Search for the Rare Decays $B \rightarrow h\nu\bar{\nu}$ With Semileptonic Tag at the Belle Experiment

Johannes Grygier

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

von der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte

Dissertation

von

Dipl.-Phys. Johannes Grygier

aus Halle (Saale)

Contents

Chapter 1

Introduction

1.1. Motivation

The prince says that the world will be saved by beauty. And Although Dostoevsky could not possibly have known anything about flavor physics when he let his character Ippolit proclaim this assessment of Myshkins philosophy, prince Myshkin might say the same thing with regard to modern physics if he was alive today. The world of particle physics has entered a very exciting stage in the history of the field. After the discovery of the Higgs boson by the two experiments CMS and ATLAS in 2012, all particles predicted by the fundamental theory of the field, the Standard Model (SM), have been confirmed by experiment. And despite being meticulously studied with ever growing precision and sophistication of the experimental techniques involved, the SM fails to provide an answer to all remaining questions the human mind poses to the machinations of nature. It, for instance, does neither provide a suitable dark matter candidate nor can it explain the observed matter-antimatter asymmetry. Hence, physicists agree that the current, though incredibly successful, theory needs to be amended to give a fully comprehensive description of nature in itself.

Such additional phenomena, often referred to as new physics (NP), will manifest themselves either in direct observation or in the effect they have on other processes causing a deviation of certain observables from the respective value predicted by the established theory. The investigation of the latter is the core principle of flavor physics and its most active subfield B physics, named after the object of study, the beauty or bottom quark. The big advantage of the indirect approach to the search for NP is a sensitivity to NP which is not limited by the beam energy of a collider such as the LHC. Flavor physics and high energy physics each form a very active frontier in the strive for NP, both promising great prospect. The probability to find hints for NP at flavor physics experiments relative to the probability of finding anything in a direct observation has risen in recent years since CMS and ATLAS have set very tight limits on the mass of new particles in many channels.

The complexity of the topic of flavor physics opens a rich field for various models to manifest themselves in different observables. A very challenging, but even more interesting example, are

decays of a B meson into a lighter charmless meson and a pair of neutrinos. Those decays have never been observed before but are sensitive to many proposed models of NP that would affect their respective branching ratio[1], notable examples are given in Section 2.2.3 on page 12. Recent results by the LHCb collaboration [A^+16a, A^+14a] in a related channel showed tensions between expectations and experimental data. This rose the general interest in similar decay topologies. $B \to h\nu\bar{\nu}$ decays are therefore a very valuable testing ground for those theories and of great interest to the scientific community.

The scope of this work is to enlarge the experimental sensitivity compared to previous results and provide the best result in each investigated channel. Chapter 2 on page 5 provides a brief introduction into the fundamental basics of the theory before it proceeds to a theoretical consideration of the decay itself. Chapter 3 on page 17 gives an overview of the experimental facility where the data for this analysis were taken, the Belle experiment at the KEKB accelerator. Chapter 4 on page 25 does the same for essential analysis tools used in this work. The analysis itself follows in Chapters 5 to 11 and is outlined below.

1.2. Analysis Outline

The analysis is designed as a search for $B \to h\nu\bar{\nu}$ decays. The emphasis is laid on the suppression of background and isolation of possible signal rather than an accurate measurement of the branching ratio which is not feasible since we do not know if those decays occur, and, in case they do, how often. Since none of the decays has yet been observed, we want to find evidence for them and design the analysis consequently to maximize the chances to do so. The general analysis procedure is outlined in Figure 1.1 on page 4 and is briefly described below. Further information, definitions, and explanations can be found in the referenced chapters. Solid lines indicate the ordering of the steps involved while dashed lines indicate important feedback from a later to an earlier stage of the analysis. The ordering of the chapters follows that of the chart, which also reflects the order in which the data is processed.

We reconstruct eight decay channels of B mesons. All channels exhibit a pair of neutrinos and a single hadron. Eight different types of hadrons are reconstructed[2]: K^+, K_S^0, K^{*+}, K^{*0}, π^+, π^0, ρ^+, and ρ^0. How the reconstruction is done is described in detail in Chapter 5 on page 31. The reconstruction of a final state consisting of one detectable particle and missing momentum requires additional information from the rest of the event. This is done by the reconstruction of a second B in one of multiple semileptonic channels. This process is referred to as tagging; the second B meson is called tag B, B_{tag}. Further informations on the tagging algorithm can be found in Section 5.2.1 on page 32.

Signal and tag candidates are used to form $\Upsilon(4S)$ candidates[3]. Correctly reconstructed $\Upsilon(4S)$ candidates encompass the whole event. For this reason, events with additional charged particles or π^0

---

[1]For the definition, please be referred to Section 9.1.3 on page 92

[2]Charge conjugate modes are implied throughout this work unless stated explicitly otherwise.

[3]See Section 3.1 on page 17 for the definition

candidates are rejected or vetoed. Subsequently, those events without additional candidates are written out and the remaining steps of the analysis proceed offline.

In a next step, certain precuts are applied which suppress background but retain a high signal efficiency. At this stage, an event might still contain multiple candidates due to particle exchange between signal and tag side. One single candidate is chosen, based on the goodness of the tag, see Section 5.2.6 on page 39. A neural net is then trained on the remaining sample where weights and correction factors, as described in Chapter 6 on page 47, are applied as a weight in the training. A cut on the net output is thereafter optimized for the final selection. Detailed information on that multivariate procedure can be found in Chapter 7 on page 53. We use the information on the background composition to define new variables designed to suppress prominent background components. This iterative process is indicated by a dashed line in Figure 1.1 on the following page.

The branching fraction of the respective $B \rightarrow h\nu\bar{\nu}$ decay is extracted from a fit to the extra energy E_{ECL} [4]. Differences between data and Monte Carlo simulations (MC) are checked by looking into the E_{ECL} sidebands, into data taken at beam energies slightly lower than the $\Upsilon(4S)$ mass, and by reconstructing $B \rightarrow D^{*}l\nu_{l}$ control channels, as described in Chapter 8 on page 77. We take the information from those comparisons to either adjust our precuts or to calculate corrections where we observe significant differences between data and MC expectations. This is marked by a dashed line in Figure 1.1 on the next page as well. The fit model and branching fraction calculation are lined out in Chapter 9 on page 89, as well as the calculation of expected results and the check of the fitting procedure by the means of toy studies. Chapter 11 on page 109 contains a description of the relevant systematic uncertainties and the methods applied to estimate those.

---

[4] For the definition, see Section 5.2.4 on page 37
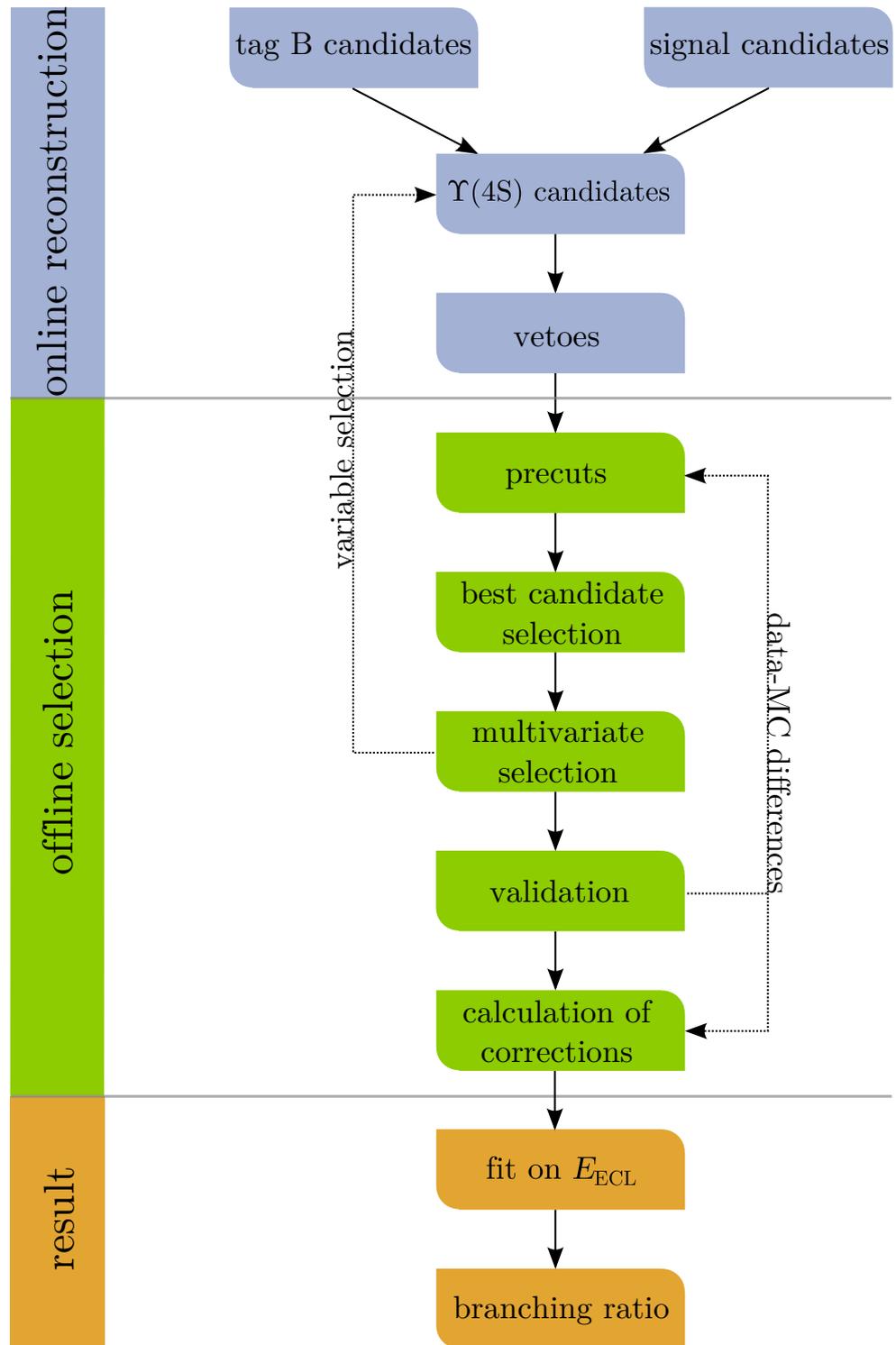
Figure 1.1.: Outline of the analysis procedure.

Chapter 2

Theoretical Basics

2.1. The Standard Model

2.1.1. Overview

The SM is one of the best established and most scrupulously checked theories in all of science. Formulated in its contemporary form in the 1970s, a great variety of experimental effort was put into effect to possibly falsify its predictions and measure its parameters. Yet, no significant deviation of experimentally accessible observables from theoretical predictions have been measured. The SM describes interactions of the most fundamental particles of nature, elementary in the sense that they are not comprised of other more fundamental particles. It consists of fermionic matter particles, bosonic force particles mediating the interactions of the former in the framework of a Lorentz invariant gauged quantum field theory, and the Higgs field. The gauge group of the SM is the following.

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{2.1}$$

The $SU(3)$ color symmetry is linked to the strong force coupling particles charged under it via gluon exchange. The unified $SU(2)$ left and $U(1)$ hypercharge symmetries describe the weak interaction via $W^\pm$ and Z^0 bosons and the electromagnetic interaction coupling photons to electric charge. Photons and Z^0 bosons emanate from the mixing of the two $T_3 = 0$ states[1], the B_μ field of the $U(1)_Y$ and the W_μ^3 field of the $SU(2)_L$. Those force carriers together with the Higgs boson comprise all fundamental bosons we know of. The Higgs Boson, so far the last fundamental particle discovered, is the remaining degree of freedom left after the $SU(2)$ symmetry of the Higgs field is broken spontaneously. This scalar field gives the massive ones ($W^\pm$ and Z^0) their mass via the aforementioned spontaneous breaking of the $SU(2)_L \otimes U(1)_Y$ down to the conserved

---

[1]Here, T_3 is the weak isospin.

$U(1)_{em}$ symmetry of the SM and enables fermionic mass terms through its non-vanishing vacuum expectation value, which are not gauge invariant for themselves.

The fermionic matter particles can be further subdivided into 2 categories, those who interact strongly (quarks) and those who do not (leptons). Quarks and leptons can be further subcategorized into three so-called families. Each family is comprised of an isospin doublet, consisting of an up-type and a down-type particle, and corresponding right-handed singlets. An overview over all particles can be found in Figure 2.1.
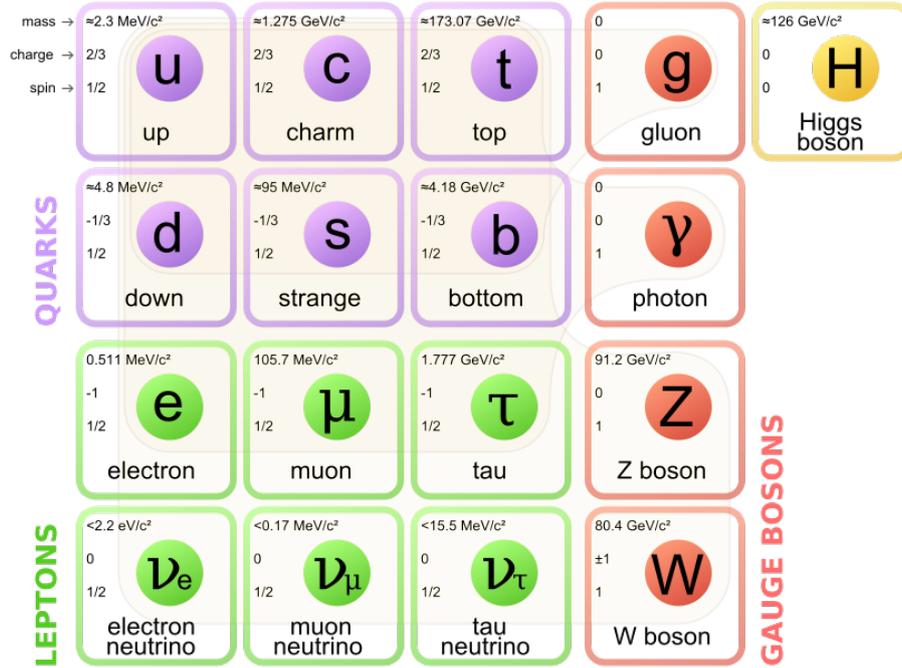
Figure 2.1.: Graphical illustration of the Standard Model with sketched interactions, source: [Wik06].

2.1.2. Flavor Structure of The Standard Model

As mentioned in Section 2.1.1 on the previous page, fermionic mass terms are not gauge invariant under $SU(2)_L$ transformations. Instead, such mass terms in the Lagrangian are generated dynamically via a Yukawa interaction of the fermions with the Higgs field in the SM. Coupling in principle can occur between all flavors.

$$\mathscr{L}^d = -\sum_{i,j=1}^{3} \bar{L}_i \Phi Y_{ij}^d d_{Rj} + h.c. \qquad \mathscr{L}^u = -\sum_{i,j=1}^{3} \bar{L}_i \Phi^C Y_{ij}^u u_{Rj} + h.c. \qquad (2.2)$$

After the breaking of electroweak symmetry, mass matrices can be defined as $M_{ij}^u = v Y_{ij}^u$ with v the vacuum expectation value of the Higgs field. Physical states are obtained by diagonalising M_{ij}^u via

the transformation $M_{ij}^u \rightarrow V_L^u M_{ij}^u V_R^u$, with two unitary matrices V_L^u and V_R^u, for the down-type quarks accordingly. The decisive issue is that up- and down-type matrices can not be diagonalised simultaneously. By convention, the up-type mass matrix is diagonalised. This implies that the mass eigenstates for the down-type quarks do not share a common base with the eigenstates of the weak interaction. The transformation between both sets of eigenstates is described by a matrix referred to as the CKM-matrix. It is itself, as a product of two unitary matrices, unitary.

$$\hat{V}_{\text{CKM}} = V_L^u \ V_L^{d\dagger} \tag{2.3}$$

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \hat{V}_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.4}$$

The matrix has four free parameters: three real mixing angles and one complex phase. Wolfenstein came up with an approximate parametrization in the form of an expansion in the absolute value of the sine of the Cabibbo angle, $\lambda = |V_{us}| \approx 0.2253$.

$$\hat{V}_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3 (\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3 (1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}\left(\lambda^4\right) \tag{2.5}$$

Here λ, A, ρ, and η are real parameters. Equation (2.5) shows a strict hierarchy between the quark generations; off-diagonal elements are suppressed by powers of λ. The result is a suppression of couplings between generations.

Elements of $\hat{V}_{\text{CKM}}$ appear in couplings of the $W^{\pm}$ but not in the couplings of any other boson as a direct consequence of the unitarity of the CKM matrix in the SM.

$$\mathcal{L}^W \propto \bar{u}_{Li}\gamma^\mu \hat{V}_{\text{CKM}}^{ij} d_{Lj} W_\mu^{\pm} \qquad \mathcal{L}_{V-A}^{Z^0} \propto \bar{d}_{Li}\gamma^\mu d_{Li} Z_\mu^0 \tag{2.6}$$

As a consequence, transitions between different flavors with the same electrical charge, or in other words $d \rightarrow d'$ or $u \rightarrow u'$, so-called flavor changing neutral currents (FCNCs) are forbidden on tree level within the SM due to the unitarity of the CKM matrix via the so called GIM mechanism.

The most relevant matrix elements in this work, V_{td} and V_{ts} are currently measured as [O$^+$14]

$$|V_{td}| = (40.0 \pm 2.7) \times 10^{-3} \qquad |V_{ts}| = (8.4 \pm 0.6) \times 10^{-3} \tag{2.7}$$

2.1.3. The systematics of Composed Particles

Because it is non-abelian, gluons, the gauge bosons of the $SU(3)_C$ symmetry, couple to each other. Since they are massless, unlike the W and Z bosons, the coupling is not reduced by a propagator term like it is in the weak interaction for lower Q^2 values. Both properties result in an antiscreening yielding in a strong coupling of colored particles, hence the name strong interaction. The result are tightly bound systems, or particles, with an additional requirement caused by the symmetry group of the strong interaction, called confinement. All stable strongly bound objects are required to be $SU(3)_C$ singlets, neutral with respect to color. Colorless objects bound by the strong force (hadrons) can either be formed by three color charges (baryons), or by a color charge and the respective anti charge. The latter states are referred to as mesons.

Mesons can be classified by their total angular momentum, parity, and, if defined, charge parity eigenvalues. This scheme can further be ordered by the isospin and flavor quantum numbers of the respective particle[2]. The most abundant particles in the final states of particle physics experiments are a strong isospin triplet called pions with $J^{PC} = 0^{-+}$. Their first excited states $(S = 1)$ are called ρ mesons with quantum numbers $J^{PC} = 1^{--}$ decaying into a pair of pions. Pseudoscalar mesons with strangeness $S = \pm 1$ and strong isospin $I = \frac{1}{2}$ are referred to as kaons. Their respective spin excitations are called K^*. Those decay into a kaon and a pion. The spin excitation is consistently marked by a superscript * for all flavored mesons. The nomenclature of those follows the heaviest constituent. In case of this being a charm quark, the respective particles are called D meson and in case of it being a beauty quark, B meson. The top quark, the heaviest of all elementary particles, does not form bound states due to its very short lifetime.

2.1.4. Problems and Challenges of Modern Physics

Albeit the SM being the best tested theory in the history of human science, there are some remaining problems. It fails for instance to provide a dark matter candidate, can not explain the huge difference in the masses of its particles, ranging from $m_e = 511\,\text{keV}$ up to $m_t = 173\,\text{GeV}$, and fails to provide enough CP violation to explain the observed baryon asymmetry in the universe. Huge efforts have thus been put into place to look for hints of NP. A possible way to do this is by trying to produce yet undiscovered particles directly on shell, another is by trying to measure quantities predictable with very low uncertainty from the SM as precisely as possible to test whether new physics processes have a significant effect on that quantity. The first approach is e.g. taken at the large multi purpose experiments ATLAS and CMS at the Large Hadron Collider (LHC) and has been very successful in the discovery of the Higgs boson. The second one is taken by flavour physics and has been very successfull as well; the discovery of B_d oscillations for instance was a strong hint towards a comparably heavy top quark.

---

[2]Flavor quantum number here refers to either strangeness, charm, or beauty.

2.2. The Decay $B \to h\nu\bar{\nu}$

2.2.1. The Decay in the SM

A transition from a given initial to a certain final state can be described by a coherent sum of decay amplitudes. Each such amplitude can be represented as a series of Feynman diagrams, such as in Figure 2.2 and calculated as a term in a perturbation series. Physical observables however, e.g. the decay rate, only depend on the absolute value squared of the transition amplitude. Equation (2.8) shows an example for an arbitrary process from the initial (IS) to the final state (FS).

$$\mathcal{R}_{\mathrm{IS}\to\mathrm{FS}} \propto |\mathcal{A}_{\mathrm{SM}} + \mathcal{A}_{\mathrm{NP}}|^2 = |\mathcal{A}_{\mathrm{SM}}|^2 + |\mathcal{A}_{\mathrm{NP}}|^2 + 2\,|\mathcal{A}_{\mathrm{SM}}|\,|\mathcal{A}_{\mathrm{NP}}|\cos\theta_{\mathrm{SM-NP}} \tag{2.8}$$

The matrix element is expressed as a sum of all terms computable within the SM, $\mathcal{A}_{\mathrm{SM}}$ and any new physics contribution $\mathcal{A}_{\mathrm{NP}}$. The interference term in Equation (2.8) is linear in the amplitude of new physics as opposed to a quadratic dependence in the time reversed direct production. If $|\mathcal{A}_{\mathrm{SM}}|$ is low, even small $|\mathcal{A}_{\mathrm{NP}}|$ can result in sizable contributions to the overall rate. Rare processes are thus a promising field to look for new physics.

A specific example are FCNCs. As discussed in Section 2.1.2 on page 6, such processes are forbidden on tree level within the SM and can only proceed via loops, most prominently via so-called penguin diagrams, see Figure 2.2a. Due to the non-diagonal CKM matrix elements and the necessary loop involved, those processes are highly suppressed. However, many models of new physics allow them to proceed on tree level by introducing additional particles with different couplings.

A notable example is the quark level transition $b \to s\nu\bar{\nu}$ which occurs in the decay $B^+ \to K^+\nu\bar{\nu}$. The dominant contributions to this decay, a penguin and a box diagram, are displayed in Figure 2.2.

(a) penguin   (b) box
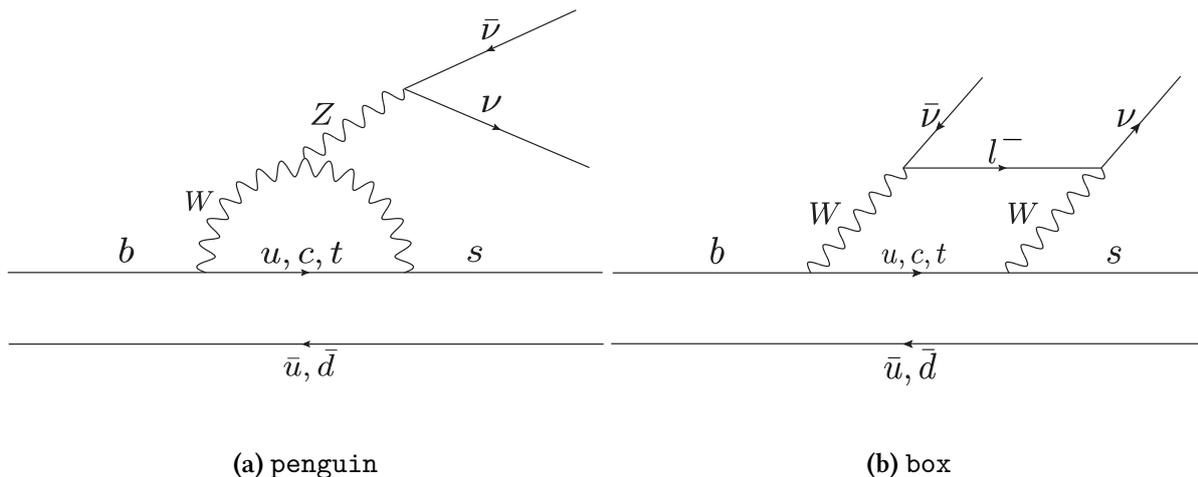
Figure 2.2.: Feynman diagrams for the decay $B \to K^{(*)}\nu\bar{\nu}$. In $B \to \pi\nu\bar{\nu}$ and $B \to \rho\nu\bar{\nu}$ decays, the final state s quark is exchanged for a d quark.

The necessary non-diagonal CKM-matrix element V_{ts}[3] and the loop suppression are the main reason for a very low predicted branching fraction; recent results are exemplified in Table 2.1 on the next page. The result only depends on V_{ts} because the contributions of the u and c quark interfere destructively due to the GIM mechanism.

The process can be described by an effective hamiltonian where all heavy fields are integrated out. The interaction is thus described by an effective 4-fermion vertex, thus factorizing a perturbatively calculable part from the expression. The quark level process within the SM is given by the following expression.

$$\mathcal{H}_{eff}^{\text{SM}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* C_L^{\text{SM}} \mathcal{O}_L \quad + \text{h.c.} \tag{2.9}$$

$$\text{with } \mathcal{O}_L^{SM} = \frac{e^2}{16\pi^2} \left(\bar{s}\gamma_\mu P_L b \right) \left(\bar{\nu}\gamma^\mu P_L \nu \right) \tag{2.10}$$

$$C_L^{\text{SM}} = -\frac{X_t}{s_w^2} \qquad X_t = 1.469 \pm 0.017 \tag{2.11}$$

Here, V_{tb} and V_{ts} are elements of the CKM-matrix, G_F is the Fermi constant, C_L^{SM} the Wilson coefficient in the SM, s_w the sine of the Weinberg angle, and X_t a numerical theoretical parameter dependent on the ratio of top and W mass. s, b and ν are the spinors of the involved quarks and neutrinos. The branching fractions of the full decays $B^+ \to K^+ \nu\bar{\nu}$ and $B^0 \to K^{*0} \nu\bar{\nu}$ on meson level are given by the following expression.

$$\frac{d\mathcal{B}\left(B^+ \to K^+ \nu\bar{\nu}\right)}{dq^2} = \tau_{B^+} 3 |N|^2 \frac{X_t^2}{s_w^4} \rho_K \left(q^2\right) \tag{2.12}$$

$$\frac{d\mathcal{B}\left(B^0 \to K^{*0} \nu\bar{\nu}\right)}{dq^2} = \tau_{B^0} 3 |N|^2 \frac{X_t^2}{s_w^4} \left[\rho_{A_1}\left(q^2\right) + \rho_{A_{12}}\left(q^2\right) + \rho_V\left(q^2\right) \right] \tag{2.13}$$

$$N = V_{tb} V_{ts}^* \frac{G_F \alpha}{16\pi^2} \sqrt{\frac{m_B}{3\pi}} \tag{2.14}$$

Here, τ_B is the B lifetime, m_B the B mass, α the fine structure constant at M_Z, the ρs are rescaled form factors which have to be taken from lattice QCD calculations. They are responsible for the largest source of uncertainty on the theoretical prediction. In Equation (2.12) and Equation (2.13), it is noticeable that the expressions for the neutral and charged cases only differ in the lifetime of the decaying B meson. The dependency on the spin of the final state hadron is contained in the soft parts, i.e. in the ρ terms.

---

[3] or V_{td} for $b \to d\nu\bar{\nu}$ decays

2.2.2. Related Decays

Diagrams of the type displayed in Figure 2.2 on page 9 describe not only $B \to h\nu\bar{\nu}$ decays but also related decays where either quarks or leptons are exchanged for other flavors. Examples are the decay $B^0 \to K^{*0}l^+l^-$ and $K \to \pi\nu\bar{\nu}$. In addition, the $\bar{b} - s$ annihilation in the decay $B_s \to l^+l^-$ exhibits the same topology and the same contributing CKM-matrix elements as the $b \to s\nu\bar{\nu}$ penguin.

The decay $B^0 \to K^{*0}l^+l^-$ has drawn a lot of attention in recent years. It is related to the $B \to h\nu\bar{\nu}$ decay by simply exchanging a pair of neutral leptons by a pair of charged leptons. However, this opens a new production channel because both diagrams in Figure 2.2 on page 9 can also proceed via a photon propagator in case of excited final states (K^* and ρ). The LHCb collaboration measured a $3.7\,\sigma$ local deviation from the SM expectation in one q^2-bin of the angular variable P_5' [A$^+$13b] and a $2.6\,\sigma$ deviation in the ratio of the branching fractions of $B^+ \to K^+e^+e^-$ to $B^+ \to K^+\mu^+\mu^-$ called R_K [A$^+$14a]. The result in P_5' was confirmed by LHCb with $3\,\text{fb}^{-1}$ worth of data [A$^+$16a] and by a completely independent Belle analysis measuring a deviation of $2.1\,\sigma$ in the same direction. Depending on the model of new physics, processes affecting $b \to sl^+l^-$ decays will also affect $b \to s\nu\bar{\nu}$ decays. Correlations between both decays can contribute to the understanding of the deviation observed by LHCb and Belle. This is for example the case for Z' explanations [BGNNS15]. Not only are $B \to h\nu\bar{\nu}$ decays suited to distinguish between several explanations for the deviations in $B^0 \to K^{*0}l^+l^-$, they are also theoretically cleaner. This is due to the fact that no dipole operator has to be taken into account and that there are no long range effects between the hadron and the neutrinos as opposed to a pair of charged leptons in the final state.

The quark level diagram of the decay $K^+ \to \pi^+\nu\bar{\nu}$ is achieved by replacing the b by an s and the s by a d quark. If new physics models couple universally to all flavors, the effect on $K^+ \to \pi^+\nu\bar{\nu}$ must be the same as the one on $B \to h\nu\bar{\nu}$. The knowledge of the decay $K^+ \to \pi^+\nu\bar{\nu}$ will make substantial progress in the next few years. The NA62 experiment at Cern, specifically designed to measure that process, started data taking in 2015 and is expected to collect sufficient statistics to observe the decay and measure its branching fraction to a precision of $10\,\%$ within the next few years [The16]. The large correlation between those decays can be observed in Figure 2.3 on page 15 where the case of minimal flavor violation is plotted.

The decay $B \to h\nu\bar{\nu}$ offers crucial information on one of the most exciting fields of particle physics today. Together with $B \to K^{(*)}l^+l^-$ and $K^+ \to \pi^+\nu\bar{\nu}$ decays, it has the power to contribute considerably to our understanding of the SM and possible extending theories.

Table 2.1.: Theoretical predictions within the SM and current best experimental limits

	$\mathscr{B}(B \to h\nu\bar{\nu})\,[10^{-5}]$		
	[BBBG09]	[BGNNS15]	limit @ $90\,\%$ C.L.
$B^+ \to K^+\nu\bar{\nu}$	0.36	0.42	1.3
$B^0 \to K^{*0}\nu\bar{\nu}$	-	0.99	5.5

2.2.3. Models of New Physics Discussed in the Context of $B \to h\nu\bar{\nu}$

Generic Right Handed Extensions In general, Equation (2.9) on page 10 can contain a right handed operator alongside the left handed one. The effective hamiltonian would be similar to Equation (2.9) on page 10 where $C_L\mathcal{O}_L$ is substituted by $C_L\mathcal{O}_L + C_R\mathcal{O}_R$ where $\mathcal{O}_R$ is expressed according to Equation (2.10) on page 10 only with a right handed projector P_R.

Although C_L and C_R are complex numbers, physical observables, foremost the $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to K^{*+}\nu\bar{\nu}$ branching fractions, only depend on two effective parameters ϵ and η.

$$\varepsilon \quad = \frac{\sqrt{|C_L|^2 + |C_R|^2}}{\left|C_L^{\mathrm{SM}}\right|} \qquad\qquad \eta \quad = \frac{-Re\left(C_L C_R^*\right)}{|C_L|^2 + |C_R|^2} \qquad (2.15)$$

$$\mathcal{R}_K = \left(1 - 2\eta\right)\varepsilon^2 \qquad\qquad \mathcal{R}_{K^*} = \left(1 + \kappa_\eta\eta\right)\epsilon^2 \qquad (2.16)$$

Where $\mathcal{R}_K$ is the ratio of the actual value of the $B^+ \to K^+\nu\bar{\nu}$ branching fraction to the SM prediction, $\mathcal{R}_{K^*}$ is the corresponding ratio for the excited state, and κ_η is a form factor dependent term whose numerical value is known. So in principle, since this work attempts to measure $\mathcal{R}_K$ and $\mathcal{R}_{K^*}$, a deviation from the SM point $(1, 0)$ in the $\varepsilon - \eta$ plain could hint to non SM right handed contributions.

Z' Models The Z boson in the penguin in Figure 2.2 on page 9 can be replaced by a heavier boson with similar properties (Z'). The Z' might be generated by various effects such as $SU(2)_L$ as a subgroup of a larger $SU(3)_L$, so called 331 models or by adding an additional gauge group [BGNNS15].

Light Invisible Scalars A scalar gauge-singlet could be included in the calculation and would enhance the branching fraction $B \to h\not{E}$ where $\not{E}$ represents missing energy. The experimental signature for such a long lived invisible scalar would be very similar to a pair of neutrinos especially for a hypothetical new particle with a low mass. This scalar, in the case of it being stable, could interact via a Higgs exchange with ordinary matter particles and could even potentially constitute a dark matter candidate.

Flavor Non-universal Coupling Experimental constraints from $B \to K^{(*)}l^+l^-$ and $B_s \to \mu^+\mu^-$ measurements would not confine possible values of $B \to h\nu\bar{\nu}$ branching fractions if the additional physics would only couple to third generation fermions. $B \to K^{(*)}\nu\bar{\nu}$ decays in fact provide the best test for such models [BGNNS15]. It is notable that the constraints from $B \to K^{(*)}l^+l^-$ and $B_s \to \mu^+\mu^-$ would be weakened even if the special coupling would be to another generation. This is due to the fact that the process is averaged over all generations in question. This implies that even if the new physics would only couple to the second generation, which might

be assumed if the $2.6\,\sigma$ deviation in R_K measured by LHCb is taken as a statistical significant evidence, the experimental constraints on the parameters of certain models become less tight.

Leptoquark Models   Leptoquarks are hypothetical particles charged under all three symmetry groups in the SM. They therefore can couple to a lepton and a quark in a single vertex and are a feature of unified theories with a larger symmetry group. $b \rightarrow s\nu\bar{\nu}$ transitions are possible on tree-level in such a scenario, as is exemplified in Figure 2.4 on page 15. It has been noted that a single leptoquark suffices in explaining recently observed deviations in flavor observables from their respective SM predictions, e.g. [BFK15]. Most of these models have in common that they do not couple universally to all lepton flavors, since they would be in contradiction with other flavor observables otherwise, most notably limits on lepton flavor violating processes.

The variety of models proposed in the literature make it difficult to make definitve statements about the effect new physics might have on $B \rightarrow h\nu\bar{\nu}$ decays. Current experimental boundaries leave a lot of room for models to either suppress or enhance the branching fractions of $B \rightarrow h\nu\bar{\nu}$ decays significantly.

2.2.4. Previous Results

There have been various previous attempts to measure $B \rightarrow h\nu\bar{\nu}$ decays, the first by the CLEO collaboration in 2000 [B$^+$01]. Back then the analysis was a modification to a search for the decay $B \rightarrow \tau\nu$. Since then, measurements explicitly dedicated to examine $B \rightarrow h\nu\bar{\nu}$ decays have been attempted by the BaBar collaboration both with semileptonic tag on a data sample of $459 \times 10^6\,B\bar{B}$ pairs [dAS$^+$10] and hadronic tag on a data sample of $471 \times 10^6\,B\bar{B}$ pairs [L$^+$13a], as well as by the Belle collaboration with hadronic tag on a datasample of $772 \times 10^6\,B\bar{B}$ pairs [L$^+$13b]. There are preceding analyses with less statistics for both experiments replaced by the aforementioned three. So far, no analysis managed to measure a significant signal and only upper limits on the branching fraction have been published. The results are summarized in Figure 2.5 on page 16. With the exception of the $B^+ \rightarrow K^+\nu\bar{\nu}$ mode, the current best limit is given by the Belle analysis utilizing the hadronic full reconstruction [L$^+$13b].

The absence of evidence for $B \rightarrow h\nu\bar{\nu}$ decays is not surprising since theoretical calculations predict very low branching fractions for the various modes [BGNNS15]. The gap between theoretical predictions and experimental exclusion limits varies strongly by channel. It is smallest for the $B^+ \rightarrow K^+\nu\bar{\nu}$ channel, where it amounts roughly to a factor of three.

2.3. Semileptonic and Hadronic *B* decays

Since the b quark is lighter than its up-type partner the t quark, B mesons can not decay within the third family and have to decay via offdiagonal transitions in the CKM-Matrix. In $(95 \pm 5)\,\%\,(B^0)$, $(97 \pm 4)\,\%\,(B^+)$ of all cases [O$^+$14] this happens via a $b \rightarrow c$ transition on quark level; in most of

such cases via a $B \to D$[45] transition on meson level. Since the absolute value of the CKM-Matrix element V_{cb} is significantly larger than that of V_{ub}, and since transitions from down-type quarks to down-type quarks are forbidden on tree level within the framework of the SM, those processes dominate by a large margin. The D meson can either be accompanied by other hadrons (hadronic decay) or by a charged lepton and the corresponding neutrino (semileptonic decay). Semileptonic B decays including either a ground state D or the first excited state D^* account for $7.12\,\%$ $(7.96\,\%)$ of all B^0 (B^+) decays [O$^+$14]. The decays used in the hadronic full reconstruction account for $10.4\,\%$ $(11.9\,\%)$ of all B^0 (B^+) decays [FKK$^+$11]. The reconstruction of semileptonic modes can be done with a higher efficiency as compared to purely hadronic modes, see Section 5.2.1 on page 32. This leaves a big chunk of the Belle dataset unexplored by the previous analysis and provides the only dataset available for the study of $B \to h\nu\bar{\nu}$ decays before Belle II will have acquired a sufficient amount of data to repeat the analysis with less statistical uncertainty.

---

[4]In this section, D meson also refers to excited states of D mesons and D_s mesons
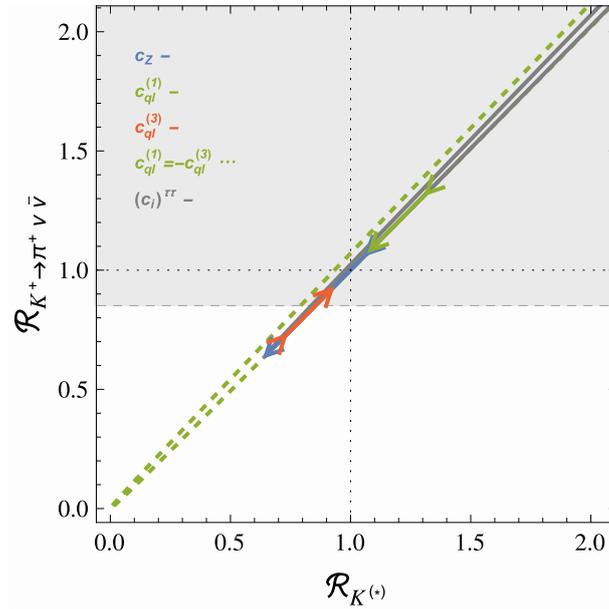
[5]and to a smaller extent charmonium states

Figure 2.3.: Correlation between the deviations from the SM predictions dependent on various Wilson coefficients for the $K \to \pi\nu\bar{\nu}$ and $B \to K^{(*)}\nu\bar{\nu}$ assuming minimal flavor violation and universal coupling. The shaded line represents experimental uncertainties, taken from [BGNNS15].
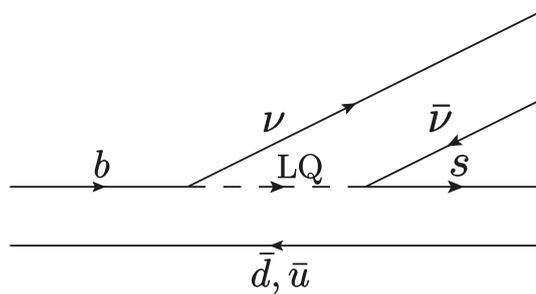

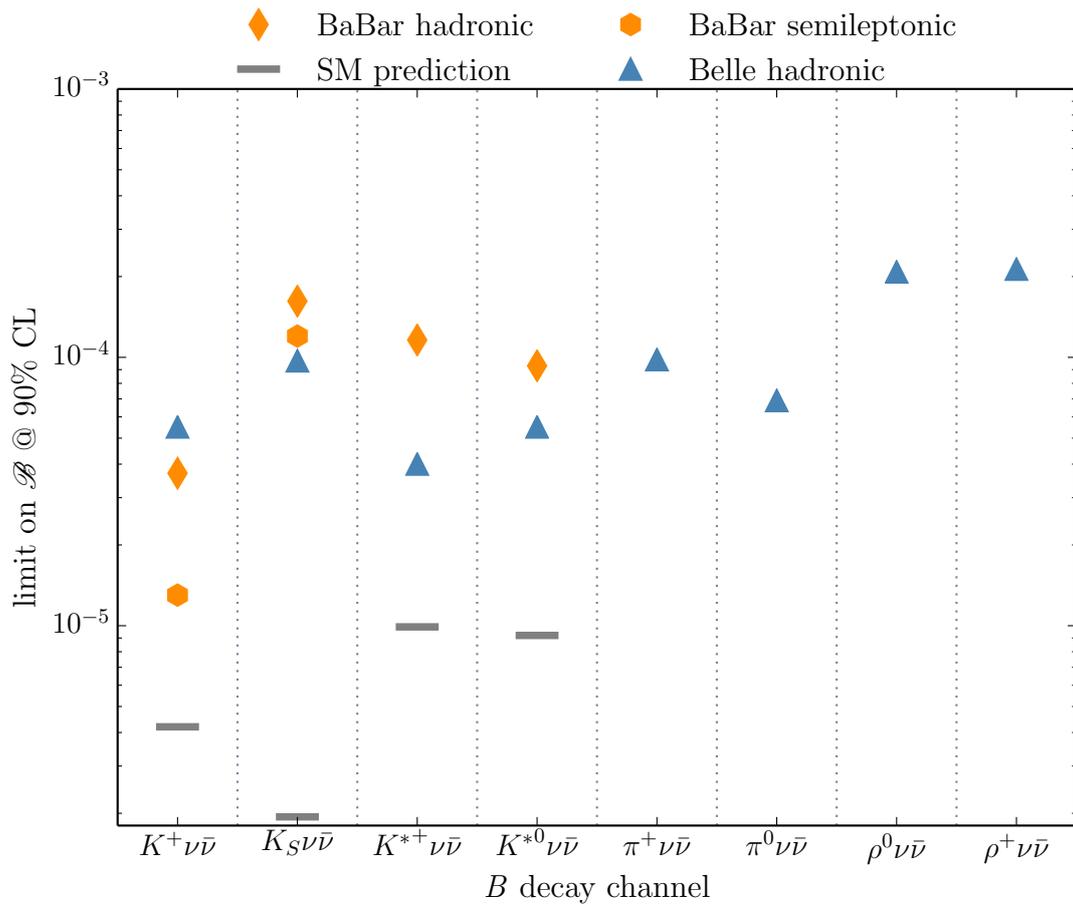Figure 2.4.: Exemplified feynmanngraph with a leptoquark on tree level.

Figure 2.5.: Summary of existing measurements.

The Belle Experiment

3.1. *B* Physics Experiments

Dedicated B physics experiments have been located at leptonic colliders (CLEO, BaBar and Belle), at a hadronic collider (LHCb), and at a hadron-lepton collider (HERA-B). B physic experiments at lepton-lepton colliders make use of the spectroscopic properties of the bottomonium system. The principle reaction is the annihilation $e^+e^- \to b\bar{b}$; the $b\bar{b}$ pair subsequently hadronizes to a bound system. The first four $b\bar{b}$ resonances with $J^{CP} = 1^{--}$ quantum numbers, the same as those of the virtual photon in the propagator of the e^+e^- annihilation, are displayed in Figure 3.1 on the following page. The masses of the first three resonances are not sufficient to produce a B meson pair, consequently they only decay into lighter hadrons or leptons. A fact that manifests itself in the narrower width of the first three $\Upsilon(\mathrm{nS})$ states in Figure 3.1 on the next page. The $\Upsilon(4\mathrm{S})$ resonance however is the first excited $b\bar{b}$ state heavier than two unexcited B mesons combined. Hence, machines like the KEKB accelerator, where the Belle experiment is located, tune the beam energies of their electron and positron beam such that the e^+e^- center of mass energy $\sqrt{s_{e^+e^-}}$ coincides with the mass of the $\Upsilon(4\mathrm{S})$.

The branching fraction of the process $\Upsilon(4\mathrm{S}) \to B\bar{B}$ has been measured to be larger than $96\,\%$ of the overall decay rate of the $\Upsilon(4\mathrm{S})$; all other known decays have branching fractions of 10^{-4} or lower. Each B meson is produced almost at rest in the $\Upsilon(4\mathrm{S})$ center of mass system (CMS) with a kinetic energy of $10\,\mathrm{MeV}$ at a mass of $5.28\,\mathrm{GeV}$. Moreover, both B meson are produced by the strong decay of a common mother particle and maintain the quantum numbers of that system and stay in an entangled system until one decays. Since the probability for hard interactions is rather low at e^+e^- colliders, in all practically relevant cases only one pair of B mesons is produced in one single event alongside no further hard interactions. This sets B factories in sharp contrast to experiments at hadronic colliders where the cross section for B meson production is strongly enhanced. The difference amounts to over a factor of a million if LHCb and Belle are compared. The production cross section for the former is measured to be $77\,\mathrm{\mu b}$ at $\sqrt{s_{pp}} = 7\,\mathrm{GeV}$ [A$^+$13a], compared to about $1.2\,\mathrm{nb}$ in the case of $e^+e^- \to \Upsilon(4\mathrm{S})$ [B$^+$14]. Nevertheless, since the hadronization of the initial b quarks does not proceed via a coherent state and a huge variety of other particles is produced at

the primary vertex, B factories can still contribute largely to the progress of the field. Processes which are better measurable at B factories than for instance at LHCb comprise foremost B decays into final states with missing energy but also inclusive measurements.

Since it is kinematically allowed to decay strongly, the decay time of the $\Upsilon(4S)$ is smaller compared to lower $b\bar{b}$ resonances and the decay width is enhanced. In turn this implies that the height over non-$b\bar{b}$ processes is significantly smaller than the one of the lighter $\Upsilon(nS)$ resonances resulting in a large contribution of non-$b\bar{b}$ events, ergo $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, $c\bar{c}$ events, referred to as *continuum* to the e^+e^- event rate at the respective center of mass energy. This continuum amounts to about three quarters of all events at the maximum of the $\Upsilon(4S)$ resonance.
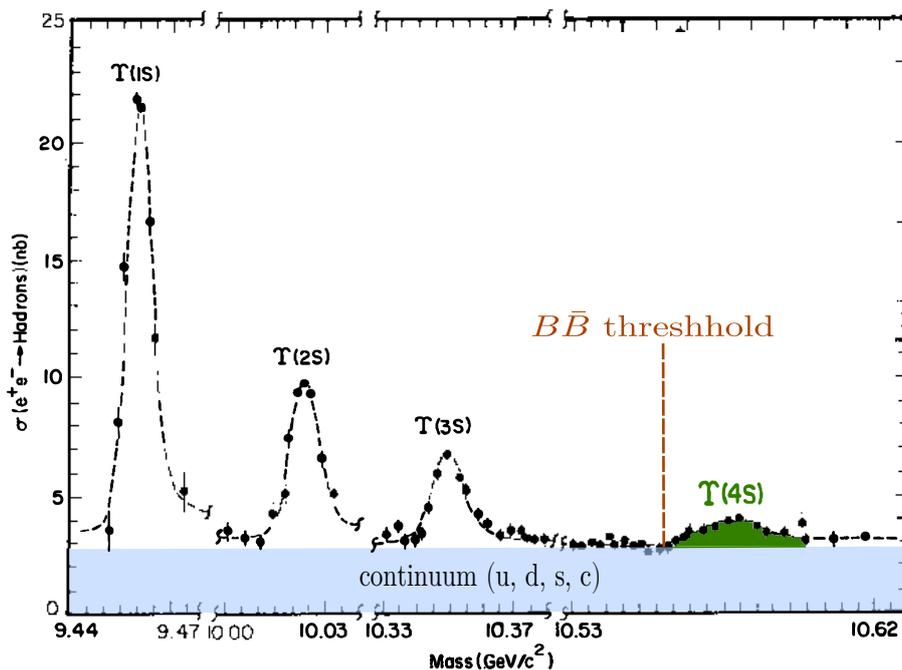

Figure 3.1.: $\Upsilon(nS)$ resonances dependent on $\sqrt{s_{e^+e^-}}$, taken from [BS93]

3.1.1. The KEKB Accelerator

The presented analysis was carried out at the Belle detector located at the KEKB accelerator in Tsukuba, Japan. KEKB is an asymmetric B factory with a high energy ring *HER* storing electrons at $8\,\mathrm{GeV}$, and a low energy ring *LER* storing positrons at $3.5\,\mathrm{GeV}$. Both collide at a single interaction point *IP* under a crossing angle of $22\,\mathrm{mrad}$. KEKB operated from 1999 to 2010 constantly increasing the instantaneous luminosity up to $2.1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Key technical details of the KEKB accelerator are summarized in Table 3.1 on the facing page, data from [B+14, KK03]. A sketch of the facility is displayed in Figure 3.2 on page 24.

Table 3.1.: Technical details of KEKB

Circumference	3016 m
Frequency	509 MHz
e^- beam energy	8.0 GeV
e^+ beam energy	3.5 GeV
Peak luminosity	$2.1 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$
Number of bunches	1584

3.2. The Belle Detector

The Belle detector is designed as a 4π general purpose detector with certain subsystems dedicated to various tasks. The detector was build to reconstruct all decay products of the $\Upsilon(4S)$ resonance, measure the momentum of each final state particle, and distinguish between different types of particles. Those comprise charged hadrons ($\pi^\pm$, $K^\pm$ $p^\pm$), charged leptons ($e^\pm$ and $\mu^\pm$), photons, and K_L^0. The following passages follow information taken from [A$^+$02, B$^+$14] where more detailed descriptions can be found.

3.2.1. The Silicon Vertex Detector (SVD)

The SVD consists of crosswise arranged silicon strip detectors. Its main purpose is vertexing, determining the decay position of B mesons as precisely as possible. When Belle started operations in 1999, it consisted of three layers of strip detectors before it was upgraded to four layers in 2003. The upgrade also increased the acceptance of the SVD to $91\,\%$ by increasing the coverage to polar angles θ of $17° < \theta < 150°$. The resolution of the second stage depending on the momentum and the polar angle can be parametrized as[1]

$$\sigma_{r\phi} = 21.9\,\mathrm{\mu m} \oplus \frac{35.5}{p\beta\sin^{\frac{3}{2}}\theta} \qquad \sigma_z = 27.8\,\mathrm{\mu m} \oplus \frac{31.9}{p\beta\sin^{\frac{5}{2}}\theta}$$

where p is the momentum and $\beta = \frac{|\mathbf{v}|}{c}$ the β factor. Since it is the part of the detector closest to the beam, it suffers from the largest radiation background. Also, it makes the SVD important for the detection of low momentum tracks for the same reason.

3.2.2. The Central Drift Chamber (CDC)

The CDC's main objective is the detection of charged particles and the measurement of their respective momentum. It is located adjacent to the SVD covering the same polar region. It consists

---

[1] Here as in the further course of this thesis, $\oplus$ indicates the square root of the quadrature sum.

of a cylindrical gas vessel filled with a mixture of $50\,\%$ ethane and $50\,\%$ helium. Therein, 50 layers of drift chambers are contained, grouped into 6 axial superlayers and 5 stereo superlayers, tilted in a small angle with respect to the beam axis to provide information on the z part of the momentum. All in all, the number of drift cells amounts to 4800. Each of the square shaped cells is formed by a sense wire in its center and eight field wires at its border. A high electric field is applied between sense and field wires. Electrons from gas atoms ionized by passing charged particles are accelerated until their energy suffices to ionize another atom, those electrons in turn ionize more atoms, leading to an avalanche of electrons. The amplification by this avalanche effect creates a measurable charge deposition at the sense wire. Single hits are combined to tracks and subsequently fitted to measure the curvature and thus the momentum. The momentum resolution of CDC only tracking is given by the following expression.

$$\frac{\sigma_{p_T}}{p_T} = 0.19\,\% \oplus \frac{0.30}{\beta}$$

In addition to measuring the momenta of charged particles, the CDC contributes to the identification of different particles, especially electrons, by measuring the energy loss per flight distance $\frac{dE}{dx}$.

3.2.3. The Aerogel Čerenkov Counter (ACC)

The ACC is one of two subsystems dedicated to particle identification. It utilizes the dependence of the Čerenkov threshold on the velocity rather than the momentum to distinguish between particles with different masses, most prominently kaons and pions. The ACC consists of 1188 modules of $1728\,\mathrm{cm}^3$ aluminum boxes filled with a silica aerogel, specially designed and produced for the Belle Experiment. Out of those 1188 boxes, 960 are installed in the barrel region, the rest in the forward region, the backward section is not instrumented. The refraction index of that aerogel is a function of the polar angle and covers a range of $1.010 < n < 1.030$. The Čerenkov signal is read out by one or two photo multiplier tubes *PMT*s. Each module faces the interaction region. The ACC is constructed to distinguish between high momentum particles with momenta above $1\,\mathrm{GeV}$ by measuring the pulse height of the individual signal, essentially counting photons.

3.2.4. The Time of Flight Counter (TOF)

The TOF is a complementary instrument to the ACC, for it provides particle separation for low momentum tracks, $p < 1.2\,\mathrm{GeV}$. The TOF measures the time it takes a particle from the interaction region to the TOF. Each of the 64 TOF modules mounted in the barrel region of the detector consists of two plastic scintillators with a PMT glued directly onto it. The timing resolution of the TOF is $100\,\mathrm{ps}$. This precise timing information is also used for the trigger.

3.2.5. The Electromagnetic Calorimeter (ECL)

The ECL consists of 8736 thallium doted CaI crystals, 6624 in the barrel, 1152 in the forward and 960 in the backward region. Its main purpose is the reconstruction and energy measurement of photons and electrons. The central axis of each crystal is directed slightly to the side of the interaction region to minimize losses through gaps between crystals. The length of one such crystal corresponds to 16.2 free interaction lengths, X_0. Each crystal is wrapped in a reflective material and read out by two PMTs glued to its rear end. Clusters are reconstructed from a crystal measuring an energy deposition of at least $10\,\text{MeV}$, called seed crystal, and an crystals within an array of 5×5 crystals around the seed crystal, measuring at least $0.5\,\text{MeV}$. Tracks from the CDC are extrapolated to the ECL to distinguish between electrons and photons. The energy resolution of the ECL can be subdivided into a noise, a leakage and a systematic term and is parametrized as follows [I$^+$00].

$$\frac{\sigma_E}{E} = \frac{0.066\,\%}{E} \oplus \frac{0.81\,\%}{\sqrt[4]{E}} \oplus 1.34\,\% \qquad E \text{ in GeV} \tag{3.1}$$

3.2.6. The Magnet

Each previously described detector component operates in a homogeneous $1.5\,\text{T}$ magnetic field generated by a $3.92\,\text{m}$ long superconducting solenoid with an effective coil radius of $1.8\,\text{m}$ whose purpose is to bend the trajectories of charged particles to measure their momentum.

3.2.7. The K_L^0 and Muon System (KLM)

The iron of the return yoke of the magnet is utilized as sampling material for Belle's muon and K_L^0 system. The detective material consists of 15 layers of resistive plate counters, glass plates coated with a highly resistive ink, separated by a gaseous[2] gap, and loaded with high voltage. Charged particles, either muons or products from hadronic interactions in the sampling material, cause discharges in between plates which are recorded as signal. The modules are assembled crosswise to provide spatial information.

3.3. Particle Identification

Being able to distinguish between various particles simultaneously with high efficiency and high purity is of utmost importance for flavor physics analyses. Moreover, since the tracking only measures the three-momentum of a charged particle, energy information can only be gained by knowledge of the mass of the particle. Five charged particles are long lived enough to cause a signal in the detector, $e^\pm$, $\mu^\pm$, $\pi^\pm$, $K^\pm$, and $p^\pm$. To separate between those, likelihood ratios have been

---

[2]Ar, butane, HFC134a in a ratio 30:8:62

constructed, utilizing the output of every subsystem with the exception of the SVD. Lepton likelihood ratios are defined as the ratio of the electron/muon likelihood relative to everything else, hadron likelihood ratios set to the following expression.

$$\text{PID}_{XY} = \frac{\prod_i \mathcal{L}_i^X}{\prod_i \mathcal{L}_i^X + \prod_i \mathcal{L}_i^Y} \tag{3.2}$$

The shape of the probability distribution functions *PDFs* of the individual particles in the respective observable of the detector subsystem has been determined in Monte Carlo studies and adjusted to resemble real data via control samples.

$K\pi$-id Kaons are separated from pions utilizing information from the CDC, the TOF, and the ACC. The kaon (pion) probability evaluates to $P_i = P_i^{\text{CDC}} \times P_i^{\text{TOF}} \times P_i^{\text{ACC}}$.

pK-id The same subdetector information as in the previous case can be used to separate kaons from protons. Again, PDFs are build from simulated data and validated on control samples.

e-id The electron likelihood utilizes five different variables to optimize separation, three out of those make use of the ECL. The first is the χ^2 value of the match between the candidates track and its cluster. The second is the ratio between the calorimetically determined energy of the candidate and its fitted momentum. The third is the ratio between the energy contained in a 3×3 array around the seed cell and the energy contained in a 5×5 array, being the maximal size for clusters in the Belle calorimeter by definition. The maximum of this ratio differs from 1 in case of electrons. Fourth comes the $\frac{dE}{dx}$ ratio mentioned in Section 3.2.2 on page 19, the fifth observable is the pulse height in the ACC.

μ-id The muon-id depends on the KLM. Tracks are extrapolated to the KLM and matched with hits in the KLM, the χ^2 value of that matching and the spatial difference between the extrapolated track and the reconstructed hit are used to construct the likelihood ratio. Y in Equation (3.2) refers to pions and kaons only in the case of muons. The muon identification efficiency suffers from the dependence on the KLM. A muon candidate formed at the interaction regions needs at least a transverse momentum of $p_\text{T} > 450\,\text{MeV}$ to reach the KLMs innermost layer, otherwise it curls inside the detector.

More detailed information on particle identification can be found in [Nak02, B$^+$14].

3.4. Triggering and Data Acquisition

The Belle Experiment pursued a larger number of physics goals besides b physics. It contributed valuably to charm physics, τ physics, and baryon spectroscopy. Bhabha and $\gamma\gamma$ events are used for callibration purposes. The Belle trigger system therefore needs more than one trigger algorithm to meet all requirements.

Events are triggered at first by the Level 1 *L1* hardware trigger which combines the output of 8 subdetectors in a hardwired logic to make a trigger decision. An event passes the L1 if it either exhibits three or more tracks, four or more isolated neutral clusters in the ECL, or a high level of energy in the ECL. Events triggered by the presence of charged tracks are subsequently processed by an online computer farm, the level 3 *L3* trigger. This tag category is called hadronic. The L3 selects hadronic events by implementing a fast tracking algorithm and judging tracks based on quality criteria such as impact parameters. The overall trigger efficiency of hadronic events was estimated to exceed $99\,\%$ after L3 [B$^+$14]. Trigger rates mostly lay in a region between $200\,\mathrm{Hz}$, at the begin of operations, and up to $500\,\mathrm{Hz}$, in some runs towards the end of operations. The trigger rate for bhabha and $\gamma\gamma$ events was attenuated artificially by a factor of 100.

The raw data from the subdetectors is processed by a computer farm in a fourth level *L4*, applying another software filter and reconstructing physical quantities from electronic signals in the process. Events surviving the criteria applied in the L1 and L3 are reconstructed into a data format utilizable for physics analysis. Tracks of charged particles are found, fitted, and refitted with a Kalman filter algorithm which takes into account effects like non-uniformities of the magnetic field, multiple scattering, energy loss, and uses more precise timing information. They are then extrapolated to other detector parts and combined with the information of those to obtain PID information, as described in Section 3.3 on page 21. In a final step charged particle candidates are written out in a Belle specific data format. Neutral particle candidates are formed from ECL clusters (γ), KLM (K_L^0), and tracks (K_S^0 and Λ^0). More information about tracking at Belle can be found in [Bel00].

The data acquisition for Belle was subject to perpetual change to cope with the ever increasing luminosity. Further information about data acquisition and triggering can be found in [SIK$^+$02].
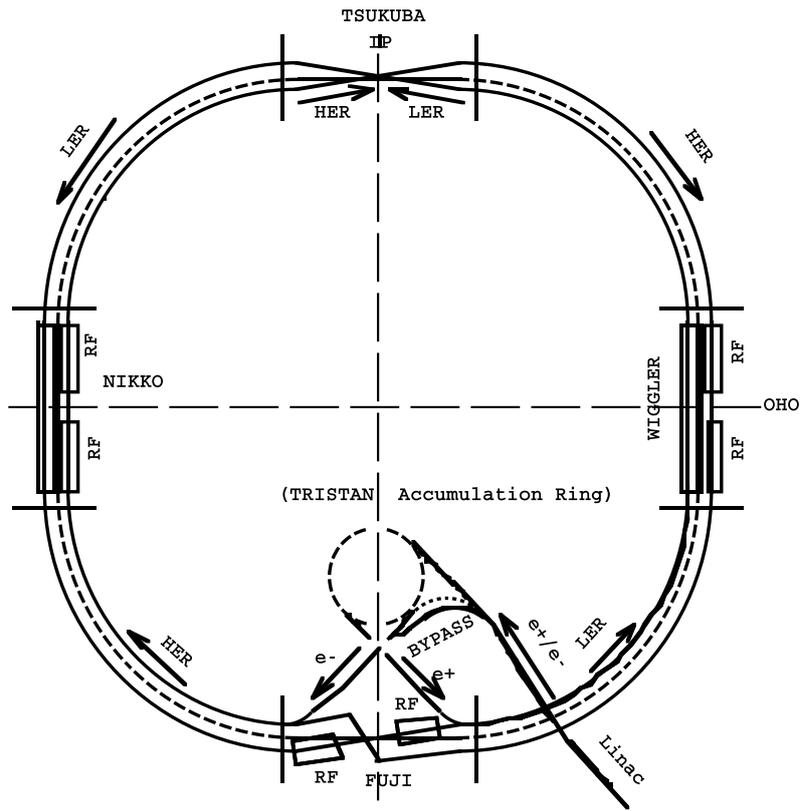
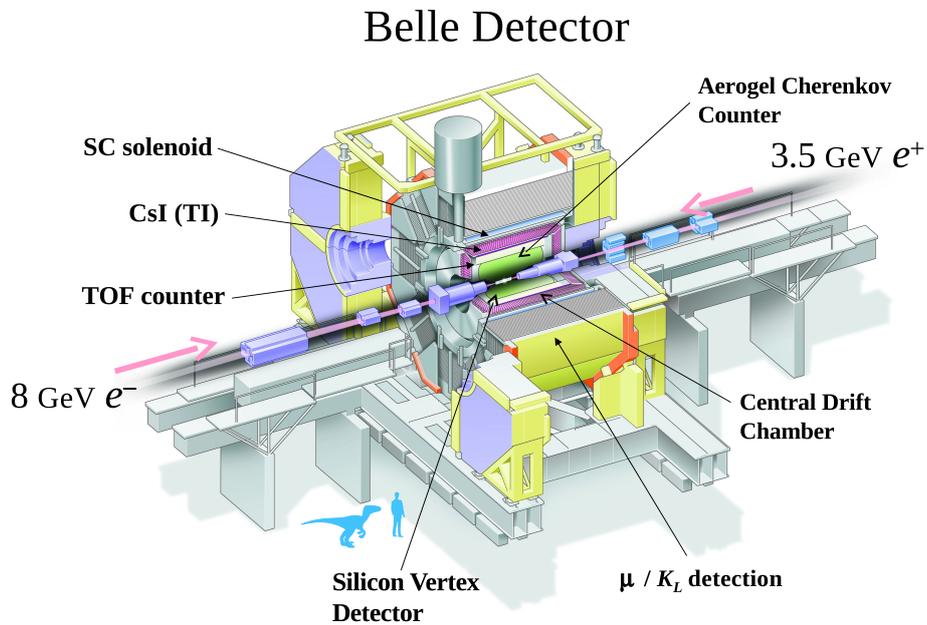Figure 3.2.: KEKB accelerator complex, taken from [KK03].

Belle Detector

Figure 3.3.: Overview over the Belle detector and its relevant subparts, courtesy by [Hei15].

Chapter 4

Tools and Methods

4.1. Multivariate Classification

The selection of signal events according to a finite set of individual variables leaves a significant amount of background events in the selected sample. In B physics, this background usually consists of a great variety of physics processes which differ from signal in various different regions of the parameter space. In rare processes like $B \to h\nu\bar{\nu}$ decays where the expected rates are very low, hence the selection is expected to be dominated by background, the separation between background and signal events is especially crucial.

It has thus been established in particle physics to make use of multivariate classification tools, the most widely used being boosted decision trees and artificial neural networks. Since this work makes ample use of artificial neural networks, this chapter gives a brief introduction to the concept and the software package used for the practical implementation in section 4.1.1.

A multivariate algorithm is a function mapping a high dimensional input vector to a single scalar $f : \mathbb{R}^n \to \mathbb{R}$. This single discriminating variable $\mathcal{N}$ combines the information of the complete input into a single variable which can subsequently be used to either cut away as much background as possible while retaining as much signal as possible, or to perform a fit on it. One big advantage of multivariate approaches is the fact that correlations between input variables are taken into account. Hence, multivariate classification is usually more powerful than a cut optimization on the same sample with the same input variables.

4.1.1. Neurobayes

Artificial neural nets have been used for decades in many fields in science. An artificial neural net consists of an arbitrary number of layers of so called nodes or neurons. The nodes in adjacent layers are connected to each other. In feed forward networks, the value of any node in the $(n-1)$-th layer is propagated to each node in the n-th layer by an activation function and is assigned a weight, adjusting the relative importance of each connection.

$$x_i^n = S \left(\sum_{j=1}^{m} \omega_{ij}^{n-1 \to n} x_j^{n-1} \right) \tag{4.1}$$

Here $\omega_{ij}^{n-1 \to n}$ is the weight of the connection from the j-th node in the $(n-1)$-th layer to the i-th node in the n-th layer and S is the activation function. A sigmoid function is often chosen as the activation function. Its main advantage being a mapping of an arbitrary real input into a compact interval.

$$S \colon \quad (-\infty, \infty) \;\to\; [0, 1] \tag{4.2}$$

This mapping exhibits a smooth behavior, asymptotically approaching $+1$ for large positive input values, and 0 for negative values with large absolute values. Here, a Fermi function is used, as exemplified below with an arbitrary positive constant c.

$$S(x) = 1 - \frac{1}{1 + \exp(-cx)} \tag{4.3}$$

However, nets with complex topologies might require an activation function whose calculation consumes less time. The presented analysis utilizes the neurobayes package [FK06] for the multivariate parts involved at three steps in the procedure, as part of the semileptonic tagging (Section 5.2.1 on page 32), as part of the continuum suppression (Section 7.1 on page 53), and for the final selection (Section 7.2 on page 55). Neurobayes is a three layer feed forward network. The layers are in that order the input layer, the hidden, and the output layer. Each layer may contain an arbitrary amount of nodes. A schematic display of such a net is exemplified in figure 4.1 on the next page. The specifics of that package will now be briefly lined out.

The optimization procedure to find the optimal configuration for the net to fit the data is called training. This training is done by an algorithm named backpropagation which adapts the weights proportional to the gradient of a loss function. The minimum of this loss function determines the optimal configuration to fit the given input. Examples for such a function are a χ^2 function and an entropy loss function E_D.

$$\chi^2 = \sum_{i=1}^{N} (t_i - o_i)^2 \tag{4.4}$$

$$E_D = \sum_{i=1}^{N} \log \left(\frac{1}{2}(1 + t_i o_i + \varepsilon) \right) \tag{4.5}$$

Where $t_i \in [-1, 1]$ is the true output, o_i the output of the network in its current status, and ε is a regularization constant. The training time is shortened by updating the weights after a fixed number of training events, by default 200, and not after the whole sample. To avoid fitting mere statistical fluctuations of the training sample into the net, a process referred to as overtraining, a weight decay is implemented. The squared sum over all weights is added to the loss function multiplied by a carefully chosen proportionality factor. Such, the net is prevented from learning unphysical and merely statistical features of the training sample. The number of regularization constants can be changed. Here, three different ones are used, see Section 7.2.2 on page 57.
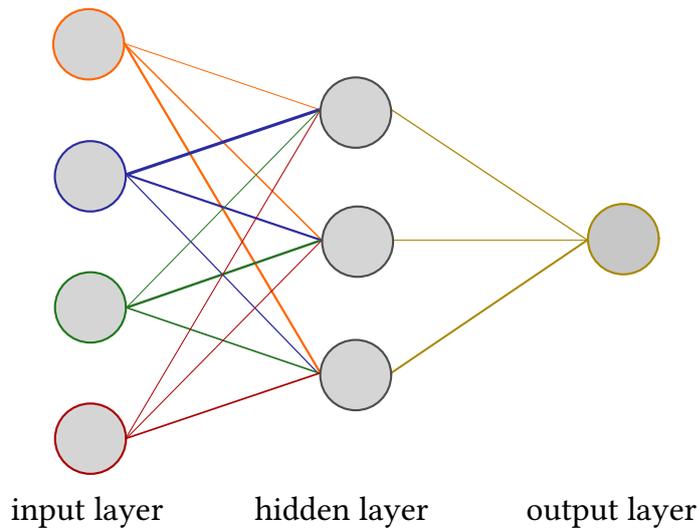
Figure 4.1.: Topology of a three layer feed forward net.

Input variables may be defined in very different numerical intervals. A sophisticated preprocessing procedure is implemented in neurobayes to enable the net to utilize all information and to assure a large separation power. This preprocessing is divided in four fundamental steps.

Equal Statistics Binning Continuous variables are rebinned into 100 bins with an unequal width. This is done such that each bin contains the same amount of input data. This treatment avoids any overproportional numerical influence of outliers to the final result.

Purity Fit B-Splines are fitted to the signal purity in each bin, smoothing out statistical fluctuations.

Normalization The input distribution is transformed to exhibit a mean of zero and a width of one by making use of the cumulative distribution. This aligning of the numerical ranges of all variables increases not only the stability of the training, it also enhances the learning speed.

Decorrelation Subsequently all input variables are ranked according to their correlation to the target. Afterwards, the input variables are decorrelated with a linear transformation diagonalizing the correlation matrix with the target variable included.

The importance of each input variable for the final training result can be assessed by a set of properties, as described below. All values for the trainings for the final selection described in section 7.2 on page 55 can be looked up in Appendix D.

Additional Significance The correlation to the target is calculated for all variables and the variable with the least correlation is removed. The Process is repeated with the $N - 1$ variables with the highest correlation to the target until only the variable with the most significance remains. The *additional significance* equals the difference between the correlation to the target of all variables with a correlation equal or higher than this variable and the correlation of all variables with a higher correlation. The value is thereafter multiplied with $\sqrt{n}$ where n is the training sample size.

Significance Only This The correlation of this variable with the target multiplied by $\sqrt{n}$.

Significance Loss The difference of the correlation of all variables to the target and that quantity without this variable, multiplied with $\sqrt{n}$.

Global Correlation The correlation of this variable to all other variables.

(a) Equalized binning

(b) Spline fits

(c) Normalization
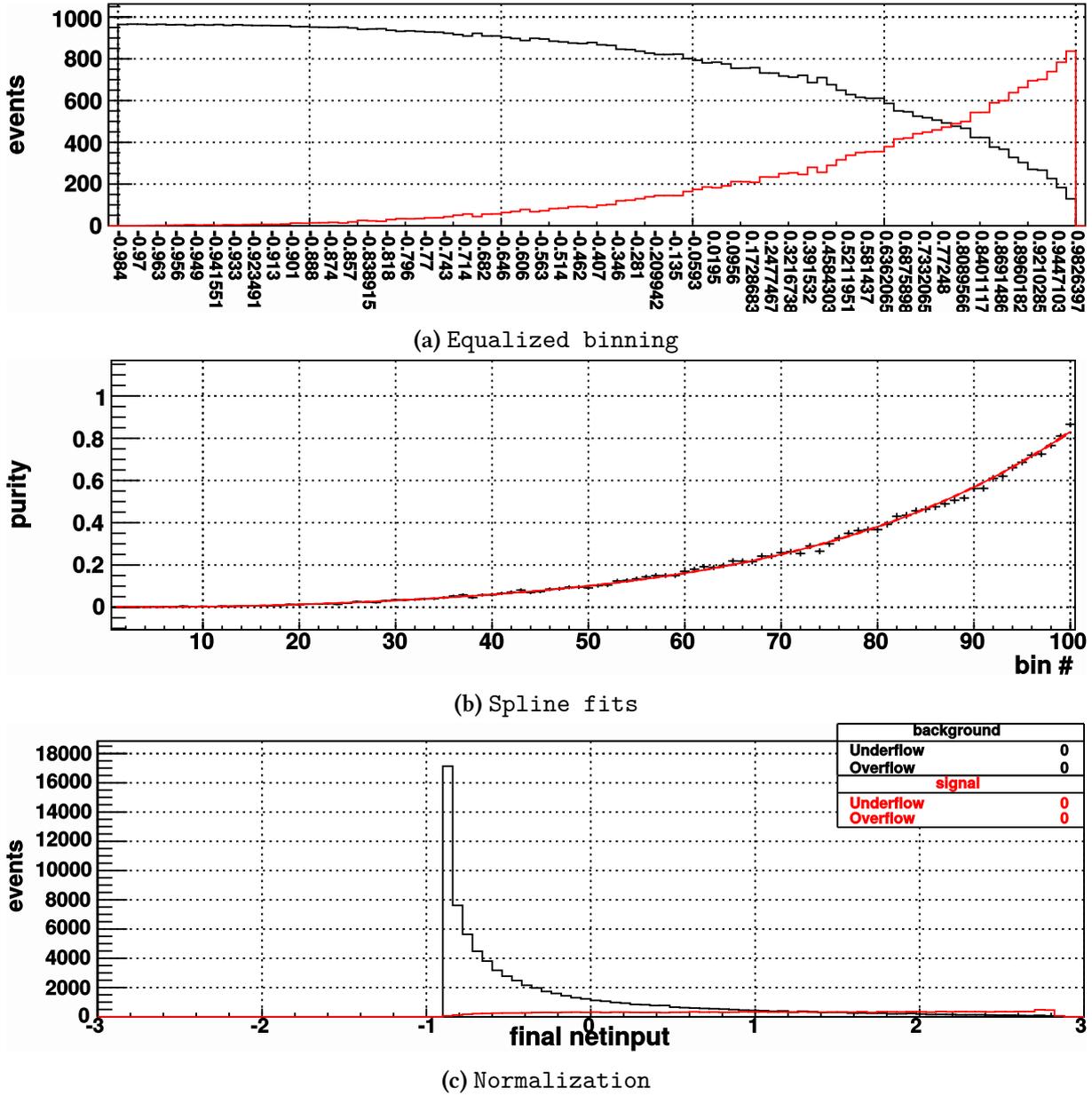
Figure 4.2.: The first two steps of the preprocessing procedure, exemplified by the continuum suppression, the most important variable in the $B^+ \rightarrow K^+ \nu\bar{\nu}$ training. The signal fraction has been set to a value of 0.2.

Chapter 5

Reconstruction

5.1. Data Sample

The analysis is performed on the complete Belle data sample recorded on the $\Upsilon(4S)$ resonance at $\sqrt{s} = 10.58\,\text{GeV}$ between 1999 and 2010 with a total integrated luminosity of $711\,\text{fb}^{-1}$, corresponding to an overall amount of $(772 \pm 11) \times 10^6 \, B\bar{B}$ pairs. For the development of this analysis, various sets of MCsamples have been used. Those MC sets are composed as follows.

generic Decays of B mesons where the b quark decays into a c quark are referred to as generic. This sample is further subdivided into a sample containing the decay of the $\Upsilon(4S)$ into a $B^+ B^-$ pair (charged) and a sample containing the decay of the $\Upsilon(4S)$ into a $B^0 \bar{B}^0$ pair (mixed). A set of simulated data equivalent to the analogous subset of the data taken over the run time of Belle is referred to as a stream[1]. Ten such streams of generic Monte Carlo are used.

continuum Events where no $\Upsilon(4S)$ is produced but instead a pair of lighter quarks, i.e. $u\bar{u}, d\bar{d}, s\bar{s}$ and $c\bar{c}$ are referred to as continuum. Only six streams of continuum Monte Carlo are available and are used to build our model.

rare Events in which the b quark of one B meson does not decay into a c quark are referred to as rare. The used data matches 20 streams of simulated events where the b quark decays into a u quark, a charged lepton and a neutrino and 50 streams for all other non $b \to c$ B decays.

signal For each signal channel, at least 10×10^6 events have been simulated in which one B meson decays into the respective signal channel while the other B meson decays generically, meaning without any specification of the channel. K^{*+}, K^{*0}, ρ^+, and ρ^0 decay channels have not been specified, each simulated signal hadron decays generically. The model according to which the samples are produced follows the phase space of three body decays.

---

[1]This definition applies to all types of background MC.

off-resonance To study backgrounds from non $B\bar{B}$ events, the Belle experiment has taken $100\,\mathrm{fb}^{-1}$ data at a center of mass energy $60\,\mathrm{MeV}$ below the nominal $\Upsilon(4S)$ mass of $10.58\,\mathrm{GeV}$. A simulation corresponding to six streams of that integrated luminosity has been utilized in concordance with those data.

5.2. Selection

The considered B meson decay is a three-body decay with two invisible daughter particles and therefore does not convey sufficient kinematical information, for instance about the hadron momentum. Because of that, the analysis depends heavily on the knowledge of the second B meson to mark (tag) the event. This second B meson is referred to as B_{tag} and is reconstructed in modes where it decays into a $D^{(*)}$ meson, a charged lepton[2], and a neutrino. A candidate is accepted if it contains a reconstructed tag candidate and the respective signal hadron, hitherto referred to as h_{sig}, but nothing else. This will be elaborated on further in this chapter.

5.2.1. Semileptonic Tagging

In contrast to fully hadronic final states of a B meson, semileptonic decays lack kinematical information, due to the neutrino in the final state. The full reconstruction utilizing hadronic decays has been well established [FKK$^+$11]; an analysis of the decay $B \to h\nu\bar{\nu}$ utilizing it has been completed [L$^+$13b]. This algorithm has been developed further to extend to semileptonic decays [Kir12] and so far two analyses have been published: one analysis on the decay $B \to \tau\nu$ [K$^+$15], and one analysis on the decay $B \to D^{*}\tau\nu$ [A$^+$16b].
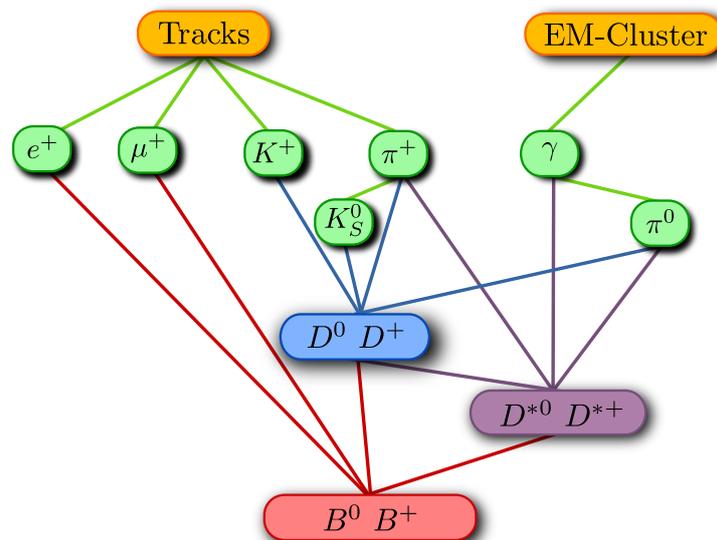
Figure 5.1.: Schematic view of the tagging algorithm, from [Kir12]

---

[2] the term lepton in this note refers to either $\mu^{\pm}$ or $e^{\pm}$ unless stated explicitly otherwise.

Analogous to the hadronic full reconstruction, a hierarchical approach to combining particle candidates to final B candidates is applied [FKK$^+$11]. Final state particle candidates are formed of tracks in the CDC and electromagnetic clusters in the calorimeter. Pairs of charged pion candidates are used to form K_S^0 candidates. π^0 candidates are formed out of a pair of photon candidates. D candidates are reconstructed out of $K^\pm$, $\pi^\pm$, π^0 and K_S^0 candidates in the next stage. D^+ mesons are reconstructed in seven different decay channels whereas D^0 mesons are reconstructed in ten different decay channels. D^* candidates are formed of a D candidate and a $\pi^\pm$ or a π^0 candidate ($D^{*\pm}$), or a D^0 candidate and a π^0 or γ candidate (D^{*0}). B_{tag} candidates are reconstructed from a D or D^* candidate and one lepton candidate. Altogether, this sums up to 216 different tag channels. The process is sketched in Figure 5.1 on the facing page.

Each step is accompanied with only very lose cuts to retain as much signal as possible[3]. In each stage, a neural net is trained whose output is taken as input in the next stage alongside with several other variables like the reconstructed mass, momentum, decay channel, spatial variables like the polar angle, and various other variables. The output of the final net trained on B_{tag} candidates, $\mathscr{N}_{\text{NB, tag}}$, is used to make the final choice on behalf of the tag side, see Section 7.2.1 on page 55 and Section 5.2.6 on page 39. The $\mathscr{N}_{\text{NB, tag}}$ distribution after the selection of the $B^+ \to K^+ \nu\bar{\nu}$ channel is displayed in Figure 5.2 on the next page. Signal in the plots refers to correctly reconstructed B candidates, cross feed to cases where the B candidate lacks a particle such as a γ or a π^0 from a $D^{*0} \to D^0 \gamma$ ($D^0 \pi^0$) decay, background to incorrectly reconstructed candidates. A listing of each reconstructed decay channel including its respective branching fraction can be found in appendix A.

This final training is designed to be uncorrelated to the angle between the momentum of the $D^{(*)}l$ system and the momentum of the B_{tag}, $\cos\theta_{B,\,D^{(*)}l}$. This variable is obtained by utilizing $M_\nu^2 = 0$ and the fact that energy and momentum conservation leave only one degree of freedom.

$$
M_\nu^2 = 0 = \left(P_B - P_{D^{(*)}l}\right)^2 = M_B^2 + M_{D^{(*)}l}^2 - 2\left(E_B E_{D^{(*)}l} - \boldsymbol{p}_B \boldsymbol{p}_{D^{(*)}l}\right)
$$

$$
\Rightarrow \quad \cos\theta_{B,\,D^{(*)}l} = \frac{2 E_B E_{D^{(*)}l} - M_B^2 - M_{D^{(*)}l}^2}{2\,|\boldsymbol{p}_B|\,\left|\boldsymbol{p}_{D^{(*)}l}\right|} \tag{5.1}
$$

Here, E denotes the zeroth component of the four momentum, $\boldsymbol{p}$ the three momentum, and M the invariant mass of the B meson or the $D^{(*)}l$ system, respectively. For correctly reconstructed B_{tag} candidates the value of $\cos\theta_{B,\,D^{(*)}l}$ corresponds to a real cosine and is consequently located in a range of $\cos\theta_{B,\,D^{(*)}l} \in [-1, 1]$. Random combinations yield a distribution peaking at $\cos\theta_{B,\,D^{(*)}l} > 1$ due to a cut on the D momentum to suppress background from $e^+ e^- \to c\bar{c}$ processes. The resulting distribution can be observed in Figure 5.2a on the following page. The event selection corresponds to the one in Figure 5.2 on the next page. A precise and more comprehensive description of the semileptonic tagging algorithm can be found elsewhere [Kir12].

---

[3] A minimal efficiency of 90 % was chosen.

The setback compared to purely hadronic tagging is the lack of precise knowledge of the B_{tag} momentum. This prevents the knowledge of the momentum direction of the signal B momentum as well and deprives one of two very valuable variables helping with the purity of selections in analyses with hadronic tags which are defined as follows.

$$M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - \boldsymbol{p}_{B_{\text{tag}}}^2} \tag{5.2}$$

$$\Delta E = E_{\text{beam}} - E_{\text{Sig}} \tag{5.3}$$

Here, E_{beam} equals half the energy in the center of mass system of the two beams which is more precisely known than the zeroth component of the four-momentum of a reconstructed B meson. Albeit this loss in purity, semileptonically tagged samples benefit from the higher branching fractions of semileptonical B meson decays as compared to those hadronic B decays used in the hadronic tagging algorithm, as discussed in Section 2.3 on page 13. In addition, both samples are statistically independent, therefor adding information to the hadronically tagged result in any case.

(a) $\cos\theta_{B, D^{(*)}l}$ **(b)** $\mathcal{N}_{\text{NB, tag}}$
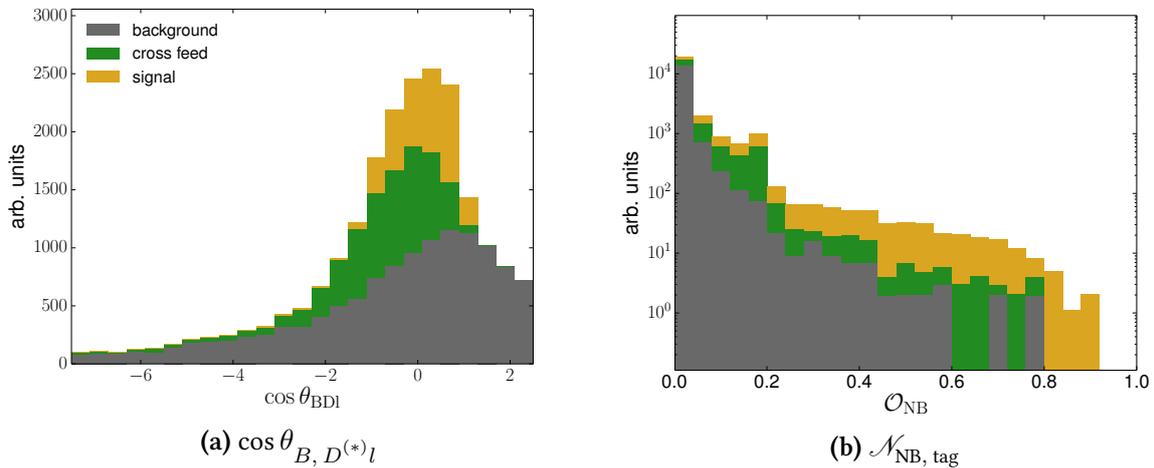
Figure 5.2.: Semileptonic tagging variables after the preselection according to the $B^+ \to K^+ \nu\bar{\nu}$ channel.

5.2.2. Signal Hadron Candidates

Signal hadron candidates are constructed from reconstructed tracks in the CDC and electromagnetic clusters in the ECL. The following criteria are applied.

5.2.2.1. Final State Particles

tracks A track is taken as a candidate if its closest approach to the interaction point IP is within 2 cm in r-direction and within 4 cm in z-direction of a cylindrical coordinate system with the IP as origin. The tracks are classified according to the criteria displayed in Table 5.1. $\mathcal{P}_{K/\pi}$ denotes the kaon likelihood given the kaon or pion hypothesis[4], $\mathcal{P}_l$ denotes the likelihood of the particle in question being a respective lepton.

Table 5.1.: Criteria for the classification of a track as a kaon or pion

	$\mathcal{P}_{K/\pi}$	$\mathcal{P}_{\pi/K}$	$\mathcal{P}_e$	$\mathcal{P}_\mu$
K	> 0.6	–	< 0.9	< 0.9
π	–	> 0.6	< 0.9	< 0.9

π^0 π^0 candidates are reconstructed from electromagnetic clusters in the ECL. Clusters are taken into account if their measured energy is above $100\,\mathrm{MeV}$ in the forward region, $50\,\mathrm{MeV}$ in the barrel region and $150\,\mathrm{MeV}$ in the backward region. Each π^0 candidate is required to have a reconstructed mass of $m_{\pi^0} \in [118\,\mathrm{MeV}, 150\,\mathrm{MeV}]$. The energy asymmetry of the two photons, the ratio between the innermost nine crystals and the maximal number of 25 crystals of a single cluster, $E_{3\times3}/E_{5\times5}$, for both photons has been considered as well. However, those variables do not add sufficient additional information, as can be seen in Figure 5.3 on the following page. The asymmetry is defined as $\left|E_{\mathrm{high}} - E_{\mathrm{low}}\right| / \left(E_{\mathrm{high}} + E_{\mathrm{low}}\right)$ where E_{low} denotes the π^0 daughter with the lower energy while E_{high} denotes the one with higher energy.

K_S^0 K_S^0 candidates are reconstructed from two oppositely charged π candidates fulfilling certain criteria, summarized in the goodKs function of the Belle analysis software frame work (BASF) yielding an excellent signal to background ratio in the final sample. The goodKs function has been widely used in Belle analyses, first described in [C+05]. Each candidate is forced to exhibit an invariant mass within $30\,\mathrm{MeV}$ of the nominal K_S^0 mass. The function applies cuts on the point of closest approach to the IP plane for the pions (d_r and d_z in Table 5.2 on page 37), the angle between the momentum of the K_S^0 candidate and the direction of the K_S^0 vertex ($\Delta\varphi$), and the flight distance of the K_S^0 up to the K_S^0 vertex, dependent on the momentum of the K_S^0 candidate.

tag B To identify (tag) the event as such, in addition to the signal particle, a second B is required, reconstructed in accordance to the algorithm described in Section 5.2.1 on page 32.

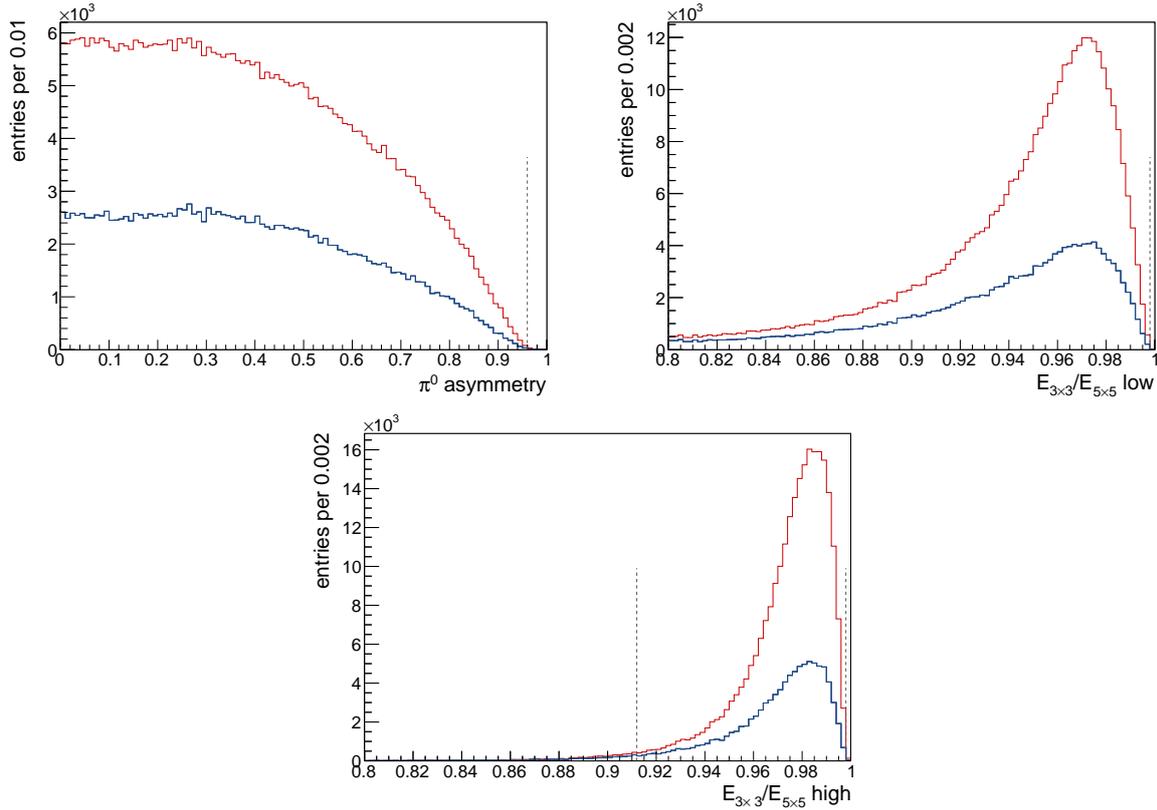[4] Analogously the pion likelihood for $\mathcal{P}_{\pi/K}$.

Figure 5.3.: π^0 related variables not used in the final selection, low and high denote the π^0 daughter with the higher or lower energy respectively. Signal is marked in red, background in blue.

5.2.2.2. Resonances

The K^* and ρ candidates are reconstructed from a kaon and a pion candidate (K^*) or a pair of pion candidates (ρ). K^{*+} candidates are reconstructed in the $K^{*+} \to K_S^0 \pi^+$ channel and the $K^{*+} \to K^+ \pi^0$ channel, K^{*0} candidates are reconstructed in the $K^{*0} \to K^- \pi^+$ channel, ρ^+ candidates in the $\rho^+ \to \pi^+ \pi^0$ channel, and ρ^0 candidates in the $\rho^0 \to \pi^+ \pi^-$ channel.

ρ candidates are required to exhibit an invariant mass within ± 250 MeV of the nominal ρ mass, K^* candidates are required to exhibit an invariant mass within ± 150 MeV of the nominal K^* mass. The vertex is fitted using a kinematic vertex fitting algorithm (kfitter). The vertex is constrained from the tracks and their respective error matrices by minimizing the χ^2 of the vertex fit.

5.2.3. Event

Since each correct signal event contains exactly three neutrinos, the knowledge of the full event as opposed to only reconstructing the signal side, in this case a single particle with unconstrained momentum, is crucial. Each candidate consists of one B_{tag} and the respective h_{sig} as described

Table 5.2.: goodKs selection cuts

	d_z [cm]	d_r [cm]	$\Delta\varphi$ [rad]	flight length [cm]
$p_{K_S^0} < 0.5\,\text{GeV}$	< 0.8	> 0.05	< 0.3	–
$0.5\,\text{GeV} > p_{K_S^0} < 1.5\,\text{GeV}$	< 1.8	> 0.03	< 0.1	> 0.08
$p_{K_S^0} > 1.5\,\text{GeV}$	< 2.4	> 0.02	< 0.03	> 0.22

earlier. A candidate is accepted if the event does not contain any additional particle, i.e. no other good track according to the criteria from Section 5.2.2.1 on page 35 is present and no additional π^0 candidate according to the respective criteria. In addition, events with a pair of raw tracks fulfilling the goodKs criteria are rejected as well. These three criteria are referred to as vetoes, charged track veto, π^0 veto, and K_S^0 veto, respectively. Since the principle of the measurement is the reconstruction of the complete event, the number of additional good particle candidates is restricted to 0. An example of a signal event is depicted in figure 5.4.
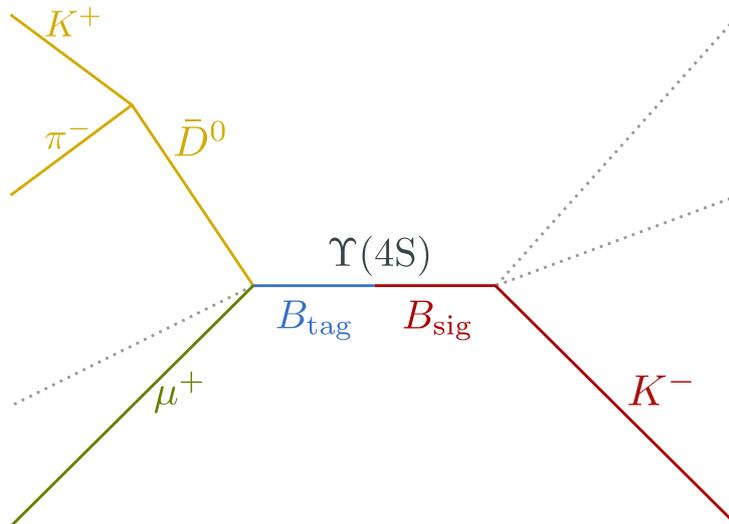
Figure 5.4.: Possible signal event topology of the $B^+ \to K^+ \nu\bar{\nu}$ channel. Neutrinos are dashed.

5.2.4. Precuts

After the online selection, various other cuts are done offline. Some of those are kinematically motivated, others practically.

$\boldsymbol{E}_{\text{ECL}}$ As mentioned in Section 5.2.3 on the preceding page, no event is allowed to contain additional good tracks, π^0 or K_S^0. This implies that no other signals should be measured anywhere in the detector, for instance not in the calorimeter. However, since the background from beam-beam interactions is always present and because particles undergo reactions such as hadronic scattering with the detector material or decay in flight

during their passage through the detector, the calorimeter nearly always measures additional hits even if the event in question is completely correctly reconstructed.

For this reasons the E_{ECL} variable has been defined. All energies of hits in the calorimeter not already assigned to a decay product of the $\Upsilon(4S)$ are summed up if they pass certain criteria. The algorithm distinguishes between isolated clusters and overlapping clusters. The common cuts can be found in Table 5.3.

For overlapping clusters additional requirements apply to the ratio of the 9 innermost crystals to all 25 crystals of the cluster $E_{3\times3}/E_{5\times5} < 0.94$ and $\mathrm{width}_{\mathrm{cluster}} > 30\,\mathrm{cm}$. The width of a cluster is defined as the root mean squared of the energy times distance of each assigned crystal normalized to the total energy of the cluster. The definition for E_{ECL} used in this work was introduced in [H$^+$13].

Table 5.3.: Photon energy cut, (E_{cut}), criteria as a function of region of the calorimeter.

ECL region	E_{cut} [GeV]
barrel	0.05
front	0.1
back	0.15

E_{miss}   The sum of all momenta of all daughter particles in the CMS of a decaying particle yields zero. The difference of the sum of the momenta of all reconstructed visible particles and zero is called missing momentum p_{miss} . The zeroth component is referred to as E_{miss}. On MC, a cut has been chosen which does not cut away any signal because the neutrino pair carries at least half the momentum of the B_{sig}: $E_{\mathrm{miss}} > 2.5\,\mathrm{GeV}$. As an example, the signal and background distribution for the $B^+ \to K^+ \nu\bar{\nu}$ channel are plotted in Figure 5.6 on page 43, with the precut value indicated by a dashed line. Figure 5.6 on page 43 also contains distributions of the next two variables, $p_{l_{\mathrm{tag}}}$ and p_{cms}.

$p_{l_{\mathrm{tag}}}$ For similar kinematic reasons, a cut on the momentum of the tag lepton in the $\Upsilon(4S)$ center of mass system of $p_{l_{\mathrm{tag}}} < 2.5\,\mathrm{GeV}$ has been chosen. The cut is tighter than the cut on p_{cms} of the signal candidate due to the massiveness of the D meson (see next paragraph).

p_{cms}   Due to conservation of momentum, the h_{sig} can not carry more than half the mass of the mother B in her rest frame. That means the maximal possible momentum of the hadron is reached when it carries half the B mass and its momentum is parallel to the momentum of the mother B. This maximal momentum in the center of mass system of the $\Upsilon(4S)$ yields a cut of $p_{\mathrm{cms}} < 2.96\,\mathrm{GeV}$.

Additionally, the resulting momentum of the h_{sig} candidate is on average more energetic for true signal candidates than for particles from other B decay chains, e.g. those of D daughters. Furthermore, there are large model uncertainties connected to low momentum hadrons in backgrounds. For those reasons, a lower cut on p_{cms} of $p_{\mathrm{cms}} > 0.5\,\mathrm{GeV}$ is applied in addition to the upper cut.

$\mathcal{N}_{\text{NB, tag}}$ The semileptonic tagging gives a final output to classify reconstructed B mesons. A cut on the final net output is chosen to reduce the background from misreconstructed B_{tag} s. Previous results [A^{+}14b] as well as this work have shown that badly reconstructed B candidates are not well modeled in the Monte Carlo. A cut on a medium level has been chosen to not loose too much signal in advance and to avoid falsely reconstructed tag side B mesons: $\mathcal{N}_{\text{NB, tag}} > 0.005$. The effect of that cut on the signal efficiency and background rejection rate is plotted in Figure 5.7 on page 43 in form of a ROC curve and as normalized distributions in $\mathcal{N}_{\text{NB, tag}}$, analogous to the aforementioned plots in Figure 5.6 on page 43.

raw tracks There are always reconstructed tracks left which do not fulfill the criteria displayed in Section 5.2.2.1 on page 35. Those tracks are referred to as *raw tracks* and might either be a hint for the event being misreconstructed and therefore a hint for the presence of additional particles or have reasons not connected to the interpretation of the event as a whole. Amongst such are for instance fake tracks, curling tracks which are found twice or background from the beams. The effect of cutting on the number of those tracks has been studied. It has been decided to allow for one raw track since the loss in efficiency would be too severe otherwise. This is illustrated in Figure 5.5 on page 42.

5.2.5. *D* veto

Since the signal as well as the tagside contain neutrinos, the mass of neither B meson is reconstructable. Hence, the tag lepton candidate can be produced by a variety of background processes as well. In case a pion from a $D \to K\pi$ decay is misidentified as a lepton, the accompanying kaon is likely to be identified as a signal candidate. To exclude such events, the invariant mass of the $K^{(*)} - l$ system is reconstructed. Decays of the described type can be observed in a significant number in the $B^{+} \to K^{+}\nu\bar{\nu}$ and $B^{0} \to K^{*0}\nu\bar{\nu}$ channels and are thus removed by excluding candidates with $1.85\,\text{GeV} < M_{h,l} < 1.87\,\text{GeV}$. The distributions are displayed in Figure 5.8 on page 44 where the cutted area is shaded for each channel, as well for the $B^{0} \to K_{\text{S}}^{0}\nu\bar{\nu}$, $B^{+} \to K^{*+}\left(\to K^{+}\pi^{0}\right)\nu\bar{\nu}$ and $B^{+} \to K^{*+}\left(\to K_{\text{S}}^{0}\pi^{+}\right)\nu\bar{\nu}$ channels where the cut is not applied because of the low statistics. The effect on the signal efficiency is not very large, the loss for the $B^{+} \to K^{+}\nu\bar{\nu}$ mode amounts to $0.9\,\%$ and the one for the $B^{0} \to K^{*0}\nu\bar{\nu}$ to $1.0\,\%$.

5.2.6. Best Candidate Selection

After the application of the cuts described in the previous section, about $14\,\%$ (generic MC) to $20\,\%$ (signal MC) of all events still contain multiple candidates. There are in principle two reasons why this can occur. First, the exchange of particles between signal and tag side, i.e. a signal particle is used to form a B_{tag} candidate while a B_{tag} decay product forms a signal candidate. This happens rather easily because the h_{sig} is not constrained in its momentum with the exception of an upper bound. Second, particles might contribute to both a signal candidate and a tag candidate. For

instance, a pion from ρ or K^* decays might be used to form a tag candidate while the other pion (kaon) forms a $B^+ \to K^+ \nu \bar{\nu}$ or $B^+ \to \pi^+ \nu \bar{\nu}$ candidate.

Since the analysis does not allow for any additional particles besides the B_{tag} and the signal candidate, those are the only reasons for multiple candidates in simulated signal decays. Consequently, the one with the highest $\mathcal{N}_{\text{NB, tag}}$ determines which candidate is chosen, being the most likely B_{tag} candidate by construction. This approach has been chosen in the past as well, e.g. in [A$^+$14b].

Table 5.4 on page 45 shows the resulting efficiency for each reconstructed channel. The efficiency of the Best Candidate Selection (BCS) is defined as the number of events where the simulated signal decay has been correctly identified as such over the overall number of reconstructed events which contain multiple candidates involving a correctly reconstructed candidate. Overall, the performance of the BCS is very satisfying. The low value, in comparison with all other channels, of the BCS efficiency in the $B^+ \to \rho^+ \nu \bar{\nu}$ channel can be explained by the relative high probability to form random π^0 candidates in combination with the larger number of $\pi^\pm$ in a typical $\Upsilon(4S)$ event as compared to the average number of $K^\pm$.

5.2.7. Optimized Cuts

As the cuts described in Section 5.2.2.2 on page 36 have been chosen to be signal inclusive, they still contain a large amount of background. To get a more sensible selection, the cuts on the invariant masses of K^*, ρ, K_S^0 and π^0 candidates have been optimized via a figure of merit (FoM) defined as follows.

$$\text{FoM} = \frac{S}{\sqrt{S + B}} \tag{5.4}$$

Here, S denotes the number of chosen real particle candidates after the cut and B the number of random combinations after the cut. The target is whether or not the candidate is a real one, not if it is a real $B \to h\nu\bar{\nu}$ candidate. The results are shown in Figure 5.9 on page 46. The optimization procedure relies on generic Monte Carlo only. The cuts on $\mathscr{N}_{\text{NB, tag}}$ and p_{cms} have been loosened to get a sufficient amount of correct candidates.
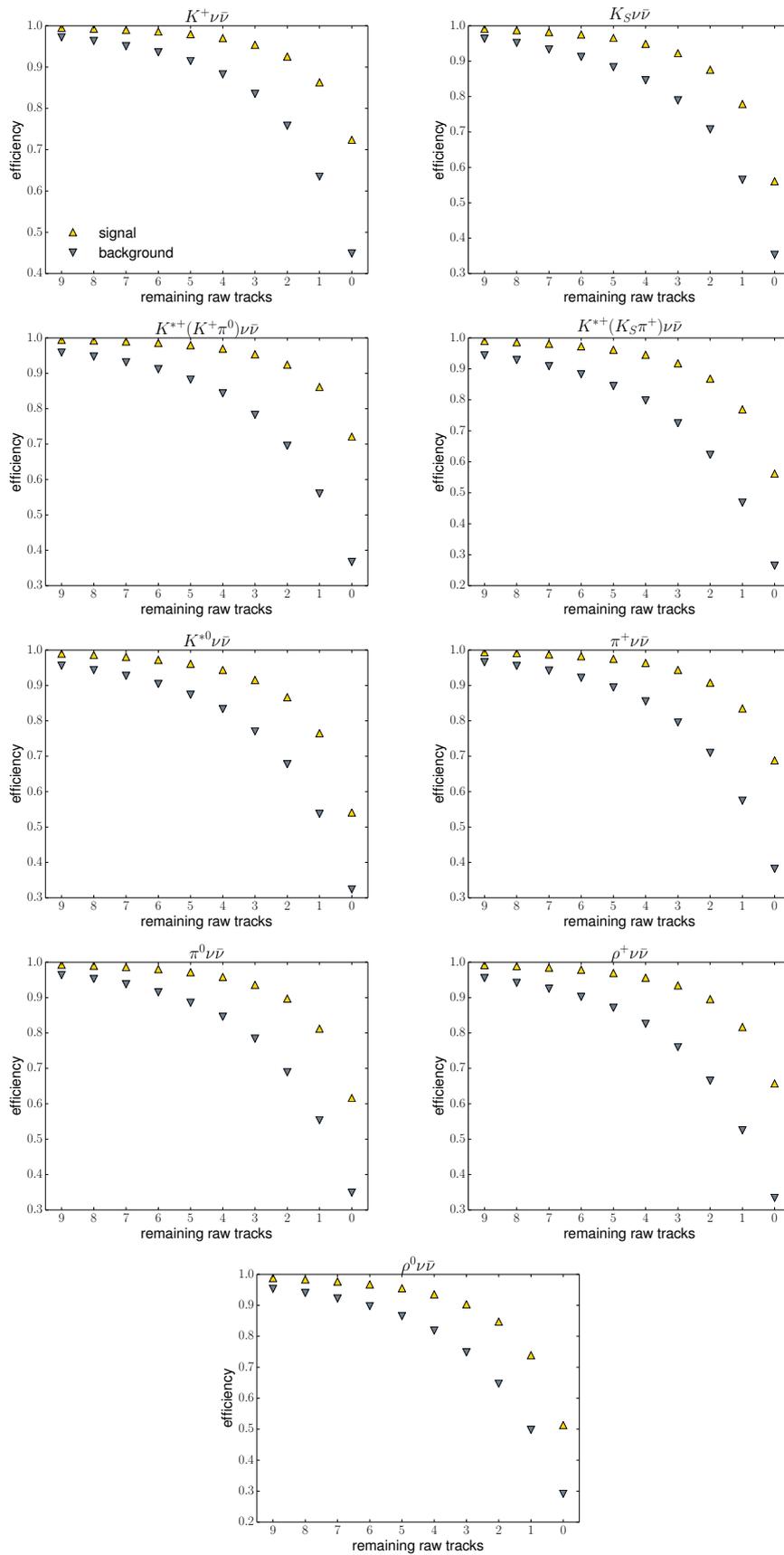
Figure 5.5.: Efficiency dependent on the number of remaining raw tracks determined on one stream of generic MC.
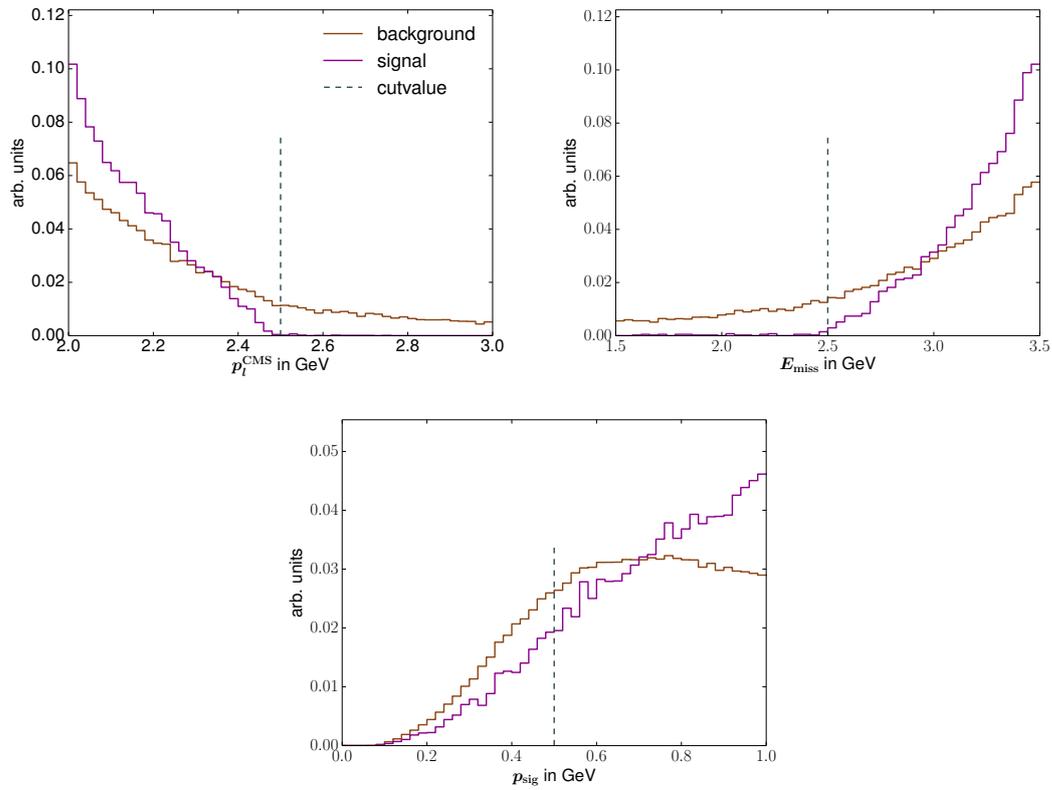
Figure 5.6.: Precuts for the $B^+ \to K^+ \nu \bar{\nu}$ channel as a showcase. Signal and background distributions have been normalized to one. The cut has been marked by a dashed line.
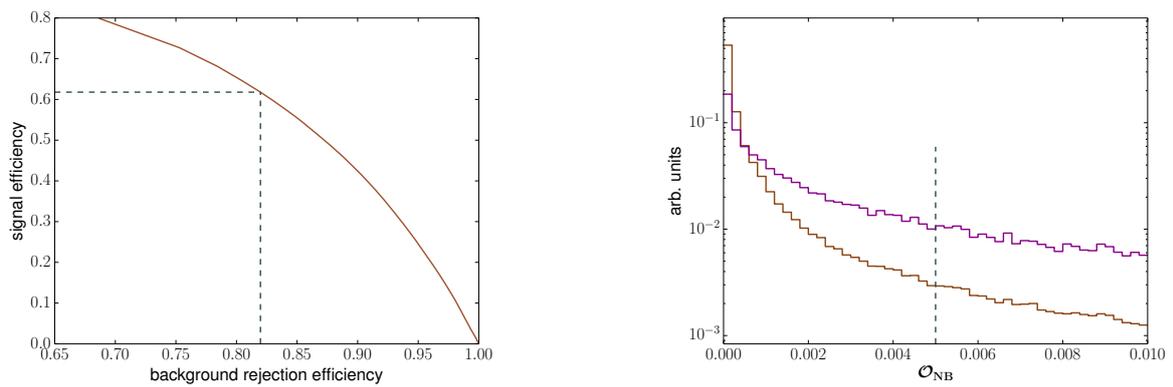
Figure 5.7.: The ROC-curve and the normalized signal and background distribution of $\mathcal{N}_{\mathrm{NB,\,tag}}$ for the $B^+ \to K^+ \nu \bar{\nu}$ channel. The cut is marked by the dashed line.

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K^{*0} \nu\bar{\nu}$

(c) $B^0 \to K_S^0 \nu\bar{\nu}$

(d) $B^+ \to K^{*+} \left(\to K^+ \pi^0 \right) \nu\bar{\nu}$

(e) $B^+ \to K^{*+} \left(\to K_S^0 \pi^+ \right) \nu\bar{\nu}$
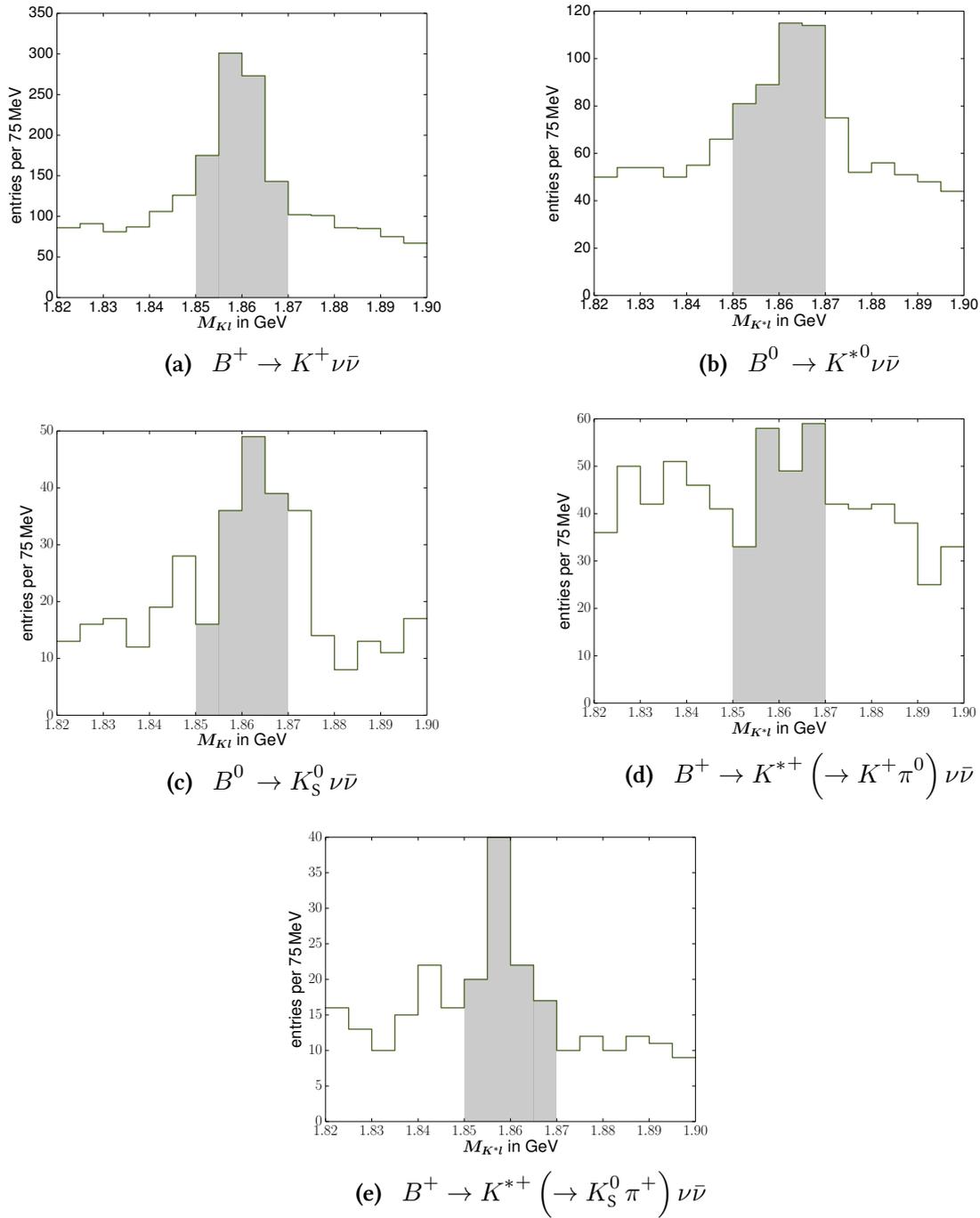
Figure 5.8.: Invariant mass distributions of $K^{(*)} - l$ combinations taken from six streams of generic Monte Carlo. The region cutted away in the $B^+ \to K^+ \nu\bar{\nu}$ and $B^0 \to K^{*0} \nu\bar{\nu}$ channel are shaded.

Table 5.4.: Efficiency of the best candidate selection

channel	ε in %
$B^+ \to K^+ \nu\bar{\nu}$	81.5
$B^0 \to K_S^0 \nu\bar{\nu}$	91.6
$B^+ \to K^{*+} \left(\to K^+ \pi^0 \right) \nu\bar{\nu}$	82.5
$B^+ \to K^{*+} \left(\to K_S^0 \pi^+ \right) \nu\bar{\nu}$	83.6
$B^0 \to K^{*0} \nu\bar{\nu}$	91.4
$B^+ \to \pi^+ \nu\bar{\nu}$	81.5
$B^0 \to \pi^0 \nu\bar{\nu}$	90.9
$B^+ \to \rho^+ \nu\bar{\nu}$	65.4
$B^0 \to \rho^0 \nu\bar{\nu}$	87.8

(a) M_{π^0}

(b) $M_{K_S^0}$

(c) $M_{K^{*+}}$

(d) $M_{K^{*+}}$

(e) $M_{K^{*0}}$

(f) M_{ρ^+}

(g) M_{ρ^0}
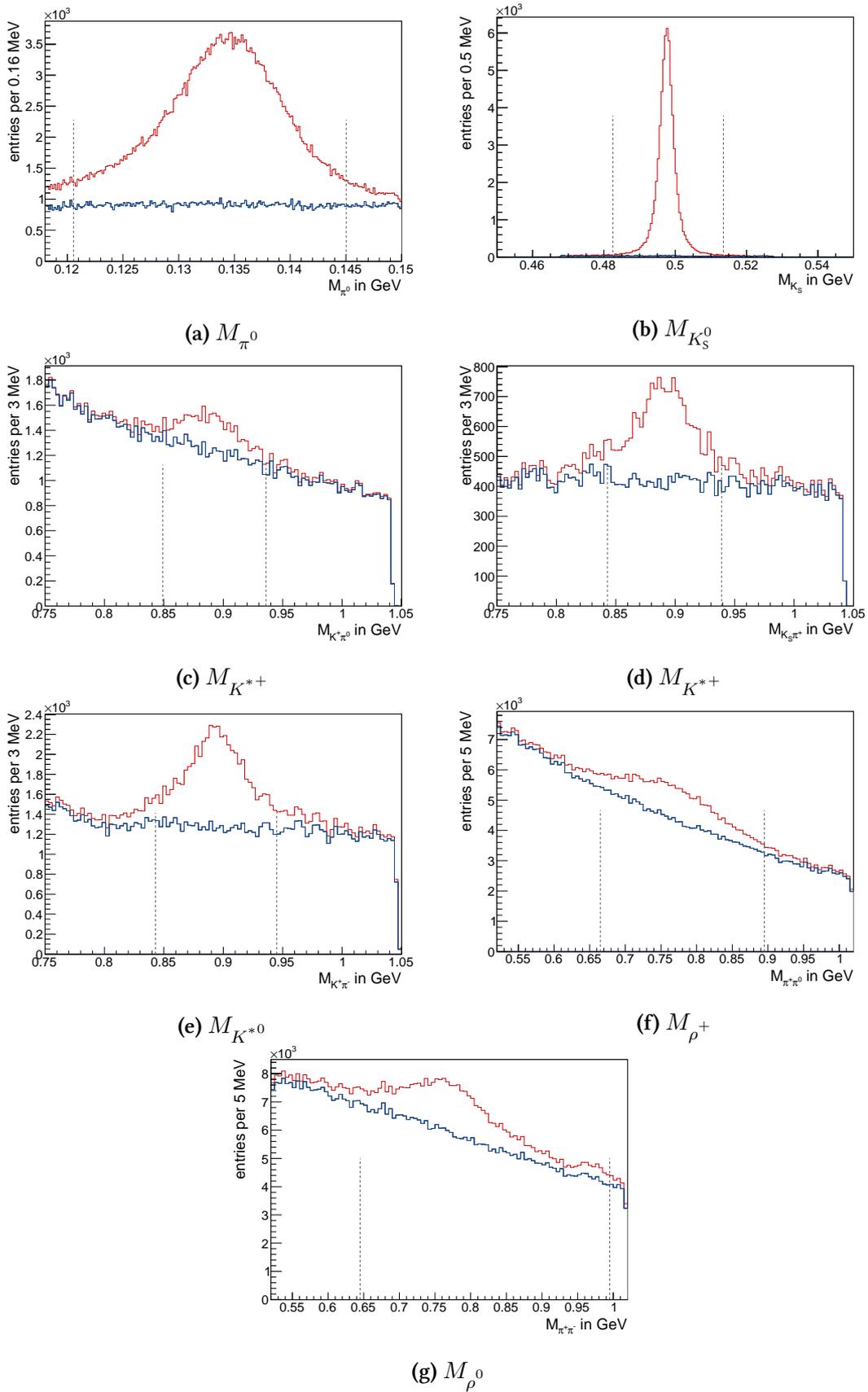
Figure 5.9.: Cuts optimized on six streams of generic Monte Carlo. Signal is red while background is blue. The region within the dashed lines is chosen.

6 Chapter

Weigths and Corrections

6.1. Rare Branching Fraction Correction

Since the MC samples provided by the collaboration have been created in 2008, the contained channels had to be updated to their respective current world average values in cases where they did change. This has not been done for all channels in the rare MC sample but only for those which turned out to contribute non-negligibly to the background remaining after the final selection. Section 7.4 on page 64 includes a discussion of those processes. A full list of all updated branching fractions and their respective values can be found in appendix B. All in all 54 channels have been updated over all eight signal channels.

6.2. Klong Veto

Neutral particles are particularly problematic for analyses which depend on the precise knowledge of the kinematics of the event. Of special concern are K_L^0 mesons. As not electromagnetically interacting particles, they are not detectable by either the tracking system or the electromagnetic calorimeter. Moreover, the interaction probability of those particles with the detector material in the KLM is very low. Thus, they often leave the detector undetected and mimic missing momentum. Even if a K_L^0 undergoes detectable reactions with the material inside the KLM, the experimental challenge is intricate. Residual hadronic reactions like those responsible for K_L^0 signals within the KLM are hard to understand theoretically, which is particularly true for low momentum K_L^0. The modeled data hence tend to not reflect the real data, or more precisely the efficiency in data differs from the efficiency in simulations.

To solve this problem, a reweighting procedure of MC has been introduced [H$^+$13] utilizing $\phi \to K_S^0 K_L^0$ with ϕs from decays from $D^0 \to \phi K_S^0$ decays. The D^0 candidates themselves were tagged by slow pions from $D^* \to D^0 \pi$ decays. Each simulated event is assigned a certain weight which depends on the simulated K_L^0 momentum and the number of K_L^0 candidates in the KLM. A KLM

cluster is taken as a K_L^0 candidate if it is not related to any extrapolated charged track within an angle of $15°$ and if it either consists of at least two hit layers or, in the case where only one layer has been hit, is connected to an ECL cluster. If there are no simulated KLM clusters at all, the event is assigned a weight of one, if there are fake clusters, the event gets a weight of 0. In case of only true[1] K_L^0 candidates, the efficiency correction for each single candidate is multiplied to form the event weight.

Data events receive a weight of either zero or one depending on whether a K_L^0 candidate was reconstructed or not.

6.3. Tag Efficiency Correction

To check how well the simulations resemble the real data regarding the efficiency of the tagging algorithm, control channels with well known branching fractions and established data-to-MC consistencies were reconstructed and combined with B_{tag} candidates to form $\Upsilon(4S)$ candidates. The control channels consist only of charged final state particles; the requirements for the tracks used to form them are the same as the ones described in Section 5.2.2.1 on page 35. Precuts are applied if their respective application is possible and a best candidate selection is executed following the same criterion as for the signal channels. Reconstructed are the following two channels.

$B^+ \to \bar{D}^0\pi^+$ D^0 candidates are reconstructed in the $K^-\pi^+$ channel. The invariant mass of the D candidate is required to be within a mass window of $30\,\text{MeV}$ around its nominal mass. The mass of the B is required to lie within a window of $100\,\text{MeV}$ around the nominal mass. Further requirements are: $|\Delta E| < 0.1\,\text{GeV}$ and $M_{\text{bc}} > 5.27\,\text{GeV}$.

$B^0 \to D^-\pi^+$ The only difference to the previous case is the charge and decay channel of the D meson. D^- candidates are reconstructed in the $K^+\pi^-\pi^-$ channel. The D mass is required to be within a mass window of $30\,\text{MeV}$ of the nominal mass. Cuts concerning $M_B, \Delta E$, and M_{bc} are the same as those in the $B^+ \to \bar{D}^0\pi^+$ channel.

These hadronic channels have been chosen because of their very high purity, despite their relatively low branching fraction of 2.68×10^{-3} for the decay $B^+ \to \bar{D}^0\pi^+$ and 4.81×10^{-3} for the decay $B^0 \to D^-\pi^+$, respectively.[2]

The ratio of the $B^+ \to \bar{D}^0\pi^+$ ($B^0 \to D^-\pi^+$) branching fraction from this tagged sample and the current world average are studied in bins of $\mathcal{N}_{\text{NB, tag}}$ such that each bin contains exactly the same number of events on data for each mode, totaling four bins in the case of $B^+ \to \bar{D}^0\pi^+$ decays and three bins in the case of $B^0 \to D^-\pi^+$ decays. Since the selection is background free, a simple counting experiment is carried through in the selected signal region to calculate the

---

[1] Clusters are called true if they are reconstructed within $\pm15°$ in both azimuthal and polar angle of the direction of the momentum of a simulated K_L^0.

[2] values from [O$^+$14]

(a) $B^+ \to \bar{D}^0 \pi^+$ **(b)** $B^0 \to D^- \pi^+$
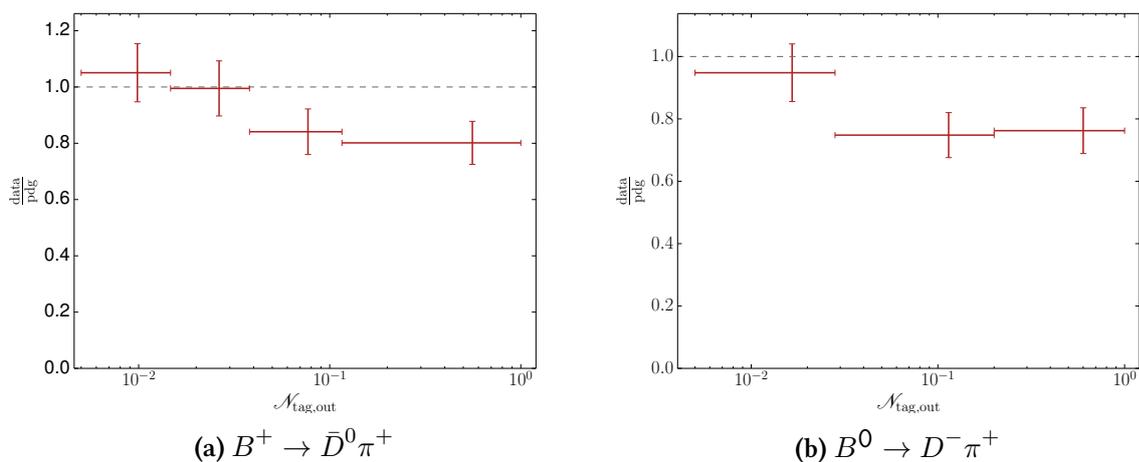
Figure 6.1.: Tag correction factors in bins of $\mathcal{N}_{\text{NB, tag}}$

respective $B \to D\pi$ signal yield. The efficiency is determined from a set of 10^7 simulated signal events for each sample. The results are plotted in Figure 6.1. It is apparent that while the branching fraction determined on the tagged sample is in good agreement with the world average value from the pdg for lower values of $\mathcal{N}_{\text{NB, tag}}$ for both neutral and charged B_{tag}, the measured branching fractions differs more from the world average values for higher values of $\mathcal{N}_{\text{NB, tag}}$. In the case of charged B_{tag} this behavior has been noted previously in a different approach to the efficiency correction [A$^+$14b]. The reason is an overestimation of the resolution of reconstructed intermediate particles on MC as compared to data. This leads to an overestimation of the $\mathcal{N}_{\text{NB, tag}}$ value on MC. The correction is applied as a weight to all correctly reconstructed tag candidates on signal and tag side.

6.4. PID Efficiency Correction

The efficiency difference of the particle identification, see Section 3.3 on page 21, between data and MC has been studied elsewhere [Nis05]. A sample of $D^{*+} \to D^0 \left(\to K^- \pi^+ \right) \pi^+$ decays has been reconstructed with different $\mathcal{P}_{K/\pi}$ cuts applied. The reconstructed $D^* - D$ mass difference distribution, $\Delta m_{D^*, D} = m_{D^*} - m_D$, has subsequently been fitted in 32 bins of the hadron momentum and 12 bins of the polar angle of the track with respect to the beam axis, resulting in 384 bins in total. The ratio of the efficiency on data and MC can be taken as a weight on the signal sample in the respective bin for the calculation of the efficiency on data, the error can be assigned as a systematic uncertainty to the final result. This study has been performed with nine $\mathcal{P}_{K/\pi}$ selection cuts applied, ranging from 0.1 to 0.9.

We find that for our sample of rather high energetic reconstructed signal hadron candidates, as opposed to hadron candidates from D decays, the kaon identification efficiency is modeled in agreement with the data while the pion identification efficiency is overestimated on MC. Since

the obtained correction factors are in good agreement with each other for each run, we average them over all runs and the respective errors to one global correction factor per channel.
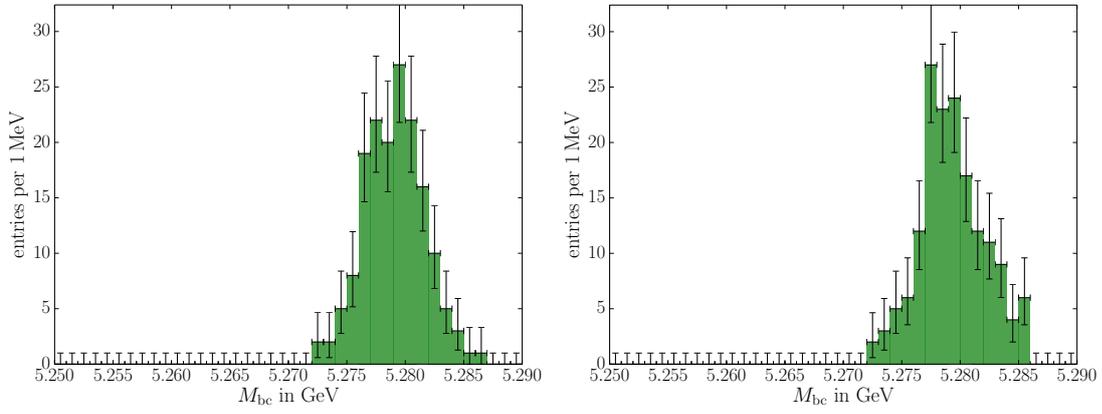
(a) $0.005 < \mathscr{N}_{\text{NB, tag}} < 0.01469$

(b) $0.01469 < \mathscr{N}_{\text{NB, tag}} < 0.03803$
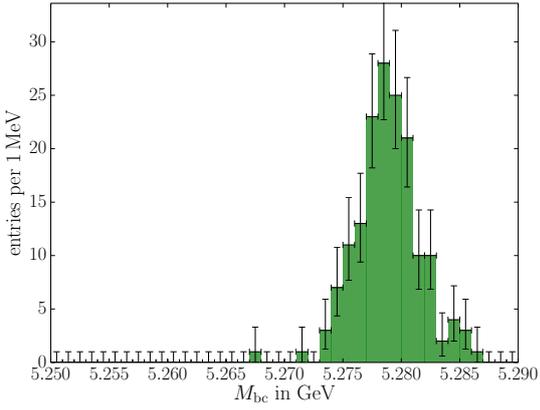
(c) $0.03803 < \mathscr{N}_{\text{NB, tag}} < 0.11582$
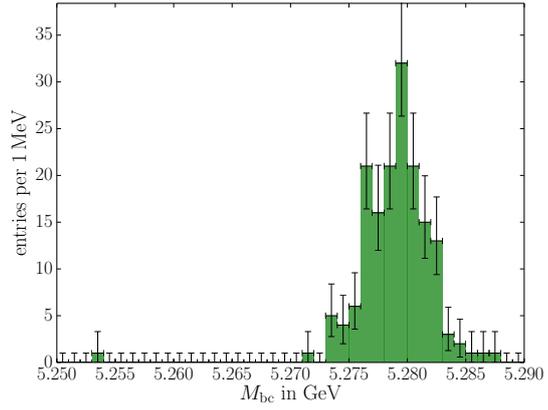
(d) $0.11582 < \mathscr{N}_{\text{NB, tag}} < 1$

Figure 6.2.: M_{bc} distributions for the corrections regarding the B^+ case.
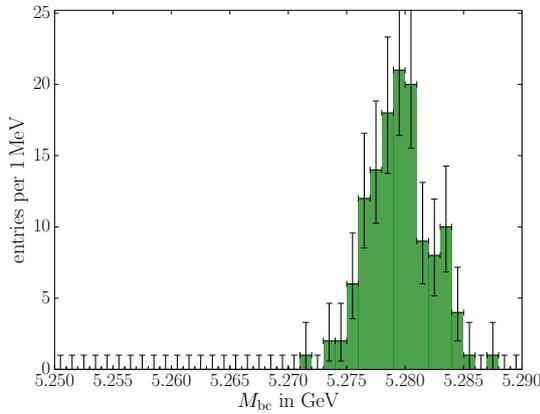
(a) $0.005 < \mathscr{N}_{\text{NB, tag}} < 0.02813$
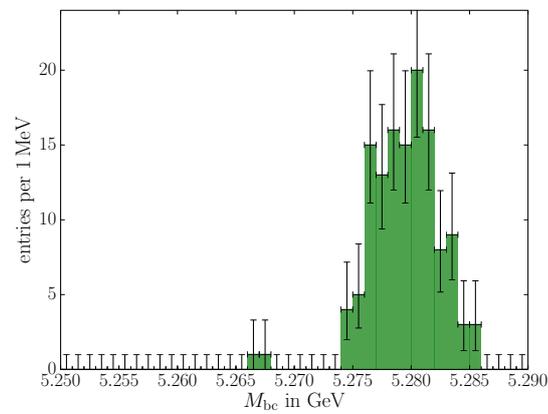
(b) $0.02813 < \mathscr{N}_{\text{NB, tag}} < 0.20016$
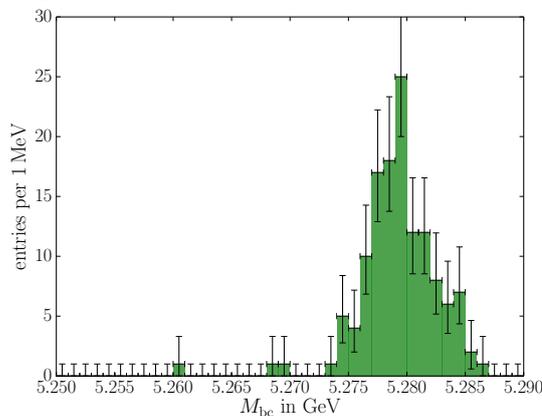
(c) $0.20016 < \mathscr{N}_{\text{NB, tag}} < 1$

Figure 6.3.: M_{bc} distributions for the corrections regarding the B^0 case.

7 Chapter

Separation

7.1. Continuum Suppression

In e^+e^- collisions at a center of mass energy of $10.58\,\text{GeV}$, $\Upsilon(4S)$ resonances are produced in only $24\,\%$ of all hadronic events[1]. In the other $76\,\%$ of all hadronic events, pairs of lighter quarks, i.e. u,d,s,c are produced. The hadrons formed out of the final state on quark level are much lighter than B mesons. This consequently yields higher momenta in the e^+e^- rest frame due to the conservation of energy because the combined mass of two B mesons is only $20\,\text{MeV}$ lower than the mass of the $\Upsilon(4S)$. This relation is visualized in Figure 7.1 on the next page via a direct comparison of a simulated signal and continuum background event.

This kinematical distinction has been utilized at all B factories to suppress the background originating from processes involving no b quarks. To obtain an optimal separation of B and continuum events, multivariate tools, first developed in [P$^+$13] and now used widely in the Belle collaboration, are employed. The multivariate classificator utilizes many shape related variables defined previously ([B$^+$14]). B events lead to a spherical event shape in the e^+e^- rest frame as opposed to continuum events which lead to a more jet like structure. Various shape variables are combined using a neural net. The training was done with the continuum background from six streams of generic Monte Carlo and a numerically equal amount of signal Monte Carlo for each reconstructed channel. A set of ROC-curves for the $B^+ \to K^+\nu\bar{\nu}$ channel with various combinations of the input variables described below are displayed Figure 7.2 on page 59. Corresponding figures for all channels can be looked up in the Appendix, Figure C.1 on page 132. The output of that neural net is hitherto referred to as $\mathscr{N}_{\text{NB, CS}}$.

A short description is now given for each variable included in the neural net.

---

[1]The term *hadronic event* here refers to a process where two quarks are produced, $e^+e^- \to q\bar{q}$.
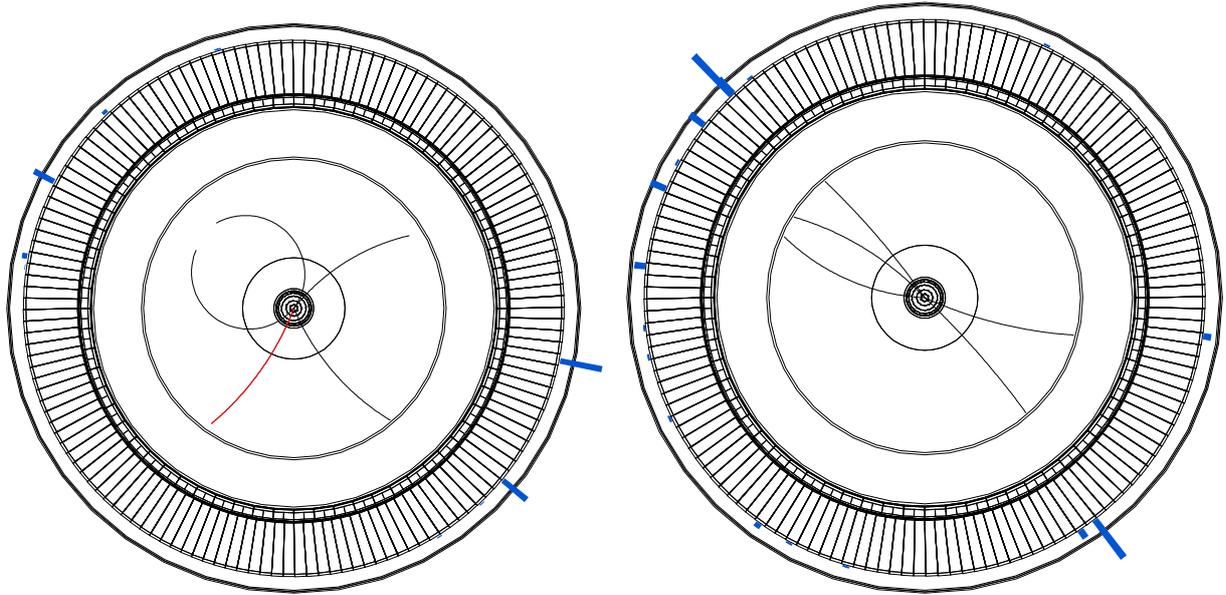
Figure 7.1.: Simulated $B^+ \to K^+ \nu \bar{\nu}$ signal event (left) and continuum event (right) in a $x - y$ projection of the inner detector up to the ECL. The simulated Kaon from the signal decay is marked in red. The high p_T products of the continuum event are clearly visible.

R_2 Fox Wolfram moments are a way to describe the topology of the event via Legendre Polynomials and are defined as follows.

$$H_k = \sum_{i,j} \frac{|\mathbf{p}_i| \, |\mathbf{p}_j| \, P_k \left(\cos \theta_{ij} \right)}{E_\mathrm{tot}^2} \tag{7.1}$$

$$R_k = \frac{H_k}{H_0} \tag{7.2}$$

The H_k are the Fox-Wolfram moments, $\mathbf{p}_i$ the three momentum of the i-th particle, P_k the k-th Legendre Polynomial, θ_{ij} the angle between the i-th and j-th particle, and E_tot the total visible energy.

The variable R_2 from the above definition Equation (7.2) proved especially useful to suppress continuum events.

KSFW moments Kakuno Super Fox-Wolfram Moments modify the normal Fox-Wolfram moments. KSFW moments split the event into signal and a rest of the event component. The sum in Equation (7.1) loops over either the signal side s or the rest of the event r such that the R_k can be combined out of both and form three different combinations $R_k^{ss}, R_k^{sr}, R_k^{rr}$ out of which several combinations are taken which proved to be helpful to separate signal from background in the past and only loosely correlated to other important kinematical variables.

$\cos \theta_\mathrm{B}$ $\Upsilon(4\mathrm{S}) \to B\bar{B}$ decays are a decay of a vector particle, the $\Upsilon(4\mathrm{S})$, into a pair of pseudo-scalar particles, the B mesons. For such a decay the angle of a pseudoscalar daughter is distributed proportional to $1 - \cos^2 \theta_\mathrm{B}$. In contrast, random combinations of continuum particles are distributed flatly in $\cos \theta_\mathrm{B}$.

Thrust Angle The thrust axis is defined as the direction maximizing the projection of final state particles momenta onto it.

$$\mathbf{T} = max \left(\frac{\sum\limits_{i} |\langle \mathbf{p}_i | \mathbf{T} \rangle|}{\sum\limits_{i} |\mathbf{p}_i|} \right) \tag{7.3}$$

For more jet-like continuum events, the cosine of the angle of the thrust axes with the beam axis tends to higher values while $B\bar{B}$ events are distributed flatly in this variable.

$\mathbf{\Delta z}$ Because of the comparably long lifetime of B mesons as compared to many lighter mesons, they also have a longer average flight distance due to the boost between the $\Upsilon(4S)$ center of mass system and the laboratory system. The signal is subtracted from the event and all remaining tracks, from the B_{tag}, fitted to one vertex. Δz is in the context of a fully reconstructed analysis therefor defined as $\Delta z = z_{\mathrm{sig}} - z_{\mathrm{tag}}$.

Cleo Cones To further suppress continuum background, the CLEO collaboration introduced variables based on the sum of the absolute values of the momenta of all particles within angular sectors around the thrust axis in intervals of $10°$, amounting to an overall number of 9 cones [A⁺96]. The cone in the direction of the thrust axis is merged with the respective cone in the opposite direction.

7.2. Network Training

To optimally separate signal from background, a neural network utilizing the Neuro Bayes algorithm [FK06] has been trained for each channel individually. The training was performed on the full MC sample consisting of ten streams of generic MC, six streams of continuum MC, rare MC weighted to the corresponding number of events, and signal MC with the signal to background ratio fixed to an arbitrary value of 0.2.

7.2.1. Variables

A large variety of input variables have been studied. Their respective importance varies from channel to channel, although the most significant variables are important for each single channel. Relevant input variables are described below, i.e. those who enter at least one net used for the final selection.

$\mathscr{N}_{\mathrm{NB,\, CS}}$ the output of the continuum suppression described in Section 7.1 on page 53 is taken as an input.

p_{cms} The momentum of the h_{sig} in the center of mass system of the $\Upsilon(4S)$.

$\cos\theta_{B,\, D^{(*)}l}$ The cosine of the angle between the B_{tag} and the $D^{(*)}l$ system as described in Section 5.2.1 on page 32.

$\mathscr{N}_{\text{NB, tag}}$ The net output of the semileptonic tagging algorithm Section 5.2.1 on page 32.

$\mathscr{N}_{\text{NB, tag}}^{\text{rank}}$ The rank the B_{tag} candidate had before all vetoes and plots. B_{tag} candidates are more likely to be random combinations if more likely candidates existed in the first place.

$p_{l_{\text{tag}}}$ The momentum of the lepton used to reconstruct the B_{tag} in the rest frame of the $\Upsilon(4S)$.

$\cos \theta_{\text{miss}}$ The cosine between the missing momentum direction and the beam axis.

$\cos \theta_{\text{thrust}}$ The thrust angle of the event is the angle between the thrust axis and the beam axis. The thrust axis is the spatial direction which maximizes the projection of all particles momenta onto it.

$M_{D_{\text{tag}}}$ The reconstructed invariant mass of the D meson used in the construction of the B_{tag} $\left(D_{\text{tag}}\right)$.

$p_{D_{\text{tag}}}$ The momentum of the D_{tag} in the center of mass frame of the $\Upsilon(4S)$.

$N_{\text{remaining}}^{\text{rawtracks}}$ As mentioned in Section 5.2.4 on page 37, we allow for one remaining raw track. This variable is one if there is one such track left, zero otherwise.

$\Delta \chi^2$ A vertex fit is executed on the whole event and on the tag side particles only. The difference of the two resulting χ^2 of the fits is referred to as $\Delta \chi^2$.

$\cos \theta_{h, l}$ The cosine of the angle between the tag lepton and the h_{sig}.

$|p_{\text{miss}}|$ The absolute value of the missing momentum in the rest frame of the $\Upsilon(4S)$.

$M_{\text{sig}}^{\text{min}}$ The minimum of the invariant masses out of all possible combinations of the signal candidate and reconstructed good tracks. If all such combinations yield high values, the origin from different particles is more likely.

$\cos \theta_{D^{(*)}l}$ The angle between the $D^{(*)}l$ system and the beam axis in the $\Upsilon(4S)$ rest frame.

$\cos \theta_{\text{tag, miss}}^{\text{lab}}$ The angle between the $D^{(*)}l$ system and the p_{miss} direction in the laboratory system.

$M_{h, l}$ The invariant mass of the signal hadron and tag lepton system. If the signal candidate comes out of the same decay as the reconstructed signal candidate, the mass tends to lower values.

$\cos \theta_{h_{\text{sig}}}$ The cosine of the angle between the h_{sig} and the beam axis in the center of mass system of the $\Upsilon(4S)$.

$\cos \theta_{\text{sig, tag}}$ The cosine of the angle between the h_{sig} and the $D^{(*)}l$ system in the rest frame of the $\Upsilon(4S)$.

$\mathcal{P}_{K/P}$ The likelihood of being a proton, given a kaon or proton, of the h_{sig}, not defined for the $B^0 \rightarrow K_S^0 \nu \bar{\nu}$, $B^0 \rightarrow \pi^0 \nu \bar{\nu}$ and $B^0 \rightarrow \rho^0 \nu \bar{\nu}$ channels. In the case of the $B^0 \rightarrow K^{*0} \nu \bar{\nu}$ channel, the value of the charged kaon is chosen.

$\mathcal{P}_{K/\pi}$ The likelihood of being a kaon, given a kaon or pion hypothesis, of the h_{sig}. In case of K^{*0} candidates, the input is taken from the kaon candidate.

Δr_{sig} The distance of the point of closest approach of the track to the IP. In cases in which the signal side is reconstructed from more than one track, the Δr_{sig} of the $K^{\pm}$ candidate $\left(B^+ \to K^{*+} \left(\to K^+ \pi^0 \right) \nu\bar{\nu} \text{ channel} \right)$ or of a random pion is chosen $\left(B^0 \to \rho^0 \nu\bar{\nu} \text{ channel} \right)$.

M_{res} The invariant mass of the resonance at the signal side if the h_{sig} is reconstructed from two daughter particles.

p_{π^0} The momentum of the π^0 from the signal side for the $B^+ \to K^{*+} \left(\to K^+ \pi^0 \right) \nu\bar{\nu}$ and $B^+ \to \rho^+ \nu\bar{\nu}$ channels.

$\cos\theta_{K\pi}$ The cosine of the angle between the daughter particles in the rest frame of the mother particle. Defined for h_{sig} channels which decay to a π^0 and a charged particle, reduces background from arbitrary $\gamma\gamma$ combinations.

7.2.2. Training Procedure

The adopted algorithm is described in Section 4.1.1 on page 25. As mentioned in that section, the preprocessing in Neuro Bayes involves fitting splines to the signal to background ratio in each variable in order to get rid of mere statistical fluctuations. If the signal over background distribution in turn exhibits discontinuous parts, or in other words real features of the respective variable, the fit can not adapt to that. If this occurs and some physical property would be cut out, the variable is split up in several parts, yielding two or more input variables covering one region each. This procedure is applied to $\cos\theta_{B, D^{(*)}l}$, $\mathcal{N}_{\text{NB, tag}}$, $\Delta\chi^2$ and PID variables.

The number of nodes in the hidden layer varies from channel to channel, depending on the available amount of background MC, such that the number of free parameters is always much smaller than the number of training events. Input variables with an added significance[2] of less than three are discarded for the final net. The entropy is taken as the loss function and a bayesian regularization procedure is chosen with three regularization parameters for the weight decay. One for the weights related to the bias node, one for all other weights from the input layer, and one for all weights connected to the output node.

The optimization is done with the BFGS algorithm [BLNZ95]. The results of all trainings can be looked up in Appendix D where all variables are ranked according to the respective separation power for each channel separately. A comparison of the distribution of the final network output ($\mathcal{N}_{\text{sel}}$) for background and scaled signal, exemplified for the $B^+ \to K^+ \nu\bar{\nu}$ channel, can be seen in Figure 7.3 on page 59.

---

[2] As defined in Section 4.1.1 on page 25.

7.3. Cut Optimization

One problem of searches for yet unmeasured processes is that the rate is unknown, this implies that the figure of merit defined in Equation (5.4) on page 41 can not be used because the signal to background ratio is equally unknown. To solve this problem, Punzi came up with a solution [Pun03]. He defined a figure of merit optimizing the sensitivity in searches, which is defined as follows.

$$\text{FoM} = \frac{\varepsilon}{\frac{n_\sigma}{2} + \sqrt{B}} \tag{7.4}$$

Here, ϵ is the signal efficiency, B is the number of background events and n_σ the number of standard deviations of the desired sensitivity. This variable is set to three since this value is commonly considered as an evidence. The advantage of this method is clear because it depends on the signal efficiency instead of the overall number of chosen signal events, which can be easily determined on signal MC and on the total amount of remaining background events which is not a problem since the background consists of well known processes.

A scan on $\mathcal{N}_\text{sel}$ has been conducted for each channel to find the optimal cut value, as can be seen in Figure 7.4 on page 62. Displayed is the FoM distribution of an optimization performed on the full training sample after the application of the precuts described in Section 5.2.4 on page 37, and a curve showing the respective signal efficiency of the cut and the efficiency of the background reduction (ROC curve). The overall efficiencies of the complete procedure including reconstruction, vetoes, precuts, and the cut on $\mathcal{N}_\text{sel}$ can be looked up in Table 7.1 on page 60.

The corresponding values for the hadronically tagged analysis are displayed in Table 7.2 on page 60. A direct comparison affirms the notion that the hadronically tagged sample is purer but less efficient. Taking both numbers into consideration, a net gain can be deduced from the data. In case of the $B^+ \to K^+ \nu\bar{\nu}$ channel for instance, the background is enhanced by a factor of about 2.5 while the efficiency is enhanced by a factor of nearly 4. Thus, the purity-efficiency ratio experiences a sizable net gain by the switch from purely hadronic to semileptonic tagging. This statement holds for all measured channels.
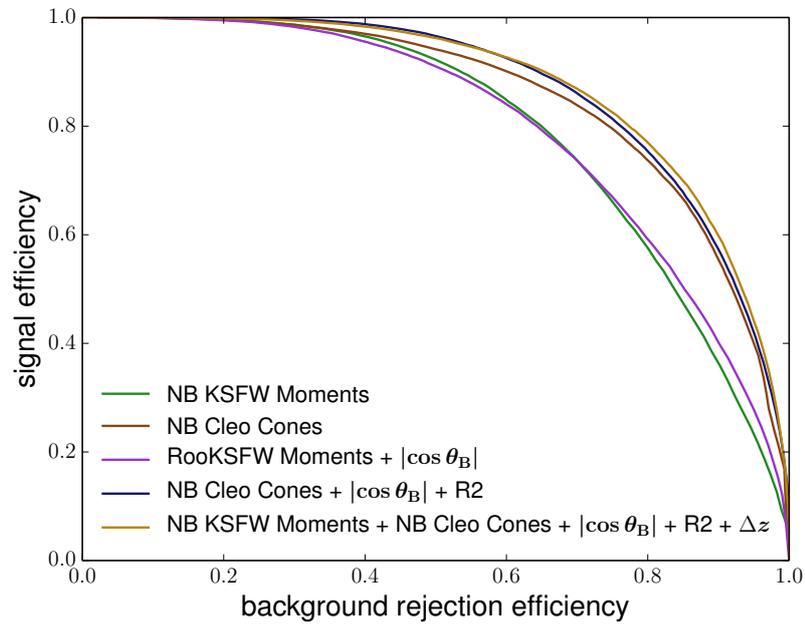
Figure 7.2.: Continuum suppression final training results, signal efficiency over background rejection efficiency for the $B^+ \to K^+ \nu \bar{\nu}$ channel as an example. The variable with the golden color, lowest in the list, has been chosen.
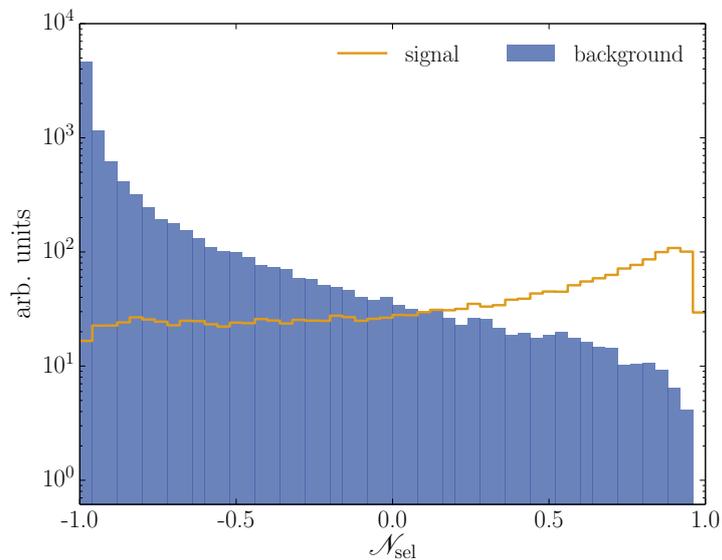
Figure 7.3.: $\mathcal{N}_{\text{sel}}$ for signal and background in the $B^+ \to K^+ \nu \bar{\nu}$ channel.

Table 7.1.: Efficiencies[a] and expected background yields[b] determined on MC after the selection.

	ε in 10^{-3}	exp. # of background
$B^+ \to K^+ \nu\bar{\nu}$	2.16	103.6
$B^0 \to K^0_S \nu\bar{\nu}$	0.91	22.4
$B^+ \to K^{*+} \left(\to K^+ \pi^0\right) \nu\bar{\nu}$	0.25	11.1
$B^+ \to K^{*+} \left(\to K^0_S \pi^+\right) \nu\bar{\nu}$	0.32	24.1
$B^0 \to K^{*0} \nu\bar{\nu}$	0.51	24.8
$B^+ \to \pi^+ \nu\bar{\nu}$	2.92	474.0
$B^0 \to \pi^0 \nu\bar{\nu}$	1.42	41.0
$B^+ \to \rho^+ \nu\bar{\nu}$	0.82	62.4
$B^0 \to \rho^0 \nu\bar{\nu}$	1.11	172.5

Table 7.2.: Efficiencies and expected number of background events of the hadronically tagged analysis from [L$^+$13b].

	ε in 10^{-3}	exp. # of background
$B^+ \to K^+ \nu\bar{\nu}$	0.57	36.8
$B^0 \to K^0_S \nu\bar{\nu}$	0.08	3.4
$B^+ \to K^{*+} \left(\to K^+ \pi^0\right) \nu\bar{\nu}$	0.18	17.2
$B^+ \to K^{*+} \left(\to K^0_S \pi^+\right) \nu\bar{\nu}$	0.10	2.4
$B^+ \to K^{*+} \nu\bar{\nu}$	0.15	19.6
$B^0 \to K^{*0} \nu\bar{\nu}$	0.14	13.8
$B^+ \to \pi^+ \nu\bar{\nu}$	0.34	101.4
$B^0 \to \pi^0 \nu\bar{\nu}$	0.17	7.3
$B^+ \to \rho^+ \nu\bar{\nu}$	0.06	33.7
$B^0 \to \rho^0 \nu\bar{\nu}$	0.13	117.0

---

[a]The quoted values for the $B^+ \to K^{*+} \nu\bar{\nu}$ subchannels include the daughter branching fractions.

[b]The scaling factors discussed in Section 8.2 on page 80 are included.

(a) $B^+ \to K^+ \nu\bar{\nu}$
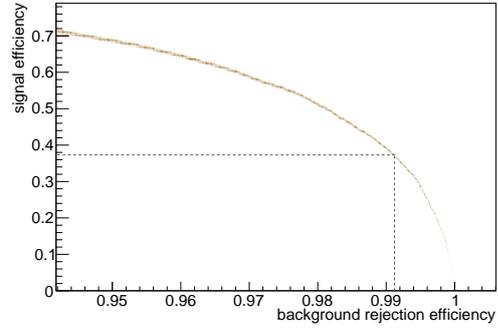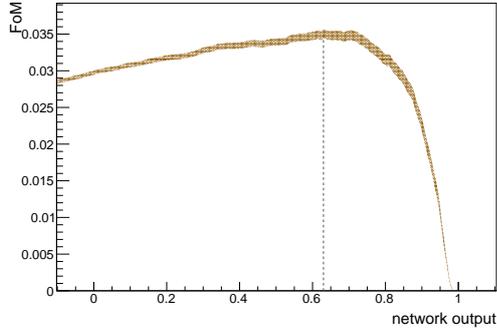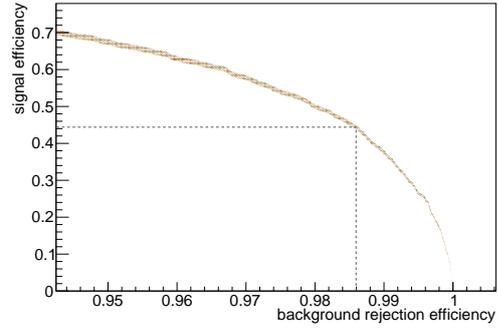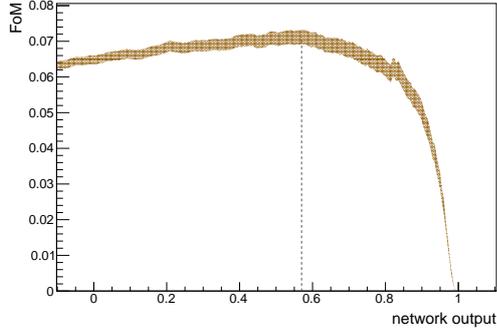
(b) $B^0 \to K^0_S \nu\bar{\nu}$
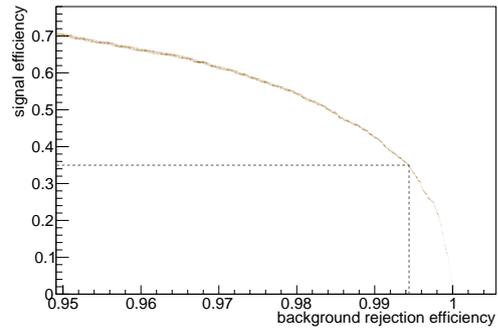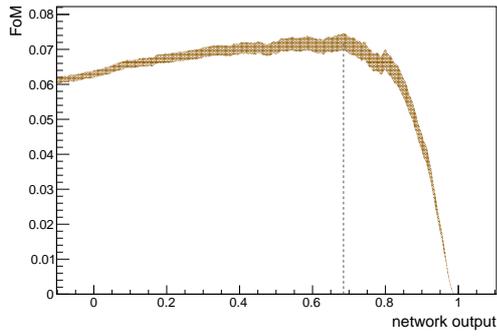
(c) $B^+ \to K^{*+} \left(\to K^+ \pi^0 \right) \nu\bar{\nu}$

(d) $B^+ \to K^{*+} \left(\to K^0_S \pi^+ \right) \nu\bar{\nu}$
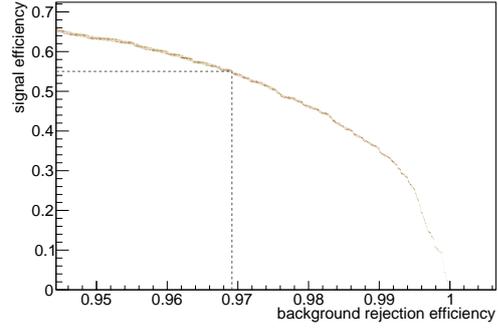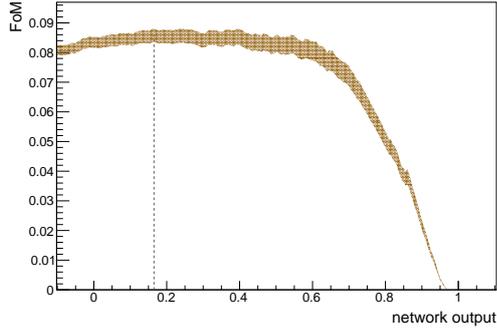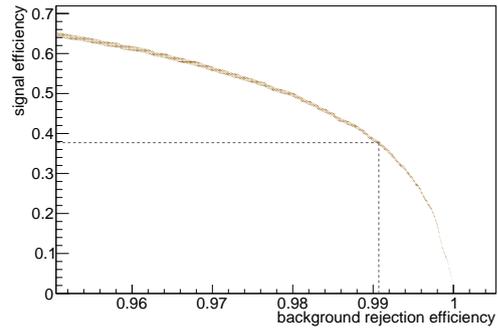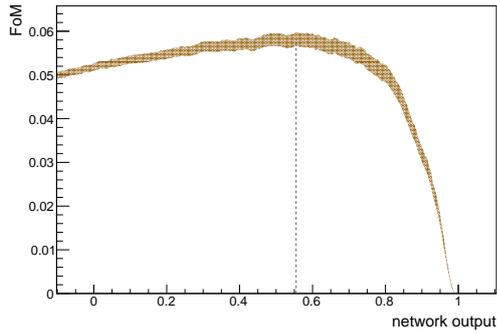
(e) $B^0 \to K^{*0} \nu\bar{\nu}$
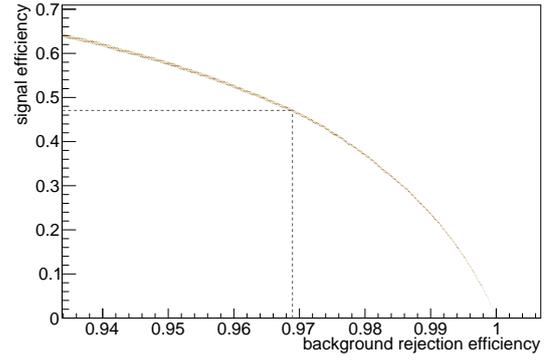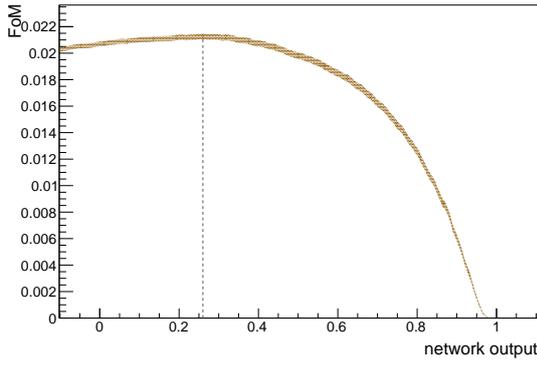
(f) $B^+ \to \pi^+ \nu\bar\nu$
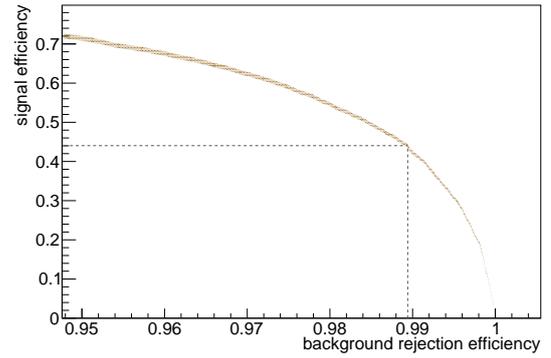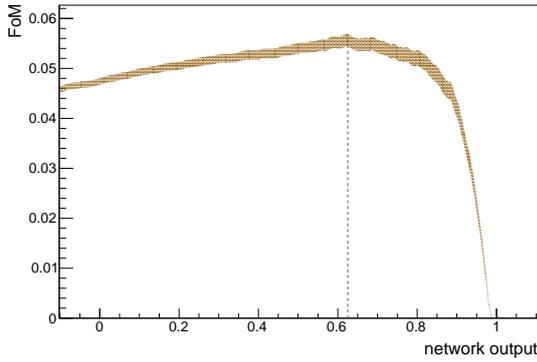
(g) $B^0 \to \pi^0 \nu\bar\nu$
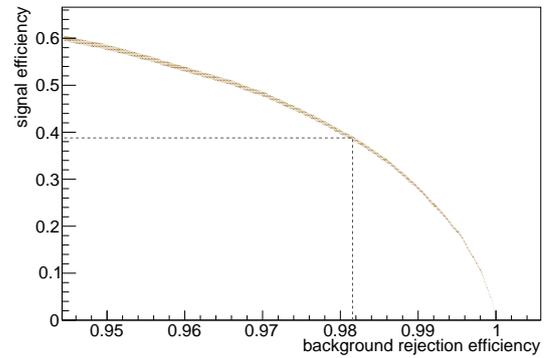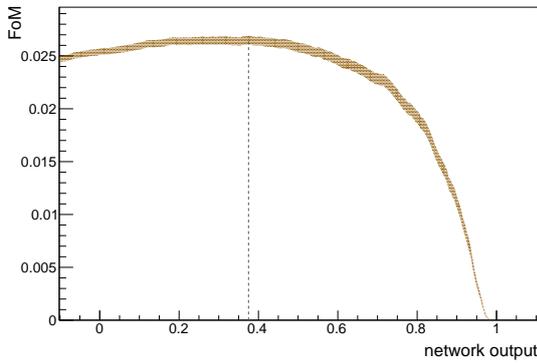
(h) $B^+ \to \rho^+ \nu\bar\nu$
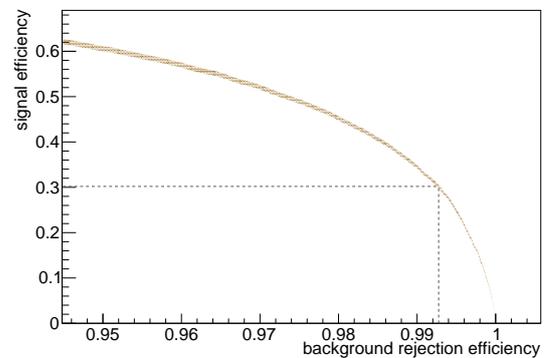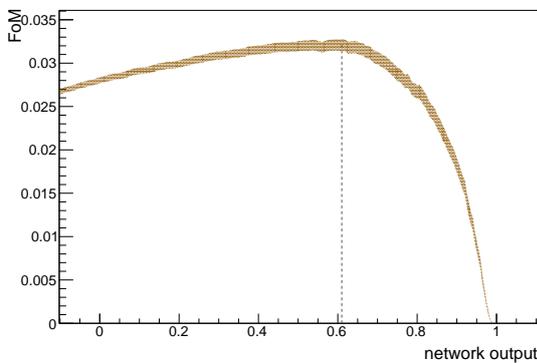
(i) $B^0 \to \rho^0 \nu\bar\nu$

Figure 7.4.: Scan on the network output for all reconstructed channels. Each left figure displays the FoM distribution while each right figure displays the according ROC curve.

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K_S^0 \nu\bar{\nu}$

(c) $B^+ \to K^{*+} \nu\bar{\nu}$

(d) $B^0 \to K^{*0} \nu\bar{\nu}$

(e) $B^+ \to \pi^+ \nu\bar{\nu}$

(f) $B^0 \to \pi^0 \nu\bar{\nu}$

(g) $B^+ \to \rho^+ \nu\bar{\nu}$

(h) $B^0 \to \rho^0 \nu\bar{\nu}$
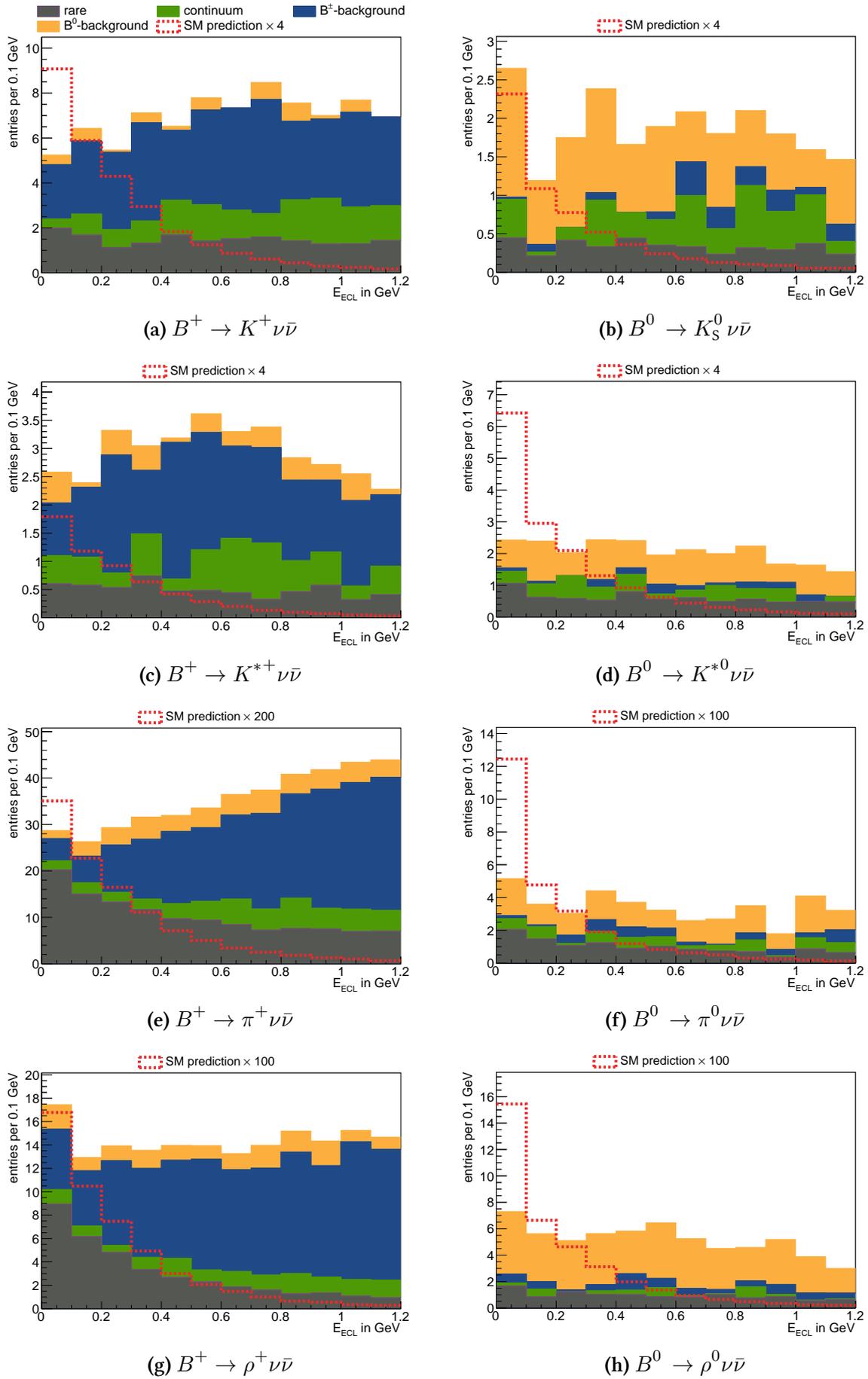
Figure 7.5.: E_{ECL} compositions for all signal modes, the signal strength is scaled relative to the values given in [BGNNS15] for the $b \to s\nu\bar{\nu}$ transitions and [HKR15] for the $b \to s\nu\bar{\nu}$ channels.

63

7.4. Background Composition

The E_{ECL} distributions resulting after applying the cut calculated in Section 7.3 on page 58 are displayed in Figure 7.5 on the preceding page. The background distributions are taken from the amount of simulated data quoted in Section 5.1 on page 31, weighted to the full amount of recorded data of $711\,\text{fb}^{-1}$. The simulated signal samples are weighted depending on the underlying transition on quark level. The respective value for the weight of each channel is given in the legend.

7.4.1. Generic

The majority of all generic background processes come from a misreconstructed signal side; the B_{tag} is correctly reconstructed in more than half of all cases for most channels as can be seen in Table 7.3 on page 66.

The processes contributing to the background are grouped into categories as can be seen in Table 7.4 on page 67. These categories include the following.

- continuum: Continuum events remaining in the final sample, almost exclusively $c\bar{c}$ events.

- events where two charged particles remain unreconstructed. This category is further subdivided into classes of charged particles because the E_{ECL} behaves differently if one or two of them is a lepton. This is mostly due to additional missing particles, in most cases π^0s. An Accompanying π^0 occur far less frequently in leptonic final states.

  - 2 leptons missing: Both missing charged particles are leptons.

  - lepton and hadron missing: One missing charged particle is a lepton, the other is a hadron.

  - 2 charged hadrons missing: Both missing charged particles are hadrons.

- wrong B type: B^+B^- events reconstructed as $B^0\bar{B}^0$ events or vice versa.

- K_{L}^0s missing: events where K_{L}^0s are not reconstructed.

- π^0s missing: events where π^0s are missing.

- no match: Events where a signal particle contains either daughter particles from tag and signal side simultaneously or even at least one particle from the beam background are labeled as no match because it is not traceable to a specific B decay. This category can only occur for unstable $B \to h\nu\bar{\nu}$ decays, e.g. in the $B^+ \to \rho^+\nu\bar{\nu}$ or the $B^0 \to K^{*0}\nu\bar{\nu}$ channel.

- other: All remaining events. Among this category are e.g. events with neutrons in the final state from $B \to J/\Psi\,(\to n\bar{n})\,X$ decays and events where even four charged particles are missed.

Albeit applying the K_L^0 veto described in Section 6.2 on page 47, major backgrounds consist of processes involving K_L^0s for the aforementioned reasons in Section 6.2 on page 47, especially in the π and ρ channels. This is mainly because the major contribution therein are decays like $B \to D\left(\to K_L^0 \pi^0\right)\pi$. This is more unlikely in the kaonic channels because the Cabibbo-favored emergence of strangeness is absorbed by a correctly reconstructed signal kaon. For those channels, the background composition is dominated by events where two charged tracks are missed. Those are mostly due to semileptonic B decays. Since they include a neutrino themselves, the direction of the missing momentum in the lab frame has shown to be not well suited to suppress such backgrounds.

Events in which a semileptonic B decay is subsequently followed by a semileptonic D decay where both leptons escape the acceptance of the detector exhibit a signal-like E_{ECL} distribution because the misreconstructed particles also evade the calorimeter. This point is demonstrated in the right hand side in Table 7.4 on page 67 as the fraction of double semileptonic decays in the signal region is roughly twice as large as the fraction of the same category in the full E_{ECL} range. The correct description of this background component is thus highly important. While semileptonic B and D^0 decays are correctly described in the used MC, this is not the case for semileptonic D^+ decays. Hence, the branching fractions for the $D^+ \to K^0 e^+ \nu_e$ and $D^+ \to K^0 e^+ \nu_e$ decays have been corrected to up-to date values, see Appendix B for a comprehensive list.

Table 7.3.: Fraction of correct tags in generic background.

channel	correct tag fraction
$B^+ \to K^+ \nu\bar{\nu}$	58.6 %
$B^0 \to K^0_S \nu\bar{\nu}$	54.7 %
$B^+ \to K^{*+}\left(\to K^+\pi^0\right)\nu\bar{\nu}$	60.0 %
$B^+ \to K^{*+}\left(\to K^0_S\pi^+\right)\nu\bar{\nu}$	39.8 %
$B^0 \to K^{*0}\nu\bar{\nu}$	64.8 %
$B^+ \to \pi^+ \nu\bar{\nu}$	48.0 %
$B^0 \to \pi^0 \nu\bar{\nu}$	48.9 %
$B^+ \to \rho^+ \nu\bar{\nu}$	60.4 %
$B^0 \to \rho^0 \nu\bar{\nu}$	49.0 %

Table 7.4.: Composition of the generic background after the application of the complete selection; plotted over the whole E_{ECL} range $0 < E_{\mathrm{ECL}} < 1.2\,\mathrm{GeV}$ (left) and for the signal window $E_{\mathrm{ECL}} < 0.3\,\mathrm{GeV}$ only (right).

	contribution in %		contribution in %
continuum	22.6	continuum	20.5
2 leptons missing	15.3	2 leptons missing	36.1
K^0_Ls and lepton missing	6.5	K^0_Ls and lepton missing	9.6
lepton and hadrons missing	24.1	lepton and hadrons missing	24.1
2 charged hadrons missing	1.7	2 charged hadrons missing	0.0
wrong B type	3.8	wrong B type	0.0
hadronic, K^0_L missing	24.1	hadronic, K^0_L missing	9.6
hadronic π^0 missing	1.0	hadronic π^0 missing	0.0
no match	0.0	no match	0.0
other	1.0	other	0.0

(a) $B^+ \to K^+ \nu\bar\nu$

	contribution in %		contribution in %
continuum	26.4	continuum	15.6
2 leptons missing	16.8	2 leptons missing	31.3
K^0_Ls and lepton missing	3.2	K^0_Ls and lepton missing	6.2
lepton and hadrons missing	32.0	lepton and hadrons missing	28.1
2 charged hadrons missing	0.0	2 charged hadrons missing	0.0
wrong B type	0.0	wrong B type	0.0
hadronic, K^0_L missing	16.0	hadronic, K^0_L missing	15.6
hadronic π^0 missing	1.6	hadronic π^0 missing	0.0
no match	0.0	no match	0.0
other	4.0	other	3.1

(b) $B^0 \to K^0_S \nu\bar\nu$

	contribution in %		contribution in %
continuum	21.0	continuum	25.0
2 leptons missing	11.3	2 leptons missing	16.7
K^0_Ls and lepton missing	8.1	K^0_Ls and lepton missing	0.0
lepton and hadrons missing	30.6	lepton and hadrons missing	33.3
2 charged hadrons missing	1.6	2 charged hadrons missing	0.0
wrong B type	0.0	wrong B type	0.0
hadronic, K^0_L missing	19.4	hadronic, K^0_L missing	8.3
hadronic π^0 missing	0.0	hadronic π^0 missing	0.0
no match	8.1	no match	16.7
other	0.0	other	0.0

(c) $B^+ \to K^{*+} \left(\to K^+ \pi^0\right) \nu\bar\nu$

	contribution in %
continuum	22.1
2 leptons missing	6.2
K_L^0s and lepton missing	2.1
lepton and hadrons missing	38.6
2 charged hadrons missing	1.4
wrong B type	2.8
hadronic, K_L^0 missing	9.0
hadronic π^0 missing	2.8
no match	12.4
other	2.8

	contribution in %
continuum	14.7
2 leptons missing	14.7
K_L^0s and lepton missing	2.9
lepton and hadrons missing	29.4
2 charged hadrons missing	2.9
wrong B type	8.8
hadronic, K_L^0 missing	2.9
hadronic π^0 missing	0.0
no match	17.6
other	5.9

(d) $B^+ \to K^{*+}\left(\to K_S^0\,\pi^+\right)\nu\bar{\nu}$

	contribution in %
continuum	22.5
2 leptons missing	10.6
K_L^0s and lepton missing	9.9
lepton and hadrons missing	21.8
2 charged hadrons missing	1.4
wrong B type	5.6
hadronic, K_L^0 missing	20.4
hadronic π^0 missing	1.4
no match	4.9
other	1.4

	contribution in %
continuum	31.6
2 leptons missing	13.2
K_L^0s and lepton missing	7.9
lepton and hadrons missing	15.8
2 charged hadrons missing	2.6
wrong B type	7.9
hadronic, K_L^0 missing	18.4
hadronic π^0 missing	0.0
no match	2.6
other	0.0

(e) $B^0 \to K^{*0}\nu\bar{\nu}$

	contribution in %
continuum	13.4
2 leptons missing	5.3
K_L^0s and lepton missing	27.4
lepton and hadrons missing	9.3
2 charged hadrons missing	1.5
wrong B type	3.9
hadronic, K_L^0 missing	37.6
hadronic π^0 missing	1.0
no match	0.0
other	0.5

	contribution in %
continuum	15.8
2 leptons missing	9.2
K_L^0s and lepton missing	32.6
lepton and hadrons missing	11.5
2 charged hadrons missing	0.3
wrong B type	4.6
hadronic, K_L^0 missing	25.0
hadronic π^0 missing	1.0
no match	0.0
other	0.0

(f) $B^+ \to \pi^+\nu\bar{\nu}$

	contribution in %
continuum	20.6
2 leptons missing	1.7
K_L^0s and lepton missing	31.8
lepton and hadrons missing	9.0
2 charged hadrons missing	6.0
wrong B type	5.2
hadronic, K_L^0 missing	18.9
hadronic π^0 missing	3.9
no match	0.4
other	2.6

	contribution in %
continuum	19.6
2 leptons missing	3.6
K_L^0s and lepton missing	44.6
lepton and hadrons missing	10.7
2 charged hadrons missing	3.6
wrong B type	3.6
hadronic, K_L^0 missing	12.5
hadronic π^0 missing	0.0
no match	0.0
other	1.8

(g) $B^0 \to \pi^0 \nu \bar{\nu}$

	contribution in %
continuum	8.9
2 leptons missing	1.1
K_L^0s and lepton missing	19.0
lepton and hadrons missing	6.3
2 charged hadrons missing	3.0
wrong B type	1.6
hadronic, K_L^0 missing	48.4
hadronic π^0 missing	1.3
no match	8.9
other	1.5

	contribution in %
continuum	10.0
2 leptons missing	1.8
K_L^0s and lepton missing	26.2
lepton and hadrons missing	9.0
2 charged hadrons missing	3.6
wrong B type	1.4
hadronic, K_L^0 missing	35.7
hadronic π^0 missing	0.5
no match	10.9
other	0.9

(h) $B^+ \to \rho^+ \nu \bar{\nu}$

	contribution in %
continuum	6.3
2 leptons missing	12.3
K_L^0s and lepton missing	40.3
lepton and hadrons missing	8.9
2 charged hadrons missing	1.2
wrong B type	0.7
hadronic, K_L^0 missing	15.7
hadronic π^0 missing	1.9
no match	10.9
other	1.7

	contribution in %
continuum	5.4
2 leptons missing	19.6
K_L^0s and lepton missing	42.9
lepton and hadrons missing	5.4
2 charged hadrons missing	0.0
wrong B type	0.9
hadronic, K_L^0 missing	8.0
hadronic π^0 missing	0.0
no match	13.4
other	4.5

(i) $B^0 \to \rho^0 \nu \bar{\nu}$

7.4.2. Rare

The fraction of background events with a correctly reconstructed B_{tag} is higher for rare decays than for generic decays, which can be seen when comparing Table 7.6 on the next page with Table 7.3 on page 66. The different rare background processes are grouped accordingly to generic backgrounds. The *2 charged missing* category is not further subdivided because it is not as important for the overall E_{ECL} distribution, neither for the full range nor for the signal range.

Two new categories are introduced which cannot occur in generic samples. Those are:

- baryonic decays: B decays involving baryons.

- tauonic decays: B decays involving one or more tau leptons.

Rare background processes are dominated by processes involving K_{L}^0s either through non resonant production, e.g. in the K_{S}^0 channel where the process $B \to K_{\text{S}}^0 K_{\text{L}}^0 K_{\text{L}}^0$ accounts for one third of the of the total amount of rare background events, albeit the process has never been observed and only a limit is known. The other possibility is via decays including mesons decaying into pairs of K_{L}^0s, e.g. the f_2'. The process $B \to K^{*+} f_2' \left(\to K_{\text{L}}^0 K_{\text{L}}^0 \right)$ is accountable for $40\,\%$ of all rare background events in the $B^+ \to K^{*+} \nu \bar{\nu}$ channel. The problem here is that, in case the K_{L}^0s leave the detector undetected, which as mentioned in Section 6.2 on page 47 is rather likely, such processes resemble the signal process. Since decays including K_{L}^0s and the signal particle only do not lead to significant amounts of energy deposed in the calorimeter, such decays peak in the E_{ECL} signal region. This is also true for $B \to \tau \nu_\tau$ decays. If the τ decays subsequently into $h \nu_\tau$, the final state exactly resembles the respective signal $B \to h \nu \bar{\nu}$ decay. However, since measurements of the decay $B \to \tau \nu_\tau$ have been published recently, the branching fraction of this process is much better known than most other background sources in the rare sample.

The effect of all this can be observed in Figure 7.5e and Figure 7.5g on page 63. The relative amount of rare decays in the final sample as compared to generic background processes is comparably high which makes reweighting the rare distribution to most recent experimental values necesssary. The rare background is clearly peaking in the signal region of $E_{\text{ECL}} < 0.3\,\text{GeV}$, significantly more in the $B^+ \to \pi^+ \nu \bar{\nu}$ and $B^+ \to \rho^+ \nu \bar{\nu}$ channels. Here, the category *other* in Table 7.7 on page 72 is mainly composed of misidentified charged tracks, for example a ρ candidate formed out of a pion and an electron misidentified as a pion from a $B \to \pi e \nu_e$ decay.

Table 7.6.: Fraction of correct tags in rare background

channel	correct tag fraction
$B^+ \rightarrow K^+ \nu\bar{\nu}$	58.8 %
$B^0 \rightarrow K^0_S \nu\bar{\nu}$	73.3 %
$B^+ \rightarrow K^{*+} \left(\rightarrow K^+ \pi^0 \right) \nu\bar{\nu}$	58.1 %
$B^+ \rightarrow K^{*+} \left(\rightarrow K^0_S \pi^+ \right) \nu\bar{\nu}$	50.8 %
$B^0 \rightarrow K^{*0} \nu\bar{\nu}$	69.4 %
$B^+ \rightarrow \pi^+ \nu\bar{\nu}$	59.4 %
$B^0 \rightarrow \pi^0 \nu\bar{\nu}$	58.7 %
$B^+ \rightarrow \rho^+ \nu\bar{\nu}$	62.8 %
$B^0 \rightarrow \rho^0 \nu\bar{\nu}$	56.2 %

Table 7.7.: Composition of the rare background after the application of the complete selection. The right hand tables display the composition in the signal window.

	contribution in %
2 charged missing	3.5
K_L^0s missing	48.1
π^0s missing	8.5
tauonic decays	14.3
baryonic decays	13.1
no match	0.0
other	12.6

(a) $K^+\nu\bar{\nu}$

	contribution in %
2 charged missing	1.0
K_L^0s missing	38.9
π^0s missing	4.7
tauonic decays	29.0
baryonic decays	11.0
no match	0.0
other	15.5

(b) $K^+\nu\bar{\nu}$

	contribution in %
2 charged missing	9.1
K_L^0s missing	74.4
π^0s missing	4.0
tauonic decays	0.3
baryonic decays	2.0
no match	1.3
other	8.9

(c) $K_S\nu\bar{\nu}$

	contribution in %
2 charged missing	10.0
K_L^0s missing	77.5
π^0s missing	3.8
tauonic decays	0.0
baryonic decays	1.9
no match	0.0
other	6.7

(d) $K_S\nu\bar{\nu}$

	contribution in %
2 charged missing	2.0
K_L^0s missing	70.3
π^0s missing	2.7
tauonic decays	11.3
baryonic decays	1.8
no match	2.0
other	9.9

(e) $K^{*+}(K^+\pi^0)\nu\bar{\nu}$

	contribution in %
2 charged missing	2.6
K_L^0s missing	63.2
π^0s missing	1.3
tauonic decays	17.1
baryonic decays	1.3
no match	0.0
other	14.5

(f) $K^{*+}(K^+\pi^0)\nu\bar{\nu}$

	contribution in %
2 charged missing	3.3
K_L^0s missing	69.3
π^0s missing	3.5
tauonic decays	3.5
baryonic decays	2.1
no match	7.9
other	10.4

(g) $K^{*+}(K_S\pi^+)\nu\bar\nu$

	contribution in %
2 charged missing	1.5
K_L^0s missing	56.5
π^0s missing	3.1
tauonic decays	6.1
baryonic decays	3.1
no match	7.6
other	22.1

(h) $K^{*+}(K_S\pi^+)\nu\bar\nu$

	contribution in %
2 charged missing	0.8
K_L^0s missing	72.5
π^0s missing	4.0
tauonic decays	0.6
baryonic decays	2.2
no match	2.3
other	17.7

(i) $K^{*0}\nu\bar\nu$

	contribution in %
2 charged missing	0.9
K_L^0s missing	62.1
π^0s missing	2.7
tauonic decays	0.9
baryonic decays	2.3
no match	3.0
other	28.1

(j) $K^{*0}\nu\bar\nu$

	contribution in %
2 charged missing	5.1
K_L^0s missing	44.2
π^0s missing	10.6
tauonic decays	32.8
baryonic decays	0.7
no match	0.0
other	6.6

(k) $\pi^+\nu\bar\nu$

	contribution in %
2 charged missing	3.4
K_L^0s missing	38.3
π^0s missing	3.7
tauonic decays	49.2
baryonic decays	0.3
no match	0.0
other	5.2

(l) $\pi^+\nu\bar\nu$

	contribution in %
2 charged missing	17.4
K^0_L s missing	61.7
π^0 s missing	5.6
tauonic decays	2.6
baryonic decays	2.2
no match	0.0
other	10.5

(m) $\pi^0 \nu \bar{\nu}$

	contribution in %
2 charged missing	21.0
K^0_L s missing	63.6
π^0 s missing	2.3
tauonic decays	2.9
baryonic decays	2.0
no match	0.0
other	8.1

(n) $\pi^0 \nu \bar{\nu}$

	contribution in %
2 charged missing	2.4
K^0_L s missing	9.8
π^0 s missing	8.0
tauonic decays	53.2
baryonic decays	0.7
no match	10.1
other	15.8

(o) $\rho^+ \nu \bar{\nu}$

	contribution in %
2 charged missing	2.7
K^0_L s missing	5.4
π^0 s missing	3.3
tauonic decays	68.5
baryonic decays	0.3
no match	7.8
other	11.9

(p) $\rho^+ \nu \bar{\nu}$

	contribution in %
2 charged missing	1.7
K^0_L s missing	47.1
π^0 s missing	10.5
tauonic decays	1.2
baryonic decays	0.7
no match	9.1
other	29.7

(q) $\rho^0 \nu \bar{\nu}$

	contribution in %
2 charged missing	2.0
K^0_L s missing	36.6
π^0 s missing	11.4
tauonic decays	1.0
baryonic decays	0.0
no match	11.6
other	37.4

(r) $\rho^0 \nu \bar{\nu}$

7.4.3. E_{ECL} Composition

Since the final fit is executed on E_{ECL}, a broader understanding thereof is of great importance. Terms contributing to the total extra energy fall into three basic categories, distinguished by the process causing the cluster in the calorimeter.

- Physics processes: Calorimeter clusters caused by a particle of the direct decay chain of the $\Upsilon(4S)$. In an analysis fully reconstructing the event, these indicate an error in the reconstrution.

- Secondary clusters: Clusters from secondary interactions of particles of the direct decay chain of the $\Upsilon(4S)$. Those include decay in flight products, backscattered tracks, hadronic reactions in the detector, and bremsstrahlung. Such clusters do not hint at any misreconstruction since the final state particle responsible is found as well.

- Beam background: clusters caused by beam background. They, again, do not indicate a misreconstruction and occur independent of the nature of the event.

The E_{ECL} distribution of each category after cuts is displayed in Figure 7.6 on the following page, exemplified for the $B^+ \to K^+ \nu \bar{\nu}$ channel. The largest share of the total E_{ECL} is indeed caused by unreconstructed primary (physics) particles. However, the peak in the first bin in Figure 7.6a on the next page shows that many background events do not deposit sufficient extra energy in the calorimeter by the means of physics events only. Since additional physics might leave its traces via secondary interactions, the resulting background E_{ECL} distribution does not peak in the signal region, as can be seen in Figure 7.5 on page 63. This finding is in good agreement with the contributing processes in Table 7.4 on page 67, where processes with two not reconstructed tracks make up a noticeable share of all background events. So, using E_{ECL} for the separation between correctly reconstructed and all other events is well justified.

However, the comparison of Figure 7.5a on page 63 and Figure 7.6a on the next page yields the surprising insight that E_{ECL} works even better than intended. E_{ECL} is designed to identify processes where clusters from $\Upsilon(4S)$ decays have not been identified as such, indicating e.g. a missing π^0 mimicking missing momentum. The peak in the first bin in Figure 7.6a on the following page, which is missing in Figure 7.5a, demonstrates that the contributions of secondary reactions (Figure 7.6b) in the detector are essential to the performance of E_{ECL} as a final fitting variable since a background peaking in the signal region leads to considerably larger statistical uncertainties. When considering the plots in Figure 7.5 on page 63, it is important to note that the selection corresponds to all chosen cuts and weights, including the preselection cut of $E_{\mathrm{ECL}} < 1.2\,\mathrm{GeV}$. This means that events can not migrate out of the plot but from the peak in the first bin of Figure 7.6a to higher E_{ECL} values yielding the rather flat distribution in Figure 7.5a.

The plots in Figure 7.6b and 7.6c also demonstrate why the signal distribution is not localized at $0\,\mathrm{GeV}$. Both processes occur as well in correctly reconstructed signal events.
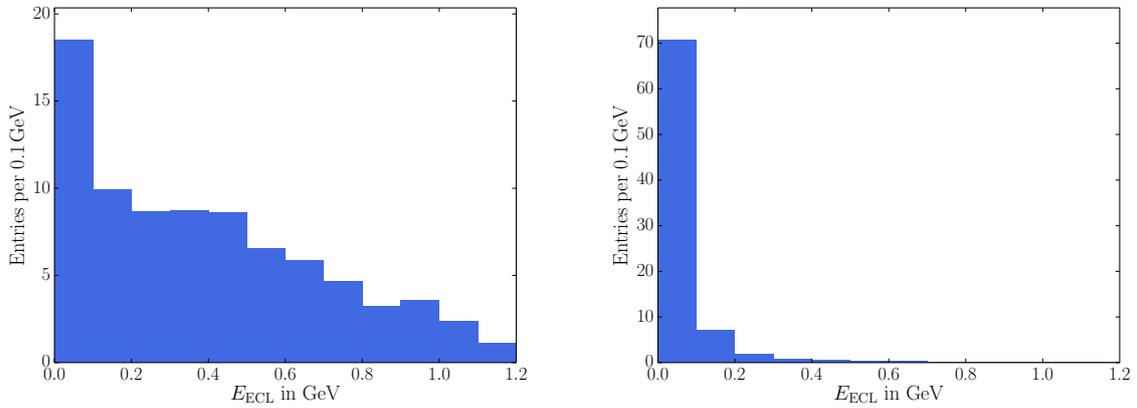
(a) E_{ECL} contribution from decay products of the primary interaction.

(b) E_{ECL} contribution from beam induced background.
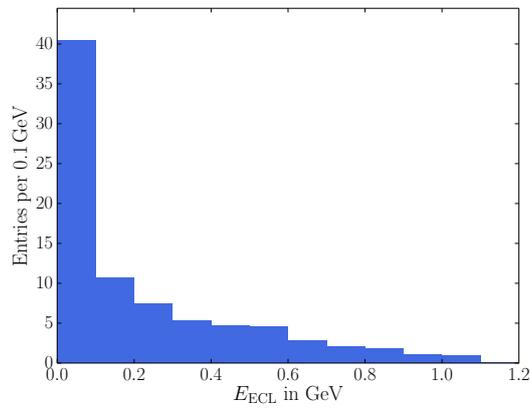
(c) E_{ECL} contribution from interactions of decay products of the primary particles in the detector.

Figure 7.6.: Contributions to the E_{ECL} background distributions of the $B^+ \to K^+ \nu\bar{\nu}$ channel. Primary particle refers to those produced in a e^+e^- collision, in most cases the $\Upsilon(4S)$.

Chapter 8

Validation

8.1. Sideband

In order to test whether the MC samples describe the data well and to estimate the background yield, we define an E_{ECL} sideband where we require $E_{\mathrm{sb}} > 0.3\,\mathrm{GeV}$. We expect the signal contribution in this E_{ECL} region to be negligible. The resulting distributions can be seen in Figure 8.1 on the following page. We take the uncertainty on the MC expectation into account when we calculate the χ^2 value. The direct comparison exhibits a rough agreement of data and MC for most channels. This statement is not true for the $B^+ \to K^+ \nu\bar{\nu}$ and also the $B^+ \to \pi^+ \nu\bar{\nu}$ channel where the data significantly overshoot the MC prediction. This is not reflected in the χ^2 value but a trend to overestimate the signal is clearly visible. In the $B^+ \to K^+ \nu\bar{\nu}$ channel for instance, the data overshoot the MC expectation in each bin. The comparison is subsequently repeated for the most important input variables and the composition of the background in the region of the parameter space where the inconsistencies occurr is analyzed.

This analysis shows an enhanced contribution of continuum background compared to the MC expectation by over a factor of two for the respective cut value on $\mathscr{N}_{\mathrm{sel}}$. It can be found that the observed discrepancy in the $B^+ \to K^+ \nu\bar{\nu}$ channel occurs in the signal regions of most variables, most prominently in the $\mathscr{N}_{\mathrm{NB,\,CS}}$ window of $0.9 < \mathscr{N}_{\mathrm{NB,\,CS}} < 1$ but also in p_{cms} ($1.5\,\mathrm{GeV} < p_{\mathrm{cms}} < 2.5\,\mathrm{GeV}$). and $\cos\theta_{B,\,D^{(*)}l}$ $\left(-1 < \cos\theta_{B,\,D^{(*)}l} < +1\right)$, see Figure 8.2 on page 79. Therefore, the processes responsible for the excess must exhibit a signal like topology. For instance, signal events can be faked by $e^+e^- \to c\bar{c}$ events where one c quark hadronizes into a correctly reconstructed D meson, while the D meson formed by the second c quark decays semileptonically where the charged lepton is added to the D candidate to form a B candidate and the kaon is taken as a signal candidate. An example for a variable hinting towards this explanation is displayed in Figure 8.4 on page 81 where the cosine of the angle between the signal kaon candidate and the tag lepton candidate is plotted. The discrepancies between data and MC occur at low K-l angles which is more likely if both come from the same decay. The deviation in the $B^+ \to \pi^+ \nu\bar{\nu}$ channel can be explained by a related process if one of the charmed mesons decays into a semileptonic final state containing a K_{L}^0.
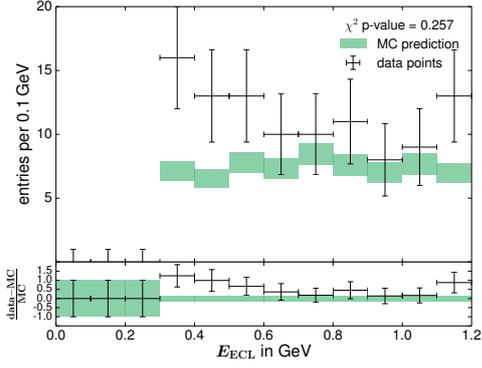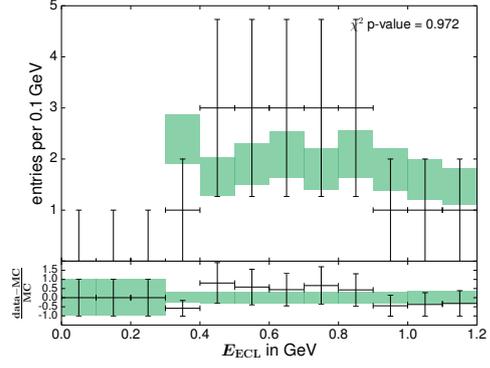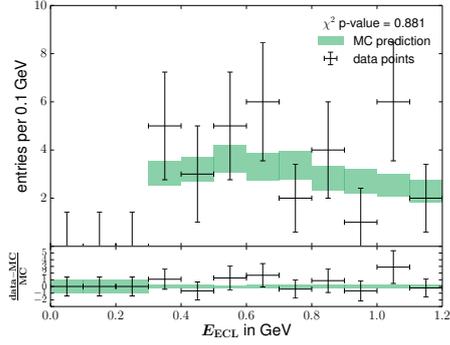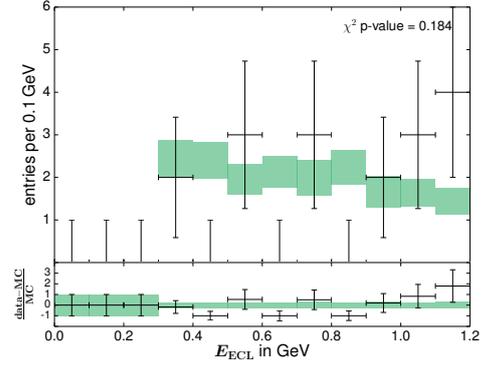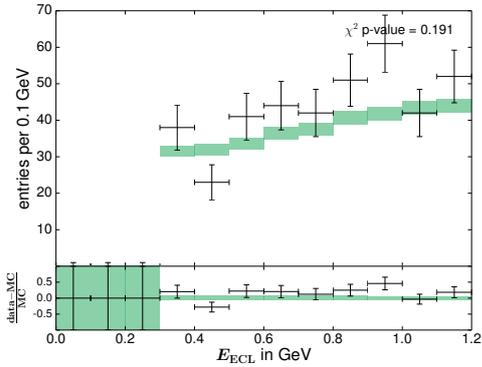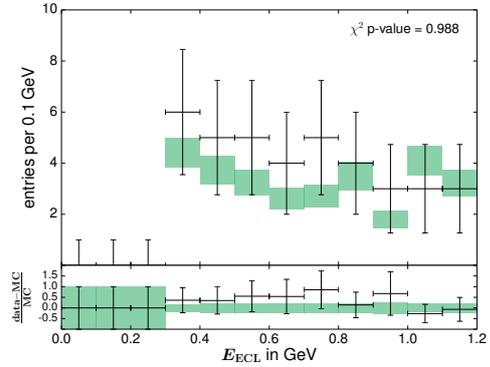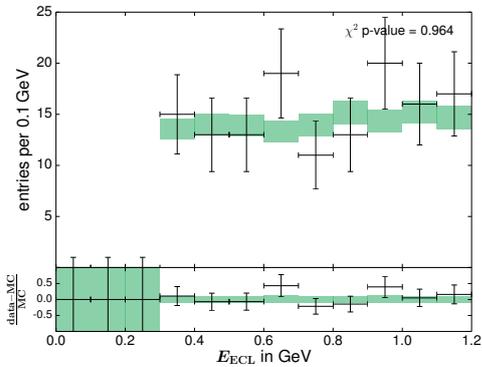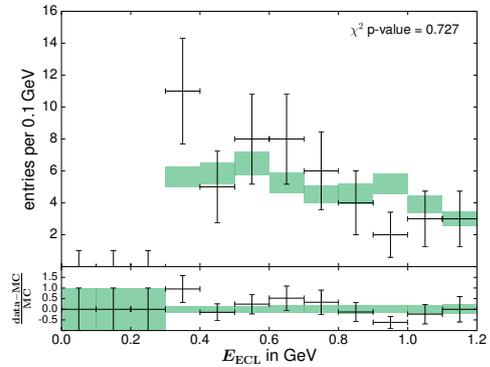
Figure 8.1.: E_{ECL} sideband distributions for all reconstructed channels, no additional weights are applied to our expectations.

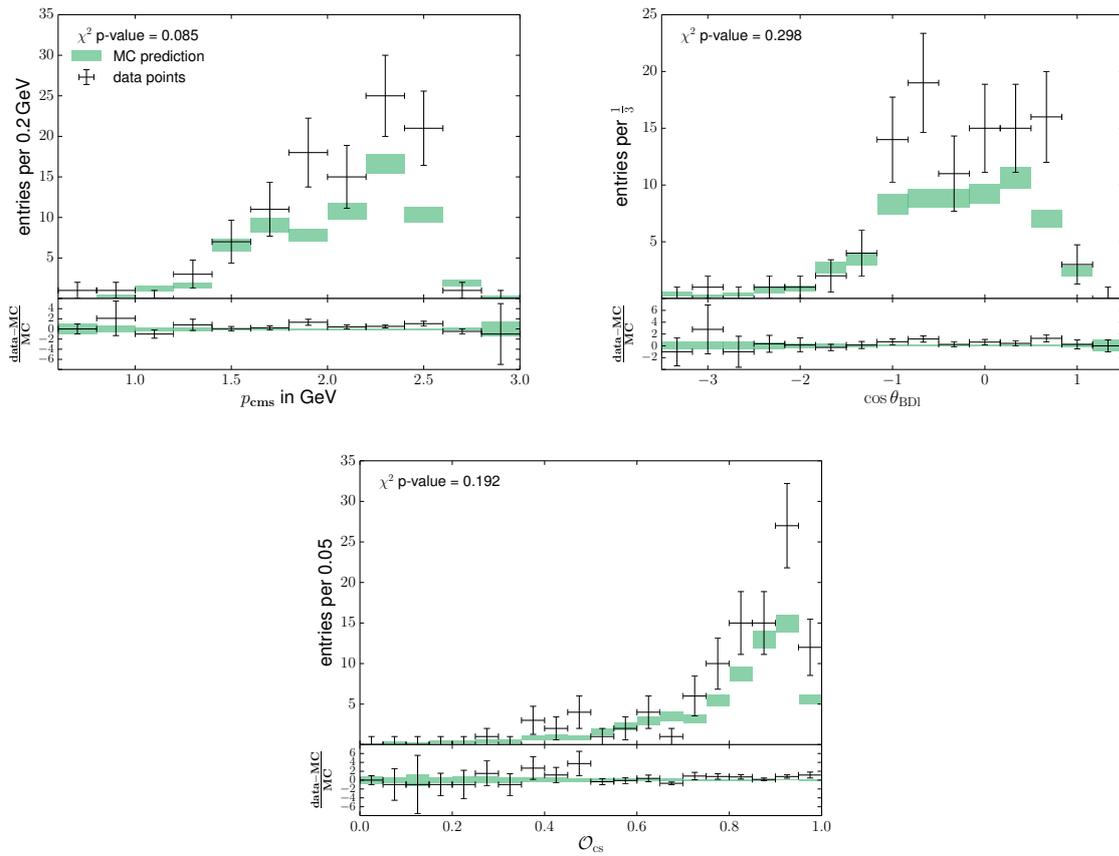Figure 8.2.: Data-MC comparison of selected variables for the $B^+ \to K^+ \nu\bar{\nu}$ sideband.

8.2. Off-Resonance

We utilize the off-resonance sample to study the influence of $e^+e^- \to c\bar{c}$ processes on the observed data-MC differences in the $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to \pi^+\nu\bar{\nu}$ on-resonance sample. However, the comparison of off-resonance data with off-resonance MC suffers from the lack of available data. All in all, the Belle detector took an amount of $100\,\mathrm{fb}^{-1}$ at off-resonance energies corresponding to one seventh of the sample taken on-resonance. The E_{ECL} distributions of this sample are displayed in Figure 8.3 on the next page. Although the statistics does not allow a conclusive statement about the compatibility of data and MC distributions, a trend towards an underestimation of data in MC is visible. We investigate this hypothesis by scanning the data-MC fraction of the number of selected events as a function of $\mathcal{N}_{\mathrm{sel}}$ in Figure 8.5 on page 82. We also plot the number of off-resonance events observed on data and mark the values corresponding to the optimal cut value in red.

The plots in Figure 8.5 on page 82 make it obvious that continuum background for semileptonically tagged selections with a one-prong signal side is generally not well modeled, independent of the cut on $\mathcal{N}_{\mathrm{sel}}$, or in other words independent of the tightness of the selection. This result affirms the notion that the data-MC disagreements are caused by mismodelled continuum events. In addition to this, the following two conclusions can be drawn as well.

- The dependence of the data-MC ratio is not flat in the cut on $\mathcal{N}_{\mathrm{sel}}$ in the vicinity of the optimal value. It is therefor not feasible to take the E_{ECL} distribution from an off-resonance sample with a looser cut on $\mathcal{N}_{\mathrm{sel}}$.

- The statistical error for cut values around the optimal one is considerably large due to the limited size of the off-resonance data sample.

Because of the first statement and despite the second, we take the data-MC ratio from the off-resonance sample and scale the on-resonance continuum MC accordingly. The E_{ECL} distribution in the off-resonance sample is displayed in Figure 8.3 on the facing page. We assign the uncertainty resulting from the limited size of the off-resonance sample as a systematic uncertainty later on. That correction factor and its statistical uncertainty can be found in Table 8.1 on the next page.

The E_{ECL} sideband distributions and the selected variables can be compared in Figure 8.6 on page 83 and Figure 8.7 on page 84. The error on the correction factor applied to the on-resonance continuum MC has been included in error band of the plots on the right side. It is important to note that the errors on the scaling factor of that background component are correlated for all bins in both figures since they are weighted with a global factor. Hence, we trust in our correction method not only for the $B^+ \to \pi^+\nu\bar{\nu}$ channel, but for the $B^+ \to K^+\nu\bar{\nu}$ channel as well, where the same argument applies.

Table 8.1.: Continuum scaling factors for the $B^+ \to K^+ \nu\bar\nu$ and $B^+ \to \pi^+ \nu\bar\nu$ channel determined on off-resonance data with the cuts on $\mathcal{N}_{\text{sel}}$ applied from the optimization in Section 7.3 on page 58.

channel	factor	error
$B^+ \to K^+ \nu\bar\nu$	2.28	0.95
$B^+ \to \pi^+ \nu\bar\nu$	2.05	0.66

(a) $B^+ \to K^+ \nu\bar\nu$

(b) $B^+ \to \pi^+ \nu\bar\nu$
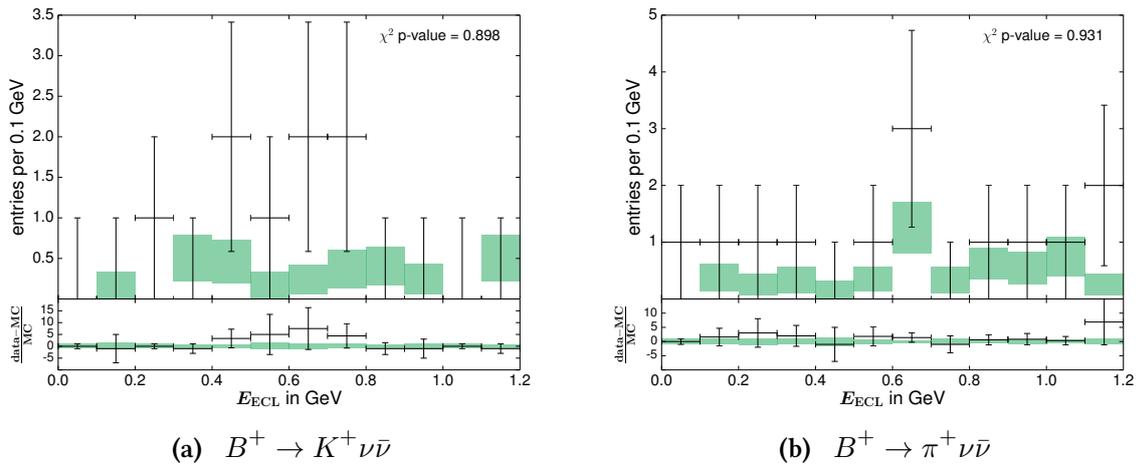
Figure 8.3.: Off-Resonance E_{ECL} distributions for MC prediction and data.
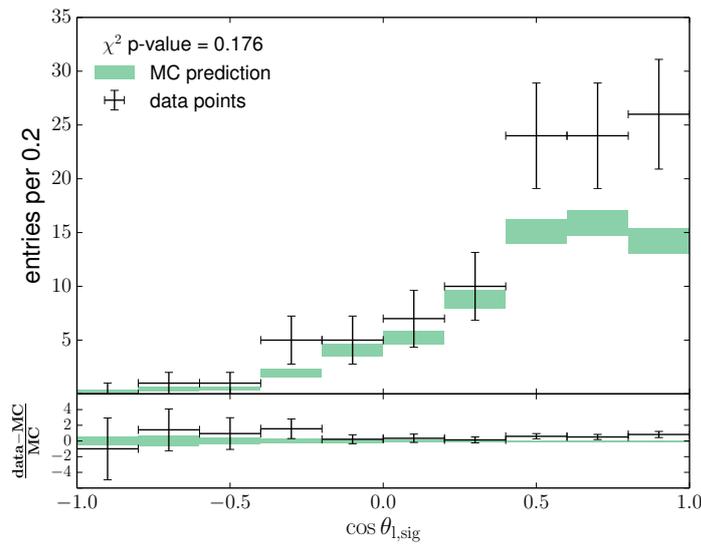
Figure 8.4.: Distribution of the angle between the signal kaon and the tagside lepton for the $B^+ \to K^+ \nu\bar\nu$ channel from the E_{ECL} sideband.
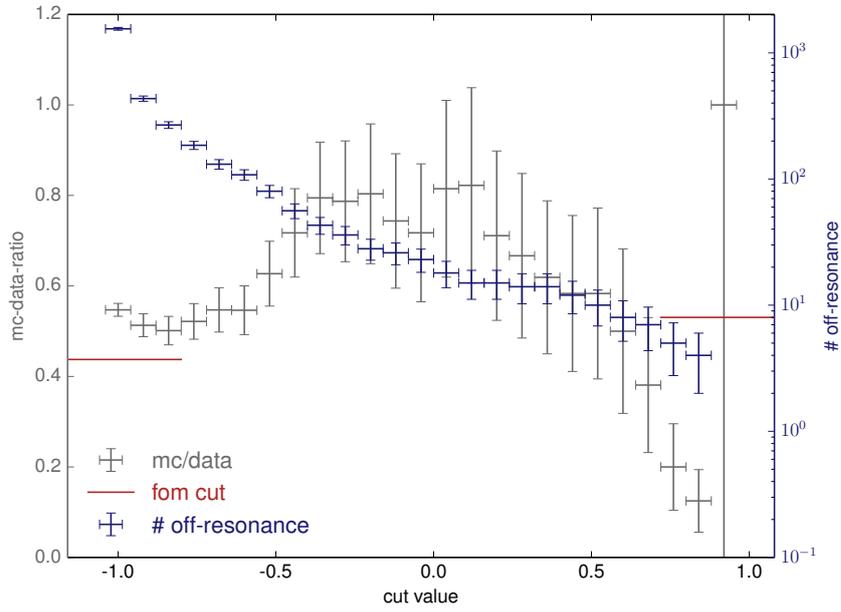
(a) $B^+ \to K^+ \nu\bar{\nu}$
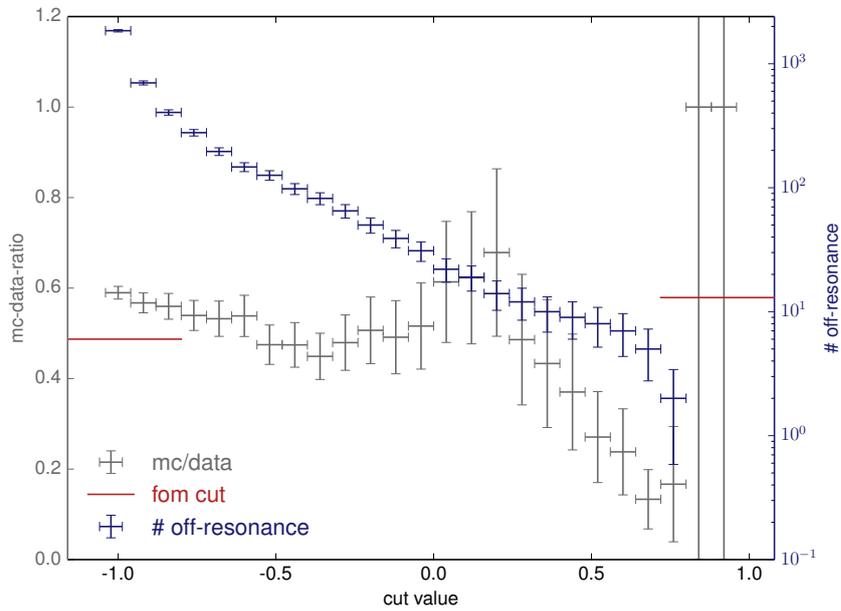
(b) $B^+ \to \pi^+ \nu\bar{\nu}$

Figure 8.5.: Data MC ratio dependent on cuts on $\mathscr{N}_{\text{sel}}$ (gray) and the overall amount of off-resonance events selected (blue). The values corresponding to the optimal selection are marked in red on both axes respectively.

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^+ \to \pi^+ \nu\bar{\nu}$
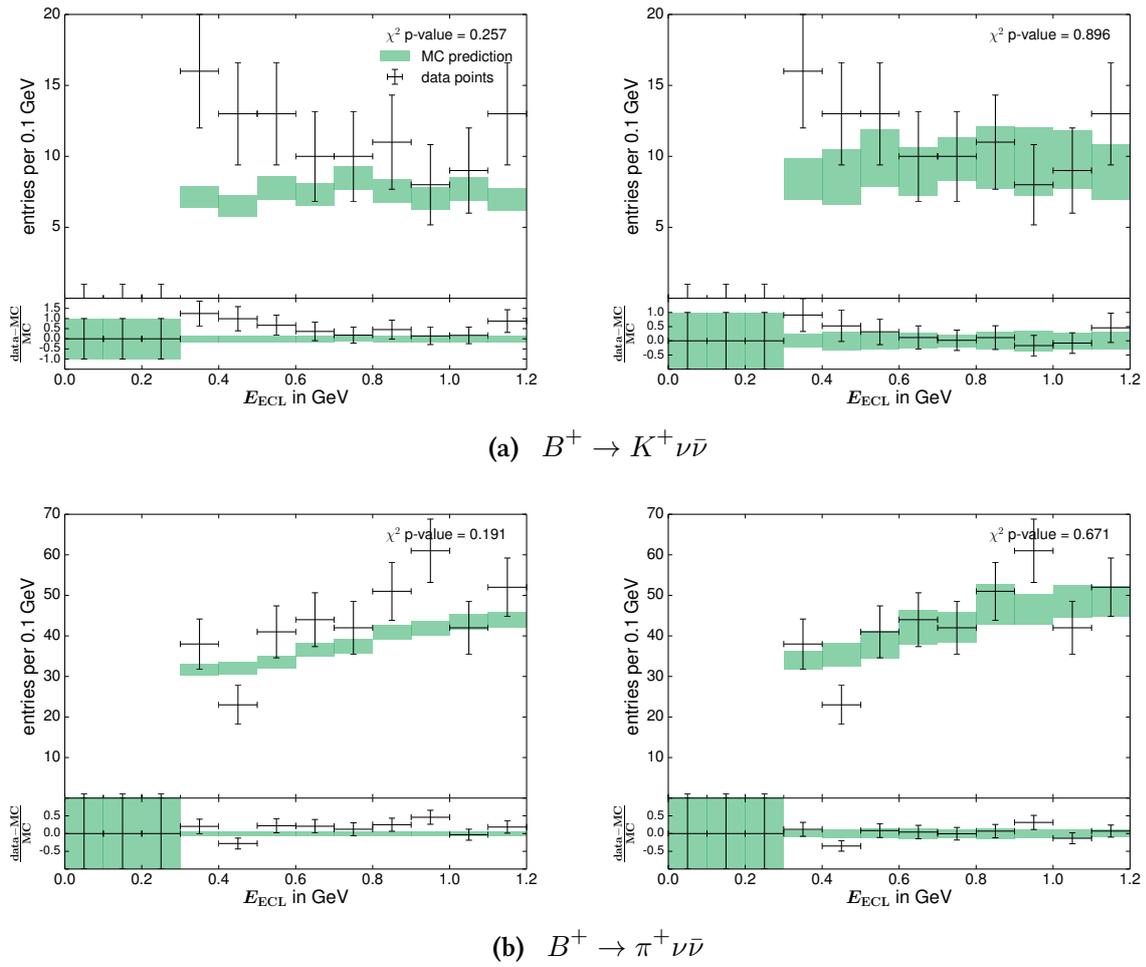
Figure 8.6.: E_{ECL} sideband distributions without reweighting (left) and with reweighted continuum simulations (right).

(a) $\cos \theta_{B, D^{(*)}l}$

(b) p_{cms}

(c) $\mathscr{N}_{\mathrm{NB, CS}}$
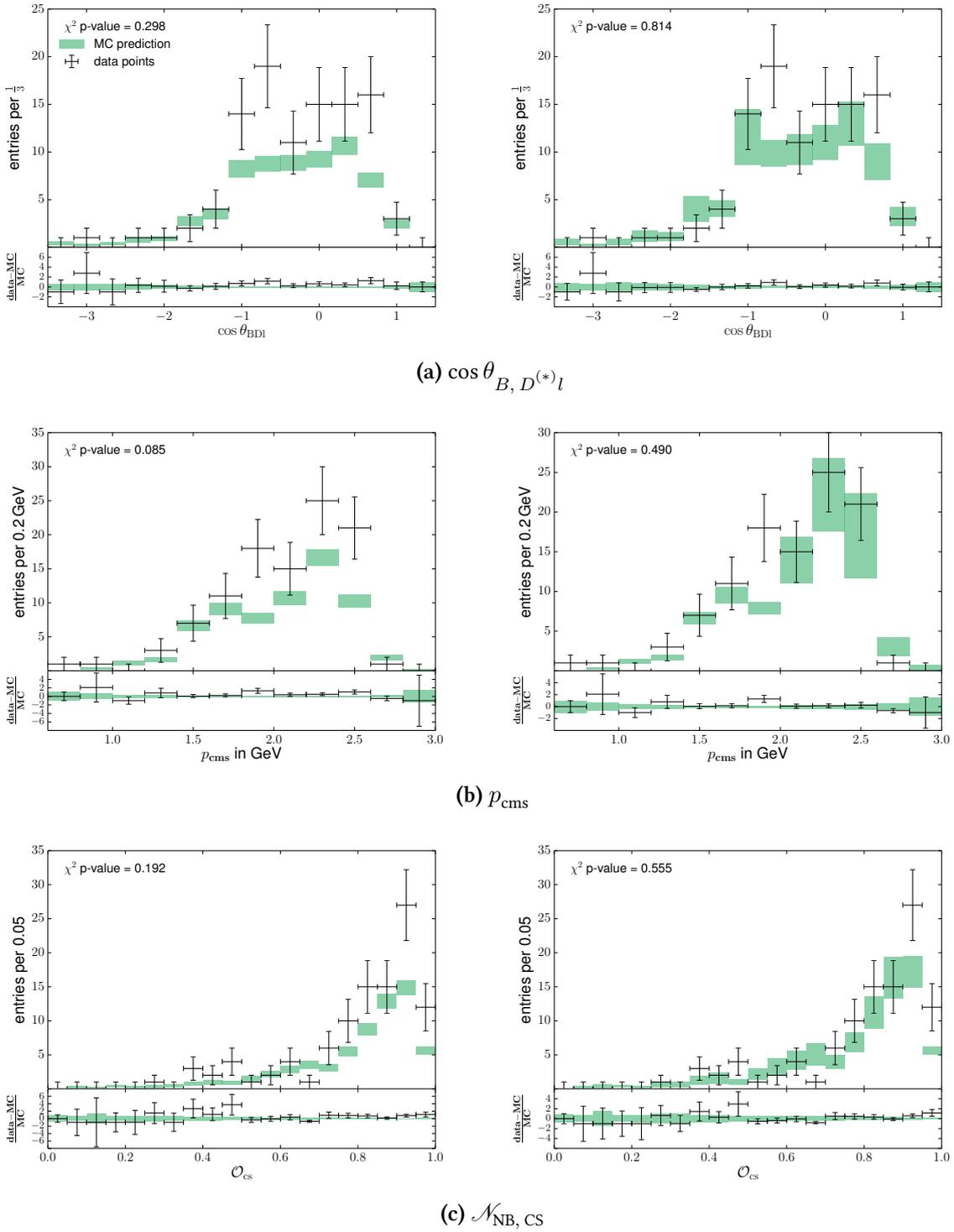
Figure 8.7.: Comparison between unweighted (left) and reweighted (right) sideband distributions in the $B^+ \to K^+ \nu \bar{\nu}$ channel.

8.3. $D^* l\nu$ channels

To further evaluate the E_{ECL} distribution for semileptonically tagged samples, tagged $B^0 \to D^{*-}l^+\nu_l$ and $B^+ \to \bar{D}^{*0}l^+\nu_l$ decays are reconstructed. D^{*+} candidates are reconstructed from $\bar{D}^0$ and π^+ candidates, $\bar{D}^{*0}$ candidates from $\bar{D}^0$ candidates and γ candidates. The $\bar{D}^0$ selection is the same as described in Section 6.3 on page 48, as well as the cuts on $\mathscr{N}_{\mathrm{NB,\,tag}}$ and $N_{\mathrm{remaining}}^{\mathrm{rawtracks}}$. The tag efficiency weights, determined in that same section, are applied as well as the K_{L}^0 veto. The additional selection cuts can be seen in Table 8.2, the resulting E_{ECL} distributions in Figure 8.8 on the following page. The cuts on the lepton momentum and the photon energy in the center of mass system of the $\Upsilon(4S)$ are set to $1.0\,\mathrm{GeV}$ and $200\,\mathrm{MeV}$ respectively due to large PID uncertainties for muons with low transverse momentum and combinatoric background for low energetic photons.

Although data differs slightly from the quantitative MC prediction, the disagreement between simulation and data is small enough to strengthen the trust in the simulations in general. Two things have to be kept in mind.

- The statistics in the $B^0 \to D^{*-}l^+\nu_l$ and $B^+ \to \bar{D}^{*0}l^+\nu_l$ control channels is considerably higher than in the $B \to h\nu\bar{\nu}$ signal channels. Whereas the number of observed events in the $B \to h\nu\bar{\nu}$ control channel amounts to 7608, the expected number of events in the $B^+ \to K^+\nu\bar{\nu}$ signal channel is only 83.6[1].

- The topologies of the decays differ. Whereas the signal side in the case of $B^+ \to K^+\nu\bar{\nu}$ channels consists only of one particle, $B^0 \to D^{*-}l^+\nu_l$ and $B^+ \to \bar{D}^{*0}l^+\nu_l$ channels require a well reconstructed D in addition to the γ, π and the lepton. Studies show that Ds are mostly correctly reconstructed, the D mass distributions in Figure 8.10 on page 88 exhibit negligible background from random combinations. Since the reconstruction of the $B \to h\nu\bar{\nu}$ channels requires only missing momentum and no additional particles, the background in the $B \to h\nu\bar{\nu}$ signal channels is too different from the $B^0 \to D^{*-}l^+\nu_l$ and $B^+ \to \bar{D}^{*0}l^+\nu_l$ channels where the selected sample consists almost exclusively of correct D and l candidates.

[1]This number does not include the scaling factor on the continuum background.

Table 8.2.: $B \to D^*l\nu_l$ selection criteria

channel	selection cut
combined	$p_{l,\,\mathrm{sig}}^{\mathrm{cms}} > 1.0\,\mathrm{Gev}$ $-4.5 < \cos\theta_{B,D^*l}^{\mathrm{sig}} < 1.5$ $-4.5 < \cos\theta_{B,D^*l}^{\mathrm{tag}} < 1.5$
$B^+ \to \bar{D}^{*0}l\nu_l$	$E_\gamma^{\mathrm{lab}} > 200\,\mathrm{MeV}$ $127\,\mathrm{MeV} < M_{D^*}\text{-}M_D < 153\,\mathrm{MeV}$
$B^0 \to D^{*-}l\nu_l$	$130\,\mathrm{MeV} < M_{D^*}\text{-}M_D < 160\,\mathrm{MeV}$

(a) $B^+ \to \bar{D}^{*0} l^+ \nu_l$ **(b)** $B^0 \to D^{*-} l^+ \nu_l$
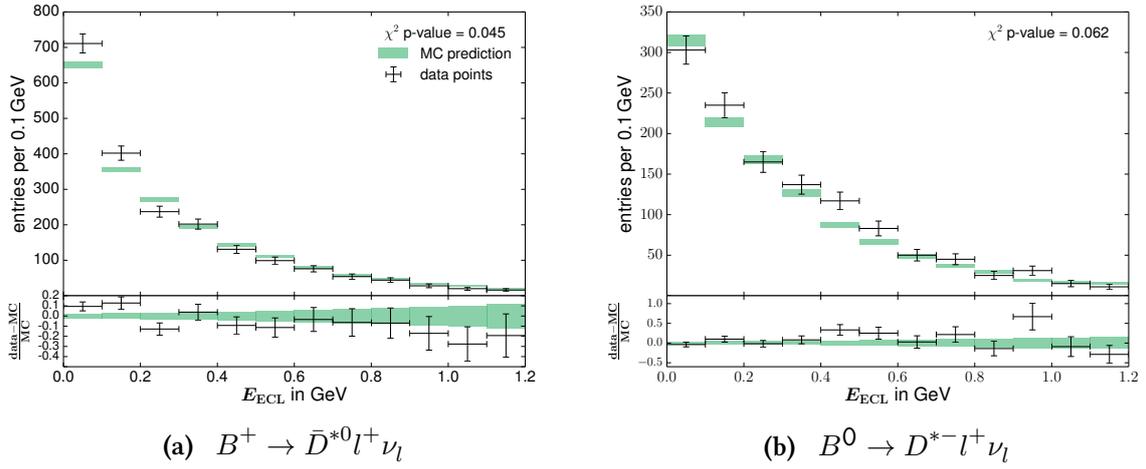
Figure 8.8.: E_{ECL} distribution for the $B \to D^* l \nu_l$ control channels.

(a) $B^+ \to \bar{D}^{*0} l^+ \nu_l$ **(b)** $B^0 \to D^{*-} l^+ \nu_l$
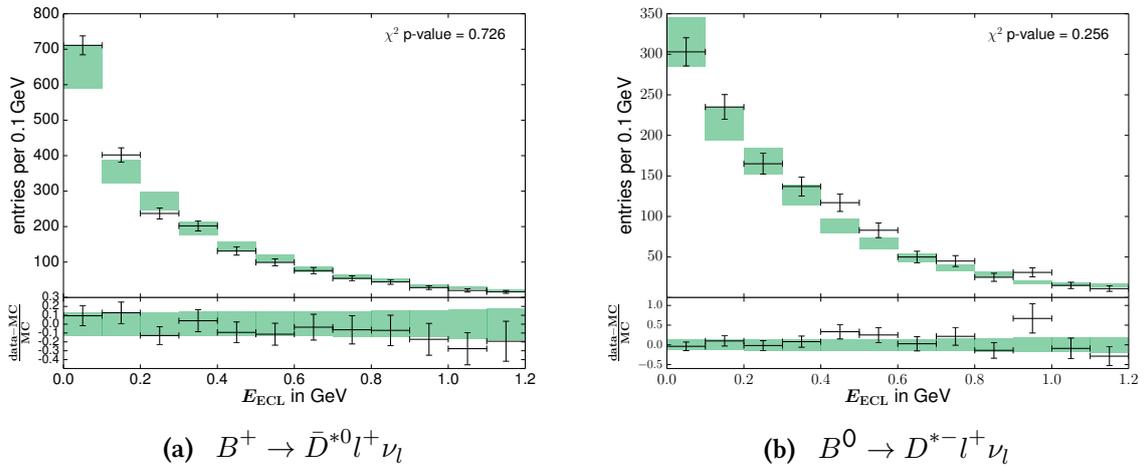
Figure 8.9.: E_{ECL} distribution for the $B \to D^* l \nu_l$ control channels, the uncertainty on the tag correction is included.

To check whether $B \rightarrow D^{**}l\nu$ pollution could lead to data-MC differences, the $\cos\theta_{B,D^*l}$ distributions for neutral and charged B mesons are plotted in Figure 8.11 on the following page. The only noteworthy deviation of the prediction from the data occurs in the unphysical region of $\cos\theta_{B,D^*l} > 1$. We conclude that the $B \rightarrow D^*l\nu$ sample is well reconstructed and understood.

Figure 8.9 on the preceding page shows the same data as Figure 8.8 on the facing page but with the systematic uncertainty on the tag side correction added to the pure statistical uncertainty of the first plot. The systematic uncertainty is calculated in the following manner: The uncertainty on the correction factor calculated in Section 6.3 on page 48 is added to the factor and the plot thereafter repeated with the new factor. The difference in each bin is taken as the systematic uncertainty of that bin in E_{ECL}. Since the correction factors are close enough to one, the uncertainty can be assumed to be symmetrical. This has been confirmed to be the case. This study shows that the data for the $B \rightarrow D^*l\nu$ control channels is compatible with the simulations within the uncertainties of the constraints of this analysis. Hence, neither tagging nor E_{ECL} calculation introduces any additional uncertainties to our signal samples other than those already accounted for.

(a) $B^+ \to \bar{D}^{*0}l^+\nu_l$

(b) $B^0 \to D^{*-}l^+\nu_l$
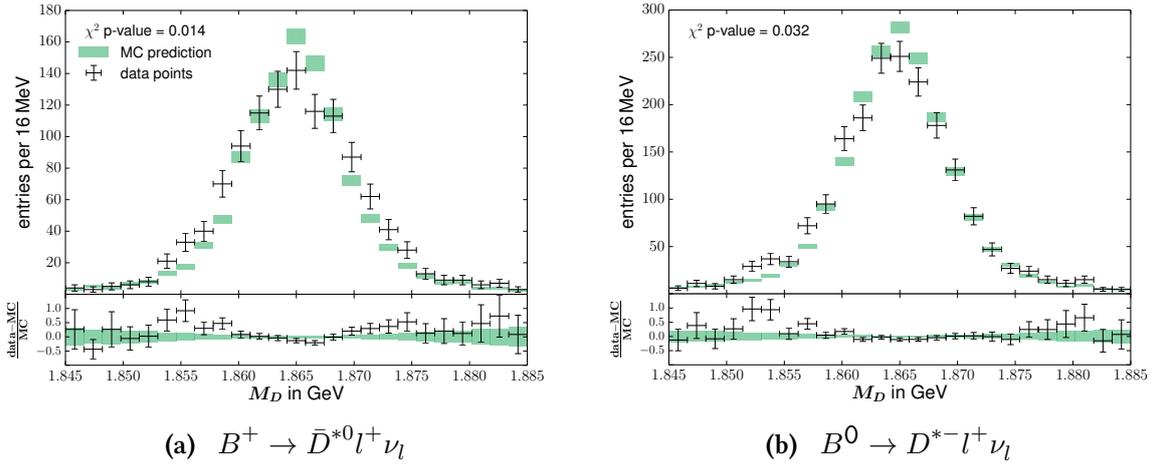
Figure 8.10.: Data MC comparison for the M_D distributions of the signal side.

(a) $B^+ \to \bar{D}^{*0}l^+\nu_l$

(b) $B^0 \to D^{*-}l^+\nu_l$
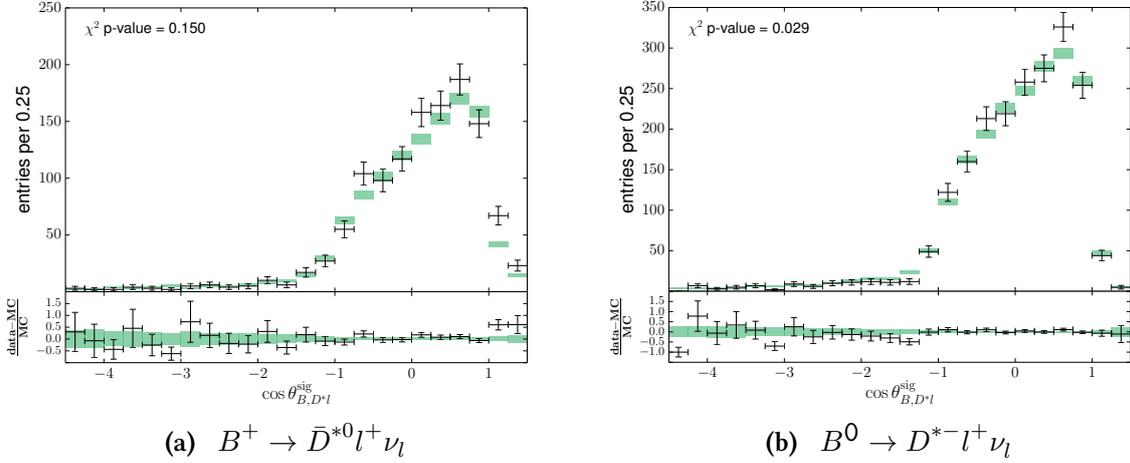
Figure 8.11.: Data MC comparison for the $\cos\theta_{B,D^*l}$ distributions of the signal side.

Chapter 9

Branching Fraction Extraction

9.1. The Model

The shapes of the signal distribution and the various background distributions are taken from Monte Carlo simulations. The signal and backgrounds are modelled via histogram templates. The relative fractions of the different background components are fixed to their respective value in the Monte Carlo simulation[1]. Generic and continuum background are taken from ten and six streams of generic Monte Carlo respectively, for rare background the full available sample is utilized. All three background components are weighted according to the full integrated luminosity.

9.1.1. Signal Yield Fits

To extract any eventual observable signal, a binned extended Maximum Likelihood fit is performed. The signal yield and overall background yield are left as free floating parameters. The likelihood of the fit model is given by the following expression

$$\mathcal{L} = \frac{n^N e^{-n}}{N!} \prod_{k=1}^{\#\text{bins}} \frac{n_k^{N_k} e^{-n_k}}{N_k!} \tag{9.1}$$

$$\text{with} \quad n_k = \sum_i n_i^k \mathcal{P}_i^k \quad \text{and} \quad n = \sum_k n_k$$

where N is the total number of observed events, n_i and $\mathcal{P}_i$ the yield and PDF of the i-th component, respectively, and N_k the number of observed events in the k-th bin. Since the amount of expected events is comparably low while background can be suppressed well, see Table 7.1 on page 60, the overall amount of expected events, especially in the $B^+ \to K^{*+} \nu \bar{\nu}$ channel is very low, see

---

[1]Where all weights are applied as described in previous chapters.

Table 7.1 on page 60. This can lead to instabilities in the fit and the calculation of the limit. A careful treatment of the initial values and boundaries of the free parameters of the fit is thus necessary. The algorithm for instance has to ensure that $\mathcal{P}_{\text{sig}}(E_{\text{ECL}}) + \mathcal{P}_{\text{bkg}}(E_{\text{ECL}}) > 0$ at any point because of the positive definiteness of PDFs. Since the total amount of data is a random variable itself, extensive toy studies have been performed, see Section 9.2 on page 93.

The significance of the observation is determined by fitting the data under the background only hypotheses B and signal plus background hypotheses $S+B$ via the following equation derived from Wilk's theorem.

$$S_{\text{obs}} = \sqrt{2\log\frac{\mathcal{L}_{S+B}}{\mathcal{L}_B}} \tag{9.2}$$

Examples of such fits are displayed in Figure 9.1 on the next page. The fits are carried out on a toy sample drawn from the final fit model. The simulated number of background events equals the number expected from Monte Carlo simulations while the number of signal events was set to zero.

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K_S^0 \nu\bar{\nu}$

(c) $B^+ \to K^{*+} \nu\bar{\nu}$

(d) $B^0 \to K^{*0} \nu\bar{\nu}$

(e) $B^+ \to \pi^+ \nu\bar{\nu}$
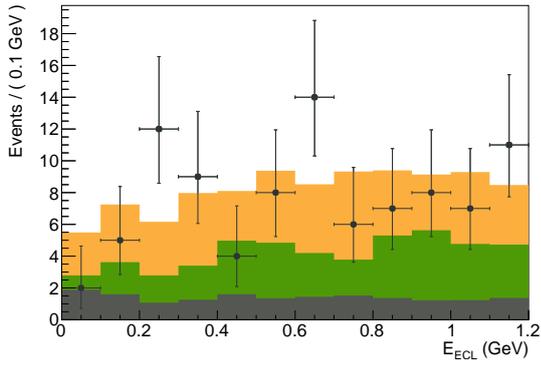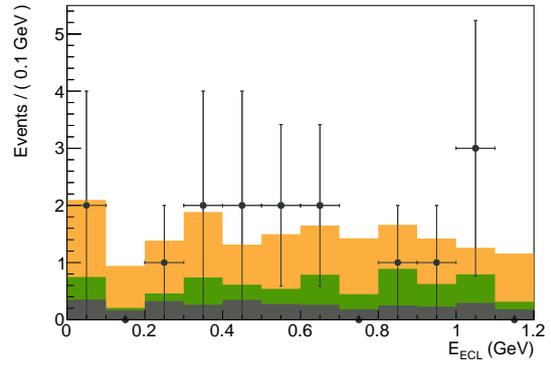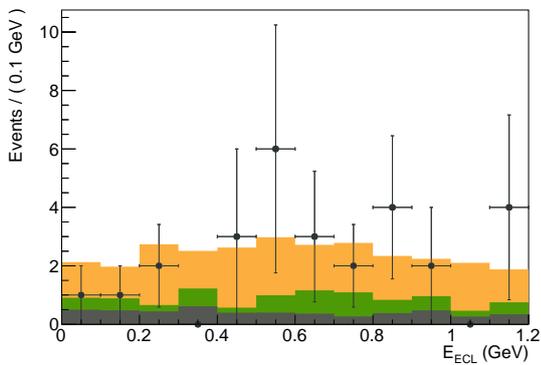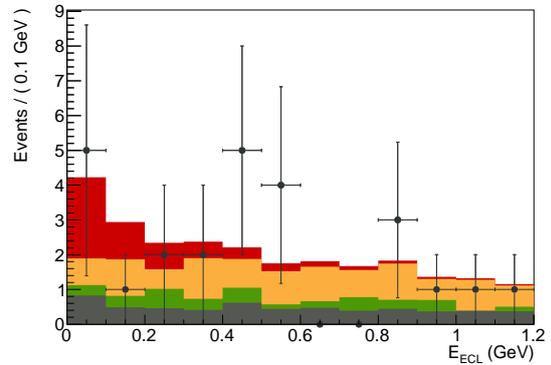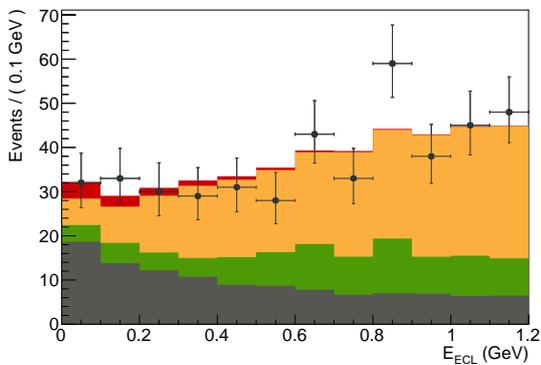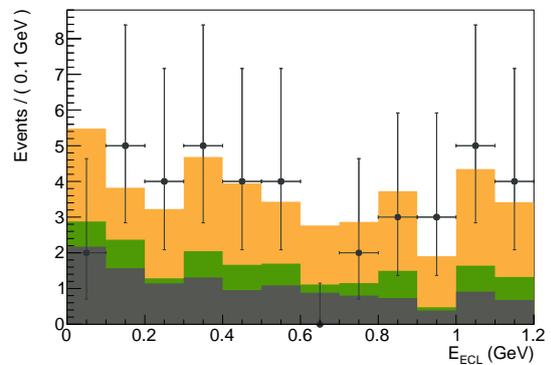
(f) $B^0 \to \pi^0 \nu\bar{\nu}$
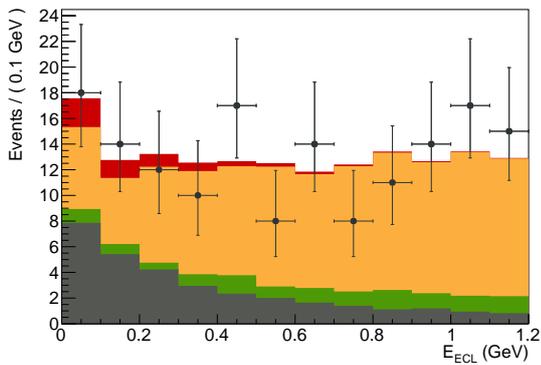
(g) $B^+ \to \rho^+ \nu\bar{\nu}$
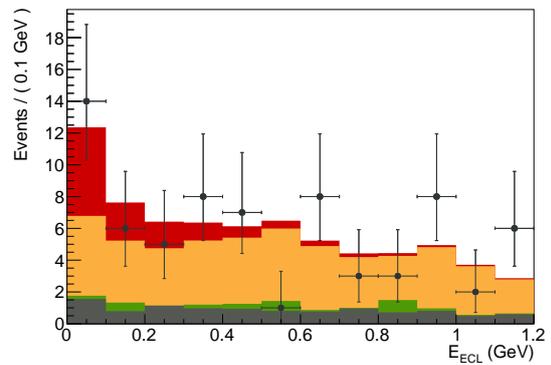
(h) $B^0 \to \rho^0 \nu\bar{\nu}$

Figure 9.1.: Fits on toy samples without simulated signal component.

9.1.2. Limit Calculation

Since this analysis is a search for yet unobserved decays, we don't know whether we will be able to see any signal. We consequently prepare for setting upper limits on the branching fractions of $B \to h \nu \bar{\nu}$ decays in case we don't observe evidence. To obtain an upper limit on the number of signal events at a given confidence level α, a Bayesian ansatz is implemented. The profiled likelihood ratio is integrated up to the point where the integral equals α times the integral over the whole range, where α corresponds to the confidence level. This procedure is equivalent to a flat positive prior.

$$\alpha \times \int_0^\infty \mathcal{L}\left(N_\text{sig}\right) \mathrm{d}N_\text{sig} = \int_0^{N_\text{limit}} \mathcal{L}\left(N_\text{sig}\right) \mathrm{d}N_\text{sig} \tag{9.3}$$

$$\mathcal{L}\left(N_\text{sig}\right) = \max_{N_\text{bkg}} \left(\mathcal{L}\left(N_\text{sig}, N_\text{bkg}\right)\right) \tag{9.4}$$

Here N_limit is the number of observed signal events excluded at the chosen confidence level, N_sig and N_bkg the signal and background yield, respectively, and $\mathcal{L}$ the likelihood ratio, i.e. the likelihood divided by the likelihood at the best fitting point. The background yield is set as a nuisance parameter of the likelihood, i.e. each point is reoptimized for the respective fixed signal yield, as is described in Equation (9.4). This is necessary since signal and background yield are correlated in the fit.

9.1.3. Branching fraction calculation

After the calculation of the limit on the number of signal events compatible with the observed data, as described in Equation (9.3), the corresponding branching fraction of the respective $B \to h \nu \bar{\nu}$ channel is calculated via the following equation.

$$\mathcal{B}^i = \frac{N_\text{limit}^i}{N_{B\bar{B}} \times \varepsilon_\text{rec}^i} \tag{9.5}$$

Here, N_limit is the obtained limit on the number of signal events, $N_{B\bar{B}}$ is the overall number of B^+B^-, $B^0\bar{B}^0$ events respectively[2], and ε_rec is the reconstruction efficiency determined on signal Monte Carlo. The index i denominates the respective channel.

---

[2]The $\Upsilon(4S)$ is assumed to decay into pairs of neutral and charged B mesons in equal rates

9.2. Toy Tests

In order to probe the method described in detail in Section 9.1 on page 89, various toy sets have been generated out of the histogram templates. For Each channel[3], a thousand datasets have been simulated from each the background distributions obtained from generic, continuum and rare MC. The normalization of the overall background is left as a free parameter following a Poisson distribution with the Monte Carlo expectation as the expected value of the distribution. This way, we include the uncertainty on the expected number of background events in the toy studies.

9.2.1. Expected Limits

On each of those signal free datasets a limit is calculated according to the method described in Section 9.1.2 on the preceding page. The obtained limits are averaged for the determination of the expected limit. The distribution of those limits can be observed in Figure 9.3 on page 95. It is noticeable that the resulting distribution of the obtained limits exhibits an asymmetric shape and a comaparable large variance and skewness. Due to the low statistics expected, and observed in the E_{ECL} sideband, see section 8.1 on page 77, in most channels, upward fluctuations of the background can have a big impact on the observed limit of a given toy sample. The influence of larger upward fluctuations of the background content in the first bin in E_{ECL} is considerable since it causes much larger exclusion limits for the signal process. The peaking rare component, although theoretically well understood, in the $B^+ \to \pi^+ \nu\bar{\nu}$ and $B^+ \to \rho^+ \nu\bar{\nu}$ channels causes a similar effect for both those channels with higher statistics.

The resulting distributions of the limits obtained in the toy studies are displayed in figure 9.3 on page 95. The mean and median of each distribution is marked as an estimator of the expected limit. From those statistical quantities, we expect the limits on each of the measured decay channels to be the best one to date. Although the applied methods yield a big improvement compared to previous work, we will most likely still not be sensitive to the predicted SM value. A comparison between the limits obtained in this work, previous results and theoretical predictions is plotted in Figure 9.2 on the next page.

---

[3] $B^+ \to K^{*+} \left(\to K^+ \pi^0 \right) \nu\bar{\nu}$ and $B^+ \to K^{*+} \left(\to K^0_{\mathrm{S}} \pi^+ \right) \nu\bar{\nu}$ channels are merged and channel in this chapter refers to the merged $B^+ \to K^{*+} \nu\bar{\nu}$ channel and not to its separately reconstructed subchannels.

Table 9.1.: Mean and median of the expected limits calculated on 1000 simulated toy distributions.

Channel	mean limit $\left[10^{-5}\right]$	median limit $\left[10^{-5}\right]$
$B^+ \to K^+ \nu\bar{\nu}$	0.85	0.80
$B^0 \to K^0_S \nu\bar{\nu}$	1.23	1.16
$B^+ \to K^{*+} \nu\bar{\nu}$	2.37	2.20
$B^0 \to K^{*0} \nu\bar{\nu}$	2.35	2.19
$B^+ \to \pi^+ \nu\bar{\nu}$	1.30	1.24
$B^0 \to \pi^0 \nu\bar{\nu}$	1.00	0.94
$B^0 \to \rho^0 \nu\bar{\nu}$	2.19	2.01
$B^+ \to \rho^+ \nu\bar{\nu}$	2.47	2.28
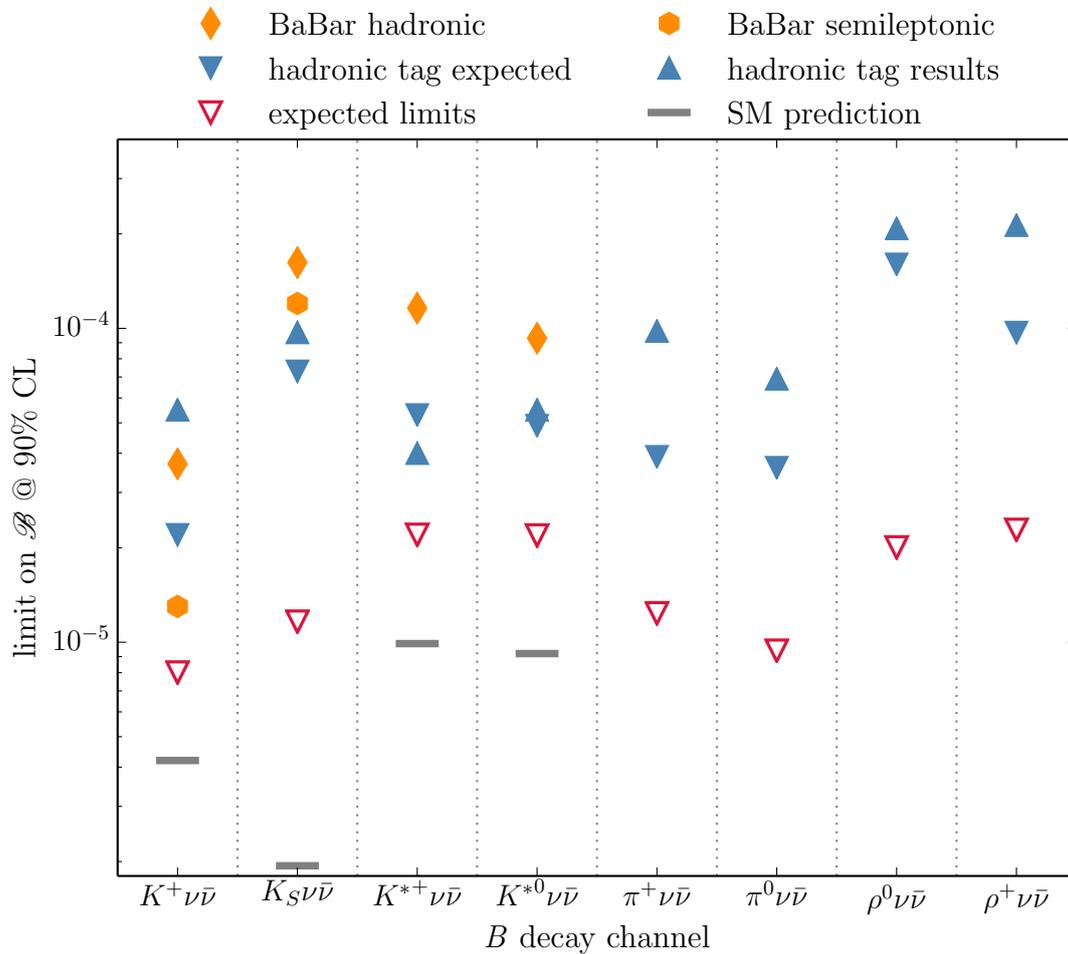
Figure 9.2.: Expected Limits (median) for this analysis compared to previous results. Theoretical values for $b \to d\nu\bar{\nu}$ channels are not plotted because the suppression compared to the $b \to s\nu\bar{\nu}$ channels and the connected compression of the y-axis.

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K_S^0 \nu\bar{\nu}$

(c) $B^+ \to K^{*+} \nu\bar{\nu}$

(d) $B^0 \to K^{*0} \nu\bar{\nu}$

(e) $B^+ \to \pi^+ \nu\bar{\nu}$

(f) $B^0 \to \pi^0 \nu\bar{\nu}$

(g) $B^+ \to \rho^+ \nu\bar{\nu}$

(h) $B^0 \to \rho^0 \nu\bar{\nu}$
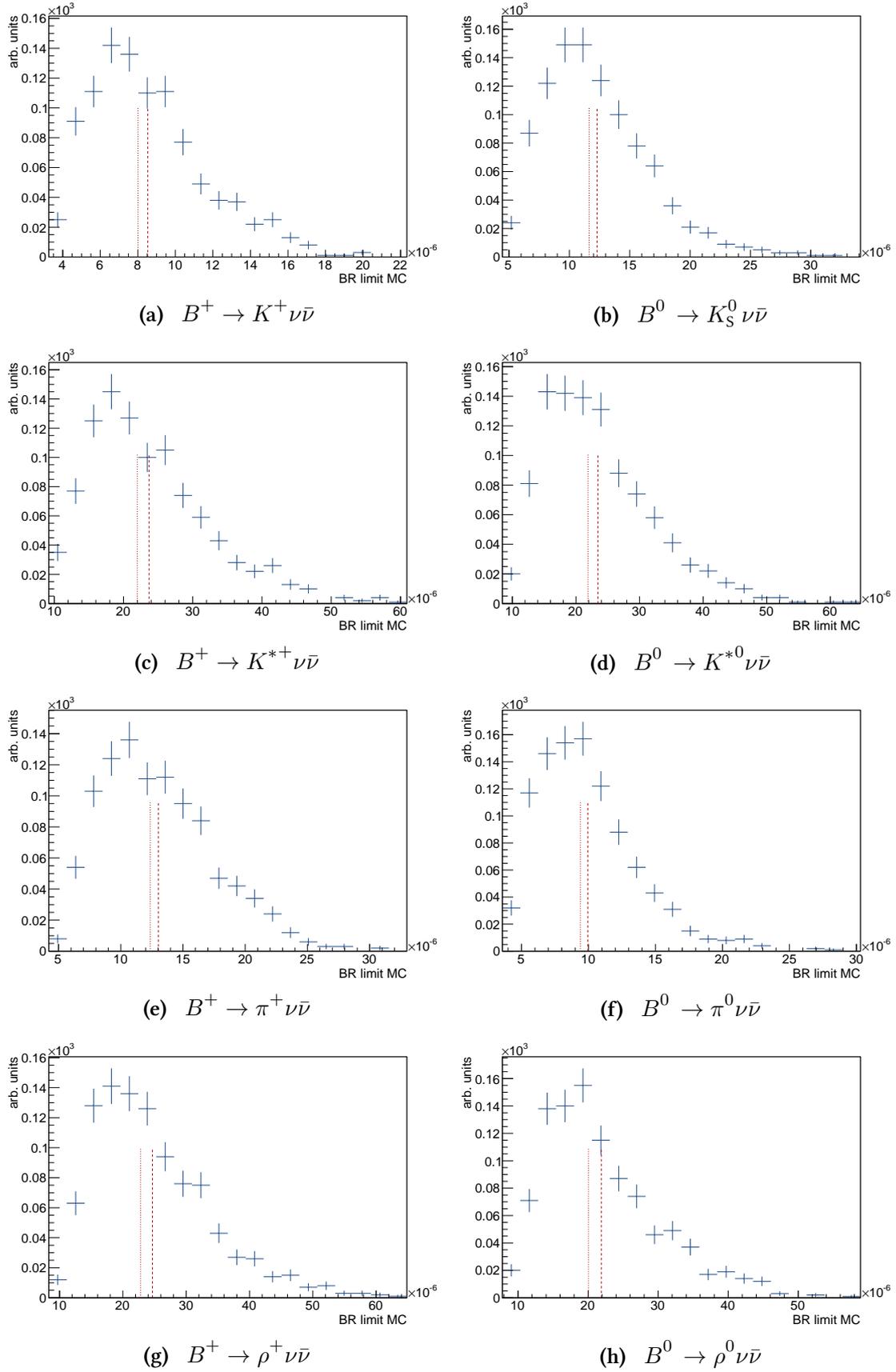
Figure 9.3.: Distribution of limits on the branching fraction of the respective channel from toys; each plot contains 1000 events. The mean is marked with a dashed, the median with a dotted line.

Table 9.2.: Ratio of fits on generated samples with $N_{\text{sig}} = 0$ yielding a fitted value of N_{sig} smaller than the limit calculated on a sample with the same estimated number of backgrounds, averaged over 1000 simulated toy samples.

channel	ratio of fits with $N_{\text{sig}}^{\text{fit}} < N_{\text{sig}}^{\text{limit}}$
$B^+ \to K^+ \nu\bar{\nu}$	92.2 %
$B^0 \to K_S^0 \nu\bar{\nu}$	93.7 %
$B^0 \to K^{*0} \nu\bar{\nu}$	92.8 %
$B^+ \to K^{*+} \nu\bar{\nu}$	93.8 %
$B^+ \to \pi^+ \nu\bar{\nu}$	92.4 %
$B^0 \to \pi^0 \nu\bar{\nu}$	92.9 %
$B^+ \to \rho^+ \nu\bar{\nu}$	92.3 %
$B^0 \to \rho^0 \nu\bar{\nu}$	92.8 %

9.2.2. Linearity Tests

In case a sizeable signal contribution is observed, it is checked whether the fitting procedure reproduces the correct results. For this purpose, 21000 distributions containing zero to twenty signal events are simulated and fitted for each channel except the $B^+ \to \pi^+ \nu\bar{\nu}$ ($B^+ \to \rho^+ \nu\bar{\nu}$) channel where 41000 31000 samples have been simulated and fitted. The function of the mean fitted values for each input signal strength is then fitted with a first order polynomial. Most channels in fact exhibit a significant bias, as can be seen in Figure 9.4 on page 99 which in turn is linear and therefore is corrected for with the parameters taken from this study according to the following equation. The numerical values of the fitted slope and intercept are listed in Table **??** on page **??**.

$$N_{\text{sig}}^{\text{corr}} = \frac{N_{\text{sig}}^{\text{fit}} - N^{(0)}}{N^{(1)}} \tag{9.6}$$

Here, $N_{\text{sig}}^{\text{fit}}$ is the fitted number of signal events, $N^{(0)}$ and $N^{(1)}$ are the coefficients of the fitted polynomial. The same formula is applied to calculated branching fractions.

To check whether the expected limit exhibits any bias, we simulate 1000 background distributions for each limit from the toy sample with a fixed background number obtained from the fit to the original toy sample. Subsequently, we fit the combined model to that distribution and count how many events exhibit a fitted signal yield smaller than the limit calculated on the original toy sample. The fit bias on the signal yield is accounted for hereby. The results of said study are displayed in Table 9.2. We conclude from the second column of that table that the expected limits are not estimated too optimistically but rather conservatively instead.

9.2.3. Probability to Find Evidence

Finally, the probability to find a signal in each given channel is evaluated. For this purpose, 10000 toy samples with the expected number of background and the expected number of signal calculated from the theory prediction of the respective branching fraction [BGNNS15, HKR15] have been generated. The fit is repeated and the number of times the significance of the fitted signal yield, see Equation (9.2) on page 90, is higher than $3\,\sigma$ is counted. The results are displayed in Table 9.3 on the following page. The probabilities to find evidence for $b \to d\nu\bar{\nu}$ SM processes is basically zero, as expected. The kaonic channels on the other hand will most likely contain real signal events. However, the probability to find sufficient signal to claim evidence for a $b \to s\nu\bar{\nu}$ decay in the course of this analysis is rather small, though not entirely negligible in the case of the $B^+ \to K^+\nu\bar{\nu}$ and the $B^0 \to K^{*0}\nu\bar{\nu}$ channels. This last statement is true for the values in the quoted reference which exhibit a theoretical uncertainty themselves. Hence, the study was repeated for an SM expectation value increased by its standard deviation. This study gives a strong hint that the first evidence of $B \to h\nu\bar{\nu}$ decays needs the increased data set of Belle II but does not render the prospect of such evidence for the SM process impossible.

Table 9.3.: Expected number of signal events and probability to find evidence for a signal (p_{evidence}), assuming the standard model calculation from [BGNNS15] for the $b \rightarrow s\nu\bar{\nu}$ channels and [HKR15] for the $b \rightarrow d\nu\bar{\nu}$ channels.

Channel	expected signal yield	p_{evidence} in %
$B^+ \rightarrow K^+ \nu\bar{\nu}$	7.0	1.5
$B^0 \rightarrow K^0_{\text{S}} \nu\bar{\nu}$	1.5	0.3
$B^+ \rightarrow K^{*+} \nu\bar{\nu}$	4.4	1.2
$B^0 \rightarrow K^{*0} \nu\bar{\nu}$	3.9	1.0
$B^+ \rightarrow \pi^+ \nu\bar{\nu}$	0.5	0.1
$B^0 \rightarrow \pi^0 \nu\bar{\nu}$	0.3	0.1
$B^0 \rightarrow \rho^0 \nu\bar{\nu}$	0.4	0.1
$B^+ \rightarrow \rho^+ \nu\bar{\nu}$	0.5	0.1

(a) `prediction`

Channel	expected signal yield	p_{evidence} in %
$B^+ \rightarrow K^+ \nu\bar{\nu}$	7.7	2.1
$B^0 \rightarrow K^0_{\text{S}} \nu\bar{\nu}$	1.7	0.3
$B^+ \rightarrow K^{*+} \nu\bar{\nu}$	4.7	1.2
$B^0 \rightarrow K^{*0} \nu\bar{\nu}$	4.2	1.8
$B^+ \rightarrow \pi^+ \nu\bar{\nu}$	0.6	0.1
$B^0 \rightarrow \pi^0 \nu\bar{\nu}$	0.3	0.1
$B^0 \rightarrow \rho^0 \nu\bar{\nu}$	0.4	0.1
$B^+ \rightarrow \rho^+ \nu\bar{\nu}$	0.5	0.1

(b) `prediction plus one` σ

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K^0_S \nu\bar{\nu}$

(c) $B^+ \to K^{*+} \nu\bar{\nu}$

(d) $B^0 \to K^{*0} \nu\bar{\nu}$

(e) $B^+ \to \pi^+ \nu\bar{\nu}$

(f) $B^0 \to \pi^0 \nu\bar{\nu}$

(g) $B^+ \to \rho^+ \nu\bar{\nu}$

(h) $B^0 \to \rho^0 \nu\bar{\nu}$
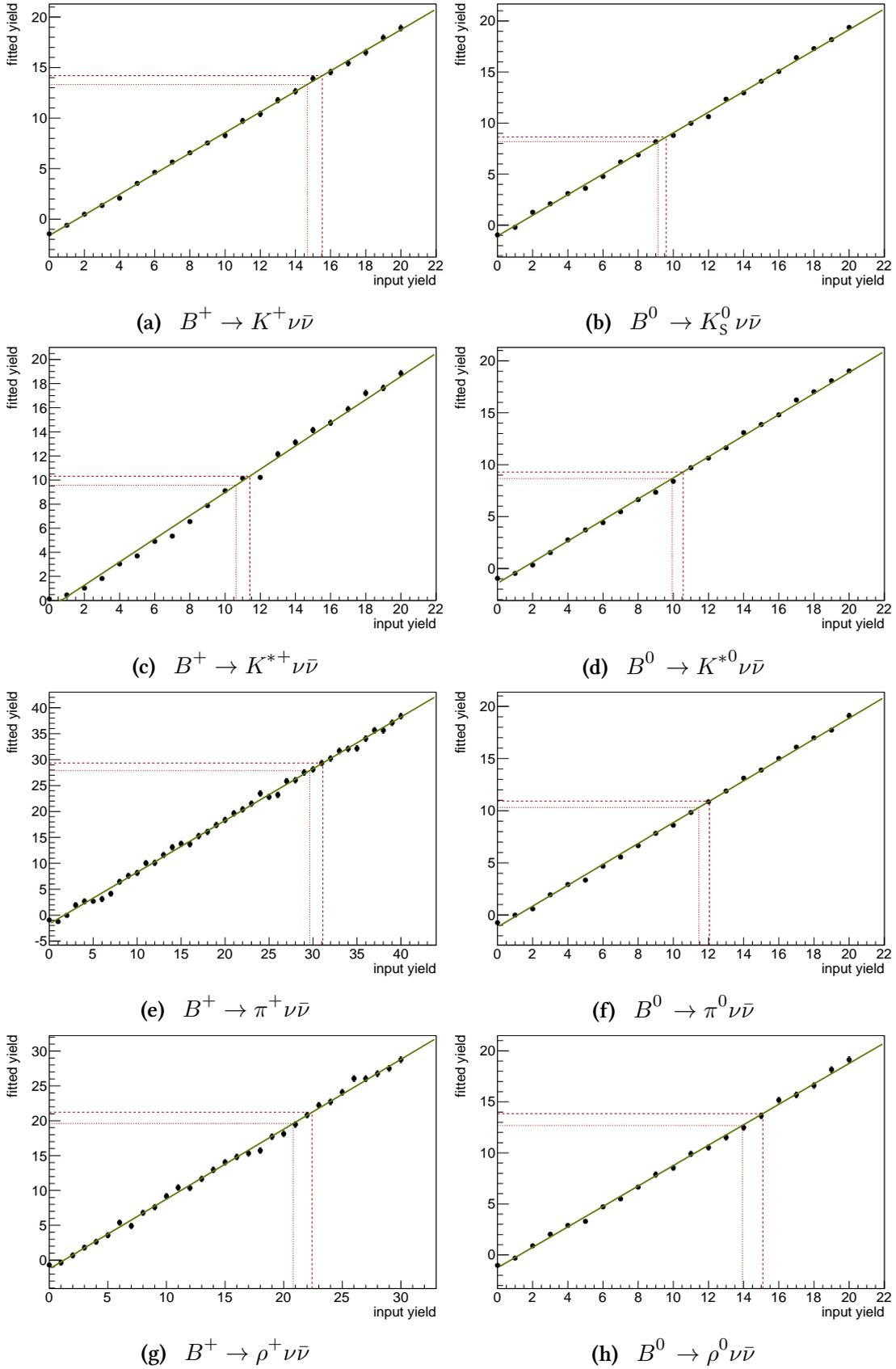
Figure 9.4.: Fit bias tests for the signal yield extraction. The positions of the mean expected limit (dashed) and of the median expected limit (dotted) are marked as an example of how a fit result would be corrected.

Chapter 10

Results

10.1. Fitted Branching Ratios

No significant signal is observed in any reconstructed channel. Table 10.1 on page 103 summarizes the results obtained on data, Figure 10.2 on page 105 shows the fits on all surveyed channels. The fit results are in good agreement with the MC expectation without signal. Only the $B^+ \to K^+ \nu\bar{\nu}$ and $B^+ \to K^{*+} \nu\bar{\nu}$ channels exhibit a deviation from the $N_{\text{sig}} = 0$ hypothesis of $2\,\sigma$ or more. An example of an event contributing to the first bin in E_{ECL} on data is depicted in Figure 10.1 on page 103. However, even in those two channels, the significance of the signal does not exceed $3\,\sigma$. The statistical error is due to the low statistics rather high throughout all channels. We compare the data and the fitted model for the final classifier output $\mathcal{N}_{\text{sel}}$ and important input variables for channels where the observed significance is greater than $1\,\sigma$ in Appendix E. None of the important variables shows a data-background model tension that can only be explained by the presence of signal.

However, we observe a concentration of continuum in the $B^+ \to K^+ \nu\bar{\nu}$ channel for higher values of $\cos\theta_{\text{thrust}}$, see Figure E.1f on page 146. The thrust axis enters the continuum suppression and the final selection net. The net suppresses the high $\cos\theta_{\text{thrust}}$ region because of the comparable high continuum contribution. This in turn depletes the region of generic and rare background and the sum yields an overall flat distribution. This indicates that first, the $\cos\theta_{\text{thrust}}$ variable could have been used to cut away continuum background peaking at higher $\mathcal{N}_{\text{NB, CS}}$ values, and second that the effect does not manifest itself in the overall background-data comparison performed on the sideband sample. This said, it must also be emphasized that the weighting procedure of the continuum MC makes the background expectation for the high $\cos\theta_{\text{thrust}}$ region fit the data very well. From this discussion, we draw the conclusion that a cut could have solved the problem of imprecise continuum equally well. The cut would have resulted in a lower efficiency and the uncertainty on the weighting procedure is accounted for as a systematic uncertainty, see chapter 11 on page 109. Hence the influence this different approach would have on the final result is well covered by the systematic uncertainties.

In Table 10.2 on page 104, we also compare the fitted and expected background yields. All results are compatible within the statistical uncertainty of the fit with MC expectations. However, a general trend towards marginally higher values of the measured background yield as compared to the expected yield is visible. This effect might be a systematic effect from the tag efficiency reweighting of the MC because the correction is $100\,\%$ correlated between all charged and all neutral channels, respectively. The cuts and reweighting procedures result in a sound description of the respective background within the precision of this analysis. We conclude that we understand the backgrounds sufficiently enough for the precision of our measurement. This strengthens the confidence in our model and result.

Figure 10.3 on page 106 shows the results of a fit to data on each of the $B^+ \to K^{*+}\nu\bar{\nu}$ subchannels. Most of the signal contribution observable in Figure 10.2c on page 105 emanates from the $K^{*+} \to K^0_S\pi^+$ subchannel. This study rules out the possibility that the observed excess results from mismodelled random π^0s or other non-considered data-MC differences since the reconstructed K^0_S sample is very pure.

In a nutshell, all results can be explained within the boundaries of the standard model without the contribution of $B \to h\nu\bar{\nu}$ decays.
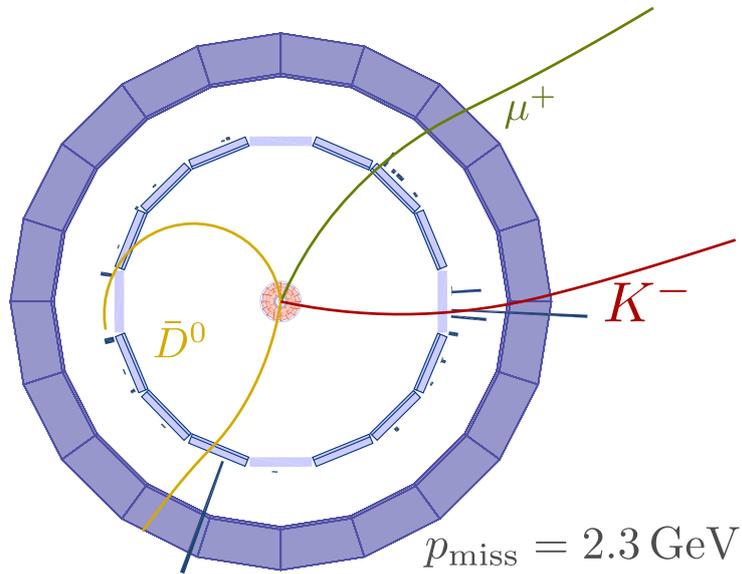
Figure 10.1.: Event with a high $\mathscr{N}_{\mathrm{sel}}$ and low E_{ECL} in the $B^+ \to K^+ \nu\bar{\nu}$ channel.

Table 10.1.: Overview over the results of the fit to data. The uncertainty is statistical only.

channel	mean $\mathscr{B}\ \left[10^{-5}\right]$	significance $[\sigma]$
$B^+ \to K^+ \nu\bar{\nu}$	1.1 ± 0.5	2.0
$B^0 \to K^0_{\mathrm{S}} \nu\bar{\nu}$	0.1 ± 0.6	0.0
$B^+ \to K^{*+} \nu\bar{\nu}$	3.7 ± 1.7	2.3
$B^0 \to K^{*0} \nu\bar{\nu}$	-0.4 ± 3.7	0.0
$B^+ \to \pi^+ \nu\bar{\nu}$	0.2 ± 0.7	0.3
$B^0 \to \pi^0 \nu\bar{\nu}$	0.0 ± 0.5	0.0
$B^+ \to \rho^+ \nu\bar{\nu}$	0.7 ± 1.4	0.4
$B^0 \to \rho^0 \nu\bar{\nu}$	1.9 ± 1.4	1.3

Table 10.2.: Expected and fitted background yields for all channels, the uncertainties are statistical only.

channel	fitted bkg yield	expected bkg yield
$B^+ \to K^+ \nu\bar\nu$	115.6 ± 13.5	103.6
$B^0 \to K_S^0 \nu\bar\nu$	25.4 ± 6.6	22.4
$B^+ \to K^{*+} \nu\bar\nu$	38.8 ± 8.8	35.2
$B^0 \to K^{*0} \nu\bar\nu$	26.9 ± 4.8	24.8
$B^+ \to \pi^+ \nu\bar\nu$	481.1 ± 26.5	474.0
$B^0 \to \pi^0 \nu\bar\nu$	50.3 ± 8.9	41.0
$B^+ \to \rho^+ \nu\bar\nu$	184.0 ± 18.1	172.5
$B^0 \to \rho^0 \nu\bar\nu$	65.3 ± 11.5	62.4

Fit Components
signal   generic
continuum   rare

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K^0_S \nu\bar{\nu}$

(c) $B^+ \to K^{*+} \nu\bar{\nu}$

(d) $B^0 \to K^{*0} \nu\bar{\nu}$

(e) $B^+ \to \pi^+ \nu\bar{\nu}$
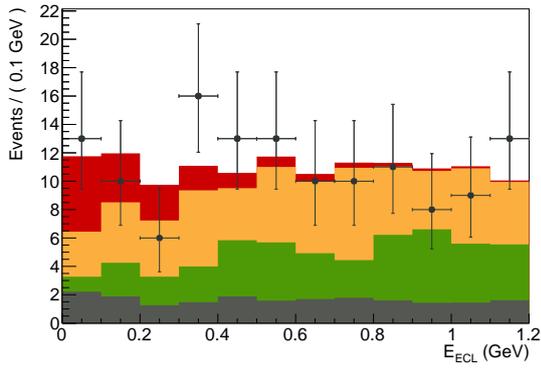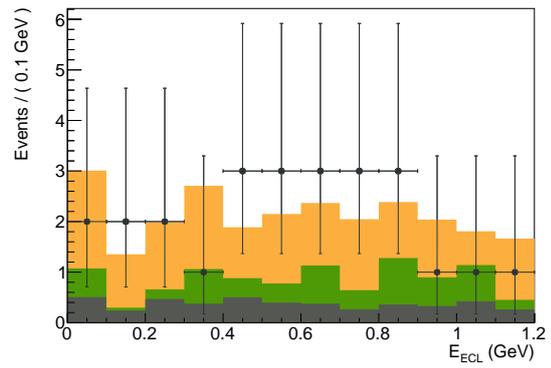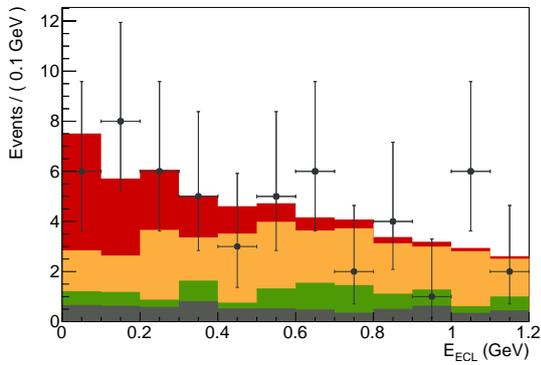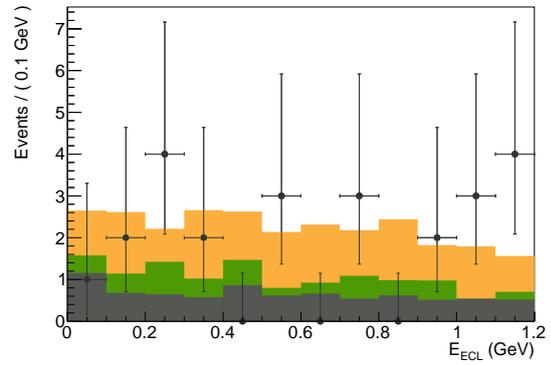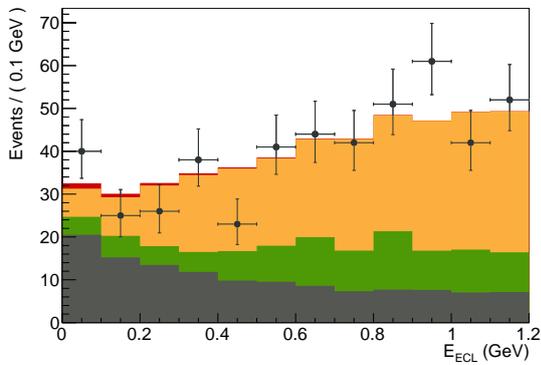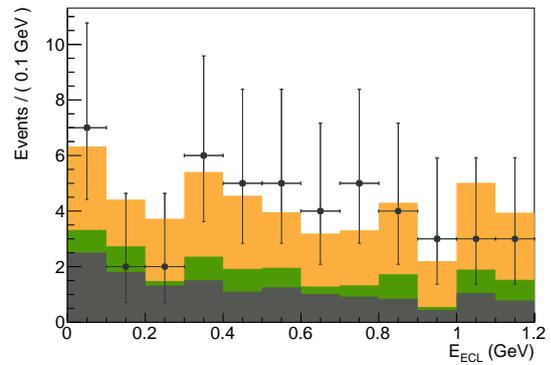
(f) $B^0 \to \pi^0 \nu\bar{\nu}$
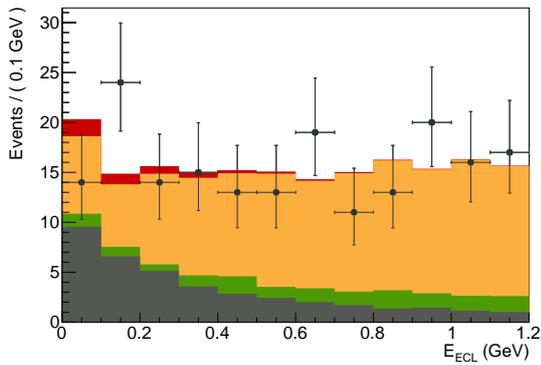
(g) $B^+ \to \rho^+ \nu\bar{\nu}$
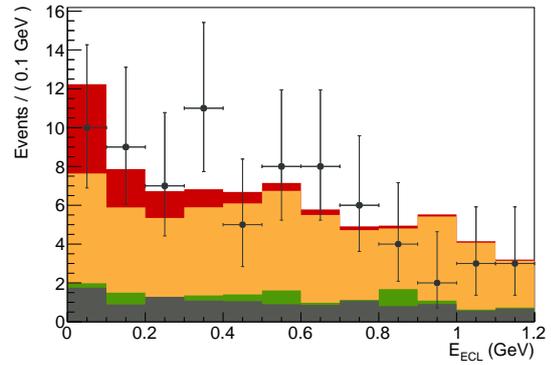
(h) $B^0 \to \rho^0 \nu\bar{\nu}$

Figure 10.2.: Fit results on data

(a) $B^+ \to K^{*+}\left(\to K^+\pi^0\right)\nu\bar{\nu}$ (b) $B^+ \to K^{*+}\left(\to K_S^0\pi^+\right)\nu\bar{\nu}$
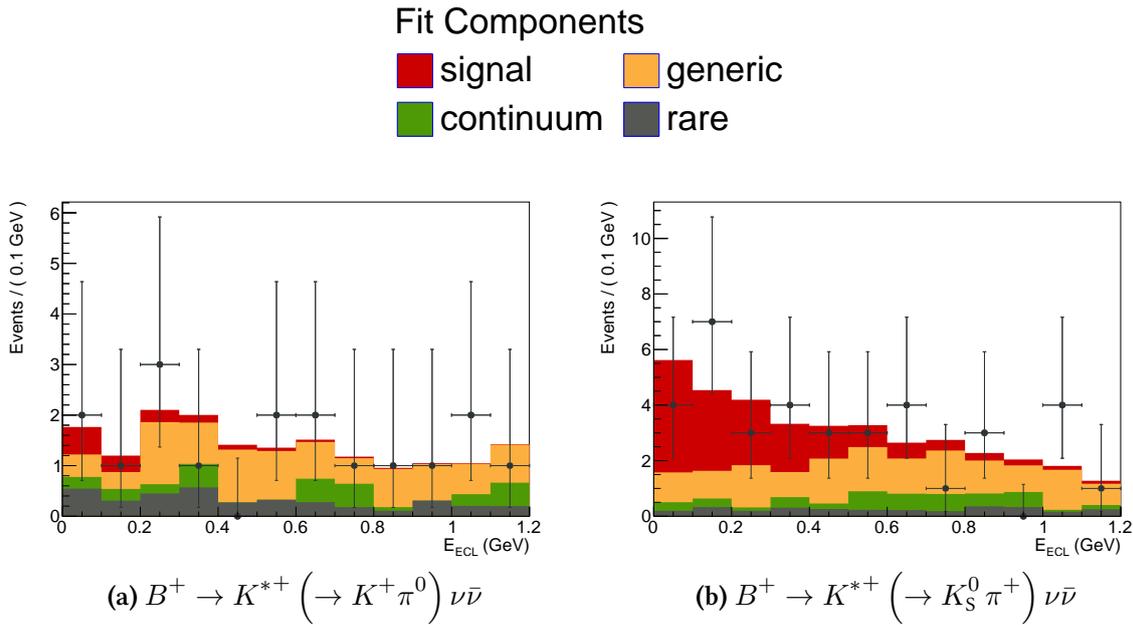
Figure 10.3.: Fit result on E_{ECL} in the K^{*+} subchannels.

10.2. Observed Limits

The limits calculated on the full data set are plotted alongside the expected values and all published previous results in Figure 10.4 on page 108 while the corresponding numerical values are displayed in Table 10.3 on page 108. We can conclude that our procedure of calculating expected limits was accurate. In cases where the observed number of signal events was in very good agreement with zero, the $B^0 \to K_S^0 \nu\bar{\nu}$, $B^+ \to \pi^+\nu\bar{\nu}$, $B^0 \to \pi^0\nu\bar{\nu}$, and $B^+ \to \rho^+\nu\bar{\nu}$ channels respectively, the observed limits basically coincide with the median of the toy distributions. Upward fluctuations cause a notable increase in the observed upper limit, as expected from the tails of the distributions in Figure 9.3 on page 95. Since the SM predicts a non-vanishing signal contribution in the $B^+ \to K^+\nu\bar{\nu}$, $B^+ \to K^{*+}\nu\bar{\nu}$, and $B^0 \to K^{*0}\nu\bar{\nu}$ channels, see Table 9.3 on page 98, no assessment whether the upward fluctuations in the signal region of the E_{ECL} distributions in the $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to K^{*+}\nu\bar{\nu}$ channel are due to signal or not simulated background components is possible. The randomly chosen simulated samples in Figure 9.1 on page 91 demonstrate that upwards fluctuations have a high likelihood to occur when a couple of channels is measured.

The most interesting case in that regard is the $B^0 \to K^{*0}\nu\bar{\nu}$ channel. We actually observe a downward fluctuation of the expected background there and the observed limit is now within a factor of two to the standard model prediction which might be helpful to rule out models of new physics largely enhancing that branching fraction. A direct comparison of the observed signal yields in the $B^+ \to K^{*+}\nu\bar{\nu}$ and $B^0 \to K^{*0}\nu\bar{\nu}$ channels demonstrates the difficulties of analyses with low statistics involved. Normal Poisson fluctuations of the background can have a relatively large impact on the observed signal, since there is no reason to assume that $B^+ \to K^{*+}\nu\bar{\nu}$ decays

behave differently from $B^0 \to K^{*0}\nu\bar{\nu}$ decays. In the SM for instance, the only difference comes from the life time difference between the B^+ and B^0 mesons.

To summarize our results, we observe the world's most stringent limits for the six rare B decays $B^0 \to K_S^0\nu\bar{\nu}$, $B^0 \to K^{*0}\nu\bar{\nu}$, $B^+ \to \pi^+\nu\bar{\nu}$, $B^0 \to \pi^0\nu\bar{\nu}$, $B^+ \to \rho^+\nu\bar{\nu}$, and $B^0 \to \rho^0\nu\bar{\nu}$. For the limit on the branching fraction of the $B^+ \to K^+\nu\bar{\nu}$ channel, we find a value of 1.8×10^{-5}, while we calculate a limit for the $B^+ \to K^{*+}\nu\bar{\nu}$ channel of 5.9×10^{-5}. Both values are statistical only, all values are quoted at a confidence level of $90\,\%$.

Table 10.3.: Observed limits without systematic uncertainties.

channel	$\mathscr{B}$ limit @ 90% CL $\left[10^{-5}\right]$
$B^+ \to K^+ \nu\bar{\nu}$	1.8
$B^0 \to K^0_S \nu\bar{\nu}$	1.2
$B^+ \to K^{*+} \nu\bar{\nu}$	5.9
$B^0 \to K^{*0} \nu\bar{\nu}$	1.7
$B^+ \to \pi^+ \nu\bar{\nu}$	1.3
$B^0 \to \pi^0 \nu\bar{\nu}$	0.9
$B^+ \to \rho^+ \nu\bar{\nu}$	2.9
$B^0 \to \rho^0 \nu\bar{\nu}$	3.9
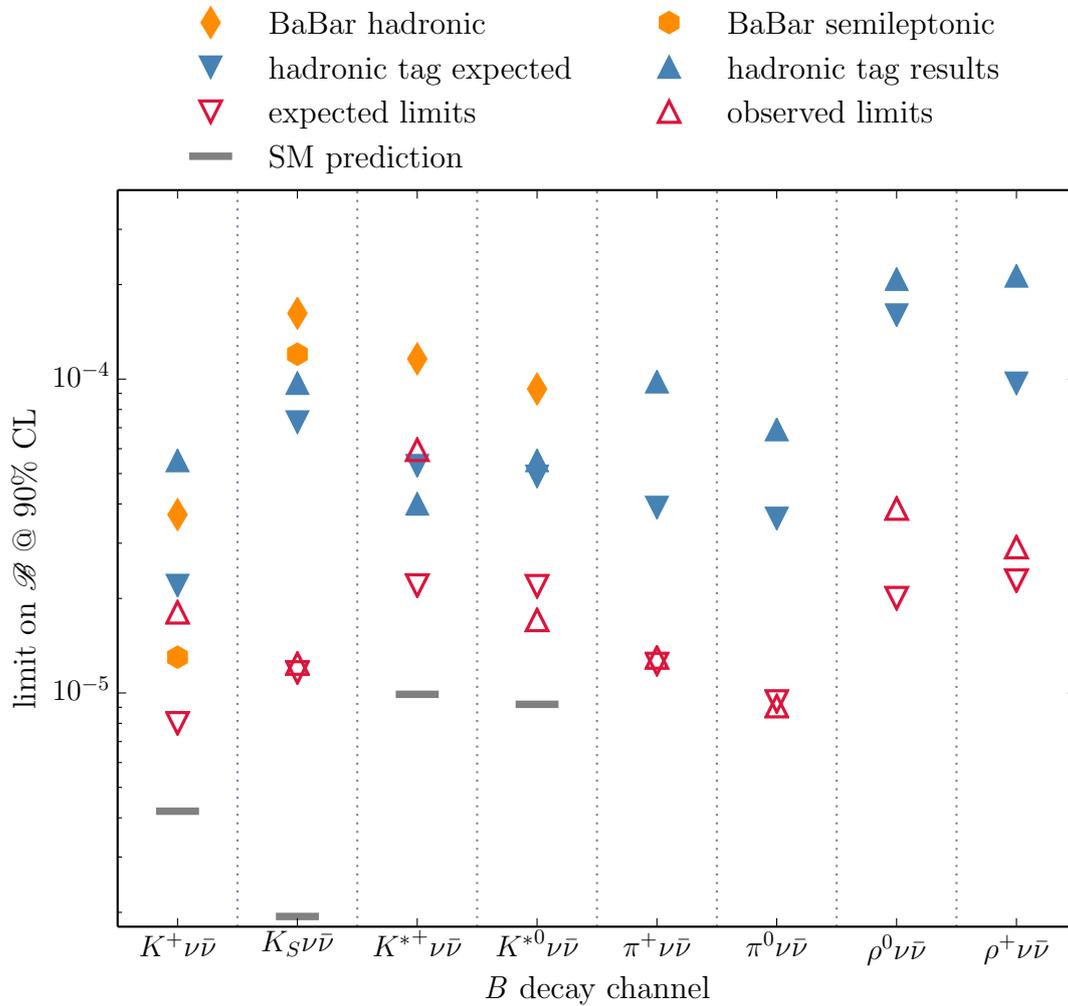
Figure 10.4.: Comparison between the observed and expected limits of this analysis and all other published work. The data points do not include the systematical uncertainty.

Chapter 11

Systematic Uncertainties

11.1. Calculation of Uncertainties

Various sources of uncertainties affect our result at various stages of the analysis. They fall into two principal categories. Those which influence the number of observed events (additive uncertainties) and those who influence the efficiency or the number of produced $B\bar{B}$ pairs (multiplicative uncertainties). The former are combined and the likelihood is then convoluted with a gaussian with zero mean and the width of the combined uncertainty. From this convoluted likelihood, the significance and limit on the number of observed events can then be calculated. The latter is used in combination with the former to calculate the limit on the branching fractions including all uncertainties.

11.1.1. Additive Uncertainties

We identify the sources of uncertainties on the signal yield listed below.

tag efficiency correction The tagging correction is applied to all simulated background events with a correctly reconstructed tag side as well as to all simulated signal events with a correctly identified tag side. The fit is repeated with the correction factor set to $\pm 1\,\sigma$ of the values listed in Section 6.3 on page 48. The differences of both results with the nominal fit result are averaged and assigned as an uncertainty.

fit bias The bias correction is also varied in an interval of $\pm 1\,\sigma$ of the nominal value, the fit is subsequently repeated.

background shape Since the MC statistics for the background model is limited, the templates used to build the background model have an uncertainty on their respective shape. The fit to data is repeated with simulated templates. The nominal templates are varied binwise according to the uncertainty of each bin. 1000 such alternative toy-models are simulated for each channel. The quantile points corresponding to a $\pm 1\,\sigma$ of the resulting distribution are taken as a systematic uncertainty.

Table 11.1.: Significance for the observation of each decay channel. The first error is statistical only, the second error is the systematic uncertainty.

channel	observed events	significance $[\sigma]$
$B^+ \to K^+ \nu\bar{\nu}$	$17.7 \pm 9.1 \pm 3.4$	1.9
$B^0 \to K^0_S \nu\bar{\nu}$	$0.6 \pm 4.2 \pm 1.4$	0.0
$B^+ \to K^{*+} \nu\bar{\nu}$	$16.2 \pm 7.4 \pm 1.8$	2.3
$B^0 \to K^{*0} \nu\bar{\nu}$	$-2.0 \pm 3.6 \pm 1.8$	0.0
$B^+ \to \pi^+ \nu\bar{\nu}$	$5.6 \pm 15.1 \pm 5.9$	0.0
$B^0 \to \pi^0 \nu\bar{\nu}$	$0.2 \pm 5.6 \pm 1.6$	0.0
$B^+ \to \rho^+ \nu\bar{\nu}$	$6.2 \pm 12.3 \pm 2.4$	0.3
$B^0 \to \rho^0 \nu\bar{\nu}$	$11.9 \pm 9.0 \pm 3.6$	1.2

relative fractions The ratio of the background components with respect to each other have been fixed in the fit. We vary this ratio by its uncertainty and assign the resulting deviation from the central value as a systematic error.

continuum correction The scaling factor of the continuum MC is varied by $\pm 1\,\sigma$ and the data refitted with the new templates. This is only done where the factor differs from one.

K^0_L**-veto** The veto efficiency correction is varied within $\pm 1\,\sigma$ and the fit is repeated.

The results are displayed for each channel and uncertainty in Table 11.2 on page 112. All values quoted in the table are absolute uncertainties on the number of observed signal events since relative values are not very meaningful for a result of basically zero. Correlations between systematic uncertainties are considered to be negligible. The measurement of the observed number of signal events including systematic uncertainties is displayed in Table 11.1. The profile likelihoods of each channel normalized to the likelihood value at the best fitting point as a function of the signal yield with and without the additive uncertainties are displayed in Figure 11.1 on page 114.

11.1.2. Multiplicative Uncertainties

We consider the following sources of uncertainties for the signal efficiency.

tag efficiency correction This systematic effect influences both the number of observed events through the background shape and the efficiency. The efficiency calculated with the correction factor plus and minus $1\,\sigma$ is assigned as an error on the efficiency. We determine a value of about $9.6\,\%$ for both neutral and charged channels.

vetoes We use the $B \to D\pi$ sample to estimate the systematic uncertainty connected to the π^0 and charged track veto. The efficiency difference between data and MC with and without the application of the veto is used for this estimation. We find this value to be $4\,\%$ in each case.

remaining raw tracks The data-MC ratio are calculated in the $B \to D\pi$ test sample without an additional raw track and with two additional raw tracks. The difference in efficiency amounts to $1.4\,\%$ and $0.8\,\%$ in the case of charged channels and to $1.3\,\%$ and $0.2\,\%$ in the case of neutral channels. We assign an uncertainty of $1\,\%$ for each channel.

PID efficiency The study used to correct for data-MC differences, see Section 6.4 on page 49, also gives the error on that factor. We obtain values for the effect on the efficiency of around $2\,\%$ for all channels with at least one charged particle in the final state.

π^0-efficiency The π^0 reconstruction efficiency has been studied by comparing the fractions of reconstructed $\eta \to \pi^+\pi^-\pi^0$ and $\eta \to 3\pi^0$ decays between data and MC [LCH03]. This study determined the uncertainty on the π^0 efficiency to be $4\,\%$. This value is assigned to each channel with a π^0 in the final state.

K_S^0-efficiency We assign a value of $2.2\,\%$ per reconstructed K_S^0 as the K_S^0 finding systematic. This follows the results found in [Whi11]. $D^* \to D\left(K_S^0\pi^+\pi^-\right)\pi$ have been reconstructed and the efficiencies on data and MC compared.

$N_{B\bar{B}}$ The uncertainty of $N_{B\bar{B}}$ has been measured to be $1.4\,\%$, where the error on the luminosity has been propagated to the number of produced $B\bar{B}$ pairs. This value is assigned as the systematic uncertainty.

tracking The uncertainty on the track finding per track has been studied in [Kop03]. To achieve this, a sample of $D^{*+} \to D^0\left(K_S^0\pi^+\pi^-\right)\pi^+$ where one K_S^0 daughter has not been reconstructed was used. The efficiency difference between data and MC for each track was found to be within $0.35\,\%$ compatible with one. This value is much smaller than other uncertainties and hence neglected, since we reconstruct channels with at most two charged particles in the final state.

MC statistics The finite amount of simulated signal decays to determine the efficiency has only a marginal effect on the efficiency calculation since we simulate at least 10×10^6 signal events per channel. This source of uncertainty will consequently be neglected.

Table 11.2.: Additive systematic uncertainties, all numbers are absolute.

	$K^+\nu\bar\nu$	$K^0_S\,\nu\bar\nu$	$K^{*+}\,\nu\bar\nu$	$K^{*0}\,\nu\bar\nu$	$\pi^+\,\nu\bar\nu$	$\pi^0\,\nu\bar\nu$	$\rho^+\,\nu\bar\nu$	$\rho^0\,\nu\bar\nu$
tag correction	0.5	0.2	0.1	1.0	1.9	0.1	0.2	0.5
fit bias	0.2	0.1	0.2	0.0	0.2	0.1	0.2	0.2
shape uncertainty	2.6	1.3	1.8	0.7	4.5	1.5	2.3	3.4
fixed fraction	0.4	0.3	0.1	0.0	1.3	0.1	0.1	1.0
continuum correction	2.0	0.0	0.0	0.0	3.1	0.0	0.0	0.0
K^0_L veto	0.2	0.2	0.1	1.0	0.6	0.4	0.6	0.0
total	3.4	1.4	1.8	1.5	5.9	1.6	2.4	3.6

Table 11.3.: Multiplicative systematic uncertainties, all are in %.

	$K^+\nu\bar\nu$	$K^0_S\,\nu\bar\nu$	$K^{*+}\,\nu\bar\nu$	$K^{*0}\,\nu\bar\nu$	$\pi^+\,\nu\bar\nu$	$\pi^0\,\nu\bar\nu$	$\rho^+\,\nu\bar\nu$	$\rho^0\,\nu\bar\nu$
tag efficiency	9.6	9.6	9.6	9.6	9.6	9.6	9.6	9.6
track veto	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
π^0 veto	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
raw tracks	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
PID efficiency	2.0	0.0	2.0	2.0	2.0	0.0	2.0	2.0
π^0 efficiency	0.0	0.0	4.0	0.0	0.0	4.0	0.0	4.0
K^0_S efficiency	0.0	2.2	2.2	0.0	0.0	0.0	0.0	0.0
$N_{B\bar B}$	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
total	11.4	11.5	12.3	11.4	11.4	12.0	11.4	12.6

The results are summarized in Table 11.3 on the preceding page. Again, the correlations between uncertainties are neglected. This enables the calculation of the final exclusion limit for the branching fraction of each channel including all systematic uncertainties. We convolute the likelihood with a gaussian according to the following formula.

$$\mathcal{L}\left(\mathcal{B}\right)_{\text{syst}} = \int_{-\infty}^{\infty} d\mathcal{B}' \mathcal{L}\left(\mathcal{B}'\right) \times g\left(\mathcal{B} - \mathcal{B}', 0, \sigma_{\text{add}} \oplus \sigma_{\text{mult}} \times \mathcal{B}'\right) \tag{11.1}$$

Here, $\mathcal{L}$ is the profile likelihood, $\mathcal{B}$ the branching fraction, σ_{add} the additive uncertainty on the branching fraction, σ_{mult} the multiplicative uncertainty, and $g\left(t, \mu, \sigma\right)$ a gaussian with mean μ and width σ evaluated at t. The limits are calculated according to Equation (9.3) on page 92. The only channel where the multiplicative systematics have a larger impact on the final value is the $B^+ \to K^{*+}\nu\bar{\nu}$ channel. This is the channel with the highest significance, the largest fitted yield relative to the statistical error. Consequently, the multiplicative uncertainties affect this channel the most. All results are listed in Table 11.4 on page 115, as well as compared to previous works on the subject.

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K_S^0 \nu\bar{\nu}$

(c) $B^+ \to K^{*+} \nu\bar{\nu}$

(d) $B^0 \to K^{*0} \nu\bar{\nu}$

(e) $B^+ \to \pi^+ \nu\bar{\nu}$

(f) $B^0 \to \pi^0 \nu\bar{\nu}$

(g) $B^+ \to \rho^+ \nu\bar{\nu}$

(h) $B^0 \to \rho^0 \nu\bar{\nu}$
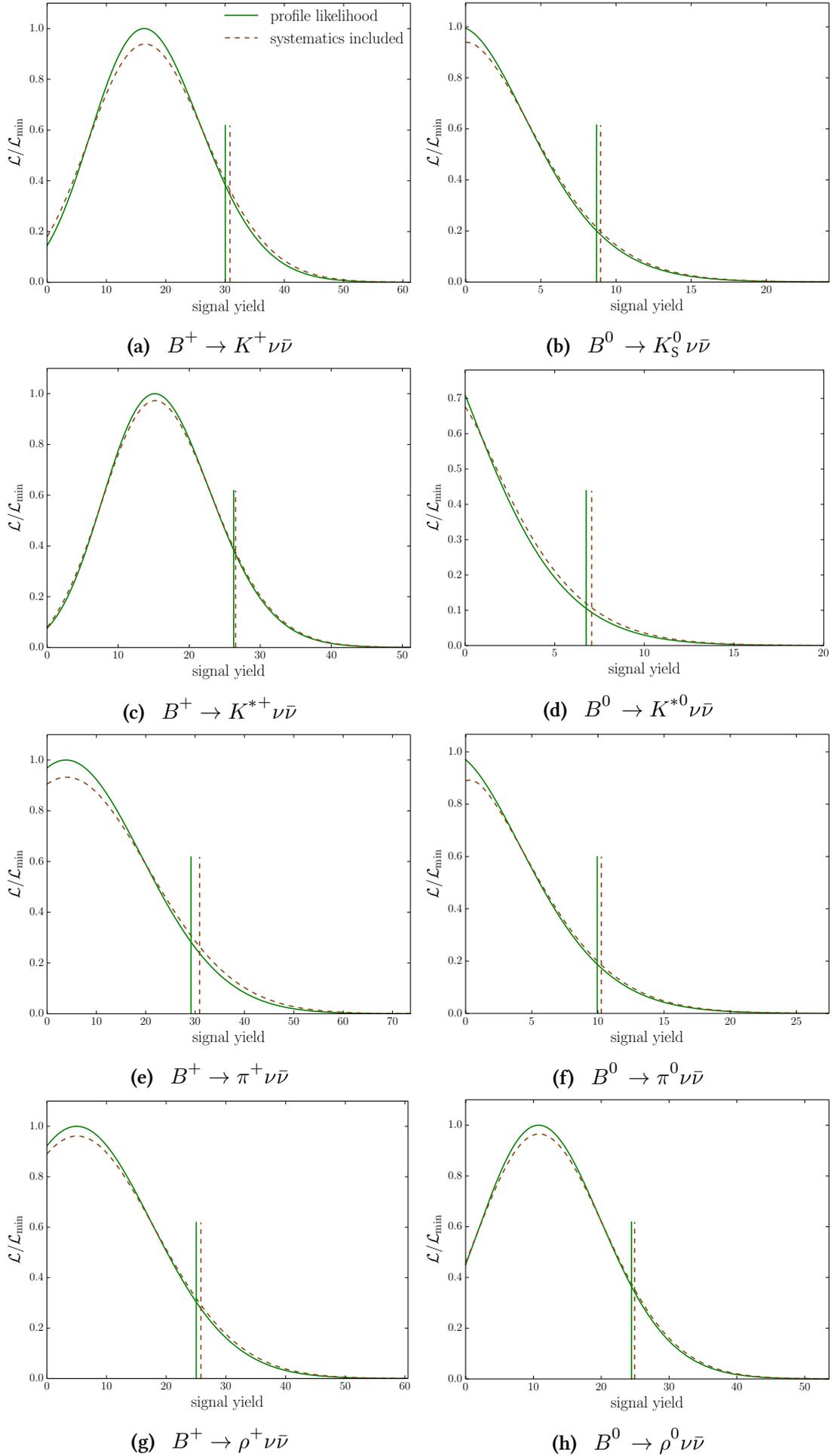
Figure 11.1.: Profile likelihood ratio with and without additive systematic uncertainties.

Table 11.4.: Limits on the branching fractions of all channels including all systematics compared to the Belle analysis with hadronic tag and the current world best limits from [O$^+$14]. Channels for which this work provides the best limit are marked in bold.

channel	$\mathscr{B}$ limit @ 90 % CL	previous Belle $\mathscr{B}$ limit @ 90 % CL	world best $\mathscr{B}$ limit @ 90 % CL
$B^+ \to K^+ \nu\bar\nu$	1.9×10^{-5}	5.5×10^{-5}	1.6×10^{-5}
$\boldsymbol{B^0 \to K^0_S \nu\bar\nu}$	1.3×10^{-5}	9.7×10^{-5}	5.6×10^{-5}
$B^+ \to K^{*+} \nu\bar\nu$	6.1×10^{-5}	4.0×10^{-5}	4.0×10^{-5}
$\boldsymbol{B^0 \to K^{*0} \nu\bar\nu}$	1.8×10^{-5}	5.5×10^{-5}	5.5×10^{-5}
$\boldsymbol{B^+ \to \pi^+ \nu\bar\nu}$	1.4×10^{-5}	9.8×10^{-5}	9.8×10^{-5}
$\boldsymbol{B^0 \to \pi^0 \nu\bar\nu}$	0.9×10^{-5}	6.9×10^{-5}	6.9×10^{-5}
$\boldsymbol{B^+ \to \rho^+ \nu\bar\nu}$	3.0×10^{-5}	21.4×10^{-5}	15×10^{-5}
$\boldsymbol{B^0 \to \rho^0 \nu\bar\nu}$	4.0×10^{-5}	20.8×10^{-5}	20.8×10^{-5}

11.2. Combination of Neutral and Charged Final States

As mentioned in Section 2.2.1 on page 9, the SM branching fractions at leading order only differ with respect to the lifetime of the B meson between charged and neutral final states. We assume the following relation to combine the results obtained in each charged decay channel with the result obtained in the corresponding neutral B decay channel.

$$\frac{\mathscr{B}\left(B^+ \to K^{(*)+}\nu\bar\nu\right)}{\mathscr{B}\left(B^0 \to K^{(*)0}\nu\bar\nu\right)} = \frac{\tau_{B^+}}{\tau_{B^0}} \approx 1.08 \tag{11.2}$$

The samples with charged and neutral B_{tag} are statistically independent and the major systematic uncertainties are independent as well since we evaluate them on fully independent samples of MC and reconstructed control channels, respectively. We therefore add the negative log-likelihoods for charged and neutral final states where we scale the branching fraction of the neutral particles by a factor of $1/1.08$ which we obtain by the means of Equation (11.2).

$$-\log\mathcal{L}\left(\mathscr{B}\left(B \to h\nu\bar\nu\right)\right) = -\log\mathcal{L}\left(\mathscr{B}\left(B \to h^+\nu\bar\nu\right)\right)$$
$$-\log\mathcal{L}\left(\mathscr{B}\left(B \to h^0\nu\bar\nu\right) \times \frac{\tau_{B^+}}{\tau_{B^0}} \times f_{K^0_S}\right) \tag{11.3}$$

Because of U-spin symmetry, this formula holds for the $B \to \pi\nu\bar\nu$ and $B \to \rho\nu\bar\nu$ channels. The factor $f_{K^0_S}$ takes into account the fact that we measure the $B^0 \to K^0_S \nu\bar\nu$ branching fraction and not the $B^0 \to K^0 \nu\bar\nu$ branching fraction and is set to two in the case of unexcited kaons and to one in all other cases. h here refers to any of K, K^*, π, or ρ. We subsequently repeat the limit calculation with the combined likelihood and obtain the upper limits displayed in Table 11.5 on the following page.

Table 11.5.: Combined limits for charged and neutral modes.

channel	$\mathscr{B}$ limit @ 90 % CL
$B \to K \nu \bar{\nu}$	1.6×10^{-5}
$B \to K^* \nu \bar{\nu}$	2.7×10^{-5}
$B \to \pi \nu \bar{\nu}$	0.8×10^{-5}
$B \to \rho \nu \bar{\nu}$	2.8×10^{-5}

11.2.1. Calculation of $\mathcal{R}_{K^{(*)}}$

As discussed in Section 2.2.3 on page 12, some models of new physics can be parametrized in terms of the ratio of the value of the $B \to K \nu \bar{\nu}$ and $B \to K^* \nu \bar{\nu}$ branching fractions and their respective SM expectation. We take the SM calculations from [BGNNS15] and their theoretical uncertainties and convolute the combined likelihood of the $B \to K \nu \bar{\nu}$ ($B \to K^* \nu \bar{\nu}$) channel with a gaussian with the mean of the measured parameter and its theoretical uncertainty as width. We subsequently calculate a limit on $\mathcal{R}_{K^{(*)}}$ similar to Equation (9.3) on page 92. The results are displayed in Table 11.6. Although we have to rely on the accuracy of theoretical predictions for our limit on $\mathcal{R}_{K^{(*)}}$, we conclude that our results leave sufficient space for models to influence the decay rate of $B \to h \nu \bar{\nu}$ decays since our results allow for an effect of over a factor of two.

Table 11.6.: Limits on $\mathcal{R}_{K^{(*)}}$ obtained via the combined charged and neutral likelihoods.

parameter	limit @ 90 % CL
$\mathcal{R}_K$	3.9
$\mathcal{R}_{K^*}$	2.7

11.3. Prospects for Belle II

The Belle experiment is currently being upgraded to an improved version with better tracking, vertex resolution, and more luminosity delivered by the likewise upgraded accelerator SuperKEKB. The Belle II collaboration expects to take a total amount of 50 times the integrated luminosity taken by Belle. The increase in both the total amount of $\Upsilon(4S)$ produced and the overall detector performance will provide analysts with the sensitivity necessary to scrutinize yet unobserved processes. Chiefly amongst, $B \to h\nu\bar{\nu}$ decays.

If we take the $B^+ \to K^+\nu\bar{\nu}$ as an example, we can study the effect an increase of a factor of fifty on the integrated luminosity would have on this analysis. The statistical uncertainty on the branching fraction of 0.5×10^{-5} is of roughly the same magnitude as theoretical predictions, [BGNNS15] e.g. predict $0.42 \pm 0.04 \times 10^{-5}$. A factor of fifty on the total amount of data would result in a statistical uncertainty of 0.07×10^{-5}. Consequently even with Belle hardware and without algorithmical improvements to be expected in the upcoming years, Belle II will be able to observe $B^+ \to K^+\nu\bar{\nu}$ decays if their branching fraction is as high as theoretically predicted. Moreover, taking all this into account, Belle II will be able to provide a measurement with uncertainties within the same order of magnitude as current theoretical uncertainties and thus will be able to make conclusive statements about several models of new physics.

Several systematic uncertainties would also significantly be reduced. Chiefly among them, all major uncertainties. The tagging uncertainty would scale proportionally to the statistical uncertainty since the error comes from the number of tagged $B^+ \to \bar{D}^0\pi^+$ and $B^0 \to D^-\pi^+$ events, whose branching fractions are very well known, and is thus limited by the size of that sample. A reduction from $9.6\,\%$ down to only $1.4\,\%$ would be a huge improvement. This argument applies to a certain extent to the systematic uncertainties related to the track and π^0 vetoes and remaining raw tracks as well. It can moreover be safely assumed that the number of produced generic, continuum, and rare Monte-Carlo events will increase at least proportionally to the increase in integrated luminosity. This would reduce the uncertainties stemming from the background shape and relative fractions, the largest additive sources for many channels.

In the case of the $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to \pi^+\nu\bar{\nu}$ channels, the largest uncertainty on the yield comes from the correction on simulated continuum. This analysis would have benefited greatly if more data taken at off-resonance energies were available. For many channels, no events recorded at off-resonance energies passing all selection criteria were recorded at all. In two channels, the $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to \pi^+\nu\bar{\nu}$ channels, a deviation between data and MC of over a factor of two was observed in the off-resonance sample, see Section 8.2 on page 80. This was accounted for by scaling the simulated continuum MC in the on-resonance sample with a factor determined on the data-MC fraction of the off-resonance sample. The amount of data was simply too small to take the shape of the continuum component from off-resonance data which would result in far smaller uncertainties for both channels. Furthermore, we can assume that this deviation occurs to some degree in other channels as well since $B^+ \to K^+\nu\bar{\nu}$ and $B^+ \to \pi^+\nu\bar{\nu}$ are the channels with the highest efficiency, with almost 3×10^{-3} in both cases, and thus the only two channels where any significant difference in the data-MC yield in the off-resonance sample was observable. This uncertainty was absorbed into the shape uncertainty in case of other channels. This will become a

much more important issue with growing amounts of data taken on-resonance and therefore growing precision. It can thus be concluded from this consideration that the accurate determination of $B \to h\nu\bar{\nu}$ branching fractions requires Belle II to acquire more data at off-resonance energies. To summarize the estimation of the expected precision and accuracy of the measurement of the branching fraction of the $B^+ \to K^+\nu\bar{\nu}$ channel, we take the theory prediction from [BGNNS15] as input to calculate the expected systematic uncertainty with those uncertainties scaled as described above. We obtain an expected uncertainty of $11\,\%$. We do not include expected improvement in for instance the veto efficiencies which is most likely to improve as well with increased luminosity. The estimation yields a value of $\mathscr{B}\left(B^+ \to K^+\nu\bar{\nu}\right) = (4.2 \pm 0.7 \pm 0.4) \times 10^{-6}$ where we used the theoretical prediction as the value.

Chapter 12

Conclusion

The scope of this analysis encompasses the measurement of eight independent quantities that have never been measured before but are of high theoretical interest and relevance for contemporary research. This analysis succeeded in increasing the sensitivity by over a factor of two in comparison to the predecessor analysis at Belle. This manifests itself in the efficiency-to-background ratio, see Table 7.1 on page 60, and in the expected limits, see Figure 9.2 on page 94, being the world's best achieved so far in all eight investigated channels. The observed limits in turn are the most stringent measured as of today in six out of those eight channels. The other two channels exhibit a significance for the presence of signal of $1.9\,\sigma$ $(2.3\,\sigma)$ in the $B^+ \to K^+ \nu\bar{\nu}$ $(B^+ \to K^{*+}\nu\bar{\nu})$ channel, respectively. Those values are still within statistical uncertainties and well below the commonly agreed upon threshold to claim evidence of $3\,\sigma$.

The sensitivity of this analysis has reached a precision at which the standard model process starts to play a role. The limits which are weaker than the current best ones are most likely affected by the presence of signal events in the selected sample. If we assume the correctness of the SM calculations from [BGNNS15], the probability to not select any real $B^+ \to K^+ \nu\bar{\nu}$ decay is only $0.11\,\%$. However, despite the fact that we achieved the sensitivity to observe real events, the luminosity taken over the run time of the Belle experiment is not high enough to provide sufficient statistics to claim evidence. Consequently, we only set upper limits on the branching fractions of the investigated decays. The results are summarized in Table 11.4 on page 115 and displayed in Figure 12.1 on page 121.

The increase in expected and observed limits is sizable and takes values of almost an order of magnitude in some channels as compared to the previous results from Belle. This was achieved by numerous improvements. First, we utilize a semileptonic as opposed to a hadronic tagging algorithm. We find that the efficiency advantage of the semileptonic reconstruction outweighs the purity advantage of the hadronic reconstruction. This is an important result for any missing energy analyses at future B factories. Second, we performed a more inclusive analysis. We set only necessary requirements on the preselection which we were able to do because of a more thorough understanding of the background processes involved. Third, we use control channels to correct

incorrect descriptions and to validate Monte Carlo simulations. Future analyses of decays with missing momentum at Belle II will benefit greatly from the insights gained by this work.

From a theory view on things, our result will be helpful to constrain parameters of models of new physics, such as those outlined in Section 2.2.3 on page 12. Especially the absence of signal in the $B^0 \rightarrow K^{*0} \nu \bar{\nu}$ channel, where our limit is very close to theoretical predictions, shrinks the parameter space for possible branching fraction enhancing models. The remaining parameter space of such theories can be tested at Belle II. If we extrapolate our results, including the systematics, we can deduce that Belle II will be precise and accurate enough to either make a precise measurement or to exclude the standard model prediction for $b \rightarrow s\nu\bar{\nu}$ channels. This statement is derived from this work and holds for semileptonic tagging. If hadronic tagging channels are added, the combined sensitivity of Belle II will be even higher.

So, no new physics has been found and the world of particle physics has not been rescued from the problems in the fundamental theory. But Dostoevsky has something to say for this situation as well. "Beauty is a terrible and awful thing! It is terrible because it has not been fathomed … . Here the boundaries meet and all contradictions exist side by side." Until Belle II.
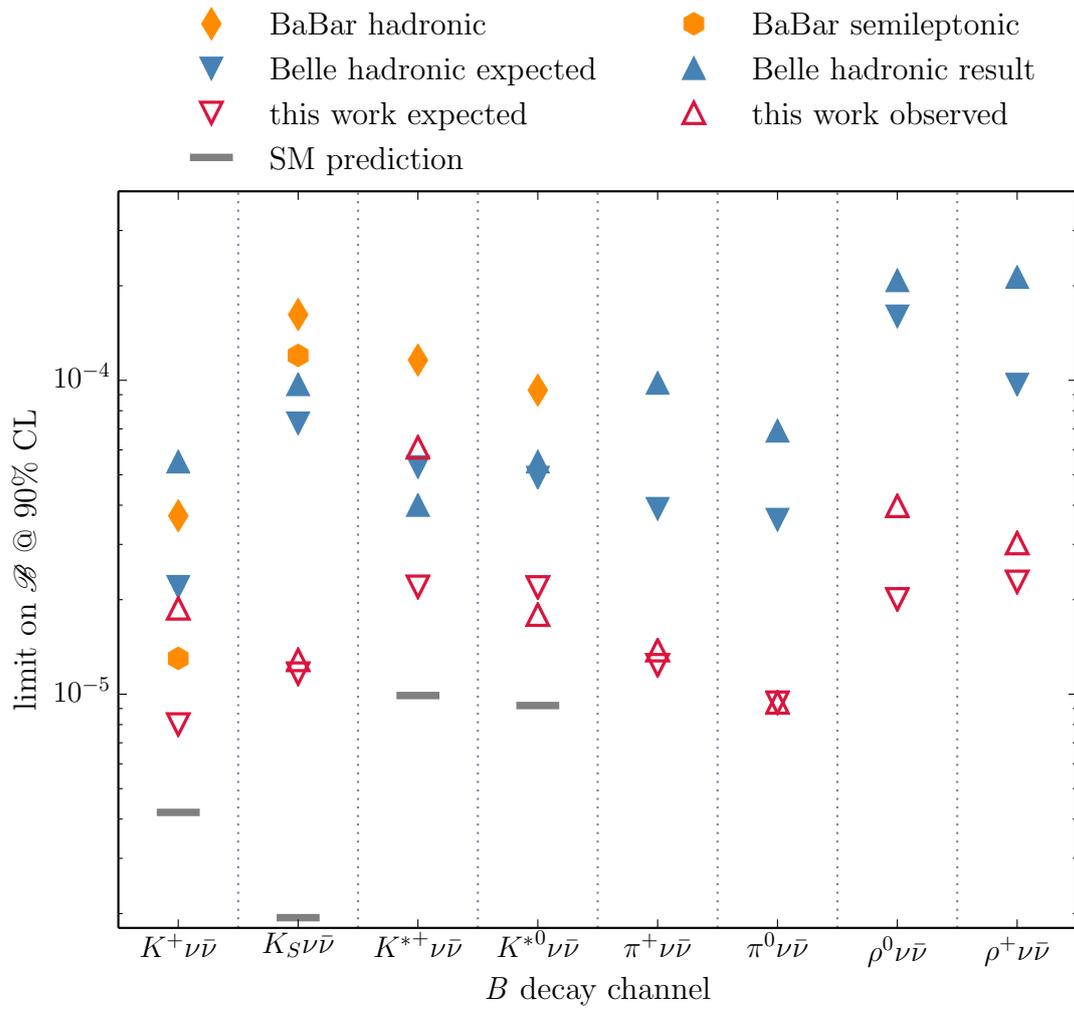
Figure 12.1.: All observed limits including systematic uncertainties in comparison with previously obtained results.

Appendix A

Reconstructed Tag-Channels

The following pages contain a list of all B, D^* and D decay channels reconstructed on the tag side.

decay channel	$\mathscr{B}_{D^0}$ [%]
$D^0 \to K^- \pi^+ \pi^0$	13.9 ± 0.5
$D^0 \to K^- \pi^+ \pi^+ \pi^-$	8.1 ± 0.2
$D^0 \to K_S^0 \pi^+ \pi^- \pi^0$	5.2 ± 0.6
$D^0 \to K^- \pi^+$	3.9 ± 0.05
$D^0 \to K_S^0 \pi^+ \pi^-$	2.8 ± 0.2
$D^0 \to \pi^+ \pi^- \pi^0$	1.4 ± 0.06
$D^0 \to K_S^0 \pi^0$	1.2 ± 0.04
$D^0 \to K_S^0 K^+ K^-$	0.4 ± 0.03
$D^0 \to K^+ K^-$	0.4 ± 0.01
$D^0 \to \pi^+ \pi^-$	0.1 ± 0.003
	37.4 ± 0.84

Table A.1.: Reconstructed D^0 Tag Channels

decay channel	$\mathscr{B}_{D^+}$ [%]
$D^+ \to K^- \pi^+ \pi^+$	9.1 ± 0.2
$D^+ \to K_S^0 \pi^+ \pi^0$	7.0 ± 0.3
$D^+ \to K^- \pi^+ \pi^+ \pi^0$	6.0 ± 0.2
$D^+ \to K_S^0 \pi^+ \pi^+ \pi^-$	3.1 ± 0.1
$D^+ \to K^+ K^- \pi^+ \pi^0$	1.5 ± 0.7
$D^+ \to K_S^0 \pi^+$	1.5 ± 0.07
$D^+ \to K^+ K^- \pi^+$	1.0 ± 0.03
	29.2 ± 0.68

Table A.2.: Reconstructed D^+ Tag Channels

decay channel	$\mathscr{B}_{D^*}$ [%]
$D^{*0} \to D^0 \pi^0$	61.9 ± 2.9
$D^{*0} \to D^0 \gamma$	38.1 ± 2.9
	$100 \pm 0.$

Table A.3.: Reconstructed D^{*0} channels

decay channel	$\mathscr{B}_{D^*}\,[\%]$
$D^{*+} \to D^0 \pi^+$	67.7 ± 0.5
$D^{*+} \to D^+ \pi^0$	30.7 ± 0.5
	98.4 ± 0.7

Table A.4.: Reconstructed D^{*+} channels

decay channel	$\mathscr{B}_{B^0}\,[\%]$
$B^0 \to D^{*-} e^+ \nu_e$	4.95 ± 0.11
$B^0 \to D^{*-} \mu^+ \nu_\mu$	4.95 ± 0.11
$B^0 \to D^- e^+ \nu_e$	2.18 ± 0.12
$B^0 \to D^- \mu^+ \nu_\mu$	2.18 ± 0.12
	14.26 ± 0.23

Table A.5.: Reconstructed B^0 channels

decay channel	$\mathscr{B}_{B^+}\,[\%]$
$B^+ \to \bar{D}^{*0} e^+ \nu_e$	5.70 ± 0.19
$B^+ \to \bar{D}^{*0} \mu^+ \nu_\mu$	5.70 ± 0.19
$B^+ \to \bar{D}^0 e^+ \nu_e$	2.26 ± 0.11
$B^+ \to \bar{D}^0 \mu^+ \nu_\mu$	2.26 ± 0.11
	15.92 ± 0.22

Table A.6.: Reconstructed B^+ channels

Appendix B

Corrected Branching Fractions

All correction factors to rare branching fractions according to latest results are listed below. Channels in which the value has not been updated since the creation of the .dec file are not listed, as well as channels which have not been observed yet and have no improved limit. In some cases sub-decay channels, e.g. $f_2' \to K_L^0 K_L^0$, are listed for the better understanding of the topology of the decay in question. However, sub-branching fractions are never included in the correction factors.

For decays which are hard to measure and not measured, isospin symmetry is used to estimate the branching fraction if possible. Those cases are specifically marked in the respective table.

channel	decay file	pdg 15	comment
$B^+ \to K^+ K^0_L \pi^0$	1.14×10^{-5}	1.20×10^{-5}	both limits
$B^+ \to K^+ n\bar{n}$	5.7×10^{-6}	5.9×10^{-6}	from $B^+ \to K^+ p^+ p^-$
$B^+ \to \tau^+ \nu_\tau$	0.93×10^{-4}	1.14×10^{-4}	
$B^+ \to f'_2 \left(\to K^0_L K^0_L \right) K^+$	1.33×10^{-6}	1.80×10^{-6}	
$B^+ \to K^+ K^0_L$	6.80×10^{-5}	6.55×10^{-5}	
$B^+ \to f_0 (1500) \left(K^0_L K^0_L \right) K^+$	4.4×10^{-6}	3.7×10^{-6}	
$B^+ \to K^0_L \pi^+$	1.16×10^{-5}	1.19×10^{-5}	
$B^+ \to K^{*+} K^{*0}$	7.1×10^{-5}	1.2×10^{-6}	
$B^+ \to K^{*+} K^0_L K^0_L$	3.13×10^{-5}	3.60×10^{-5}	
$B^+ \to \Phi (1020) \left(K^0_S K^0_L \right) K^+$	8.3×10^{-6}	8.8×10^{-6}	
$B^0 \to K^+ K^- K^0_L$	9.42×10^{-6}	1.65×10^{-5}	
$B^+ \to K^{*+} \gamma$	4.03×10^{-5}	4.21×10^{-5}	

Table B.1.: $B^+ \to K^+ \nu\bar{\nu}$

channel	decay file	pdg 15	comment
$B^0 \to K^0_L K^0_S K^{*0}$	3.42×10^{-5}	2.75×10^{-5}	from $B^0 \to K^+ K^- K^{*0}$
$B^0 \to \Phi (1020) \left(\to K^0_L K^0_S \right) K^0_S$	4.15×10^{-6}	3.65×10^{-6}	
$B^0 \to K^0_S n\bar{n}$	1.0×10^{-6}	1.33×10^{-6}	from $B^0 \to K^0_S p^+ p^-$
$B^0 \to \Phi (1020) \left(\to K^0_L K^0_S \right) K^0_L$	4.15×10^{-6}	3.65×10^{-6}	

Table B.2.: $B^0 \to K^0_S \nu\bar{\nu}$

channel	decay file	pdg 15	comment
$B^+ \to K^{*+} K^{*0}$	7.1×10^{-5}	1.2×10^{-5}	was an upper limit
$B^+ \to K^{*+} \gamma$	4.03×10^{-5}	4.20×10^{-5}	
$B^+ \to K^{*+} K^0_L K^0_L$	3.13×10^{-5}	3.60×10^{-5}	from $B^+ \to K^{*+} K^+ K^-$
$B^+ \to \tau^+ \nu_\tau$	0.93×10^{-4}	1.14×10^{-4}	
$B^+ \to K^{*+} n\bar{n}$	5.7×10^{-6}	3.6×10^{-6}	from $B^+ \to p^+ p^-$
$B^+ \to K^+ K^0_L \pi^0$	1.14×10^{-5}	1.20×10^{-5}	both limits

Table B.3.: $B^+ \to K^{*+} \nu\bar{\nu}$

channel	decay file	pdg 15	comment
$B^0 \to K^{*0} K^0_L K^0_L$	5.7×10^{-6}	6.88×10^{-6}	from $B^+ \to K^{*+} K^+ K^-$
$B^0 \to K^{*0} \gamma$	4.01×10^{-5}	4.33×10^{-5}	
$B^0 \to K^{*0} n\bar{n}$	5.60×10^{-6}	1.24×10^{-6}	from $B^0 \to K^{*0} p^+ p^-$

Table B.4.: $B^0 \to K^{*0} \nu\bar{\nu}$

decay	$\mathscr{B}$ decay file	$\mathscr{B}$ pdg 15	comment
$B^+ \to K^0_L \pi^+$	1.16×10^{-5}	1.19×10^{-5}	
$B^+ \to \tau^+ \nu_\tau$	9.30×10^{-5}	1.14×10^{-4}	
$B^+ \to K^0_L K^0_L \pi^+$	3.2×10^{-6}	5.1×10^{-7}	both limits
$B^+ \to K^{*0} \pi^+$	1.07×10^{-5}	1.01×10^{-5}	
$B^+ \to \pi^0 \mu^+ \nu_\mu$	7.3×10^{-5}	7.8×10^{-5}	
$B^+ \to \pi^+ n\bar{n}$	1.50×10^{-6}	1.62×10^{-5}	from $B^+ \to \pi^+ p^+ p^-$
$B^+ \to \eta \mu^+ \nu_\mu$	8.4×10^{-5}	3.8×10^{-5}	
$B^+ \to \pi^0 e^+ \nu_e$	7.3×10^{-5}	7.8×10^{-5}	
$\bar{B}^0 \to \pi^+ e^- \nu_e$	1.36×10^{-4}	1.45×10^{-4}	

Table B.5.: $B^+ \to \pi^+ \nu\bar{\nu}$

decay	$\mathscr{B}$ decay file	$\mathscr{B}$ pdg 15	comment
$B^0 \to \rho^- \mu^+ \nu_\mu$	2.77×10^{-4}	2.94×10^{-4}	
$B^+ \to \tau^+ \nu_\tau$	9.3×10^{-5}	1.14×10^{-4}	
$B^+ \to \mu^+ \nu_\mu \pi^0$	7.3×10^{-5}	7.8×10^{-5}	
$B^0 \to K^{*0}(1680)\left(\to K^0_L \pi^0\right)\pi^0$	1.01×10^{-5}	7.50×10^{-6}	both limits
$B^+ \to K^{*+} \pi^0$	6.9×10^{-6}	8.2×10^{-6}	
$B^0 \to \rho^- e^+ \nu_e$	2.77×10^{-4}	2.94×10^{-4}	
$B^+ \to e^+ \nu_e \pi^0$	7.3×10^{-5}	7.8×10^{-5}	
$B^0 \to K^0_L f_2(1270)\left(\to \pi^0 \pi^0\right)$	1.25×10^{-6}	1.35×10^{-6}	

Table B.6.: $B^0 \to \pi^0 \nu\bar{\nu}$

decay	$\mathscr{B}$ decay file	$\mathscr{B}$ pdg 15	comment
$B^+ \to \tau^+ \nu_\tau$	9.3×10^{-5}	1.14×10^{-4}	
$B^+ \to a_1^+\left(\to \rho^+ \pi^0\right)K^0_L$	1.745×10^{-5}	1.750×10^{-5}	

Table B.7.: $B^+ \to \rho^+ \nu\bar{\nu}$

decay	$\mathscr{B}$ decay file	$\mathscr{B}$ pdg 15	comment
$B^0 \to K_L^0 \eta' \left(\rho^0 \gamma \right)$	3.25×10^{-5}	3.30×10^{-5}	
$B^0 \to K_L^0 \rho^0$	2.7×10^{-6}	2.35×10^{-6}	
$B^0 \to K^{*0} (1430) \left(\to K_L^0 \pi^0 \right) \rho^0$	2.7×10^{-6}	2.35×10^{-6}	
$B^+ \to \rho^0 \mu^+ \nu_\mu$	1.49×10^{-4}	1.58×10^{-4}	
$B^0 \to \rho^- \mu^+ \nu_\mu$	2.77×10^{-4}	2.94×10^{-4}	
$B^0 \to \pi^- \mu^+ \nu_\mu$	1.36×10^{-4}	1.45×10^{-4}	
$B^0 \to K^{*0} \rho^0$	5.6×10^{-6}	3.9×10^{-6}	
$B^0 \to K^{*0} K_L^0 K_L^0$	1.14×10^{-5}	0.90×10^{-5}	from $B^0 \to K^{*0} K^+ K^-$

Table B.8.: $B^0 \to \rho^0 \nu \bar{\nu}$

channel	decay file	pdg 15	comment
$D^+ \to K^0 e^+ \nu_e$	6.8×10^{-2}	8.83×10^{-2}	
$D^+ \to K^0 \mu^+ \nu_\mu$	6.8×10^{-2}	9.3×10^{-2}	

Table B.9.: Corrected D Fractions

Appendix C

Continuum Suppression Performance

(a) $B^+ \to K^+ \nu\bar{\nu}$

(b) $B^0 \to K^0_S \nu\bar{\nu}$

(c) $B^+ \to K^{*+} \left(\to K^0_S \pi^+ \right) \nu\bar{\nu}$

(d) $B^+ \to K^{*+} \left(\to K^+ \pi^0 \right) \nu\bar{\nu}$

(e) $B^0 \to K^{*0}\nu\bar{\nu}$

(f) $B^+ \to \pi^+\nu\bar{\nu}$

(g) $B^0 \to \pi^0\nu\bar{\nu}$

(h) $B^0 \to \rho^0\nu\bar{\nu}$

— NB KSFW Moments
— NB Cleo Cones
— RooKSFW Moments + $|\cos\theta_B|$
— NB Cleo Cones + $|\cos\theta_B|$ + R2
— NB KSFW Moments + NB Cleo Cones + $|\cos\theta_B|$ + R2 + $\triangle z$

(i) $B^+ \to \rho^+\nu\bar{\nu}$
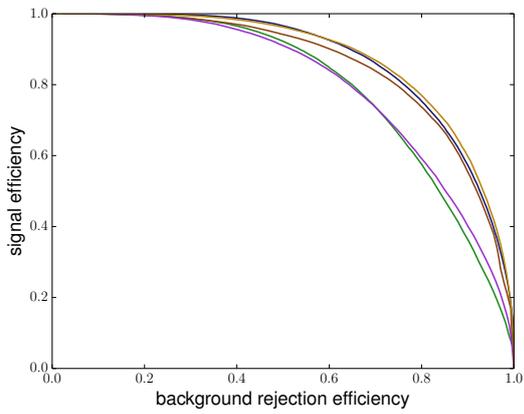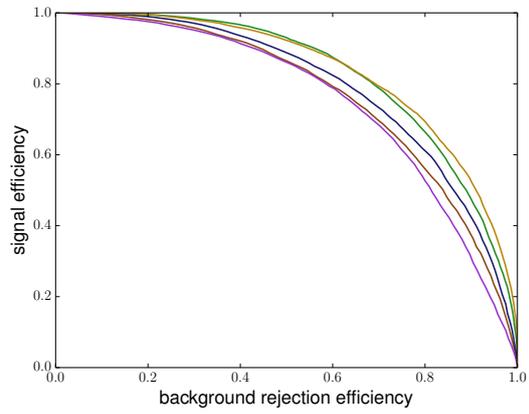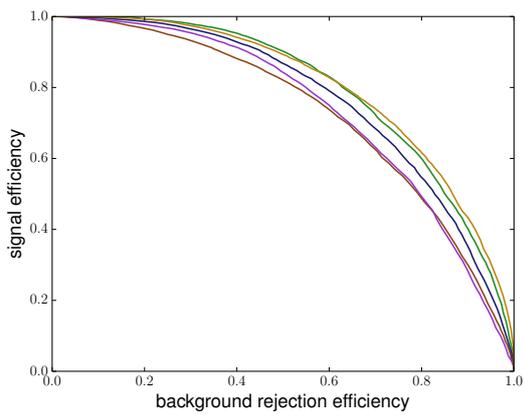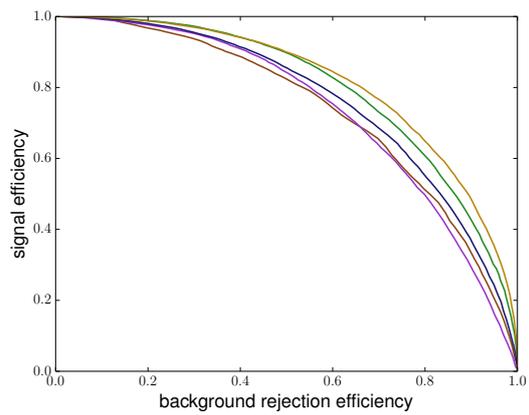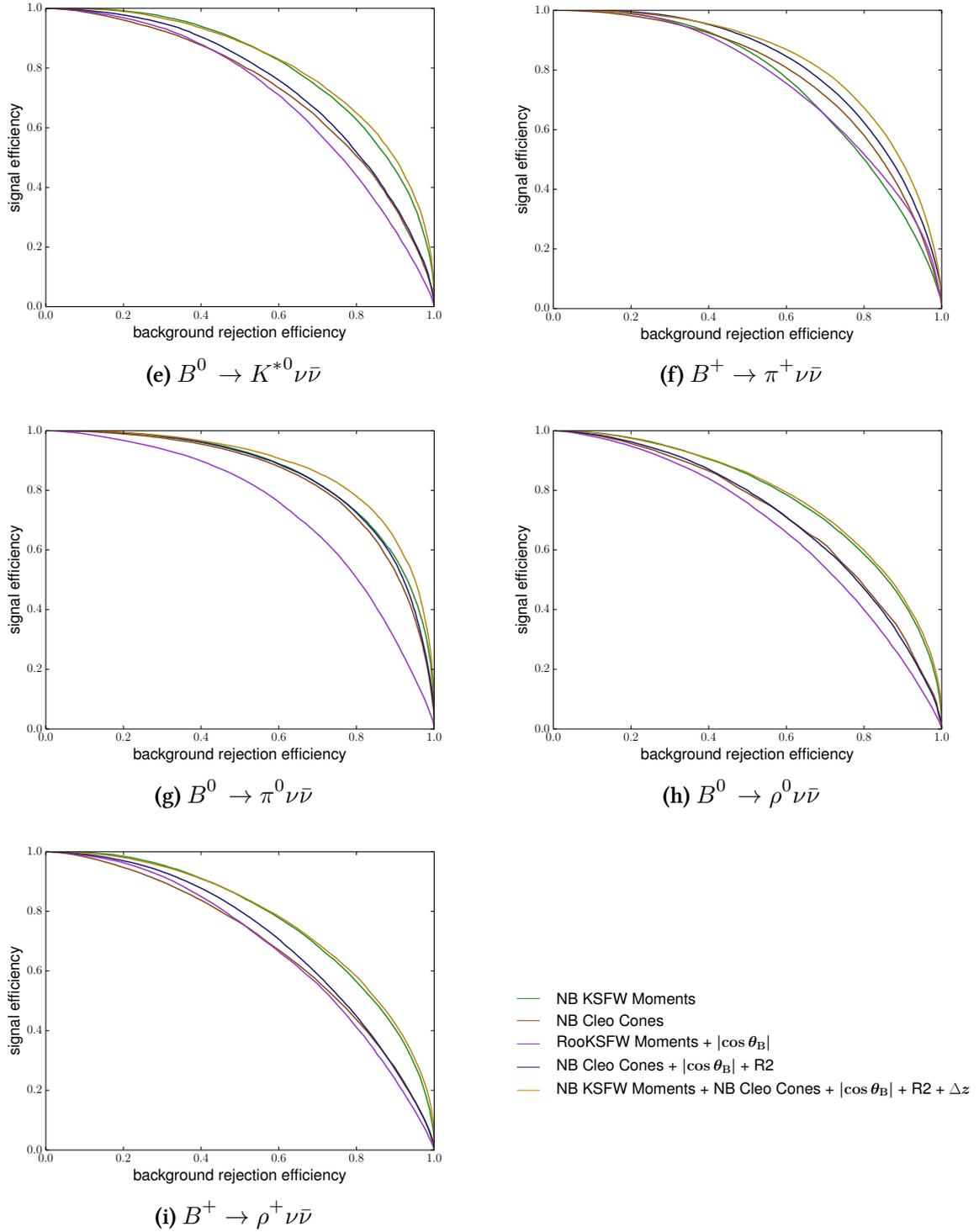
Figure C.1.: Continuum suppression final training results, signal efficiency over background rejection efficiency. The variable with the golden color has been chosen.

Appendix D

Network Properties

We list our best fit model for each channel with the input variables from Section 7.2.1 on page 55 characterized by the properties described in Section 4.1.1 on page 25. All variables entering the final selection are displayed. The resulting distributions are displayed in Figure D.1 on page 143 as well.

variable	significance	only	without	global correlation in %
$\mathscr{N}_{\text{NB, CS}}$	174.28	174.28	33.82	80.10
p_{cms}	71.22	151.68	31.38	89.20
$\cos\theta_{B,\,D^{(*)}l}^{>1}$	44.50	100.39	26.92	77.50
$\cos\theta_{B,\,D^{(*)}l}$	40.07	69.84	22.43	49.50
$\cos\theta_{\text{miss}}$	34.63	131.00	9.99	97.20
$N_{\text{remaining}}^{\text{rawtracks}}$	32.35	38.70	26.20	30.30
$\cos\theta_{\text{thrust}}$	26.91	26.54	15.33	73.70
$p_{l_{\text{tag}}}$	26.71	89.54	9.39	83.30
$M_{D_{\text{tag}}}$	25.49	28.03	19.57	31.90
$\Delta\chi^2$	19.27	73.49	16.49	40.40
$M_{\text{sig}}^{\text{min}}$	15.59	112.81	14.03	57.40
$\cos\theta_{D^{(*)}l}$	15.01	100.21	11.78	93.70
$\cos\theta_{h,\,l}$	10.68	93.48	8.32	90.40
$\mathscr{N}_{\text{NB, tag}}^{\text{rank}}$	10.85	8.10	7.81	38.40
$M_{h,\,l}$	9.58	131.63	11.44	76.80
M_{miss}	7.22	95.88	9.96	95.40
$p_{D_{\text{tag}}}$	8.42	57.10	7.35	73.40
$\mathscr{N}_{\text{NB, tag}}$	8.25	45.11	9.00	33.40
$\mathcal{P}_{\mu}^{l_{\text{tag}}}$	8.17	26.95	7.55	37.20
$\mathscr{N}_{\text{NB, tag}}^{high}$	8.31	55.34	8.41	40.90
$\cos\theta_{\text{sig, tag}}^{\text{lab}}$	6.77	24.87	8.93	53.20
$\cos\theta_{\text{sig, tag}}$	6.79	31.49	7.62	59.90
$\cos\theta_{\text{tag, miss}}$	4.99	128.87	6.12	98.00
$\cos\theta_{\text{sig, miss}}^{\text{lab}}$	5.83	95.32	5.14	89.90
E_{miss}	5.78	136.58	5.62	93.90
$M_{l^{\text{tag}}l'}$	4.90	94.30	4.82	67.30
Δr_{sig}	4.76	51.52	4.46	50.30
$\cos\theta_{D^{(*)}l}$	3.86	58.15	4.10	63.50
$\cos\theta_{\text{sig, miss}}$	4.03	109.69	4.52	96.00
$d_{\text{h, IP}}$	3.64	69.54	3.60	40.50
$\mathcal{P}_{K/\pi}$	3.69	67.20	3.52	48.50

Table D.1.: network training configuration of $B^+ \to K^+ \nu\bar{\nu}$

variable	significance	only	without	global correlation in %
$\mathcal{N}_{\text{NB, CS}}$	65.14	65.14	17.36	72.90
$\cos\theta_{B,\,D^{(*)}l}^{>1}$	29.98	43.19	16.49	76.50
M_{miss}	35.45	38.49	9.89	95.40
$\cos\theta_{\text{sig, miss}}$	14.18	39.07	3.64	96.60
$\cos\theta_{\text{thrust}}$	16.02	13.50	6.98	82.60
p_{cms}	11.95	57.84	13.80	85.40
$\mathcal{N}_{\text{NB, tag}}^{high}$	13.07	32.30	8.98	43.00
$M_{h,\,l}$	10.27	53.20	8.06	70.80
$M_{D_{\text{tag}}}$	10.06	12.33	8.14	33.80
$\cos\theta_{B,\,D^{(*)}l}$	8.40	31.41	8.45	51.70
$|p_{\text{miss}}|$	6.92	19.59	6.82	45.40
$\cos\theta_{D^{(*)}l}$	3.69	31.55	5.77	91.90
$\cos\theta_{\text{tag, miss}}$	6.58	45.61	4.40	97.10
E_{miss}	6.69	49.17	6.24	95.10
$\mathcal{N}_{\text{NB, tag}}$	5.79	13.67	5.00	24.50
$N_{\text{remaining}}^{\text{rawtracks}}$	5.67	9.74	6.11	22.20
$\cos\theta_{\text{sig, tag}}^{\text{lab}}$	5.46	9.31	5.94	66.00
$p_{D_{\text{tag}}}$	5.20	25.90	3.76	71.10
$\cos\theta_{h,\,l}$	4.48	33.74	4.17	91.10
$\mathcal{P}_{\mu}^{l_{\text{tag}}}$	4.30	8.63	4.56	41.70
$p_{l_{\text{tag}}}$	3.90	31.80	4.02	82.20
$\cos\theta_{\text{sig, tag}}$	3.78	8.98	4.25	72.80

Table D.2.: network training configuration of $B^0 \to K_S^0 \nu\bar{\nu}$

variable	significance	only	without	global correlation in %
$\mathcal{N}_{\text{NB, CS}}$	70.86	70.86	19.21	72.50
p_{cms}	45.45	68.96	16.54	83.40
$\cos\theta_{B,\,D^{(*)}l}^{>1}$	29.30	43.25	19.09	74.10
$\cos\theta_{K\pi}$	23.59	36.08	14.92	51.00
$p_{l_{\text{tag}}}$	17.07	32.96	5.50	83.60
$|p_{\text{miss}}|$	18.53	25.27	11.06	79.40
$N_{\text{remaining}}^{\text{rawtracks}}$	15.93	23.42	13.35	33.00
$M_{D_{\text{tag}}}$	15.34	20.77	11.42	37.80
$\cos\theta_{\text{tag, miss}}$	11.79	47.34	5.76	96.50
$\cos\theta_{\text{thrust}}$	10.92	15.85	5.58	82.80
$\cos\theta_{D^{(*)}l}$	11.39	32.00	6.75	89.10
$\cos\theta_{B,\,D^{(*)}l}$	9.57	31.16	7.27	51.70
M_{res}	8.12	15.80	7.82	10.00
$M_{\text{sig}}^{\text{min}}$	7.47	42.95	6.07	51.20
$\cos\theta_{\text{sig, miss}}$	7.41	45.16	3.18	95.90
$\mathcal{N}_{\text{NB, tag}}^{high}$	6.96	26.96	6.80	41.60
$\mathcal{N}_{\text{NB, tag}}$	6.51	25.17	6.05	35.20
$M_{\pi^0}^{D_{\text{tag}}}$	6.00	8.45	6.25	31.60
$\mathcal{N}_{\text{NB, tag}}^{\text{rank}}$	6.03	7.18	5.63	40.90
M_{π^0}	5.81	19.60	5.76	20.60
$\Delta\chi^2$	5.58	24.20	5.28	25.30
$\cos\theta_{h,\,l}$	4.07	35.24	3.75	85.10
$\cos\theta_{D^{(*)}l}$	3.23	18.78	3.02	57.10

Table D.3.: network training configuration of $B^+ \to K^{*+}\left(\to K^+\pi^0\right)\nu\bar{\nu}$

variable	significance	only	without	global correlation in %
$\cos\theta_{B,\,D^{(*)}l}^{>1}$	29.06	29.06	12.93	74.10
M_{miss}	38.34	27.24	8.03	94.30
$\mathscr{N}_{\mathrm{NB,\,CS}}$	25.71	39.99	9.14	72.80
p_{cms}	11.43	40.79	12.74	80.30
$\cos\theta_{\mathrm{thrust}}$	9.57	10.73	4.38	81.30
$\cos\theta_{\mathrm{miss}}$	9.91	32.55	2.60	96.90
$M_{D_{\mathrm{tag}}}$	10.04	11.73	6.61	39.30
$\cos\theta_{B,\,D^{(*)}l}$	7.54	21.50	7.22	50.90
$N_{\mathrm{remaining}}^{\mathrm{rawtracks}}$	5.76	11.55	5.08	25.50
$\mathscr{N}_{\mathrm{NB,\,tag}}^{high}$	5.60	17.69	5.13	40.50
$\mathscr{N}_{\mathrm{NB,\,tag}}$	5.22	15.31	4.52	35.60
$p_{D_{\mathrm{tag}}}$	4.64	16.87	2.69	72.30
$\cos\theta_{h,l}$	5.02	23.87	4.97	83.80
M_{res}	4.90	9.00	4.75	9.30
$\cos\theta_{D^{(*)}l}$	4.62	20.73	4.35	87.70
$\cos\theta_{D^{(*)}l}$	3.98	12.93	3.08	63.60
$|p_{\mathrm{miss}}|$	3.27	12.98	3.03	38.70

Table D.4.: network training configuration of $B^+ \to K^{*+}\left(\to K_{\mathrm{S}}^0\,\pi^+\right)\nu\bar{\nu}$

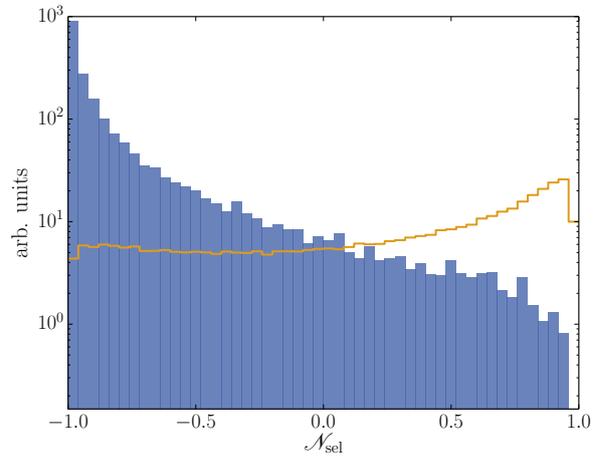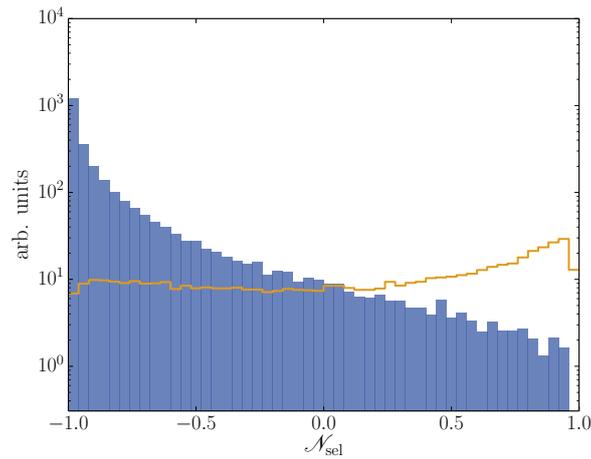| variable	significance	only	without	global correlation in %
p_{cms}	179.41	179.41	35.45	91.70
$\cos\theta_{\mathrm{miss}}$	80.95	120.78	8.31	96.80
$\cos\theta_{B,\,D^{(*)}l}^{\,>1}$	56.25	104.67	40.46	72.40
$\cos\theta_{B,\,D^{(*)}l}$	57.87	86.33	29.82	48.50
$\cos\theta_{\mathrm{thrust}}$	51.52	41.76	18.57	87.40
$N_{\mathrm{remaining}}^{\mathrm{rawtracks}}$	46.90	62.92	38.30	33.30
$\mathscr{N}_{\mathrm{NB,\,CS}}$	38.77	181.46	22.08	80.40
$\Delta\chi^2$	31.49	85.16	26.71	37.20
$M_{D_{\mathrm{tag}}}$	28.61	47.61	22.71	35.10
$p_{l_{\mathrm{tag}}}$	26.14	88.62	13.27	77.80
$\cos\theta_{h_{\mathrm{sig}}}$	18.99	73.48	8.63	88.50
$\mathscr{P}_{\mu}$	17.82	16.97	17.66	7.50
$\mathscr{N}_{\mathrm{NB;\,tag}}$	17.65	61.46	16.03	35.90
$\mathscr{N}_{\mathrm{NB,\,tag}}^{high}$	17.56	65.60	14.04	41.60
$\mathscr{N}_{\mathrm{NB,\,tag}}^{\mathrm{rank}}$	14.46	20.37	12.53	35.10
$\cos\theta_{h,\,l}$	13.44	85.21	9.69	86.00
M_{miss}	11.13	136.89	10.43	96.20
Δr_{sig}	9.52	58.09	7.27	57.10
$|p_{\mathrm{miss}}|$	9.10	35.87	10.14	29.90
$M_{h,\,l}$	8.22	117.93	8.44	78.00
$p_{D_{\mathrm{tag}}}$	8.05	58.51	7.17	65.80
$\mathscr{P}_{\mu}^{l_{\mathrm{tag}}}$	6.40	20.98	5.60	32.80
$\cos\theta_{D^{(*)}l}^{\mathrm{lab}}$	5.87	91.86	4.61	92.90
$\cos\theta_{\mathrm{sig,\,miss}}$	4.87	84.89	5.81	86.80
$\mathscr{P}_{K/\pi}$	4.75	99.21	5.42	56.60
$\cos\theta_{D^{(*)}l}$	4.53	50.29	4.59	56.70
$M_{l^{\mathrm{tag}}l'}$	4.15	95.80	4.07	67.10
$\cos\theta_{\mathrm{sig,\,tag}}^{\mathrm{lab}}$	4.18	23.75	4.34	43.50
$M_{\pi^0}^{D^*}$	3.14	7.97	3.10	56.20

Table D.5.: network training configuration of $B^+ \to \pi^+ \nu\bar{\nu}$

variable	significance	only	without	global correlation in %
$\mathscr{N}_{\text{NB, CS}}$	121.03	121.03	35.25	77.30
p_{cms}	48.41	113.88	19.61	91.50
$\cos\theta_{B,\,D^{(*)}l}^{>1}$	36.96	61.47	24.24	72.80
$\cos\theta_{B,\,D^{(*)}l}$	25.28	54.08	13.58	53.00
$\mathscr{N}_{\text{NB, tag}}^{high}$	21.12	52.43	13.70	50.20
$\cos\theta_{\text{thrust}}$	17.20	27.99	4.93	91.70
$M_{D_{\text{tag}}}$	12.41	28.14	12.95	37.00
$p_{l_{\text{tag}}}$	13.45	43.80	7.93	76.10
$N_{\text{remaining}}^{\text{rawtracks}}$	13.46	22.22	13.50	23.50
$\mathscr{N}_{\text{NB, tag}}$	10.48	27.05	8.89	28.50
$\cos\theta_{\text{tag, miss}}$	8.16	61.51	6.24	97.90
$\cos\theta_{D^{(*)}l}$	9.07	42.64	7.50	94.20
$\lvert p_{\text{miss}}\rvert$	9.25	72.58	7.49	88.20
$\cos\theta_{\text{sig, tag}}^{\text{lab}}$	7.72	9.78	5.81	66.80
$\cos\theta_{h,l}$	7.75	43.08	5.99	84.10
$M_{\pi^0}^{D_{\text{tag}}}$	7.31	6.12	6.52	32.80
$\cos\theta_{D^{(*)}l}$	6.70	27.30	4.62	59.90
$\cos\theta_{\text{sig, tag}}$	6.66	7.36	6.79	76.20
$\cos\theta_{\text{sig, miss}}$	5.04	45.40	5.81	96.70
$p_{D_{\text{tag}}}$	4.31	38.07	4.90	64.10
$M_{h,l}$	4.30	67.34	4.01	70.60
$\Delta M_{D^*D}^{\text{tag}}$	4.01	11.28	4.13	50.20
$\cos\theta_{\text{sig, miss}}^{\text{lab}}$	3.64	37.74	1.80	88.80
$\mathcal{P}_\mu^{l_{\text{tag}}}$	3.01	12.58	2.77	39.60

Table D.6.: network training configuration of $B^0 \to \pi^0 \nu\bar{\nu}$

variable	significance	only	without	global correlation in %
$\mathcal{N}_{\text{NB, CS}}$	129.48	129.48	40.31	70.30
p_{cms}	80.57	123.78	27.95	84.50
$\cos\theta_{B,\,D^{(*)}l}^{>1}$	61.29	81.14	38.81	68.90
$\cos\theta_{K\pi}$	45.35	50.11	27.28	42.10
$\cos\theta_{B,\,D^{(*)}l}$	39.04	67.53	19.96	48.90
$N_{\text{remaining}}^{\text{rawtracks}}$	36.61	51.49	29.34	34.20
$|p_{\text{miss}}|$	29.56	71.67	28.69	80.10
$\mathcal{N}_{\text{NB, tag}}$	26.37	54.49	17.02	37.10
$\mathcal{N}_{\text{NB, tag}}^{high}$	25.61	54.97	16.19	41.10
$\cos\theta_{\text{sig, miss}}$	24.90	68.69	8.01	96.00
$\cos\theta_{\text{thrust}}$	23.02	38.67	10.40	78.10
$M_{D_{\text{tag}}}$	20.10	50.81	22.07	37.60
$\Delta\chi^2$	16.98	48.92	15.83	24.10
$p_{l_{\text{tag}}}$	16.95	54.22	8.71	75.70
M_{res}	16.79	31.61	16.22	9.40
$M_{\pi^0}^{D_{\text{tag}}}$	13.49	15.01	11.83	34.00
$\mathcal{N}_{\text{NB, tag}}^{\text{rank}}$	11.21	17.32	9.70	37.60
$\cos\theta_{h,l}$	10.58	58.43	9.25	77.90
$\cos\theta_{D^{(*)}l}$	10.35	25.87	10.09	48.20
M_{π^0}	8.25	28.21	8.43	19.40
$\cos\theta_{\text{tag, miss}}^{\text{lab}}$	8.04	58.83	5.59	79.30
$M_{h,l}$	6.15	57.07	9.16	74.70
E_{miss}	7.35	91.29	6.23	95.10
$\cos\theta_{\text{sig, miss}}^{\text{lab}}$	7.13	55.23	6.21	83.30
$p_{D_{\text{tag}}}$	6.52	46.99	6.37	63.20
$\cos\theta_{\text{thrust}}^{\text{sig,tag}}$	5.48	35.14	5.41	78.00
$\cos\theta_{\text{tag, miss}}$	4.20	76.00	4.16	97.30
$\cos\theta_{D^{(*)}l}$	4.22	50.99	4.48	91.40
$\cos\theta_{\text{sig, tag}}$	4.92	20.34	3.99	59.30
$\Delta z_{l^{\text{tag}}h^{\text{sig}}}$	4.73	26.29	4.73	19.10
$M_{l^{\text{tag}}l'}$	3.52	43.78	3.77	61.40
$\Delta M_{D^*D}^{\text{tag}}$	3.06	15.62	4.25	64.60

Table D.7.: network training configuration of $B^+ \to \rho^+\nu\bar{\nu}$

variable	significance	only	without	global correlation in %
p_{cms}	140.14	140.14	38.53	84.60
$\mathscr{N}_{\mathrm{NB,\,CS}}$	80.05	133.52	40.82	66.70
$\cos\theta_{B,\,D^{(*)}l}^{>1}$	60.56	81.27	42.71	68.50
M_{res}	41.88	55.36	37.07	14.30
$\cos\theta_{B,\,D^{(*)}l}$	37.65	65.74	15.41	52.50
$\mathscr{N}_{\mathrm{NB,\,tag}}^{high}$	31.37	69.95	17.50	52.80
$|p_{\mathrm{miss}}|$	20.11	41.89	15.07	72.70
$\cos\theta_{\mathrm{sig,\,miss}}$	23.25	60.58	4.17	95.50
$\cos\theta_{K\pi}$	20.89	17.19	17.22	18.30
χ^2_{sig}	20.84	34.87	17.00	26.20
$\cos\theta_{\mathrm{thrust}}$	19.74	34.24	9.61	73.50
$M_{D_{\mathrm{tag}}}$	17.73	44.11	16.64	36.90
M_{miss}	18.30	99.44	7.55	95.60
$N_{\mathrm{remaining}}^{\mathrm{rawtracks}}$	16.28	32.27	16.13	21.60
$\mathscr{N}_{\mathrm{NB,\,tag}}$	15.04	39.72	11.63	31.80
$\mathscr{N}_{\mathrm{NB,\,tag}}^{\mathrm{rank}}$	10.67	25.17	11.24	19.30
$\cos\theta_{h,\,l}$	9.37	50.47	8.74	78.80
$\cos\theta_{D^{(*)}l}$	9.88	25.08	8.75	50.20
$\Delta z_{l^{\mathrm{tag}}h^{\mathrm{sig}}}$	10.16	34.21	8.96	30.30
$p_{l_{\mathrm{tag}}}$	9.19	45.85	8.29	73.10
$\Delta M_{D^*D}^{\mathrm{tag}}$	8.06	24.10	7.96	46.90
$M_{h,\,l}$	5.99	62.81	6.37	69.00
$p_{D_{\mathrm{tag}}}$	6.49	49.48	6.09	61.90
$\Delta\chi^2$	5.95	33.15	5.73	19.30
$\cos\theta_{\mathrm{thrust}}^{\mathrm{sig,tag}}$	5.49	31.31	5.83	73.30
$\cos\theta_{\mathrm{miss}}$	5.02	69.97	3.81	95.70
$\cos\theta_{D^{(*)}l}$	4.44	37.71	5.81	88.40
$\cos\theta_{\mathrm{sig,\,tag}}$	4.87	17.68	3.35	57.50
$\cos\theta_{\mathrm{tag,\,miss}}^{\mathrm{lab}}$	3.92	44.86	4.24	71.40
$\cos\theta_{\mathrm{sig,\,miss}}^{\mathrm{lab}}$	4.44	48.63	5.55	82.60
$\cos\theta_{h_{\mathrm{sig}}}$	3.63	58.78	3.68	89.30

Table D.8.: network training configuration of $B^0 \to \rho^0 \nu\bar{\nu}$
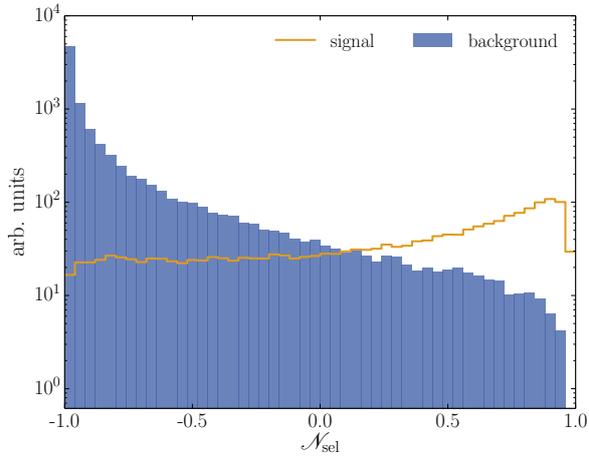
(a) $B^+ \rightarrow K^+ \nu\bar{\nu}$
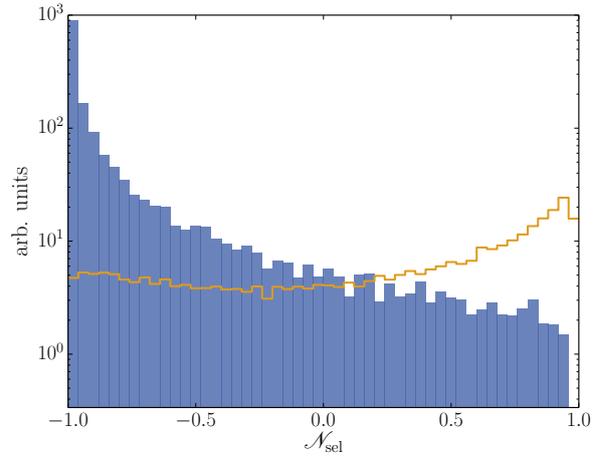
(b) $B^0 \rightarrow K_S^0 \nu\bar{\nu}$
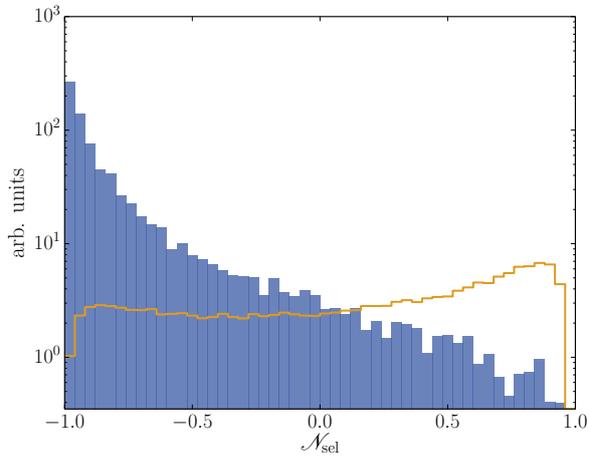
(c) $B^+ \rightarrow K^{*+} \left(\rightarrow K_S^0 \pi^+ \right) \nu\bar{\nu}$

(d) $B^+ \rightarrow K^{*+} \left(\rightarrow K^+ \pi^0 \right) \nu\bar{\nu}$

(e) $B^0 \rightarrow K^{*0} \nu\bar{\nu}$

(f) $B^+ \to \pi^+ \nu\bar{\nu}$

(g) $B^0 \to \pi^0 \nu\bar{\nu}$

(h) $B^0 \to \rho^0 \nu\bar{\nu}$

(i) $B^+ \to \rho^+ \nu\bar{\nu}$
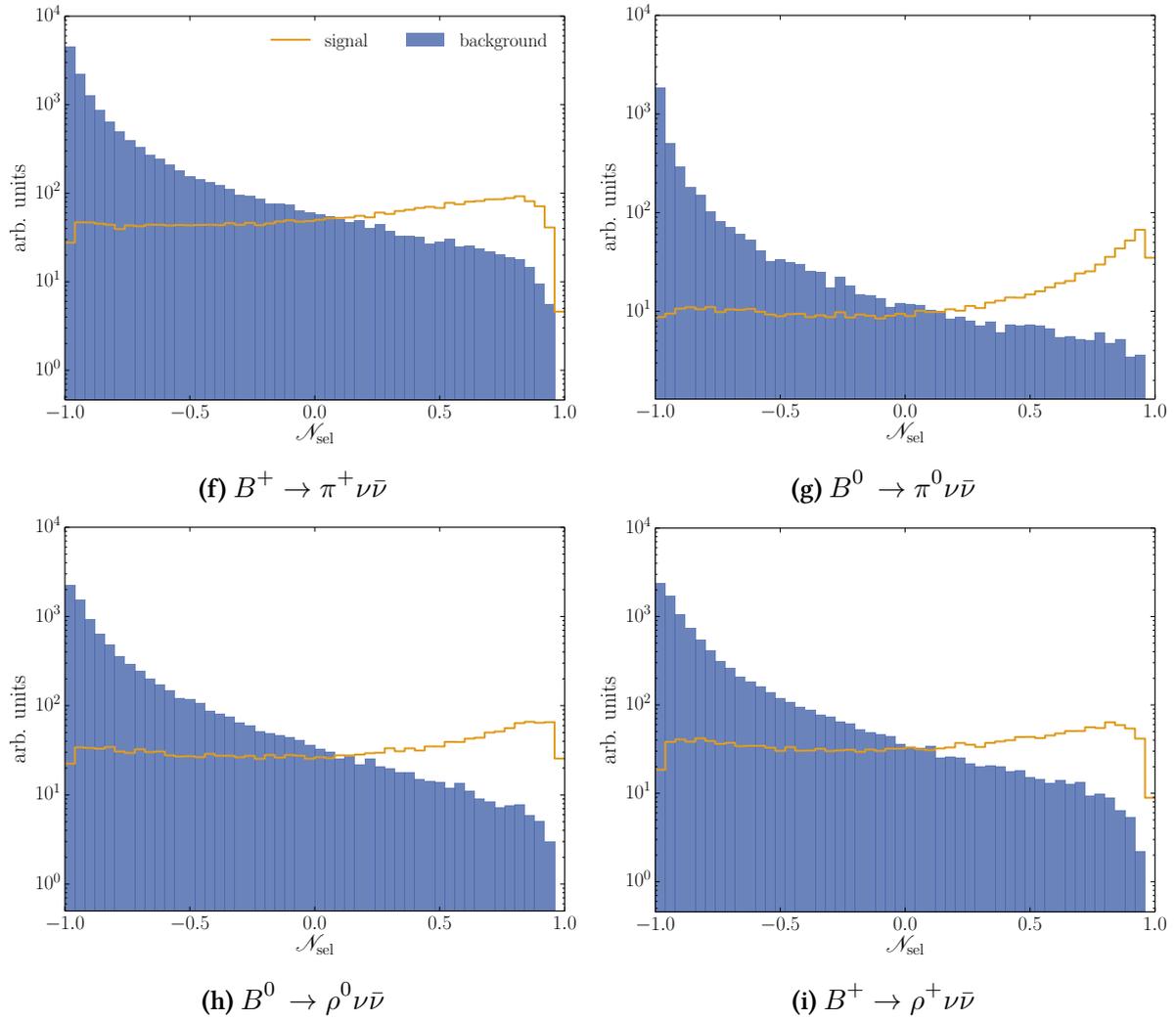
Figure D.1.: Background and weighted signal distributions of all trainings. The signal to background ratio has been arbitrarily fixed to a value of 0.2.

Appendix E

Data-MC Comparison for the Full Selected Sample

We compare the distributions of the MC predictions with those measured on data for the final net output $\mathscr{N}_{\text{sel}}$ and key input variables of the respective net. We show those distributions for the three channels in which we observe a significance of more than $1\,\sigma$. This is done to check whether the possible signal is compatible with the data.
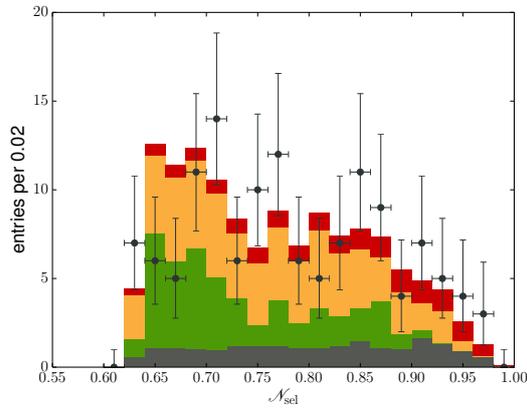
(a) `final net output`
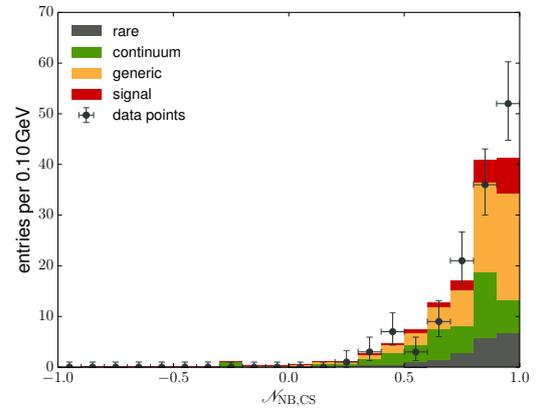
(b) `continuum suppression`
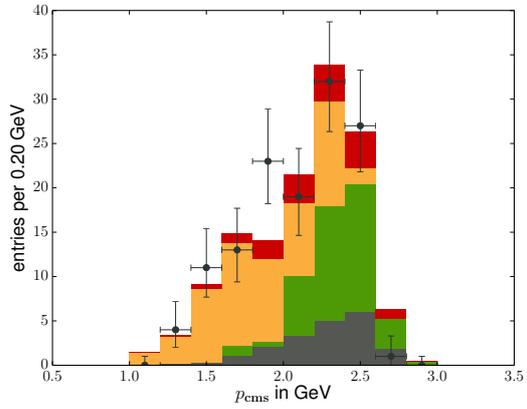
(c) `momentum of the signal particle`
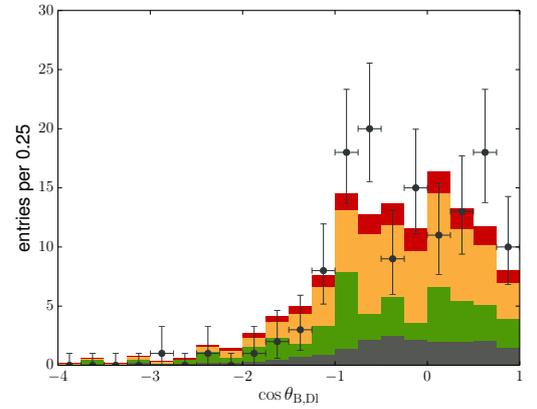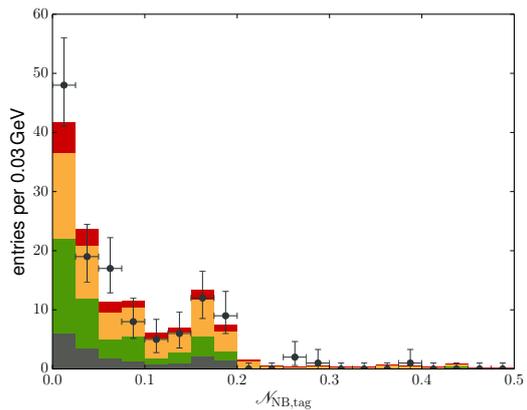
(d) $\cos\theta_{B,\,D^{(*)}l}$

(e) `output of the tagging network`
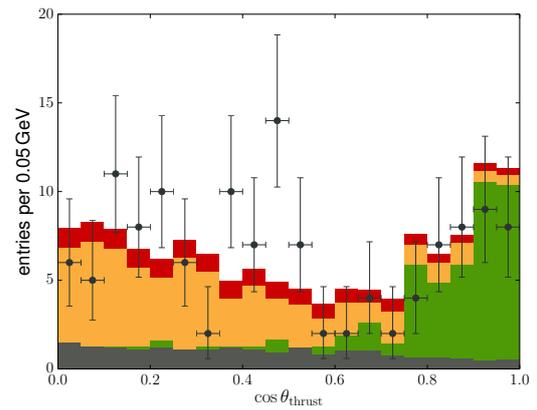
(f) `cosine of the thrust axis.`

Figure E.1.: data-MC comparison for the $B^+ \to K^+ \nu\bar{\nu}$ channel
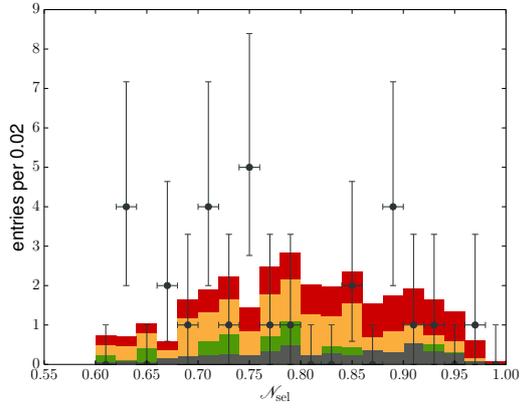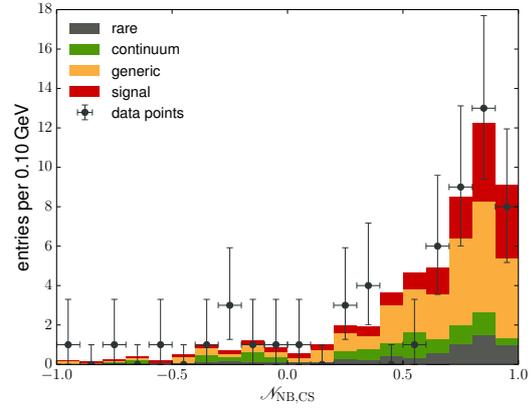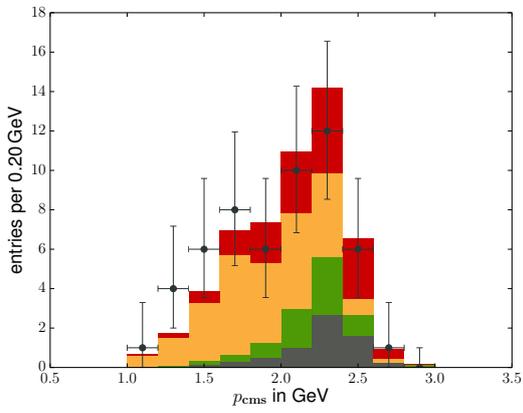
(a) `final net output`

(b) `continuum suppression`

(c) `momentum of the signal particle`
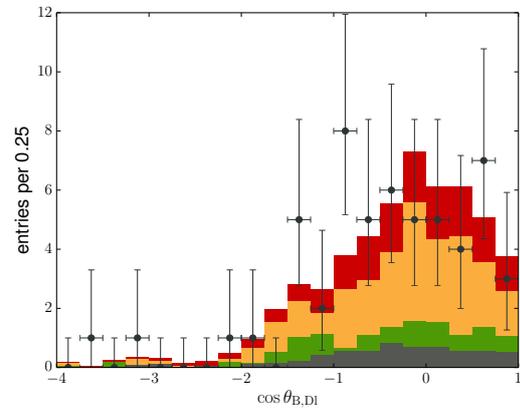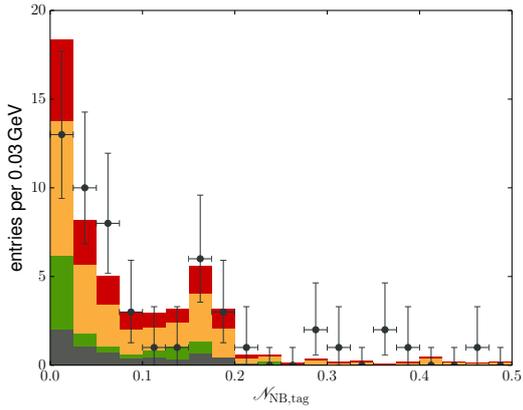
(d) $\cos\theta_{B, D^{(*)}l}$

(e) `output of the tagging network`
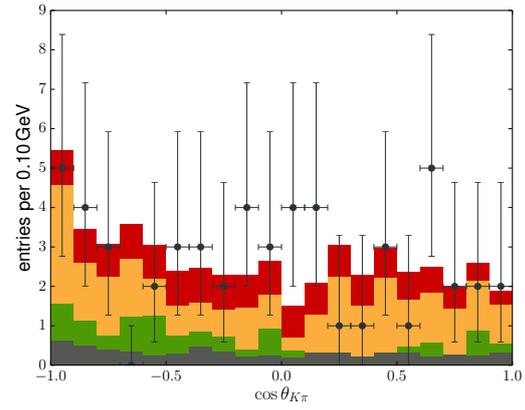
(f) `cosine between the momenta of the` ρ `daughters`

Figure E.2.: data-MC comparison for the $B^+ \to K^{*+}\nu\bar{\nu}$ channel
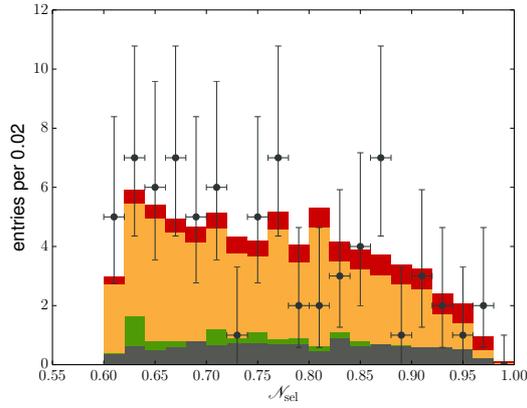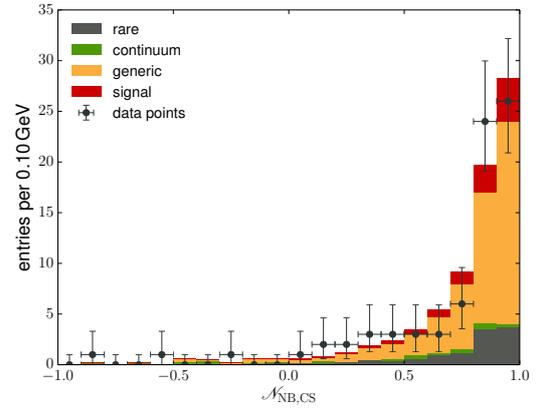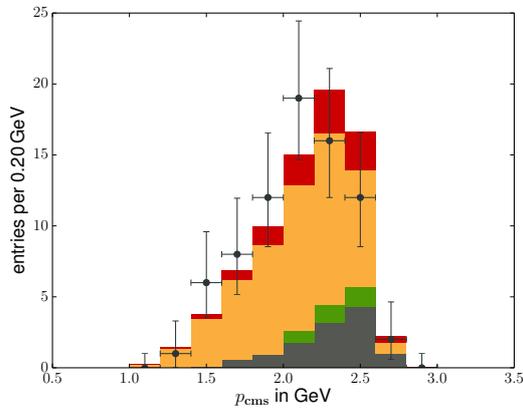
(a) `final net output`

(b) `continuum suppression`

(c) `momentum of the signal particle`
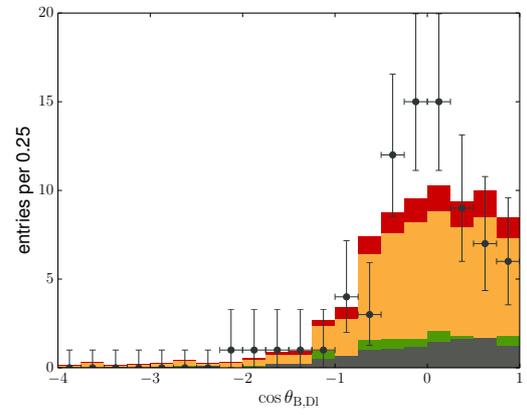
(d) $\cos\theta_{B, D^{(*)}l}$
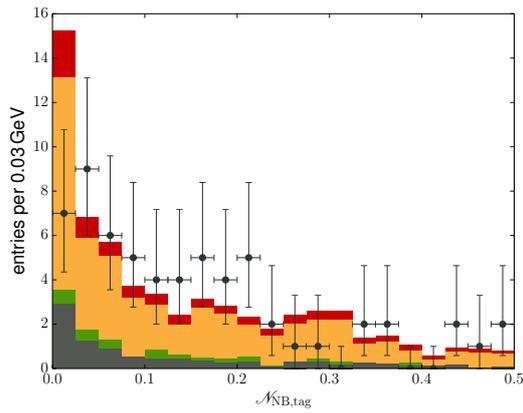
(e) `output of the tagging network`
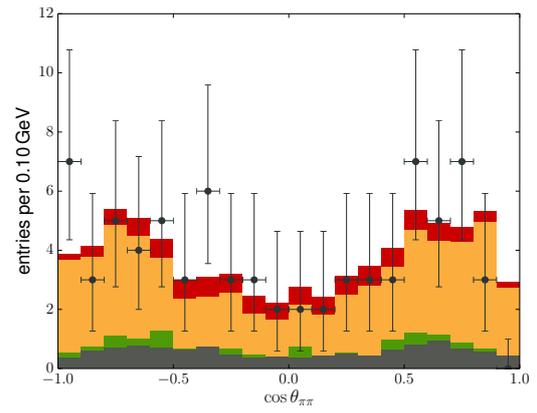
(f) `cosine between the momenta of the` ρ `daughters`

Figure E.3.: data-MC comparison for the $B^0 \rightarrow \rho^0 \nu\bar{\nu}$ channel

Bibliography

[A$^+$96]   D. M. Asner et al. Search for exclusive charmless hadronic *B* decays. *Phys. Rev. D*, 53:1039–1050, Feb 1996.

[A$^+$02]   A. Abashian et al. The belle detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 479(1):117 – 232, 2002. Detectors for Asymmetric B-factories.

[A$^+$13a]   R Aaij et al. Measurement of B meson production cross-sections in proton-proton collisions at $\sqrt{s}$ = 7 TeV. *JHEP*, 08:117, 2013.

[A$^+$13b]   R Aaij et al. Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$. *Phys. Rev. Lett.*, 111:191801, 2013.

[A$^+$14a]   Roel Aaij et al. Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays. *Phys. Rev. Lett.*, 113:151601, 2014.

[A$^+$14b]   A. Abdesselam et al. Measurement of the branching fraction of $B^+ \to \tau^+\nu_\tau$ decays with the semileptonic tagging method and the full Belle data sample. In *8th International Workshop on the CKM Unitarity Triangle (CKM 2014) Vienna, Austria, September 8-12, 2014*, 2014.

[A$^+$16a]   Roel Aaij et al. Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity. *JHEP*, 02:104, 2016.

[A$^+$16b]   A. Abdesselam et al. Measurement of the branching ratio of $\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method. 2016.

[B$^+$01]   T. E. Browder et al. A Search for B —> tau nu. *Phys. Rev. Lett.*, 86:2950–2954, 2001.

[B$^+$14]   A. J. Bevan et al. The Physics of the B Factories. *Eur. Phys. J.*, C74:3026, 2014.

[BBBG09]   M. Bartsch, M. Beylich, G. Buchalla, and D. N. Gao. Precision Flavour Physics with B —> K nu anti-nu and B —> K l+ l-. *JHEP*, 11:011, 2009.

[Bel00]   Belle Tracking Group. Charged particle tracking in belle, July 2000.

[BFK15]   Damir Bečirević, Svjetlana Fajfer, and Nejc Košnik. Lepton flavor nonuniversality in b→s$\ell^+\ell^-$ processes. *Phys. Rev.*, D92(1):014016, 2015.

[BGNNS15]   Andrzej J. Buras, Jennifer Girrbach-Noe, Christoph Niehoff, and David M. Straub. $B \to K^{(*)} \nu \overline{\nu}$ decays in the Standard Model and beyond. *JHEP*, 02:184, 2015. v1.

[BLNZ95]   Richard H. Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[BS93]   D Besson and T Skwarnicki. Upsilon spectroscopy: Transitions in the bottomonium system. *Annual Review of Nuclear and Particle Science*, 43(1):333–378, 1993.

[C$^+$05]   K. F. Chen et al. Time-dependent CP-violating asymmetries in b —> s anti-q q transitions. *Phys. Rev.*, D72:012004, 2005.

[dAS$^+$10]   P. del Amo Sanchez et al. Search for the Rare Decay $B \to K \nu \bar{\nu}$. *Phys. Rev.*, D82:112002, 2010.

[FK06]   M. Feindt and U. Kerzel. The neurobayes neural network package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(1):190 – 194, 2006. Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics ResearchA-CAT 05X International Workshop on Advanced Computing and Analysis Techniques.

[FKK$^+$11]   M. Feindt, F. Keller, M. Kreps, T. Kuhr, S. Neubauer, D. Zander, and A. Zupanc. A hierarchical neurobayes-based algorithm for full reconstruction of b mesons at b factories. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 654(1):432 – 440, 2011.

[H$^+$13]   K. Hara et al. Evidence for $B^- \to \tau^- \overline{\nu}_\tau$ with a hadronic tagging method using the full data sample of belle. *Phys. Rev. Lett.*, 110:131801, Mar 2013.

[Hei15]   Manuel Heider. *Dalitz analysis of $B^- \to D^0 \pi^- \pi^0$ and $B^0 \to D^- \pi^+ \pi^0$ at the Belle experiment*. PhD thesis, KIT, 2015.

[HKR15]   Christian Hambrock, Alexander Khodjamirian, and Aleksey Rusov. Hadronic effects and observables in $B \to \pi \ell^+ \ell^-$ decay at large recoil. *Phys. Rev.*, D92(7):074020, 2015.

[I$^+$00]   H. Ikeda et al. A detailed test of the csi(tl) calorimeter for belle with photon beams of energy between 20 mev and 5.4 gev. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 441(3):401 – 426, 2000.

[K$^+$15]   B. Kronenbitter et al. Measurement of the branching fraction of $B^+ \to \tau^+ \nu_\tau$ decays with the semileptonic tagging method. *Phys. Rev.*, D92(5):051102, 2015.

[Kir12]   Karsten Kirchgessner. Semileptonische markierungsseiten-rekonstruktion, 2012.

[KK03]   S. Kurokawa and E. Kikutani. Overview of the {KEKB} accelerators. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 499(1):1 – 7, 2003. KEK-B: The {KEK} B-factory.

[Kop03]   P. Koppenburg. A measurement of the track finding efficiency using partially recon-structed d^* decays, may 2003. bn 621.

[L$^+$13a]   J. P. Lees et al. Search for $B \to K^{(*)}\nu\overline{\nu}$ and invisible quarkonium decays. *Phys. Rev.*, D87(11):112005, 2013.

[L$^+$13b]   O. Lutz et al. Search for $B \to h^{(*)}\nu\bar{\nu}$ with the full Belle $\Upsilon(4S)$ data sample. *Phys. Rev.*, D87(11):111103, 2013.

[LCH03]   S.W. Lin, P. Chang, and H.C. Huang. Update of π^0 systematics using inclusive η $\left(78\,\mathrm{fb}^{-1}\right)$, may 2003. bn 645.

[Nak02]   E Nakano. Belle {PID}. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 494(1−3):402 − 408, 2002. Proceedings of the 8th International Conference on Instrumentatio n for Colliding Beam Physics.

[Nis05]   S. Nishida. Study of kaon and pion identification using inclusive d^* sample, January 2005. bn 779.

[O$^+$14]   K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.

[P$^+$13]   M. Prim et al. Angular analysis of $B^0 \to \phi K^*$ decays and search for CP violation at Belle. *Phys. Rev.*, D88(7):072004, 2013.

[Pun03]   Giovanni Punzi. Sensitivity of searches for new signals and its optimization. *eConf*, C030908:MODT002, 2003.

[SIK$^+$02]   Soh Yamagata Suzuki, Ryosuke Itoh, Hyun-Woo Kim, Hong-Joo Kim, Hyun-Ok Kim, Mikihiko Nakao, Eunil Won, and Masanori Yamauchi. Belle {DAQ} system upgrade at 2001. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 494(1−3):535 − 540, 2002. Proceedings of the 8th International Conference on Instrumentatio n for Colliding Beam Physics.

[The16]   The NA62 collaboration. 2016 na62 status report to the cern spsc. http://na62.web.cern.ch/NA62/Documents/, 2016.

[Whi11]   E. White. Determination of K^0_S efficiency and systematic uncertainty., july 2011. bn 1207.

[Wik06]   Wikimedia Commons. Standard model of elementary particles, June 2006. published by MissMJ under CC 3.0.

Danksagung