Search for $B_s \rightarrow \phi \pi^0$ Decays at the Belle Experiment

Master Thesis of

Fabian Fichter

At the Department of Physics
Institute of Experimental Nuclear Physics (IEKP)

Reviewer: Prof. Dr. Michael Feindt
Second reviewer: Prof. Dr. Thomas Müller
Advisor: Dr. Pablo Goldenzweig

Duration: October 2015 – October 2016
## Contents

1. Introduction ........................................ 1

2. The Belle Experiment .............................. 3
   2.1. The KEKB Accelerator .......................... 3
   2.2. The Belle Detector ............................ 3

3. Data Analysis Techniques ......................... 7
   3.1. Monte Carlo Methods .......................... 7
   3.2. Boosted Decision Trees ........................ 7
   3.3. Kernel Density Estimation ..................... 8

4. Reconstruction ..................................... 9
   4.1. Decay Chain .................................. 9
   4.2. Preselection .................................. 11

5. Continuum Suppression ............................ 15

6. Best Candidate Selection .......................... 19

7. Reconstruction Efficiency and Yield Estimation 21

8. Fit Model .......................................... 23
   8.1. Signal ....................................... 23
   8.2. Continuum Background ......................... 25
   8.3. B\_B, and non-B\_B Backgrounds ............... 25
   8.4. Peaking Backgrounds ............................ 25
   8.5. Summary ..................................... 26

9. Ensemble Tests .................................... 29

10. Control Sample .................................... 31

11. Outlook ........................................ 33

12. Summary and Conclusion .......................... 35

Bibliography ........................................ 37

List of Figures ....................................... 40

List of Tables ....................................... 41

Appendix ............................................. 43
   A. Distributions of Training Variables ............ 43
B. Distributions of Training Variables for the Control Sample . . . . . . . . . . . . . . . . . . . 46
1. Introduction

At present, the theoretical predictions of the Standard Model (SM) are in very good agreement with results from particle physics experiments. However, the SM cannot consistently explain phenomena like gravity, dark matter, and the asymmetry between matter and antimatter in the observable universe. Therefore, it is desirable to perform measurements that are sensitive to deviations between the SM and possible New Physics (NP) models that could explain some of these phenomena.

Decays dominated by electroweak penguin processes are sensitive to several models of physics beyond the SM. One of these decays is $B_s \rightarrow \phi \pi^0$ shown in Fig. 1.1.

Figure 1.1.: Electroweak penguin diagram of $B_s \rightarrow \phi \pi^0$ [1].

Theoretical predictions of its branching fraction within the SM are $1.6^{+1.1}_{-0.3} \cdot 10^{-7}$ in the framework of QCD factorization [1] and $(1.94 \pm 1.14) \cdot 10^{-6}$ in the framework of flavor symmetry [2]. The branching fraction could be enhanced by NP particles, e.g., $Z'$ or SUSY particles, by up to an order of magnitude [1].

Both the tension between perturbative and non-perturbative theoretical predictions within the SM and the potential NP effects make $B_s \rightarrow \phi \pi^0$ an interesting decay to study experimentally.

At the Belle experiment an integrated luminosity of $(121.4 \pm 0.8) \text{fb}^{-1}$ has been collected at the $\Upsilon(5S)$ resonance corresponding to $(6.53 \pm 1.07) \cdot 10^6 B_s^{(*)} \overline{B}_s^{(*)}$ pairs [3].
In this work we perform a blind analysis on simulated data equivalent to the full Belle $\Upsilon(5S)$ dataset as a preparation for the extraction of results from the real dataset. The analysis is performed with basf2 (Belle Analysis Framework 2) [4, 5], the software framework developed for the Belle II experiment, which is planned to start operation in 2018.
2. The Belle Experiment

This chapter gives a brief introduction to the setup of the Belle experiment at KEK in Tsukuba, Japan. Based on Refs. [6–8] it describes the KEKB accelerator, a B-factory, and the Belle detector, located at the collision point of KEKB. For a more detailed description and technical details, see the mentioned references.

2.1. The KEKB Accelerator

The KEKB accelerator is a circular $e^+e^-$ collider designed for experiments at the luminosity frontier. As shown in Fig. 2.1 it consists of two rings called high-energy ring (HER) and low-energy ring (LER), where electrons and positrons are accelerated separately to 8 GeV and 3.5 GeV, respectively. This gives a center of mass energy of 10.58 GeV allowing resonant production of $\Upsilon(4S)$, which mainly decays into pairs of $B$ mesons, because its mass is slightly above the production threshold of $2m_B$. The energy asymmetry of the beams leads to a boosted center-of-mass system (CMS) which allows measurements of $B$ meson decay times. These are important for CP violation measurements.

For some runs the beam energies were adjusted to collect data at other $\Upsilon$ resonances or to do energy scans. During these runs an integrated luminosity of $121.4 \text{ fb}^{-1}$ at the $\Upsilon(5S)$ resonance has been collected. Since the $\Upsilon(5S)$ mass exceeds the production threshold of $B_s(\ast)$ meson pairs, the data of these runs can be used to analyze the physics of $B_s$ decays.

2.2. The Belle Detector

In the following, we briefly describe the components of the Belle detector shown in Figs. 2.2 and 2.3.

**Magnetic field and iron yoke**

A magnetic field is required to measure the momentum and charge of particles and is provided by a 1.5 T superconducting solenoid. The solenoid is surrounded by a multi-purpose iron yoke. Besides functioning as a return yoke for the magnetic field, it also serves as an absorber material for the KLM and supports other detector components.

**Silicon vertex detector (SVD)**

The purpose of the SVD is to measure the vertex positions of $B$ meson decays,
Figure 2.1.: Setup of the KEKB accelerator rings.

Figure 2.2.: Central vertical cross-section of the Belle detector.
2.2. The Belle Detector

Figure 2.3.: Perspective illustration of the Belle detector.

especially for the observation of CP asymmetries. Therefore, the double-sided silicon-strip detectors are placed very close to the collision point just outside of the beampipe and must hence withstand much radiation.

Central drift chamber (CDC)
The CDC is designed to track charged particles and to provide information for their identification by measuring the energy loss $dE/dx$. It has 50 cylindrical layers, and in total there are 8400 drift cells with varying size, optimized using beam test results. It uses a helium-ethane gas mixture to minimize multiple Coulomb scattering while retaining a good $dE/dx$ resolution.

Time-of-flight counter (TOF)
The TOF uses plastic scintillators to distinguish between pion and kaon tracks with momenta below about 1.2 GeV/c. Additionally, because of its good time resolution of 100 ps it is also used for triggering.

Aerogel Cherenkov counter (ACC)
The ACC provides additional information for particle identification, especially for the discrimination between $\pi^\pm$ and $K^\pm$ in the momentum range of $1.2 \text{ GeV}/c < p < 3.5 \text{ GeV}/c$. In total it has 1188 silica aerogel threshold Cherenkov counter modules, where the refractive indices of the aerogels are between 1.01 and 1.03. The Cherenkov light is detected by photo-multiplier tubes (PMTs).

Electromagnetic calorimeter (ECL)
The ECL uses CsI(Tl) (thallium activated caesium iodide) crystals as scintillation material and silicon photodiodes to detect the scintillation light. It was mainly designed to accurately measure the energy and position of photons with energies up to 4 GeV, but also provides important information for electron identification.

$K_L^0$ and $\mu$ detection system (KLM)
The purpose of the KLM is to identify $K_L^0$ mesons and muons. There, iron plates serve as an absorber material, and glass-electrode-resistive plate counters (RPCs) are used for the detection of charged particles, with absorber material and detectors arranged in alternating layers.
Extreme forward calorimeter (EFC)

The EFC is a calorimeter that covers the region of small angles relative to the beampipe and complements the ECL. Since it is close to the beampipe, it consists of radiation-hard bismuth germanium oxide (BGO) crystals.
3. Data Analysis Techniques

In this chapter, we describe statistical methods that are used to analyze the data and extract the results.

3.1. Monte Carlo Methods

If an analysis procedure is optimized on the same data the final results are extracted from, then there is the danger of unintentionally introducing biases by optimizing the analysis for the results to match the expectations. In order to avoid such biases, the analysis should be optimized on datasets independent of, but equivalent to, the dataset used for the measurement. In a so-called blind analysis [6, pp. 160-163] the optimization is performed without looking at the signal region in real data.

In order to get datasets equivalent to the real data, we need to simulate the relevant physics processes of the experiment. Since particle physics processes are inherently probabilistic, pseudorandom number generators are a central part of the simulation software. Because of that, the simulation software and the resulting data are usually referred to as Monte Carlo (MC) generators and MC data, respectively.

Analyses for the Belle experiment use EvtGen [9] to simulate the particle physics processes. Its output is the information that describes the state of all intermediate and final state particles, such as energies and momenta. This information is then processed by a detector simulation that models the interaction of particles with the detector material. In the case of Belle, GEANT3 [10] is used for this purpose.

3.2. Boosted Decision Trees

A Boosted Decision Tree (BDT) uses a forest of decision trees to classify data. At each non-terminal node of a decision tree a cut on one variable is applied, where the cuts are determined during a supervised training. Each of the decision trees is limited to a shallow depth in order to avoid overtraining, and hence only has small separation power. After the training of a decision tree, the training data is reweighted using a so-called boosting algorithm in order to focus on data that is classified incorrectly by the previous tree(s). By iteratively adding decision trees and boosting steps, a strong classifier can be built from many weak classifiers. The details of the algorithm and implementation used for this work are given in Ref. [11].
3.3. Kernel Density Estimation

When there is no phenomenologically motivated probability density function (PDF) shape for a fit component, we need a generic model that creates a smooth PDF from a data sample \((x_1, x_2, ..., x_n)\). A kernel density estimator (KDE) [12] provides a generic, non-parametric estimate for a PDF by the sum of so-called kernels, where a kernel \(K\) is a non-negative function with mean zero that integrates to one. The KDE is given by

\[
P(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right),
\]

where the bandwidth \(h\) is the smoothing parameter.

In this work we use an implementation of adaptive Gaussian KDEs in RooFit [13]. This specific KDE uses standard normal distributions as kernels and adapts the bandwidth parameter \(h\) for each \(x_i\).
4. Reconstruction

For this analysis we use six streams of centrally produced uds ($e^+e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}$), charm ($e^+e^- \rightarrow c\bar{c}$), bsbs ($e^+e^- \rightarrow \Upsilon(5S) \rightarrow B^*_s \bar{B}^*_s$), and nonbsbs ($e^+e^- \rightarrow \Upsilon(5S) \rightarrow B^*(\pi), \Upsilon(4S)\gamma$) background MC events, where one stream is equivalent to the full Belle $\Upsilon(5S)$ on-resonance dataset of experiments 43, 53, 67, 69, and 71. Additionally, we have generated one million MC events for each of the three signal channels and the peaking backgrounds $B_s \rightarrow \phi\eta$, $B_s \rightarrow \phi K^0_S$, and $B_s \rightarrow K^+K^-\pi^0$.

For this analysis we use basf2 (Belle Analysis Framework 2) [4], which is being developed for the upgrade of the Belle experiment. Therefore, the above mentioned MC data is converted using the Belle-to-Belle II (B2BII) conversion tool, a module of basf2.

4.1. Decay Chain

At the Belle experiment, $B_s$ and $\bar{B}_s$ are produced pairwise via the decay of resonantly produced $\Upsilon(5S)$. Since not only $\Upsilon(5S) \rightarrow B_s\bar{B}_s$, but also $\Upsilon(5S) \rightarrow B^*_s\bar{B}^*_s$ and $\Upsilon(5S) \rightarrow B^0_s\bar{B}^0_s$ are kinematically allowed, we have three decay modes containing $B^*_s$, with production fractions $f_{B^*_s, B_s} = (87.0 \pm 1.7)\%$ and $f_{B^*_s, \bar{B}_s} = (7.3 \pm 1.4)\%$ [14]. $B^*_s$ dominantly decay into $B_s$ by radiating a photon with energy equal to the mass difference between $B^*_s$ and $B_s$, which is $48.7^{\pm2.7}_{-2.1}$ MeV [15]. These photons do not carry enough energy to be reconstructed efficiently, so the three decay modes cannot be distinguished by their final state particles. However, since the photon is not taken account of in the reconstruction, the decay modes are separated in the beam-constrained mass $M_{bc}$ and $\Delta E = E\gamma - E_B$. Here $p_B^*$ and $E_B^*$ are the momentum and energy of the reconstructed $B_s$ candidate, and $E_B = \sqrt{s}/2$ the beam energy, each calculated in the center-of-mass frame. The parametrized peak values of $M_{bc}$ and $\Delta E$ for the three cases are given in Table 4.1. Fig. 4.1 visualizes the different peak values of the three channels with a scatter plot of signal MC events in the $M_{bc}-\Delta E$ plane.

In this work we search for $B_s \rightarrow \phi\pi^0$ decays where $\phi \rightarrow K^+K^-$ and $\pi^0 \rightarrow \gamma\gamma$, since these are most frequent decay modes of $\phi$ and $\pi^0$, with branching fractions of $f_{K^+K^-} = (48.9 \pm 0.5)\%$ and $f_{\gamma\gamma} = (98.823 \pm 0.034)\%$, respectively [15]. This means that we have two charged kaons and two photons in the final state.
Table 4.1.: Parametrized peak values of $M_{bc}$ and $\Delta E$ for each production channel [16].

<table>
<thead>
<tr>
<th>Channel</th>
<th>$M_{bc}$</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_s^<em>\bar{B}_s^</em>$</td>
<td>$m_{B_s^*}$</td>
<td>$\sqrt{E_B^* - \left(\frac{m_{B_s^<em>}^2 - m_{B_s}^2}{4E_B^</em>}\right)^2}$</td>
</tr>
<tr>
<td>$B_s^*\bar{B}_s$</td>
<td>$\sqrt{\frac{m_{B_s}^2 + m_{B_s^<em>}^2}{2} - \left(\frac{m_{B_s}^2 - m_{B_s^</em>}^2}{4E_B^*}\right)^2}$</td>
<td>$-\frac{m_{B_s}^2 - m_{B_s^<em>}^2}{4E_B^</em>}$</td>
</tr>
<tr>
<td>$B_s\bar{B}_s$</td>
<td>$m_{B_s}$</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 4.1.: Scatter plot of signal MC events in the $M_{bc}$-$\Delta E$ plane. The production channels are weighted according to their branching fractions.
4.2. Preselection

In this section, we give the requirements that are used to identify candidates of a $B_s \rightarrow \phi \pi^0$ decay. We start with the final state particles and move up to $B_s$ via the intermediate states $\phi$ and $\pi^0$.

4.2.1. Final State Particles

The decay channel we study has photons in the final state, hence we select photon candidates from reconstructed ECL clusters. To reduce the background, we require a minimum photon energy of $E_\gamma > 50 \text{ MeV}$ for clusters in the barrel ECL and $E_\gamma > 100 \text{ MeV}$ in the endcap ECL, where the barrel ECL covers the polar angle region $32.2^\circ < \theta < 128.7^\circ$, and the two endcaps cover $12.4^\circ < \theta < 31.4^\circ$ and $130.7^\circ < \theta < 155.1^\circ$ [8]. The $E_\gamma$ distribution of signal MC events is shown in Fig. 4.2.

In addition to the photons, there are two kaons in the final state. These are reconstructed from charged tracks. The tracks are required to satisfy $dr < 0.2 \text{ cm}$ and $|dz| < 4 \text{ cm}$, where $dr$ and $dz$ are the closest distances of a track from the interaction point in radial and beam direction, respectively. They are also referred to as impact parameters. Furthermore, we demand that the likelihood ratio of kaon candidates must satisfy $L_{K/\pi} > 0.6$. The likelihood ratio is calculated from the particle likelihoods reported by the CDC, TOF, and ACC using

$$L_{K/\pi} = \frac{L_{K}^{\text{CDC}} \cdot L_{K}^{\text{TOF}} \cdot L_{K}^{\text{ACC}}}{L_{K}^{\text{CDC}} \cdot L_{K}^{\text{TOF}} \cdot L_{K}^{\text{ACC}} + L_{\pi}^{\text{CDC}} \cdot L_{\pi}^{\text{TOF}} \cdot L_{\pi}^{\text{ACC}}} \quad (4.1)$$

as given in [6, p. 68]. Additionally, we apply the cut $p_T > 0.1 \text{ GeV/c}$ on the transverse momentum of kaon candidates.

4.2.2. $\phi$ and $\pi^0$ Reconstruction

The reconstructed photons from $\pi^0 \rightarrow \gamma \gamma$ decays are expected to have an invariant mass close to the $\pi^0$ mass $m_{\pi^0} = (134.9766 \pm 0.0006) \text{ MeV/c}^2$ [15], so we apply the cut
0.10 GeV/c² < M_{\gamma\gamma} < 0.16 GeV/c² to all \pi^0 candidates. Fig. 4.3 shows the distribution of M_{\gamma\gamma} for signal MC. The absolute value of the \pi^0 candidates’ three-momentum is required to satisfy \( p > 0.1 \text{ GeV}/c \).

The \( \phi \) candidates are reconstructed by combining a pair of reconstructed kaons with opposite charge. Only candidates with an invariant mass in a window around the \( \phi \) mass \( m_\phi = (1019.461 \pm 0.019) \text{ MeV}/c^2 \) [15] pass the selection. The window is chosen to be 1.005 GeV/c² < M_{K^+K^-} < 1.035 GeV/c². The signal MC M_{K^+K^-} distribution is shown in Fig. 4.4.

Figure 4.3.: Invariant mass distribution of \( \pi^0 \to \gamma\gamma \) candidates for signal MC events.

Figure 4.4.: Invariant mass distribution of \( \phi \to K^+K^- \) candidates for signal MC events.
4.2.3. $B_s$ mesons

The $\phi$ and $\pi^0$ candidates that pass the cuts mentioned above are then combined to $B_s$ candidates. Considering Table 4.1 with the nominal masses $m_{B_s^*} = 5415.4^{+2.4}_{-2.1}$ MeV/$c^2$ and $m_{B_s} = (5366.77 \pm 0.24)$ MeV/$c^2$ [15], we decide that only candidates satisfying $M_{bc} > 5.30$ GeV/$c^2$ and $-0.4$ GeV < $\Delta E$ < 0.1 GeV pass the selection. Plots of the respective distributions can be found in Section 8.1, where they are used to fit PDF models.

Table 4.2 summarizes all the cuts candidates must satisfy to be considered.

| $K^\pm$ | $|dr| < 0.2$ cm |
|----------|-----------------|
|          | $|dz| < 4$ cm    |
|          | $p_T > 0.1$ GeV/$c$ |
|          | Number of SVD Hits $\geq 1$ |
|          | $L_{K/\pi} > 0.6$ |
| $\phi$  | $1.005$ GeV/$c^2 < M_{K^+K^-} < 1.035$ GeV/$c^2$ |
| $\gamma$ | $E_\gamma > 50$ MeV for barrel ECL |
|          | $E_\gamma > 100$ MeV for endcap ECL |
| $\pi^0$ | $0.10$ GeV/$c^2 < M_{\gamma\gamma} < 0.16$ GeV/$c^2$ |
|          | $p > 0.1$ GeV/$c$ |
| $B_s$    | $M_{bc} > 5.30$ GeV/$c^2$ |
|          | $-0.4$ GeV < $\Delta E$ < 0.1 GeV |
5. Continuum Suppression

The dominant background comes from the production of light quarks via $e^+e^- \rightarrow q\bar{q}$, where $q \in \{u,d,s,c\}$. In order to reduce this continuum background, we train a multivariate classifier using event shape variables to separate $B_s$ candidates of signal and continuum background MC. This approach is promising, because the topology of continuum events is very different from events where $\Upsilon(5S) \rightarrow B_s^{(*)} \bar{B}_s^{(*)}$, since the three-momenta of the initial light quarks in continuum events are large and back-to-back. This means that there are two distinct jets in continuum events, whereas signal events, where $e^+e^- \rightarrow \Upsilon(5S) \rightarrow B_s^{(*)} \bar{B}_s^{(*)}$, are largely isotropic due to the small initial momenta of the $B_s^{(*)}$ mesons.

In order to use this information to separate continuum and signal events, we train a BDT using FastBDT [11] with the following input variables:

**Thrust axis**

The thrust axis [6, 17] is defined as the normalized vector $\vec{T}$ that maximizes the quantity

$$T = \frac{\sum_i |\vec{p}_i \cdot \vec{T}|}{\sum_i |\vec{p}_i|},$$

(5.1)

called thrust, where $p_i$ are the momenta of a set of final state particles in an event. The thrust axis of a $B_s$ candidate is defined by restricting the set of particles to the daughters of the $B_s$ candidate. Analogously, to get the thrust axis of the so-called rest of event (ROE), we only sum over the particles that do not belong to the $B_s$ candidate.

In continuum events with two jets of light quarks the ROE thrust typically is close to one, since high-eneregetic light quarks produce narrow jets due to their small mass. In contrast, events with $B_s$ mesons are almost isotropic and therefore have a smaller thrust, which is why the ROE thrust is useful for continuum suppression. Furthermore, we use $\cos \theta_B$ with the angle $\theta_B$ between the thrust axes of the $B_s$ candidate and the ROE, as well as $\cos \theta_z$, where $\theta_z$ is the angle between the thrust axis of the $B_s$ candidate and the beam axis.

**KSFW moments**

The Fox-Wolfram moments [18] $H_l$ ($l \in \mathbb{N}_0$) are defined by

$$H_l = \sum_{i,j} \frac{|\vec{p}_i| \cdot |\vec{p}_j|}{E^2_{\text{tot}}} P_l (\cos \phi_{ij}) ,$$

(5.2)
where $\phi_{ij}$ is the angle between the hadrons $i$ and $j$ with momenta $\vec{p}_i$ and $\vec{p}_j$, respectively. $E_{\text{tot}}$ is the total energy seen in an event and $P_l$ is the Legendre polynomial of order $l$. The Fox-Wolfram moments are not directly used for continuum suppression, but we use the ratio $R_2$ of the zeroth and second Fox-Wolfram moment

$$R_2 = \frac{H_2}{H_0}. \quad (5.3)$$

In addition, modified Fox-Wolfram moments [6, 19], also known as Kakuno-Super-Fox-Wolfram (KSFW) moments, are included in the training. The moments are decomposed into charged, neutral and missing particle components named $H^{sc}_l$, $H^{sn}_l$, and $H^{smo}_l$, where $l \in \{0, 2, 4\}$. The indices $s$ and $o$ indicate that the definition includes a double sum, where one runs over the daughters of the $B_s$ candidate and the other over the particles of the ROE. We also use the moments $H^{mo}_l$, which only sum over ROE particles and where $l \in \{0, 1, 2, 3, 4\}$. Additionally, we include the missing mass squared $M_{\text{miss}}^2$ and the sum of transverse energies $E_T$ in the event. The mentioned variables are explained in detail in Ref. [6, pp. 114-115].

**CLEO Cones**

The so-called CLEO cones are event shape variables that were introduced by the CLEO collaboration [6, 20]. They are a set of nine variables that describe the momentum flow in intervals of $10^\circ$ in respect to the thrust axis of the $B_s$ candidate. We exclude the first two CLEO cones because of correlations with $M_{bc}$ that could introduce a bias into the fit of the $M_{bc}$ distributions.

Fig. 5.1 shows the distributions of the BDT output $C$ for signal and background events of the training and test samples, respectively. The results of training and test samples are consistent, so there is no overtraining.

To decide on the cut value of the BDT output, we use the figure of merit (FOM)

$$\text{FOM} = \frac{\epsilon_{\text{sig}}}{\sqrt{B + \frac{\epsilon_{\text{sig}}}{2}}}, \quad (5.4)$$

**Figure 5.1.:** Continuum suppression classifier distributions of signal and background events for training and test samples, respectively.
suggested by Punzi [21], with signal efficiency $\epsilon_{\text{sig}}$ and expected number of continuum background events $B$ in the signal region, where the signal region is defined by $5.35 < M_{bc} < 5.43$ and $-0.2 < \Delta E < 0.1$ to comprise the three production channels. The parameter $a$ is the required significance in terms of standard deviations from the null hypothesis. We choose $a = 5$, whereby it should be noted that the results are not very sensitive to changes of $a$. The FOM does not depend on the signal cross section, so we can optimize it without the need to make assumptions about the $B_s \to \phi\pi^0$ branching fraction. Fig. 5.2 shows the FOM and the signal efficiency in dependence of the cut on $C$. Since there is a large slope in the signal efficiency at the point of maximal FOM, we decide to use a more loose cut at $C = 0.9$. This cut is applied to all candidates passing the selection criteria in Table 4.2. The continuum suppression retains 65.3% of signal MC events and rejects 96.1% of continuum background MC.

**Figure 5.2.** Figure of merit and signal efficiency in dependence of the cut on the continuum suppression classifier output.
6. Best Candidate Selection

Of the signal events passing all the previously mentioned selection criteria 2.3% contain more than one $B_s$ candidate. We have compared three different criteria to select the best $B_s$ candidate for events with multiple candidates:

**Smallest $\delta M_{\pi^0}$**

Since most of the events with multiple candidates are due to multiple $\pi^0$ candidates, choosing the candidate with smallest $\pi^0$ mass discrepancy

$$\delta M_{\pi^0} = |M_{\gamma\gamma} - m_{\pi^0}|$$

should be a suitable criterion for best candidate selection.

**Smallest mass $\chi^2$**

In order to be able to select the correct candidate when there are multiple $\phi$ candidates, we can slightly extend the first method and choose the candidate with smallest mass $\chi^2$, where

$$\chi^2 = \left( \frac{M_{\gamma\gamma} - m_{\pi^0}}{\sigma_{\gamma\gamma}} \right)^2 + \left( \frac{M_{K^+K^-} - m_{\phi}}{\sigma_{K^+K^-}} \right)^2.$$  \hfill (6.2)

**Largest BDT output**

For a more elaborate best candidate selection, we can train a BDT to separate true and false $B_s$ candidates using signal MC events. We then choose the best candidate by using the candidate with the largest BDT classifier output. As mentioned above, events with more than one $B_s$ candidate usually arise from multiple $\pi^0$ candidates, and thus the choice of BDT inputs is geared to choosing the correct $\pi^0$ candidate. The inputs are the photon energies $E_\gamma$, the ratios E9E25 of the energies of the central 3x3 over the central 5x5 cells of the photons’ ECL clusters, the invariant mass and momentum of the $\pi^0$ candidate $m_{\gamma\gamma}$ and $p_{\gamma\gamma}$, and the helicity angle $\cos \theta_H^{\pi^0\rightarrow\gamma\gamma}$.

Applying these criteria to signal MC events gives the results summarized in Table 6.1. As you can see, the BDT yields the best separation of the considered selection criteria. Fig. 6.1 compares its performance on the training and test samples. There is no indication of overtraining, so we use the BDT for best candidate selection. Unless stated otherwise any further mention of best candidate selection refers to the BDT.
Table 6.1.: Correctly reconstructed events (in %) for the subset of events with multiple candidates and in total, including events with only one candidate, for three different selection methods and each production channel.

<table>
<thead>
<tr>
<th>Channel</th>
<th>evts. w/ mult. cands. (%)</th>
<th>total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\delta M_{x0}$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>$B^+_s \overline{B}^+_s$</td>
<td>78.2</td>
<td>80.0</td>
</tr>
<tr>
<td>$B^+_s \overline{B}^-_s$</td>
<td>78.2</td>
<td>79.8</td>
</tr>
<tr>
<td>$B^-_s \overline{B}^-_s$</td>
<td>76.1</td>
<td>78.2</td>
</tr>
</tbody>
</table>

Figure 6.1.: Best candidate selection classifier distributions of true and false $B_s$ candidates for training and test samples, respectively.
In this chapter, we use the results of applying the above mentioned reconstruction and selection criteria to signal and background MC in order to estimate the expected number of events in the real data set. For the estimates of continuum, bsbs, and nonbsbs backgrounds we average the results of six streams of generic MC.

The expected yields of signal and peaking backgrounds are calculated using the reconstruction efficiencies of one million MC events. By way of example, we carry out the calculation of the signal yield:

\[ Y_{\phi\pi^0} = N_{B_s^{(*)}B_s^{(*)}} \cdot B_{B_s^{(*)}\rightarrow\phi\pi^0} \cdot \prod_i B_i \cdot \epsilon_{\text{PID}} \cdot \sum_j \epsilon_j f_j, \]  

(7.1)

where

- \( Y_{\phi\pi^0} \) is the expected signal yield.
- \( N_{B_s^{(*)}B_s^{(*)}} \) is the number of produced \( B_s^{(*)}\overline{B}_s^{(*)} \) pairs.
- \( B_{B_s^{(*)}\rightarrow\phi\pi^0} \) is the signal branching fraction, where we use the range of theoretical SM predictions.
- \( B_i \) are the branching fractions of the daughter decays \( \phi \rightarrow K^+K^- \) and \( \pi^0 \rightarrow \gamma\gamma \).
- \( \epsilon_{\text{PID}} \) is the correction factor on the efficiency of the cut on \( L(K/\pi) \) to account for differences between MC and data.
- \( \epsilon_j \) are the reconstruction efficiencies of signal MC for the three production channels \( B_s^{(*)}\overline{B}_s^{(*)} \), \( B_s^{(*)}\overline{B}_s \), and \( B_s\overline{B}_s \).
- \( f_j \) are the production fractions of the three signal channels.

The resulting yields of the calculations explained above are summarized in Table 7.1. It should be noted that events where the MC generator information contains a signal or peaking background decay are removed from the bsbs component.

The range of expected SM branching fractions for \( B_s \rightarrow \phi\pi^0, B_s \rightarrow \phi\eta, \) and \( B_s \rightarrow \phiK_s^0 \) are taken from Ref. [2]. As we do not know of a theoretical prediction for \( B_s \rightarrow K^+K^-\pi^0 \), we cannot calculate an estimated yield. However, since the reconstruction efficiency is
Table 7.1.: Efficiencies and estimated yields calculated from MC events for signal components, continuum background, $B_s \bar{B}_s$ and non-$B_s \bar{B}_s$ backgrounds, and peaking backgrounds.

<table>
<thead>
<tr>
<th>type</th>
<th>$\varepsilon$ (%)</th>
<th>expected yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_s^* \bar{B}_s^*$</td>
<td>22.17</td>
<td>0.05 - 1.17</td>
</tr>
<tr>
<td>$\phi\pi^0$ $B_s^* \bar{B}_s$</td>
<td>22.14</td>
<td>0.00 - 0.10</td>
</tr>
<tr>
<td>$B_s \bar{B}_s$</td>
<td>22.05</td>
<td>0.00 - 0.07</td>
</tr>
<tr>
<td>evtgen-uds,-charm</td>
<td>790.67</td>
<td></td>
</tr>
<tr>
<td>evtgen-bsbs</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>evtgen-nonbsbs</td>
<td>5.17</td>
<td></td>
</tr>
<tr>
<td>$\phi\eta$</td>
<td>0.30</td>
<td>0.00 - 0.03</td>
</tr>
<tr>
<td>$\phi K_S^0$</td>
<td>0.61</td>
<td>0.00 - 0.01</td>
</tr>
<tr>
<td>$K^+ K^- \pi^0$</td>
<td>0.03</td>
<td>$&lt;&lt; 1$</td>
</tr>
</tbody>
</table>

very low, especially due to the cut on $M_{K^+ K^-}$, its branching fraction would have to be orders of magnitude larger than the signal branching fraction in order to give a noticeable contribution.

Since the expected yields of the peaking backgrounds are orders of magnitude smaller than the signal contribution, the peaking backgrounds are not considered in our fit and will, in a future fit to real data, be treated as systematic errors.
8. Fit Model

In order to extract the signal yield from the data, we perform an extended unbinned maximum likelihood fit \[22\]. We fit to the observables \( M_{bc} \) and \( \Delta E \), so the likelihood \( L \) is given by

\[
L = e^{-\sum_j Y_j \cdot \prod_i P(M_{bc}^i, \Delta E^i)} = e^{-\sum_j Y_j \cdot \prod_i \left( \sum_j Y_j p_j(M_{bc}^i, \Delta E^i) \right)},
\]  
\[ (8.1) \]

where \( P \) is the total PDF, \( p_j \) its components, \( Y_j \) the yields of the components, and \( N \) the total number of events. The initial plan was to fit more dimensions, but due to fit instabilities and biases, we found ourselves constrained to switch to a two-dimensional fit in order to obtain a robust fit procedure.

The PDFs of all components are factorized into one-dimensional PDFs, i.e.,

\[
P_j(M_{bc}, \Delta E) = P_j(M_{bc}) \cdot P_j(\Delta E).
\]  
\[ (8.2) \]

The following sections explain which PDF shapes are used to model the \( M_{bc} \) and \( \Delta E \) distributions of the respective components and show the fits to MC data.

8.1. Signal

Since there are three signal channels, the signal PDF \( P_{sig} \) is the sum of three components weighted by their respective production fractions \( f_i \) and efficiencies \( \epsilon_i \):

\[
P_{sig} = \frac{1}{N} \left( f_{B_d^0 \bar{B}_s^0} \epsilon_{B_d^0 \bar{B}_s^0} P_{B_d^0 \bar{B}_s^0} + f_{B_s^+ \bar{B}_d^-} \epsilon_{B_s^+ \bar{B}_d^-} P_{B_s^+ \bar{B}_d^-} + \left( 1 - f_{B_d^0 \bar{B}_s^0} - f_{B_s^+ \bar{B}_d^-} \right) \epsilon_{B_d^0 \bar{B}_s^0} P_{B_d^0 \bar{B}_s^0} \right),
\]  
\[ (8.3) \]

with the normalization factor

\[
N = f_{B_s^+ \bar{B}_d^-} \epsilon_{B_s^+ \bar{B}_d^-} + f_{B_d^0 \bar{B}_s^0} \epsilon_{B_d^0 \bar{B}_s^0} + \left( 1 - f_{B_d^0 \bar{B}_s^0} - f_{B_s^+ \bar{B}_d^-} \right) \epsilon_{B_d^0 \bar{B}_s^0}.
\]  
\[ (8.4) \]

For each of the three components we use a double Gaussian with common mean to describe the \( M_{bc} \) distribution, and the sum of a Crystal Ball function \[23\] and a Gaussian with common mean to describe the \( \Delta E \) distribution. The PDFs are fitted to the signal MC samples, and their resulting one-dimensional projections are shown in Fig. 8.1. We use the
Figure 8.1.: Signal MC distributions in $M_{bc}$ and $\Delta E$ with the respective fitted PDFs for the production channels $B_s^* B_s^*$, $B^*_s B_s$, and $B_s B_s$ (from top to bottom).
previously mentioned production fractions from Ref. [14], and the efficiencies from Table 7.1. The parameter values of the components’ PDFs are determined from the fits to the signal MC samples. For a fit to real data they will be corrected by differences between data and MC obtained from a control sample described in Chapter 10. All parameters of the signal PDF, including the production fractions, are fixed in fits of the total PDF.

8.2. Continuum Background

The continuum background is modeled by an ARGUS function [24] for the $M_{bc}$ distribution and a first order Chebyshev polynomial for the $\Delta E$ distribution. The results of the fit to continuum background MC are shown in Fig. 8.2.

![Figure 8.2: $M_{bc}$ and $\Delta E$ distributions as well as the fitted PDFs for continuum background MC events.](image)

All parameters of the PDF are floated except for the cutoff of the ARGUS function, which is fixed to the beam energy in the center-of-mass frame $E_B = 5.434$ GeV/c$^2$.

8.3. $B_s B_s$ and non-$B_s B_s$ Backgrounds

The expected yields of the bsbs and nonbsbs backgrounds are low according to Table 7.1 and consist of many different decays, so we combine them into one component. As stated in Chapter 7, the peaking backgrounds are removed from the bsbs component. Since there is no obvious physically motivated PDF shape we could use to model the distributions, we decide to use a non-parametric, generic PDF. Both the $M_{bc}$ and $\Delta E$ distributions are modeled with a one-dimensional KDE (see Section 3.3). The KDEs and the MC distributions they are generated from, are illustrated in Fig. 8.3.

In fits of the total PDF this component is fixed to the KDE obtained from the fit to MC data.

8.4. Peaking Backgrounds

As mentioned before, the peaking backgrounds are not included in the fit, because the expected yields are very small. Nevertheless, the $M_{bc}$ and $\Delta E$ distributions of the peaking backgrounds for the channel with the largest production fraction ($B_s \overline{B}_s^*$) are illustrated in Fig. 8.4 together with suitable fitted PDFs as they will be used in determining the systematic error of these components. The respective PDF models are displayed in Table 8.1.
8. Fit Model

Figure 8.3.: $B_s$ and non-$B_s$ distributions and PDFs in $M_{bc}$ and $\Delta E$.

<table>
<thead>
<tr>
<th>$M_{bc}$</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_s \rightarrow \phi \eta$</td>
<td>bifurcated Gaussian + Gaussian, 2nd order Chebyshev polynomial</td>
</tr>
<tr>
<td>$B_s \rightarrow \phi K_S^0$</td>
<td>Gaussian KDE, Crystal Ball + Gaussian</td>
</tr>
<tr>
<td>$B_s \rightarrow K^+ K^- \pi^0$</td>
<td>double Gaussian, Crystal Ball + Gaussian</td>
</tr>
</tbody>
</table>

8.5. Summary

Table 8.1.: Summary of the peaking background PDFs for the fit variables $M_{bc}$ and $\Delta E$.

Table 8.2.: Summary of the PDFs used to model the shape of signal and background distributions in the fit variables $M_{bc}$ and $\Delta E$.

<table>
<thead>
<tr>
<th>$M_{bc}$</th>
<th>$\Delta E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal</td>
<td>double Gaussian, Crystal Ball + Gaussian</td>
</tr>
<tr>
<td>Continuum</td>
<td>ARGUS, 1st order Chebyshev polynomial</td>
</tr>
<tr>
<td>$B_s$, non-$B_s$</td>
<td>Gaussian KDE, Gaussian KDE</td>
</tr>
</tbody>
</table>

8.5. Summary

Table 8.2 summarizes the PDFs used to model the $M_{bc}$ and $\Delta E$ distributions of the fit components, which comprise the signal channels, the continuum background, and the $B_s$ and non-$B_s$ background.
Figure 8.4: Distributions and PDFs of the peaking backgrounds in $M_{bc}$ and $\Delta E$. From top to bottom these are $B_s \rightarrow \phi \eta$, $B_s \rightarrow \phi K_S^0$, and $B_s \rightarrow K^+ K^- \pi^0$. 
9. Ensemble Tests

In this chapter, we validate our fit procedure by performing ensemble tests. Therefore, we need many MC datasets each equivalent to the real data. Because of the limited number of available background MC, we sample the fitted background distributions to get sets of $M_{bc}$ and $\Delta E$ values. The number of background events is fluctuated according to a Poisson distribution with mean value equal to the number of expected background events. The signal candidates are sampled from generated signal MC events to account for the correlation of $M_{bc}$ and $\Delta E$.

We then apply the fit procedure described in Chapter 8 to each of these datasets. The fitted signal yield is compared to the one used to generate the dataset by calculating the pull value, which is given by

$$
\frac{Y^{\text{fit}} - Y^{\text{in}}}{\sigma^{\text{fit}}_Y},
$$

where $Y^{\text{fit}}$ and $\sigma^{\text{fit}}_Y$ are the value and uncertainty of the fitted signal yield, and $Y^{\text{in}}$ is the input signal yield. By design the pull distribution is a standard normal distribution for an unbiased fitter with correct error estimates. When we fit a Gaussian to the pull distribution of a parameter, the mean value and the standard deviation quantify the bias and the accuracy of the error estimation.

Fig. 9.1 shows the results of an ensemble test with 10000 pseudo-experiments, where the input signal yield is chosen to be $Y^{\text{in}}_{\text{sig}} = 1$. It can be seen that the fitter is stable with a moderate bias towards negative yields and slight overestimation of the fit error.

The procedure is repeated for input yields $Y^{\text{in}}_{\text{sig}} \in [0, 20]$ with 500 pseudo-experiments per input yield value. The results of this linearity test are given in Fig. 9.2. The plot of fitted yield over input yield includes a linear fit, where for an unbiased fitter we would expect slope $p_1 = 1$ and offset $p_0 = 0$. The deviation can be used to get an estimate for the systematic error of the fit.
Figure 9.1.: Distributions of fitted yield, fit error, pull, and fit likelihood obtained from 10000 pseudo-experiments with input signal yield $Y_{\text{sig}}^{\text{in}} = 1$.

Figure 9.2.: Fitted yield, fit error, and fit bias of a linearity test for input signal yields $Y_{\text{sig}}^{\text{in}} \in [0, 20]$ obtained from 500 pseudo-experiments per input yield value.
10. Control Sample

In order to be able to extract a small signal with a fit, we have to fix the signal PDF shape. Since we determine the PDF shape from MC events, we would have to account for the differences between the data and MC in the case of a fit to real data. Therefore, we examine a high-statistics channel with a similar final state that could be used as a control sample to correct the $M_{bc}$ and $\Delta E$ resolution obtained from MC. We decide to use $B_0^d \to K^*\pi^0$ from $\Upsilon(4S) \to B_0^dB_0^d$ decays at the $\Upsilon(4S)$ resonance, where $K^* \to K^+\pi^-$ and $\pi^0 \to \gamma\gamma$, with two tracks and two photons in the final state. We use six streams of background MC, each equivalent to the real data of experiments 51 and 55, to study the background.

The cuts that are applied during the reconstruction can be found in Table 10.1. Where possible the cuts are the same as for the target decay. As for $B_s \to \phi\pi^0$, we train BDTs

| Table 10.1.: Summary of preselection cuts for the control sample. |
|-------------------|-----------------|-----------------|-----------------|
| $K^+, \pi^-$      | $|dr| < 0.2 \text{ cm}$                         |
|                   | $|dz| < 4 \text{ cm}$                               |
|                   | $p_T > 0.1 \text{ GeV/c}$                            |
|                   | Number of SVD Hits $\geq 1$                          |
|                   | $L_{K/\pi} > 0.6$ for $K^+$                           |
|                   | $L_{K/\pi} < 0.4$ for $\pi^-$                        |
| $K^*$             | $0.85 \text{ GeV/c}^2 < M_{K^+K^-} < 0.93 \text{ GeV/c}^2$ |
| $\gamma$          | $E_\gamma > 50 \text{ MeV}$ for barrel ECL          |
|                   | $E_\gamma > 100 \text{ MeV}$ for endcap ECL          |
| $\pi^0$           | $0.10 \text{ GeV/c}^2 < M_{\gamma\gamma} < 0.16 \text{ GeV/c}^2$ |
|                   | $p > 0.1 \text{ GeV/c}$                              |
| $B_0^d$           | $M_{bc} > 5.20 \text{ GeV/c}^2$                      |
|                   | $-0.4 \text{ GeV} < \Delta E < 0.1 \text{ GeV}$     |

for continuum suppression and best candidate selection. The variables for the continuum suppression training are taken over from the target decay mode and can be found in Chapter 5. The training for best candidate selection differs from the $B_s \to \phi\pi^0$ case (see Chapter 6) in that the helicity angle of $\phi \to K^+K^-$ is replaced by the one of $K^* \to K^+\pi^-$. 


whereas all other training variables are the same. The distributions of the training variables for continuum suppression and best candidate selection are shown in Appendix B.

The fit procedure is also borrowed from the target decay mode. We fit the dimensions \(M_{bc}\) and \(\Delta E\) and use the same PDF shapes to model signal and continuum background, with the difference that we only have one signal channel. Fig. 10.1 shows the fit of the signal PDF to the \(M_{bc}\) and \(\Delta E\) distributions of reconstructed candidates in signal MC events.

**Figure 10.1.** \(M_{bc}\) and \(\Delta E\) signal MC distributions of the control sample as well as the fitted PDF shapes.

The fitted continuum background PDF shape is displayed in Fig. 10.2. The endpoint of the ARGUS function is set to 5.289 GeV/c\(^2\) to account for the different beam energy compared to \(\Upsilon(5S)\) data.

**Figure 10.2.** \(M_{bc}\) and \(\Delta E\) signal MC distributions of the control sample as well as the fitted PDF shapes.

A future fit of the combined PDF to \(\Upsilon(4S)\) real data will then provide corrections on the resolutions in \(M_{bc}\) and \(\Delta E\) for the \(B_s \rightarrow \phi \pi^0\) PDF. The \(M_{bc}\) and \(\Delta E\) peak values will be corrected using the study of the \(B_s \rightarrow D_s^- \pi^+\) control sample in Ref. [25].


11. Outlook

In the previous chapters, we displayed the current state of the analysis. In the following, we want to give an outlook and summarize the remaining steps that are to be done before unblinding.

Although the bsbs component is very small, we could study the $M_{K^+K^-}$ sidebands to check whether there is crossfeed from other resonances. This check can provide useful information for the estimation of the systematic error.

In order to account for the differences between MC and real data, a fit to a control sample as outlined in Chapter 10 will be performed. The resulting corrections on the resolution in $M_{bc}$ and $\Delta E$ will then be applied to the signal PDF obtained from the fits to MC samples. The calibration of the $M_{bc}$ and $\Delta E$ mean values will be taken from an Υ(5S) control sample. After the corrections of the resolutions and mean values, the signal PDF should provide a suitable model for the signal shape in real data.

Finally, the sources of systematic errors will be studied in detail. The following sources of systematic errors should be considered:

**Number of $B_s^*(s)\bar{B}_s^*(s)$ pairs**

The number of $B_s^*(s)\bar{B}_s^*(s)$ pairs in the Belle Υ(5S) dataset and the production fractions $f_{B_s^*(s)}$ are not exactly known, so they introduce a systematic uncertainty.

**Signal reconstruction efficiency**

Since the reconstruction efficiency depends on many selection criteria, this comprises several systematic errors. The efficiency of the cut on the likelihood ratio $L(K/\pi)$ to identify kaons slightly differs in MC and data, so we have to account for the uncertainty of the correction factor. Furthermore, the efficiency of the cut on the continuum suppression classifier is estimated using signal MC events, and thus we should consider the discrepancy between MC and data as a systematic uncertainty. Additionally, statistical fluctuations due to the limited number of MC events in the calculation of the signal reconstruction efficiencies result in a systematic error, which, however, is small.

**Daughter branching fractions**

The uncertainties of the branching fractions of $\phi \to K^+K^-$ and $\pi^0 \to \gamma\gamma$ also contribute as a systematic uncertainty.
PDF shape
The parameters of the signal PDF are determined from fits to MC and are corrected using control samples. Both the uncertainty of the fit and the uncertainty of the calibration have to be taken into account as a systematic uncertainty.

Fit bias
The fit stability and accuracy has been studied in Chapter 9 using ensemble tests. The fit bias obtained from the ensemble tests can be used to estimate the systematic uncertainty of the fit procedure.
12. Summary and Conclusion

In this work we have performed a blind analysis of the decay channel $B_s \rightarrow \phi\pi^0$ on MC data equivalent to the Belle $\Upsilon(5S)$ real data with an integrated luminosity of $121.4 \text{fb}^{-1}$.

After a cut-based preselection we train multivariate classifiers to separate signal and background events as well as to choose the correct candidate in case an event contains multiple candidates. We then display the fit procedure — a 2D unbinned maximum likelihood fit in $M_{bc}$ and $\Delta E$ — and perform ensemble tests to validate that the fit is stable and that there is no significant bias. Following this, we suggest $B^0_d \rightarrow K^*\pi^0$ as a control sample that could be used to correct the PDF of the target decay to account for differences between the MC simulation and real data. Finally, we outline the next steps towards unblinding the data.

This analysis was presented at the October 2016 Belle General Meeting where internal referees have been requested to begin the review process prior to fitting real data.

In summary, the present work lays the ground for the search for $B_s \rightarrow \phi\pi^0$ decays using Belle $\Upsilon(5S)$ data. A measurement of this decay could provide an interesting probe in the search for physics beyond the Standard Model.
Bibliography

[1] L. Hofer, D. Scherer, and L. Vernazza, “$B_s \rightarrow \phi \rho^0$ and $B_s \rightarrow \phi \pi^0$ as a handle on isospin-violating New physics,” JHEP, vol. 02, p. 080, 2011, 1011.6319.


[25] S. Esen, “BF measurement of $B^0_s$ Decays to $D^{(*)}_sD^{(*)}_s$ with 121.4 fb$^{-1}$,” *Belle Note 1203*. 

---

---

---
List of Figures

1.1. Electroweak penguin diagram of $B_s \rightarrow \phi \pi^0$ [1] ........................... 1

2.1. Setup of the KEKB accelerator rings ........................................... 4
2.2. Central vertical cross-section of the Belle detector ......................... 4
2.3. Perspective illustration of the Belle detector ................................ 5

4.1. Scatter plot of signal MC events in the $M_{bc}$-$\Delta E$ plane. The production channels are weighted according to their branching fractions ........................................... 10
4.2. Photon energy distribution for signal MC events .......................... 11
4.3. Invariant mass distribution of $\pi^0 \rightarrow \gamma \gamma$ candidates for signal MC events .................................................. 12
4.4. Invariant mass distribution of $\phi \rightarrow K^+ K^-$ candidates for signal MC events .................................................. 12

5.1. Continuum suppression classifier distributions of signal and background events for training and test samples, respectively .................................................. 16
5.2. Figure of merit and signal efficiency in dependence of the cut on the continuum suppression classifier output .................................................. 17

6.1. Best candidate selection classifier distributions of true and false $B_s$ candidates for training and test samples, respectively .................................................. 20

8.1. Signal MC distributions in $M_{bc}$ and $\Delta E$ with the respective fitted PDFs for the production channels $B_s \overline{B_s}^*$, $B_s^* \overline{B_s}$, and $B_s \overline{B_s}$ (from top to bottom) .................. 24
8.2. $M_{bc}$ and $\Delta E$ distributions as well as the fitted PDFs for continuum background MC events .................................................. 25
8.3. $B_s B_s$ and non-$B_s B_s$ distributions and PDFs in $M_{bc}$ and $\Delta E$. .................................................. 26
8.4. Distributions and PDFs of the peaking backgrounds in $M_{bc}$ and $\Delta E$. From top to bottom these are $B_s \rightarrow \phi \eta$, $B_s \rightarrow \phi K_S^0$, and $B_s \rightarrow K^+ K^- \pi^0$. .................................................. 27

9.1. Distributions of fitted yield, fit error, pull, and fit likelihood obtained from 10000 pseudo-experiments with input signal yield $Y_{\text{sig}}^{\text{in}} = 1$ .................................................. 30
9.2. Fitted yield, fit error, and fit bias of a linearity test for input signal yields $Y_{\text{sig}}^{\text{in}} \in [0, 20]$ obtained from 500 pseudo-experiments per input yield value .................................................. 30

10.1. $M_{bc}$ and $\Delta E$ signal MC distributions of the control sample as well as the fitted PDF shapes .................................................. 32
10.2. $M_{bc}$ and $\Delta E$ signal MC distributions of the control sample as well as the fitted PDF shapes .................................................. 32

A.1. Distributions of continuum suppression variables for signal and background (1) .................................................. 43
A.2. Distributions of continuum suppression variables for signal and background (2) .................................................. 44
A.3. Distributions of best candidate selection variables for true and false candidates in signal MC events. .............................................. 45
B.4. Distributions of continuum suppression variables for signal and background of the control sample (1). .............................................. 46
B.5. Distributions of continuum suppression variables for signal and background of the control sample (2). .............................................. 47
B.6. Distributions of best candidate selection variables for true and false candidates in signal MC events of the control sample. ......................... 48
List of Tables

4.1. Parametrized peak values of $M_{bc}$ and $\Delta E$ for each production channel [16]. 10
4.2. Summary of preselection cuts. 13

6.1. Correctly reconstructed events (in %) for the subset of events with multiple candidates and in total, including events with only one candidate, for three different selection methods and each production channel. 20

7.1. Efficiencies and estimated yields calculated from MC events for signal components, continuum background, $B_s\bar{B}_s$ and non-$B_s\bar{B}_s$ backgrounds, and peaking backgrounds. 22

8.1. Summary of the peaking background PDFs for the fit variables $M_{bc}$ and $\Delta E$. 26
8.2. Summary of the PDFs used to model the shape of signal and background distributions in the fit variables $M_{bc}$ and $\Delta E$. 26

10.1. Summary of preselection cuts for the control sample. 31
Appendix

A. Distributions of Training Variables

A.1. Continuum Suppression

Figure A.1.: Distributions of continuum suppression variables for signal and background (1).
Figure A.2.: Distributions of continuum suppression variables for signal and background (2).
A. Distributions of Training Variables

A.2. Best Candidate Selection

Figure A.3.: Distributions of best candidate selection variables for true and false candidates in signal MC events.
B. Distributions of Training Variables for the Control Sample

B.1. Continuum Suppression

Figure B.4.: Distributions of continuum suppression variables for signal and background of the control sample (1).
B. Distributions of Training Variables for the Control Sample

Figure B.5.: Distributions of continuum suppression variables for signal and background of the control sample (2).
B.2. Best Candidate Selection

Figure B.6.: Distributions of best candidate selection variables for true and false candidates in signal MC events of the control sample.
Ich versichere die Arbeit selbstständig angefertigt und alle benutzten Hilfsmittel vollständig angegeben zu haben.

Karlsruhe, Oktober 2016

(Fabian Fichter)
Danksagung

An dieser Stelle möchte ich allen danken die meine Masterarbeit ermöglicht oder mich bei ihrer Erstellung unterstützt haben.

Zuallererst danke ich Prof. Michael Feindt für die Möglichkeit meine Arbeit in seiner Arbeitsgruppe schreiben zu können.

Außerdem bedanke ich mich bei Prof. Thomas Müller für das Übernehmen des Korreferats.

Mein besonderer Dank gilt Dr. Pablo Goldenzweig, der sich als Betreuer dieser Arbeit stets Zeit für Fragen nahm.

Meinen Eltern möchte ich für all die Unterstützung während meines Studiums danken.