

Extending the Full Event Interpretation to the $\Upsilon(5S)$ Resonance

Moritz Bauer

Master's thesis

29th February 2020

Institut für Experimentelle Teilchenphysik (ETP)

Advisor: Prof. Dr. Günter Quast
Coadvisor: Prof. Dr. Florian Bernlochner

Editing time: 1st March 2019 – 29th February 2020

Erweiterung der Full Event Interpretation auf die $\Upsilon(5S)$ -Resonanz

Moritz Bauer

Masterarbeit

29. Februar 2020

Institut für Experimentelle Teilchenphysik (ETP)

Referent: Prof. Dr. Günter Quast
Korreferent: Prof. Dr. Florian Bernlochner

Bearbeitungszeit: 1. März 2019 – 29. Februar 2020

Ich versichere wahrheitsgemäß, die Arbeit selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde.

Karlsruhe, 29. Februar 2020

.....
(**Moritz Bauer**)

Contents

1. Introduction	1
2. B Factories and the Belle Detector	3
2.1. B Mesons and Υ Resonances	3
2.2. Belle Experiment	5
2.2.1. KEKB	5
2.2.2. Belle Detector	7
3. Methods and Algorithms	11
3.1. Software Setup	11
3.1.1. Belle II Analysis Software Framework and B2BII Conversion Package	11
3.2. B Meson Tagging at Belle and Belle II	12
3.2.1. Method	12
3.2.2. Full Reconstruction	14
3.2.3. Full Event Interpretation	15
3.3. Variables Used for Particle Candidates Reconstructed by the FEI	19
4. Full Event Interpretation at $\Upsilon(5S)$	25
4.1. Implementation of B_s^0 meson and its decay channels	25
4.1.1. Pre-Cuts	27
4.1.2. Decay Channel $B_s^0 \rightarrow J/\psi\phi$	28
4.1.3. Using $\Delta E'$ Instead of ΔE	28
4.1.4. Correlation between M_{bc} and Variables Used to Select B_s^0 Meson Candidates	29
4.2. Datasets Used to Train the FEI at $\Upsilon(5S)$	32
5. Validation of the Full Event Interpretation at $\Upsilon(5S)$	33
5.1. Continuum Suppression	33
5.2. Performance on Monte Carlo Data	34
5.3. Performance on Recorded Data	39
5.3.1. Unbinned Maximum Likelihood Fit in M_{bc}	40
5.3.2. Estimating the Total Tag-side Efficiency and Purity	48
5.3.3. Estimating the tag-side efficiency and purity of individual channels	51
5.3.4. Performance on Off-Resonance Data	51
5.4. Calibration attempt with $B_s^0 \rightarrow X\ell\nu$	52

6. Conclusion and Outlook	59
A. Components of fit with reduced range	61
B. Additional channels	65
C. Correlation between features and M_{bc} for trained channels	67
D. Template fits with varying number of bins	71

1. Introduction

One of the key aims of physics is to create theories that describe the behavior of nature. Among these theories, the Standard Model of particle physics has proven to be enormously successful. Using only 19 free parameters, this quantum field theory is able to describe a wide array of interactions at high energies. To this date, experiments at particle colliders have all been consistent with the Standard Model. There are, however, indications of physics beyond the Standard Model. Certain phenomena we observe, for example those at astronomical scales, require additional answers. Furthering our understanding of the Standard Model and searching for new physics beyond which answer these open questions is thus a common goal of experimental high-energy particle physics.

At particle colliders, this search can be conducted in two ways: At the energy frontier and at the precision frontier. Experiments at the energy frontier try to find particles at increasingly higher collision energies, currently the highest energies are reached by the Large Hadron Collider (LHC). The Belle Experiment, however, operates at the precision frontier. It takes advantage of the clean experimental environment provided by electron-positron colliders. Obtaining precise measurements requires sophisticated analysis tools. Among these, machine learning methods have contributed greatly to increased experimental precision. These methods can employ multiple variables better than traditional methods which require selections fine-tuned using analysts' domain knowledge. Machine learning methods allow computationally advantageous exploitation of the large parameter space in modern analyses. This makes analyses possible which would be statistically limited with traditional methods.

B factories such as the Belle Experiment aim to produce mostly B mesons, composite particles containing a bottom quark. As these are only produced in pairs, the Full Event Interpretation exploits multivariate method to allow analysts to use information from both of these mesons. While the Belle Experiment has mostly recorded data at the $\Upsilon(4S)$ bottomonium resonance, it has also recorded a smaller dataset at the $\Upsilon(5S)$ resonance. The existing Full Event Interpretation is only designed to operate at the $\Upsilon(4S)$ resonance. This thesis presents the adaptation and application of the Full Event Interpretation to the $\Upsilon(5S)$ bottomonium resonance. This will allow analyses which use all information of pairs of B_u^+ , B_d^0 or B_s^0 mesons produced in $\Upsilon(5S)$ decays.

Chapter 2 presents the bottomonium resonances and the experimental setup at the Belle Experiment. In Chapter 3, the software setup and the multivariate algorithm are presented. Chapter 4 presents the adaptation of the algorithm to allow reconstructing B_s^0 mesons. Chapter 5 covers the validation and calibration of the algorithm with a dataset recorded by the Belle detector.

2. B Factories and the Belle Detector

In this chapter I present the physics motivation for research at B factories. I briefly describe the KEKB accelerator and the Belle detector in this chapter, for a detailed description see “The Physics of the B Factories” [1] and the “KEKB B Factory design report” [32]. I focus especially on the $\Upsilon(5S)$ resonance and its differences from the $\Upsilon(4S)$ resonance commonly investigated at B factories.

2.1. B Mesons and Υ Resonances

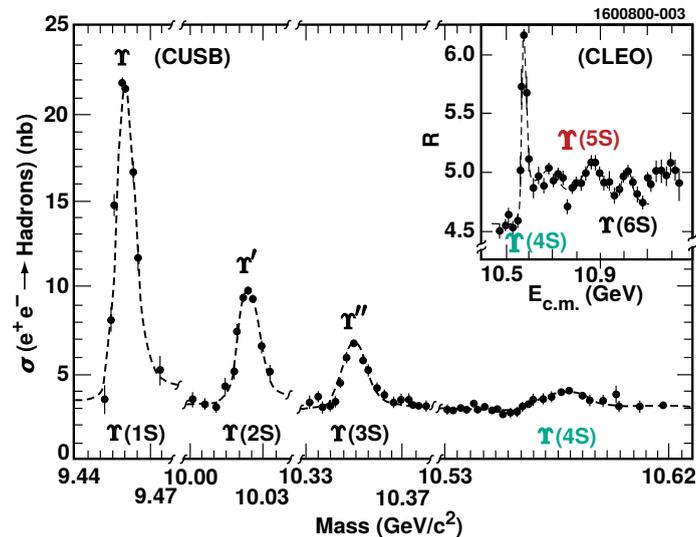


Figure 2.1.: The bottomonium resonances $\Upsilon(1S)$, $\Upsilon(2S)$, $\Upsilon(3S)$ and $\Upsilon(4S)$ in the total hadronic cross section σ_{had} . Given in nb as measured by CUSB [37]. The figure in the upper right corner shows additional results for $\Upsilon(4S)$, $\Upsilon(5S)$ and $\Upsilon(6S)$ obtained by the CLEO Collaboration [8]. These are given as the ratio $R = \sigma_{\text{had}}/\sigma_{\mu\mu}$. This plot was initially published in Ref. [9]. This figure is adapted from [14].

The heavy quarks charm and bottom can form bound states containing a quark and antiquark of the same flavor. Due to the high mass of both quarks, these states behave non-relativistic and, in analogy to the positronium state formed by electron and positron,

are referred to as quarkonia. The state formed by a $b\bar{b}$ pair is called “bottomonium”.

At electron-positron colliders, bottomonium states are created via virtual photons which have the quantum numbers $J^{PC} = 1^{--}$. This limits the $b\bar{b}$ states created in this process to ones with $^{2S+1}L_J = 3_1^S$ [9]. Six of these so-called $\Upsilon(nS)$ resonances, shown in Figure 2.1, have been measured at energies between 9.46–10.99 GeV [46].

The resonances $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ can only decay via OZI-suppressed strong force decays or via the weak or electromagnetic interaction. For Υ states above the “open-bottom threshold”, i.e. the energy needed to create a pair of B mesons, the OZI suppression is lifted. This makes the decay to two B mesons the dominating decay mode for $\Upsilon(4S)$ and heavier resonances.

The mass of the $\Upsilon(4S)$ resonance is only 20 MeV above the open-bottom threshold. Consequently, no other particles are produced when the resonance decays and the B mesons are almost at rest in the bottomonium frame. Because B mesons provide many interesting physics cases, there are specialized experiments that operate almost exclusively at the Υ resonances. These experiments are referred to as “B factories” in this work. Examples for this type of experiment design are the Belle, BaBar and Belle II Experiments. While the LHCb experiment at the Large Hadron Collider (LHC) can also be described as a B factory it is excluded from the following description due to the completely different setup of experiments at hadron colliders.

The initial physics motivation for the construction of B factories was the investigation of CP violation as predicted by the CKM mechanism [33].

To precisely measure the CP violating contribution to the processes $B_d^0 \rightarrow J/\psi K_S^0$, one has to measure the distance between two neutral B-meson decay vertices to account for B-meson mixing, the conversion of B^0 mesons to \bar{B}^0 mesons. Since the difference between the decay lengths of two B mesons in their center-of-mass frame is only in the order of a few 10 μm , B factories have asymmetric beam energies to boost the bottomonium system in the direction of the higher-energetic beam. This results in longer lifetimes in the inertial reference frame of the detector and thus increased distance between the decay vertices as measured by the detector.

Besides measuring CP violation, B factories are also able to perform other precision measurements of theory predictions in the B meson sector, thus testing the Standard Model of particle physics for consistency [13]. This is made possible by the combination of lepton colliders’ comparatively clean decay spectrum and very high luminosities.

While B factories usually operate at the $\Upsilon(4S)$ resonance to exploit the higher cross section and more favorable decay spectrum, the higher-energy $\Upsilon(5S)$ resonance has unique access to the s-quark sector. At this resonance, which was found at 10.86 GeV by the CLEO experiment [8], the bottomonium also decays to B_s^0 mesons. These mesons provide the opportunity to test whether the Standard Model behaves as expected when exchanging a first generation up or down quark for a second-generation strange quark [11, Chapter 23].

In contrast to the decay to B mesons at the $\Upsilon(4S)$ resonance, the $\Upsilon(5S)$ resonance does not only decay to B_s^0 in the ground state. Instead, the bottomonium decays preferentially (87%) to two excited B_s^{*0} mesons, to one excited B_s^{*0} and one ground-state B_s^0 (7.3%) or

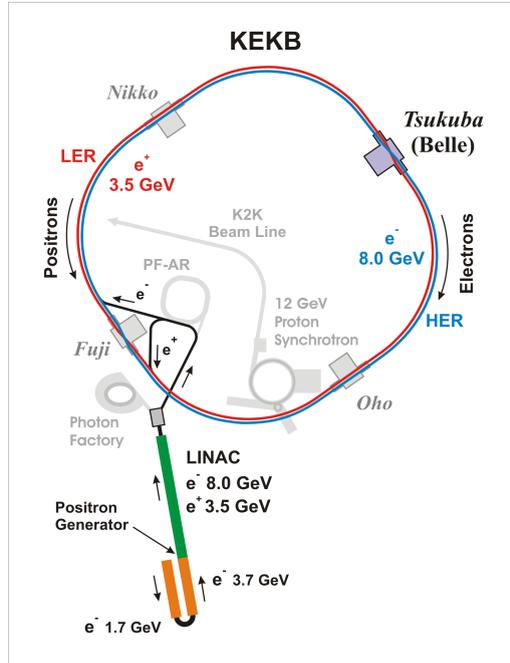


Figure 2.2.: A sketch of the KEKB accelerator complex in Tsukuba, Japan. LINAC refers to the linear accelerator, LER to the low energy ring and HER to the high energy ring. Taken from Ref. [31].

to two non-excited B_s^0 mesons (no up-to-date measurement¹) [6].

The excited B_s^{*0} mesons each decay to a ground-state B_s^0 meson and a 50 MeV photon. This photon is difficult to detect which is why it is usually not reconstructed. The challenges this missing photon presents to the method of Full Reconstruction are elaborated on in Chapter 4.

2.2. Belle Experiment

The Belle Experiment consists of the KEKB accelerator and the Belle detector. During their operation from June 1999 to June 2010, KEKB and the Belle detector were able to produce and collect an integrated luminosity of 1041 fb^{-1} , a large dataset used in a wide range of analyses.

2.2.1. KEKB

The KEKB collider at the High Energy Accelerator Research Organisation (KEK) in Tsukuba, Japan (shown in Figure 2.2) provides the particle beams used for the Belle detector. It was constructed in the tunnel previously used for the TRISTAN collider and consists of two storage rings, one high-energy ring (HER) for 8 GeV electrons and one low-energy ring (LER) for 3.5 GeV positrons. The two rings cross at two points, one

¹The most recent measurement of decays to $B_s^{(*)0}$, [6], which supersedes [36], does not contain a measurement of the decay $\Upsilon(5S) \rightarrow B_s^0 B_s^0$. $\mathcal{B}(\Upsilon(5S) \rightarrow B_s^0 B_s^0)$ is assumed to be 5.7%

(referred to as Interaction Point; IP) is surrounded by the Belle Experiment while the other one is needed to guarantee that both particle rings have the same circumference [35]. A single linear accelerator is used to fill the two KEKB rings with electrons and positrons. While the electrons are injected directly from this linear accelerator into the HER, the positrons to fill the LER are created by bombarding a tungsten single-crystal target with electrons [45].

The beams in the HER and the LER collide inside the Belle detector, the center-of-mass energy for each beam collision or “event” is described by

$$E_{\text{CMS}} = 2\sqrt{E_{\text{HER}}E_{\text{LER}}}. \quad (2.1)$$

In contrast to hadron colliders such as the LHC, E_{CMS} is also the exact energy of each interaction at electron-positron colliders. This makes them particularly useful for missing-energy studies. These are physics analyses in which all detectable particles are used to study undetectable particles such as neutrinos or potential dark-matter candidates. The precise control over the initial state also allows lepton colliders like KEKB to select between the different bottomonium resonances.

The luminosity in e^-e^+ accelerator rings is given by

$$\mathcal{L} = \frac{N_b n_{e^-} n_{e^+} f}{A_{\text{eff}}}, \quad (2.2)$$

with the number of electrons/positrons in each bunch n_{e^-/e^+} , the number of bunches N_b , the circulation frequency f and the effective beam-intersection area at the interaction point A_{eff} [11, Chapter 1].

KEKB significantly reduced this beam-intersection area compared to previous experiments and achieved a beam size of $A_{\text{eff}} = 80 \times 1 \mu\text{m}^2$. This way, KEKB was able to reach a peak luminosity of $2.11 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [2], approximately twice its design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [35].

During most of its operation, the Belle Experiment was operating at the $\Upsilon(4S)$ resonance since all Υ resonances at higher masses also have lower cross-sections. For this reason alone, most of the data collected by the Belle Experiment (711 fb^{-1}) was recorded at $\Upsilon(4S)$.

In addition to the $\Upsilon(4S)$ dataset, the Belle Experiment has also produced and collected samples at other center-of-mass energies, including a dataset at the $\Upsilon(5S)$ resonance with an integrated luminosity of 121.4 fb^{-1} . Further details can be found in Table 2.1. Taking the cross section for $e^-e^+ \rightarrow b\bar{b}$ and the branching fraction

$$\mathcal{B}(\Upsilon(5S) \rightarrow B_s^{(*)0}\bar{B}_s^{(*)0}) = 0.172 \pm 0.030 \quad (2.3)$$

as measured by [6] into account, one derives that

$$(7.11 \pm 1.30) \times 10^6 \quad (2.4)$$

$B_s^{(*)0}\bar{B}_s^{(*)0}$ pairs have been produced by the KEKB collider.

Table 2.1.: A detailed overview of all datasets recorded by Belle. The largest dataset at $\Upsilon(4S)$ is used for most analyses, the second-largest at $\Upsilon(5S)$ is of particular importance for the subject presented in this thesis. On-resonance measurements are taken with beam energies as close as possible to the peak of a bottomonium resonance, off-resonance measurements are taken at energies below the peak. The on-resonance measurements are analyzed to study B physics, the off-resonance measurements are needed to study the background of on-resonance measurements. Adapted from Ref. [11, Chapter 3].

Resonance	Integrated luminosity in fb^{-1}		Number of Υ decays or $B\bar{B}/B_s^{(*)0}\bar{B}_s^{(*)0}$ pairs/ 10^6
	On-resonance	Off-resonance	
$\Upsilon(1S)$	5.7	1.8	102.0 ± 2.0
$\Upsilon(2S)$	24.9	1.7	158.0 ± 4.0
$\Upsilon(3S)$	2.9	0.2	11.0 ± 0.3
$\Upsilon(4S)$	711.0	89.4	$620.0 \pm 9.0 B\bar{B}$
$\Upsilon(5S)$	121.4	1.7	$7.1 \pm 1.3 B_s^{(*)0}\bar{B}_s^{(*)0}$

2.2.2. Belle Detector

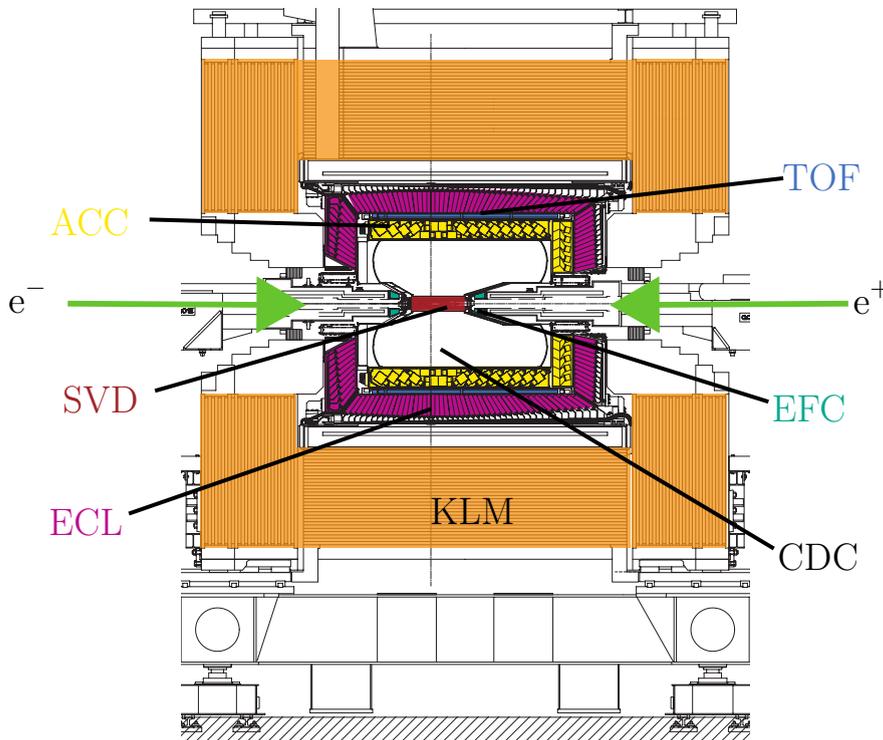


Figure 2.3.: The Belle detector with its seven sub-detectors. Each detector is described in the text, for further detail see Ref. [1]. Adapted from Ref. [13].

The Belle detector, the only experiment at KEKB, is an almost hermetic particle detector. This means that it is designed to enclose as much of the interaction point as possible, allowing it to capture all particles produced in bottomonium decays. The Belle detector was explicitly designed with a focus on CP violation since Kobayashi and Maskawa had predicted this phenomenon in the B meson system [13]. This design, however, still resulted in a general purpose detector which for example was able to contribute to precision measurements of $|V_{ub}|$ and $|V_{cb}|$.

The detector itself consists of seven sub-detectors: Two tracking detectors (Silicon Vertex Detector and Central Drift Chamber), two calorimeters (Electromagnetic Calorimeter and Extreme Forward Calorimeter), a K_L^0 and muon detection system (KLM) and an Aerogel Cherenkov counter as well as a time-of-flight detector for particle identification. Between the ECL and the KLM a superconducting solenoid is placed. This provides the 1.5 T magnetic field used to determine the charge and momentum of charged particles.

The following briefly summarizes the functions of each detector. A more thorough description of the initial design can be found in Ref. [1], additional information can be found in Ref. [11, chapter 2].

The **Silicon Vertex Detector (SVD)**, consisting of four layers of double-sided silicon-strip detectors (DSSDs), is Belle's innermost sub-detector. Initially, the SVD only had three layers, in a configuration referred to as SVD1. In 2003, SVD1 was replaced by the more radiation-tolerant, four-layered SVD2.

The SVD is crucial in precisely determining the decay vertices of B mesons to observe time-dependent CP asymmetries, one of Belle's main design goals. In addition, the SVD is used to determine the decay vertices of D mesons and τ leptons which significantly improves their reconstruction.

The SVD2's inner layer is mounted at a distance of 5 mm from the beam pipe. The other layers are mounted at distances of 29 mm, 55 mm and 88 mm. This close proximity to the interaction point is needed to accurately resolve the origin of particle tracks but comes at the cost of high radiation damage to the detector. The readout chip installed in the SVD1 when it first started operation was not able to tolerate more than 200 krad. This prevented significant increases in KEKB's luminosity. Subsequent upgrades of the readout chips in the SVD1 and its successor SVD2 eventually increased the radiation tolerance to 20 Mrad.

The main tracking detector, the **Central Drift Chamber (CDC)**, is a gaseous detector with 50 cylindrical layers reaching from an inner radius of 80 mm to an outer radius of 880 mm. The CDC allows Belle to precisely determine the curvature of charged particle tracks, mostly due to its large size and high number of layers as described by the Glückstern Formula [23]. With this curvature, the important parameters momentum and energy loss per distance dE/dX of each track can be determined. Using the Bethe formula [10], one can then identify particles with dE/dX . This is especially useful for particles which do not leave the CDC and can therefore not be identified using the PID system.

The CDC is segmented into 8400 drift cells with a maximum drift distance between 8 and 10 mm. Since B mesons mostly decay into particles with momenta lower than 1 GeV, multiple Coulomb-scattering is the dominant effect limiting momentum resolution. For this reason, the gas mixture chosen for the CDC is helium-ethane in a 1:1 ratio. This mixture has a low average proton number which limits multiple scattering while still providing a large energy-loss resolution due to the high ethane content.

Belle’s **electromagnetic calorimeter (ECL)**, is mainly used to detect photons from various B meson decays. In addition, the ECL measures the energy of electrons which aids in electron particle identification (PID). The ECL consists of 8736 CsI(Tl) crystals, each with a silicon photodiode readout. These crystals almost point at the interaction point. A small tilt with respect to the interaction point is chosen to avoid photons escaping through gaps between the crystals. The number of crystals was chosen to maximise the granularity of the calorimeter while guaranteeing that about 80% of the energy deposited by a photon from the center of the crystal is contained within the crystal. The ECL covers a polar angle between 17.0° and 150.0° which corresponds to a geometric acceptance of 91%.

A second calorimeter, the **Extreme Forward Calorimeter(EFC)**, covers the forward region from 6.4° to 11.5° as well as the backward region from 163.3° to 171.2° . These regions require extreme radiation hardness which is why Bismuth Germanate was chosen as scintillator. The EFC is crucial for processes such as $B \rightarrow \tau\nu$ since these have decay products in the aforementioned regions. In addition, the EFC also fulfills several other functions such as reducing backgrounds in the CDC by acting as a beam mask as well as monitoring the luminosity for the Belle Experiment.

The **K_L^0 and muon (KLM)** system in Belle uses resistive plate counters (RPCs) to detect charged particles. Alternating with layers of RPCs are layers of iron, which, in addition to providing a return yoke for the magnetic flux of the solenoid, serve as absorber material with 3.9 hadronic interaction lengths for K_L^0 mesons. These neutral mesons shower in the ECL, the solenoid or in the iron plates which allows the RPCs to measure them. The KLM can also be used to distinguish between charged hadrons (π^\pm or K^\pm) and muons since muons are minimum ionizing particles which lose less energy per distance and can thus penetrate further outward through the detector.

To distinguish between particle types, especially between kaons and pions, Belle employs an **Aerogel Cherenkov Counter (ACC)**. This sub-detector consists of silica aerogels with different refractive indices between 1.01 and 1.03, depending on their polar angle region. The Cherenkov light emitted by these aerogels is detected by fine-mesh photomultiplier tubes. These were chosen because they can operate in the strong magnetic field (1.5 T) in which the ACC is installed. The light yield from the ACC is used among other measurements to construct a likelihood probability for correctly identifying pions or kaons.

The other sub-detector used solely for particle identification is the **time-of-flight detector system (TOF)**. Using 128 plastic scintillation counters, this sub-detector can achieve a time resolution of 100 ps for particles with a momentum below 1.2 GeV. The time at the interaction point, t_{IP} , is well known at Belle. Together with the momentum from the tracking system this information can be used to test different particle hypotheses for each track. Besides particle identification, the TOF is also used to provide timing signals to the trigger system. To achieve the desired trigger rate of less than 70 kHz, the TOF is also equipped with 64 thin trigger scintillation counters (TSCs) which mitigate false triggers from conversion of photons in the plastic scintillator material [38, Chapter 5].

Since only a small fraction of all particle collisions (“events”) produce interesting processes, most collider experiments employ a **trigger system**. This system uses relatively coarse criteria to select interesting events and discard the rest. Since KEKB reaches a beam

crossing rate of 509 MHz and Belle's data acquisition system can process an event rate of 500 Hz, the trigger system must reduce the event rate by a factor of 10^6 . Belle achieves this by using a two-stage trigger system consisting of the Level-1 hardware trigger and the Level-3 software trigger. The hardware trigger's Global Decision Logic (GDL) collects trigger events from each sub-detector, e.g. trigger signals for charged particles from CDC and TOF. By correlating several trigger signals, the GDL decides whether to send the event to the Level-3 trigger. This high-level trigger system uses the Belle Analysis Framework (`basf`) [25] to construct events by collecting all data from the different sub-detectors, the same way an offline reconstruction would. If these reconstructed events pass final selection criteria, they are saved for analysis.

3. Methods and Algorithms

In this section I describe the software setup at the Belle II experiment. I also describe B meson tagging, an analysis method used to study decays to invisible particles with high precision. In addition, I describe two multivariate algorithms used to implement this method at the Belle and Belle II Experiments.

3.1. Software Setup

3.1.1. Belle II Analysis Software Framework and B2BII Conversion Package

To reproducibly analyze the large amounts of data generated by the Belle II detector, the Belle II collaboration uses the Belle II Analysis Software Framework (`basf2`) [34]. It operates in a modular manner to analyze data created by the Belle II detector. Each event is processed independently by a number of modules that only retrieve from and write to a shared storage, the `DataStore`. These modules are written in the C++ programming language and compiled at install time or distributed as pre-compiled binaries via the CernVM File System. Using a Python interface, users can create chains of `basf2` modules that process events sequentially to arrive at some higher-level representation of the measured data. This software design allows physicists to quickly change analysis workflows in Python but also guarantees fast data processing with pre-compiled software. All studies conducted in this work are based on a development version of `basf2` with commit hash `5a2374a6ec`.

As the Belle II Experiment uses a completely new software framework, data recorded with the Belle Experiment must be converted before it can be used by `basf2`. This conversion is achieved with the B2BII software package [22]. In principle, the raw data from the Belle Experiment could be reprocessed with the clustering, track-finding and -fitting algorithms used in Belle II, this would require calibration studies to validate the algorithms for the detector. Instead, B2BII uses tracks and clusters created with `basf` and converts them to the `basf2` conventions. B2BII also provides interfaces to detector-specific variables to allow for a comparison between results obtained with `basf2` analysis methods and results from the Belle Experiment. Specific differences between variables used in Belle and Belle II are elaborated on in Section 3.3.

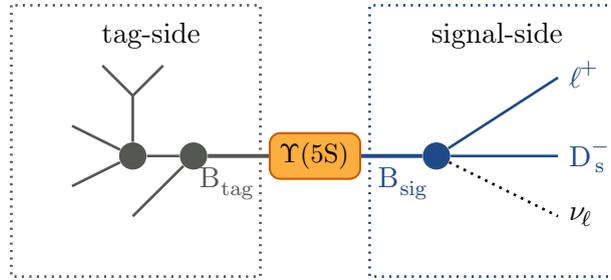


Figure 3.1.: The concept of B tagging explained schematically. Shown on the right side is the signal decay $B_s^0 \rightarrow D_s^- \ell^+ \nu_\ell$ which cannot be reconstructed completely as the neutrino is invisible to the detector. Instead, the full information is inferred from the known beam energy and from the second B meson which can be reconstructed completely. This figure is adapted from Ref. [27].

3.2. B Meson Tagging at Belle and Belle II

3.2.1. Method

To analyze decays of B mesons into states with at least one neutrino, additional information is required since neutrinos cannot be detected directly at collider experiments. It is possible to infer this missing information by analyzing all other measured decay particles. In the following, this method is called “B-tagging”. It is also referred to as “Recoil B-meson reconstruction” in e.g. Ref. [11, Chapter 7].

As B mesons are always created in pairs with no additional particles at B factories, one can reconstruct both independently. Both reconstructed mesons are then combined into the full event. Any discrepancies between this reconstructed $\Upsilon(4S)$ or $\Upsilon(5S)$ resonance and the initial beam conditions must then be attributed to undetected particles. If only one undetected particle is produced in the B decay, one can completely calculate its four-momentum with this method but even for multiple invisible particles B-tagging greatly simplifies physics analyses.

Comparing B-tagging performance

Since there are several ways to reconstruct the B_{tag} , Ref. [29] defines three performance indicators to compare them:

- **Tagging efficiency** is the ratio of tagged B mesons to the total number of B mesons in the data sample. As only a fraction of B decay channels is used for B-tagging, this has a theoretical upper limit given by the sum of the branching fractions of all channels used for the reconstruction of the B_{tag} .
- **Tag-side efficiency** is the ratio of *correctly reconstructed* and tagged B mesons to the total number of B mesons in the data sample. It is the product of branching fractions of the implemented channels and the efficiency of the algorithm.
- **Purity** is the ratio of the two other indicators or alternatively the number of correctly tagged B mesons out of all tagged ones.

There are two main approaches to B-tagging, inclusive reconstruction and exclusive reconstruction. The following describes the two ends of a spectrum of tag-side reconstruction methods, many analyses use partially inclusive reconstruction methods which feature properties of both approaches.

Inclusive reconstruction

In the purely inclusive reconstruction approach, one regards all particles not used to reconstruct the B_{sig} as belonging to the B_{tag} and does not construct different particle hypotheses for the tracks and clusters in the detector. Instead, all tracks remaining after signal selection are treated as pions and all calorimeter clusters without a track as photons. In the simplest version of inclusive reconstruction the momentum of the B_{tag} is then reconstructed from the sum of the momenta of all of these tracks.

The inclusive approach has an efficiency of close to 100% since it does not require measuring and correctly identifying all decay particles of the tag-side B meson. Inclusive reconstruction does, however, suffer from low discriminating power since only statistical knowledge about distributions of physical quantities can be used to check the B_{tag} for consistency with a B meson decay. Examples for such quantities are the beam-constrained mass M_{bc} and the energy difference ΔE [11, Chapter 7] are defined as:

$$M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - \mathbf{p}_{\text{B}}^2} \quad (3.1)$$

$$\Delta E = E_{\text{B}} - E_{\text{beam}} \quad (3.2)$$

All parameters used to calculate these two quantities are measured in the center-of-mass frame. Variables with the subscript B are properties of reconstructed signal B mesons, i.e. E_{B} is the energy of B_{sig} and \mathbf{p}_{B} is its momentum.

Inclusive reconstruction is used for decays with clear signatures and very low branching fractions such as $B_{\text{u}}^+ \rightarrow \mu\nu_{\mu}$, a rare decay process with $\mathcal{B} < 8.6 \times 10^{-7}$ and a clear signal definition due to the high-energy muon [7].

Exclusive reconstruction

The second main category of algorithms for tagging B mesons, exclusive reconstruction, uses known decay modes of B mesons to explicitly reconstruct the B_{tag} meson. The efficiency of exclusive reconstruction algorithms is restricted by two factors: The incomplete number of decay channels used to reconstruct B mesons and the requirement that all daughter particles of the B_{tag} within the event have to be found and correctly identified. For decays with many final-state particles, this second point is especially relevant since the tagging efficiency of all tracks must be multiplied to give the overall probability of detecting all final-state particles. Together these two effects limit the efficiency of exclusive reconstruction algorithms to $\mathcal{O}(1\%)$.

Exclusive reconstruction can again be subdivided into two categories: semileptonic and hadronic reconstruction. Semileptonic reconstruction uses B meson decays with a lepton-neutrino pair for tagging. These are more abundant than hadronic decays but since

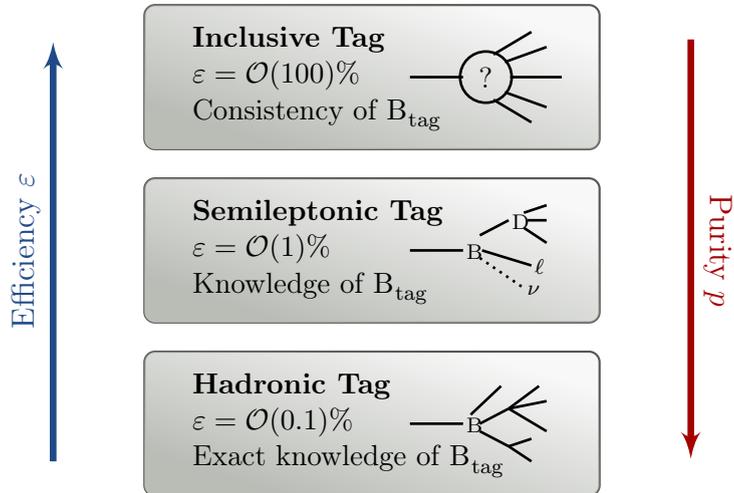


Figure 3.2.: Comparison of the three methods used for recoil B-tagging, ranked by efficiency and purity. Taken from [29].

the neutrino is not measured in the detector the precision of semileptonic reconstruction is inherently limited. In missing-energy studies this is especially problematic since the assumption that all missing energy can be attributed to a single particle is no longer valid.

The second category, hadronic reconstruction, only uses decays without neutrinos, this limits the efficiency to $\mathcal{O}(0.1\%)$ but allows exact reconstruction of the B_{tag} . An illustration of the three different approaches to recoil B-tagging can be seen in Figure 3.2.

3.2.2. Full Reconstruction

The Full Reconstruction (FR) is a hierarchical multivariate algorithm that exclusively reconstructs B mesons [16] in 1104 decay channels. Training a classifier for each of these decay chains would be computationally intensive and would also be statistically limited since one would have to generate training data for each of these decays. Instead, one factorizes the decay process into stages and exploits the fact that intermediate particles have similar features no matter which decay they originate from.

In each of the four stages of the Full Reconstruction algorithm, particles are combined within loose selection windows and a classifier from the NeuroBayes package [17] is applied. This classifier uses, in addition to kinematic features, the product of the classifier outputs of the daughter particles to assign a signal probability to each possible particle. These possible particles, referred to as “candidates”, must then have a higher signal probability than a predefined threshold to be passed to the next stage. In this manner, the Full Reconstruction is able to reconstruct B mesons in 1104 decay chains with only 71 classifiers.

The Full Reconstruction is able to reach maximum efficiencies of 0.18 % for B^0 and 0.28 % for B^+ [16]. This corresponds to an improvement by a factor of two compared to the previous method employed by Belle which used a series of cuts on the invariant masses of reconstructed particles. In addition to limiting the efficiency, the cut-based method fixes the selection to a working point in the efficiency-purity plane. Multivariate methods like

the FR offer more flexibility as analysts can cut on the classifier output to select the desired working point along the efficiency-purity curve.

Several attempts were made to adapt the Full Reconstruction algorithm to the $\Upsilon(5S)$ resonance but to this date, no results using fully-reconstructed B_s^0 mesons have been published. The Full Reconstruction in Belle was superseded by the Full Event Interpretation in Belle II.

3.2.3. Full Event Interpretation

The Full Event Interpretation (FEI) is a completely new implementation following the same general principle of the Full Reconstruction. This section aims to give an overview of the algorithm and its usage at Belle and Belle II. The full algorithm is described in detail in References [29], [30] and [41].

The Full Event Interpretation algorithm uses approximately 200 Stochastic Gradient Boosted Decision Tree classifiers (SGBDTs) [20] from the `FastBDT` package to recombine final-state particles to B mesons. The BDT classifiers are arranged in six stages, each combining particles to reconstruct progressively heavier mesons. Figure 3.5 shows the relationships between the different intermediate and final-state particles [28].

In detail, the six reconstruction and classification stages of the FEI at the $\Upsilon(4S)$ resonance are:

- Stage 0: Construction of long-lived charged (e^- , μ^- , π^+ , K^+ , p) and neutral (γ and K_L^0) particles from tracks, calorimeter and KLM clusters, and pre-vertexed V0 objects
- Stage 1: Reconstruction of J/ψ , π^0 and Λ candidates
- Stage 2: Reconstruction of K_S^0 and Σ^+ candidates
- Stage 3: Reconstruction of different D as well as Λ_c^+ candidates
- Stage 4: Reconstruction of different D^* candidates
- Stage 5: Reconstruction of B_u^+ and B_d^0 candidates

By default, the FEI is trained without K_L^0 and without baryons. K_L^0 are usually omitted due to low momentum and direction resolution compared to other final-state particles since they can only be seen in the KLM subdetector. However, the FEI still includes the option to use decays including K_L^0 mesons as 29% of all B decays contain them. The resulting tag is less accurate but the larger efficiency worth the tradeoff in some analyses [29]. Reconstruction of baryons is also disabled by default, decay channels including Λ and Σ baryons as well as protons were added later in the development of the FEI as their reconstruction efficiency was thought to be low in Belle II. To ensure backwards compatibility between old trainings and newer software they have to be enabled explicitly.

In each of the FEI's reconstruction stages, all particle candidates from the previous stages are combined to heavier hadrons. Following this, loose pre-cuts on e.g. the beam-constrained mass M_{bc} or the energy difference ΔE are applied. Combinations which fail to pass the cuts are discarded and are not processed any further. This decreases the computing load significantly but rarely removes any desired particles.

After reconstruction, a vertexing algorithm is applied to determine the probability of the recombined particles originating from a common decay vertex. By default, this is the `kFitter` algorithm, one of `basf2`'s three fitting algorithms [29, Chapter 4]. While this algorithm is less precise than e.g. the `TreeFitter` algorithm, it is significantly faster. Due to the large amount of combinations which have to be fit, speed is a significant requirement of fitters suitable for the FEI. For this reason, a fast Kalman-filter based fitting algorithm, `FastFit`, was developed for the FEI. It is, however, not the default as it is not included in `basf2`. Next, a SGBDT classifier from the `FastBDT` package is applied to the recombined particle candidates and the 5–20 candidates (depending on the particle species) with the highest classifier output are propagated to the next stage. This approach, usually called a “best-candidate selection”, differs from the candidate selection in the Full Reconstruction in which specific, hand-selected classifier output thresholds are defined for each decay channel. The best-candidate selection distributes available resources more evenly between the different events and decay channels since no single event can generate large amounts of possible candidates which would have to be fit in the next stage. In addition, the best-candidate selection limits the total amount of memory used for the reconstruction of each event which guarantees smooth application without any fine-tuning to match the hardware infrastructure [29]. The top-level classifier which is used to assign a signal probability to B meson candidates provides 20 candidates for analysis use. While it is possible to simply use the candidate with the highest classifier output, this is not always the best strategy since it does not exploit the information of the signal side. An analysis-specific signal-side selection can be used to discard wrong B_{tag} candidates or to change their ranking. As shown in Figure 3.3, a perfect selection of the B_{tag} could improve the efficiency by 92%. This serves as an illustration of the potential for improvement in this final ranking.

The improved classifier, the best-candidate selection and the additional channels lead to larger tag-side efficiency and higher purity of the Full Event Interpretation compared to the Full Reconstruction. This is also illustrated in Figure 3.4 which shows that the FEI out-performs the FR in most cases [27]. This is especially true for regions of high efficiency in which most analyses set their working point.

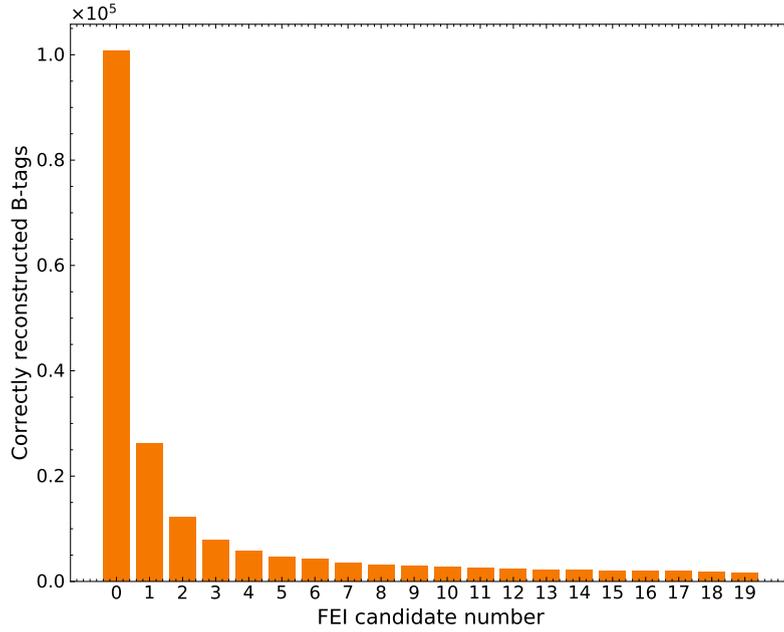


Figure 3.3.: FEI classifier rank of correctly reconstructed B_u^+ candidates. 48% of all correctly reconstructed B tags are not ranked as the top candidate and can only be used if some secondary selection criterion is applied, e.g. by requiring no additional tracks after reconstruction of the second B meson.

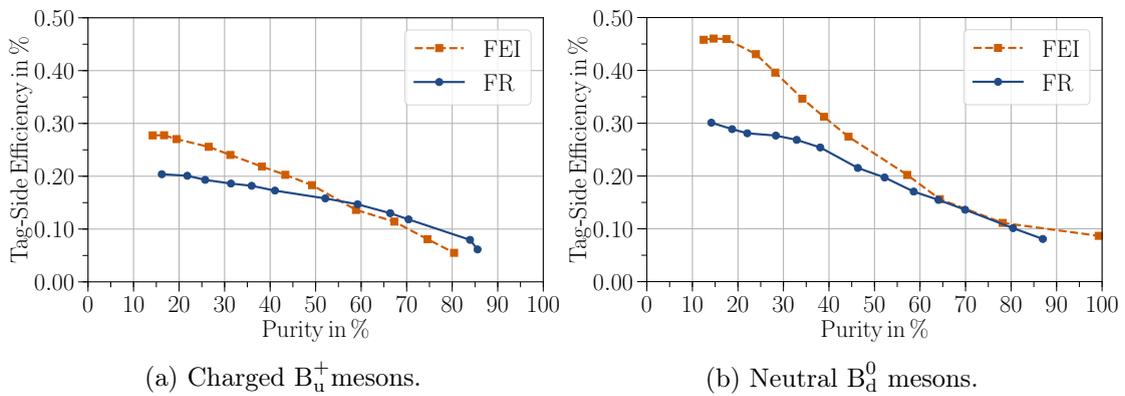


Figure 3.4.: Receiver operating characteristic of the Full Reconstruction and the Full Event Interpretation applied to data. The tag-side efficiency and purity are measured by fitting the beam-constrained mass spectrum. Taken from Ref. [27].

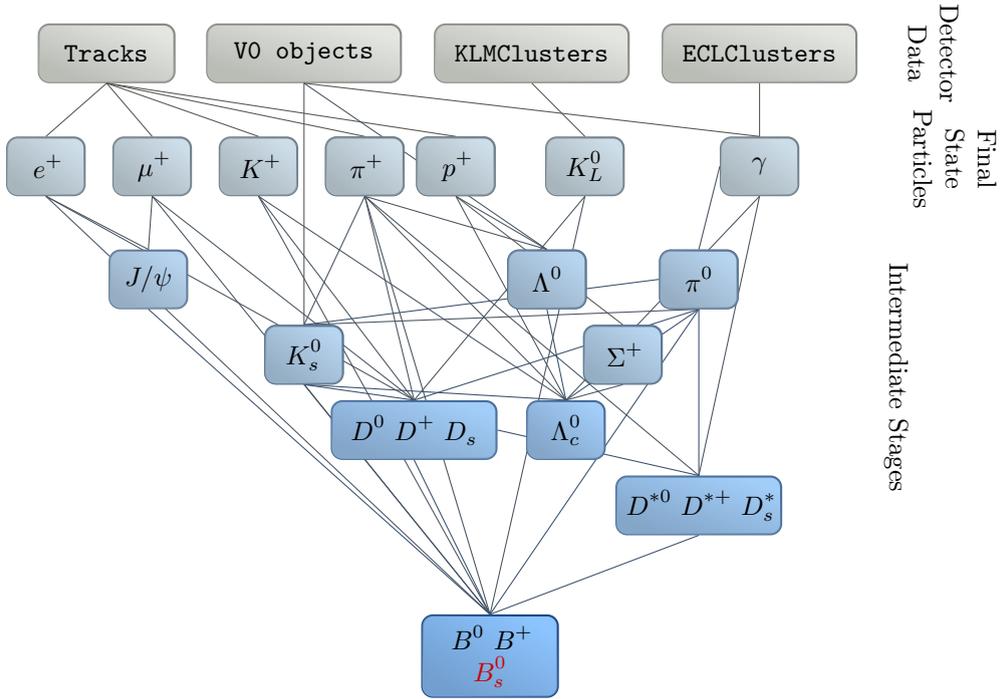


Figure 3.5.: The hierarchical nature of the Full Event Interpretation, shown with B_s^0 mesons as top-level particles in addition to B_u^+ and B_d^0 mesons. Particle candidates are reconstructed in six stages and a BDT classifier is applied at each stage to rank the candidates. Initial candidates are reconstructed from tracks and clusters, intermediate particles from combinations of candidates from previous stages. The classifier in each stage uses the classifier output of the previous stage as input. Adapted from Ref. [29].

Training procedure

As described in Ref. [30] and [29, Chapter 4], a map-reduce approach is used to perform the distributed training of the FEI. This means that the most computing-intensive step of training the FEI, the particle combination and vertex fitting, is performed by independent cluster nodes whose output is then combined. To do this efficiently, the total amount of training data and the amount of B mesons therein is counted in the beginning by a central scheduling agent. This agent then distributes subsets of the total available training data to each computing node. The nodes then reconstruct particle candidates from lighter particles in their dataset and perform a time-intensive vertex fit on the candidates. Following this vertex fit, the internal `basf2 DataStore` is written to a cache directory and the reconstructed particle candidates from all cluster nodes are merged by the central scheduler. Then the classifier is trained on the total dataset and the gained weight files are submitted to a local database. Each step repeats this procedure, using the cached `DataStore` of the previous steps and the trained classifier weight files as its input.

One total training of the FEI requires approximately 100 Million simulated data events to also provide training data for reconstruction channels with low expected branching fractions. On the KEKCC cluster, one full training run takes between 2–5 days using 500 cluster nodes.

3.3. Variables Used for Particle Candidates Reconstructed by the FEI

Multivariate methods like Boosted Decision Trees require variables to classify events. In contrast to other discrimination methods such as the likelihood or Fisher’s linear discriminant [18], features used in BDTs can have non-linear correlations with each other [24]. This greatly simplifies the feature selection for the FEI.

Charged stable particles

The charged particles which can be reconstructed directly from tracks in the detector are e^\pm , μ^\pm , π^\pm , K^\pm , and p/\bar{p} . The same features are used for all of them since all candidates are reconstructed directly from particle tracks.

The features used in the BDT classifier are:

Particle identification (PID) variables Belle and Belle II use individual likelihoods on different hypotheses from multiple subdetectors to give combined likelihoods on the particle species. These likelihoods are saved in PID variables which in Belle II give the combined likelihood ratio of a particle hypothesis over all possible hypotheses. This likelihood ratio, i.e. the posterior probability of the PID calculation, can then be interpreted as a prior probability by the classifier. To account for the differences between Belle and Belle II, slightly different PID variables are used as features in BDTs trained on Belle data. Belle does not have global PID variables. Instead, there are electron ID (eID; using information from the CDC, the ECL and the ACC), muon ID (muID; using the KLM), and binary likelihoods from the ACC, the TOF and the CDC. The classifier for Belle data uses eID, muID and binary likelihood ratios of π^\pm over K^\pm , K^\pm over π^\pm , p over π^\pm , and π^\pm over p.

Kinematic parameters Tracks originating from primary interactions have usually higher momenta than tracks created by secondary particles, i.e. particles created by interaction with the detector. To account for this correlation, the absolute momentum, the transverse momentum and the z -component of the momentum are also used as features in the BDT.

Track-fit results The goodness of the track fit indicates how well the individual measurements in detector layers correspond to the assumed trajectory of a charged particle in a magnetic field. This is used to separate track candidates from each other and from background.

Track impact parameter [12] For each track candidate, the point of closest approach (POCA) is calculated. This is the point closest to the interaction point when projecting the track on the plane orthogonal to the beam. The signed radial component and the z -component of this point are used in the BDT. This especially helps to distinguish between particles from inside the detector and particle from outside.

Pre-cut rank Before applying the classifier, all particle candidates of a type are ranked according to their particle ID. The position in this ranking is also used in the classifier.

Photons

Photons are only detected in the calorimeter so the photon classifier relies on the properties of electromagnetic showers in the ECL. For high-energetic photons, these shower are detected in multiple crystals in the calorimeter which are then combined by a clustering algorithm. The features used are:

Kinematic parameters The energy of the photon candidate, its momentum in transverse direction and its momentum in z -direction are used in the BDT.

Cluster region and shape The BDT classifier uses the region in which the clusters are created (forward calorimeter, backward calorimeter or barrel) as well as their shape. Explicitly, the shape is described by the number of crystals in a cluster and the ratio E_9 over E_{21} . E_9 is here the sum of the energy deposited in the 3×3 crystals around the center of a cluster, E_{21} the sum of the energy deposited in the 5×5 crystals (without corners) around the center. The showers created by photons in calorimeters are wider than those created by hadrons, thus E_9 over E_{21} is larger for photons than hadrons.

Cluster timing The signal recorded from the crystal with the highest energy deposition in a cluster is used to determine the creation time of each cluster. Subtracting the time of flight should yield the time of interaction t_0 for photons originating from the IP.

Pre-cut rank Before applying the classifier, all photon candidates are ranked according to their energy. The position in this ranking is also used in the classifier.

K_L^0 mesons

Long-lived Kaons are only detected in the KLM subdetector where they produce showers. These showers are reconstructed by a clustering algorithm and their geometric features can be used to extract momentum and energy. The K_L^0 classifier uses the energy of each cluster as well as the cluster timing information.

π^0 mesons

In Belle II, neutral π mesons are reconstructed from two photons during analysis. This is not the case for Belle data where π^0 s are already pre-reconstructed in the dataset. This difference requires different features to be used in their classification.

The BDT uses:

Energy, momentum and mass The energy and mass of pions as well as their transversal and z -component momentum are used for Belle II as well as converted Belle data. In addition, the deviation of the reconstructed mass from the nominal pion mass is also used, this provides established physics knowledge that the classifier would otherwise have to learn.

Angle between the photons The angle between the two photons is a powerful feature since it can distinguish between the combination of two unrelated photons and the combination of two photons coming from a two-body decay such as $\pi^0 \rightarrow \gamma\gamma$.

Classifier output of daughters Since the FEI reconstructs neutral pions from photons in Belle II, the photon classifier output is used as an input to the neutral pion classifier. This is not available for converted Belle data as neutral pions are already pre-reconstructed.

Goodness of vertex fit As pions are already vertex fit with a custom fit procedure during B2BII conversion, they have a χ^2 probability assigned to this mass-constrained fit. Such a feature is not used for Belle II data.

Pre-cut rank All candidates are ranked by the deviation of their (invariant) mass from the nominal mass of π^0 mesons. This rank is also used in the classifier.

 K_S^0 mesons

Similar to neutral pions, neutral short-lived kaons have to be treated differently depending on whether the FEI is run on converted Belle or on Belle II data. In both Belle and Belle II data, K_S^0 mesons are already pre-vertexed during tracking, but in Belle II they can also be reconstructed directly from pions.

The K_S^0 classifier uses:

Energy, mass and mass difference The energy measured for the K_S^0 meson in the center-of-mass frame is used in the classifier. The mass of the reconstructed candidate as well as its absolute deviation from the nominal K_S^0 mass are used in the classifier for both directly reconstructed as well as pre-reconstructed kaons.

Vertex information A vertex fit is performed on both kaons reconstructed during tracking as well as those reconstructed in the FEI. This fit returns vertex parameters: The 3D distance from the interaction point to the vertex, the radial as well as the z -component of the vertex position, the significance of the distance, and the goodness of fit.

Angular variables The angle between the momenta of the two daughters and the angle between the momentum and the vertex vector of the kaon candidate are both used as features in the BDT.

Daughter variables For kaons reconstructed by the FEI in Belle II data, the BDT can also access all variables of the kaon’s daughters. The variables used as features in the BDT are the momenta of the daughters in the kaon rest frame, the radial and z -component of their distance to the point of closest approach (POCA) as well as the output of the BDT used to classify them.

Belle K_S^0 finder variables The Belle experiment used two different methods to classify K_S^0 mesons, the cut-based “goodKs” selection and the NeuroBayes-based multivariate “nisKsFinder” [3]. The output of these two methods is available for Belle data converted by B2BII and is used in the K_S^0 classifier in this case.

Pre-cut rank Similar to neutral pions, all K_S^0 candidates are ranked by the deviation of their (invariant) mass from the nominal mass of K_S^0 mesons. This rank is used in the classifier.

Intermediate particles (D , D^* and J/ψ mesons)

Mass and mass difference To optimally exploit the mass constraint of the reconstructed intermediate mesons, the intermediate BDTs use the mass, the signed mass difference to the nominal mass as well as the absolute value of this difference.

Released energy and energy difference The energy released in the decay and the difference to the nominal released energy are also used as BDT features. The released energy is calculated by subtracting the invariant masses of all daughters from the intermediate meson’s invariant mass.

Invariant mass of daughter combinations The daughter candidates are combined in all possible combinations and the invariant mass of each combination is used as a feature in the BDT.

Output of the previous classifiers During the best-candidate selection, the output of previous stages already influences the following stages. This is also done explicitly by passing the classifier output of each daughter and the product of all daughter outputs to the intermediate-meson classifier.

Vertex-fitting information A powerful feature is the goodness of the vertex fit to each B candidate, estimated with the χ^2 value. Additionally, the χ^2 value of each daughter (for which the vertex fit already happened in the previous step) is also used as a feature.

Kinematic variables The momentum of daughters in the rest frame of the intermediate meson, the angle between the intermediate meson’s momentum vector and the position of the vertex of each daughter, the angles between any two daughter’s momentum vectors and the angle between momentum and vertex vectors are all used in the BDT.

Reconstruction information To discriminate between the different decay channels used for each daughter, the ID assigned to each daughter’s decay channel is passed to the classifier.

B mesons

The same variables can be used for neutral as well as charged B mesons since their kinematic properties and masses are very similar due to isospin symmetry. For the same reason, the features used to reconstruct semileptonic and hadronic B decays are also the same. An important constraint for the feature selection is that features must not be correlated with the beam-constrained mass M_{bc} . This is because M_{bc} is used in many analyses and is also used to evaluate the FEI's performance. Any strong correlation could bias analyses and would certainly skew the FEI's evaluation. For this reason, the invariant masses as well as the mass and mass difference used in the intermediate-meson classifiers cannot be used here. The same applies to the energy and energy difference.

In addition to the features used to classify intermediate mesons, the following features are used in the classifier for B mesons at $\Upsilon(4S)$:

Explicit vertex information In addition to the goodness of fit, the information from the vertex fit is also used explicitly. This includes the (x,y,z) coordinates, the 3D, 2D and transverse distance relative to the interaction point and the significance of the distance relative to the IP.

Energy difference The energy difference ΔE mentioned in Section 3.2.1 is a powerful discriminating variable for particles created almost at rest relative to the beam's center-of-mass. ΔE and M_{bc} are weakly correlated via the beam energy but for $\Upsilon(4S)$ this is accepted since using ΔE increases the classifiers performance significantly.

Particle rank All particle candidates are also ranked by the product of all daughters' classifier outputs and this rank is passed to the classifier.

4. Full Event Interpretation at $\Upsilon(5S)$

In this section I describe the development of the FEI for the application at the $\Upsilon(5S)$ resonance. I also describe the different measures implemented to improve the FEI's performance and their individual impact in simulated data.

4.1. Implementation of B_s^0 meson and its decay channels

The decay products of the $\Upsilon(5S)$ resonance include neutral B_s^0 mesons in addition to charged B_u^+ and neutral B_d^0 mesons. These heavier B mesons consisting of a bottom quark and a strange quark have different kinematic properties and unique decay channels. An example of a hadronic B_s^0 decay can be seen in Figure 4.2.

To fully reconstruct B_s^0 mesons, 46 new hadronic B_s^0 decay channels covering approximately 18.9% of all B_s^0 decays are added to the FEI. The most important criterion for their selection is the branching fraction, since training the classifier for each channel requires a statistically representative sample of correctly reconstructed mesons. For this reason, 38 of the 46 channels contain a D_s or D_s^* meson. Due to energy conservation, bottom quarks cannot decay to the heavier top quark on tree-level so they must decay to an up-type quark from another quark family as shown in Figure 4.1a. These decays are CKM suppressed by $\mathcal{O}(|V_{cb}|) = 10^{-2}$ and $\mathcal{O}(|V_{ub}|) = 10^{-3}$. Penguin decays containing virtual top quarks allow $b \rightarrow s$ transitions as shown in Figure 4.1b but are suppressed by $\mathcal{O}(|V_{ts}|) = 10^{-2}$. Since $|V_{ub}|$ is so much smaller, the decays to charmed mesons dominate the B_s^0 meson's branching fraction.

While the implemented channels are the hadronic decay channels with the highest branching fractions, none of them have a branching fraction higher than 1.5%. The large phase space available to B_s^0 meson decays leads to a large amount of hadronic decay channels of which over 100 are known [46]. Semileptonic decay channels with electrons or muons would offer much higher efficiencies since their individual branching fractions are much higher, but the incomplete tagging information these channels offer is undesired in many analyses. For this reason only hadronic channels are considered for the FEI at $\Upsilon(5S)$.

Not all channels could be trained with the available central generic Monte Carlo data for $\Upsilon(5S)$. This is mostly due to insufficient statistics for correctly reconstructed events as the classifier cannot be trained if less than 500 events are successfully reconstructed. Since training with a larger MC sample could solve this problem in the future, the channels are

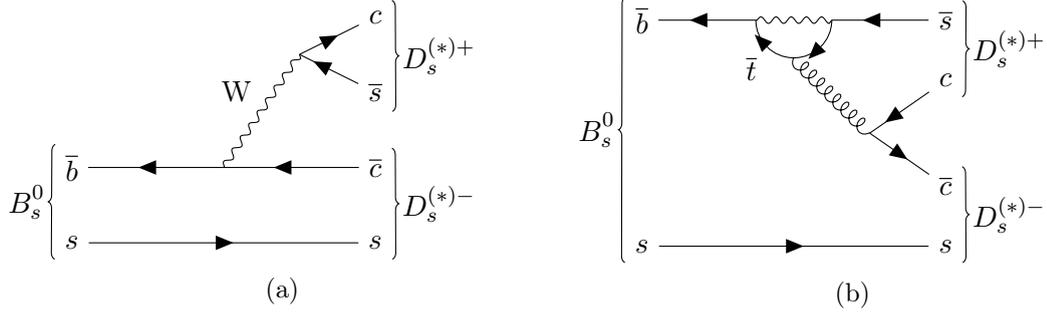


Figure 4.1.: Leading-order tree (a) and penguin (b) Feynman graphs for $B_s^0 \rightarrow D_s^{(*)+} D_s^{(*)-}$ decays. These three decays (both D_s excited, one excited, and none excited) account for 4.5% of all B_s^0 decays [46] and have the highest individual hadronic branching fractions.

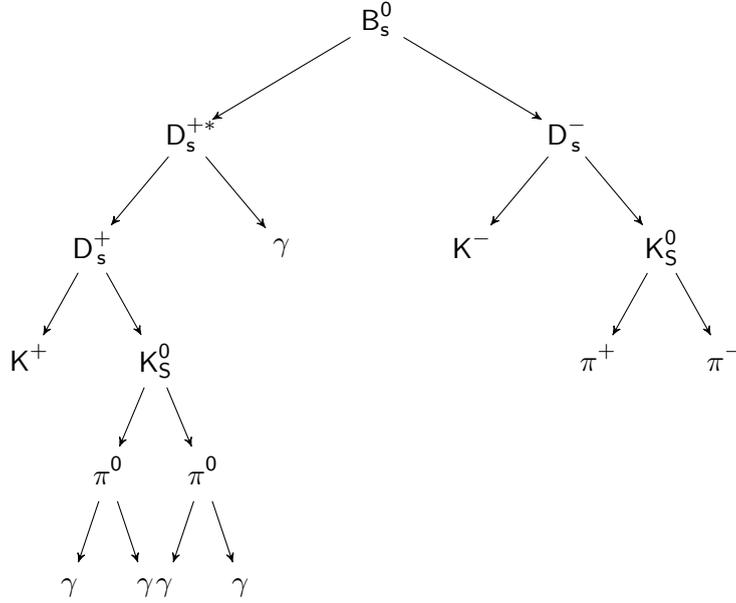


Figure 4.2.: An example for a fully hadronic decay of B_s^0 to final-state particles detected in the Belle detector. The FEI can reconstruct this decay at the $\Upsilon(5S)$ resonance. All six stages of the FEI are needed here to reconstruct this B_s^0 meson.

still implemented even though presently they and other channels with untrained classifiers always return a signal probability of zero. Channels with J/ψ mesons are an exception. This meson is reconstructed in a very pure sample which does not leave enough background events to train the classifier (see Section 4.1.2). The 19 successfully trained channels can be seen in Table 4.1. These channels account for approximately 10.54% of all B_s^0 decays according to the Belle decay file used to generate the Monte Carlo dataset. The branching fractions in this file are based on the 2007 edition of the Review of Particle Physics [48].

Table 4.1.: All additional, successfully trained channels in the FEI at $\Upsilon(5S)$ accounting for 10.54% of all B_s^0 decays. Channels 18 and 34 occur via an intermediate $K^*(892)$ resonance that is not reconstructed. The channels 7, 8, and 10 use a looser pre-cut of $|\Delta E'| < 0.5$ to allow training with a statistically significant sample. The columns labeled N_{Sig} and N_{Bkg} show the number of correctly and incorrectly reconstructed B_s^0 candidates. Additional channels that were added but could not be trained, i.e. had less than 500 events in either background or signal, can be found in Appendix B.

Decay Mode ID	Decay Process	MC BF in %	N_{Sig}	N_{Bkg}
0	$B_s^0 \rightarrow D_s^- D_s^+$	0.86	3148	1747
1	$B_s^0 \rightarrow D_s^{*+} D_s^-$	0.90	2090	3039
2	$B_s^0 \rightarrow D_s^{*-} D_s^{*+}$	1.97	2281	934
3	$B_s^0 \rightarrow D_s^+ D^-$	0.17	918	3779
4	$B_s^0 \rightarrow D_s^{*-} D_s^+$	0.27	576	767
5	$B_s^0 \rightarrow D_s^{*+} D^-$	0.17	507	3101
7	$B_s^0 \rightarrow D_s^- K^+$	0.02	795	1272
8	$B_s^0 \rightarrow D_s^- \pi^+$	0.28	13654	2221
10	$B_s^0 \rightarrow D_s^{*-} \pi^+$	0.27	8385	2596
14	$B_s^0 \rightarrow D_s^- \pi^+ \pi^+ \pi^-$	0.47	19063	49632
15	$B_s^0 \rightarrow D_s^- D^0 K^+$	0.96	752	42480
17	$B_s^0 \rightarrow D_s^- \pi^+ \pi^0$	0.09	25078	11630
18	$B_s^0 \rightarrow D_s^- D^0 K^+ \pi^0$	0.25	514	316956
22	$B_s^0 \rightarrow D_s^- D^{*0} K^+$	0.50	769	19269
28	$B_s^0 \rightarrow D_s^{*-} D^0 K^+$	0.50	1017	24593
30	$B_s^0 \rightarrow D_s^{*-} D^{*0} K^+$	1.50	1115	9925
32	$B_s^0 \rightarrow D_s^{*-} \pi^+ \pi^+ \pi^-$	0.77	27738	36660
33	$B_s^0 \rightarrow D_s^{*-} \pi^+ \pi^0$	0.10	26173	10487
34	$B_s^0 \rightarrow D_s^{*-} D^0 K^+ \pi^0$	0.50	593	135622

4.1.1. Pre-Cuts

The FEI rejects particle combinations which fail to match loose selection criteria at each stage. This is mostly done to limit the unnecessary use of computing resources in the vertexing step after the recombination but also presents the BDT classifier with more

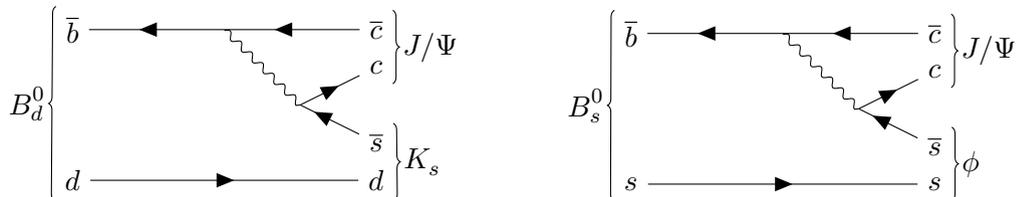


Figure 4.3.: The “golden mode” for B_d^0 mesons and its equivalent for B_s^0 mesons. These modes are color-suppressed but are easily reconstructed due to the clear charmonium decay.

relevant background data, thereby improving its performance. For the final stage in which B mesons are reconstructed, the FEI at $\Upsilon(4S)$ discards candidates with a beam-constrained mass of less than 5.2 GeV as well as those with an absolute energy difference of more than 0.5 GeV. These parameters are explained in Section 3.2.1.

For the FEI at $\Upsilon(5S)$, these pre-cuts have to be adapted to account for the higher mass of B_s^0 mesons. The cut on M_{bc} is raised to 5.3 GeV. This separates between B_s^0 and lighter B mesons and prevents these lighter mesons from being reconstructed in the B_s^0 meson channels. The cut on ΔE was replaced by a cut on a modified variable, $\Delta E'$, which was tightened to $|\Delta E'| < 0.1$ GeV. The variable is defined in Section 4.1.3, the effect of tightening the cut is investigated in Section 5.2.

4.1.2. Decay Channel $B_s^0 \rightarrow J/\psi\phi$

The color-suppressed decay of B_s^0 mesons to a J/ψ meson and a ϕ meson is the $\Upsilon(5S)$ analogue to the process $B_d^0 \rightarrow J/\psi K_s^0$, the “golden mode” at B factories (both shown in Figure 4.3). The process has this name because it has an easily reconstructable decay signature consisting of the charmonium decaying to two high-energy leptons. In the FEI at $\Upsilon(4S)$, the golden modes are implemented for B_d^0 and B_u^+ . The same was done in this thesis, however the ϕ is not reconstructed explicitly since the additional constraint gained by doing so is not needed due to the already clear signature. Even without this constraint, the FEI is not able to train a classifier for J/ψ modes at the time of this work. The exceptionally clear selection does not provide enough background (i.e. misreconstructed J/ψ) to train the classifier. Attempts to solve this problem by allowing more electron candidates in earlier selection steps to lead to less pure J/ψ selections were not successful. A general solution to problems like these could be to bypass the classifier entirely for very pure modes but the impact of this was not studied.

4.1.3. Using $\Delta E'$ Instead of ΔE

The FEI at $\Upsilon(4S)$ uses the energy difference ΔE as an important input feature for the final B meson classifier. It should be noted that a small correlation exists between ΔE and M_{bc} , something which should normally be avoided since M_{bc} is used for later evaluation of the method. Since ΔE has high discriminating power and the correlation is only due to the common inputs, it is still used in the final classifier.

For completely reconstructed B mesons at $\Upsilon(4S)$, the distribution of ΔE peaks at 0 GeV. This is not necessarily the case for $\Upsilon(5S)$. The $\Upsilon(5S)$ resonance does not only decay to B_s^0

but also to an excited state B_s^{*0} which decays in the process $B_s^{*0} \rightarrow B_s^0 \gamma$. Since the photon is usually not reconstructed, the reconstructed B_s^0 mesons are distributed over three peaks. In order of their contribution to the branching fraction, these peaks are: $\Upsilon(5S) \rightarrow B_s^{*0} \bar{B}_s^{*0}$, $\Upsilon(5S) \rightarrow B_s^0 \bar{B}_s^{*0}$ and its charge conjugate, and $\Upsilon(5S) \rightarrow B_s^0 \bar{B}_s^0$. The distribution of all three peaks in ΔE is illustrated in Figure 4.4.

This presents two possible disadvantages: The classifier might, during training, learn only the central value of the $B_s^{*0} \bar{B}_s^{*0}$ peak and lead to a selection bias against the less common decays to $B_s^0 \bar{B}_s^{*0}$ and $B_s^0 \bar{B}_s^0$. Additionally, the discriminating power of ΔE could be weaker since the distribution is spread over a wider range of decays.

To curtail these effects, a new variable was defined for use in the FEI at $\Upsilon(5S)$. This variable is defined as

$$\Delta E' = \Delta E + M_{bc} - m_{B_s} \quad (4.1)$$

and centers all three decay peaks around zero (see Figure 4.5). On the whole dataset, the absolute correlation coefficient between M_{bc} and $\Delta E'$ is $|r| = 0.006$, much smaller than the correlation between M_{bc} and ΔE ($|r| = 0.673$). This is also illustrated in Figure 4.6.

4.1.4. Correlation between M_{bc} and Variables Used to Select B_s^0 Meson Candidates

The other variables used in the B_s^0 meson classifier are also evaluated with regard to their correlation to M_{bc} . While this evaluation has already been performed for the FEI at $\Upsilon(4S)$, the differing decay spectrum of B_s^0 mesons could lead to unexpected correlations between variables that are unproblematic in the $\Upsilon(4S)$ case. Across all decay channels, features related to the previous stage's output have the strongest correlation to M_{bc} . The maximum correlation exists between M_{bc} and the D_s^- classifier output for decay channel 7 ($B_s^0 \rightarrow D_s^- K^+$) with a correlation coefficient of 0.38. Two-body decays also exhibit a correlation between the momenta of daughters and M_{bc} , these variables reach a maximum correlation coefficient of 0.31. This is expected as there is a direct relation between the mass of a reconstructed particle and the momentum of its daughters. An overview over all maximum correlation coefficients for different numbers of B_s^0 daughters can be found in Appendix C. Overall, these different correlations, accumulated by the Boosted Decision Tree, can lead to a non-uniform event selection with respect to the beam-constrained mass. A simple solution would be to avoid variables with strong correlation entirely, however this comes at the cost of lower classification power. A better solution to this problem could be boosting to uniformity as demonstrated by Ref. [42]. This feature is already implemented in `FastBDT` but is currently not used in the FEI as its impact has to be studied thoroughly.

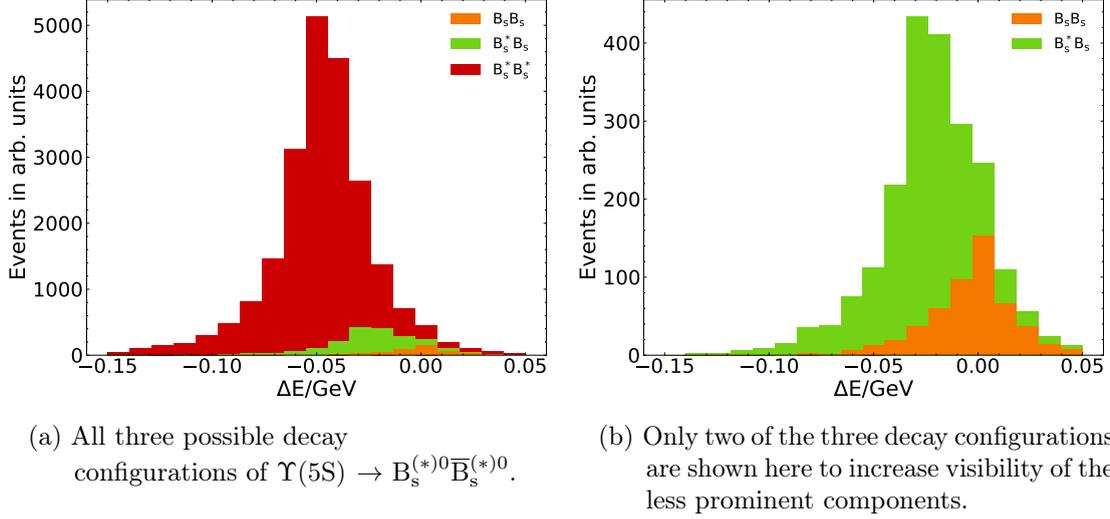


Figure 4.4.: ΔE of reconstructed B_s^0 mesons for simulated decays of $\Upsilon(5S)$ to $B_s^{*0} \bar{B}_s^{*0}$, $B_s^{*0} \bar{B}_s^0$, and $B_s^0 \bar{B}_s^0$.

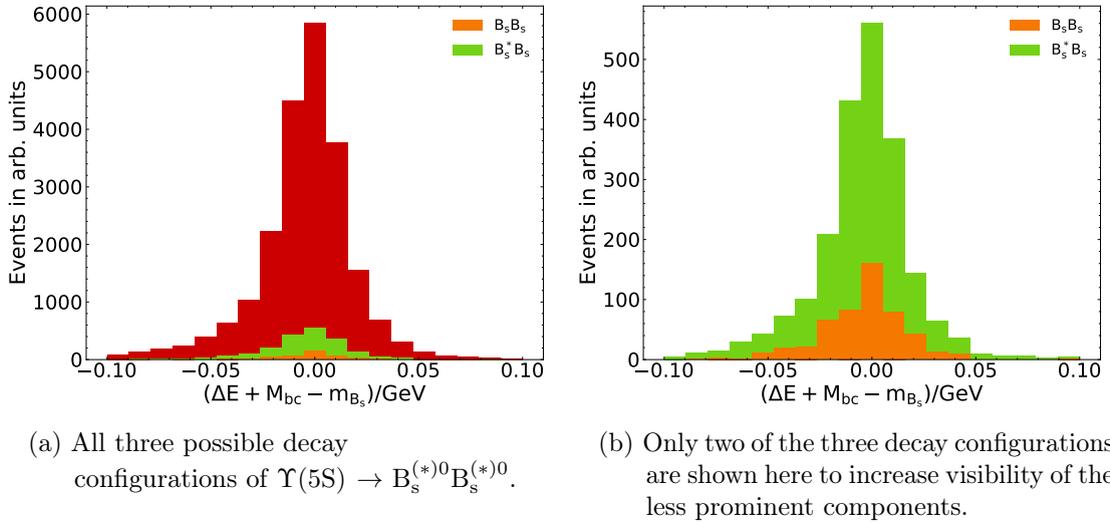


Figure 4.5.: $\Delta E' = \Delta E + M_{bc} - m_{B_s}$ of reconstructed B_s^0 mesons for simulated decays of $\Upsilon(5S)$ to $B_s^{*0} \bar{B}_s^{*0}$, $B_s^{*0} \bar{B}_s^0$, and $B_s^0 \bar{B}_s^0$.

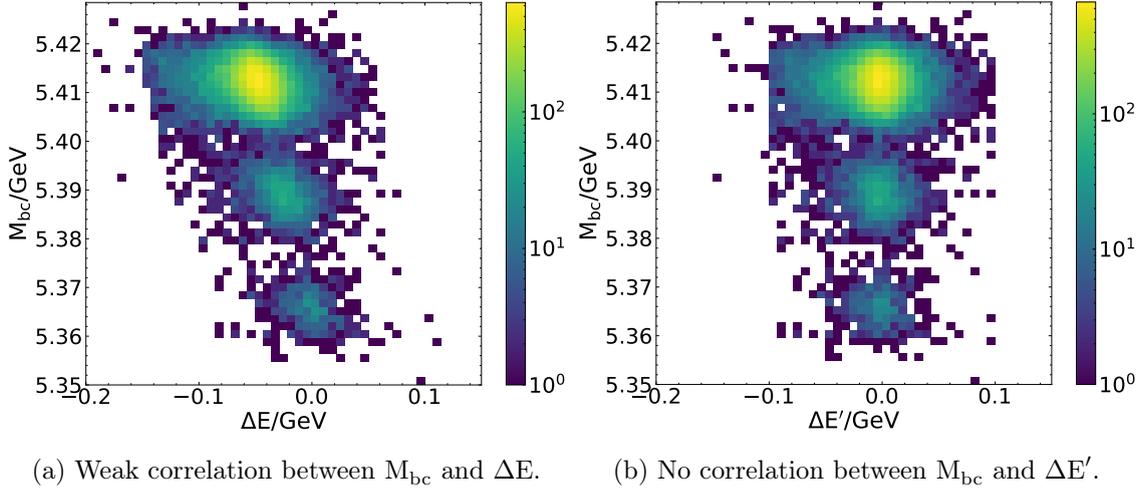


Figure 4.6.: Two-dimensional histogram of beam-constrained mass M_{bc} and $\Delta E/\Delta E'$ for correctly reconstructed B_s^0 mesons. The three hotspots for all three decay configurations are clearly visible. Switching from the feature ΔE to $\Delta E'$ avoids correlation to M_{bc} .

Table 4.2.: Centrally-produced generic Monte Carlo samples for $\Upsilon(5S)$ at Belle. $B^{(*)}$ refers to both (potentially excited) B_u^+ and B_d^0 mesons. Charm and UDS MC was not used for training.

Name	Decay processes	Number of events / 10^6
BsBs	$\Upsilon(5S) \rightarrow B_s^{(*)0} \bar{B}_s^{(*)0}$	42.8
NonBsBs	$\Upsilon(5S) \rightarrow B^{(*)} \bar{B}^{(*)}(\pi)$, $\Upsilon(5S) \rightarrow \Upsilon(4S)\gamma$	178.9
Charm	$e^- e^+ \rightarrow c \bar{c}$	1533.7
UDS	$e^- e^+ \rightarrow (u\bar{u}, d\bar{d}, s\bar{s})$	954.4

4.2. Datasets Used to Train the FEI at $\Upsilon(5S)$

The centrally available MC dataset is divided into 6 equally sized subsets. Each of these has the same luminosity as the available recorded data and is referred to as a “stream”. The FEI at $\Upsilon(5S)$ is trained using streams 1–5 of events containing approximately 35M pairs of $B_s^{(*)0}\bar{B}_s^{(*)0}$. In addition, streams 1 and 2 of events containing approximately 60M B_d^0 and B_u^+ pairs. Stream 0 for all types of simulated data is used for later validation. This guarantees statistical independence between training and validation sets.

The FEI is not trained on continuum Monte Carlo, i.e. simulated non- $\Upsilon(5S)$ events. There are two reasons for this: Firstly, there are two dedicated continuum suppression algorithms in `basf2`, one using deep-learning methods [49] and one based on boosted decision trees. Both can easily be integrated in any analysis and should be adaptable for $\Upsilon(5S)$. Secondly, the treatment of intermediate mesons reconstructed in continuum data is not trivial. While reconstructed intermediate meson candidates can correspond to physical particles, their kinematic properties can be very different. The intermediate classifiers could learn those properties at the cost of worse identification of intermediate mesons originating from B_s^0 decays.

5. Validation of the Full Event Interpretation at $\Upsilon(5S)$

To give a lower boundary on the FEI’s efficiency at the $\Upsilon(5S)$ resonance, the FEI is used to reconstruct on B_s^0 meson in each event. This meson – B_{tag} – is chosen from the list of candidates provided by the FEI by always selecting the candidate with the highest classifier output \mathcal{P}_{FEI} . This is not always the best selection strategy since no information from the second B_s^0 meson – B_{sig} – is used. Here, however, the B_{sig} is not reconstructed to avoid introducing any analysis-specific biases. To estimate the performance of the FEI, the M_{bc} distribution of candidates reconstructed by the FEI is then approximated with analytic expressions for correct B_s^0 candidates and incorrect B_s^0 candidates. The normalization parameters of this approximation can then be used to calculate the tag-side efficiency and purity of selections made with the FEI.

5.1. Continuum Suppression

Any analysis at Belle must consider the impact of non-resonant meson production in which mesons lighter than B mesons are produced. The contribution from these light and charmed mesons is called “continuum”. As particle candidates reconstructed from continuum do not originate from the decay of a single particle, their distribution does not have a clearly defined peak in variables such as M_{bc} . At the $\Upsilon(4S)$ resonance, no measures are taken to reduce the impact of continuum data in the validation of the FEI. Most continuum events are already rejected by the selections applied in the FEI and the resulting B candidates are pure enough to study the FEI’s efficiency. Due to the lower cross section of $e^-e^+ \rightarrow \Upsilon(5S)$, lower branching fraction of $\Upsilon(5S) \rightarrow B_s^0 B_s^0$ and lower number of successfully trained channels, the continuum component has a much higher relative contribution to B_s^0 candidates. To allow validation of the FEI at $\Upsilon(5S)$, additional measures have to be taken to limit the impact of continuum background.

As no B_{sig} is reconstructed, not all common purity-enhancing techniques are available in this validation. To also avoid any selections that would not generalize to all analyses, only topological variables, i.e. variables describing the overall distribution of final-state particles in the detector, are used to suppress the continuum contribution. As all B mesons are created with relatively low momentum their decay products are spread out evenly in the detector. The continuum background consists of decays of lighter mesons with higher

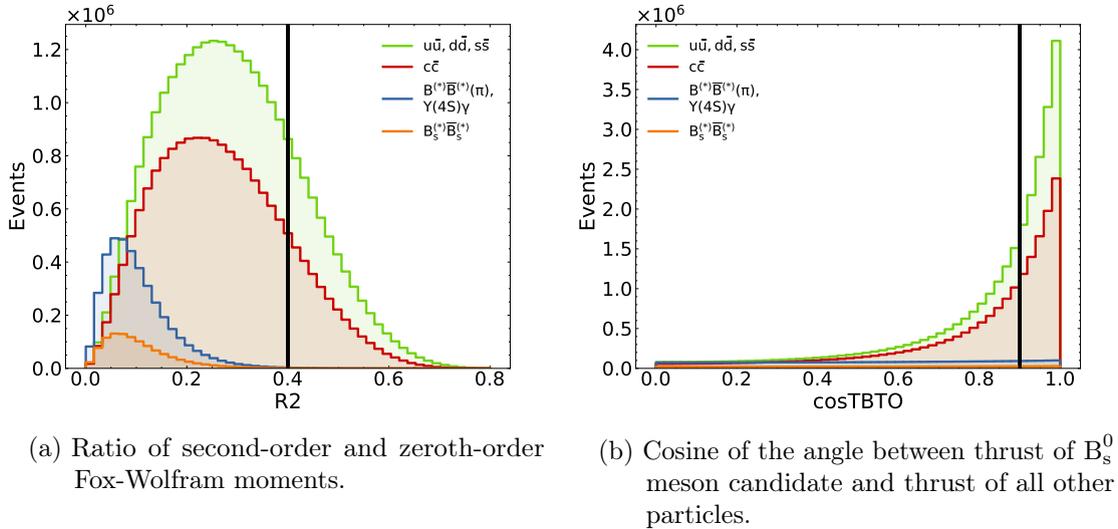


Figure 5.1.: Distribution of topological continuum-suppression variables for all four MC datasets. B_s^0 meson candidates with R2 and $\cos\theta_{B,\text{other}}^{\text{thrust}}$ higher than the values indicated by the black line are discarded.

momentum and thus higher boost of the decay products, which is why the final-state products are often concentrated in two or more regions in the detector. To exploit this shape difference, the event shape is described using rotationally invariant Fox-Wolfram polynomials [19]. The ratio of second-order to zeroth-order polynomials is referred to as R2. Its distribution is shown in Figure 5.1a. Requiring all B_s^0 candidates produced by the FEI to have an R2 value below 0.4 discards 19.3% of continuum background while retaining 99.76% of events containing pairs of B_s^0 , B_u^+ or B_d^0 mesons.

A second way to exploit the shape difference between B events and continuum events is using the cosine of the angle between the thrust axis of the B_s^0 meson candidate and the thrust axis of all other particles in the event. This distribution $\cos\theta_{B,\text{other}}^{\text{thrust}}$ is shown in Figure 5.1b. Requiring all B_s^0 meson candidates to have $\cos\theta_{B,\text{other}}^{\text{thrust}} < 0.9$ discards 85.31% of the remaining background and retains 88.26% of events containing some kind of B meson.

5.2. Performance on Monte Carlo Data

Signal definitions

In simulated data, a Monte Carlo matching algorithm determines whether the hypothesis used to reconstruct a decay chain is consistent with the true chain in which the decay occurred. This consistency is subject to interpretation. In some analyses very exact reconstruction is required while for example in B-tagging entirely correct reconstruction is not always necessary. The default signal definition of `basf2` requires correct hypotheses for a reconstructed particle and all of its final-state daughters. The only exception are photons produced as final-state radiation as these usually carry only a small portion of the

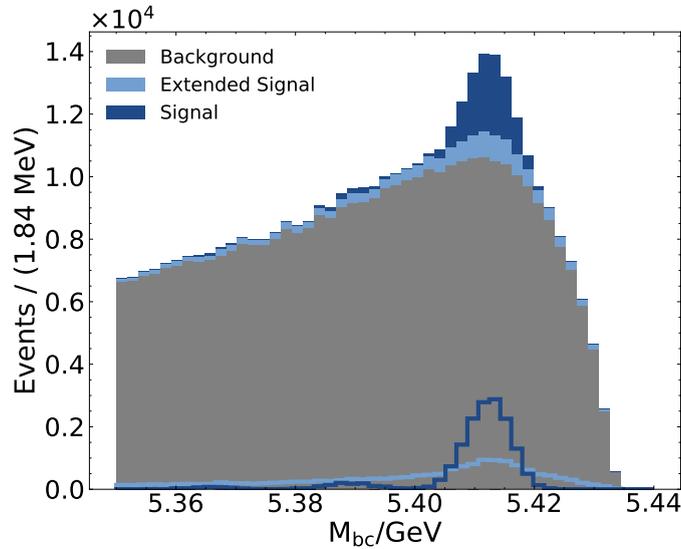


Figure 5.2.: The two signal definitions used in `basf2`'s Monte Carlo matching and their behavior in the FEI at $\Upsilon(5S)$. The “Signal” definition requires entirely correct reconstruction with the exception of missing intermediate resonances and missing photons produced in final-state radiation. The “Extended Signal” definition also allows the misidentification of charged final state particles in addition to the more stringent Signal definition.

overall momentum. A second signal definition, referred to as “extended signal definition”, also allows misidentification of charged final-state particles. As shown in Figure 5.2, this additional contribution peaks in a region similar to the default signal definition and can thus be a useful contribution to the tag-side efficiency of the FEI.

Tag-side efficiency over purity on Monte Carlo data

On Monte Carlo data, the performance of the FEI can be evaluated easily since one has access to the true decay chain of each B meson. By requiring that a reconstructed event matches the generated decay, one can define a sample of correctly reconstructed B_s^0 mesons for which M_{bc} (defined in Section 3.2.1) is then evaluated at different thresholds of the Signal Probability classifier. Figure 5.3 shows two examples of the distribution of M_{bc} at different thresholds of \mathcal{P}_{FEI} . Figure 5.4 shows the tag-side efficiency over the purity of the final FEI training procedure for B_s^0 , B_d^0 and B_u^+ mesons.

Including intermediate resonances

Many B_s^0 meson decay channels include short-lived resonances such as K^* or ρ . The FEI can be used to reconstruct these resonances and use them in turn to reconstruct B_s^0 mesons. In the FEI at $\Upsilon(4S)$, explicit reconstruction of resonant particles is not implemented since the constraint gained by this is small. This is because the resonances do not travel in space so an additional vertex fit cannot increase the purity of the selection. In addition, explicit reconstruction severely decreases the overall efficiency due to the best-candidate selection [29, Appendix C].

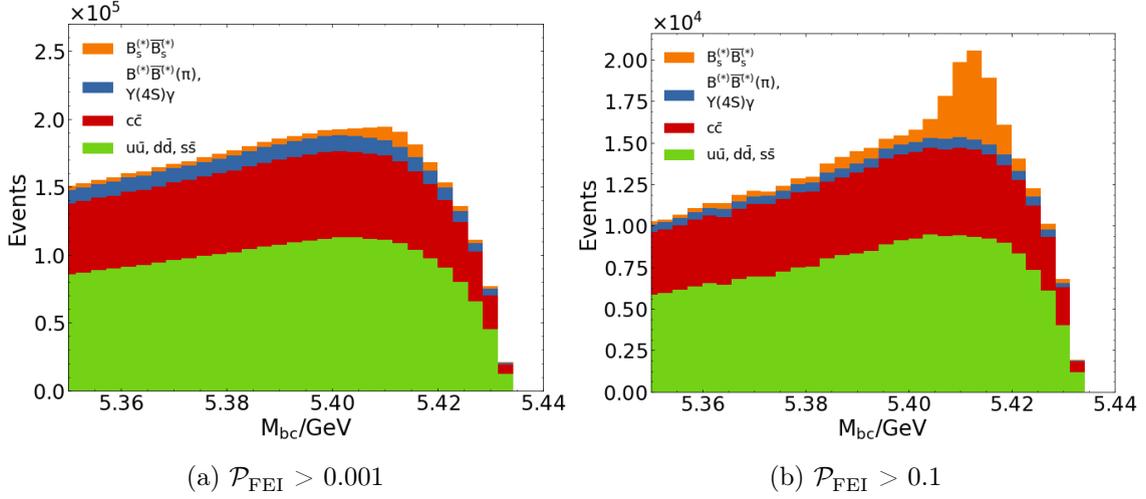


Figure 5.3.: Beam-constrained mass M_{bc} of reconstructed B_s^0 mesons for two different cuts on the FEI classifier \mathcal{P}_{FEI} . For higher values of \mathcal{P}_{FEI} , the desired B_s^0 meson component (orange) is enhanced while the B_d^0 and B_u^+ component (blue) is reduced. The continuum contribution (red, green) is suppressed by requiring $R2 < 0.4$ and $\theta_{\text{thrust}_{B,\text{other}}} < 0.9$ to enhance visibility. Further suppression of the continuum component is analysis-specific so a conservative approach is chosen here.

Whether these assumptions also hold true for the FEI at $\Upsilon(5S)$ must be investigated. Based on previous work by S. Wehle [21], the newly implemented particle hypotheses are ϕ , η , η' , ρ^0 and $K^*(892)$. The effect of implementing these new decay channels is then evaluated on Monte Carlo data. The receiver operating characteristic in Figure 5.5 shows that additional intermediate resonances lead to a significantly lower tag-side efficiency. Due to the large decrease in efficiency the explicit reconstruction of these resonant states is not pursued any further. Instead, their main decay products were added directly as decay channels of higher-level particles.

Impact of training with $\Delta E'$ instead of ΔE

As explained in Section 4.1.3, the final B meson classifier in the FEI at $\Upsilon(4S)$ uses the energy difference ΔE as an important feature. The impact of replacing this feature with the linear combination $\Delta E'$ is evaluated here. As can be seen in Figure 5.6, using $\Delta E'$ increases the tag-side efficiency in all purity bins on Monte Carlo data. Therefore, all further validation was conducted with this feature.

Impact of B_s^0 pre-cut range

Each stage of the FEI is preceded by a set of selection criteria which limit which candidates are recombined to heavier mesons. These are loose constraints which mostly limit required computing resources. They do, however, have a second effect: They limit which incorrect candidates are shown to the final classifier, thereby increasing its efficiency in relevant

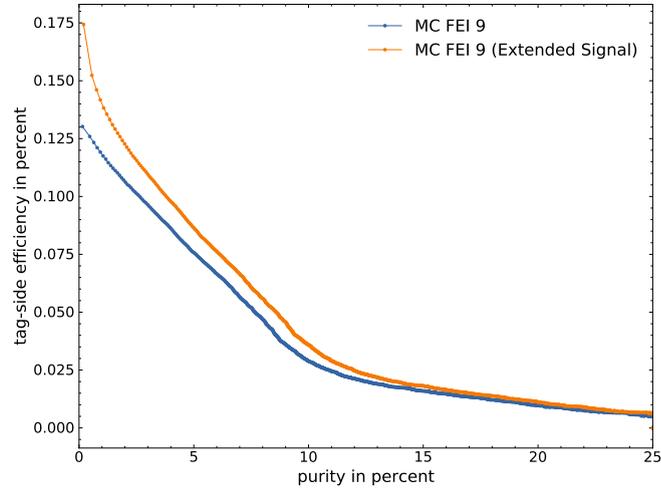
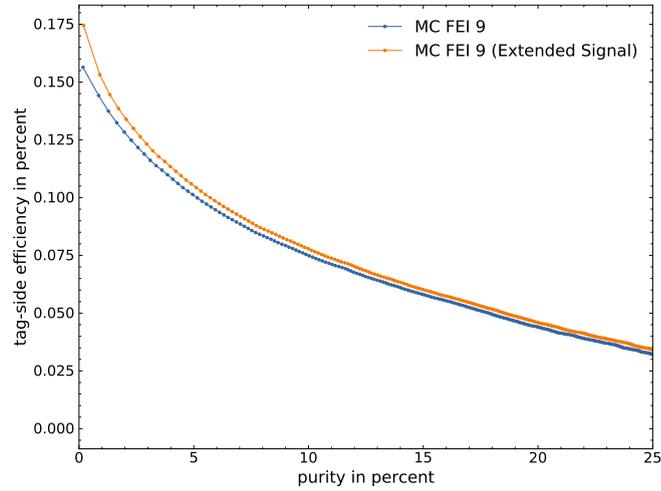
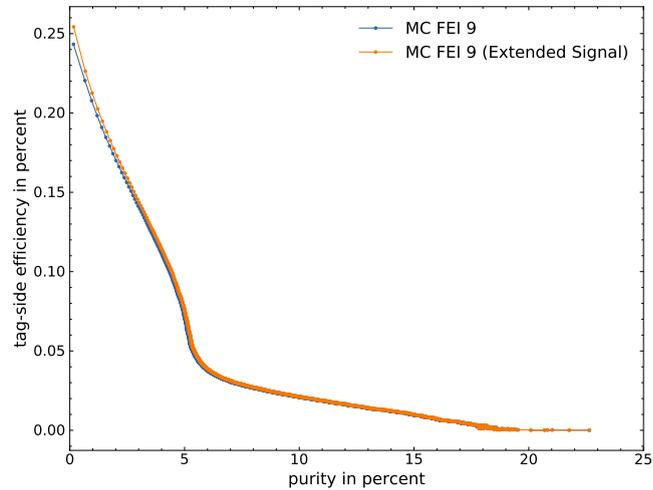
(a) B_s^0 mesons.(b) B_d^0 mesons.(c) B_u^+ mesons.

Figure 5.4.: Tag-side efficiency over purity for reconstructed B mesons on Monte Carlo data. “Extended Signal” refers to a signal definition which in addition also allows misidentification of charged particles, i.e. a pion and a kaon. Since perfect reconstruction of the B_{tag} is not always required, this signal definition is a valid alternative.

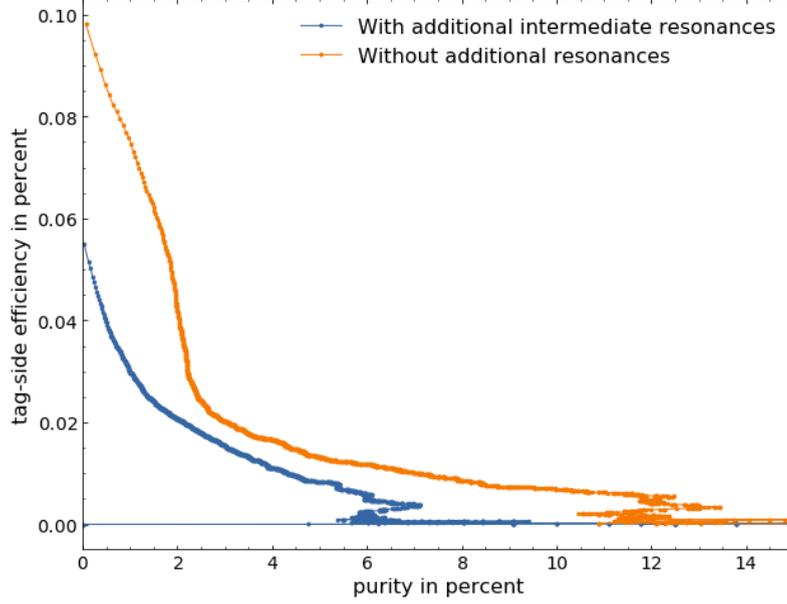


Figure 5.5.: Receiver-operating characteristic on Monte Carlo data of an FEI training with additional intermediate resonances compared to one without them. In low efficiency regions only few B_s^0 are considered which leads to large fluctuations in purity. Both trainings differ from the final training since this path was not pursued any further.

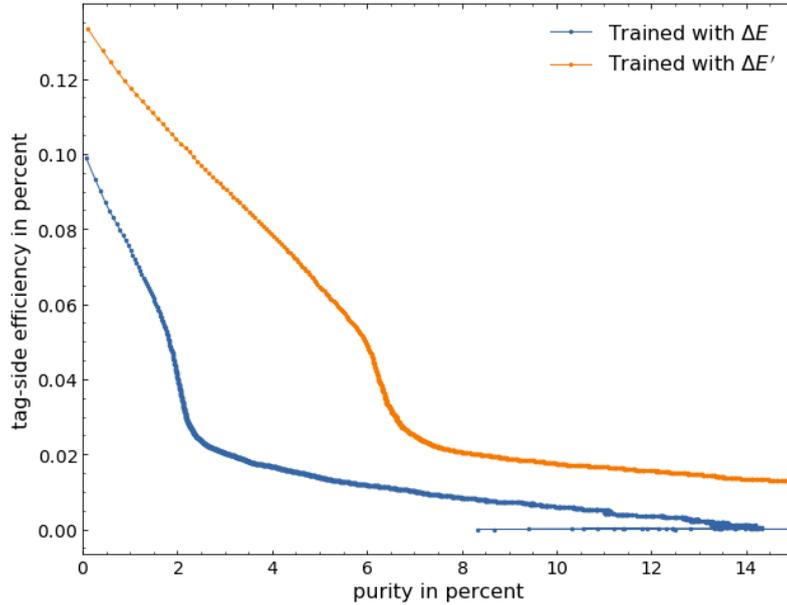


Figure 5.6.: Receiver-operating characteristic of a B_s^0 classifier trained with ΔE feature compared to one trained with $\Delta E'$. Since all three decay configurations of B_s^0 mesons share the same central value of $\Delta E'$, the classifier can use this feature to better distinguish between signal and background.

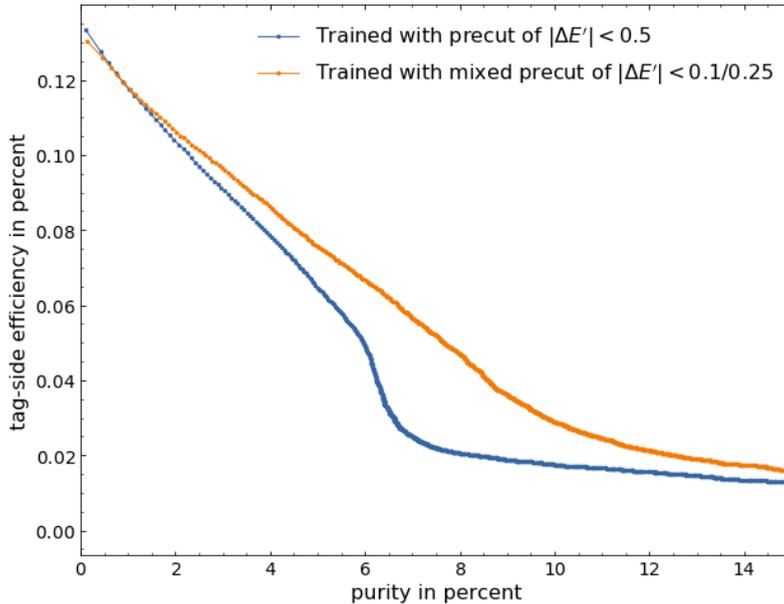


Figure 5.7.: Receiver-operating characteristics of two classifiers trained either with a cut of $|\Delta E' < 0.5|$ and with a mixed cut of $|\Delta E' < 0.1/0.25|$.

regions. For B_s^0 mesons, the pre-cut was initially set to $|\Delta E'| < 0.5$ which is the default for ΔE . Tightening the pre-cut to $|\Delta E'| < 0.25$ increased the signal efficiency slightly (see Figure 5.7), so an even tighter cut of $|\Delta E'| < 0.1$ was investigated. This cut increases the tag-side efficiency in high-purity regions but causes some channels to fail in the training step of the FEI since not enough background data could be provided to train the classifier. The pre-cut lead to a very pure sample which does not contain enough “wrong” candidates to be statistically representative.

Since discarding these channels is undesirable behavior, the pre-cut was loosened for the decay channels $B_s^0 \rightarrow D_s^- K^+$, $B_s^0 \rightarrow D_s^- \pi^+$, and $B_s^0 \rightarrow D_s^{*-} \pi^+$. The impact of the two pre-cut ranges can be seen in Figure 5.7. While the maximum tag-side efficiency is slightly lower for the FEI trained with the tighter pre-cut, the tighter pre-cut significantly improves the behavior in the purity region between 6 and 8%. This behavior makes it easier for analysts to choose a working point on the efficiency-purity plane.

5.3. Performance on Recorded Data

Since the dataset created by Monte Carlo simulation can only approximate the real data, it is important to evaluate the performance of methods like the FEI on real, measured datasets. Data containing approximately 7.11×10^6 decays of $\Upsilon(5S)$ resonances to $B_s^{(*)0} \bar{B}_s^{(*)0}$ pairs is used for this evaluation [6]. This dataset was recorded by the Belle Experiment and is processed with the B2BII conversion module for `basf2`.

To determine the necessary key indicators, namely the tag-side efficiency and purity, the total number of correctly reconstructed B_s^0 mesons in the dataset has to be estimated by

maximum likelihood estimation. Maximum likelihood estimation is a method in which the parameters of an analytical description of a dataset are gained by maximising a cost function corresponding to agreement between the measured values and the analytic function. The method used for maximum likelihood estimation in this work is extended unbinned maximum likelihood estimation (EUML). The unbinned maximum likelihood estimation is able to fully exploit all information in the dataset since the resolution is not decreased by binning. An *extended* unbinned maximum likelihood estimation also includes the size of the sample as part of the result of the experiment and is thus able to account for the variation between different sample sizes [15].

The EUML fit is applied to the beam-constrained mass M_{bc} since correctly reconstructed B_s^0 mesons produce three distinct peaks at around 5.4 GeV in this observable. To validate each component of the eventual fit to the beam-constrained mass in data, the fit is first applied to M_{bc} of selected subsets of Monte Carlo data. Some parameters in the later fit are also fixed to this MC expectation they cannot be estimated in recorded data.

This procedure is then repeated for data with different cut values of the FEI classifier value \mathcal{P}_{FEI} . The normalizations returned by each of these fits can then be used to calculate a point in the tag-side-efficiency–purity plane.

5.3.1. Unbinned Maximum Likelihood Fit in M_{bc}

Fit of correctly reconstructed mesons in MC

To model the correctly reconstructed mesons in M_{bc} , three Gaussian functions corresponding to the three decay configurations to B_s^0 mesons are chosen. The result of the EUML fit can be seen in Figure 5.8.

The optimal fit parameters, namely the means of the Gaussians and their normalizations, are compatible with the values used to generate the Monte Carlo data (see Table 5.1). This indicates that neither the fit nor the FEI are biased towards the main peak.

The centralized MC used for this study is older than the newest measurement of the mass and decay width of B_s^0 mesons by the Belle Experiment [6], therefore the parameters from the fit to MC cannot be used directly in the fit to recorded data. Instead, correction factors of $\mu_{data} - \mu_{MC} = 4.47 \text{ MeV}$ and $\sigma_{data}/\sigma_{MC} = 1.04$, found in internal Belle note 1203 [44], must be applied to the result obtained by the fit to simulated data.

Similar measures have to be taken with regard to the relative normalization of the peaks. While this normalization is not fixed in the fit to Monte Carlo data, the result is not used. Instead, the relative normalization between the three signal peaks is obtained from [6].

Fit of incorrectly reconstructed mesons in MC

The FEI does not always reconstruct B_s^0 mesons correctly. This means that a decay of a B_s^0 meson occurs in an event and a candidate is provided by the FEI for this event but that the decay chain of this candidate contains an error, for example a misidentified final state particle or a particle missing entirely. To model these misreconstructed B_s^0 mesons, the sum of a Crystal Ball function and an ARGUS distribution (both in Ref. [11, Chapter 7.1]) is used. The Crystal Ball function is commonly used to describe lossy processes in high-energy

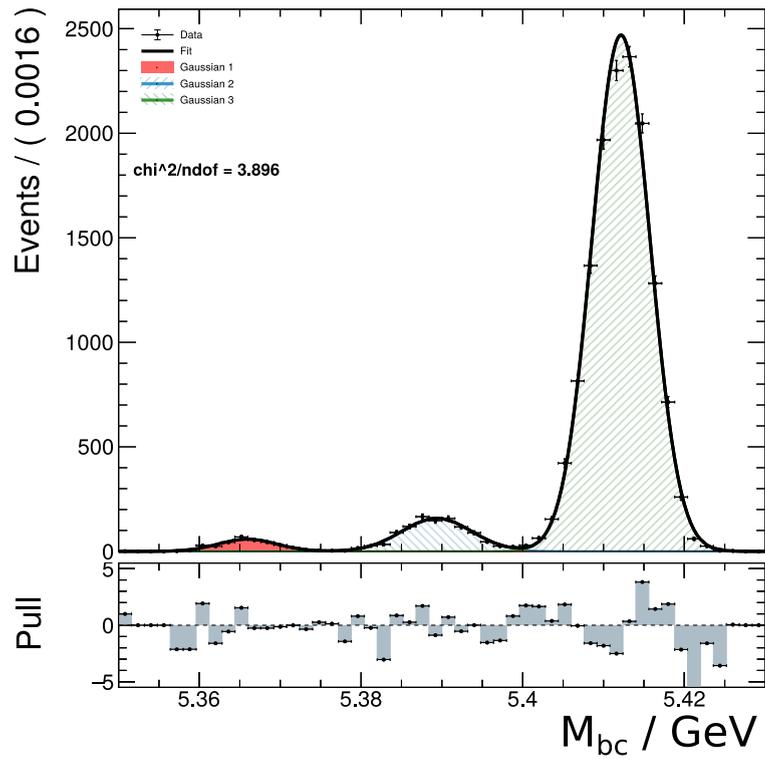


Figure 5.8.: EURL fit to M_{bc} with the sum of three Gaussian distributions to correctly reconstructed B_s^0 mesons in Monte Carlo data with $\mathcal{P}_{\text{FEI}} > 0.1$. All parameters of the Gaussian distributions are left floating and are optimized by the fit.

Table 5.1.: Parameters returned by a fit to M_{bc} of correctly reconstructed particles compared to parameters used to generate the Monte Carlo dataset. Correctly reconstructed particles are selected with the `isSignal` definition from Monte Carlo data. The Monte Carlo dataset was generated with values obtained from [5]. μ_2 has no explicit value in Monte Carlo data as the peak does not originate from a single particle mass and is created by the decay to one B_s^0 and one B_s^{*0} meson. As the smearing caused by the detector resolution is much larger than the decay width of B mesons, the B_s^0 and $B_s^{(*)0}$ mesons are simulated with a simplified decay width of zero.

Parameter	Value from fit	Value in MC
μ_1 / GeV	5.366	5.37
μ_2 / GeV	5.390	-
μ_3 / GeV	5.412	5.41
σ_1 / GeV	0.004	0
σ_2 / GeV	0.004	0
σ_3 / GeV	0.004	0
N1 / %	2.138	2.60
N2 / %	7.004	7.30
N3 / %	90.858	90.10

physics and is appropriate here since missing particles introduce a tail for values of M_{bc} smaller than the Gaussian peak from correctly reconstructed B_s^0 mesons.

The contribution from misreconstructed particle candidates must be estimated entirely from Monte Carlo data since they peak in the same region as the correctly reconstructed particles and are therefore indistinguishable for the purposes of the fit.

Fit of continuum background and other B mesons

Since the distribution of background from B_u^+ and B_d^0 decays and the distribution of continuum background have a similar shape, they are approximated with a single analytic function.

Fit with a Cruiff function

Following the example of the FEI at $\Upsilon(4S)$, an attempt was made to model the continuum background with an ARGUS function from the `proffit` [40] cost-function builder. This proved ultimately unsuccessful since the minimization software `Minuit` [26], accessed via the `iminuit` interface [39], was unable to find a stable minimum. Instead, a Cruiff function from `proffit` is chosen to estimate the non- $B_s^0\bar{B}_s^0$ and continuum contribution. This function is a split Gaussian with non-gaussian tails, defined in Ref. [4] as

$$f(x) = \exp\left\{\frac{(x-m)^2}{2\sigma_{L,R}^2 + \alpha_{L,R}(x-m)^2}\right\}.$$

Here, σ_L and α_L describe the function for $x < m$. For $x > m$ it is described by σ_R and α_R . The Cruiff function has thus five free parameters to be found by the fit. Due to the larger

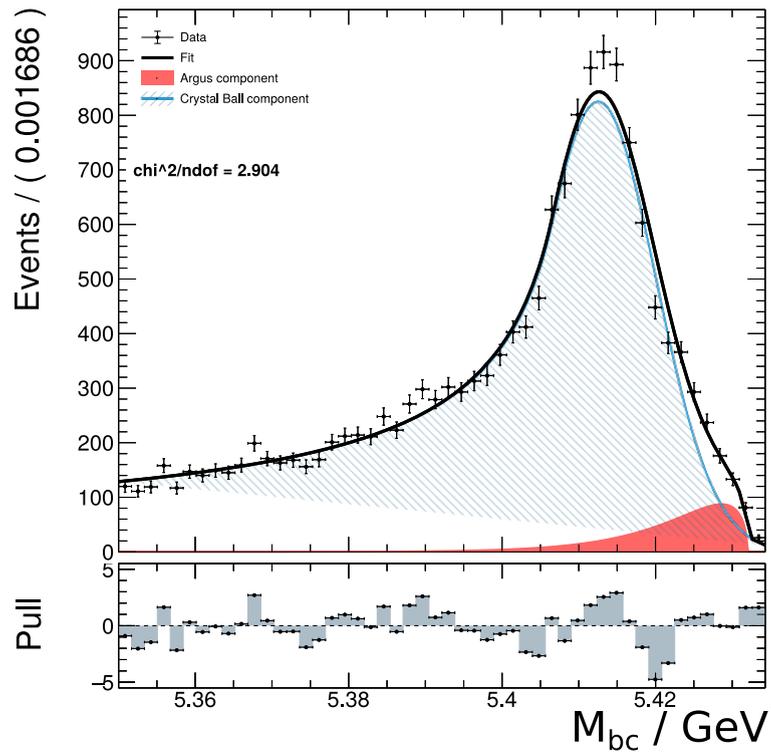


Figure 5.9.: EURL fit to M_{bc} of incorrectly reconstructed B_s^0 mesons in Monte Carlo data with $\mathcal{P}_{\text{FEI}} > 0.1$. All parameters of this distribution in recorded data must be estimated in Monte Carlo data as both the correctly reconstructed and incorrectly reconstructed B_s^0 mesons peak in the same region of M_{bc} .

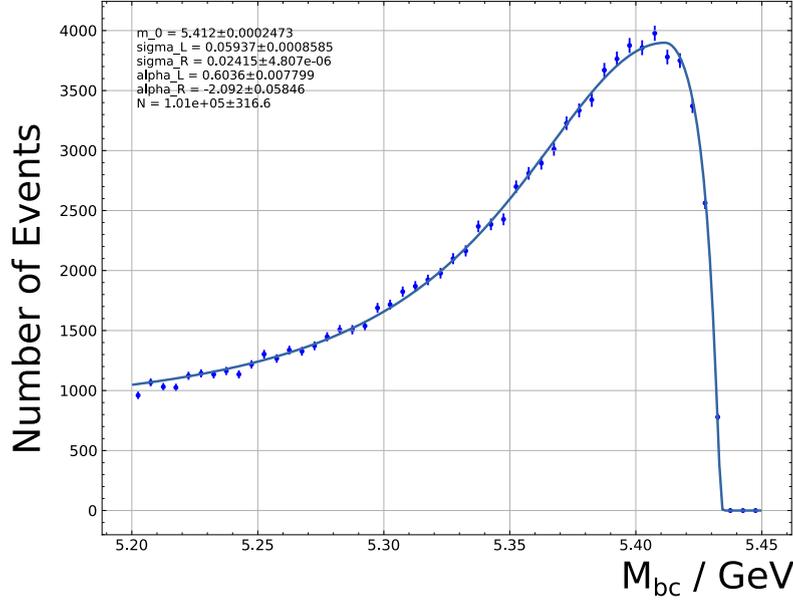


Figure 5.10.: Unbinned maximum likelihood fit to $B_d^0 \bar{B}_d^0$, $B_u^+ \bar{B}_u^-$ and continuum MC background with a Cruijff function. Only events with $\mathcal{P}_{\text{FEI}} > 0.1$, $5.2 \text{ GeV} < M_{bc} < 5.46 \text{ GeV}$, $R2 < 0.4$ and $\theta_{\text{thrust}_{B,\text{other}}} < 0.9$ are considered.

number of parameters, the Cruijff function can also approximate the distribution of M_{bc} in the region between 5.2 GeV and 5.3 GeV.

The result of the successful fit to MC in the range 5.2–5.46 GeV can be seen in Figure 5.10.

While the more common ARGUS distribution is completely described by the shape parameter and the kinematic cut-off, the Cruijff function also has a parameter specifying the location of the peak. This is problematic because the correctly reconstructed B_s^0 mesons peak in the same location and thus the two distributions cannot be fit independently. Instead, one has to assume that both distributions always peak at the same value, an assumption that is not easily verifiable in recorded data.

In principle one could also apply the same method used for misreconstructed B_s^0 mesons and fix the normalization and shape of the Cruijff function from Continuum and NonBsBs MC. However, this is also problematic as the Monte Carlo simulation is usually less accurate for the continuum contribution than for the $B\bar{B}$ contribution.

Fit with an ARGUS distribution in a smaller range of M_{bc}

While the fit with the ARGUS distribution did not succeed with the `probfitt` package in the initially chosen range of M_{bc} , reducing the range from 5.2–5.46 GeV to 5.35–5.43 GeV and switching to the `Roofit` [47] package allowed the fit to converge. As Figure 5.11 shows, the ARGUS distribution cannot completely approximate the behavior of the M_{bc} distribution in the range 5.35–5.39 GeV, however the normalization parameter returned by the fit only deviates by 1.3% from the value obtained by counting. Additionally, the ARGUS distribution is completely described by the shape parameter c_{ARGUS} and the kinematic cut-off $m_{0,\text{ARGUS}}$.

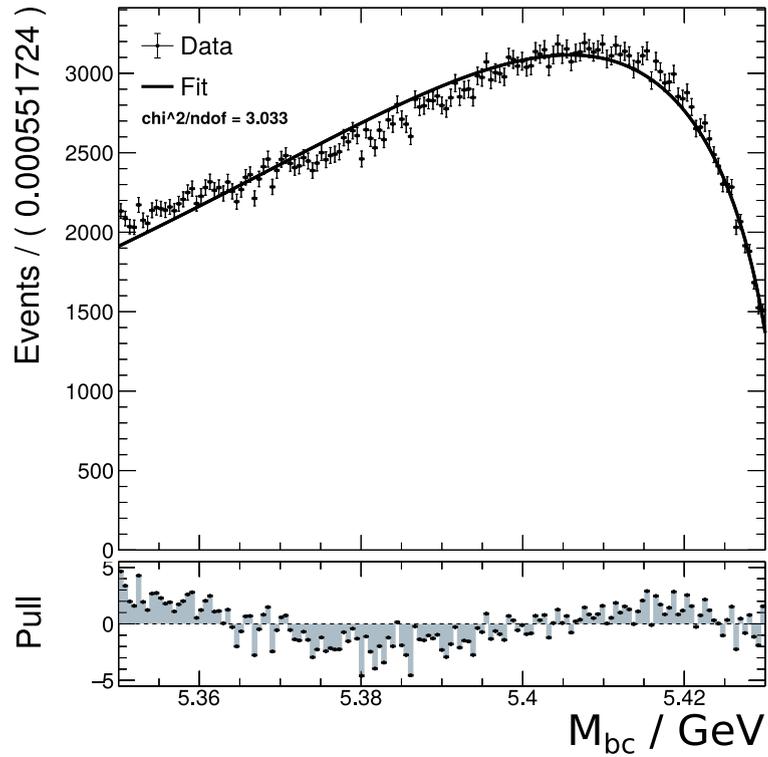


Figure 5.11.: Extended unbinned maximum likelihood fit of $B_d^0 \bar{B}_d^0$, $B_u^+ \bar{B}_u^-$, and continuum MC background with an ARGUS distribution. Only events with $\mathcal{P}_{\text{FEI}} > 0.1$, $5.35 < M_{bc} < 5.43$, $R_2 < 0.4$ and $\theta_{\text{thrust}_{B,\text{other}}} < 0.9$ are considered. The ARGUS distribution cannot completely approximate the background distribution, however the normalization parameter only deviates from the true value by 1.3%.

Table 5.2.: Fixed parameters of an example fit to recorded data in the range 5.375–5.43 GeV. The means μ and the widths σ of the Gaussian distributions are numbered from low to high values of M_{bc} . Parameters with subscripts **cb** and **ARGUS** are parameters of the Crystal Ball and ARGUS functions used to describe the distribution of incorrectly reconstructed B_s^0 mesons. N_{cb} and N_{ARGUS} are the normalization parameters of these functions.

Parameter	Value
μ_1 / GeV	5.37
μ_2 / GeV	5.39
σ_1 / GeV	3.46×10^{-3}
σ_2 / GeV	4.20×10^{-3}
μ_{cb} / GeV	5.42
σ_{cb} / GeV	7.57×10^{-3}
α_{cb}	7.99×10^{-1}
n_{cb}	6.90×10^{-1}
N_{cb}	1.53×10^4
$m_{0,ARGUS}$ / GeV	5.43
c_{ARGUS}	-6.56×10^2
N_{ARGUS}	4.29×10^2

Fit to combined MC and to data

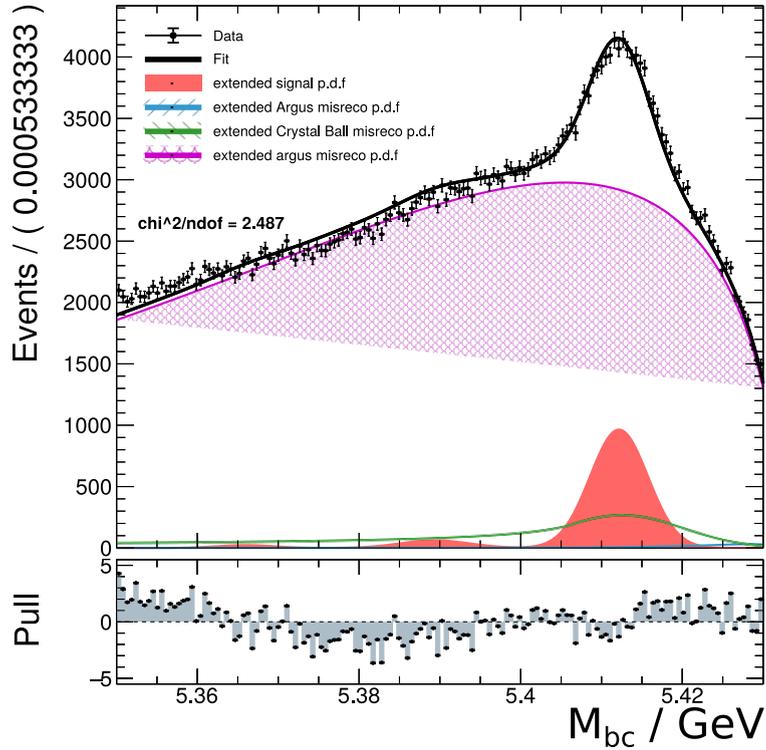
To evaluate the performance of the fitting procedure, the fit is first applied to combined Monte Carlo data from all four categories (BsBs, NonBsBs, Charm, and UDS). The fit function is composed of three components:

Correctly reconstructed B_s^0 are modeled with three Gauss distributions with mean, width, and relative normalization fixed to the MC expectation. The overall normalization parameter is left floating to be determined by the fit.

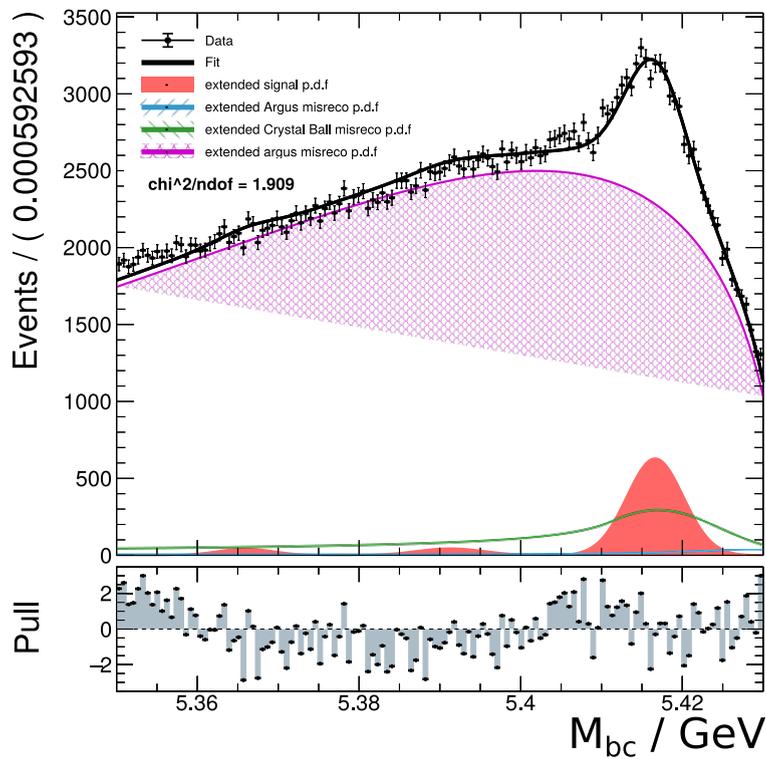
Incorrectly reconstructed B_s^0 are described with an ARGUS and a Crystal Ball function with all parameters fixed to the MC expectation.

$B_d^0 \bar{B}_d^0$, $B_u^+ \bar{B}_u^-$, and Continuum background is described with an ARGUS function with shape and normalization parameters found by the fit. The kinematic cut-off is fixed to 5.432 GeV.

This combined fit function has three free parameters. The result of a fit to the complete $\Upsilon(5S)$ dataset is shown for simulated data in Figure 5.12a and for recorded data in Figure 5.12b. These are only an example for $\mathcal{P}_{FEI} > 0.1$, the fit procedure must be repeated for multiple cuts on \mathcal{P}_{FEI} . For each cut on the classifier, the normalization parameters of the signal and background functions can then be used to gain an efficiency-purity curve (“Receiver-operating characteristic”).



(a) Simulated data



(b) Recorded data

Figure 5.12.: EURL fits to simulated data combined from all Monte Carlo samples and to data. The fit function is the sum of three Gauss distributions (red) with mean and width fixed from a fit to MC, an ARGUS (blue) and a Crystal Ball (green) function entirely fixed from the fit to MC, and an ARGUS function (pink) with floating shape and normalization.

Table 5.3.: Floating parameters of example fits in the range 5.375–5.43 GeV to combined simulated data and to recorded data. Only the overall normalization parameter of the signal distribution and the shape and normalization of the background distribution are varied in the fit. All other parameters are obtained from Monte Carlo. The relative normalization between the three signal peaks is obtained from [6].

Parameter	Value in MC	Error	Value in data	Error in data
c_{bkg}	-5.06×10^1	2.34×10^{-1}	-4.48×10^1	2.80×10^{-1}
N_{bkg}	3.78×10^5	7.35×10^2	2.94×10^5	6.56×10^2
N_{sig}	1.79×10^4	4.13×10^2	1.08×10^4	3.70×10^2

Fitting only two peaks in M_{bc}

While the fit to all three peaks converges in general, its χ^2 value is usually high since the ARGUS function used to approximate the continuum background is not able to completely describe the beam-constrained mass in the range 5.35–5.40 GeV. Reducing the range even further to 5.375–5.43 GeV leads to much higher goodness-of-fit values as shown in Figure 5.13. Since the peak created by the decay to two ground-state B_s^0 mesons is below the new M_{bc} threshold, it is now not considered in the fit. This results in a lower signal yield, however the effect on the overall tag-side efficiency is small as only $\approx 5.7\%$ of all decays to $B_s^0 \bar{B}_s^0$ occur without an intermediate resonance. Figure 5.14a shows a fit to this smaller range. Neglecting the $B_s^0 \bar{B}_s^0$ peak also has the added benefit of more reliable fitting in low-purity regions, i.e. for loose cuts on \mathcal{P}_{FEI} . In these regions, separating the $B_s^0 \bar{B}_s^0$ peak from background fluctuations is difficult and the fits become unstable. The fixed parameters in this fit are shown in Table 5.2, the three parameters optimized by the fit are shown in Table 5.3.

In all further studies, the fit to two peaks in the smaller range is used as this procedure has a higher goodness-of-fit. The individual components of this fit in the smaller range can be found in Appendix A.

5.3.2. Estimating the Total Tag-side Efficiency and Purity

To validate the method used to obtain the tag-side efficiency and purity, it is first applied to the simulated dataset. The tag-side-efficiency–purity-curve obtained in this way is then compared to the one from Monte Carlo truth, i.e. obtained by requiring the reconstructed event to be identical (isSignal) or almost identical (isExtendedSignal) with the simulated event. As evident from Figure 5.15, the method provides a good estimate of the tag-side efficiency in regions of purity above 10% but overestimates the efficiency for purities below 5% as the fit cannot separate the peaking background contribution and the $B_s^{(*)0} \bar{B}_s^{(*)0}$ signal peak. Stronger continuum suppression can mitigate this problem and increase the purity but given that no B_s^0 signal channel is selected in this method, the available methods are limited and could lead to a misrepresentation of the actual tag-side efficiency.

Using the same method, i.e. fitting the signal peaks and using the normalizations to calculate the tag-side efficiency and purity, the performance of the FEI is now estimated. Figure 5.16

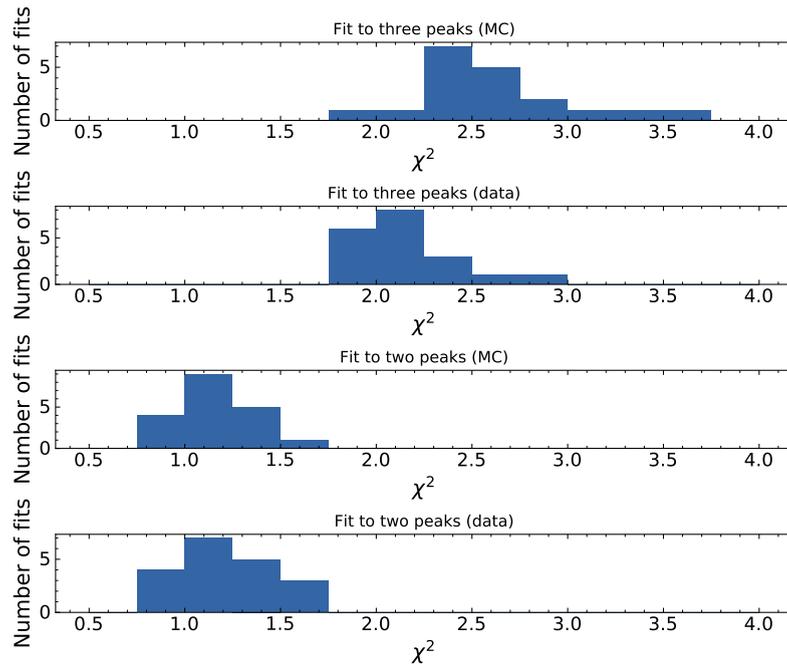
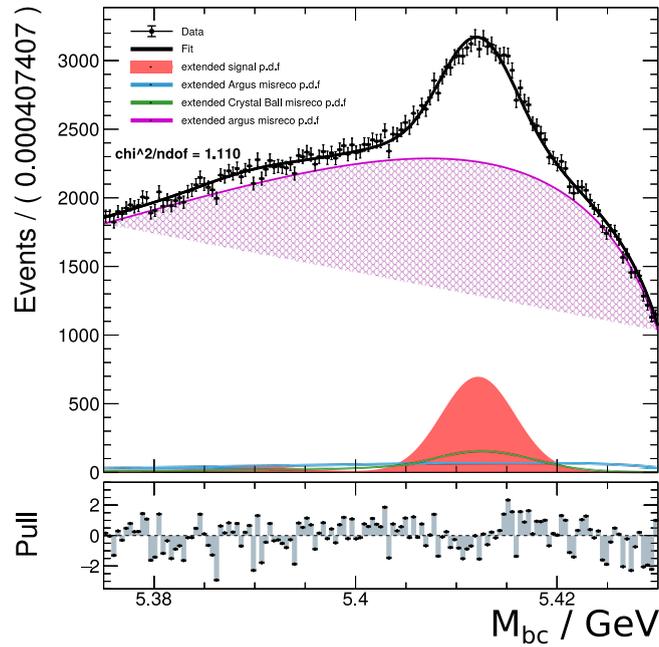
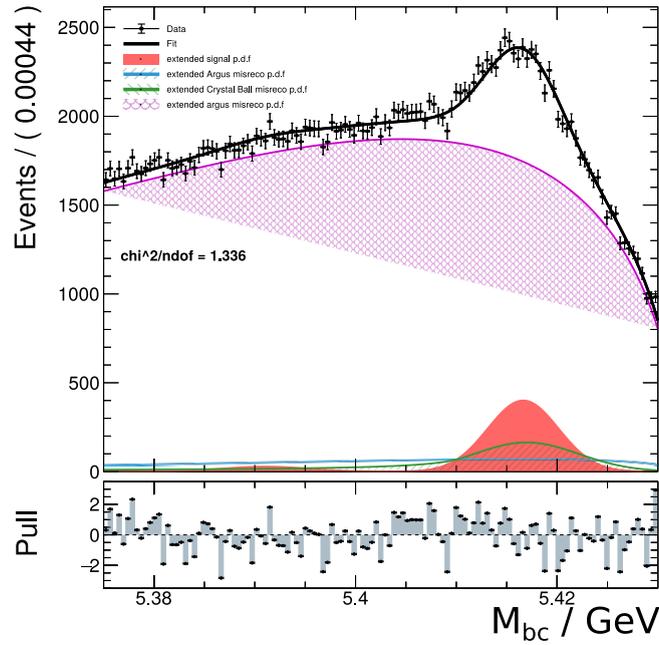


Figure 5.13.: χ^2 value of EURL fits to M_{bc} for different cuts on \mathcal{P}_{FEI} . The top two histograms show fits for the range 5.35–5.43 GeV (to Monte Carlo data and to recorded data). The bottom two histograms show fits for the range 5.375–5.43 GeV (again to MC data and recorded data).



(a) Simulated data



(b) Recorded data

Figure 5.14.: EURL fits to simulated data combined from all Monte Carlo samples and to data in the range 5.35–5.43 GeV. The fit function is the sum of two Gauss distributions with mean and width fixed from a fit to MC, an ARGUS and a Crystal Ball function entirely fixed from the fit to MC, and an ARGUS function with floating shape and normalization.

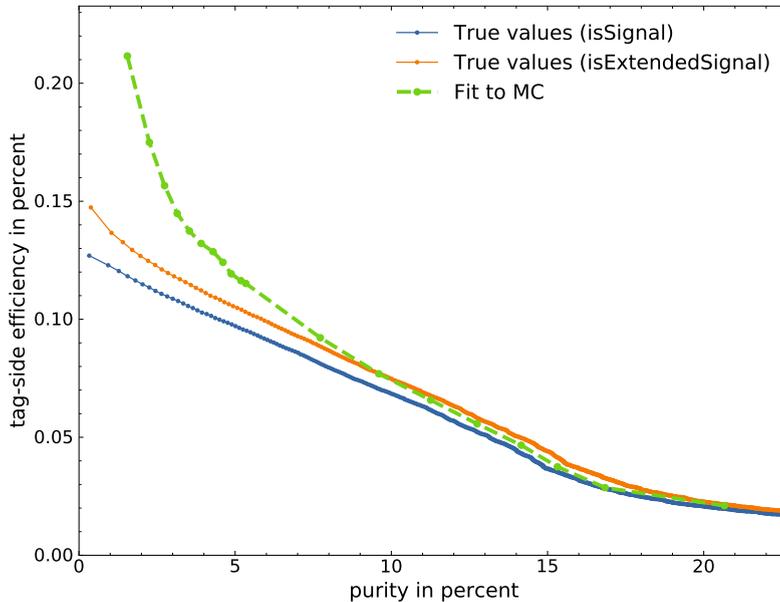


Figure 5.15.: Comparison of tag-side efficiency over purity obtained by fit and by counting, both in Monte Carlo data. Agreement between the curves is good above 10% purity. For purities below 5% the fit overestimates the efficiency significantly.

shows the two curves as estimated from data. As is apparent from this figure, there is a large difference between the curves for Monte Carlo data and recorded data. A part of this difference can be attributed to the FEI itself. As the method is trained on Monte Carlo data which can only approximate the kinematic distributions found in recorded data, it will always perform better on Monte Carlo data. Another part can certainly be attributed to the fitting method which, especially in low-purity regions, is not able to completely separate continuum background and B_s^0 mesons.

5.3.3. Estimating the tag-side efficiency and purity of individual channels

To investigate the difference between data and MC, the tag-side efficiency and purity of individual channels is also investigated. This is only possible for some channels as the fit cannot reliably approximate M_{bc} in channels with very low branching fractions or low reconstruction efficiencies. The fit becomes unstable, especially in low-purity regions. Figure 5.17 shows five channels for which individual fits are successful. Their behavior differs significantly but in four out of five channels, the efficiency on Monte Carlo data is higher than on recorded data. This is consistent with the trend in the complete dataset.

However, as is evident from Figure 5.17e, estimating the tag-side efficiency in regions below 5% purity becomes difficult. The fit assigns almost all peaking behavior to the continuum contribution as it cannot distinguish between signal and background anymore.

5.3.4. Performance on Off-Resonance Data

To evaluate the FEI's performance on continuum background, it is applied to off-resonance data. This is data recorded with the beam energy set to 60 MeV below the $\Upsilon(4S)$ resonance,

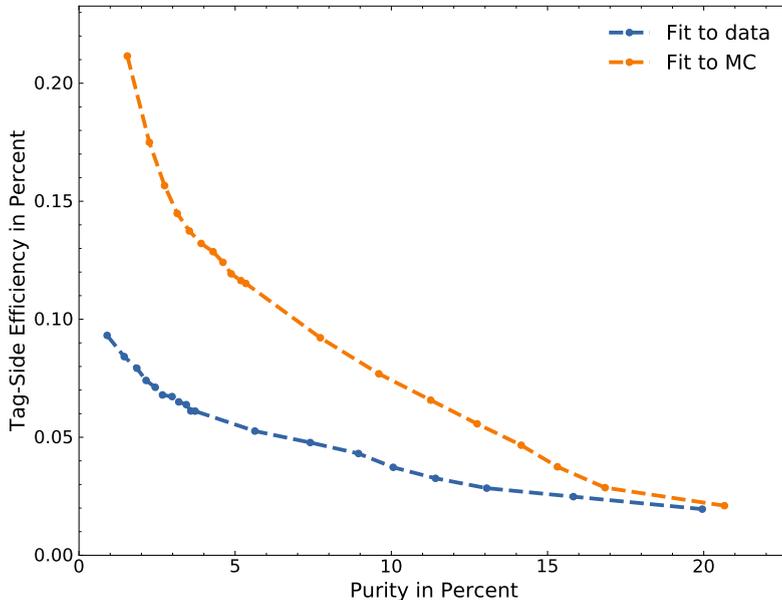


Figure 5.16.: Comparison of tag-side efficiency over purity for Monte Carlo data and measured data. Both curves are obtained with EUML fits to M_{bc} as described in Section 5.3.1. The efficiency on independent Monte Carlo data is higher as the FEI classifiers are trained on MC and not on recorded data. Tag-side efficiency and purity are defined in Section 3.2.1.

i.e. 350 MeV below the $\Upsilon(5S)$ resonance. This difference must be accounted for in the beam-constrained mass, it is scaled by the ratio of the two beam energies:

$$M'_{bc} = \frac{10.87 \text{ GeV}}{10.52 \text{ GeV} - 0.06 \text{ GeV}} M_{bc}. \quad (5.1)$$

In addition to the scaling of M_{bc} , the data is also scaled by the luminosity factor to account for the different sizes of the off-resonance sample and the $\Upsilon(5S)$ sample. This factor is

$$N_{\text{lumi}} = \frac{121.06 \text{ fb}^{-1}}{20.26 \text{ fb}^{-1}} = 5.97. \quad (5.2)$$

As can be seen from Figure 5.18, the FEI is not negatively influenced by the lower beam energy. No peaking behavior is visible and by counting, the number of events deviates by 27.29 % from the on-resonance case in Monte Carlo.

5.4. Calibration attempt with $B_s^0 \rightarrow X l \nu$

Calibration of the FEI's performance is necessary to correct for the differences between selection efficiencies when applying the FEI to simulated and recorded data. This is done by deriving a calibration factor ϵ from the ratio between the number of signal events in

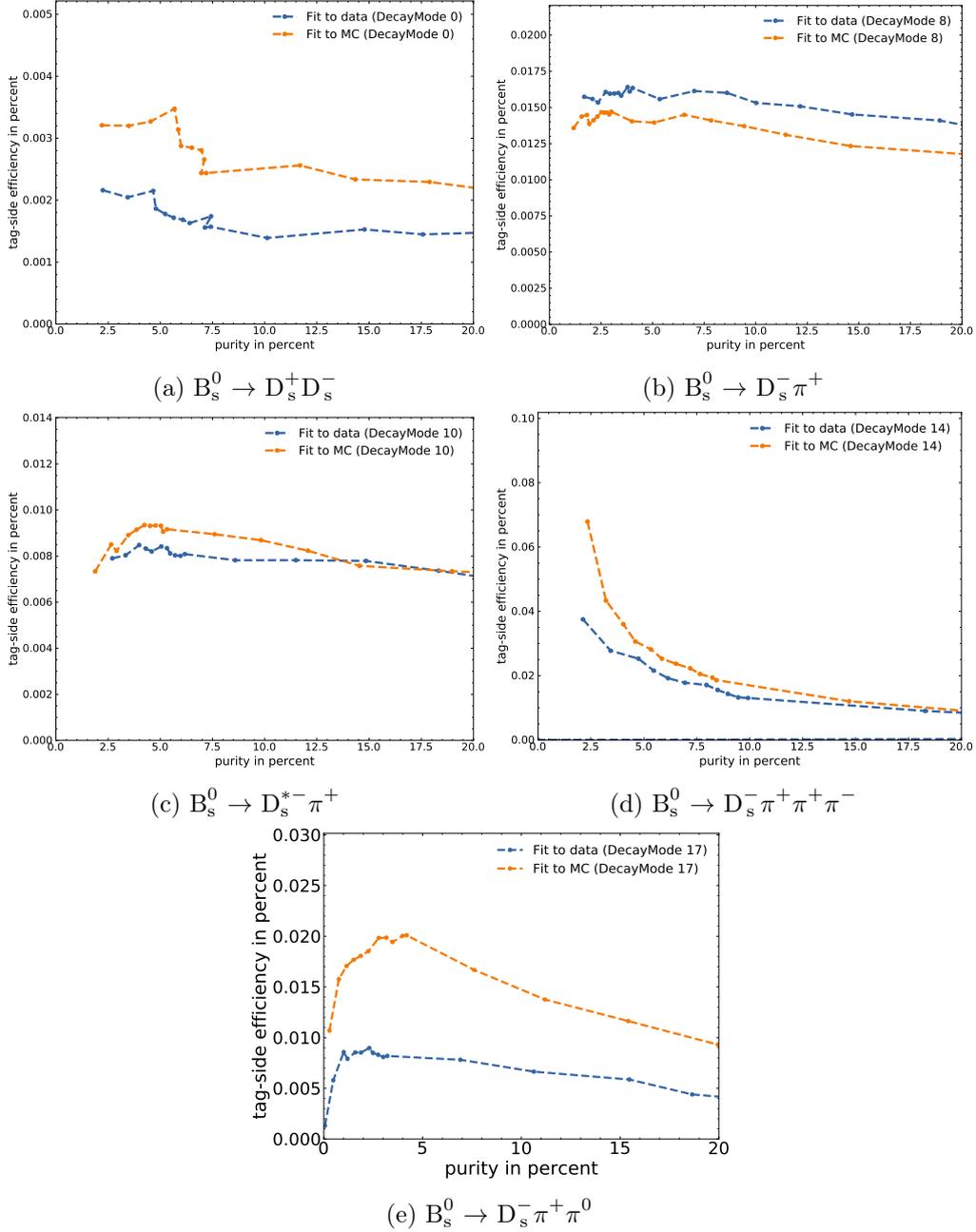


Figure 5.17.: Tag-side efficiency over purity for individual channels.

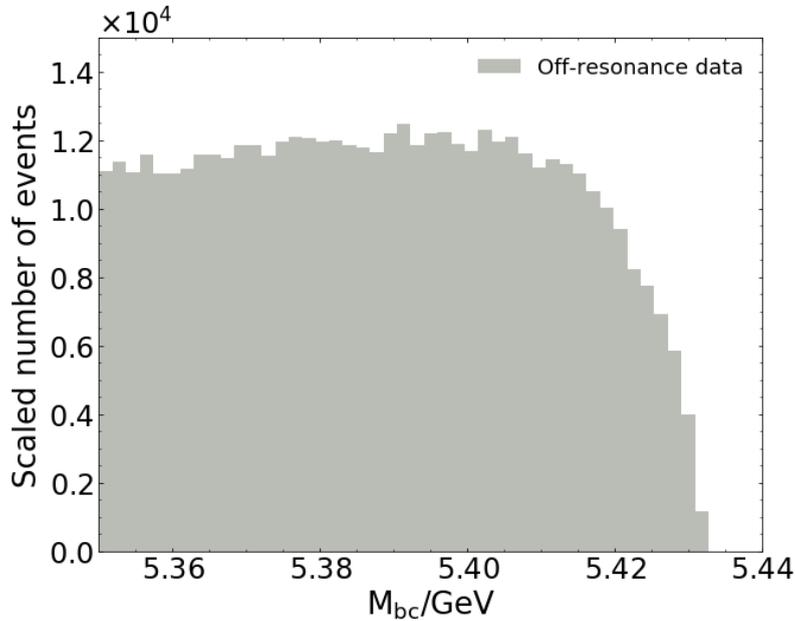


Figure 5.18.: Off-resonance data as shaped by the FEI selection process, scaled to account for the lower beam energy and luminosity.

Monte Carlo data N_{MC} and the number of signal events in recorded data N_{data} . N_{data} must be derived from a fit to data, N_{MC} can be obtained from a fit to an independent MC sample or by counting in combination with requiring correct reconstruction via Monte Carlo matching. For the FEI at $\Upsilon(4S)$, the calibration factor was derived in Ref. [43] using the Missing Mass Squared of several semileptonic B meson decay channels. The calibration uses exclusive signal definitions for each of these channels which allows the use of the completeness constraint by rejecting events with any additional final-state particles. This procedure is not available to the FEI at $\Upsilon(5S)$. The limited size of the $\Upsilon(5S)$ dataset at the Belle Experiment allows only inclusive signal-side measurements for which no completeness constraint can be applied.

To separate between signal and background contributions, the momentum of the signal-side lepton in the CMS frame p_{lep}^* is chosen. To extract the signal contribution, three different distributions are modeled with templates. The selections used in the event reconstruction are listed in Table 5.4. While leptons are accepted if their momentum in the CMS frame is above 0.3 GeV, the template fit does not converge in this region. To mitigate this, the fit is constrained to the region $p_{\text{lep}}^* = 1.2\text{--}2.5$ GeV.

The signal template is constructed by requiring the signal-side lepton to originate directly from the decay of a B_s^0 or \bar{B}_s^0 meson. One background template is constructed by inverting this criterion in the $B_s^0\bar{B}_s^0$ and $B_u^+\bar{B}_u^-/B_d^0\bar{B}_d^0$ MC samples. A second background template is constructed from Continuum MC as a different shape is expected for this background. Alternatively, this second background template is constructed from off-resonance data which by definition cannot contain any signal leptons. The off-resonance dataset is also a more reliable approximation of the continuum contribution than the Monte Carlo dataset as

Table 5.4.: Selections applied to reconstruct the B_{tag} and a lepton. Variables marked with * are in the CMS frame.

Component	Variable	Cut
B_{tag}	\mathcal{P}_{FEI}	> 0.01
	BCS with highest \mathcal{P}_{FEI}	
B_{tag} ROE mask	p^* / GeV	≤ 3.2
	nCDCHits (tracks)	> 0
	p / GeV (cluster)	≥ 0.05
Continuum suppression	R2	< 0.4
	$\theta_{\text{thrust}_{B,\text{other}}}$	< 0.9
lepton	dr / cm	< 0.5
	dz / cm	< 2
	(e mu)IDBelle	$> (0.5 0.9)$
	p^* / GeV	> 1.2
	BCS with highest p^*	

described in Section 5.3.4. The off-resonance template is scaled with the luminosity factor N_{lumi} but the lepton momentum is not scaled to account for the lower beam energy as the effect of this scaling is not well studied. The normalization of these templates is then varied in a maximum likelihood fit to approximate the distribution in data. An additional Gaussian constraint is placed on the normalization of the continuum template as otherwise the fit completely removes the signal contribution.

Using a continuum template from off-resonance data

This method generally has lower systematic uncertainties as the off-resonance template describes the continuum contribution at the resonance better than the continuum MC template. The off-resonance template is, however, much smaller than the on-resonance sample and must therefore be scaled with N_{lumi} . This also scales statistical fluctuations which do not exist in the much larger on-resonance dataset. This effect is visible in Figure 5.19. Additionally, the region in which the background normalization is constrained is much larger in the scaled dataset as the standard Poisson error for counting is scaled by the luminosity factor. Together, these statistical limitations result in very different calibration factors depending on the number of bins used in the fit (see Table 5.5).

Using a continuum template from Monte Carlo data

As Figure 5.20 shows, the template generated from Monte Carlo data does not suffer from the same statistical limitations as the off-resonance template. Table 5.5 shows that the calibration factors calculated with this method are much less dependent on the number of bins in the fit. The mean calibration factor achieved with this method is

$$\eta = 1.67. \quad (5.3)$$

Further studies are needed to find a reliable calibration factor for the FEI at $\Upsilon(5S)$.

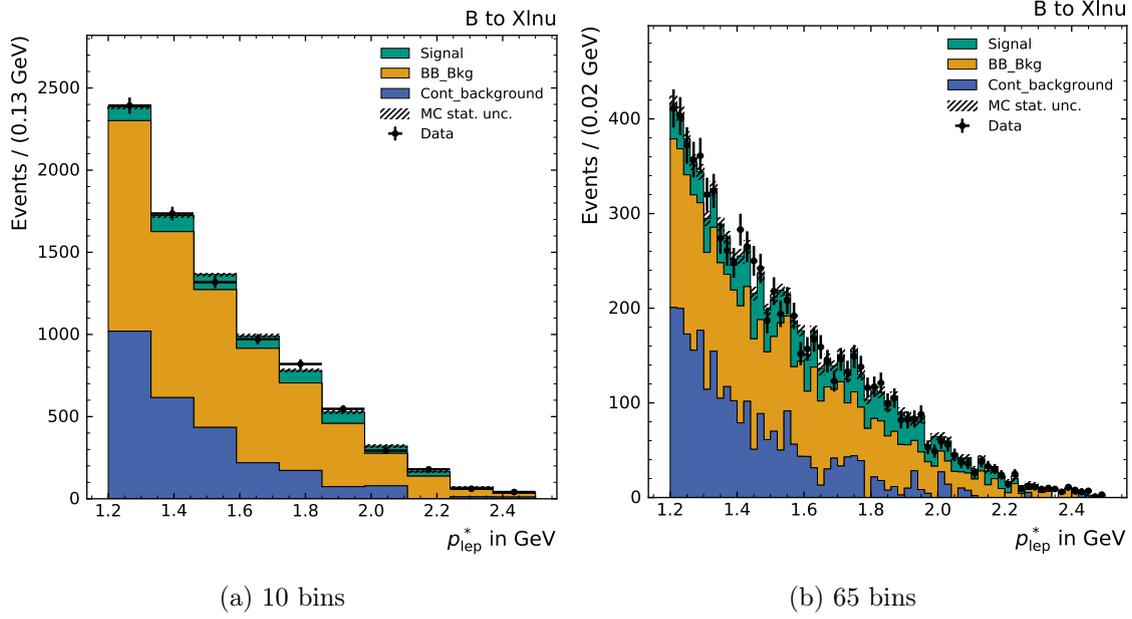


Figure 5.19.: Fit to p_{lep}^* with two templates constructed from Monte Carlo data and a continuum template from scaled off-resonance data. The continuum normalization is constrained with $\sigma = \sqrt{N_{\text{offres}}} \cdot N_{\text{lumi}}$. The fit is repeated for different numbers of bins to evaluate its stability.

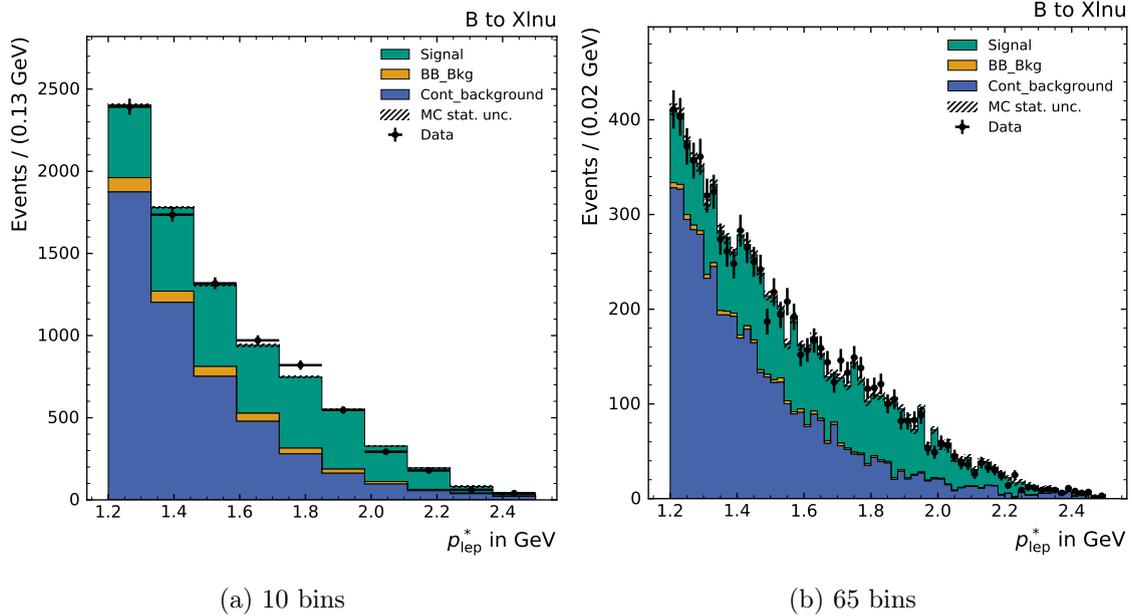


Figure 5.20.: Fit to p_{lep}^* with three templates constructed from Monte Carlo data. A Gaussian constraint with $\sigma = \sqrt{N_{\text{cont}}}$ is placed on the normalization of the continuum template.

Table 5.5.: Calibration factors calculated with template fits to p_{lep}^* . “MC” and “Off-Res.” refer to different continuum background templates. The number of bins N_{bins} is varied to test the stability of the procedure, the calibration factor should be independent of N_{bins} . The large standard deviation of factors calculated with an off-resonance template indicates that this is not the case here. The specified uncertainties are statistical uncertainties returned by the fit, systematic uncertainties are not determined.

N_{bins}	MC	Off-Res.
10	1.73 ± 0.19	0.33 ± 0.24
15	1.79 ± 0.20	0.42 ± 0.25
20	1.61 ± 0.14	0.95 ± 0.19
25	1.49 ± 0.18	0.76 ± 0.18
30	1.77 ± 0.20	0.61 ± 0.13
35	1.66 ± 0.19	0.35 ± 0.26
40	1.54 ± 0.17	0.90 ± 0.19
45	1.75 ± 0.20	0.76 ± 0.20
50	1.53 ± 0.20	0.64 ± 0.18
55	1.73 ± 0.19	0.88 ± 0.18
60	1.63 ± 0.21	0.73 ± 0.19
65	1.86 ± 0.20	0.91 ± 0.22
mean	1.67	0.69
std.	0.12	0.22

6. Conclusion and Outlook

To obtain as much information as possible from the decay of two B mesons, B-tagging algorithms like the Full Event Interpretation are used. This algorithm uses multivariate methods to reconstruct tag-side B mesons. Physicists are then able to combine these tag-side B mesons with B mesons reconstructed in a signal decay channel and use the additional information from the tag-side to measure decays on the signal-side. So far, the Full Event Interpretation has only been used to reconstruct B mesons created in the decay of $\Upsilon(4S)$ bottomonium resonances.

This master's thesis presents the application of the Full Event Interpretation to the $\Upsilon(5S)$ resonance and the validation of this method. The newly implemented decay channels will allow measurements of decay processes originating from B_s^0 , B_d^0 or B_u^+ mesons. Several improvements of to the initial implementations were made that increased the efficiency in Monte Carlo data significantly.

The implementation was validated with 121.4 fb^{-1} of data recorded at the $\Upsilon(5S)$ resonance by the Belle Experiment. Using the method presented in this thesis, the efficiency of the reconstruction of B_s^0 mesons was evaluated and a maximum tag-side efficiency of

$$\epsilon_{tse} = 0.1 \% \tag{6.1}$$

was determined.

Using the decay $B_s^0 \rightarrow X\ell\nu$, an efficiency calibration of the FEI was attempted. A calibration factor of $\eta = 1.67$ was found using a simulated continuum template but further studies using off-resonance templates are needed to obtain a reliable result.

While the work presented in this thesis proves the viability of the FEI at $\Upsilon(5S)$, there is still potential for improvement. A few options are mentioned below:

Measures to allow training of all channels The FEI was tested and validated with all available centrally produced B_s^0 Monte Carlo samples. This sample was, however, not big enough to train classifiers for all implemented decay channels. Generating a Monte Carlo sample 2–3 times as large as the current centrally produced Monte Carlo sample might allow training of channels for which currently not enough B_s^0 mesons are reconstructed correctly. For pure channels containing J/ψ mesons, other measures have to be found. These could include better sampling of meson candidates or less stringent pre-selections.

Better continuum suppression Analyses at beam energies beyond $\Upsilon(5S)$ suffer from increased continuum backgrounds whose impact has to be mitigated. Ways to improve the validation presented in this thesis could be the use of multivariate continuum suppression methods. This would introduce a dependency on the multivariate method but could improve the stability of the validation procedure.

Improved calibration procedure The calibration attempt presented in this thesis suffered from statistical fluctuations in the continuum template which are caused by the limited size of the available off-resonance sample. A better fitting method could try to smooth this template to mitigate this effect. Alternatively, the Monte Carlo continuum template could be reweighted to better match the distribution in data.

Larger recorded dataset Belle II, Belle's successor, could record a larger dataset at the $\Upsilon(5S)$ resonance. As the FEI is already implemented in `basf2`, this future dataset could be easily analyzed with the software presented in this thesis. As B-tagging analyses are usually statistically limited, a larger dataset could enable many measurements with the FEI.

A. Components of fit with reduced range

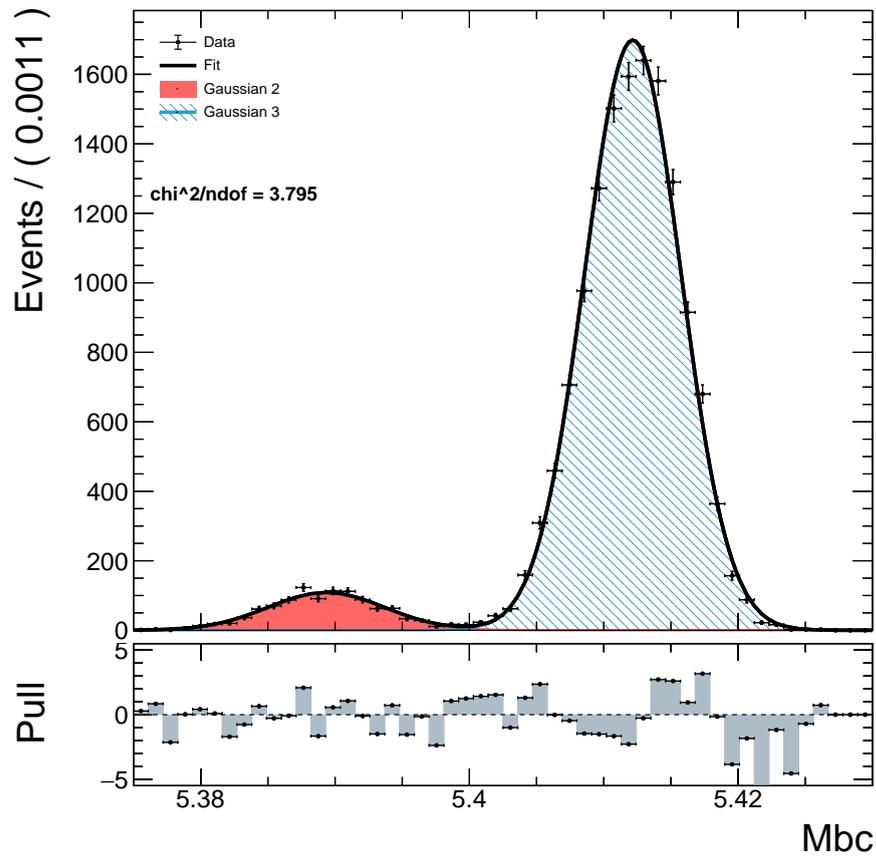
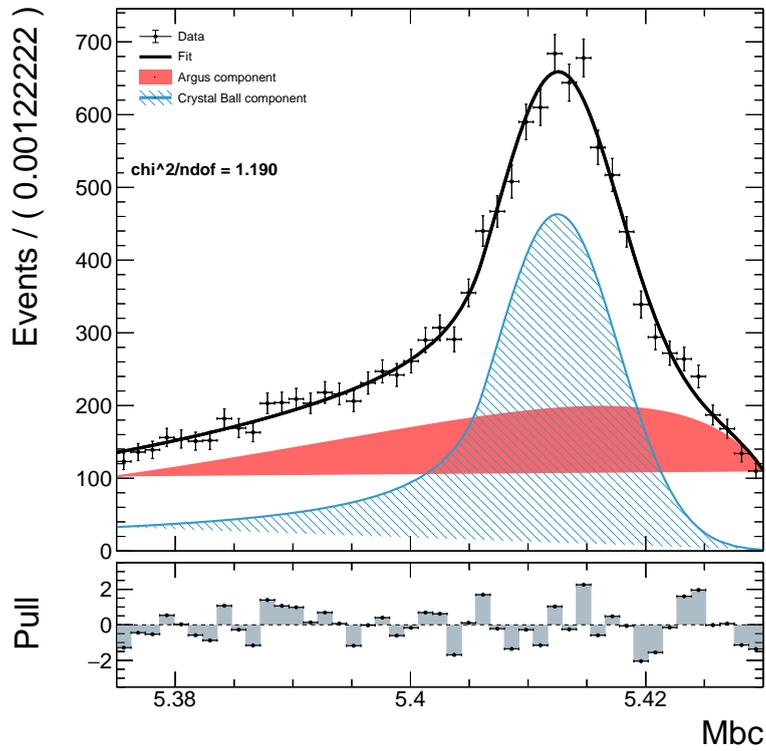
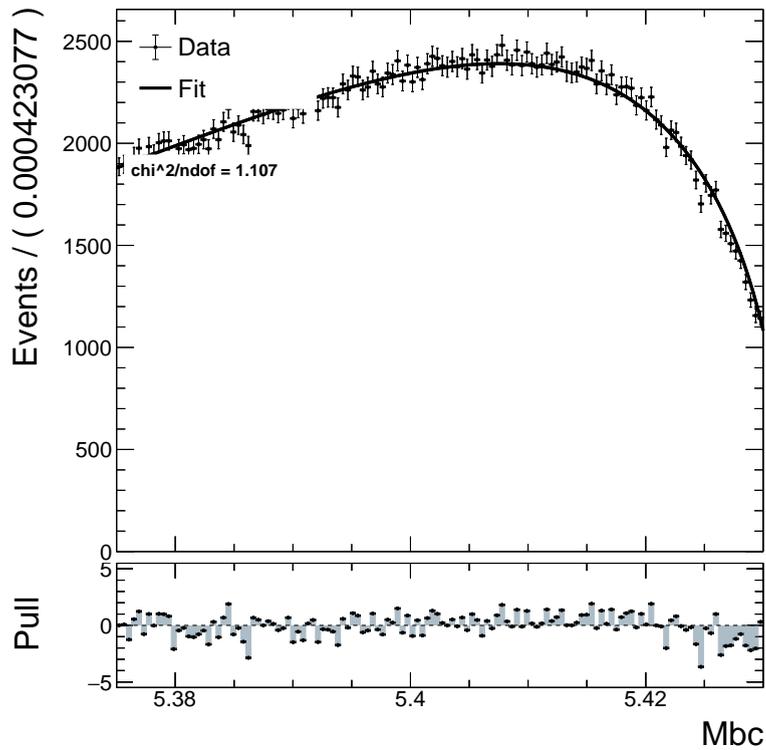


Figure A.1.: Signal component of the fit consisting of two Gaussian distributions in a reduced range of M_{bc} .



(a) Misreconstructed $B_s^{(*)0}$ component of the fit, described by an ARGUS function and a Crystall Ball function.



(b) Continuum and $B\bar{B}$ background component of the fit, described by an ARGUS function.

Figure A.2.: Background components of the fit to Monte Carlo data in a reduced range of M_{bc} .

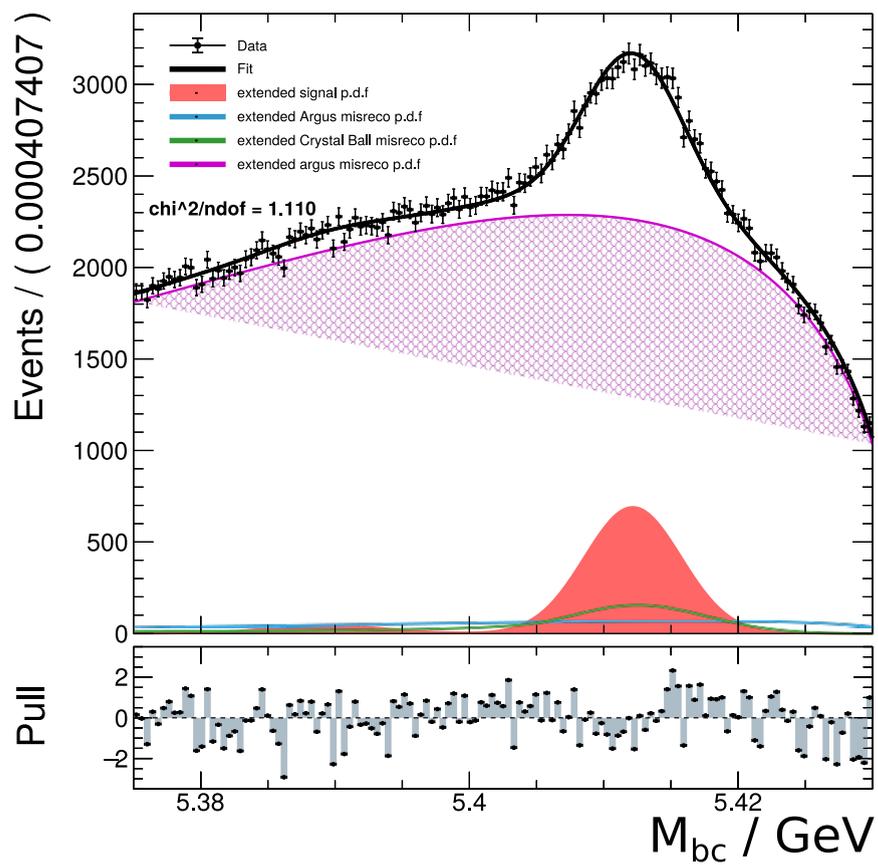


Figure A.3.: Fit to combined Monte Carlo data with all components in a reduced range of M_{bc} .

B. Additional channels

Table B.1.: B_s^0 decay channels that were implemented in the FEI at $\Upsilon(5S)$. Not all channels could be trained with the available central MC sample.

Trained	Not trained
$B_s^0 \rightarrow D_s^- D_s^+$	$B_s^0 \rightarrow D_s^{*+} D_s^{*-}$
$B_s^0 \rightarrow D_s^{*+} D_s^-$	$B_s^0 \rightarrow D_s^{*-} K^+$
$B_s^0 \rightarrow D_s^{*-} D_s^{*+}$	$B_s^0 \rightarrow \bar{D}^{*0} K_S^0$
$B_s^0 \rightarrow D_s^+ D^-$	$B_s^0 \rightarrow \bar{D}^0 K_S^0$
$B_s^0 \rightarrow D_s^{*-} D_s^+$	$B_s^0 \rightarrow D_s^- D^+ K_S^0$
$B_s^0 \rightarrow D_s^{*+} D^-$	$B_s^0 \rightarrow D_s^- D^0 K_S^0 \pi^+$
$B_s^0 \rightarrow D_s^- K^+$	$B_s^0 \rightarrow D_s^- D^+ K^+ \pi^-$
$B_s^0 \rightarrow D_s^- \pi^+$	$B_s^0 \rightarrow D_s^- D^+ K_S^0 \pi^0$
$B_s^0 \rightarrow D_s^{*-} \pi^+$	$B_s^0 \rightarrow D_s^- D^{*+} K_S^0$
$B_s^0 \rightarrow D_s^- \pi^+ \pi^+ \pi^-$	$B_s^0 \rightarrow D_s^- D^{*0} K^+ \pi^0$
$B_s^0 \rightarrow D_s^- D^0 K^+$	$B_s^0 \rightarrow D_s^- D^{*0} K_S^0 \pi^+$
$B_s^0 \rightarrow D_s^- \pi^+ \pi^0$	$B_s^0 \rightarrow D_s^- D^{*+} K^+ \pi^-$
$B_s^0 \rightarrow D_s^- D^0 K^+ \pi^0$	$B_s^0 \rightarrow D_s^- D^{*+} K_S^0 \pi^0$
$B_s^0 \rightarrow D_s^- D^{*0} K^+$	$B_s^0 \rightarrow D_s^{*-} D^+ K_S^0$
$B_s^0 \rightarrow D_s^{*-} D^0 K^+$	$B_s^0 \rightarrow D_s^{*-} D^{*+} K_S^0$
$B_s^0 \rightarrow D_s^{*-} D^{*0} K^+$	$B_s^0 \rightarrow J/\psi K^+ K^-$
$B_s^0 \rightarrow D_s^{*-} \pi^+ \pi^+ \pi^-$	$B_s^0 \rightarrow D_s^{*-} D^0 K_S^0 \pi^+$
$B_s^0 \rightarrow D_s^{*-} \pi^+ \pi^0$	$B_s^0 \rightarrow D_s^{*-} D^+ K^+ \pi^-$
$B_s^0 \rightarrow D_s^{*-} D^0 K^+ \pi^0$	$B_s^0 \rightarrow D_s^{*-} D^+ K_S^0 \pi^0$
	$B_s^0 \rightarrow D_s^{*-} D^{*0} K^+ \pi^0$
	$B_s^0 \rightarrow D_s^{*-} D^{*0} K_S^0 \pi^+$
	$B_s^0 \rightarrow D_s^{*-} D^{*+} K^+ \pi^-$
	$B_s^0 \rightarrow D_s^{*-} D^{*+} K_S^0 \pi^0$

C. Correlation between features and M_{bc} for trained channels

Table C.1.: Mean and maximum of correlation between M_{bc} and variables used to classify B_s^0 candidates with two daughters

Variable	Mean	Maximum	Max. channel
daughter(1,SignalProbability)	0.24	0.39	7
daughter(0,SignalProbability)	0.23	0.35	0
daughterProductOf(SignalProbability)	0.20	0.32	7
formula(deltaE+Mbc-5.3669)	0.11	0.27	10
daughterAngle(0,1)	0.11	0.15	0
useRestFrame(daughter(0, p))	0.11	0.31	10
useRestFrame(daughter(1, p))	0.09	0.18	7
chiProb	0.06	0.11	7
daughter(1,decayModeID)	0.06	0.10	4
daughter(0,decayModeID)	0.06	0.15	4
preCut_rank	0.06	0.10	7
useRestFrame(daughter(0, distance))	0.05	0.25	10
significanceOfDistance	0.05	0.10	7
daughter(0, chiProb)	0.05	0.14	7
useRestFrame(daughter(1, distance))	0.05	0.16	10
daughter(1, chiProb)	0.04	0.08	0
dx	0.03	0.06	10
distance	0.03	0.08	10
dy	0.03	0.17	10
dr	0.03	0.15	10
decayAngle(0)	0.02	0.07	7
decayAngle(1)	0.02	0.07	7
dz	0.02	0.06	10
cosAngleBetweenMomentumAndVertexVector	0.01	0.01	4

Table C.2.: Mean and maximum of correlation between M_{bc} and variables used to classify B_s^0 candidates with three daughters

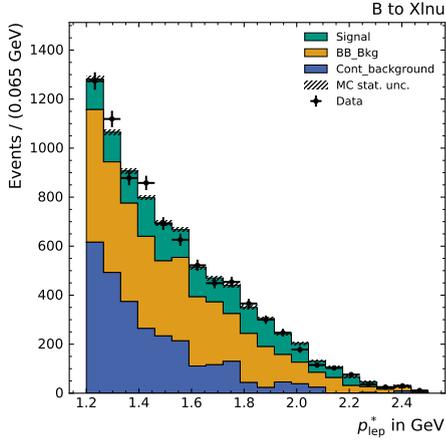
Variable	Mean	Maximum	Max. channel
daughter(0,SignalProbability)	0.15	0.34	17
daughterProductOf(SignalProbability)	0.14	0.30	17
daughter(1,SignalProbability)	0.11	0.23	33
daughter(2,SignalProbability)	0.10	0.27	33
daughterAngle(0,2)	0.10	0.23	33
daughterAngle(1,2)	0.09	0.24	17
preCut_rank	0.08	0.13	33
useRestFrame(daughter(1, p))	0.07	0.20	33
useRestFrame(daughter(0, p))	0.06	0.21	17
daughterAngle(0,1)	0.06	0.07	17
useRestFrame(daughter(2, p))	0.06	0.17	33
formula(deltaE+Mbc-5.3669)	0.04	0.12	33
daughter(0,decayModeID)	0.03	0.12	17
decayAngle(0)	0.03	0.06	17
daughter(0, chiProb)	0.02	0.08	17
decayAngle(1)	0.02	0.03	17
useRestFrame(daughter(1, distance))	0.01	0.04	33
decayAngle(2)	0.01	0.04	17
useRestFrame(daughter(2, distance))	0.01	0.01	30
chiProb	0.01	0.01	28
daughter(1, chiProb)	0.01	0.01	30
useRestFrame(daughter(0, distance))	0.01	0.01	22
daughter(1,decayModeID)	0.01	0.02	30
daughter(2, chiProb)	0.01	0.01	30
significanceOfDistance	0.01	0.01	28
cosAngleBetweenMomentumAndVertexVector	0.00	0.01	30
dy	0.00	0.01	30
distance	0.00	0.01	30
dz	0.00	0.01	30
dx	0.00	0.00	15
dr	0.00	0.00	22

Table C.3.: Mean and maximum of correlation between M_{bc} and variables used to classify B_s^0 candidates with four daughters

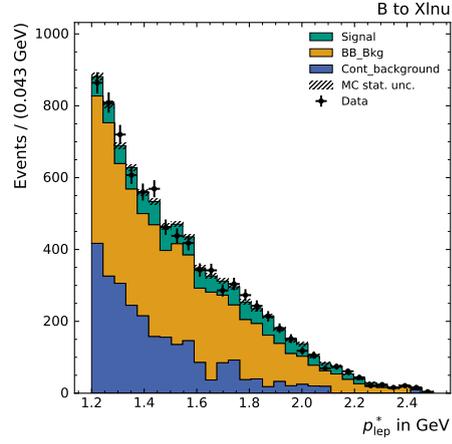
Variable	Mean	Maximum	Max. channel
daughterProductOf(SignalProbability)	0.14	0.28	32
daughter(0,SignalProbability)	0.14	0.28	32
daughter(1,SignalProbability)	0.08	0.18	32
chiProb	0.07	0.08	32
useRestFrame(daughter(0, p))	0.07	0.15	14
daughterAngle(0,1)	0.06	0.09	14
daughter(3,SignalProbability)	0.06	0.12	32
daughterAngle(0,2)	0.06	0.10	32
preCut_rank	0.06	0.09	34
daughterAngle(1,2)	0.05	0.08	32
daughterAngle(0,3)	0.05	0.10	32
daughterAngle(1,3)	0.05	0.09	32
daughter(2,SignalProbability)	0.04	0.09	32
useRestFrame(daughter(3, distance))	0.04	0.05	32
daughterAngle(2,3)	0.04	0.09	32
significanceOfDistance	0.03	0.04	32
formula(deltaE+Mbc-5.3669)	0.03	0.07	32
useRestFrame(daughter(2, distance))	0.03	0.06	32
decayAngle(0)	0.03	0.05	32
useRestFrame(daughter(1, distance))	0.02	0.06	32
daughter(0, chiProb)	0.02	0.05	14
decayAngle(3)	0.02	0.03	18
dx	0.02	0.03	14
daughter(0,decayModeID)	0.02	0.04	14
useRestFrame(daughter(1, p))	0.02	0.04	32
dy	0.01	0.02	32
useRestFrame(daughter(0, distance))	0.01	0.03	32
useRestFrame(daughter(3, p))	0.01	0.03	14
decayAngle(1)	0.01	0.02	32
distance	0.01	0.02	14
dr	0.01	0.02	32
decayAngle(2)	0.01	0.02	32
useRestFrame(daughter(2, p))	0.01	0.02	14
daughter(1, chiProb)	0.01	0.01	32
daughter(2, chiProb)	0.00	0.01	32
daughter(3, chiProb)	0.00	0.01	32
dz	0.00	0.01	32
cosAngleBetweenMomentumAndVertexVector	0.00	0.00	34
daughter(1,decayModeID)	0.00	0.00	18

D. Template fits with varying number of bins

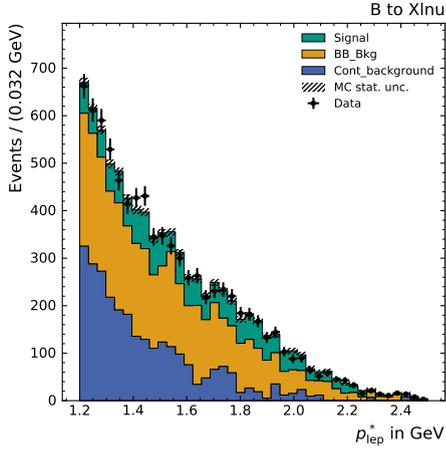
To determine a calibration factor for the FEI at $\Upsilon(5S)$, template fits were used to obtain the normalization of the momentum of reconstructed leptons (see Section 5.4). The calibration factors calculated with this method can be found in Table 5.5. Figure D.1 shows the lepton distribution and the result of a template fit for varying numbers of bins when using a continuum template generated with off-resonance data. Figure D.2 shows the same for a continuum template generated with Monte Carlo data.



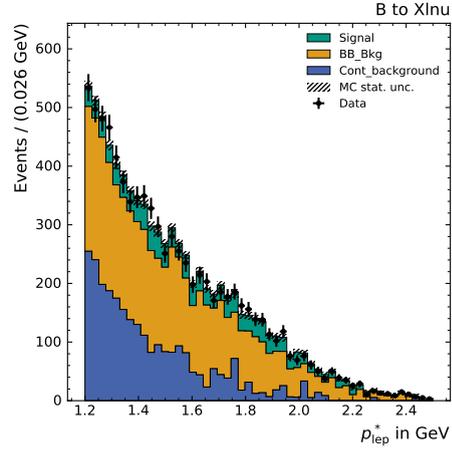
(a) 20 bins



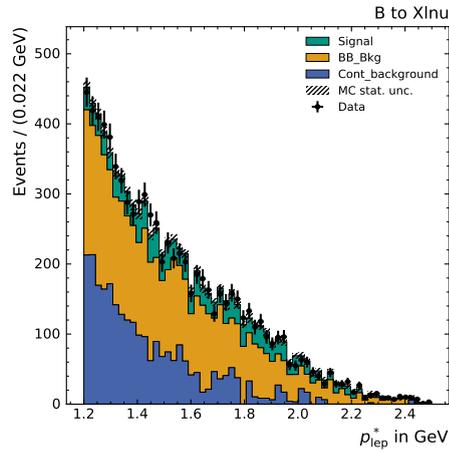
(b) 30 bins



(c) 40 bins

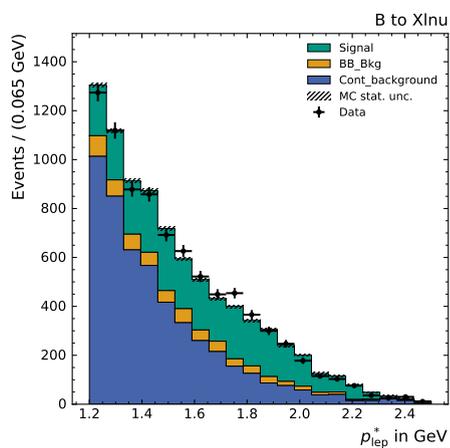


(d) 50 bins

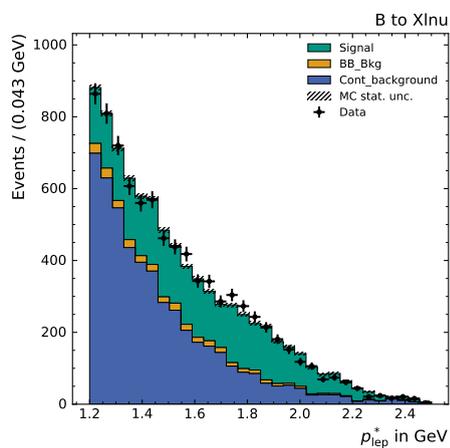


(e) 60 bins

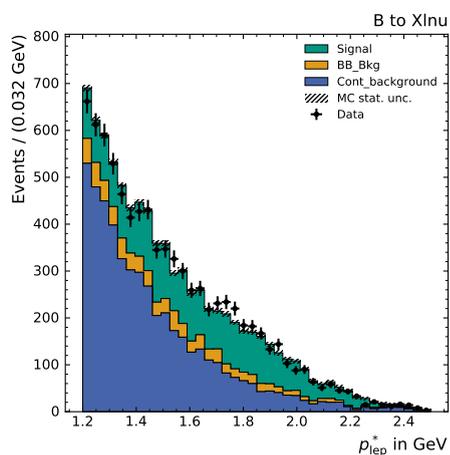
Figure D.1.: Result of a template fit using an off-resonance template to describe the continuum contribution.



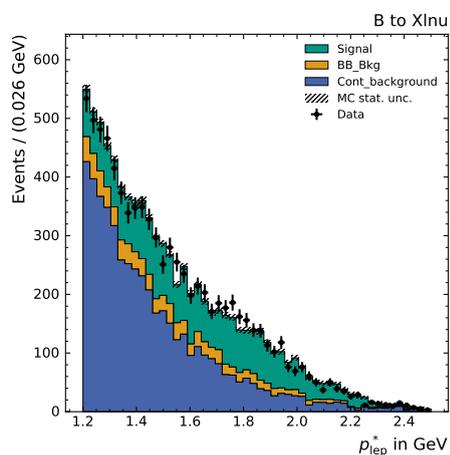
(a) 20 bins



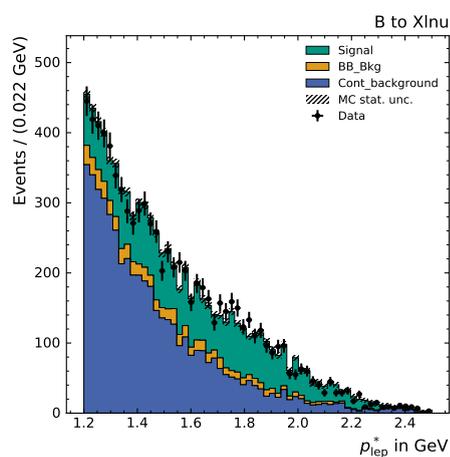
(b) 30 bins



(c) 40 bins



(d) 50 bins



(e) 60 bins

Figure D.2.: Result of a template fit using a Monte Carlo template to describe the continuum contribution.

Bibliography

- [1] A. Abashian et al. “The Belle Detector”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. Detectors for Asymmetric B-Factories 479.1 (Feb. 21, 2002), pp. 117–232. ISSN: 0168-9002. DOI: [10.1016/S0168-9002\(01\)02013-7](https://doi.org/10.1016/S0168-9002(01)02013-7).
- [2] Tetsuo Abe et al. “Achievements of KEKB”. In: *Progress of Theoretical and Experimental Physics* 2013.3 (Mar. 1, 2013). DOI: [10.1093/ptep/pts102](https://doi.org/10.1093/ptep/pts102).
- [3] Akimasa Ishikawa, Hiroshi Nakano, and Kazutaka Sumisawa. *Ks Selection with NeuroBayes and nisKsFinder Class*. Nov. 20, 2012.
- [4] BABAR Collaboration et al. “Study of $B \rightarrow X\gamma$ decays and determination of $|V_{td}/V_{ts}|$ ”. In: *Physical Review D* 82.5 (Sept. 10, 2010), p. 051101. DOI: [10.1103/PhysRevD.82.051101](https://doi.org/10.1103/PhysRevD.82.051101).
- [5] Belle Collaboration et al. “Measurement of the Decay $B_s^0 \rightarrow D_s^- \pi^+$ and Evidence for $B_s^0 \rightarrow D_s^\mp K^\pm$ in e^+e^- Annihilation at $\sqrt{s} \approx 10.87$ GeV”. In: *Physical Review Letters* 102.2 (Jan. 13, 2009), p. 021801. DOI: [10.1103/PhysRevLett.102.021801](https://doi.org/10.1103/PhysRevLett.102.021801).
- [6] Belle Collaboration et al. “Precise measurement of the branching fractions for $B_s^0 \rightarrow D_s^{(*)+} D_s^{(*)-}$ and first measurement of the $D_s^{*+} D_s^{*-}$ polarization using e^+e^- collisions”. In: *Physical Review D* 87.3 (Feb. 12, 2013), p. 031101. DOI: [10.1103/PhysRevD.87.031101](https://doi.org/10.1103/PhysRevD.87.031101).
- [7] Belle Collaboration et al. “Search for $B^+ \rightarrow \mu^+ \nu_\mu$ and $B^+ \rightarrow \mu^+ N$ with inclusive tagging”. In: *Physical Review D* 101.3 (Feb. 21, 2020), p. 032007. DOI: [10.1103/PhysRevD.101.032007](https://doi.org/10.1103/PhysRevD.101.032007).
- [8] D. Besson et al. “Observation of New Structure in the e^+e^- Cross Section above the $\Upsilon(4S)$ ”. In: *Physical Review Letters* 54.5 (Feb. 4, 1985), pp. 381–384. DOI: [10.1103/PhysRevLett.54.381](https://doi.org/10.1103/PhysRevLett.54.381).
- [9] D Besson and T Skwarnicki. “Upsilon Spectroscopy: Transitions in the Bottomonium System”. In: *Annual Review of Nuclear and Particle Science* 43.1 (1993), pp. 333–378. DOI: [10.1146/annurev.ns.43.120193.002001](https://doi.org/10.1146/annurev.ns.43.120193.002001).
- [10] H. Bethe. “Zur Theorie Des Durchgangs Schneller Korpuskularstrahlen Durch Materie”. In: *Annalen der Physik* 397.3 (1930), pp. 325–400. ISSN: 1521-3889. DOI: [10.1002/andp.19303970303](https://doi.org/10.1002/andp.19303970303).
- [11] A. J. Bevan et al. “The Physics of the B Factories”. In: *The European Physical Journal C* 74.11 (Nov. 19, 2014), p. 3026. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-014-3026-9](https://doi.org/10.1140/epjc/s10052-014-3026-9).

- [12] Nils Braun. “Combinatorial Kalman Filter and High Level Trigger Reconstruction for the Belle II Experiment”. DOI: <https://doi.org/10.5445/IR/1000089317>.
- [13] Jolanta Brodzicka et al. “Physics Achievements from the Belle Experiment”. In: *Progress of Theoretical and Experimental Physics* 2012.1 (Jan. 1, 2012). DOI: [10.1093/ptep/pts072](https://doi.org/10.1093/ptep/pts072).
- [14] David G Cassel. *CLEO B Physics*, p. 17.
- [15] Glen Cowan. *Statistical Data Analysis*. Oxford University Press, 1998. 218 pp. ISBN: 978-0-19-850156-5.
- [16] 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. 21, 2011), pp. 432–440. ISSN: 0168-9002. DOI: [10.1016/j.nima.2011.06.008](https://doi.org/10.1016/j.nima.2011.06.008).
- [17] Michael Feindt. *A Neural Bayesian Estimator for Conditional Probability Densities*. arXiv:physics/0402093. Feb. 18, 2004.
- [18] R. A. Fisher. “The Statistical Utilization of Multiple Measurements”. In: *Annals of Eugenics* 8.4 (1938), pp. 376–386. ISSN: 2050-1439. DOI: [10.1111/j.1469-1809.1938.tb02189.x](https://doi.org/10.1111/j.1469-1809.1938.tb02189.x).
- [19] Geoffrey C. Fox and Stephen Wolfram. “Event Shapes in E+e-Annihilation”. In: *Nuclear Physics B* 149.3 (Mar. 26, 1979), pp. 413–496. ISSN: 0550-3213. DOI: [10.1016/0550-3213\(79\)90003-8](https://doi.org/10.1016/0550-3213(79)90003-8).
- [20] Jerome H. Friedman. “Stochastic Gradient Boosting”. In: *Computational Statistics & Data Analysis. Nonlinear Methods and Data Mining* 38.4 (Feb. 28, 2002), pp. 367–378. ISSN: 0167-9473. DOI: [10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
- [21] *Full Reconstruction of the $\Upsilon(5S)$ Resonance and Analysis of the Tetraquark State Z_B at Belle* / *publish.etp*. URL: <https://publish.etp.kit.edu/record/21036> (visited on 02/10/2020).
- [22] Moritz Gelb et al. “B2BII: Data Conversion from Belle to Belle II”. In: *Computing and Software for Big Science* 2.1 (Nov. 8, 2018), p. 9. ISSN: 2510-2044. DOI: [10.1007/s41781-018-0016-x](https://doi.org/10.1007/s41781-018-0016-x).
- [23] R. L. Gluckstern. “Uncertainties in Track Momentum and Direction, Due to Multiple Scattering and Measurement Errors”. In: *Nuclear Instruments and Methods* 24 (July 1, 1963), pp. 381–389. ISSN: 0029-554X. DOI: [10.1016/0029-554X\(63\)90347-1](https://doi.org/10.1016/0029-554X(63)90347-1).
- [24] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. Tech. rep. July 7, 2009.
- [25] R. Itoh. “BASF - BELLE Analysis Framework”. In: 1997.
- [26] F. James and M. Roos. “Minuit – a System for Function Minimization and Analysis of the Parameter Errors and Correlations”. In: *Computer Physics Communications* 10 (Dec. 1975), pp. 343–367. DOI: [10.1016/0010-4655\(75\)90039-9](https://doi.org/10.1016/0010-4655(75)90039-9).
- [27] T. Keck et al. “The Full Event Interpretation”. In: *Computing and Software for Big Science* 3.1 (Feb. 25, 2019), p. 6. ISSN: 2510-2044. DOI: [10.1007/s41781-019-0021-8](https://doi.org/10.1007/s41781-019-0021-8).

- [28] Thomas Keck. “FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment”. In: *Computing and Software for Big Science* 1.1 (Sept. 29, 2017), p. 2. ISSN: 2510-2044. DOI: [10.1007/s41781-017-0002-8](https://doi.org/10.1007/s41781-017-0002-8).
- [29] Thomas Keck. *Machine Learning at the Belle II Experiment: The Full Event Interpretation and Its Validation on Belle Data / by Thomas Keck*. 2018. ISBN: 978-3-319-98249-6.
- [30] Thomas Keck. “The Full Event Interpretation for Belle II”. Karlsruhe Institute of Technology (KIT), 2014.
- [31] *KEKB (Accelerator)*. In: *Wikipedia*. Page Version ID: 924548096. Nov. 4, 2019.
- [32] *KEKB B Factory Design Report - INSPIRE-HEP*. URL: <http://inspirehep.net/record/401380/> (visited on 02/09/2020).
- [33] Makoto Kobayashi and Toshihide Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1, 1973), pp. 652–657. ISSN: 0033-068X. DOI: [10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [34] T. Kuhr et al. “The Belle II Core Software”. In: *Computing and Software for Big Science* 3.1 (Nov. 28, 2018), p. 1. ISSN: 2510-2044. DOI: [10.1007/s41781-018-0017-9](https://doi.org/10.1007/s41781-018-0017-9).
- [35] S. Kurokawa and E. Kikutani. “Overview of the KEKB Accelerators”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. KEK-B: The KEK B-Factory 499.1 (Feb. 21, 2003), pp. 1–7. ISSN: 0168-9002. DOI: [10.1016/S0168-9002\(02\)01771-0](https://doi.org/10.1016/S0168-9002(02)01771-0).
- [36] R. Louvot et al. “Measurement of the Decay $B_s^0 \rightarrow D_s - \pi^+$ and Evidence for $B_s^0 \rightarrow D_s \pm K^\pm$ in e^+e_- Annihilation at $\sqrt{s} = 10.87\text{-GeV}$ ”. In: *Phys.Rev.Lett.* 102 (2009), p. 021801. DOI: [10.1103/PhysRevLett.102.021801](https://doi.org/10.1103/PhysRevLett.102.021801).
- [37] D. M. J. Lovelock et al. “Masses, Widths, and Leptonic Widths of the Higher Upsilon Resonances”. In: *Physical Review Letters* 54.5 (Feb. 4, 1985), pp. 377–380. DOI: [10.1103/PhysRevLett.54.377](https://doi.org/10.1103/PhysRevLett.54.377).
- [38] National Lab for High Energy Physics. *A Study of CP Violation in B Meson Decays. Technical Design Report*. KEK-95-1. National Lab. for High Energy Physics, 1995.
- [39] Piti Ongmongkolkul. *Scikit-Hep/Iminuit*. Version v1.3.8. Scikit-HEP Project.
- [40] Piti Ongmongkolkul. *Scikit-Hep/Probit*. Version 1.1.0. Scikit-HEP Project.
- [41] Christian Pulvermacher. “Analysis Software and Full Event Interpretation for the Belle II Experiment”. DOI: [10.5445/IR/1000050704](https://doi.org/10.5445/IR/1000050704).
- [42] Alex Rogozhnikov et al. “New Approaches for Boosting to Uniformity”. In: *Journal of Instrumentation* 10.03 (Mar. 30, 2015), T03002–T03002. ISSN: 1748-0221. DOI: [10.1088/1748-0221/10/03/T03002](https://doi.org/10.1088/1748-0221/10/03/T03002).
- [43] Judith Schwab. “Calibration of the Full Event Interpretation for the Belle and the Belle II Experiment”. Karlsruhe Institute of Technology (KIT), 2017.
- [44] Sevda Esen. *BF measurement of B_s^0 Decays to $D_s^{(*)+}D_s^{(*)-}$ with 121.4fb^{-1}* .

- [45] T. Suwada et al. “Enhancement of the Positron Intensity by a Tungsten Single-Crystal Target at the KEKB Injector Linac”. In: *2007 IEEE Particle Accelerator Conference (PAC)*. 2007 IEEE Particle Accelerator Conference (PAC). June 2007, pp. 2778–2780. DOI: [10.1109/PAC.2007.4440573](https://doi.org/10.1109/PAC.2007.4440573).
- [46] M. Tanabashi et al. “Review of Particle Physics”. In: *Physical Review D* 98.3 (Aug. 17, 2018), p. 030001. ISSN: 2470-0010, 2470-0029. DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001).
- [47] Wouter Verkerke and David P. Kirkby. “The RooFit Toolkit for Data Modeling”. In: *eConf C0303241* (2003), MOLT007.
- [48] W.-M. Yao et al. and (Particle Data Group). “The Review of Particle Physics”. In: *J. Phys. G* 33, 1 (2006).
- [49] Dennis Weyland. “Continuum Suppression with Deep Learning Techniques for the Belle II Experiment”. Karlsruhe Institute of Technology (KIT), 2017.