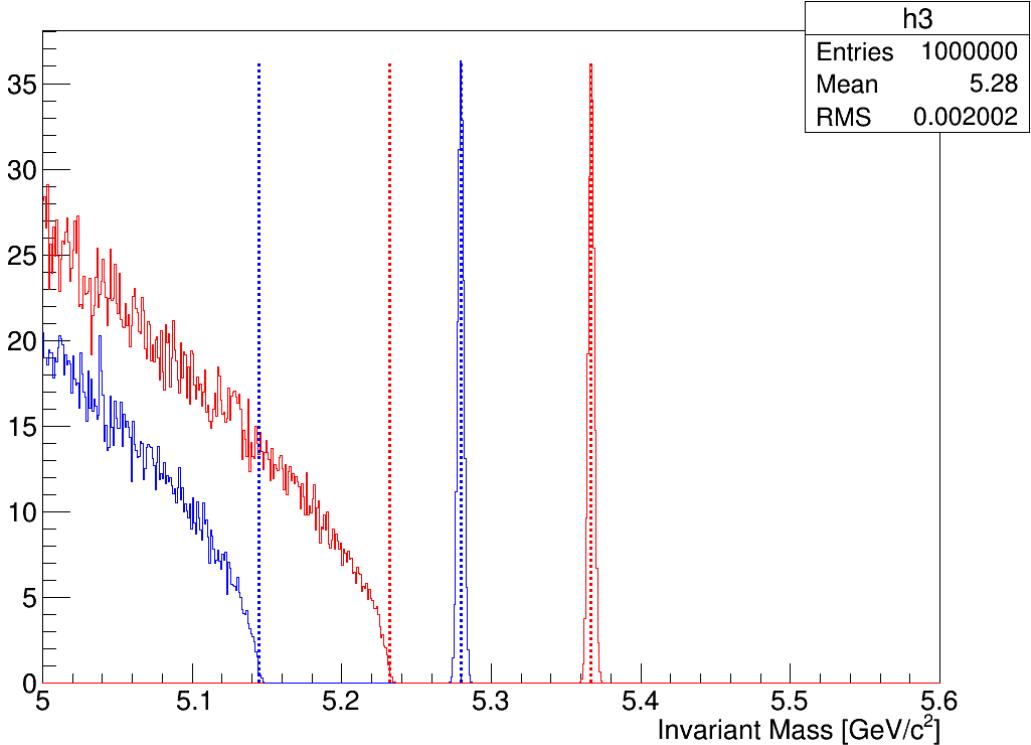


Figure 42: Invariant mass spectra of partially reconstructed simulated $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ and $B_s^0 \rightarrow K^\pm \pi^\mp \pi^0$ decays, together with the correct mass peaks of the B^0 and B_s^0 mesons

Conclusion

The minimum mass defect when reconstructing $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ events as $B^0 \rightarrow K^\pm \pi^\mp$ events occurs when \mathbf{p}_{π^0} is zero in the center of mass frame of the mother B^0 meson, and, when this is the case, amounts exactly to the mass of the π^0 . Moreover, these findings are not exclusive to $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ decays, but hold for any three-body decay reconstructed as a two-body decay, i.e.

the minimum mass defect equals the mass of the daughter particle omitted in the reconstruction and occurs when the omitted daughter particle is emitted at rest in the center of mass frame of the mother particle. The general expressions for the mass defect and squared masses defect, for any three-body decay reconstructed as a two-body decay are given by

$$m_{defect}^2 = -m_d + 2m_p E_{d,cm}, \quad (85)$$

and

$$m_{defect} = m_p - \sqrt{m_p + m_{\pi^0} - 2m_{B^0} E_{\pi^0,cm}}, \quad (86)$$

where the subscript p denotes the mother particle, the subscript d denotes the missed daughter particle, and the subscript cm specifies that the quantity is measured in the center of mass frame of the mother particle. The (theoretical) findings in this paper were confirmed by simulation, for both the $B^0 \rightarrow K^\pm \pi^\mp \pi^0$ decay and the $B_s^0 \rightarrow K^\pm \pi^\mp \pi^0$ decay. These findings are of great value for the improved understanding of the shapes and features of partially reconstructed channels showing up as backgrounds in reconstructed invariant mass spectra, such as the $B^0 \rightarrow K^\pm \pi^\mp$ mass spectrum.

References

- Aad, G., Abajyan, T., Abbott, B., Abdallah, J., Khalek, S. A., Abdelalim, A., ... others (2012). Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1), 1–29.
- Aaij, R., Adeva, B., Adinolfi, M., Adrover, C., Affolder, A., Ajaltouni, Z., ... others (2013). Search for the lepton-flavor-violating decays $B_s^0 \rightarrow e^\pm \mu^\mp$ and $B^0 \rightarrow e^\pm \mu^\mp$. *Physical review letters*, 111(14), 141801.
- Aaij, R., Beteta, C. A., Adeva, B., Adinolfi, M., Adrover, C., Affolder, A., ... others (2012). Search for the rare decays and $b0 \mu+ \mu-$. *Physics Letters B*, 708(1), 55–67.
- Adrover, C., et al. (2013). *Search for the b0 s $\mu+ \mu-$ and b0 $\mu+ \mu-$ decays with 3 fb- 1 at lhcb*. LHCb-ANA-2013-032.
- Anderlini, L., Contu, A., Zhang, Y., Vazquez Gomez, R., Muller, D., Jones, C. R., ... others (2016). *The pidcalib package* (Tech. Rep.).
- Aubert, B., et al. (2007). Improved Measurements of the Branching Fractions for $B^0 \rightarrow \pi^+ \pi^-$ and $B^0 \rightarrow K^+ \pi^-$, and a Search for $B^0 \rightarrow K^+ K^-$. *Phys. Rev.*, D75, 012008. doi: 10.1103/PhysRevD.75.012008
- Bettini, A. (2014). *Introduction to elementary particle physics*. Cambridge University Press.
- Bornheim, A., et al. (2003). Measurements of charmless hadronic two body B meson decays and the ratio $B(B \rightarrow D K) / B(B \rightarrow D \pi)$. *Phys. Rev.*, D68, 052002. ([Erratum: Phys. Rev.D75,119907(2007)]) doi: 10.1103/PhysRevD.75.119907,10.1103/PhysRevD.68.052002
- Buckley, A. G. (2006). *A study of b k π decays with the lhcb experiment* (Unpublished doctoral dissertation). University of Cambridge.
- Chakdar, S. (2016). New physics at the tev scale. *arXiv preprint arXiv:1604.07358*.
- Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C., & Zaritsky, D. (2006). A direct empirical proof of the existence of dark matter. *The Astrophysical Journal Letters*, 648(2), L109.
- Crystal ball function wikipedia webpage*. (2016). https://en.wikipedia.org/wiki/Crystal_Ball_function. (Accessed: 10-10-2016)
- Duh, Y. T., et al. (2013). Measurements of branching fractions and direct CP asymmetries for $B \rightarrow K\pi$, $B \rightarrow \pi\pi$ and $B \rightarrow KK$ decays. *Phys. Rev.*, D87(3), 031103. doi: 10.1103/PhysRevD.87.031103
- Kappert, R. (2016). *Electron identification efficiency in $B_{(s)}^0 \rightarrow e\mu$* . (Unpublished master's thesis). Rijksuniversity of Groningen.
- Koopman, R. (2015). *Ageing of the lhcb outer tracker & b-hadron production*

- and decay at $\sqrt{s} = 7$ tev* (Unpublished doctoral dissertation). Amsterdam U.
- Kuindersma, H. (2016). *Trigger efficiencies at lhcb in the search for $b_s \rightarrow e\mu$* (Unpublished master's thesis). Rijksuniversity of Groningen.
- Lhcb wikipedia webpage.* (2016). <https://upload.wikimedia.org/wikipedia/commons/3/36/Lhcview.jpg>. (Accessed: 09-18-2016)
- McDonald, A. B., Klein, J. R., & Wark, D. L. (2003). Solving the solar neutrino problem. *Scientific American*, 288(4), 40–49.
- Morrison, J. (2010). *Modern physics: for scientists and engineers*. Academic Press.
- Myers, S. (2013). The large hadron collider 2008–2013. *International Journal of Modern Physics A*, 28(25), 1330035.
- Olive, K. A. (2016). Review of Particle Physics. *Chin. Phys.*, C40(10), 100001. doi: 10.1088/1674-1137/40/10/100001
- Pivk, M. (2006). svlot: A quick introduction. In *Statistical problems in particle physics, astrophysics and cosmology: Proceedings of phystat05, oxford, uk, 12-15 september 2005* (p. 173).
- Tolk, S. (2016). *Discovery of rare b decays* (Unpublished doctoral dissertation). Rijksuniversity of Groningen.
- Vos, C. (2015). *Overview of the theoretical status of clfv* (Unpublished master's thesis). Rijksuniversity of Groningen.
- Zonneveld, J. (2016). *Dependence of b meson fragmentation fraction ratio f_s/f_d on centre-of-mass energy* (Unpublished master's thesis). Amsterdam U.