Measurement of the branching fractions of the semileptonic decays $B^+ \to \eta \ell^+ \nu_\ell$ and $B^+ \to \eta' \ell^+ \nu_\ell$ with signal-side only reconstruction at the Belle experiment

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsprogramm ProPhys
der Georg-August University School of Science (GAUSS)

vorgelegt von

Uwe Gebauer

aus Frankfurt am Main

Göttingen, 2021

Betreuungsausschuss

Prof. Dr. Ariane Frey
Prof. Dr. Stan Lai

Mitglieder der Prüfungskommission:

Referent:	Prof. Dr. Ariane Frey II. Physikalisches Institut, Georg-August-Universität Göttingen
Koreferentin:	Prof. Dr. Stan Lai II. Physikalisches Institut, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Dr. Baida Achkar
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Laura Covi
Institut für Theoretische Physik, Georg-August-Universität Göttingen

PD Dr. Jörn Große-Knetter
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Wolfram Kollatschny
Institut für Astrophysik, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 11.06.2021

Referenz: II.Physik-UniGö-Diss-2021/03

Measurement of the branching fractions of the semileptonic decays $B^+ \to \eta \ell^+ \nu_\ell$ and $B^+ \to \eta' \ell^+ \nu_\ell$ with signal-side only reconstruction at the Belle experiment

Abstract

This thesis presents a measurement of the branching fractions of the decays $B^+ \to \eta \ell^+ \nu_\ell$ and $B^+ \to \eta' \ell^+ \nu_\ell$. The entire data sample collected by the Belle experiment at the energy of the $\Upsilon(4S)$ resonance is used, which contains in total $772 \times 10^6 B\overline{B}$ pairs. The η meson is reconstructed in two decay channels and the η' meson in one decay channel. Only the decay products of one of the two B mesons from the $B\overline{B}$ pair are explicitly reconstructed to achieve a high efficiency. Instead of explicitly reconstructing the remainder of the event, the neutrino in the decay is inferred using conservation laws to allow reconstruction of the B^+. Background events are reduced using boosted decision trees. A binned maximum likelihood fit of MC distributions to the data sample is used to extract the signal yield in the end. The entire reconstruction chain keeps the dependence on the transferred momentum q^2 as low as possible to reduce effects of the modelling of the decay. The determined branching fractions are $\mathcal{B}(B^+ \to \eta \ell^+ \nu_\ell) = (2.83 \pm 0.55 \pm 0.34) \times 10^{-5}$ and $\mathcal{B}(B^+ \to \eta' \ell^+ \nu_\ell) = (2.79 \pm 1.29 \pm 0.30) \times 10^{-5}$ with the statistical uncertainty first and the systematic second.

Messung des Verzweigungsverhältnisses der semileptonischen Zerfälle $B^+ \to \eta \ell^+ \nu_\ell$ und $B^+ \to \eta' \ell^+ \nu_\ell$ mit Rekonstruktion nur der Signal-Seite am Belle-Experiment

Zusammenfassung

Diese Dissertation präsentiert eine Messung des Verzweigungsverhältnisses der Zerfälle $B^+ \to \eta \ell^+ \nu_\ell$ und $B^+ \to \eta' \ell^+ \nu_\ell$. Dabei wird der gesamte an der Energie der $\Upsilon(4S)$-Resonanz gemessene Datensatz des Belle-Experiments verwendet. Dieser enthält $772 \times 10^6 B\overline{B}$-Paare. Das im Zerfall entstehende η-Meson wird in zwei und das η'-Meson in einem Zerfallskanal rekonstruiert. Nur die Zerfallskette eines der beiden B-Mesonen wird explizit rekonstruiert, wobei das Neutrino mittels Erhaltungssätzen aus dem Rest des Ereignisses abgeleitet wird. Hiermit wird eine größtmögliche Rekonstruktionseffizienz erzielt. Der Anteil an Untergrundereignissem im rekonstruierten Datensatz wird mittels Boosted Decision Trees verringert, bevor die Anzahl an Signalereignissen im Datensatz mittels eines Binned-Maximum-Likelihood-Fits ermittelt wird, bei dem die einzelnen MC Verteilungen an den Datensatz gefittet werden. Die gesamte Analyse hält die Abhängigkeit der Selektion vom Impulsübertrag q^2 so gering wie möglich, um Auswirkungen der Zerfallsmodellierung zu vermeiden. Die gemessenen Verzweigungsverhältnisse sind
$\mathcal{B}(B^+ \to \eta \ell^+ \nu_\ell) = (2.83 \pm 0.55 \pm 0.34) \times 10^{-5}$ und
$\mathcal{B}(B^+ \to \eta' \ell^+ \nu_\ell) = (2.79 \pm 1.29 \pm 0.30) \times 10^{-5}$ mit statistischer und systematischer Unsicherheit.

Contents

CHAPTER 1

Introduction

Particle physics has been improving the understanding of the behaviour of the world on the smallest known scales for over a hundred years, starting from the discovery of the electron in 1897 up to the Higgs boson in 2012. While the Standard Model describing the known fundamental particles and their interactions has been very successful, comparing it to the world around us points to many open questions. One of them concerns the asymmetry of matter and antimatter. While they are produced in equal amounts in the early universe, evidence from astrophysics [1] confirms that the observable universe is dominated by matter.

Three criteria are necessary for such an asymmetry, as determined by Sakharov [2]:

1. Violation of baryon number conservation,
2. CP violation,
3. Interactions out of equilibrium.

The first two are especially accessible through particle physics experiments. While no experimental evidence against baryon number violation has been found so far, CP violation was experimentally verified in the weak interaction, described by the CKM mechanism [3]. However, when compared to astrophysical measurements, the magnitude of CP violation in the Standard Model is still too small to produce the observed amount of matter in the universe.

Nonetheless, as the only known source of CP violation, the weak interaction warrants closer inspection. One of the facilities specifically built to detect and study this effect is the Belle experiment, whose focus lies on studying the behaviour and decay of B mesons via the weak interaction. Out of these, the decays of the type $b \to u$ are of special

interest as their heavy suppression in the *CKM* mechanism makes them sensitive to additional processes not yet included in the Standard Model. The Belle experiment is a particle detector situated at the asymmetric KEKB accelerator, producing B mesons in pairs from electron and positron annihilation and measuring their decay products. It is a so-called B-Factory optimized to produce a large dataset of B meson decays.

This thesis measures the decays[1] $B^+ \to \eta \ell^+ \nu_\ell$ and $B^+ \to \eta' \ell^+ \nu_\ell$ with the dataset taken at the Belle experiment. In these, the ℓ^+ is either an electron or a muon, the ν_ℓ is the corresponding neutrino and the $\eta^{(\prime)}$ is a charmless hadron. The world averages [4] of the branching ratios from previous measurements, combined from Belle, Babar and CLEO, are $\mathcal{B}\left(B^+ \to \eta \ell^+ \nu_\ell\right) = (3.9 \pm 0.5) \times 10^{-5}$ and $\mathcal{B}\left(B^+ \to \eta' \ell^+ \nu_\ell\right) = (2.3 \pm 0.8) \times 10^{-5}$. As the low branching fractions lead to a small amount of signal events expected in the dataset, the yield is maximised by explicitly reconstructing only one B meson out of the original pair. The $b \to u$ transition can be measured in different processes and via different methods, however, the results from these are not always compatible. To keep the effect of uncertainties in the theoretical description low, special importance is laid on reconstructing the entire kinematic spectrum as uniformly as possible.

The thesis begins with an overview of the relevant theoretical foundations of particle physics in Chapter 2. Afterwards, the Belle detector is presented in Chapter 3 followed by an introduction to various statistical methods used throughout this thesis in Chapter 4. In Chapter 5 the actual reconstruction of the decay process $B^+ \to \eta^{(\prime)} \ell^+ \nu_\ell$ is presented introducing the methods to form candidates for the involved particles out of the detector measurements. Out of the vast amount of candidates, a subset with a higher concentration of correctly reconstructed signal decays is selected in Chapter 6. The fit procedure to determine the signal yield from the selected sample is introduced in Chapter 7. In Chapter 8 several validation steps are summarized to confirm the analysis procedure, followed by the introduction of the sources of systematic uncertainty considered in Chapter 9. The thesis concludes with the presentation of the measured branching fractions and their discussion in Chapter 10.

[1] Unless explicitly stated otherwise, whenever a process or particle is mentioned with a specific charge throughout this thesis the charge-conjugated process, that is the process where every particle is exchanged with its anti-particle, is implicitly included.

CHAPTER 2

Theory

This chapter introduces the various theoretical concepts and backgrounds important for the physics investigated at B-Factories. It starts by explaining the Standard Model and its particle content in Section 2.1, introducing the fermions included and subsequently the bosons together with their corresponding interactions. The fundamental discrete symmetries are introduced in Section 2.2, followed by their violations at the hands of the weak force in Section 2.3. In Section 2.4 the *CKM* mechanism is presented as an explanation of these violations. The last part, Section 2.5, narrows the focus towards the decay processes of interest for this thesis, starting from the general behaviour of B mesons.

2.1. The Standard Model of Particle Physics

The Standard Model of Particle Physics (*SM*) [5] is the theoretical description of all currently known elementary particles and three of the four fundamental interactions. It takes the form of a quantum field theory with the combined symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ from the strong, the weak and the electromagnetic interactions. While the *SM* is a very well tested theory consistent with overwhelming experimental evidence for its predictions, it does not contain everything. The fourth interaction, gravity, is not included in the *SM*. The strength of gravity is negligible at typical reactions the *SM* is used to describe due to the small number of involved particles and their low individual masses. When comparing the *SM* predictions with astrophysics, further discrepancies become visible. The *SM* currently provides no mechanism to generate the matter density observed, and also contains no candidate to explain dark matter as observed.

Elementary particles are defined as all particles not possessing any further substructure and assumed as point-like particles. The elementary particles included in the *SM* are grouped in two categories depending on their spin. The first category are fermions with half-integer spin, which for the known fundamental fermions always is 1/2. The fermions contain all particles making up ordinary matter, and are themselves divided into quarks (q) and leptons (ℓ). Both of these two groups contain six[1] fermions, divided into three generations, which consist of two fermions each. Between the three generations, the particles, while distinct, differ only in mass but are identical otherwise. For all fermions, a corresponding antiparticle exists with the same mass but opposite charge. The fermions included in the *SM* with some of their properties are listed in Table 2.1.

Generation	Type	Particle	Charge [e]	Mass [MeV/c^2]
1.	Quarks	Up (u)	2/3	$2.16^{+0.49}_{-0.26}$
		Down (d)	-1/3	$4.67^{+0.48}_{-0.17}$
	Leptons	Electron (e)	-1	0.5109989461(31)
		Electron neutrino (ν_e)	0	$< 1.1 \times 10^{-6}$
2.	Quarks	Charm (c)	2/3	1270 ± 20
		Strange (s)	-1/3	93^{+11}_{-5}
	Leptons	Muon (μ)	-1	105.6583745(24)
		Muon neutrino (ν_μ)	0	< 0.19
3.	Quarks	Top (t)	2/3	172760 ± 300
		Bottom (b)	-1/3	4180^{+30}_{-20}
	Leptons	Tau (τ)	-1	1776.86 ± 0.12
		Tau neutrino (ν_τ)	0	< 18.2

Table 2.1.: Fermions included in the Standard Model of Particle Physics. The quarks come in three colour variations not separately listed. For all particles a corresponding antiparticle with opposite charge exists. For the neutrinos only upper bounds are known. Values taken from Reference [4].

The quarks are the only fermions coupling to all three fundamental forces included in the *SM*. The first generation consists of the up and down quark, with the charm and strange quarks in the second, and the top (or truth) and bottom (or beauty) quarks in the third generation. All quarks carry one of the three colours of the strong interaction. The confinement rule of the strong interaction requires all quarks to form bound, colourless states called hadrons[2]. These bound states generally fall into either baryons containing three quarks (qqq) all with different colours, or mesons ($q\overline{q}$) containing a quark-antiquark pair with cancelling colour and anti-colour. While the individual quarks carry a fractional electric charge, these combinations always result in integer charges. While all

[1] Strictly speaking each quark is further subdivided into three varieties of different colour.
[2] The top quark is an exception, as it decays too fast for bound states to form.

mesons are unstable and eventually decay, the stable[3] proton (uud) together with the neutron (udd) makes up the core of ordinary matter as nucleons.

The leptons consists of a charged lepton and a corresponding neutrino at every generation. While the charged leptons interact both weakly and electromagnetically, the neutrinos, being electrically neutral, only interact weakly. The charged leptons are the electron, muon and tau in order of generation and mass. The electron, which is also the first elementary particle to have been discovered, is part of every atom. As the neutrinos interact only weakly, they pass through matter without much interaction and do not form bound states. The neutrinos are assumed to be massless in the *SM*, however the detection of neutrino oscillations [6] has shown this can not be correct for at least two of them. So far, only upper limits could be set for these masses.

The other type of particles are bosons with integer spin. Three types of spin-1 gauge bosons act as carriers of the three interactions included in the *SM*, together with the spin-0 Higgs boson. The bosons are summarised in Table 2.2.

Name	Interaction	Charge [e]	Spin	Mass [GeV/c^2]
Photon (γ)	Electromagnetic	0	1	0
Z^0	Weak	0	1	91.1876 ± 0.0021
$W^{\pm}$		± 1	1	80.379 ± 0.012
Gluon (g)	Strong	0	1	0
Higgs (H)	Higgs field	0	0	125.10 ± 0.14

Table 2.2.: Bosons included in the Standard Model of Particle Physics. The total of eight gluons are identical except for their colour charge. Values taken from Reference [4].

2.1.1. Electromagnetic interaction

The electromagnetic interaction, described by Quantum Electrodynamics (QED), is mediated by the massless photon. It couples to all electrically charged particles with the fundamental vertex shown in Figure 2.1. As the photon itself is uncharged, photons can not couple to themselves.

The electromagnetic interaction is the only one included in the *SM* with an unlimited range in vacuum. Its gauge group is $U(1)$, the unitary group of order 1. The electromagnetic interaction, together with gravity, is responsible for all macroscopic events on

[3] No proton decay has been measured in several searches, leading to a current limit [4] for the life time above 3.6×10^{29} years.

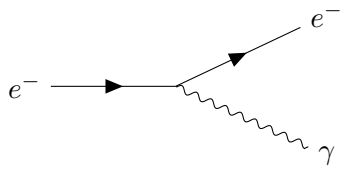


Figure 2.1.: The fundamental QED vertex. The electron can be replaced by any charged particle.

scales larger than individual atoms. Particles which primarily decay via the electromagnetic interaction have average lifetimes of the order of 10^{-16} s.

2.1.2. Strong interaction

The strong interaction is described by Quantum Chromodynamics (QCD) and has the gauge group $SU(3)_C$, the special unitary group of order 3, where the index C stands for the colour charge. It is mediated by eight gluons. While they are massless similar to the photons, they themselves carry the colour charge they couple to. This allows the gluons to couple to themselves, adding additional vertices consisting only of gluons. These are shown in Figure 2.2. There are three colour charges, together with corresponding anti-colours.

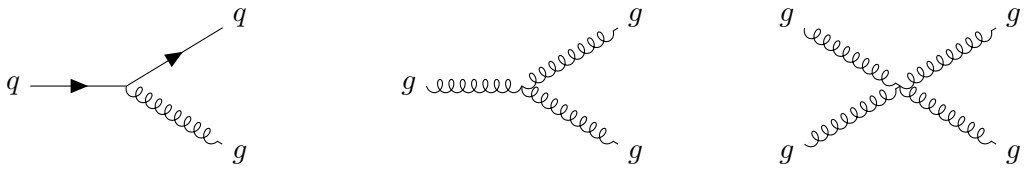


Figure 2.2.: The fundamental QCD vertices.

Although the gluons are massless, the effective range of the strong interaction is very limited. Unlike the electromagnetic interaction, the coupling strength of the strong interaction increases with distance to objects carrying net colour charge. While inside bound states, the quarks experience asymptotic freedom without much effect of the strong interaction, increasing energy is needed the further a colour-charged part is separated. This causes all produced quarks to quickly hadronise into colour-neutral bound states.

The strong interaction is the strongest interaction in the *SM* and whenever available strong processes dominate over the other. As a result, bound states which can decay via a strong process are generally the shortest lived at average lifetimes in the order of 10^{-23} s.

2.1.3. Weak interaction

The weak interaction, as described by Quantum Flavourdynamics (QFD) with the symmetry group $SU(2)_L$, is special in several ways. It is mediated by two types of massive gauge bosons, the electrically charged $W^{\pm}$-*bosons* and the neutral Z^0-*boson*. They couple to all types of fermions. While the Z^0 couples to a pair of a particle and the corresponding antiparticle, the $W^{\pm}$ couples to a pair of either two different leptons or quarks. The $W^{\pm}$-*bosons* can also couple to both themselves as well as to one or two photons or Z^0. The fundamental vertices of the weak interaction are shown on Figure 2.3, including those involving both weak and the electromagnetic gauge bosons.

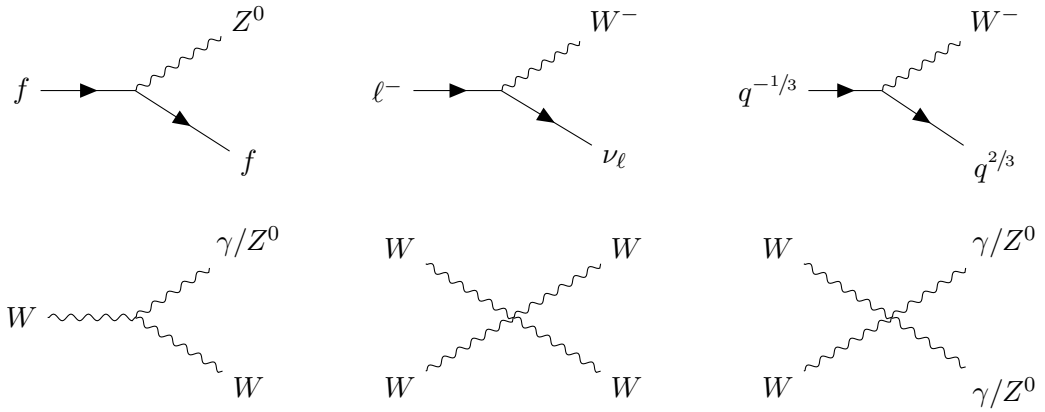


Figure 2.3.: The fundamental QFD vertices.

The $W^{\pm}$ have the unique feature of being able to change the flavour of the quarks they couple to. They can couple to a pair of quarks from different generations, providing the only way for the quarks of the second and third generation to decay into quarks of lower generations. These couplings across generations are however suppressed compared to couplings within a generation. The mechanism is discussed further in Section 2.4. The Z^0 can only couple to quarks within the same generation.

Another feature unique to the $W^{\pm}$ and Z^0 is that their coupling depends on the chirality of a particle, the eigenvalue to γ^5. For massless particles, this coincides with the helicity, which depends on the relative directions of its spin and movement, if both have the same direction it is right-handed, and of they are opposite it is left-handed, while for other particles the helicity only serves as an approximation [5]. For the Z^0 the coupling strength differs depending on the handedness, but the $W^{\pm}$ only couple to left-handed particles (and right-handed antiparticles) at all. This leads to the grouping of the left-handed leptons and quarks of each generation into doublets that couple to the $W^{\pm}$, while their right-handed counterparts form singlets. This difference on handedness is further elaborated in Section 2.3. The mass of the weak gauge bosons not only restricts the range of the interaction, but is also the reason the weak interaction is the weakest in the

SM at low energy scales, and mostly visible when the strong and electromagnetic one are impossible or heavily suppressed, corresponding to particles having lifetimes higher than 10^{-13} s.

2.1.4. Higgs mechanism

A theoretical problem in the *SM* was the inclusion of a mass term for the $W^{\pm}$ and Z^0 bosons. In the context of the electroweak unification [7–9], combining both the weak and the electromagnetic interaction, this is solved by introducing the Higgs field with four degrees of freedom [10, 11]. In a spontaneous symmetry breaking three of these give mass to three of the gauge bosons, forming the $W^{\pm}$ pair and the Z^0. The fourth stays massless and forms the photon. The massive fermions gain their mass by their Yukawa coupling to the Higgs field.

The remaining degree of freedom forms the Higgs boson, the only spin-0 particle in the *SM*. It is the most recently discovered fundamental particle [12, 13], with discovery in 2012.

2.2. Symmetries

Every physical system or interaction can be characterized by its symmetries. Symmetries are operations on a system leaving its physical behaviour invariant, that is the transformed system behaves indistinguishably from the original one. In classical physics, these are continuous transformations such as translations in time, space and also rotations. Their importance comes from Noether's theorem [14], which relates continuous symmetries to fundamental conservation laws.

In particle physics, three additional discrete symmetries are important: Parity (P), Charge Conjugation (C) and Time Reversal (T). These will be explained in more detail in the following. A common point between all three discrete symmetries is that they are inversions, and applying them two times returns to the original state. Therefore, every symmetry operator X out of P, C, T must fulfil

$$X^2 \;=\; I \tag{2.1}$$

with the identity operator I. Accordingly, the only possible eigenvalues are ± 1.

Parity Parity is the inversion of a system in its spatial coordinates. The parity operation (P) is defined as the inversion around the origin:

$$P(\vec{x}) \;=\; -\vec{x}. \tag{2.2}$$

From this, inversions around arbitrary points or mirroring around arbitrary planes can be reached by additional translations and rotations. This inversion does not affect pseudovectors, leaving direction of angular momentum unchanged. As a result, parity changes the handedness of a particle from left to right-handed and vice versa. The fermions are arbitrarily assigned the eigenvalue $P|f\rangle = +1|f\rangle$, while antifermions have $P|\bar{f}\rangle = -1|\bar{f}\rangle$. The eigenvalues of composite hadrons are determined by multiplication of the components together with any orbital angular momentum contributions.

Charge Conjugation The charge conjugation (C) symmetry inverts not only the electric charge, but all additive quantum numbers including flavour and colour charge, while mass and momentum stay unchanged. This turns a particle into its antiparticle. Charge inversion does not change the handedness of a particle. Eigenstates of C are particles who are their own antiparticle, which can only be the case for neutral particles.

Time Reversal The last of the set of discrete symmetries is time reversal (T). In a particle reaction, this results in the exchange of the ingoing and the outgoing states. Unlike for the other two symmetries no particles are eigenstates of T.

CPT As a fundamental theorem of quantum field theory, the CPT-theorem [15] requires that the combination of all three symmetries together is absolutely conserved. Any violations in CPT symmetry would imply violations of Lorentz invariance. Another consequence of the CPT theorem is that particles and their corresponding antiparticles have the same masses.

2.3. Symmetry violations

The three discrete symmetries are individually conserved in both the electromagnetic and strong[4] interactions. While it originally was assumed that they would also be conserved in the weak interaction, when first measured experimentally by Wu [16] in 1956 it was found that the weak interaction maximally violated parity. The experiment used the beta decay of spin polarized Cobalt nuclei:

$$^{40}\mathrm{Co} \rightarrow {}^{40}\mathrm{Ni} + e^- + \bar{\nu}_e \tag{2.3}$$

Due to conservation of angular momentum, the electron and neutrino have to be emitted along the polarization axis. If P was a symmetry of the weak interaction, the electrons would be emitted equally often parallel and anti-parallel to the Cobalt spin. However, it was found that electrons were only emitted anti-parallel to the polarization, which

[4] While the strong interaction theoretically allows violations, they seem to not be realized in nature.

resulted in left-handed electrons. This meant the weak interaction only acts on one preferred handedness, violating the parity symmetry which would require equal treatment.

From the discovery of parity violation follows immediately that charge symmetry is also violated, as can be seen in the decay of charged pions:

$$\pi^+ \to \mu^+ \nu_\mu. \tag{2.4}$$

Conservation of angular momentum requires the decay products of the pion to have the same handedness. As the neutrino is (almost) massless, it needs to be left-handed for the weak interaction to couple to it, resulting in the anti-muon also being left-handed. Applying charge conjugation would yield the decay

$$\pi^- \to \mu^- \bar{\nu}_\mu \tag{2.5}$$

with the decay products still being left-handed as before. While chirality is Lorentz-invariant, the helicity is not. For massive particles, the handedness is therefore not an exact quantum number, allowing the chirality-based coupling of the weak force to particles of opposite helicity. This coupling to the wrong handedness, which applies to the muon in the original process, is suppressed for particles with lower masses[5], becoming impossible for massless particles as chirality and helicity become equal. In the charge inverted process in Equation (2.5) however, the neutrino would have the wrong handedness, and this process is not seen.

Additionally applying P inverts the handedness as well and returns the observed decay of the π^-. This gave hope that the true symmetry was in fact the combination of the two, CP. Neutral kaons gave Cronin and Fitch [17] the opportunity to test this in an experiment. Neutral kaons are produced in the strong eigenstates $K^0(d\bar{s})$ and $\overline{K}^0(s\bar{d})$ with a well-defined quark content. However, like other neutral mesons they can convert into each other via the weak interaction, shown in general in Figure 2.4. This produces mixed states.

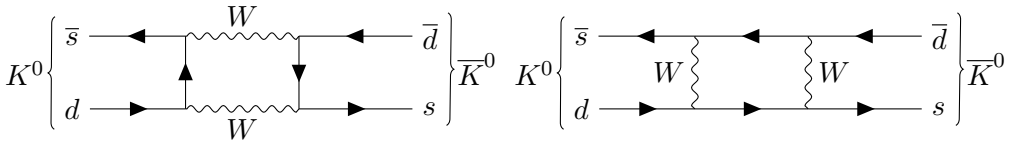


Figure 2.4.: Oscillation diagrams for neutral kaons.

The two strong states K^0 and $\overline{K}^0$ are not eigenstates of CP, but instead related via $CP\,|K^0\rangle = +1\,|\overline{K}^0\rangle$. In general, mixed states can be constructed out of the two as

$$|\pm\rangle = N\left(p\,|K^0\rangle \pm q\,|\overline{K}^0\rangle\right), \tag{2.6}$$

[5]This is incidentally also the reason why the pion preferentially decays into a muon instead of the lighter electron.

with an appropriate normalisation factor N. From this relation the actual CP-states for the neutral kaons immediately follow as:

$$|K_1\rangle = \frac{1}{\sqrt{2}}\left(|K^0\rangle + |\overline{K}^0\rangle\right), \tag{2.7}$$

$$|K_2\rangle = \frac{1}{\sqrt{2}}\left(|K^0\rangle - |\overline{K}^0\rangle\right). \tag{2.8}$$

The K_1 has the CP-eigenvalue 1 and quickly decays to two pions. However, the K_2 with its eigenvalue of -1 does not have this decay channel available. It instead decays to three pions with much less remaining phase space, and is therefore much longer lived. In the experiment, a beam of neutral kaons was allowed to propagate long enough for the short-lived component to decay down to a negligible level. Nonetheless, a small amount of decays into two pions was still observed. The explanation lies in the long-lived state K^0_L not being exactly the CP-eigenstate K_2, but having a small contribution of the CP-opposite state:

$$|K^0_\mathrm{L}\rangle = \frac{1}{\sqrt{1+|\epsilon|^2}}\left(|K_2\rangle + \epsilon\,|K_1\rangle\right). \tag{2.9}$$

The strength of this CP-violation due to the mixing of the neutral states was found [4] to be $|\epsilon| = (2.228 \pm 0.011) \times 10^{-3}$.

In general, in this type of CP-violation the decay rates of the final states are identical for both mesons, and the violation happens at the mixing state, where the oscillation rate is not equal for both directions. This causes the discrepancy between the weak and the CP-eigenstates. It happens whenever the mixing is not perfectly equal, that is $|q/p| \neq 1$.

For decays of an initial, mixed state i^0_{phys} and its charge-conjugate partner to a state with a charged lepton $\ell^\pm$ used to determine the decaying flavour, the asymmetry can be defined as:

$$A_{SL}(t) \coloneqq \frac{\frac{\mathrm{d}\Gamma}{\mathrm{d}t}(\bar{i}^0_{phys}(t) \to \ell^+ X) - \frac{\mathrm{d}\Gamma}{\mathrm{d}t}(i^0_{phys}(t) \to \ell^- X)}{\frac{\mathrm{d}\Gamma}{\mathrm{d}t}(\bar{i}^0_{phys}(t) \to \ell^+ X) + \frac{\mathrm{d}\Gamma}{\mathrm{d}t}(i^0_{phys}(t) \to \ell^- X)} = \frac{1-|q/p|^4}{1+|q/p|^4}. \tag{2.10}$$

It is interesting to note that although the decay rates are time-dependent due to the oscillation, the resulting asymmetry itself is not and can be expressed as a constant only depending on the factors q and p.

Direct CP-violation in decays has also been observed. Here, the decay rate differs after every particle is exchanged with its antiparticle.

$$\mathcal{B}\left(i \to f\right) \neq \mathcal{B}\left(\bar{i} \to \bar{f}\right). \tag{2.11}$$

For the decay amplitudes A_f of the decay to the final state f, and $\overline{A}_{\overline{f}}$ for the process with all particles exchanged with their antiparticles, this is equivalent to $|\overline{A}_{\overline{f}}/A_f| \neq 1$. The decay amplitudes are not invariant under the exchange of particles with antiparticles.

For neutral kaons, this effect is even smaller than that of the mixing mentioned above, with the factor determining the size, ϵ', being [4] $\epsilon'/\epsilon = (1.66 \pm 0.23) \times 10^{-3}$. However, direct CP-violation is possible for charged mesons and baryons too, which can not oscillate. For the general decay of a charged meson, the fully time-independent asymmetry is:

$$A_{f^\pm} \coloneqq \frac{\Gamma(i^- \to f^-) - \Gamma(i^+ \to f^+)}{\Gamma(i^- \to f^-) + \Gamma(i^+ \to f^+)} = \frac{|\bar{A}_{f^-}/A_{f^+}|^2 - 1}{|\bar{A}_{f^-}/A_{f^+}|^2 + 1}. \tag{2.12}$$

A third type of CP-violation exists in the interference of direct decay and mixing. For this interference to be possible, mixing between the particle and antiparticle state must occur, therefore this type is also only possible for neutral mesons. In an interference process, the decay can both happen directly as $i \to f$ and with an intermediate oscillation step as $i \to \bar{i} \to f$. The asymmetry for the decay into an CP-eigenstate f_{CP} is:

$$A_{f_{CP}}(t) \coloneqq \frac{\frac{\mathrm{d}\Gamma}{\mathrm{d}t}(\bar{i}^0_{phys}(t) \to f_{CP}) - \frac{\mathrm{d}\Gamma}{\mathrm{d}t}(i^0_{phys}(t) \to f_{CP})}{\frac{\mathrm{d}\Gamma}{\mathrm{d}t}(\bar{i}^0_{phys}(t) \to f_{CP}) + \frac{\mathrm{d}\Gamma}{\mathrm{d}t}(i^0_{phys}(t) \to f_{CP})}. \tag{2.13}$$

Unlike the other two types of CP-violation, this asymmetry can not be expressed in a time-independent form. It is the dominant source of CP-violation for B mesons, and will be discussed in more detail together with them in Section 2.5.

If the CPT theorem holds, violation of CP directly leads to violation of time-symmetry. Long after the others, this was the last symmetry violation to be experimentally discovered [18].

2.4. The *CKM*-mechanism

With the third discovered quark, the strange quark, the question appeared how to explain its observed decay and coupling to the other quarks. The solution from Cabibbo was that the strange quark mixes with the down quark in the decay via the weak interaction [19]. This was soon extended to the *GIM* (Glashow, Illiopoulos and Maiani)-mechanism, a 2×2 rotation matrix [20] relating the mass eigenstates of the down and strange quarks in the weak current:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}. \tag{2.14}$$

The matrix predicted a fourth quark, later discovered as the charm, to complement the up-quark. The second up-type quark also provided a mechanism to suppress decays

of down-strange pairs by cancellation effects, explaining why these were not observed. While the mixing allows the mass eigenstates of up and strange to couple in a vertex to the $W^{\pm}$; in the neutral current, the mixing matrix appears twice, with the two appearances cancelling each other. The 2×2 rotation matrix has only one parameter giving the strength of the mixing, the *Cabibbo angle* θ_C.

The *GIM*-mechanism can not explain CP-violation. This is only possible with an extension from two to three generations of quarks, as was proposed by the Cabibbo-Kobayashi-Maskawa (*CKM*) mechanism [3], before even the charm quark was experimentally observed. The confirmation of CP-violation for B mesons as described by the *CKM* mechanism lead to half of the Nobel prize in physics for Kobayashi and Maskawa in 2008. With the additional third generation of quarks introduced by the *CKM*-mechanism, the mixing matrix, now called the *CKM*-matrix (V_{CKM}), is extended to a 3×3 matrix[6].

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.15}$$

As a general complex 3×3 matrix, it starts with 18 free, real parameters, while a 3-dimensional rotation matrix only has three angles as parameters. After constraining the matrix to be unitary and absorbing parameters into quark phases where possible, not only the three angles but an additional fourth parameter, a complex phase, remains. This additional phase is what allows for CP-violation with three generations of quarks, as now the amplitude of a process after taking the CP-inverse and summing up the constituent amplitudes can result in a different strength.

The entries of the matrix, and from them the four fundamental parameters, can only be determined experimentally. Each of the matrix entries gives the relative coupling strength of the corresponding pair of quarks to the $W^{\pm}$. For a pair of quarks i and j, the corresponding matrix element enters the vertex factor for the coupling to the $W^{\pm}$ as:

$$\frac{-ig_w}{2\sqrt{2}} \gamma^{\mu}(1-\gamma^5) V_{ij}. \tag{2.16}$$

Multiple definitions of the four parameters and how to relate them to the entries are possible, one of them is [21]:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}\mathrm{e}^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}\mathrm{e}^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}\mathrm{e}^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}\mathrm{e}^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}\mathrm{e}^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{2.17}$$

with $c_{ij} = \cos(\theta_{ij})$ and $s_{ij} = \sin(\theta_{ij})$. The three angles in this parametrisation correspond to the coupling strength between the three generations. The angle θ_{12} corresponds

[6] In principle it is arbitrary if the down or up-type quarks are mixed.

to the Cabibbo angle θ_C. While the three angle terms are all real, the phase δ in the term $e^{-i\delta}$ results in complex matrix elements.

Another possibility is the Wolfenstein parametrisation [22]. It uses a Taylor expansion in the factor $\lambda = \sin(\theta_{12})$. With terms up to the order of λ^3 it is:

$$V_{CKM} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + O(\lambda^4). \tag{2.18}$$

The strength of CP-violation in this parametrisation is measured by the factor of $\rho - i\eta$[7].

The magnitudes of the matrix elements according to measurement [4] are currently:

$$|V_{CKM}| = \begin{pmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{pmatrix}. \tag{2.19}$$

The structure of the matrix preferring couplings within the same generation is clearly visible.

One way to validate the CKM mechanism is to use the fact that V_{CKM} is required to be a unitary matrix, which results in range of restrictions the elements must fulfil, depending on which direction the matrix is traversed:

$$\sum_i V_{ij} V_{ik}^* = \delta_{jk} \tag{2.20}$$

$$\sum_j V_{ij} V_{kj}^* = \delta_{ik}. \tag{2.21}$$

The relations summing up to zero can be visualized as triangles in the complex plane, called unitarity triangles. As both the length of the sides and the three angles of each triangle can be measured independently, this allows to overconstrain the equations. The most commonly used triangle is

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = 0, \tag{2.22}$$

as all the sides are in the same order λ^3 in the Wolfenstein parametrisation. The other triangles have one side much smaller than the others, which not only makes visualization more difficult but also requires different precision levels for the different sides to give a useful measurement overall.

Instead of the form from Equation (2.22), the triangle is often shown with the bottom side normalised and two of the corners at $(0, 0)$ and $(0, 1)$. The apex gives the CP-violation strength in this formulation, which can be seen in Figure 2.5.

[7] Often the parameters $\bar{\rho}$ and $\bar{\eta}$ are used instead, which are modified to ensure the unitarity of the Wolfenstein parametrisation in higher orders of λ.

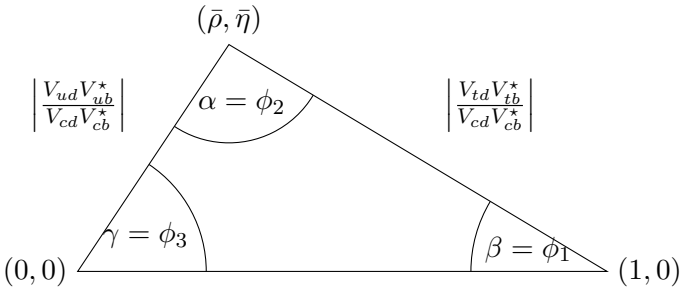


Figure 2.5.: The unitarity triangle for b quarks.

The current result on the measurements to constrain various parameters is shown in Figure 2.6. All measurements are compatible with each other, forming a closed triangle and showing no discrepancies from the *CKM*-mechanism.

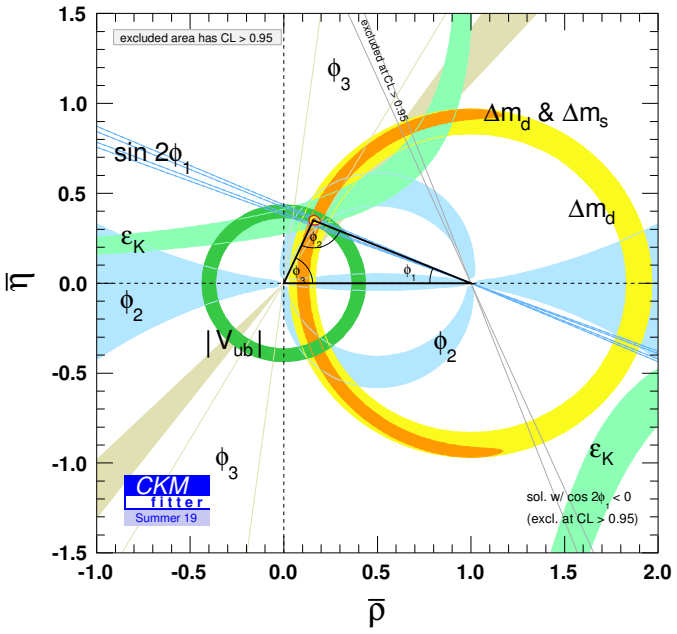


Figure 2.6.: Summary of all current measurements and fits to the unitarity triangle, from [23]

2.5. B mesons

The b quark is the lighter quark of the third generation pair, and the heaviest quark forming bound states. Depending on the type of lighter quark, there are four ground-state mesons, the $B_d^0(d\bar{b})$, $B_u^+(u\bar{b})$, $B_s^0(s\bar{b})$ and $B_c^+(c\bar{b})$[8]. The B_u^+ and B_d^0 are often simply called B^+ and B^0 and are, together with their charge-conjugated partners, the two most important for this analysis.

The large mass of B mesons yields a large variety of decay channels. Via the unitarity triangle shown in Figure 2.5 they are useful to test the theory and detect possible deviations due to new physics or effects not yet accounted for. As the decay of both the $B^\pm$ and the B^0 involve one of the suppressed CKM elements $|V_{cb}|$ and $|V_{ub}|$, they have a relatively long lifetime of $\tau_{B^\pm} = 1638 \pm 4\,\text{fs}$ and $\tau_{B^0} = 1519 \pm 4\,\text{fs}$.

The neutral B mesons, like the lighter mesons mentioned before, can oscillate between the B^0 and the $\overline{B}^0$ states via the diagrams shown in Figure 2.7. Unlike the neutral kaons, the two states resulting from the mixing have almost the same lifetime and instead differ in mass. This oscillation plays an important role in the dominant type of CP-violation for neutral B mesons, which is the interference between the direct and mixing types. The oscillation amplitudes are also useful in measuring the CKM matrix elements $|V_{tb}|$ and $|V_{td}|$, as the internal quark lines are dominantly t quarks.

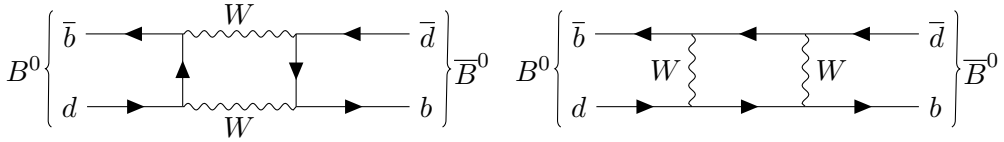


Figure 2.7.: Oscillation Diagrams for neutral B mesons.

This interference occurs in decays to final states accessible by both B^0 and $\overline{B}^0$, to allow a B^0 to either directly decay or decay to the same state after an intermediate oscillation into a $\overline{B}^0$. As such, the final state must be an CP-eigenstate. One such decay channel commonly used is to the final state $J/\psi K_S^0$. The Feynman diagram both with and without the oscillation step for $\overline{B}^0 \to J/\psi K_S^0$ can be seen in Figure 2.8.

The neutral B mesons, B^0 and $\overline{B}^0$, form mixed states similar to the kaons described in Section 2.3. Instead of the lifetime, the main difference between the two mixed states is their mass, consequently they are labelled B_H (heavy) and B_L (light) with the mass difference Δm. For this channel, the asymmetry from Equation (2.13) then takes the

[8] To each of them a charge-conjugated antiparticle is implied

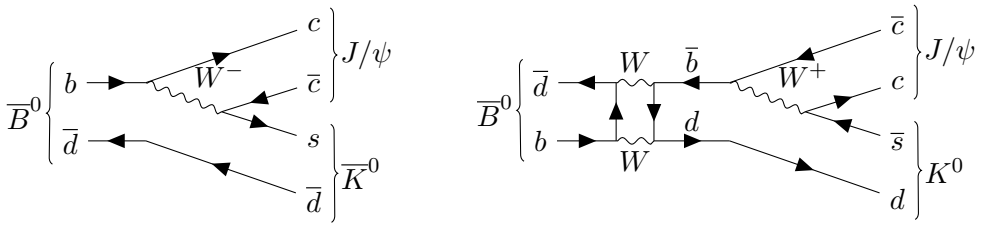


Figure 2.8.: Diagram for $\overline{B}^0 \to J/\psi K^0_S$, once direct (left) and once with and intermediate oscillation (right). The K^0_S is produced via either a K^0 or a $\overline{K}^0$.

form

$$A_{J/\psi K^0_S}(t) \;=\; \sin(\Delta m t)\sin(2\phi_1). \tag{2.23}$$

The time t is the duration after the decaying B meson was in a defined quark state, either after production or tagged via the decay of another correlated B meson out of a coherent production via the $\Upsilon(4S)$ resonance, as is done at B-Factories. This asymmetry provides an opportunity to determine the angle ϕ_1, which enters the asymmetry due to a phase difference of $2\phi_1$ between the direct decay and the decay after mixing.

Unlike for direct or mixing CP-violation, the asymmetry in Equation (2.23) disappears when integrated over time. The requirement to resolve the asymmetry time-dependently to measure the size of CP-violation was a major factor in the design of the B-Factories like Belle, which will be described in Chapter 3.

2.5.1. Semileptonic B meson decays

All B mesons eventually decay. As they are the lightest mesons containing a b quark, their only way of decay involves the decay of this b via the weak interaction, explaining their relatively long lifetime compared to particles with strong decay modes available. In general the decay modes can be grouped into three categories depending on the particles in the final state. They are fully hadronic decays, fully leptonic decays and semileptonic decays. The general decay patterns can be seen in Figure 2.9.

In both hadronic and semileptonic decays the b decays into an up-type quark of the first two generations, due to the difference in size in the CKM-matrix elements involved in the coupling, $|V_{cb}| \gg |V_{ub}|$, the decay into a charm quark is much more likely. Due to the coupling ratio of the W, the hadronic decay is the most likely. However, the hadronic decay channels come with several drawbacks for analyses. The large amount of quarks produced complicate theoretical modelling due to QCD effects. Also, as the W couples to quarks on both its creation and decay, two CKM matrix elements enter in the decay amplitude, making precise measurements more difficult.

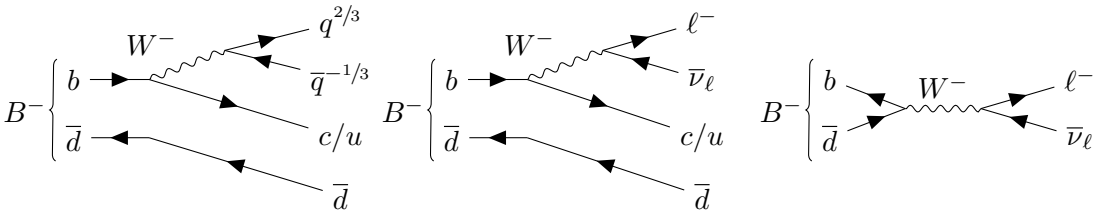


Figure 2.9.: From left to right: Fully hadronic, semileptonic and purely leptonic decay of a B^- meson at tree level.

In contrast, purely leptonic decays are processes without any QCD contribution on the side of the decay products, and therefore theoretically easier to handle. They also only involve one *CKM* matrix element. However, they are helicity suppressed and have a low branching fraction, requiring large datasets for analysis. Semileptonic decays then are in an intermediate position. Their branching fraction is in between the purely hadronic and leptonic types, and like the leptonic decays they only involve one *CKM* matrix element. Also, the leptons created in the W decay do not couple to the strong interaction, greatly reducing, although not eliminating, the QCD effects on the kinematics. This makes them useful to study especially the elements $|V_{cb}|$ and $|V_{ub}|$ involved in b quark decays.

2.5.2. Charmless semileptonic decays

While the majority of B mesons decay into mesons containing a charm quark, a small amount decays in a so-called charmless decay via a direct $b \to u$ transition. As this decay has to cross two generations, it is strongly suppressed by the small size of $|V_{ub}|$. One use of these decays is to measure $|V_{ub}|$ via semileptonic decays, just as $b \to c$ transitions are used in measuring $|V_{cb}|$. The large suppression of charmless decays can be seen when comparing the inclusive branching fractions [24] of all semileptonic decays containing a meson with a c quark (X_c) with that for semileptonic decays so mesons where the b turned into an u (X_u):

$$\mathcal{B}\left(B \to X_c \ell^+ \nu_\ell\right) = (10.65 \pm 0.16)\%, \tag{2.24}$$

$$\mathcal{B}\left(B \to X_u \ell^+ \nu_\ell\right) = (2.13 \pm 0.30) \times 10^{-3}. \tag{2.25}$$

In the case of $B \to X_u \ell^+ \nu_\ell$ decays, the inclusive branching fraction measured above amounts to about twice as much as the sum of all currently exclusively measured individual channels. The matrix element $|V_{ub}|$ can be determined from both the inclusive measurements, accepting all final states containing a $b \to u$ transition, as well as the exclusive measurements explicitly reconstructing specific decay channels. However, while the inclusive measurements results in a value of $|V_{ub}|_{\text{incl}} = (4.32 \pm 0.12_{\text{exp}} \pm 0.13_{\text{theo}}) \times 10^{-3}$, the measurement using the channel $\overline{B} \to \pi \ell^- \overline{\nu}_\ell$ results in the significantly lower value of $|V_{ub}|_{\text{excl}} = (3.67 \pm 0.15) \times 10^{-3}$. This discrepancy shows that there still is improvement

in the understanding of charmless decays needed, both in the theoretical description and in determining the exact composition of the decay modes.

2.5.3. Semileptonic decays to η and η' mesons

Two possible charmless mesons in the $B \to X_u \ell^+ \nu_\ell$ channel are the $\eta(u\bar{u} + d\bar{d} - 2s\bar{s}/\sqrt{6})$ and the $\eta'(u\bar{u} + d\bar{d} + s\bar{s}/\sqrt{3})$ mesons. Both are neutral, pseudoscalar mesons. Their masses [4] are:

$$m_\eta = 547.862 \pm 0.017\,\text{MeV} \tag{2.26}$$

$$m_{\eta'} = 957.78 \pm 0.06\,\text{MeV} \tag{2.27}$$

The decay process can be seen in Figure 2.10. As the $\eta^{(\prime)}$ mesons are both neutral, the only semileptonic decays they can be produced in are those of charged B mesons. The averages of previous measurements [4] of the branching fractions for the two decays are $\mathcal{B}(B^+ \to \eta \ell^+ \nu_\ell) = (3.9 \pm 0.5) \times 10^{-5}$ and $\mathcal{B}(B^+ \to \eta' \ell^+ \nu_\ell) = (2.3 \pm 0.8) \times 10^{-5}$.

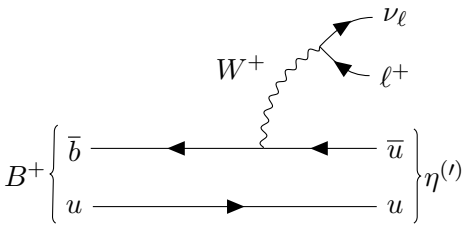


Figure 2.10.: Diagram of the decay process $B^\pm \to \eta^{(\prime)} \ell^\pm \nu_\ell$.

Decay	Branching Fraction [%]
$\eta \to \gamma\gamma$	39.41 ± 0.20
$\eta \to \pi^0\pi^0\pi^0$	32.68 ± 0.23
$\eta \to \pi^+\pi^-\pi^0$	22.92 ± 0.28

Table 2.3.: The three most common decay modes [4] of the η meson.

Decay	Branching Fraction [%]
$\eta' \to \pi^+\pi^-\eta$	42.5 ± 0.5
$\eta' \to \rho^0\gamma$	29.5 ± 0.4
$\eta' \to \pi^0\pi^0\eta$	22.4 ± 0.5

Table 2.4.: The three most common decay modes [4] of the η' meson.

The three most common decay modes for both mesons are listed in Table 2.3 and Table 2.4. In principle all decays of the type $B \to X_u \ell^+ \nu_\ell$ are suited to determine $|V_{ub}|$, however for the $B^\pm \to \eta^{(\prime)} \ell^\pm \nu_\ell$ decays, the necessary theoretical description is not available with sufficient precision. Nonetheless, understanding these decays and measuring their branching fractions precisely is an important step not only to be prepared when theoretical calculations advance, but also to determine the composition of the inclusive $B \to X_u \ell^+ \nu_\ell$ range. The quark transition in the decays is described using form factors. These model the energy distribution between the resulting particles, depending on the energy transfer to the virtual W-boson. Various theoretical models exist to describe the shape, which needs to be verified with experimental results, which are also necessary to derive the overall normalization, as well as other factors depending on the model.

Due to its lower mass compared to mesons containing c quarks, decays to $\eta^{(\prime)}$ can transfer more momentum to the produced leptons. However large amounts of transferred momentum q result in very low-energy, and therefore difficult to measure, hadrons. As a result most previous measurements had limited the acceptance in q to below some threshold, although knowledge over the entire range is important to validate the theoretical description. For the determination of the *CKM* matrix element, the description over the entire range is needed.

CHAPTER 3

Experiment

In the search for CP-violation in B mesons, two experiments were built, Belle and Babar. The data analysed in this work was taken at the Belle experiment located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan. It consists of the KEKB [25] e^+e^- accelerator and storage ring paired with the Belle detector [26] at the single interaction point. As a B-Factory, the setup is optimized to produce and measure as many B mesons as possible. This is accomplished by operating the collider at the $\Upsilon(4S)$ resonance of 10.58 GeV [4]. The $\Upsilon(4S)$ is the lightest of the $b\bar{b}$ states heavy enough to decay via the strong interaction, almost exclusively producing coherent pairs of B^+B^- and $B^0\overline{B}^0$ mesons[1] at about the same rate. The production of a B^+B^- pair is shown in Figure 3.1.

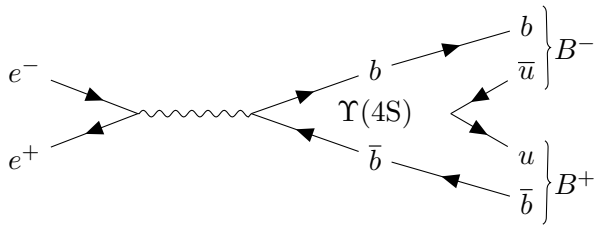


Figure 3.1.: Reaction forming an $\Upsilon(4S)$ from the two beams, followed by the decay to a B^+B^- pair.

To measure the CP-violation according to Equation (2.23), the two B mesons have to be reconstructed individually and their decay-time measured. However, due to the very small difference in mass between the $\Upsilon(4S)$ and the $B\overline{B}$ pair, the B mesons are almost

[1] These are referred to collectively as $B\overline{B}$

at rest in the frame of reference of the $\Upsilon(4S)$, making it very challenging to separate decay vertices between them. A unique feature of the B-Factories provides a solution. By using an asymmetric collider, the $B\overline{B}$ system is strongly boosted in the frame of the detector, which leads to better separation between the vertices due to relativistic effects. The asymmetric design also provides an advantage in measuring the decay time. While measuring the time passed between the two decays is impossible, the distance between the vertices can be measured with much better precision. With the known boost factor this difference directly translates into a temporal difference between the decays.

Following the construction and commissioning phase, the Belle experiment took data in the years from 1999 to 2010. After the end of operation, the detector was upgraded to Belle II [27], together with an upgrade of the accelerator to SuperKEKB [28]. This upgraded setup started data taking in 2019. An overview of the results from the Belle experiment, together with those of the other B-Factory Babar, can be found in [29, 30].

This chapter starts with a short introduction to the accelerator KEK in Section 3.1, before discussing the construction of the Belle detector in Section 3.2, its particle identification system in Section 3.3 and finally the dataset taken over the operation period in Section 3.4. A more detailed description of accelerator and detector can be found in [25, 26].

3.1. KEKB

The KEKB accelerator complex generates the colliding beams for the Belle experiment by accelerating electrons and positrons in several steps. It developed out of its predecessor TRISTAN, whose general layout and tunnel it inherited. The general layout of the accelerator can be seen in Figure 3.2. Electrons are accelerated in the linear accelerator $(LINAC)$[2] up to their collision energy. Partway during the acceleration, some of these electrons are collided with a target to generate positrons, which are also accelerated to serve as the collision partners.

After reaching their required energies, the electron and positron beams are injected into a pair of storage rings with a circumference of 3016 m. The storage rings are placed side-by-side in the tunnel. Due to the asymmetric design of the accelerator, these rings are operated at different energies. The electrons are injected into the high energy ring (HER) and the positrons into the low energy ring (LER). When operating at the $\Upsilon(4S)$ resonance with $\sqrt{s} = 10.58\,\text{GeV}$, the HER has an energy of 8.0 GeV and the LER of 3.5 GeV, which corresponds to a Lorentz boost factor of $\beta\gamma = 0.425$ for the centre-of-mass(CM) frame. Other energy settings have also been used.

The decay channels used to investigate CP-violation have a low overall branching fraction. A statistically significant measurement therefore requires the collection of a large

[2] Despite the name it contains a 180° turn necessary to fit inside the KEK area.

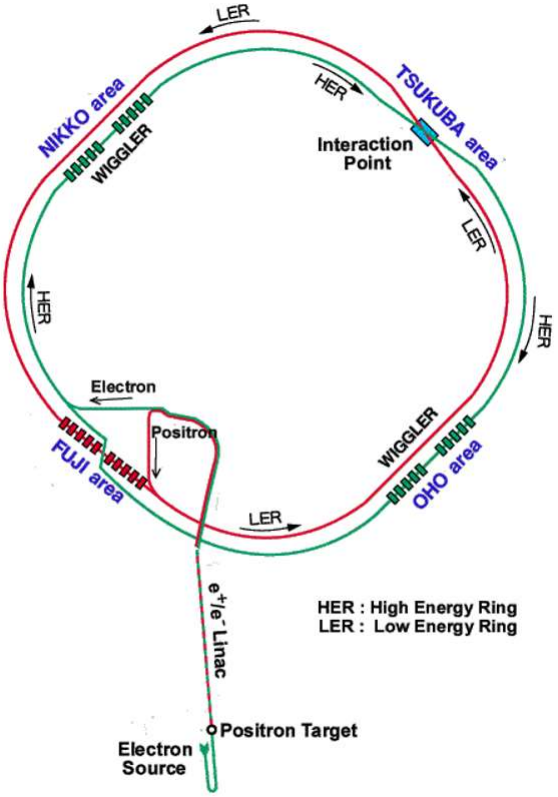


Figure 3.2.: Schematic view of the KEKB installation, from Ref. [29]

data sample of collisions. The amount of collision data an accelerator setup accumulates is measured with the luminosity:

$$\mathcal{L} = \frac{N_b n_{e^-} n_{e^+} f}{A_{eff}}. \tag{3.1}$$

In both rings, the electrons and positrons circulate in discretely spaced groups called bunches. Each ring is designed to contain $N_b = 1584$ such bunches, with the number of particles contained in a bunch $n_{e^\pm}$, typically between 10^{10} and 10^{11}. The bunches circulate with the frequency f, and A_{eff} is the effective cross-section between the bunches at their point of interaction.
The original design luminosity of the KEKB accelerator was $\mathcal{L} = 1 \times 10^{34}/(\mathrm{cm}^2\,\mathrm{s})$. After several improvements over the time of operation more than twice of that design goal could be achieved with $\mathcal{L} = 2.1 \times 10^{34}/(\mathrm{cm}^2\,\mathrm{s})$. During beam operation, the electrons and positrons continually deplete and the beam currents decrease. KEKB used a continuous injection scheme to constantly replenish spent particles.

The complex housing KEKB has four experimental halls inherited from its predecessor TRISTAN. Only one of these contains a detector, built around the only interaction

point at Tsukuba Hall.

3.2. Belle

The Belle detector is a general-purpose detector covering almost the entire solid angle range, situated at the only interaction point (*IP*) of the KEKB accelerator complex. At this point, the two beams collide under an angle of 22 mrad producing either the desired $\Upsilon(4S)$ meson or a variety of background processes. These background processes can themselves be used as signal for a variety of other studies not involving a $\Upsilon(4S)$. Before reaching the collision, the beams pass through a final set of focusing magnets. Due to the collision angle the interaction region is limited to only the point where the beams intersect, allowing a precise determination of the *IP* for analyses.

The detector consists of three main segments, the *barrel* in the central region with the various sub-detectors arranged in layers, and the forward and backward *endcaps* measuring particles travelling at angles close to the beam pipe. The forward direction is defined as the direction of the electron beam, and coincides with the boost direction of the CM system. The boosted CM system results in an asymmetric detector design to account for the difference in particles expected between the forward and backward sides. The detector is surrounded by a solenoid magnet producing a 1.5 T magnetic field oriented in parallel to the positron beam direction.

To measure CP-violation, both a full coverage to detect all resulting particles, as well as a high precision is required. Good vertexing capabilities are necessary to distinguish the decay vertices of the two B mesons and determine their distance, together with particle identification (`PID`) capabilities to determine the B meson flavour. The high luminosity requires the inner detectors to have an especially high tolerance for radiation damage to not significantly degrade in performance during the operation. Conversely, to minimize the effect of interactions with the detector material by the passing particles on the resolution, the material budget in terms of radiation length X_0 has to be kept as low as possible especially for the inner detectors.

The global coordinate system of Belle is defined around an z axis along the inverse of the positron beam direction. Due to the small crossing angle this is different from the electron beam axis. The x and y axes point outward and upward of the accelerator rings respectively. Additionally, polar coordinates are used with the angles polar angle θ with respect to the z axis and the azimuthal angle ϕ. The radius is defined as $r = \sqrt{x^2 + y^2}$.

In the following the individual sub-detectors making up Belle are described in more detail. They are, beginning from the innermost sub-detector, the Silicon Vertex Detector (`SVD`), the Central Drift Chamber (`CDC`), the Aerogel Cherenkov Counter (`ACC`), the Time-of-Flight (`TOF`) system, the Electromagnetic Calorimeter (`ECL`) and the K_L^0 and

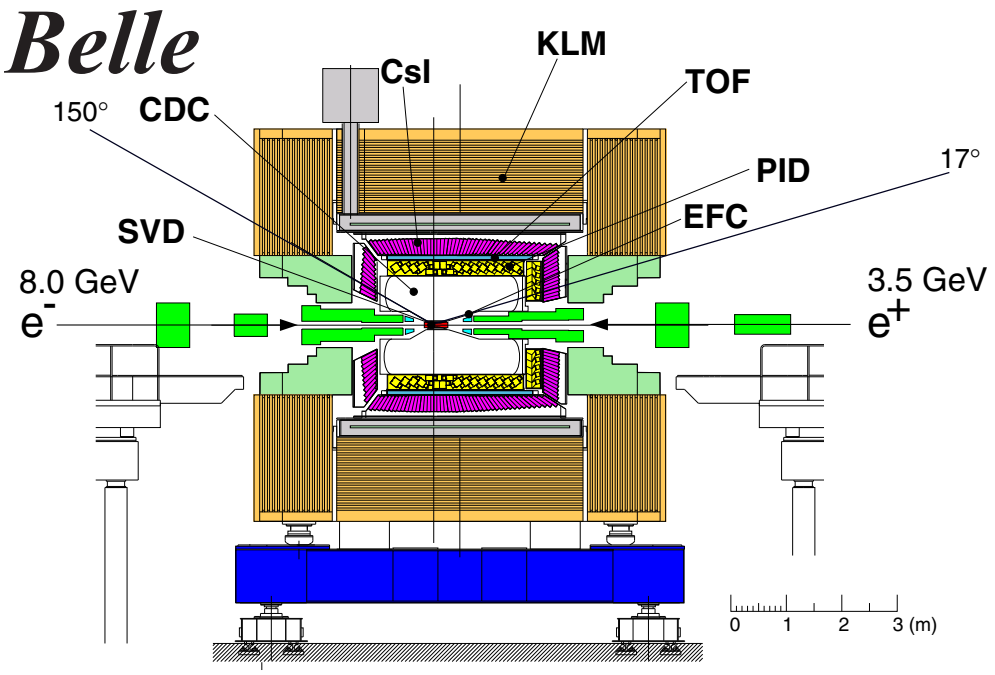


Figure 3.3.: Schematic side view of the Belle detector, showing locations of the components, taken from Ref. [26].

Muon Detector (KLM). An overview of the entire layout can be seen in Figure 3.3. The angular acceptance of all subsystems, when not stated otherwise, is $17° < \theta < 150°$.

3.2.1. Silicon Vertex Detector

The innermost sub-detector is the Silicon Vertex Detector (SVD). It is situated directly outside of the beam pipe, which is made out of beryllium to reduce the effects of Coulomb scattering, together with a gold coating to reduce synchrotron radiation. The beam pipe has two layers with cooling in between to counteract the energy absorbed from the beam.

The SVD uses layers of double sided strip detectors (DSSD) to measure the point of traversal of particles to reconstruct the trajectory of charged particles and from it the decay vertex. Particles passing through the strips release electrons from the silicon, which can be measured as an electric current. The modules of each layer contain strips on both sides rotated by 90° relative to each other. The combination of the two sides allows the determination of the point of traversal of particles.

Over the time of its operation, Belle used two different SVD configurations. The original design, called SVD1, had three layers of DSSD modules. The three layers are positioned

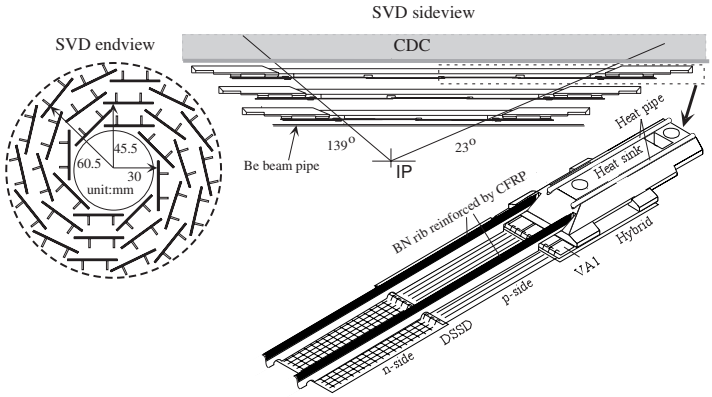


Figure 3.4.: Layout of the modules of `SVD1`, and schematic of one such module. Taken from Ref. [26].

at radii of 30 mm, 45.5 mm and 60.5 mm from the *IP* with some overlap between the modules of each layer to prevent gaps. The layout can be seen in Figure 3.4. The angular acceptance of `SVD1` was smaller than that of the detector in general at only $23° < \theta < 139°$. The original design was not sufficiently radiation hard and needed frequent exchanges of damaged parts.

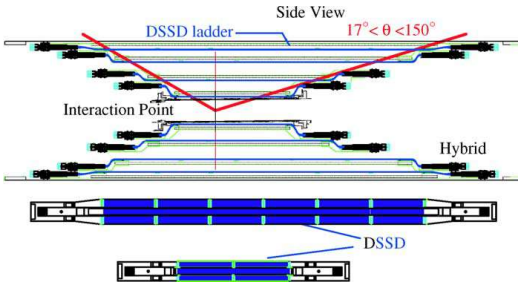


Figure 3.5.: Layout of the upgraded `SVD2`. Taken from Ref. [31].

In the summer of 2003, it was replaced with the `SVD2` [31] of a new design. Besides being more durable, it added a fourth layer. Due to a new beam pipe of smaller diameter, the innermost layer could be placed closer to the *IP* than before, with the radii of the layers now being 20 mm, 43.5 mm, 70 mm and 88 mm. The configuration is shown in Figure 3.5. The `SVD2` also improved the angular acceptance to $17° < \theta < 150°$. The impact parameter resolution of both `SVD` configurations was determined with cosmic rays and can be seen in Figure 3.6. The resolution is strongly dependent on the momentum of the measured particle. Especially in the z direction `SVD2` provides a significant improvement over the previous `SVD1`. About 85% of the integrated luminosity was taken with this second configuration. This is sufficient to achieve the 100 µm resolution on the decay vertex distance between the $B\overline{B}$ necessary for measurements of CP-violation.

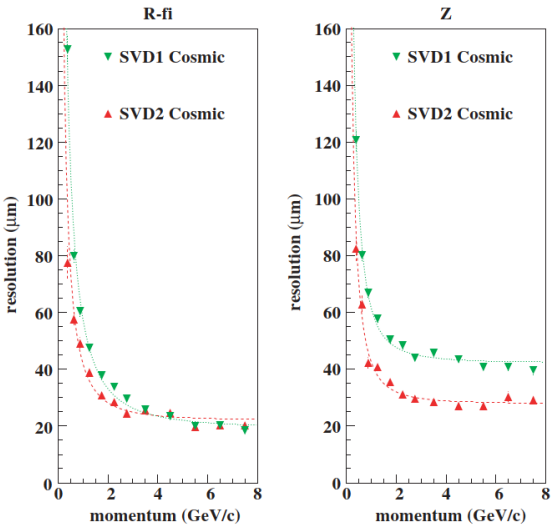


Figure 3.6.: Resolution of the SVD1 and SVD2, on the left in $r\phi$ and on the right in z. Taken from Ref. [31].

Hits on the SVD are matched to charged particle tracks extrapolated inwards from the following detector, the CDC. These added hits increase the precision of the track extrapolation and give a better determination of the original vertex inside the beam pipe. When multiple particles pass through the SVD close enough together, an ambiguity can develop between the strips in both directions, giving rise to additional hits unrelated to real particles. These can be resolved as only combinations related to an actual particle passing through should have a corresponding CDC track pointing towards them.

3.2.2. Central Drift Chamber

The Central Drift Chamber (CDC) is used to measure the trajectory of charged particles and, together with the SVD, forms the Belle tracking system. The CDC is inside the 1.5 T magnetic field. The Lorentz force induces a curvature on the trajectory of particles which is used to determine the momentum. It also contributes to identifying the particle type by measuring the energy loss $\mathrm{d}E/\mathrm{d}x$ and provides information to trigger events on charged tracks.

The CDC is a typical drift chamber, consisting of a gas-filled barrel with wires inside both for applying an electric field and measuring current. It occupies the region from an inner radius of 80 mm to an outer radius of 880 mm. In parallel to the upgrade to SVD2, the inner configuration was changed to accommodate the new, larger SVD2 detector. This changed the inner radius to 110 mm. The original design before this change is shown in Figure 3.7. The gas used to fill the CDC is a mixture of 50%He and 50%C_2H_6(ethane).

A gas mixture with a low atomic number Z was chosen to minimize multiple scattering and the resulting effects on the particle trajectory.

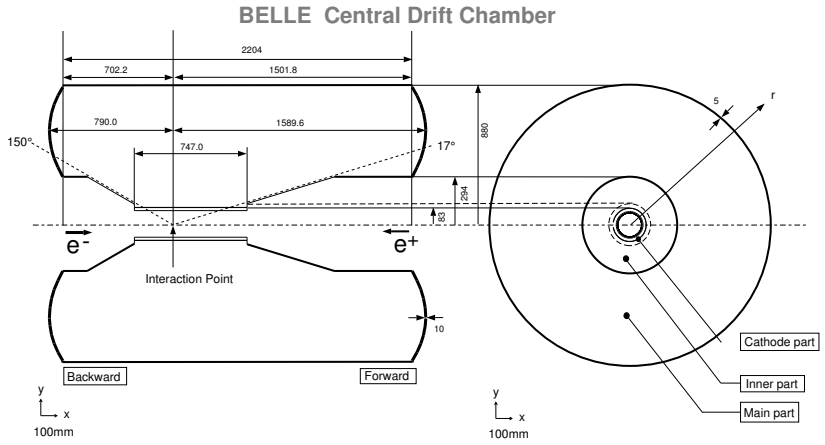


Figure 3.7.: Overview of the Belle drift chamber design, taken from Ref. [26].

The CDC contains a total of 8400 drift cells, each consisting of one anode sense wire, surrounded by eight cathode field wires. The sense wires used for the actual measurement are made out of 30 µm thick, gold-plated tungsten, while the field wires are 126 µm thick aluminium. The wires are grouped into 50 cylindrical layers. Between three and six layers form groups, alternating between axial layers parallel to the beam pipe and stereo layers being positioned at a small angle. The angled stereo wires are used to gain additional information on the z coordinate. The three innermost layers are formed out of cathode strips. These are segmented and used to improve the resolution in z and to trigger on events.

If a charged particle traverses the gas volume, it ionizes the gas particles and frees electrons. These electrons get accelerated towards the sense wires, turning into a cascade due to collisions and further ionizations with the gas until they reach the wire, where the induced current of these electrons is measured. The initial energy of the electrons, and therefore the strength of the cascade reaching the wire, depends on the type of the particle. The timing until the electrons reach the sense wire is used to determine the distance between the wire and the trajectory at its closest approach. The CDC reaches a spatial resolution of around 130 µm.

Particles traversing the CDC lose energy when interacting with the gas by scattering and ionization. This energy loss $\mathrm{d}E/\mathrm{d}x$ for a given momentum is different for each type of particle due to their different masses and can be measured proportional to the current induced by the ionization. The different behaviour of the particles typically measured by the CDC is compared in Figure 3.8. The $\mathrm{d}E/\mathrm{d}x$ measurement forms an important part

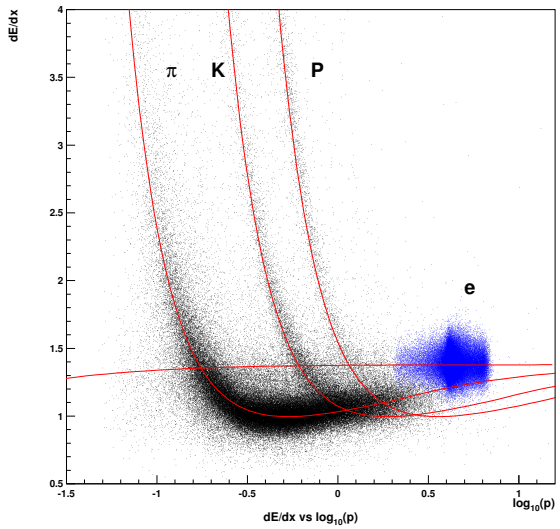


Figure 3.8.: $^{\mathrm{d}E}/_{\mathrm{d}x}$ behaviour of different particle types in the CDC, taken from Ref. [26].

of the PID.

3.2.3. Aerogel Cherenkov Counter

Directly outside the CDC come two PID systems primarily used to distinguish $\pi^{\pm}$ from $K^{\pm}$. The first is the Aerogel Cherenkov Counter (ACC) for particles with high momenta above $1.2\,\mathrm{GeV}/c$, up to a limit of about $4\,\mathrm{GeV}/c$. It consists of two parts, the barrel covering an angular acceptance of $33.3° < \theta < 127.9°$, and the forward endcap covering $13.6° < \theta < 33.4°$.

The barrel part contains a total of 960 sensor modules, arranged in 16 sectors along the θ angle. Each sector consists of 60 identical modules arranged symmetrically around ϕ. The forward endcap consists of 288 modules, arranged in 5 layers around the beam pipe. All modules are oriented to point towards the *IP*. The general layout of the ACC can be seen in Figure 3.9.

The detection uses Cherenkov light emitted by particles travelling faster than the speed of light in the traversed medium. If the material has the refractive index n, the threshold velocity to emit Cherenkov light is $\beta = {}^{v}/_{c} = {}^{1}/_{n}$. An individual ACC module contains five layers of silica aerogel material with a refractive index between 1.01 and 1.03, depending on the polar angle. Photons caused by passing particles are detected with one or two fine-mesh photomultiplier tubes capable of operating inside the 1.5 T magnetic field of Belle. The construction of a barrel and endcap module is shown in

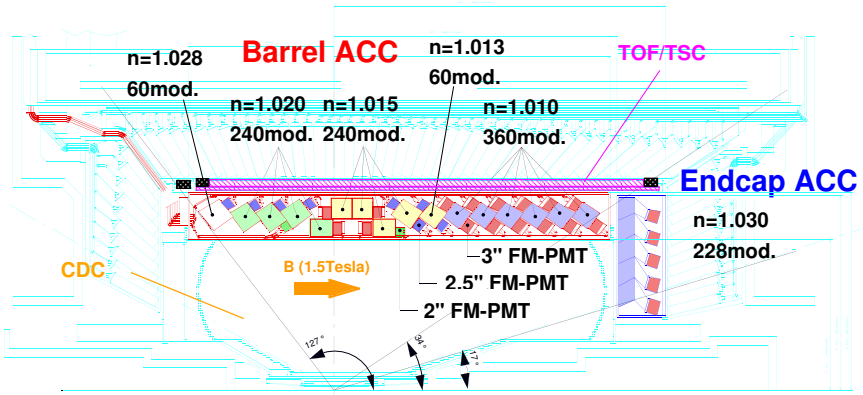


Figure 3.9.: Arrangement of the ACC modules, taken from Ref. [26].

Figure 3.10.

The refractive index of the aerogel is chosen such that $\pi^{\pm}$ in the ACC energy range emit Cherenkov light while $K^{\pm}$ do not. If tracks found with the CDC and SVD point towards towards an area inside the ACC acceptance, whether a module in that direction detected sufficient photons gives a clear distinction about what type of particle caused the track.

3.2.4. Time-of-Flight counter

The Time-of-Flight (TOF) system is the second of the PID detectors, complementing the ACC in distinguishing particles between $\pi^{\pm}$ and $K^{\pm}$. It is effective for particles with momenta below 1.2 GeV/c, extending on the ACC range. The TOF only contains a barrel part with an angular acceptance of $34° < \theta < 120°$. In the forward direction therefore particle identification has to rely on ACC. Particles in the momentum range covered by TOF make up 90% of the decay products in an average Belle event. The position of the TOF directly outward of the ACC can be seen in Figure 3.9.

The TOF determines particle type by precisely measuring the time it takes a particle originating from the *IP* to reach the TOF detector modules, which are about 1.2 m away. The exact flight length and momentum the particle is derived from the tracking detectors CDC and SVD. The time of interaction in the beam pipe is determined from the accelerator timing and needs to be precisely calibrated to the TOF readout to determine the exact difference. Adding this time of flight to the trajectory information allows determination of the mass and therefore type of the measured particle. The minimal energy needed to

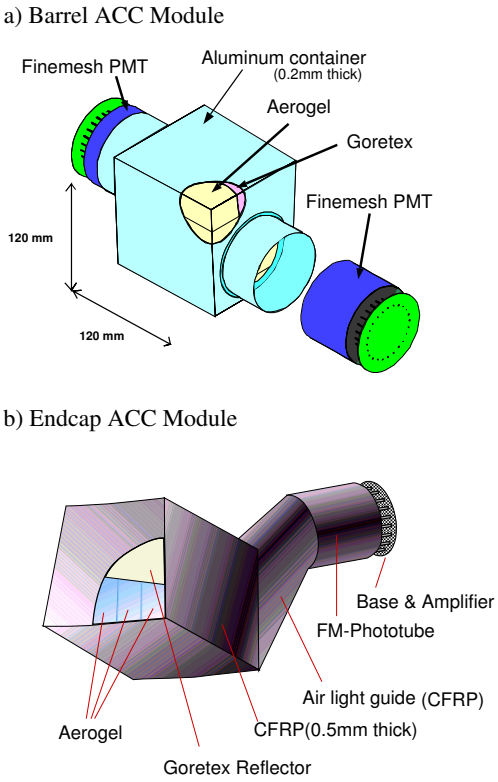


Figure 3.10.: Schematic of two different types of individual ACC modules, taken from Ref. [26].

reach the TOF is 0.28 GeV, particles with even lower momentum stay inside the CDC and PID must rely on the $\mathrm{d}E/\mathrm{d}x$ measurement alone.

The TOF consists of 128 plastic scintillators read out by fine-mesh photomultiplier tubes at each end. These scintillators are grouped in pairs into 64 TOF modules, together with an additional scintillator located inwards. This additional scintillator is used for the Trigger Scintillation Counter (TSC) as part of the trigger system and only read out my one photomultiplier tube located at the backwards side. The layout of one such module can be seen in Figure 3.11. The TSC is separated from the TOF with a 1.5 cm gap. This gap prevents background events arising from photon interactions in the TSC to reach the TOF scintillators, allowing background events to be reduced by requiring a signal in both. The time resolution achieved by the TOF is around 100 ps.

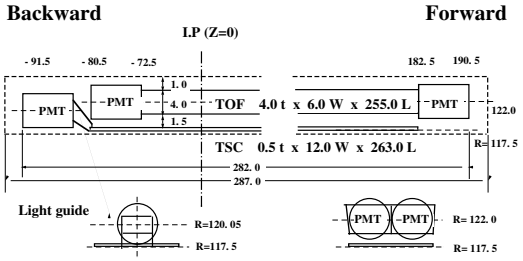


Figure 3.11.: Measurements and general arrangement of the TOF and TSC, taken from Ref. [26].

3.2.5. Electromagnetic Calorimeter

Electrons and photons are stopped by the material of the Electromagnetic Calorimeter (ECL). This interaction allows to measure their energy. Photons can only be measured by the ECL and occur in a wide range of energies at Belle, from about 4 GeV directly in B meson decays down to below 0.5 GeV at the end of a decay chain. Separating individual photons requires a high granularity of the detector. The ECL is also used in separating electrons from other charged hadrons.

The ECL consists of a central barrel part and two endcaps for particles at angles closer to the beam pipes. The barrel with an inner radius of 1.25 m is situated directly outside the TOF, which is attached to it. The barrel length is 3.0 m and it provides an angular acceptance of $32.2° < \theta < 128.7°$. The forward endcap is located 2.0 m forward of the IP and provides an additional angular acceptance $12.4° < \theta < 31.4°$. The backward endcap is located 1.0 m backwards from the IP and has an acceptance of $130.7° < \theta < 155.1°$. The endcaps are used to measure particle on trajectories leaving the CDC at the front or back. Between the barrel and the endcaps are small gaps in acceptance to make room for cables to the inner detectors to pass, reducing the overall acceptance by about 3%. The general layout can be seen in Figure 3.12.

The ECL is an array of 8736 modules using thallium-doped caesium iodide CsI(Tl) crystals as scintillation material and silicon photodiodes to measure the scintillation light. Unlike in the inner detectors, where a low material budget is desired, the ECL scintillator material is optimized for the opposite. The crystal length of 30 cm corresponds to 16.2 X_0, here a large value is desired to stop the particles and measure their total energy inside the ECL. The modules are oriented to point approximately towards the IP. The slight deviation allows for some overlap between modules to increase the acceptance, as otherwise particles could pass through the gaps between the modules undetected.

Electrons and charged hadrons entering the scintillator crystals emit bremsstrahlung photons reducing their energy, while photons convert to electron-positron pairs. These processes repeat, creating a shower of particles until they fall below an energy threshold

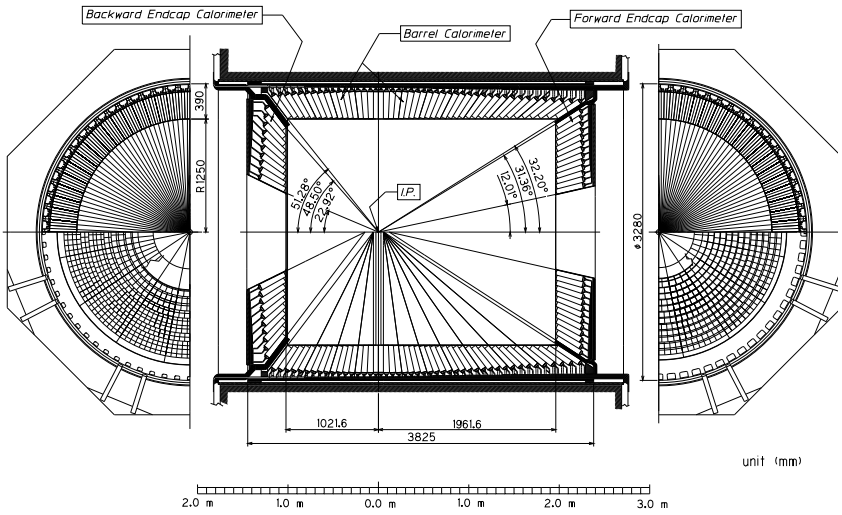


Figure 3.12.: The layout of the Belle electromagnetic calorimeter, taken from Ref. [26].

and instead cause the emission of visible light by the interaction with the scintillator. This light is then measured by the photodiodes. In general the particles created during the shower process are not limited to only one crystal cell, but spread out over the neighbouring cells as well. These cells together form a *cluster* in the ECL.

The resolution in energy and incident angle measurement depends on the energy of the measured particle, improving with increasing energy. Starting at 0.1 GeV it is 4% in energy and about 13 mrad in angle, improving to 1.6% and 3 mrad for energies of 8 GeV.

Similar to the ECL is the Extreme Forward Calorimeter (EFC). It is a separate system located outside the main Belle detector, close to the beam pipe to measure particles at very small angles, mostly coming from Bhabha scattering and beam backgrounds. Measuring the rate of Bhabha events is important as it is directly linked to the luminosity of the accelerator. While the general principle is the same as for the ECL, the crystals are made out of Bismuth Germanate ($Bi_4Ge_3O_{12}$), which is more resistant to the high radiation levels in this region.

3.2.6. K_L^0 and Muon Detector

The outermost sub-detector is the KLM, primarily measuring K_L^0 and $\mu^\pm$. These interact weakly at the energies typical for the Belle experiment and are not stopped inside the previous components. Even the additional material of the KLM is not enough to fully stop these particles however and they leave the detector. The energy of muons can be determined with the momentum measurement from the tracking detectors, the KLM giv-

ing additional hits for the trajectory and providing PID information as no other charged particle should reach that far. However, K_L^0 leave no track, so only a lower bound on their energy can be determined from the energy they deposit in the KLM and ECL together with a measurement of their direction.

The KLM, like the ECL, consists of a barrel and two endcaps and also serves as the flux return for the 1.5 T solenoid magnet. The angular coverage is $20° < \theta < 155°$, of which the barrel covers $45° < \theta < 125°$. The KLM is build of a structure of alternating 4.7 cm thick iron plates and detection modules. While the barrel has 15 detection modules, the endcaps have only 14. The iron serves as interaction material for the passing particles, and the 14 iron plates used in all regions of the detector add 3.9 interaction lengths to the 0.8 interaction lengths of the ECL for the K_L^0 to react and deposit energy.

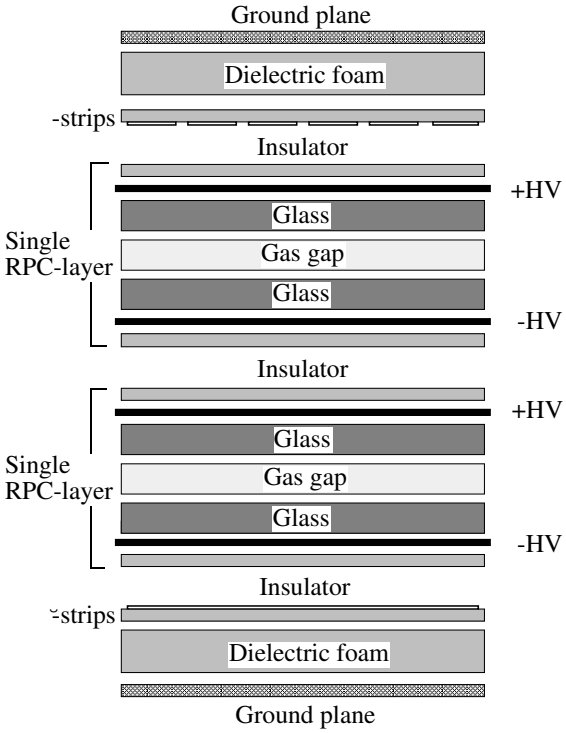


Figure 3.13.: Configuration of one KLM detection layer, taken from Ref. [26].

A detection module consists of two resistive plate counters, each composed of two glass electrodes with a gas-filled gap in between and readout strips at the outer ends. The arrangement can be seen in Figure 3.13. Particles passing through the detection module ionize the gas, and the ionized gas in turn causes and electric discharge between

the electrodes. This discharge is measured and with the orientation of the readout strips gives the point of traversal. Particles need a momentum of at least 600 MeV/c to reach the KLM. The angular resolution achieved is below 10 mrad.

3.2.7. Trigger and Data Quality Monitoring

It is important to save all physics events useful for a later analysis. However, most of the signals measured by the detector components are background events or noise. Attempting to save these as well would overwhelm both the bandwidth of the data acquisition (DAQ) as well as the data storage. The trigger system (TRG) is responsible to select which events are worth saving and which can be discarded immediately.

The trigger works in several stages, starting with the level 1 (L1) trigger implemented in hardware using custom electronics. It takes inputs from several sub-detectors, the CDC and TSC to determine the presence of charged particles and from the ECL for photons, as well as the KLM. These inputs are used by the Global Decision Logic (GDL) to select events with indications of particles high enough in energy and sufficiently away from the beam direction to indicate more than beam background.

This selection is followed by the level 3 (L3) trigger after all data is provisionally read out. It uses a simplified form of the event reconstruction optimized for speed. This further inspects events chosen by L1 due to criteria indicating charged tracks in more detail, and reject further background. Events passing both trigger levels are saved for further analysis, where additional selections are applied depending on their purpose. The trigger efficiency for hadronic events is above 99.5%.

If the conditions of one of the sub-detectors deteriorates during operation, all data recorded from then on is taken under suboptimal conditions, if that sub-detector still records usable data at all. It is therefore important to detect problematic behaviour as early as possible, so the cause can be rectified and the quality of the data improved. This is the task of the Data Quality Monitoring (DQM). It uses a subset of the recorded events to immediately calculate quantities indicating detector performance, while also providing a reference for nominal conditions. This comparison is used to quickly discover and investigate any deviations.

3.3. Particle identification

Most measurements at Belle rely on reconstructing decay chains originating from B mesons out of the numerous particles measured by the detector in each event. The five charged particles with sufficient lifetime to be candidates for the tracks measured are $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$ and $p^{\pm}$. The Particle Identification (PID) assigns one of these particle

hypotheses to the tracks measured in the detector by combining information from several sub-detectors. For a track, the likelihood of a particle of each type giving the measured response is formed, and ratios of these likelihoods are used in analyses.

3.3.1. Hadronic likelihood

For the three types of hadrons $\pi^\pm$, $K^\pm$ and $p^\pm$, likelihood ratios are built comparing the probability to be one type over the other [32]. The likelihood of a candidate being a $\pi^\pm$ over a $K^\pm$ is:

$$\mathscr{P}_{\pi/K} = \frac{\mathscr{L}_\pi}{\mathscr{L}_\pi + \mathscr{L}_K}. \tag{3.2}$$

For other combinations of particles the likelihood is built analogously. The individual likelihoods $\mathscr{L}_i$ for particle type i are

$$\mathscr{L}_i = \mathscr{L}_i^{\texttt{CDC}} \times \mathscr{L}_i^{\texttt{TOF}} \times \mathscr{L}_i^{\texttt{ACC}} \tag{3.3}$$

combining likelihoods from three different sub-detectors. The specific measurements used are:

- The energy loss in the CDC. The different behaviour for the particle types can be seen in Figure 3.8. In case a track is outside of the angular acceptance of the ACC and TOF, or has not enough energy to reach those, $^{\mathrm{d}E}/_{\mathrm{d}x}$ is the only measurement available.
- The Cherenkov light yield measured by the ACC.
- The flight time measured by the TOF.

3.3.2. Electron likelihood

The electron probability [33] is defined as the likelihood ratio of a track being an electron against not being one:

$$\mathscr{P}_e = \frac{\mathscr{L}_e}{\mathscr{L}_e + \mathscr{L}_{\mathrm{non}-e}}. \tag{3.4}$$

These likelihoods use five measurements, two of which are also used in the hadronic likelihoods:

- The energy loss $^{\mathrm{d}E}/_{\mathrm{d}x}$ in the CDC
- The Cherenkov light yield measured by the ACC

Additionally, three measurements based on the ECL are included:

- The shape of the shower measurement. Electrons produce more compact showers spread over fewer crystals than hadrons.

- The ratio of the energy absorbed by the ECL to the track momentum determined in the CDC. As most electrons get stopped in the ECL and deposit all their energy, this should be close to one, while hadrons pass through the ECL without depositing their entire energy.
- The quality of the matching between the CDC track and the associated ECL cluster. A well reconstructed trajectory should point directly toward its associated cluster.

3.3.3. Muon likelihood

The muon identification [34] is defined as the likelihood ratio of a track being a muon against the hadrons $\pi^{\pm}$ and $K^{\pm}$:

$$\mathscr{P}_{\mu} = \frac{\mathscr{L}_{\mu}}{\mathscr{L}_{\mu} + \mathscr{L}_{\pi} + \mathscr{L}_{K}}. \tag{3.5}$$

Muon candidates are all CDC tracks that have associated hits in the KLM. Two quantities correlating these two are used to build the likelihood:

- The matching quality between the trajectory extrapolated outwards of the CDC and the KLM hits. If the trajectory is well determined, it should pass close to correctly assigned hits.
- The penetration range determined from the momentum in the CDC compared with the range required to reach the associated part of the KLM.

3.3.4. Neutral particles

While not strictly part of the PID, neutral particles have to be assigned as well. All clusters in the ECL without an associated track pointing towards it are reconstructed as photons, while similarly all KLM hits without a corresponding track are reconstructed as K^0_{L}.

3.4. Dataset

During the operational period from 1999 to 2010 the Belle experiment took data mainly at the $\Upsilon(4S)$ resonance. However, significant data sets where also taken with other accelerator settings, investigating the lighter resonances $\Upsilon(1S)$, $\Upsilon(2S)$ and $\Upsilon(3S)$ as well as the heavier $\Upsilon(5S)$, which can decay into $B_s\overline{B_s}$ pairs. So called off-resonance data was taken with the beam energy set just slightly below each of these resonances, to study the background components not involving an intermediate Υ resonance. All data taken together reached an integrated luminosity over $1\,\mathrm{ab}^{-1}$.

This analysis uses the entire dataset available to study $B\overline{B}$ pairs at the $\Upsilon(4S)$ resonance, which has an integrated luminosity of $711\,\mathrm{fb}^{-1}$. This corresponds to

$$N_{B\overline{B}} = 772 \pm 11 \times 10^6 \tag{3.6}$$

$B\overline{B}$ pairs [30]. Additionally, 79.4/fb of data taken 60 MeV below the $\Upsilon(4S)$ resonance is used to validate the simulation of background processes.

CHAPTER 4

Statistical Methods

Analysing the data taken at experiments in particle physics requires a variety of statistical tools and methods. They broadly fall into two categories. The first are methods to classify events into categories and techniques to construct quantities useful for that purpose, often using simulated data. The second method concerns generating such simulated data, called *Monte Carlo* (MC), to develop the analysis and compare the expectation with the measured result.

This chapter starts with introducing boosted decision trees (`BDT`) as a method to categorize events in Section 4.1. Following is a description of the fit method used to extract the amount of signal events in the final sample in Section 4.2. At Belle, B mesons are always produced in pairs. In Section 4.3 different reconstruction methods for the second B meson in an event are compared. Afterwards, methods to identify events without such a pair of B mesons are introduced in Section 4.4. Finally, Section 4.5 discusses the simulated data used throughout this analysis.

4.1. Boosted decision trees

One of the most important steps in reconstruction analyses such as this work is to categorize events into either signal events, containing the process of interest, or background. To improve on the traditional method of using sequential cuts on variables one after another, multivariate algorithms (`MVA`) [35] are used. Instead of only a single variable at each step, they can use the full variable space including the correlations between variables to develop a prediction which category a given event belongs to.

Before such an `MVA` can be used to classify events, it needs to be trained first. This

training determines the parameter set used later in the classification. One commonly used way is supervised machine learning with a set of events, usually from a simulation, for which the category they belong to is known in advance. Besides the category, the events need to have a set of variables quantifying the features the `MVA` can use in learning how to distinguish the categories.

When training any type of `MVA` classifier, it is important for the method to have enough flexibility to learn the features of the input variables, but stop learning once all real features are accounted for and only statistical fluctuations of the training set are left. Further training to build a more complex classification model including these fluctuations is called overtraining. While overtraining can lead to better classification performance on the data set used during the training, this does not translate to the same performance on statistically independent data sets.

To prevent this type of overtraining, the sample used by the `MVA` is split in two. The first half is used by the `MVA` to learn the features to develop the classification model. The separation achieved during the training phase is compared with that achieved on the second half. As long as actual features of the underlying distribution are learnt, the performance on both should be similar, however once overtraining sets in, further improvements can only be seen on the training sample and the performance begins to diverge.

The following explains the `MVA` classifier used in this thesis, boosted decision trees, in more detail. Out of the many different configurations possible, this chapter will mostly focus on what is used in the application following in Chapter 6.

4.1.1. Decision trees

A decision tree (`DT`) consists of a series of one-dimensional splits of the dataset. The starting point is a root node containing the entire data set. At each node, the variable and split value on it that gives the largest improvement in separation is determined, and used to separate the remaining sample into two new nodes. The concept is visualized in Figure 4.1. Multiple definitions of this separation are possible, a common one and used here is the Gini index $p \times (1 - p)$ with the purity p. The purity is defined as N_{Sig}/N_{All}, the fraction of signal events (N_{Sig}) in the sample (N_{All}). The lower the value of the Gini index for a given node is, the more it is dominated by events of one particular category, as it is by definition symmetric in both categories. The best split is the one where the sum of the Gini indices of the two newly created nodes, weighted by the amount of events in them, is lowest.

This procedure is repeated until a predefined depth in number of splits from the root node is reached. Nodes without further splits are taken as signal or background nodes depending on their contents. Such nodes without further splits are called leafs. When

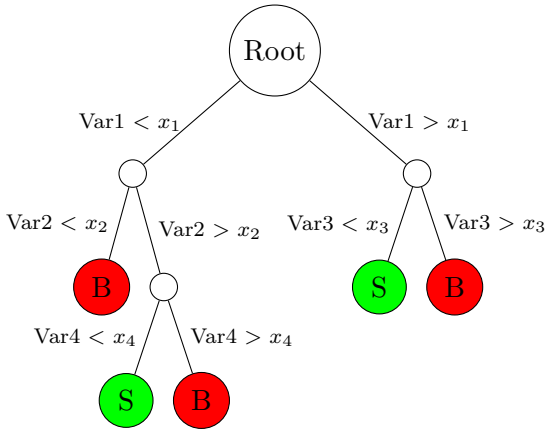


Figure 4.1.: Schematic view of a single decision tree, growing downwards from the root node at the top.

using such a decision tree to classify data, each event passes through the splits after each node until it reaches a leaf and is assigned to the corresponding category. A single decision tree therefore gives a binary response for one of the two categories, without any measure of certainty.

Single decision trees need to be trained to a great depth to reach significant classification power. However, this results not only in many branches where the final splits only marginally improve the separation, but also makes them vulnerable to overtraining. Such splits are therefore usually removed after training finished by a procedure called pruning.

4.1.2. Boosting

As the performance achievable with a single decision tree is limited, several methods to improve performance are available. Instead of single trees, they use boosting methods to build a collection of decision trees, called a forest. The results of the trees in the forest are combined to give the classifier response. These work best with a large number of individually weak trees of a shallow depth, containing only a few splits each. As the individual trees are not grown to full size, the need for pruning also disappears.

The method used in this work is AdaBoost [36]. It sequentially trains the trees making up the forest. For each tree, the events in the training sample are weighted depending on the results of the previous trees, giving a higher weight to events previously misclassified. This way, later trees preferentially learn to classify the events where trees earlier in the procedure failed to give an accurate result.

When beginning with the first tree, all events are weighted equally. Until a specified number M of trees is reached, the following procedure is repeated:

1. Train a decision tree with the current set of weighted events.

2. Determine the fraction *err* of events which are misclassified by the newly trained tree and calculate the factor

$$\alpha_i = \left(\frac{1 - err}{err} \right) \tag{4.1}$$

   associated with this tree.

3. The weights of all events misclassified by the newly trained tree are multiplied by α^β, with the learning rate β.

4. The weights are normalized by rescaling the sum of all weights back to the original value.

The optimal value of the learning rate β depends on the training set and the number of trees to be trained. Events which misclassified repeatedly continuously grow in weight. Later trees therefore start to concentrate their training on these events, until they get classified correctly and their weight decreases again with every normalization step.

The output of a forest of M trees is then determined by combining the individual binary responses of the trees. For a set of variables x, each tree provides such a response $b_i(x)$. In the average over the trees, the factor α_i determined during the training is used to weight the trees. The influence on the output is higher for trees correctly classifying a larger fraction of the training events:

$$y_{\texttt{BDT}}(x) = \frac{1}{M} \times \sum_i^M \ln(\alpha_i) \times b_i(x). \tag{4.2}$$

Unlike individual decision trees, a `BDT` gives a response $y(x)$ in a range of values. The ideal point above which events are classified as signal has to be determined separately depending on the needs of the analysis. One way to determine such a value is to maximise a Figure-of-Merit (`FoM`) using the number of signal (N_{Sig}) and background (N_{Bkg}) events in the sample passing the requirement. Multiple definitions are available, the one used later in this thesis in Chapter 6 is:

$$\texttt{FoM} \;=\; \frac{N_{Sig}}{\sqrt{N_{Sig} + N_{Bkg}}}. \tag{4.3}$$

For a sample size following a Poisson distribution, the term $\sqrt{N_{Sig} + N_{Bkg}}$ corresponds to the uncertainty of N_{Sig}. Maximizing this `FoM` therefore means maximizing the precision in the yield of signal events.

4.2. Binned maximum likelihood fit

After all selection criteria are applied, the remaining sample often still contains a significant contribution of background events. When measuring branching fractions it is however important to determine the actual amount of signal events in the sample. If the shape of both the signal as well as the background components in some measurement variable is known well enough, these shapes can be fitted to the measured data to determine what fraction each of them contribute to the selected sample. Together with the overall measured yield, these fractions directly translate to the desired number of events for each contribution.

If the shapes of the contributions are not analytically known, they have to be derived from simulated data (MC). Such simulated events can be used to create histograms in the fitting variables. While the data sample always only has a limited amount of measured events and therefore an associated statistical uncertainty, with simulated data filling the histograms to be used as templates for the shape, the same is true for them. In regular binned maximum likelihood fits, this statistical uncertainty stemming from the number of simulated events in each histogram bin in the MC distribution is neglected. However, when fitting histograms with many bins and complex distributions, there can often be regions of the distribution with very low yield and correspondingly a low number of simulated events per bin, making this uncertainty relevant for the fit.

One solution to take the limited MC statistics into account during a binned maximum likelihood is the method developed by Barlow and Beeston in [37], which is also used in this analysis in Chapter 7. For the fit the data sample as well as the m MC shapes are binned in identical histograms with n bins each. The weights of the m predictions are determined by a fit to the data sample. For each bin i there are d_i measured data events, and f_i events from the simulated MC predictions. The f_i are a function of the fractions P_j that each MC template j has in the data sample. These P_j, together with their uncertainties, are ultimately the quantities of interest, as they together with the size of the signal sample give the yield for each of the contributions.

For each bin, the template j has a generated number of events a_{ij}, giving a total number of events generated for the contribution as

$$N_j \;=\; \sum_{i=1}^{n} a_{ij}. \tag{4.4}$$

Similarly the total number of measured data events is:

$$N_D \;=\; \sum_{i=1}^{n} d_i. \tag{4.5}$$

A normalization factor $\frac{N_D}{N_j}$ is introduced to absorb the difference in measured data events and number of events used to simulate the template. With this, the number of predicted

events in the bin i follows as:

$$f_i = \sum_{j=1}^{m} P_j \frac{N_D}{N_j} a_{ji} = \sum_{j=1}^{m} p_j a_{ji}. \tag{4.6}$$

$$\text{with } p_j = P_j \frac{N_D}{N_j} \tag{4.7}$$

The factors p_j, which decide the event numbers in each bin in Equation (4.6), are to be fitted via maximisation of a likelihood function[1]. If the number of MC events per bin were high enough for their uncertainty to be negligible, this would simply consist of the Poisson distributions for getting the measured number of events d_i, which carries its own uncertainty, from the predicted f_i:

$$\mathscr{L} = \prod_{i=1}^{n} \frac{f_i^{d_i}}{d_i!} \mathrm{e}^{-f_i} \tag{4.8}$$

$$\Rightarrow \ln \mathscr{L} = \sum_{i=1}^{n} d_i \ln f_i - f_i \tag{4.9}$$

However, the predicted number of events a_{ji} for each bin might not perfectly represent the underlying distribution due to the finite size of the simulation. Instead, they are derived from an unknown, ideal prediction A_{ji} for the case of unlimited MC sample size. In general, the a_{ji} generated in the simulation originate from the unknown A_{ji} via a binomial distribution. If they can be assumed small compared to the overall sample size, that is $A_{ji} \ll N_j$, a Poisson distribution can safely be assumed instead. Instead of a fixed value as in the pure binned likelihood fit, the bin contents now vary, and the fitter can adjust by varying the A_{ji}. The resulting bin content and likelihood function used are:

$$f_i = \sum_{j=1}^{m} p_j A_{ji} \tag{4.10}$$

$$\ln \mathscr{L} = \sum_{i=1}^{n} d_i \ln(f_i) - f_i + \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ji} \ln(A_{ji}) - A_{ji} \tag{4.11}$$

The fit now has two sets of variables it needs to optimise, the p_j and the much more numerous A_{ji}. However, only the p_j are ultimately of interest. The maximisation of the likelihood function uses the derivative set to zero for both sets of variables:

$$\frac{\partial \ln \mathscr{L}}{\partial p_j} = 0 \Rightarrow \sum_{i=1}^{n} (\frac{d_i}{f_i} - 1) A_{ji} = 0 \; \forall \; j \tag{4.12}$$

$$\frac{\partial \ln \mathscr{L}}{\partial A_{ji}} = 0 \Rightarrow \frac{d_i p_j}{f_i} - p_j + \frac{a_{ji}}{A_{ji}} - 1 = 0 \; \forall \; i, j \tag{4.13}$$

[1] Constant terms are omitted as they do not affect the maximisation.

Besides the n equations in Equation (4.12), the A_{ji} introduce an additional $n \times m$ equations in Equation (4.13). However, solving for A_{ji},

$$A_{ji} = \frac{a_{ij}}{1 + p_j \left(1 - \frac{d_i}{f_i}\right)}, \tag{4.14}$$

shows that they only depend on the f_i for given p_j. The problem simplifies therefore to n equations. The fitting procedure now consists of alternatively determining a new set of p_j from Equation (4.12), then using these to determine new A_{ji} in Equation (4.13). These alternate steps repeat until a maximum in the likelihood function is reached.

If the MC templates contain bin-dependent weights for corrections or due to merging of sources of different statistics, all the a_{ji} and A_{ji} acquire another factor of w_{ji} in front of them. These are the average of all the weights of the events of source j in the bin i, this treatment assures that while the shape correction of the weights takes effect the information of the number of generated events is conserved. In case a bin contains no events at all for the source with the highest overall contribution p_j, the corresponding A_{ji} can take higher values even though the generated a_{ji} is zero.

4.3. B meson pair reconstruction

At a B-Factory like Belle, B mesons are always produced in pairs. In this analysis, however, the signal process of interest is only the decay of one B meson (B_{sig}), whose decay products are mixed together in the detector with the second B meson (B_{tag}) produced. Several options exist for treatment of this B_{tag} with a different trade-off between efficiency and precision. The most common approaches are listed in the following together with their advantages and disadvantages.

4.3.1. Full reconstruction

The full reconstruction [38] approach attempts to reconstruct the exact decay chain of the B_{tag}. Before any selection of the signal side is started, neural networks [39] attempt to reconstruct B decays in a large number of decay channels.

Two variants of the full reconstruction exist, the hadronic and the semileptonic full reconstruction. They both start with the same initial steps and reconstruct a $B \to D^{(*)}X$ decay, for both charged and neutral B and $D^{(*)}$. The reconstruction proceeds in four stages with a specially trained neural network each, starting with the final state particles, followed by D mesons and J/ψ, D^* and in the final stage B candidates in $B \to D^{(*)}X$ together with a quality indicator for each candidate. The hadronic and semileptonic full reconstruction differ in their choice of X for the final step, in the hadronic case it is

another hadron, while the semileptonic one selects either an electron or muon. Additional neural networks can be used to identify background events not from a $B\overline{B}$ process.

The decay channel used in the reconstruction determines from which type of B meson the B_{tag} was, from which the flavour of the B_{sig} at that exact point in time directly follows. Additionally, as the original $\Upsilon(4S)$ state is well known from the initial beam conditions, the reconstructed B_{tag} momentum directly gives access to the momentum of the B_{sig}. The B_{sig} momentum can either serve as a cross-check if all decay products are visible and therefore give a second determination of this momentum, or it can be used to access the momentum of an invisible particle in the signal side decay, usually either neutrino or K_{L}^0. For the semileptonic full reconstruction, the original decay contains a neutrino and therefore the B momentum is only constrained, but not exactly determined.

As the B_{tag} decay is explicitly reconstructed, all particles used do not need to be considered for the B_{sig} decay, greatly reducing the combinatorial possibilities. If, after both sides are reconstructed, any unused particles remain in the event that do not convincingly come from background sources unrelated to the $\Upsilon(4S)$, at least one of the reconstructions must be wrong by not considering this particle. This results in a very high purity.

The main disadvantage of the full reconstruction is its low efficiency. The decay channels used only cover a part of the B decay width, and the reconstruction itself also has limited efficiency, especially if a high reliability in the candidates is desired. Even when accepting all candidates, the overall efficiency only reaches 0.18% in $B^0\overline{B}^0$ and 0.28% in B^+B^- events for the hadronic case. Adding the semileptonic reconstruction increases the efficiency at the cost of less precise determination of the B momentum.

4.3.2. Untagged analysis

Instead of attempting to reconstruct the B_{tag}, the untagged approach concentrates only on the signal decay. This allows all events to be used when building candidates and thereby greatly increases the available statistics. However, as all final state particles in the event have to be considered, the amount of combinations to consider is much larger than when using a full reconstruction before. As only one half of the $\Upsilon(4S)$ decay is reconstructed, the particles remaining unused do not help in judging the accuracy of the reconstruction, and the potential for mixing particles from both B mesons is larger.

In exchange for the highest efficiency in reconstructing the signal events, this approach comes with the lowest purity, making suppressing backgrounds a challenge. It is therefore mostly suited when either the sample size is small and efficiency is important, or when other effective methods to isolate signal events are available. Otherwise there is the possibility that the increase in efficiency is lost again during background discrimination.

Without reconstructing the B_{tag}, there is no direct way to access the missing neutrino momentum in semileptonic signal processes. However, the loose neutrino reconstruction [40] is a method to get an approximation of the neutrino even in an untagged analysis. This method to infer the neutrino momentum is used in this analysis.

Loose Neutrino Reconstruction

The detectors at the B-Factories almost fully enclose the interaction region, detecting outgoing particles. On the other hand, being built around e^+e^- colliders, the $\Upsilon(4S)$ momentum is also exactly known as the sum of the ingoing electron and positron. From conservation of momentum it immediately follow that the sum of incoming and outgoing momenta should be identical. Any difference between these can be attributed to particles that are not measured by the detector, like a neutrino.

In the loose neutrino reconstruction [40] this difference, called *missing momentum* (p_{miss}) 4-momentum with its invariant mass as the *missing mass* (m_{miss}), is determined:

$$p_{\text{miss}} = p_{\Upsilon(4S)} - \left(\sum_{i}^{\text{Particles}} E_i, \Sigma \vec{p}_i \right) \tag{4.15}$$

$$m^2_{\text{miss}} = |p_{miss}|^2. \tag{4.16}$$

However, the missing momentum is always the sum of all unmeasured particles in the event, which can not be further disentangled with the information available. A clear identification of the missing momentum with a neutrino is therefore only possible if a single neutrino is the only particle that escaped detection in an event. As neutrinos are nearly massless, the missing mass allows to test this assumption, as any additional missing particles shift it to higher values.

4.4. Continuum suppression

A common source of background events at a B-Factory are interactions of the type $e^+e^- \to q\overline{q}$, where q is any of the four lighter quarks u, d, s or c. They are collectively known as continuum background. These events differ in shape from those containing an intermediate $\Upsilon(4S)$ resonance, which allows specified variables to be used to identify them. In the centre-of-mass frame, both B mesons are almost at rest. Their decay products have no preferred direction and are distributed isotropically. In $q\overline{q}$ events however the two quarks have lower mass and are produced with significant momenta in the centre-of-mass frame, and produce two jets of particles travelling in the same general direction. A comparison of the general shapes is shown in Figure 4.2. In the following some variables used throughout this analysis are introduced.

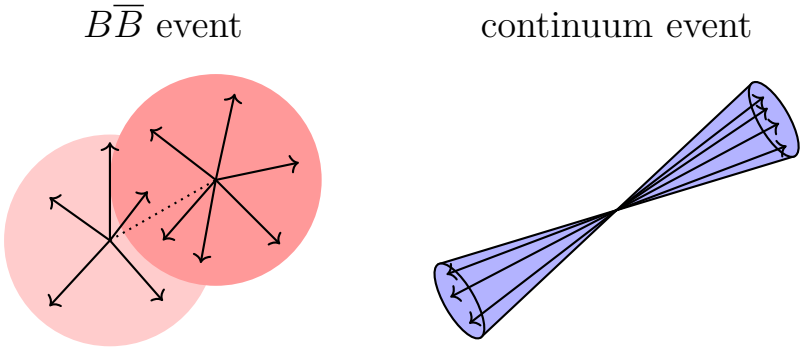


Figure 4.2.: Comparison of the typical shape of events in the centre-of-mass frame of reference, where either a $B\overline{B}$ pair or a pair of lighter quarks was produced.

4.4.1. Thrust angle

For a group of N particles, each of which has the momentum $\vec{p}_i$, the thrust in the direction of a unit vector $\vec{t}$ is defined as:

$$T := \frac{\sum_{i=1}^{N} |\vec{t} \cdot \vec{p}_i|}{\sum_{i=1}^{N} |\vec{p}_i|}. \tag{4.17}$$

The thrust axis is then the vector $\vec{t}$ which results in the maximal value for T. After all particles have been assigned to either the signal or the tag side, the thrust axis in the centre-of-mass frame can be determined for both individually. In events originating from a $B\overline{B}$ pair, the two thrust axes are uncorrelated, while for $q\overline{q}$ events they are strongly correlated to the direction of motion of the original two quarks, which travel in opposite directions.

The angle between the two axes, or rather its cosine $|\cos(\theta_{thrust})|$, can therefore be used to distinguish events. For $B\overline{B}$ it should be evenly distributed, while for $q\overline{q}$ events it peaks it high values [41].

4.4.2. Fox-Wolfram Moments

The general shape of an event with N particles can be quantified by the Fox-Wolfram moments [42]. These make no distinction between signal or tag side particles. They are defined as the sum over all particle pairs, together with the k-th Legendre polynomial P_k:

$$H_k := \sum_{i,j}^{N} |\vec{p}_i||\vec{p}_j| P_k(\cos\theta_{ij}). \tag{4.18}$$

As the lowest order Legendre Polynomial $P_0 = 1$ is constant, the lowest order Fox-Wolfram moment is used to normalize the other moments. In particular, the second-order moment with $P_2 = 1/2(3x^2 - 1)$ is often used in this way:

$$R_2 = \frac{H_2}{H_0}. \tag{4.19}$$

Its value increases as the shape of the event gets closer to two collimated groups of particles travelling in opposite directions. This feature makes R_2 a good measure on the shape typical for continuum events.

4.4.3. Modified Fox-Wolfram moments

The modified Fox-Wolfram moments, also known as the Kakuno Super Fox-Wolfram (KSFW) moments [43], have been developed to more precisely take into account relationships between the signal and tag side of an event, as well as the different types of particles. They fall into two types.

The first type relates signal to tag side particles. The index j selects one of three types of tag-side particles l for the sum, either charged (c) or neutral (n) particles, or the missing momentum (m) of the event, which is treated as a separate category. The index k always runs over all signal particles. There are two types of moments, which differ in whether they take into account the charges of the particles:

$$R^{so}_{j,i} = \sum_{k,l} Q_k Q_l |\vec{p}_l| P_i(\cos\theta_{kl}),\ i \in 1,3 \tag{4.20}$$

$$R^{so}_{j,i} = \sum_{k,l} |\vec{p}_l| P_i(\cos\theta_{kl}).\ i \in 0,2,4 \tag{4.21}$$

The terms for $i \in 1,3$ vanish except when j takes on charged particles.

The second type considers only particles on the tag side. Here, no distinction is made between charged and uncharged particles in the sum:

$$R^{oo}_{i} = \sum_{k,l} Q_k Q_l |\vec{p}_k||\vec{p}_l| P_i(\cos\theta_{kl}),\ i \in 1,3 \tag{4.22}$$

$$R^{oo}_{i} = \sum_{k,l} |\vec{p}_k||\vec{p}_l| P_i(\cos\theta_{kl}).\ i \in 0,2,4 \tag{4.23}$$

For $i \in 0,2,4$, these reproduce the regular Fox-Wolfram momenta, except for considering only tag-side particles.

Originally, these moments are combined into one discriminant with the Fisher coefficients $\alpha_{i,j}$, β_i and γ. Disregarding the vanishing $R^{so}_{j,i}$ terms, this gives a total of 17

coefficients, with the sum of the transverse momentum of all particles added:

$$KSFW = \sum_j \sum_{i=0}^{4} \alpha_{i,j} \frac{R^{so}_{j,i}}{E_{\text{beam}} - \Delta E} + \sum_{i=0}^{4} \beta_i \frac{R^{oo}_i}{E_{\text{beam}} - \Delta E} + \gamma \sum^{Particles} |p_t| \qquad (4.24)$$

However, in this analysis the moments are used individually instead of this combination. In the original application this $KSFW$ is further split into seven bins in the missing mass.

4.5. Monte Carlo Simulation

For the development of the analysis and comparison to the expected behaviour, as well as to extract the signal contribution according to Section 4.2, Monte Carlo samples with known features and classification of the contained particles are needed. The generation of such MC samples proceeds in several steps.

At first, the initial beam collision producing an $\Upsilon(4S)$ decaying into a $B\overline{B}$ pair and its subsequent decay is simulated by EVTGEN [44] and PYTHIA [45]. The Belle detector is modelled in GEANT3 [46]. In this model the decay products of the initial collision are propagated through the detector and their interaction with detector material, including energy loss via ionisation and creation of secondary particles via bremsstrahlung or pair production, is simulated. Final state radiation is afterwards added by PHOTOS [47, 48]. The particle trajectories and interactions are used to simulate the specific response for each sub-detector. Additional background is added using data taken without collision events. These backgrounds include cosmic rays and beam effects like Bhabha scattering and interaction with the beam and gas particles. The simulated detector responses are then processed as if they were real data and can be used for analyses.

4.5.1. MC samples used

For use in Belle analyses, several MC samples have been prepared. Each sample contains events of one type of interaction that can occur in real data. The samples are further divided into *streams*, where one stream corresponds to the number of events of the sample's type expected in the Belle dataset (see Section 3.4) taken at the $\Upsilon(4S)$ resonance. The MC simulates changes in detector configuration and conditions to follow the developments in real data over time. Multiple streams were generated for each sample to decrease the statistical uncertainty associated with their use.

The most common outcome after an $\Upsilon(4S)$ decays to a $B\overline{B}$ pair is a further decay of the both the b quarks via a $b \to c$. Ten streams of such decays, involving at least one charmed hadron, that is a hadron containing a c quark, have been generated. This type of events are collectively called $b \to c$. The signal decay is included in a special sample where

one part of the $B\overline{B}$ pair decays semileptonically as $B \to X_u \ell^+ \nu_\ell$ in a $b \to u$ transition. This sample, called $b \to u$ for the component it is defined by, has a size corresponding to twenty streams, however, the partial branching fraction of charmless semileptonic decays of B mesons is very small. The sample is therefore not split into twenty, but only two subsets in the analysis, each corresponding to ten times the total Belle luminosity. Additionally, there is a sample (continuum) containing $e^+e^- \to q\overline{q},\ q \in u, d, s, c$ processes not involving an $\Upsilon(4S)$ resonance. The size of this sample corresponds to six streams. This type of events is an important background to consider in the signal extraction. The sizes are summarized in Table 4.1.

Type	Size [Streams]
$b \to c$	10
continuum	6
$b \to u$	20

Table 4.1.: MC sample sizes used in this analysis. One streams corresponds to the expectation for that type at the total measured Belle luminosity.

Branching fractions used in the MC reflect the measured or calculated knowledge at the time of its generation. The used values have been updated to the most recent averages [4] for the semileptonic B decays to D, D^*, π and ω, while the $B \to \rho\ell\nu$ decay instead takes the result of the Belle measurement [49]. These decays together make up the most important sources of background.

The $b \to u$ sample contains an inclusive part to fill the discrepancy between the sum of all measured exclusive modes and the inclusive measurement of the $B \to X_u \ell^+ \nu_\ell$ branching fraction. Its size has been adjusted so the total size of the sample corresponds to the most recent inclusive value [24]. This analysis assumes a branching fraction of the $\Upsilon(4S)$ to B^+B^- and $B^0\overline{B}^0$ of 51.3% and 48.7%. Besides the branching fractions, the form factors involved in the simulation of the semileptonic decays to D, D^* and D^{**} (D_1, D_0, D_0', D_2) have also been updated [50] to the newest values [24].

CHAPTER 5

Decay chain reconstruction

After an event has been recorded by the detector, basic steps of data processing and track finding are common to all analyses. Besides these, every analysis needs to implement their own scheme to reconstruct the particles in the desired decay modes and to distinguish them from background.

In this analysis, the decay $B^+ \to \eta \ell^+ \nu_\ell$ is reconstructed in two channels:

- $\eta \to \gamma\gamma$,
- $\eta \to \pi^+\pi^-\pi^0$.

The branching fractions of these decays are listed in Table 2.3. The second most frequent decay channel, $\eta \to \pi^0\pi^0\pi^0$, is not considered as the final state of six photons would be too difficult to distinguish from combinatorial background in an untagged analysis. For the $B^+ \to \eta' \ell^+ \nu_\ell$ decay, two of the channels from Table 2.4 are used:

- $\eta' \to \pi^+\pi^-\eta$,
- $\eta' \to \rho^0\gamma$,

Here the third channel, $\eta' \to \pi^0\pi^0\eta$, is not considered for just the same reason. The η in the $\eta' \to \pi^+\pi^-\eta$ is reconstructed in the two channels $\eta \to \gamma\gamma$ and $\eta \to \pi^+\pi^-\pi^0$, and the following η' reconstruction is treated as two independent channels, $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ and $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$ within this analysis. Throughout the reconstruction of a decay chain, each original track or cluster can only be used once. In this thesis, the notation $\eta^{(\prime)}$ stands for the neutral hadron originating from the B meson in the $B^\pm \to \eta^{(\prime)} \ell^\pm \nu_\ell$ decay.

In this chapter, candidates for these decays are constructed from the measured particles in each event. As these are rare decays, it is important to get a large sample size of events to keep statistical uncertainties low. In this first step only loose selection criteria are applied, prioritizing a high acceptance over background removal. Increasing the signal fraction in the selected sample by removing events considered background is done afterwards in Chapter 6. All selection criteria are chosen so as to avoid introducing a bias in the q^2 distribution wherever possible, especially to avoid restricting the selected candidates to a specific part of the range.

The chapter starts with identifying the final state particles used in Sections 5.1 and 5.2, that is photons, electrons, muons and pions. This is followed by the reconstruction of $\eta^{(\prime)}$ candidates in Sections 5.3 and 5.4. In Section 5.5, the candidates reconstructed in the previous sections are combined in a first step of background rejection. Afterwards, the kinematics of the final missing particle, the neutrino, is estimated in Section 5.6. After all signal particles are identified, they are combined to a $B^{\pm}$ in Section 5.7 and the variables used later in the signal extraction are introduced. In Section 5.8, several methods to determine the transferred momentum q^2 are compared.

From here onward, all unmarked quantities are measured in the rest frame of reference of the detector, while those marked with "*" are in the centre-of-mass frame, which corresponds to the frame of reference of the $\Upsilon(4S)$. Any other frames of reference used will be mentioned explicitly. Some of the distributions shown are very similar for the different reconstruction channels. In these cases, usually only the channel $\eta \to \gamma\gamma$ is shown with the main text, while the other channels are provided in Appendix A for reference.

5.1. Photon identification

Photon candidates to be used in the reconstruction are all well measured energy clusters inside the ECL without a charged track pointing to the same direction, which would indicate a charged particle showering instead of a photon produced in the initial reaction. All photons must be above the ECL energy threshold for reliable measurement. This threshold is different depending on whether the photon is measured in the barrel region or the endcaps due to direction-dependent beam backgrounds:

$$17^\circ < \theta < 32^\circ : \; E_\gamma > 100\,\mathrm{MeV}, \tag{5.1}$$

$$32^\circ < \theta < 130^\circ : \; E_\gamma > 50\,\mathrm{MeV}, \tag{5.2}$$

$$130^\circ < \theta < 150^\circ : \; E_\gamma > 150\,\mathrm{MeV}. \tag{5.3}$$

From the accepted photons, π^0 candidates are built automatically using a mass-range around the nominal mass by the Belle detector software. The π^0 candidates are constrained to their nominal mass of $135\,\mathrm{MeV}/c^2$ [4] before being used in the reconstruction.

5.2. Charged final state particles

The analysis contains a variety of charged particles, which need to be identified from the tracks measured by the detector. Assigning a particle type uses the PID explained in Section 3.3. Before assigning a specific type however, the following constraints common to all tracks are applied:

When a charged particle has a transverse momentum (p_T) below $275\,\mathrm{MeV}/c$, it can move in a spiral inside the CDC instead of passing through and leaving the CDC. When such a track passes through the interaction point (IP) again, it can be reconstructed as two or even more separate tracks. These duplicates have to be removed. For all pairs of tracks in the momentum range, their momentum difference

$$\delta p \;=\; |\vec{p}_1 - \vec{p}_2| \tag{5.4}$$

is determined, together with their relative angle. Pairs with $\delta p < 100\,\mathrm{MeV}/c$ and an angle below 15°, if they are both reconstructed with the same charge, or 165° if oppositely charged, are considered duplicates. Which of the pair is kept is decided using the distances parallel (dz) and transversal (dr) to the beam axis of their point of closest approach to the IP. These together form

$$\chi^2 \;=\; |5 \cdot dr|^2 + |dz|^2\,, \tag{5.5}$$

giving higher weight to the parallel distance, and from the pair of tracks only the one with the lower value is kept.

Charged particles far away from the IP are also removed. While these could be real particles, they are very likely to originate not from the $\Upsilon(4S)$ decay, but are background from other processes. All tracks to be used must therefore lie within a certain distance:

$$|dr| \leq 0.5\,\mathrm{cm}, \tag{5.6}$$

$$|dz| \leq 2\,\mathrm{cm}. \tag{5.7}$$

As all particles in an event originate from the electrically neutral $\Upsilon(4S)$, the sum of charges of all particles should be neutral as well. However, due to either misreconstructed tracks, tracks not reconstructed at all or additional background tracks from other sources, this is not always the case. To remove events clearly missing particles while also keeping as many as possible for the reconstruction, a total charge of all accepted tracks of up to ± 2 is allowed, all events outside this range are discarded.

5.2.1. Leptons

Common to all signal decays is the presence of exactly one charged lepton from the original $B^+ \to \eta^{(\prime)}\ell^+\nu_\ell$ decay. This charged lepton can be either an electron or a muon[1]. As

[1] From now on, lepton always means either an electron or muon for simplicity.

it originates directly from the B^+ decay to particles which together are comparatively light, it often inherits a large part of the original B^+ momentum. Therefore, a general requirement of at least $1.3\,\mathrm{GeV}/c$ in the CM frame for both types is applied to remove leptons from secondary decays, which have lower energy.

Lepton candidates can only be reliably identified if they have enough momentum in the detector frame of reference to reach the outermost sub-detector important for their respective `PID`. This is the `ECL` for electrons and the `KLM` for muons, and the resulting requirements are $p_e > 0.4\,\mathrm{GeV}/c$ and $p_\mu > 0.8\,\mathrm{GeV}/c$. As muons are measured further out in the detector, they need to have a higher momentum than electrons. Additionally, they are required to lie within the detector angular acceptance with $17° < \theta_e < 150°$ and $25° < \theta_\mu < 145°$. Out of these candidates, electrons and muons are selected using the `PID` introduced in Sections 3.3.2 and 3.3.3. The chosen requirements are $\mathscr{P}_e > 0.5$ and $\mathscr{P}_\mu > 0.9$.

Due to conservation laws, leptons are always either produced in pairs or, more often in the processes relevant as backgrounds, together with a neutrino. However, a second neutrino in the event would make the loose neutrino reconstruction approach, as detailed in Section 4.3.2, infeasible. The number of neutrinos in an event can, however, not be directly measured. Instead, only events with exactly one lepton satisfying the above criteria are accepted.

5.2.2. Hadrons

Charged pions are the most common source of tracks in the Belle experiment, and appear in all signal decays except $\eta \to \gamma\gamma$. The requirement for a charged track to be accepted as a pion is $\mathscr{P}_{\pi/K} > 0.6$ according to the `PID` from Section 3.3.1. The track already identified as a lepton in Section 5.2.1 is not considered. Due to the requirement of only one lepton candidate per event, all other tracks necessarily failed a comparison against a lepton hypothesis.

For some parts of the analysis, kaons are needed as well. These are taken as all remaining tracks which satisfy $\mathscr{P}_{K/\pi} > 0.6$. As $\mathscr{P}_{K/\pi} = 1 - \mathscr{P}_{\pi/K}$ by definition, no tracks are assigned to both the kaon and pion candidates.

5.3. Reconstructing the η meson

Reconstruction of η mesons is done separately in two decay channels, $\eta \to \gamma\gamma$ and $\eta \to \pi^+\pi^-\pi^0$. Their branching fractions are listed in Table 2.3. Reconstructed η candidates are required to have an invariant mass close to the nominal mass of $548\,\mathrm{MeV}/c^2$ [4]. As the precision of the reconstructed mass is dependent on the reconstruction channel,

the range of accepted masses is determined individually for each channel by fitting a Gaussian to the reconstructed mass distribution. As a compromise between accepting most of the signal events while removing unnecessary background, in general a window of 3σ around the fitted peak position is accepted.

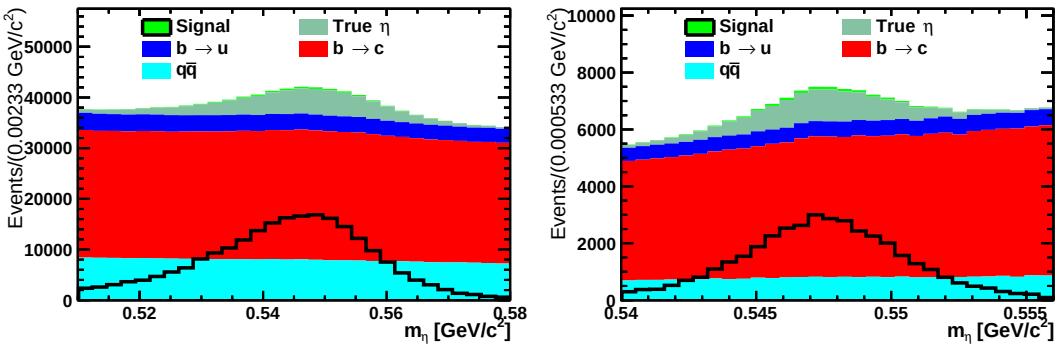
5.3.1. $\eta \to \gamma\gamma$

Candidates in the decay channel $\eta \to \gamma\gamma$ are reconstructed as combinations of two photons in the event. However, the number of photons in a typical Belle event is large, which leads to a lot of combinations close to the mass of an η without an actual η decay involved. Most of these background photons originate from decays of π^0, which is a common particle to appear in many decays. Therefore, a veto [51] is used against photons likely to originate from a π^0 decay. For all combinations of photons in the event, the invariant mass is determined and if a combination lies in the range $110\,\mathrm{MeV}/c^2 < m_{\gamma\gamma} < 160\,\mathrm{MeV}/c^2$, both photons are not considered for the $\eta \to \gamma\gamma$ reconstruction.

From all remaining photons, η candidates are built by combining the two 4-momenta. Candidates with an invariant mass within $510\,\mathrm{MeV}/c^2 < m_{\gamma\gamma} < 580\,\mathrm{MeV}/c^2$ are considered for further analysis. The reconstructed mass distribution on simulated data is shown in Figure 5.1a. The majority of the background events consist of random combinations lying inside the acceptance window, these are equally distributed throughout the range. There are also events containing correctly reconstructed η mesons, which do not originate from a $B^+ \to \eta\ell^+\nu_\ell$ process and share the same mass distribution as those that do. These can not be distinguished at this step.

5.3.2. $\eta \to \pi^+\pi^-\pi^0$

η candidates in this channel are combinations of two pions of opposite charge with one π^0. All combinations are required to have a mass inside $540\,\mathrm{MeV}/c^2 < m_{\pi^+\pi^-\pi^0} < 555\,\mathrm{MeV}/c^2$ to be accepted as candidates. Additionally, the three pion system needs to pass a vertex fit with a result of $\chi^2/\mathit{d.o.f.} < 3$ to assure they are compatible with originating from a common decay vertex. The mass distribution of accepted candidates is shown in Figure 5.1b. Similar to the $\eta \to \gamma\gamma$ channel, the background events separate into η mesons from other sources, and combinatoric background showing no peak at the nominal η mass.



(a) $\eta \to \gamma\gamma$ (b) $\eta \to \pi^+\pi^-\pi^0$

Figure 5.1.: Reconstructed η-masses on MC, after all requirements within this chapter have been applied. Background containing a correctly identified η not from a signal decay is listed separately. The background types correspond to the different MC sets. The signal contribution is also superimposed with arbitrary scale.

5.4. Reconstructing the η' meson

In addition to the two $B^+ \to \eta\ell^+\nu_\ell$ channels, the decay $B^+ \to \eta'\ell^+\nu_\ell$ is also reconstructed. The main approach is to use the previously reconstructed η-candidates to reconstruct the process via $\eta' \to \pi^+\pi^-\eta$. Afterwards, the reconstruction in $\eta' \to \rho^0\gamma$ is also attempted. As in Section 5.3, candidates are required to lie within 3σ of the reconstructed mass peak, with the nominal mass being $958\,\mathrm{MeV}/c^2$ [4].

In the channel $\eta' \to \pi^+\pi^-\eta$, the resolution in mass depends on the resolution of the reconstructed η, which in turn depends on the decay channel. The channel is therefore split into $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ and $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$. While the general procedure is identical for both, the mass requirements are determined separately.

5.4.1. $\eta' \to \pi^+\pi^-\eta$

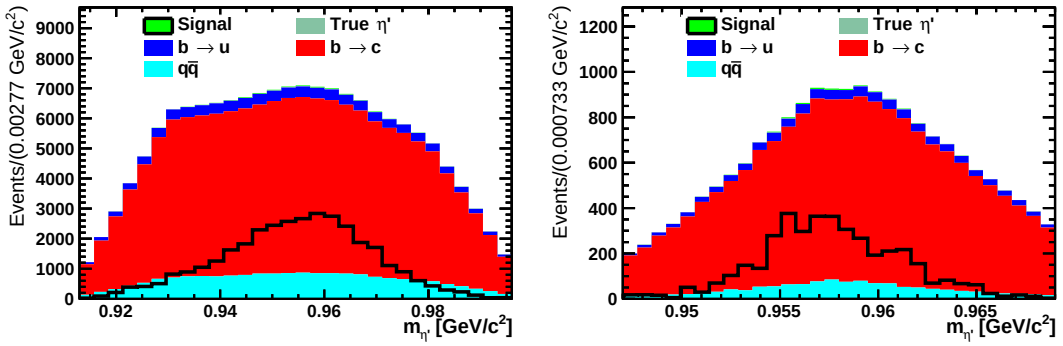
η' candidates are reconstructed by matching η candidates with two additional, oppositely charged pions. Besides the requirement on the reconstructed mass, wrong combinations are further removed by a requirement on the mass difference between the η' candidate and the associated η, determined similarly to the mass requirements. This mass difference is defined as

$$\Delta m = m_{\eta'} - m_\eta. \tag{5.8}$$

The resulting mass requirements are for the channel $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ $913\,\mathrm{MeV}/c^2 < m_{\eta'} < 996\,\mathrm{MeV}/c^2$ and $400\,\mathrm{MeV}/c^2 < \Delta m < 420\,\mathrm{MeV}/c^2$, while in the channel $\eta' \to$

$\pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$ they are $947\,\mathrm{MeV}/c^2 < m_{\eta'} < 969\,\mathrm{MeV}/c^2$ and $404\,\mathrm{MeV}/c^2 < \Delta m < 417\,\mathrm{MeV}/c^2$. The final criterion is another vertex fit including the newly added pions, again required to succeed with $\chi^2/_{d.o.f.} < 3$.

The resulting mass distribution is shown in Figure 5.2 and the Δm distributions in Figure 5.3. Due to the combination of multiple requirements on reconstructed masses, the background develops a peaking structure in the same region as the signal. However, a method to reduce background candidates sufficiently without such an effect is not available. Unlike in the $B^+ \to \eta\ell^+\nu_\ell$ channels, the contribution from true η' with a different origin is small.



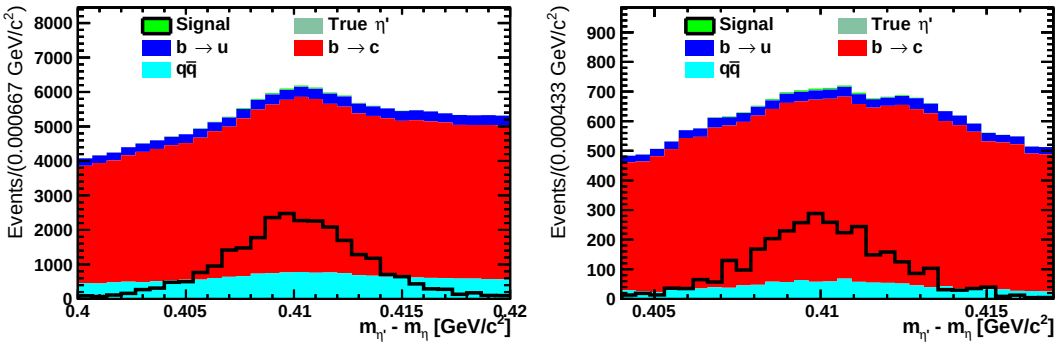
(a) $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ (b) $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$

Figure 5.2.: Reconstructed η'-masses on MC, after all requirements within this chapter have been applied. Background containing a correctly identified η not from a signal decay is listed separately. At this stage the signal contribution is barely visible due to its small size. The signal contribution is also superimposed with arbitrary scale.

5.4.2. $\eta' \to \rho^0\gamma$

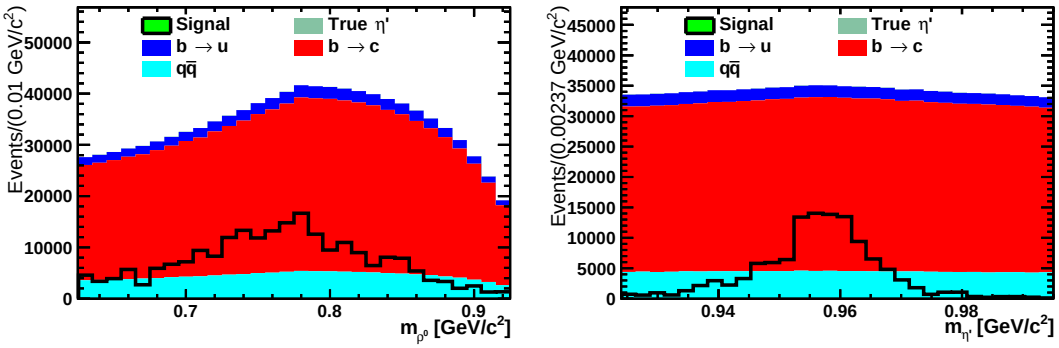
Independent from the η candidates, the channel $\eta' \to \rho^0\gamma$ is also reconstructed. Here at first a ρ^0 candidate is built by combining two oppositely charged $\pi^\pm$, into which the ρ^0 almost exclusively decays [4]. Due to the very high width of the ρ^0, instead of the 3σ mass window used for the other reconstructions, here the mass has to lie inside the world average measured width [4] of $[625\,\mathrm{MeV}/c^2, 925\,\mathrm{MeV}/c^2]$.

These ρ^0 candidates are then combined with all photons not affected by the π^0-veto in Section 5.3.1, and the resulting η' candidates are required to lie within $924\,\mathrm{MeV}/c^2 < m_{\eta'} < 995\,\mathrm{MeV}/c^2$. As in the other channels, a vertex fit requirement of $\chi^2/_{d.o.f.} < 3$ is also applied. The reconstructed masses are shown in Figure 5.4. The ρ has a very large



(a) $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ (b) $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$

Figure 5.3.: Reconstructed $\Delta m = m_{\eta'} - m_\eta$ on MC, after all requirements within this chapter have been applied. Background containing a correctly identified η' not from a signal decay is listed separately. The signal contribution is also superimposed with arbitrary scale.



(a) $\rho^0 \to \pi^+\pi^-$ (b) $\eta' \to \rho^0\gamma$

Figure 5.4.: Reconstructed ρ^0 and η' masses in the channel $\eta' \to \rho^0\gamma$ on MC, after all requirements within this chapter have been applied. Background containing a correctly identified η' not from a signal decay is listed separately. The signal contribution is also superimposed with arbitrary scale.

width, and the reconstructed η' mass also has larger width than the other η' channels. This channel has the largest background contribution.

5.5. Combination requirements

Each event now contains exactly one lepton candidate and a variety of $\eta^{(\prime)}$ candidates. In case the lepton is an electron, it is possible that it emitted a bremsstrahlung photon before discovery, reducing the electron energy. Recovery is attempted by searching for the photon with the smallest angle to the electron among those with $E_\gamma < 1\,\text{GeV}$. If the angle of this photon to the electron is below 5°, and it is not previously used in the $\eta^{(\prime)}$ reconstruction, it is added to the 4-momentum of the electron.

At this stage, the vast majority of candidates are background, either wrong combinations of final state particles that just by chance happen to have the same mass as an $\eta^{(\prime)}$ but are otherwise unrelated, or correctly reconstructed $\eta^{(\prime)}$ coming from other decay processes. Adding the lepton allows to already remove some additional candidates.

The $\eta^{(\prime)}$ and the lepton together can be considered as a pseudoparticle, which, for correctly reconstructed particles, only differs from the B^+ in the missing neutrino. Its momentum is:

$$p_{\ell\eta^{(\prime)}} = p_\ell + p_{\eta^{(\prime)}}. \tag{5.9}$$

Instead of a direct reconstruction, the B^+ energy and momentum $E_B^\star$, $|\vec{p}_B^\star|$ can be approximated using the centre-of-mass energy $\sqrt{s}$ by assuming it at rest in the CM frame, as its momentum is low:

$$E_B^\star = \frac{\sqrt{s}}{2}, \tag{5.10}$$

$$|\vec{p}_B^\star| = \sqrt{\frac{{E_B^\star}^2}{c^2} - \frac{m_B^2}{c^4}} = \sqrt{\frac{s}{4c^2} - \frac{m_B^2}{c^4}}. \tag{5.11}$$

With the missing neutrino being (nearly) massless, the angle between the $\eta^{(\prime)} - \ell^+$ combination and the B^+ can be determined [52]:

$$(P_B^\star - P_{\ell\eta^{(\prime)}}^\star)^2 = m_\nu^2 \approx 0, \tag{5.12}$$

$$\Rightarrow 0 = m_B^2 + m_{\ell\eta^{(\prime)}}^2 - 2E_B^\star E_{\ell\eta^{(\prime)}}^\star + 2\vec{p}_B^\star \cdot \vec{p}_{\ell\eta^{(\prime)}}^\star, \tag{5.13}$$

$$\Rightarrow \cos(\theta_{B,\ell\eta^{(\prime)}}^\star) = \frac{2E_B^\star E_{\ell\eta^{(\prime)}}^\star - m_B^2 - m_{\ell\eta^{(\prime)}}^2}{2|\vec{p}_B^\star||\vec{p}_{\ell\eta^{(\prime)}}^\star|}. \tag{5.14}$$

For correctly reconstructed events, the $\theta_{B,\ell\eta^{(\prime)}}^\star$ is a physical angle and therefore, barring detector inaccuracies, it follows that $|\cos(\theta_{B,\ell\eta^{(\prime)}}^\star)| < 1$. For background events the assumption in Equation (5.12) is violated and, correspondingly, Equation (5.14) could take

any value. Both to remove this region almost entirely populated by background, and to allow the determination of the momentum transfer q^2 in Section 5.8 using $\theta^{\star}_{B,\ell\eta^{(\prime)}}$, all events with $|\cos(\theta^{\star}_{B,\ell\eta^{(\prime)}})| > 1$ are discarded. The distributions for signal and background can be seen in Figures 5.5 and A.1, showing the large amount of background events that can be removed.

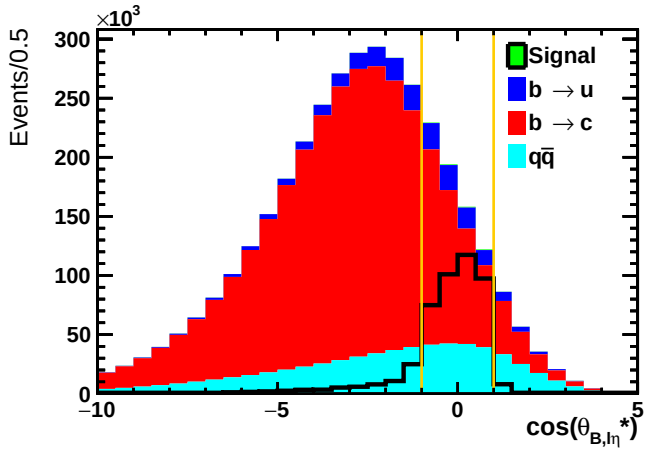


Figure 5.5.: Reconstructed $\cos(\theta^{\star}_{B,\ell\eta^{(\prime)}})$ in the $\eta \to \gamma\gamma$ channel on MC, with all other cuts throughout this section applied. Only events inside the yellow lines are accepted. The signal contribution is also superimposed with arbitrary scale.

An event can still contain multiple candidates in the same reconstruction channel. After a final combined vertex fit of the entire decay chain including the lepton, for each channel only the candidate with the lowest resulting $\chi^2/_{d.o.f.}$ is kept. In case no candidate succeeds the fit, the event is discarded.

5.6. Neutrino reconstruction

Unlike the other final state particles in the $B^+ \to \eta^{(\prime)}\ell^+\nu_\ell$ decay, the neutrino leaves the detector without further interaction and can not be measured directly. However, its momentum can be inferred with the method presented in Section 4.3.2. This method equates the missing momentum, which is the difference between initial $\Upsilon(4S)$ momentum and sum of all outgoing momenta, with the neutrino momentum.

To determine p_{miss} via Equation (4.15), the sum of all particles in the event is needed. In this summation, the requirements on the distance of tracks to the *IP* is weakened from those in Equation (5.6) to $dr < 1.5\,\text{cm}$ and $dz < 10\,\text{cm}$. The decay chain of the tag side is not explicitly reconstructed. However, the particle type of the final state particles needs to be decided to correctly assign their momentum. All remaining tracks that fulfil

the $K^{\pm}$ requirements in Section 5.2.2 are treated as such, otherwise they are considered as $\pi^{\pm}$. The remaining energy in the ECL not associated to one of these charged particles is added as photons.

If the sum in Equation (4.15) correctly adds to the neutrino in the event, the result in Equation (4.16) should be $m^2_{\text{miss}} = 0$. Any further unmeasured particles however lead to a higher missing mass. Even without additional particles the detector resolution can cause the missing mass to shift. To remove clearly wrong candidates early, only events with $|m^2_{\text{miss}}| < 7\,\text{GeV}^2$ are accepted. The distribution in the accepted range is shown in Figures 5.6 and A.2. Due to missing particles from the B_{tag} decay, there is a significant contribution of higher values both for signal and background.

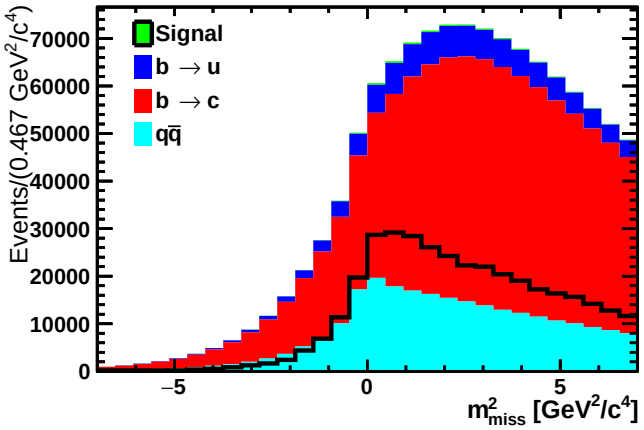


Figure 5.6.: Accepted m^2_{miss} in the $\eta \to \gamma\gamma$ channel on MC, with all selections throughout this section applied. The signal contribution is also superimposed with arbitrary scale.

The resolution of the Belle detector is significantly better for momentum than for energy. After the summation of all charged and neutral particles, the neutrino is therefore constrained to be massless by adjusting the energy of p_{miss} from Equation (4.15):

$$p_\nu = (|\Sigma\vec{p}_i|\,, \Sigma\vec{p}_i)\,. \tag{5.15}$$

This improves the following reconstruction of the B meson.

5.7. B meson reconstruction

With candidates for the $\eta^{(\prime)}$, the lepton and the neutrino reconstructed, a B^+ candidate can now be built. For this, no requirement on the invariant mass is made, as it would

restrict the fit range described in the following. A very small number of events contain candidates with a negative invariant mass, these are discarded.

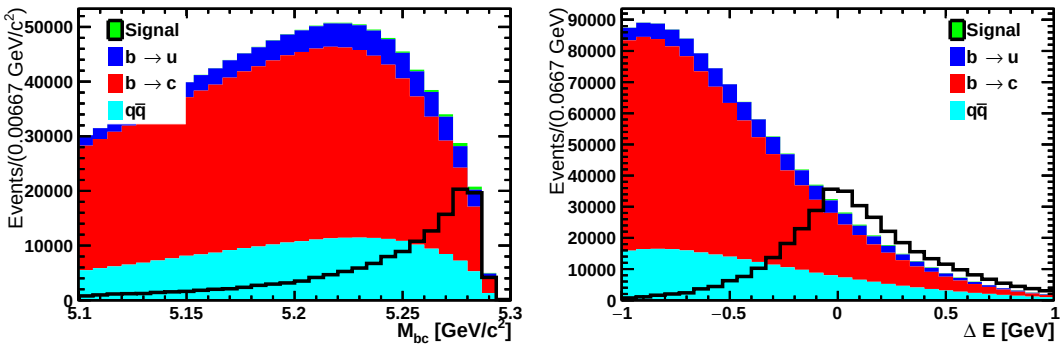
Continuum background among the remaining candidates is removed using the Fox-Wolfram moments explained in Section 4.4.2. Here, only events with the ratio R_2 (Equation (4.19)) below 0.4 are kept. This first reduction allows the `BDT` in Chapter 6 to concentrate on the remaining background events, which are more difficult to separate.

The signal extraction in Chapter 7 uses a two-dimensional fit in two variables. These are the beam-constrained mass (M_{bc}) and the energy difference (ΔE). For M_{bc} the mass of the reconstructed B^+ is determined using the sum of the reconstructed momenta, but substituting the reconstructed energy with the beam energy. ΔE, on the other hand, is the difference between the reconstructed energy of the B^+ candidate and the beam energy. The two variables are defined as:

$$M_{\mathrm{bc}} = \sqrt{E^{\star 2}_{beam}/c^4 - \vec{p}^{\,\star 2}_B/c^2}, \tag{5.16}$$

$$\Delta E = E^{\star}_B - E^{\star}_{beam}, \tag{5.17}$$

where $E^{\star}_{beam}$ is the energy of one of the beams in the CM system, equivalent to half the centre-of-mass energy. In a perfectly reconstructed event, M_{bc} would come out at the nominal mass of a B^+, and the energy difference would vanish. The fit to determine the signal fraction only uses the limited range of $5.1\,\mathrm{GeV}/c^2 < M_{\mathrm{bc}} < 5.3\,\mathrm{GeV}/c^2$ and $-1\,\mathrm{GeV} < \Delta E < 1\,\mathrm{GeV}$. All candidates outside are discarded at this step to allow the `BDT` training in the following section to concentrate on events useful for the signal determination later. The distributions are shown in Figures 5.7, A.3 and A.4, showing how signal events peak at the expected value.



(a) M_{bc} (b) ΔE

Figure 5.7.: The M_{bc} and ΔE distribution in the channel $\eta \to \gamma\gamma$ in the range to be used later for the fit on MC, with all selections throughout this section applied. The signal contribution is also superimposed with arbitrary scale.

5.8. Momentum transfer q^2

One important quantity in describing the kinematics of a semileptonic decay is the momentum transfer q^2. In the decay $b \to uW^-$, the W^- further decays into the charged lepton and neutrino, which therefore carries information about the relative distribution of the energy of the original particle. The general definition is:

$$q^2 = (p_\ell + p_{\nu_\ell})^2. \tag{5.18}$$

The resolution of the only indirectly reconstructed neutrino is the limiting factor in the precision in Equation (5.18). This section introduces two different methods available to increase the precision.

5.8.1. Neutrino correction

One such method [53] uses the inferred neutrino as the starting point. This neutrino is then adjusted using Equations (5.16) and (5.17) and their expected values for an ideal reconstruction.

Firstly, the energy of the neutrino is rescaled by a factor α chosen so that the energy difference ΔE vanishes:

$$\alpha E_\nu + E_\ell + E_{\eta^{(\prime)}} - E_{beam} = 0. \tag{5.19}$$

After the energy, the direction of the neutrino is corrected as well. As the neutrino is massless, the same scaling factor α applies to the momentum as well. This corrected neutrino is used as part of the B^+ momentum in Equation (5.16):

$$M_{\mathrm{bc}} = \sqrt{E_{beam}^2/c^4 - (\alpha\vec{p}_\nu + \vec{p}_\ell + \vec{p}_{\eta^{(\prime)}})^2/c^2}. \tag{5.20}$$

The direction of the neutrino vector is set to the direction closest to the original reconstructed one that fulfils $M_{\mathrm{bc}} = m_{B^+}$ for the nominal mass of the B^+. This corrected neutrino is used in Equation (5.18) to determine q^2_α.

5.8.2. Cone averaging

Another approach [54] is to estimate the momentum of the B^+ instead of the neutrino to calculate the momentum transfer with the reconstructed $\eta^{(\prime)}$ as:

$$q^2 = (p_{\ell^+} + p_{\nu_\ell})^2 = (p_{B^+} - p_{\eta^{(\prime)}})^2 \tag{5.21}$$

While the energy of the $B^\pm$ can be determined from the $\Upsilon(4S)$ via momentum conservation, the direction of its momentum can not. However, the angle from Equation (5.14) allows to constrain the momentum to a cone around the combined $\eta^{(\prime)} + \ell^+$ direction.

Instead of further constraining the direction, q^2 is determined for four equally spaced directions along this cone, with the first determined randomly:

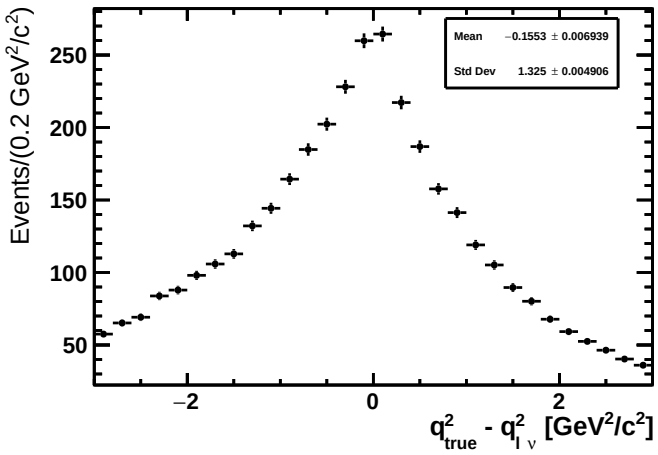
$$\phi_1 = \phi, \tag{5.22}$$
$$\phi_2 = \phi + \pi/2, \tag{5.23}$$
$$\phi_3 = \phi + \pi, \tag{5.24}$$
$$\phi_4 = \phi + 3\pi/2. \tag{5.25}$$

The final value is determined by averaging the four values of q^2, weighted with the angle $\theta_{B\phi_i}$ to the beam axis of each B^+ direction.

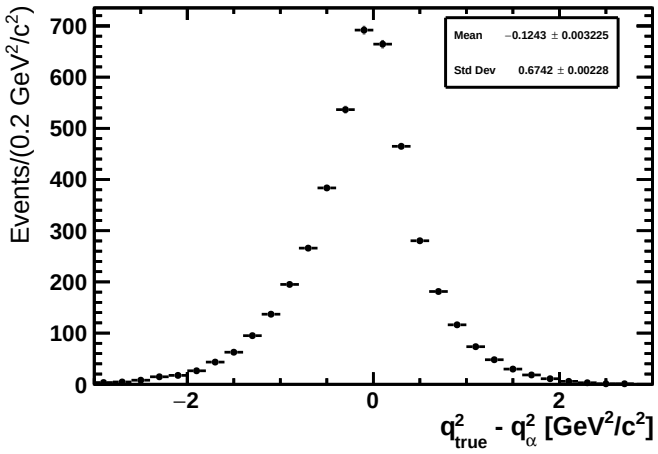
$$q^2_{cone} = \frac{\sin^2\theta_{B\phi_1} q^2_{\phi_1} + \sin^2\theta_{B\phi_2} q^2_{\phi_2} + \sin^2\theta_{B\phi_3} q^2_{\phi_3} + \sin^2\theta_{B\phi_4} q^2_{\phi_4}}{\sin^2\theta_{B\phi_1} + \sin^2\theta_{B\phi_2} + \sin^2\theta_{B\phi_3} + \sin^2\theta_{B\phi_4}} \tag{5.26}$$

5.8.3. Comparison

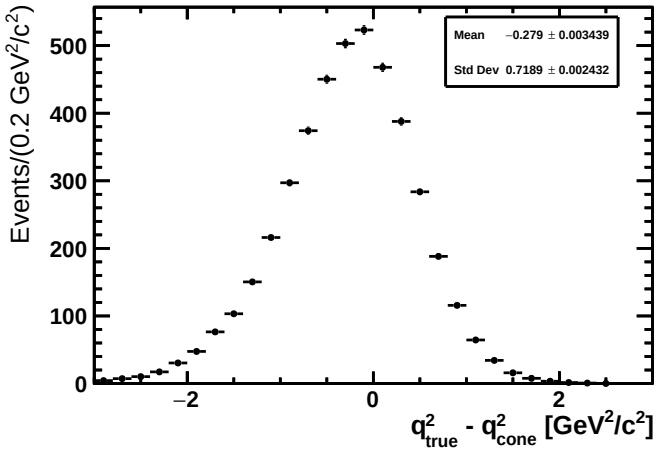
The three methods to determine q^2 are compared in Figure 5.8 by showing the difference to the generated value. The comparison uses correctly reconstructed events in the channel $\eta \to \gamma\gamma$, because they have the highest number of events. Both methods clearly improve the accuracy compared to using the inferred neutrino from Equation (5.15) directly. The level of precision achieved between both other methods is similar, however the neutrino correction shown in Figure 5.8b seems to be slightly better in accuracy than the averaging method shown in Figure 5.8c when comparing the RMS.



(a) Inferred ν directly



(b) q^2_α



(c) q^2_{cone}

Figure 5.8.: Comparison of the accuracy of the three methods to determine q^2, using correctly reconstructed events in the channel $\eta \to \gamma\gamma$.

CHAPTER 6

Background reduction with boosted decision trees

Signal candidates have been reconstructed in Chapter 5. However, as the $B^{\pm} \to \eta^{(\prime)}\ell^{\pm}\nu_{\ell}$ decays only make up a small fraction of the total amount of events, the vast majority of these candidates are actually composed of various backgrounds. While some background rejection was already performed, further selection is necessary to increase the signal fraction. This chapter describes the use of Boosted Decision Trees (described in Section 4.1) to increase the fraction of events actually containing a signal decay in the sample. Everything described in this chapter is done independently for each of the signal decay channels.

This chapter begins with a detailed description of the setup used by the BDTs and their training variables in Section 6.1. Afterwards, the performance of the resulting BDTs is evaluated in Section 6.2.

6.1. Configuration

For the training of the BDT the background is grouped into two categories, events originating from $\Upsilon(4S)$ decays and those from continuum processes. The former corresponds to the combined $b \to c$ and $b \to u$ components of the MC. These two categories differ in the general kinematic distributions of the events, as discussed in Section 4.4. In each reconstruction channel two separate BDTs are trained to make use of these difference, one using $B\overline{B}$ and the other continuum type background in the training. In the selection both BDTs are used in combination to classify every event.

To train the BDTs, a sample with known categorisation is needed. For the $b \to c$ and continuum MC, one stream of each is used. For the $b \to u$ ten streams worth of events

are used because of the low amount of signal events, and these types of MC only being provided in two sets worth ten streams each, as detailed in Section 4.5.1. The streams used in the training are not reused in further analysis steps to ensure statistical independence and prevent biases from overtraining effects.

6.1.1. Training variables

There are several, often conflicting, requirements on the variables used to classify events. They need to have a sufficient difference in distribution between the signal events and at least some part of the background to be useful in distinguishing the two. However, to prevent introducing a bias or dependence on the decay model and preserving the entire range of q^2 as uniformly as possible, variables correlated to q^2 should be avoided. If the BDT selection removes part of the range of the correlated variable, due to the correlation a part of the q^2 range could be removed as well. Moreover, as only a few perfectly uncorrelated variables exist and sufficient separation power is necessary to extract the signal component, the selection of variables to use contains a large number of compromises.

The features used in the classification are the same for all reconstructed $\eta^{(\prime)}$ decay channels. The BDT trained on continuum uses additional, specialized variables in addition to the full set used by the $B\overline{B}$ BDT. Two of the variables use a special pseudo-particle Y, calculated similarly to the neutrino in Section 5.6 from only the tag-side particles, that is leaving out the signal lepton and $\eta^{(\prime)}$ candidate in the sum in Equation (4.15). This pseudo-particle serves as an approximation to the second B meson and especially its direction vector. The variables used by both BDTs are:

- The total number of particles in the event satisfying the criteria in Section 5.1 and Section 5.2.
- The missing invariant mass m_{miss} of the event as determined in Equation (4.16).
- The energy asymmetry between the daughter particles of the $\eta^{(\prime)}$ decay. In the $\eta \to \gamma\gamma$ channel, these are the two photons. In the $\eta \to \pi^+\pi^-\pi^0$ and the two $\eta' \to \pi^+\pi^-\eta$ channels there are three daughter particles, the asymmetry uses the $\pi^+\pi^-$ pair. Finally, in the $\eta' \to \rho^0\gamma$ channel the two daughters are the ρ^0 and the γ. The energy asymmetry is defined as:

$$A_\eta = \frac{E_{d1} - E_{d2}}{E_{d1} + E_{d2}}. \tag{6.1}$$

- The difference between the transferred momentum squared q^2 calculated with the adjusted neutrino as described in Section 5.8.1, and using the inferred neutrino from Section 5.6 without any adjustments in Equation (5.18). For correctly reconstructed events, both are distributed around the correct value, and the adjustment

of the neutrino improves the precision. For background however, the inferred neutrino does not correspond to any real particle, causing the modification to move the neutrino towards an unrelated direction. The distribution is shown in Figure 6.1.

- The angle between the $\eta^{(\prime)}$ candidate and the Y representing the other side of the event.
- The angle between the lepton candidate and the Y.
- The number of distant tracks, these are tracks failing the requirement in Equations (5.6) and (5.7) by being too far away from the *IP*.
- The number of $K^{\pm}$ candidates in the event as determined in Section 5.2.2.

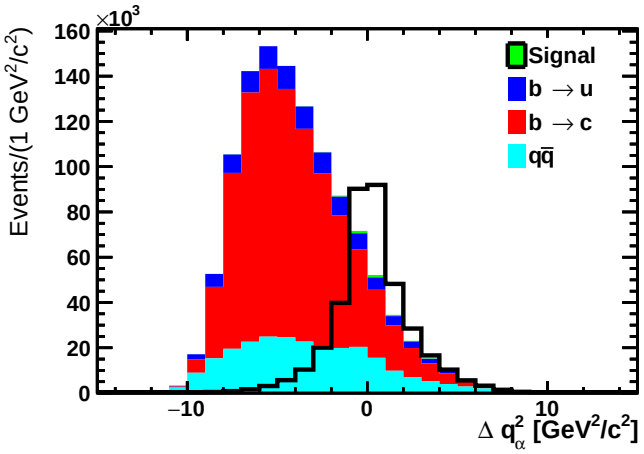


Figure 6.1.: The Δq^2 distribution of the decay mode $\eta \to \gamma\gamma$ used in the `BDT` training. Signal events are symmetrical around zero, while background events peak at much smaller values. The signal contribution is also superimposed with arbitrary scale.

The `BDT`s trained with continuum events as the background source use the following additional variables:

- The cosine of the angle between the thrust axes of the particles reconstructed for the signal decay and the remaining particles in the event, as described in Section 4.4.1. The distribution with the background concentrated at high values can be seen in Figure 6.2.
- The modified Fox-Wolfram moments introduced in Section 4.4.3. Some of the individual moments are strongly correlated with q^2 and therefore not used. The remaining moments going into the `BDT` training are: $R^{so}_{c,0}$, $R^{so}_{c,1}$, $R^{so}_{c,2}$, $R^{so}_{c,3}$, $R^{so}_{c,4}$, $R^{so}_{n,0}$, $R^{so}_{n,2}$, $R^{so}_{n,4}$, R^{oo}_{1}, R^{oo}_{2}, R^{oo}_{3}, R^{oo}_{4} and R_{mm}, which is similar to the missing mass. Not used are the three moments using the missing energy and R^{oo}_{0}.

The distributions for all used variables are compared between signal and background in Figure 6.3 and Appendix A.2.1.

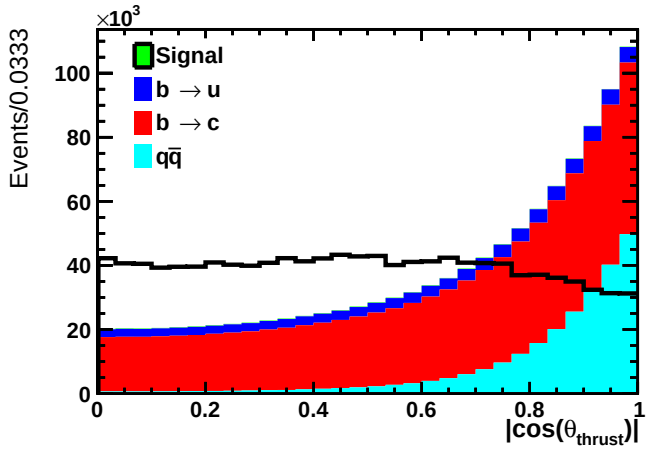


Figure 6.2.: The $|\cos(\theta_{thrust})|$ distribution in the $\eta \to \gamma\gamma$ decay channel used in the BDT training. The signal contribution is also superimposed with arbitrary scale.

The correlations between the variables used are shown in Figure 6.4 and Appendix A.2.2. Removing any of these decreases performance, as variables might be useful for some background components and correlated for others.

6.1.2. BDT settings

While the BDTs are trained completely separately, they use the same settings to improve comparability between channels. The only exceptions are number of events used in the training, which is directly limited by the number of events available, and the number of trees, which is different for each channel. While a higher number of trees allows the BDT to learn more features of the data sample, eventually all underlying features the BDT can learn are indeed learned, and the BDT begins to overtrain. The limiting factor in this analysis is the amount of signal in the sample. The number of events used is listed in Table 6.1. In most cases all available data in the sample were included in the training, however in cases with very many events in the $B\overline{B}$ background, a limit was set due to hardware constraints in the training.

As the amount of signal training events is the same for both the $B\overline{B}$ and the continuum BDT, they have the same amount of trees in the BDT. Here, the number of trees for each channel is listed in Table 6.2. The ideal numbers are determined experimentally as the point where further increase leads to a higher amount of overtraining in the resulting classifier.

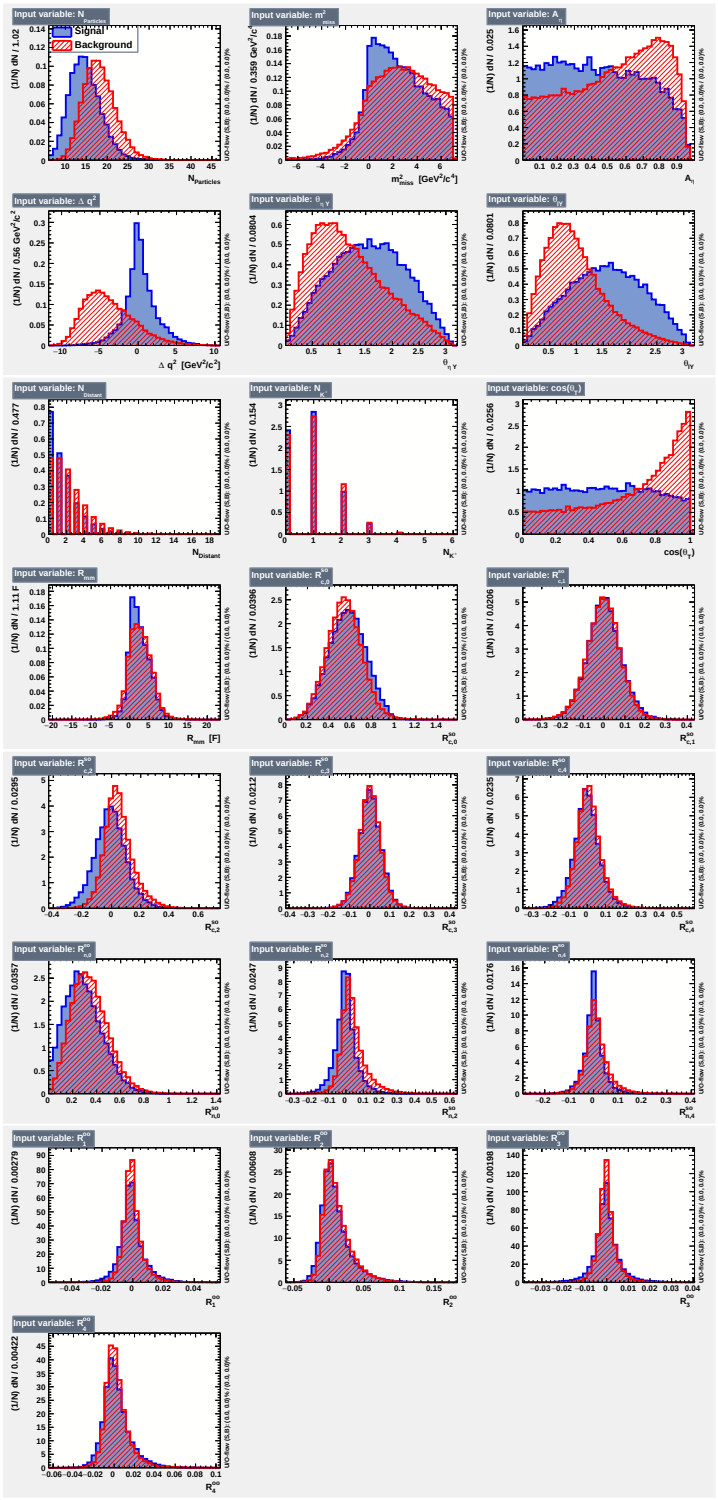
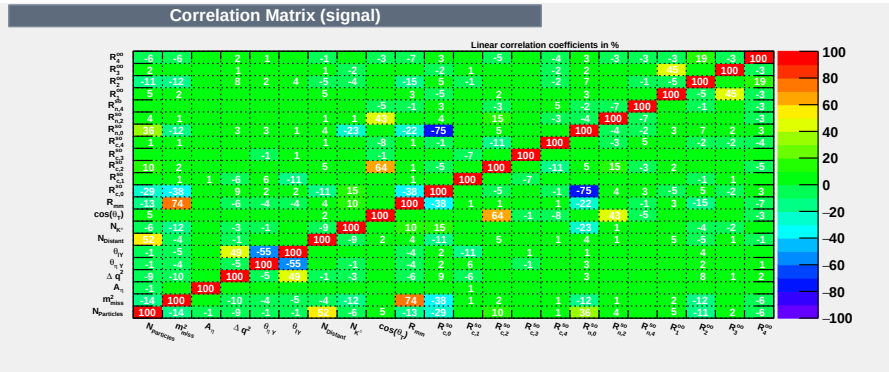
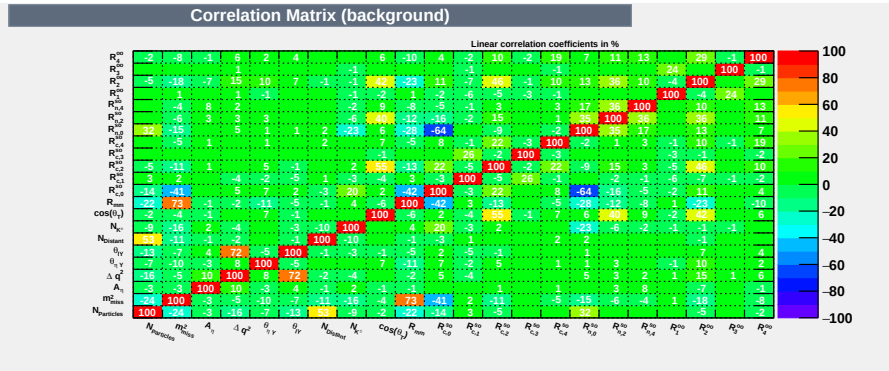


Figure 6.3.: BDT training variables in the channel $\eta \to \gamma\gamma$



(a) Signal sample



(b) Background sample

Figure 6.4.: Correlations between the BDT training variables for signal and background events in the reconstruction channel $\eta \to \gamma\gamma$.

The large number of trees used was found to work best with very shallow individual trees, each of which has a depth of only two layers. The learning rate β (see Section 4.1.2) is set to 0.25. Before the training, the input variables are automatically transformed into uniform shapes by first flattening and then transformation into a Gaussian to help the comparison in the training procedure.

Channel	Signal	$B\overline{B}$ background	Continuum background
$\eta \to \gamma\gamma$	22328	100000	131989
$\eta \to \pi^+\pi^-\pi^0$	6300	100000	14515
$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$	2287	97811	9916
$\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$	335	11584	718
$\eta' \to \rho^0\gamma$	522	10000	23375

Table 6.1.: Number of training events of each category for the BDT. The sample for testing has the same size.

Channel	Trees
$\eta \to \gamma\gamma$	2000
$\eta \to \pi^+\pi^-\pi^0$	1000
$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$	400
$\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$	200
$\eta' \to \rho^0\gamma$	200

Table 6.2.: Number of trees contributing in each reconstruction channel.

6.2. BDT performance and application

The Figure-of-Merit (FoM) defined in Equation (4.3) is maximised to select events for further analysis steps, with all events failing the resulting requirement discarded. In this case, the FoM is a two-dimensional function of the two BDT selections. To prevent any overtraining effects, the FoM is determined on the statistically independent MC sample not used in the BDT training. The resulting distributions are shown in Figures 6.5 and A.13.

The value of the FoM is directly dependent on the assumption for the $\mathcal{B}\left(B^+ \to \eta^{(\prime)}\ell^+\nu_\ell\right)$ used in generating the sample. For comparison, the FoM has also been determined assuming the current world average [4] for the signal branching fractions. The results for both are compared in Table 6.3. While the resulting FoM changes significantly due to the different relative amount of signal and background, the position of the maximum, which determines the BDT selection, stays largely unaffected. The resulting selection values for the BDT are listed in Table 6.4.

In the FoM results, a clear hierarchy between the channels can be seen. In channels with a high FoM a lower statistical uncertainty is expected, showing that the channel $\eta \to \gamma\gamma$ is the most promising for a good result. On the other hand, especially the two channels $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$ and $\eta' \to \rho^0\gamma$ only have a very low FoM.

The efficiencies achieved with the resulting selection are listed in Table 6.5. Together,

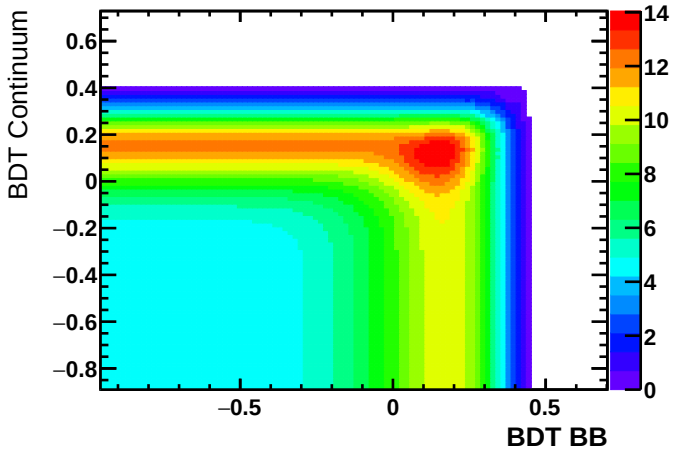


Figure 6.5.: Figure-of-Merit as a function of the two BDT outputs in the channel $\eta \to \gamma\gamma$.

Channel	FoM MC	FoM PDG
$\eta \to \gamma\gamma$	14.1	6.65
$\eta \to \pi^+\pi^-\pi^0$	8.46	3.97
$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$	3.18	1.47
$\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$	1.43	0.663
$\eta' \to \rho^0\gamma$	0.296	0.134

Table 6.3.: Achieved Figure-of-Merit of the BDT selection.

the efficiencies of the initial event selection are also listed. The overall efficiency is determined as the fraction of generated events with the signal decay retained throughout all selection steps:

$$\epsilon = \frac{N^{MC}_{Signal,Reco}}{N^{MC}_{Signal,Gen}}. \tag{6.2}$$

For partial efficiencies, the generated and reconstructed number of events are replaced with the number of events before and after the selection step.

The efficiencies show a separation in two groups of channels. The three channels $\eta \to \gamma\gamma$, $\eta \to \pi^+\pi^-\pi^0$ and $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ all have efficiencies in the same order of magnitude. While the selection efficiency is highest in the $\eta \to \gamma\gamma$ channel, this is balanced by a lower efficiency of the BDT. The two channels $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$ and $\eta' \to \rho^0\gamma$ however have a significantly lower efficiency.

Channel	$B\overline{B}$ BDT	Continuum BDT
$\eta \to \gamma\gamma$	0.138	0.124
$\eta \to \pi^+\pi^-\pi^0$	0.134	0.086
$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$	0.167	0.124
$\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$	0.232	0.092
$\eta' \to \rho^0\gamma$	0.222	0.251

Table 6.4.: Selection for each type of BDT used, all events of a channel which are above both values are accepted for further analysis.

Channel	Event selection	BDT	Combined
$\eta \to \gamma\gamma$	9.38	31.1	2.92
$\eta \to \pi^+\pi^-\pi^0$	4.49	45.3	2.03
$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$	5.23	42.6	2.23
$\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$	1.38	34.9	0.482
$\eta' \to \rho^0\gamma$	0.708	30.6	0.217

Table 6.5.: Selection efficiencies for the selection in Chapter 5 and the BDT selection in %

CHAPTER 7

Signal determination

After applying the BDT requirement, the remaining sample still contains more background than signal events. The amount of signal events in the remaining sample is determined with a binned maximum likelihood fit as described in Section 4.2 to calculate the branching fraction of the $\eta^{(\prime)}$ decays.

This chapter begins with an explanation of the specific setup of the fit for this analysis in Section 7.1. In Section 7.2 the translation from the fit result to a branching fraction is given. The reliability of the fit is evaluated using the MC in Section 7.3. Finally, the data sample is fitted in Section 7.4 and compared to a restricted fit of the background region.

7.1. Fit setup

Each decay channel is fitted separately. The data sample is binned in two dimensions in the variables M_{bc} and ΔE (Equations (5.16) and (5.17)) and fitted to MC templates of the background sources. The fit range in both variables is:

$$5.1\,\mathrm{GeV}/c^2 < M_{\mathrm{bc}} < 5.3\,\mathrm{GeV}/c^2 \tag{7.1}$$

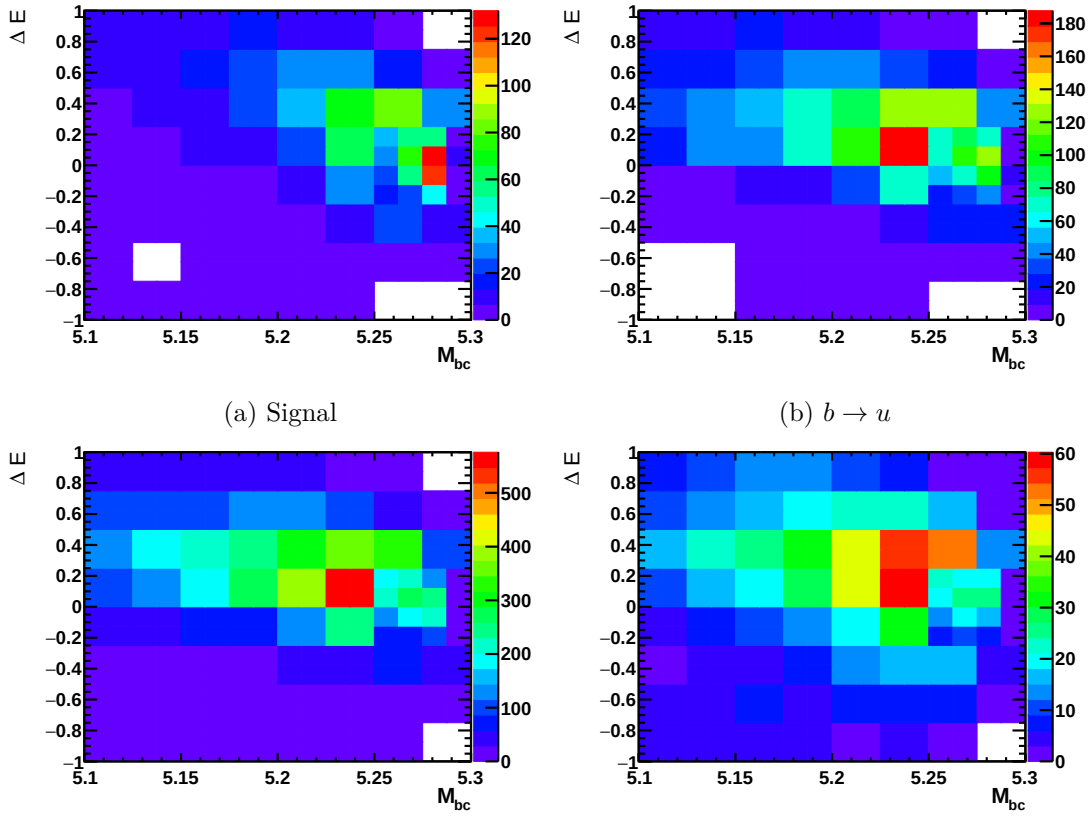
$$-1\,\mathrm{GeV} < \Delta E < 1\,\mathrm{GeV} \tag{7.2}$$

The fit range is wide to make the most out of the available data sample. By extending the range into regions without many signal events, the fit can better estimate contribution purely from background events.

The sample is split eight times in both variables with a width of 25 MeV in M_{bc} and 250 MeV in ΔE. The signal events are concentrated in the area $5.25\,\mathrm{GeV} < M_{\mathrm{bc}}$

and $-0.25\,\mathrm{GeV} < \Delta E < 0.25\,\mathrm{GeV}$. The size of this area is mostly determined by the resolution of the neutrino reconstruction. In this area a further division of bins in both variables is conducted, turning one bin into four bins. This results in an increased precision of the distribution of the signal events for the fit process. At the end the total number of bins is 76.

Four MC components are used in the fit. One contains events correctly reconstructed in the signal decay. The other three correspond to the three types of MC listed in Section 4.5.1, with the signal component removed from $b \to u$. Only signal events with all particles assigned correctly are included in the signal sample, events containing a signal decay but wrong selection of particles, or with a different channel of signal decay are part of the $b \to u$ sample. The amount of such events together is about 10% of the correct signal events. The distribution of these four components in the fitted bins is shown in Figure 7.1 and Appendix A.3.1.



(a) Signal

(b) $b \to u$

(c) $b \to c$

(d) continuum

Figure 7.1.: Distribution of the four fit components over the fit area in the decay channel $\eta \to \gamma\gamma$.

The $b \to u$ contribution is fixed to the expected contribution from previous measurements listed in Equation (2.25). The other three components, that is signal, $b \to c$ and continuum, are freely floated and determined by the fit.

7.2. Branching fraction

The fit procedure as described in Section 4.2 gives the fraction P_{Signal} of signal events in the data sample. Combined with the overall number of data events N^{data} remaining in the sample after the selection, this gives the number of signal events measured as $N^{data} \times P_{Signal}$. For the other fit components the yield is determined similarly with the corresponding fit fractions P_j.

Together with the efficiency ϵ from Equation (6.2) and the sample size $N_{B\overline{B}}$ from Section 3.4 this gives the desired branching fraction $\mathcal{B}\left(B^{\pm} \to \eta^{(\prime)}\ell^{\pm}\nu_{\ell}\right)$ as:

$$\begin{aligned} N^{data}P_{Signal} &= 4N_{B\overline{B}}\mathcal{B}\left(\Upsilon(4\mathrm{S}) \to B^{+}B^{-}\right) \times \\ &\quad \times \mathcal{B}\left(B^{+} \to \eta^{(\prime)}\ell^{+}\nu_{\ell}\right)\mathcal{B}\left(\eta^{(\prime)} \to X\right)\epsilon, \qquad (7.3) \\ \Rightarrow \mathcal{B}\left(B^{+} \to \eta^{(\prime)}\ell^{+}\nu_{\ell}\right) &= \frac{N_{data}P_{Signal}}{4N_{B\overline{B}}\mathcal{B}\left(\Upsilon(4\mathrm{S}) \to B^{+}B^{-}\right)\mathcal{B}\left(\eta^{(\prime)} \to X\right)\epsilon}. \qquad (7.4) \end{aligned}$$

As both parts of the $B^{+}B^{-}$ pair could potentially decay in the signal mode, and the lepton ℓ^{+} could be both an electron and a muon[1], a factor of four appears in the branching fraction of the entire decay chain. The branching fractions of $\eta^{(\prime)}$ to the reconstructed final states, $\mathcal{B}\left(\eta^{(\prime)} \to X\right)$ are listed in Tables 2.3 and 2.4. The branching fraction $\mathcal{B}\left(\Upsilon(4\mathrm{S}) \to B^{+}B^{-}\right)$ is taken as $51.3 \pm 0.6\%$ [4].

7.3. Fit validation on MC samples

Before the fit is performed on data, its behaviour is tested using the MC samples available. Out of these, pseudo-data is generated by merging all fit components together at the appropriate ratio and randomly varying the contents of each bin interpreting the original bin content as the expected value of a Poisson distribution. For this section, 10000 such pseudo-data sets are generated.

Pull distributions are a way to check the reliability of the fit against variations in the data set. The pull [29] is defined as the residual of the fit normalized to the fit uncertainty:

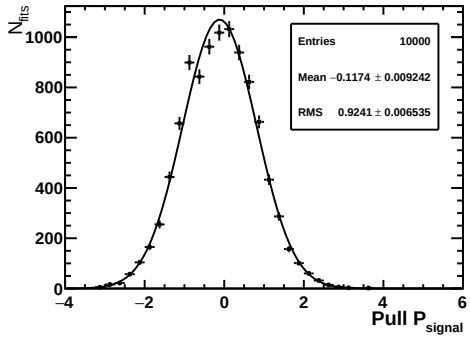
$$Pull = \frac{P_{fit} - P_{input}}{\sigma_{fit}}. \qquad (7.5)$$

[1] This analysis assumes lepton universality.

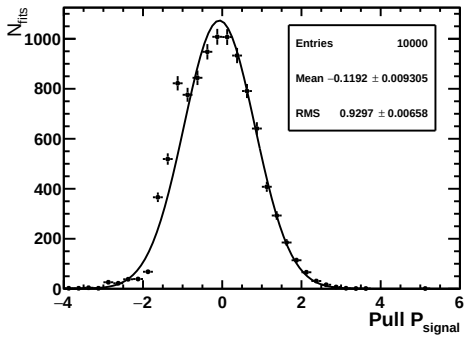
As the pseudo-data is generated from known MC distributions, the original input fraction is precisely known. Considering the variation when generating each new sample, the fit result should vary equally in both directions. If the uncertainties are determined accurately, with a width close to one. The distribution over all fits to pseudo-data is shown in Figure 7.2 for the signal component.

From each of these fits, the fit quality $\chi^2/_{d.o.f.}$ can be determined, with a the number of degrees of freedom being 72. The distributions for all channels are summarized in Figure 7.3. The $\chi^2/_{d.o.f.}$ distribution looks similar for all channels. While in an ideal case the maximum would lie at one, in these cases it lies at slightly lower values. This can be explained by the fitted pseudo-data ultimately originating from the templates used in the fit. Although they are varied, this correlation leads to a higher similarity.

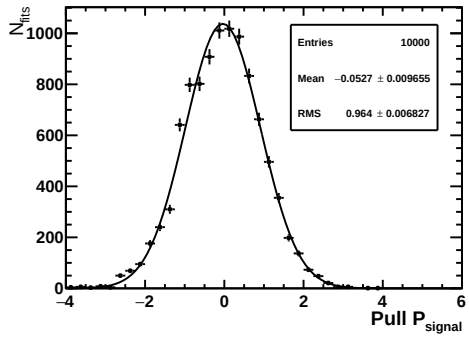
The pull distributions show the channels separating into two groups. The three channels $\eta \to \gamma\gamma$, $\eta \to \pi^+\pi^-\pi^0$ and $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ are reasonably close to the ideal distributions. However, the channel $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$ shows a very asymmetric distribution without a recognizable Gaussian shape, while $\eta' \to \rho^0\gamma$ has a mean significantly offset from one, showing a fit with an expected bias. This observation conforms with the observed efficiencies in Table 6.5, where the problematic two channels are an order of magnitude lower. As a consequence, these two channels are removed from further analysis and for the decay $B^+ \to \eta'\ell^+\nu_\ell$ only the reconstruction channel $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ is used.



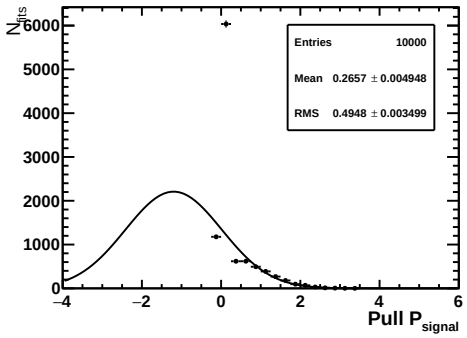
(a) $\eta \to \gamma\gamma$



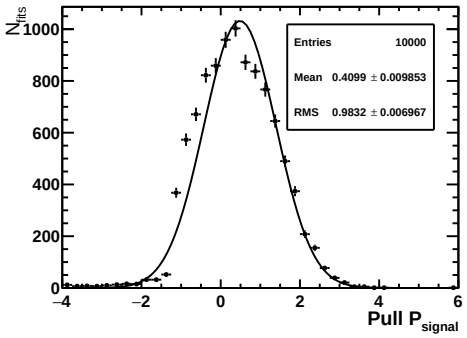
(b) $\eta \to \pi^+\pi^-\pi^0$



(c) $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

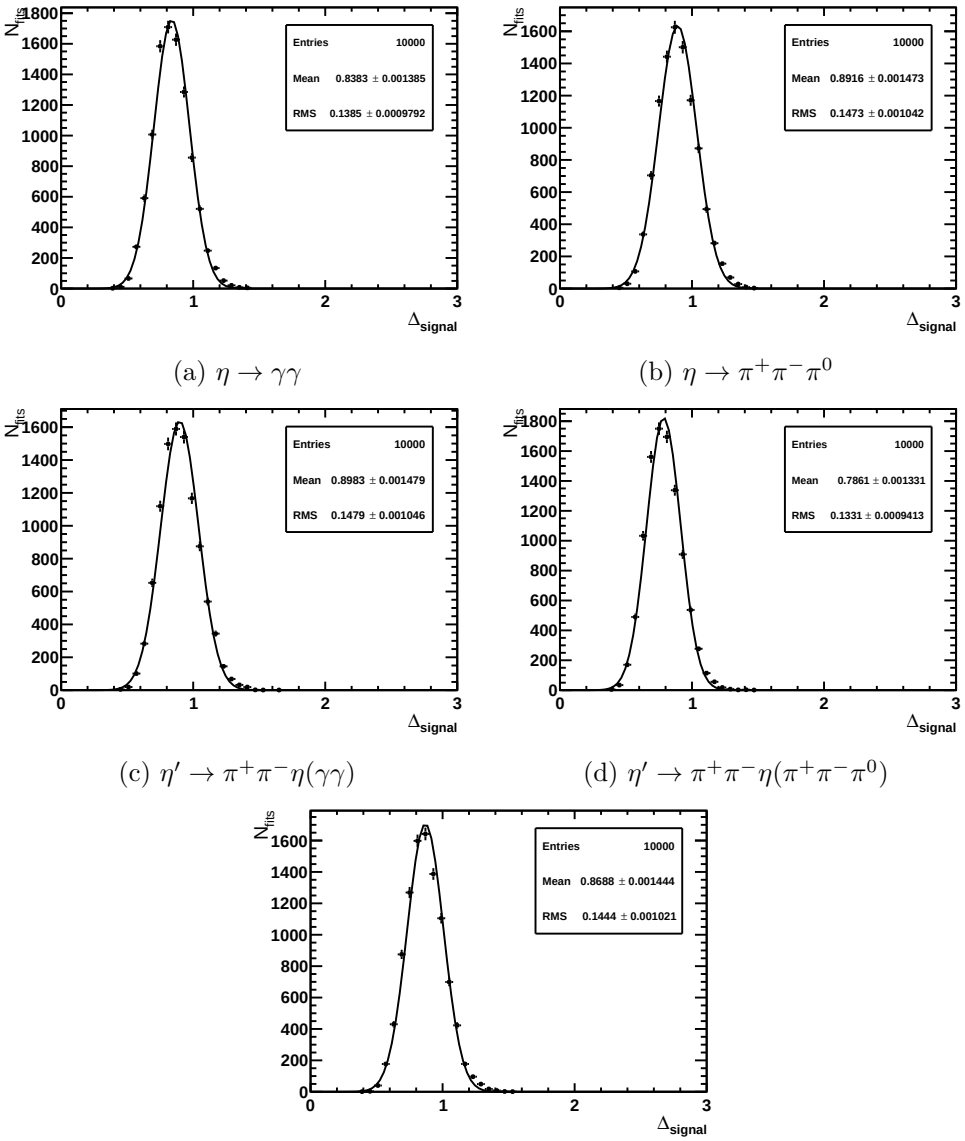


(d) $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$



(e) $\eta' \to \rho^0\gamma$

Figure 7.2.: Pull distributions for all $\eta^{(\prime)}$ reconstruction channels.



(a) $\eta \to \gamma\gamma$

(b) $\eta \to \pi^+\pi^-\pi^0$

(c) $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

(d) $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$

(e) $\eta' \to \rho^0\gamma$

Figure 7.3.: $\chi^2/_{d.o.f.}$ distributions from the fits to pseudo-data.

7.4. Fit on data

After the additional validation summarized in Chapter 8 was finished, the MC components were fitted to the reconstructed data sample. The fitted yields are summarized in Table 7.1. While the fit quality in both $B^+ \to \eta \ell^+ \nu_\ell$ channels is very similar, the $\eta \to \gamma\gamma$ channel has the highest overall yield of events, and also the lowest relative uncertainty in the yield of signal events. In Figure 7.4, the projections on the two fit variables M_{bc} and ΔE are shown with the fitted yield applied to the MC.

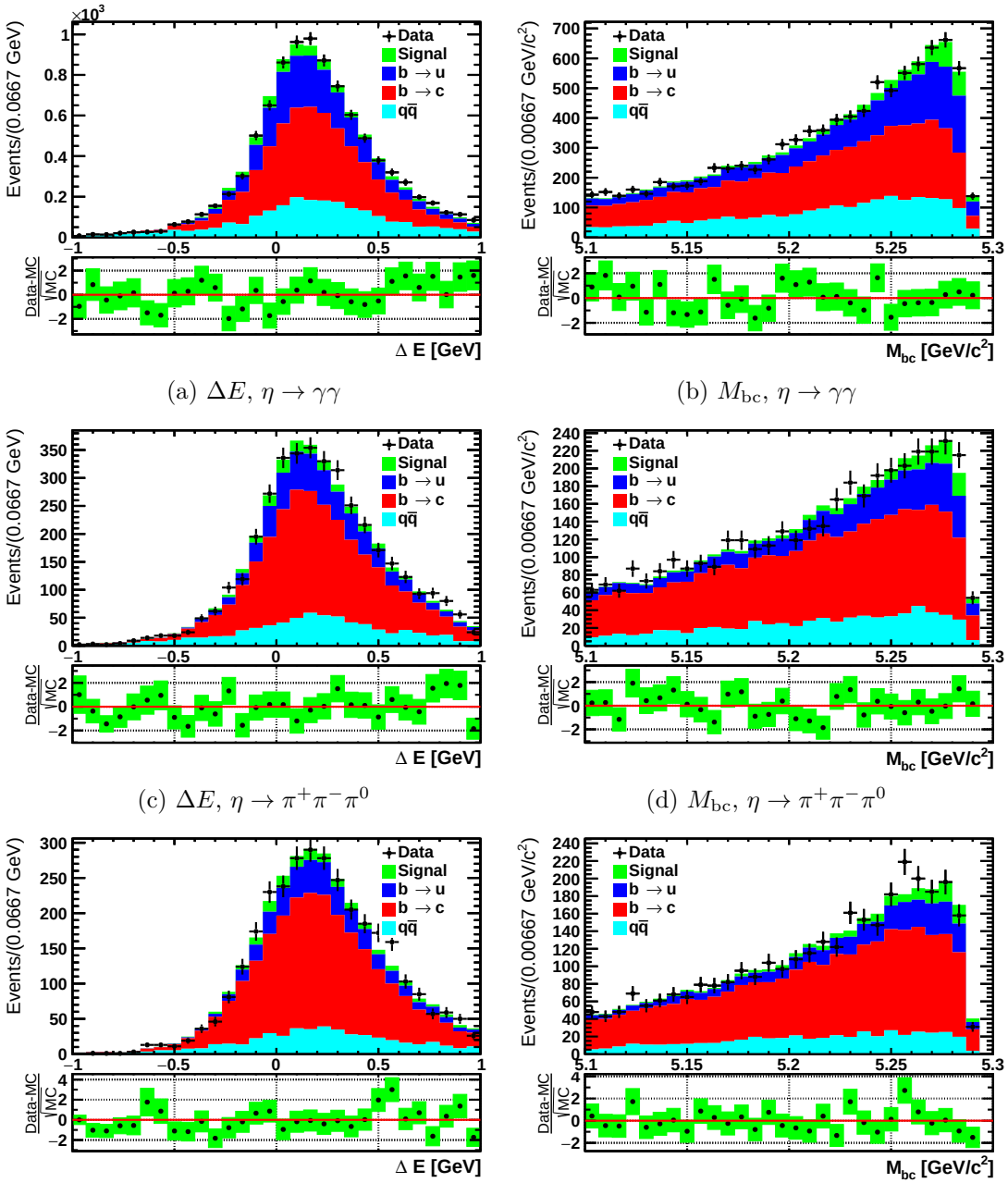
Component	$\eta \to \gamma\gamma$	$\eta \to \pi^+\pi^-\pi^0$	$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$
Signal	530 ± 116	196 ± 77	166 ± 76
$b \to u$	2219 ± 47	674 ± 26	459 ± 22
$b \to c$	4337 ± 233	2262 ± 147	2078 ± 150
continuum	2285 ± 221	692 ± 137	479 ± 129
Data	9373	3828	3185
$\chi^2/_{d.o.f.}$	88.0/72	86.3/72	64.4/72

Table 7.1.: Fitted event yields and quality, together with the total yield of the data sample. Uncertainties are those of the fit procedure, except for the fitted $b \to u$ component, which instead has a Poissonian uncertainty as $\sqrt{N}$ associated.

7.4.1. Background fit

As a validation of the fit result, the same fit procedure has been repeated in a region dominated by background events. Compared to the nominal fit, the background area removes all bins in the signal-rich region $5.25\,\mathrm{GeV}/c^2 < M_{\mathrm{bc}}$ and $|\Delta E| < 0.25\,\mathrm{GeV}$. Additionally, all bins are removed where the signal contribution is expected to be above 5%. The signal template is not fitted in this region. The resulting event yields are listed in Table 7.2 and compared to the yields the nominal fit would expect in this area.

The yields of background events by both method are compatible within their uncertainties for all three channels, increasing confidence in the fit result over the full range. As the nominal fit does contain a small yield of signal events in the background region too, the remaining background yields are generally lower than for the background-only fit. The signal component is however very small and does not compromise comparability between the two fits.



(a) ΔE, $\eta \to \gamma\gamma$

(b) M_{bc}, $\eta \to \gamma\gamma$

(c) ΔE, $\eta \to \pi^+\pi^-\pi^0$

(d) M_{bc}, $\eta \to \pi^+\pi^-\pi^0$

(e) ΔE, $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

(f) M_{bc}, $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

Figure 7.4.: Fit variables comparing data to the MC templates scaled to the fit result. The difference between data and MC events is normalized to $\sqrt{MC}$, corresponding to the MC yield uncertainty assuming Poisson statistics.

Template	$\eta \to \gamma\gamma$		$\eta \to \pi^+\pi^-\pi^0$		$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$	
	Nominal	Bkg only	Nominal	Bkg only	Nominal	Bkg only
Signal	85 ± 19	0	36 ± 14	0	54 ± 25	0
$b \to u$	696 ± 15	693 ± 26	229 ± 9	228 ± 15	213 ± 10	213 ± 15
$b \to c$	2001 ± 107	2133 ± 120	1072 ± 70	1177 ± 75	1205 ± 87	1205 ± 96
continuum	972 ± 94	848 ± 111	313 ± 62	235 ± 66	296 ± 80	338 ± 90
Data	3678		1642		1759	
$\chi^2/_{d.o.f.}$	88.0/72	28.7/34	86.3/72	50.4/37	64.4/72	51.4/49

Table 7.2.: Yields in the background-only region, both by the fit exclusive to this region as well as the nominal fit. The nominal yield and uncertainty of the $b \to u$ component are those of the full range scaled to the background region and therefore differ slightly from the background-only yield.

CHAPTER 8

MC validation

Before the fit was applied to data in Section 7.4, several studies were conducted to confirm the reliability of the MC and the assumptions made during the reconstruction. These are collected in this chapter.

It starts with sideband studies in the $\eta^{(\prime)}$ mass in Section 8.1 to validate the background description, followed by a more specific look at the continuum background in Section 8.2. Afterwards, the dependence of the reconstruction on q^2 is evaluated in Section 8.3. A control mode is presented in Section 8.3.1.

8.1. Sidebands

An accurate description of the background distributions is important for extracting the signal component. To validate this description, so called *sidebands* are used, samples containing almost only background events. These are reconstructed in the same way as in Chapter 5 and use the `BDT`s from Chapter 6. However, to remove the signal component, the mass selection of $\eta^{(\prime)}$ candidates is changed.

Sidebands are useful to study the agreement of MC and data without introducing a bias. When studying the agreement on signal distributions, the size and shape of the observed data can influence the selections chosen. This could lead to optimizing the selection around statistical fluctuations and biasing the result. As the sideband does not contain significant signal, such a risk is not present.

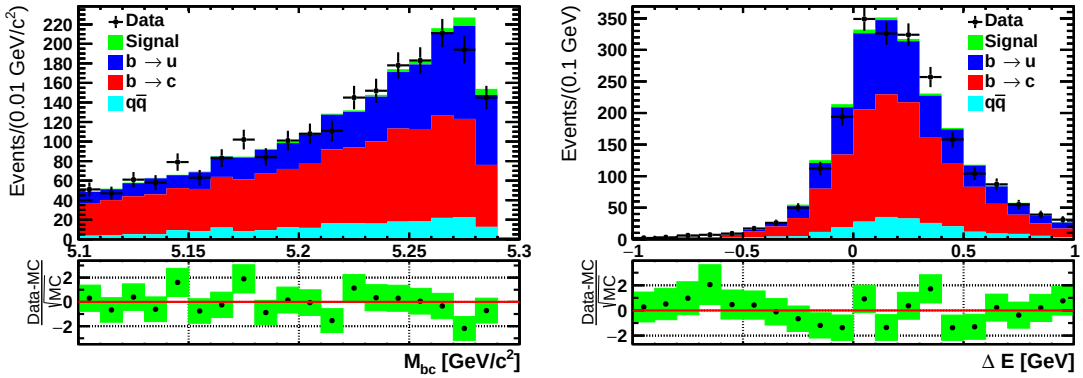
For the two channels $\eta \to \gamma\gamma$ and $\eta \to \pi^+\pi^-\pi^0$, the sideband uses events with a mass between 3 and 5σ around the reconstructed peak. Table 8.1 summarises the ac-

cepted ranges. For $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ the sideband is defined in the mass difference Δm (Equation (5.8)). All candidates with more than 6σ above the reconstructed peak, here above $430\,\mathrm{MeV}/c^2$, are used as sideband candidates. Events are not considered for the sideband if they also contain an accepted signal candidate in the same channel.

Channel	Lower Range [MeV/c^2]	Higher Range [MeV/c^2]
$\eta \to \gamma\gamma$	$[487, 510]$	$[580, 603]$
$\eta \to \pi^+\pi^-\pi^0$	$[535, 540]$	$[556, 561]$

Table 8.1.: Mass ranges of the $B^+ \to \eta\ell^+\nu_\ell$ sidebands.

The most important variables to have good agreement between data and MC are those used in the signal determination, specifically the fit variables M_{bc} and ΔE. In the two channels $\eta \to \gamma\gamma$ and $\eta \to \pi^+\pi^-\pi^0$ the comparison is shown in the Figures 8.1 and 8.2. In both of these good agreement between MC and data can be observed.

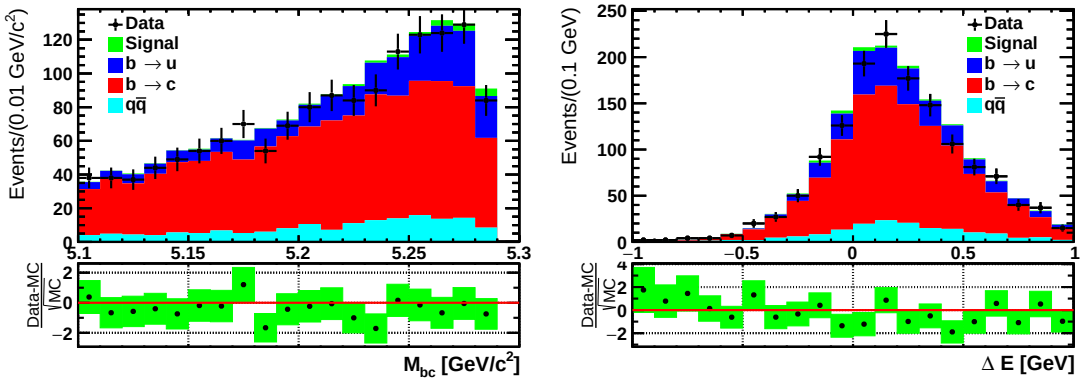


(a) M_{bc} (b) ΔE

Figure 8.1.: Fit variables in the sideband for the reconstruction channel $\eta \to \gamma\gamma$.

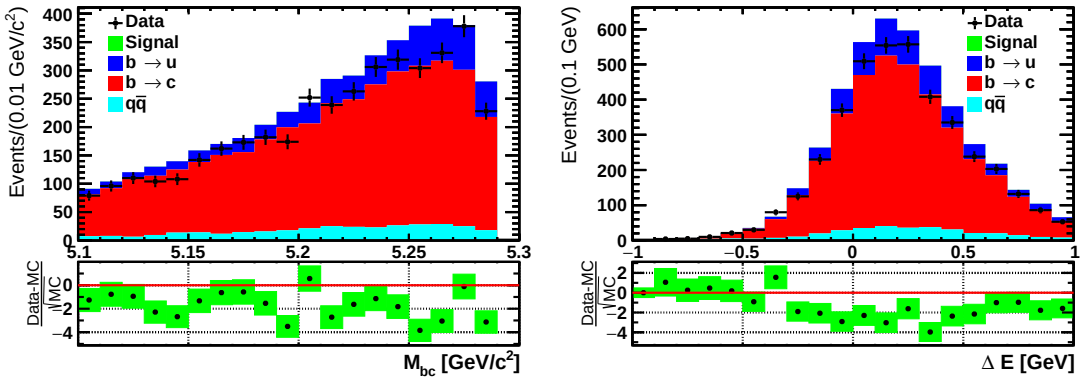
In the channel $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$, shown in Figure 8.3, the distribution is slightly different. Especially in Figure 8.3a the data can be observed to consistently lie below the MC. However, no structure can be seen beyond fluctuations, indicating this is purely a problem of overall normalisation. As only the shape of the components is relevant for the fit procedure, a slight discrepancy in normalisation does not affect the result. This is consistent with the results in Section 7.4.1 showing no discrepancy when fitting only the background to the data sample.

While agreement in all variables is desirable, MC generators are always only an approximation of the true behaviour. Figure 8.4 shows $\theta_{\eta^{(\prime)}Y}$, one of the variables used in the `BDT`. While here a generally good agreement is seen, small discrepancies in the



(a) M_{bc} (b) ΔE

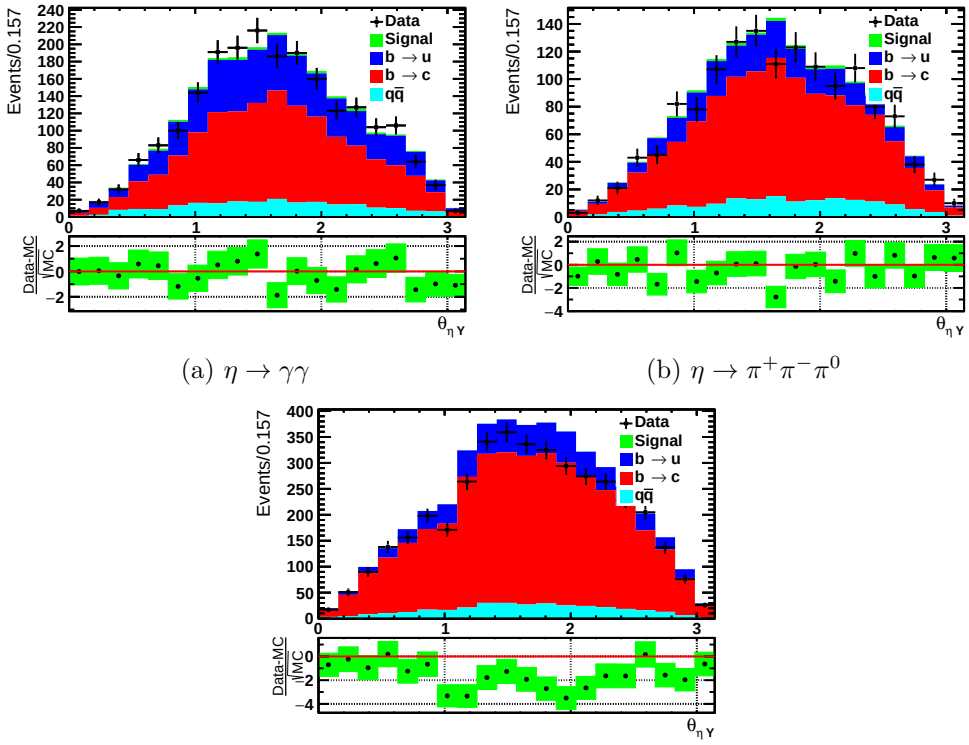
Figure 8.2.: Fit variables in the sideband for the reconstruction channel $\eta \to \pi^+\pi^-\pi^0$.



(a) M_{bc} (b) ΔE

Figure 8.3.: Fit variables in the sideband for the reconstruction channel $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$.

training variables are not a problem in this analysis. If the BDT output was later used as a fit variable, agreement would be more important and could bias the result. However, if the BDT is only used to separate events, discrepancies only lead to a loss in performance due to non-optimal selections. As with all other selection steps, discrepancies can affect the efficiency estimation however.



(a) $\eta \to \gamma\gamma$ (b) $\eta \to \pi^+\pi^-\pi^0$

(c) $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

Figure 8.4.: Angle between the $\eta^{(\prime)}$ and the pseudoparticle representing the other side.

8.2. Continuum MC comparison

At the energy of the $\Upsilon(4S)$ resonance, the cross sections for continuum processes going to one of the four lighter quarks are about three times as large as the cross section for the production of an $\Upsilon(4S)$ resonance:

$$\sigma_{\Upsilon(4S)} = 1.1 \times 10^{-9}\,\mathrm{b}, \tag{8.1}$$

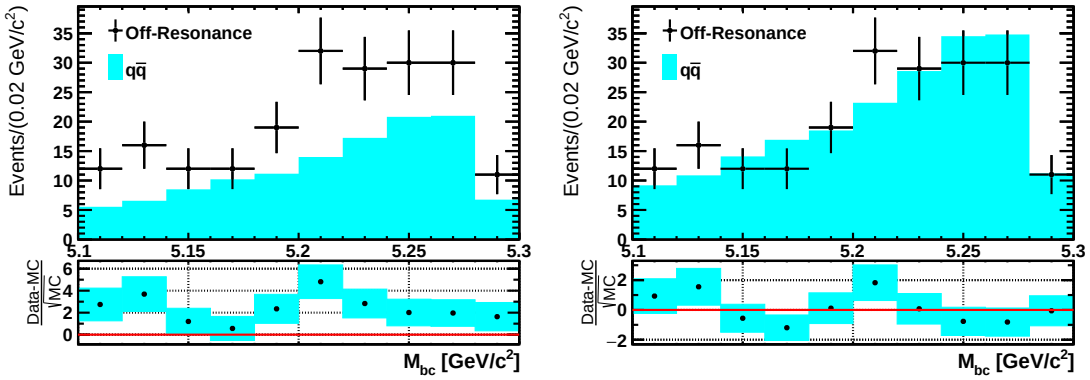
$$\sigma_{uds} = 2.05 \times 10^{-9}\,\mathrm{b}, \tag{8.2}$$

$$\sigma_{charm} = 1.25 \times 10^{-9}\,\mathrm{b}. \tag{8.3}$$

The heaviest of the four is listed separately, as it has the highest chance to produce heavier particles mimicking signal processes. In the MC the charm component is also generated separately from the combined *uds* component.

The data taken slightly below the $\Upsilon(4S)$ resonance can be used to validate the description of this part of the MC separately from the rest. The difference in event shape explained in Section 4.4 allows for an effective determination of these events, but with their large cross section, they stay a relevant part of the background.

When comparing the MC with the off-resonance data in Figure 8.5a, a clear difference can be seen with significantly more data events. This is not unexpected, as the off-resonance data contains more processes than just the $e^+e^- \to q\bar{q}$ events. One such additional process is $e^+e^- \to \tau^+\tau^-$.



(a) Without rescaling

(b) With rescaling

Figure 8.5.: Off-resonace data compared to continuum MC for M_{bc} in the $\eta \to \gamma\gamma$ channel

When scaling the continuum MC to attain the same integral as the off-resonance data in Figure 8.5b, the agreement improves. As the continuum part is a separate component freely floating in the fit as detailed in Section 7.1, the effect of the overall scaling does not affect the result.

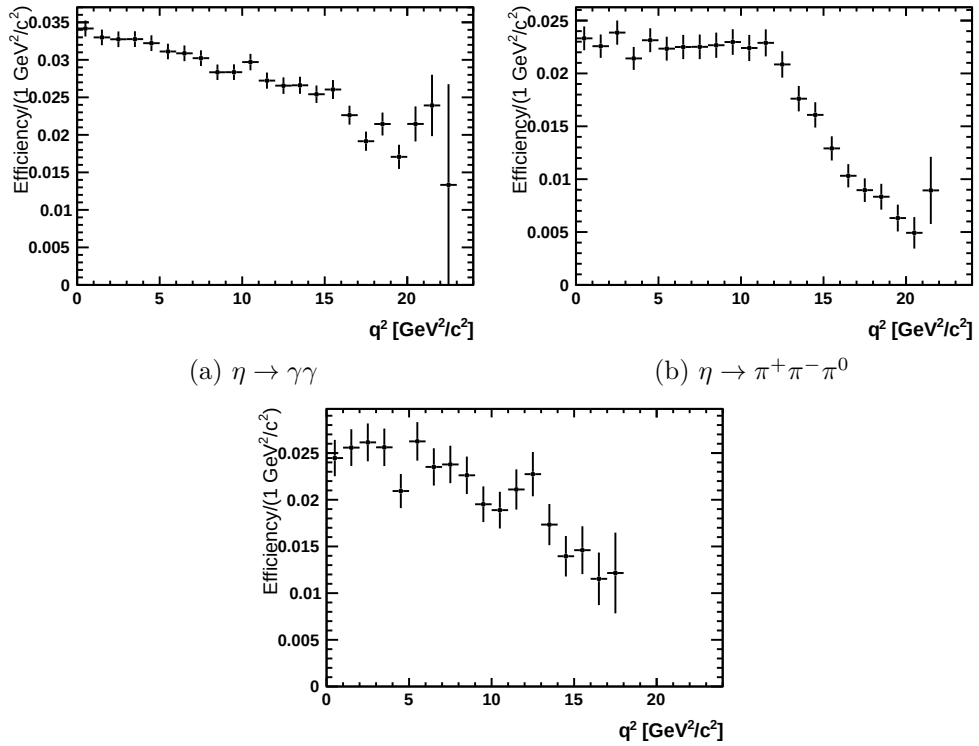
The off-resonance data could in principle be used to independently determine the contribution of off-resonance effects. However, due to the efficiency of the `BDT` in reducing continuum background, the remaining sample size is very small, leading to high uncertainties on the yield from off-resonance data and making this approach infeasible.

For the other two reconstruction channels, the comparison is shown in Figures A.18 and A.19. The observed discrepancy in yield is smaller than in the $\eta \to \gamma\gamma$ channel. The continuum parts not described by the MC are in part removed by the additional requirements on the charged hadrons.

8.3. Efficiency dependence of q^2

The analysis aims to keep the entire q^2 range intact throughout the reconstruction to avoid modelling errors, as the description of the decay kinematics often depends on the transferred momentum. Examining the selection efficiencies relative to q^2 in Figure 8.6

gives an indication to what extent this has been achieved.



(a) $\eta \to \gamma\gamma$ (b) $\eta \to \pi^+\pi^-\pi^0$

(c) $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

Figure 8.6.: Signal efficiencies over the entire event selection chain as a function of q^2, determined on MC samples.

The first observation is that in all channels, the entire range is retained with the efficiency never falling to zero. While the ideal case would be a flat distribution, giving an efficiency independent of the momentum transfer, this is not reached. The form of the efficiency dependence varies for the different decay channels. For $\eta \to \gamma\gamma$, a slight linear decrease towards higher values can be observed. The other two channels show the same general structure. At low values, the efficiency is flat and independent of q^2, followed by a decrease above a certain threshold. This effect is caused by the detection efficiency of the pions included in both channels being dependent on their energy, which in turn depends on q^2.

8.3.1. Control mode

The signal extraction method is validated by reconstructing a control mode, a process that is similar to the nominal signal process. It can be reconstructed in the same way

while being much better known from previous analysis. Ideally suited are decays with a high enough branching fraction that reconstruction can be applied with as little modification as possible. As the control mode is well known, its reconstruction should reproduce the previous measurements.

The decay process chosen as control mode is $B^+ \to D^0 \ell^+ \nu_\ell$, as it combines a relatively high and well measured branching fraction with an event topology somewhat similar to the signal decays. The D^0 is reconstructed in the decay $D^0 \to K^- \pi^+$.

Reconstruction

The D^0 candidates are reconstructed as pairs of one kaon and one pion of opposite charge. The pair is required to have a reconstructed invariant mass in a 3σ window around the fitted maximum mass similar to the signal modes. This results in a requirement of $1850\,\mathrm{MeV}/c^2 < m_{K\pi} < 1879\,\mathrm{MeV}/c^2$. All following steps, including the `BDT` training, are not changed compared to the signal modes.

The main background source in the control mode are D mesons originating from other sources, mostly from semileptonic decays $B \to D^* \ell^+ \nu_\ell$, with the D^* subsequently decaying into a D and an unreconstructed pion. While the D is correctly selected, it is incorrectly associated directly with the original B meson, as neither an attempt to reconstruct D^* mesons not to veto them is made. Another source for D mesons are $e^+e^- \to c\bar{c}$ processes, making up almost the entire continuum component. This can be seen in the resulting mass peak in Figure 8.7. This mass peak also shows one of the deficiencies of the MC used. The simulated peak is significantly narrower than on real data. The selection width chosen is wide enough that the difference on the efficiency between data and MC is however negligible.

Control mode result

The fit in the control mode uses the same setup as for the signal modes in Chapter 7. The fit result for the branching fraction is:

$$\mathcal{B}\left(B^+ \to D^0 \ell^+ \nu_\ell\right) \;=\; (2.536 \pm 0.036 \pm 0.087)\%, \tag{8.4}$$

where the first uncertainty is statistical from the fit and the second is systematic. The systematic uncertainties follow the determination for signal events detailed in the Chapter 9, with the individual breakdown in Table 9.1. As the $B^+ \to D^0 \ell^+ \nu_\ell$ decay has a very high statistic, the uncertainty is dominated by systematic effects.

The current world average [4] for this decay is $\mathcal{B}\left(B^+ \to D^0 \ell^+ \nu_\ell\right) = 2.35 \pm 0.09\%$, or

$$\mathcal{B}\left(B^\pm \to D^0 \ell^\pm \nu_\ell\right) = 2.29 \pm 0.09\% \tag{8.5}$$

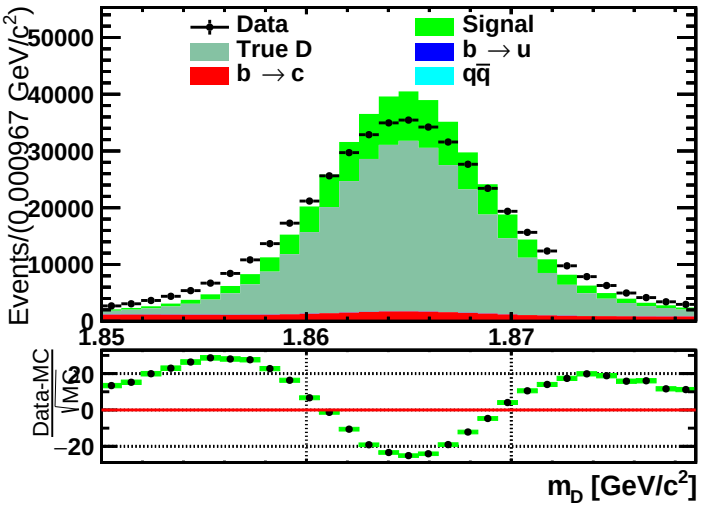


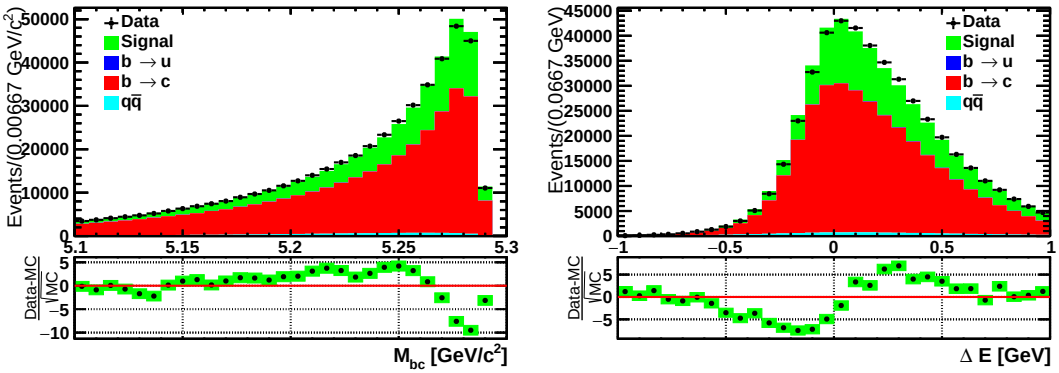
Figure 8.7.: The reconstructed D mass. The true D meson category contains all correctly reconstructed $D^0 \to K^-\pi^+$ processes in which the D does not originate from a $B^+ \to D^0\ell^+\nu_\ell$ decay. These D mesons make up the majority of the background. The MC simulates a mass peak narrower than what is found in data.

when rescaling to $\mathcal{B}(\Upsilon(4S) \to B^+B^-) = 51.3\%$, as is assumed in this analysis. Within their combined uncertainties, these two values are just slightly incompatible.

A previous analysis of this channel by Belle measured [55] $\mathcal{B}\left(B^+ \to D^0\ell^+\nu_\ell\right) = 2.58 \pm 0.04 \pm 0.15\%$ and this is very well compatible with the obtained result. This comparison suggests a systematic source of the discrepancy to the measurement average. There is also the possibility of additional systematic uncertainties specific to the $B^+ \to D^0\ell^+\nu_\ell$ decay being neglected, possibly related to the imperfect MC simulation.

The fit variables scaled to the result are shown in Figure 8.8. Even with the fit result applied, there is still a significant amount of discrepancy to be seen between the data and the MC. This is considered to be due to defects in the MC, as the mass peak in Figure 8.7 is also not well modelled. This kind of discrepancy is however not seen in the sidebands of the signal channels in Section 8.1.

To account for this discrepancy to the world average a systematic uncertainty will be applied for the signal modes as detailed in Section 9.4.



(a) M_{bc} (b) ΔE

Figure 8.8.: Fit variables in the control mode $B^+ \to D^0 \ell^+ \nu_\ell$ with the fit result applied.

CHAPTER 9

Systematics

Besides the statistical uncertainty of the result, which originates from an uncertainty in the fit as described in Chapter 7, there are also a variety of systematic uncertainties. These are listed here together with their determination.

The chapter begins with uncertainties related to the description of the initial interactions and particle decays in Section 9.1, followed by uncertainties on the detector response to traversing particles in Section 9.2. Effects of the limited size of simulated data are described in Section 9.3. Finally, an uncertainty due to the control mode discrepancy is introduced in Section 9.4. The size of all uncertainties is collected in Table 9.1.

9.1. MC description

The generation of the MC relies on previous measurements and theoretical descriptions of the decay properties of the involved particles. Both of these always have an associated uncertainty. As most of these translate into uncertainties in the composition of the templates in the fit, the uncertainty of the end result is generally determined by varying each value individually within one standard deviation and repeating the fit. The difference to the nominal fit result is interpreted as the uncertainty of the result.

The main background processes are semileptonic decays of the type $B \to X\ell^+\nu_\ell$. Therefore, the main contributing decay modes are updated to the current averages in Section 4.5.1 by reweighting the corresponding events in the sample. For each of these decays, a corresponding uncertainty is assigned from the uncertainty of the new weight.

A separate case is the inclusive $B \to X_u \ell^+ \nu_\ell$ measurement. Here only the inclusive part of the $b \to u$ is varied to achieve the overall variation in size in the $b \to u$ component. The exclusive modes are not affected by the original rescaling in Section 4.5.1 and are evaluated separately.

The branching fraction $\mathcal{B}\left(B^+ \to \eta^{(\prime)} \ell^+ \nu_\ell\right)$ is determined from the event yield and the full decay chain. Part of this chain are the initial probability to produce a B^+B^- pair, $\mathcal{B}\left(\Upsilon(4\mathrm{S}) \to B^+B^-\right)$, and the branching fractions of the $\eta^{(\prime)}$ to the various final states. Their associated uncertainties are evaluated directly within Equation (7.4).

In this analysis, the main unknown about the continuum contribution is its overall size, however as the normalization is determined in the fit this has no effect on the analysis. The remaining uncertainty comes from the relative cross section of the charm and *uds* components. By varying the two contributions by 20% relative to each other the effect of this uncertainty is estimated.

9.1.1. Form factors

The kinematic behaviour of particle decays is described using different models, which use a certain number of measured parameters, called form factors. This modelling can lead to two kind of uncertainties: either an uncertainty on the form factors used, or the theoretical model itself being imprecise and not accurately describing the underlying physics.

For most of the particle transitions described, the models used differ. This leads to two general ways of assigning uncertainties depending on whether the decay is via an initial $b \to c$ or $b \to u$ transition. All reweightings and uncertainties concern semileptonic decays, which dominate the background.

The decays $B \to D\ell^+\nu_\ell$ and $B \to D^*\ell^+\nu_\ell$ are modelled using HQET (Heavy Quark Effective Theory) [56] in the MC. HQET describes bound states between a heavy and a light quark assuming the masses of the b and c quarks involved in the decay as infinite, a limit in which they behave identical except for spin and flavour, and later adds corrections for the masses. This approach is, however, only valid for masses larger than the strong coupling Λ_{QCD} and HQET can not be used to describe $b \to u$ decays as the u is too light.

The description of the decays $B \to D\ell^+\nu_\ell$ uses one form factor, while $B \to D^*\ell^+\nu_\ell$ uses three form factors. These have been updated [50] to the most recent values [4]. The decays $B \to D^{**}\ell^+\nu_\ell$, where the D^{**} can be one of the four states, D_1, D_0, D_0' and D_2, are described with the ISGW2 [57] model in the MC samples. These decays have been reweighted to conform to the more accurate LLSW [58] model describing the decays with three form factors. To determine the uncertainty, each form factor was individually varied within its uncertainty in the reweighting procedure.

The signal decays $B^+ \to \eta \ell^+ \nu_\ell$ and $B^+ \to \eta' \ell^+ \nu_\ell$ use the ISGW2 [57] model in the MC. The uncertainty on the modelling of the kinematics is determined by reweighting [59] the signal contribution, with the form factors updated to the Ball07 [60] model. The $b \to u$ background decay $B^+ \to \omega \ell^+ \nu_\ell$ uses the Ball98 [61] in the MC, and is reweighted using the same procedure to the Ball05 [62] model for comparison.

The comparison models for the $b \to u$ processes are based on LCSR [63] (Light Cone Sum Rule). The difference in fit result between the two models is assigned as the uncertainty. The shape parameters [64] describing the inclusive component in the $b \to u$ MC are also considered.

9.2. Experimental conditions

The last step in the MC generation is the simulation of the behaviour of the detector reacting to the MC particles. As the description is not always perfect, discrepancies in detection efficiency can occur. Separate studies are available to compare the MC behaviour with data for different particle types, leading to corrections on the MC events if divergences are found. These corrections lead to associated uncertainties on their size. However, even in cases where no divergence could be observed, an uncertainty is provided representing the precision of the comparison.

9.2.1. Particle detection

The reconstruction efficiency for charged tracks was studied in D^* decays [65]. In this study, the D^* was only partially reconstructed, leaving one charged particle out whose presence was inferred from the vertices. The detection efficiency is determined by whether or not a corresponding track to the inferred particle has been measured.

No significant difference in efficiency between data and MC was observed. The uncertainty of this measurement results in the assignment of a fully correlated uncertainty of 0.35% per charged track with a momentum $p > 200\,\mathrm{MeV}/c$ in the reconstructed state. All tracks considered in this analysis fulfil this criterion.

Besides charged particles, photons are the main type of particles to be detected. Their detection efficiency is determined in a study using radiative Bhabha events [66]. This study determined the momentum of the radiative photon in $e^+e^- \to e^+e^-\gamma$ reactions by using the missing momentum of the combined e^+e^- pair, and compared the efficiency in finding the accompanying photon between real and simulated data. No correction results from this, but an uncertainty of 2% per photon in the final state is introduced. This uncertainty is not applied to the Bremsstrahlung photon used to correct the electron momentum in Section 5.5.

Photons pairs used to reconstruct π^0 candidates are treated differently. Here, the efficiency was studied separately [67] using decays of η mesons containing different numbers of π^0, without restrictions on the production channel of the η. This results in a lower uncertainty of 2.5% assigned per π^0 used in the reconstruction, instead of 2% for each of the photons.

The particle identification procedures, both for leptons and hadrons, behave differently on data and in MC predictions. The difference for hadrons was determined in $D^{+*} \to D^0\pi^+$ decays, followed by $D^0 \to K^-\pi^+$. For charged leptons, $e^+e^- \to e^+e^-\ell^+\ell^-$ events were used. Corrections [68, 69] are provided to compensate for this discrepancy. Each of these corrections comes with an uncertainty, the size of which is depending on the momentum and angle inside the detector of the corrected particles.

The total number of $B\overline{B}$ pairs listed in Equation (3.6) is derived from the measured integrated luminosity. There, the luminosity is measured via Bhabha events. This measurement is the main source of uncertainty on the listed value. This uncertainty is propagated through the calculation of the branching fraction to the end result.

While events containing additional neutrinos are mostly removed by requiring exactly one lepton candidate, this does not constrain the number of K^0_L. As these are only partially measured by the `KLM` system, they constitute a second unmeasured particle compromising the assumptions made in Section 5.6, or fake a muon candidate. As a conservative estimation of the uncertainty, the fraction of events containing at least one K^0_L in the fit templates is varied by 20%.

Events accepted in the reconstruction should only contain one lepton coming directly from the signal B^+ decay. However, in background events the used lepton candidate can also originate from a subsequent decay of a B daughter, continuum or beam backgrounds, or not be a lepton at all, as hadrons can be misidentified as leptons. These parts of the MC are not well modelled. To estimate an upper bound on the effect, an uncertainty is assigned after increasing and decreasing the weight of all events with a lepton candidate not directly out of a semileptonic B decay by 20%.

9.3. Event yield

The efficiency in reconstructing an event is important to determine the branching fraction out of the measured yield. This efficiency is summarized in Table 6.5. While the determination would be exact in the case of unlimited statistics, the limited amount of events available in any MC sample causes an uncertainty in the efficiency. While the number of generated events can be treated as exactly known, the yield of accepted events during the reconstruction is interpreted as following a Poissonian distribution with the

uncertainty as

$$N^{MC}_{Signal,Reco} \pm \sqrt{N^{MC}_{Signal,Reco}}. \tag{9.1}$$

The same situation appears in the $b \to u$ component, fixed in the fit procedure. Here too the uncertainty on the yield is taken as an Poissonian uncertainty. The accompanying uncertainty is determined by varying the fixed yield within this Poissonian uncertainty before repeating the fit.

9.4. Control mode

The branching fraction determined in the control mode $B^+ \to D^0 \ell^+ \nu_\ell$ in Section 9.4 does not exactly agree with the reference value. The level of disagreement between the measurements of the branching fraction is quantified as

$$\sigma_D \;=\; |\mathcal{B}\,(Fit) - \mathcal{B}\,(Ref)| - \sqrt{\sigma^2_{\text{stat}} + \sigma^2_{\text{syst}} + \sigma^2_{\text{Ref}}}. \tag{9.2}$$

Here, $\mathcal{B}\,(Ref)$ is the reference branching fraction (Equation (8.5)) with the uncertainty σ_{Ref}. $\mathcal{B}\,(Fit)$ is the measured branching fraction as quoted in Equation (8.4) with the uncertainties σ_{stat} and σ_{syst}. Interpreting this as a general uncertainty of the fit, an additional uncertainty of 5% is added to the determination of the $B^+ \to \eta^{(\prime)} \ell^+ \nu_\ell$ branching fractions.

Source	$\eta \to \gamma\gamma$	$\eta \to \pi^+\pi^-\pi^0$	$\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$	$D^0 \to K^-\pi^+$
Statistical	22	39	46	1.4
Combined Systematic	11	14	11	3.4
$\mathcal{B}(B^+ \to X_{Bkg})$	2.4	1.7	1.3	0.52
$\mathcal{B}(\Upsilon(4S) \to B^+B^-)$	1.2	1.2	1.2	1.2
$\mathcal{B}\left(\eta^{(\prime)} \to X\right)$	0.51	1.2	1.7	1
Continuum	0.2	0.62	0.63	0.093
$B \to D^{(*,**)}\ell^+\nu_\ell$ form factors	0.82	1.1	1.3	0.36
$B^+ \to \eta^{(\prime)}\ell^+\nu_\ell$ form factors	3.0	2.9	0.14	0.00039
$B^+ \to \omega\ell^+\nu_\ell$ form factors	0.81	2.1	2	0.00039
$b \to u\ell^-\overline{\nu}_\ell$ shape	0.39	0.15	0.21	0.00039
Charged tracks	0.35	1.1	1.1	1.1
γ detection	4.0	2.5	4.0	0
Electron ID	1.6	1.6	1.5	1.5
Muon ID	2.1	2.1	2	1.9
First $\pi^\pm/K^\pm$ `PID`	0	0.97	1.1	0.89
Second $\pi^\pm$ `PID`	0	1.3	2.2	0.86
$N_{B\overline{B}}$	1.4	1.4	1.4	1.4
Background with K^0_L	3.5	8.6	3.8	0.36
Misidentified leptons	4.3	5.5	2.3	0.27
Efficiency determination	0.86	1.3	2.3	0.11
$b \to u\ell\nu$ yield	4.1	5.2	4.4	0.0051
Control Mode	5.0	5.0	5.0	-

Table 9.1.: All systematic uncertainties in %, with the statistical uncertainty for comparison.

CHAPTER 10

Results and Conclusion

This chapter presents the final branching fractions determined for the signal decay modes in Section 10.1. Afterwards, the results are discussed and compared to previous measurements in Section 10.2. Finally, Section 10.3 gives a very brief outlook into the future that further studies could bring.

10.1. Results

The yield of events in the $B^+ \to \eta^{(\prime)}\ell^+\nu_\ell$ decays was determined in Section 7.4. This measurement allows to determine the branching fractions of these decays. The results for the $B^+ \to \eta\ell^+\nu_\ell$ decay are:

$$\mathcal{B}\left(B^+ \to \eta\ell^+\nu_\ell\right) = \tag{10.1}$$

$$\textit{using } \eta \to \gamma\gamma \;:\; (2.91 \pm 0.64 \pm 0.32) \times 10^{-5}, \tag{10.2}$$

$$\textit{using } \eta \to \pi^+\pi^-\pi^0 \;:\; (2.65 \pm 1.04 \pm 0.37) \times 10^{-5}. \tag{10.3}$$

The first uncertainty is statistical from the fit, while the second is systematic as listed in Table 9.1. The results from both measurement channels are in good agreement with each other. This allows to average the two results, leading to the overall branching fraction for the η meson of:

$$\mathcal{B}\left(B^+ \to \eta\ell^+\nu_\ell\right) = (2.83 \pm 0.55 \pm 0.34) \times 10^{-5}. \tag{10.4}$$

While the statistical uncertainties are considered as uncorrelated in the average, the systematic uncertainties are treated as fully correlated between the two results. This is a conservative approach likely to slightly overestimate the uncertainty, however the

measurement is dominated by the statistical uncertainty and the effect is very small.

Only the $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ mode of the $B^+ \to \eta'\ell^+\nu_\ell$ was usable to determine the branching fraction. The result is:

$$\mathcal{B}\left(B^+ \to \eta'\ell^+\nu_\ell\right) \;=\; (2.79 \pm 1.29 \pm 0.30) \times 10^{-5}. \tag{10.5}$$

10.2. Discussion

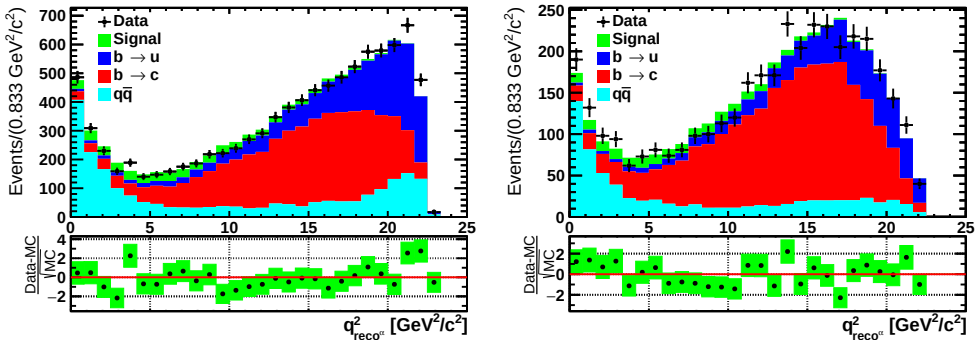
The current world averages [4] for the two measured decay channels are $\mathcal{B}\left(B^+ \to \eta\ell^+\nu_\ell\right) = (3.9 \pm 0.5) \times 10^{-5}$ and $\mathcal{B}\left(B^+ \to \eta'\ell^+\nu_\ell\right) = (2.3 \pm 0.8) \times 10^{-5}$. While the η' result is very compatible, the η result, while still compatible, is lower in this analysis. However, all the measurements of these channels come with a large uncertainties.

This analysis is the second measurement of the $B^+ \to \eta^{(\prime)}\ell^+\nu_\ell$ branching fraction at Belle. It complements the previous measurement [70]. Both measurements are compatible and independent due to the difference in methodology. The earlier measurement uses the full reconstruction approach discussed in Section 4.3.1 to achieve a higher purity, compared to this analysis which focused on a high efficiency. The $B^+ \to \eta\ell^+\nu_\ell$ decay has been measured multiple times by the Babar experiment [71–73] with similar precision. One of their measurements [73] diverges, while the others are compatible with the result in this analysis.

For the $B^+ \to \eta'\ell^+\nu_\ell$ decay, the result presented here is compatible with the newest measurement by Babar [71], which is also more precise, while their earlier result [73] has a significantly larger uncertainty. An earlier measurement by CLEO [74] is an order of magnitude larger and incompatible, but has a very large uncertainty.

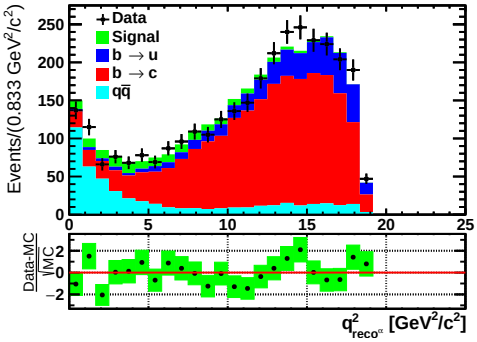
The low purity in the selected data sample of the presented analysis is directly connected to the approach keeping the entire range of q^2 intact. As can be seen in Figure 10.1, the background contribution concentrates at high values of q^2, where at the same time the signal contribution is expected to be lower with the used model. A determination of partial branching fractions in at least the lower range of q^2 was considered, but discarded as no validation method with a similar behaviour and range in q^2 was available in place of the $B^+ \to D^0\ell^+\nu_\ell$ decay used.

Several systematic uncertainties are significant. Besides detector effects like the photon identification, one of the larger effects is that assigned to the control mode discrepancy, which is the largest single contribution in $\eta \to \gamma\gamma$ and $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$. For $\eta \to \pi^+\pi^-\pi^0$ the largest systematic uncertainty comes from the K^0_L backgrounds. This suggests that in the case of better statistical uncertainties further studies to better quantify this effect and possibly to remove such backgrounds might be useful. However, within this analysis the statistical uncertainty dominates in all reconstructed modes.



(a) $\eta \to \gamma\gamma$

(b) $\eta \to \pi^+\pi^-\pi^0$



(c) $\eta \to \pi^+\pi^-\pi^0$

Figure 10.1.: Reconstructed distribution of q^2, with the MC scaled to the fitted yield.

10.3. Outlook

As this measurement is limited by the size of the dataset, significant improvements in precision can be expected after the dataset collected by the Belle II experiment exceeds that of Belle. With higher statistics and improved measurement precision, the first task would be to find selection criteria that can improve the purity of the sample, which might open up further decay channels of the $\eta^{(\prime)}$ for reconstruction. The improved detector is also expected to reduce the systematic uncertainties related to particle detection efficiencies.

Besides pure measurements of branching fractions, the expected Belle II dataset could allow further measurements. One clear candidate would be a measurement of the branching fraction in bins of q^2 to measure the decay form factors. As can be seen in Figure 10.1, especially the bins of high q^2 pose a challenge for future analyses. To measure this region with high precision against the large background contamination a novel approach might be needed.

With improvements in the theoretical descriptions, and with them the quality of the

MC, the $B^+ \to \eta^{(\prime)} \ell^+ \nu_\ell$ decays could also provide an additional source to measure the CKM matrix element $|V_{ub}|$ and help to clear up the tension between measurements.

Bibliography

[1] L. Canetti, M. Drewes, and M. Shaposhnikov. “Matter and antimatter in the universe”. In: *New Journal of Physics* 14.9 (Sept. 2012), p. 095012.

[2] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Soviet Physics Uspekhi* 34.5 (May 1991), pp. 392–393.

[3] M. Kobayashi and T. Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657.

[4] P. A. Zyla et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020), p. 083C01.

[5] D. Griffiths. *Introduction to Elementary Particles*. Wiley, Dec. 1987.

[6] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Physical Review Letters* 81.8 (Aug. 1998), pp. 1562–1567.

[7] S. L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (Feb. 1961), pp. 579–588.

[8] S. Weinberg. “A Model of Leptons”. In: *Physical Review Letters* 19.21 (Nov. 1967), pp. 1264–1266.

[9] A. Salam. “Weak and electromagnetic interactions”. In: *Selected Papers of Abdus Salam*. WORLD SCIENTIFIC, May 1994, pp. 244–254.

[10] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Physical Review Letters* 13.9 (Aug. 1964), pp. 321–323.

[11] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Physical Review Letters* 13.16 (Oct. 1964), pp. 508–509.

[12] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29.

[13] S. Chatrchyan et al. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61.

[14] E. Noether. “Invariante Variationsprobleme”. ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257.

[15] G. Lüders. “On the Equivalence of Invariance under Time Reversal and under Particle-Antiparticle Conjugation for Relativistic Field Theories”. In: *Det Kongelige Danske Videnskabernes Selskab, Matematisk-Fysiske Meddelelser* 28.5 (1954), pp. 1–17.

[16] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Physical Review* 105 (4 Feb. 1957), pp. 1413–1415.

[17] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Physical Review Letters* 13 (4 July 1964), pp. 138–140.

[18] J. P. Lees et al. “Observation of Time-Reversal Violation in the B^0 Meson System”. In: *Physical Review Letters* 109 (21 Nov. 2012), p. 211801.

[19] N. Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Physical Review Letters* 10.12 (June 1963), pp. 531–533.

[20] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Physical Review D* 2.7 (Oct. 1970), pp. 1285–1292.

[21] L.-L. Chau and W.-Y. Keung. “Comments on the Parametrization of the Kobayashi-Maskawa Matrix”. In: *Physical Review Letters* 53.19 (Nov. 1984), pp. 1802–1805.

[22] L. Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. In: *Physical Review Letters* 51.21 (Nov. 1983), pp. 1945–1947.

[23] J. Charles et al. “CP violation and the CKM matrix: assessing the impact of the asymmetric B factories”. In: *The European Physical Journal C* 41.1 (May 2005). updated results and plots available at: http://ckmfitter.in2p3.fr, pp. 1–131.

[24] Y. S. Amhis et al. “Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018”. In: *The European Physical Journal C* 81 (2021). updated results and plots available at `https://hflav.web.cern.ch/`, p. 226.

[25] S. Kurokawa and E. Kikutani. “Overview of the KEKB accelerators”. In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment* 499.1 (Feb. 2003), pp. 1–7.

[26] A. Abashian et al. “The Belle detector”. In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment* 479.1 (Feb. 2002), pp. 117–232.

[27] T. Abe. *Belle II Technical Design Report*. 2010.

[28] K. Akai, K. Furukawa, and H. Koiso. “SuperKEKB Collider”. In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment* 907 (Nov. 2018), pp. 188–199.

[29] A. J. Bevan et al. "The Physics of the B Factories". In: *European Physical Journal C: Particles and Fields* 74.11 (Nov. 2014), p. 3026.

[30] J. Brodzicka et al. "Physics Achievements from the Belle Experiment". In: *Progress of Theoretical and Experimental Physics* 2012.1 (Dec. 2012), pp. 4D001–.

[31] Z. Natkaniec et al. "Status of the Belle silicon vertex detector". In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors, and Associated Equipment* 560.1 (May 2006), pp. 1–4.

[32] E. Nakano. "Belle PID". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 494.1-3 (Nov. 2002), pp. 402–408.

[33] K. Hanagaki et al. "Electron identification in Belle". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 485.3 (June 2002), pp. 490–503.

[34] A. Abashian et al. "Muon identification in the Belle experiment at KEKB". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 491.1-2 (Sept. 2002), pp. 69–82.

[35] O. Behnke et al. *Data Analysis in High Energy Physics*. Wiley-VCH GmbH, June 26, 2013.

[36] Y. Freund and R. E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (Aug. 1997), pp. 119–139.

[37] R. J. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Computer Physics Communications* 77 (1993), pp. 219–228.

[38] M. Feindt et al. "A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 654.1 (Oct. 2011), pp. 432–440.

[39] M. Feindt and U. Kerzel. "The NeuroBayes neural network package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 559.1 (Apr. 2006), pp. 190–194.

[40] B. Aubert et al. "Measurement of the $B^0 \to \pi\ell\nu$ Form-Factor Shape and Branching Fraction, and Determination of $|V_{ub}|$ with a Loose Neutrino Reconstruction Technique". In: *Physical Review Letters* 98.9 (Feb. 2007).

[41] S. Brandt et al. "The principal axis of jets — an attempt to analyse high-energy collisions as two-body processes". In: *Physics Letters* 12.1 (Sept. 1964), pp. 57–61.

[42] G. C. Fox and S. Wolfram. "Observables for the Analysis of Event Shapes in e^+e^--Annihilation and Other Processes". In: *Physical Review Letters* 41.23 (Dec. 1978), pp. 1581–1585.

[43] S. H. Lee et al. "Evidence for $B^0 \to \pi^0\pi^0$". In: *Physical Review Letters* 91.26 (Dec. 2003).

[44] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1-2 (Apr. 2001), pp. 152–155.

[45] T. Sjöstrand, S. Mrenna, and P. Skands. "PYTHIA 6.4 physics and manual". In: *Journal of High Energy Physics* 2006.05 (May 2006), pp. 026–026.

[46] R. Brun et al. "GEANT3". In: (Sept. 1987).

[47] E. Barberio, B. van Eijk, and Z. Wąs. "Photos — a universal Monte Carlo for QED radiative corrections in decays". In: *Computer Physics Communications* 66.1 (July 1991), pp. 115–128.

[48] E. Barberio and Z. Wa s. "PHOTOS - a universal Monte Carlo for QED radiative corrections: version 2.0". In: *Computer Physics Communications* 79.2 (Apr. 1994), pp. 291–308.

[49] A. Sibidanov et al. "Study of exclusive $B \to X_u \ell \nu$ decays and extraction of $|V_{ub}|$ using full reconstruction tagging at the Belle experiment". In: *Physical Review D* 88.3 (Aug. 2013).

[50] A. Abdesselam et al. *Precise determination of the CKM matrix element $|V_{cb}|$ with $\bar{B}^0 \to D^{*+} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle*. Conference Note. 2017.

[51] B. R. Ko et al. "Search for CP Violation in the Decays $D^0 \to K^0_S P^0$". In: *Physical Review Letters* 106.21 (May 2011).

[52] B. Aubert et al. "Determination of the form factors for the decay $B^0 \to D^{*-} \ell^+ \nu_\ell$ and of the CKM matrix element $|V_{cb}|$". In: *Physical Review D* 77.3 (Feb. 2008).

[53] D. M. Asner et al. "A Study of Exclusive Charmless Semileptonic B Decays and Extraction of $|V_{ub}|$ at CLEO". In: *Physical Review D* 76.1 (July 2007).

[54] H. Ha et al. "Measurement of the decay $B^0 \to \pi^- \ell^+ \nu$ and determination of $|V_{ub}|$". In: *Physical Review D* 83.7 (Apr. 2011).

[55] R. Glattauer et al. "Measurement of the decay $B \to D \ell \nu_\ell$ in fully reconstructed events and determination of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{cb}|$". In: *Physical Review D* 93.3 (Feb. 2016).

[56] M. Neubert. "Heavy quark effective theory". In: *Subnucl. Ser.* 34 (1997). Ed. by A. Zichichi, pp. 98–165.

[57] D. Scora and N. Isgur. "Semileptonic meson decays in the quark model: An update". In: *Physical Review D* 52.5 (Sept. 1995), pp. 2783–2812.

[58] A. K. Leibovich et al. "Semileptonic B decays to excited charmed mesons". In: *Physical Review D* 57.1 (Jan. 1998), pp. 308–330.

[59] D. Côté et al. "Reweighting of the form factors in exclusive $B \to X \ell \nu_\ell$ decays". In: *The European Physical Journal C* 38.1 (Dec. 2004), pp. 105–113.

[60] P. Ball and G. W. Jones. "$B \to \eta'$ form factors in QCD". In: *Journal of High Energy Physics* 2007.08 (Aug. 2007), pp. 025–025.

[61] P. Ball and V. M. Braun. "Exclusive semileptonic and rare B meson decays in QCD". In: *Physical Review D* 58.9 (Sept. 1998), p. 094016.

[62] P. Ball and R. Zwicky. "$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules revisited". In: *Physical Review D* 71.1 (Jan. 2005), p. 014029.

[63] T. Huang et al. "Semileptonic $B(B_s, B_c)$ decays in the light-cone QCD sum rules". In: *International Journal of Modern Physics A* 23.21 (Aug. 2008), pp. 3237–3245.

[64] F. D. Fazio and M. Neubert. "$B \to X_u \ell \overline{\nu_\ell}$ decay distributions to order α_s". In: *Journal of High Energy Physics* 1999.06 (June 1999), pp. 017–017.

[65] B. Bhuyan. "High P_T Tracking Efficiency Using Partially Reconstructed D^* Decays." In: *Belle Note* 1165 (2010).

[66] H. W. Kim. "Study of High Energy Photon Detection Efficiency Using Radiative Bhabha". In: *Belle Note* 499 (2002).

[67] S. W. Lin. "Update of π^0 systematics Using Inclusive eta (78/fb)". In: *Belle Note* 645 (2003).

[68] L. Hinz. "Lepton ID efficiency correction and systematic error". In: *Belle Note* 954 (2006).

[69] S. Nishida. "Study of kaon and pion identification using inclusive D^* sample". In: *Belle Note* 779 (2005).

[70] C. Beleño et al. "Measurement of the decays $B \to \eta \ell \nu_\ell$ and $B \to \eta' \ell \nu_\ell$ in fully reconstructed events at Belle". In: *Physical Review D* 96.9 (Nov. 2017).

[71] J. P. Lees et al. "Branching fraction and form-factor shape measurements of exclusive charmless semileptonic B decays, and determination of $|V_{ub}|$". In: *Physical Review D* 86.9 (Nov. 2012).

[72] B. Aubert et al. "Measurement of the $B^+ \to \omega \ell^+ \nu$ and $B^+ \to \eta \ell^+ \nu$ branching fractions". In: *Physical Review D* 79.5 (Mar. 2009).

[73] B. Aubert et al. "Measurements of $B \to \{\pi, \eta, \eta'\}\, \ell \nu_\ell$ Branching Fractions and Determination of $|V_{ub}|$ with Semileptonically Tagged B Mesons". In: *Physical Review Letters* 101.8 (Aug. 2008).

[74] N. E. Adam et al. "Study of Exclusive Charmless Semileptonic B Decays and $|V_{ub}|$". In: *Physical Review Letters* 99.4 (July 2007).

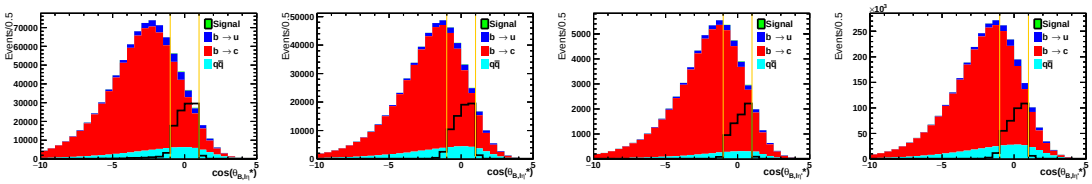
Appendices

APPENDIX A

Additional plots

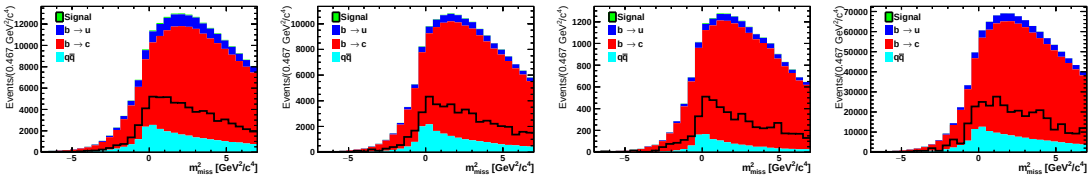
Throughout the main body of this thesis, only a subset of plots is shown at times to save space. In these cases, the remaining plots are collected here for reference.

A.1. Reconstruction



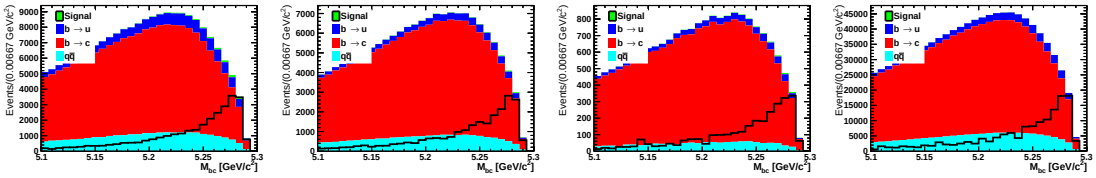
(a) $\eta \to \pi^+\pi^-\pi^0$ (b) $\eta' \to \pi^+\pi^-\eta(2\gamma)$ (c) $\eta' \to \pi^+\pi^-\eta(3\pi)$ (d) $\eta' \to \rho^0\gamma$

Figure A.1.: Reconstructed $\cos(\theta^{\star}_{B,\ell\eta^{(\prime)}})$ on MC, with all other cuts throughout Chapter 5 applied. Only events inside the yellow lines are accepted. The signal contribution is also superimposed with arbitrary scale.



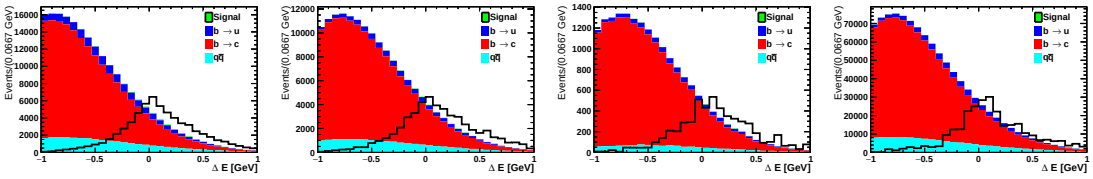
(a) $\eta \to \pi^+\pi^-\pi^0$ (b) $\eta' \to \pi^+\pi^-\eta(2\gamma)$ (c) $\eta' \to \pi^+\pi^-\eta(3\pi)$ (d) $\eta' \to \rho^0\gamma$

Figure A.2.: Accepted m^2_{miss} range on MC, with all selections throughout Chapter 5 applied. The signal contribution is also superimposed with arbitrary scale.



(a) $\eta \to \pi^+\pi^-\pi^0$ (b) $\eta' \to \pi^+\pi^-\eta(2\gamma)$ (c) $\eta' \to \pi^+\pi^-\eta(3\pi)$ (d) $\eta' \to \rho^0\gamma$

Figure A.3.: The M_{bc} distribution in the range to be used later for the fit on MC, with all selections throughout Chapter 5 applied. The signal contribution is also superimposed with arbitrary scale.



(a) $\eta \to \pi^+\pi^-\pi^0$ (b) $\eta' \to \pi^+\pi^-\eta(2\gamma)$ (c) $\eta' \to \pi^+\pi^-\eta(3\pi)$ (d) $\eta' \to \rho^0\gamma$

Figure A.4.: The ΔE distribution in the range to be used later for the fit on MC, with all selections throughout this section applied. The signal contribution is also superimposed with arbitrary scale.

A.2. BDT

A.2.1. BDT training variables

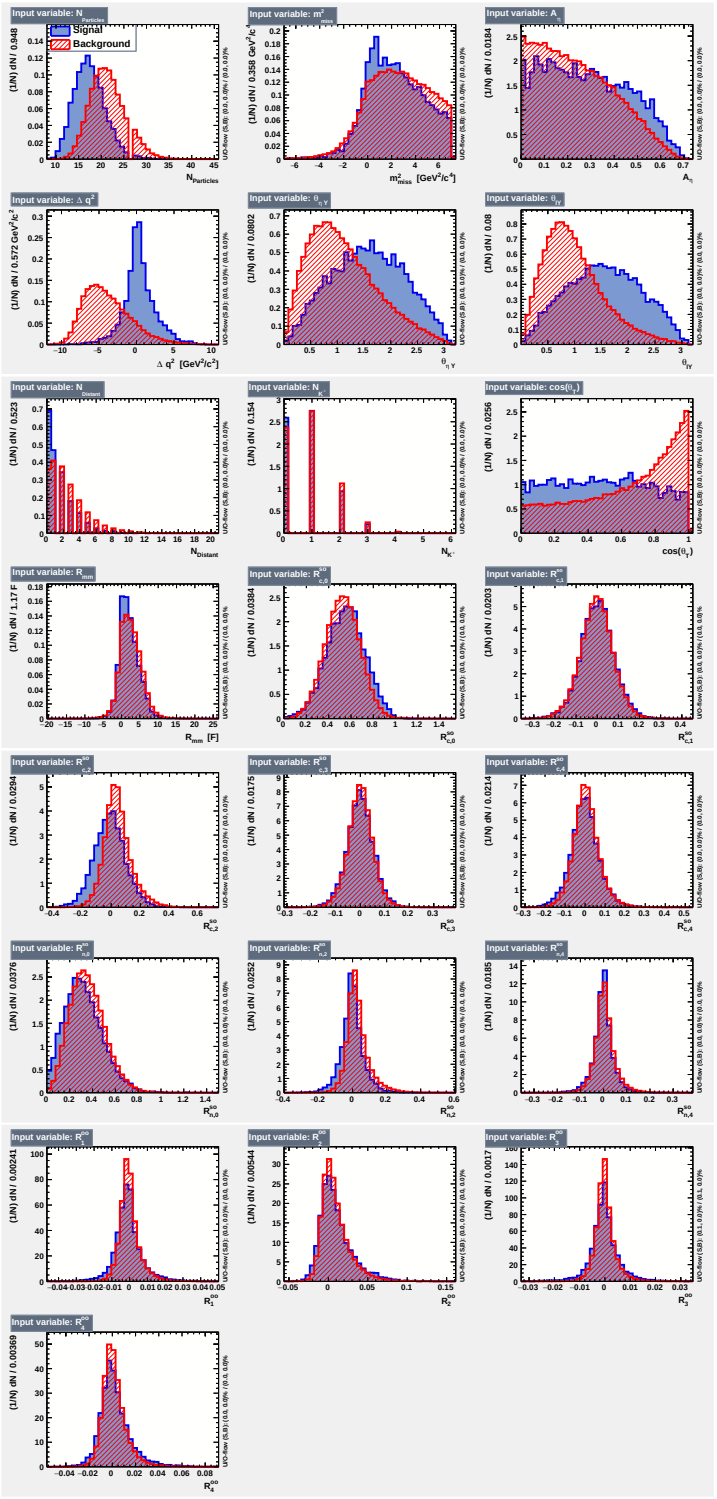


Figure A.5.: BDT training variables in the channel $\eta \to \pi^+\pi^-\pi^0$

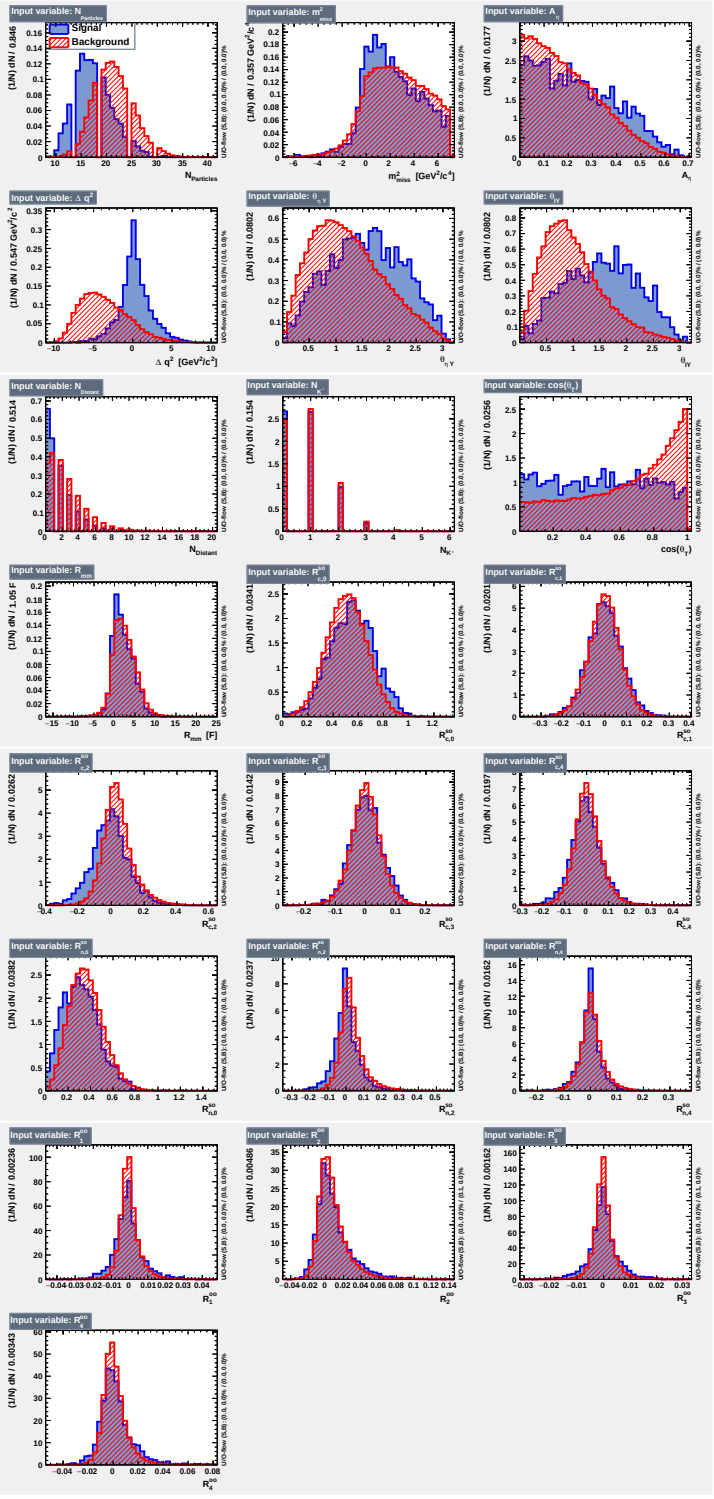


Figure A.6.: BDT training variables in the channel $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

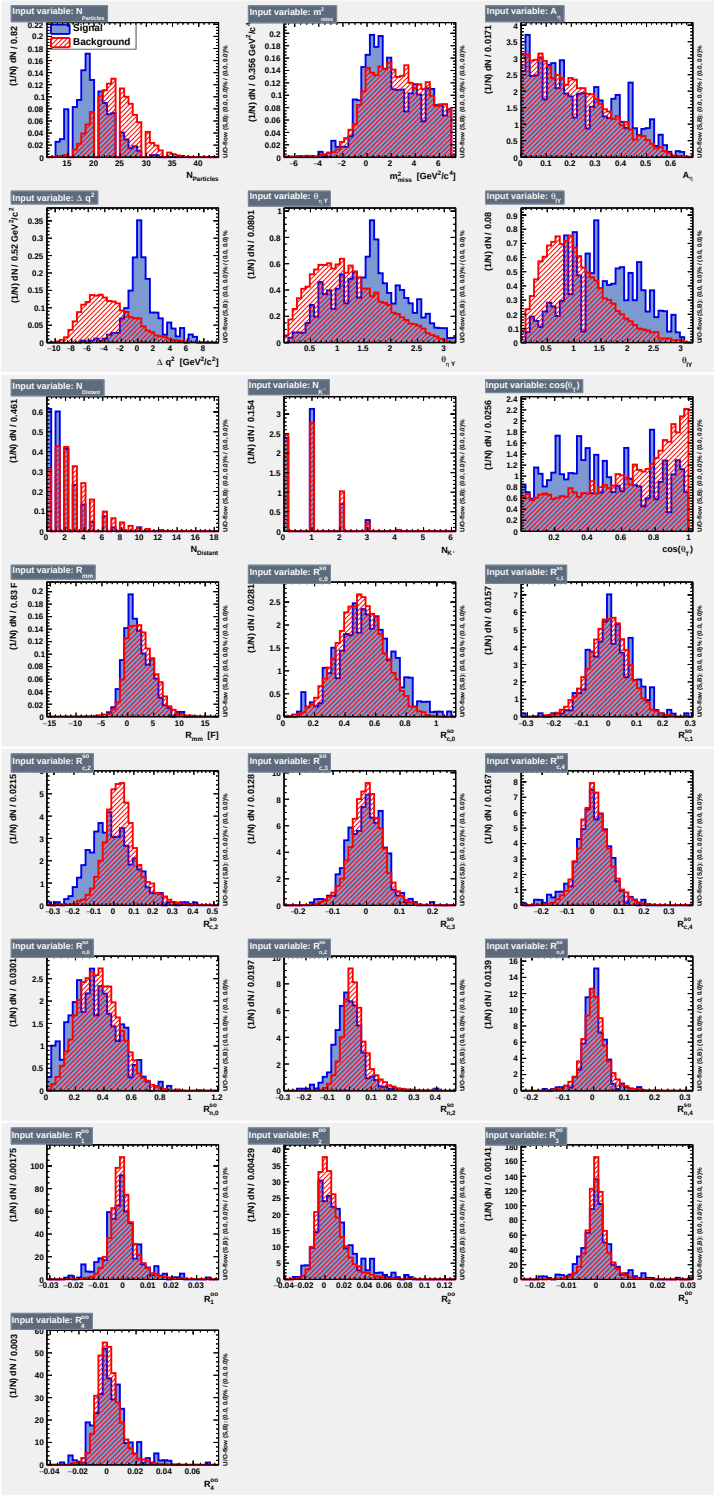


Figure A.7.: BDT training variables in the channel $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$

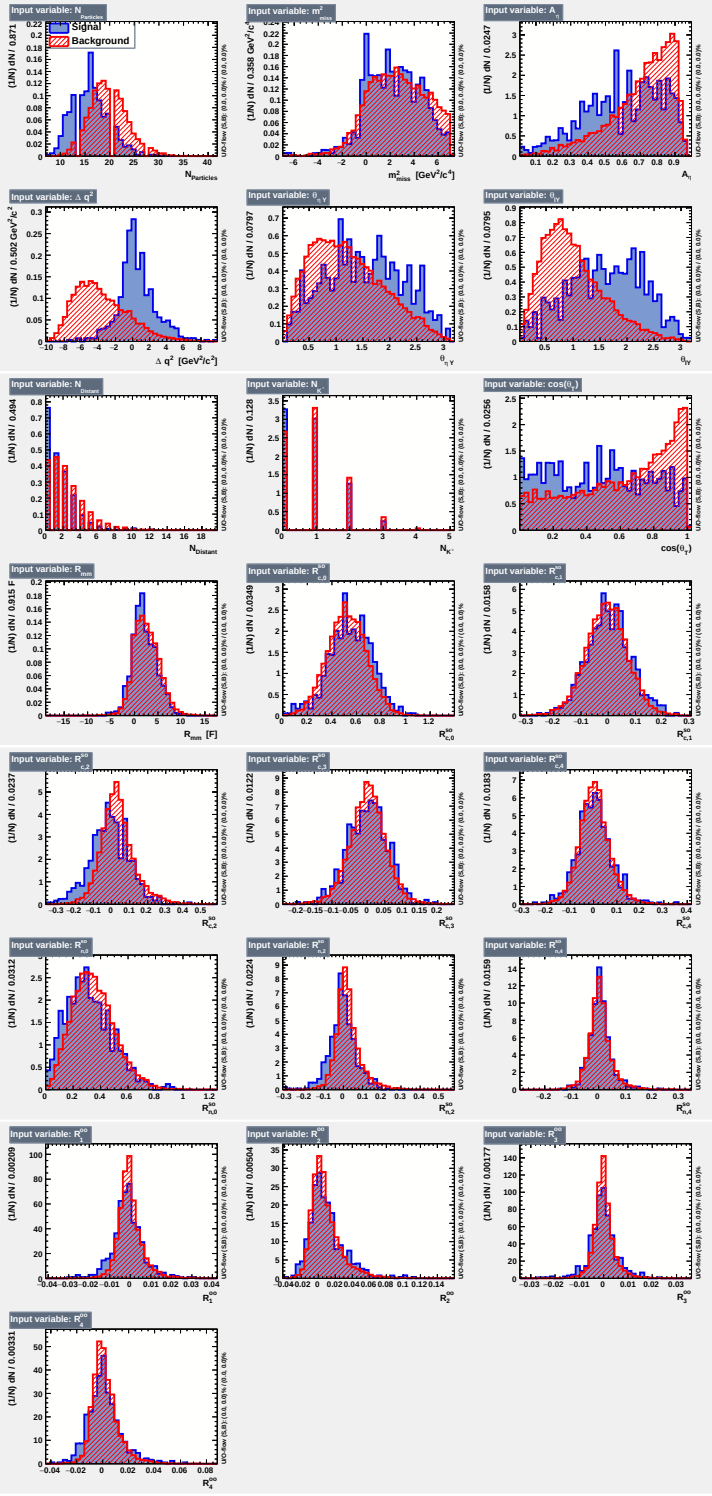
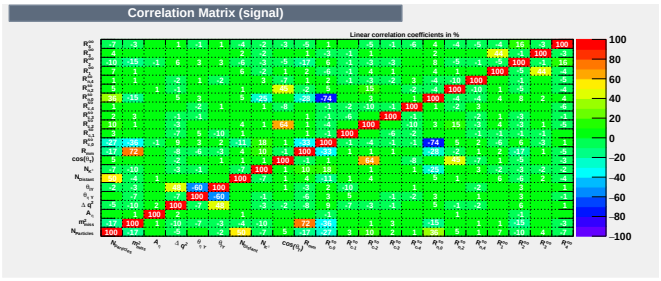
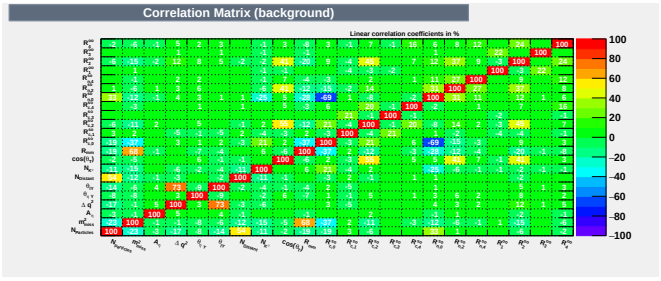


Figure A.8.: BDT training variables in the channel $\eta' \to \rho^0\gamma$

A.2.2. Correlation coefficients

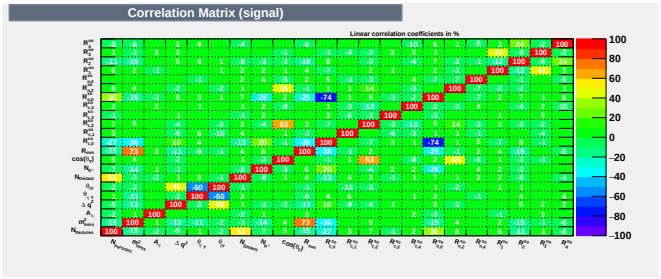


(a) Signal sample

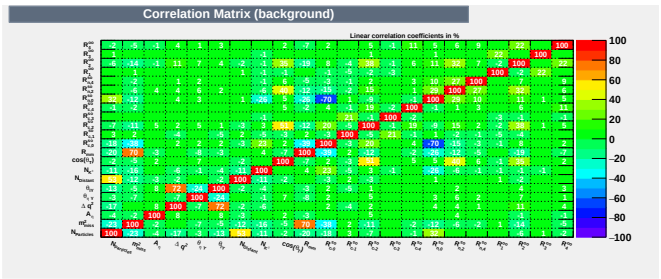


(b) Background sample

Figure A.9.: Correlations between the BDT variables in the channel $\eta \to \pi^+\pi^-\pi^0$.

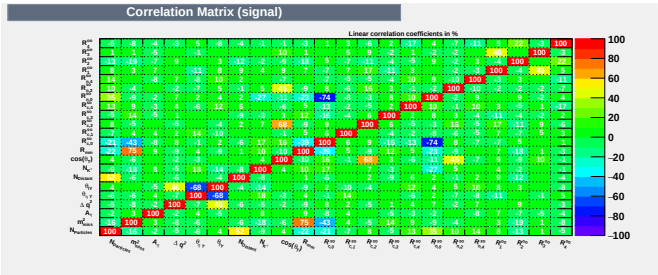


(a) Signal sample

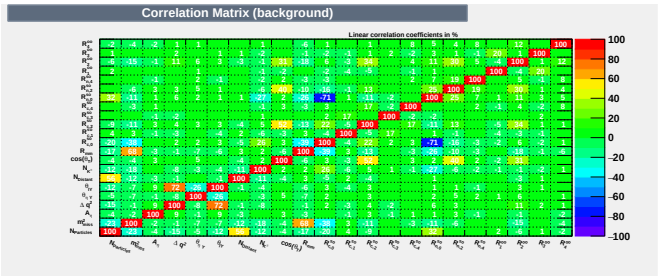


(b) Background sample

Figure A.10.: Correlations between the BDT variables in the channel $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$.

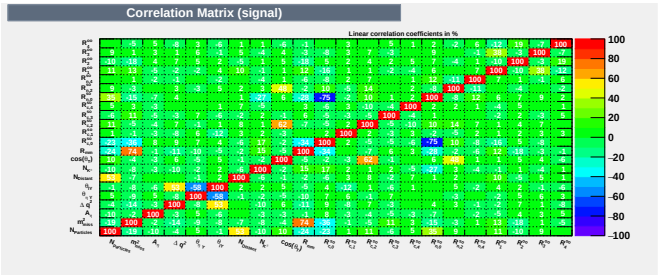


(a) Signal sample

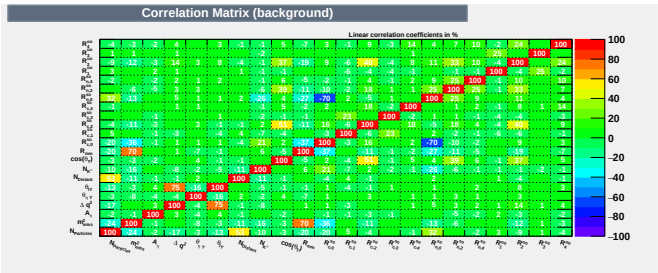


(b) Background sample

Figure A.11.: Correlations between the BDT variables in the channel $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$.



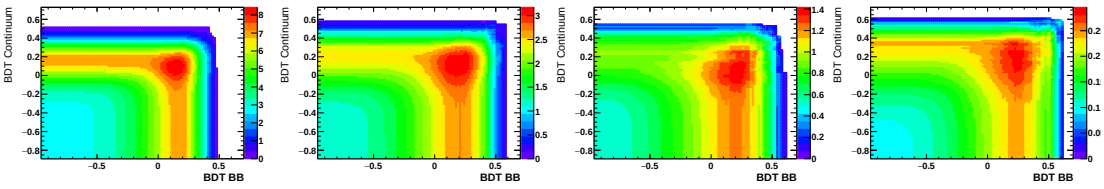
(a) Signal sample



(b) Background sample

Figure A.12.: Correlations between the BDT variables in the channel $\eta' \to \rho^0\gamma$.

A.2.3. BDT figure of merit

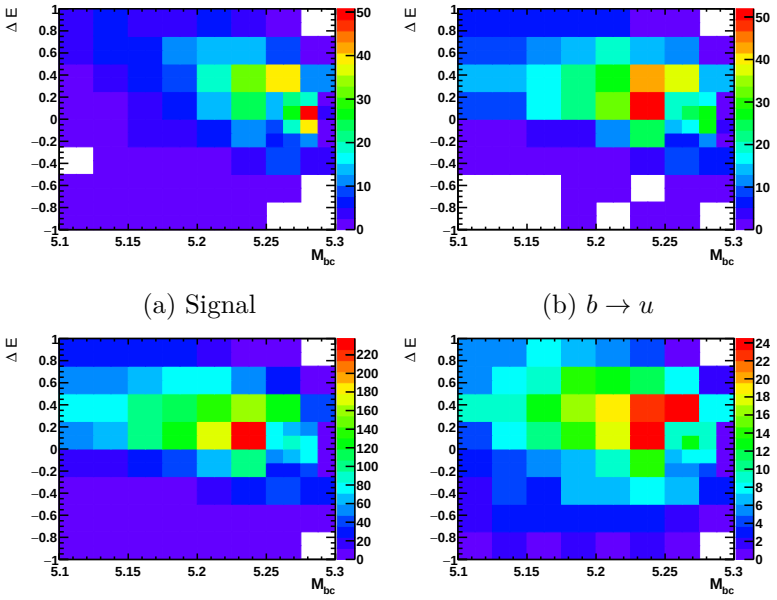


(a) $\eta \to \pi^+\pi^-\pi^0$ (b) $\eta' \to \pi^+\pi^-\eta(2\gamma)$ (c) $\eta' \to \pi^+\pi^-\eta(3\pi)$ (d) $\eta' \to \rho^0\gamma$

Figure A.13.: Figure-of-Merit as a function of the two BDT outputs.

A.3. Fit

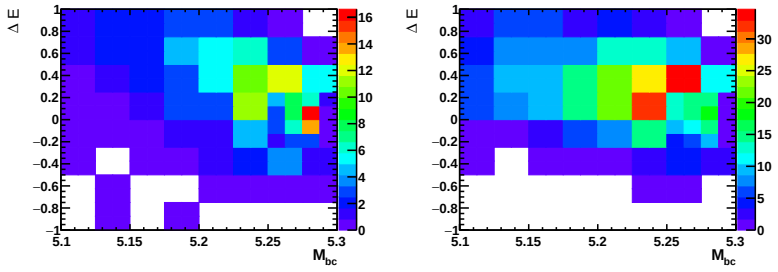
A.3.1. Fit variable distribution



(a) Signal (b) $b \to u$

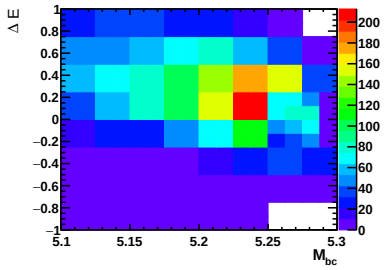
(c) $b \to c$ (d) continuum

Figure A.14.: Distribution of the four fit components in the channel $\eta \to \pi^+\pi^-\pi^0$

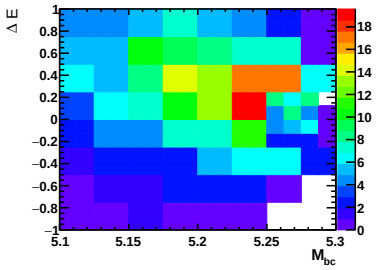


(a) Signal

(b) $b \to u$

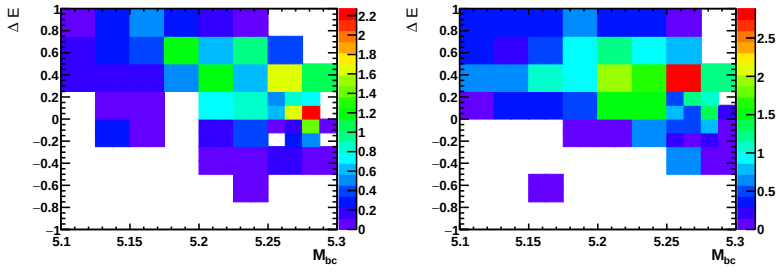


(c) $b \to c$



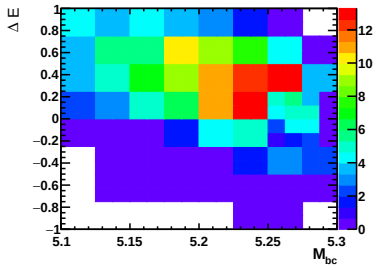
(d) continuum

Figure A.15.: Distribution of the four fit components in the channel $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$

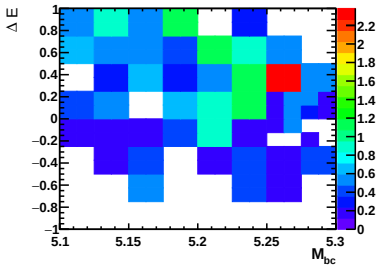


(a) Signal

(b) $b \to u$

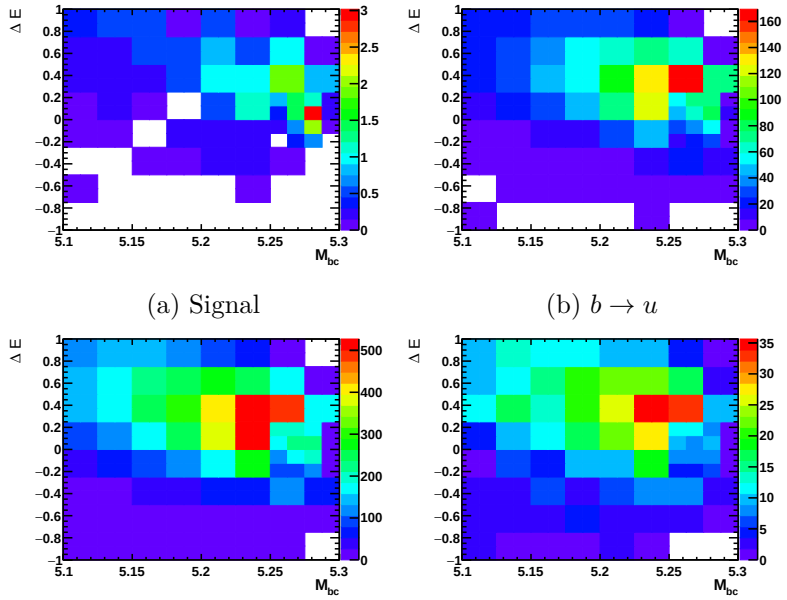


(c) $b \to c$



(d) continuum

Figure A.16.: Distribution of the four fit components in the channel $\eta' \to \pi^+\pi^-\eta(\pi^+\pi^-\pi^0)$



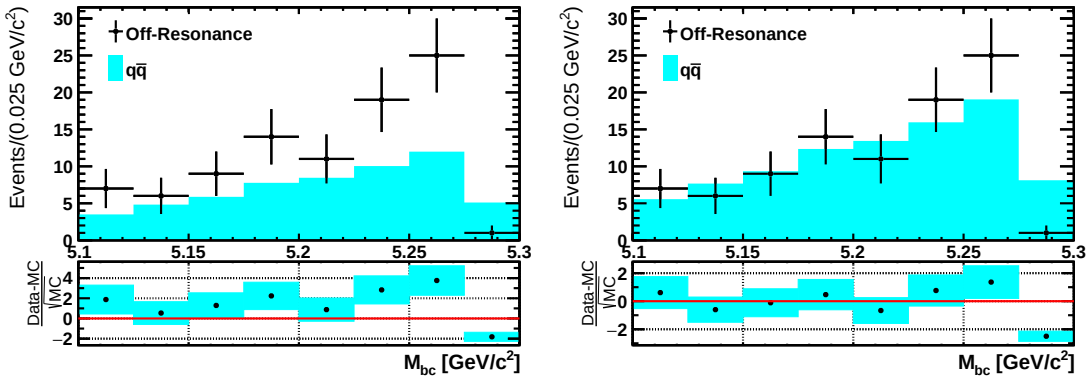
(a) Signal (b) $b \to u$

(c) $b \to c$ (d) continuum

Figure A.17.: Distribution of the four fit components in the channel $\eta' \to \rho^0\gamma$

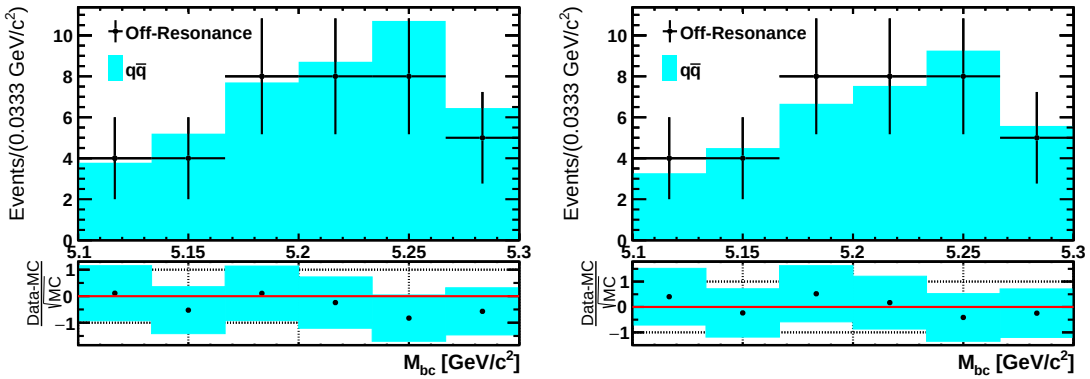
A.4. Validation

A.4.1. Continuum comparison



(a) Without rescaling (b) With rescaling

Figure A.18.: Off-resonace data compared continuum MC for M_{bc} in the $\eta \to \pi^+\pi^-\pi^0$ channel



(a) Without rescaling

(b) With rescaling

Figure A.19.: Off-resonace data compared continuum MC for M_{bc} in the $\eta' \to \pi^+\pi^-\eta(\gamma\gamma)$ channel

Acknowledgements

I would like to start by thanking my supervisor Prof. Dr. Ariane Frey for letting me work in her group. Her support through the years and her patience when things didn't go as planned were always valuable and allowed me to finish this thesis. I would also like to thank Prof. Dr. Stan Lai for being my co-referee, as well as the other members of the examination committee Dr. Baida Achkar, Prof. Dr. Laura Covi, PD Dr. Jörn Große-Knetter and Prof. Dr. Wolfram Kollatschny.

I also want to thank the other members of our group, Ulf, Harrison, Philipp, Benjamin, Noreen, Marike and César. You were all important and the many fun hours at or after work helped balance the more stressful times. Special thanks go to Ulf and Noreen for reading through this thesis.

Additional thanks go to the many colleagues I met during my visits to KEK, leading to many fruitful discussions.