

ANGULAR ANALYSIS OF $B \rightarrow K^* \ell \ell$
AND SEARCH FOR $B^+ \rightarrow K^+ \tau \tau$
AT THE BELLE EXPERIMENT

DISSERTATION
ZUR ERLANGUNG DES DOKTORGRADES
AN DER FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND NATURWISSENSCHAFTEN
FACHBEREICH PHYSIK
DER UNIVERSITÄT HAMBURG

vorgelegt von
Diplom-Physiker Simon Wehle
aus Hannover

Hamburg, 2016

Gutachter der Dissertation:	Prof. Dr. Caren Hagner Dr. Alexandre Glazov Prof. Dr. Tim Gershon	Universität Hamburg DESY University of Warwick
Gutachter der Disputation:	Prof. Dr. Caren Hagner Dr. Alexandre Glazov Prof. Dr. Peter Schleper Prof. Dr. Gudrid Moortgat-Pick Prof. Dr. Günter Sigl	Universität Hamburg DESY Universität Hamburg DESY/Universität Hamburg Universität Hamburg
Datum der Disputation:		9. September 2016

Abstract

Rare decays of B mesons are an ideal probe to search for phenomena beyond the Standard Model of particle physics, since contributions from new particles can affect the decays on the same level as Standard Model predictions. The rare decay of $B \rightarrow K^{(*)}\ell^+\ell^-$ offers the quark transition $b \rightarrow s\ell^+\ell^-$, a flavor changing neutral current which is forbidden at tree level in the Standard Model. Higher order processes such as penguin or W^+W^- box diagrams allow for these processes, leading to branching ratios of less than one in a million. Various extensions to the Standard Model predict influences of new physics, which can enhance or suppress branching ratios or lead to changes in angular distributions of the decay products. In order to produce a sufficient amount of B mesons, the so-called B -factories were designed. One of them is the asymmetric-energy e^+e^- collider KEKB with the Belle experiment, located in Tsukuba, Japan. During 1999 to 2010 the experiment recorded a total luminosity of 1 ab^{-1} of electron positron collisions.

In this thesis a comprehensive study of the flavor changing neutral current process $b \rightarrow s\ell^+\ell^-$ in B meson decays with accompanying kaons is presented in two separate analyses. All three lepton modes, e^+e^- , $\mu^+\mu^-$ and $\tau^+\tau^-$ are investigated to search for evidence of physics beyond the Standard Model. The measurements are performed using the full Belle data sample of 772×10^6 $B\bar{B}$ pairs, recorded at the $\Upsilon(4S)$ resonance energy.

The first analysis in this thesis covers the muon and electron modes in the decay of $B^0 \rightarrow K^*(892)^0\ell^+\ell^-$. To maximize signal efficiency and purity, neural networks are developed sequentially from the bottom to the top of the decay chain, transferring each time the output probability to the subsequent step such that most effective selection cuts are applied in the last stage based on all information combined. Reconstructed signal yields of both channels exceed previous B -factory results of Belle and BaBar measurements, enabling a full angular analysis in this decay with the Belle data for the first time. In total 117.6 ± 12.4 signal candidates for $B^0 \rightarrow K^*(892)^0\mu^+\mu^-$ and 69.4 ± 12.0 signal events for $B^0 \rightarrow K^*(892)^0e^+e^-$ are observed in the data. The branching ratios of both modes are extracted and found to be in agreement with previous measurements. With the combined data of both channels, the differential decay rate is extracted in three angular dimensions in five bins of q^2 , the di-lepton invariant mass squared. A series of spatial transformations is applied to reduce the number of free parameters of the differential decay rate from eight to three. With four

different transformations the fit is independently sensitive to observables P'_4, P'_5, P'_6 and P'_8 , which are optimized regarding uncertainties arising from form-factors. Altogether 20 independent three-dimensional maximum likelihood fits are performed extracting $P'_{4,5,6}$ or P'_8 , the K^* longitudinal polarization F_L and the transverse polarization asymmetry A_T . Results in the region of $q^2 < 8 \text{ GeV}^2/c^4$ are compared with Standard Model predictions and overall agreement is observed. One measurement is found to deviate by $\sim 2.1\sigma$ from the predicted value in the same direction and in the same q^2 region where the LHCb collaboration reported the so called P'_5 anomaly [1, 2].

The second analysis in this thesis is dedicated to the τ mode of $b \rightarrow s\ell^+\ell^-$ in the search for $B^+ \rightarrow K^+\tau^+\tau^-$. This mode is particularly interesting as new particles could couple to the high mass of τ leptons stronger than to e and μ . However, due to several neutrinos being present in the final state, it is difficult to find. The full reconstruction technique is used, which is unique for e^+e^- colliders and makes it possible to find signatures of the decay. On the control channel, $B^+ \rightarrow K^+\tau^+\tau^-(K_S)$, it is shown that the reconstruction and background suppression methods work as expected and that they deliver consistent results in both data and Monte Carlo. This analysis demonstrates using simulated events and control channels on the Belle dataset, that the current upper limit can be improved by more than one order of magnitude compared to the current value to

$$\mathcal{B}^{Projected}(B^+ \rightarrow K^+\tau^+\tau^-) < 3.17 \times 10^{-4}, \quad (1)$$

at 90% confidence level including systematic uncertainties determined by an extensive study of all the sources. This limit will come close to the prediction of models developed in the context of Minimal Lepton Flavor Violation [3].

Zusammenfassung

Für die Erforschung der Physik jenseits des Standardmodells der Teilchenphysik sind seltene Zerfälle von B -Mesonen von großer Bedeutung. Sie eignen sich besonders gut, um Vorhersagen des Standardmodells zu testen und Hinweise für Neue Physik zu finden. Neue Physik kann bei seltenen Zerfällen in der gleichen Größenordnung wie Standardmodell-Prozesse auftreten und damit einen signifikanten Beitrag zum Signal darstellen. Der seltene Zerfall $B^0 \rightarrow K^{*}(892)^0 \ell^+ \ell^-$ erfolgt über den Flavor Changing Neutral Current $b \rightarrow s \ell^+ \ell^-$. Diese Quark-Zerfälle sind im Standardmodell nicht direkt im Zerfallsdiagramm möglich, sondern erfolgen nur durch Prozesse höherer Ordnung wie Pinguin-Diagramme oder $W^+ W^-$ Box-Diagramme, wodurch sie stark unterdrückt werden und daher nur sehr selten auftreten. Die Verzweungsverhältnisse sind dabei in der Größenordnung oder kleiner als eins zu einer Million. Erweiterungen des Standardmodells prognostizieren verschiedene Einflüsse Neuer Physik, die mit dem Zerfall im Standardmodell interferieren und zu veränderten Verzweungsverhältnissen oder zu Änderungen der Winkelverteilung der Zerfallsprodukte führen können. Der Zerfall von $B \rightarrow K^{(*)} \ell^+ \ell^-$ bietet vier unabhängige kinematische Messgrößen, aus denen eine Vielzahl an weiteren Beobachtungsgrößen bestimmt werden kann, um die Vorhersagen des Standardmodells zu überprüfen. Das Belle Experiment am japanischen Teilchenbeschleuniger KEKB in Tsukuba ist als sogenannte « B -Fabrik» ausgelegt und hat zwischen 1999 und 2010 Daten aufgenommen. An dem asymmetrischen Elektron-Positron-Beschleuniger werden die Strahlen nahe der $\Upsilon(4S)$ Resonanzenergie bei einem Lorentz-Boost von $\beta\gamma = 0.425$ zur Kollision gebracht. Für die Analysen in dieser Arbeit wird der volle Belle-Datensatz von 772×10^6 B -Mesonenpaaren verwendet.

Im Rahmen dieser Dissertation wird der Zerfall $b \rightarrow s \ell^+ \ell^-$ anhand der Messung der Winkelverteilung von $B^0 \rightarrow K^{*}(892)^0 \ell^+ \ell^-$ und der Bestimmung des oberen Limits des Verzweungsverhältnisses von $B^+ \rightarrow K^+ \tau^+ \tau^-$ untersucht. In der ersten Analyse werden sechs Zustandsgrößen in fünf Intervallen der invarianten Masse des Leptonen-Paares q^2 in einer Winkelanalyse gemessen. Insgesamt werden 20 unabhängige Messungen durchgeführt, in denen die zu bestimmenden Größen mit einem dreidimensionalen Maximum-Likelihood-Fit aus den Daten gewonnen werden. 2013 hat das LHCb Experiment in einer dieser Größen eine Abweichung von der Standardmodell-Vorhersage von 3.7σ gefunden [1]. In dem hier präsentierten Resultat wird eine Abweichung im gleichen kinematischen Bereich bestätigt. Insgesamt kann eine solche

Abweichung entweder durch Physik jenseits des Standardmodells erklärt werden oder durch einen unerwartet hohen Einfluss der hadronischen Charmonium-Resonanz auf die Genauigkeit der Theorievorhersage [2].

In der zweiten Analyse dieser Arbeit wird der τ -Kanal von $b \rightarrow s\ell^+\ell^-$ anhand der Bestimmung des oberen Limits des Verzweigungsverhältnisses von $B^+ \rightarrow K^+\tau^+\tau^-$ untersucht. Bei dem Zerfall entstehen 2 bis 4 Neutrinos, die nicht detektiert werden können. Um dennoch Signalkandidaten zu finden, wird die Methode der vollständigen Rekonstruktion verwendet, die einzigartig an e^+e^- Beschleunigern ist. Es konnte durch simulierte Ereignisse demonstriert werden, dass die Erwartung des oberen Limits des Verzweigungsverhältnisses bei Belle mit diesen Methoden auf $\mathcal{B}(B^+ \rightarrow K^+\tau^+\tau^-) < 3.17 \times 10^{-4}$ gesetzt werden kann – damit wäre eine Messung um etwa eine Größenordnung präziser als die aktuell beste Messung.

Contents

I	Introduction	1
1	Introduction to $B \rightarrow K^{(*)}\ell^+\ell^-$	5
1.1	Overview of the Standard Model	5
1.1.1	The Electroweak Interaction	6
1.1.2	Flavor Changing Neutral Currents	9
1.1.3	Theoretical Framework of Effective Hamiltonians	9
1.1.4	Effective Hamiltonian for $b \rightarrow s\ell^+\ell^-$	10
1.1.5	Decay Rates	11
1.2	The Differential Decay Rate	11
1.2.1	Decay Topology	11
1.2.2	Definition of the Differential Decay Rate	13
1.2.3	Definition of Observables	15
1.2.4	Increasing the Statistical Sensitivity	16
1.3	Overview of Recent Measurements	17
1.4	Beyond the Standard Model	17
2	Experimental Setup	21
2.1	Fundamentals of Particle Collisions	21
2.2	The Belle Experiment	22
2.3	The Belle Detector	23
2.3.1	Triggers and Data Acquisition	26
2.3.2	Particle Identification	26
2.3.3	Example Decay	27
2.4	Luminosity and Data Samples	27
2.5	Simulated Datasets	27
3	Analysis Techniques	31
3.1	The Concept of Probability	31
3.1.1	Probability Density Distributions	33
3.2	Dependencies in Datasets	33
3.2.1	Flat Distributions	34
3.2.2	Analyzing Correlation	36

3.3	Parameter Estimation	38
3.3.1	Maximum Likelihood Method	38
3.4	Classification Methods	39
3.4.1	Cut based Methods	40
3.4.2	Fisher's linear discriminant	41
3.4.3	Neural Networks	41
3.4.4	The NeuroBayes Framework	43
3.4.5	Boosted Decision Trees	44
3.5	Full Reconstruction Technique	45
 II Angular Analysis of $B \rightarrow K^* \ell \ell$		49
 4 Reconstruction of $B \rightarrow K^{(*)} \ell^+ \ell^-$		51
4.1	Analysis Overview	51
4.1.1	Data	51
4.2	Particle Selection	52
4.3	Event Selection	53
4.4	Background Suppression Strategy	55
4.4.1	Sources of Backgrounds	56
4.5	Neural Network Trainings	56
4.5.1	The Prior Probability	57
4.5.2	Primary Particles - Stage One	57
4.5.3	Training of K^* - Stage Two	59
4.5.4	B Meson Trainings - Final Stage	62
4.6	Continuum Suppression	63
4.6.1	Fox-Wolfram Moment R_2	63
4.6.2	Super Fox-Wolfram Moments	64
4.7	Best Candidate Selection	64
4.8	Peaking Backgrounds	64
4.8.1	Rare B Decays	65
4.9	Evaluation of the Reconstruction Performance	65
4.9.1	Best Cut Estimation	65
4.9.2	Efficiency	69
 5 Signal Yields and Data Evaluation		73
5.1	Signal Yields	73
5.2	Branching Ratios and Efficiency Evaluation	74
5.2.1	Extraction of $\mathcal{B}(B^0 \rightarrow K^{*0}(892) \ell^+ \ell^-)$	75
5.2.2	Extraction of $\mathcal{B}(B \rightarrow K^{*0} J/\psi)$ and $\mathcal{B}(B^0 \rightarrow K^{*0} \psi(2S))$	77
5.3	Data and Monte Carlo Comparison	79
5.3.1	Variables of the Neural Network	79
5.3.2	Angular Reconstruction Efficiency	79

6	Angular Analysis	85
6.1	Fit Components	85
6.1.1	Signal	85
6.1.2	Background	88
6.2	Fit Procedure	89
6.3	Monte Carlo Toy Study	91
6.3.1	Background Parametrization	91
6.3.2	Simultaneous Fitting	92
6.3.3	Neural Network Cut Optimization	93
6.3.4	Fixing A_T or F_L	94
6.3.5	Estimating the Sensitivity in the Muon Mode	94
6.3.6	Conclusions from Toy Studies	94
6.4	Acceptance and Efficiency	97
7	Systematic Study	103
7.1	Data to Monte Carlo Differences	103
7.2	Fit Bias	105
7.3	Efficiency Correction	106
7.4	Peaking Backgrounds	107
7.5	Other Sources of Systematic Uncertainties	107
7.5.1	Cross-Feed	107
7.5.2	K-Star S-Wave Component	107
7.5.3	CP Asymmetries	108
7.6	Summary	109
8	Results and Discussions	111
8.1	Angular Analysis Results	111
8.1.1	Standard Model Predictions	111
8.1.2	Extraction	111
8.2	Discussion and Outlook	112
III	Search for $B^+ \rightarrow K^+ \tau^+ \tau^-$	123
9	Motivation and Analysis Overview	125
10	Reconstruction and Analysis	127
10.1	Event Selection	127
10.1.1	Best Candidate Selection	128
10.2	Pre-Selection Cuts	128
10.2.1	Background Composition	128
10.3	Background Suppression	129
11	Limit Estimation	135
11.1	Branching Ratio	135

11.2	Limit Calculation	136
11.3	Limit Optimization	136
12	Systematic Uncertainties	139
12.1	Data MC Comparison	139
12.1.1	Tag Side Correction	139
12.1.2	E_{ECL} Sideband	141
12.1.3	EECL Distribution	141
12.1.4	Control Channel on Data	141
12.2	Other Uncertainties	142
12.2.1	Number of BB Pairs	142
12.2.2	Particle Identification	142
12.2.3	Tracking	143
12.2.4	Generator	143
12.2.5	Statistical Uncertainties from Simulation	143
12.3	Summary	143
13	Results	145
IV	Conclusion	147
14	Conclusion	149
A	Appendix $B \rightarrow K^{(*)}\ell^+\ell^-$	167
A.1	Data – Monte Carlo Comparison	167
A.2	Fit Projections	167
B	Appendix $B^+ \rightarrow K^+\tau^+\tau^-$	191
B.1	Variables for the Classifier	191
B.2	Data MC Comparison	191

Part I

Introduction

Preamble

From the beginning of documented history mankind has tried to understand its surrounding nature. Over the last centuries, vast advancements were made in mathematics, physics and technology. The idea arose to find a universal underlying theory, capable of explaining all processes in our universe. The current knowledge about the structure of matter and its interactions is formulated in the Standard Model of particle physics, which was formed during the past decades based on experimental observations and their implications. It is capable of explaining the subatomic phenomena within the boundary of the field of experimental particle physics to a great precision. However, mainly from cosmological observations, it has become clear that the Standard Model is incomplete. Up to this point, we are still far away from a theory, which is capable of explaining everything in our universe. Many alternative theories deliver extensions to the Standard Model, in most cases by introducing new particles none of which could be found in experiments so far. Actually, particle physics does not suffer particularly from a lack of competing theories but rather from a lack of uncharted territories where new physics may be found. In the worst case, effects of new physics enter at an energy scale which is far beyond reach for humanity – making it impossible to further explore the fundamental principals of our universe.

There are two frontiers searching for new physics. The high energy frontier like experiments at the Large Hadron Collider probe the structure of matter at energies allowing for the direct production of exotic heavy particles. The other frontier looks for indirect effects of new physics in high precision measurements, where new particles can appear in intermediated steps of the decays. One of the leading roles in the high precision frontier is taken by the Belle experiment, dedicated specifically to measurements related to decays of B mesons. The experiment is located in Tsukuba, Japan, at the KEKB accelerator, and recorded a data sample containing about 772 million $B\bar{B}$ pairs during its operation from 1999 to 2010. It is designed as a so-called B -factory, an e^+e^- collider operating at the $\Upsilon(4S)$ resonance energy, allowing for the pairwise production of B mesons in a clean experimental environment.

Recent observations in particle physics exhibit a variety of tensions with the Standard Model – although none of these observations provides strong indications towards one specific physics scenario, but in their entirety they offer a hint that new physics might be within the reach of the present experiments. One of the most promising areas to find deviations from the Standard Model are flavor changing neutral currents like the

quark transition of $b \rightarrow s\ell^+\ell^-$. In the Standard Model, these decays cannot occur directly, but only as higher order processes involving heavy bosons leading to a strong suppression. In this case, relative contributions of any effects from new particles or new phenomena increases and can be identified and examined in high precision.

In this thesis, the analysis of the $b \rightarrow s\ell^+\ell^-$ transition in decays of $B \rightarrow K^{(*)}\ell^+\ell^-$ at the Belle experiment is presented in two separated studies: an analysis of the angular distributions of in the decay of $B^0 \rightarrow K^*(892)^0\ell^+\ell^-$ in the ee and $\mu\mu$ final states and the search for $B^+ \rightarrow K^+\tau^+\tau^-$. During the last years, several observations were made in this transition, which might hint to new physics. For example the LHCb collaboration reported a discrepancy in the angular distribution of the decay $B^0 \rightarrow K^*(892)^0\mu^+\mu^-$ with a significance of 3.4σ with respect to the Standard Model prediction [2]. In part II of this thesis, this decay is analyzed with the Belle dataset. The extraction of this process is an experimental challenge as the decay occurs only in about one of a million B decays – in fact it is among the smallest B meson branching fractions ever observed at B -factories [4]. With the use of consecutive neural networks and advanced data analysis techniques, it is demonstrated that enough statistics can be obtained to perform a full angular analysis in this channel for the first time at B -factories. In contrast to the LHCb measurement also the di-electron channel is used to access the three dimensional differential decay rate. Another interesting tension related to this decay was discovered in the ratio of the branching fractions of $B^+ \rightarrow K^+\mu^+\mu^-$ and $B^+ \rightarrow K^+e^+e^-$. Whereas the Standard Model predicts lepton-universality, thus a ratio of nearly one, LHCb observed a 2.6σ deviation from this value [5]. In this light, the tau decay mode is particularly interesting, because effects from new physics might additionally couple to the high mass of τ leptons stronger than to e and μ . In the second analysis of this thesis, in part III, a search for $B^+ \rightarrow K^+\tau^+\tau^-$ is performed. The final state of the decay contains at least 2 neutrinos, which cannot be seen in the detectors making this decay particularly difficult to analyze. The clean experimental environment at Belle, however, provides the possibility to reconstruct the entire event and find a signal of the decay using the so-called full reconstruction technique.

Contrary to searches in the high energy frontier, new discoveries in the flavor sector will most likely arise through a conjunction of results throughout a variety of different decays. For example by adding a single new scalar particle to the Standard Model, a TeV-scale leptoquark, it is demonstrated in Ref. [6] that most tensions in the flavor sector could be explained. In particular, $b \rightarrow s\ell^+\ell^-$ transitions covered in this thesis are affected by this theory. Additionally, excesses in the decay rate $B \rightarrow D^*\tau\bar{\nu}$ which are found by BaBar, Belle and LHCb deviating by 3.5σ from the Standard Model [7–9] could be explained. An interesting connection between these tensions and a possible resonance in the di-photon spectrum in measurements by ATLAS and CMS experiments at 750 GeV is drawn in Ref. [10]. This would be a perfect paradigm where precision and high energy frontier work together to determine the nature of this possible manifestations of new physics. The following analyses are dedicated to add further pieces to this overall picture.

1. Introduction to $B \rightarrow K^{(*)}\ell^+\ell^-$

In this thesis the decay $B \rightarrow K^{(*)}\ell^+\ell^-$ is analyzed experimentally. An angular analysis of $B^0 \rightarrow K^{*}(892)^0\ell^+\ell^-$ is performed in part II and a search for $B^+ \rightarrow K^+\tau^+\tau^-$ is described in part III. Beforehand, this chapter provides a theoretical introduction to the decays. First, a brief overview of the Standard Model of particle physics is presented. The theoretical description of the decay is detailed and observables that can be measured experimentally are presented. Finally, the need for physics beyond the Standard Model is discussed and implications for the measured quantities are investigated for a variety of new physics scenarios.

1.1 Overview of the Standard Model

The Standard Model (SM) of particle physics is a relativistic quantum field theory describing the interplay between elementary particles and the electromagnetic, weak and strong force. It is the agreed common theory of particle physics and result of several decades of measurements, discoveries and predictions. The theory aggregates the unification of the electromagnetic and weak force to the electroweak interaction by the Glashow-Weinberg-Salam (GWS) [11, 12] model and the Quantum Chromo Dynamics (QCD) describing the strong force. The SM is able to describe observable phenomena in particle collisions by large detail but is not a «theory of everything», as particularly gravitation is not included in this theory, although being a fundamental force.

The SM is formulated as a quantum gauge theory with three local groups

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y, \quad (1.1)$$

where C indicates the color charge of the strong interaction, L is the left chirality of the weak interaction and Y the hypercharge of the electroweak force. Within

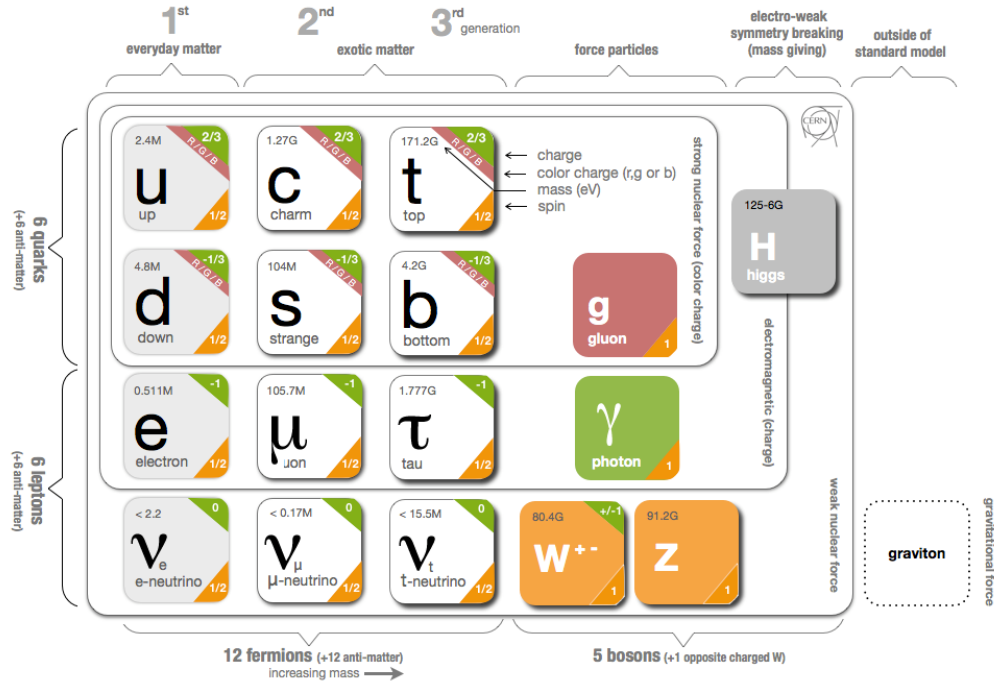


Figure 1.1: Constituents of the Standard Model. Credit: Cern.

the SM elementary particles are considered to be point-like, with no intrinsic sub-structure. There are two kinds of particles: matter- and forces-particles – fermions and bosons respectively. Interactions are mediated by bosonic particles, the gauge bosons. They couple to the individual charge of the gauge groups, which can involve self-coupling if they carry charge themselves. The weak force is mediated by the exchange of heavy $W^\pm$ and Z^0 bosons, the strong and electromagnetic force by massless gluons and photons respectively. The fermions are divided into leptons and quarks. Each of them appears in three generations of tuples, where in each tuple, the partners are separated by one unit of electrical charge. The generations share the same physical proprieties and differ only by increasing mass. Quarks carry color, weak and electromagnetic charge, making them interact via all three forces. Leptons participate in weak interactions and, except for the neural neutrinos, additionally in electromagnetic processes. An overview of the particles of the SM is depicted in fig. 1.1.

1.1.1 The Electroweak Interaction

For the electroweak interaction, fermions of the SM get arranged into doublet and singlet states of the weak isospin, where left-handed particles appear in doublets and right-handed particles in singlets. Fermions of the Standard Model with their

corresponding quantum numbers regarding the electroweak force are shown in table 1.1. The gauge symmetry of the left-handed isospin group is $SU(2)_L$ and $U(1)_Y$ is the symmetry group of the weak hypercharge Y , which is defined by the Gell-Mann–Nishijima formula as

$$Y = 2 \cdot (Q - T_3), \quad (1.2)$$

where Q is the electric charge of a particle and T_3 the third component of the weak isospin. The $W^\pm$ bosons allow for transitions between the doublets as they carry $T_3 = \pm 1$, thus only left-handed particles are affected. The GWS theory describes the electroweak interactions as spontaneously broken symmetry groups

$$SU(2)_L \otimes U(1)_Y \xrightarrow{SSB} U(1)_{EM}. \quad (1.3)$$

This spontaneous symmetry breaking (SSB) is introduced by the Higgs mechanism and causes the four primordial massless gauge bosons to mix into the massive weak mediators $W^\pm, Z^0$, the massless γ and it is responsible for the masses of the fermions [13].

Table 1.1: Fermions of the Standard Model with their corresponding quantum numbers regarding the electroweak force. The particles are divided into singlets and doublets of the weak isospin, denoted with R for right-handed and L for left-handed respectively. T is the weak isospin with its third component T_3 , Y is the weak hypercharge and Q the electrical charge.

(1)	(2)	(3)	T	T_3	Y	Q
$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$\frac{1}{2}$	$+1/2$ $-1/2$	-1	0 -1
e_R	μ_R	τ_R	0	0	-2	-1
$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$\frac{1}{2}$	$+1/2$ $-1/2$	$1/3$	$+2/3$ $-1/3$
u_R	c_R	t_R	0	0	$+4/3$	$+2/3$
d_R	s_R	b_R	0	0	$-2/3$	$-1/3$

In weak currents, transitions between the generations are not possible. For quarks however decays like $b \rightarrow c + W^-$ can proceed. This is possible because the quark mass eigenstates appear as a rotation of the weak eigenstates of the quarks. The rotation is usually described for the down-type quarks, denoting the flavor eigenstate as q' , which is a superposition of the mass eigenstates. This principal was first introduced by Cabibbo [14] for two generations of quarks and generalized by Kobayashi and Maskawa [15] to three generations¹. Hence, the CKM matrix describes the rotation

¹In 2008 Makoto Kobayashi and Toshihide Maskawa were awarded the Nobel Prize in Physics, where the Belle and BaBar experiment were explicitly mentioned for the effort of proving their predictions.

by

$$\begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{weak}} = \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{mass}}. \quad (1.4)$$

The CKM matrix is complex and unitary if there are no more than three generations of quarks. It offers four free parameters: three mixing angles and one complex phase. The elements V_{ij} appear at each vertex of the charged current, where i and j are the flavors of the corresponding quarks. In the $\mathcal{CP}$ -conjugated processes, the complex-conjugated element V_{ij}^* accounts for the coupling. In the SM the complex phase leads to $V_{ij} \neq V_{ij}^*$ and provides the only known mechanism that leads to $\mathcal{CP}$ violation. The diagonal elements of the matrix correspond to the flavor transition within one generation, where the magnitude for this is close to one. Off diagonal elements are suppressed: $|V_{us}|$ and $|V_{cd}|$ are about 0.22, $|V_{cb}|$ and $|V_{ts}|$ are of the order 4×10^{-2} and $|V_{td}|$ and $|V_{ub}|$ of the order 5×10^{-3} . As a consequence the quark transition $b \rightarrow c$ is suppressed by the CKM mechanism, which leads to the relatively long lifetime of B mesons. The unitarity of the CKM matrix can be expressed by

$$\sum_i V_{ij} V_{ik}^* \quad \text{and} \quad \sum_j V_{ij} V_{ik}^* = \delta_{jk} \quad (1.5)$$

for all generations i, j . The relations can be visualized as triangles in the complex plane, where a particularly interesting relation is given by

$$V_{ud} V_{ub}^* + V_{cd} V_{cb}^* + V_{td} V_{tb}^* = \delta_{db} = 0, \quad (1.6)$$

as many of the elements can be extracted from measurements related to B decays. From this relation one can construct the so-called Unitary Triangle by dividing each side by the experimentally best known term $V_{cd} V_{cb}^*$. The angles in this triangle are defined as

$$\alpha = \phi_2 = \arg \left(-\frac{V_{td} V_{tb}^*}{V_{ud} V_{ub}^*} \right), \quad (1.7)$$

$$\beta = \phi_1 = \arg \left(-\frac{V_{cd} V_{cb}^*}{V_{td} V_{tb}^*} \right), \quad (1.8)$$

$$\gamma = \phi_3 = \arg \left(-\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*} \right). \quad (1.9)$$

The current constraints on the Unitary Triangle are depicted in fig. 1.2 in the $\bar{\rho}\bar{\eta}$ -plane, where $\bar{\rho}$ and $\bar{\eta}$ are defined as

$$\bar{\rho} + i\bar{\eta} = -\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*}. \quad (1.10)$$

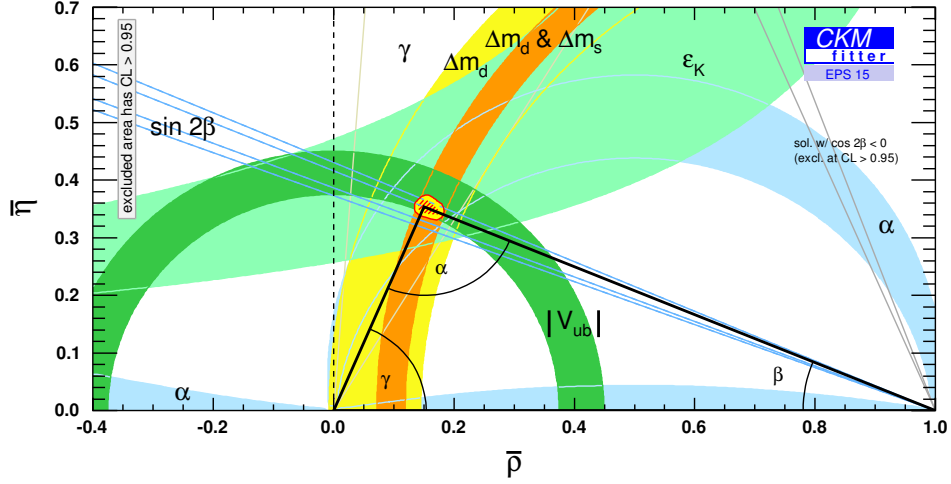
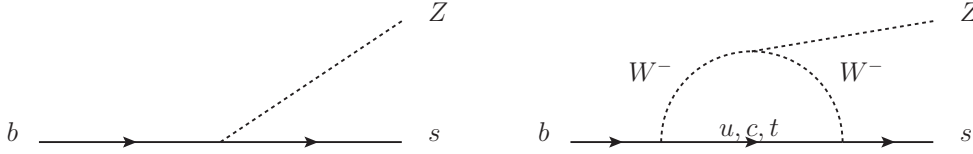


Figure 1.2: Constraints on the Unitary Triangle by the CKMfitter group [16].


 Figure 1.3: Example Feynman diagrams for the flavor changing neutral current $b \rightarrow s$. The left diagram is forbidden in the Standard Model.

1.1.2 Flavor Changing Neutral Currents

In the Standard Model neutral weak interactions, mediated by the Z boson are not capable to perform transition between quark flavors, such as $b \rightarrow s$ or $s \rightarrow d$. These Flavor Changing Neutral Currents (FCNC) are forbidden in the SM on tree-level and can only appear via higher order penguin or box diagrams. An example for a forbidden and allowed FCNC is depicted in fig. 1.3. By far the dominant decay mode of B mesons is the $b \rightarrow c$ transition and already $b \rightarrow u$ is suppressed compared to this mode by $[V_{ub}/V_{cb}]^2 \sim 0.01$. These suppressed modes are particularly suited for searches of new physics, because influences of new operators can add a large relative contribution to the SM amplitude.

1.1.3 Theoretical Framework of Effective Hamiltonians

Decays of B mesons can involve different energy scales, which are related to the underlying mediators. The electroweak scale is mediated by heavy W and Z bosons and the interaction distance is relative to the mass of its propagator, hence $\delta x \sim 1/W$. The energy scale of the hadronization is in the order of $\delta x \sim 1/m_b$. In theoretical

frameworks it is complicated to calculate decay amplitudes at several energy scales simultaneously. The solution is to separate the short distance processes, which can be calculated with perturbative techniques, from long distance processes. Resulting from this is an effective theory, where the heavy fields get integrated out, leaving the hadronic effects separated. The matrix element for a quark transition from state M to F is given by

$$\mathcal{A}(M \rightarrow F) = \langle F | \mathcal{H}_{eff} | M \rangle = \frac{G_F}{\sqrt{2}} \sum_i V_{CKM}^i \mathcal{C}_i(\mu) \langle F | \mathcal{O}_i(\mu) | M \rangle, \quad (1.11)$$

where $\mathcal{H}_{eff}$ is the effective Hamiltonian for the transition, G_F is the Fermi constant, V_{CKM} is the CKM matrix element of the transition and μ is the renormalization scale. Now the physical processes are split into two sets of operators: the short distance effects are included in the so called Wilson coefficients $\mathcal{C}_i(\mu)$ and the long distance effects in operators $\mathcal{O}_i(\mu)$, which are handled by non perturbative theory. Wilson coefficients can be calculated including effects of new physics. With measurements of Wilson coefficients one can probe the predictions of the SM with a high precision and differentiate between new physics scenarios.

1.1.4 Effective Hamiltonian for $b \rightarrow s \ell^+ \ell^-$

The $b \rightarrow s$ transition can proceed through electromagnetic, gluonic and weak penguin and box diagrams. For one-loop processes the effective Hamiltonian is given by the operator product expansion:

$$\mathcal{H}_{eff} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_{i=1}^{10} \mathcal{C}_i(\mu) \mathcal{O}_i(\mu), \quad (1.12)$$

where $V_{qq'}$ is the corresponding CKM matrix element. The local operators $\mathcal{O}_1$ and $\mathcal{O}_2$ are current-current operators, $\mathcal{O}_{3-6}$ are QCD penguin operators, $\mathcal{O}_7$ and $\mathcal{O}_8$ are electromagnetic and chromomagnetic operators and $\mathcal{O}_9$ and $\mathcal{O}_{10}$ are the vector and axial vector component of the electroweak penguin operator respectively.

For the $b \rightarrow s \ell^+ \ell^-$ transition only the operators $\mathcal{O}_7$, $\mathcal{O}_9$ and $\mathcal{O}_{10}$ contribute in leading order. They are described as

$$\mathcal{O}_7 = \frac{e}{16\pi^2} \bar{s}_\alpha \sigma_{\mu\nu} (m_s L + m_b R) b_\alpha F^{\mu\nu}, \quad (1.13)$$

$$\mathcal{O}_9 = \frac{e^2}{16\pi} \bar{s}_\alpha \gamma^\mu L b_\alpha \bar{\ell} \gamma_\mu \ell, \quad (1.14)$$

$$\mathcal{O}_{10} = \frac{e^2}{16\pi} \bar{s}_\alpha \gamma^\mu L b_\alpha \bar{\ell} \gamma_\mu \gamma_5 \ell, \quad (1.15)$$

where α represents the color index, e is the electromagnetic coupling constant, L and R refer to the projection operators $(1 - \gamma_5)/2$ and $(1 + \gamma_5)/2$ respectively, $\sigma^{\mu\nu} = [\gamma^\mu, \gamma^\nu]$, the symbols s and b are the fields for a strange and bottom quark, ℓ is the field for a lepton and $F_{\mu\nu}$ is the electromagnetic field tensor. The right-handed counterpart of these operators are denoted with $\mathcal{O}'_7$, $\mathcal{O}'_9$ and $\mathcal{O}'_{10}$, where $L \leftrightarrow R$ are interchanged. The underlying process of these operators are visualized in fig. 1.4.

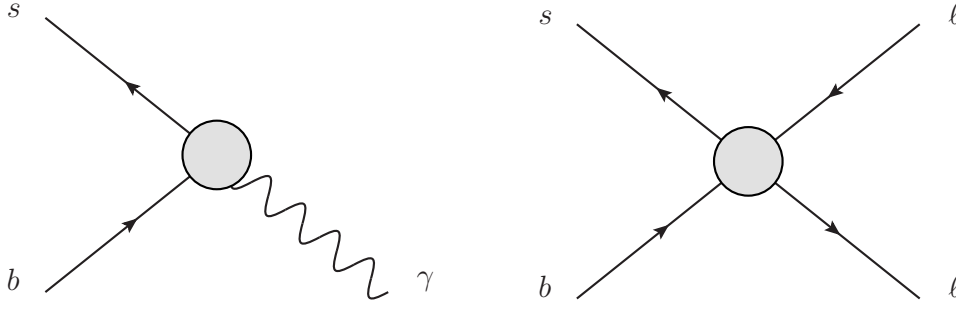


Figure 1.4: Feynman diagrams for the underlying process of the operators $\mathcal{O}'_7$ (left), $\mathcal{O}'_9$ and $\mathcal{O}'_{10}$ (right).

1.1.5 Decay Rates

In quantum mechanics transition probabilities or decay rates between two energy eigenstates $|i\rangle$ and $|f\rangle$ of a quantum system can be described by perturbation theory in first order with Fermi's Golden Rule:

$$\Gamma_{i \rightarrow f} = \frac{2\pi}{\hbar} |\mathcal{M}|^2 \rho, \quad (1.16)$$

where ρ is the density of final states, $\mathcal{M} = \langle f | H' | i \rangle$ is the matrix element and H' the corresponding Hamiltonian. Decay rates can be measured in experiments and directly related to the matrix element.

1.2 The Differential Decay Rate

Three Feynman diagrams contribute in the lowest order to the decay amplitude of $B \rightarrow K^{(*)} \ell^+ \ell^-$: Two electroweak penguin diagrams exchanging a photon or Z boson and a box diagram with a W boson loop. They are depicted in fig. 1.5. In these diagrams non Standard Model particles can occur and enhance or suppress the amplitude of the decay. In fig. 1.5(d) a possible contribution of a super-symmetric charged Higgs is shown. Not only the decay amplitude can be changed by new physics operators, also the angular distributions of the decay products can be affected due to short distance interactions.

1.2.1 Decay Topology

The decay can be completely described with four independent kinematic variables. A common choice is $q^2 = M_{\ell\ell}^2$ and three angles $\cos \theta_\ell$, $\cos \theta_K$ and ϕ , illustrated in fig. 1.6. The angle θ_ℓ is defined as the angle between the direction of the ℓ^+ (ℓ^-) and the direction of the dilepton system in the B ($\bar{B}$) rest frame. The angle θ_K is defined between the direction of the kaon and the direction of the K^* in the B ($\bar{B}$)

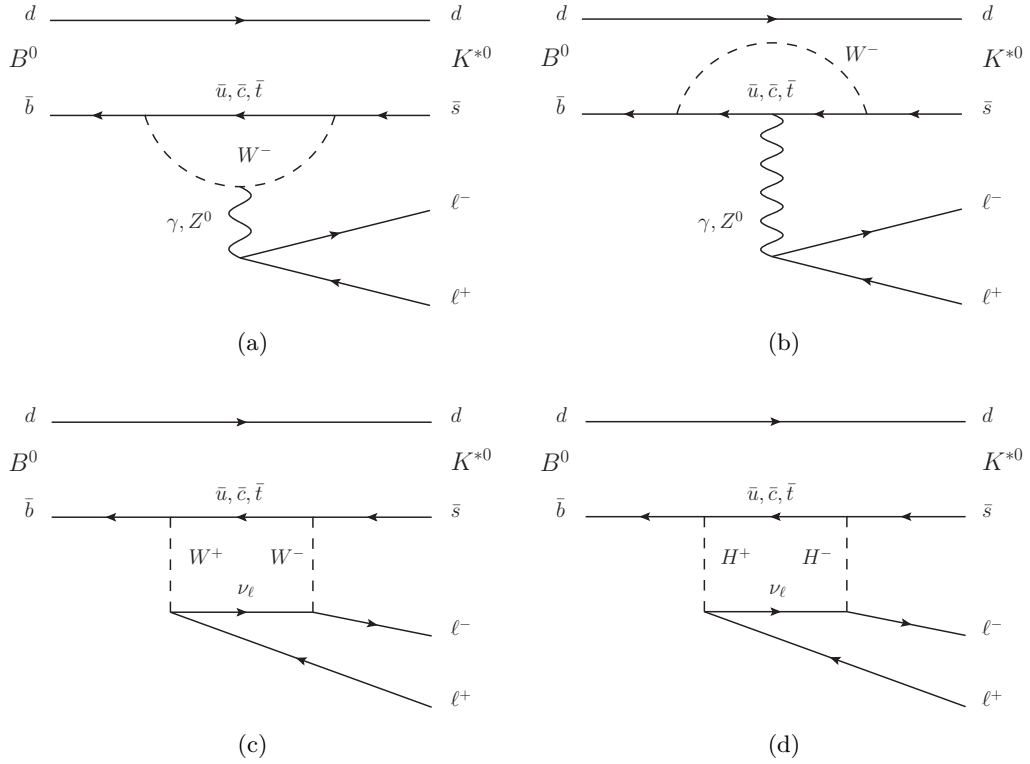


Figure 1.5: Standard Model Feynman graphs (a,b,c) for the decay $B \rightarrow K^{(*)}\ell^+\ell^-$ featuring penguin and box processes and (d) one non-SM scenario including charged Higgs replacing the W boson loop.

rest frame. Finally, the angle ϕ is determined as the angle between the decay planes of the $\ell^+\ell^-$ and the K^* decay planes. Definitions of the angles follows Ref. [17].

Explicitly the angles for the B^0/B^+ decays are defined as:

$$\cos \theta_\ell = \left(\hat{p}_{\ell^+}^{(\ell^+\ell^-)} \right) \cdot \left(\hat{p}_{\ell^+\ell^-}^{(B)} \right) = \left(\hat{p}_{\ell^+}^{(\ell^+\ell^-)} \right) \cdot \left(-\hat{p}_B^{(\ell^+\ell^-)} \right), \quad (1.17)$$

$$\cos \theta_K = \left(\hat{p}_K^{(K^*)} \right) \cdot \left(\hat{p}_{K^*}^{(B)} \right) = \left(\hat{p}_K^{(K^*)} \right) \cdot \left(-\hat{p}_B^{(K^*)} \right) \quad (1.18)$$

and for $\bar{B}^0/B^-$:

$$\cos \theta_\ell = \left(\hat{p}_{\ell^-}^{(\ell^+\ell^-)} \right) \cdot \left(\hat{p}_{\ell^+\ell^-}^{(\bar{B})} \right) = \left(\hat{p}_{\ell^-}^{(\ell^+\ell^-)} \right) \cdot \left(-\hat{p}_{\bar{B}}^{(\ell^+\ell^-)} \right), \quad (1.19)$$

$$\cos \theta_K = \left(\hat{p}_K^{(K^*)} \right) \cdot \left(\hat{p}_{K^*}^{(\bar{B})} \right) = \left(\hat{p}_K^{(K^*)} \right) \cdot \left(-\hat{p}_{\bar{B}}^{(K^*)} \right), \quad (1.20)$$

where the notation $\hat{p}_a^{(f)}$ refers to the direction of momentum of particle a in rest frame f . The angle ϕ is defined for B^0/B^+ as:

$$\cos \phi = \left(\hat{p}_{\ell^+}^{(B)} \times \hat{p}_{\ell^-}^{(B)} \right) \cdot \left(\hat{p}_K^{(B)} \times \hat{p}_\pi^{(B)} \right), \quad (1.21)$$

$$\sin \phi = \left[\left(\hat{p}_{\ell^+}^{(B)} \times \hat{p}_{\ell^-}^{(B)} \right) \times \left(\hat{p}_K^{(B)} \times \hat{p}_\pi^{(B)} \right) \right] \cdot \hat{p}_{K^*}^{(B)} \quad (1.22)$$

and for $\bar{B}^0/B^-$ as:

$$\cos \phi = \left(\hat{p}_{\ell^-}^{(\bar{B})} \times \hat{p}_{\ell^+}^{(\bar{B})} \right) \cdot \left(\hat{p}_K^{(\bar{B})} \times \hat{p}_\pi^{(\bar{B})} \right), \quad (1.23)$$

$$\sin \phi = \left[\left(\hat{p}_{\ell^-}^{(\bar{B})} \times \hat{p}_{\ell^+}^{(\bar{B})} \right) \times \left(\hat{p}_K^{(\bar{B})} \times \hat{p}_\pi^{(\bar{B})} \right) \right] \cdot \hat{p}_{K^*}^{(B)}. \quad (1.24)$$

1.2.2 Definition of the Differential Decay Rate

The measured decay in the experiment is not just $B \rightarrow K^{(*)}\ell^+\ell^-$ but $B \rightarrow K^*(\rightarrow K\pi)\ell^+\ell^-$. Additional information about the polarization of the K^* can be obtained from the angle between the K and π . The differential decay rate can be obtained by squaring the matrix element, summing over all spins of the final state particles and constraining kinematics of the four-body decay. This procedure is described in detail

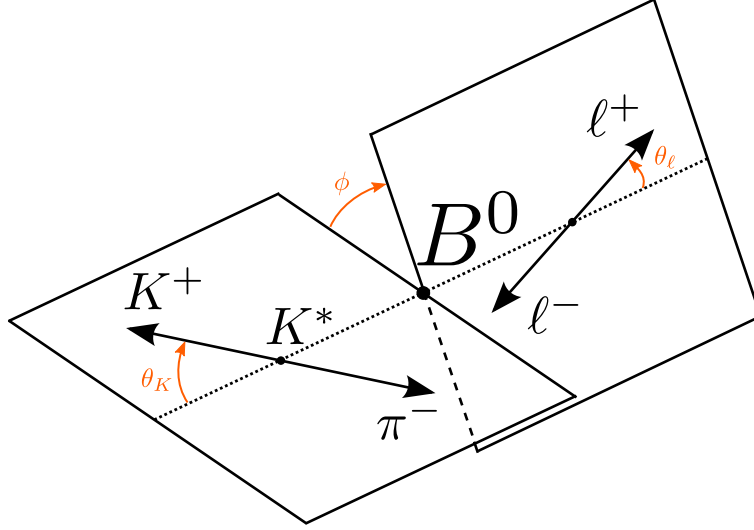


Figure 1.6: Definition of the angles in the decay $B^0 \rightarrow K^{*0} \ell^+ \ell^-$.

in Ref. [18] resulting in the differential decay rate

$$\begin{aligned} \frac{1}{d\Gamma/dq^2} \frac{d^4\Gamma}{d\cos\theta_\ell d\cos\theta_K d\phi dq^2} = & \frac{9}{32\pi} \left(I_1^s \sin^2\theta_K + I_1^c \cos^2\theta_K \right. \\ & + (I_2^s \sin^2\theta_K + I_2^c \cos^2\theta_K) \cos 2\theta_\ell \\ & + I_3 \sin^2\theta_K \sin^2\theta_\ell \cos 2\phi \\ & + I_4 \sin 2\theta_K \sin 2\theta_\ell \cos \phi \\ & + I_5 \sin 2\theta_K \sin \theta_\ell \cos \phi \\ & + I_6 \sin^2\theta_K \cos \theta_\ell \\ & + I_7 \sin 2\theta_K \sin \theta_\ell \sin \phi \\ & + I_8 \sin 2\theta_K \sin 2\theta_\ell \sin \phi \\ & \left. + I_9 \sin^2\theta_K \sin^2\theta_\ell \sin 2\phi \right), \end{aligned} \quad (1.25)$$

where the angular coefficients $I_i^{(a)}$ are functions of q^2 only and can be expressed in terms of the K^* transversity amplitudes [18]. In this notation the q^2 dependencies are completely separated from the angular variables. The coefficients $I_i^{(a)}$ are all physical observables and contain the complete information that can be extracted from the measurement. They are functions of Wilson coefficients, containing information about the short-distance effects and can be affected by new physics.

Definitions in eq. (1.25) are valid for the decay $B \rightarrow K^{*0} \ell^+ \ell^-$. The $\mathcal{CP}$ conjugated decay $\bar{B}^0 \rightarrow \bar{K}^{*0} \ell^+ \ell^-$ has to be considered separately. The differential decay rate for the combined measurement of B^0 and $\bar{B}^0$ decays can be written as

$$\frac{d^4(\Gamma + \bar{\Gamma})}{d\cos\theta_\ell d\cos\theta_K d\phi dq^2} = \frac{9}{32\pi} \sum_{i=1}^9 (I_i + \bar{I}_i) f_i(\cos\theta_\ell, \cos\theta_K, \phi), \quad (1.26)$$

where f_i contain the angular dependencies as in eq. (1.25) and $\bar{I}_i^{(a)}$ equals $I_i^{(a)}$ with all weak phases conjugated and the sign flipped for $\bar{I}_{5,6,8,9}^{(a)}$. In the context of this thesis measurements will be performed only for the $\mathcal{CP}$ averaged quantities.

1.2.3 Definition of Observables

The $\mathcal{CP}$ averaged $I_i^{(a)}$ terms can be combined into symmetric S_i and asymmetric A_i terms regarding their sign under $\mathcal{CP}$ transformation:

$$S_i^{(a)} = \frac{I_i^{(a)} + \bar{I}_i^{(a)}}{d(\Gamma + \bar{\Gamma}) / dq^2} \quad (1.27)$$

$$A_i^{(a)} = \frac{I_i^{(a)} - \bar{I}_i^{(a)}}{d(\Gamma + \bar{\Gamma}) / dq^2}. \quad (1.28)$$

Other important observables in measurements of $B \rightarrow K^{(*)} \ell^+ \ell^-$ are the forward-backward asymmetry A_{FB} , the longitudinal polarization of the K^{*0} , F_L and the transverse polarization asymmetry $A_T^{(2)}$, which are defined as,

$$A_{FB} = \frac{3}{4} \frac{I_6}{I_1^c + 4I_2^s} = \frac{3}{4} S_6, \quad (1.29)$$

$$F_L = \frac{I_1^c}{I_1^c + 4I_2^s} = S_2^c = S_2^s - 1, \quad (1.30)$$

$$A_T^{(2)} = \frac{2I_3}{4I_2^s} = \frac{2S_3}{1 - F_L}. \quad (1.31)$$

Although all information of the decay is included in the $I_i^{(a)}$ terms, in experimental measurements one tends to use different combinations as observables in order to cancel as many theoretical uncertainties on $I_i^{(a)}$ as possible. Especially the observables P'_i , defined as

$$P'_{i=4,5,6,8} = \frac{S_{j=4,5,7,8}}{\sqrt{F_L(1 - F_L)}}, \quad (1.32)$$

are considered to be largely free from form-factor uncertainties [19].

Angular Projections

By knowing the total differential decay rate, one could in principle determine all S_i observables experimentally by fitting the probability density function to data. However, with low signal yields, which is a common problem examining rare B decays, it can be necessary to integrate over one or two angles. This provides a projection of the differential decay rate onto the remaining angles. The three possible one

dimensional projections are

$$\frac{1}{\Gamma} \frac{d^2\Gamma}{d \cos \theta_\ell dq^2} = \frac{3}{4} F_L (1 - \cos^2 \theta_\ell) + \frac{3}{8} (1 - F_L) (1 + \cos^2 \theta_\ell) + A_{FB} \cos \theta_\ell, \quad (1.33)$$

$$\frac{1}{\Gamma} \frac{d^2\Gamma}{d \cos \theta_K dq^2} = \frac{3}{2} F_L \cos^2 \theta_K + \frac{3}{4} (1 - F_L) (1 - \cos^2 \theta_K), \quad (1.34)$$

$$\frac{1}{\Gamma} \frac{d^2\Gamma}{d\phi dq^2} = \frac{1}{2\pi} \left(1 + \frac{1}{2} (1 - F_L) A_T^{(2)} \cos \phi + A_{Im} \sin 2\phi \right), \quad (1.35)$$

where A_{Im} is defined as

$$A_{Im} = \frac{I_9}{I_1^c + 4I_2^s}. \quad (1.36)$$

In the next subsection a different way of reducing the dimensions of the problem is presented.

1.2.4 Increasing the Statistical Sensitivity

The full differential decay rate of $B \rightarrow K^{*0} (\rightarrow K^\pm \pi^\mp) \ell^+ \ell^-$ with $\mathcal{CP}$ averaged observables can be expressed using definitions from above and Ref. [18] by

$$\begin{aligned} \frac{1}{d\Gamma/dq^2} \frac{d^4\Gamma}{d \cos \theta_\ell d \cos \theta_K d\phi dq^2} = \frac{9}{32\pi} & \left[\frac{3}{4} (1 - F_L) \sin^2 \theta_K + F_L \cos^2 \theta_K \right. \\ & + \frac{1}{4} (1 - F_L) \sin^2 \theta_K \cos 2\theta_\ell \\ & - F_L \cos^2 \theta_K \cos 2\theta_\ell + S_3 \sin^2 \theta_K \sin^2 \theta_\ell \cos 2\phi \\ & + S_4 \sin 2\theta_K \sin 2\theta_\ell \cos \phi + S_5 \sin 2\theta_K \sin \theta_\ell \cos \phi \\ & + S_6 \sin^2 \theta_K \cos \theta_\ell + S_7 \sin 2\theta_K \sin \theta_\ell \sin \phi \\ & \left. + S_8 \sin 2\theta_K \sin 2\theta_\ell \sin \phi + S_9 \sin^2 \theta_K \sin^2 \theta_\ell \sin 2\phi \right]. \end{aligned} \quad (1.37)$$

In total there are eight free parameters, which can be derived by a direct fit to the data. The expected statistics for Belle are not sufficient for performing an eight-dimensional fit to the data. In the following a folding technique is described, lowering the dimension of the problem thereby increasing the statistical sensitivity. The folding is applied to specific regions in the three-dimensional angular space, exploiting the symmetries of cosine and sine to cancel terms in the equation. As a consequence the number of free parameters in the fit is reduced without losing experimental sensitivity. The procedure is explained in detail in Refs. [1] and [20]. With the following transformations to the dataset one can be independently sensitive to the observable of interest:

$$P'_4, S_4 : \begin{cases} \phi \rightarrow -\phi & \text{for } \phi < 0 \\ \phi \rightarrow \pi - \phi & \text{for } \theta_\ell > \pi/2 \\ \theta_\ell \rightarrow \pi - \theta_\ell & \text{for } \theta_\ell > \pi/2, \end{cases} \quad (1.38)$$

$$P'_5, S_5 : \begin{cases} \phi \rightarrow -\phi & \text{for } \phi < 0 \\ \theta_\ell \rightarrow \pi - \theta_\ell & \text{for } \theta_\ell > \pi/2, \end{cases} \quad (1.39)$$

$$P'_6, S_7 : \begin{cases} \phi \rightarrow \pi - \phi & \text{for } \phi > \pi/2 \\ \phi \rightarrow -\pi - \phi & \text{for } \phi < -\pi/2 \\ \theta_\ell \rightarrow \pi - \theta_\ell & \text{for } \theta_\ell > \pi/2, \end{cases} \quad (1.40)$$

$$P'_8, S_8 : \begin{cases} \phi \rightarrow \pi - \phi & \text{for } \phi > \pi/2 \\ \phi \rightarrow -\pi - \phi & \text{for } \phi < -\pi/2 \\ \theta_K \rightarrow \pi - \theta_K & \text{for } \theta_\ell > \pi/2 \\ \theta_\ell \rightarrow \pi - \theta_\ell & \text{for } \theta_\ell > \pi/2. \end{cases} \quad (1.41)$$

Each of the transformations vanishes all terms of eq. (1.37) except for the first five and the corresponding S_i term. The number of free parameters of each transformed decay rate is consequently reduced to three: F_L , S_3 and one of the observables $S_{4,5,7,8}$ or $P'_{4,5,6,8}$.

1.3 Overview of Recent Measurements

The decay $B \rightarrow K^{(*)} \ell^+ \ell^-$ has been studied by several experiments. The measurements of Belle, BaBar and CDF, see Refs. [21–23], observed 230, 60 and 164 signal events respectively. They extracted observables in projections of one of the three angles; in particular A_{FB} and F_L .

The LHCb experiment was first to provide a full angular measurement in all three angles with the extraction of $A_{FB}, F_L, A_T^{(2)}, A_T^{Re}, S_3, S_9$ and A_9 , described in Ref. [17]. The signal yield is 883 ± 34 . In a separate analysis of the same dataset, LHCb extracted $P'_{4,5,6,8}$ together with F_L and $A_T^{(2)}$ in six bins of q^2 [1]. The folding technique described in section 1.2.4 was applied to increase the statistical sensitivity. In 23 of the 24 measurements agreement with the standard model was found. One measurement deviated from the Standard Model expectations by 3.7σ . The so-called P'_5 anomaly was found in the variable P'_5 in the q^2 region $4.30 < q^2 < 8.68 \text{ GeV}^2/c^4$.

With three times more integrated luminosity LHCb published an update on the full angular analysis [2]. For the first time enough signal candidates were available to fit the full angular decay rate directly. Within a slightly different q^2 region the deviation in P'_5 could be verified and a global deviation of 3.4σ from the Standard Model expectation was announced. The LHCb result for all bins of P'_5 is depicted in fig. 1.7.

1.4 Beyond the Standard Model

Although the Standard Model can claim remarkable successes in particle physics, it lacks to explain a variety of important phenomena. It is considered to be a

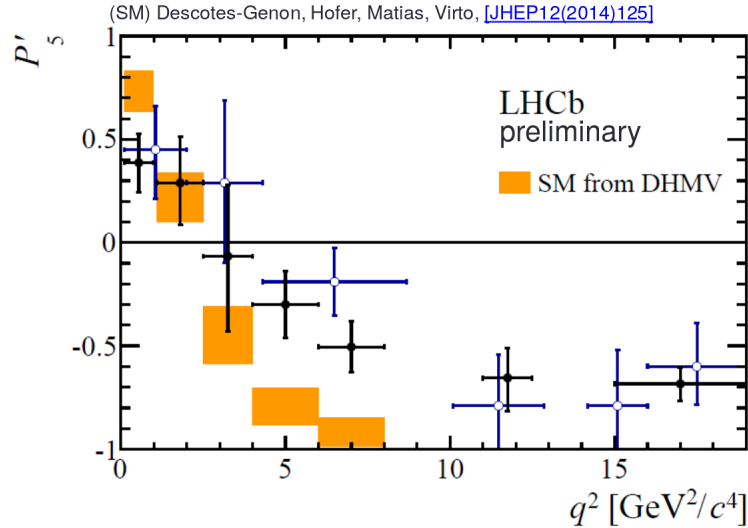


Figure 1.7: P'_5 measurement by LHCb with results from 2013 [1] and from 2015 [2] shown in blue and black respectively.

approximation of a more general theory that is only valid at the accessible energies at the electroweak scale.

One of the most arguable flaws of the Standard Model are the number of free parameters. In total 19 parameters have to be determined from experiments. There are six quark masses, three² lepton masses, three mixing angles in the CKM matrix, one $\mathcal{CP}$ -violating phase, three gauge coupling constants, the QCD vacuum angle, the Higgs vacuum expectation value and the Higgs mass. For a global theory it would be desirable that measurable constants could be obtained from theoretical expectations.

The most stringent arguments however arise from cosmological observations. It has been observed from the rotation speed of galaxies, that it is not in equilibrium with the centrifugal force. From these observations it can be calculated, that the actual mass of the galaxy has to be far greater than the visible mass. The existence of an unknown form of gravitationally acting matter was proposed to account for this mass – the so-called dark matter. Effects from gravitational lensing in collisions of galaxies have supported this assumption furthermore. Additionally, there has to be a form of energy, called dark energy, which accounts for the accelerated expansion of the universe. In total it has been measured, that more than 95% of the universe consist of dark matter and dark energy [24]. In the Standard Model, there is no candidate for a dark matter particle or an explanation for dark energy – the upper limit for the mass of neutrinos is too low to account for the observed phenomena.

Another unexplained phenomenon is the imbalance between matter and anti-matter

²Through neutrino mixing one knows that neutrinos must be massive – a phenomenon which is not foreseen in the SM.

in the universe. The theory of the Big Bang assumes equal production of matter and antimatter. However, there is no anti-matter observed in the universe today. $\mathcal{CP}$ violating processes could serve as an explanation, but the mechanism included in the Standard Model could only account for a small fraction of the observed discrepancy.

New physics is clearly needed for an explanation of these phenomena. Low energy supersymmetry (SUSY) can explain a variety of problems with the SM. But in SUSY the amount of particles is doubled, as every fermion gets a bosonic and every boson gets a fermionic supersymmetric partner. For the following new physics scenarios predictions for the S_i observables are calculated by [18]:

Models with Minimal Flavor Violation (MFV)

Models with constrained Minimal Flavor Violation (CMFV) belong to the simplest extensions of the Standard Model. All operators and also the $\mathcal{CP}$ -violating observables in these scenarios are SM-like. The flavor violating transitions are governed by the CKM.

General Minimal Supersymmetric Standard Models (GMSSM)

In general MSSM no special flavor and $\mathcal{CP}$ -violation is assumed. These models feature a huge amount of free parameters. In calculations performed by Ref. [18], new physics effects are created dominantly by complex contributions to the Wilson coefficient C'_7 .

Flavor-Blind Minimal Supersymmetric Standard Model (FBMSSM)

In this model the CKM matrix remains the only source of flavor violation [25]. New $\mathcal{CP}$ violating but flavor conserving phases are added to the soft sector.

The observables S_4 , S_5 and S_6^2 are displayed for these models in fig. 1.8. They are particularly interesting by featuring a zero crossing point in q^2 , which is a powerful measure in discriminating against various models.

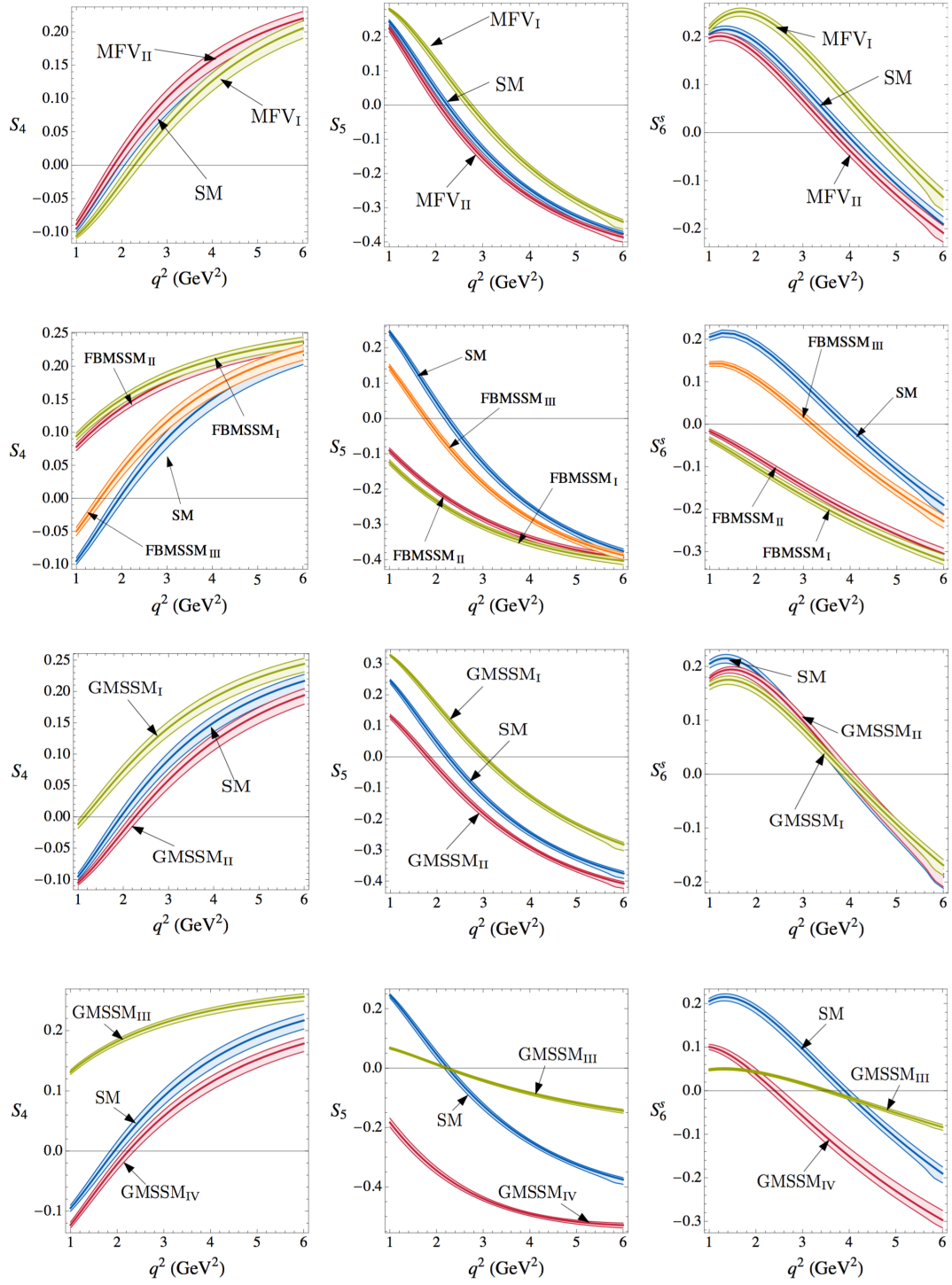


Figure 1.8: New physics scenarios for the observables S_4 , S_5 and S_6^2 , taken from Ref. [18].

2. Experimental Setup

The analyses in this thesis are performed within the Belle Collaboration with data recorded by the Belle detector. The Belle experiment is located at the asymmetric-energy e^+e^- collider KEKB in Tsukuba, Japan. It was designed as a so-called B -factory, an experiment aiming to record as many B mesons as possible in the precise and clean environment of an electron positron collider. The center-of-mass energy of the collider corresponds to the mass of the $\Upsilon(4S)$ resonance, which decays almost exclusively into a B mesons pair. Although hadron colliders are able to produce B mesons much more abundantly, only in e^+e^- collisions it is possible to produce them without any further particles from the primary interaction. This circumstance allows for unique measurements and clean experimental environments.

2.1 Fundamentals of Particle Collisions

To understand the nature of matter and its interactions one can use particle collisions with energy densities which are close to the state of the very early universe.

Particle collisions serve two main objectives. On the one hand heavy matter and new particles can be created from pure energy according to Einsteins famous relation

$$E = mc^2 \tag{2.1}$$

and on the other hand one can use high energetic particles as a probe to examine the structure of matter. This can be derived from De Broglie's theory of the duality between matter and light, where the wavelength λ of a particle is proportional to its momentum p :

$$\lambda = \frac{h}{p} = \frac{h \cdot c}{E_{kin} \cdot (E_{kin} + 2m_0c^2)}, \tag{2.2}$$

where h is the Planck constant, c is the speed of light, m_0 is the rest mass and E_{kin} is the kinetic energy.

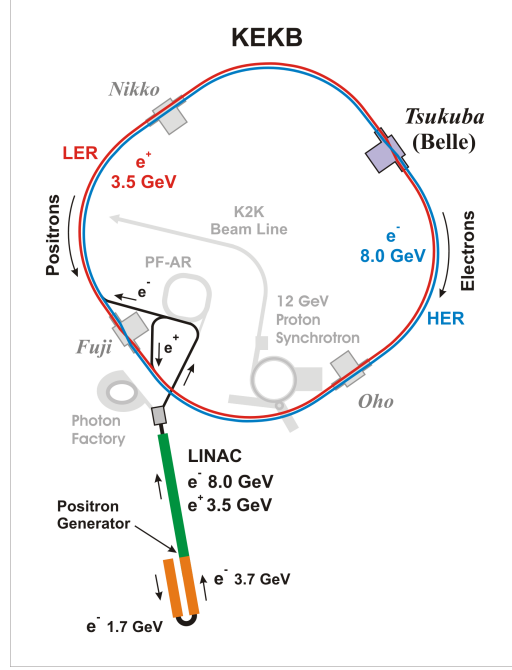


Figure 2.1: The KEKB accelerator facility at The High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan¹.

The resolution one can achieve is proportional to the wavelength $\lambda \approx d$.

Several quantities are important for physical analyses. The center of mass energy determines what matter can be created and the luminosity relates to the rate of physical interactions per time. Particularly, the interaction rate

$$\frac{dN}{dt} = L \cdot \sigma \quad (2.3)$$

is determined by the luminosity L of the accelerator and the cross section (σ) of the reaction. The luminosity is defined by

$$L = \frac{1}{4\pi} \frac{f \cdot N_1 \cdot N_2}{\sigma_x \cdot \sigma_y}, \quad (2.4)$$

where f is the frequency of interactions i.e. the bunch crossing rate, $N_{1,2}$ are the numbers of particles in each bunch and $\sigma_{x,y}$ is the width of the elliptical distribution of the particles in the bunch.

2.2 The Belle Experiment

The KEKB accelerator [26] is an asymmetric-energy e^+e^- circular collider with a circumference of ~ 3 km, serving electrons in the high energy ring (HER) at 8 GeV

¹Figure by Krib (Own work), via Wikimedia Commons.

and positrons in the low energy ring (LER) at 3.5 GeV. The layout of the accelerator is displayed in fig. 2.1. The center of mass energy $\sqrt{s}$ is chosen slightly above the mass-threshold of two B mesons at the $\Upsilon(4S)$ resonance energy:

$$\sqrt{s} = \sqrt{4 \cdot E_{LER} \cdot E_{HER}} = 10.58 \text{ GeV} \simeq m_{\Upsilon(4S)}, \quad (2.5)$$

where $E_{LER}(E_{HER})$ refers to the energy of the LER (HER). Due to the asymmetric layout of the beams the initial states of the reaction are boosted into the HER direction, resulting in a Lorentz-boost $\beta\gamma$ of the system equivalent to

$$\beta\gamma = \frac{E_{HER} - E_{LER}}{\sqrt{s}} = 0.425. \quad (2.6)$$

The boost leads to an increased lifetime of the B mesons in the laboratory frame, allowing more precise measurements of its flight distance from which analyses of $\mathcal{CP}$ -violation can benefit.

The KEKB accelerator holds two world records for luminosity. First in peak luminosity

$$L = 21,083 \cdot 10^{33} \frac{1}{\text{cm}^2 \cdot \text{s}} \quad (2.7)$$

and the second in the largest recorded integrated luminosity,

$$\int L dt = 1040 \cdot 10^{39} \frac{1}{\text{cm}^2} = 1040 \frac{1}{\text{fb}}. \quad (2.8)$$

2.3 The Belle Detector

The Belle detector is a hermetic detector optimized for energy and momentum resolution in the energy range from a few GeV down to some 100 MeV, located around at the Tsukuba interaction region at KEKB. It features several layers of track finding and particle identification layers within and around a 1.5 Tesla superconducting solenoid, depicted in fig. 2.2. The detector is designed to record particle collisions at a high luminosity and allows for high precision measurement of time dependent $\mathcal{CP}$ -violating decays. Due to the asymmetric layout of the beams at KEKB, also the detector features an asymmetric design to account for the boost of the produced particles into the HER direction. The Detector is described in detail in [27, 28], main parts will be discussed below.

SVD The Silicon Vertex Detector (SVD) is the innermost layer and the first element for precise tracking of charged particles. The silicon detector from the first period of data taking (SVD1) surrounded the interaction point covering $23^\circ < \theta < 139^\circ$. Due to damage from radiation, it was replaced in 2004 by an updated version (SVD2). The new detector covers an angle of $17^\circ < \theta < 150^\circ$, consisting of four layers of silicon strip detectors. In the presented analyses, information from the SVD is used for vertex fitting of the secondary particles in B meson decays, serving as background suppression method.

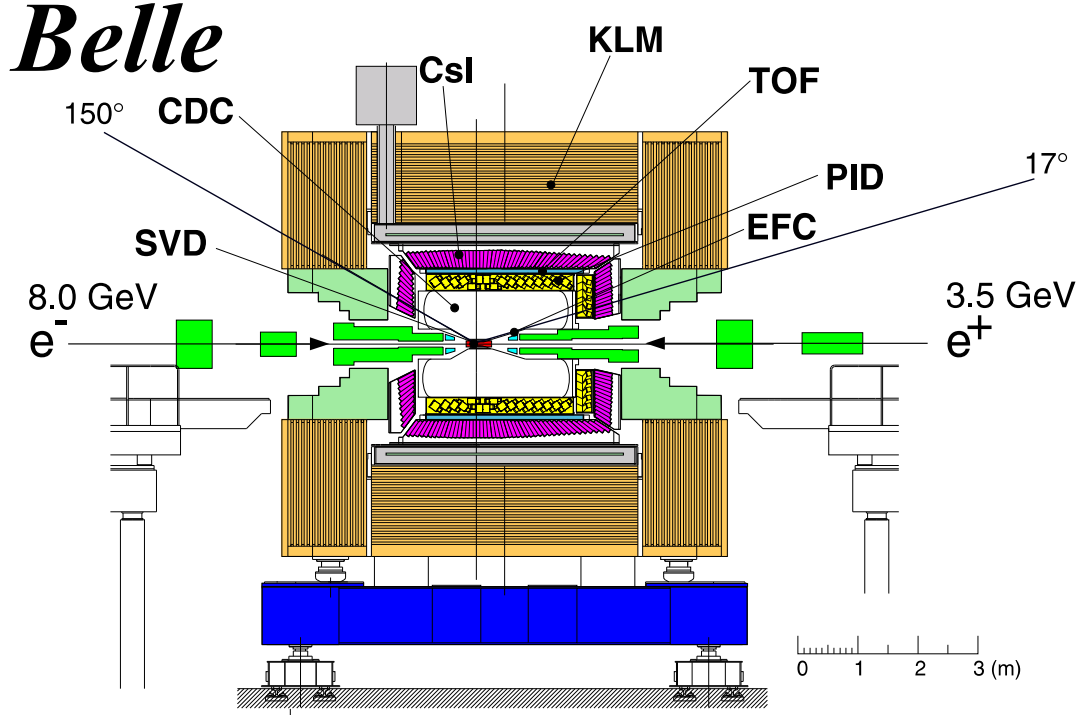


Figure 2.2: A schematic view of the Belle detector, from [27].

CDC The main tracking element is the Central Drift Chamber (CDC). Its main purpose is to find tracks and measure the momentum of charged particles. Moreover, it provides information about the specific energy loss dE/dx , which is an important part in particle identification. The CDC consists of 50 cylindrical layers of drift cells organized in 11 super-layers. It covers an angle of $17^\circ < \theta < 150^\circ$ along the beam axis. The CDC is filled with a gas mixture of 50% helium and 50% ethane.

ACC The Aerogel Cherenkov Counter (ACC) is designed for particle identification of high energetic charged particles ($p > 1.2 \text{ GeV}/c$). It surrounds the CDC along the barrel side and also in forward direction. The ACC is made of 960 elements with the dimension of $12 \times 12 \times 12 \text{ cm}^3$. Cherenkov light from the aerogel radiators is detected in photomultipliers attached to the side of the device. Its main purpose is to identify charged pions, kaons and protons.

TOF The Time Of Flight (TOF) counter is located around the ACC on the barrel side, covering a polar angular range of $34^\circ < \theta < 120^\circ$. It allows for a measurement of the velocity of low energetic ($p < 1.2 \text{ GeV}/c$) particles and serves as a part of the particle identification system. The TOF and ACC array can be seen in fig. 2.3.

ECL The Cesium iodide calorimeter is the electro-magnetic calorimeter (ECL) of the Belle detector. It consists of 6624 Thallium doped Cesium iodide CsI(Tl)

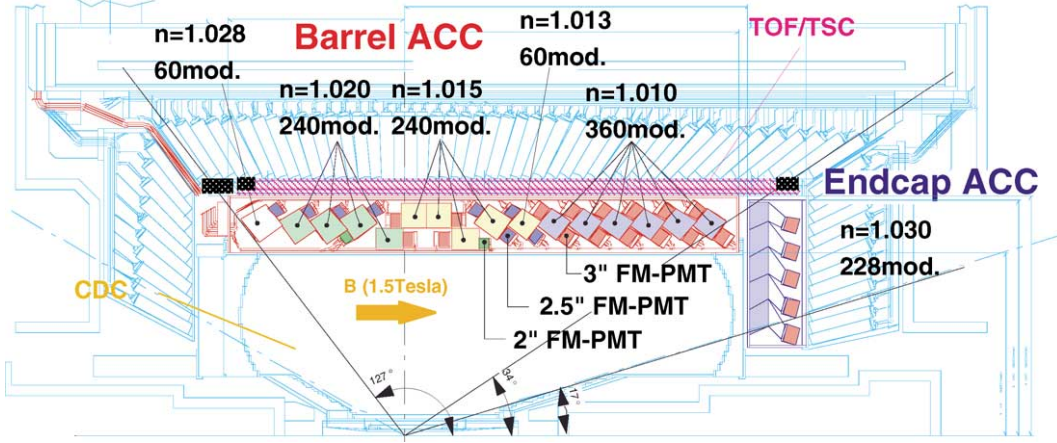


Figure 2.3: A side view the ACC and TOF of the Belle detector from [27].

crystal elements around the barrel and 2112 elements on the end-caps. Each element has a cross section of $6 \times 6 \text{ cm}^2$ a length of 30 cm corresponding to 16.2 radiation lengths (X_0) for photons and electrons. The crystals are wrapped into a thin layer of goretex Teflon with a photo-diode mounted onto their end. The main objective of the ECL is to measure the energy deposition of electrons and photons. High energetic electrons and photons produce electromagnetic showers in the scintillating crystals by bremsstrahlung and pair-production. The size of the shower scales with the particle's energy. The scintillating material gets excited by the secondary particles of the shower and emits light during the de-excitation process, which gets detected by the photo-diodes and scales according to the energy of the primary particle. The ECL covers a polar angle of $17^\circ < \theta < 150^\circ$. For the measurement of particles in the extreme forward direction, an Extreme Forward Calorimeter (EFC) is located around the beam pipe the polar angle of $6.4^\circ < \theta < 11.5^\circ$ and $163.3^\circ < \theta < 172.2^\circ$.

Magnet The superconducting solenoid is made from a Niob Titanium Copper (Ni-Ti/Cu) composition. It creates a nearly homogeneous magnetic field of 1.5 Tesla inside the barrel, forcing charged particles into curved tracks, consequently allowing for momentum measurements.

KLM The K_L and muon detection system (KLM) is the most outer part of the detector, serving as return path for the magnetic flux and as an absorber and detector material for K_L mesons and muons. The KLM detector consists of two end-cap regions and a central barrel region. The latter is formed by eight detector blocks and flux-return plates. Each detector is made of glass Resistive Plate Counter (RPC) modules, separated by iron plates. Muons with a momentum above 600 MeV/c can reach the KLM and can be separated from K_L candidates since they are charged and leave a track in the CDC pointing to hits in the ECL and KLM.

2.3.1 Triggers and Data Acquisition

The data acquisition system (DAQ) of the Belle detector is controlled by a trigger system, deciding whether to store an event or not. The system is optimized for selecting hadronic decays $e^+e^- \rightarrow q\bar{q}$, two photon processes, $e^+e^- \rightarrow \tau^+\tau^-$, Bhabha scattering or muon pairs. Background processes from synchrotron radiation, interactions between the beam and residual gas or events from cosmic rays are tried to be suppressed. At peak luminosity of $\mathcal{L} = 2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ the event rate for signal and background processes were 200 Hz and 600 Hz respectively. The trigger system has to decide fast over the quality of an event based on the signals in the detector and is organized in several steps, so-called levels. At the lowest stage is the Level 1 trigger, which is fed directly by all sub-components of the detector except for the SVD. The trigger information of each sub-detector is combined 1.85 μs after the beam crossing in the Global Decision Logic (GDL) and the trigger decision is produced 2.2 μs after the collision. In total the GDL features four triggers for hadronic events. The two-track trigger, three-track trigger, an ECL cluster separation trigger and a trigger based on the total energy deposition. With a combination of these, the GDL offers more than 99 % efficiency for hadronic events. In the beginning of the Belle experiment this was the only trigger system, however it became soon clear, that for a manageable amount of data the reduction had to be improved. For this reason the trigger Levels 3 and 4 were implemented. The Level 3 trigger system is realized by a fast track finding algorithm running on an online farm of computers with the main purpose to suppress beam background events by vetoing tracks that do not originate from the interaction point. After this step, the raw data is stored on tapes. In the last step the software trigger Level 4 is applied to raw data offline. The main software reconstruction algorithms are applied and data summary tapes, DSTs, are produced, containing only physics events with an average size of 4 GB $\times$ pb. Altogether, the trigger efficiency for hadronic events is nearly 100% while half of the background processes are rejected. A detailed description of the system can be found in Ref. [27].

2.3.2 Particle Identification

The rest mass of particles cannot be measured directly by detector components. It can be calculated from the momentum and energy of the particle, which is calculated from the precise measurement of its momentum and rest-mass under a particle hypothesis. With combination of information of different detector components the particle identification (PID) system delivers probabilities of measured charged tracks to be an electron, muon, kaon or pion. A likelihood is constructed with information of the energy loss in the CDC (dE/dx), TOF information, number of photons in the ACC and ECL clusters in the path of the corresponding track and KLM clusters. For separating different particle hypotheses likelihood ratios are constructed with

$$\text{PID}_{\alpha \text{ vs. } \beta} = \frac{\prod_i \mathcal{L}_{\alpha}^i}{\prod_i \mathcal{L}_{\alpha}^i + \prod_i \mathcal{L}_{\beta}^i}, \quad (2.9)$$

where $\prod_i \mathcal{L}_{\alpha}^i$ represents the likelihood for particle kind α in detector component i .

2.3.3 Example Decay

In fig. 2.4 an event display of a sample decay, which is examined within this thesis is shown. The event which was recorded on 16th November of 1999 on data with the Belle detector shows a candidate for the decay $B^0 \rightarrow K_S J/\psi$ with a high signal probability. Two oppositely charged pions could be matched to daughters of a K_S meson with the help of CDC tracking and SVD vertexing. Additionally, two oppositely charged muons were found with the invariant mass compatible with the J/ψ resonance. The response of various detector components is colored in the plots. Blue circles indicate the drift time on wires in the CDC.

2.4 Luminosity and Data Samples

Belle operated from 1st of June 1999 until end of June 2010. The total integrated luminosity reached 1040fb^{-1} , displayed in fig. 2.5. The majority of data was taken at the center of mass energy corresponding to the mass of the $\Upsilon(4S)$ resonance. Off-resonance data was taken 60 MeV below this energy for 10% of the running time to collect data containing non $B\bar{B}$ background. Apart from that, a noticeable amount of data was taken at the energy of the $\Upsilon(5S)$, which allows additionally for the production of $B_s\bar{B}_s$ pairs. In total Belle recorded $(772 \pm 11) \times 10^6$ $B\bar{B}$ pairs at the $\Upsilon(4S)$ resonance. A summary of the amount of data taken by the Belle experiment at the corresponding resonances is presented in table 2.1.

Table 2.1: Summary of the integrated luminosity recorded by Belle broken down to the resonance energy.

Resonance	On-peak luminosity (fb^{-1})	Number of resonance decays
$\Upsilon(1S)$	5.7	102×10^6
$\Upsilon(2S)$	24.9	158×10^6
$\Upsilon(3S)$	2.9	11×10^6
$\Upsilon(4S)$ SVD1	140.0	152×10^6 $B\bar{B}$
$\Upsilon(4S)$ SVD2	571.0	620×10^6 $B\bar{B}$
$\Upsilon(5S)$	121	7.1×10^6 $B_s\bar{B}_s$

2.5 Simulated Datasets

In this thesis, both analyses are performed on simulated data in the first place on so-called Monte Carlo (MC) data. The analysis procedures are established and tested within the simulated environment. The software packages EvtGen [29] and PYTHIA [30] are used to simulate the particle decays. In this step the decay chain is simulated and all intermediated and final state particles are determined. Final state radiation is calculated by the PHOTOS package [31]. In this thesis this is crucial for the electron modes of $B \rightarrow K^{(*)}\ell^+\ell^-$. The simulation and detector response is afterwards

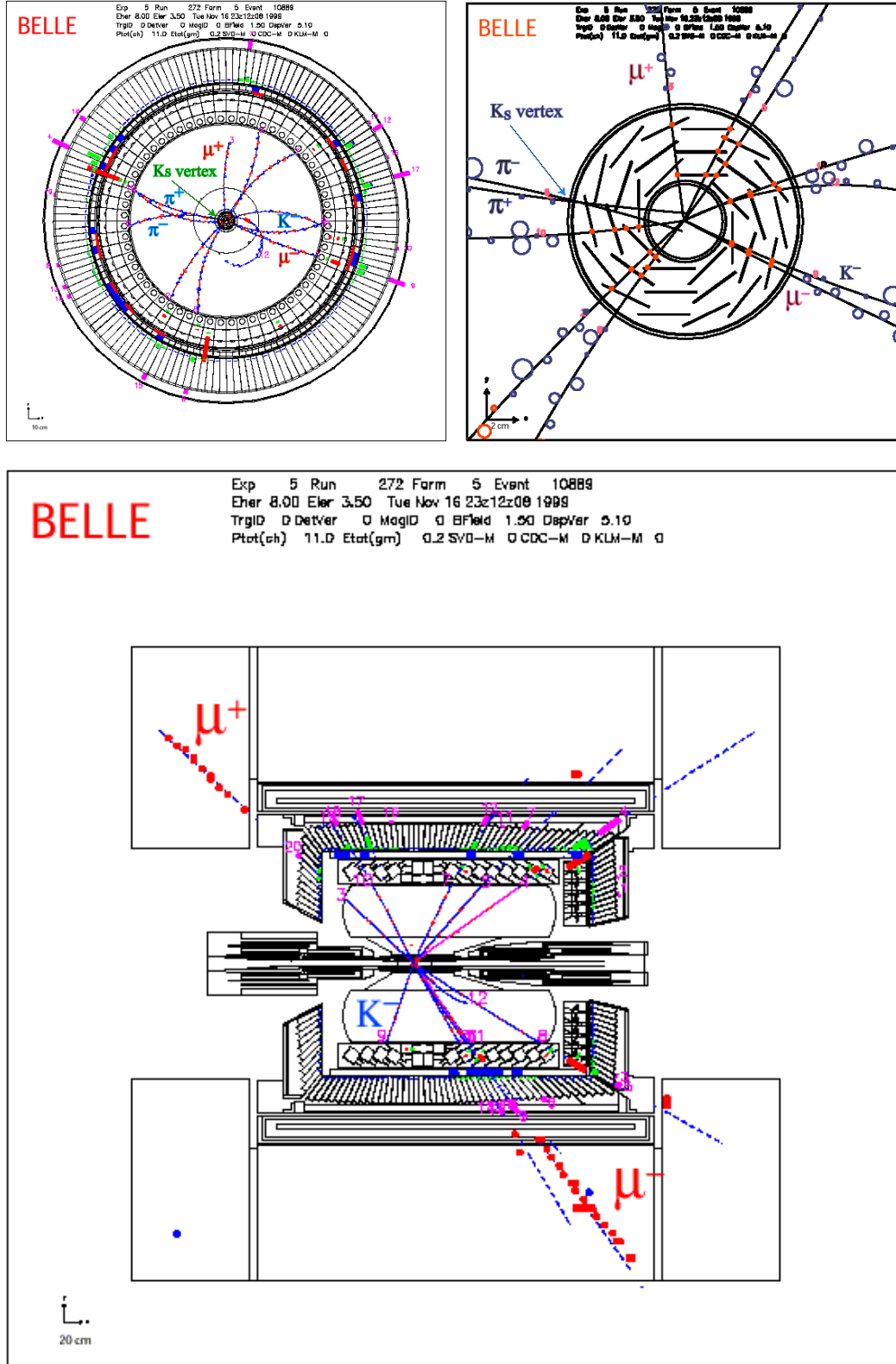


Figure 2.4: Belle event display of a fully reconstructed $B \rightarrow K_S J/\psi$ event where the J/ψ decays into two muons. The top plots show the x-y plane and the bottom plot displays the y-z side-view of the detector.

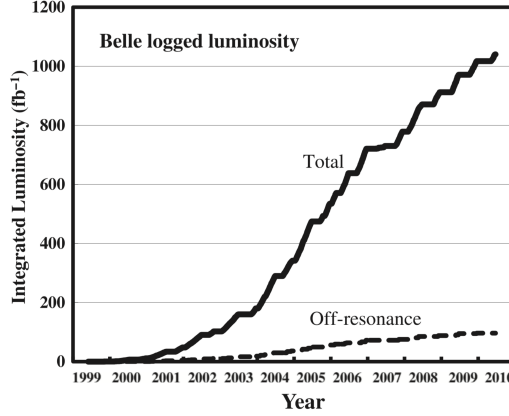


Figure 2.5: Belle integrated luminosity. Taken from [27].

performed with the software package GEANT3 [32]. The Belle experiment provides officially produced MC samples, aiming to resemble the recorded dataset in both size and composition.

Generic MC

For studying the backgrounds large samples of $b \rightarrow c$ transitions were simulated containing all known decays with their appropriate branching ratio. There are two kinds of this MC depending on the spectator quark², hence the charge of the B meson: the charged (B^+B^-) and mixed ($B^0\bar{B}^0$) generic MC. Belle provides ten independent sets of MC, referred to as stream of generic MC, each of them corresponding to the total integrated luminosity of the Belle dataset. The **Continuum MC** contains light quark fragmentation $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$). The hadronisation of decays are simulated with PYTHIA and JETSET [33]. The total statistics of the continuum MC corresponds to four times the Belle dataset. In table 2.2 all individual fractions of the generic MC are listed. They serve as the basis of the background composition in both described analyses.

Table 2.2: Composition of the Belle generic MC.

Name	Process
mixed	$\Upsilon(4S) \rightarrow B^0\bar{B}^0$, with generic B^0 decay
charged	$\Upsilon(4S) \rightarrow B^+\bar{B}^-$
charm	continuum $e^+e^- \rightarrow c\bar{c}$
uds	continuum $e^+e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}$

²A spectator quark is the light quark component of the B meson, in $\Upsilon(4S)$ decays this can be a u or d quark, defining the charge of the compound.

$b \rightarrow \ell \nu_\ell$

The $b \rightarrow \ell \nu_\ell$ MC contains $B \rightarrow X_u \ell \nu_\ell$ decays, where X_u originates from a $b \rightarrow u$ transition, e.g. a π or ρ meson. The available sample corresponds to 20 times the luminosity of Belle.

Rare MC

The rare MC contains all known and calculated B decays, which are not included in the other samples, e.g. also $b \rightarrow s \ell^+ \ell^-$ transitions. For the context of this thesis however, the total amount of signal candidates in this sample is not sufficient, so that special signal MC is created.

Signal MC

For both analyses a particular large set of simulated events is created in order to study the signal processes and train multivariate methods. Details of the generated signal samples are presented in the corresponding sections of the analyses.

3. Analysis Techniques

Both measurements in this thesis offer improvements in statistical sensitivity compared to previous analyses due to the use of advanced multivariate data analysis methods. In this chapter the most important techniques for data analysis are presented. A lot of improvements in reconstruction efficiency can be made by considering conditions and correlations in-between features of the datasets. In the first parts dependencies within datasets are discussed, followed by parameter estimation and classification methods. The probabilities play an important role throughout the analyses and are discussed in the beginning of this chapter.

3.1 The Concept of Probability

There are basically two antagonizing frontiers in statistical analyses: Bayesians and Frequentists. The origin of the dispute is the very definition of probability itself. Both opinions share axioms for the foundation of probability theory that were formed by A. Kolmogorow. They state that for the probability $\mathcal{P}$ for some event E , denoted as $\mathcal{P}(E)$ the following axioms have to be satisfied: First, the probability is a non-negative real number:

$$\mathcal{P}(E) \in \mathbb{R}, \mathcal{P}(E) \geq 0 \quad \forall E \in \Omega, \quad (3.1)$$

where Ω is the space of possible events. The second axiom states that the probability that at least one of the elements of all possible events will occur is one:

$$\mathcal{P}(\Omega) = 1. \quad (3.2)$$

This is sometimes also referred to as the unitarity of probabilistic theory. The third axiom is the assumption of countable additivity. It states that the probability of the occurrence of one of the incompatible events $\mathcal{A}$ and $\mathcal{B}$ is equal to the sum of their individual probabilities:

$$\mathcal{P}(\mathcal{A} \cup \mathcal{B}) = \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B}) \quad \text{if } \mathcal{A} \cap \mathcal{B} = \emptyset. \quad (3.3)$$

These three axioms define handling and behavior of probability. The exact method of how to obtain probabilities is not determined. Two approaches are detailed hereafter.

Frequentists Approach

The Frequentists approach defines probability as follows: If an event can be realized in n distinguishable ways, of which none is preferred over the others, and k realizations have the attribute $\mathcal{A}$, then the probability for the appearance of $\mathcal{A}$ is

$$\mathcal{P}(\mathcal{A}) = \frac{k}{n}. \quad (3.4)$$

One could argue, that the definition of probability with itself was a circular reasoning, but in many cases there are underlying symmetries, allowing this kind of statement. In all other cases one has to refer to the empirical definition of probability: If in n observations, k carry the attribute $\mathcal{A}$, the probability for the appearance of $\mathcal{A}$ is

$$\mathcal{P}(\mathcal{A}) = \lim_{n \rightarrow \infty} \frac{k}{n}. \quad (3.5)$$

With this definition, one could argue again, that the limes of infinity is impossible to reach, but in many practical situations, the probability converges fast to a certain value. However, the Frequentists approach has its limitations – for example the answer to the simple question, if it will be raining tomorrow cannot be described with this approach.

Bayesian Approach

The Bayesian definition of probability is at first glance rather subjective, as it states, that probability is the degree of believe of the occurrence of the attribute $\mathcal{A}$ in an observation. The belief in a hypothesis is expressed in a prior probability, which is updated in the light of new, relevant data. In the limes of infinite number of observations, the Bayesian definition approaches the Frequentists empirical definition of probability. The Bayesian approach has advantages, when it comes to situations with no underlying symmetries and no way of repeating the problem.

The true strengths of the Bayesian approach manifests in the Bayes theorem, which relates conditional probabilities. The conditional probabilities $\mathcal{P}(\mathcal{A}|\mathcal{B})$ refers to the probability for $\mathcal{A}$ under the condition that $\mathcal{B}$ has occurred. The Bayes theorem states:

$$\mathcal{P}(\mathcal{A}|\mathcal{B}) = \mathcal{P}(\mathcal{B}|\mathcal{A}) \cdot \frac{\mathcal{P}(\mathcal{A})}{\mathcal{P}(\mathcal{B})}. \quad (3.6)$$

In particle physics an important application of the Bayes theorem is the determination of the probability to observe a signal event in a given data sample. From simulated data (Monte Carlo) the probability of $P(\text{data}|\text{signal})$ can be derived, hence one knows the data under the condition for signal or background classes. The final goal is to make conclusions of the probability for a signal event under the observation of the measured data $\mathcal{P}(\text{signal}|\text{data})$. With the help of the Bayes theorem and the proper prior probability, one can easily calculate $\mathcal{P}(\text{signal}|\text{data})$.

3.1.1 Probability Density Distributions

Random variables can be classified as either continuous or as discrete, where their distribution can be described by an underlying probability density function (PDF). Discrete random variables can adopt discrete values r_i with the corresponding probability $P(r_i) = P_i \in [0, 1]$, with the sum over all possible realizations of the random variable has to be one: $\sum_{i=a}^b P_i = 1$. In physics, one rather faces continuous random variables, such as lifetime or reconstructed mass of a particle. The random variable x is described by the PDF $f(x)$ and the probability to measure x in the interval $a < x < b$ is given by

$$P(a < x < b) = \int_a^b f(x) dx \quad (3.7)$$

where the PDF is a non-negative, normalized function:

$$f(x) \geq 0 \text{ and } \int_{-\infty}^{\infty} f(x) dx = 1. \quad (3.8)$$

Important indicators of a PDF are the estimate value E and variance V . The expected value is defined as

$$E[x] = \int_{-\infty}^{\infty} x f(x) dx, \quad (3.9)$$

which is equal to the mean value $\langle x \rangle$ and fulfills $E[ax] = \langle ax \rangle = a \langle x \rangle$ for a constant a . The variance of the distribution is defined as the second central momentum μ_2 :

$$V[x] = \mu_2 = E[(x - \langle x \rangle)^2] = \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 f(x) dx. \quad (3.10)$$

In the following sections random variables in multiple dimensions or sets of different random variables are used. In many cases these random variables are not independent of each other. Probability density functions in multiple dimensions have to be also normalized and positive. Mean and variance for the n -dimensional function $f(x_1, x_2, \dots, x_n)$ are calculated below

$$\langle x_i \rangle = E[x_i] = \int x_i f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n, \quad (3.11a)$$

$$V[x_i] = \int (x_i - \langle x_i \rangle)^2 f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \quad (3.11b)$$

Additionally, in multiple dimensions, one can calculate the covariance between two variables x_i and x_j :

$$\sigma_{x_i x_j} = \int (x_i - \langle x_i \rangle) (x_j - \langle x_j \rangle) f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \quad (3.12)$$

3.2 Dependencies in Datasets

Multidimensional maximum likelihood fits rely heavily on correct treatment of underlying correlation. Several methods are used in this thesis to determine correlations in probability density distributions used in the fit. The algorithms used are inspired by the framework presented in Ref. [34].

Linear Correlation

The linear correlation between two features of the same dataset can be described by the coefficient r , also known as Pearson's correlation coefficient. For a dataset with N entries and features x and y , it can be calculated from the data by

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}, \quad (3.13)$$

where

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{and} \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (3.14)$$

are the arithmetic means, of the distributions x and y . The range of r is by construction within $[-1, 1]$, where $r = 0$ corresponds to no linear dependence and $r = 1(-1)$ to 100% linear (anti-) correlation. A sample distribution for $r = 0.8$ is shown in figs. 3.1(a) and 3.1(b). Two linearly uncorrelated distributions are displayed in figs. 3.1(c) and 3.1(e), where $r = 0$. However, one can observe in fig. 3.1(c) that the shape of x_2 is dependent on the region of y_2 . This is additionally demonstrated in fig. 3.2 by displaying subsets of x_2 in different regions of y_2 . Both distributions offer non-linear correlations. One has to keep in mind, that this type of conditional distributions cannot be found by looking at the linear correlation of a dataset, which could be critical for maximum likelihood fits. In the following a method for automatic detection of non-linear correlations is introduced.

3.2.1 Flat Distributions

For some applications it can be useful to transform a non discrete distribution x to x_f which is flat in for example $[0, 100]$. This can have benefits for instance if x has large tails and one tries to calculate divisions between histograms. Statistical fluctuations could lead to divisions by zero, whereas in flat distributions, always the same number of events is expected in each bin. One example is a likelihood ratio between two classes in the same feature, where in tails of the distribution statistical fluctuations are large. One way to calculate a flat distribution is to use different bin sizes in a way that all bins contain the average of N/n_{bins} entries, where N is the size of the dataset and n_{bins} is the number of bins used. Values for the bin edges can be derived by calculating percentiles of x . This way, however, the distribution stays the same but the resulting histograms do not suffer from numerical instabilities. A different way is to transform each value x_i of the sample so that the transformation $x \rightarrow x_f$ becomes a flat distribution. This can be achieved by transforming the original distribution with its Cumulative Distribution Function CDF . If the CDF is not known, one can estimate it by creating a histogram h_{orig} with n_{bins} bins and event yields n_i^{orig} in bin

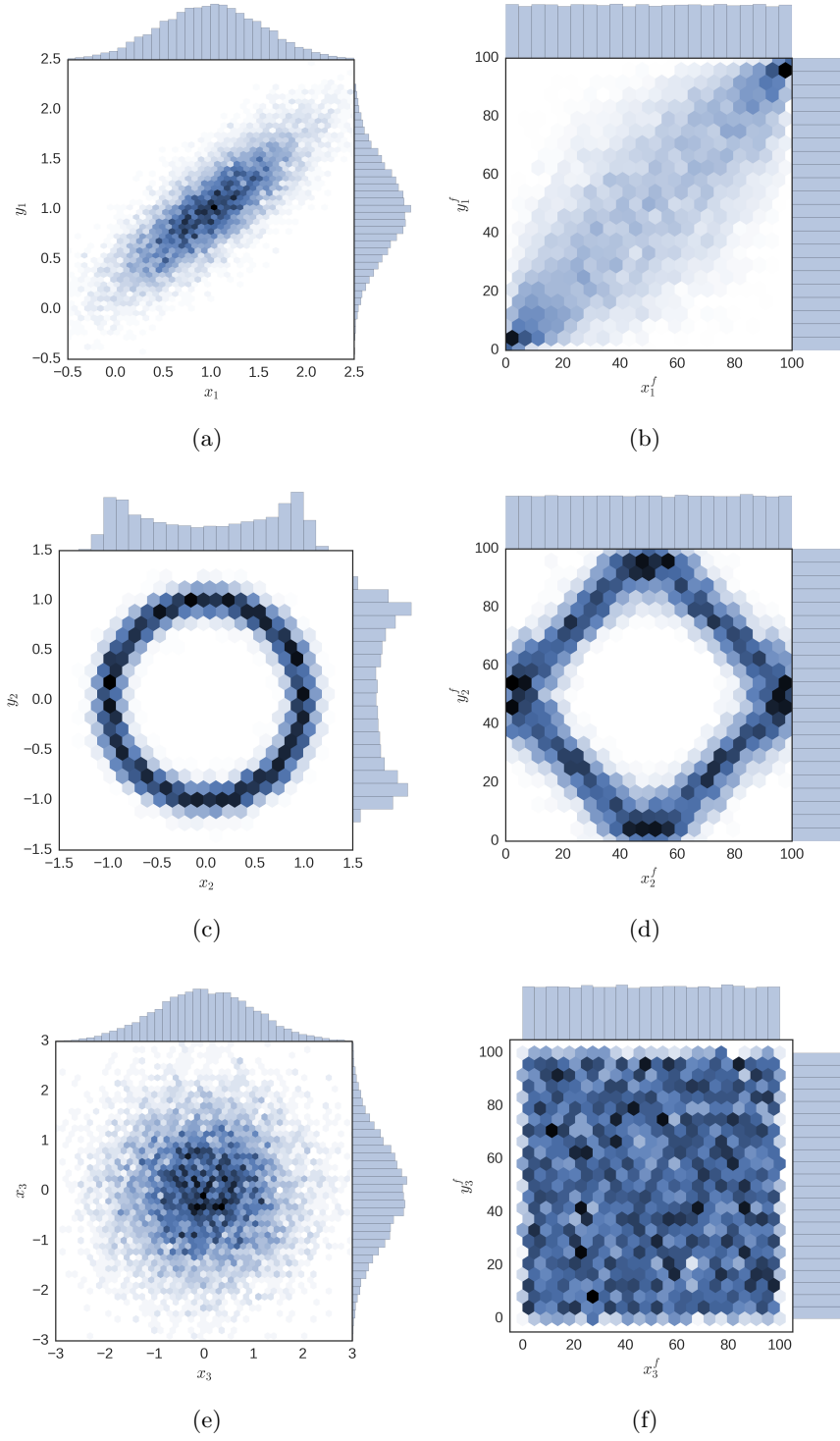


Figure 3.1: Distributions of different types of correlations between two observables (left column) with the corresponding flat transformation (right column). Plot (a) shows a correlated dataset with $r = 0.8$ and (c) and (e) both show distributions without linear correlation $r = 0.0$.

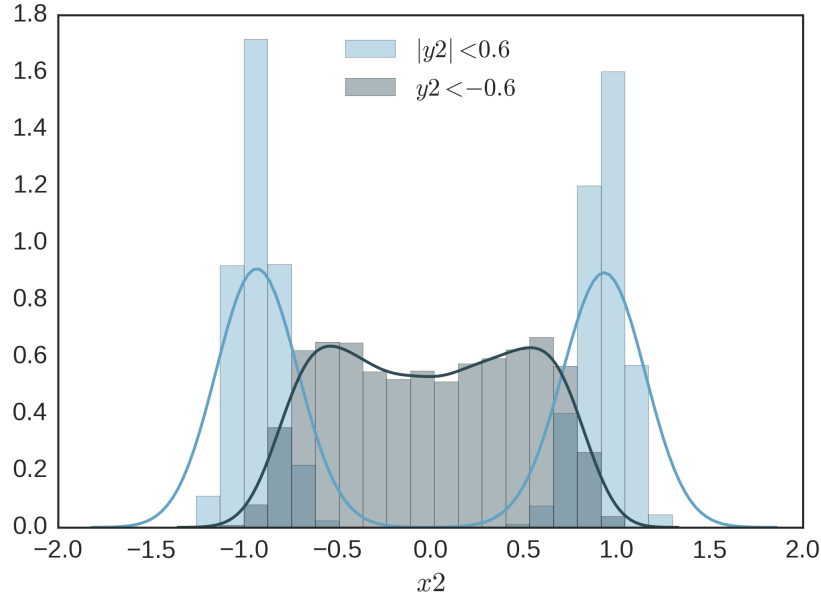


Figure 3.2: Demonstration of the dependency between x_2 and y_2 although both distributions have no linear correlations, i.e. $r = 0$.

i , by successively adding all previous contents to each bin to derive h_{CDF} with bin content

$$n_i^{CDF} = \frac{1}{N} \sum_{j=1}^{j \leq i} x_j^{\text{orig}}, \quad (3.15)$$

for $i = 1, 2, \dots, n_{\text{bins}}$. Usually the normalization for h_{CDF} is $\frac{1}{N}$ but also other factors can be useful, like $100/N$ for extracting percentiles of x . The CDF can be constructed from h_{CDF} by a spline fit or linear interpolation between the content of each bin. In applications throughout this thesis a spline fit is used for the estimation of the CDF . The transformation to a flat distribution can be calculated by $x_f = CDF(x)$. This procedure is demonstrated in fig. 3.3.

3.2.2 Analyzing Correlation

One can analyze correlations by looking at the two-dimensional distributions of two observables. In order to find linear and non-linear correlations, a robust, automated technique is needed.

Flat correlation plots

If one considers two observables x and y from the same dataset with N entries, one can transform both of them to their corresponding flat distribution x^f and y^f . A test statistic can be constructed from a two dimensional scatter-plot to test their linear and non-linear dependency. From the scatter-plot, one can create a $n \times n$ histogram

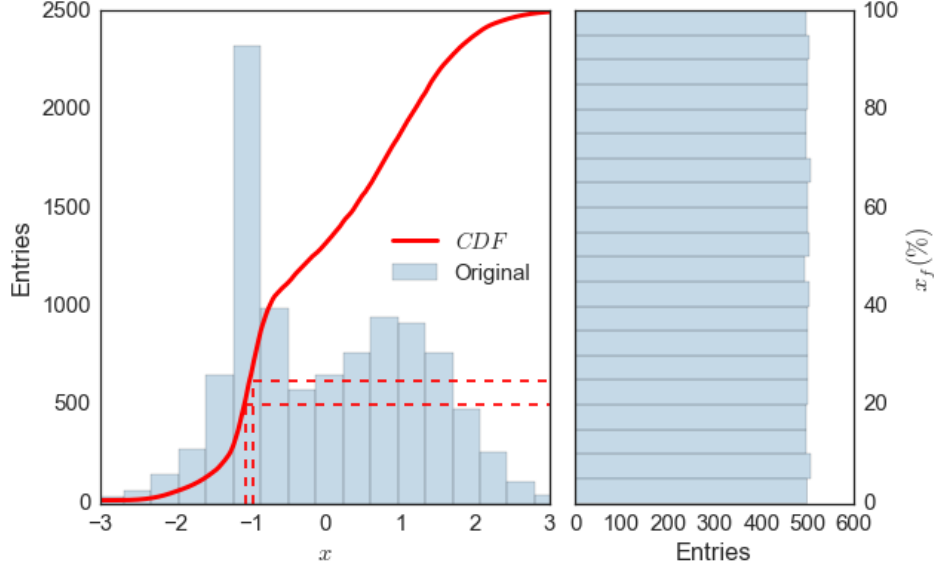


Figure 3.3: Flattening of a distribution with its CDF. The left plot shows the original distribution x and the cumulative distribution function CDF . The right plot is the flat transformation of x which can be calculated by $x_f = CDF(x)$.

H with bins of equal size. For an uncorrelated dataset, one expects the scattered distribution to be uniformly distributed in the two dimensional plane and thus one can expect in each bin of the histogram $n_{exp} = N/n^2$ events. If n is chosen large enough (i.e $n_{exp} \gtrsim 25$), one can estimate the statistical error of each bin content with the Poisson error $\sigma_{n_{exp}} = \sqrt{N/n^2}$. From this average, a χ^2 test-statistic can be calculated:

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^n \frac{(n_{ij} - n_{exp})^2}{\sigma_{n_{exp}}^2}, \quad (3.16)$$

where n_{ij} is the bin content $H(i, j)$. The probability for both distributions being uncorrelated follows a χ^2 distribution with $n^2 - (2n - 1)^1$ degrees of freedom. This procedure is adapted and described in detail in Ref. [34]. Examples of scatter plots for two observables with different types of correlations and the transformation to a flat distribution can be seen in fig. 3.1. The flat correlation plot for these three types is displayed in fig. 3.4. The color of the plot indicates deviations from the expected signal yield n_{exp} in units of the standard deviation σ with respect to the statistical error $\sigma_{n_{exp}}$. Regions in green match the expected yield, blue (red) color signals significantly less (more) events then expected from a flat hypothesis.

With this test one can see by eyes if the distribution between two features of a dataset

¹Constraints for the number of events in each row/column due to the flat transformation remove one degree of freedom.

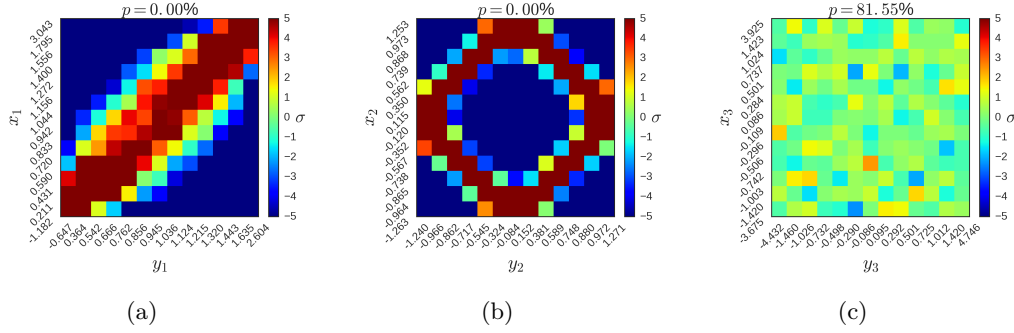


Figure 3.4: Flat correlation of correlated and uncorrelated distributions. The distribution (a) shows a linear correlation from fig. 3.1(a). In the second plot (b) the distribution of fig. 3.1(c) is shown in (c) the uncorrelated distribution from fig. 3.1(e) is displayed as a flat correlation plot. The deviation from a flat expectation is shown colored and expressed in standard deviations σ .

are correlated or not. Moreover, the test produces a test-statistic, the probability p for both features providing a two dimensional flat distribution. On the basis of this quantity one can for instance select variables for a multidimensional maximum likelihood fit. This technique is used throughout the thesis in the analysis section for the angular analysis of $B \rightarrow K^{(*)} \ell^+ \ell^-$.

3.3 Parameter Estimation

A common problem in statistical physics is, that the true values of parameters of a physical problem are generally not known. The estimated set of parameters $\hat{a}$ differs from the true values a_0 by statistical and systematic errors. In many cases the measured data can be regarded as a set of random variables $\vec{x}$, following a probability density function (PDF) $f(\vec{x}|a_0)$. The estimation process tries to estimate the underlying parameters $\hat{a}$ and their errors. Prominent examples are the estimation of the mean value of a distribution or the lifetime τ of an atomic state.

A parameter estimator has to fulfill in general four criteria:

- **Consistency:** $\lim_{N \rightarrow \infty} \hat{a} = a_0$
- **Unbiasedness:** $E[\hat{a}] = a_0$
- **Efficiency:** the variance of $\hat{a}$ has to be as small as possible
- **Robustness:** $\hat{a}$ has to be unaffected by wrong data or hypothesizes

3.3.1 Maximum Likelihood Method

The maximum likelihood method is a general technique for parameter estimation. In case of data following a normal distribution this method is equal to the χ^2 -test, see for instance [35].

More generally, the common task is to determine a set of parameters $\theta = \theta_1, \dots, \theta_n$ for a sample of random variables $\mathbf{x} = \vec{x}_1, \dots, \vec{x}_m$ following a function $f(\vec{x}_i|\theta)$. Usually, $f(\vec{x}_i|\theta)$ is chosen to be a probability density function $\mathcal{P}(\vec{x}_i|\theta)$, which is positive and normalized for all possible parameters θ within their boundaries Ω :

$$\forall \theta : \int_{\Omega} \mathcal{P}(\vec{x}|\theta) d\vec{x} \equiv 1 \quad (3.17)$$

$$\mathcal{P}(\vec{x}_i|\theta) \geq 0. \quad (3.18)$$

For an ensemble of N independent data points, the likelihood function $L(\vec{x}|\theta)$ of the total set is the product of the probability of each data point:

$$L(\vec{x}_1, \dots, \vec{x}_n|\theta) = \prod_{i=1}^N f(\vec{x}_i|\theta). \quad (3.19)$$

With the maximum likelihood method, the estimated parameters $\hat{\theta}$ are chosen to maximize $L(\vec{x}|\theta)$:

$$L(\vec{x}_1, \dots, \vec{x}_n|\hat{\theta}) \geq L(\vec{x}_1, \dots, \vec{x}_n|\theta) \quad \forall \theta. \quad (3.20)$$

For numerical stability in computational algorithms, the logarithm $\mathcal{L}(\theta)$ is calculated and the negative log-likelihood function is minimized. Furthermore, the data might consist of several categories with individual number of expected events and which each category following their own PDF. The parameters can then be obtained by an extended maximum likelihood fit

$$\mathcal{L}(\theta) = \ln L(\theta) = \sum_{j=1}^N \left\{ \sum_{i=1}^{\mathcal{C}} N_i \mathcal{P}_i(\vec{x}_j|\theta) \right\} - \sum_{i=1}^{\mathcal{C}} N_i, \quad (3.21)$$

where $\mathcal{C}$ is the number of categories, $\mathcal{P}_i$ is the PDF of the i^{th} category and N_i is the signal count of the category. In the context of this thesis, this method is used to determine observables from the angular distribution in the decay $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$.

3.4 Classification Methods

The general task of a binary classification is to discriminate between two hypotheses H_0 and H_1 from a set of random variables $\mathbf{x}$ following a multidimensional PDF $f(\mathbf{x})$. The goal is to find a function $\mathbf{x} \rightarrow t$ which reduces the multidimensional problem to a lower dimensional test statistic t , following the distribution $g(t)$. It is chosen in a way to allow for a maximal discrimination between the two hypotheses in $g(t)$, i.e. to maximize the separation between the distributions $g(t|H_1)$ and $g(t|H_0)$. The following sections detail various methods for the construction of the function $g(t)$. The classification is performed by a cut t_{cut} on t where the hypothesis H_1 is accepted for $t > t_{\text{cut}}$ and otherwise rejected. If the classification method is not able to completely separate both classes in t , which is the usual case, one is bound to miss classify some events. These mistakes are called errors of the first and second kind. If

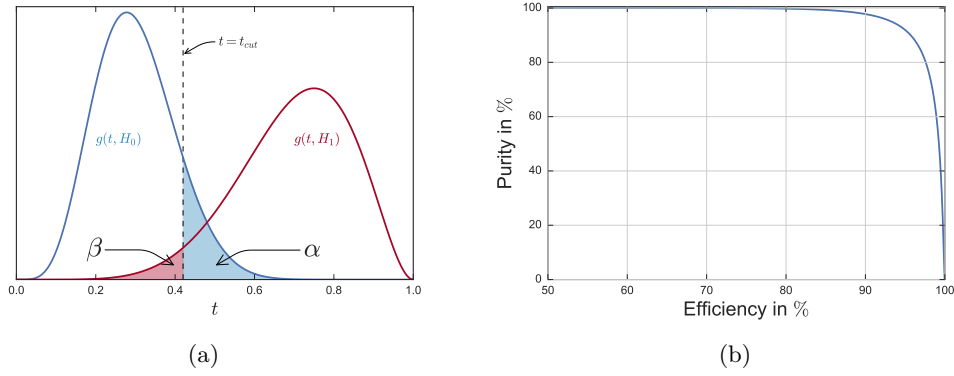


Figure 3.5: Example of a test statistic with the corresponding purity vs. efficiency plot on the right side.

one accepts the wrong hypothesis, one speaks of errors of the first kind, α . Errors of the second kind are, if one rejects the correct hypothesis, with the corresponding probability β . The probabilities for this can be calculated by

$$\alpha = \int_{t_{cut}}^{\infty} g(t, H_0) dt$$

$$\beta = \int_{-\infty}^{t_{cut}} g(t, H_1) dt.$$

This is depicted in fig. 3.5(a). Important measures of a classification method are the purity and efficiency:

$$\text{efficiency} = \frac{N_{(\text{true}|\text{selected})}}{N_{\text{true}}}$$

$$\text{purity} = \frac{N_{(\text{true}|\text{selected})}}{N_{(\text{true}|\text{selected})} + N_{(\text{false}|\text{selected})}}$$

The strengths of a classifier can be displayed in the purity vs efficiency plot, shown in figure 3.5(b). The optimal working point would be (1,1), which obviously can not be reached in non-separable problems as always errors of the first and second kind are made. In all other cases the optimal working point is dependent on the weighting of the errors of first and second kind i.e. the compromise between purity and efficiency. If one compares different methods with each other, the method which provides a higher efficiency at a given purity is preferred.

3.4.1 Cut based Methods

A cut based analysis is the simplest method to separate two classes in a multidimensional variables space $\mathbf{x}$. Sequential cuts are applied to each input dimension in order

to isolate the desired $\mathbf{x} \in H_0$ class from the background class $\mathbf{x} \in H_1$. This way no transformation $\mathbf{x} \rightarrow t$ is needed, but all applied cuts determine a fixed purity/efficiency working point. Moreover, this method does not consider any correlations of the input variables. Cut based methods are a commonly used technique, because they are simple to implement and can be easily comprehended. Cuts are used within this analysis to remove background events in unphysical regions. In general it is tried within this thesis to apply as least cuts as possible where signal candidates would get lost.

3.4.2 Fisher's linear discriminant

The Fisher's linear discriminant, often referred to as Linear Discriminant Analysis LDA, is a simple method to separate classes in a vector space. One constructs a hyperplane in the vector-space $\mathbf{x}$ in order to separate two classes by their mean in units of their variance. To construct the test-statistic t , the variable vectors ($\mathbf{x} = x_1, x_2, \dots, x_n$) are assigned to the corresponding hypothesis: $(\mathbf{x}^{(1)}, \mathbf{x}^{(0)})$. In a next step, the mean of the vectors $\bar{\mathbf{x}}^{(1)}$ and the covariance matrix of each class is calculated

$$V_{ij}^{(1)} = \frac{1}{N} \sum \left(\mathbf{x}_i^{(1)} - \bar{\mathbf{x}}_i^{(1)} \right) \cdot \left(\mathbf{x}_j^{(1)} - \bar{\mathbf{x}}_j^{(1)} \right). \quad (3.22)$$

Accordingly one calculates $\bar{\mathbf{x}}^{(0)}$, $V_{ij}^{(0)}$ and the mean covariance matrix $V_{ij} = \frac{1}{2}(V_{ij}^{(1)} + V_{ij}^{(0)})$. The test statistic t is then calculated by

$$t = \sum_{i=1}^N f_i x_i - \frac{1}{2} \sum_{i=1}^N f_i (\bar{\mathbf{x}}_i^{(1)} + \bar{\mathbf{x}}_i^{(0)}) \quad (3.23a)$$

with

$$f_i = \sum_k \left(V^{-1} \right)_{ik} (\bar{\mathbf{x}}_k^{(1)} - \bar{\mathbf{x}}_k^{(0)}). \quad (3.23b)$$

It can be shown, that this algorithm delivers optimal results in the case of Gaussian distributed input variables X_i and linear correlation [35]. LDAs are one of the simples multivariate methods, they are used within the presented analyses as a reference model for demonstrating the benefits of more advanced techniques.

3.4.3 Neural Networks

Datasets can offer higher order, i.e. non-linear correlations between the signal or background hypothesis $H_{0,1}$ in the data $\mathbf{x}$. One method to detect and benefit from this correlation are artificial neural networks. They were developed in an attempt to simulate the process behind biological nerve cells. Neurons in human brains for example gather electrical stimulations from other brain cells. They fire themselves an electrical pulse to other neurons, if the input signal exceeds a certain threshold. The strength of the signal depends on the connection in between two neurons. It

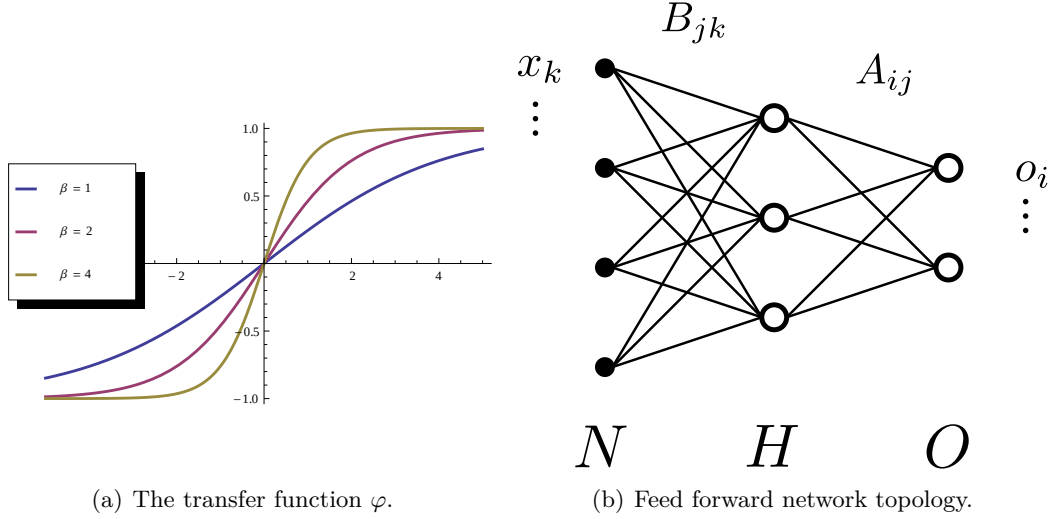


Figure 3.6: Topology of a Neural Network (right). The transfer function (left) with different learning gradients β is used in each perceptron to map the sum of the input to the range $[-1, 1]$.

is believed that the adjustment of these connections implies learning. The artificial neuron is called perceptron and offers the same structure. It sums up incoming signals x_i multiplied by a weight w_i and determines the outgoing signal o with the transfer function φ :

$$o(x) = \varphi \left(\sum x_i w_i \right). \quad (3.24)$$

A common choice for the transfer function is the sigmoid function:

$$\varphi = \tanh(\beta x) = \frac{2}{1 + e^{-2\beta x}} - 1. \quad (3.25)$$

It maps the interval $(-\infty, \infty)$ to $[-1, 1]$ and its gradient β is a key factor in the learning process of the network. Variations of the parameter β can be seen in fig. 3.6(a). This function is approximately linear in the region around zero and turns into a function with exponential character at growing $|x|$, in this region, where φ approaches ± 1 , the perceptron is called saturated. The optimal training conditions are provided in the linear region of φ . The array of several perceptrons forms an artificial neural network, where the topology of this arrangement determines the type of neural network. It has been shown that feed forward networks deliver optimal solutions for classification problems and are widely used in high energy physics. Feed forward networks consist of an input layer, one or several hidden layers H and an output layer, see fig. 3.6(b) for an example topology. The figure shows a network with N input knots, H perceptrons in one hidden layer and O output knots. The weights between the layers B_{jk} and A_{ij} have to be determined during the learning process and refer to the expertise of the classifier. For the learning process one needs data with known truth $(\vec{x}, \vec{t})$. The truth $\vec{t}$ can be received from simulated data

(Monte Carlo simulations) or historical data and is another important factor for the quality of the training. The most frequently used algorithm for the training of neural networks is called back-propagation of errors. It is an iterative process: Each input $\vec{x}_i$ is propagated through the network and the output $\vec{o}_i$ is calculated. The network output is compared to the truth and an error function $E_D(\vec{o}_i, \vec{t}_i)$ is calculated. Finally the weights are updated in order to minimize the error function. The update of the weights can be performed after each event (on-line learning), after the whole training sample (batch learning) or after a certain number of events (quasi on-line learning). A common choice of the error function is the χ^2 error

$$E_D^{\chi^2} = \frac{1}{2} \sum_i (o_i - t_i)^2. \quad (3.26)$$

Alternatively the entropy loss function can be used

$$E_D = \sum_i \ln(1 + o_i t_i). \quad (3.27)$$

The training of the networks is a crucial part of this algorithm and source of many mistakes. As the learning process is a high dimensional minimization process (the quantity of weights $\{w\}$ is equal to this dimension), the global minimum is difficult to reach. It is more likely that the minimization algorithm gets stuck at a local minimum. From this point of view it is advisable to keep N and H and thus $\{w\}$ as low as possible. Another problem is the overfitting or overtraining of the network. If the size of the training sample is in the order of $\{w\}$, the network can learn the sample by rote. At this point the error function can even reach zero but the ability to generalize is lost. For this reason a common technique is to split the dataset into a training sample and a test sample. During the training one compares the error function on the training and test sample. At the point of overfitting, the error function on the test sample stops decreasing and rises.

Preprocessing of the input variables can furthermore improve the training. Various preprocessing steps will be discussed alongside the next section with the NeuroBayes framework as it performs an excellent preprocessing.

3.4.4 The NeuroBayes Framework

The NeuroBayes framework [36] combines robust preprocessing, decorrelation of the dataset and a final neural network for classification. It has shown exceptional results for physical and non-physical problems in terms of robustness, separation power, and performance. The preprocessing steps are performed automatically, so that one can use the algorithm out of the box without modification of the input data.

The first step of the classification is the flattening of each input variable. The variables are drawn into a binned histogram, where the bin size of the histogram is adjusted, so that each bin contains the same amount of events. This stabilizes numerical calculations for further preprocessing steps and allows for a robust estimation of

the underlying probability density function. In the next step the PDF for signal and background is obtained and the purity is calculated for each input variable. All variables are then transformed to their corresponding purity. The last step transforms each purity distribution to a Gaussian shape with mean zero. This step can be achieved by transforming the data with the inverse error function.

The classification is performed by a decorrelation procedure similar to a principal component analysis, in a way that the classifier output owns maximal correlation to the signal and background class. This way all linear correlations are taken into account and the result is optimal in the absence of higher order correlations [36].

The optional last step of the classification is the use of a neural network. This aims to learn residual higher order correlations which are not taken into account from the decorrelation procedure. For the most part in this thesis the neural network mode does not improve the result significantly.

3.4.5 Boosted Decision Trees

In recent years, Boosted Decision Trees became more and more popular and the method of choice for many classification tasks. The key point of their success is the combination of many weak classifier with a technique called boosting. This method shall be explained briefly in the following.

A decision tree classifier divides the training data into two regions, dominated by either signal or background. This is done by consecutively cutting at one variable a time, leaving a signal enhanced or suppressed subset of the data. The cut value is chosen to maximize the separation gain, where commonly the entropy difference is calculated. The method can be represented by a tree like structure, with a cut on a feature at each branch where the number of consecutively performed cuts defines the depth of the tree. As a result, the tree commonly delivers the signal fraction in the accordant region. An example of a decision tree with depth 2 from a training in part III of this thesis is shown in fig. 3.7.

Simple decision trees, especially those with a large depth, are susceptible for overfitting and mostly not powerful classifiers. They became favored as a so-called weak-learners in boosted learning algorithms by limiting their depth. Boosting algorithms combine many weak-learners to one classifier with great separation power. In the training, weights are assigned to the data and the first weak-learner is trained. Afterwards event weights for miss reconstructed events are increased and a new weak-learner is trained. This procedure is usually repeated many times. The final output is derived as a weighted sum of the outputs of all weak-learners in the ensemble.

Boosting algorithms can be classified by the way the weights for the dataset are determined and the ensemble is structured. In this thesis Stochastic Gradient Boosts are used which follow the implementation based on Ref. [37]. The new event weights are calculated as the negative gradient of the loss function L :

$$w_i = -\frac{\partial L}{\partial \sigma(\vec{x}_i)}, \quad (3.28)$$

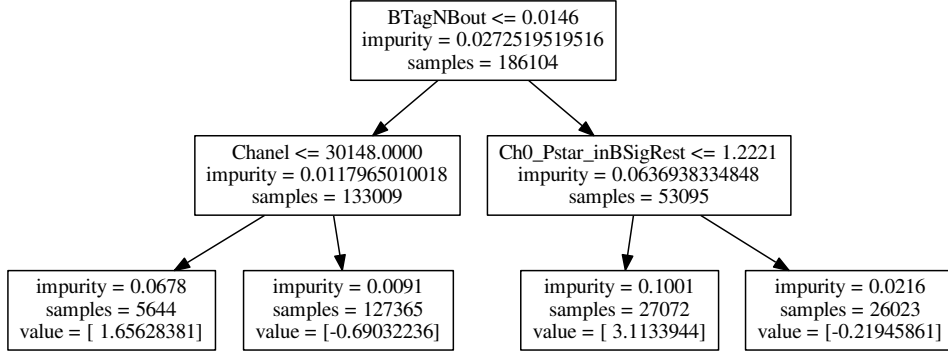


Figure 3.7: Example of a decision tree from a training in part III of this thesis. Cuts are performed to variables of the training-set. The last row states the number of entries in each category, the corresponding impurity and the output value of the tree.

where $\sigma(\vec{x}_i)$ refers to the current classifier output. As a consequence, events which are easy to classify become irrelevant after a few iterations and the method can concentrate on difficult regions. In stochastic learning, the training dataset for each weak-learner is additionally a random subset of the whole training set. This speeds up the training and helps the method to generalize. Important parameters to tune in the training are the depth of the weak-learners, the number of trees in the ensemble and the learning rate. In part III Stochastic Gradient Boosted Decision Trees are used.

3.5 Full Reconstruction Technique

The full reconstruction method is a unique technique suitable for B factories, which allows for the reconstruction of decays which are not fully visible in the detector. This method is used in part III of this thesis to reconstruct signal candidates for $B^+ \rightarrow K^+ \tau^+ \tau^-$. One of the most important advantages of an electron positron collider is that the initial state of the collision is precisely known, since the four-vector of both colliding beams is precisely known. The precise knowledge of the initial state allows for constraints for missing energy and momentum. This method is essential if the final state of interest contains neutrinos, as they can not be directly detected. The $\Upsilon(4S)$ is just above the mass threshold for the decay into two B mesons, so that it can decay directly via the strong interactions into $B^0 \bar{B}^0$ or $B^+ B^-$ in more than 96% of the cases. For the full reconstruction method, the pairwise appearance of B mesons is exploited to find otherwise (partly) hidden decays. A successful reconstruction of one B meson (the tag-side) implies the existence of exactly one other B meson (signal-side). For this it is needed that the tag-side B is reconstructed

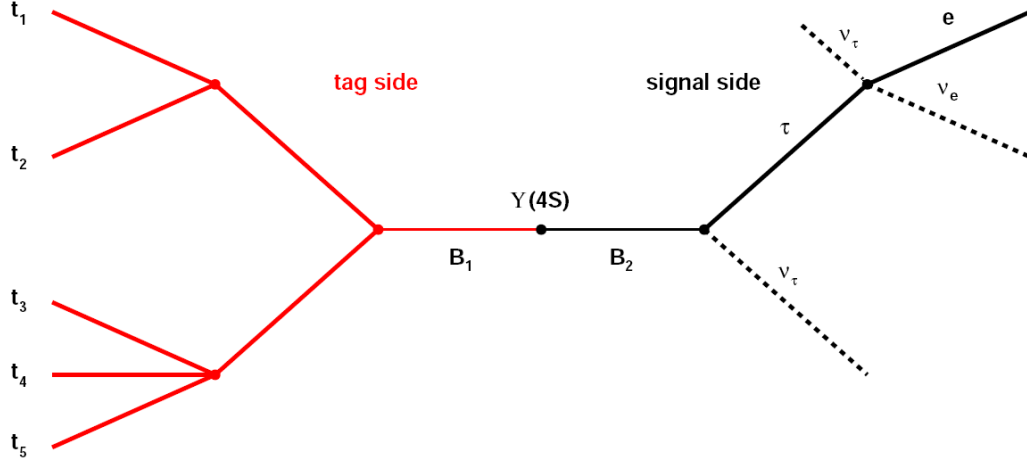


Figure 3.8: Schematic diagram of the full reconstruction.

in purely hardronic modes, hence without neutrinos in the final state. At the KEKB accelerator this is possible because pileup, which means multiple interactions per bunch crossing, is negligible. As the four-momenta $\mathcal{P}$ of the electrons in the HER (High Energy Ring) and LER (Low Energy Ring) are precisely known, energy and momentum conservation laws allow for calculating the four-momentum of the signal side B meson, even without reconstructing one single of its tracks in the detector:

$$\mathcal{P}_{e^+} + \mathcal{P}_{e^-} = \mathcal{P}_{\text{tag-side}} + \mathcal{P}_{\text{signal-side}}. \quad (3.29)$$

In addition to the four-momentum of the signal-side B meson, one can assign all tracks in the detector and hits in the ECL/KLM, which are not used for the reconstruction of the tag-side B meson, to the signal-side B meson. It could be shown that with this technique it is even possible to reconstruct signal-side B mesons with no detectable decay products like the decay $B^0 \rightarrow \nu\bar{\nu}$ [38]. Other examples for measurements which can be performed with this technique are

$$B^+ \rightarrow \tau^+ \nu_\tau, \quad (3.30a)$$

$$B^+ \rightarrow D^{(*)} \tau^+ \nu_\tau, \quad (3.30b)$$

$$B^+ \rightarrow K^+ \nu\nu. \quad (3.30c)$$

In Figure 3.8 this procedure is schematically demonstrated for the decay $B^+ \rightarrow \tau \nu_\tau$. The tag-side B is usually reconstructed in purely hardronic modes, for instance $B \rightarrow DK$. In the illustrated example it was possible to reconstruct the tag-side B with combinations of tracks $t_{1,2,3,4,5}$. The only remaining track originates from the decay of the signal-side B meson in the decay $B \rightarrow \tau \nu$. The precise knowledge of the four-momentum of the signal-side B meson allows for calculating the missing momentum.

In the context of this thesis the decay $B^+ \rightarrow K^+ \tau^+ \tau^-$ is reconstructed using the full reconstruction technique. The neural network based framework at Belle [39] uses in

total 1104 exclusive hadronic decay modes of B mesons for the reconstruction using 71 NeuroBayes [40] neural networks for classification. Compared to the previous cut based method, the efficiency could be increased by almost a factor of 2.

The total efficiency of this technique is however limited by the fraction of pure hadronic decays. The majority of B mesons decays semileptonically, thus these events can not be used for a full reconstruction. In total, one can correctly reconstruct B^+ or B^0 candidates in 0.28% or 0.18% of all $B\bar{B}$ events respectively.

Part II

Angular Analysis of $B \rightarrow K^* \ell \ell$

4. Reconstruction of $B \rightarrow K^{(*)}\ell^+\ell^-$

The first analysis in this thesis covers the muon and electron modes of $b \rightarrow s\ell^+\ell^-$ with the reconstruction of $B \rightarrow K^{(*)}\ell^+\ell^-$. In total 8 decay modes are exclusively reconstructed in 12 final state configurations. This chapter details the reconstruction procedures and background suppression methods. An angular analysis is performed with the combined data of the decay modes $B^0 \rightarrow K^*(892)^0\mu^+\mu^-$ and $B^0 \rightarrow K^*(892)^0e^+e^-$. For this, an in-depth study is performed with simulated data presented in chapter 6. All procedures are tested in advance with Monte Carlo toy studies, detailed in section 6.3. Sources of systematic uncertainties are evaluated in chapter 7. Finally, results on data are extracted, validated and compared to Standard Model predictions and previous measurements, presented in chapter 8.

4.1 Analysis Overview

The decay $B \rightarrow K^{(*)}\ell^+\ell^-$ is reconstructed exclusively in 12 final states, where $\ell = e, \mu$. An overview of the used decay channels is given in table 4.1. The charged conjugated (cc.) mode is always implied if not explicitly stated otherwise. The reconstruction procedure aims to reach an optimal signal efficiency and purity in order to gain sufficient statistics for a full angular analysis. For this reason there are no strong selection constraints applied on primary particles. All information of the decay is gathered at the stage of the final reconstruction of the B meson candidate, where multivariate data analysis techniques are used to determine its quality. The procedures and data analysis tools used are explained in section 4.4. In case of multiple candidates per event, the decay channel with the highest probability is selected.

4.1.1 Data

This analysis is performed on the full Belle dataset corresponding to 711 fb^{-1} (detailed in section 2.4). Beforehand the analysis is performed on simulated data where both the official Belle generic MC is used and dedicated sets of signal MC are generated.

Table 4.1: Decay channels of the exclusive reconstruction of $B \rightarrow K^{(*)}\ell^+\ell^-$, where $\ell = e, \mu$ and cc. is implied.

B^+	B^0
$B^+ \rightarrow K^{*+}(K^+\pi^0)\ell^+\ell^-$	$B^0 \rightarrow K^{*0}(K_S\pi^0)\ell^+\ell^-$
$B^+ \rightarrow K^{*+}(K_S\pi^+)\ell^+\ell^-$	$B^0 \rightarrow K^{*0}(K^+\pi^-)\ell^+\ell^-$
$B^+ \rightarrow K^+\ell^+\ell^-$	$B^0 \rightarrow K_S\ell^+\ell^-$

Simulated signal events are generated with a decay model in the EvtGen generator using a routine to implement $b \rightarrow s\ell\ell$ decays according to Ref. [41]. A large set of signal events is generated for all modes shown in table 4.1. Additionally, a so-called phase-space MC is generated, in which all angular distributions of the decay products are uniformly distributed.

Table 4.2: Branching fraction of $B \rightarrow K^{(*)}\ell^+\ell^-$ and expected number of signal events in Belle MC based on PDG ([4]).

Type	$\mathcal{B}$	Expected signal candidates
$B^+ \rightarrow K^+e^+e^-$	$(5.5 \pm 0.7) \times 10^{-7}$	424
$B^+ \rightarrow K^+\mu^+\mu^-$	$(4.49 \pm 0.23) \times 10^{-7}$	346
$B^+ \rightarrow K^*(892)^+e^+e^-$	$(1.55^{+0.4}_{-0.31}) \times 10^{-6}$	1195
$B^+ \rightarrow K^*(892)^+\mu^+\mu^-$	$9.6 \pm 1.0) \times 10^{-7}$	741
$B^0 \rightarrow K^0e^+e^-$	$(1.6^{+1.0}_{-0.8}) \times 10^{-7}$	123
$B^0 \rightarrow K^0\mu^+\mu^-$	$(3.4 \pm 0.5) \times 10^{-7}$	262
$B^0 \rightarrow K^*(892)^0e^+e^-$	$(1.03^{+0.19}_{-0.17}) \times 10^{-6}$	795
$B^0 \rightarrow K^*(892)^0\mu^+\mu^-$	$(1.05 \pm 0.10) \times 10^{-6}$	811

Two streams of MC are used for the training of the background suppression classifiers. One independent stream of generic MC is used for evaluating all methods. For calculating the expected purity of the candidates sample, generic MC is mixed with the number of signal candidates that are expected in the Belle dataset. These expectations are calculated based on the branching ratios from PDG [4] scaled according to the total number of $B\bar{B}$ pairs in the Belle dataset. The result can be seen in table 4.2.

4.2 Particle Selection

The charged particles $e^\pm, \mu^\pm, \pi^\pm$ and $K^\pm$ and the neutral particles K_S, π^0 and γ can be directly reconstructed in the detector. They are the first stage in the reconstruction and in the following are referred to as primary particles.

For all charged tracks loose impact parameter constraints are applied to the nominal interaction point in the radial direction of $|dr| < 1.0$ cm and along the beam direction

of $|dz| < 5.0$ cm. Belle provides a particle identification (PID) calculated from a variety of signals in the detector components. Soft cuts are applied on the particle identification system to remove highly unlikely candidates. The PID system is described in section 2.3.2. Electrons are identified by the electron ID (eid). The eid provides a likelihood ratio $\mathcal{P}_{eid}(e) = L(e)/(L(e) + L(\text{hadron}))$. A soft cut is applied and all charged tracks satisfying $\mathcal{P}_{eid}(e) > 0.1$ are accepted as electrons. High energetic electrons can emit photons from bremsstrahlung. To recover the original momentum of the electrons, a search for photons in a cone of 0.05 radians around the initial momentum direction of the track is performed. If photons are found in this region, their momenta are added to the electron. This is referred to as bremsstrahlung recovery process. For $\mu^\pm$ candidates there is a muon-id (muid) separating muons from other particles detected in the KLM. Again, a soft cut is applied and all charged particle candidates with $\mathcal{P}_{muid}(\mu) > 0.1$ are accepted as muons. Candidates for $K^\pm$ are separated against pions through information from the TOF, ACC and CDC. A loose selection of $\mathcal{P}(K/\pi) = L(K)/(L(K) + L(\pi)) > 0.1$ is required. For the $\pi^\pm$ candidates no PID selection is applied. Candidates for K_S are formed from two oppositely charged tracks with the mass assumption of a charged pion. Candidates for neutral pions are reconstructed from two photons, each required to have $E_\gamma > 30$ MeV. Furthermore, candidates are required to have an invariant mass of $115 \text{ MeV}/c^2 < M_{\gamma\gamma} < 153 \text{ MeV}/c^2$. An overview of all particle selection criteria is shown in table 4.3.

Table 4.3: Summary of the particle selection criteria.

	Selection	cut value
Charged tracks	$ dr $	< 1.0 cm
	$ dz $	< 5.0 cm
	p_t	$< 10.0 \text{ GeV}/c$
$e^\pm$ candidate	$\mathcal{P}_{eid}(e)$	> 0.1
$\mu^\pm$ candidate	$\mathcal{P}_{muid}(\mu)$	> 0.1
$K^\pm$ candidate	$\mathcal{P}(k/\pi)$	> 0.1
$\pi^\pm$ candidate	no selection	
K_S candidate	K_S finder class	accept status good K_S
π^0 candidate	E_γ	$> 30 \text{ MeV}$
π^0 candidate	$M_{\gamma\gamma}$	$115 \text{ MeV}/c^2 < M_{\gamma\gamma} < 153 \text{ MeV}/c^2$

4.3 Event Selection

In the event selection loose cuts are applied on masses and energies to exclude unphysical regions for $B \rightarrow K^{(*)}\ell^+\ell^-$ and vetoes are applied to remove irreducible sources of background. After this step the main suppression of background is performed, detailed in the next sections.

In the second stage of the reconstruction, K^* candidates are formed in four channels from $K^\pm, K_S, \pi^\pm$ and π^0 candidates. The reconstructed channels are:

$$\begin{aligned} K^{*0} &\rightarrow K^+\pi^-, \\ K^{*0} &\rightarrow K_S\pi^0, \\ K^{*+} &\rightarrow K^+\pi^0 \text{ and} \\ K^{*+} &\rightarrow K_S\pi^+. \end{aligned}$$

For these candidates, an invariant mass cut of $0.6 \text{ GeV}/c^2 < M_{K^*} < 1.4 \text{ GeV}/c^2$ is applied and a vertex fit is performed, which is used for background suppression later on.

In the final stage of the reconstruction $K^{(*)}$ candidates are combined together with oppositely charged lepton pairs to form B meson candidates. Large amounts of random combinations are rejected by cuts on kinematic variables. Two independent variables can be constructed using constraints from the condition that in $\Upsilon(4S)$ decays B mesons are produced pairwise and carry half the center-of-mass (CM) frame beam energy, E_{Beam} , each¹. They are the beam constrained mass, M_{bc} , and the energy difference, ΔE , in which signal features a distinct distribution that can discriminate against background. The variables are defined in the $\Upsilon(4S)$ rest frame as

$$M_{\text{bc}} \equiv \sqrt{E_{\text{Beam}}^2 - |\vec{p}_B|^2} \text{ and} \quad (4.1)$$

$$\Delta E \equiv E_B - E_{\text{Beam}}, \quad (4.2)$$

where E_B and $|\vec{p}_B|$ are the energy and momentum of the reconstructed candidate respectively. Correctly reconstructed candidates are located around the nominal B mass in M_{bc} and feature ΔE of around zero. Candidates are selected satisfying $5.22 < M_{\text{bc}} < 5.3 \text{ GeV}/c^2$ and $-0.10 (-0.05) < \Delta E < 0.05 \text{ GeV}$ for $\ell = e$ ($\ell = \mu$).

Large contributions of irreducible background arises from charmonium decays $B \rightarrow K^{(*)}J/\psi$ and $B \rightarrow K^{(*)}\psi(2S)$, where the $c\bar{c}$ state decays into two leptons. These decays own the same signature as the desired signal and are vetoed with cuts on the invariant mass of the lepton pair

$$-0.25 \text{ GeV}/c^2 < M_{ee(\gamma)} - M_{J/\psi} < 0.08 \text{ GeV}/c^2, \quad (4.3)$$

$$-0.15 \text{ GeV}/c^2 < M_{\mu\mu} - M_{J/\psi} < 0.08 \text{ GeV}/c^2, \quad (4.4)$$

$$-0.20 \text{ GeV}/c^2 < M_{ee(\gamma)} - M_{\psi(2S)} < 0.08 \text{ GeV}/c^2 \text{ and} \quad (4.5)$$

$$-0.10 \text{ GeV}/c^2 < M_{\mu\mu} - M_{\psi(2S)} < 0.08 \text{ GeV}/c^2. \quad (4.6)$$

These veto regions are displayed in fig. 4.1. In the electron case, additionally photon energies of detected photons from the bremsstrahlung recovery process are added.

¹Assuming the environment of a B -factory such as Belle and BaBar.

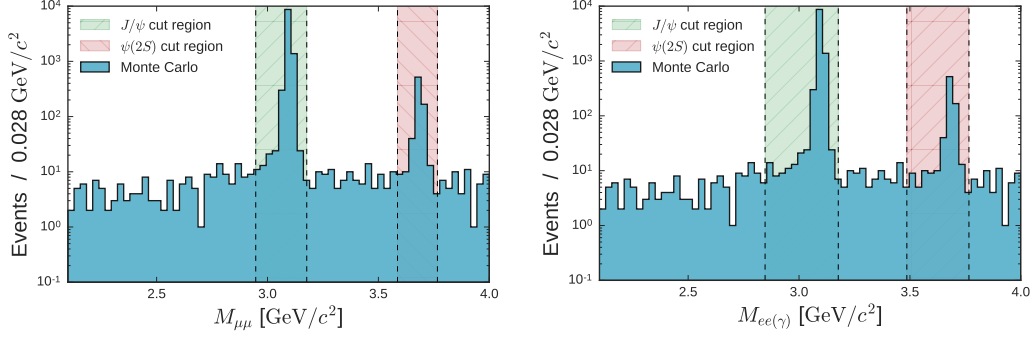


Figure 4.1: Veto regions for charmonium background for the di-electron (right) and di-muon (left) channels.

Di-electron background can also arise from photon conversion $\gamma \rightarrow e^+e^-$ and π^0 Dalitz decays ($\pi^0 \rightarrow e^+e^-\gamma$). In order to eliminate this source of background the constraint $M_{e(\gamma)e(\gamma)} > 0.14 \text{ GeV}/c^2$ is required.

For the B meson candidates, a vertex fit is performed, which is used for background suppression and also to define the distance between the two leptons along the beam direction $\Delta z_{\ell\ell}$. All cuts and vetoes are listed in table 4.4.

Table 4.4: Summary of the event selection criteria.

Selection			
Charmonium veto	$-0.25 < M_{ee(\gamma)} - M_{J/\psi} < +0.08 \text{ GeV}/c^2$		
	$-0.15 < M_{\mu\mu} - M_{J/\psi} < +0.08 \text{ GeV}/c^2$		
	$-0.20 < M_{ee(\gamma)} - M_{\psi(2S)} < +0.08 \text{ GeV}/c^2$		
	$-0.10 < M_{\mu\mu} - M_{\psi(2S)} < +0.08 \text{ GeV}/c^2$		
π^0 Dalitz $/\gamma \rightarrow e^+e^-$	$M_{ee} > 0.14 \text{ GeV}/c^2$		
K^* mass	$0.6 /c^2 < M_{K^*} < 1.4 \text{ GeV}/c^2$		
Beam constrained mass	$5.22 < M_{bc} < 5.3 \text{ GeV}/c^2$		
Energy difference	$-0.10 < \Delta E(\ell = e) < +0.05 \text{ GeV}$		
	$-0.05 < \Delta E(\ell = \mu) < +0.05 \text{ GeV}$		

4.4 Background Suppression Strategy

In this analysis, only soft cuts are applied in the first place to keep most signal candidates. With this approach it is tried to eliminate background as late as possible in the reconstruction of B mesons. This leads to a large amount of different sources of possible backgrounds. Multivariate data analysis techniques are used to combine all available information of a B meson candidate in order to separate signal from background.

A hierarchical framework is used in the reconstruction, starting in the first stage with the primary particles from tracks ($e, \mu, K^\pm$, and $\pi^\pm$) and the neutral particles (K_S, π^0 and γ). In the second step combinations to K^* particles are performed and in the last stage the final B meson candidates are constructed.

In each stage, all particle candidates are analyzed with a neural network (NeuroBayes [36]) and an output NB_{out} is assigned. This output is chosen to correspond to a Bayesian probability in the range $[0, 1]$ where 1 corresponds to the candidate being true signal. In this manner a network output of $NB_{out} = 0.9$ corresponds to a probability of 90% for the candidate to be a true signal event. To transfer quality information about the primary particles in the detector to higher stage composite particles (K^* and B) the network output of the secondary or children particles of each candidate is included to its neural network input. One of the most important variables in the training of B meson candidates is the combined probability for all children corresponding to their correct hypothesis,

$$NB_{\text{Prod}} = \prod_{\text{child } i} NB_{out,i}. \quad (4.7)$$

4.4.1 Sources of Backgrounds

In the reconstruction several sources of background are considered. In the particle and event selection, cuts are introduced, preserving almost all the signal and excluding regions without signal. The probability of a signal event is in the order of $\mathcal{O}(10^{-6})$, which means that a huge amount of background events exists for each signal candidate.

In this analysis the following sources of backgrounds are considered:

Continuum In continuum events, e^+e^- annihilates into light quark pairs $u\bar{u}, d\bar{d}, s\bar{s}$ as well as events containing charm quarks $c\bar{c}$. These initial quark pairs however exhibit a large energy release, forming back to back jet-like structures.

Combinatorial Combinatorial background arises from wrong combination of tracks in B decays. This is the dominant source of background.

Peaking A process is considered as peaking background when it mimics the signal shape in M_{bc} . For the peaking background several sources have to be considered: First, irreducible background exists from $B \rightarrow K^*J/\psi$ and $B \rightarrow K^*\psi(2S)$ events, which passes the q^2 vetoes. Secondly, double mistag events from $B \rightarrow K^*\pi\pi$ can occur, where both pions are misidentified as muons.

Cross-feed Here a candidate is assigned to the wrong decay channel hypothesis by missing some decay products or misreconstruction of one of its children. This source is analyzed in the systematics in chapter 7 in detail.

4.5 Neural Network Trainings

The neural network trainings are performed in several stages since the composite particles like K^* or B mesons use the network output of their children as an input. Consequently, the networks for the primary particles have to be trained first.

4.5.1 The Prior Probability

The framework uses a prior probability for each candidate for being signal or background. Rare decay channels have a lower prior probability than decay channels with high branching fractions.

When training the neural networks signal MC is mixed with generic MC. The trainings are performed with almost equal sizes for the signal and background class. As a result the signal to background ratio is much higher than expected in real data. Consequently, in the trainings the probability $P_t(S)$ for signal and background $P_t(B)$ is different to the probability for signal (background) in the data $P_p(S)$ ($P_p(B)$). The trained classifier delivers a probability o_t valid only with the prior probabilities for signal and background from the training sample. However, one is interested in the probability o_p , which correspond to a Bayesian probability of the particle to be true with respect to the branching fraction of the decay in real data. Predictions from PDG are used and the number of $B\bar{B}$ pairs in the Belle data sample for the expected amount of signal candidates, see table 4.2.

To account for this one can calculate a correction with the Bayes theorem, which described in section 3.1. The calculation of the correction for this probability can be found in Ref. [42] resulting in

$$o_p = \frac{1}{1 + (\frac{1}{o_t} - 1) \frac{P_p(B)}{P_p(S)} \frac{P_t(S)}{P_t(B)}}. \quad (4.8)$$

This formula transforms the NeuroBayes output with the unweighted signal to background ratio into the correct Bayesian probability.

4.5.2 Primary Particles - Stage One

In the first stage, the primary particles and K_S are reconstructed:

$$e, \mu, \pi^\pm, \pi^0, K^\pm, K_S, \gamma.$$

The classifiers in this stage are trained on generic MC to separate wrong and right hypotheses of the particle type. This stage uses the same expertise as in the well tested neural network based full reconstruction, widely used at Belle [42]. The classifiers use kinematic variables as input as well as variables derived from the particle identification system, for instance TOF and KLM information and energy loss in the CDC. A detailed description of all variables and the training procedure can be found in Ref. [42]. Lists of these particles alongside their probability of being the right hypothesis are passed into the next stage. The output of the classifiers on generic MC can be seen in fig. 4.2.

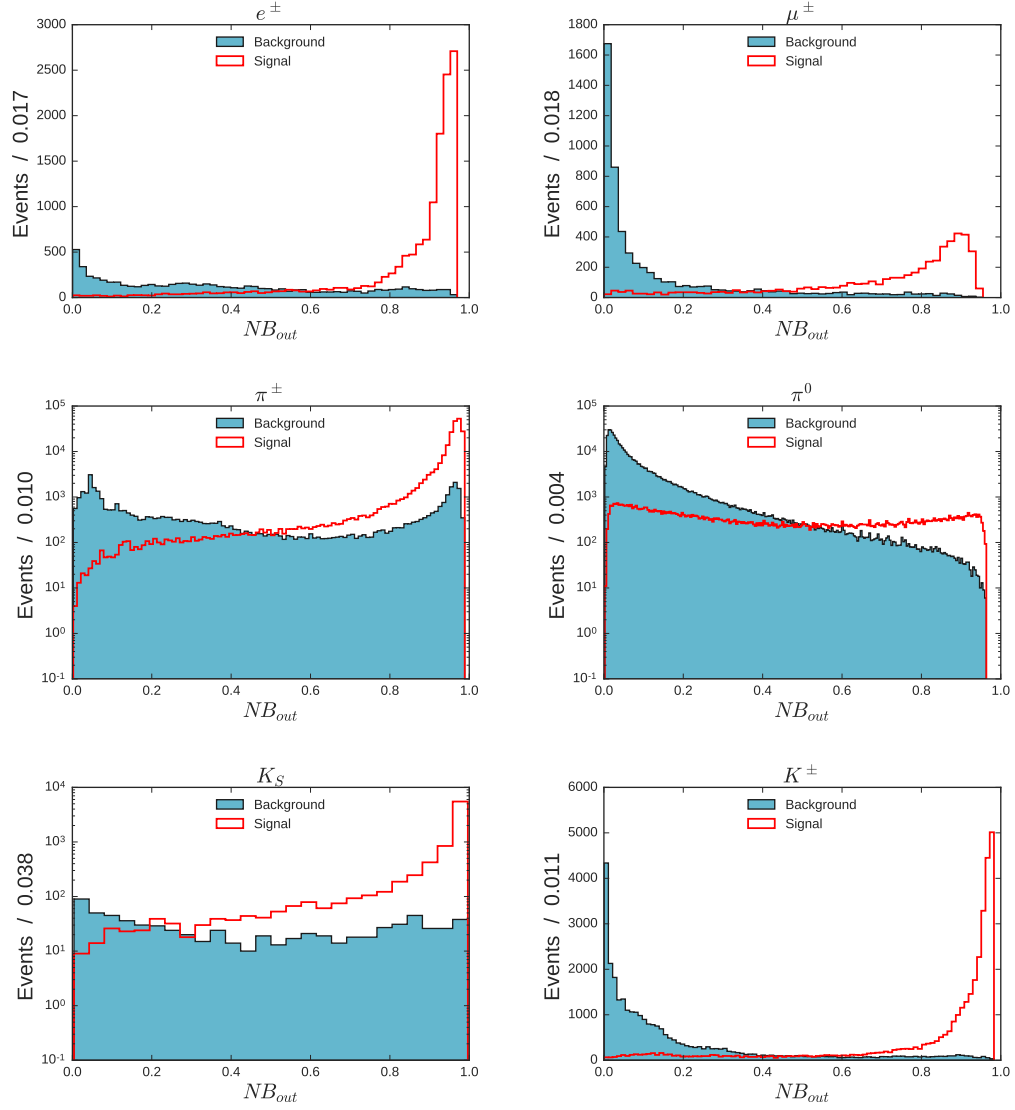


Figure 4.2: Network output for the classifiers in stage 1 for signal and background on a set of 10,000 generic MC events after particle selection and before event selection.

4.5.3 Training of K^* - Stage Two

In the second stage, K^* candidates are reconstructed in the decay channels

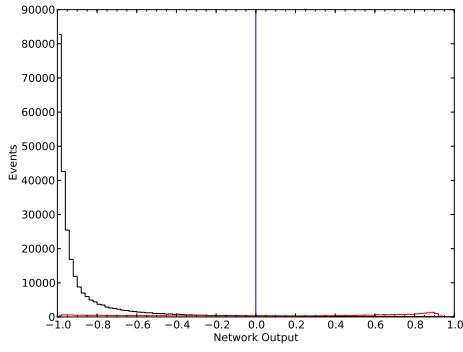
$$\begin{aligned} K^{*+} &\rightarrow K^+\pi^0, \\ K^{*+} &\rightarrow K_S\pi^+, \\ K^{*0} &\rightarrow K_S\pi^0 \text{ and} \\ K^{*0} &\rightarrow K^+\pi^-. \end{aligned}$$

Also in this stage, a NeuroBayes classifier is trained to calculate a probability for each hypothesis. The signal class consists of truly reconstructed K^* from the signal MC and the background class of all misreconstructed K^* of the generic MC. This means that the prior probability for each K^* decay mode has to be adjusted to the correct ratio in the MC. Variables used in the trainings are described in table 4.5. The most important variables in the training are the momenta of the decay products (children of the K^*), the product of the network output of the children and the χ^2 value of the vertex fit. The classifier does not have information about the mass of the K^* candidate. This information is only used in the classification of the B meson candidate.

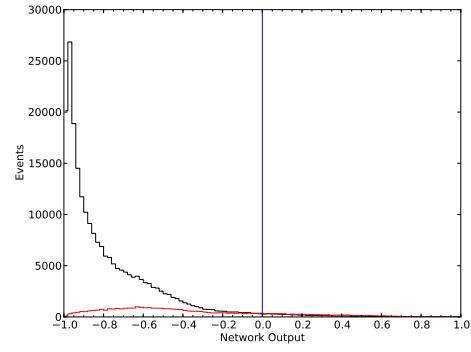
Table 4.5: Variables used in the neural networks for the K^* candidates.

Variable	Description
p_tot	p_{tot} (total momentum) of the candidate
ChX_NBout	Network output of child X
ChX_Pseudo_Hel_Ang	Pseudo helicity angle of child X
ChX_ptot	p_{tot} of child X
sum_Chld_NB	Sum of the NeuroBayes outputs
prod_Chld_NB	Product of the NeuroBayes outputs
mom_dir_dev	Momentum direction deviation
sig_dist_to_IP	Significance of distance to interaction point
chi2	χ^2 value of the vertex fit of the daughters
dist_to_IP	Distance to interaction point

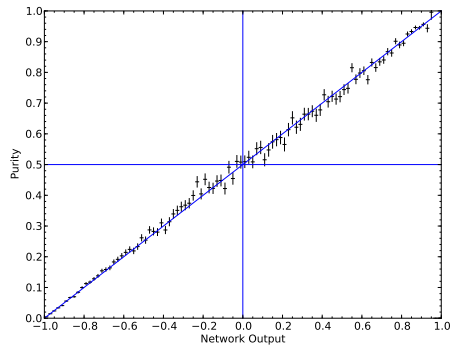
In figs. 4.3 and 4.4 the result of the classification can be seen as well as the purity vs. the network output. The purity aligns well with the diagonal in the latter. In this case, the transformed network output NB_{out} , which is rescaled to $[0, 1]$ can be interpreted as a Bayesian probability for a signal candidate. In figs. 4.3(d) and 4.4(d) one can observe slight deviations of the distribution from the diagonal, which is not severe for the analysis.



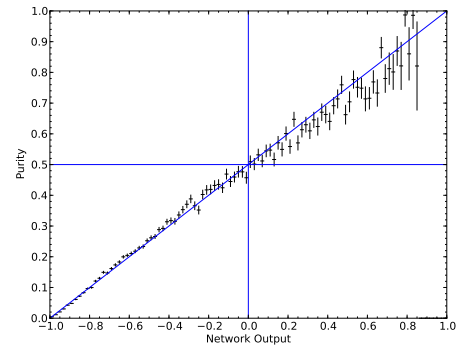
(a) NB_{out} distribution for $K^{*+} \rightarrow K^+\pi^0$.



(b) NB_{out} distribution for $K^{*+} \rightarrow K_S\pi^+$.



(c) Network diagonal plot for $K^{*+} \rightarrow K^+\pi^0$.



(d) Network diagonal plot for $K^{*+} \rightarrow K_S\pi^+$.

Figure 4.3: Network output for signal (red) and background (black) in the first row and purity vs. network output (black data points) in the second row for the trainings of NeuroBayes for the charged K^* modes, produced by the NeuroBayes trainings-report.

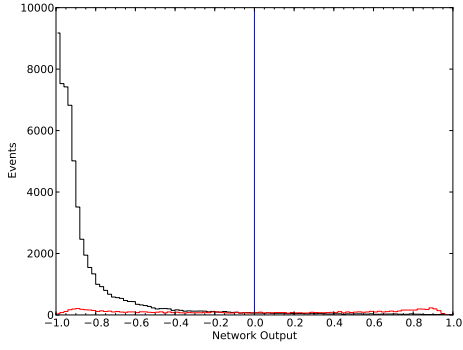
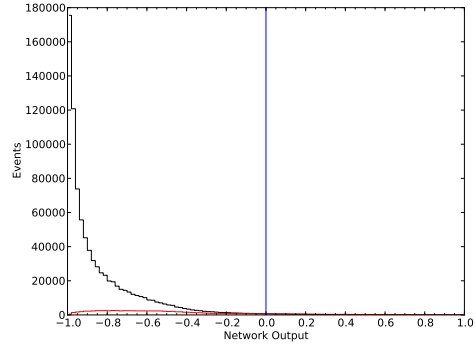
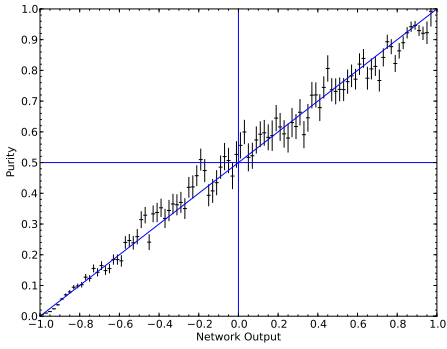
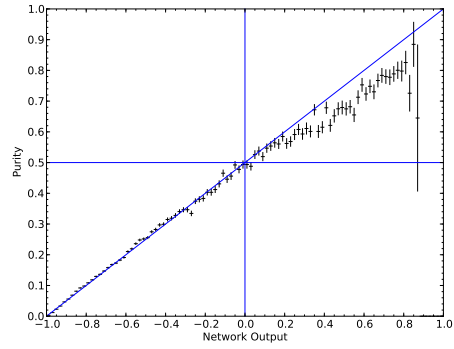

 (a) NB_{out} distribution for $K^{*0} \rightarrow K_S \pi^0$.

 (b) NB_{out} distribution for $K^{*0} \rightarrow K^+ \pi^-$.

 (c) Network diagonal plot for $K^{*0} \rightarrow K_S \pi^0$.

 (d) Network diagonal plot for $K^{*0} \rightarrow K^+ \pi^-$.

Figure 4.4: Network output for signal (red) and background (black) in the first row and purity vs. network output (black data points) in the second row for the trainings of NeuroBayes for the neutral K^* modes, produced by the NeuroBayes trainings-report.

4.5.4 B Meson Trainings - Final Stage

In the final stage of the background reduction, a neural network is trained for each decay mode in the reconstruction for the following channels:

$$\begin{aligned}
 B^+ &\rightarrow K^+ e^+ e^-, \\
 B^+ &\rightarrow K^+ \mu^+ \mu^-, \\
 B^+ &\rightarrow K^*(892)^+ e^+ e^-, \\
 B^+ &\rightarrow K^*(892)^+ \mu^+ \mu^-, \\
 B^0 &\rightarrow K^0 e^+ e^-, \\
 B^0 &\rightarrow K^0 \mu^+ \mu^-, \\
 B^0 &\rightarrow K^*(892)^0 e^+ e^- \text{ and} \\
 B^0 &\rightarrow K^*(892)^0 \mu^+ \mu^-.
 \end{aligned}$$

At this point of the reconstruction, all information of the quality of the reconstructed candidates is available and fed into the network. The training of the final stage of the reconstitution follows the same procedure as for the K^* trainings. Two oppositely charged leptons are combined with a $K^{(*)}$ meson for B -meson candidates. Candidates passing the event selection criteria (see table 4.4) are evaluated by the neural networks and assigned the final network output on which signal events are selected.

Compared to the trainings of K^* , this stage has a larger set of variables:

ΔE : the energy difference between the B candidate and half of the beam energy, defined in eq. (4.2).

NB_{child_i} : NeuroBayes output of the i th child of the B candidate. It can be seen as a probability that the hypothesis for the child is correct.

$\prod_i NB_{\text{child}_i}$: Product of the NeuroBayes output of the children of the B candidate. It can be seen as the combined probability that all children have been assigned the correct hypothesis. This variable is one of the most important variables in the training of the main classifier.

$\sum_i NB_{\text{child}_i}$: Sum of the NeuroBayes outputs of the children of the B meson candidate.

qr_{NN} : Result of the neural network flavor tagging algorithm.

qr_{LR} : Result of the multidimensional likelihood ratio flavor tagger.

$\cos \theta_B$: Cosine of the angle θ_B between the B candidate and the beam direction.

$\Delta \vec{P} \vec{X}$: The momentum direction deviation, i.e. the difference between the momentum vector and the flight direction of the B candidate in the xy plane.

χ^2 : Value of the vertex fit of the candidate.

d_{IP} : Distance of the candidate to the interaction point.

$\sigma(d_{IP})$: Significance of the distance to the interaction point, derived from the error of the vertex fit and the beam spot size.

$\Delta z_{\ell\ell}$: Distance between the two leptons in the z direction. For true signal candidates the value is around zero.

$p_{t,i}$: Transverse momentum of child i .

E_{vis} : Visible energy of the event. For the calculation, all tracks are assumed to be pions and the energies of all photons are summed up. Semileptonic decays tend to have lower visible energy due to neutrinos in the final state.

M_{miss} : Missing mass of the event is calculated by the mass of $\mathcal{P}_{Beam} - \mathcal{P}_{Event}$, where $\mathcal{P}_{Beam}$ is the four-vector of the center-of-mass system, and $\mathcal{P}_{Event}$ is the four-vector of all tracks (assumed to be pions) and photons in the Event.

KSWF moments : In total 18 Super Fox-Wolfram moments are used at Belle. They are described in section 4.6.

R_2 : Ratio of the zeroth and the second Fox-Wolfram moments (see section 4.6).

4.6 Continuum Suppression

Continuum events contribute significantly to the background in this analysis. Several observables and methods are used for continuum suppression. The information is combined into the main neural networks for background suppression.

All observables use the fact that events from continuum have different kinematic shapes. The beam energy at the $\Upsilon(4S)$ mass is close to the mass-threshold for the production of two B mesons. Consequently, in $\Upsilon(4S)$ decays the B mesons are produced nearly at rest in the center-of-mass system (CMS), so that these events tend to be nearly isotropic in the detector. In contrast, events from continuum exhibit a large energy release and have a preferred spatial distribution, forming back to back jet-like structures.

In the following paragraphs the used variables are explained.

4.6.1 Fox-Wolfram Moment R_2

The Fox-Wolfram moments [43] use all tracks in the event and describe the angular distributions using Legendre polynomials. The moments are calculated by

$$H_k = \sum_{i,j}^N \frac{|\vec{p}_i| |\vec{p}_j| P_k(\theta_{ij})}{E_{vis}^2}, \quad (4.9)$$

where N is the number of charged particles in the event, $|\vec{p}_l|$ the momentum of the charged particle l , P_k is the k th Legendre polynomial, θ_{ij} the angle between the i th and j th particle and E_{vis} the total visible energy of the event. The ratio

$$R_k = \frac{H_k}{H_0} \quad (4.10)$$

can be formed, where R_2 has proven to be especially effective in separating $B\bar{B}$ from continuum events.

4.6.2 Super Fox-Wolfram Moments

The Super Fox-Wolfram moments [44] are a modification of the normal Fox-Wolfram moments. The summation over all particles is split into groups of the combination of final state particles belonging to the B meson candidate (s) and the remaining tracks (o). With this separation one can form in principle infinite sets of variables of the three types $R_k^{ss}, R_k^{so}, R_k^{oo}$. In total 18 variables of this kind are used in this analysis for continuum suppression.

4.7 Best Candidate Selection

Due to a loose pre-cut selection, multiple candidates for a $B^\pm$ (B^0) meson are observed in 31% (32%) of the events on signal MC. In this case, one has to choose one candidate, based on certain criteria, which will be discussed in this section. The mass of the B meson cannot be used as it is correlated with the M_{bc} distribution, which will be fitted and should therefore remain unbiased. A common way is to choose the ΔE variable for the best candidate selection. A correctly reconstructed B features ΔE around zero. If multiple candidates occur, one takes the candidate with ΔE closest to zero. The framework described in the previous section also allows for the selection of the most probable candidate, based on the neural network output NB_{out} . In this case one takes the candidate with the maximum network output. The neural network rank is calculated across all decay channels of a particle, for instance for all four modes of the B^0 . This means, if one requires the B meson to have the highest rank, events for some decay channels are missed because only one of the four decay channels has rank one. The result of the neural network based best candidate selection is presented in table 4.6. One can observe that the neural network output for the best candidate selection delivers the best results. Consequently, $\max(NB_{out})$ is required for the signal candidate. This selection performs better across all decay channels compared to a selection on ΔE and does not introduce a correlation to M_{bc} , which is demonstrated in section 4.9.

4.8 Peaking Backgrounds

Potential peaking background in M_{bc} can arise in B decays, when one or more of the selected particles are misidentified. For instance if a π^+ is misidentified as a ℓ^+ . The largest contribution comes from $B \rightarrow D(K^*\pi)\pi$, where both pions are misidentified and accepted as muons. A cut on the mass of the K^+/ℓ^- system around the nominal D meson mass could veto this contribution. To search for this source of background, the truth of the lepton hypothesis in the background sample is monitored on simulated data. One can see in fig. 4.5 that there is no peaking background from particles wrongly classified as leptons.

Table 4.6: Best candidate selection efficiency as derived from signal MC before prior probability is adjusted.

Channel	Random	NBout	ΔE
$B^+ \rightarrow K^+ e^+ e^-$	0.335	0.969	0.714
$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.376	0.990	0.945
$B^+ \rightarrow K^*(892)^+ e^+ e^-$	0.297	0.828	0.732
$B^+ \rightarrow K^*(892)^+ \mu^+ \mu^-$	0.318	0.841	0.744
$B^0 \rightarrow K^0 e^+ e^-$	0.400	0.907	0.778
$B^0 \rightarrow K^0 \mu^+ \mu^-$	0.310	0.993	0.942
$B^0 \rightarrow K^*(892)^0 e^+ e^-$	0.400	0.907	0.778
$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	0.417	0.920	0.832

4.8.1 Rare B Decays

The Belle «rare MC» contains $b \rightarrow s$ processes enhanced by a factor of 50 compared to the total luminosity. The signal decay $B \rightarrow K^{(*)} \ell^+ \ell^-$ is also included and thus has to be excluded when looking at peeking background from this source. All contributions in M_{bc} and $q^2 \equiv M_{\ell^+ \ell^-}^2$ from rare MC after removing the signal can be seen in fig. 4.6. The dominant source of background is $B \rightarrow K^* \pi \pi$ where both pions are misidentified as muons. For this reason almost no background of this source is observed in the $B^0 \rightarrow K^*(892)^0 e^+ e^-$ sample. In the $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ sample 297 candidates are observed, which correspond to 5.94 expected candidates on the whole Belle dataset.

4.9 Evaluation of the Reconstruction Performance

The reconstruction efficiency is the main quantity for the reconstruction performance. However, in addition to a high reconstruction efficiency it is also required that the neural network output is not correlated with the beam constrained mass M_{bc} . This variable is used to extract signal and background yields in the analysis and must remain unbiased. Generic background events do not peak in M_{bc} and can be well distinguished from signal events. If, however, the neural network learns correlations to the mass of the B meson from the input variables, background can peak and look signal-like in M_{bc} . In fig. 4.7 one can see that this is not the case and NB_{out} and M_{bc} are uncorrelated.

4.9.1 Best Cut Estimation

The best cut on the neural network for each decay channel is selected by a Figure Of Merit (FOM), defined as

$$FOM = \frac{n_{sig}}{\sqrt{n_{sig} + n_{bkg}}}, \quad (4.11)$$

where n_{sig} and n_{bkg} are the estimated numbers of signal and background events in the signal region $M_{bc} > 5.27 \text{ GeV}/c^2$. The result is displayed in fig. 4.8. To estimate

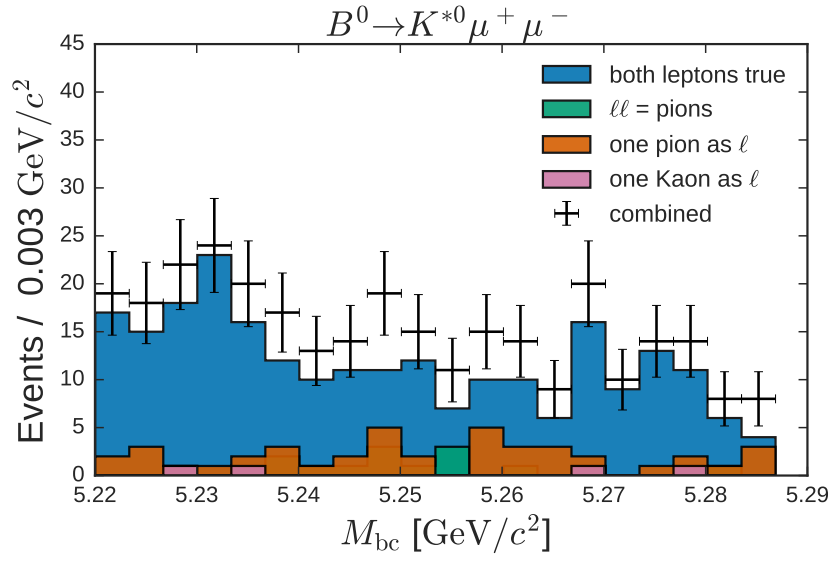


Figure 4.5: Background for $B^0 \rightarrow K^{*}(892)^0 \mu^+ \mu^-$ on generic MC in M_{bc} with different truth for the selected leptons.

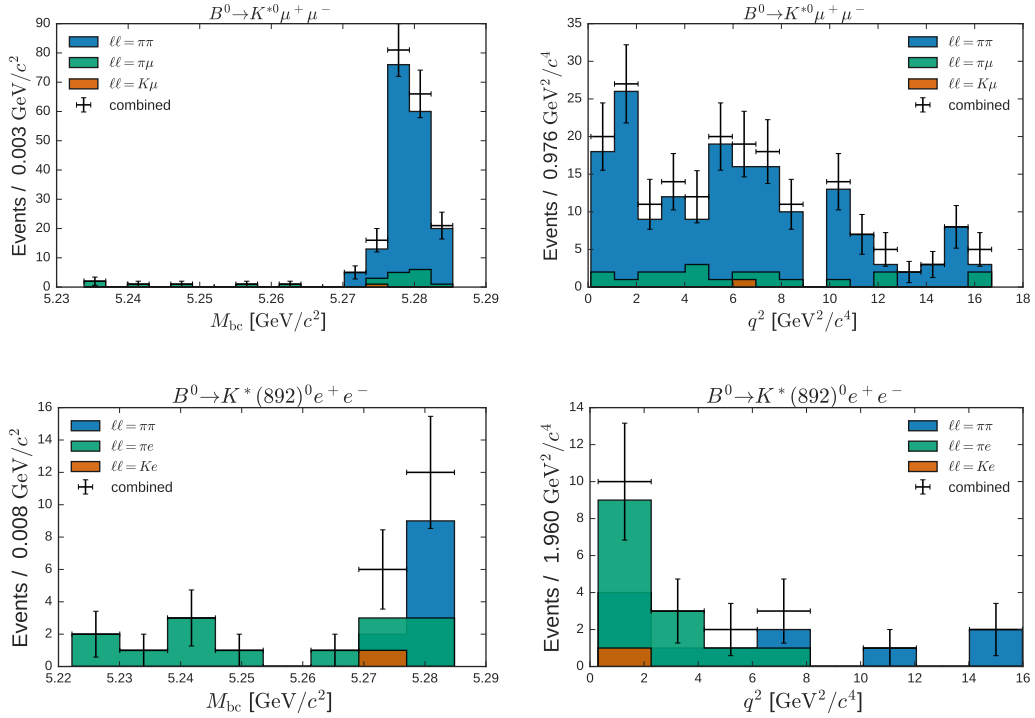


Figure 4.6: Peaking background on the rare MC dataset in M_{bc} and $q^2 \equiv M_{\ell^+\ell^-}^2$.

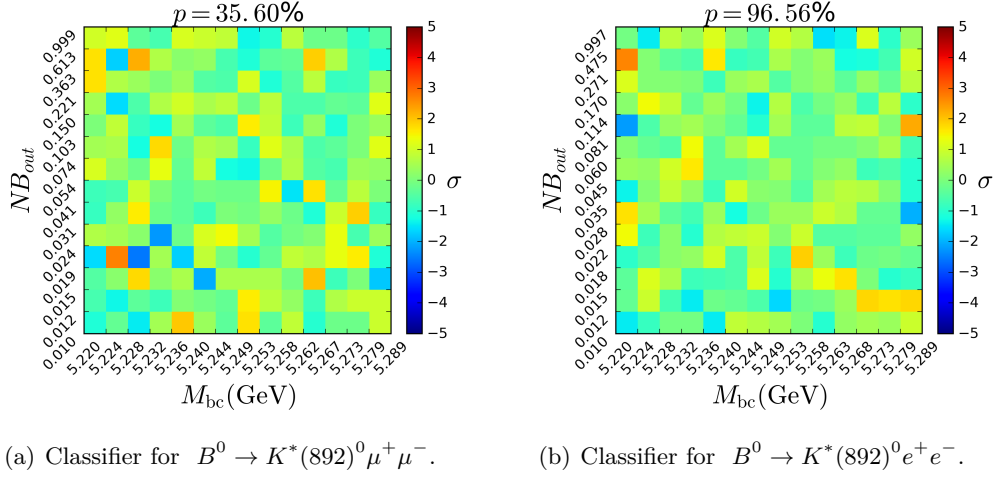


Figure 4.7: Flat correlation plots (see section 3.2.2 for further information about this kind of plots) for the neural network output vs. M_{bc} for two example channels. There is no significant correlation observable.

the number of signal candidates n_{sig} the branching fractions from Ref. [4] are used and the estimated numbers of signal events on the Belle dataset are calculated with respect to the reconstruction efficiency. The number of background events n_{bkg} are obtained after the final signal selection on generic MC, which compares to the Belle data-set in size and composition.

The resulting total reconstruction efficiency for each decay channel is presented in table 4.7.

Table 4.7: Neural network cut for the highest figure of merit (FOM). The corresponding FOM, efficiency and expected signal and background events in the signal region $M_{bc} > 5.27 \text{ GeV}/c^2$ are displayed. The efficiency includes all reconstruction and acceptance effects.

Channel	NB_{out} Cut	FOM	Efficiency [%]	$n_{\text{sig}}^{\text{exp}}$	$n_{\text{bkg}}^{\text{exp}}$
$B^+ \rightarrow K^+ e^+ e^-$	0.92	9.00	26.34	111	78
$B^+ \rightarrow K^*(892)^+ e^+ e^-$	0.51	2.05	2.94	35	176
$B^+ \rightarrow K^*(892)^+ \mu^+ \mu^-$	0.67	2.67	3.54	26	122
$B^+ \rightarrow K^+ \mu^+ \mu^-$	0.93	9.33	33.00	114	66
$B^0 \rightarrow K^0 e^+ e^-$	0.93	3.43	15.37	18	22
$B^0 \rightarrow K^*(892)^0 e^+ e^-$	0.91	4.90	5.37	42	47
$B^0 \rightarrow K^0 \mu^+ \mu^-$	0.97	6.11	17.60	46	21
$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	0.83	7.05	14.28	116	199

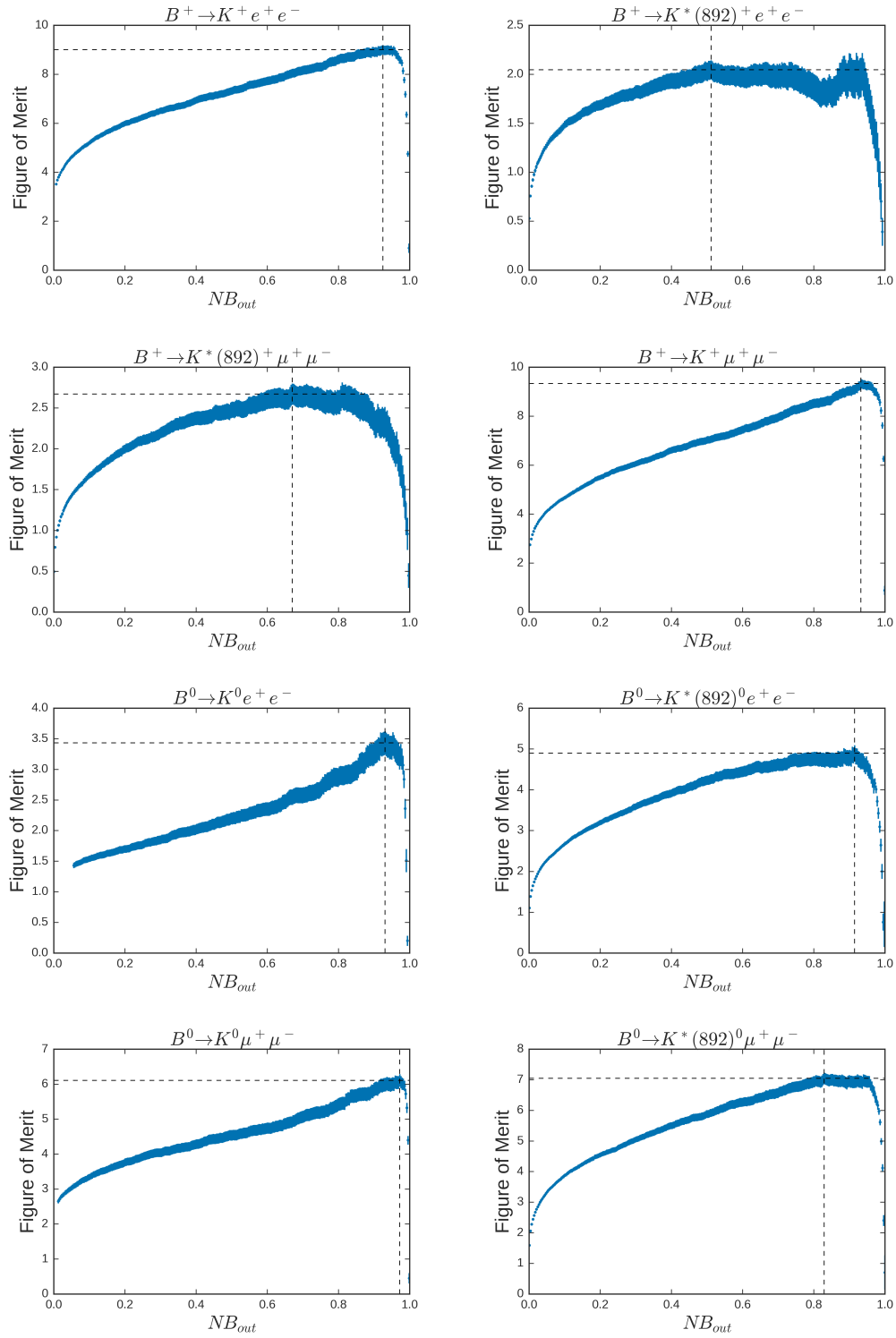


Figure 4.8: Estimation of the best cut on the basis of the Figure Of Merit (FOM) for each decay channel.

4.9.2 Efficiency

The reconstruction efficiency is first evaluated on MC and tested on data in the control region of $B \rightarrow K^{(*)}J/\psi$. The efficiency is defined as $\epsilon \equiv N_{\text{rec}}/N_{\text{gen}}$, where N_{rec} is the number of reconstructed events and N_{gen} is the number of generated events. All generated candidates are accounted for in N_{rec} , regardless if they are cut by event selection vetos or if tracks are missing in the detector. In this manner the calculation of the efficiency accounts for reconstruction and acceptance effects. The efficiency and signal purity of each decay channel is dependent on the cut on the neural network of the decay channel. Different working points of each neural network can be selected for a desired signal efficiency/purity.

The final neural network output for signal and background events is displayed in fig. 4.9 with the corresponding cut for the best FOM. For the angular analysis in chapter 6 cut values for $B^0 \rightarrow K^{*}(892)^0 e^+ e^-$ and $B^0 \rightarrow K^{*}(892)^0 \mu^+ \mu^-$ are optimized for the sensitivity to angular observables, described in section 6.3.3, and cut values do not correspond to the best FOM here. The efficiency in bins of q^2 is of interest for this analysis, which is shown in table 4.8. The efficiencies and errors of the efficiency in bins of q^2 are determined with

$$\epsilon = \frac{N_{\text{rec}}^{\text{bin}}}{N_{\text{gen}}^{\text{bin}}} \quad \text{and} \quad \sigma_{\epsilon} = \sqrt{\frac{N_{\text{rec}}^{\text{bin}} (N_{\text{gen}}^{\text{bin}} - N_{\text{rec}}^{\text{bin}})}{N_{\text{gen}}^{\text{bin}^3}}} \quad (4.12)$$

where $N_{\text{rec}}^{\text{bin}}$ is the number of reconstructed events and $N_{\text{gen}}^{\text{bin}}$ the number of generated events in the corresponding bin.

The estimation of the efficiency on data is discussed in the next chapter.

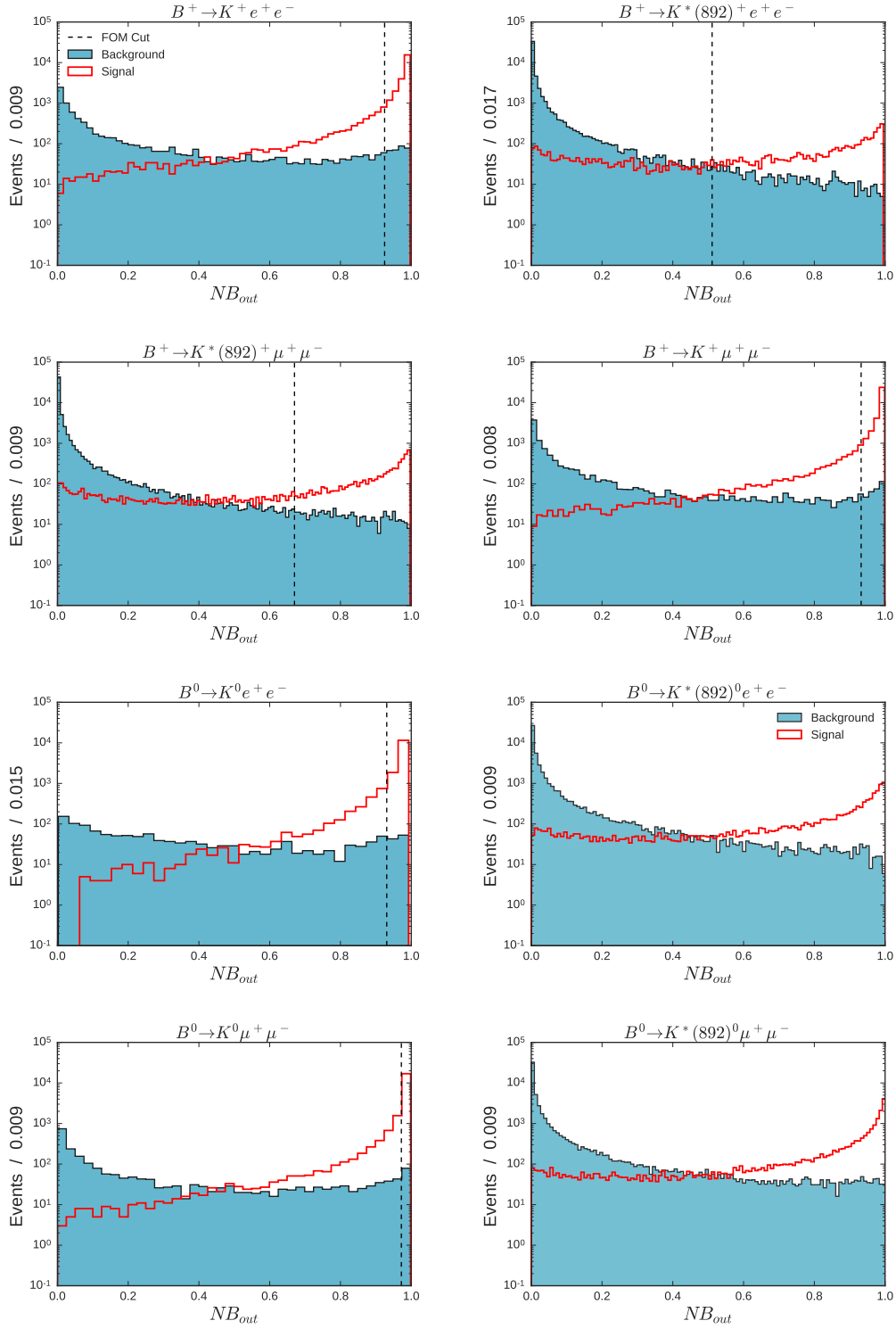


Figure 4.9: Performance of the neural networks for the classification of B mesons. The data for signal corresponds to correctly reconstructed events in the signal MC and background corresponds to generic MC for two times the expected size of the Belle dataset.

Table 4.8: Efficiency in bins of q^2 in %. The binning is described in the angular analysis in chapter 6. The efficiency is calculated based on the number of generated candidates within the q^2 range of the particular bin.

	0	1	2	3	4	combined
$B^+ \rightarrow (K^+ \pi^0) e^+ e^-$	5.529	5.271	5.789	3.371	2.413	4.582 ± 0.133
$B^+ \rightarrow (K^0 \pi^+) e^+ e^-$	4.880	4.440	4.739	2.550	2.452	3.921 ± 0.087
$B^+ \rightarrow (K^+ \pi^0) \mu^+ \mu^-$	4.733	3.881	5.728	7.042	3.596	4.884 ± 0.118
$B^+ \rightarrow (K^0 \pi^+) \mu^+ \mu^-$	4.260	3.550	4.885	5.135	3.482	4.203 ± 0.078
$B^+ \rightarrow K^*(892)^+ e^+ e^-$	5.091	4.712	5.080	2.823	2.437	4.137 ± 0.073
$B^+ \rightarrow K^*(892)^+ \mu^+ \mu^-$	4.414	3.658	5.160	5.777	3.515	4.426 ± 0.065
$B^0 \rightarrow (K^+ \pi^-) e^+ e^-$	18.256	16.745	18.232	11.371	12.902	15.780 ± 0.164
$B^0 \rightarrow (K^0 \pi^0) e^+ e^-$	1.844	1.931	2.088	0.971	0.839	1.587 ± 0.079
$B^0 \rightarrow (K^+ \pi^-) \mu^+ \mu^-$	21.621	17.881	25.676	27.819	25.752	23.442 ± 0.164
$B^0 \rightarrow (K^0 \pi^0) \mu^+ \mu^-$	2.800	2.071	3.322	3.196	2.123	2.663 ± 0.089
$B^0 \rightarrow K^*(892)^0 e^+ e^-$	12.697	11.704	12.820	7.875	8.910	11.003 ± 0.115
$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	15.289	12.635	18.133	19.658	17.938	16.511 ± 0.118

5. Signal Yields and Data Evaluation

The evaluation of the data is performed in several steps. First, signal yields for the electron and muon modes of $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$ are extracted on the whole range of q^2 . Additionally, the data is split into bins of q^2 and signal and background fractions are determined for each bin by a fit to M_{bc} . These results are used in MC simulations in chapter 6 for performing toy studies with correct statistics. The data is furthermore used to perform several cross-checks aimed to verify the signal reconstruction efficiency. For this, branching ratios are calculated for both decay channels and the well known J/ψ and $\psi(2S)$ resonance branching fractions are determined within the veto regions in q^2 .

5.1 Signal Yields

The yields for signal and background classes are extracted by an unbinned extended maximum likelihood fit to the M_{bc} distribution of reconstructed events for $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$. Signal is parametrized by an empirically determined function introduced by the Crystal Ball Collaboration [45]. The so-called Crystal Ball function accounts for radiative tails in the distribution and for the calorimeter resolution. It is defined as

$$\mathcal{P}^{CB}(M_{bc}, m_0, \sigma, \alpha, n) = \begin{cases} e^{-\frac{(M_{bc}-m_0)^2}{2\sigma^2}} & \text{if } M_{bc} > m_0 - \alpha\sigma \\ \left(\frac{n}{\alpha}\right)^n e^{-\frac{\alpha^2}{2}} \left(\frac{m_0-M_{bc}}{\sigma} + \frac{n}{\alpha} - \alpha\right)^{-n} & \text{if } M_{bc} \leq m_0 - \alpha\sigma, \end{cases} \quad (5.1)$$

where m_0 and σ are the mean and width of the distribution. The Crystal Ball function can be interpreted as a Gaussian shape which transits to a power-law tail, below a certain threshold, which is defined by the parameters α and n . All shape parameters are determined by a fit to data in the control channel $B \rightarrow K^* J/\psi$ in the corresponding q^2 veto region and fixed in the extraction of $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$. The background distribution is parametrized by an empirically determined shape

introduced by the ARGUS Collaboration [46]. This so-called ARGUS shape is defined as

$$\mathcal{P}^{ARGUS}(M_{bc}, m_0, \alpha) = M_{bc} \sqrt{1 - \left(\frac{M_{bc}}{m_0}\right)^2} e^{-\alpha(1 - (M_{bc}/m_0)^2)}, \quad (5.2)$$

where α describes the slope and m_0 the cutoff value of the distribution. Due to kinematic constraints, m_0 corresponds to half the center-of-mass energy ($m_0 = E_{\text{Beam}}/2$) in the reconstruction of B mesons from $\Upsilon(4S)$ decays at Belle.

The neural network cuts are optimized for the angular analysis, which is described in section 6.3.3 and do not correspond to the best FOM. On the total range of q^2 there are 118 ± 12 signal candidates extracted for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ and 69 ± 12 for $B^0 \rightarrow K^*(892)^0 e^+ e^-$. The result of the fits is depicted in fig. 5.1. The signal and background yields in bins of q^2 are determined on the combined sample of the electron and muon channel. The fit results in bins of q^2 are displayed in fig. 5.2, with the corresponding yields listed in table 5.1. Here, additionally the results for a sample containing only the muon mode is presented. These values are used to simulate accurate yields in the Monte Carlo toy studies detailed in section 6.3.

Table 5.1: Fitted yields for signal and background events in the binning of q^2 for both the combined and muon channels.

q^2 range [GeV ² /c ⁴]	$\ell = \mu, e$		$\ell = \mu$	
	n_{sig}	n_{bkg}	n_{sig}	n_{bkg}
1.00 – 6.00	49.5 ± 8.4	30.3 ± 5.5	23.8 ± 5.6	11.7 ± 3.4
0.10 – 4.00	30.9 ± 7.4	26.4 ± 5.1	12.2 ± 4.1	7.5 ± 2.7
4.00 – 8.00	49.8 ± 9.3	35.6 ± 6.0	29.7 ± 6.6	14.9 ± 3.9
10.09 – 12.90	39.6 ± 8.0	19.3 ± 4.4	31.9 ± 6.3	10.2 ± 3.2
14.18 – 19.00	56.5 ± 8.7	16.0 ± 4.0	36.8 ± 7.1	9.8 ± 3.1

5.2 Branching Ratios and Efficiency Evaluation

Several crosschecks are performed on data in order to validate the reconstruction efficiency determination and background suppression procedure. Neural networks in this analysis are trained on simulated data and are susceptible to differences between data and MC. The reconstruction efficiency is derived from MC and assumed to be the same for real data. With the following crosschecks for this assumption are performed. An evaluation of the efficiency for the angular analysis $B \rightarrow K^{(*)} \ell^+ \ell^-$ is performed by extracting the branching fractions of processes with the signature $B^0 \rightarrow (K^+ \pi^-) \ell^+ \ell^-$. The results are compared with other measurements and world averages from PDG. First, the branching ratio $\mathcal{B}(B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-)$ is extracted for electrons and muons. A second validation is performed in vetoed regions of the invariant mass of the di-lepton system q^2 . The sidebands are defined in regions with irreducible backgrounds from $B \rightarrow K^* J/\psi$ and $B \rightarrow K^* \psi(2S)$, where these decays

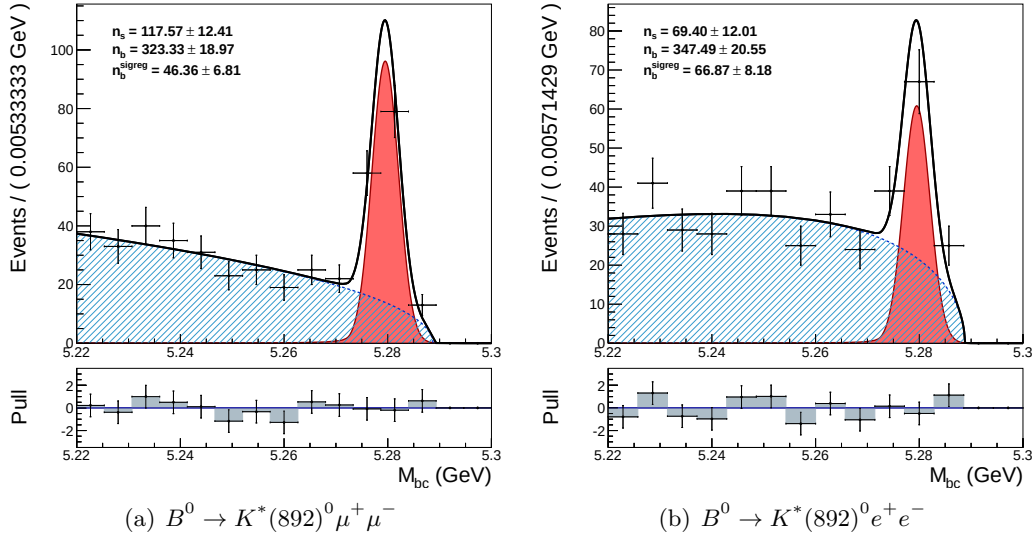


Figure 5.1: Signal extraction for $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$ on the total range of q^2 . Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

are dominant. The decays $B \rightarrow K^* J/\psi(\psi(2S))$ have been studied with high precision in the past, which makes them an ideal probe for crosschecks.

5.2.1 Extraction of $\mathcal{B}(B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-)$

The efficiencies for the electron and muon channels are extracted on a large set of simulated events with the same selection as on data. Veto regions for the electron channel are larger due to an inferior resolution in q^2 caused by bremsstrahlung of electrons. All cut regions are depicted in fig. 5.3 and listed in table 4.4. Cuts on q^2 preserve 58.75% (77.23%) of signal candidates for electrons (muons) on generator level. The total reconstruction efficiency for $B^0 \rightarrow K^*(892)^0 e^+ e^-$ is $7.52 \pm 0.03\%$ and $12.16 \pm 0.04\%$ for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$, with neural network cuts optimized for the angular analysis. These efficiencies are calculated on the whole range of q^2 and include reconstruction and acceptance effects.

The following results are extracted from data:

$$\mathcal{B}(B \rightarrow K^*(892)^0 e^+ e^-) = (1.19 \pm 0.21) \times 10^{-6}, \quad (5.3)$$

and

$$\mathcal{B}(B \rightarrow K^*(892)^0 \mu^+ \mu^-) = (1.25 \pm 0.14) \times 10^{-6}, \quad (5.4)$$

and for the ratio of the muon to electron channel

$$\frac{\mathcal{B}(B \rightarrow K^*(892)^0 \mu^+ \mu^-)}{\mathcal{B}(B \rightarrow K^*(892)^0 e^+ e^-)} = 1.055 \pm 0.216 \quad (5.5)$$

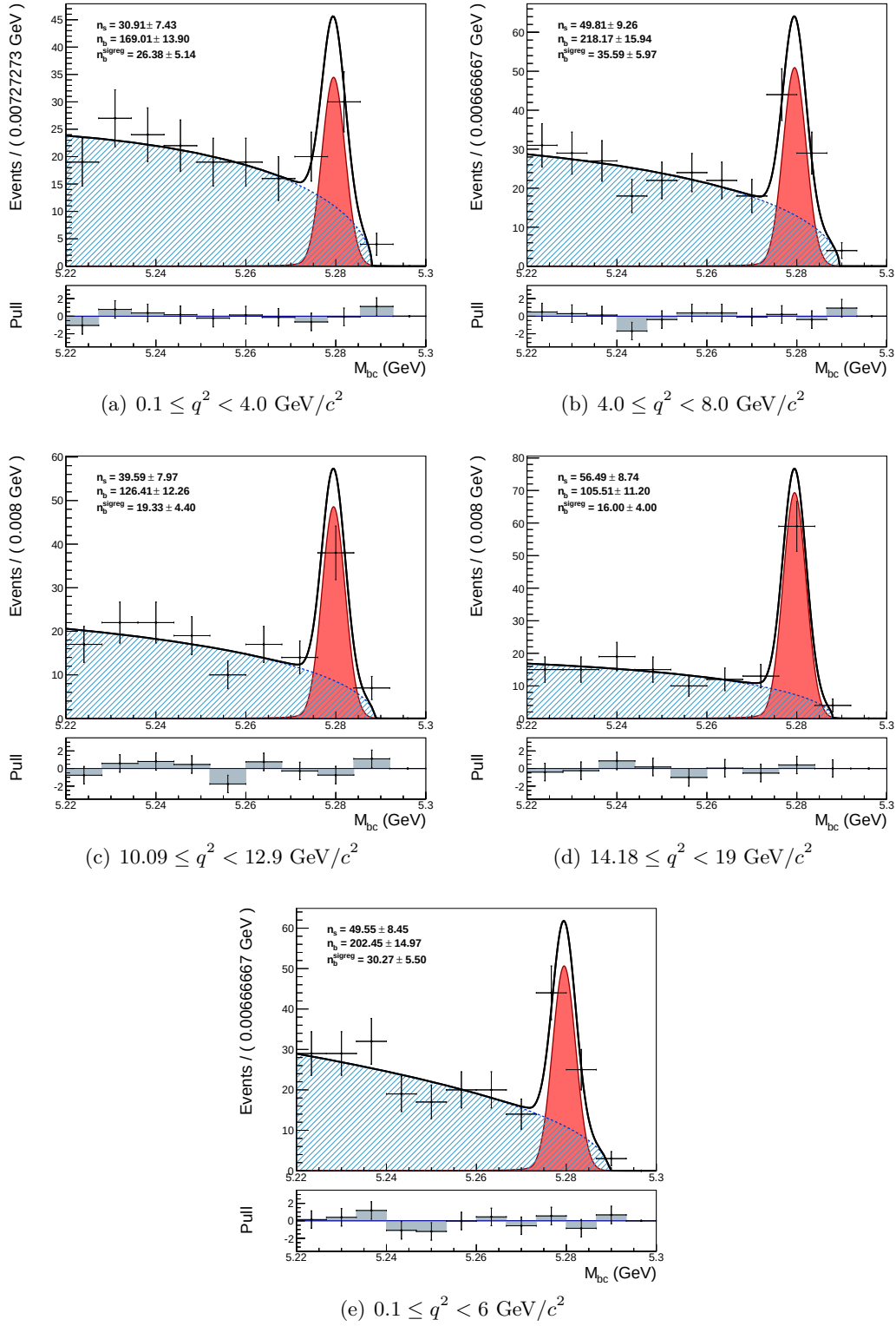


Figure 5.2: Signal extraction for $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$, $\ell = e, \mu$ in bins of q^2 . Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

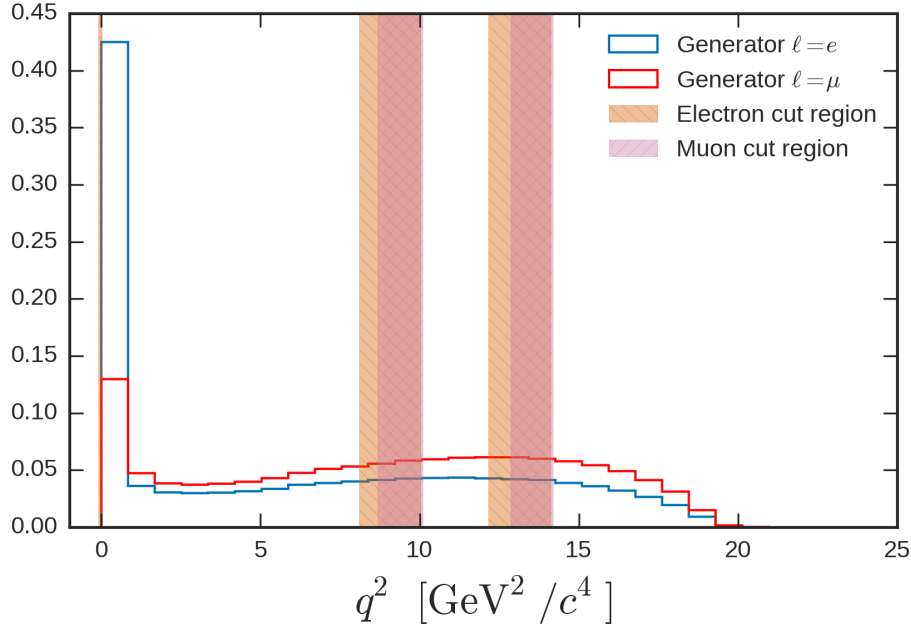


Figure 5.3: Comparison of generator distribution for $B^0 \rightarrow K^*(892)^0 e^+ e^-$ and $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$. The corresponding veto regions for the charmonium resonances are superimposed.

where quoted errors on these measurements are statistical only. The PDG [4] values for the measured quantities are

$$\mathcal{B}(B \rightarrow K^*(892)^0 e^+ e^-)^{PDG} = (1.03^{+0.19}_{-0.17}) \times 10^{-6}, \quad (5.6)$$

and

$$\mathcal{B}(B \rightarrow K^*(892)^0 \mu^+ \mu^-)^{PDG} = (1.05 \pm 0.10) \times 10^{-6}. \quad (5.7)$$

The extracted branching ratios are in agreement with averages from PDG within errors. Both measurements are slightly above the world average from PDG, which might be due to small pollution from peaking background that is not corrected for in this study. For a proper branching ratio measurement this would need to be included or be put into the systematic uncertainties. This measurement however aims to verify the reconstruction efficiency determination, which is satisfied by the result for both channels.

5.2.2 Extraction of $\mathcal{B}(B \rightarrow K^{*0} J/\psi)$ and $\mathcal{B}(B^0 \rightarrow K^{*0} \psi(2S))$

A second validation of the efficiency estimation of the classifiers in the analysis is performed by measuring the well known branching fractions of $B \rightarrow K^* J/\psi$ and $B \rightarrow K^* \psi(2S)$. In a first step the efficiency for these processes is derived from signal

Monte Carlo. On data the signal yield for the individual process is extracted and the corresponding branching ratio is calculated. Two sidebands within the veto regions in q^2 are defined, where the individual decay of the charmonium resonance is dominant and decays from $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$ can be neglected. The first region is

$$9.2 < q^2 < 9.8 \text{ GeV}^2/c^4, \quad (5.8)$$

with the dominant decay of $B \rightarrow K^* J/\psi$. The second is

$$13.0 < q^2 < 14.0 \text{ GeV}^2/c^4, \quad (5.9)$$

with the dominant decay of $B \rightarrow K^* \psi(2S)$. The efficiency is evaluated in the sidebands by calculating the ratio of the number of truly reconstructed signal candidates on signal MC with the corresponding neural network selection and the number of generated signal events in the particular region. The resulting efficiencies are listed in table 5.2.

Table 5.2: Reconstruction efficiency in the q^2 sidebands.

Channel / Sideband	J/ψ [%]	$\psi(2S)$ [%]
$B^0 \rightarrow K^*(892)^0 e^+ e^-$	11.79 ± 0.21	11.26 ± 0.17
$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	16.17 ± 0.22	16.25 ± 0.18

The signal component in M_{bc} is fitted with a Crystal Ball function and the background with an ARGUS shape. The yields are extracted with an unbinned extended maximum likelihood fit. All fit results are shown in fig. 5.4. The branching fraction for the processes are calculated by:

$$\mathcal{B}(B \rightarrow X_{c\bar{c}} K^*) = \frac{N_{obs}}{\epsilon_s \cdot f_{\ell\ell} \cdot N_{B\bar{B}}}, \quad (5.10)$$

where N_{obs} is the extracted signal yield, ϵ_s is the signal efficiency and $N_{B\bar{B}}$ is the number of recorded B meson pairs at Belle. An additional factor $f_{\ell\ell}$ from PDG [4] is taken into account describing the fraction of the $c\bar{c}$ state decaying into lepton pairs, detailed in table 5.3.

Table 5.3: Fraction $f_{\ell\ell}$ of $X_{c\bar{c}}$ into leptons from PDG.

$\mathcal{B}(X_{c\bar{c}} \rightarrow \ell^+ \ell^-)$	$X_{c\bar{c}} = J/\psi$ [%]	$X_{c\bar{c}} = \psi(2S)$ [%]
$\ell = e$	5.971 ± 0.032	0.789 ± 0.017
$\ell = \mu$	5.961 ± 0.033	0.79 ± 0.09

The resulting branching fractions are listed in table 5.4. The error on $\psi(2S) \rightarrow \mu\mu$ is by a factor of 5 larger than the corresponding electron mode. Hence, the error on the branching fraction for $B^0 \rightarrow K^* \psi(2S) (\rightarrow \mu\mu)$ is larger compared to the electron channel, despite higher statistics in this channel. The PDG values for the decays are

$$\mathcal{B}(B^0 \rightarrow K^* J/\psi) = (1.32 \pm 0.06) \times 10^{-3} \quad (5.11)$$

and

$$\mathcal{B}(B^0 \rightarrow K^* \psi(2S)) = (5.9 \pm 0.4) \times 10^{-4}. \quad (5.12)$$

All results are in agreement within statistical errors. Systematic errors are not taken into account.

Table 5.4: Extracted Branching Fraction. $X_{c\bar{c}}$ is either J/ψ or $\psi(2S)$.

$\mathcal{B}(B^0 \rightarrow K^* X_{c\bar{c}})$	$X_{c\bar{c}} = J/\psi [\times 10^{-3}]$	$X_{c\bar{c}} = \psi(2S) [\times 10^{-4}]$
e^+e^- Mode	1.36 ± 0.04	5.5 ± 0.33
$\mu^+\mu^-$ Mode	1.24 ± 0.03	6.0 ± 0.7

5.3 Data and Monte Carlo Comparison

The comparison between data and Monte Carlo is an important cross-check to verify whether simulated distributions agree with measured ones. Since all methods of the analysis are trained and tested on simulated data the importance of this step must not be underestimated. First a comparison of the neural network input variables is performed. Secondly sidebands are used in the charmonium veto regions of q^2 , defined in section 5.2.2, in order to verify the angular dependence of the reconstruction efficiency.

5.3.1 Variables of the Neural Network

The input variables after the neural network selection are compared between data and Monte Carlo in figs. A.1 and A.2. In the MC sample, the number of background and signal events are composed according to fitted yields from data with values from table 5.1. A p -value of a Kolmogorov-Smirnov statistic on 2 samples is stated for each variable. The most important quantity is the neural network output, which is presented in fig. 5.5. It accumulates all information of the input. The observed distributions agree very well, hence one can apply the trained expertise to data. This is consistent with the observed agreement of extracted branching fractions to the world averages, described in sections 5.2.1 and 5.2.2.

5.3.2 Angular Reconstruction Efficiency

The angular dependence of the reconstruction efficiency is evaluated in the q^2 veto regions for $B \rightarrow K^* J/\psi$ and $B \rightarrow K^* \psi(2S)$ between data and MC. The result is displayed in figs. 5.6 and 5.7. A systematic discrepancy in the ratio data over MC in form of a linear trend can be observed in $\cos \theta_K$. The consequences and implications of this will be analyzed in the context of the systematic error estimation in section 7.1.

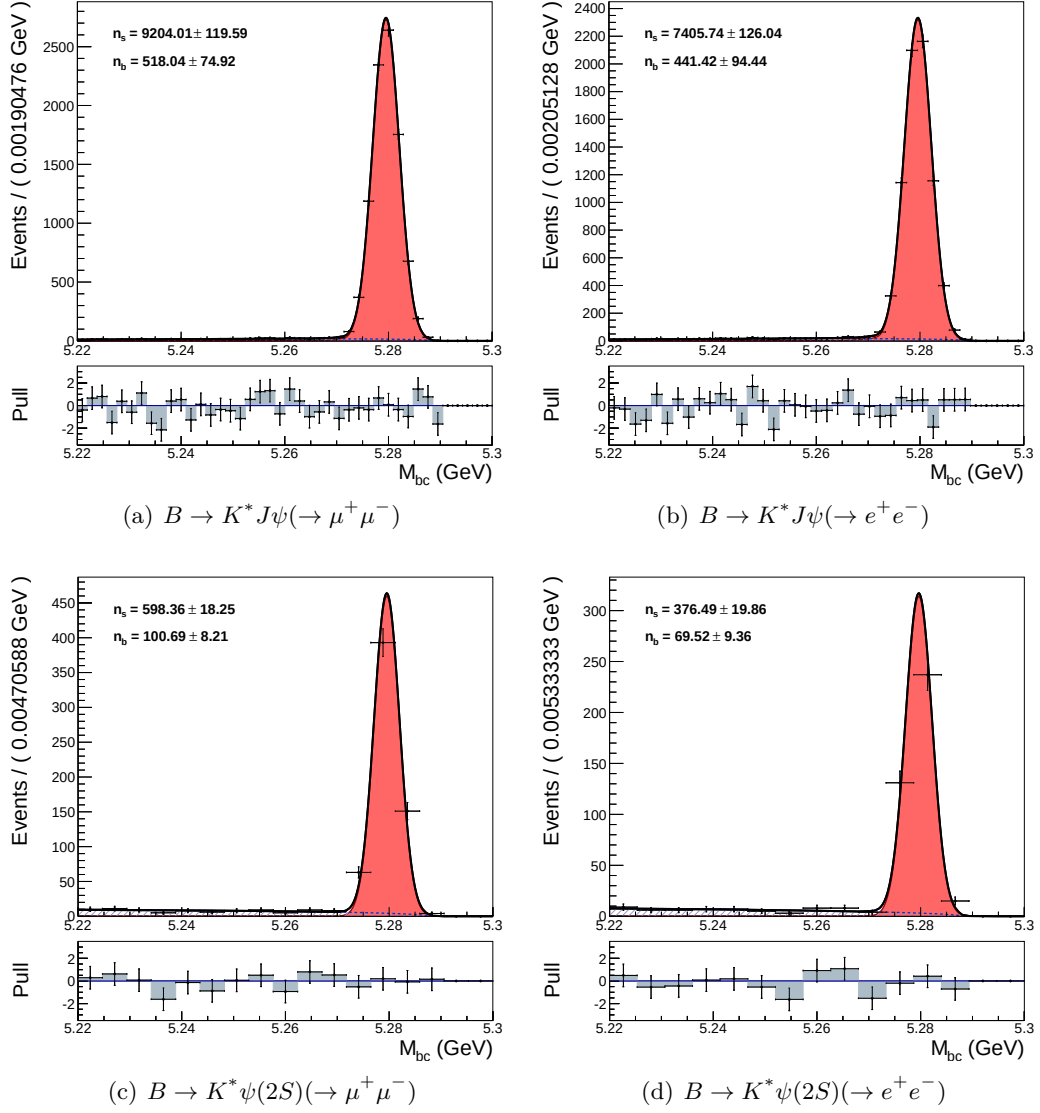


Figure 5.4: Extraction of the signal yields for $B \rightarrow J/\psi \ell \ell$ and $B \rightarrow \psi(2S) \ell \ell$ in the corresponding q^2 sidebands. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

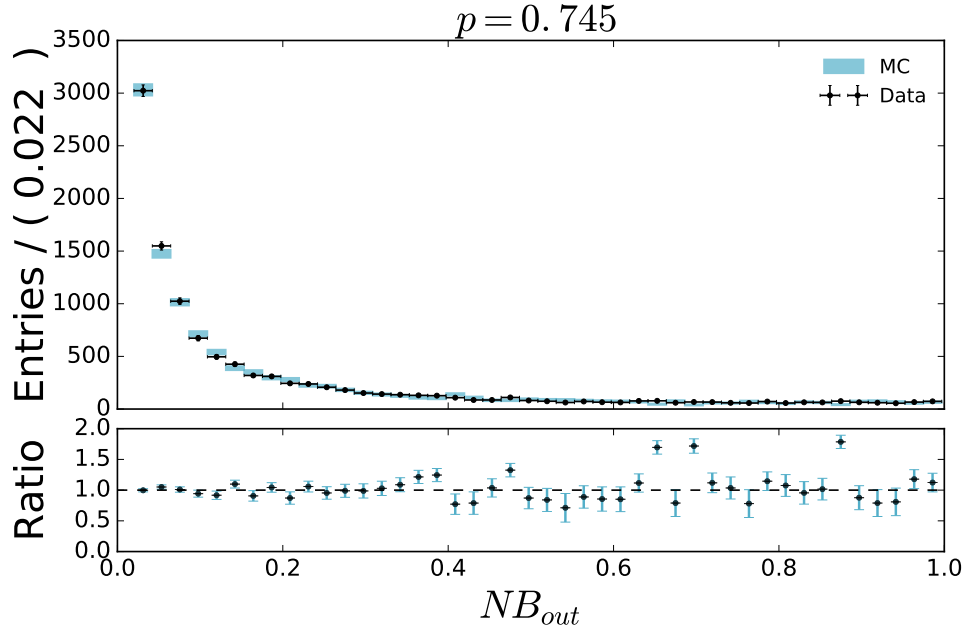


Figure 5.5: Neural network output comparison between data and Monte Carlo from the reconstruction of $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ in the M_{bc} sideband $M_{bc} < 5.27 \text{ GeV}/c^2$. The p-value of a Kolmogorov-Smirnov statistic on 2 samples is stated.

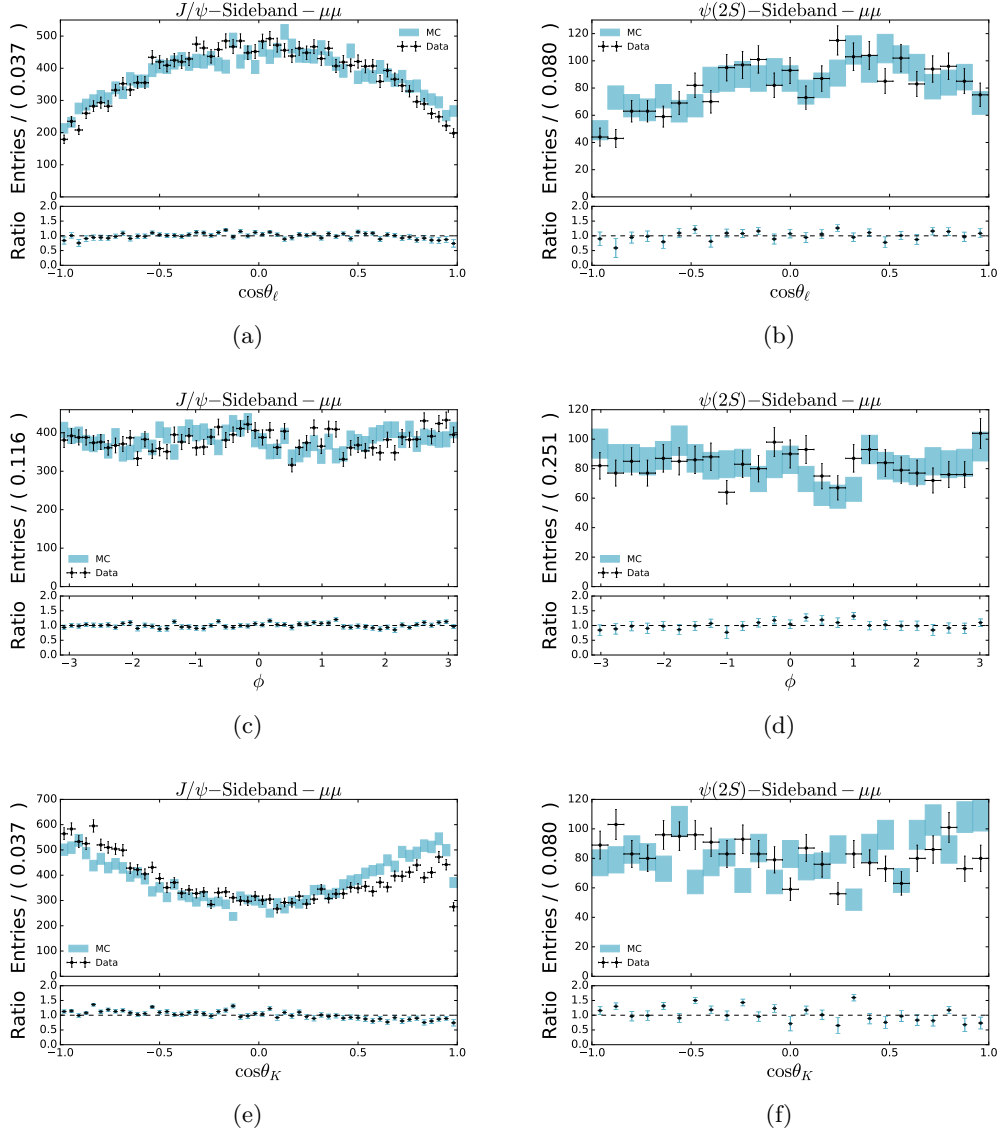


Figure 5.6: Data vs. MC comparison of the reconstruction of $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ in the $9.4 \leq q^2 < 9.8 \text{ GeV}^2/c^4$ and $13 \leq q^2 < 14 \text{ GeV}^2/c^4$ veto regions.

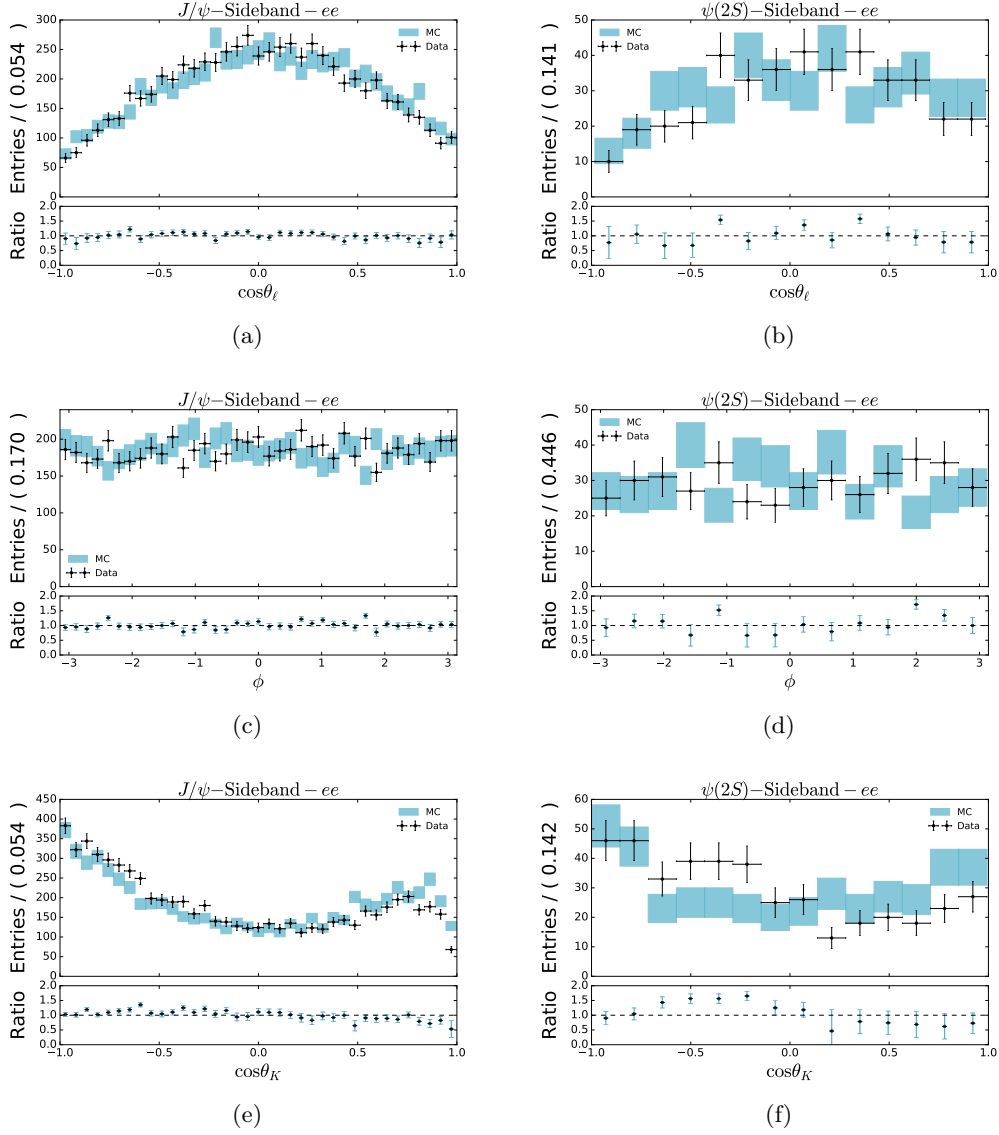


Figure 5.7: Data vs. MC comparison of the reconstruction of $B^0 \rightarrow K^*(892)^0 e^+ e^-$ in the $9.4 \leq q^2 < 9.8 \text{ GeV}^2/c^4$ and $13 \leq q^2 < 14 \text{ GeV}^2/c^4$ veto regions.

6. Angular Analysis

The angular analysis is performed for $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$ including the electron and muon mode. As detailed in the introduction, section 1.2.1, the decay is kinematically described by three angles θ_ℓ, θ_K and ϕ and the invariant mass squared of the lepton pair q^2 . The binning in q^2 is depicted in fig. 6.1 and detailed in table 5.1 alongside the measured signal and background yields. Uncovered regions in the q^2 spectrum arise from vetoes against backgrounds of the charmonium resonances $J/\psi \rightarrow \ell^+ \ell^-$ and $\psi(2S) \rightarrow \ell^+ \ell^-$. A zeroth bin is defined in $1.0 < q^2 < 6.0 \text{ GeV}^2/c^4$, which is considered to be the theoretically cleanest [18].

6.1 Fit Components

Two main components are considered in the fit: the signal distribution and combinatorial background. These components are combined to the final probability density function and are described in the following sections.

6.1.1 Signal

For this analysis, the folding technique described in section 1.2.4 is applied. In this manner, for each of the extractions of P'_4, P'_5, P'_6 and P'_8 a folded dataset is produced to which a dedicated differential decay rate is fitted. In total four different signal probability functions are used together with their individual dataset transformation, which are presented in the following.

Using the transformation from eq. (1.38) one can extract P'_4 from a fit to the differential

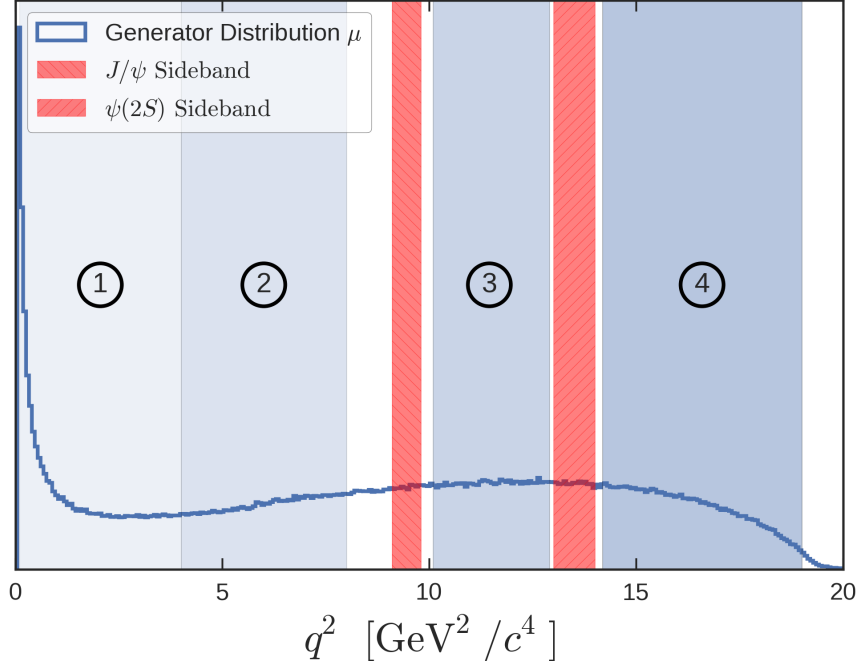


Figure 6.1: Binning in $q^2 \equiv M_{\ell^+\ell^-}^2$ with generated distribution for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$. The cut veto for the charmonium resonances is indicated in red.

decay rate

$$\begin{aligned} \frac{1}{d\Gamma/dq^2} \frac{d^4\Gamma}{d\cos\theta_\ell d\cos\theta_K d\phi dq^2} = & \frac{9}{8\pi} \left[\frac{3}{4}(1 - F_L) \sin^2\theta_K + F_L \cos^2\theta_K \right. \\ & + \frac{1}{4}(1 - F_L) \sin^2\theta_K \cos 2\theta_\ell \\ & - F_L \cos^2\theta_K \cos 2\theta_\ell + S_3 \sin^2\theta_K \sin^2\theta_\ell \cos 2\phi \\ & \left. + \sqrt{F_L(1 - F_L)} P'_4 \sin 2\theta_K \sin 2\theta_\ell \cos \phi \right]. \quad (6.1) \end{aligned}$$

As a proof-of-concept it is demonstrated that the data transformation procedure leads folded PDFs. For this, simulated data is generated from the total differential decay rate eq. (1.37). Afterwards this data is folded using definitions from eq. (1.38) and compared with the folded PDF of eq. (6.1). The result and data are displayed in fig. 6.2 and one can observe that the folding procedure together with the folded probability density function delivers coherent results, as the data is described perfectly.

Similarly, with the transformation from eq. (1.39) one can extract P'_5 with the

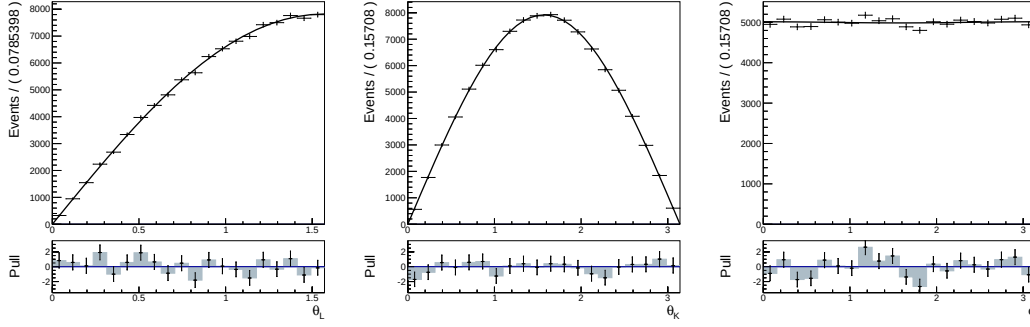


Figure 6.2: Simulated data according to eq. (1.37) is folded for the extraction of P'_4 using definitions of eq. (1.38). The line indicates the fit result for the transformed PDF eq. (6.1).

differential decay rate

$$\begin{aligned} \frac{1}{d\Gamma/dq^2 d\cos\theta_\ell d\cos\theta_K d\phi dq^2} \frac{d^4\Gamma}{dq^2} = & \frac{9}{8\pi} \left[\frac{3}{4}(1-F_L)\sin^2\theta_K + F_L\cos^2\theta_K \right. \\ & + \frac{1}{4}(1-F_L)\sin^2\theta_K\cos 2\theta_\ell \\ & - F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi \\ & \left. + \sqrt{F_L(1-F_L)}P'_5\sin 2\theta_K\sin\theta_\ell\cos\phi \right]. \quad (6.2) \end{aligned}$$

With the transformation from eq. (1.40) one can extract P'_6 with the differential decay rate

$$\begin{aligned} \frac{1}{d\Gamma/dq^2 d\cos\theta_\ell d\cos\theta_K d\phi dq^2} \frac{d^4\Gamma}{dq^2} = & \frac{9}{8\pi} \left[\frac{3}{4}(1-F_L)\sin^2\theta_K + F_L\cos^2\theta_K \right. \\ & + \frac{1}{4}(1-F_L)\sin^2\theta_K\cos 2\theta_\ell \\ & - F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi \\ & \left. + \sqrt{F_L(1-F_L)}P'_6\sin 2\theta_K\sin\theta_\ell\sin\phi \right]. \quad (6.3) \end{aligned}$$

Correspondingly, P'_8 can be extracted using eq. (1.41), which leads to the differential decay rate

$$\begin{aligned} \frac{1}{d\Gamma/dq^2 d\cos\theta_\ell d\cos\theta_K d\phi dq^2} \frac{d^4\Gamma}{dq^2} = & \frac{9}{8\pi} \left[\frac{3}{4}(1-F_L)\sin^2\theta_K + F_L\cos^2\theta_K \right. \\ & + \frac{1}{4}(1-F_L)\sin^2\theta_K\cos 2\theta_\ell \\ & - F_L\cos^2\theta_K\cos 2\theta_\ell + S_3\sin^2\theta_K\sin^2\theta_\ell\cos 2\phi \\ & \left. + \sqrt{F_L(1-F_L)}P'_8\sin 2\theta_K\sin 2\theta_\ell\sin\phi \right]. \quad (6.4) \end{aligned}$$

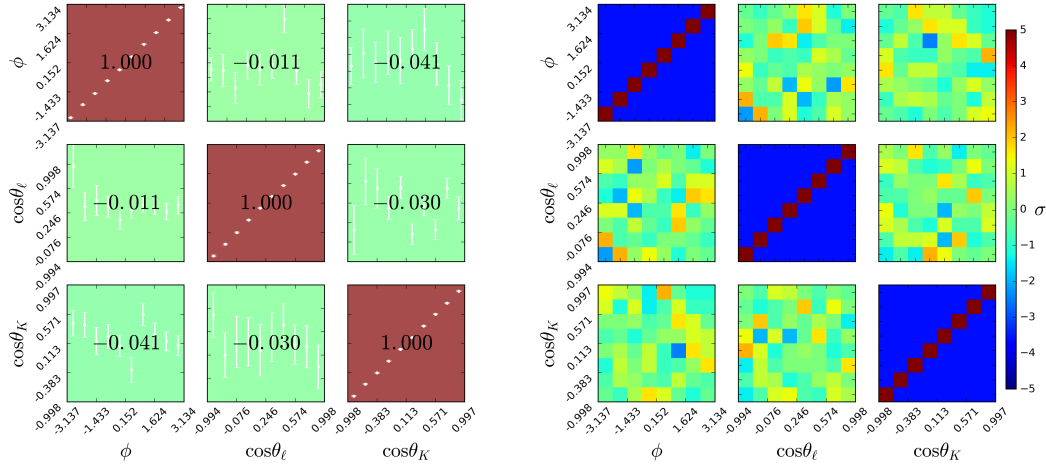


Figure 6.3: Correlation among the angular variables in the background sample displayed as correlation matrix (left) and flat correlation matrix (right), see details for flat correlation plots in section 3.2.2. In the correlation matrix a color is displayed corresponding to the amount of correlation, where green signals a correlation coefficient of around zero.

6.1.2 Background

For parameterizing the background, several methods are evaluated. As a first method the direct product of three second order polynomials, one for each dimension, is used

$$f_{\text{bkg}}^{\text{pol}}(q^2, \cos \theta_\ell, \cos \theta_K, \phi) = \left(\sum_{i=0,1,2} a_i \cos^i \theta_\ell \cdot \sum_{j=0,1,2} b_j \cos^j \theta_K \cdot \sum_{k=0,1,2} c_k \phi^k \right), \quad (6.5)$$

where a_i, b_j and c_k are q^2 dependent. This introduces in total nine additional parameters to the fit. Moreover, the polynomials have to be able to describe the shape of the background. The procedure of a direct product of each angular variable assumes that they are uncorrelated and independent. The correlation among the variables in the background sample can be seen in fig. 6.3. Additionally, non linear correlations are investigated by flat correlation plots (explained in section 3.2.2) in fig. 6.3(b). They are found to be insignificant. As no correlations are observed, the factorization approach in eq. (6.5) is assumed to be justified.

The second method which is evaluated are template histograms. For a full histogram in three dimensions an insufficient number of background events is available. Assuming again that the angular variables in the background are uncorrelated the three-dimensional PDF is build by multiplying the histograms of each projection of the angular variables:

$$f_{\text{bkg}}^{\text{hist}}(q^2, \cos \theta_\ell, \cos \theta_K, \phi) = \left(h_1(q^2, \cos \theta_\ell) \cdot h_2(q^2, \cos \theta_K) \cdot h_3(q^2, \phi) \right). \quad (6.6)$$

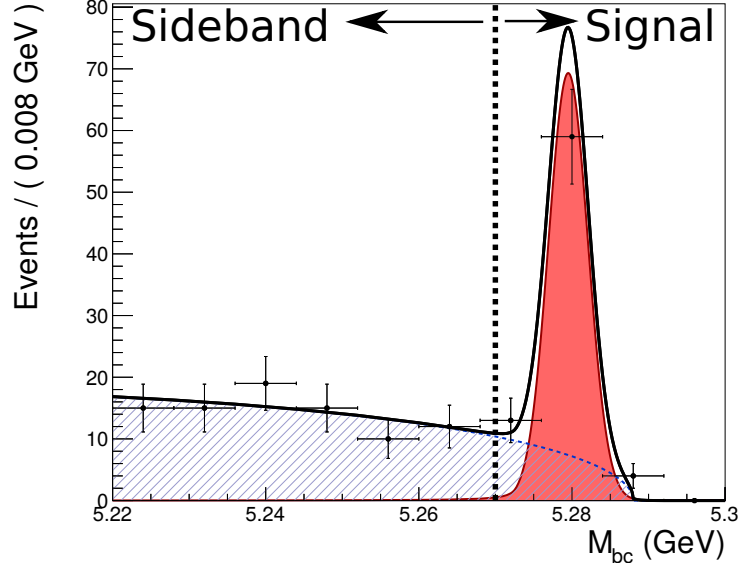


Figure 6.4: Definition of the signal and sideband regions in M_{bc} in an example fit for the data of q^2 bin 4.

This method is fast and robust even if the background shape is complicated. However, it introduces systematic deviations from the true distribution due to statistical fluctuations in the sideband. To compensate for this, the histograms are smoothed with an algorithm introduced in Ref. [35], which takes Poisson errors for bins with a small amount of entries into account. The method aims to optimize the pull distribution from the smoothed histogram to the original histogram with a least square minimization which accounts for statistical fluctuations in the bins. Both methods are tested in pseudo experiments with simulated data in section 6.3.

6.2 Fit Procedure

The angular variables cannot be used to determine the fraction of signal and background events without biasing the angular distributions. Thus, a separate fit in the beam constrained mass M_{bc} has to be performed. The fit in M_{bc} is done first to determine the signal to background ratio. The M_{bc} distribution is split into a signal (upper) and sideband (lower) region by introducing a cut at 5.27 GeV, as displayed in fig. 6.4. With the fit over the total range of M_{bc} , one can determine the number of signal events by a Crystal Ball function [45] and the number of background events with an ARGUS shape [46]. The corresponding number of background events in the signal region is extracted by integrating the background function over the signal region. This is performed in chapter 5 and the yields for both classes are shown in table 5.1.

The shape of the background for the angular observables can be estimated on the M_{bc} sideband. This is beneficial as no Monte Carlo information is needed. In this

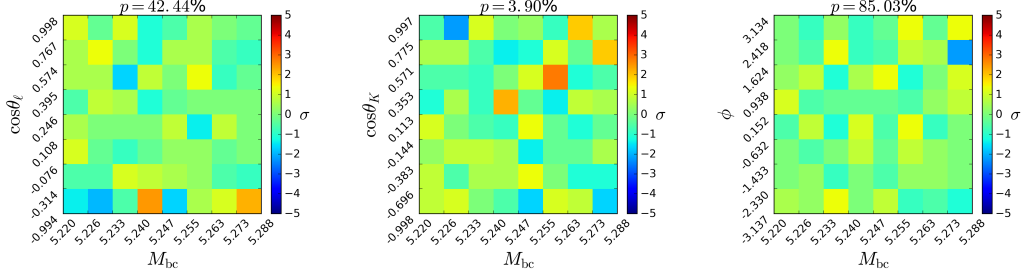


Figure 6.5: Flat correlation plots of the angular variables with M_{bc} in the background sample with the probability p for a flat hypothesis (see details for this kind of plot in section 3.2.2).

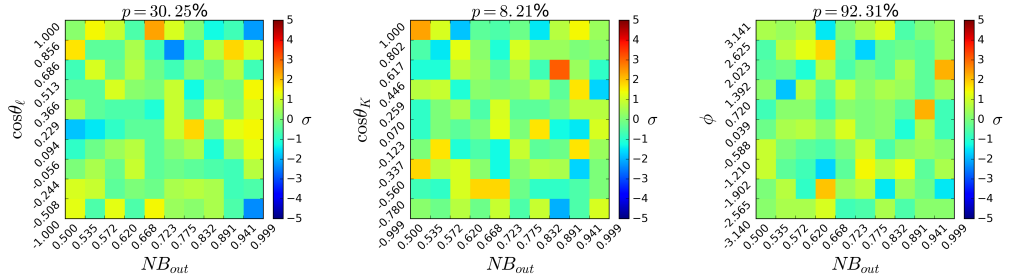


Figure 6.6: Flat correlation plots of the angular variables with NB_{out} in the background sample with the probability p for a flat hypothesis (see details for this kind of plot in section 3.2.2).

procedure, the underlying assumption is that the angular observables are uncorrelated with M_{bc} in the background sample. The correlation displayed in fig. 6.5 shows no severe correlations can be observed. The statistics in the sideband is higher and allows for a better estimation of the background shape and furthermore stabilizes the fit in the signal region.

To further improve the estimation of the shape of the background, the background sample can be enlarged by lowering the cut on the neural network for the sideband region. For this, the shape of the background must not change with cuts on the neural network. With a slightly decreased cut on the network one can more than double the statistics for the determination of the shape of the background. In fig. 6.6 one can see that the shapes of the background are insensitive to a small reduction of the cut on the neural network output NB_{out} .

The final fits are performed in the signal region of M_{bc} in three dimensional unbinned extended maximum likelihood fits in bins of q^2 , each featuring three free parameters: P' , $A_T^{(2)}$ and F_L . Signal and background yields are determined from M_{bc} and the shape of the background is derived from angular distributions on the sideband of M_{bc} .

and fixed in the final fit.

Within the statistics of Belle, it is not possible to adopt the same binning as used in Refs. [47, 48]. Instead of three, two bins in the low q^2 region below the J/ψ resonance in the range of $[0.1, 8] \text{ GeV}^2/c^4$ are chosen. The theoretically preferred region is in the range of $[1, 6] \text{ GeV}^2/c^4$ in which most of the assumptions for the Standard Model prediction are assumed to be valid [18]. Outside of this region uncertainties for the theoretical predictions are less well known or unclear due to hadronic interferences with the charmonium resonances. This region has overlaps with both bins one and bin two and will be additionally fitted as a 0th bin.

6.3 Monte Carlo Toy Study

In Monte Carlo toy studies various aspects of the fit procedure are tested and evaluated including the background parametrization, tests with simultaneous fitting, optimization of the cut on the neural networks and the evaluation of the expected sensitivity. In the toy study, Monte Carlo data is generated with the known truth of the observables in the fit. The fit is then performed 10000 times and the results are averaged across all fits. At the end, the deviation of the fit result from the simulated truth, divided by the fit error, is examined. This is called a pull distribution of a variable v :

$$\text{Pull}(v) = \frac{v_{fit} - v_{true}}{\text{error}(v_{fit})}, \quad (6.7)$$

where v_{fit} is the value derived from the fit and v_{true} is the simulated truth. If the fit procedure is not biased or corrupted and errors are calculated correctly, the pull should be a Gaussian distribution with mean zero and width of one. Any deviation from this expectation can be interpreted as a systematic uncertainty or may hint to flaws in the fit procedure. For each toy experiment, signal MC and background MC statistics are generated according to the measured yields from data, shown in table 5.1.

Small systematic deviations from the simulated values are observed due to low statistics and low signal to background ratios in some measurements. The following toy studies examine a variety of procedures all aiming to minimize both a fit bias and the statistical error, which lead to more precise results in the final measurement. The focus in the following studies is to optimize the results for P'_5 and P'_4 .

6.3.1 Background Parametrization

In a first evaluation the effect of different background parametrization is monitored. The shape of the background is determined on sidebands of M_{bc} and the shape is fitted with smoothed histograms and polynomials, described in section 6.1.2. The result is presented in fig. 6.7 for the variables P'_4 and P'_5 as an example. Both methods perform equally well, since the errors are compatible between both methods and no significant difference in a fit bias can be observed. The smoothed histograms exhibit slightly more stability in the fits for which reason they are preferred in this analysis.

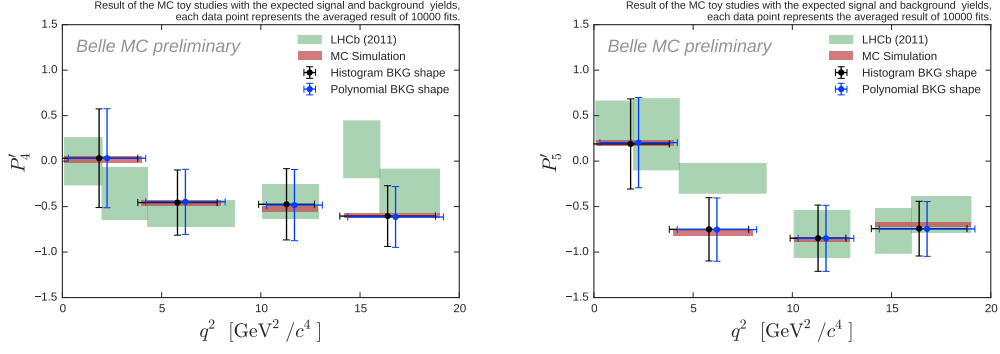


Figure 6.7: Result of the toy studies for two different background parametrization, where each data point represents the averaged result of 10000 fits to pseudo data.

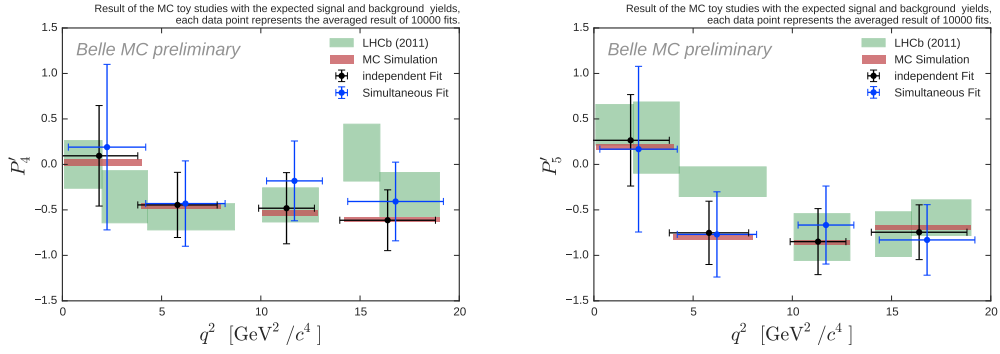


Figure 6.8: Result for simultaneous fitting. In each bin the measurements of P'_4 and P'_5 are fitted simultaneously and share the same values for F_L and A_T , where each data point represents the averaged result of 10000 fits to pseudo data.

6.3.2 Simultaneous Fitting

With the folding procedure, described in section 1.2.4, one can in principle perform a simultaneous fit with four different datasets, where F_L and A_T are constrained to the same value for the extractions of $P'_{4,5,6,8}$. However, in this case the fit stability is not sufficient due to the small amount of data as for all four extractions one needs to use independent angular observables because they have different transformations and boundaries. But it is possible to fit the variables (P'_4, P'_5) and (P'_6, P'_8) together in an independent fit since the transformations of the angles lead to the same boundaries in the angular space. The result is displayed in fig. 6.8. It is observed, however, that a substantial amount of fits fails, which leaves the final result biased and worse in terms of statistical sensitivity.

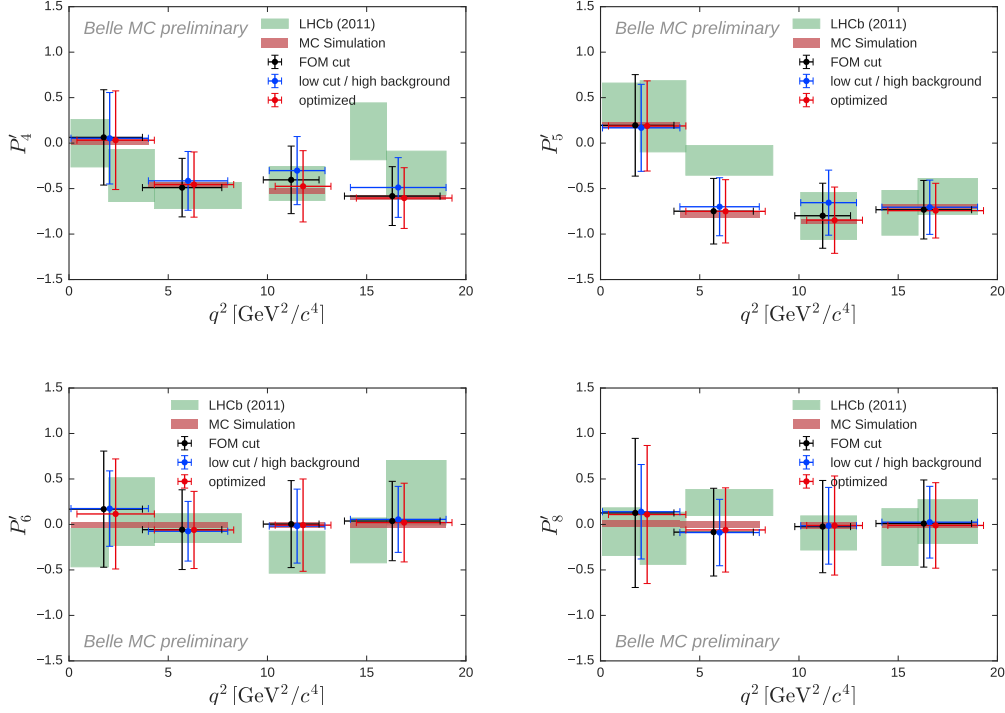


Figure 6.9: Result for different cuts for the neural network which lead to varied signal yields and purities in each measurement. All data points represents the averaged result of 10000 fits to pseudo data.

6.3.3 Neural Network Cut Optimization

The most important optimization covers the neural network cut and hence the signal to background ratio for both $B^0 \rightarrow K^*(892)^0 e^+ e^-$ and $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$. One can allow the classifier to run at different working points in purity vs. efficiency, leading to varied signal to background yields in the fit. As a default, the best cut estimation is based on the FOM (section 4.9.1). However, the best cut according to the figure of merit is not necessarily the best working point for the angular fit. Two different alternatives are tested in the first place: Cutting slightly softer on the neural network output leads to more signal events but worse purity and cutting harder leads to less total signal events but with a higher purity. The result is displayed in fig. 6.9. It shows clearly that a higher purity, thus a tighter neural network cut, is superior as the observed deviations from the simulated truth becomes smaller. It can be observed that the statistical sensitivity increases with lower cut on the neural network. However, this effect is negated by additional systematic uncertainties for the fit bias and peaking backgrounds, which are not displayed.

Particularly interesting is the result of the measurement of the second bin of P'_5 . For this bin, a grid search in the cuts for the electron (c_{ee}) and muon ($c_{\mu\mu}$) channel is performed. For each pair of cuts, signal and background yields are extracted from

the M_{bc} distribution on data. This creates a two dimensional map of the signal to background ratio in dependence of c_{ee} and $c_{\mu\mu}$. With this information, a toy study is performed to gain the statistical sensitivity and systematic error from a fit bias. Additional uncertainties for peaking backgrounds are added to the sensitivity estimation. The sensitivity is optimized by minimizing the total error of P'_5 , which is the linear addition of the statistical error from the toy study e_{stat} , the systematic error from the fit bias e_{bias} and the estimated error from peaking background $\hat{e}_{peaking}$. The latter is estimated from rare MC. From this procedure, the estimated sensitivity for each pair of cuts is obtained. The result can be seen in fig. 6.10. This procedure is computationally intensive for which reason the search is performed first in a coarse grid on a large range of the neural network outputs to find the area of the global minimum, depicted in fig. 6.10(a). Afterwards the minimum is localized in this area in a second search on a finer grid, displayed in fig. 6.10(b).

This optimized cut also delivers persistently good results regarding the fit bias for other bins and P' observables, which can be observed in fig. 6.9.

6.3.4 Fixing A_T or F_L

In some measurements a fit bias can be observed, for instance in third bin of P'_5 , see fig. 6.9. A test is performed to control this bias by fixing the values for F_L and/or A_T to the values in the simulation of the datasets. In fig. 6.11 one can see that no significant improvement can be achieved with this procedure compared to the result shown in fig. 6.9.

6.3.5 Estimating the Sensitivity in the Muon Mode

The statistical sensitivity of just using the channel $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ is evaluated separately. Since the statistics of the electron channel is found to be insufficient for stable fit results, a difference between a measurement of combined and muon only data would be an interesting observation. The result of the toy study based only on the data of the muon channel is depicted in fig. 6.12. One can observe that the sensitivity and the fit bias is significantly worse due to the reduced number of events in the fits.

6.3.6 Conclusions from Toy Studies

In summary, the following conclusions from the toy studies have been made for the final fit. The background parametrization is based on kernel density smoothed template histograms. Each P'_i observable is fitted individually together with A_T and F_L , as simultaneous fitting turned out to be worse. The neural network cut is determined with a grid-search for both channels simultaneously, optimizing the measurement for P'_5 in the second bin of q^2 .

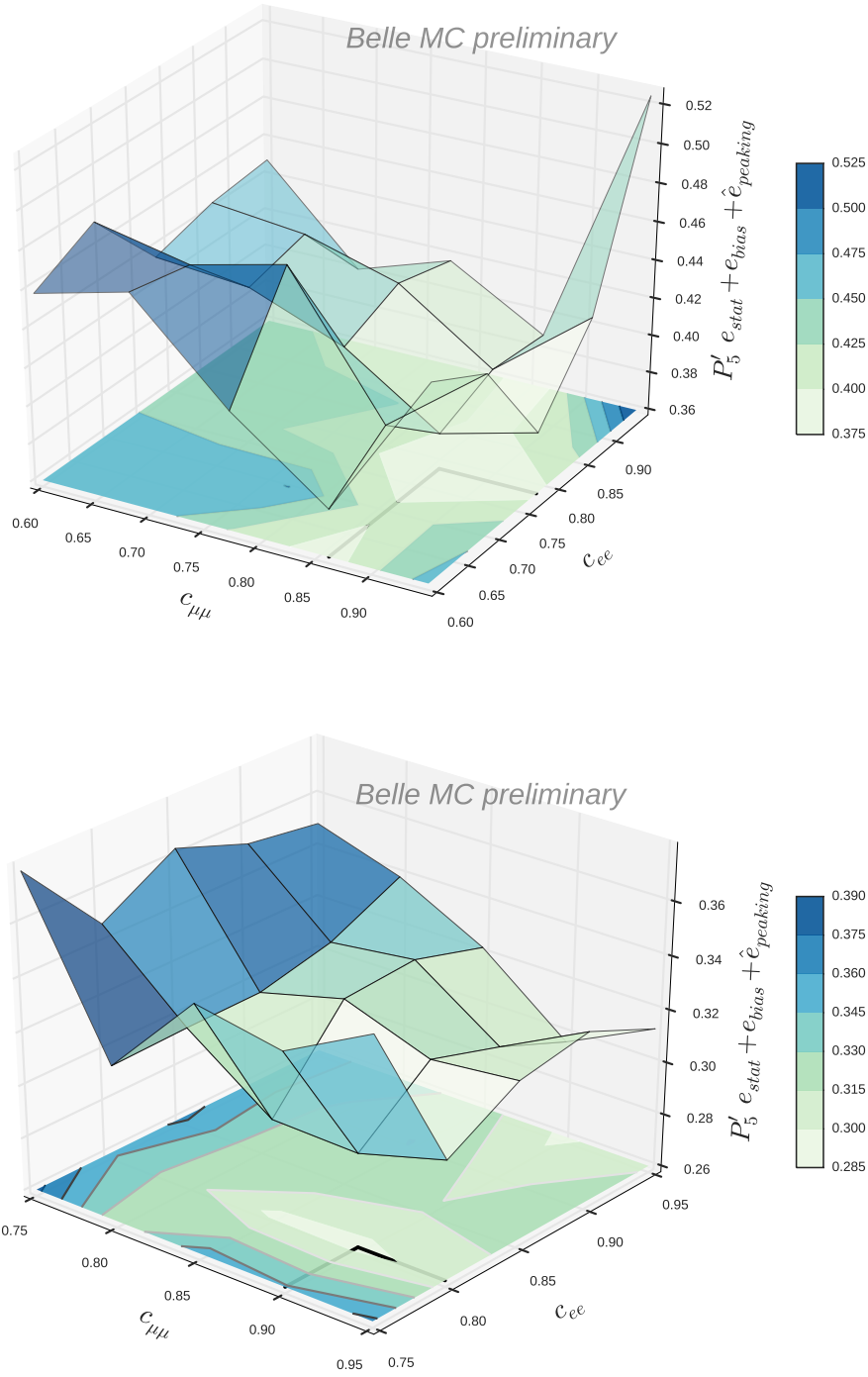


Figure 6.10: Cut optimization for the measurement of P'_5 in bin 2. The estimated total error $e_{stat} + e_{bias} + \hat{e}_{peaking}$ is displayed on the z-axis in dependence of the cuts for the electron (c_{ee}) and muon ($c_{\mu\mu}$) channel on a large range of cuts (top) and in the region with the global minimum (bottom).

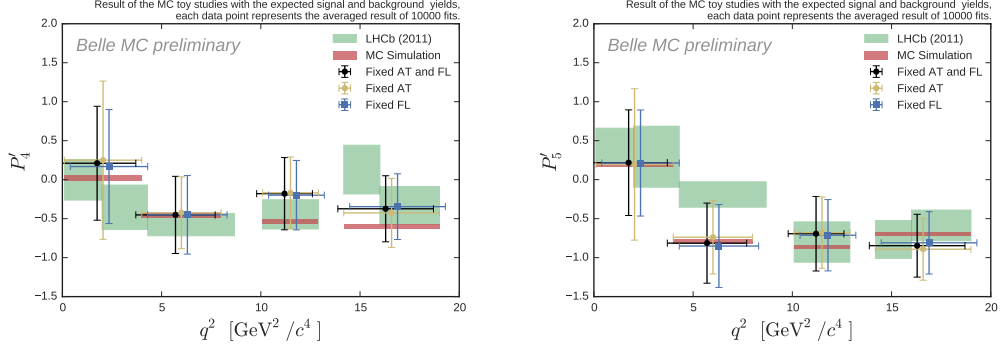


Figure 6.11: Result of the MC toy studies with fixed values for A_T or F_L and the expected signal and background yields for the FOM cut.

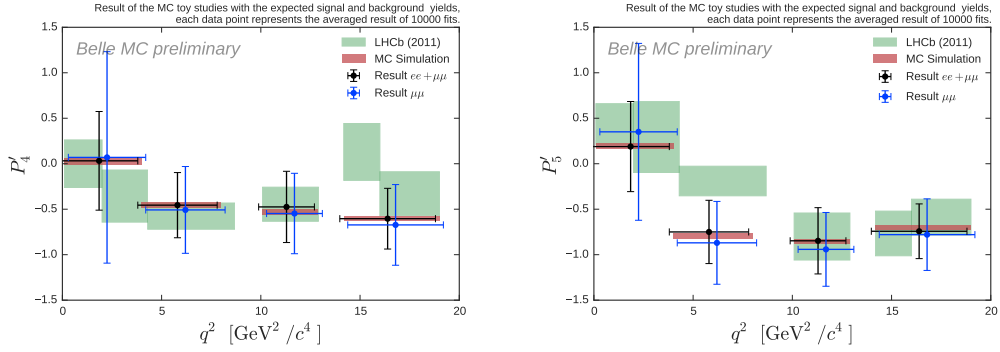


Figure 6.12: Estimation of the sensitivity for the muon mode only, displayed in addition to the combined sample.

6.4 Acceptance and Efficiency

To account for acceptance and efficiency effects in the fit, weights are assigned to the data. For the fit with efficiency, each event is weighted by the inverse of its combined efficiency. The dependence of the reconstruction efficiency is determined across all three angular observables and q^2 in each bin of q^2 . The total reconstruction efficiency is calculated from the direct product of the relative efficiencies f_{eff}^{fit} in the corresponding variables by

$$f_{eff}^{bin}(\cos \theta_\ell, \cos \theta_K, \phi, q^2) = f_{eff}^{fit}(\cos \theta_\ell) \otimes f_{eff}^{fit}(\cos \theta_K) \otimes f_{eff}^{fit}(\phi) \otimes f_{eff}^{fit}(q^2), \quad (6.8)$$

assuming that the efficiency is uncorrelated in the three-dimensional angular space and uncorrelated to q^2 . This assumption is investigated in fig. 6.13, which does not show significant correlations between the reconstruction efficiency among the angular variables. The correlation to q^2 , however, is significant and analyzed separately.

In the following, the method is presented for calculating the relative reconstruction efficiency $f_{eff}^{fit}(\mathcal{A})$ for an observable $\mathcal{A}$. This method calculates the difference between generated and reconstructed distributions on signal MC. One could in principle achieve this by dividing a histogram of the generated distribution for $\mathcal{A}$ by the reconstructed one. However, bins with low amount of statistics can lead to instabilities in this procedure. In this analysis, numerical stability is achieved by using flat distributions, with similar techniques as described in section 3.2.1. For this, the generated distribution x_{gen} from the signal MC is used and the cumulative distribution (CDF) is estimated. The CDF is estimated by a histogram, fitted with a spline-fit s_{gen}^{CDF} (see section 3.2.1). In the next step the reconstructed distribution x_{rec} , which includes reconstruction and acceptance effects, is transformed with the spline fit of the CDF. The distribution of the reconstruction efficiency is then defined as

$$x_{eff} = s_{gen}^{CDF}(x_{rec}), \quad (6.9)$$

where the axis corresponds to percentiles of the generated distribution x_{gen} . The final reconstruction efficiency is obtained by a spline-fit $f_{eff}^{fit}(\mathcal{A})$ to the distribution of x_{eff} . This method fits orthogonal splines to the data in a way that the pull between the fit and the data points becomes ideal, i.e. that it is a Gaussian with width one and mean zero. All fits for the efficiency in bins of q^2 are depicted in fig. 6.14. In these plots, the data points correspond to the distribution x_{eff} for the individual observable and the fit f_{eff}^{fit} is displayed additionally. In the case of uncorrelated efficiencies within the observables the generated and reconstructed distributions would be equal, consequently a flat distribution for x_{eff} would be expected. This is approximately the case for the observable ϕ in this measurement. Large variations of the efficiency across $\cos \theta_\ell$ and $\cos \theta_K$ can be observed. Especially for small angles between the leptons (θ_ℓ), the reconstruction efficiency is low for low values of q^2 .

Correlation to q^2

In fig. 6.13 strong non-linear correlations between q^2 and $\cos \theta_\ell$ are observed although their linear correlation coefficient is consistent with zero. As the efficiency is deter-

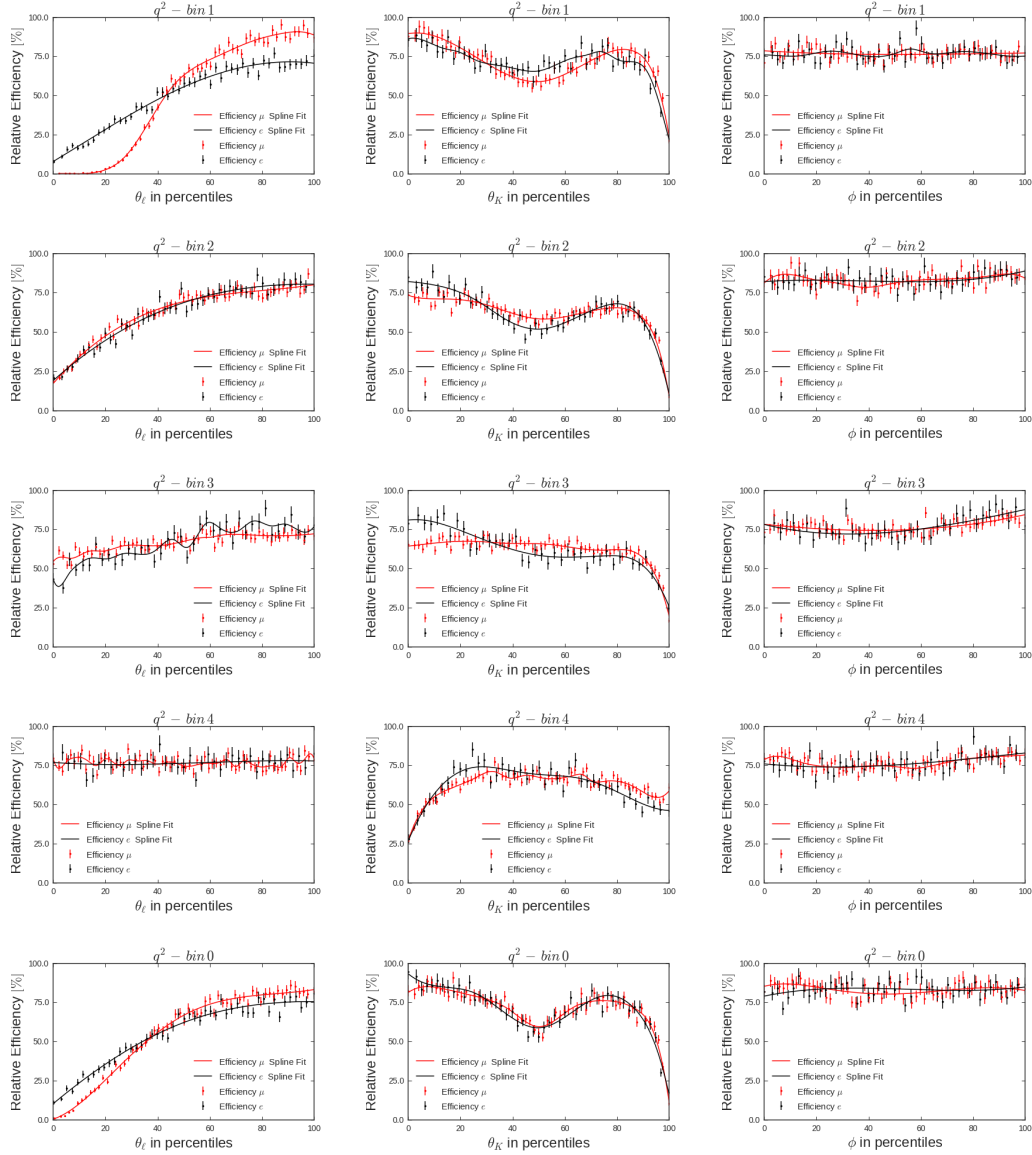


Figure 6.14: Results for the fits of the efficiencies. Both the fits for the electron (orange) and muon (blue) efficiency are superimposed over the ratio of generated and reconstructed events (data points).

mined in each bin, the average across changes within one bin is taken into account. In fig. 6.15 one can see the variation of the reconstruction efficiency of $\cos \theta_\ell$ in each bin. In the range of $2 < q^2 < 8 \text{ GeV}^2/c^4$ the transition of the reconstruction efficiency in $\cos \theta_\ell$ is strong, affecting in particular bins 0 and 1. For this correlation a systematic error is assigned, see chapter 7. For the angles $\cos \theta_K$ and ϕ no significant correlations within the bins are observed.

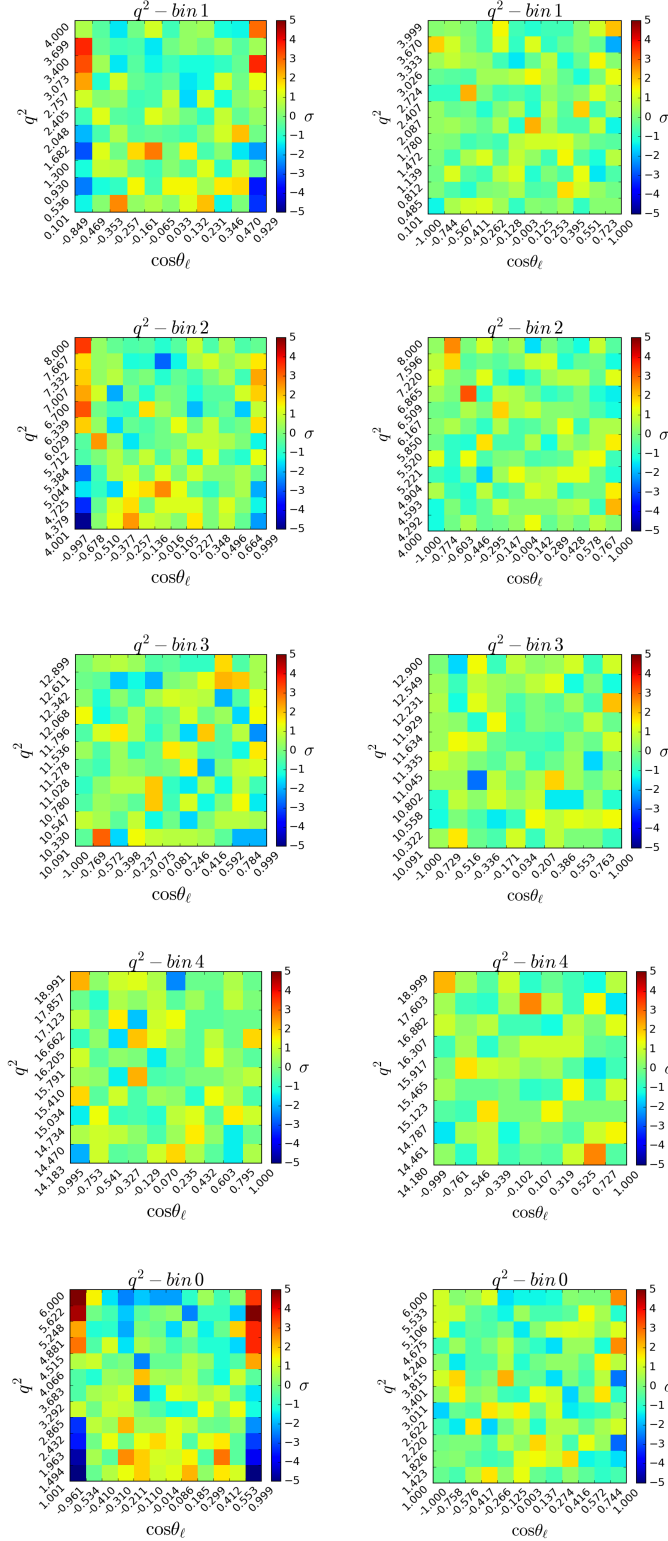


Figure 6.15: q^2 dependence of the reconstruction efficiency in $\cos\theta_\ell$ for each bin of q^2 for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ (left) and $B^0 \rightarrow K^*(892)^0 e^+ e^-$ (right) (see details for this kind of plot in section 3.2.2).

7. Systematic Study

For the angular analysis sources of systematic uncertainty are considered, if they introduce an angular or q^2 dependent bias to the distributions of signal or background candidates. Systematic uncertainties are examined using pseudo-experiments with large signal yields in order to minimize statistical fluctuations and compare the nominal with a varied model. The average variation between the two models is taken as systematic uncertainty.

7.1 Data to Monte Carlo Differences

A difference between data and Monte Carlo will influence the neural network classifier and thus may lead to a bias in the acceptance function.

From the data to MC comparison in the q^2 sidebands in the J/ψ region, only small deviations from the fits are expected, see figs. 5.6 and 5.7. As one can see in these plots, significant deviations occur only in $\cos\theta_K$. These deviations are modeled by applying a linear fit to the observed ratio between data and MC, depicted in fig. 5.6(e). This difference is subtracted twice from the fit for the efficiency on MC in order to simulate a mismodeling of the true efficiency. Thus, the fit for the efficiency weights does deviate from the actual reconstruction efficiency on MC by twice the deviation observed between data and MC. The result of the fit with a biased efficiency correction leads to modified results, where the observed deviation from the expected truth is taken into account as a systematic uncertainty. For possible influences in the other two variables ϕ and $\cos\theta_\ell$, the smoothness of the acceptance spline fit is increased, making it less accurately follow features in the data. The result of this modified acceptance function can be seen in fig. 7.1.

The bias on the angular variables for this modified acceptance function can be seen in fig. 7.2. This difference is taken into account as systematic error. The assigned errors are listed in table 7.1.

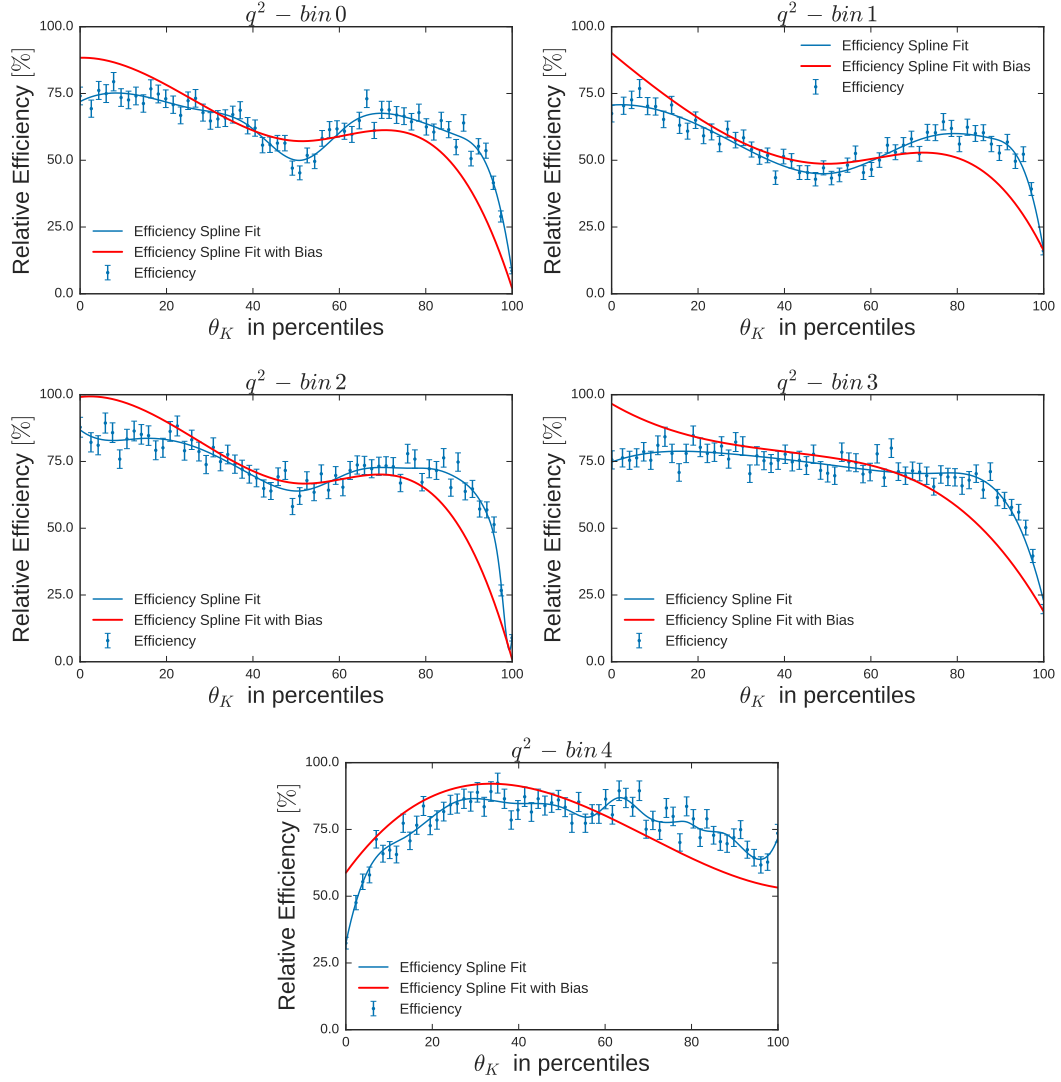


Figure 7.1: Fit for the acceptance function for $\cos\theta_K$ with a linearly depended bias. The bias simulates twice the observed deviation between data and MC in this variable. From a fit with the modification of the efficiency function the systematic uncertainty is derived for the data MC discrepancy.

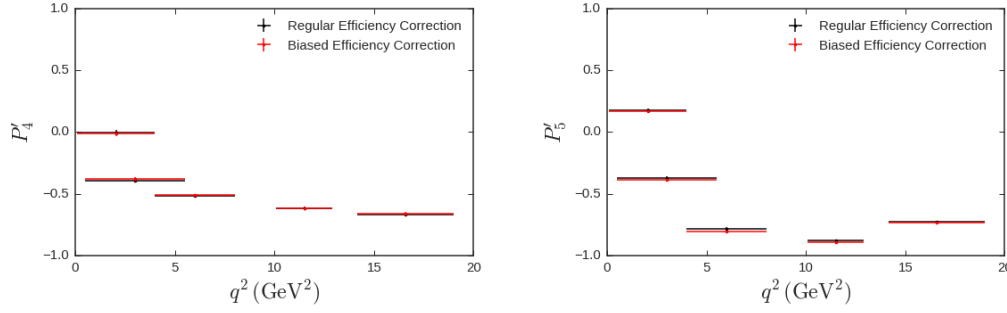


Figure 7.2: Fit to the simulated data with efficiency and acceptance effects from the neural network and with added bias to the $\cos\theta_K$ acceptance fit.

Table 7.1: Systematic uncertainty for the difference between data and MC.

bin	P'_4	P'_5	P'_6	P'_8
0	0.011	-0.011	-0.130	-0.143
1	-0.009	-0.007	-0.138	-0.163
2	0.002	-0.021	-0.165	-0.153
3	-0.000	-0.014	-0.220	-0.195
4	0.005	-0.003	-0.234	-0.232

7.2 Fit Bias

Due to the limited statistics, a fit bias in some bins of the angular analysis is observed. In 10000 pseudo experiments on simulated data the fit for each measurement is performed and the results are compared to the simulated values. The information of the mean of the pull distribution from the toy study is used for each measurement to determine a systematic bias on the measurement. The central values of the measurements are not corrected for the bias but the absolute value of the deviation is assigned as systematic error. The assigned corrections for each measurement are displayed in table 7.2.

Table 7.2: Fit bias for all measurements in the statistically preferred binning. The values are the absolute deviations from the simulated values.

bin	P'_4	P'_5	P'_6	P'_8
0	0.032	-0.003	0.073	0.034
1	0.011	-0.006	0.119	0.103
2	0.001	0.043	-0.056	-0.058
3	0.056	0.013	-0.003	-0.005
4	-0.003	-0.046	0.027	-0.005

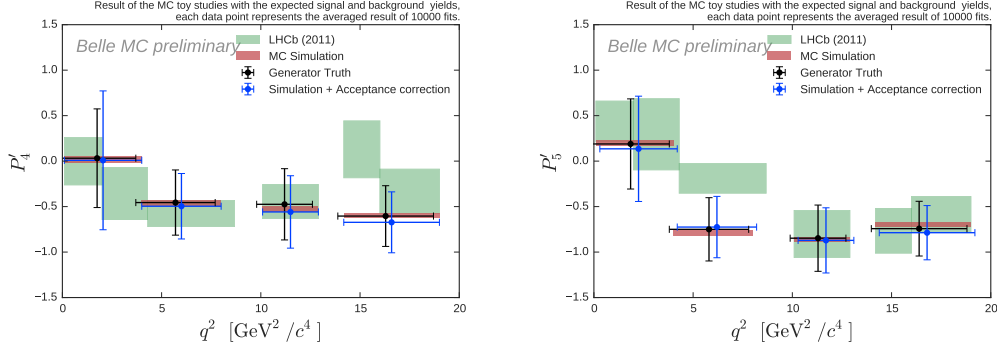


Figure 7.3: Toy Study With Efficiency: Displayed is the comparison between a toy study with data on generator level and data on reconstruction level with added weights according to the efficiency fit with the NN cut from the best figure of merit.

7.3 Efficiency Correction

For the fit of the reconstruction efficiency function a factorization of the efficiencies in the angular observables and q^2 is assumed. In figs. 6.13 and 6.15 one can see, that there are slight correlations between q^2 and $\cos\theta_\ell$ within the range of the bins.

With this correlation the reconstructed dataset including a correction for the efficiency will deviate slightly from the dataset from the generator. The deviation in results between a dataset of simulated data with efficiency correction weights and a dataset based on generator truth is evaluated. The individual bias can be seen in table 7.3.

Table 7.3: Bias for all measurements from correcting for the reconstruction efficiency.

bin	P'_4	P'_5	P'_6	P'_8
0	0.148	0.040	-0.084	-0.032
1	0.024	0.021	-0.043	0.082
2	0.060	-0.010	-0.068	-0.036
3	0.088	0.022	-0.022	-0.042
4	0.065	0.033	-0.019	0.010

Another possibility is to perform the toy study with data on reconstruction level and applied weights and combine the measured deviation into the systematic error to account for the fit bias and the acceptance function together. The result of the toy study with efficiency correction deviates in some bins slightly from the result on generator level, see fig. 7.3. With this method, the overall systematic error would be smaller compared to the case when one accounts for the fit bias and acceptance function bias separately.

7.4 Peaking Backgrounds

Peaking backgrounds are estimated for each q^2 bin in the measurement using MC. Peaking components are not considered in the fit to M_{bc} , which adds a systematic error on the estimation of the number of signal events in the signal region. In section 4.8 it is estimated that the only dominant contribution can arise from $B^0 \rightarrow K^* \pi \pi$ where both pions were misidentified as muons. In total less than 6 events are expected for the muon channel only. The impact of the peaking component is simulated by repeating the toy study and subsidizing 6 events from the signal with events from the peaking background in each bin. The mean deviation of the procedure is ± 0.027 for the value of $P'_{4,5,6,8}$, which corresponds to approximately 2 – 5% of the statistical error.

7.5 Other Sources of Systematic Uncertainties

The following sources of systematic uncertainties have been considered and are found to be negligible due to their small contribution.

7.5.1 Cross-Feed

The signal cross-feed is calculated for the B^0 decay channels ($B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$, $B^0 \rightarrow K^*(892)^0 e^+ e^-$, $B^0 \rightarrow K^0 \mu^+ \mu^-$, $B^0 \rightarrow K^0 e^+ e^-$) with the following procedure. Signal MC is generated for each of the channels. For each set of signal MC the corresponding neural network cut is applied and the remaining yields of the other three channels is counted. With this yield, one can calculate the efficiency for the cross-feed component. Together with the estimated number of candidates in the Belle dataset (table 4.2), one can consequently calculate the number of estimated cross-feed events. The numbers are presented in table 7.4. No relevant amount of cross-feed is observed.

Table 7.4: Estimated number of cross-feed events in the data.

decay channel	cross-feed	efficiency [%]	expected number of events
$B^0 \rightarrow K^*(892)^0 e^+ e^-$	$B^0 \rightarrow K^0 e^+ e^-$	0.069	0.085
$B^0 \rightarrow K^*(892)^0 e^+ e^-$	$B^0 \rightarrow K^0 \mu^+ \mu^-$	0.001	0.003
$B^0 \rightarrow K^*(892)^0 e^+ e^-$	$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	0.001	0.008
$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	$B^0 \rightarrow K^0 e^+ e^-$	0.002	0.002
$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	$B^0 \rightarrow K^*(892)^0 e^+ e^-$	0.000	0.000
$B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$	$B^0 \rightarrow K^0 \mu^+ \mu^-$	0.028	0.073

7.5.2 $K^*(892)$ S-Wave Component

The parametrization on eq. (1.37) does not include a potential S-wave contribution from $K^*(892)$ decays. For the S-wave contribution, the fraction F_S of the S-wave component in the K^{*0} mass window has to be considered in the differential decay

rate. Following Ref. [48] and denoting the right side of eq. (1.37) as W_P , one can add correction terms by

$$(1 - F_S)W_P + \frac{9}{32\pi}(W_S + W_{SP}), \quad (7.1)$$

where

$$W_S = \frac{2}{3}F_S \sin^2 \theta_\ell \quad (7.2)$$

and

$$\begin{aligned} W_{SP} = & \frac{4}{3}A_S \sin^2 \theta_\ell \cos \theta_K + A_S^{(4)} \sin \theta_K \sin 2\theta_\ell \cos \phi + A_S^{(5)} \sin \theta_K \sin \theta_\ell \cos \phi \\ & + A_S^{(7)} \sin \theta_K \sin \theta_\ell \sin \phi + A_S^{(8)} \sin \theta_K \sin 2\theta_\ell \sin \phi, \end{aligned} \quad (7.3)$$

which includes all the interference terms, $A_S^{(i)}$, of the S-wave with the K^{*0} transversity amplitudes, defined in [49].

The fraction F_S is searched for in our data by fitting the invariant mass of the $K\pi$ pair. The background shape is fixed on MC for random combinations. The resonant part of the K^{*0} decay mode is described by a non-relativistic Breit-Wigner shape with free shape parameters in the fits. For the S-wave component of the K^{*0} decay a third order Chebychev polynomial is used to describe discrepancies between the total data and a combination of the Breit-Wigner shape and the background shape. In order to get sufficient statistics, the fit is performed on a data sample with lowered neural network cut. The result is displayed in fig. 7.4(a). In the next step, the shape for the S-wave component is fixed and a fit to the final distribution of the K^* mass is performed, where the yield for the S-wave fraction is determined in the fit. The result $F_S = 0.000 \pm 0.024$ is consistent with zero and the fit result is displayed in 7.4(b).

7.5.3 $\mathcal{CP}$ Asymmetries

In eq. (1.37) the number of B^0 and $\bar{B}^0$ decays are assumed to be equal as the measured parameters correspond to symmetric or antisymmetric $\mathcal{CP}$ -averages terms. If there is a production, detection or direct $\mathcal{CP}$ asymmetry observed, the measured $\mathcal{CP}$ -symmetric parameters, defined in eq. (1.28), must be corrected by

$$S_i^{\text{obs}} = S_i - A_i(\mathcal{A}_{\mathcal{CP}} + \kappa\mathcal{A}_P + \mathcal{A}_D), \quad (7.4)$$

where $\mathcal{A}_{\mathcal{CP}}$ is the direct $\mathcal{CP}$ asymmetry between $B^0 \rightarrow K^{*0}\ell^+\ell^-$ and $\bar{B}^0 \rightarrow \bar{K}^{*0}\ell^+\ell^-$, $\mathcal{A}_P$ is the production asymmetry, which can be neglected at Belle and $\mathcal{A}_D$ is the detection asymmetry, which can be caused due to different interaction cross-sections with matter for K^+ and K^- mesons. The measured direct $\mathcal{CP}$ asymmetry is consistent with zero [50]. Also the yields of B^0 and $\bar{B}^0$ events are statistically equal in the signal region of our measurement with 153 and 150 events respectively. Together with the small theoretical values for the $\mathcal{CP}$ asymmetric parameters $A_i^{(s)}$ ($\lesssim \mathcal{O}(10^{-3})$) [18]) influences of this kind are neglected.

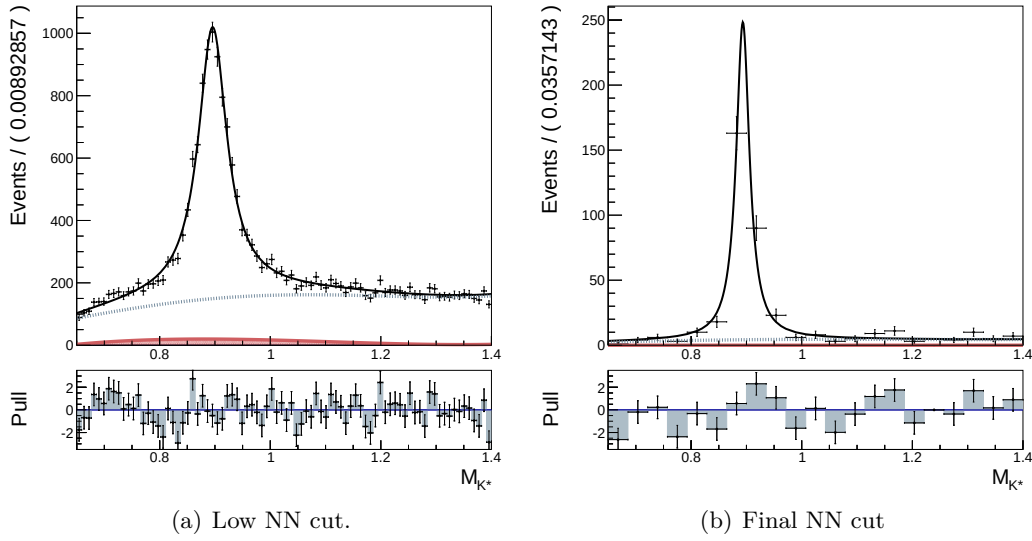


Figure 7.4: The $K^*(892)$ invariant mass in the decay of $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ for a low and a high cut on the neural network (NN). Combinatorial (dashed blue), potential S-wave signal (red filled) and total (solid) fit distributions are superimposed on the data points.

7.6 Summary

All sources of included systematics are summarized separately for P'_4 , P'_5 , P'_6 and P'_8 in tables 7.5 to 7.8. The total systematic uncertainty is calculated as the square root of the quadratic sum of all systematic uncertainties assuming they are uncorrelated.

Table 7.5: Summary of all considered systematic uncertainties for P'_4 .

Bin	0	1	2	3	4
Peaking Background	0.0855	0.0646	0.0366	0.0457	0.0358
Data/MC Difference	0.0109	0.0088	0.0020	0.0003	0.0047
Efficiency Correction	0.1475	0.0241	0.0599	0.0877	0.0650
Fit Bias	0.0316	0.0114	0.0007	0.0558	0.0027
Total	0.1738	0.0704	0.0702	0.1135	0.0744

Table 7.6: Summary of all considered systematic uncertainties for P'_5 .

Bin	0	1	2	3	4
Peaking Background	0.0901	0.0636	0.0078	0.0498	0.0131
Data/MC Difference	0.0112	0.0067	0.0208	0.0142	0.0029
Efficiency Correction	0.0397	0.0205	0.0098	0.0215	0.0327
Fit Bias	0.0031	0.0061	0.0430	0.0127	0.0460
Total	0.0992	0.0675	0.0494	0.0575	0.0580

Table 7.7: Summary of all considered systematic uncertainties for P'_6 .

Bin	0	1	2	3	4
Peaking Background	0.0170	0.0513	0.0229	0.0215	0.0026
Data/MC Difference	0.1298	0.1378	0.1655	0.2201	0.2341
Efficiency Correction	0.0835	0.0432	0.0683	0.0218	0.0192
Fit Bias	0.0735	0.1189	0.0562	0.0027	0.0268
Total	0.1718	0.1939	0.1890	0.2222	0.2364

Table 7.8: Summary of all considered systematic uncertainties for P'_8 .

Bin	0	1	2	3	4
Peaking Background	0.1242	0.0161	0.0395	0.0518	0.0255
Data/MC Difference	0.1433	0.1630	0.1531	0.1955	0.2316
Efficiency Correction	0.0319	0.0824	0.0359	0.0418	0.0099
Fit Bias	0.0337	0.1033	0.0579	0.0048	0.0047
Total	0.1952	0.2105	0.1722	0.2065	0.2332

8. Results and Discussions

8.1 Angular Analysis Results

The angular analysis is based on methods, which are evaluated with toy studies based on simulated events, as detailed in chapter 6. The results of the measurement are compared with Standard Model predictions.

8.1.1 Standard Model Predictions

The measurements are compared with Standard Model predictions based upon different theoretical calculations. Values from «DHMV» refer to the soft-form-factor method of Ref. [51]. The form factors are computed from QCD sum rules on the light-cone (LCSRs) with B meson distribution amplitudes of Ref. [52]. «BSZ» corresponds to using QCD form factors computed from LCSRs with K^* distribution amplitudes described in [53]. Predictions for the large q^2 bin are calculated using Lattice QCD with QCD form factors from Refs.[54, 55] the labels do not refer to DHMV and BSZ here. All provided values are calculated by the authors of Refs. [19, 51, 56] by doing a Gaussian scan to all parameters and taking the mean for the central value and one standard deviation as an error bar.

The third set of theoretical predictions is provided by the methods and authors of Refs. [57, 58] whose framework is specially tailored to the low q^2 region. It is referred to as «Camalich et.al.».

No predictions are made for the region close to the $c\bar{c}$ resonances J/ψ and $\psi(2S)$. Here, many of the assumptions made for the Standard Model predictions are thought to be violated due to interferences with the hadronic state.

8.1.2 Extraction

The procedure of the extraction of the final results is detailed in chapter 6 and follows directly insights gained from pseudo experiments. All observables $P'_{4,5,6,8}$ are

extracted from three-dimensional unbinned maximum likelihood fits in four bins of q^2 and the additional zeroth bin. Each P'_i is fitted together independently with F_L and A_T . Considering also the zeroth bin, which exhibits overlap with the range of the first and second bin, 20 independent measurements are performed. The background shape in each fit is determined on data in the M_{bc} sideband $M_{bc} < 5.27 \text{ GeV}/c^2$ and the signal to background fraction is determined by a fit to M_{bc} in each q^2 bin, see section 5.1 and table 5.1. A sample fit result for P'_5 in bin 2 is displayed in fig. 8.1 with the corresponding projections. The results are shown in figs. 8.2 to 8.6 together with available Standard Model predictions and LHCb measurements. All fit projections are depicted in appendix A.2.

Additionally, there are further derived quantities, calculated from the fit results. The $\mathcal{CP}$ -symmetric observables S_i are calculated from

$$S_{4,5,7,8} = P'_{4,5,6,8} \sqrt{F_L(1 - F_L)} \quad (8.1)$$

and

$$S_3 = \frac{1}{2}(1 - F_L)A_T. \quad (8.2)$$

Their values are displayed in fig. 8.7 together with the LHCb measurements. The result of all parameters in the fits are stated in table 8.1 together with the respective distance to Standard Model prediction. No significant $\mathcal{CP}$ asymmetry is observed. The yields of B^0 and $\bar{B}^0$ events are statistically equal in the signal region of the measurement with 153 and 150 events respectively.

8.2 Discussion and Outlook

The results of this analysis are compared to three different Standard Model predictions, described above. For each measured quantity, the distance to all three predictions is calculated in measures of the combined error, which is calculated as the square root of the quadratic sum of the theory error, the statistical error of the measurement and the systematic uncertainty. For P'_5 a deviation corresponding to 2.1σ is observed to the DHMV Standard Model prediction in the q^2 range $4.0 < q^2 < 8.0 \text{ GeV}^2/c^4$. Among the P' observables this is the largest observed deviation in the measurement. The pull distributions between predictions and measurements is shown in fig. 8.8. From fig. 8.8(a) one can clearly observe a distinct separation from this measurement to the others. For all three theoretical models the mean of the pull distribution is compatible with zero. Distributions for DHMV and BSZ are slightly shifted towards the deviation of the P'_5 anomaly. Additionally, their width are smaller than one, which means that at least one of the error sources is overestimated. This is an indication that the discrepancy in this bin is conservatively estimated.

The discrepancy in P'_5 supports measurements by LHCb [1], where a 3.7σ tension was observed in the region $4.30 < q^2 < 8.68 \text{ GeV}^2/c^4$. This analysis was originally intended to examine the same q^2 region. Theorists however expressed skepticism about unknown

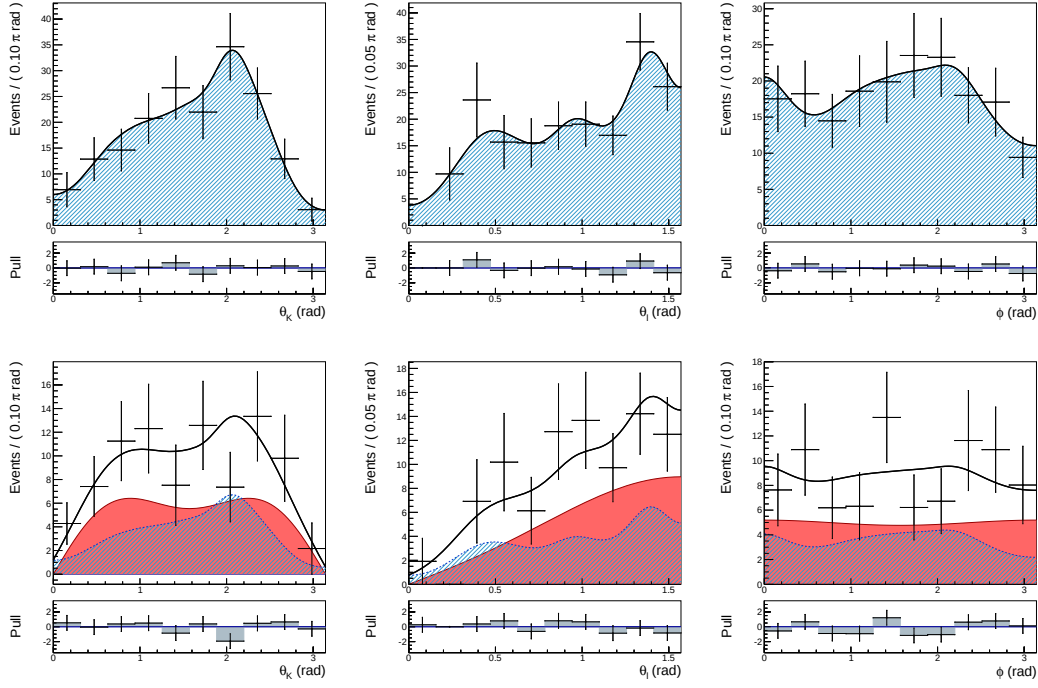


Figure 8.1: Projections for the fit result of P'_5 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

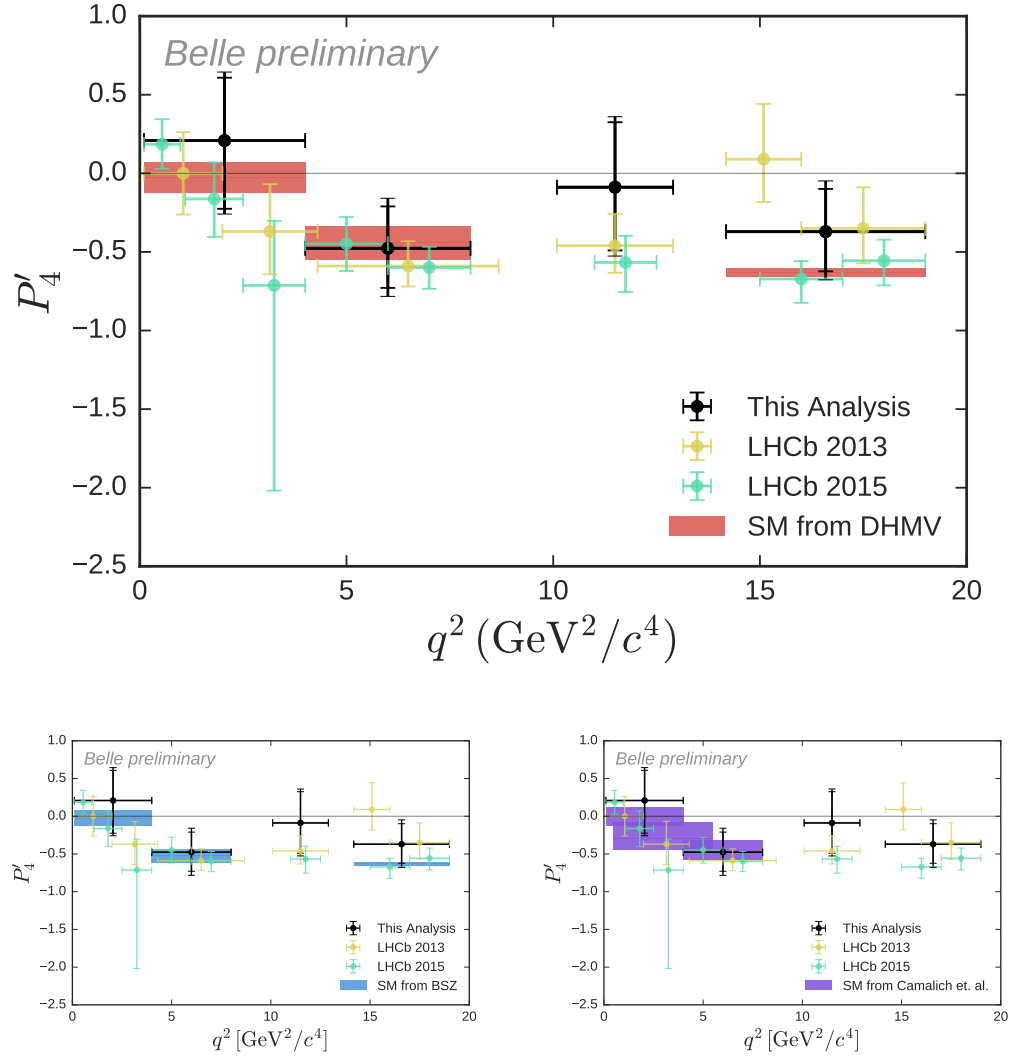


Figure 8.2: Result for P'_4 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.

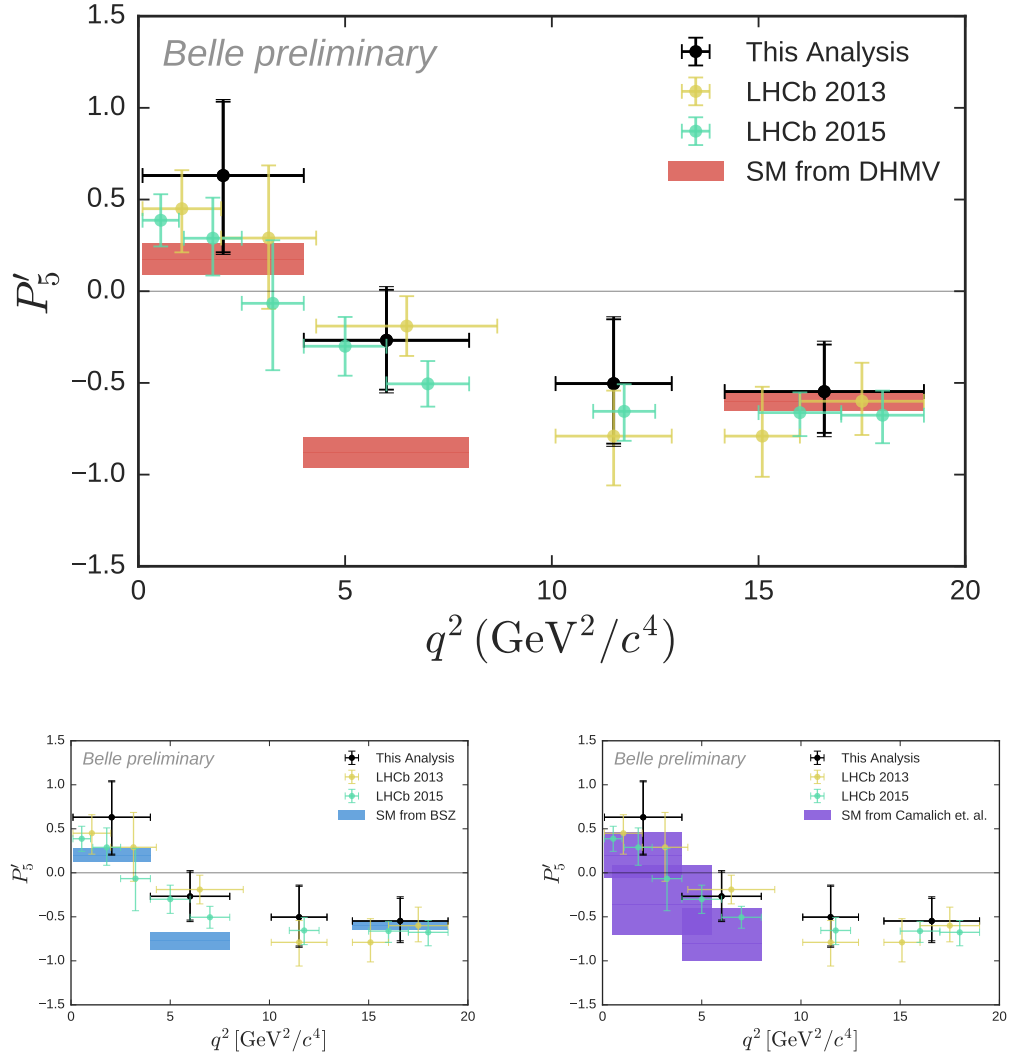


Figure 8.3: Result for P'_5 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.

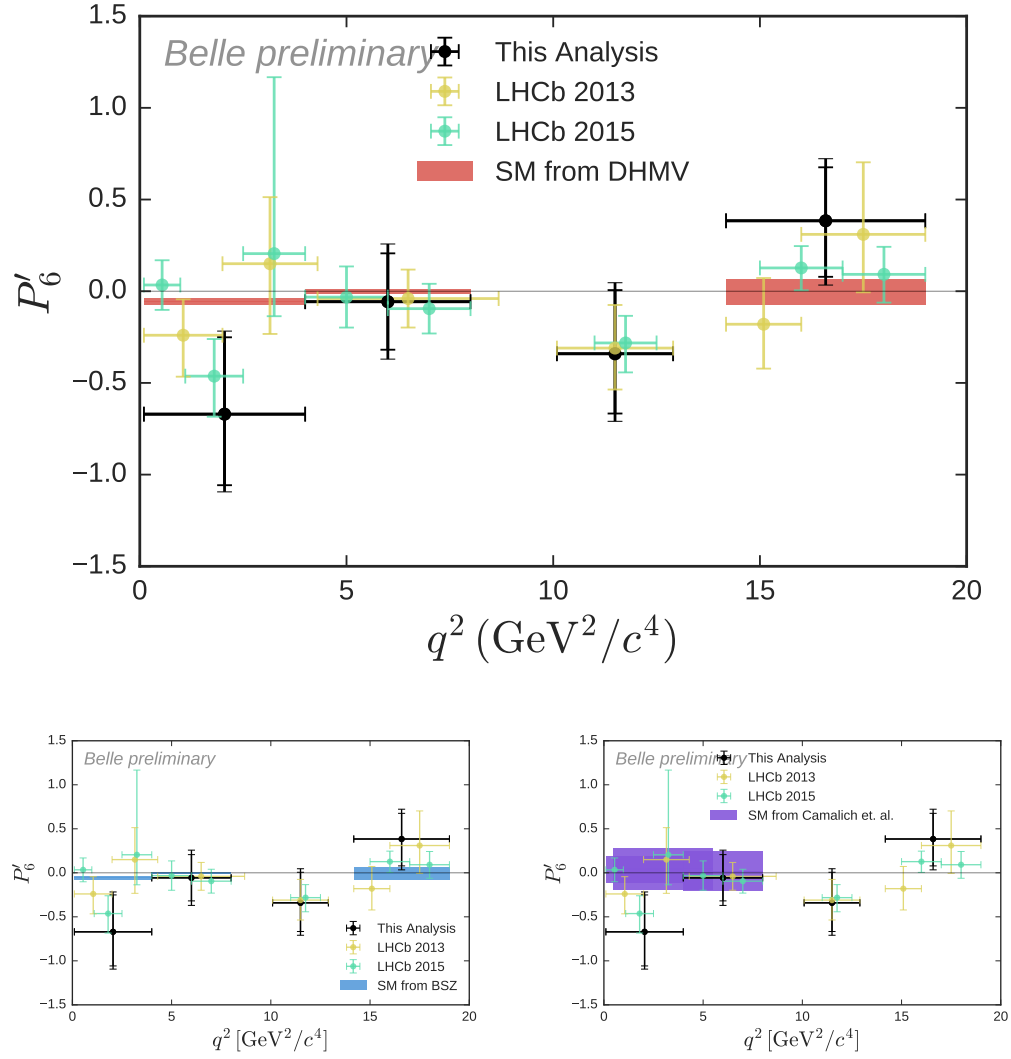


Figure 8.4: Result for P'_6 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.

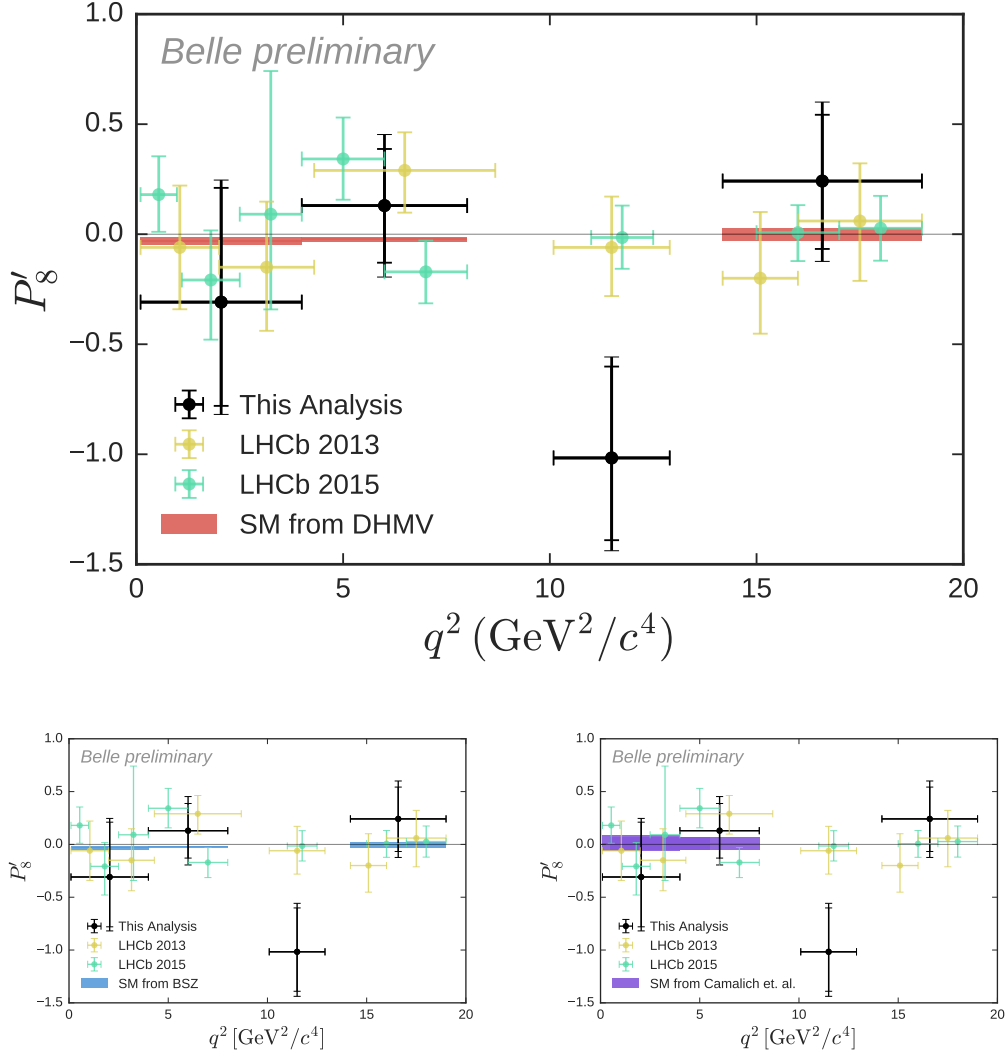


Figure 8.5: Result for P'_8 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.

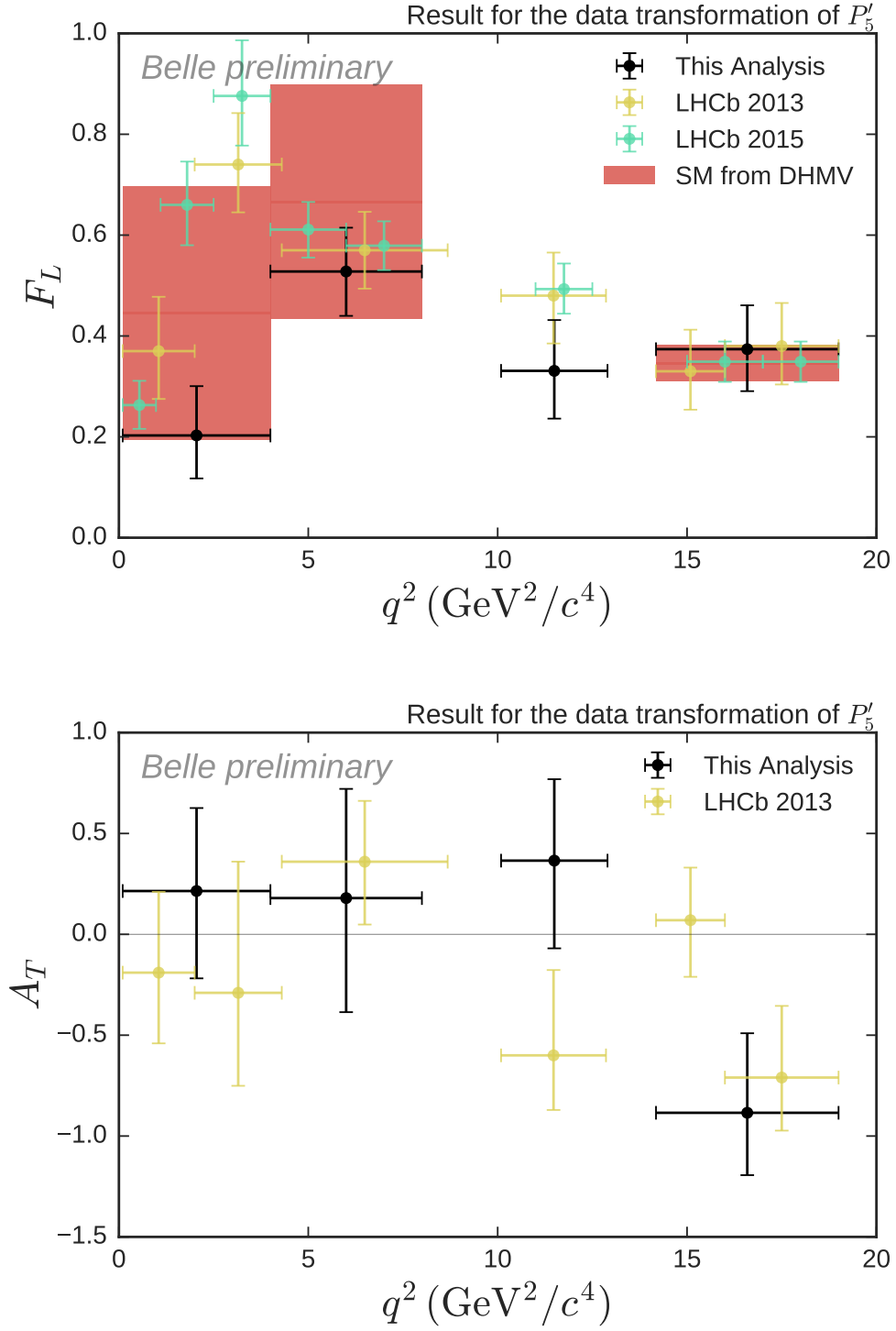


Figure 8.6: Result for F_L and A_T in the fit for P'_5 compared to Standard Model predictions by [19]. The error corresponds to the statistical only. Results from LHCb [1, 2] are shown for comparison.

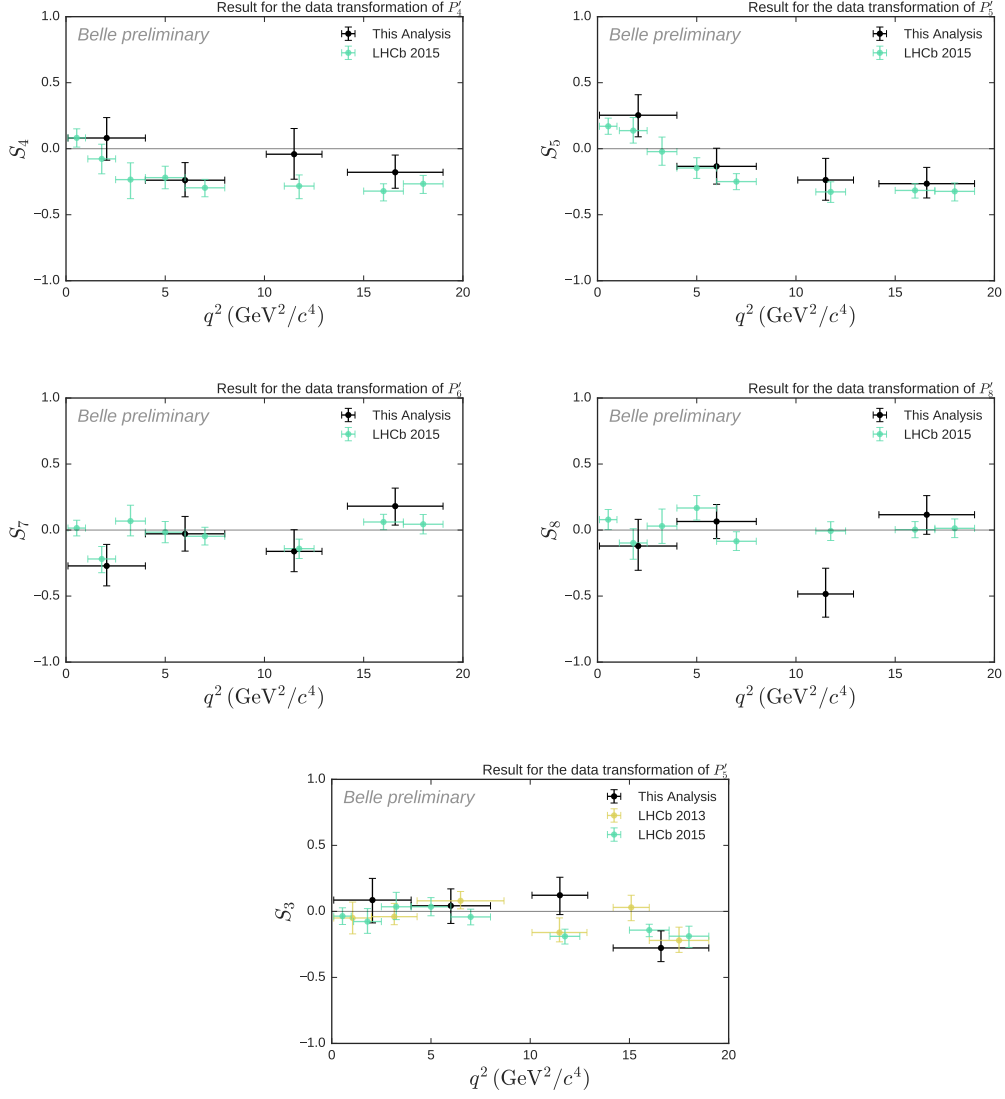


Figure 8.7: Result for derived quantities $S_{4,5,7,8}$ and S_3 . The error corresponds to the statistical only. Results from LHCb [1, 2] are shown for comparison.

theory errors originating from the hadronic J/ψ resonance at $q^2 > 8.0 \text{ GeV}^2/c^4$. LHCb performed an update on the analysis [2] with three times the integrated luminosity. In the update a discrepancy was observed in $4.0 < q^2 < 6.0 \text{ GeV}^2/c^4$ and $6.0 < q^2 < 8.0 \text{ GeV}^2/c^4$ at the level of 2.8σ and 3.0σ respectively. The global discrepancy to the DHMV Standard Model prediction in the measurement is stated with 3.4σ .

The largest discrepancy between the measurements performed by LHCb and this results can be observed in bin 3 of the observable P'_8 . Combining both total errors quadratically, the observed distance corresponds to 1.7σ for the LHCb result from 2013 which serves equal binning in this measurement and can consequently be compared. Considering all 20 independent measurements, a deviation of this magnitude in one measurement can be expected.

It has to be remarked that this Belle measurement is performed on a dataset consisting of 177.2 ± 16.6 events and includes muon and electron modes. Compared to the latest LHCb measurement this sample is smaller by a factor of 12. The LHCb detector is specialized in detecting charged hadronic and muonic final states, which makes $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ an ideal channel for the experiment. Furthermore the B cross section is large in $p\bar{p}$ -collisions at 8/13 TeV. Despite the differences in statistics the discrepancy measured in this analysis is significant and the central value aligns with both LHCb results.

In contrast to LHCb, the Belle detector is more efficient in reconstructing neutral final states and electrons. This analysis contains $B^0 \rightarrow K^*(892)^0 e^+ e^-$ decays in addition to the muon mode, which is currently not possible for LHCb. For the future it is planned to add charged decay modes $B^+ \rightarrow K^*(892)^+ e^+ e^-$ and $B^+ \rightarrow K^*(892)^+ \mu^+ \mu^-$ to the fit. It is expected that this will increase statistics by about 30%. The theory values will have to be slightly modified for that case. However, as isospin symmetry breaking is supposed to be almost negligible, the corrections are expected to be small.

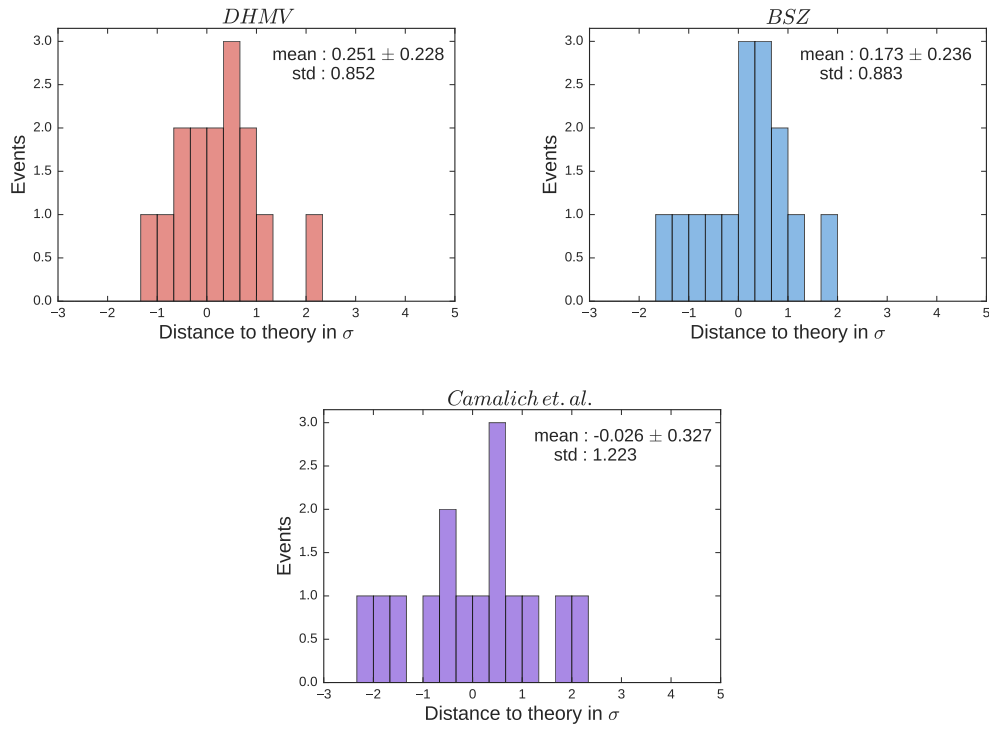


Figure 8.8: Distribution of the distance between measurements and theory predictions for the models "DHMV", "BSZ" and "Camalich et.al.".

Table 8.1: Results of the angular analysis. Observables are compared to Standard Model predictions for the analyzed q^2 regions, detailed in section 8.1.1. Systematic errors are only calculated and presented for $P'_{4,5,6,8}$, where the first errors is the statistical and the second the systematic error. Values for A_T and F_L are presented from the fit and data transformation of P'_5 .

q^2 Range in GeV^2/c^4	Observable	Measurement	Theory "DHMV"	Significance of deviation	Theory "BSZ"	Significance of deviation	Theory "Camalich et.al."	Significance of deviation
[1.00, 6.00]	P'_4	$-0.095^{+0.302}_{-0.309} \pm 0.174$	-	-	-	-	$-0.300^{+0.090}_{-0.080}$	0.56σ
	P'_5	$0.385^{+0.276}_{-0.285} \pm 0.099$	-	-	-	-	$-0.360^{+0.190}_{-0.170}$	2.09σ
	P'_6	$-0.202^{+0.278}_{-0.270} \pm 0.172$	-	-	-	-	$0.040^{+0.100}_{-0.100}$	0.71σ
	P'_8	$0.440^{+0.311}_{-0.320} \pm 0.195$	-	-	-	-	$0.010^{+0.020}_{-0.019}$	1.14σ
	F_L	$0.339^{+0.088}_{-0.084}$	-	-	-	-	$0.720^{+0.090}_{-0.120}$	2.56σ
	A_T	$0.168^{+0.354}_{-0.367}$	-	-	-	-	-	-
[0.10, 4.00]	P'_4	$0.208^{+0.400}_{-0.434} \pm 0.070$	-0.026 ± 0.098	0.52σ	-0.029 ± 0.103	0.53σ	$-0.010^{+0.060}_{-0.060}$	0.49σ
	P'_5	$0.631^{+0.403}_{-0.419} \pm 0.067$	0.175 ± 0.086	1.05σ	0.199 ± 0.077	1.00σ	$0.200^{+0.110}_{-0.110}$	0.98σ
	P'_6	$-0.670^{+0.419}_{-0.387} \pm 0.194$	-0.055 ± 0.018	1.33σ	-0.055 ± 0.018	1.33σ	$0.040^{+0.060}_{-0.060}$	1.53σ
	P'_8	$-0.309^{+0.519}_{-0.472} \pm 0.210$	-0.030 ± 0.017	0.50σ	-0.031 ± 0.016	0.50σ	$0.006^{+0.033}_{-0.033}$	0.56σ
	F_L	$0.190^{+0.099}_{-0.086}$	0.445 ± 0.251	0.95σ	0.535 ± 0.178	1.69σ	$0.530^{+0.120}_{-0.130}$	2.08σ
	A_T	$0.154^{+0.397}_{-0.417}$	-	-	-	-	-	-
[4.00, 8.00]	P'_4	$-0.477^{+0.266}_{-0.252} \pm 0.070$	-0.441 ± 0.106	0.12σ	-0.521 ± 0.087	0.16σ	$-0.490^{+0.070}_{-0.060}$	0.05σ
	P'_5	$-0.267^{+0.275}_{-0.269} \pm 0.049$	-0.881 ± 0.082	2.15σ	-0.770 ± 0.100	1.72σ	$-0.810^{+0.170}_{-0.140}$	1.68σ
	P'_6	$-0.057^{+0.264}_{-0.262} \pm 0.189$	-0.003 ± 0.011	0.17σ	-0.002 ± 0.008	0.17σ	$0.020^{+0.110}_{-0.110}$	0.23σ
	P'_8	$0.130^{+0.257}_{-0.259} \pm 0.172$	-0.022 ± 0.010	0.49σ	-0.020 ± 0.007	0.48σ	$0.007^{+0.020}_{-0.021}$	0.39σ
	F_L	$0.525^{+0.086}_{-0.087}$	0.666 ± 0.232	0.57σ	0.682 ± 0.153	0.90σ	$0.640^{+0.130}_{-0.150}$	0.67σ
	A_T	$0.173^{+0.542}_{-0.575}$	-	-	-	-	-	-
[10.09, 12.90]	P'_4	$-0.088^{+0.414}_{-0.402} \pm 0.114$	-	-	-	-	-	-
	P'_5	$-0.504^{+0.351}_{-0.327} \pm 0.057$	-	-	-	-	-	-
	P'_6	$-0.341^{+0.347}_{-0.326} \pm 0.222$	-	-	-	-	-	-
	P'_8	$-1.017^{+0.415}_{-0.374} \pm 0.207$	-	-	-	-	-	-
	F_L	$0.348^{+0.098}_{-0.093}$	-	-	-	-	-	-
	A_T	$0.245^{+0.385}_{-0.405}$	-	-	-	-	-	-
[14.18, 19.00]	P'_4	$-0.371^{+0.272}_{-0.252} \pm 0.074$	-0.632 ± 0.026	0.99σ	-0.632 ± 0.026	0.99σ	-	-
	P'_5	$-0.547^{+0.257}_{-0.225} \pm 0.058$	-0.601 ± 0.051	0.23σ	-0.601 ± 0.051	0.23σ	-	-
	P'_6	$0.384^{+0.292}_{-0.306} \pm 0.236$	-0.004 ± 0.069	0.99σ	-0.004 ± 0.069	0.99σ	-	-
	P'_8	$0.242^{+0.301}_{-0.309} \pm 0.233$	-0.001 ± 0.028	0.63σ	-0.001 ± 0.028	0.63σ	-	-
	F_L	$0.363^{+0.088}_{-0.085}$	0.346 ± 0.036	0.18σ	0.346 ± 0.036	0.18σ	-	-
	A_T	$-0.696^{+0.383}_{-0.317}$	-	-	-	-	-	-

Part III

Search for $B^+ \rightarrow K^+ \tau^+ \tau^-$

9. Motivation and Analysis Overview

The second analysis in this thesis focuses on the decay $B^+ \rightarrow K^+ \tau^+ \tau^{-1}$, which is experimentally only very poorly constrained [59]. Compared to the electron and muon mode of $b \rightarrow s \ell^+ \ell^-$ the tau decay mode is more difficult to analyze, due the presence of at least 2 neutrinos in the final state. As a consequence, one cannot detect the full decay chain in the detector, since neutrinos remain hidden and it is not possible to detect signal candidates in invariant mass or energy distributions because information is missing. In order to reconstruct the decay, the full reconstruction technique is used, which is explained in section 3.5. With this method a second B meson is searched for in the same event and fully reconstructed. This reduces the reconstruction efficiency drastically, as the efficiency for finding a B meson in purely hadronic final states is about a few permille.

Despite these experimental challenges, the decay mode offers interesting aspects to be examined. It belongs to a class of rare decays, which are sensitive to a variety of new physics scenarios. Recent results in $b \rightarrow s \ell^+ \ell^-$ related measurements, as for instance the P'_5 deviation which is covered in the first analysis of this thesis, make it a promising candidate for new insights. Especially the so called R_K anomaly is related to the tau decay mode and might reveal deviations to Standard Model predictions. The R_K anomaly was found by LHCb [5] in the ratio between the electron and muon branching fraction

$$R_K \equiv \frac{\mathcal{B}(B^+ \rightarrow K^+ \mu \mu)}{\mathcal{B}(B^+ \rightarrow K^+ e e)} = 0.745^{+0.090}_{-0.074}(\text{stat}) \pm 0.036(\text{syst}), \quad (9.1)$$

which is supposed to be one in the Standard Model [60] to a high precision due to lepton universality. Unlike the angular observables discussed in the previous part this quantity is almost free of theoretical uncertainties as most hadronic effects cancel in this ratio. The observed deviation correspond to 2.6σ . In particular the

¹The charge conjugated mode is always implied if not especially stated.

branching ratio of $B^+ \rightarrow K^+ \tau^+ \tau^-$, may deliver important hints regarding lepton flavor universality and may provide evidence for the existence of new physics in the flavor sector.

The boundaries for $b \rightarrow s \tau \tau$ are poorly constrained and a comprehensive study of new physics effects in this channel is performed in Ref. [61]. In some theoretical models the τ modes of $B \rightarrow K^{(*)} \ell^+ \ell^-$ are preferred for searching for new physics. New particles could couple to the mass of the comparably heavy τ lepton, and thus enhance the sensitivity by a factor of $|m_\tau/m_\mu|^2 \simeq 286$. The only experimental study for $B^+ \rightarrow K^+ \tau^+ \tau^-$ was performed by BaBar and an upper limit of 3.3×10^{-3} [59] at 90% confidence was set. This limit is about four orders of magnitude above Standard Model prediction for the branching ratio, which is $\mathcal{B}(B^+ \rightarrow K^+ \tau \tau)^{SM} < 1.44(15) \times 10^{-7}$ [62]. The Minimal Lepton Flavor Violation (MLFV) scenario is examined in Ref. [3] and concludes that effects from new physics could enhance the branching ratio to

$$\mathcal{B}(B \rightarrow K \tau^- \tau^+)^{MLFV} < 2 \times 10^{-4}, \quad (9.2)$$

which is already close to the current experimental limits.

In this analysis the full Belle dataset is used and it is performed blinded, which means that all methods are evaluated and tested on simulated data before real data is examined. The neural network based full-reconstruction technique at Belle [42] is used to reconstruct a tag-side B meson. Using kinematic constraints from the knowledge of the initial state of the $e^+ e^-$ pair and the tag-side B , the signal B meson is reconstructed. After removing all tracks and hits from the signal and tag-side B candidates from the event, the remaining energy deposition in the electromagnetic calorimeter (ECL) is accumulated and the sum of the extra energy E_{ECL} is calculated. For signal events one does not expect additional energy in the calorimeter, hence a peaking distribution at zero in E_{ECL} . A counting experiment is performed in a signal window in this variable after the final selection to calculate the upper limit on $B^+ \rightarrow K^+ \tau^+ \tau^-$. In chapter 10 the reconstruction of the signal and the suppression of backgrounds is detailed. With the expected number of signal and background events and the sensitivity for signal candidates the expected upper limit is calculated in chapter 11. Finally, systematic uncertainties are analyzed in chapter 12.

Signal MC

Two sets of signal Monte Carlo are generated for the analysis. In each, the decay $B^+ \rightarrow K^+ \tau^+ \tau^-$ is simulated 10 million times. Each event of this sample contains a signal candidate and a generically decaying B meson. Two different generator models are used with form-factor calculations are provided in Ref. [63] and Ref. [41].

10. Reconstruction and Analysis

10.1 Event Selection

The first step is to search for a tag-side B meson in each event, the B_{tag} . This is done by a neural network based full-reconstruction method [42]. The algorithm delivers candidates for $B^\pm$ mesons in each event using many exclusive hadronic decay channels. If one or more tag-side candidates are found by this algorithm, the rest of the event is used to find the signal-side B meson, B_{sig} , in the decay $B^+ \rightarrow K^+ \tau^+ \tau^-$. For each B_{tag} candidate all belonging tracks and hits in the calorimeter are removed from the event and the signal side selection is applied. Furthermore, it is required that no additionally π^0 and exactly three additional charged tracks remain in the event. One of the charged tracks has to be identified as a $K^\pm$ with the opposite charge than the B_{tag} and a particle identification likelihood ratio $\mathcal{R}_K \equiv \mathcal{L}_K / (\mathcal{L}_K + \mathcal{L}_\pi)$ greater than 0.6. The remaining two charged tracks are required to have opposite charge without further constraints. With this selection, the τ is identified in single-prong τ decays $\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$, $\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$ and $\tau^+ \rightarrow \pi^+ \nu_\tau$, summarized in table 10.1. In order to lower the amount of multiple candidates per events, only the two most probable B_{tag} candidates are accepted, based on their neural network output.

Table 10.1: List of the reconstructed τ decay channels.

Decay channel	Branching fraction [%]
$\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e$	17.4 ± 0.4
$\tau^+ \rightarrow \mu^+ \nu_\tau \bar{\nu}_\mu$	17.8 ± 0.4
$\tau^+ \rightarrow \pi^+ \nu_\tau$	10.8 ± 0.6
Sum	46.0 ± 0.8

10.1.1 Best Candidate Selection

More than one tag-side candidate is observed in 20% of the events on signal MC. For the remaining two oppositely charged tracks besides the $K^\pm$, there are six possibilities for the mass hypothesizes: $ee, e\mu, e\pi, \mu\mu, \mu\pi$ and $\pi\pi$ when τ decays are reconstructed in $\tau^+ \rightarrow e^+ \nu_\tau \bar{\nu}_e, \tau^+ \rightarrow \mu \nu_\tau \bar{\nu}_\mu$ or $\tau^+ \rightarrow \pi^+ \nu_\tau$. The analysis consists of six categories, one for each of the charged final states of the two taus. All of these decay-modes are treated combined in the analysis. The mass hypothesis is assigned for each of the two tracks using pre-trained neural networks from the neural network based full-reconstruction framework [42]. These NeuroBayes classifiers were trained separating i.a. e, μ and $\pi^\pm$ from other charged primary particles. The mass hypothesis with the highest corresponding neural network output is accepted. The information of the individual decay-channel is taken into the classifying stage as a dedicated variable. All event selection criteria are summarized in table 10.2.

Table 10.2: Summary of the event selection criteria.

	Selection
$K^\pm$	$\mathcal{R}_K \equiv \mathcal{L}_K / (\mathcal{L}_K + \mathcal{L}_\pi) > 0.6$ $\text{charge}(K) = - \text{charge}(B_{tag})$
B_{Tag}	Rank of the $B_{tag} \leq 2$
Number of remaining charged Tracks	3
Number of remaining π^0	0

10.2 Pre-Selection Cuts

This analysis aims to use as few cuts as possible and rather classifies signal and background using multivariate methods. However, some pre-cuts are performed in variables removing large fractions of background while conserving the majority of the signal. One of the most important pre-cuts is the cut on the neural network output of the B_{tag} , $\mathcal{NB}(B_{tag}) > 0.0001$, which preserves almost the entire signal while removing more than 50% of the background. The variable q^2 , described in eq. (10.1), can be interpreted as the constrained invariant mass of the τ pair. It is calculated using the momenta of the $\Upsilon(4S)$, the B_{tag} and the reconstructed $K^\pm$ momentum and is able to reject background in the region $q^2 < 12 \text{ GeV}^4/c^2$ almost without losing any signal candidates. Furthermore, a signal window in M_{bc} of the B_{tag} and the E_{ECL} variable is defined. All the cuts are listed in table 10.3. The corresponding signal and background efficiencies are displayed in fig. 10.1.

10.2.1 Background Composition

After the event selection is performed, the data consists of both the desired decay $B^+ \rightarrow K^+ \tau^+ \tau^-$ and miss-reconstructed background events. The majority of the background originates from generic $b \rightarrow c$ transitions like $B^+ \rightarrow D^0 (\rightarrow K^+ \pi^-) \ell^+ \nu_\ell$

Table 10.3: Summary of pre-cuts used in the analysis. See section 10.3 for a definition of the variables. The efficiencies are determined independently.

Cut	Lost signal [%]	Lost background [%]
$q^2 > 12$	0.10	18.43
$M_{bc}(B_{tag}) > 5.27$	14.45	75.96
$\mathcal{NB}(B_{tag}) > 0.0001$	2.43	55.61
$E_{ECL} < 1.5$	2.45	25.94

decays. As charged $B^\pm$ mesons are selected, the largest background component is due to random combinations in the **charged** (B^+B^-) generic MC. Without a tight selection on the neural network output of the B_{tag} , also the other components like **mixed** ($B^0\bar{B}^0$) generic MC, as well as continuum events $e^+e^- \rightarrow q\bar{q}$ ($q = u, d, s, c$), contribute (**uds** and **charm** MC).

Continuum In continuum events, e^+e^- annihilates into light quark pairs $u\bar{u}, d\bar{d}, s\bar{s}$ as well as events containing charm quarks $c\bar{c}$. The initial quark pair however exhibits a large energy release, forming back to back jet-like structures. The full-reconstruction is used with continuum suppression, which combines the quality of the B_{tag} meson with event shape parameters (modified Fox Wolfram Moments [44]) and allows for rejecting this kind of background by cutting on the network output of the B_{tag} .

Combinatorial Combinatorial background arises from wrong combination of tracks in $B\bar{B}$ events.

Peaking Irreducible peaking background arises from missing K_L in the event. This source can be controlled on data reconstructing an additional K_S and removing it from the event. This study is described in chapter 12.

10.3 Background Suppression

For the background suppression a Boosted Decision Tree (BDT) with a gradient boost is chosen. These algorithms have shown exceptionally good results in recent years and are described in section 3.4.5. They can be used without preprocessing the data. Independent data samples for training and testing of the classifier are used to ensure that no over-fitting is performed.

A set of 35 variables which provide reasonable separation between signal and background events is available for classification, depicted in figs. B.1 and B.2. Subsets of variables are chosen to accord to their importance to the classifier. Only the best variables regarding their separation power are kept. A larger set of variables may show better performance but is vulnerable to over-fitting and differences between data and Monte Carlo might have an influence.

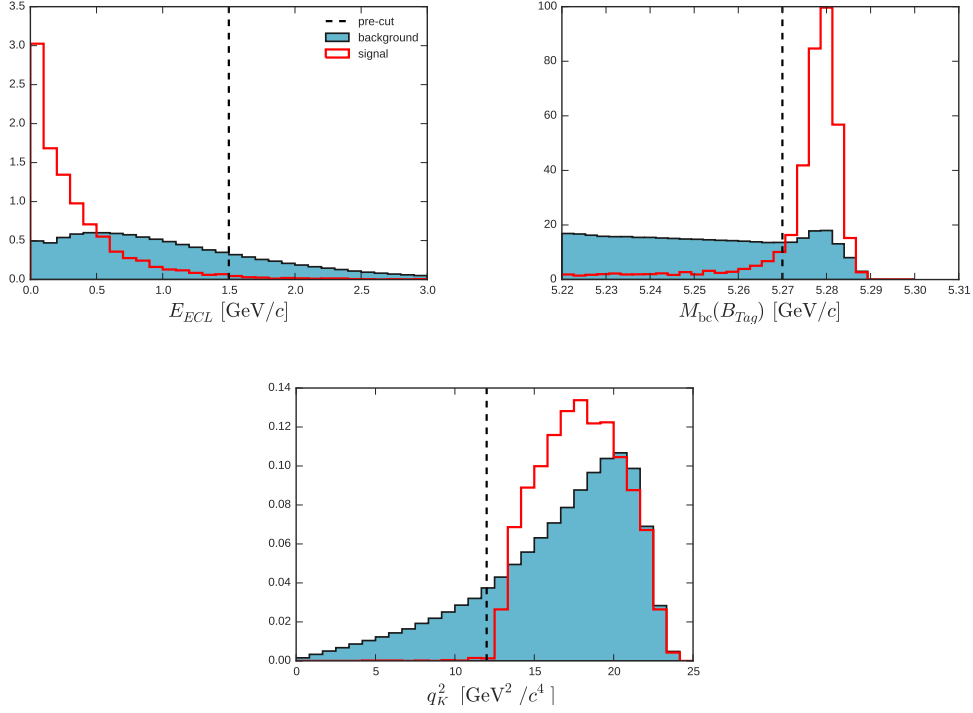


Figure 10.1: Pre-cuts on simulated data. The distributions are arbitrarily normalized.

The agreement between Monte Carlo and data is tested on a sideband in E_{ECL} , described in detail in sections 12.1 and 12.1. Due to increasing systematic errors more sensitivity might be lost than gained by the information in variables with bad agreement. Consequently, variables with poor data/MC agreement are removed from the training set.

In total 14 variables are kept for the separation between signal and background:

$\mathcal{NB}(B_{tag})$: The NeuroBayes output of the B_{tag} candidate.

$M_{K^+\tau^-}$: Invariant mass of the K^+ and charged daughter of the τ^- .

$\hat{p}_{\tau^+}$: The momentum of the positively charged τ in the rest frame of the signal B candidate.

decay channel : Decay hash value corresponding to the six possibilities for the mass hypotheses of the charged children of the τ pair ($ee, e\mu, e\pi, \mu\mu, \mu\pi$ and $\pi\pi$).

$\mathcal{NB}(\tau^+ \times \tau^-)$: The product of the NeuroBayes outputs of the children of both τ .

ΔE^{tag} : The beam constrained energy of the B_{tag} candidate.

q^2 : The constrained invariant mass of the τ pair, defined as

$$q^2 \equiv (\vec{p}_{(\Upsilon(4S))} - \vec{p}_{B_{tag}} - \vec{p}_{K^+})^2, \quad (10.1)$$

where $\vec{p}_{(\Upsilon(4S))}$ is the momentum of the $\Upsilon(4S)$, $\vec{p}_{tag}$ the momentum of the B_{tag} and $\vec{p}_K$ the momentum of the $K^\pm$.

$M_{\tau^+\tau^-}$: The reconstructed invariant mass of the τ pair.

M_{bc}^{tag} : The beam constrained mass of the B_{tag} candidate.

$\theta_{\tau^-}^{hel}$: The pseudo helicity angle of the τ^- .

$\sigma(d_{B_{tag}})$: The significance of the distance to the B_{tag} candidate, derived from the error of the vertex fit.

χ^2 : χ^2 value of the vertex fit of the candidate.

d_{IP} : Distance of the candidate to the interaction point.

Q : Defined as the reconstructed mass of the B candidate subtracted by the reconstructed mass of the children: $Q \equiv M_B - M_{K^+} - M_{\tau^+} - M_{\tau^-}$.

The input variables are displayed in figs. 10.2 and 10.3.

In addition to the BDT the performance of a NeuroBayes neural network and a Fisher discriminant (LDA) is tested for comparison, which have also both proven to deliver reliable results in high energy physics. Figure 10.4 shows that the BDT delivers the best results, as it serves higher efficiencies for all chosen purity levels.

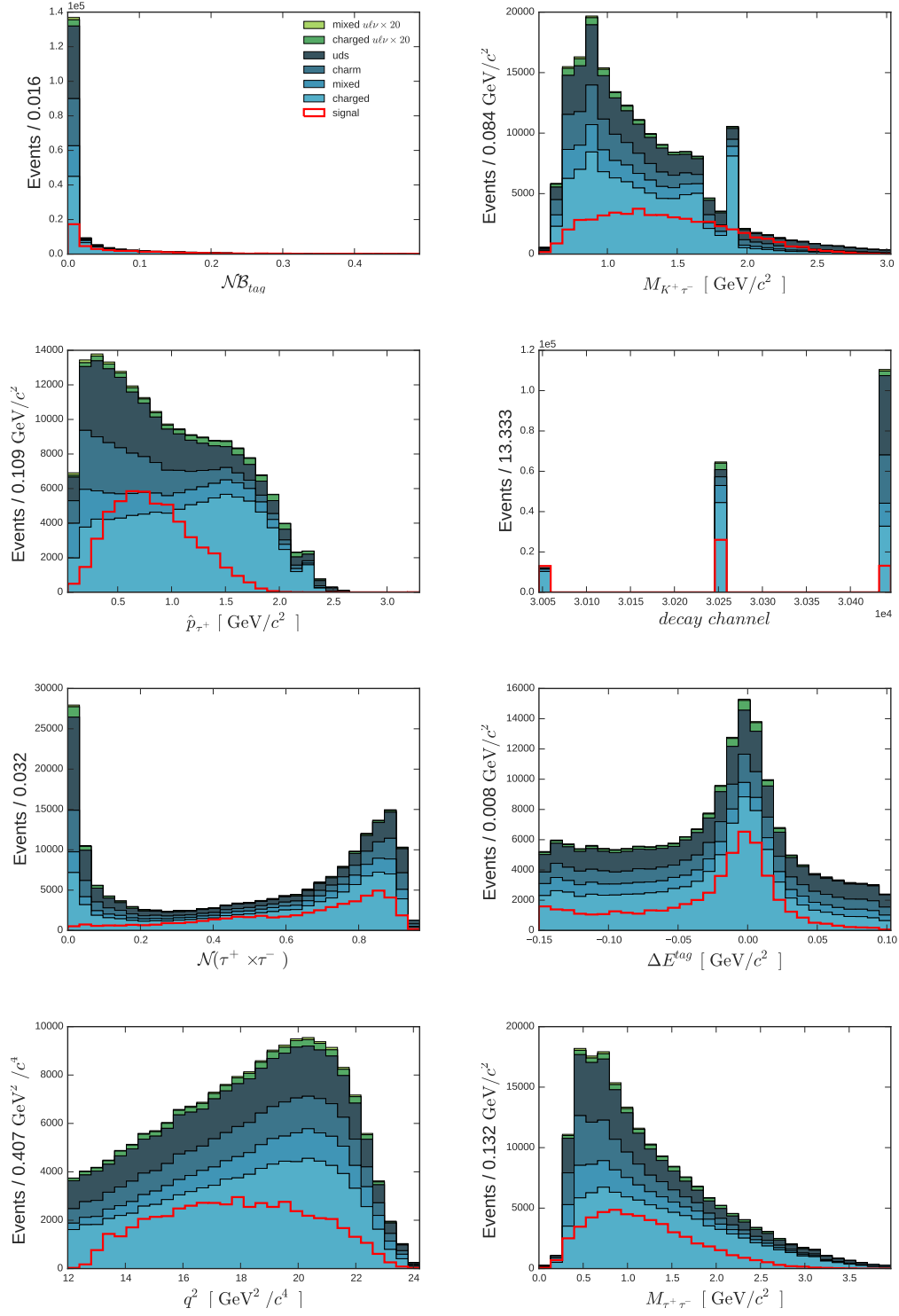


Figure 10.2: Input variables for the boosted decision tree. The signal distribution is shown with a red line and the background is displayed as a stacked plot for several sources (blue tones for generic MC and green for rare $u\bar{u}\nu$ MC). The description of the variables can be found in section 10.3.

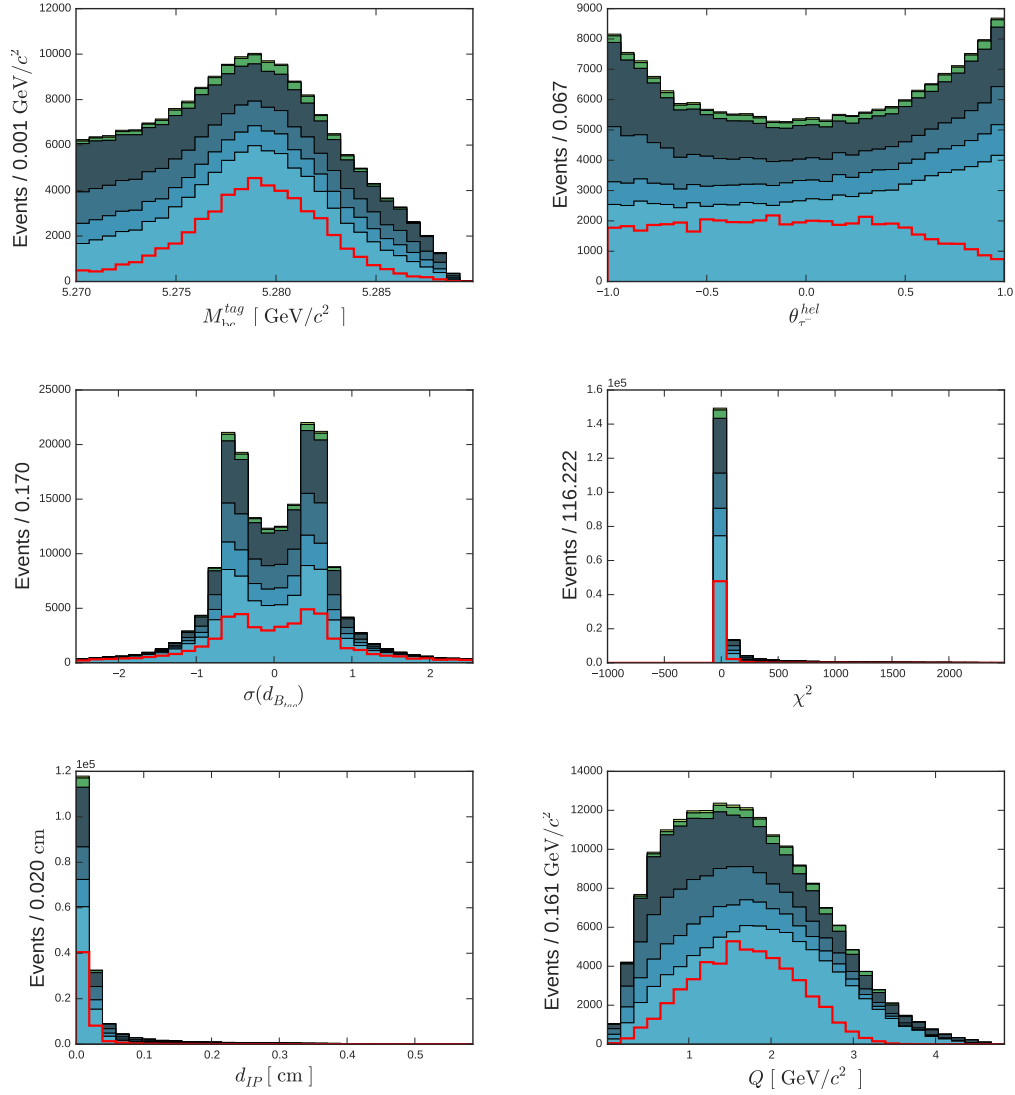


Figure 10.3: Input variables for the boosted decision tree. Continuation of fig. 10.2.

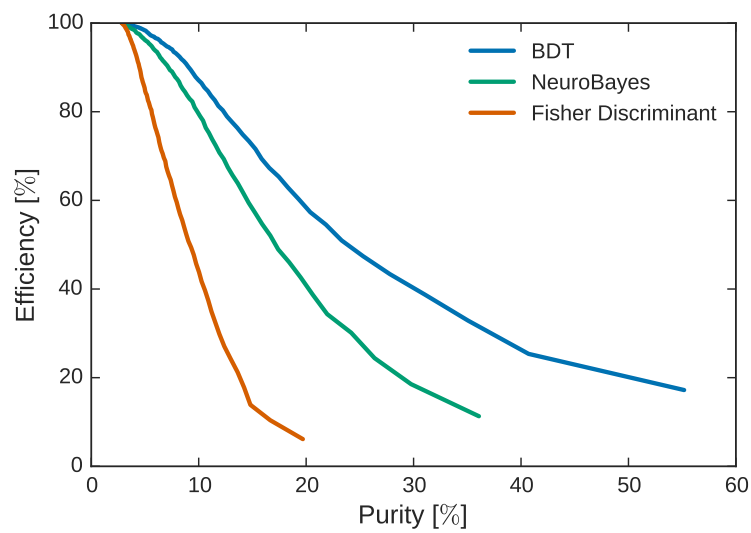


Figure 10.4: Performance comparison of different classification algorithms on an independent test-dataset.

11. Limit Estimation

The upper limit on $B^+ \rightarrow K^+ \tau^+ \tau^-$ is calculated with a counting experiment in a signal window in E_{ECL} under the assumption that no signal is observed. The cut parameters c_E on E_{ECL} and c_{BDT} on the output of the BDT are optimized in order to obtain the most stringent upper limit. Both cut parameters influence the amount of expected background events in the signal region and the reconstruction efficiency ϵ_s for signal candidates.

11.1 Branching Ratio

The branching ratio for a decay channel d is calculated by:

$$\mathcal{B}(d) = \frac{N_d^{true}}{\mathcal{N}}, \quad (11.1)$$

where $\mathcal{N}$ is the total number of decays and N_d^{true} is the true amount of signal events from the desired decay channel. Since this number is unknown in analyses which are subjected to acceptance and efficiency effects, it can be estimated by correcting the number of observed events N_d^{obs} by the reconstruction and acceptance efficiency ϵ_s :

$$N_d^{true} \simeq \frac{N_d^{obs}}{\epsilon_s}. \quad (11.2)$$

The observed $B^\pm$ candidates in this analysis corresponds to

$$\mathcal{N} = 2 \cdot f_{B^+} \cdot N_{B\bar{B}}, \quad (11.3)$$

where $N_{B\bar{B}}$ is the number of recorded B meson pairs in the Belle dataset and $f_{B^+} = 51.4 \pm 0.6\%$ [4] is the fraction of $B^+ B^-$ pairs. An upper limit is calculated for the maximum number of observed signal candidates N^{UL} with a certain confidence,

assuming the absence of a signal. Combined, the upper limit on the branching fraction can be derived by

$$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-) < \frac{N^{UL}}{\epsilon_s \cdot 2 \cdot f_{B^+} \cdot N_{B\bar{B}}}. \quad (11.4)$$

11.2 Limit Calculation

Upper limits are calculated in dependency of a certain degree of confidence. In case of absent backgrounds they can be derived from the equation:

$$1 - \alpha = \sum_{k=0}^n P(k; \mu), \quad (11.5)$$

where $100(1 - \alpha)\%$ is the chosen confidence level, μ is the unknown true expectation value and P the underlying probability density function. The ROOT [64] implementation TRolke is used, suggested by W. Rolke [65]. It uses as Profile Likelihood method with a Gaussian background model and Poisson signal distribution:

$$L(\mu, b, \sigma | x, y) = \frac{(\mu + b)^x}{x!} \exp(-(\mu + b)) \times \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - b)^2}{2\sigma^2}\right), \quad (11.6)$$

where b is the expected background with uncertainty σ , x and y are the measured values for signal and background. The uncertainties on x and y are treated as nuisance parameters and are modeled with a Gaussian probability density function.

11.3 Limit Optimization

The cut parameters c_E on E_{ECL} and c_{BDT} on the output of the BDT are optimized in order to minimize $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-)^{UL}$. For each pairs of cuts c_{BDT} and c_E , the corresponding signal efficiency ϵ_s and the expected amount of background events in the signal region is calculated. The systematic uncertainties, which are detailed in chapter 12, are also evaluated for each cut, as they are partly dependent on c_{BDT} . With these values N^{UL} and consequently the estimation of the upper limit on the branching fraction can be derived. The best upper limit is found at $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-) < 3.96 \times 10^{-4}$ at 95% confidence level (C.L.) for $c_{BDT} = 0.553$ and $c_E = 0.206$ GeV, including the uncertainties on N^{UL} and ϵ_s . The expected number of background events in the signal window is 8.83 and the signal efficiency is $\epsilon_s = 2.709 \times 10^{-5}$. The distributions for E_{ECL} and the boosted decision tree output are displayed in fig. 11.1. The result for the limit optimization is shown in fig. 11.2.

Additionally, the limits are calculated with the Bayesian Markov-Chain based tool, Theta [66], and a scan over the upper limit with different amount of observed events while expecting 8.83 events is performed. The scan for the observed upper limit for different yields on data is depicted in fig. 11.3.

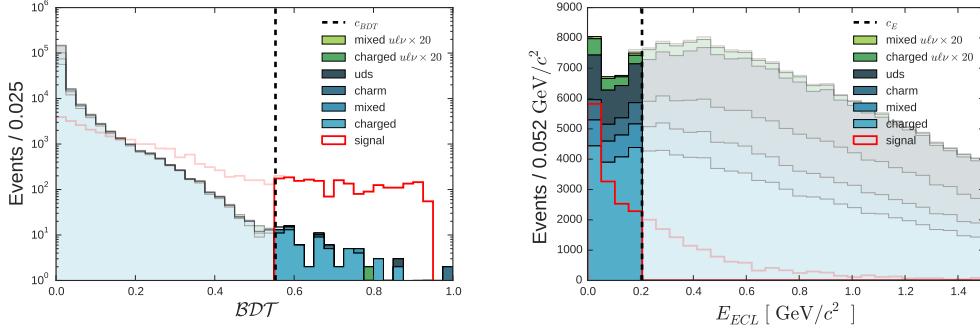


Figure 11.1: The output of the BDT and the variable E_{ECL} prior to the cut for the best upper limit.

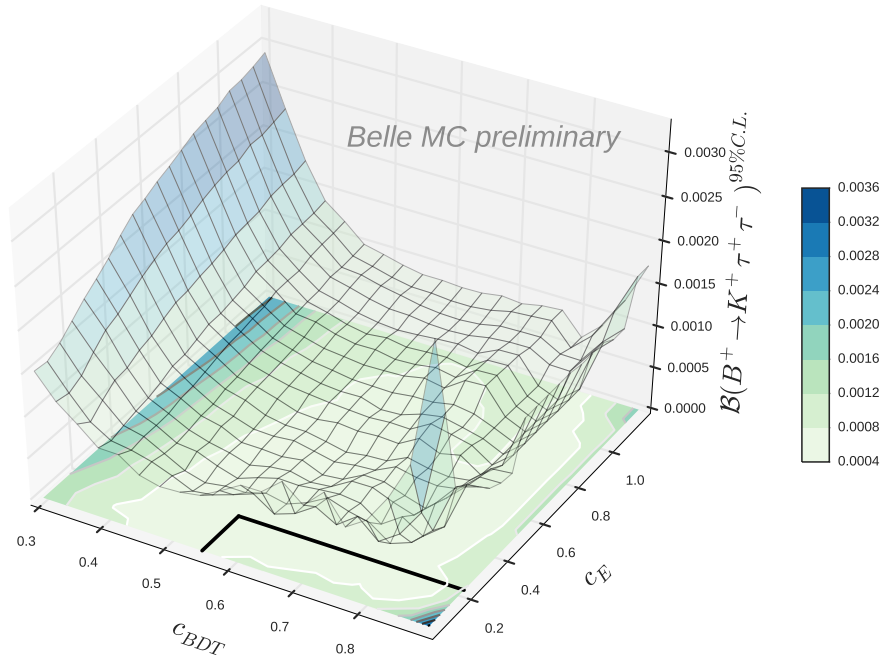


Figure 11.2: Calculation of the lowest upper limit on $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-)^{UL}$ at 95% C.L. in dependence on the cut c_E on E_{ECL} and c_{BDT} on the output of the BDT.

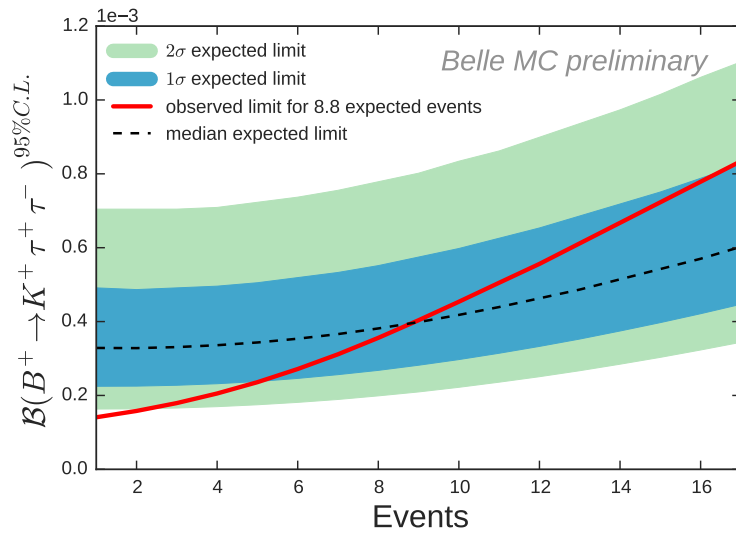


Figure 11.3: Observed and expected limits for different numbers of observed events, calculated with Bayesian methods [66] at 95% C.L.

12. Systematic Uncertainties

For this analysis systematic uncertainties which affect the branching fraction measurement are covered. The upper limit is derived from

$$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-) < \frac{N^{UL}}{\epsilon_s \cdot 2 \cdot f_{B^+} \cdot N_{B\bar{B}}} = \frac{\tilde{N}^{UL}(N_{exp}, \epsilon_s, \sigma_{\epsilon_s})}{\mathcal{N}}. \quad (12.1)$$

This implies that two main quantities can have errors: The number of observed events, $\mathcal{N} = 2 \cdot f_{B^+} \cdot N_{B\bar{B}}$ and the expected upper limit for the observed signal candidates N^{UL} . The latter includes all systematic uncertainties from the signal efficiency ϵ_s and the expected amount of background events N_{exp} . They can be treated as nuisance parameter in the calculation of the upper limit and lead to an adjusted N_{sys}^{UL} including uncertainties. The errors are dominated by the statistical error on the number of expected background events. For this reason conservatively large estimates on the systematic errors can be assumed without large effects on the total upper limit.

12.1 Data MC Comparison

12.1.1 Tag Side Correction

Large differences between yields of the data and MC samples are observed in the analysis, depicted in fig. 12.1(a). The reason for this discrepancy arises from systematic differences of the reconstruction efficiency from the NeuroBayes full reconstruction. This effect has been studied on a variety of control channels on data and can be corrected by introducing correction-weights for the B_{tag} candidate based on the neural network output $\mathcal{NB}(B_{tag})$ and the decay channel of the tag-side B . The distribution of the weights is displayed in fig. 12.1(c) which has a mean value of 0.81, indicating that the reconstruction efficiency of the tag-side is overestimated by the framework. When weighting each event in the MC class by the corresponding weight, almost perfect alignment and equal total yields for the M_{bc} distribution of the B_{tag} candidate can be observed in fig. 12.1(b). The systematic error of this procedure for the efficiency of the B_{tag} is estimated with 4.6%.

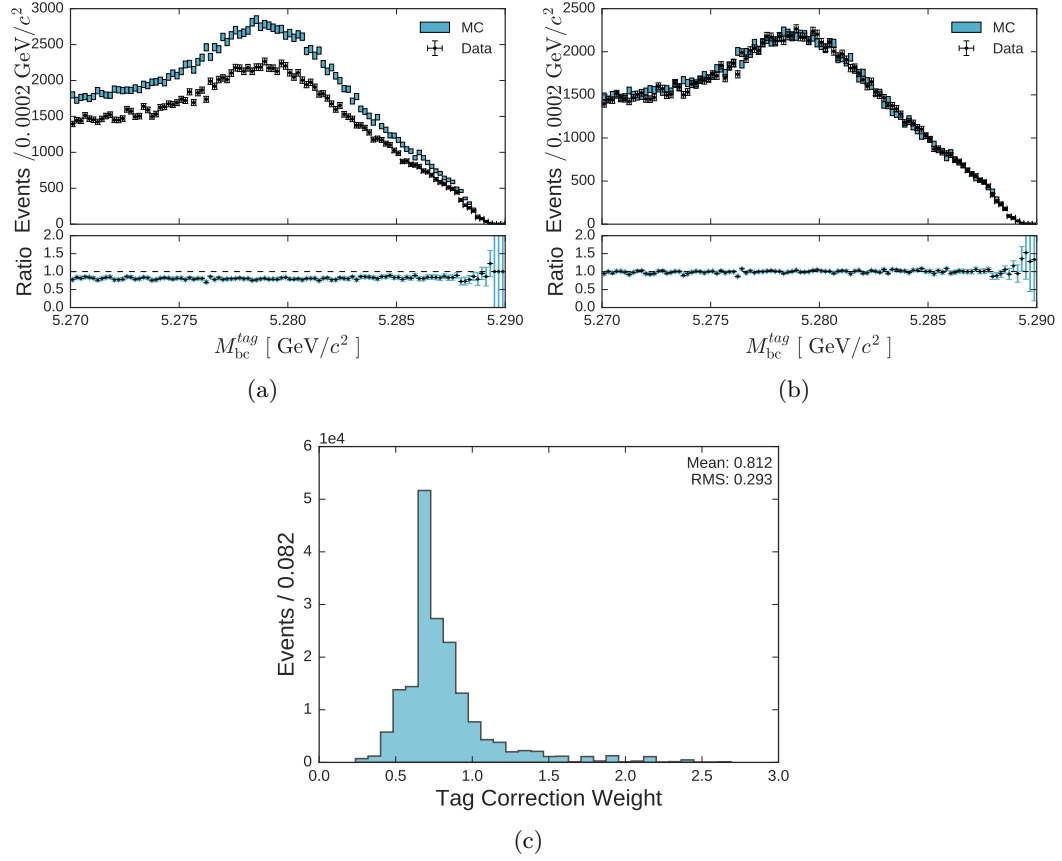


Figure 12.1: Impact of the Monte Carlo efficiency correction for the B_{tag} candidate. Displayed is M_{bc} of the B_{tag} with (b) and without (a) weights (c).

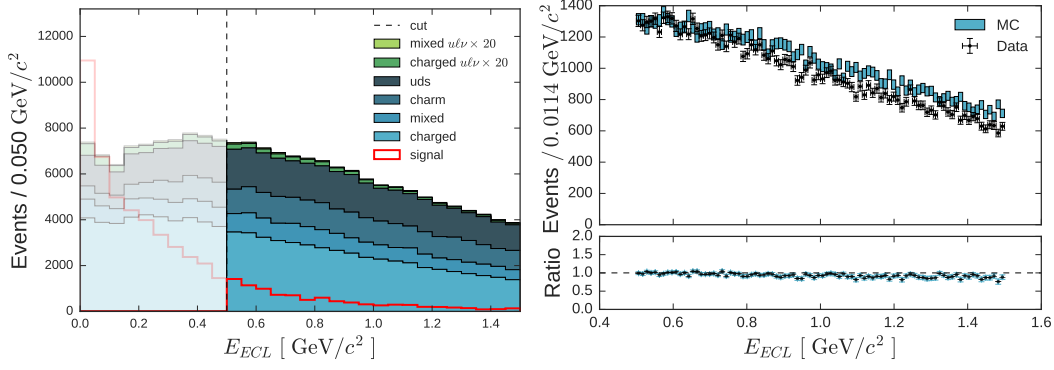


Figure 12.2: E_{ECL} sideband for all different types of MC (left) and the comparison with data (right) after the Monte Carlo efficiency correction for the B_{tag} candidate is applied.

12.1.2 E_{ECL} Sideband

A comparison between generic MC and data is performed on a sideband in E_{ECL} , where $E_{ECL} > 0.5$ GeV/c². The sideband region is displayed in fig. 12.2. A systematic discrepancy in E_{ECL} (fig. 12.2(b)) is observed, which has to be analyzed on the full range for determining systematic effects.

The agreement of data and MC for all variables which are suited for background suppression are compared in appendix B.2. In order to assign an uncertainty to the agreement between data and Monte Carlo, the mean of the classifier output for data and MC on the sideband of E_{ECL} is compared. The observed difference is $0.067 \pm 5.734\%$, dominated by statistical fluctuations. The BDT cut c_{BDT} is varied within this error and systematic uncertainties are applied for observed differences in the limit. For the signal efficiency 6% and for the expected background 10% uncertainty is applied (corresponding to 2 events).

12.1.3 E_{ECL} Distribution

Before looking into data in the signal region of E_{ECL} region, one can only extrapolate the observed difference in the E_{ECL} distribution into the signal region, fig. 12.2, and estimate a yield difference. A preliminary uncertainty of 10% is applied to the candidates in the signal region. When opening the box in the signal region, the distribution with increasing cut on the classifier can be monitored. It is possible, that some backgrounds are not modeled correctly on MC and soft cuts on the BDT output can make the distribution more similar.

12.1.4 Control Channel on Data

The distribution of E_{ECL} can also be studied with the control process $B^+ \rightarrow K^+ \tau^+ \tau^- (K_S)$, where the K_S daughters are removed from the event. The con-

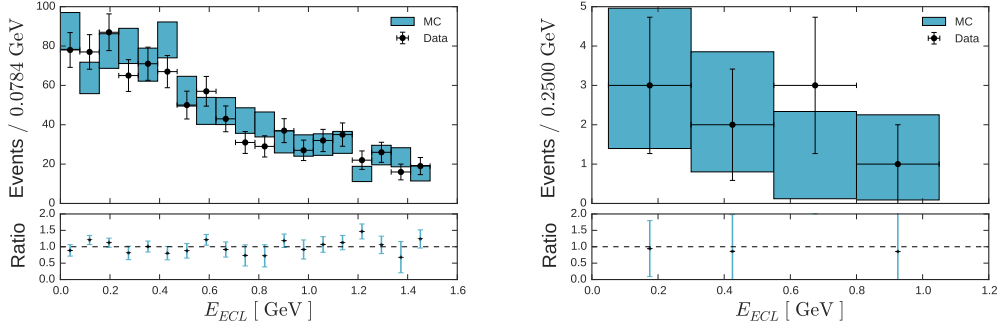


Figure 12.3: E_{ECL} distribution for $B^+ \rightarrow K^+ \tau^+ \tau^- (K_S)$, where the K_S daughters are removed from the event with the final cut $c_{BDT} = 0.553$ (right) and $c_{BDT} > 0.1$ (left) for comparison.

tol process is reconstructed on the full range of E_{ECL} and the BDT output is applied to the events. The result is shown in fig. 12.3.

With the final selection of $c_{BDT} = 0.553$ in total 9 events are observed while expecting 9.11 events from simulations. The distributions are displayed in fig. 12.3(b). Excellent agreement between data and MC can be observed. This result leads to the conclusion that the used methods work as expected. Trained expertise from MC can be successfully applied to real data.

12.2 Other Uncertainties

12.2.1 Number of $B^+ B^-$ Pairs

The number of observed candidates can be estimated by $\mathcal{N} = 2 \cdot f_{B^+} \cdot N_{B\bar{B}}$, where $N_{B\bar{B}} = 771 \cdot 10^6 \pm 1.4\%$ is the B meson pairs in the Belle dataset and $f_{B^+} = 0.516 \pm 0.006$ [4] is the fraction for the decay $\Upsilon(4S) \rightarrow B^+ B^-$. Combining both uncertainties, 1.8% error for the normalization of the signal is assigned.

12.2.2 Particle Identification

For the charged kaon track, a particle identification value of $\mathcal{R}_K > 0.6$ is required. A conservative systematic error of 2.2% is applied for this selection. For the remaining two charged tracks the mass hypothesis (e, μ or $\pi^\pm$) based on the probability output of trained neural networks is applied. Randomly switching the hypothesizes to the second best option, changes the mean of the output of the main classifier by 3.6%. This results in 5.09% uncertainty for the tau daughters. Combined with the kaon identification 5.54% uncertainty for the total particle identification is taken into account.

12.2.3 Tracking

The Belle standard systematic uncertainty of 0.35% error for each charged track is applied, resulting in 1.05% total tracking uncertainty.

12.2.4 Generator

Two sets of signal MC are used with different decay model for the generator with form-factor calculations by Ball et al.[63] and Ali et al. [41]. The classifier is trained on one dataset and the upper limit is estimated on the other one. When switching the datasets for training and limit estimation, a difference of $1.715 \pm 6.949\%$ is observed. For the generator model a conservative uncertainty of 8.66% is applied.

12.2.5 Statistical Uncertainties from Simulation

The dominant source of systematics arise from statistical uncertainties for the estimation of the signal efficiency and the number of expected background events. The signal efficiency is calculated by

$$\epsilon_s = \frac{N_{sig}}{N_{gen}}, \quad (12.2)$$

where N_{sig} is the number of correctly selected candidates and N_{gen} is the number of generated signal events. The statistical error on the estimation of the signal efficiency can be calculated with a binomial error

$$\sigma_{\epsilon_s} = \frac{\sqrt{\frac{N_{sig}(N_{gen}-N_{sig})}{N_{gen}}}}{N_{gen}}. \quad (12.3)$$

The difference to Poisson errors is however negligible in the light of small signal efficiencies. For ϵ_s , 6.13% statistical uncertainty is applied and 33.7% for N_{exp} .

12.3 Summary

All sources of systematic errors are summarized in table 12.1. The errors are treated as nuisance parameters in the calculation of the upper limit.

Table 12.1: List of covered systematic uncertainties.

	Source	Error [%]
N_{exp}	Tracking	1.05
	PID	5.54
	B_{tag} correction	4.60
	E_{ECL} distribution	10.00
	Data - Monte Carlo discrepancy	10.00
	Statistical error	33.64
	Total	37.21
ϵ_s	$N_{B^+B^-}$	1.80
	Tracking	1.05
	PID	5.54
	B_{tag} correction	4.60
	Data - Monte Carlo discrepancy	6.00
	Generator	8.66
	Statistical error	6.18
	Total	14.33

13. Results

The estimated upper limit $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-) < 3.97 \times 10^{-4}$ is obtained at the 95% confidence level using Belle Monte Carlo datasets and Frequentists methods. With a Bayesian Markov Chain Monte Carlo calculator a limit of $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-) < 4.61 \times 10^{-4}$ is observed. All results are summarized in table 13.1 and the expected distribution in E_{ECL} is presented in fig. 13.1.

Table 13.1: Determined parameters for the best upper limit on MC including systematic uncertainties.

Description	Value
Cut on the BDT	$c_{BDT} = 0.553$
Cut on E_{ECL}	$c_E = 0.206 \text{ GeV}$
Efficiency for signal events	$\epsilon_s = 2.709 \times 10^{-5}$
Expected number of background events	8.84
Upper limit for signal candidates	$N^{UL95\%} = 8$
Upper limit at 90% C.L.	$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-) < 3.17 \times 10^{-4}$
Upper limit at 95% C.L.	$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-) < 3.96 \times 10^{-4}$

On the control channel $B^+ \rightarrow K^+ \tau^+ \tau^- (K_S)$ it is demonstrated on data that the reconstruction and background suppression methods work as expected.

It could be shown, that this analysis is capable of improving the current upper limit for $B^+ \rightarrow K^+ \tau^+ \tau^-$ by one order of magnitude compared the previous analysis from BaBar. In addition, New Physics scenarios, in particular Minimal Lepton Flavor Violation models, are in reach of the upper limit prediction [3].

The final result on data can be obtained immediately, however in order to comply with the Belle publication guidelines an internal review process has to be performed first, which was not yet granted in favor of the angular analysis.

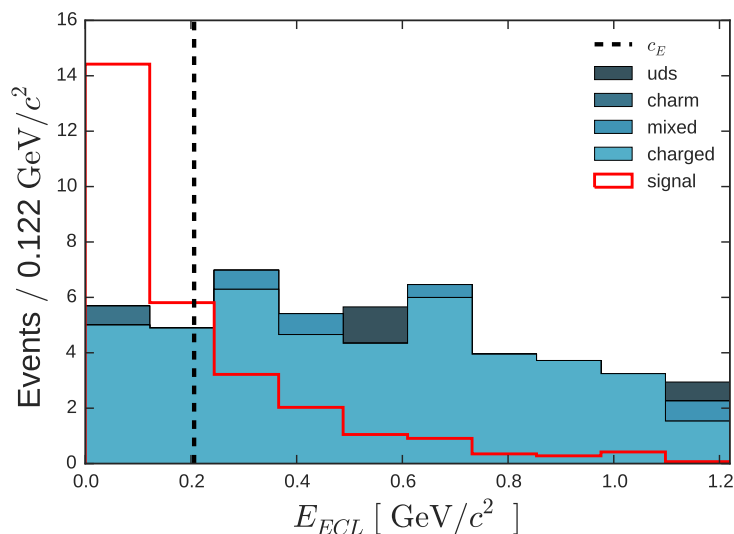


Figure 13.1: Expected amount of background for the final result. The signal distribution has arbitrary normalization.

Prospects for Belle II

In a brief analysis the sensitivity for Belle II is estimated using the same methods.

The goal for Belle II is an increase in statistics by the factor 50, see [67] for more details. Together with a higher instantaneous luminosity also more beam background is expected. To compensate for this, a new tracking and particle identification system is installed into the Belle II detector. A major advantage is also provided with the new software framework for Belle II, which will lead to better reconstruction efficiency and background suppression. The Full Event Interpretation framework [68] will supersede the full reconstruction technique and is expected to have more than twice the reconstruction efficiency compared to Belle.

Assuming a linear scaling of the classification methods and systematic errors, which is reasonably conservative, the projected Belle II result is

$$\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-)^{BelleII} < 2 \times 10^{-5}, \quad (13.1)$$

based on the methods used in this analysis.

With this prediction Belle II will be able to challenge predictions made in the framework of MLFV models with the measurement of $B^+ \rightarrow K^+ \tau^+ \tau^-$.

Part IV

Conclusion

14. Conclusion

In this thesis I present a comprehensive study of the flavor changing neutral current process of $b \rightarrow s \ell^+ \ell^-$ in B meson decays with accompanying kaons in two separate analyses. All three lepton modes, $e^+ e^-$, $\mu^+ \mu^-$ and $\tau^+ \tau^-$ are investigated to search for evidence of physics beyond the Standard Model. The measurements are performed using the full Belle data sample of 772×10^6 $B\bar{B}$ pairs, recorded at the $\Upsilon(4S)$ resonance.

The first analysis in this thesis covers the muon and electron modes in the decay of $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$. The reconstruction efficiency is sufficient to gather enough signal candidates to perform a fit to the differential decay rate in three dimensions, which is sensitive to effects from new physics. Reconstructed signal yields of both channels exceed previous B -factory results of Belle and BaBar measurements, enabling a full angular analysis in this decay with the Belle data for the first time. To maximize signal efficiency and purity, neural networks are developed sequentially from the bottom to the top of the decay chain, transferring each time the output probability to the subsequent step such that most effective selection cuts are applied in the last stage based on all information combined. In total 117.6 ± 12.4 signal candidates for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ and 69.4 ± 12.0 signal events for $B^0 \rightarrow K^*(892)^0 e^+ e^-$ are reconstructed. With the reconstruction efficiency and signal yields the branching ratios of both modes are extracted and compared to previous measurements in order to validate the reconstruction procedure. The measurements are in agreement with values from PDG [4]. With the combined data of both channels a full angular analysis in three dimensions in five bins of the di-lepton invariant mass squared, q^2 , is performed. A data transformation technique is applied to reduce the dimension of the angular decay rate from eight to three dimensions. By this means the fit is independently sensitive to observables P'_4 , P'_5 , P'_6 and P'_8 , which are optimized regarding uncertainties from form-factors. Altogether 20 independent measurements are performed extracting $P'_{4,5,6}$ or P'_8 , the K^* longitudinal polarization F_L and the transverse polarization asymmetry A_T . The results in the region $q^2 < 8 \text{ GeV}^2/c^4$

are compared with Standard Model predictions and overall agreement is observed. One measurement is found to deviate by $\sim 2.1\sigma$ from the predicted value in the same direction and in the same q^2 region where the LHCb collaboration reported the so called P'_5 anomaly [1, 2]. From a global perspective the deviations in P'_5 can be explained by unexpectedly large hadronic contributions from the charmonium resonance which are not accounted for in the theory prediction or by a yet unknown particle [2]. Further investigations from theoretical and experimental side have to be performed in order to unveil the nature of the anomaly in P'_5 . With this result a second independent measurement shows that Standard Model predictions might not be valid in this observable and new physics could be around the corner.

The second analysis in this thesis is dedicated to the τ mode of $b \rightarrow s\ell^+\ell^-$ with the search for $B^+ \rightarrow K^+\tau^+\tau^-$. This mode is particularly interesting as new physics could couple to the high mass of the tau. However, due to several neutrinos being present in the final state, it is difficult to reconstruct. For this decay, the full reconstruction technique is used, which is unique for e^+e^- accelerators and makes it possible to find signatures of undetected neutrinos in the decay. In this analysis it could be demonstrated using simulated events and control channels on the Belle dataset, that this analysis is able to improve the upper limit by more than one order of magnitude compared to the current value to

$$\mathcal{B}^{Projected}(B^+ \rightarrow K^+\tau^+\tau^-) < 3.17 \times 10^{-4} \quad (14.1)$$

at 90% confidence level including systematic uncertainties determined by an extensive study of all the sources. On the control channel, $B^+ \rightarrow K^+\tau^+\tau^-(K_S)$, it is demonstrated on data that the reconstruction and background suppression methods work as expected and that they deliver consistent results on both data and Monte Carlo. With this measurement Belle will be able to set the lowest experimental limits to this branching ratio. Furthermore, the limit will come close to the prediction of models developed in the context of Minimal Lepton Flavor Violation, which will help to constrain these models [3].

The next major step for the analyses described in this thesis will be the upcoming Belle II experiment with improved vertexing, reconstruction and particle identification system, expected to collect 50 ab^{-1} of integrated luminosity at the SuperKEKB collider. Both analyses will benefit from the increased statistics and may reveal more hints for new physics. In a sensitivity study I demonstrated that with methods described in this thesis it will be possible to set tight constraints for Minimal Lepton Flavor Violating models in the decay of $B^+ \rightarrow K^+\tau^+\tau^-$ with the data of Belle II.

Bibliography

- [1] R. Aaij et al. “Measurement of Form-Factor-Independent Observables in the Decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ ”. In: *Phys. Rev. Lett.* 111.19 (Nov. 2013). DOI: 10.1103/physrevlett.111.191801. URL: <http://dx.doi.org/10.1103/PhysRevLett.111.191801>.
- [2] Roel Aaij et al. “Angular analysis of the $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ decay”. In: (2015). arXiv: 1512.04442 [hep-ex].
- [3] Rodrigo Alonso et al. “Lepton universality violation and lepton flavor conservation in B -meson decays”. In: *JHEP* 10 (2015), p. 184. DOI: 10.1007/JHEP10(2015)184. arXiv: 1505.05164 [hep-ph].
- [4] K Nakamura et al. “Review of Particle Physics”. In: *Journal of Physics G: Nuclear and Particle Physics* 37.7A (2010), p. 075021. URL: <http://stacks.iop.org/0954-3899/37/i=7A/a=075021>.
- [5] R. Aaij et al. “Test of Lepton Universality Using $B^+ \rightarrow K^+ \ell^+ \ell^-$ Decays”. In: *Phys. Rev. Lett.* 113.15 (Oct. 2014). DOI: 10.1103/physrevlett.113.151601. URL: <http://dx.doi.org/10.1103/PhysRevLett.113.151601>.
- [6] Martin Bauer et al. “One Leptoquark to Rule Them All: A Minimal Explanation for $R_{D^{(*)}}$, R_K and $(g-2)_\mu$ ”. In: (2015). arXiv: 1511.01900 [hep-ph].
- [7] A. Bozek et al. “Observation of $B^+ \rightarrow \bar{D}^{*0} \tau^+ \nu_\tau$ and evidence for $B^+ \rightarrow \bar{D}^0 \tau^+ \nu_\tau$ at Belle”. In: *Phys. Rev. D* 82 (7 Oct. 2010), p. 072005. DOI: 10.1103/PhysRevD.82.072005. URL: <http://link.aps.org/doi/10.1103/PhysRevD.82.072005>.
- [8] J. P. Lees et al. “Evidence for an excess of $\bar{B} \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ decays”. In: *Phys. Rev. Lett.* 109 (2012), p. 101802. DOI: 10.1103/PhysRevLett.109.101802. arXiv: 1205.5442 [hep-ex].
- [9] Roel Aaij et al. “Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \tau^- \bar{\nu}_\tau) / \mathcal{B}(\bar{B}^0 \rightarrow D^{*+} \mu^- \bar{\nu}_\mu)$ ”. In: *Phys. Rev. Lett.* 115.11 (2015). [Addendum: *Phys. Rev. Lett.* 115,no.15,159901(2015)], p. 111803. DOI: 10.1103/PhysRevLett.115.159901, 10.1103/PhysRevLett.115.111803. arXiv: 1506.08614 [hep-ex].
- [10] Martin Bauer et al. “Flavor Anomalies, the Diphoton Excess and a Dark Matter Candidate”. In: (2015). arXiv: 1512.06828 [hep-ph].

- [11] S. L. Glashow. “Partial Symmetries of Weak Interactions”. In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.
- [12] Steven Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.
- [13] Peter W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.
- [14] Nicola Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10 (1963). [648(1963)], pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.
- [15] Makoto Kobayashi et al. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.
- [16] J. Charles et al. “Current status of the Standard Model CKM fit and constraints on $\Delta F = 2$ New Physics”. In: *Phys. Rev. D* 91.7 (2015), p. 073007. DOI: 10.1103/PhysRevD.91.073007. arXiv: 1501.05013 [hep-ph].
- [17] R. Aaij et al. “Differential branching fraction and angular analysis of the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ ”. In: *JHEP* 08 (2013), p. 131. DOI: 10.1007/JHEP08(2013)131. arXiv: 1304.6325.
- [18] Wolfgang Altmannshofer et al. “Symmetries and Asymmetries of $B \rightarrow K^{*} \mu^+ \mu^-$ Decays in the Standard Model and Beyond”. In: *JHEP* 01 (2009), p. 019. DOI: 10.1088/1126-6708/2009/01/019. arXiv: 0811.1214 [hep-ph].
- [19] Sebastien Descotes-Genon et al. “Optimizing the basis of $B \rightarrow K^{*} \ell \ell$ observables in the full kinematic range”. In: *JHEP* 05 (2013), p. 137. DOI: 10.1007/JHEP05(2013)137. arXiv: 1303.5794 [hep-ph].
- [20] M De Cian. “Track Reconstruction Efficiency and Analysis of $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ at the LHCb Experiment”. PhD thesis. University of Zurich, Mar. 2013. URL: <http://dx.doi.org/10.5167/uzh-85745>.
- [21] J.-T. Wei et al. “Measurement of the Differential Branching Fraction and Forward-Backward Asymmetry for $B \rightarrow K^{(*)} l^+ l^-$ ”. In: *Phys. Rev. Lett.* 103 (17 Oct. 2009), p. 171801. DOI: 10.1103/PhysRevLett.103.171801. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.103.171801>.
- [22] B. Aubert et al. “Angular distributions in the decay $B \rightarrow K^{*} l^+ l^-$ ”. In: *Phys. Rev. D* 79 (3 Feb. 2009), p. 031102. DOI: 10.1103/PhysRevD.79.031102. URL: <http://link.aps.org/doi/10.1103/PhysRevD.79.031102>.
- [23] T. Aaltonen et al. “Measurements of the Angular Distributions in the Decays $B \rightarrow K^{(*)} \mu^+ \mu^-$ at CDF”. In: *Phys. Rev. Lett.* 108 (2012), p. 081807. DOI: 10.1103/PhysRevLett.108.081807. arXiv: 1108.0695 [hep-ex].
- [24] M. Kowalski et al. “Improved Cosmological Constraints from New, Old and Combined Supernova Datasets”. In: *Astrophys. J.* 686 (2008), pp. 749–778. DOI: 10.1086/589937. arXiv: 0804.4142 [astro-ph].

-
- [25] A. Bartl et al. “General flavor blind MSSM and CP violation”. In: *Phys. Rev. D* 64 (2001), p. 076009. DOI: 10.1103/PhysRevD.64.076009. arXiv: hep-ph/0103324 [hep-ph].
- [26] S. Kurokawa et al. “Overview of the KEKB accelerators,” in: *Nucl. Instrum. Methods Phys. Res., Sect. A* **499**, 1 (2003) 1 (2003).
- [27] 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 (2002). Detectors for Asymmetric B-factories, pp. 117–232. ISSN: 0168-9002. DOI: [http://dx.doi.org/10.1016/S0168-9002\(01\)02013-7](http://dx.doi.org/10.1016/S0168-9002(01)02013-7). URL: <http://www.sciencedirect.com/science/article/pii/S0168900201020137>.
- [28] Jolanta Brodzicka et al. “Physics achievements from the Belle experiment”. In: *Progress of Theoretical and Experimental Physics* 2012.1 (2012). DOI: 10.1093/ptep/pts072. eprint: <http://ptep.oxfordjournals.org/content/2012/1/04D001.full.pdf+html>. URL: <http://ptep.oxfordjournals.org/content/2012/1/04D001.abstract>.
- [29] David J. Lange. “The EvtGen particle decay simulation package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1–2 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: [http://dx.doi.org/10.1016/S0168-9002\(01\)00089-4](http://dx.doi.org/10.1016/S0168-9002(01)00089-4). URL: <http://www.sciencedirect.com/science/article/pii/S0168900201000894>.
- [30] Torbjorn Sjostrand et al. “High-energy physics event generation with PYTHIA 6.1”. In: *Comput. Phys. Commun.* 135 (2001), pp. 238–259. DOI: 10.1016/S0010-4655(00)00236-8. arXiv: hep-ph/0010017 [hep-ph].
- [31] Elisabetta Barberio et al. “PHOTOS: A Universal Monte Carlo for QED radiative corrections in decays”. In: *Comput. Phys. Commun.* 66 (1991), pp. 115–128. DOI: 10.1016/0010-4655(91)90012-A.
- [32] R Brun et al. *GEANT 3: user’s guide Geant 3.10, Geant 3.11; rev. version*. Geneva: CERN, 1987. URL: <https://cds.cern.ch/record/1119728>.
- [33] Torbjorn Sjostrand. “High-energy physics event generation with PYTHIA 5.7 and JETSET 7.4”. In: *Comput. Phys. Commun.* 82 (1994), pp. 74–90. DOI: 10.1016/0010-4655(94)90132-5.
- [34] M. Feindt et al. “An algorithm for quantifying dependence in multivariate data sets”. In: *Nuclear Instruments and Methods in Physics Research A* 698 (Jan. 2013), pp. 84–89. DOI: 10.1016/j.nima.2012.09.043. arXiv: 1207.0981 [physics.data-an].
- [35] V. Blobel. *Statistische und Numerische Methoden der Datenanalyse*. 1998.
- [36] M. Feindt. “A Neural Bayesian Estimator for Conditional Probability Densities”. In: *ArXiv Physics e-prints* (Feb. 2004). eprint: physics/0402093.

- [37] Jerome H. Friedman. “Stochastic gradient boosting”. In: *Computational Statistics & Data Analysis* 38.4 (Feb. 2002), pp. 367–378. DOI: 10.1016/S0167-9473(01)00065-2. URL: [http://dx.doi.org/10.1016/S0167-9473\(01\)00065-2](http://dx.doi.org/10.1016/S0167-9473(01)00065-2).
- [38] C.-L. Hsu et al. “Search for B^0 decays to invisible final states at Belle”. In: *Physical Review D* 86.3 (Aug. 2012). DOI: 10.1103/PhysRevD.86.032002. URL: <http://dx.doi.org/10.1103/PhysRevD.86.032002>.
- [39] 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. DOI: 10.1016/j.nima.2011.06.008. URL: <http://dx.doi.org/10.1016/j.nima.2011.06.008>.
- [40] M. Feindt. “A Neural Bayesian Estimator for Conditional Probability Densities”. In: *ArXiv Physics e-prints* (Feb. 2004). eprint: [physics/0402093](https://arxiv.org/abs/physics/0402093).
- [41] Ahmed Ali et al. “Helicity analysis of the decays $B \rightarrow K^* \ell^+ \ell^-$ and $B \rightarrow \rho \ell \nu_\ell$ in the large energy effective theory”. In: *Eur.Phys.J.* C25 (2002), pp. 583–601. DOI: 10.1140/epjc/s2002-01040-3. arXiv: [hep-ph/0205254](https://arxiv.org/abs/hep-ph/0205254) [hep-ph].
- [42] Michael Feindt et al. “A Hierarchical NeuroBayes-based Algorithm for Full Reconstruction of B Mesons at B Factories”. In: *Nucl.Instrum.Meth.* A654 (2011), pp. 432–440. DOI: 10.1016/j.nima.2011.06.008. arXiv: 1102.3876 [hep-ex].
- [43] Geoffrey C. Fox et al. “Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes”. In: *Phys. Rev. Lett.* 41 (23 Dec. 1978), pp. 1581–1585. DOI: 10.1103/PhysRevLett.41.1581. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.41.1581>.
- [44] S. H. Lee et al. “Evidence for $B^0 \rightarrow \pi^0 \pi^0$ ”. In: *Phys. Rev. Lett.* 91 (26 Dec. 2003), p. 261801. DOI: 10.1103/PhysRevLett.91.261801. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.91.261801>.
- [45] Tomasz Skwarnicki. “A STUDY OF THE RADIATIVE CASCADE TRANSITIONS BETWEEN THE UPSILON-PRIME AND UPSILON RESONANCES”. In: (). DESY-F31-86-02.
- [46] H. Albrecht. “Measurement of the polarization in the decay $B \rightarrow J/\psi K^*$ ”. In: *Physics Letters B* 340 (Dec. 1994), pp. 217–220. DOI: 10.1016/0370-2693(94)01302-0.
- [47] Sebastien Descotes-Genon et al. “Optimizing the basis of $B \rightarrow K^* \ell^+ \ell^-$ observables in the full kinematic range”. In: *JHEP* 1305 (2013), p. 137. DOI: 10.1007/JHEP05(2013)137. arXiv: 1303.5794 [hep-ph].
- [48] R Aaij et al. “Differential branching fraction and angular analysis of the decay $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ ”. In: *Phys.Rev.Lett.* 108 (2012), p. 181806. DOI: 10.1103/PhysRevLett.108.181806. arXiv: 1112.3515 [hep-ex].

- [49] Joaquim Matias. “On the S-wave pollution of $B \rightarrow K^* \ell^+ \ell^-$ observables”. In: *Phys. Rev. D* 86 (2012), p. 094024. DOI: 10.1103/PhysRevD.86.094024. arXiv: 1209.1525 [hep-ph].
- [50] R. Aaij et al. “Measurement of the CP Asymmetry in $B^0 \rightarrow K^{*0} \mu^+ \mu^-$ Decays”. In: *Phys. Rev. Lett.* 110 (3 Jan. 2013), p. 031801. DOI: 10.1103/PhysRevLett.110.031801. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.110.031801>.
- [51] Sébastien Descotes-Genon et al. “On the impact of power corrections in the prediction of $B \rightarrow K^* \mu^+ \mu^-$ observables”. In: *JHEP* 12 (2014), p. 125. DOI: 10.1007/JHEP12(2014)125. arXiv: 1407.8526 [hep-ph].
- [52] A. Khodjamirian et al. “Charm-loop effect in $B \rightarrow K^{(*)} \ell^+ \ell^-$ and $B \rightarrow K^* \gamma$ ”. In: *JHEP* 09 (2010), p. 089. DOI: 10.1007/JHEP09(2010)089. arXiv: 1006.4945 [hep-ph].
- [53] Aoife Bharucha et al. “ $B \rightarrow V \ell^+ \ell^-$ in the Standard Model from Light-Cone Sum Rules”. In: (2015). arXiv: 1503.05534 [hep-ph].
- [54] Ronald R. Horgan et al. “Lattice QCD calculation of form factors describing the rare decays $B \rightarrow K^* \ell^+ \ell^-$ and $B_s \rightarrow \phi \ell^+ \ell^-$ ”. In: *Phys. Rev. D* 89.9 (2014), p. 094501. DOI: 10.1103/PhysRevD.89.094501. arXiv: 1310.3722 [hep-lat].
- [55] R. R. Horgan et al. “Rare B decays using lattice QCD form factors”. In: *PoS LATTICE2014* (2015), p. 372. arXiv: 1501.00367 [hep-lat].
- [56] Sébastien Descotes-Genon et al. “Global analysis of $b \rightarrow s \ell \ell$ anomalies”. In: (2015). arXiv: 1510.04239 [hep-ph].
- [57] S. Jäger et al. “On $B \rightarrow V \ell \ell$ at small dilepton invariant mass, power corrections, and new physics”. In: *JHEP* 05 (2013), p. 043. DOI: 10.1007/JHEP05(2013)043. arXiv: 1212.2263 [hep-ph].
- [58] Sebastian Jäger et al. “Reassessing the discovery potential of the $B \rightarrow K^* \ell^+ \ell^-$ decays in the large-recoil region: SM challenges and BSM opportunities”. In: *Phys. Rev. D* 93.1 (2016), p. 014028. DOI: 10.1103/PhysRevD.93.014028. arXiv: 1412.3183 [hep-ph].
- [59] K. Flood. “New Results in Radiative Electroweak Penguin Decays at Babar, SLAC-PUB-15513”. In: (2010). DOI: SLAC-PUB-15513.
- [60] Christoph Bobeth et al. “Angular distributions of $\bar{B} \rightarrow \bar{K} \ell^+ \ell^-$ decays”. In: *JHEP* 12 (2007), p. 040. DOI: 10.1088/1126-6708/2007/12/040. arXiv: 0709.4174 [hep-ph].
- [61] Christoph Bobeth et al. “New Physics in $\Gamma_1 2^s$: $(\bar{s}b)(\bar{\tau}\tau)$ Operators”. In: *Acta Phys. Polon. B* 44 (2013), pp. 127–176. DOI: 10.5506/APhysPolB.44.127. arXiv: 1109.1826 [hep-ph].

- [62] Chris Bouchard et al. “Standard Model Predictions for $B \rightarrow K\ell^+\ell^-$ with Form Factors from Lattice QCD”. In: *Phys. Rev. Lett.* 111.16 (2013). [Erratum: *Phys. Rev. Lett.* 112, no. 14, 149902 (2014)], p. 162002. DOI: 10.1103/PhysRevLett.112.149902, 10.1103/PhysRevLett.111.162002. arXiv: 1306.0434 [hep-ph].
- [63] Ahmed Ali et al. “A Comparative study of the decays $B \rightarrow (K, K^*)\ell^+\ell^-$ in standard model and supersymmetric theories”. In: *Phys. Rev. D* 61 (2000), p. 074024. DOI: 10.1103/PhysRevD.61.074024. arXiv: hep-ph/9910221 [hep-ph].
- [64] R. Brun et al. *ROOT web page*, <http://root.cern.ch/>. 2001.
- [65] Wolfgang A. Rolke et al. “Limits and confidence intervals in the presence of nuisance parameters”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 551.2–3 (2005), pp. 493–503. ISSN: 0168-9002. DOI: <http://dx.doi.org/10.1016/j.nima.2005.05.068>. URL: <http://www.sciencedirect.com/science/article/pii/S016890020501291X>.
- [66] *Theta is a framework for template-based statistical modeling and inference, focussing on problems in high-energy physics*. URL: <http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto/>.
- [67] T. Abe et al. “Belle II Technical Design Report”. In: (2010). arXiv: 1011.0352 [physics.ins-det].
- [68] Christian Pulvermacher et al. “An automated framework for hierarchical reconstruction of B mesons at the Belle II experiment”. In: *J. Phys. Conf. Ser.* 608.1 (2015), p. 012048. DOI: 10.1088/1742-6596/608/1/012048. arXiv: 1410.3259 [hep-ex].

List of Figures

1.1	Constituents of the Standard Model. Credit: Cern.	6
1.2	Constraints on the Unitary Triangle by the CKMfitter group [16]. . .	9
1.3	Example Feynman diagrams for the flavor changing neutral current $b \rightarrow s$. The left diagram is forbidden in the Standard Model.	9
1.4	Feynman diagrams for the underlying process of the operators $\mathcal{O}'_7$ (left), $\mathcal{O}'_9$ and $\mathcal{O}'_{10}$ (right).	11
1.5	Standard Model Feynman graphs (a,b,c) for the decay $B \rightarrow K^{(*)}\ell^+\ell^-$ featuring penguin and box processes and (d) one non-SM scenario including charged Higgs replacing the W boson loop.	12
1.6	Definition of the angles in the decay $B^0 \rightarrow K^*\ell^+\ell^-$	14
1.7	P'_5 measurement by LHCb with results from 2013 [1] and from 2015 [2] shown in blue and black respectively.	18
1.8	New physic scenarios for the observables S_4 , S_5 and S_6^2 , taken from Ref. [18].	20
2.1	KEKB	22
2.2	A schematic view of the Belle detector, from [27].	24
2.3	A side view the ACC and TOF of the Belle detector from [27].	25
2.4	Belle event display of a fully reconstructed $B \rightarrow K_S J/\psi$ event where the J/ψ decays into two muons. The top plots show the x-y plane and the bottom plot displays the y-z side-view of the detector.	28
2.5	Belle integrated luminosity. Taken from [27].	29
3.1	Distributions of different types of correlations between two observables (left column) with the corresponding flat transformation (right column). Plot (a) shows a correlated dataset with $r = 0.8$ and (c) and (e) both show distributions without linear correlation $r = 0.0$	35
3.2	Demonstration of the dependency between x_2 and y_2 although both distributions have no linear correlations, i.e. $r = 0$	36
3.3	Flattening of a distribution with its CDF. The left plot shows the original distribution x and the cumulative distribution function CDF . The right plot is the flat transformation of x which can be calculated by $x_f = CDF(x)$	37

3.4	Flat correlation of correlated and uncorrelated distributions. The distribution (a) shows a linear correlation from fig. 3.1(a). In the second plot (b) the distribution of fig. 3.1(c) is shown in (c) the uncorrelated distribution from fig. 3.1(e) is displayed as a flat correlation plot. The deviation from a flat expectation is shown colored and expressed in standard deviations σ	38
3.5	Example of a test statistic with the corresponding purity vs. efficiency plot on the right side.	40
3.6	Topology of a Neural Network (right). The transfer function (left) with different learning gradients β is used in each perceptron to map the sum of the input to the range $[-1, 1]$	42
3.7	Example of a decision tree from a training in part III of this thesis. Cuts are performed to variables of the training-set. The last row states the number of entries in each category, the corresponding impurity and the output value of the tree.	45
3.8	Schematic diagram of the full reconstruction.	46
4.1	Veto regions for charmonium background for the di-electron (right) and di-muon (left) channels.	55
4.2	Network output for the classifiers in stage 1 for signal and background on a set of 10,000 generic MC events after particle selection and before event selection.	58
4.3	Network output for signal (red) and background (black) in the first row and purity vs. network output (black data points) in the second row for the trainings of NeuroBayes for the charged K^* modes, produced by the NeuroBayes trainings-report.	60
4.4	Network output for signal (red) and background (black) in the first row and purity vs. network output (black data points) in the second row for the trainings of NeuroBayes for the neutral K^* modes, produced by the NeuroBayes trainings-report.	61
4.5	Background for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ on generic MC in M_{bc} with different truth for the selected leptons.	66
4.6	Peaking background on the rare MC dataset in M_{bc} and $q^2 \equiv M_{\ell^+ \ell^-}^2$	66
4.7	Flat correlation plots (see section 3.2.2 for further information about this kind of plots) for the neural network output vs. M_{bc} for two example channels. There is no significant correlation observable.	67
4.8	Estimation of the best cut on the basis of the Figure Of Merit (FOM) for each decay channel.	68
4.9	Performance of the neural networks for the classification of B mesons. The data for signal corresponds to correctly reconstructed events in the signal MC and background corresponds to generic MC for two times the expected size of the Belle dataset.	70

5.1	Signal extraction for $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$ on the total range of q^2 . Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	75
5.2	Signal extraction for $B^0 \rightarrow K^*(892)^0 \ell^+ \ell^-$, $\ell = e, \mu$ in bins of q^2 . Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	76
5.3	Comparison of generator distribution for $B^0 \rightarrow K^*(892)^0 e^+ e^-$ and $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$. The corresponding veto regions for the charmonium resonances are superimposed.	77
5.4	Extraction of the signal yields for $B \rightarrow J/\psi \ell \ell$ and $B \rightarrow \psi(2S) \ell \ell$ in the corresponding q^2 sidebands. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	80
5.5	Neural network output comparison between data and Monte Carlo from the reconstruction of $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ in the M_{bc} sideband $M_{bc} < 5.27 \text{ GeV}/c^2$. The p-value of a Kolmogorov-Smirnov statistic on 2 samples is stated.	81
5.6	Data vs. MC comparison of the reconstruction of $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ in the $9.4 \leq q^2 < 9.8 \text{ GeV}^2/c^4$ and $13 \leq q^2 < 14 \text{ GeV}^2/c^4$ veto regions.	82
5.7	Data vs. MC comparison of the reconstruction of $B^0 \rightarrow K^*(892)^0 e^+ e^-$ in the $9.4 \leq q^2 < 9.8 \text{ GeV}^2/c^4$ and $13 \leq q^2 < 14 \text{ GeV}^2/c^4$ veto regions.	83
6.1	Binning in $q^2 \equiv M_{\ell^+ \ell^-}^2$ with generated distribution for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$. The cut veto for the charmonium resonances is indicated in red.	86
6.2	Simulated data according to eq. (1.37) is folded for the extraction of P'_4 using definitions of eq. (1.38). The line indicates the fit result for the transformed PDF eq. (6.1).	87
6.3	Correlation among the angular variables in the background sample displayed as correlation matrix (left) and flat correlation matrix (right), see details for flat correlation plots in section 3.2.2. In the correlation matrix a color is displayed corresponding to the amount of correlation, where green signals a correlation coefficient of around zero.	88
6.4	Definition of the signal and sideband regions in M_{bc} in an example fit for the data of q^2 bin 4.	89
6.5	Flat correlation plots of the angular variables with M_{bc} in the background sample with the probability p for a flat hypothesis (see details for this kind of plot in section 3.2.2).	90
6.6	Flat correlation plots of the angular variables with NB_{out} in the background sample with the probability p for a flat hypothesis (see details for this kind of plot in section 3.2.2).	90
6.7	Result of the toy studies for two different background parametrization, where each data point represents the averaged result of 10000 fits to pseudo data.	92

6.8	Result for simultaneous fitting. In each bin the measurements of P'_4 and P'_5 are fitted simultaneously and share the same values for F_L and A_T , where each data point represents the averaged result of 10000 fits to pseudo data.	92
6.9	Result for different cuts for the neural network which lead to varied signal yields and purities in each measurement. All data points represents the averaged result of 10000 fits to pseudo data.	93
6.10	Cut optimization for the measurement of P'_5 in bin 2. The estimated total error $e_{stat} + e_{bias} + \hat{e}_{peaking}$ is displayed on the z-axis in dependence of the cuts for the electron (c_{ee}) and muon ($c_{\mu\mu}$) channel on a large range of cuts (top) and in the region with the global minimum (bottom).	95
6.11	Result of the MC toy studies with fixed values for A_T or F_L and the expected signal and background yields for the FOM cut.	96
6.12	Estimation of the sensitivity for the muon mode only, displayed in addition to the combined sample.	96
6.13	Correlation among the efficiency of the angular variables in the background sample displayed as flat correlation matrix (see details for this kind of plot in section 3.2.2).	98
6.14	Results for the fits of the efficiencies. Both the fits for the electron (orange) and muon (blue) efficiency are superimposed over the ratio of generated and reconstructed events (data points).	99
6.15	q^2 dependence of the reconstruction efficiency in $\cos\theta_\ell$ for each bin of q^2 for $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ (left) and $B^0 \rightarrow K^*(892)^0 e^+ e^-$ (right) (see details for this kind of plot in section 3.2.2).	101
7.1	Fit for the acceptance function for $\cos\theta_K$ with a linearly depended bias. The bias simulates twice the observed deviation between data and MC in this variable. From a fit with the modification of the efficiency function the systematic uncertainty is derived for the data MC discrepancy.	104
7.2	Fit to the simulated data with efficiency and acceptance effects from the neural network and with added bias to the $\cos\theta_K$ acceptance fit.	105
7.3	Toy Study With Efficiency: Displayed is the comparison between a toy study with data on generator level and data on reconstruction level with added weights according to the efficiency fit with the NN cut from the best figure of merit.	106
7.4	The $K^*(892)$ invariant mass in the decay of $B^0 \rightarrow K^*(892)^0 \mu^+ \mu^-$ for a low and a high cut on the neural network (NN). Combinatorial (dashed blue), potential S-wave signal (red filled) and total (solid) fit distributions are superimposed on the data points.	109

8.1	Projections for the fit result of P'_5 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	113
8.2	Result for P'_4 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.	114
8.3	Result for P'_5 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.	115
8.4	Result for P'_6 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.	116
8.5	Result for P'_8 compared to Standard Model predictions from various sources described in section 8.1.1. Results from LHCb [1, 2] are shown for comparison.	117
8.6	Result for F_L and A_T in the fit for P'_5 compared to Standard Model predictions by [19]. The error corresponds to the statistical only. Results from LHCb [1, 2] are shown for comparison.	118
8.7	Result for derived quantities $S_{4,5,7,8}$ and S_3 . The error corresponds to the statistical only. Results from LHCb [1, 2] are shown for comparison.	119
8.8	Distribution of the distance between measurements and theory predictions for the models "DHMV", "BSZ" and "Camalich et.al.".	121
10.1	Pre-cuts on simulated data. The distributions are arbitrarily normalized.	130
10.2	Input variables for the boosted decision tree. The signal distribution is shown with a red line and the background is displayed as a stacked plot for several sources (blue tones for generic MC and green for rare $ul\nu$ MC). The description of the variables can be found in section 10.3.	132
10.3	Input variables for the boosted decision tree. Continuation of fig. 10.2.	133
10.4	Performance comparison of different classification algorithms on an independent test-dataset.	134
11.1	The output of the BDT and the variable E_{ECL} prior to the cut for the best upper limit.	137
11.2	Calculation of the lowest upper limit on $\mathcal{B}(B^+ \rightarrow K^+ \tau^+ \tau^-)^{UL}$ at 95% C.L. in dependence on the cut c_E on E_{ECL} and c_{BDT} on the output of the BDT.	137
11.3	Observed and expected limits for different numbers of observed events, calculated with Bayesian methods [66] at 95% C.L.	138
12.1	Impact of the Monte Carlo efficiency correction for the B_{tag} candidate. Displayed is M_{bc} of the B_{tag} with (b) and without (a) weights (c).	140

12.2	E_{ECL} sideband for all different types of MC (left) and the comparison with data (right) after the Monte Carlo efficiency correction for the B_{tag} candidate is applied.	141
12.3	E_{ECL} distribution for $B^+ \rightarrow K^+ \tau^+ \tau^- (K_S)$, where the K_S daughters are removed from the event with the final cut $c_{BDT} = 0.553$ (right) and $c_{BDT} > 0.1$ (left) for comparison.	142
13.1	Expected amount of background for the final result. The signal distribution has arbitrary normalization.	146
A.1	Data vs. MC comparison of the Neural Network input variables after the final cut on the network. In the MC component, the number of background and signal events are composed according to fitted yields from data. The p-value of a Kolmogorov-Smirnov statistic on 2 samples is stated for each variable.	168
A.2	Continuation of fig. A.1.	169
A.3	Projections for the fit result of P'_4 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	170
A.4	Projections for the fit result of P'_4 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	171
A.5	Projections for the fit result of P'_4 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	172
A.6	Projections for the fit result of P'_4 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	173
A.7	Projections for the fit result of P'_4 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	174
A.8	Projections for the fit result of P'_5 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	175
A.9	Projections for the fit result of P'_5 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	176

- A.10 Projections for the fit result of P'_5 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 177
- A.11 Projections for the fit result of P'_5 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 178
- A.12 Projections for the fit result of P'_5 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 179
- A.13 Projections for the fit result of P'_6 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 180
- A.14 Projections for the fit result of P'_6 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 181
- A.15 Projections for the fit result of P'_6 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 182
- A.16 Projections for the fit result of P'_6 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 183
- A.17 Projections for the fit result of P'_6 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 184
- A.18 Projections for the fit result of P'_8 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 185
- A.19 Projections for the fit result of P'_8 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points. 186

A.20	Projections for the fit result of P'_8 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	187
A.21	Projections for the fit result of P'_8 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	188
A.22	Projections for the fit result of P'_8 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.	189
B.1	Input variables for the boosted decision tree.	192
B.2	Input variables for the boosted decision tree.	193
B.3	Data vs. MC comparison in the E_{ECL} sideband.	194
B.4	Data vs. MC comparison in the E_{ECL} sideband.	195

List of Tables

1.1	Fermions of the Standard Model with their corresponding quantum numbers regarding the electroweak force. The particles are divided into singlets and doublets of the weak isospin, denoted with R for right-handed and L for left-handed respectively. T is the weak isospin with its third component T_3 , Y is the weak hypercharge and Q the electrical charge.	7
2.1	Summary of the integrated luminosity recorded by Belle broken down to the resonance energy.	27
2.2	Composition of the Belle generic MC.	29
4.1	Decay channels of the exclusive reconstruction of $B \rightarrow K^{(*)}\ell^+\ell^-$, where $\ell = e, \mu$ and cc. is implied.	52
4.2	Branching fraction of $B \rightarrow K^{(*)}\ell^+\ell^-$ and expected number of signal events in Belle MC based on PDG ([4]).	52
4.3	Summary of the particle selection criteria.	53
4.4	Summary of the event selection criteria.	55
4.5	Variables used in the neural networks for the K^* candidates.	59
4.6	Best candidate selection efficiency as derived from signal MC before prior probability is adjusted.	65
4.7	Neural network cut for the highest figure of merit (FOM). The corresponding FOM, efficiency and expected signal and background events in the signal region $M_{bc} > 5.27 \text{ GeV}/c^2$ are displayed. The efficiency includes all reconstruction and acceptance effects.	67
4.8	Efficiency in bins of q^2 in %. The binning is described in the angular analysis in chapter 6. The efficiency is calculated based on the number of generated candidates within the q^2 range of the particular bin. . . .	71
5.1	Fitted yields for signal and background events in the binning of q^2 for both the combined and muon channels.	74
5.2	Reconstruction efficiency in the q^2 sidebands.	78
5.3	Fraction $f_{\ell\ell}$ of $X_{c\bar{c}}$ into leptons from PDG.	78
5.4	Extracted Branching Fraction. $X_{c\bar{c}}$ is either J/ψ or $\psi(2S)$	79

7.1	Systematic uncertainty for the difference between data and MC. . . .	105
7.2	Fit bias for all measurements in the statistically preferred binning. The values are the absolute deviations from the simulated values. . . .	105
7.3	Bias for all measurements from correcting for the reconstruction efficiency.	106
7.4	Estimated number of cross-feed events in the data.	107
7.5	Summary of all considered systematic uncertainties for P'_4	109
7.6	Summary of all considered systematic uncertainties for P'_5	110
7.7	Summary of all considered systematic uncertainties for P'_6	110
7.8	Summary of all considered systematic uncertainties for P'_8	110
8.1	Results of the angular analysis. Observables are compared to Standard Model predictions for the analyzed q^2 regions, detailed in section 8.1.1. Systematic errors are only calculated and presented for $P'_{4,5,6,8}$, where the first errors is the statistical and the second the systematic error. Values for A_T and F_L are presented from the fit and data transformation of P'_5	122
10.1	List of the reconstructed τ decay channels.	127
10.2	Summary of the event selection criteria.	128
10.3	Summary of pre-cuts used in the analysis. See section 10.3 for a definition of the variables. The efficiencies are determined independently.	129
12.1	List of covered systematic uncertainties.	144
13.1	Determined parameters for the best upper limit on MC including systematic uncertainties.	145

A. Appendix $B \rightarrow K^{(*)} \ell^+ \ell^-$

A.1 Data – Monte Carlo Comparison

The result for the data Monte Carlo comparison for all input variables of the neural network are depicted in figs. A.1 and A.2.

A.2 Fit Projections

Fit projections for the angular analysis are presented in figs. A.3 to A.22.

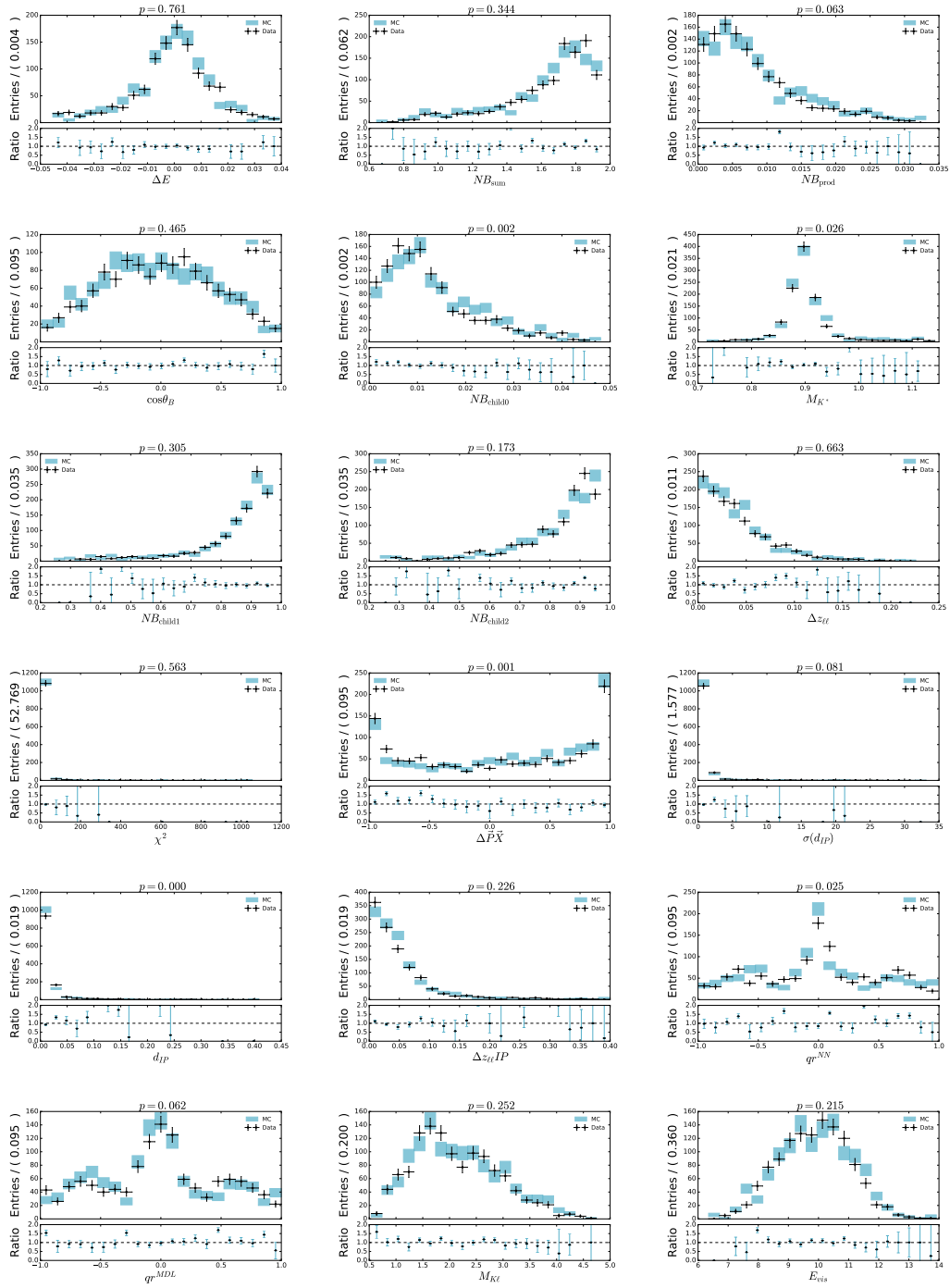


Figure A.1: Data vs. MC comparison of the Neural Network input variables after the final cut on the network. In the MC component, the number of background and signal events are composed according to fitted yields from data. The p-value of a Kolmogorov-Smirnov statistic on 2 samples is stated for each variable.

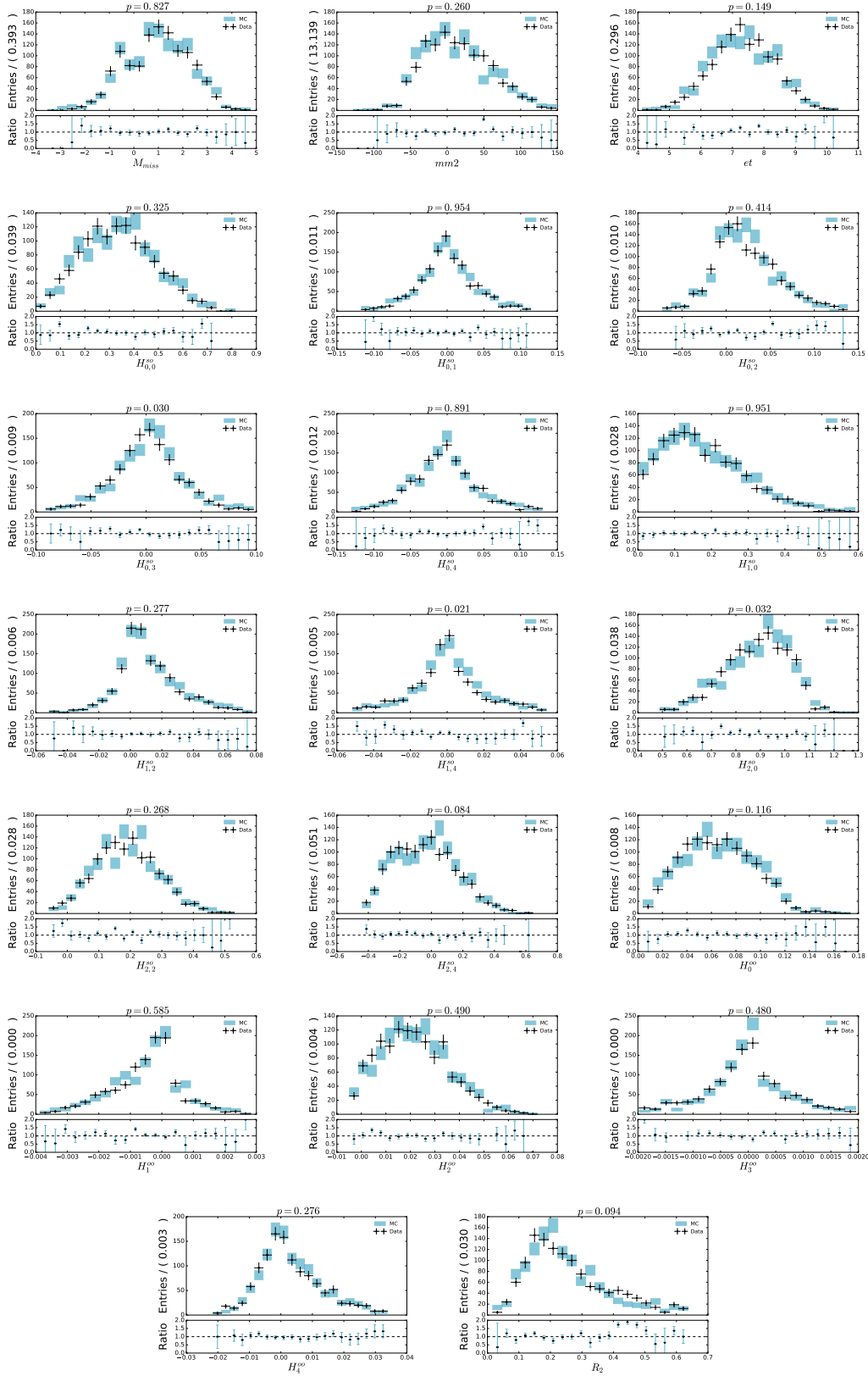


Figure A.2: Continuation of fig. A.1.

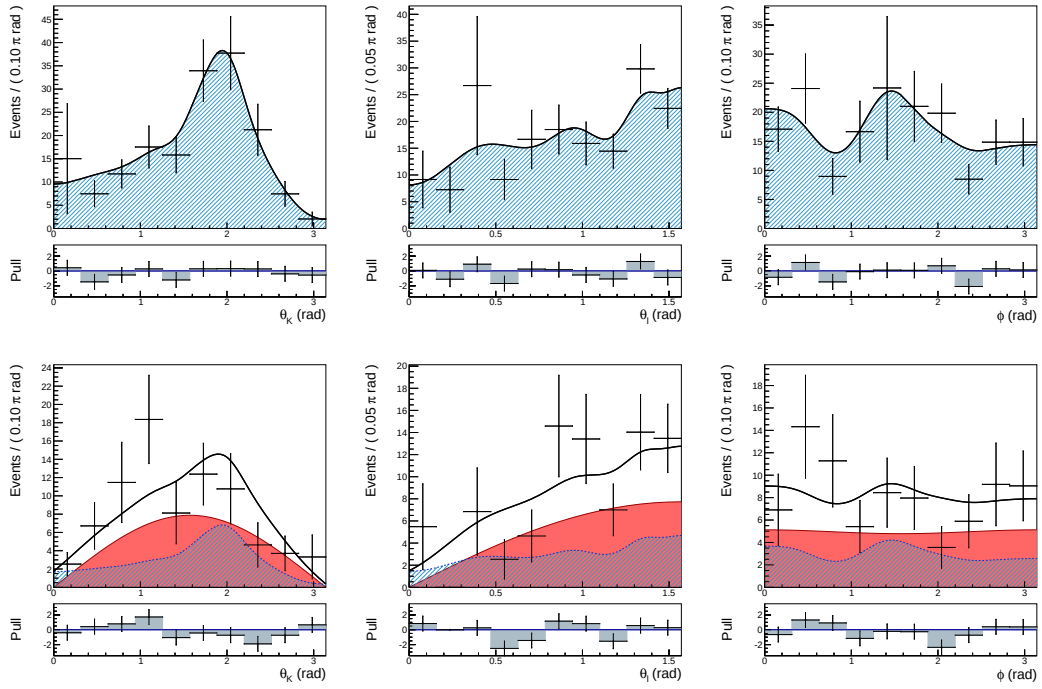


Figure A.3: Projections for the fit result of P'_4 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

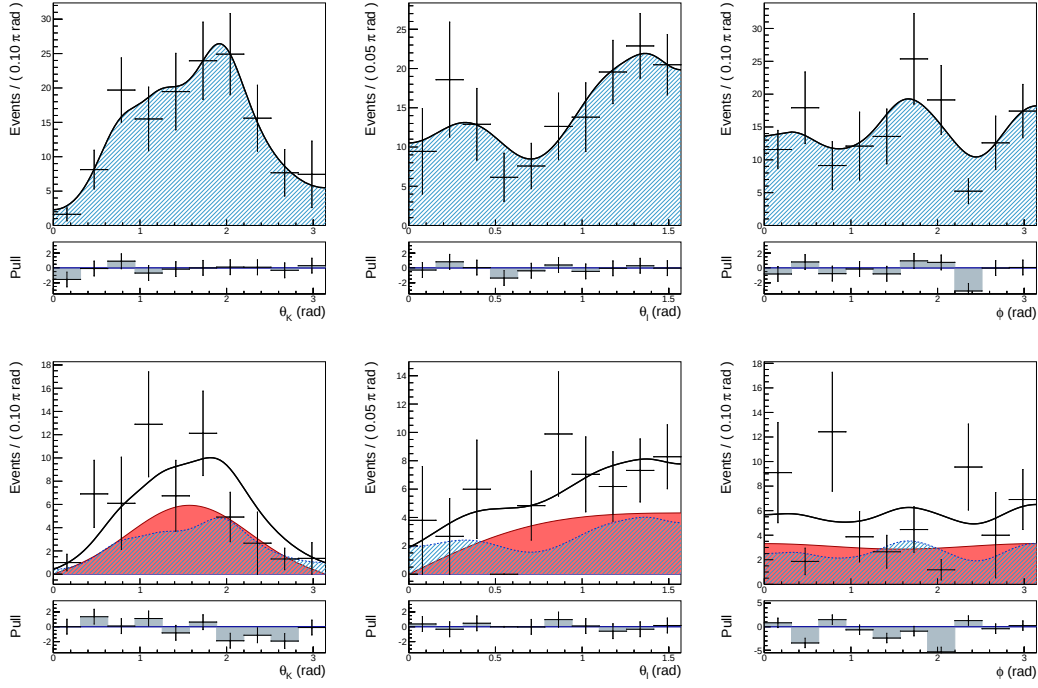


Figure A.4: Projections for the fit result of P'_4 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

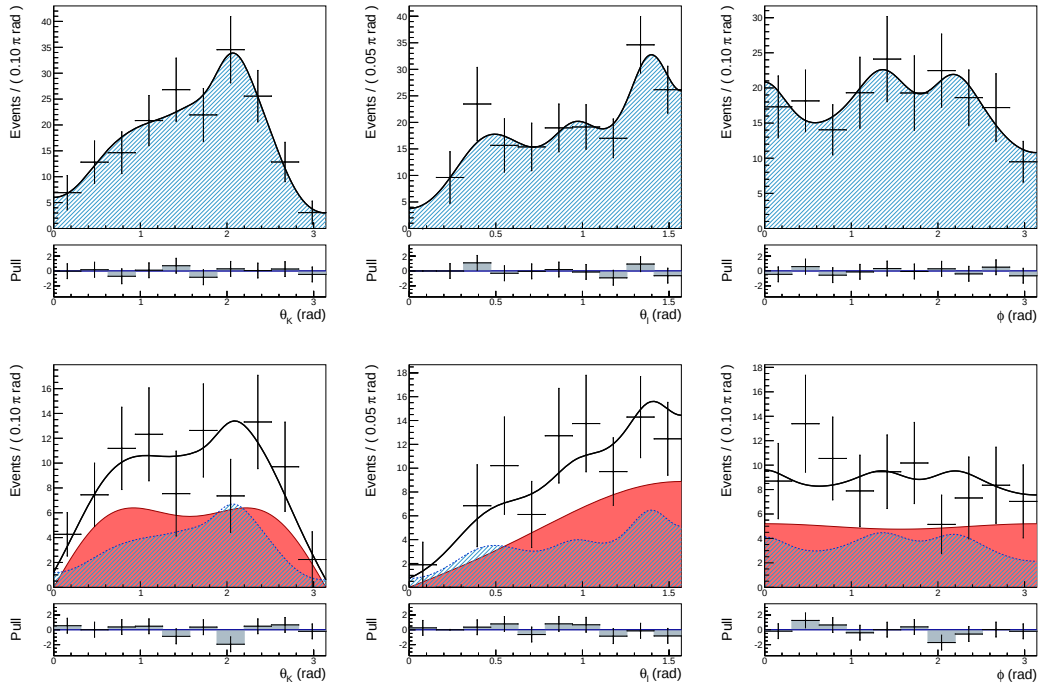


Figure A.5: Projections for the fit result of P'_4 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

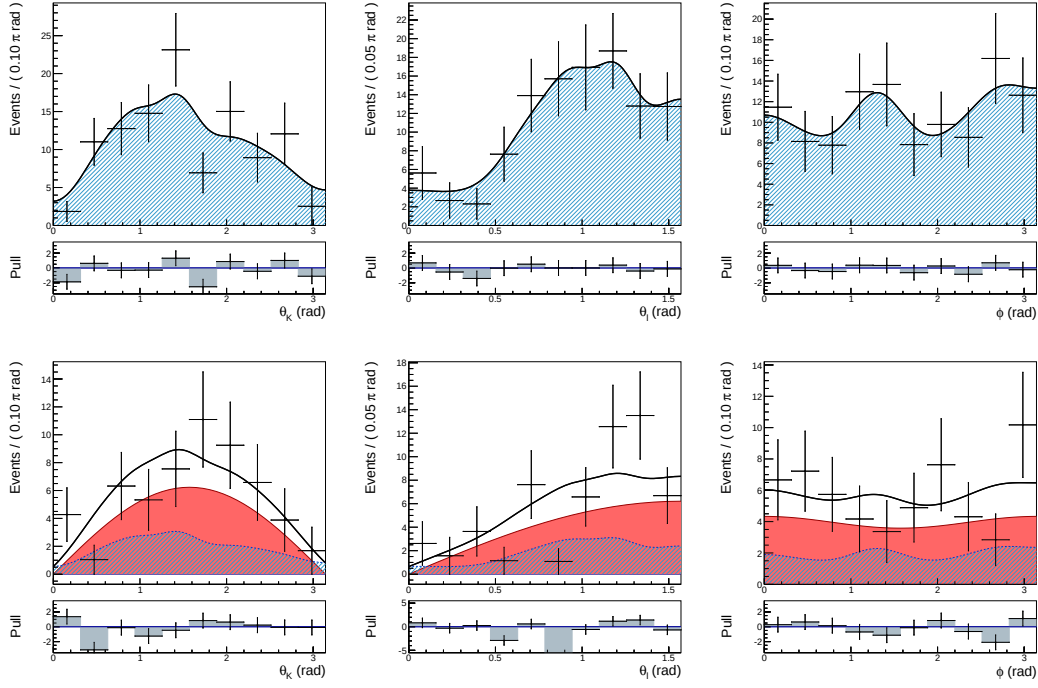


Figure A.6: Projections for the fit result of P'_4 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

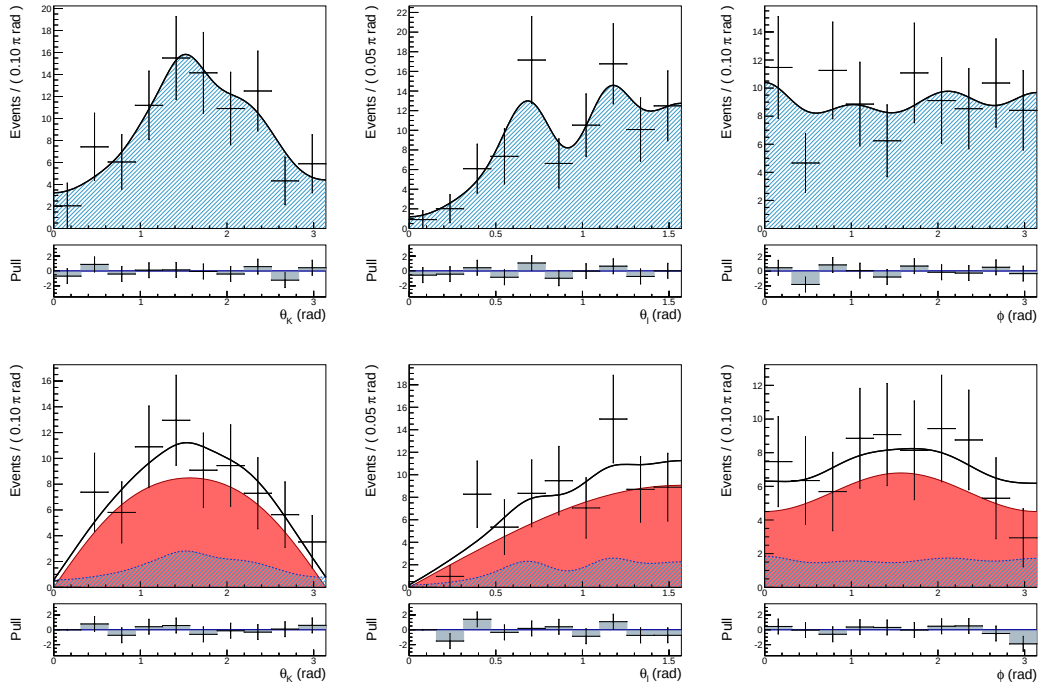


Figure A.7: Projections for the fit result of P'_4 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

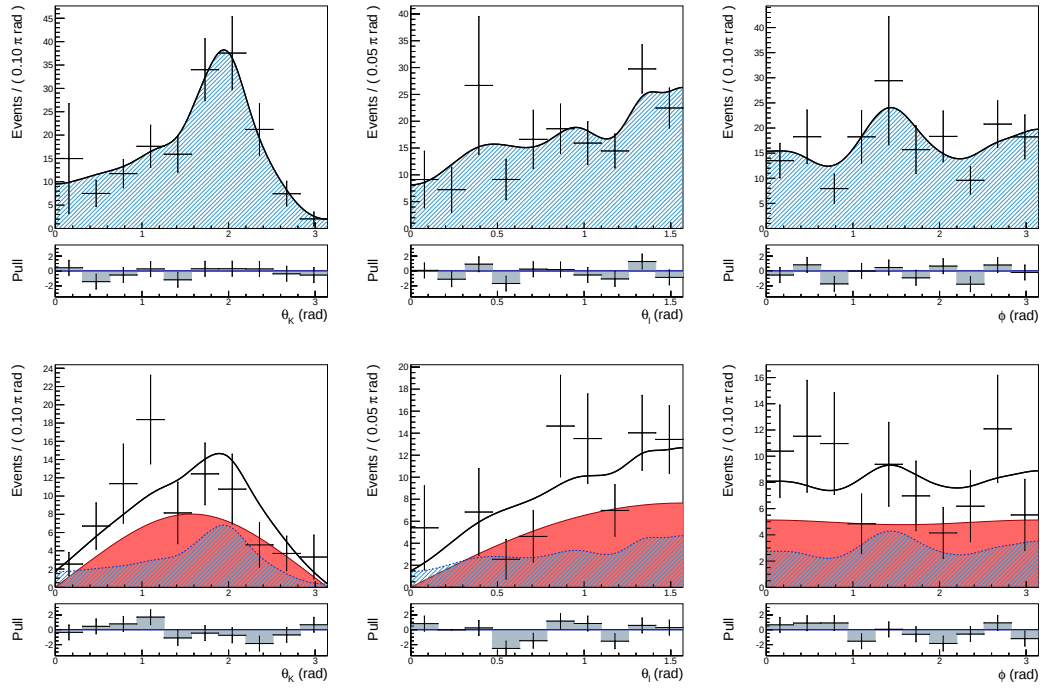


Figure A.8: Projections for the fit result of P'_5 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

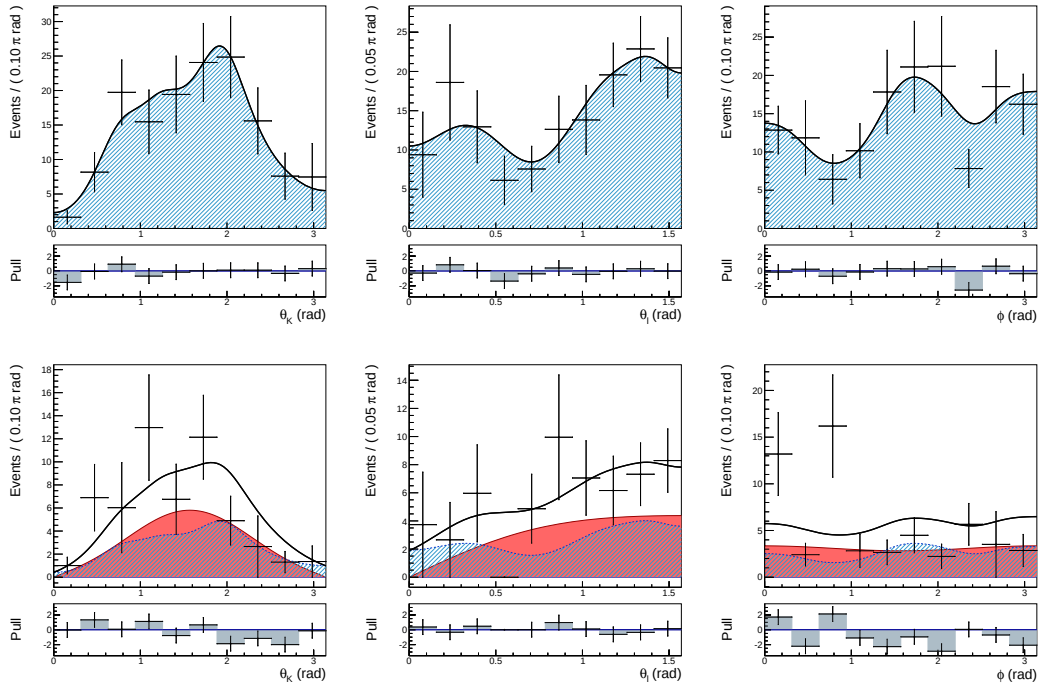


Figure A.9: Projections for the fit result of P'_5 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

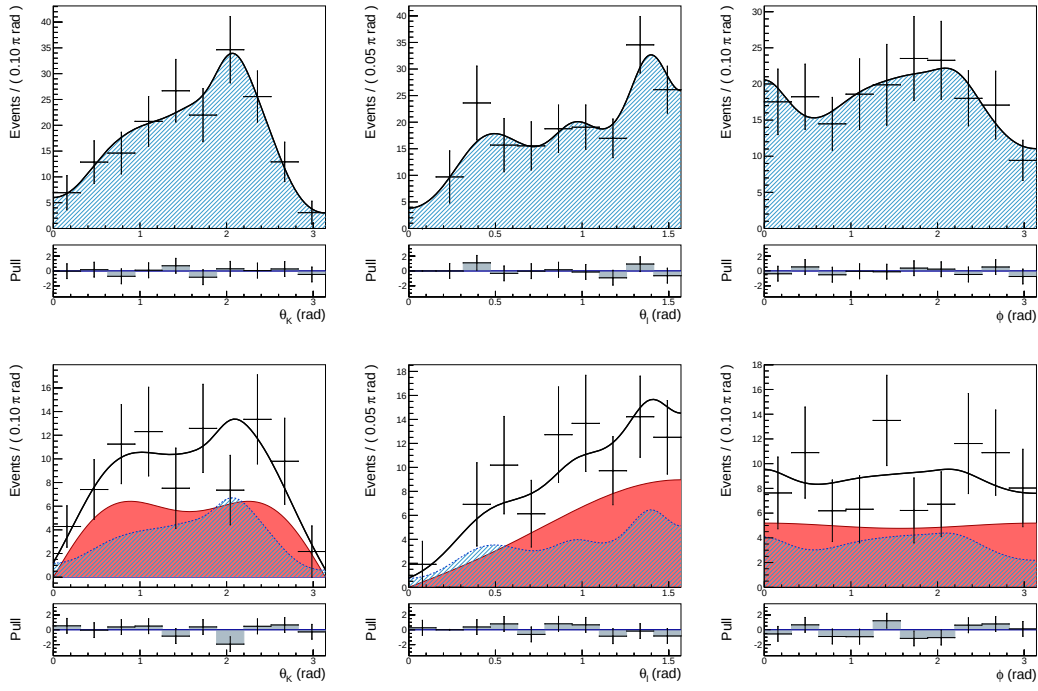


Figure A.10: Projections for the fit result of P'_5 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

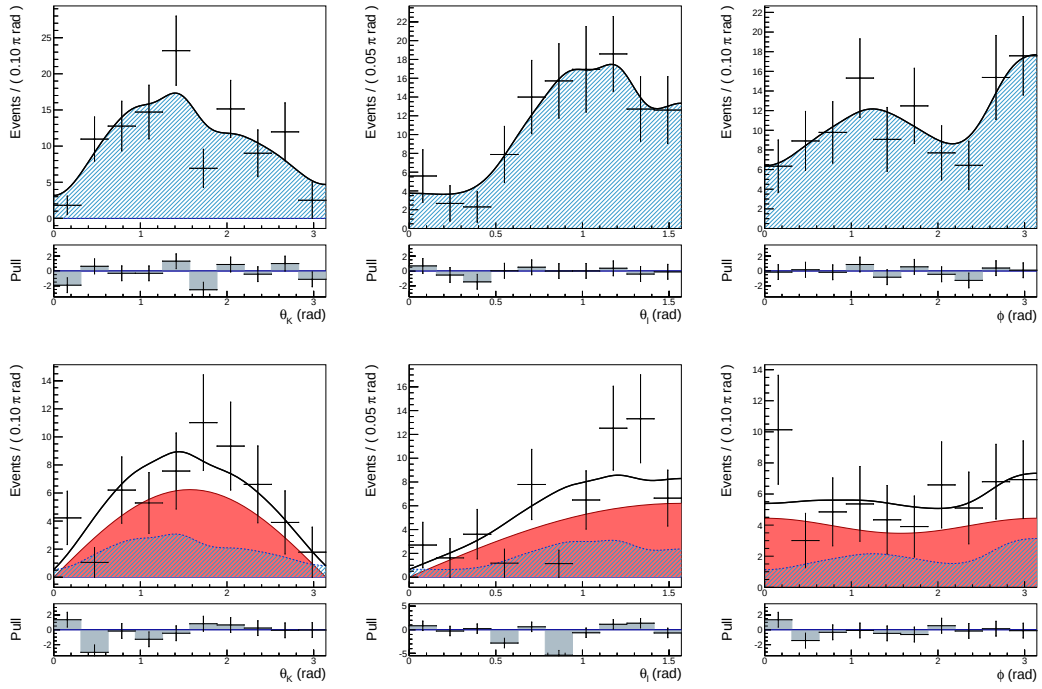


Figure A.11: Projections for the fit result of P'_5 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

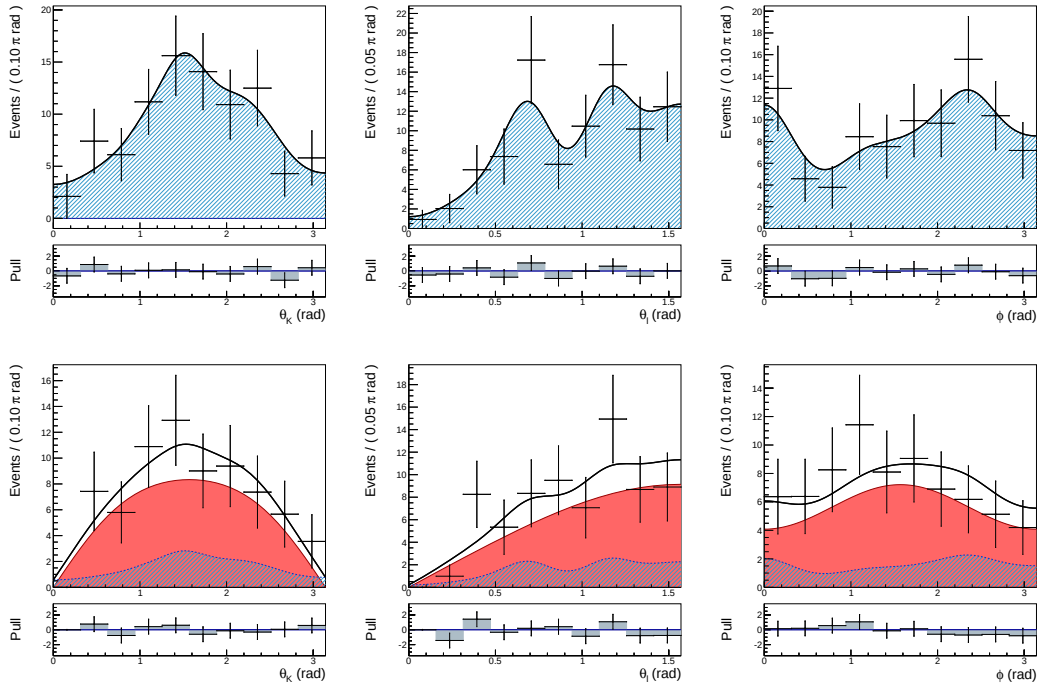


Figure A.12: Projections for the fit result of P'_5 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

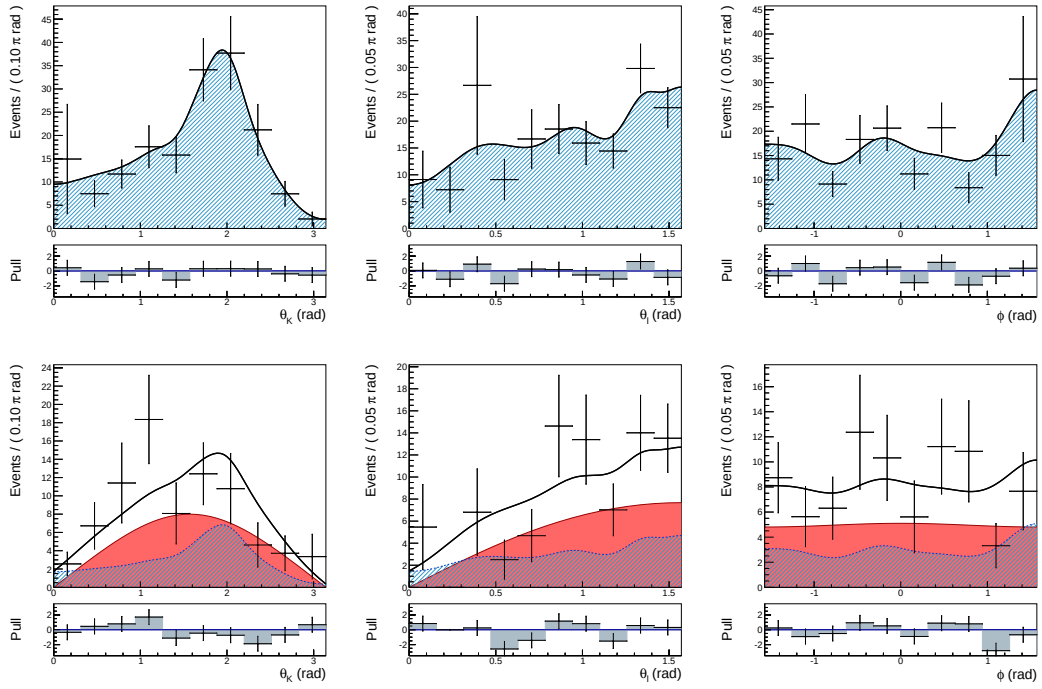


Figure A.13: Projections for the fit result of P'_6 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

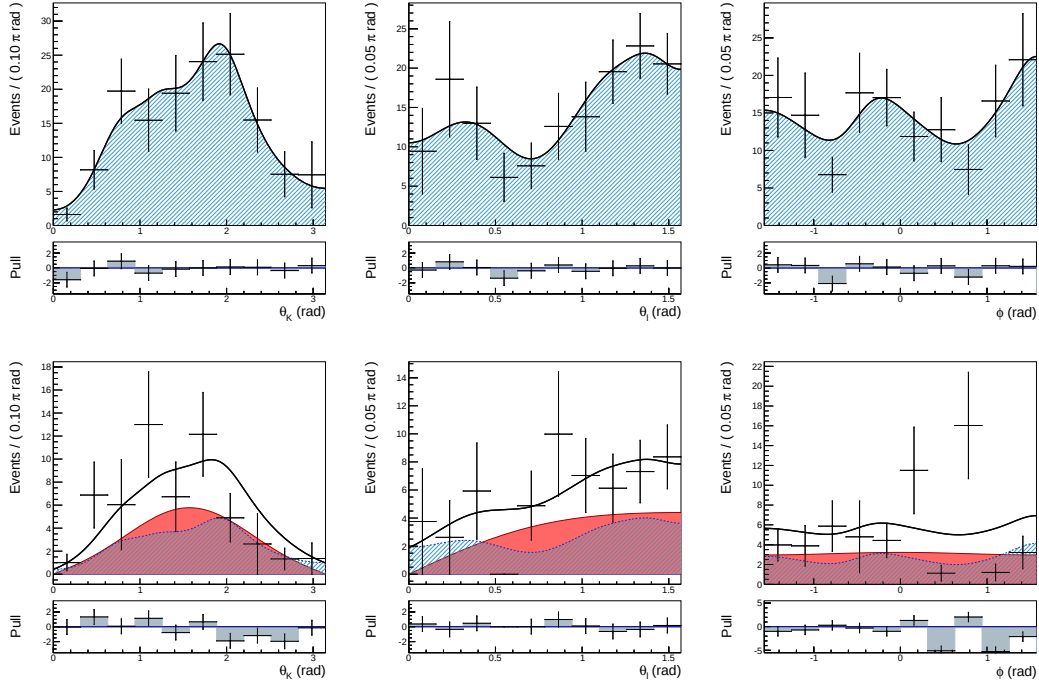


Figure A.14: Projections for the fit result of P'_6 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

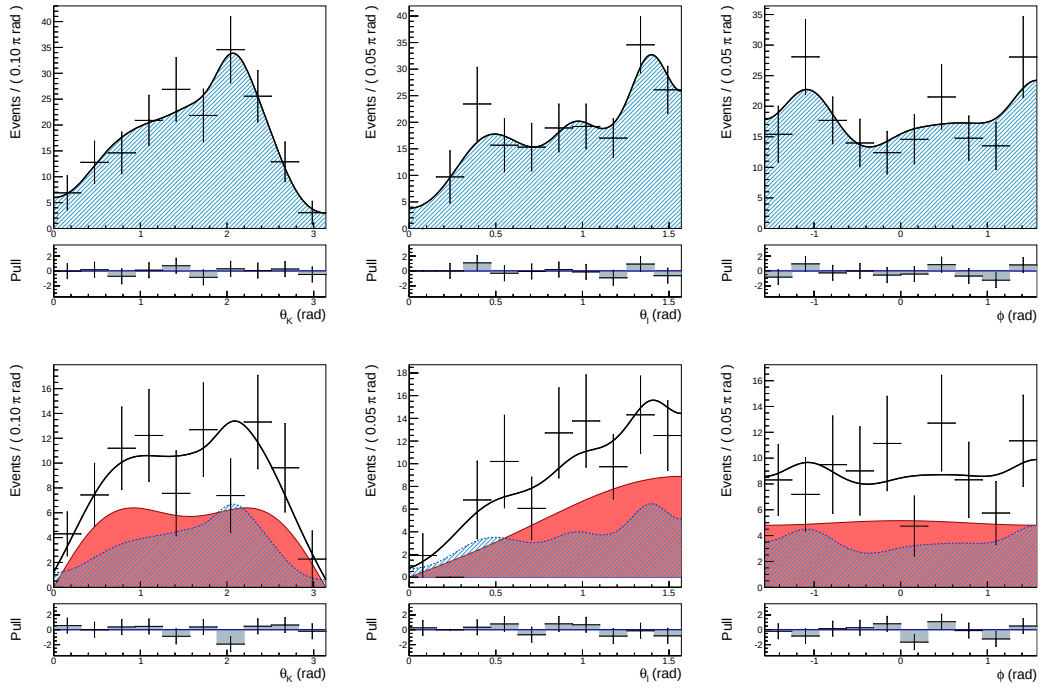


Figure A.15: Projections for the fit result of P'_6 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

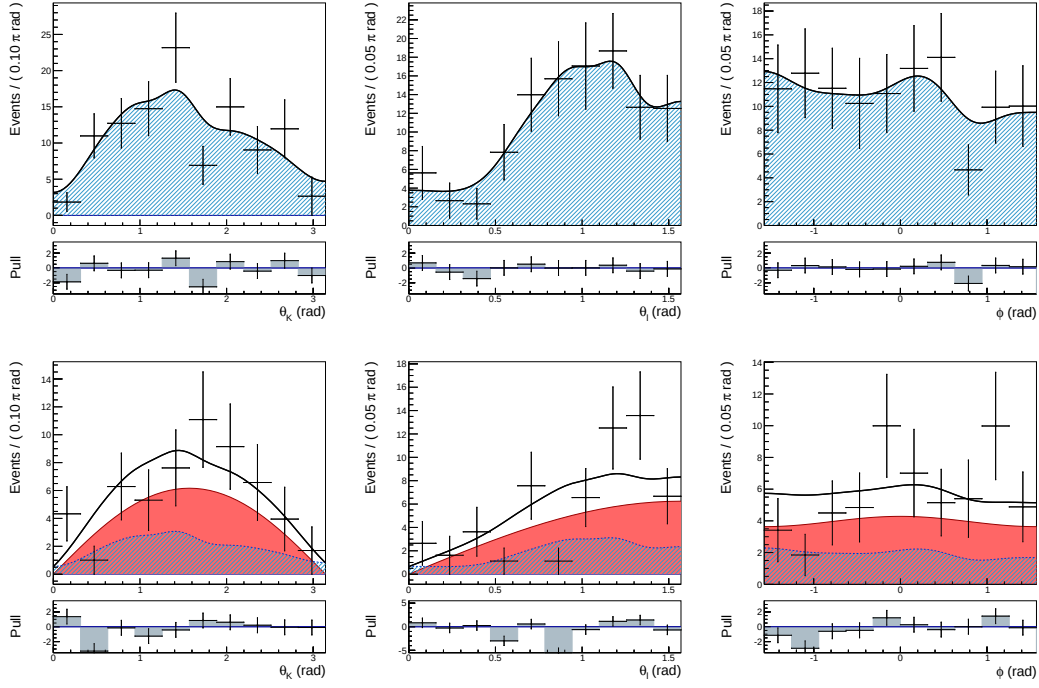


Figure A.16: Projections for the fit result of P'_6 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

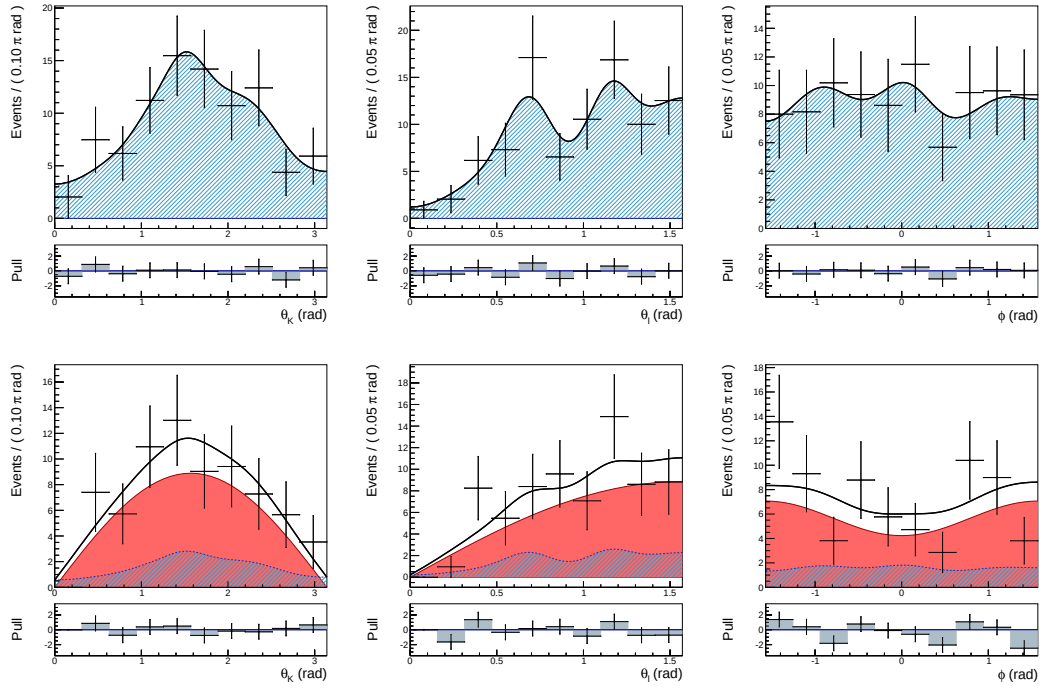


Figure A.17: Projections for the fit result of P'_6 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

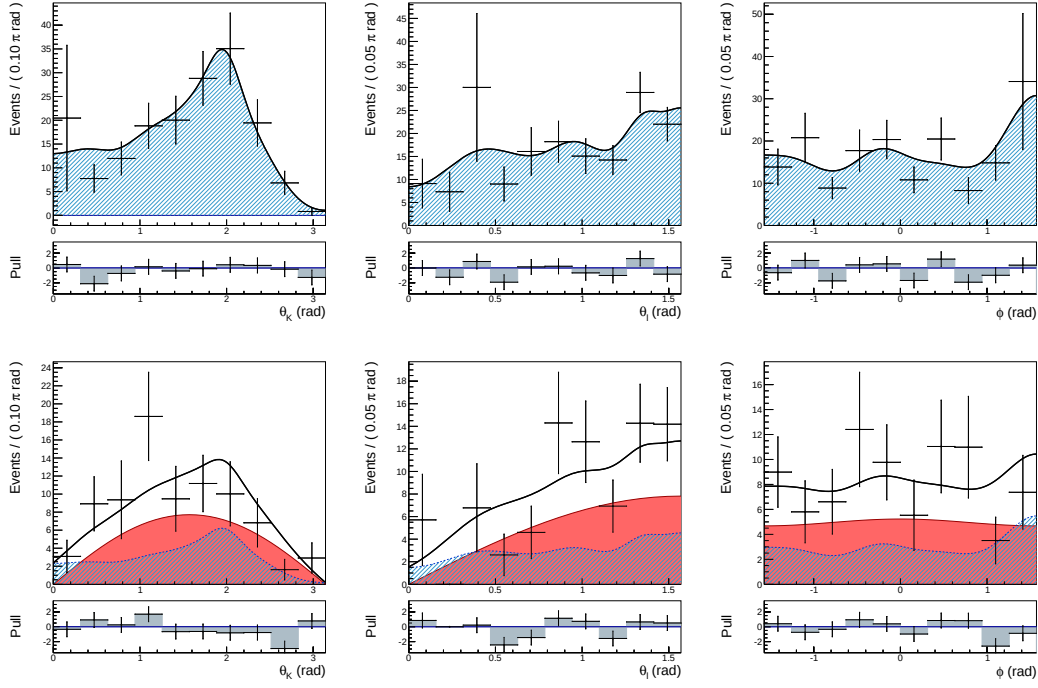


Figure A.18: Projections for the fit result of P'_8 in bin 0. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

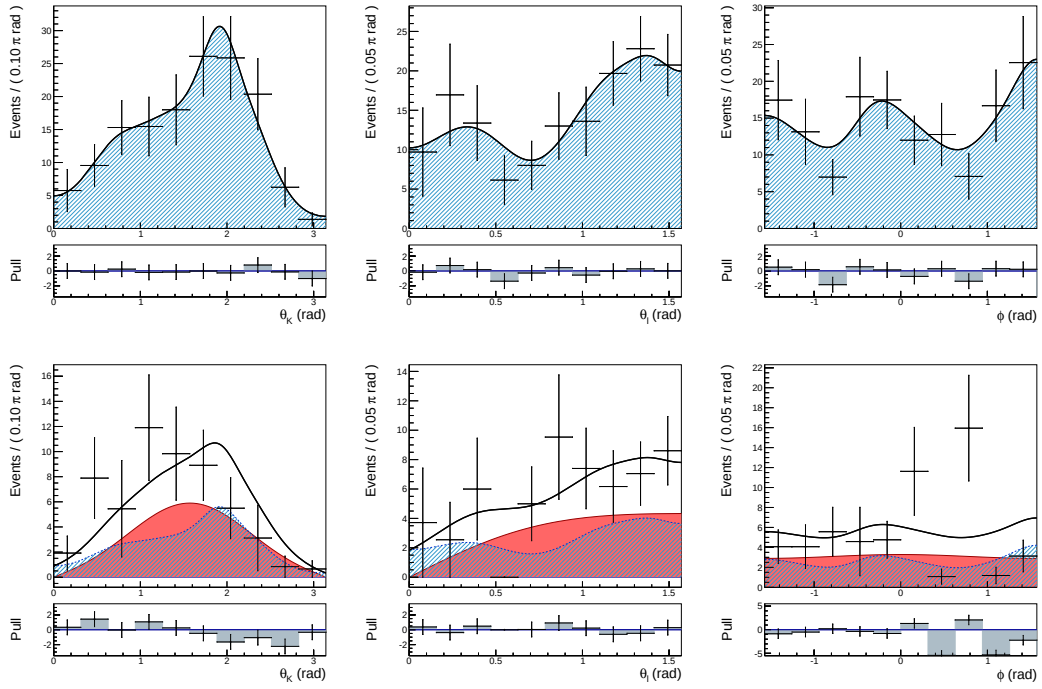


Figure A.19: Projections for the fit result of P'_8 in bin 1. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

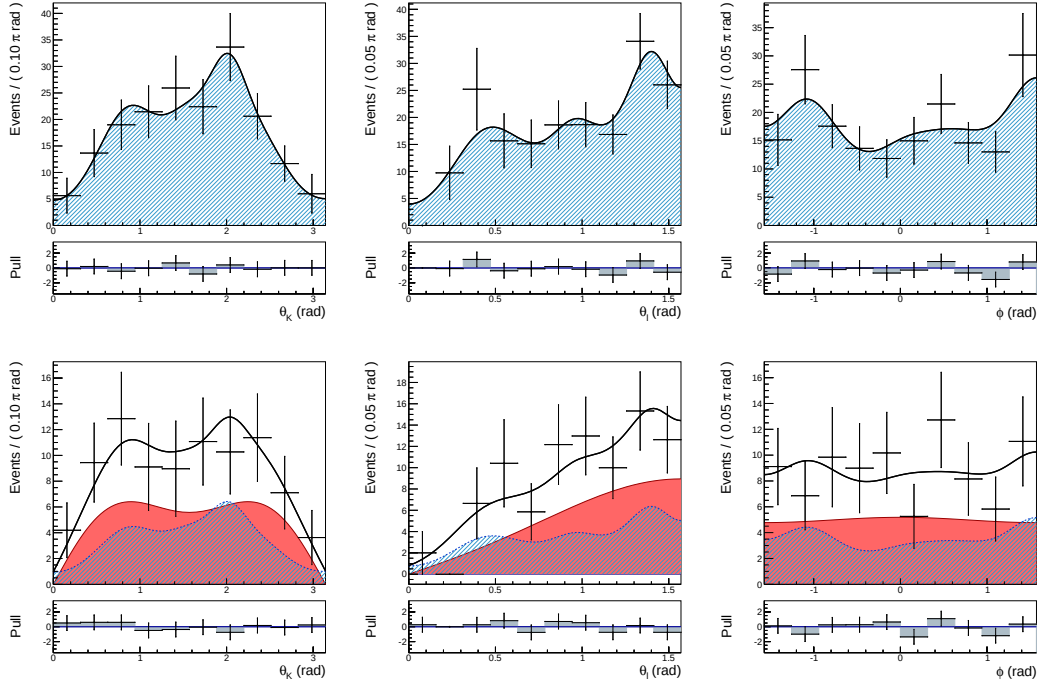


Figure A.20: Projections for the fit result of P'_8 in bin 2. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

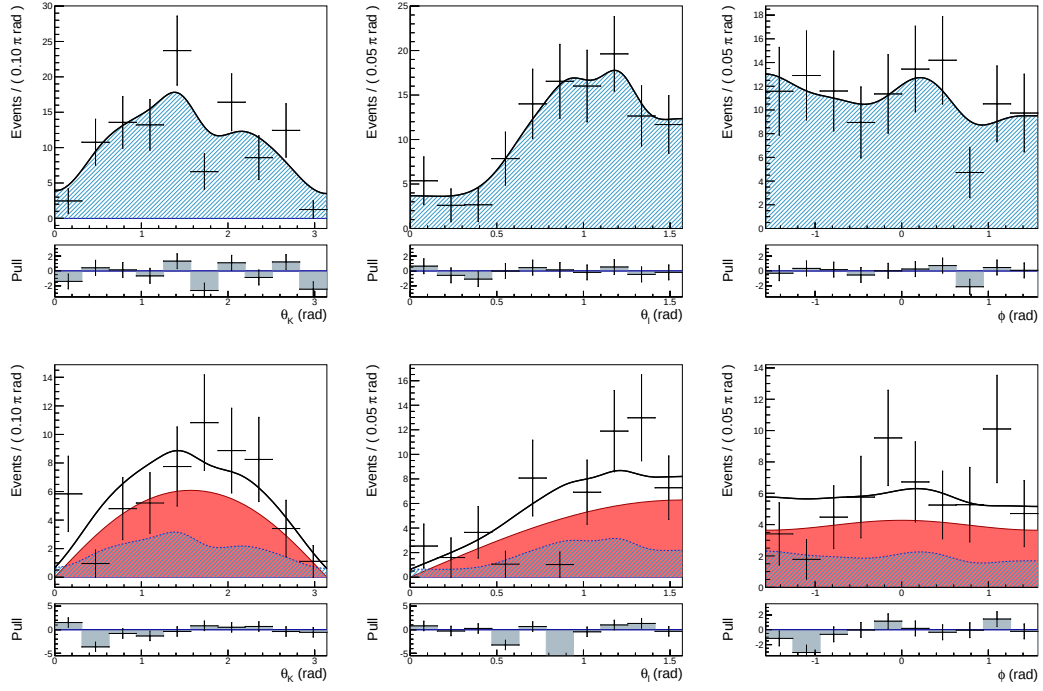


Figure A.21: Projections for the fit result of P'_8 in bin 3. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

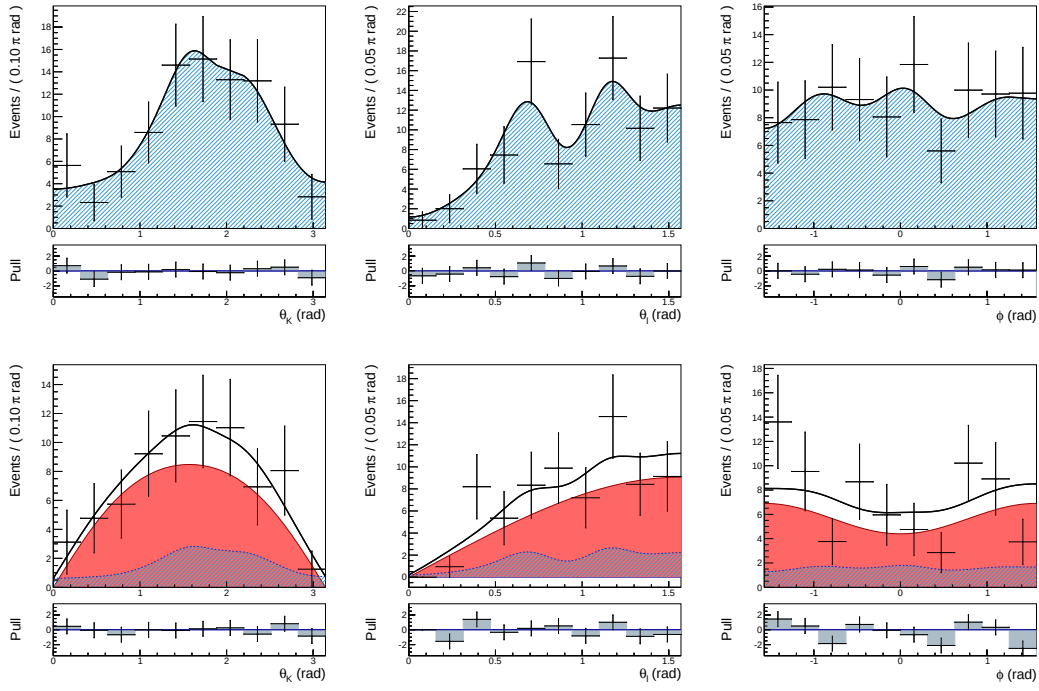


Figure A.22: Projections for the fit result of P'_8 in bin 4. Fit to the M_{bc} sideband for the determination of the background shape (top) and signal region (bottom) are displayed. Combinatorial (dashed blue), signal (red filled) and total (solid) fit distributions are superimposed on the data points.

B. Appendix $B^+ \rightarrow K^+ \tau^+ \tau^-$

B.1 Variables for the Classifier

All variables, which we can use for classification between signal and background are displayed in figs. B.1 and B.2. The variables are transformed to a flat distribution, where the x-axis corresponds approximately to the percentile of the distribution. The variables which are used for the analysis are explained in section 10.3.

B.2 Data MC Comparison

Data MC comparison for all variables is performed. The result is shown in figs. B.3 and B.4. Bad agreement is observed in variables related to missing mass and energy. Also, numbers of missing γ and K_L differ. Surprisingly, there is a discrepancy in the ratio for the zeroth and second Fox Wolfram moment R_2 . This variable is useful to suppress background from continuum events. But as the full reconstruction with continuum suppression is used, it is not problematic to remove this variable from the trainings set.

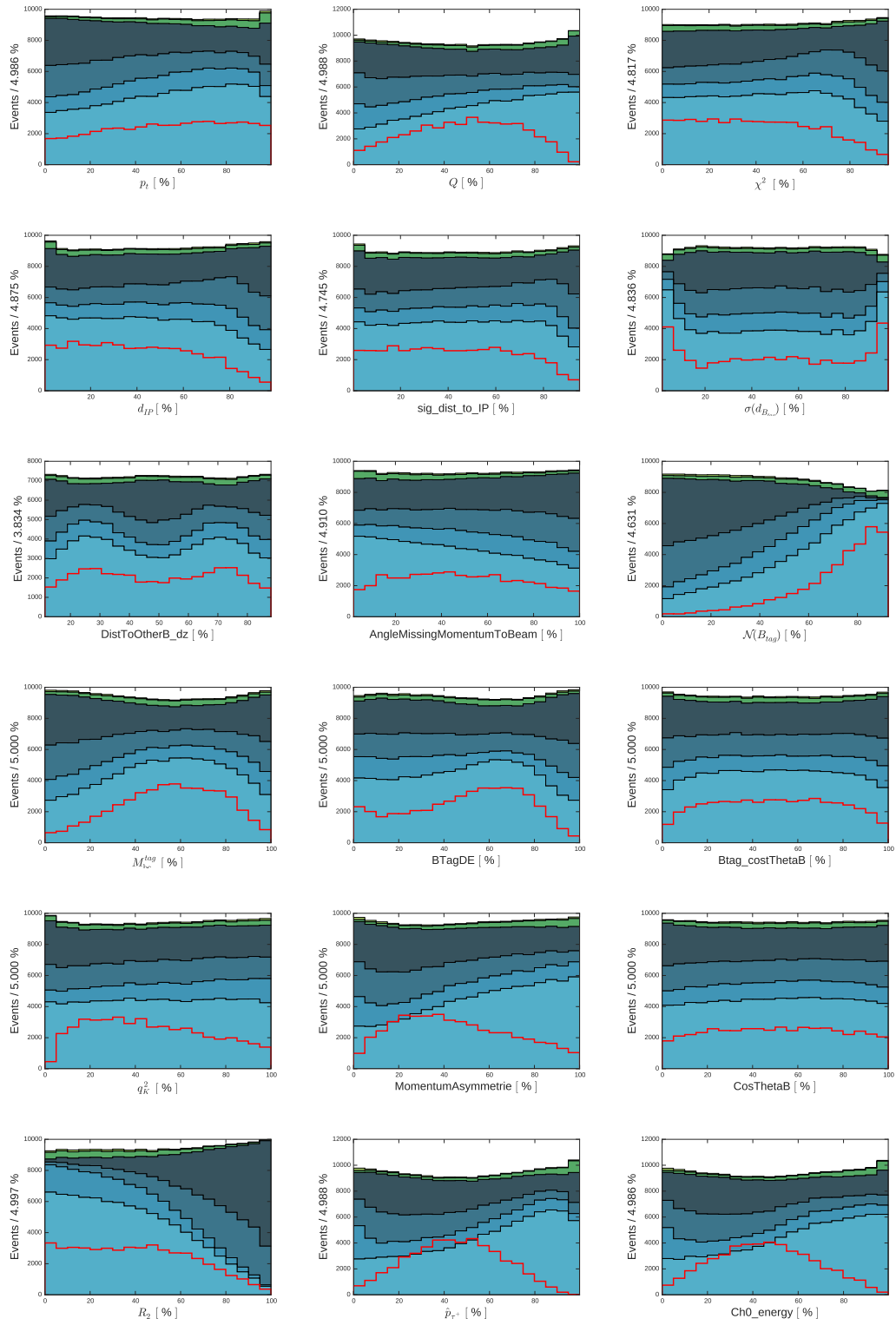


Figure B.1: Input variables for the boosted decision tree.

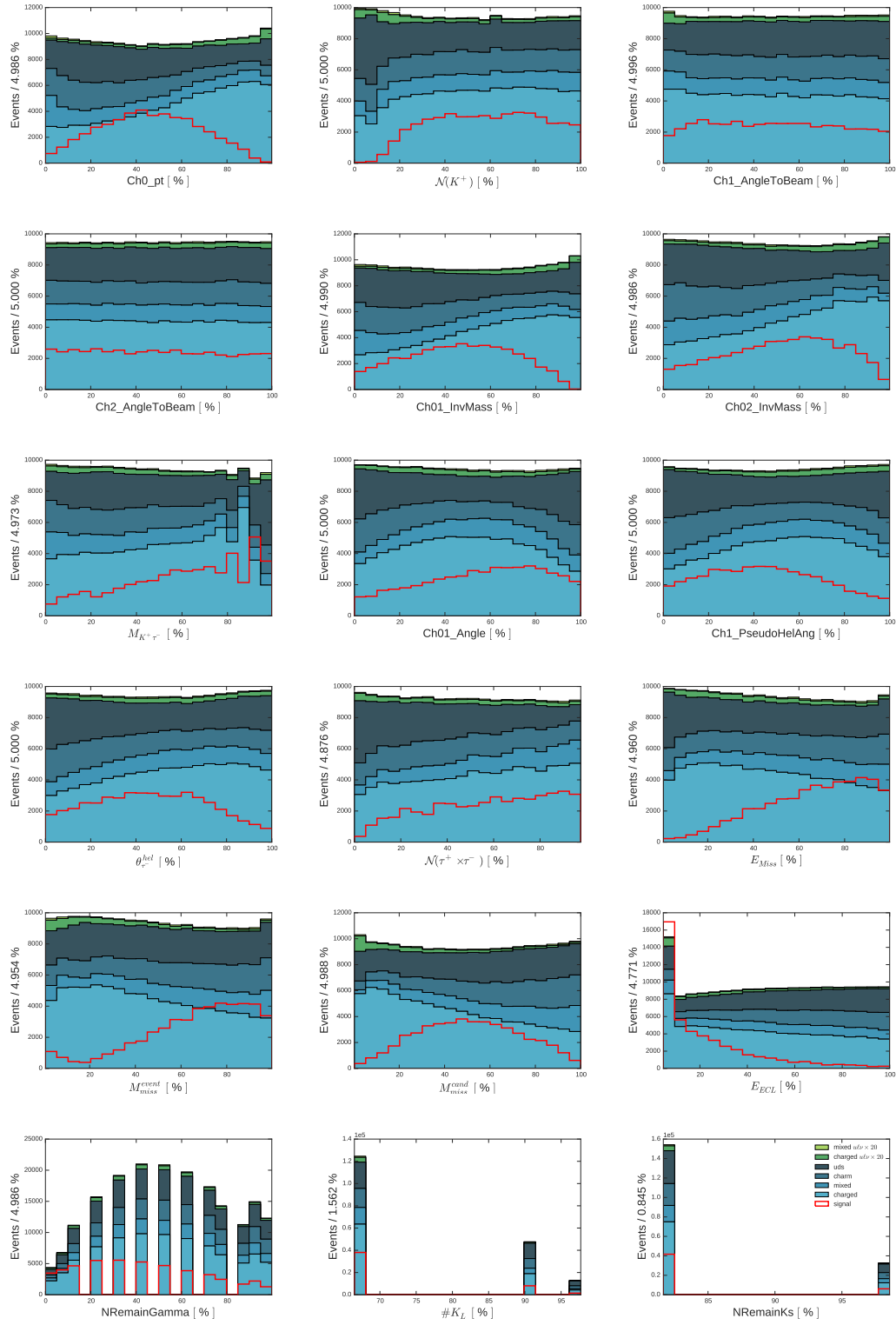


Figure B.2: Input variables for the boosted decision tree.

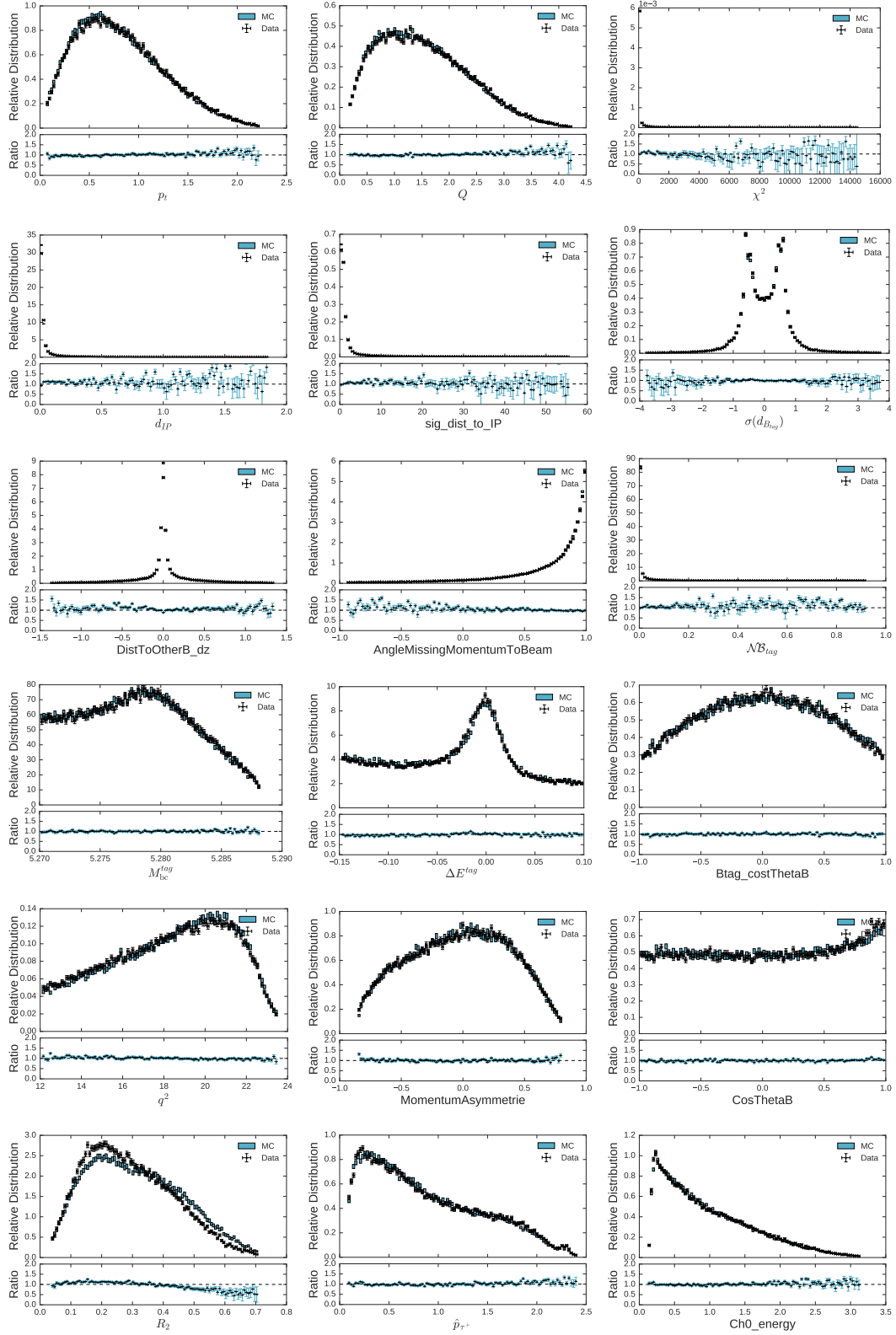
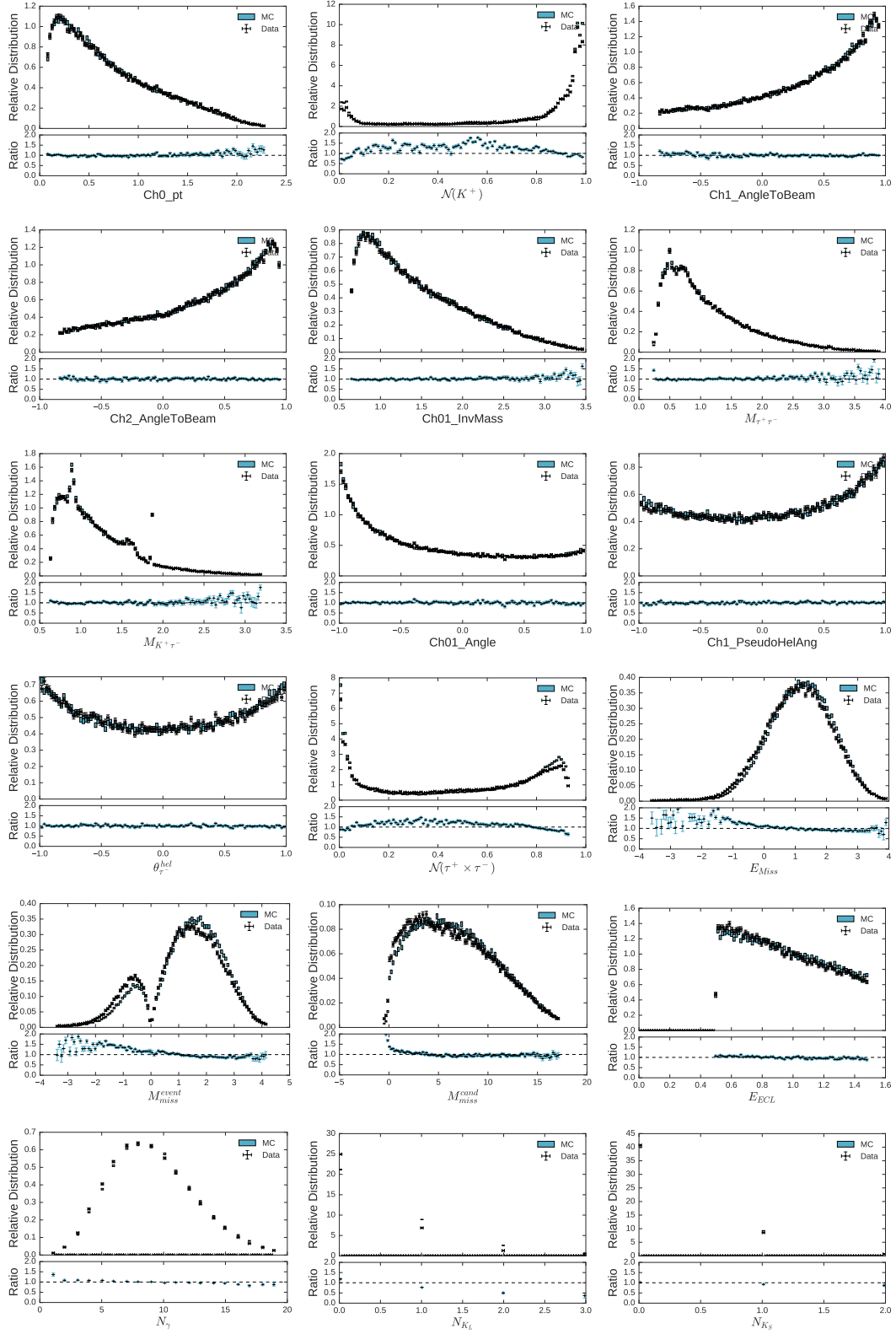


Figure B.3: Data vs. MC comparison in the E_{ECL} sideband.


 Figure B.4: Data vs. MC comparison in the E_{ECL} sideband.

Danksagungen

Diese Doktorarbeit wäre ohne die Hilfe vieler Menschen nicht möglich gewesen. Ich möchte mich daher ganz herzlich bei allen Personen bedanken die mir diese Arbeit ermöglicht haben und mich tatkräftig unterstützt haben.

Mein besonderer Dank geht an Carsten Niebuhr für viele lehrreiche Diskussionen, Rat, Unterstützung und die Möglichkeit diese Doktorarbeit in der Belle Gruppe am Desy durchzuführen.

Bei Prof. Caren Hagner bedanke ich mich ganz herzlich für die Betreuung der Arbeit als Gutachterin. And I thank Dr. Alexandre Glazov for taking over the position as second evaluator of the thesis.

Especially I want to thank Sergey Yashenkov for the help and support with my analysis and for the organization within the collaboration! Ebenso bedanke ich mich bei Jo-Frederik Krohn und Leonhard Koch, die als Summer Studenten am Desy an den Analysen teilgenommen haben.

I want to thank the whole Desy Belle group for the nice atmosphere and help, whenever it was needed. Special thanks deserve Serguei Levonian, Sergey Yashenkov, Carsten Niebuhr, Wulfrin Bartel, Maiko Takahashi, Ami Rostomyan, Gianluca Inguglia, Jo-Frederik Krohn and Oliver Frost for inspiring suggestions and their efforts in proofreading.

Abschließend möchte ich mich ganz herzlich bei meiner Familie sowie Elina-Maria für die Unterstützung, Motivation und Verständnis in den nicht immer leichten Phasen dieser Arbeit bedanken.