Search for $B^+ \to \ell^+\nu_\ell\gamma$ decays with hadronic tagging using the full Belle data sample

Andreas Heller

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. Andreas Heller

aus Verl

Contents

4

1 Introduction

The Standard Model (SM) is the foundation of particle physics which provides a framework for all known particles and their interactions. It has been able to correctly reproduce experimental results at particle colliders for the last forty years. However, experimental findings and theoretical considerations indicate that an extension of the SM into a more complete model is necessary. The first evidence for physics beyond the SM is the discovery of neutrino masses through neutrino oscillation. Another important phenomenon which cannot be understood within the SM is dark matter since no known particle is a viable candidate to describe all of its properties. Additional important theoretical concerns which arise in the SM are: the fine tuning problem of the Higgs mass; the baryogenesis in the early universe; the missing concept of a Grand Unified Theory which also includes quantum gravity; and the hierarchy of the particle masses as well as the weak mixing angles. Many compelling arguments are given for the search of physics beyond the SM. Extensions of the SM postulate new particles to account for the unexplained effects.

The Belle experiment and its successor, the Belle II experiment, are designed to investigate the B meson system to the highest precision. The most important result by Belle was the measurement of CP violation in the B system in several decay channels. The CP violation is caused by an irreducible complex phase in the Cabibbo-Kobayashi-Maskawa (CKM) matrix which contains the weak mixing angles between the quarks. This mixing mechanism is studied intensely to look for new CP-violating phases which would be a sign of New Physics. Additionally, branching fractions and polarizations in B decays provide information about the flavor and spin structure of the weak decay and ultimately about New Physics. New particles which couple to the SM particles are expected to be too massive to be produced directly at the low center-of-mass energy. However, they can contribute to the decays as virtual particles in higher-order loop processes. These contributions will result in small deviations to the SM predictions.

The biggest obstacle for precise theoretical predictions is posed by strong interactions at low energies which are present in all B decays. These cannot be calculated perturbatively due to the size of the strong coupling constant which is of order one resulting in a non-convergent power series. This is solved by moving the non-perturbative parts of the calculation into parameters which are determined in the experiment. Non-perturbative calculations of these quantities are, in many cases, difficult and have large errors.

With the measurement of the $B^+ \to \ell^+ \nu_\ell \gamma$ decay, a yet unmeasured parameter of the B meson can be determined which is called λ_B. This is needed in a QCD

factorization approach [1] which provides an important ansatz to compute hadronic B decays in two mesons. Specifically, the parameter is an input in the computation of charmless hadronic decays where the B meson decays in two light mesons with a quark content of up, down, or strange. This comprises many decay channels where the parameter introduces a sizable uncertainty. Especially the calculations for color-suppressed modes $B \to \pi\pi, \pi\rho$, and $\rho\rho$ have large errors which are dominated by this parameter and experimental data hints to values of $\lambda_B \sim 200\,\mathrm{MeV}$. Non-perturbative calculations with QCD-sum-rules yield disagreeing results of $\lambda_B \sim (350-500)\,\mathrm{MeV}$ [2]. The measurement of $B^+ \to \ell^+\nu_\ell\gamma$ is the only way to resolve this tension experimentally and ultimately determine whether inconsistencies persist in the QCD factorization approach.

The best limit for the decay is reported by the BaBar collaboration with a branching fraction of $\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) < 14 \times 10^{-6}$ resulting in a limit of about $\lambda_B > 120\,\mathrm{MeV}$ at 90% confidence level. This limit is insufficient to provide stringent constraints for the theory. The BaBar analysis is performed with a simple cut-based method which has room for improvement to achieve a more significant result. This work presents the first measurement of this decay by Belle.

I performed an analysis of the $B^+ \to \ell^+\nu_\ell\gamma$ decay with the full Belle dataset of $772 \times 10^6 B\bar{B}$ pairs. Here, the charged lepton is either an electron or a muon and two separate analyses are performed with different energy requirements on the signal photon.

The thesis is structured as follows. A discussion of the QCD factorization approach as well as a summary of the latest result of the calculation for the $B^+ \to \ell^+\nu_\ell\gamma$ decay is given in Section 2. Section 3 contains different analysis techniques which are essential to the analysis. The signal selection is described in Section 4. A detailed description of the fit model which is used to extract the signal as well as an examination of the fit bias is given in Section 5. Before the fit on data, sideband distributions are examined in Section 6 and the systematic error of the procedure is estimated in Section 7. For the systematic error, a control channel of $B^+ \to K^*(892)\gamma$ is analyzed to obtain the error on a neural network estimator which is crucial to the analysis. The measurement on data is described in Section 8, which is followed by a discussion with respect to the BaBar result mentioned above. Finally, the conclusion to the analysis is given in Section 9.

2 Theoretical overview

Calculations of B decays always contain low-energy QCD effects which have to be absorbed in hadronic parameters. Accordingly, the precision of the predictions depend on the knowledge of these parameters. One such parameter, λ_B, is needed in the QCD factorization approach [1] which provides an important ansatz to compute B decays into two mesons. The parameter is called the *inverse first moment of the B meson light-cone distribution amplitude* and is defined as

$$\frac{1}{\lambda_B} = \int_0^\infty d\omega \frac{\Phi_{B^+}(\omega)}{\omega}. \tag{1}$$

Here, ω describes the energy of the light quark and $\Phi_{B^+}(\omega)$ denotes the wave function of the quark inside the B meson in a highly boosted system. (The wave function has only one parameter since the energy of both quarks is fixed by the total B meson energy which leaves one degree of freedom.) The λ_B parameter is the integral of this distribution amplitude divided by the quark energy.

Theoretical calculations with light-cone sum rules yield values of $\lambda_B \sim (350 - 500)$ MeV [2]. This is in tension with measurements of $B \to \pi\pi, \pi\rho$, and $\rho\rho$ decays which favor lower values of $\lambda_B \sim 200$ MeV [3]. These decay modes are particularly sensitive to λ_B since they have relatively large color-suppressed amplitudes which are strongly influenced by the value of λ_B. In table 2.1 the measured branching fractions are compared to the theoretical predictions obtained with $\lambda_B = 200$ MeV and $\lambda_B = 400$ MeV. The lower value for λ_B improves the agreement between theory and experiment for nearly all decay modes. In general, results from the QCD factorization are in good agreement with experimental measurements which proves the correctness of the procedure.

A notable exception is the $B^0 \to \pi^0\pi^0$ decay which shows a large discrepancy between theory and experiment. However, this has to be taken with a grain of salt since the decay is experimentally challenging. A preliminary result by Belle on the final data set reports a branching fraction of $\mathcal{B}(B^0 \to \pi^0\pi^0) = (0.89 \pm 0.12 \pm 0.10) \times 10^{-6}$ [5] which is more than 3σ below the BaBar measurement of $\mathcal{B}(B^0 \to \pi^0\pi^0) = (1.83 \pm 0.21 \pm 0.13)$ [6]. Consequently, the unpublished Belle result will reduce the tension with the theoretical prediction of 0.63×10^{-6}.

As of today, only a limit of $\lambda_B > 120$ MeV exists at 90% confidence level. This is calculated from the most stringent limit of the branching fraction of $\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) < 14 \times 10^{-6}$ reported by the BaBar collaboration [7]. Accordingly, λ_B is the dominant theoretical uncertainty for color-suppressed decay modes. In order to understand the importance of this parameter, the QCD factorization approach is described in the following.

Table 2.1: Predictions for branching fractions of B decays into light and unflavored states computed with QCD factorization (QCDF) with a comparison to experimental results. The theoretical predictions are given for two different values of λ_B where the first is obtained from non-perturbative calculations and the second is chosen to optimize the agreement between experiment and theory. The branching fractions are given in multiples of 10^{-6} [3], [4].

Mode	QCDF ($\lambda_B = 400\,\mathrm{MeV}$)	QCDF ($\lambda_B = 200\,\mathrm{MeV}$)	Experiment
$\pi^-\pi^0$	$6.2^{+2.4}_{-2.0}$	5.46	5.5 ± 0.4
$\rho_L^-\rho_L^0$	$21.0^{+8.5}_{-7.3}$	21.3	24.0 ± 2.0
$\pi^-\rho^0$	$9.3^{+4.0}_{-3.2}$	10.4	8.3 ± 1.2
$\pi^0\rho^-$	$15.1^{+5.7}_{-5.0}$	11.9	$10.9^{+1.4}_{-1.5}$
$\pi^+\pi^-$	$9.0^{+3.8}_{-3.3}$	5.21	5.1 ± 0.2
$\pi^0\pi^0$	$0.35^{+0.37}_{-0.21}$	0.63	1.9 ± 0.2
$\pi^+\rho^-$	$22.8^{+9.1}_{-8.0}$	13.2	15.7 ± 1.8
$\pi^-\rho^+$	$11.5^{+5.1}_{-4.3}$	8.4	7.3 ± 1.2
$\pi^\pm\rho^\mp$	$34.3^{+11.5}_{-10.0}$	21.6	23.0 ± 2.3
$\pi^0\rho^0$	$0.52^{+0.76}_{-0.42}$	1.64	2.0 ± 0.5
$\rho_L^+\rho_L^-$	$30.3^{+12.9}_{-11.2}$	22.3	24.2 ± 3.1
$\rho_L^0\rho_L^0$	$0.44^{+0.66}_{-0.37}$	1.33	0.73 ± 0.28

2.1 QCD factorization for exclusive non-leptonic B decays

The results presented in this section are taken from Ref. [1] which elaborates on all issues in great detail.

The low-energetic B meson system contains strong long-distance effects which cannot be calculated perturbatively. For B decays into two mesons, the hadronic matrix element $\langle M_1 M_2 | H_{\mathrm{weak}} | \bar{B} \rangle$ of the weak transition $\bar{B} \to M_1 M_2$ has to be determined, where the M_i are mesons lighter than the B meson. Approaches to the problem try to find valid approximations which allow to compute as much of this matrix element as possible in a perturbative manner.

The simplest approximation can be made by assuming a strong ordering of the energy scales given by $M_W \gg m_b$ where M_W is the mass of the weak W boson and m_b is the mass of the bottom quark. The weak propagator contracts to a point-like
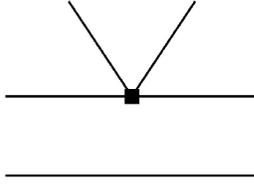
Figure 2.1: Feynman diagram showing the naive factorization approach of the B meson decaying into two mesons. Any exchange of gluons between the high-momentum meson M_1 and the (BM_2) system is assumed to be negligible. The Feynman diagrams in this section show in the lower line the spectator quark of the B meson and in the upper line the bottom quark. The non-horizontal lines show the quarks of the meson which is emitted from the weak decay [1].

interaction with a four-fermion vertex because the bottom quark, which sets the maximum energy scale of the decay, is significantly lighter than the weak gauge boson. This approximation is always justified in B meson decays. The strength of the transition is determined by the decay constant of the meson and the Fermi constant G_F which contains the weak coupling and propagator term in the low-energy limit.

An additional energy scale ordering can be identified with $m_b \gg \Lambda_{\mathrm{QCD}}$ where Λ_{QCD} is the scale parameter of the QCD which has a value of about $220\,\mathrm{MeV}$. For energies far above this scale, the QCD becomes perturbative where values of about $1\,\mathrm{GeV}$ are considered safe. This approximation states that, if the energy associated with an interline is of the order m_b, the strong interaction becomes perturbative and the binding effects of the hadron can be neglected. In contrast to the previous assumption, this approximation does not apply to all possible internal lines since dynamics which are lower in energy than the bottom quark mass can also enter the decay amplitude.

To able to factorize the two meson decay, one of the mesons must have a high momentum. This is only given for light mesons which contain the three lightest quark flavors $u, d,$ and s. Combining this with the considerations about the strong ordering $M_W \gg m_b \gg \Lambda_{\mathrm{QCD}}$, one arrives in lowest order at the so called naive factorization, diagrammed in Fig. 2.1, where the transition amplitude simplifies to

$$\langle M_1 M_2 | H_{\mathrm{weak}} | \bar{B} \rangle \rightarrow \langle M_2 | H_{\mathrm{weak}} | 0 \rangle \, \langle M_1 | H_{\mathrm{weak}} | \bar{B} \rangle \, .$$

Here, H_{weak} is the effective Hamiltonian of the weak interaction and M_2 is the light meson which interacts with the B meson decay only through the weak decay vertex. This assumes that no cross talk involving gluons takes place between M_1 and M_2,
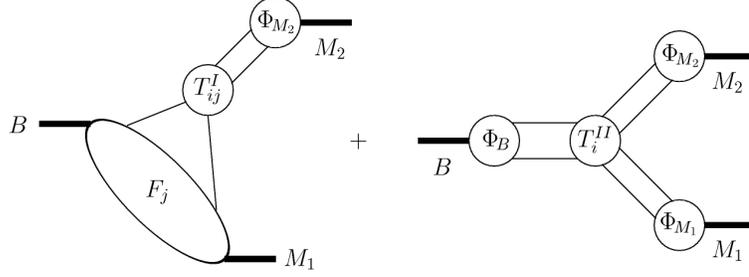
Figure 2.2: Graphical representation of the factorization formula given in Eq. 2 with its two factorization topologies [1].

as well as between M_2 and $\bar{B}$. The gluons are assumed to be high-energetic with a virtuality of the order m_b and thus suppressed. To put it in other terms, the M_2 meson is produced as a compact object with a high momentum relative to the remaining particles which suppresses further interactions. In consequence, the decay factorizes into a product of the decay constant of M_2 and the form factor $F^{B\to M_1}(m_2^2)$, where m_2 is the mass of the light meson. This approach is found to be too simplistic for most decays and many theoretical arguments support that observation.

In an extensive study of all one-gluon exchange corrections to the naive factorization, it is found that only hard gluon vertices contribute in the heavy quark limit where soft contributions are suppressed by powers of $\Lambda_{\mathrm{QCD}}/m_b$. The more sophisticated factorization formula is given by

$$\langle M_1 M_2 | H_{\mathrm{weak}}^i | \bar{B} \rangle = \sum_j F_j^{B\to M_1}(m_2^2) \int_0^1 du\, T_{ij}^I \Phi_{M_2}(u) + (M_1 \leftrightarrow M_2)$$

$$+ \int_0^1 d\xi\, du\, dv\, T_i^{II}(\xi, u, v) \Phi_B(\xi) \Phi_{M_1}(u) \Phi_{M_2}(v)$$

if M_1 and M_2 are both light, (2)

$$\langle M_1 M_2 | H_{\mathrm{weak}}^i | \bar{B} \rangle = \sum_j F_j^{B\to M_1}(m_2^2) \int_0^1 du\, T_{ij}^I \Phi_{M_2}(u)$$

if M_1 is heavy and M_2 is light.

Here, $F_j^{B\to M_{1,2}}(m_{2,1}^2)$ denotes the $B \to M_{1,2}$ form factor with $m_{1,2}$ being the light meson masses, Φ_X is the light-cone distribution amplitude (LCDA) of the quarks within meson X, and T_{ij}^I and $T_i^{II}(\xi, u, v)$ are hard scattering kernels where ξ, u, and v are the fractions of the meson momenta which are carried by the light quarks. These
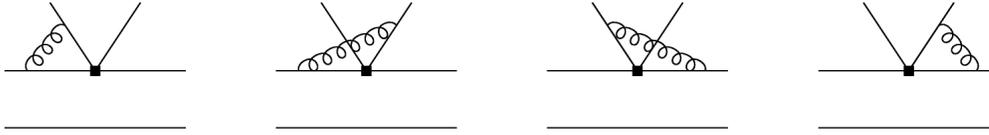
Figure 2.3: Leading order Feynman diagrams where the gluon attaches the bottom quark and the quarks of the emitted meson [1].

kernels contain the perturbatively calculable contributions of the decay which comprise the weak transition and high-energetic strong contributions. The strong contributions within the kernels can in principle be calculated to arbitrary order in α_s. For several decay modes calculations to second order in the strong coupling are available [2]. The remaining terms in Eq. 2 are non-perturbative which means they contain the soft effects. All terms are given in the light-cone distribution (the distribution which depends only on the longitudinal momentum) since the transverse momenta of the quarks inside the mesons are negligible when the relative momentum between the systems is large enough.

The formula is graphically represented in Fig. 2.2 for the two terms occurring in the equation. It is important to note that meson systems, which are described by form factors and LCDAs, interact only through the hard scattering kernels with each other and no soft long-distance contributions are present.

The left diagram represents the part of the equation containing the form factors. Here, the energetic meson emitted from the weak decay interacts with the bottom quark through a hard gluon exchange. This leading order process is shown in the Feynman diagrams in Fig. 2.3. It can be shown that the soft gluon contributions between the meson and the bottom quark cancel out if all of them are taken into account. This is connected to the color-transparency argument by Bjorken [8] which states that soft gluons only interact with the color-dipole moment (which is a higher order interaction) of the emitted color-singlet meson, hence they are suppressed. The hard gluons transfer a large momentum between the bottom quark and the light meson which makes it sensitive to the light meson LCDA. This results in a complicated convolution of the scattering amplitude T_{ij}^I with the LCDA of the light meson Φ_X (see Eq. 2.)

In the case that the spectator quark of the B meson can go to either final state meson, as for example in $B^+ \to \pi^0 K^+$, the first diagram in Fig. 2.2 has to be taken into account twice. Here, the two mesons M_1 and M_2 can be exchanged where each uses a different form factor $F_j^{B \to M_{1,2}}(m_{2,1}^2)$ together with the according LCDA of the emitted meson. In the decay $\bar{B}^0 \to \pi^+ K^-$ the spectator has to end up in the π^+ and
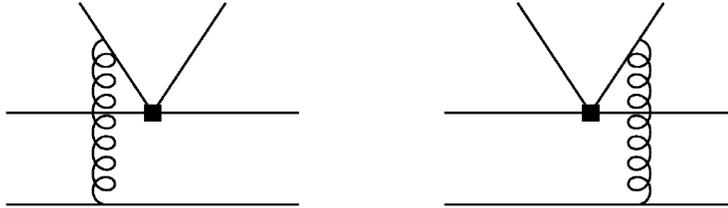
11

Figure 2.4: Higher-order Feynman diagrams showing the spectator scattering which is a correction of $\mathcal{O}(\alpha_s)$ to the leading order term. Here, the constituents of the emitted meson exchange a gluon with the spectator quark [1].

the switched diagram is absent.

The second diagram in Fig. 2.2, which is connected to the scattering kernel $T_i^{II}(\xi, u, v)$, represents a hard spectator interaction. Here, the light spectator quark of the B meson interacts with the quarks inside the energetic meson as shown in Fig. 2.4. The gluons connect quarks with large relative momentum and therefore the interaction between them is hard. This topology is similar to the leading order diagrams. Accordingly, the same color-transparency argument made above applies also here in order to explain that soft gluon contributions are suppressed. Since the hard gluon attaches to the light quarks in all mesons, the contribution depends on the LCDAs of all mesons including that of the B meson. The spectator scattering is a correction of order α_s to the leading term, that means that the whole scattering kernel $T_i^{II}(\xi, u, v)$ is a correction to T_{ij}^I.

The hard spectator interaction is power-suppressed in the heavy quark limit if one final state meson is heavy (containing a charm quark). This is covered in the second case of Eq. 2.

Higher order decay topologies contribute in the form of the penguin and chromomagnetic dipole diagrams which are shown in Fig. 2.5. These transitions enter both scattering kernels T_{ij}^I and $T_i^{II}(\xi, u, v)$ as corrections which are one order in α_s above the leading term. In the first case, a quark loop is attached to the bottom quark from which a gluon is emitted where in the second case the gluon is directly attached to the contracted weak decay vertex. Also here it can be shown that all internal lines, the quark loop as well as the gluon, are high-energetic. The gluon has to be an energetic particle since it decays in two high-energetic quarks. The quark pair has to be energetic since one quark has to form a high-momentum meson with the low-energetic spectator quark of the B meson. Furthermore, the two quarks are produced in an energy-symmetric state, therefore the second quark also has to be
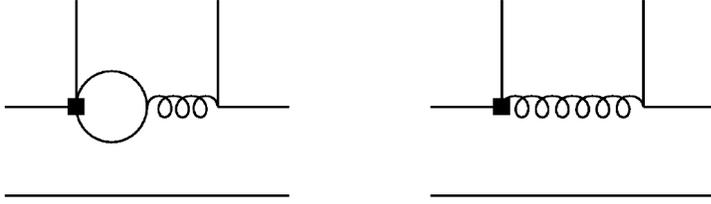
Figure 2.5: Higher-order Feynman diagrams which are a correction of $\mathcal{O}(\alpha_s)$ to the leading order processes. Here, a gluon is emitted from the weak decay vertex of the bottom quark, either directly or with an intermediate penguin process. The gluon hadronizes into two quarks which form two separate mesons with the spectator quark of the B meson and the decay product of the bottom quark [1].

energetic.

All remaining decay topologies are power suppressed by a factor $\Lambda_{\mathrm{QCD}}/m_b$ and need not be taken into account in this approximation. In lowest order of α_s the QCD factorization formula reduces to the naive factorization approach shown in Fig. 2.1.

The QCD factorization does not hold if one final state meson is heavy and the spectator quark of the B meson goes to the light meson, as for example in $\bar{B}^0 \to \pi^0 D^0$. In that constellation the relative momentum between the final state mesons is small and the emitted meson cannot be factorized from the $\bar{B}^0 \to \pi^0$ transition since the soft gluon contributions from Fig. 2.3 and Fig. 2.5 do not cancel.

To conclude, the factorization approach identifies additional contributions of hadronic decays which are perturbatively calculable. The form of the matrix element simplifies since specific hadronic quantities are identified providing the possibility to measure them or perform a calculation in lattice and QCD-sum-rules. As described above, the B meson LCDA is needed to compute the spectator scattering term which is only present in charmless transitions. The parameter dominates the theoretical uncertainty of this contribution. It can be shown that up to order α_s this LCDA only appears in the integral form given in Eq. 1.

A fundamental result of the factorization formula is that the leading order terms are real. Hence, complex strong phases which introduce direct CP asymmetries are predicted to be at most of order $\mathcal{O}(\alpha_s)$ or $\mathcal{O}(\Lambda_{\mathrm{QCD}}/m_b)$.

13

2.2 The $B^+ \to \ell^+\nu_\ell\gamma$ decay

The following results are taken from Ref. [9] in which the $B^+ \to \ell^+\nu_\ell\gamma$ decay is calculated up to next-to-leading order.

The $B^+ \to \ell^+\nu_\ell\gamma$ decay consists of a weak transition accompanied by a photon emitted by one of the charged particles. The constituent quarks of the B meson annihilate into a virtual W boson which in turn decays into a lepton-neutrino pair. Here, only electrons and muons are considered for the charged lepton.

Helicity suppression

The purely weak $B^+ \to \ell^+\nu_\ell$ decay is strongly helicity-suppressed since the parity violating weak current couples only to left-handed particles and right-handed anti-particles. Due to the large mass difference between the leptons and the bottom quark, both the anti-neutrino and the charged lepton are almost exclusively right-handed. This results in a very small coupling to the weak current which is an effect called helicity suppression. The $B^+ \to \ell^+\nu_\ell$ decay has a branching fraction of

$$\mathcal{B}(B^+ \to \ell^+\nu_\ell) = \frac{G_F^2 m_B m_\ell^2}{8\pi}(1 - \frac{m_\ell^2}{m_B^2})^2 f_B^2 |V_{ub}|^2 \tau_B,$$

which is proportional to the squared mass of the charged lepton m_ℓ^2. The less massive the lepton the larger its velocity which in turn determines the size of its left-handed spinor component which couples in the weak transition. Therefore the branching fraction decreases with smaller lepton masses despite the fact that lighter final states have a larger kinematic phase space. Furthermore, the transition amplitude depends on the Fermi constant G_F, the matrix element V_{ub} which describes the weak mixing of up and bottom quark states, the B meson mass m_B, the lifetime of the B meson τ_B, and the decay constant for the B meson f_B.

The full decay

Adding a photon to the final state of the purely weak decay removes the helicity suppression. Even though the electromagnetic interaction introduces an additional factor of α_{em} into the equation, the overall branching fraction is expected to be larger than in the photon-less case.

In leading order, the photon is emitted from the up quark. The emission changes the B meson into an excited virtual B^{*+}/B_1^+ state with spin one. The lepton pair can now couple to the spin one state which results in the correct handedness for
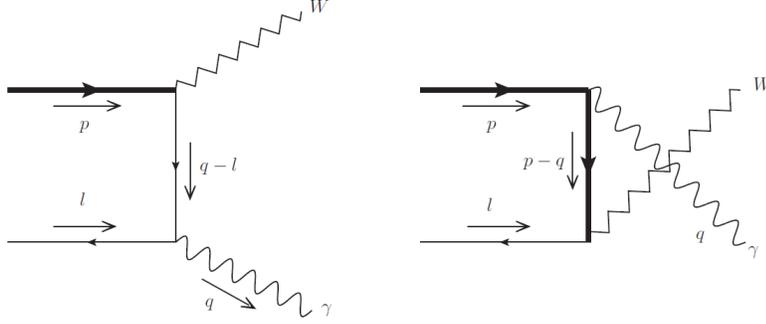
Figure 2.6: Feynman diagrams for the tree decay contributions. The left diagram shows the leading order term where the photon is emitted by the up quark. Here, the propagator joining the W and photon vertex is light. The bottom quark emission in the right diagram is suppressed since the propagator is now heavy [9].

both leptons in the weak decay, with a left-handed particle and right-handed anti-particle. From this follows, that the photon emission from the charged lepton is still suppressed since the B meson does not change into a spin one state before annihilating. Additionally, also the bottom quark photon emission is of sub-leading order, as explained below.

The weak transition of the decay, which is low-energetic and therefore point-like, is precisely calculable. The photon emission is incorporated through a factorization where $E_\gamma \gg \Lambda_{\mathrm{QCD}}$ has to be satisfied. The reason is identical to the QCD factorization in the previous section. The strong interactions inside the hadron become perturbative if the momentum transfer of the photon onto a quark is larger than the QCD scale. Values for E_γ above $1\,\mathrm{GeV}$ are considered safe in the calculation. Massless leptons are assumed for the calculation which applies to electrons and muons but not to tauons. Due to this assumption, the results for the two decay modes are identical.

The decay is calculated in heavy-quark expansion where the decay width is determined in orders of $1/m_b$. The double differential decay rate is given by

$$\frac{d^2\Gamma}{dE_\gamma dE_\ell} = \frac{\alpha_{em} G_F^2 |V_{ub}|^2}{16\pi^2} m_B^3 (1-x_\gamma)\left[(1-x_\nu)^2 (F_A + F_V)^2 + (1-x_\ell)^2 (F_A - F_V)^2\right], \quad (3)$$

where $x_{\gamma,e,\nu} = 2E_{\gamma,e,\nu}/m_B$ and $0 \leq x_i \leq 1$, F_A is the axial form factor and F_V is the vector form factor.

As mentioned above, the decay proceeds through virtual excited B^+ states. The vector form factor describes the coupling to B^{*+} with a quantum number of 1^- where the spins of the two constituent quarks are in parallel. The axial form factor describes

15

the transition through a B_1^+ state with a quantum number of 1^+. This state has an orbital angular momentum of one and spins which are anti-parallel.

In the following, the sub-leading photon emission of the lepton is added to the axial form factor since it has the same tensor structure as the hadronic axial coupling.

Since the form factors are independent of the lepton energy, an integration over E_ℓ can be performed which yields

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{em} G_F^2 |V_{ub}|^2}{48\pi^2} m_B^4 (1 - x_\gamma) x_\gamma^3 \left[F_A^2 + F_V^2 \right]. \tag{4}$$

The form factors are given by

$$F_V(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + \left[\xi(E_\gamma) + \frac{Q_u m_B f_B}{(2E_\gamma)^2} + \frac{Q_b m_B f_B}{2E_\gamma m_b} \right],$$

$$F_A(E_\gamma) = \frac{Q_u m_B f_B}{2E_\gamma \lambda_B(\mu)} R(E_\gamma, \mu) + \left[\xi(E_\gamma) - \frac{Q_u m_B f_B}{(2E_\gamma)^2} - \frac{Q_b m_B f_B}{2E_\gamma m_b} + \frac{Q_\ell f_B}{E_\gamma} \right]. \tag{5}$$

Here, $Q_{\ell,u,b}$ are the electric charges of the lepton, up quark, and bottom quark, respectively, and $R(E_\gamma, \mu)$ is the radiative correction calculated at the energy scale μ. The first term in the form factors, containing λ_B, represents the leading order contribution describing the photon emission by the light quark. The leading order term is corrected for higher-order radiative effects, with the $R(E_\gamma, \mu)$ factor containing mass corrections for the up quark. The remaining terms in square brackets are $1/m_b$ corrections which are: higher-order contributions for the hard and soft photon emission of the up quark (Q_u and the $\xi(E_\gamma)$-term, respectively); the photon emission by the bottom quark, which is suppressed due to its higher mass (Q_b-term); and the photon emission by the lepton, which is only present in the axial form factor (Q_ℓ-term). The radiative corrections contained in $R(E_\gamma, \mu)$ reduce the leading order amplitude by about $20 - 25\%$. The remaining $1/m_b$ corrections have considerable parametric uncertainties. However, using central values for the parameters, the terms suppressed $1/m_b$ reduce the decay amplitude by about half the amount of the radiative corrections.

In leading order, the photon is emitted from the up quark which makes the decay sensitive to the first moment of the LCDA, λ_B, which is the parameter needed in the QCD factorization described in the previous section. Since no other hadrons contribute in the decay, the LCDA is the only non-perturbative hadronic quantity at leading order. In other words, the photon probes the LCDA of the B meson which makes the decay ideal to determine λ_B.

The photon emission of the up and bottom quark is shown in Fig. 2.6. The suppression of the bottom quark emission can be explained by the propagator connecting

the weak decay vertex and the photon line. Here, the propagator has hard virtuality being of the order $1/m_b$.

On the other hand, the propagator for the up quark emission is light; thus it has terms which are not suppressed by $1/m_b$. In addition, the light propagator also contains sub-leading $1/m_b$ terms for the up quark photon emission. Two of these next-to-leading order terms are soft, non-local and therefore difficult to include in the calculation. These are summarized in the $\xi(E_\gamma)$-term in Eq. 5. It is found that the form factor in $B \to \pi$ transitions is similar to this contribution. This is used in Ref. [10] to reduce the large theoretical uncertainty introduced by this term. Nevertheless, the theoretical error on the parameter λ_B is dominated by the soft effects in the up quark emission.

Old model as cross check
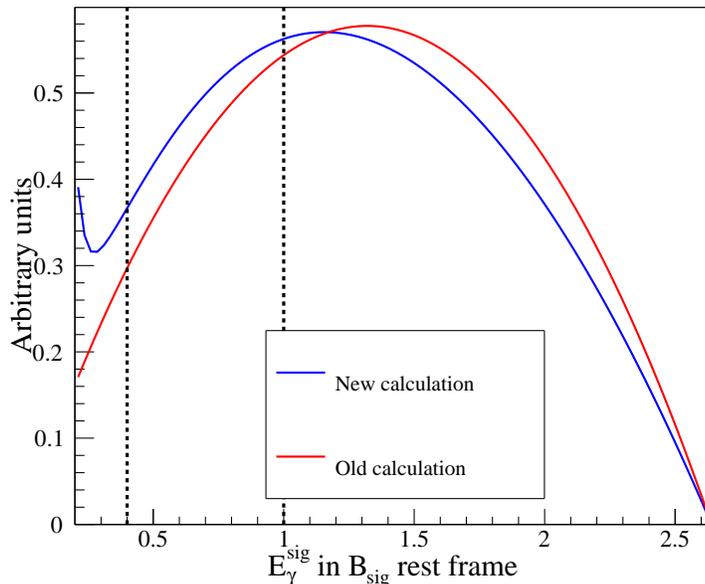
Figure 2.7: Photon energy spectrum for the latest next-to-leading order calculation from Ref. [9] and an leading order calculation from Ref. [11] without detector simulation. The difference between the curves is used as a measure of the systematic model uncertainty. The vertical lines show two energy thresholds at $400\,\mathrm{MeV}$ and $1\,\mathrm{GeV}$ which are used for the nominal and secondary analysis, respectively (see section 3.2.1).

17

An older calculation from Ref. [11] is used to estimate the theoretical error for the signal model. This calculation misses next-to-leading order terms of the result above and the form factors read

$$F_A = F_V = \frac{1}{E_\gamma} \left(Q_u \frac{1}{\lambda_B} - \frac{Q_b}{m_b} \right).$$

The expression omits constant factors which have no influence on the shape of the spectrum since the normalization is not of interest here.

The shapes of the two theoretical predictions in Fig. 2.7 exhibit only minor differences. The additional next-to-leading order terms introduce a divergence at low energies at which the model is expected to break down.

As mentioned above, the new model predicts an absolute decrease of the branching fraction by about 30% which is mainly due to radiative corrections of the leading order up quark contribution.

3 Tools

In this section several analysis techniques are introduced that are used in various places of the analysis.

3.1 Particle ID

The $B^+ \to \ell^+ \nu_\ell \gamma$ signal consists of two detectable particles. An important part of the selection rests upon the correct identification (ID) of the charged lepton. A significant contribution to this selection is given by the electron and muon ID variables of the Belle experiment which combine information of several detector components.

The Belle detector is a large-solid-angle magnetic spectrometer that consists of a silicon vertex detector, a 50-layer central drift chamber (CDC), an array of aerogel threshold Cherenkov counters (ACC), a barrel-like arrangement of time-of-flight scintillation counters (TOF), and an electromagnetic calorimeter (ECL) comprising CsI crystals located inside a superconducting solenoid coil that provides a 1.5 T magnetic field. An iron flux return located outside the coil is instrumented to detect K_L^0 mesons and to identify muons (KLM). A detailed description of the Belle detector can be found in [12],[13].

The Belle experiment was located at the KEKB asymmetric-energy e^+e^- collider. The collider operates at the $\Upsilon(4S)$ resonance with a center-of-mass energy of $10.58\,\mathrm{GeV}/c^2$, where the resonance decays almost exclusively to $B\bar{B}$ pairs. A detailed description of the KEKB accelerator is given in [14],[15]

Probability density functions (PDF) are determined for each particle type and sub-detector, and the gathered information is then combined to likelihood functions. The particle IDs are defined by likelihood ratios in which the likelihood of a particle type is divided by the likelihoods of its dominant backgrounds.

The electron likelihood function is formed from the following information [16]:

- The dE/dx energy loss in the CDC.

- The energy over momentum E/p which is obtained from the energy deposition in the ECL and the curvature of the track in the CDC, respectively.

- The transverse shower shape in the ECL defined by the ratio E_9/E_{25} where the energy deposition of a 3×3 and 5×5 cell array are divided with the highest-energetic cell located in the center.

- The χ^2 value obtained from matching the CDC track to the shower position in the ECL.

- The photon yield in the ACC which measures the velocity of the charged particle.

The electron likelihood is divided by the likelihood function of all other particle types, which are formed from the same information.
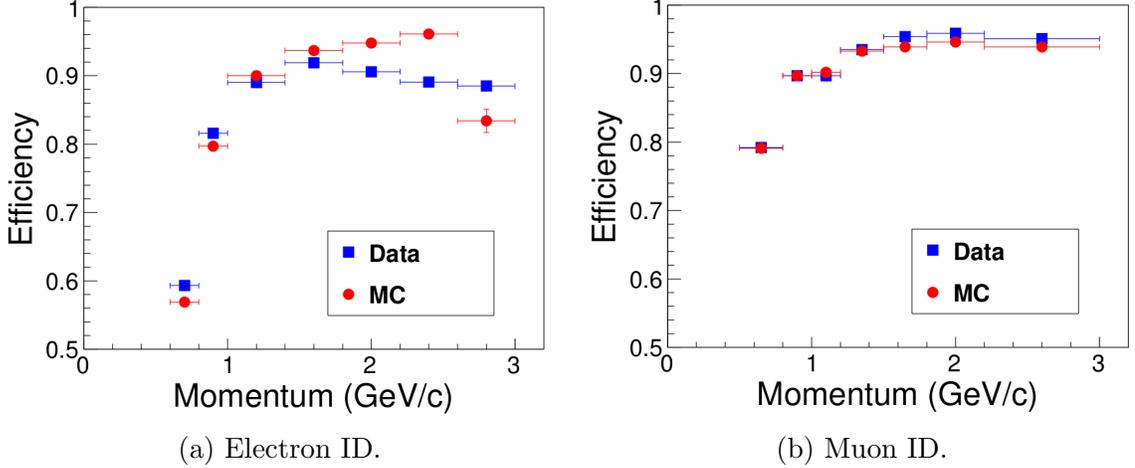

(a) Electron ID.

(b) Muon ID.

Figure 3.1: Efficiency of the Belle lepton ID determined in $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ processes for a hard selection on the variables [17].

Muons usually penetrate the whole detector with little energy deposition, and a different approach is taken to identify these particles [18]. Hits in the KLM, the outermost part of the detector, are used to identify muons. For that, CDC tracks are extrapolated to the KLM and the squared distances from the extrapolated track to the hits in the KLM are summed to compute a χ^2 value. Additionally, the difference between the expected and measured penetration depth of the track in the KLM is used in the likelihood. The main backgrounds for the muon are K_L^0 and π^+ mesons. For that reason, the muon ID is determined to be the likelihood ratio dividing the muon likelihood by the likelihoods of K_L^0, π^+ and muon hypotheses

$$\text{Muon ID} = \frac{\mathcal{L}_{\mu^+}}{\mathcal{L}_{\mu^+} + \mathcal{L}_{\pi^+} + \mathcal{L}_{K_L^0}}.$$

The π^+ and K_L^0 likelihoods are formed from the same detector information as the muon likelihood.

The efficiency of the electron and muon ID variables is shown for $e^+e^- \rightarrow e^+e^-\ell^+\ell^-$ processes in Fig. 3.1 where a hard selection is applied. The efficiency increases with

20

momentum for both variables, and it remains approximately constant above $1\,\text{GeV}/c$ where the muon ID has a higher efficiency. A small systematic difference between data and MC is observed which is discussed in section 4.4.2 in the context of this analysis.

The kaon ID is used in the full reconstruction in section 3.4 and also in the control channel study of $B^+ \to K^*(892)\gamma$ decays in section 7.3. This likelihood variable combines information from the TOF with the energy loss in the CDC and the ACC photon yield. The latter two pieces of information are also used for the electron ID. The main background for K_L^0 mesons are π^+ mesons and accordingly, a likelihood ratio for the kaon ID is defined as follows

$$\text{Kaon ID} = \frac{\mathcal{L}_{K_\text{L}^0}}{\mathcal{L}_{K_\text{L}^0} + \mathcal{L}_{\pi^+}} = 1 - \text{Charged pion ID}.$$

The ID for charged pions is simply one minus the kaon ID which assumes that these two particles mutually constitute their only background.

3.2 MC samples

The analysis is performed blindly, this means that it is optimized on Monte Carlo (MC) samples only. The decays are simulated with the EvtGen software package [19], and the simulation of the detector response is performed with GEANT3 [20]. Beam background is recorded by the experiment and added to the events in the MC.

3.2.1 Signal MC

The form factors given in Eq. 5 are used to define the $B^+ \to \ell^+\nu_\ell\gamma$ decay model in the EvtGen package. The parameters for the form factors, which are taken from Ref. [9], are listed in table 8.3. Using central values for all parameters, the photon energy spectrum in Fig. 2.7 is obtained in the rest frame of the B_sig meson. Two photon energy thresholds are marked in the plot for which separate analyses is performed. The nominal analysis uses only photons above $1\,\text{GeV}$ for which the theoretical model is valid. Additionally, a secondary analysis is performed which uses a looser selection with photon energies larger than $400\,\text{MeV}$. This lower threshold is chosen to veto regions where the theoretical model begins to diverge.

The signal MC is generated with 2×10^6 events for the electron and muon channel. The analysis is optimized for an assumed signal branching fraction of 5×10^{-6}, which corresponds to the number favored by theory [2] given the value of $\lambda_B \approx 200\,\text{MeV}$.

21

The signal MC is weighted by

$$\text{weight} = \frac{N_{B\bar{B}} \times \mathcal{B}(B^+ \to \ell^+ \nu_\ell \gamma)}{\text{Number of generated events}}$$

$$= \frac{771.6 \times 10^6 \times 5 \times 10^{-6}}{2 \times 10^6} = 0.001929,$$

where $N_{B\bar{B}}$ is the number of B meson pairs in the full Belle data sample.

3.2.2 Background MC

The background is estimated with samples of $b \to c$ MC, $b \to u\ell^+\nu_\ell$ MC, rare MC, and $q\bar{q}$ $(q = u, d, s, c)$ MC. The first three samples contain decays that proceed via the $\Upsilon(4S)$ resonance which results in a $B\bar{B}$ pair. The last sample contains all remaining non-resonant processes where lighter quark flavors are produced and the branching fraction increases with the quark mass. Together, these samples contain all possible decays occurring at the $\Upsilon(4S)$ energy, and each sample covers a specific group of decays. For the analysis, the $b \to c$ and $q\bar{q}$ $(q = u, d, s, c)$ processes are summarized into the so-called generic MC.

For the generic MC ten streams of resonant $b \to c$ and six streams of $q\bar{q}$ continuum events are available, where one stream corresponds to the number of events contained in the recorded Belle data set.

The $b \to u\ell^+\nu_\ell$ MC contains processes where the bottom quark of a B meson decays semi-leptonically to an up quark. The MC can be divided into three classes of decays: well understood exclusive decays where the resulting up quark forms a meson X_u of the type π^0, η, π^+, ω, ρ^+, η', or ρ^0; decays containing intermediate strong resonances: f_{0-2}, a_{0-2}^0, a_{0-2}^+, h_1, and b_1; and inclusive decays with an unspecified X_u^0 or X_u^+ meson which sums up all remaining decay amplitudes. This MC sample contains 20 streams of data.

Old values are used for some of the branching fraction in the $b \to u\ell^+\nu_\ell$ MC. Therefore the most important decays are weighted to the current world averages provided by Heavy Flavor Averaging Group [21]. The biggest difference is found for the $B^+ \to \eta\ell^+\nu_\ell$ decay mode which is weighted down to 56% of its original value.

The largest backgrounds for the analysis are $B^+ \to \pi^0\ell^+\nu_\ell$ and $B^+ \to \eta\ell^+\nu_\ell$ decays which are contained in the $b \to u\ell^+\nu_\ell$ sample. Due to the importance of this background, high-luminosity samples are generated for these two decays with 1×10^7 events for each of the four lepton-meson combinations. This corresponds to about 100 times the statistics found in the Belle data set.

The rare MC consists of resonant decays with $b \to s$ transitions. It also includes all rare $B\bar{B}$ processes not contained in the other MC samples, including the signal

decay $B^+ \to \ell^+ \nu_\ell \gamma$ with branching fractions of 5×10^{-6}. Fifty streams of the rare MC are used.

3.3   Continuum suppression

The majority of events, which occur when operating a collider at the energy of the $\Upsilon(4S)$ resonance, are $q\bar{q}$ continuum processes where lighter $u, d, s,$ or c quark pairs are produced. This is an unwanted background to the $B\bar{B}$ events. The B meson pair is produced at low momentum which results in a spherical event structure. The event topology for continuum events on the other hand, has a jet-like structure due to the lower mass of the initial quark pair. Several event shape variables are defined and combined in a neural network to distinguish the two event types based on their decay topology.

The first set of event shape variables are Fox-Wolfram-Moments (FWM) [22] which are given by

$$H_\ell \equiv \sum_{i,j} |\vec{p}_i||\vec{p}_j|P_\ell(\cos\theta_{ij}). \tag{6}$$

Here, the indices i and j run over all charged tracks in the event, $|\vec{p}_i|$ denotes the momentum of particle i, and $P_\ell(\cos\theta_{ij})$ denotes the Legendre polynomials of ℓ-th order. These depend on the cosine of the opening angle between particles i and j. For jet-like events, H_ℓ tends towards values of one if ℓ is even and to values of zero if ℓ is odd. The FWMs are normalized to the zeroth order moment to obtain the reduced FWMs as $R_\ell = H_\ell/H_0$. Only the reduced R_2 moment is used for the continuum suppression.

The Super-Fox-Wolfram-Moments (SFWM) [23] are an improved version of the FWMs developed specifically for $B\bar{B}$ events in which a B meson candidate is already identified. The reduced FWMs are separated into different categories where either all combined particles originate from the B meson candidate (R_ℓ^{ss}), all particles are taken from the rest of the event (R_ℓ^{oo}), and mixed cases where particles are taken form both the B meson candidate and the rest of the event (R_ℓ^{so}). In total 18 different SFWMs are used for the network separation.

Other variables used in the network are the angle of the reconstructed B meson candidate and the beam axis as well as the angle between thrust axes of the B meson candidate and the remaining particles in the event. The thrust axis is defined as the vector $\vec{n}$ maximizing the sum of the projections of the particle momenta on itself

$$T = \max_{|\vec{n}|=1} \frac{\sum_i |\vec{n} \cdot \vec{p}_i|}{\sum_i |\vec{p}_i|}. \tag{7}$$

CLEO cones [24] are variables which sum the absolute values of the momenta of the particles lying in a cone around the thrust axis of the B meson candidate. The cones are divided in intervals of $10°$ starting with $0° - 10°$, $10° - 20°$ and so on.

Additional variables which can be used to identify $B\bar{B}$ events are: the flavor tagging output from the Hamlet flavor tagger [25] and Δz which is the distance of the fitted decay vertices of two B mesons in beam direction.

A continuum suppression is used in two instances of the analysis. It is integrated into the network output of the full reconstruction in section 3.4, where only the SFWMs and the angle between the thrust axis of the B_{tag} and the beam axis are used. Another continuum suppression is used for the selection of the $B^+ \rightarrow K^*(892)\gamma$ control channel in section 7.3. Here, all variables described in this section are included, and the network is trained to separate signal MC from continuum $q\bar{q}$ events.

3.4   Hadronic tagging

The analysis is based on the calculation of the squared missing mass of the $B^+ \rightarrow \ell^+\nu_\ell\gamma$ signal decay. In order to be able to compute this variable with a high resolution, as described in section 4.1, the whole visible event has to be reconstructed. This means, in addition to the particles of the signal decay, the other tag-side B meson (B_{tag}) has to be fully reconstructed in a hadronic decay channel without any additional neutrinos in it. After that, the whole event is known, except for the signal neutrino whose mass can now be identified with the missing mass of the event.

The reconstruction of the B_{tag} has the additional advantage, that it allows for a very efficient signal-side selection since it cleans the event of all tag-side particles before the signal-side particles are selected.

The EKP full reconstruction [26] is used to obtain the B_{tag} in the hadronic decay channel. The B_{tag} meson decays in multiple steps into the final decay products, and the algorithm addresses this decay topology with a hierarchical reconstruction scheme. Decaying particles are reconstructed from daughter candidates where the daughters are used independently from their specific decay modes. This avoids the explicit reconstruction of the B_{tag} candidates whole decay chain into a certain final state. Each particle type is selected with a NeuroBayes neural network (NN) [27] which is trained separately for each of its immediate decay channels. Building the decay chain in this way leads to a large reduction of computing time since the number of classification problems reduces from 1104 to 71.

In order to reduce the computing time, the NN outputs are used to perform a highly efficient pre-selection. For all selections on the outputs it is required that the same number of additional background events are added for each additional signal

event.

The charged B_{tag} meson candidate is reconstructed in one of 17 final states

- $\bar{D}^{(*)}X_{\mathrm{h}ad}$ (seven states)
- $\bar{D}^{(*)0}D_s^{(*)+}$ (four states)
- $\bar{D}^0K^+$
- $D^-\pi^+\pi^+$

- $J/\psi K^+$
- $J/\psi K^+\pi^0$
- $J/\psi K^+\pi^+\pi^-$
- $J/\psi K_S^0\pi^+$

with $X_{\mathrm{h}ad}$ being a set of one to four pions, of which one can be neutral.

Additional selections are applied to intermediate particles to further reduce the computation time. The D meson momenta are selected to be $p^*(D) < 2.6\mathrm{GeV}/c$ in the $\Upsilon(4S)$ rest frame, as high momenta are a sign for non-resonant $q\bar{q}$ processes with a jet-like decay topology. Mass windows for all D mesons are selected to be $40\,\mathrm{MeV}$ and $60\,\mathrm{MeV}$ around their nominal mass, where the tighter selection is used for modes with higher background levels. For excited D^* states, instead of the reconstructed mass, the mass differences between excited and ground state are used in the selection. These differences are selected to be $\Delta\mathrm{M}(D^{*+} \in (135, 150)\,\mathrm{MeV}/c^2$, $\Delta\mathrm{M}(D^{*0} \in (132, 152)\,\mathrm{MeV}/c^2$ and $\Delta\mathrm{M}(D_s^{*,+} \in (110, 175)\,\mathrm{MeV}/c^2$. The J/ψ particles are reconstructed from e^+e^- or $\mu^+\mu^-$ decays. Two charged tracks are used to reconstruct a K_S^0 candidate whose mass must be within $30\,\mathrm{MeV}$ of the nominal mass. Neutral pions are reconstructed from pairs of photons, each with an energy of at least $30\,\mathrm{MeV}$ and an invariant mass falling within a $19\,\mathrm{MeV}$ window of the nominal pion mass. Photons are identified as energy depositions in the calorimeter above $20\,\mathrm{MeV}$ without an associated track. Charged tracks are identified as pions or kaons using the Belle kaon ID. Charged-track quality is improved by requiring that $|dz| < 4.0\,\mathrm{cm}$ and $dr < 2.0\,\mathrm{cm}$, where $|dz|$ and dr are the distances of closest approach of the track to the interaction point along the beam axis and in the transverse plane, respectively.

The NN output of the final B_{tag} meson is used for the signal selection in the analysis. Important input variables include: the network outputs of the daughter particles; the reconstructed masses of the daughters; $\Delta E = E_{B_{\mathrm{tag}}} - E_{\mathrm{beam}}$, which is the difference between the B_{tag} candidate energy and the beam energy in the center-of-mass system (CMS); the mass difference $\Delta\mathrm{M}(D^*)$ for D^* meson decay channels; the angles between the daughters in the B_{tag} meson rest frame; the momentum of the daughters in the lab frame; and $\cos\Theta_B$, the cosine of the angle between the beam and the B_{tag} direction. The network output can be interpreted as the probability that the B_{tag} candidate is correctly reconstructed, which means all particle hypotheses of the decay chain are correct.
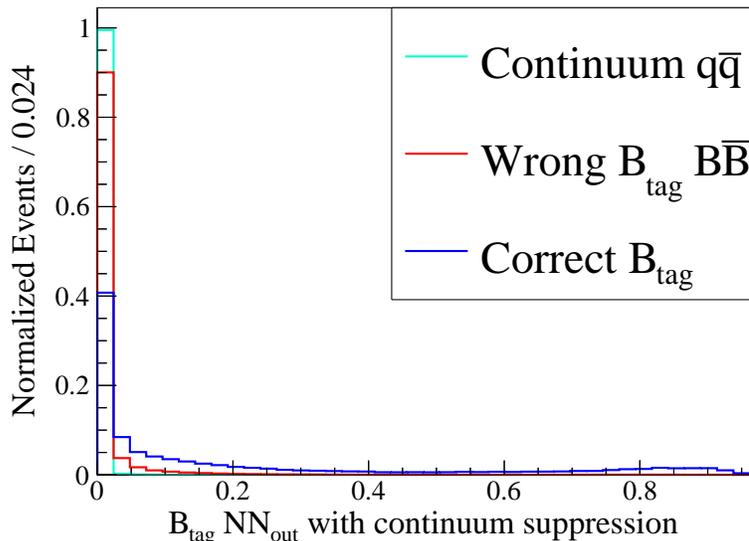
Figure 3.2: NN output of the full reconstruction for the B_{tag} meson with continuum suppression for Generic MC. The network output represents the Bayesian probability for the B meson to be correctly reconstructed. Correctly and wrongly reconstructed B_{tag} candidates are identified with the MC matching described in section 3.5.

Additional event shape variables are used in the B_{tag} network to suppress continuum background. Variables used are Fox-Wolfram moments and the thrust axis of the B_{tag} meson candidate relative to the beam axis, further described in section 3.3. In the case of multiple B_{tag} candidates, the candidate with the highest network output is selected. The network output is shown in Fig. 3.2 with its the separation between correctly and wrongly reconstructed $B\bar{B}$ decays. The separation improves for $q\bar{q}$ continuum background.

The two variables ΔE and $M_{\text{bc}} = \sqrt{E_{\text{beam}}^2 - \vec{p}_{B_{\text{tag}}}^2}$ can be defined from the reconstructed momentum of the B_{tag}. The M_{bc} mass is calculated with the beam energy which delivers a better estimate of the B_{tag} energy, as shown in Fig. 4.2a. Loose selections are made on these variables in the full reconstruction which require $\Delta E \in (-150, 100)\text{MeV}$ and $M_{\text{bc}} \in (5.22, 5.30)\text{GeV}/c^2$. The M_{bc} mass is not correlated to any other variable. The distribution shown in Fig. 3.3 has a signal region above $5.27\,\text{GeV}/c^2$ and a sideband below that value which consists of combinatoric $B\bar{B}$ and $q\bar{q}$ continuum background. The sideband is used for data-MC consistency checks.
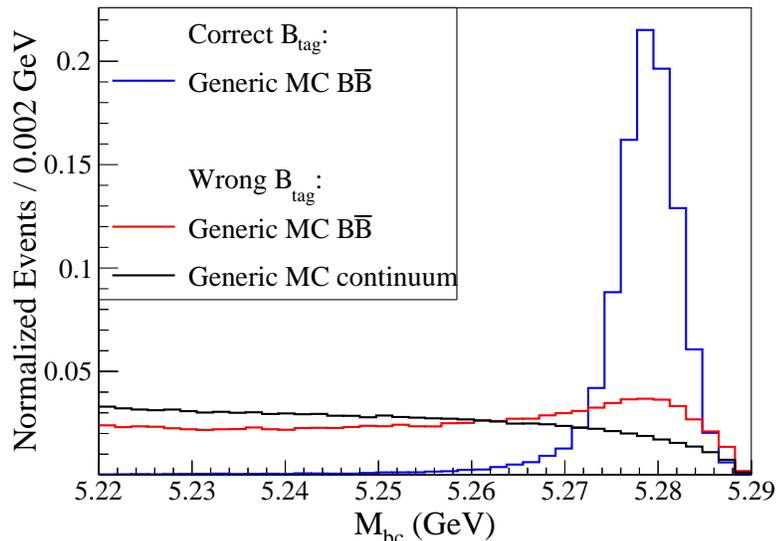
Figure 3.3: Beam-constrained B_{tag} mass for correctly and wrongly reconstructed $B\bar{B}$ events for the generic MC. Correctly and wrongly reconstructed B_{tag} candidates are identified with the MC matching described in section 3.5.

The decay modes in the hadronic full reconstruction cover about 12% of the B^+ branching fraction. Nevertheless, the efficiency is low and has a dependence on the signal-side decay channel. For the simple $B^+ \to \ell^+ \nu_\ell \gamma$ process, a relatively high tagging efficiency of 0.6% is found for signal MC, and the efficiency drops to about 0.2% for generic MC. Observed efficiency differences between data and MC are described in section 4.4.1.

3.5   MC matching

Two MC matching variables for the signal and tag-side are defined which identify correctly reconstructed events in the MC. The MC flag assumes positive values if the reconstruction is deemed correct and negative values otherwise, where different errors in the reconstruction are codified by different values.

Signal events are assumed to be well reconstructed if they produce a peak in the squared missing mass variable, described in section 4.1. The signal-side reconstruction is assumed to be wrong if any of the signal particle candidates has a wrong particle hypothesis. Furthermore, it is checked that no identical particles are among the signal children and that all children have the same mother particle. Photons from

27
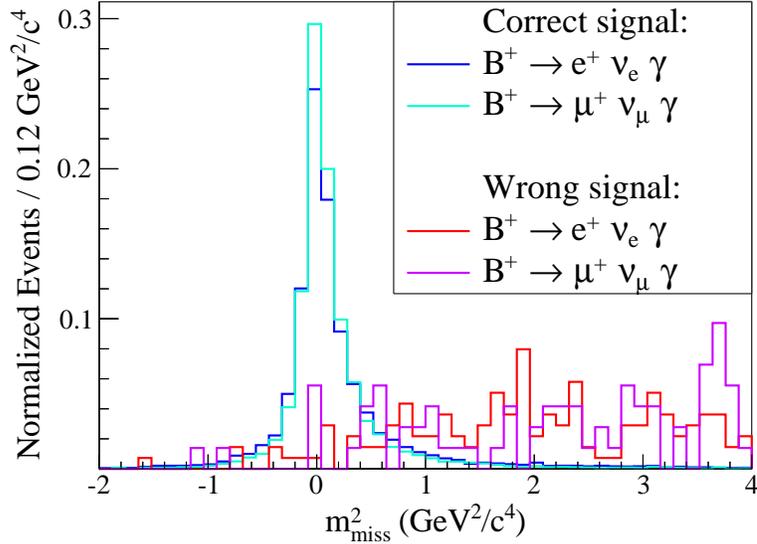
Figure 3.4: Comparison of m^2_{miss} for correctly and wrongly matched signal MC for both signal channels. The distributions peak only for correctly matched MC.

final state radiation are missed in the reconstruction of the signal-side, nevertheless these events produce a peak in the missing mass. Consequently, missing photons are allowed in correctly flagged signal decays. This effect is commonly found in the $B^+ \rightarrow e^+\nu_e\gamma$ signal channel where one or two bremsstrahlung photons are radiated from the electron. If a particle other than a photon is missing in the reconstruction, the reconstruction is assumed to be incorrect. The effectiveness of the signal MC matching is shown in the missing mass distribution in Fig. 3.4.

The definition of a correctly reconstructed B_{tag} candidate in the MC is taken from Ref. [28]. This tag-side MC matching algorithm separates peaking from non-peaking events in the M_{bc} mass of the B_{tag}, which is defined in section 3.4. It checks if all particles candidates used in the B_{tag} reconstruction have the same B meson as their common mother. This identifies most of the badly reconstructed B_{tag} candidates in the MC. It includes cross-feed in $B\bar{B}$ events, continuum events with $e^+e^- \rightarrow q\bar{q} \rightarrow (u,d,s,c)$, and events which reconstruct other background like beam-background photons. Events are considered sufficiently well reconstructed if final state particles are missing or if particles have a wrong hypothesis but still originate from the B_{tag}. The performance of the tag-side flag is demonstrated in Fig. 3.3 which shows the M_{bc} mass distribution for generic MC events. The matching works reasonably well. Correctly flagged $B\bar{B}$ events peak at the B meson mass of

28

about $5.28\,\text{GeV}/c^2$ and other events flagged as background exhibit a flat shape in the distribution.

3.6   Correlation analysis

The statistical correlation of two variables is checked with an algorithm described in [29]. Here, the variables are binned so that each bin contains the same amount of events; this is also called flattening. After flattening each variable separately, their two-dimensional distribution is expected to be flat as well if they are uncorrelated. The uniformity of the two-dimensional distribution is checked in a χ^2 test which sums the squared differences between the bin values and the mean. The mean is the total number of events divided by the bin number. In order to obtain a χ^2 distribution for uncorrelated variables, the event numbers in the bins should follow a Gaussian distribution. Therefore, the algorithm chooses the bin number so that at least 25 events are contained in each bin. With the χ^2 statistic, the significance of the correlation between the variables can be determined.

The algorithm is used to determine the correlation between the fit variables of the missing mass and the network output defined in section 4.3. The missing mass is fitted in bins of the network output and the correlation determines whether the shapes of the MC will differ among the bins. A strong correlation of $> 8\sigma$ is found for the signal, $b \to u\ell^+\nu_\ell$, $B^+ \to \pi^0\ell^+\nu_\ell$, and $B^+ \to \eta\ell^+\nu_\ell$ MC samples. The generic MC shows a smaller correlation of 2.2σ in the muon channel, and no correlation is found in the electron channel. According to these findings, the shapes of the MC have to be fitted in each network output bin separately.

4 Signal-side analysis

The signal-side selection is performed after the tag-side has been fully reconstructed. This means that only the signal-side particles are expected in the detector which allows for an efficient signal selection which is described in section 4.2. After that, the remaining background is suppressed with a NN which is described in section 4.3. The squared missing mass, introduced in section 4.1, is used for the signal extraction which is described in section 5. The NN mentioned above is used to bin the missing mass distribution where the number of bins is optimized for signal fit significance.

4.1 Missing mass

The missing neutrino information in the $B^+ \to \ell^+ \nu_\ell \gamma$ signal decay makes a signal identification more difficult. The best discriminator between signal and background is given by the squared missing mass, which corresponds to the squared neutrino mass. It is calculated as

$$m^2_{\mathrm{miss}} = (p_{B_{\mathrm{sig}}} - p_\ell - p_\gamma)^2,$$

where the four-momenta of the daughter lepton and photon are subtracted from that of the B_{sig} candidate. The four-momentum of the B_{sig} meson is obtained from: the four-momentum of the hadronically reconstructed B_{tag} meson, the measured boost of the $\Upsilon(4S)$ resonance, and the two-body decay topology of the resonance. Combining this information, the two B mesons have the same energy and opposite three-momenta in the rest frame of the $\Upsilon(4S)$.

For correctly reconstructed tag- and signal-side decays, the m^2_{miss} variable peaks around zero with a width which is given by the energy and momentum resolution of the three particles used for its calculation. The three particle resolutions which affect the resolution of m^2_{miss} are shown in Fig. 4.1a for signal MC. Here, the momentum resolution of the individual particles types are shown for correct particle hypotheses where the correct B_{tag} is identified according to section 3.5. The best resolution is found for muons followed by electrons. The momentum of these particles is determined from a CDC track fit which has the highest accuracy. Electrons experience energy loss through bremsstrahlung, this can be seen in the asymmetric deviation to momenta lower than the true values. The worst resolution is found for photons since their momentum or energy is reconstructed from the energy deposition in the ECL which has the lowest precision. The B_{tag} momentum resolution is in between the resolution of leptons and photon, since it is a combination of multiple particle resolutions of the hadronic decay chain.
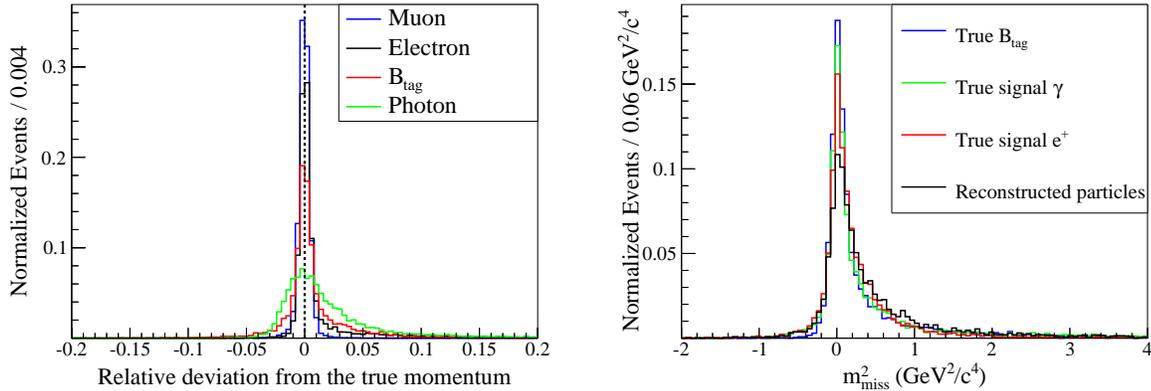
(a) Difference between reconstructed and true particle momenta for correctly identified signal and B_{tag} candidates, where the difference is divided by the true momentum.

(b) m^2_{miss} distribution for signal MC of the electron channel, where one of the input four-vectors for the m^2_{miss} calculation is set to the MC truth and the remaining ones are taken from the reconstruction. All particles of signal and tag-side are selected to be correctly identified.

Figure 4.1: Reconstruction resolution of the particles used to calculate m^2_{miss} (a), and its impact on the resolution of m^2_{miss} (b).

The energy is calculated from the momentum and therefore its resolution is determined by that of the momentum. For the B_{tag}, however, momentum and energy are reconstructed independently. Effects of the resolution on m^2_{miss} are shown in Fig. 4.1b for signal MC of the electron channel. Here, multiple m^2_{miss} distributions are plotted, where in each distribution a different four-momentum which is used in the m^2_{miss} calculation is set to its true value. As expected, the worse the four-momentum resolution in Fig. 4.1a, the bigger its impact on the width of the m^2_{miss}. Also the B_{tag} four-momentum fits into that pattern when accounting for the additional energy resolution which the other particles miss. These distributions serve as an illustration and are not directly comparable to the m^2_{miss} distributions of the signal in the analysis, since a significant amount of events with wrong B_{tag} particles pass the selection.

To improve the resolution of m^2_{miss}, bremsstrahlung photons are taken into account in the electron channel. For that, one bremsstrahlung photon below $1\,\text{GeV}$, lying in a five degree cone around the signal electron trajectory, is added to the four-vector of the signal electron. Additionally, the reconstructed energy of the B_{tag} particle is substituted with the beam energy in both signal channels (this is analogous to the

31
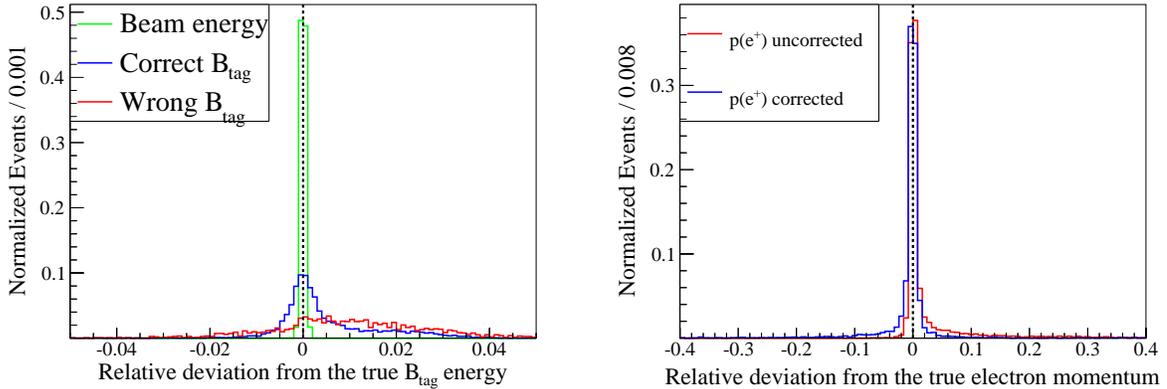
(a) Deviation of the true B_{tag} energy from the beam energy, and the reconstructed energy of wrong and correct tag-side candidates.

(b) Deviation of the reconstructed electron momentum with and without bremsstrahlung correction from its true momentum.

Figure 4.2: Improvement of the B_{tag} energy estimate through beam energy substitution (a) and the electron momentum by the addition of bremsstrahlung (b). The difference between the reconstructed and the true value is divided by the true value.

M_{bc} variable in section 3.4). The improvements in the resolution of the quantities can be seen in Fig. 4.2. The estimate of the B_{tag} energy improves significantly, regardless of whether the tag-side decay chain is correctly or wrongly reconstructed. This is important, since a significant portion of the wrongly reconstructed tag-side particles in signal MC passes the event selection of the analysis. The resolution of the electron momentum shows a smaller improvement, where the one-sided tail from the bremsstrahlung is reduced and the distribution becomes more symmetric around the true value. The corrections make the m_{miss}^2 distribution in Fig. 4.3 more peaking, thus increasing the significance of the signal. Furthermore, the shapes of the background change little in the signal region.

4.2 Signal selection

An efficient signal-side selection is determined on the normalized MC distributions, where requirements are applied successively and each plot shows a distribution with all previously described selections. All background samples described in section 3.2.2 and the correctly reconstructed signal MC, satisfying the MC matching condition described in section 3.5, are used to determine the selection. Misreconstructed signal decays constitute a negligible background after selection and will not be shown in
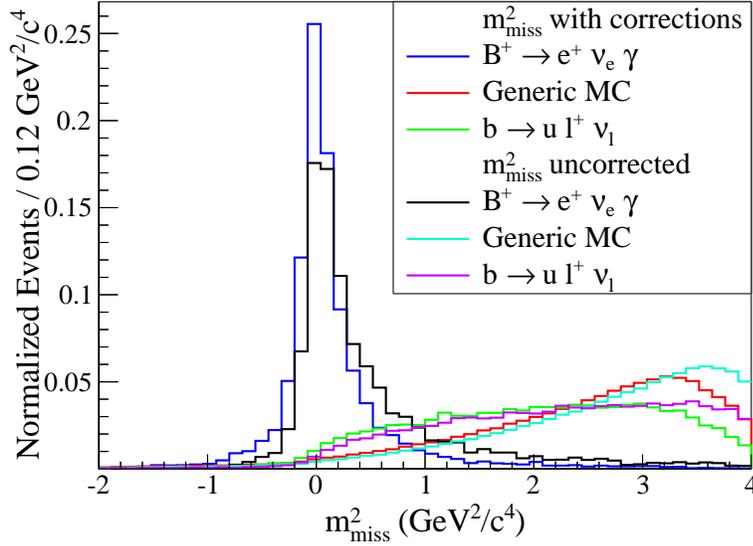
Figure 4.3: Comparison of m^2_{miss} distributions with and without beam energy and bremsstrahlung correction for correctly identified signal MC of the electron channel. The most important background sources of $b \to u\ell^+\nu_\ell$, generic $B\bar{B}$ and $q\bar{q}$ continuum decays are also shown.

the plots. Where relevant, the generic MC is split up into its resonant $B\bar{B}$ and non-resonant $q\bar{q}$ contribution. The selection thresholds are indicated by perpendicular lines.

Due to the similarity of the two signal decays $B^+ \to e^+\nu_e\gamma$ and $B^+ \to \mu^+\nu_\mu\gamma$, all requirements except for the lepton identification (ID) are chosen to be identical for both signal channels. No requirements are made on variables too strongly correlated to m^2_{miss} which is fitted to extract the signal. This includes, most importantly, the momentum of the signal lepton which carries most of the energy of the signal decay.

The signal lepton candidate is identified from charged tracks with the impact parameter requirements $|dz| < 4.0\,\text{cm}$ and $dr < 2.0\,\text{cm}$. For photons, ECL clusters without an associated charged track and an energy of at least $20\,\text{MeV}$ are taken into account. This primary particle selection is identical to the one used for the tag-side reconstruction.

A loose skimming selection is applied to reduce the size of the large background MC samples. It is required that:

- The remaining energy in the ECL, which is defined below, is below $1.2\,\text{GeV}$.

- At least one charged track with an electron ID or muon ID above 0.4 is found in the event.

- $m^2_{\text{miss}} \in (-2.0, 4.0)\,\text{GeV}^2/c^4$.

- The charges of the two B mesons are opposite, where the charge of the signal-side B meson (B_{sig}) is determined by the signal lepton candidate, and the B_{tag} charge is obtained from the full reconstruction.

(a) Electron channel. (b) Muon channel.
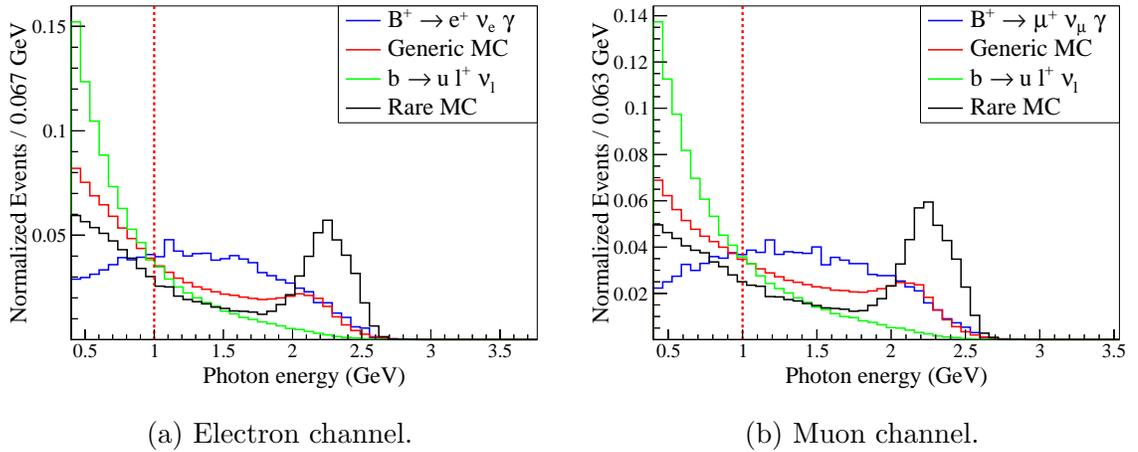
Figure 4.4: The initial distribution of the signal photon candidate energy in the B_{sig} rest frame. Thresholds are at $1\,\text{GeV}$ for the nominal and $400\,\text{MeV}$ for the secondary analysis. The peak in the rare MC originates from two-body decays of the form $b \to s\gamma$, where the strange-quark forms an excited kaon.

As discussed in section 3.2.1, in the nominal analysis the minimal signal photon energy has to be at least $1\,\text{GeV}$ in the B_{sig} rest frame for the theoretical model to be valid. However, believing the model to approximately reproduce the true spectrum down to energies of $400\,\text{MeV}$, about 30% of the signal is found to be removed by that requirement, which can be seen in Fig. 4.4. This selection can potentially lead to a reduction in signal significance, and therefore a secondary analysis is performed with signal photons above an energy of $400\,\text{MeV}$.

The efficiency for correctly reconstructed signal events in MC is 0.6% after full reconstruction and the skimming selections. A complete summary of the event numbers and selection efficiencies in MC for all selections is given in table 4.1.

34

Best photon candidate

In signal decays a high energetic photon is produced directly in the B meson three body decay. Therefore the highest energetic photon in the event is selected as the signal candidate. This selection only has an effect on the secondary analysis where the signal photon can have energies down to 400 MeV.

No charged tracks left

After full reconstruction and signal-side selection, no remaining charged track is expected for correctly reconstructed signal events. In some cases, a charged track can be reconstructed twice if it is low energetic and curls in the detector. This will only happen for lower energetic tag-side tracks. Therefore, fake tracks caused by curlers are accounted for in the track veto. Fake tracks are defined as follows:

- All charged tracks are compared to the remaining signal-side track where already used tag-side tracks are not matched to each other.

- The tracks are required to be in parallel, i.e. the cosine of the angle between two tracks has to be above 0.999.

- The transverse momentum difference is required to be below 30 MeV.

- If only one of the tracks has SVD hits this track is kept.

- Otherwise, the track with lower distance to the IP defined by $(dr/\sigma_{dr})^2 + (dz/\sigma_{dz})^2$ is kept, where the impact parameters are divided by their errors.

Fig. 4.5 shows the number of remaining tracks subtracted by the number of curling tracks, and the quantity is required to be zero. Taking into account the curling track gives about a one percent higher signal, than directly requiring the remaining charged tracks to be zero.

Lepton ID

The Belle particle ID variables, described in section 3.1, are used to select the signal lepton candidate, where the ID is required to be greater than 0.8 for both the muon and the electron channel. The variables provide an efficient and pure lepton selection which can be seen in the distributions in Fig 4.6. Since at this point nearly no wrong lepton candidates are found for signal events, the selection serves as a background veto. Nearly all of the rare MC is vetoed with this condition. The systematic error introduced by this cut is estimated in section 7.

(a) Electron channel.

(b) Muon channel.
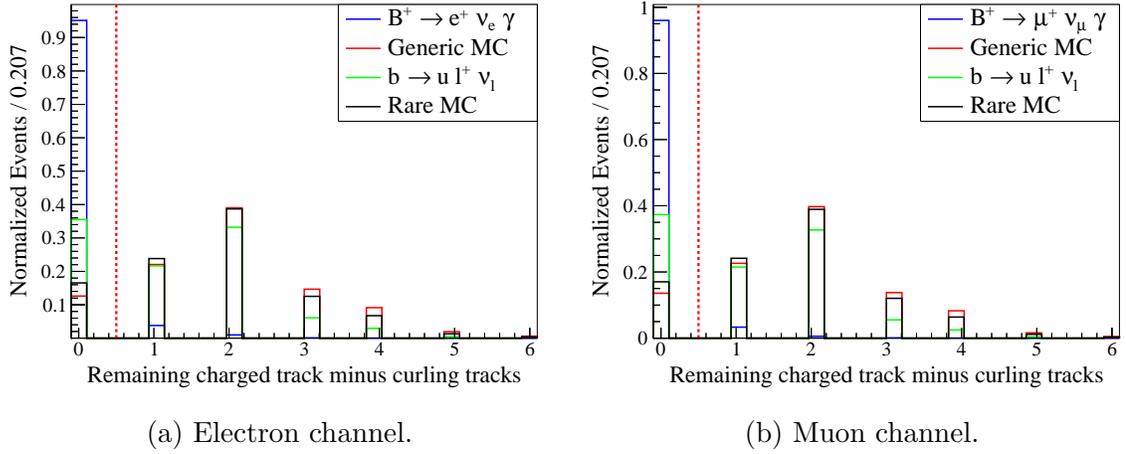
Figure 4.5: Remaining charged tracks subtracted by the number of curling tracks which is required to be zero.

(a) Electron ID.

(b) Muon ID.
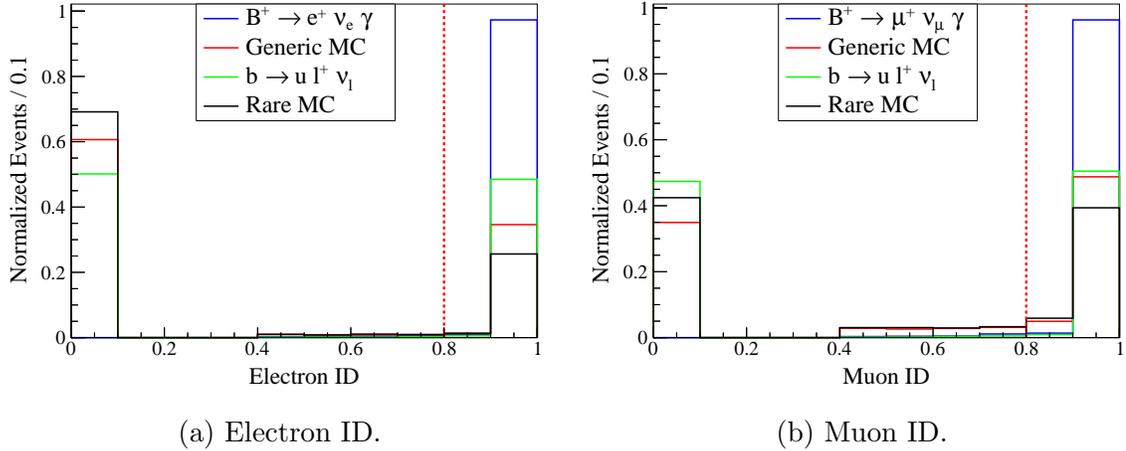
Figure 4.6: Lepton ID variables for the signal lepton candidate which is selected to be larger than 0.8 for both channels.

Extra energy in the ECL

For correctly reconstructed signal events little remaining energy is expected to be left in the ECL. However, even for well reconstructed signal, which peaks in m^2_{miss}, remaining energy can be found. This can be due to several sources: unreconstructed particles (mainly photons) from the tag-side where the *btag* candidate is still sufficiently well

reconstructed, bremsstrahlung from electrons, secondary particles created by particle interaction with the detector, and beam background.

The energy clusters in the ECL are summed to obtain the extra energy in the ECL variable. Only clusters above certain energy thresholds are taken into account which reduces "unphysical" background not associated with the $B\bar{B}$ process. The Belle collaboration has a standard definition for these energy thresholds which are proven to optimize the "physical" content of the quantity. They read

- $E_{\mathrm{cluster}} > 100\,\mathrm{MeV}$ for the forward end-cap $(0.548 > \Theta > 0.216)$

- $E_{\mathrm{cluster}} > 50\,\mathrm{MeV}$ for the barrel $(2.246 > \Theta > 0.562)$

- $E_{\mathrm{cluster}} > 150\,\mathrm{MeV}$ for the backward end-cap $(2.281 > \Theta > 2.707)$,

where Θ is the polar angle of the photon in the detector.

A plot showing the distribution of the variable is given in Fig. 4.7. It is required for the energy to be below 900 MeV. For entries exactly at zero, no cluster above the thresholds is detected in the event. Additionally, a gap is found between the zero and the minimal cluster counting threshold of 50 MeV. Background tends to have more remaining energy in the ECL than signal. Due to its good separation, the variable is used in the network training in section 4.3.

(a) Electron channel.　　　　　　　　(b) Muon channel.
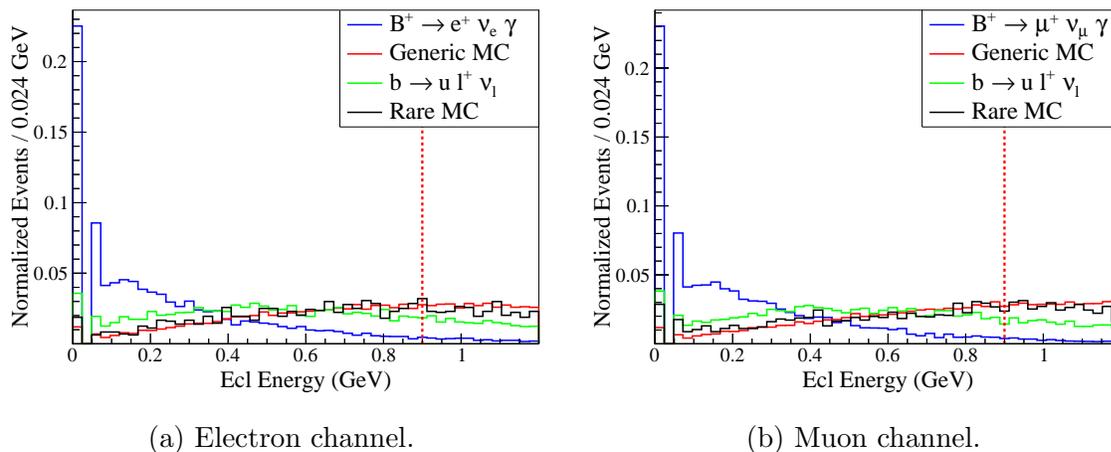
Figure 4.7: Extra energy in the ECL which is required to be below 900 MeV.

Angle between signal photon and lepton

Bremsstrahlung emitted by the signal electron candidate is in a few cases identified as the signal photon candidate for the background of the electron channel. In this case, the angular separation between the signal photon and lepton candidate ($\cos\Theta_{\gamma\ell}$) is small since the bremsstrahlung is emitted along the direction of the electron. The applied selection demands the cosine of the angle to be below 0.6 which is shown in Fig. 4.8, where the angle is computed in the rest frame of B_{sig}. As expected, no peak is observed for small angles in the muon channel.

(a) Electron channel. (b) Muon channel.
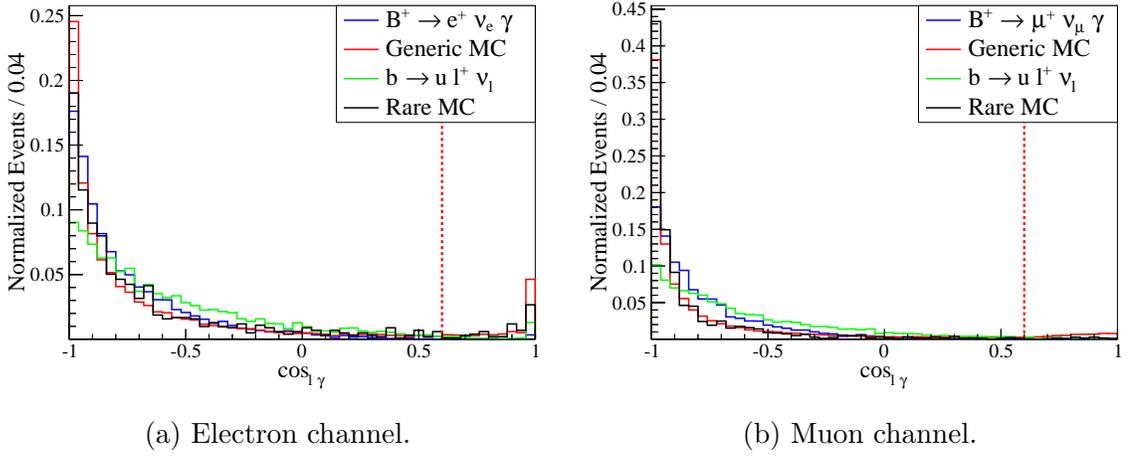
Figure 4.8: $\cos\Theta_{\gamma\ell}$ angle which is selected to be below 0.6.

Pion mass veto

A major background source of $B^+ \to \pi^0 \ell^+ \nu_\ell$ is suppressed with a selection on a pion mass variable. The π^0 decays into two photons where one of these photons is identified as the signal photon candidate. In order to find the second photon, the invariant π^0 mass is calculated from the signal photon candidate and the remaining photons in the calorimeter, and only the best combination closest to the nominal π^0 mass of about $135\,\text{MeV}/c^2$ is kept.

This procedure has the risk of producing artificial peaks in the mass spectrum through mere combinatorics, since the probability to produce a candidate close to the nominal mass increases with the number of possible combinations. The mass distribution changes, when only photons above a certain energy are combined with the signal photon candidate; this changes the number of possible combinations in the

event. No candidate is computed if no photon above the energy threshold is found. A rising energy threshold will reduce the number of remaining photons to be combined with the signal photon, thus reducing the probability for an artificial mass peak in the distribution. This also increases the signal efficiency, however, it will also discard photons from π^0 decays and reduce the power of the background veto.

For this selection, a high signal efficiency is demanded from the variable and therefore a high energy threshold of $100 \, \mathrm{MeV}/c^2$ is applied on the remaining photons in the calorimeter. The distribution is shown in Fig. 4.9 where all events in a window of $20 \, \mathrm{MeV}/c^2$ around the nominal π^0 mass are rejected. The signal MC shape is flat at the nominal mass, moreover, an η peak from $B^+ \to \eta \ell^+ \nu_\ell$ decays is reconstructed at the mass of $550 \, \mathrm{MeV}/c^2$. The variable does not optimize for η mesons, which is a sign that the variable contains physical information instead being a result of combinatorics.

Several of these variables with lower energy thresholds on the remaining photon in the ECL are used in the network training in section 4.3.

(a) Electron channel. (b) Muon channel.
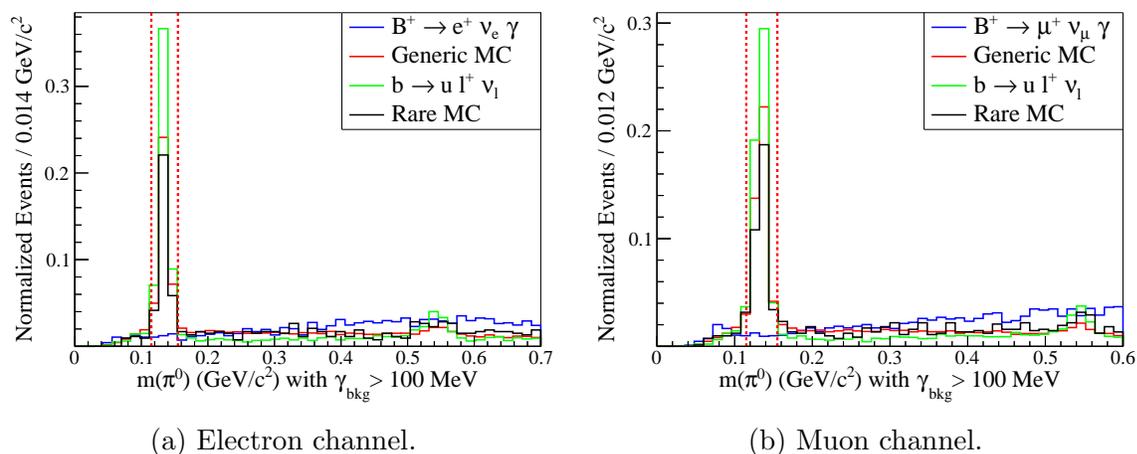
Figure 4.9: Invariant π^0 mass (m_{π^0}) reconstructed from the signal photon candidate and remaining photons in the ECL with energies above $100 \, \mathrm{MeV}$, where a mass window of $m_{\pi^0} \in (115, 155) \, \mathrm{MeV}/c^2$ is selected.

M_{bc} mass

The beam-constrained B_{tag} mass M_{bc}, which is defined in section 3.4, is required to be larger than $5.27 \, \mathrm{GeV}/c^2$ as shown in Fig. 4.10. A peak for the correctly reconstructed B_{tag} mesons is visible at the mass $5.28 \, \mathrm{GeV}/c^2$ and the selection removes non-resonant

$q\bar{q}$ background and events with wrongly reconstructed tag-sides. The mass region below $5.27\,\mathrm{GeV}/c^2$ is used as a sideband for data-MC consistency checks.
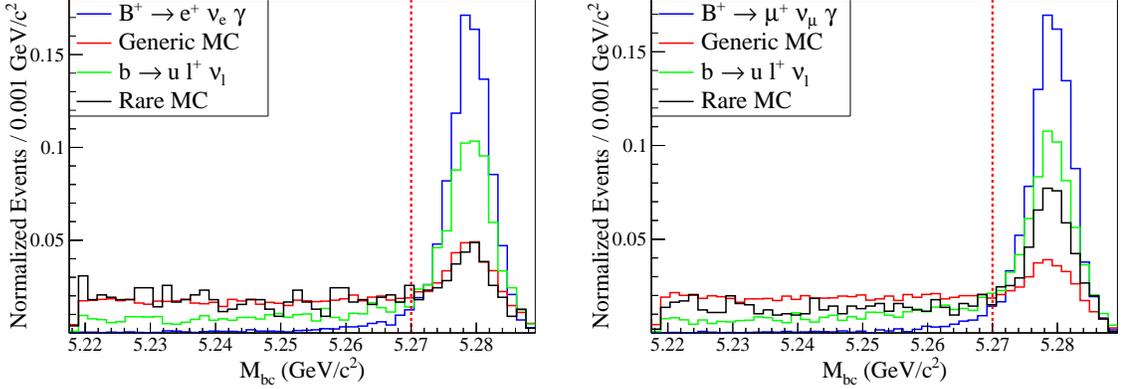
(a) Electron channel. (b) Muon channel.

Figure 4.10: Beam-constrained B_{tag} mass M_{bc} which is required to be above $5.27\,\mathrm{GeV}/c^2$.

4.2.1 Improvement of data-MC consistency

The following selections are mainly determined to reduce observed discrepancies between data and MC which are found in the $q\bar{q}$ continuum background. Specific selections remove these inconsistencies. Due to these discrepancies, continuum background is vetoed in general by using a continuum suppression. After this selection few background events are expected from continuum, as seen in Fig. 4.17. Additionally, the yield of the continuum component is a free parameter in the measurement, as described in section 5.

The contribution of the rare MC is negligible from this point on and therefore not shown anymore. Since the deviations are observed in the continuum background, the generic MC sample is displayed separately with its resonant $B\bar{B}$ and non-resonant $q\bar{q}$ contributions.

Missing momentum

The missing momentum (p_{miss}) is defined analogously to the m^2_{miss} variable, where the beam momentum is subtracted by the B_{tag}, signal photon, and signal lepton momentum.

40
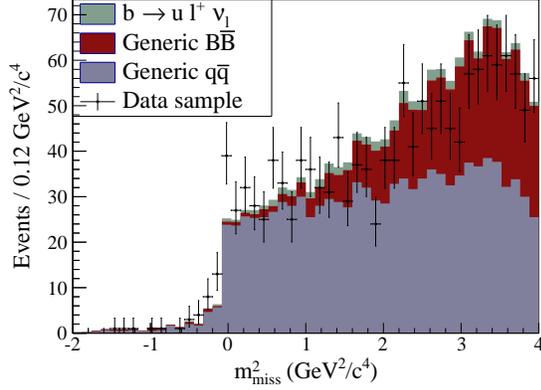
(a) Muon channel before requiring high p_{miss}.
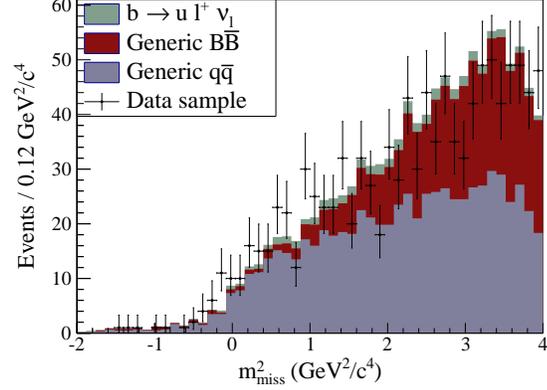
(b) Muon channel after requiring high p_{miss}.
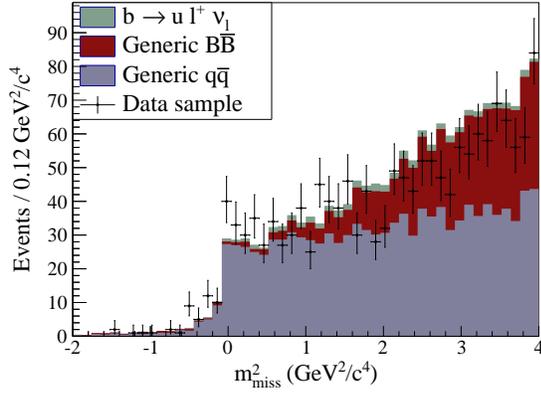
(c) Muon channel showing m^2_{miss} computed using the beam energy without requiring high p_{miss}.
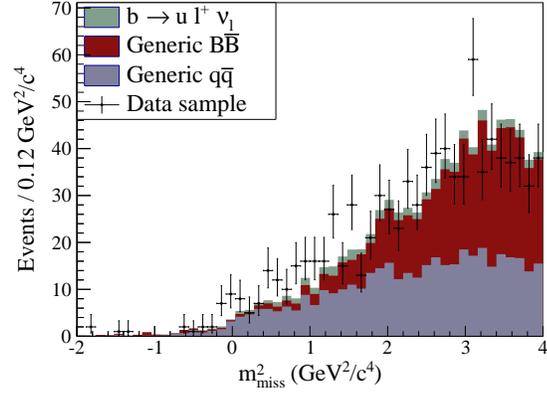
(d) Electron channel.

Figure 4.11: m^2_{miss} distributions in the sideband of $M_{\text{bc}} < 5.27\,\text{GeV}/c^2$. A discrepancy is found for the muon channel if the reconstructed B_{tag} energy is used in the m^2_{miss} calculation (a), and it vanishes by requiring p_{miss} to be above $800\,\text{MeV}/c$ (b). A smaller discrepancy is found if the beam energy is used instead of the B_{tag} energy for the m^2_{miss} calculation (c), and no discrepancy is found for the electron channel (d).

The variable is introduced to the selection to remove badly modeled background observed in the sideband of $M_{\text{bc}} < 5.27\,\text{GeV}/c^2$. Here, the m^2_{miss} variable calculated with the reconstructed B_{tag} energy instead of the beam energy (see section 4.1) shows a discrepancy between data and MC in the muon channel, as seen in Fig. 4.11a. The deviation is less pronounced when using the beam energy for the m^2_{miss} calculation,

41

(a) Electron channel.
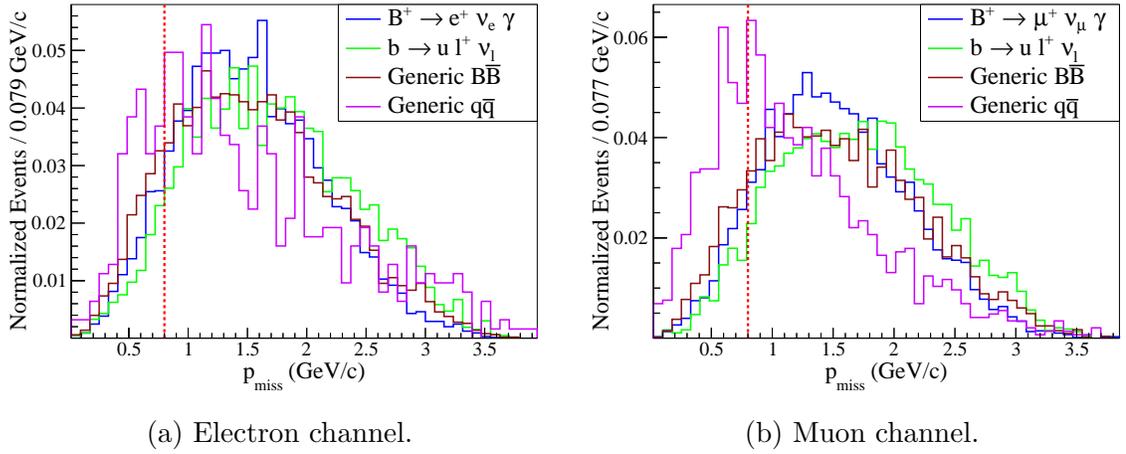
(b) Muon channel.

Figure 4.12: Missing momentum which is selected to be larger than $800 \,\mathrm{MeV}/c$.

(a) Electron channel.
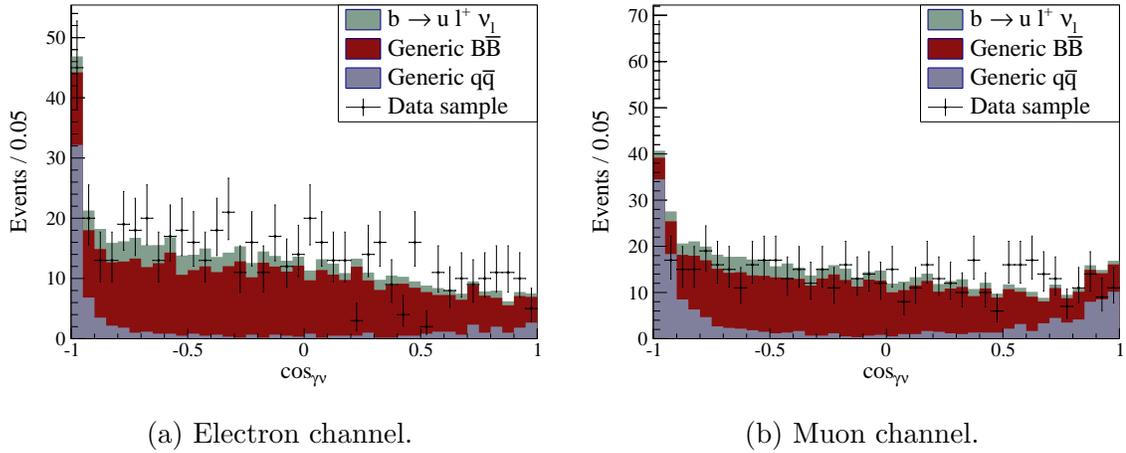
(b) Muon channel.

Figure 4.13: $\cos\Theta_{\gamma\nu}$ angle in the sideband of $m_{\mathrm{miss}}^2 > 0.3 \,\mathrm{GeV}^2/c^4$. A discrepancy is found in the muon channel for large angles.

as shown in Fig. 4.11c.

The discrepancy is caused by non-resonant $q\bar{q}$ processes in which the signal-side muon candidate is in fact a misidentified charged pion or kaon. The signal-side has a large invariant mass similar to the wrong B_{tag} candidate which results in a m_{miss}^2 value near zero. The whole event is reconstructed in a balanced way which also results in a p_{miss} value near zero. Therefore the discrepancy disappears in Fig. 4.11b after a high p_{miss} larger than $800 \,\mathrm{MeV}/c$ is required. Additionally, the shape becomes smoother,

(a) Electron channel.   (b) Muon channel.
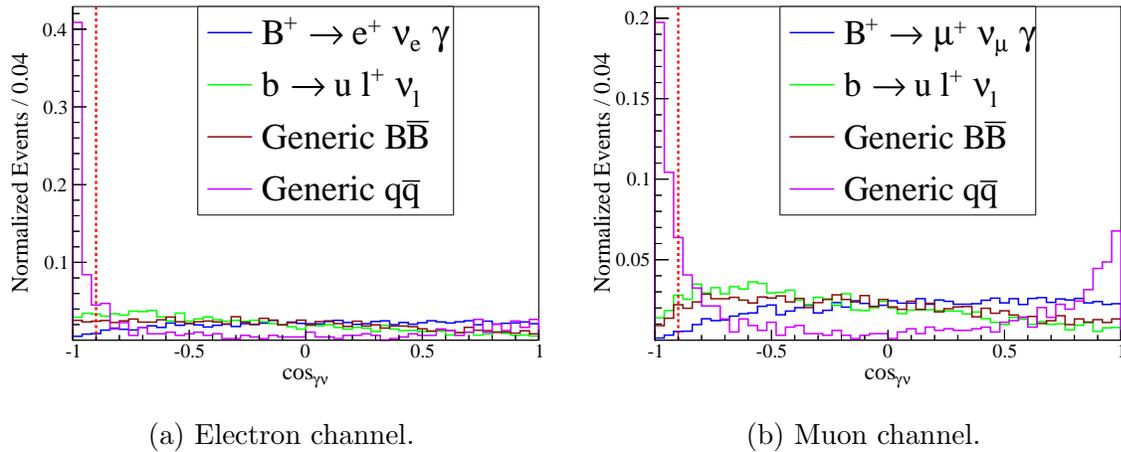
Figure 4.14: $\cos\Theta_{\gamma\nu}$ angle which is selected to be larger than -0.9.

losing its step-like feature around zero.

Since the electron cannot be misidentified with a pion or kaon, no discrepancy is observed in the electron channel in Fig. 4.11d. The distributions for p_{miss} are shown in Fig. 4.12.

Angle between the signal photon candidate and the missing momentum

The angle between the signal photon and p_{miss} ($\cos\Theta_{\gamma\nu}$) has a data-MC discrepancy in the sideband of $m^2_{\mathrm{miss}} > 0.3\,\mathrm{GeV}^2/c^4$, which is shown in Fig. 4.13. Again, the discrepancy is only observed in the muon channel where the particles are flying back to back.

Continuum background dominates for small and large angles, as shown in Fig. 4.14. This can be explained with the jet structure of these events which allows for two likely possibilities for the angle: either the signal photon candidate is taken from the same jet as the p_{miss}, or from the opposite jet. Therefore it is concluded that the discrepancy originates from continuum processes.

The discrepancy is vetoed by requiring the cosine of the angle to be larger than -0.9 for both the electron and the muon channel. Due to the helicity structure of $B^+ \to \ell^+\nu_\ell\gamma$ decays, the lepton tends to fly in opposite direction of the neutrino and photon; therefore little signal is rejected by this requirement.

43

$B_{\mathbf{tag}}$ NN output with continuum suppression

The NN output of the full reconstruction from section 3.4 is used to suppress continuum background. The output contains event shape variables in order to identify jet-like events, as described in section 3.3.

The continuum network output in Fig. 4.15 is shown for the generic MC split up into its $B\bar{B}$ and $q\bar{q}$ continuum component. The variable can be interpreted as a Bayesian probability for the B_{tag} meson to be a correct candidate. As seen in Fig. 4.15c and 4.15d in the range up to 10^{-3}, the network distinguishes continuum from $B\bar{B}$ events where the former event type has smaller values. It is required that the output value is above 2×10^{-4}.

The distribution of the NN output in the sideband of $M_{\text{bc}} < 5.27\,\text{GeV}/c^2$ in Fig 4.16 shows a good agreement between data and MC for both signal channels. The systematic error introduced by this selection is estimated in section 7.

Summary

The selection efficiencies for each step and all MC samples are listed in table 4.1. The overall signal selection efficiency is 47% and 45% for the electron and muon channel, respectively. The expected event numbers from the background MC samples are: 328 (299) for $b \to c$ decays, 78 (76) for $b \to u\ell^+\nu_\ell$ decays, and 17 (6) events from non-resonant $q\bar{q} \to (u, d, s, c)$ processes for the muon (electron) channel. The contribution from rare MC is found to be negligible. The m^2_{miss} distribution after the selection with the MC weighted to the expected data yields is shown in Fig 4.17.
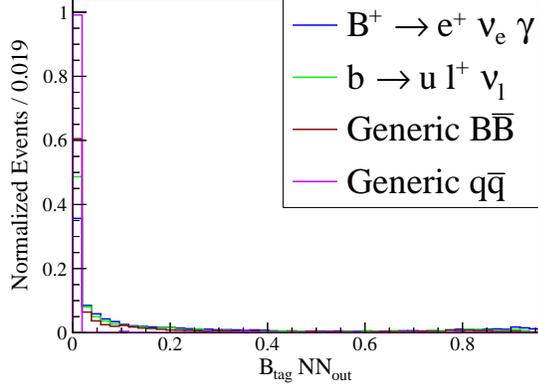
(a) Electron channel in the whole range.
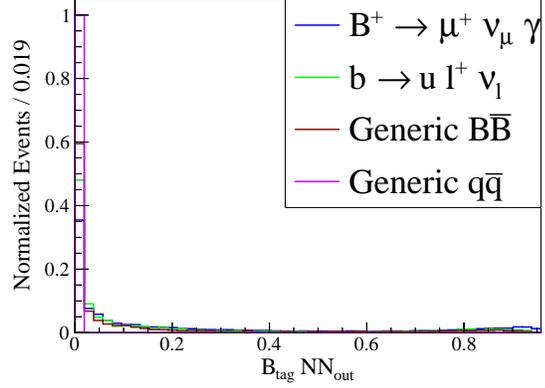
(b) Muon channel in the whole range.
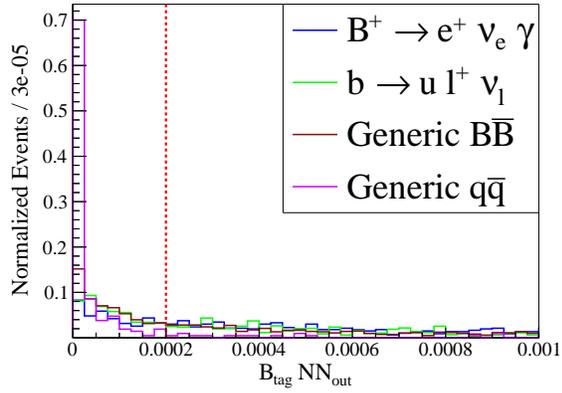
(c) Electron channel in the sub-range.
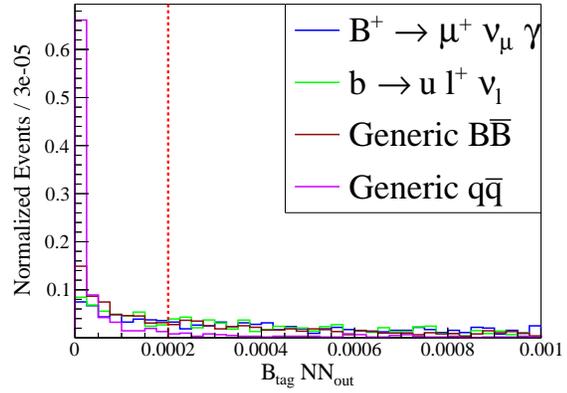
(d) Muon channel in the sub-range.

Figure 4.15: B_{tag} NN output containing event shape variables shown for the whole value range in (a) and (b). For better visibility the sub-range up to 10^{-3} is also shown in (c) and. (d).
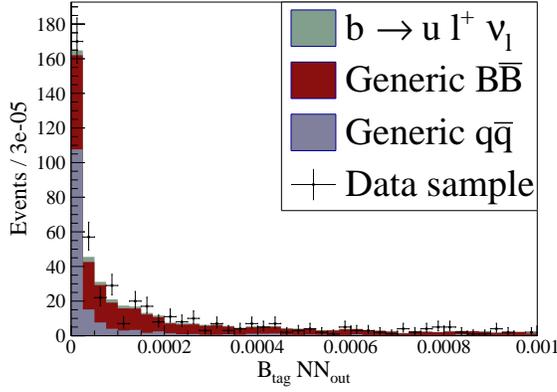
(a) Electron channel.
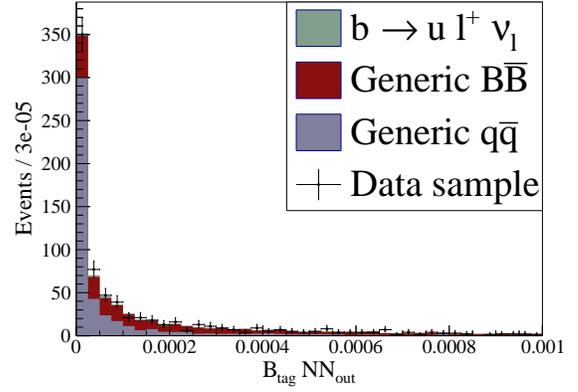
(b) Muon channel.

Figure 4.16: B_{tag} NN output with event shape variables in the side band of $M_{\mathrm{bc}} < 5.27\,\mathrm{GeV}/c^2$. A good agreement between data and MC can be seen for both channels.
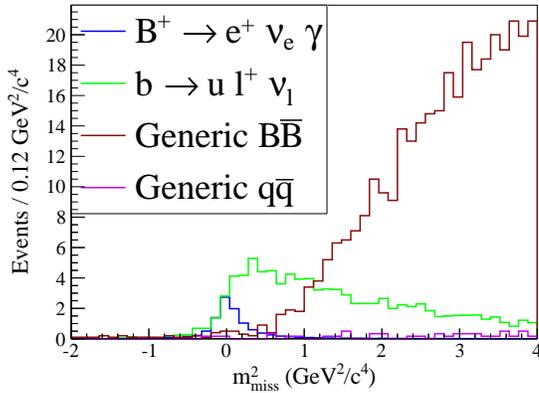
(a) Electron channel.
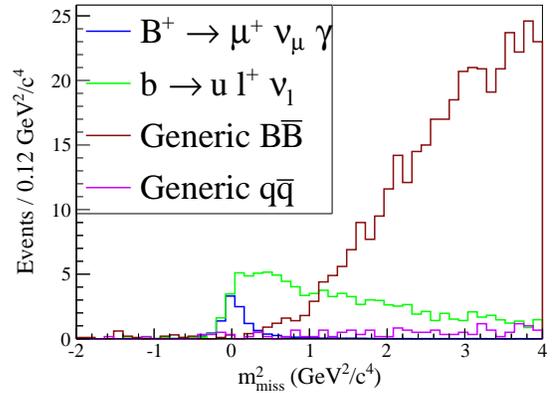
(b) Muon channel.

Figure 4.17: m^2_{miss} distribution after all selections with the MC samples weighted to the luminosity of the data sample. The signal MC is shown with a branching fraction of 5×10^{-6}.

Table 4.1: Event numbers and efficiencies for the signal selection of the electron and muon channel. The overall efficiencies do not include skimming and full reconstruction which has an efficiency of 0.14% for signal MC.

	Muon channel							
	Event numbers				Efficiencies			
MC sample	Signal	Generic	$b \to u\ell^+\nu_\ell$	Rare	Signal	Generic	$b \to u\ell^+\nu_\ell$	Rare
Initial	8084	526422	28990	30642	-	-	-	-
Best. Photon	8081	522023	28915	30563	1	0.99	1	1
Tracks	7768	81692	16105	5847	0.96	0.16	0.56	0.19
Lepton ID	7593	43010	8020	2619	0.98	0.53	0.5	0.45
Energy Ecl	7348	27716	6536	1775	0.97	0.64	0.81	0.68
$\cos\Theta_{\gamma\ell}$	7339	26424	6458	1750	1	0.95	0.99	0.99
Pion mass	7275	18969	4277	1464	0.99	0.72	0.66	0.84
M_{bc}	6817	6759	2971	738	0.94	0.36	0.69	0.5
Missing Mom.	6093	5536	2743	527	0.89	0.82	0.92	0.71
$\cos\Theta_{\gamma\nu}$	6048	5053	2598	261	0.99	0.91	0.95	0.5
B_{tag} NN output	5805	3873	2415	215	0.96	0.77	0.93	0.82
Overall efficiencies	-	-	-	-	0.7181	0.0074	0.0833	0.007

	Electron channel							
	Event numbers				Efficiencies			
MC sample	Signal	Generic	$b \to u\ell^+\nu_\ell$	Rare	Signal	Generic	$b \to u\ell^+\nu_\ell$	Rare
Initial	7940	564627	30067	31133	-	-	-	-
Best. Photon	7872	559586	29973	31047	0.99	0.99	1	1
Tracks	7503	85908	16501	6016	0.95	0.15	0.55	0.19
Lepton ID	7380	30116	8128	1437	0.98	0.35	0.49	0.24
Energy Ecl	7107	20136	6597	1014	0.96	0.67	0.81	0.71
$\cos\Theta_{\gamma\ell}$	7071	18800	6448	949	0.99	0.93	0.98	0.94
Pion mass	6991	13777	4217	780	0.99	0.73	0.65	0.82
M_{bc}	6519	5661	2969	262	0.93	0.41	0.7	0.34
Missing Mom.	5809	4799	2715	218	0.89	0.85	0.91	0.83
$\cos\Theta_{\gamma\nu}$	5710	4278	2498	107	0.98	0.89	0.92	0.49
B_{tag} NN output	5467	3442	2293	79	0.96	0.8	0.92	0.74
Overall efficiencies	-	-	-	-	0.6885	0.0061	0.0763	0.0025

4.3 Network training

After selection, the most relevant background source is given by $b \to u\ell^+\nu_\ell$ processes where the up-quark forms a π^0 or η meson, which in turn decay into two photons. The light mesons are produced with a high momentum, and thus the boosted system can decay with a large energy asymmetry giving one of the photons most of its energy. This photon is identified as the signal photon candidate, and together with a correct lepton from the B decay, the signal signature is reproduced. About 60% (20%) of the background originates from $B^+ \to \pi^0\ell^+\nu_\ell$ ($B^+ \to \eta\ell^+\nu_\ell$) in the signal region of $m_{\text{miss}}^2 < 1\text{GeV}^2/c^4$. Due to the dominance of these decays, dedicated MC samples are produced with 2×10^7 events for each meson decay channel. Fig. 4.18 shows the distribution of the dominant backgrounds.

(a) Electron channel. (b) Muon channel.
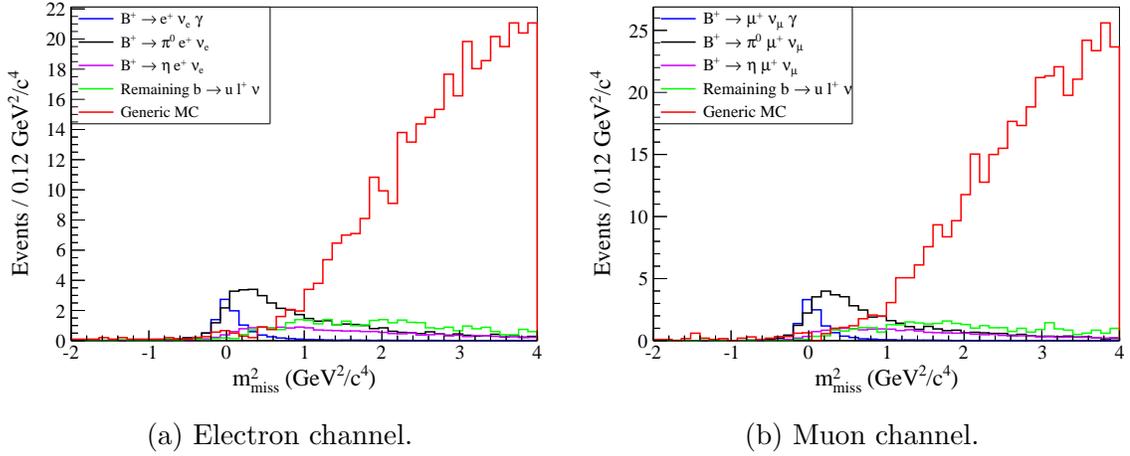
Figure 4.18: m_{miss}^2 distribution where the background sources of $B^+ \to \pi^0\ell^+\nu_\ell$ and $B^+ \to \eta\ell^+\nu_\ell$ dominate in the signal region below $1\text{GeV}^2/c^4$.

The remaining background originates from $B\bar{B}$ events with $b \to c$ transitions. This includes several final states, but also here about 50% of the signal photon candidates originate from π^0 decays.

To further separate signal and background, π^0 and η vetoes are computed to identify the second photon of the meson decay in the ECL. Invariant meson masses, as already introduced in section 4.2, are computed from the signal photon candidate which is combined with the remaining photons in the ECL and the invariant mass closest to nominal mass is kept. Many of theses variables are computed with different energy thresholds on the photons in the ECL. The thresholds, which are increased in 10 MeV steps, lie between 20 and 100 MeV for the π^0 and 20 to 300 MeV for the

η masses. Different photon combinations end up in the mass spectrum for different thresholds, since only the one combination closest to the nominal mass is taken into account. This leads to different invariant mass spectra with complementary information.

To illustrate the effect of the energy thresholds on the remaining photons in the ECL, two π^0 mass spectra are shown in Fig 4.19. The first plot shows the mass with the minimum 20 MeV threshold on the ECL photons and the second plot has a threshold at 60 MeV. The distributions show the background samples of the generic MC, $B^+ \to \pi^0 \ell^+ \nu_\ell$ decays, $B^+ \to \eta \ell^+ \nu_\ell$ decays, and remaining decays from the $b \to u \ell^+ \nu_\ell$ MC sample. With the lower energy threshold for the remaining photons, all samples, including the signal MC, gather around the nominal π^0 mass of about 135 MeV where events containing a photon from a π^0 peak more strongly. A decrease in combinatorics can be seen when the energy threshold is increased. The shape of the signal MC and the $b \to u \ell^+ \nu_\ell$ sub-sample which does not contain any π^0 of η decays become flat at the π^0 mass. Additionally, a peak for the η mass at 550 MeV shows up for the $B^+ \to \eta \ell^+ \nu_\ell$ sample.

(a) Electron channel.   (b) Muon channel.
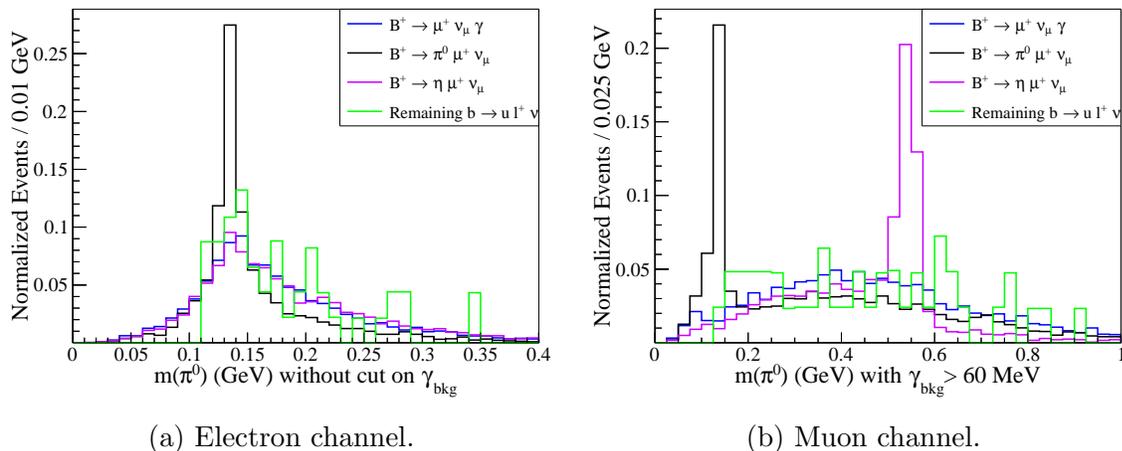
Figure 4.19: π^0 mass computed from the signal photon candidate and a remaining photon in the ECL. The masses are plotted for two different energy thresholds on the ECL photons where the artificial peak around the nominal π^0 mass disappears for the higher threshold.

A second set of meson masses is computed with the same energy thresholds as used for the extra energy in the ECL variable described in section 4.2. These take the different background levels in the barrel, forward, and backward direction of the calorimeter into account. Again, several variables are defined, where the values for

49

the thresholds of the ECL variable are scaled by factors between 0.5 and 1.5 which are increased in steps of 0.1.

All variables are combined in a NN which is formed with the NeuroBayes package [27]. This software computes each input variable's significance to the training. Only the most significant variables in the network are retained. Apart from the meson masses, included variables are: the extra energy in the ECL, $\cos\Theta_{\gamma\ell}$, and $\cos\Theta_{\gamma\nu}$. For the secondary analysis, different variables are used in the network training and the angles among the signal particle candidates are excluded. This aims to reduce the signal-model dependence of the network output for the secondary analysis, since it contains a kinematic region for which the theoretical prediction is unreliable. The complete list of variables used for the two trainings can be found in table 4.2.

(a) Electron channel.　　　　　　　　　(b) Muon channel.
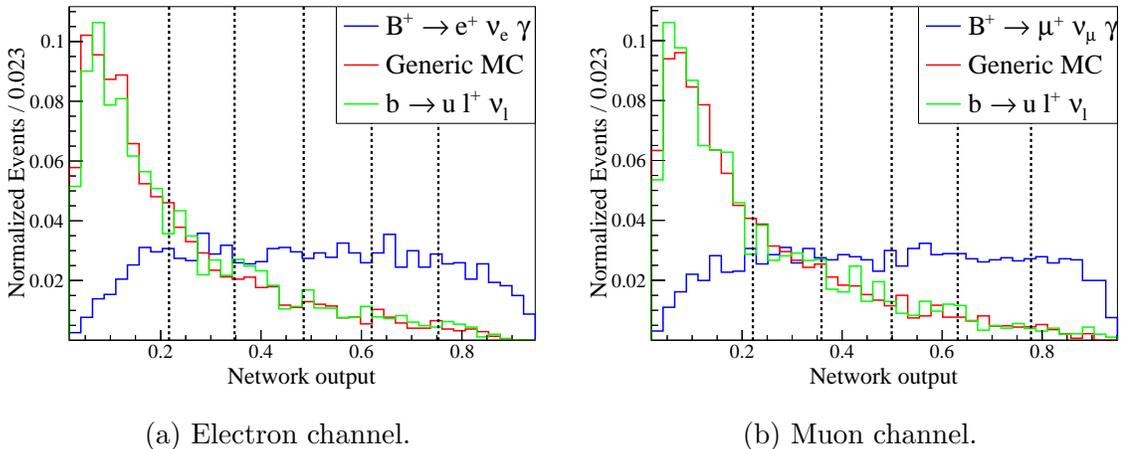
Figure 4.20: NN output for the independent MC samples. The vertical lines indicate the bin boundaries used for the binned fit described in section 5.

The training is carried out jointly for the electron and muon channel in the signal region of $m^2_{\mathrm{miss}} < 1\,\mathrm{GeV}^2/c^4$. Signal MC is trained against the $b \to u\ell^+\nu_\ell$ MC and the dedicated $B^+ \to \pi^0\ell^+\nu_\ell$ MC. Since the $b \to u\ell^+\nu_\ell$ MC has a lower luminosity than the $B^+ \to \pi^0\ell^+\nu_\ell$ MC and contains several smaller background sources, it is weighted more strongly in the training. It is found that this training constellation gives the best separation where no improvement is observed by adding the generic MC.

Training and analysis are performed on separate samples to assess the independence of the network training to random sample fluctuations and ensure an unbiased optimization of the analysis. Therefore, each sample is split in two parts of equal size by dividing the events randomly to a training and an independent sample. The separation power of the network output is shown in Fig. 4.20, and the agreement of
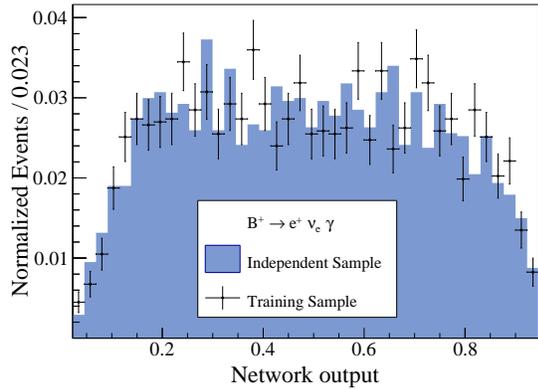
(a) Electron channel signal MC.
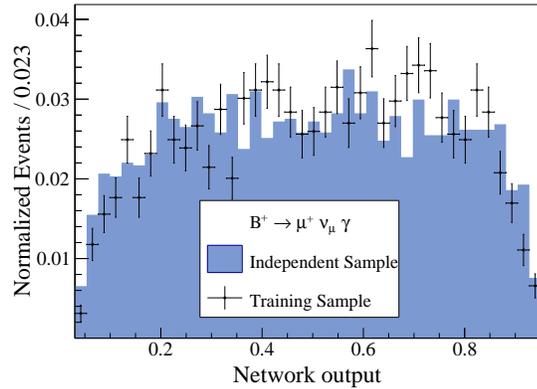
(b) Muon channel signal MC.
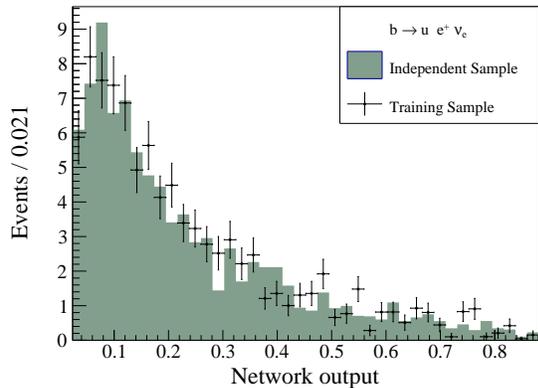
(c) Electron channel $b \to u\ell^+\nu_\ell$ MC.
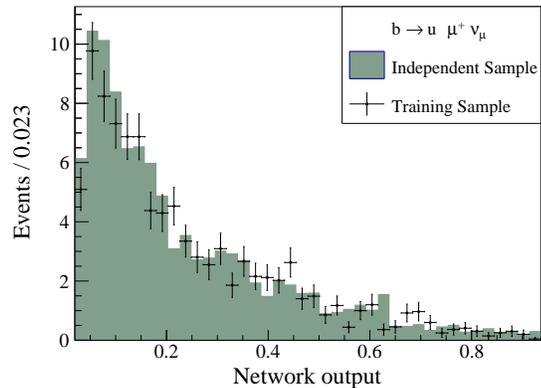
(d) Muon channel $b \to u\ell^+\nu_\ell$ MC.

Figure 4.21: Network output comparison between the training and independent sample for the training components signal MC and $b \to u\ell^+\nu_\ell$ MC for both signal channels.

the network outputs between training and independent sample is shown in Fig. 4.21 for the signal and $b \to u\ell^+\nu_\ell$ MC. The outputs for the secondary analysis are given in appendix A.4. The network outputs are in good agreement for all MC components proving that the network is not over-trained.

More sophisticated variables have been tested to improve the separation of the training, but none of them delivered higher signal significance. A veto for high momentum π^0 and η mesons decaying into photons, proposed in [30], has been included in the training. Here, the signal photon candidate is combined with all remaining

photons in the ECL to determine two-dimensional PDFs giving the probability of a true meson being reconstructed. The PDFs, which are determined separately for each of the three ECL segments, depend on the energy of the signal photon candidate and the mass of the reconstructed meson.

A second test uses a NN to identify true photons in the ECL, which are then combined with signal photon candidate to compute the meson veto variables.

These tests show that the combination of variables used in the analysis provide a good separation, and no significant improvement can be achieved for the expected statistics of the signal.

Data-MC differences for the network output are checked in sidebands in section 6 where no significant deviation between data and MC is found. Additionally, the systematic error on the network output is determined in the $B^+ \to K^*(892)\gamma$ control channel in section 7.

Table 4.2: NN variables used in the joint training of the electron and muon channel. Different variables are used for nominal and secondary analysis which are sorted by significance to the training, according to NeuroBayes [27].

Nominal analysis ($E_\gamma^{\text{sig}} > 1\,\text{GeV}$)	Secondary analysis ($E_\gamma^{\text{sig}} > 400\,\text{MeV}$)
$\cos\Theta_{\gamma\ell}$	$m(\pi^0)$ with $\text{E}(\gamma_{\text{ECL}}) > 40\,\text{MeV}$
Extra energy in the ECL	Extra energy in the ECL
$m(\pi^0)$ with $\text{E}(\gamma_{\text{ECL}}) > 40\,\text{MeV}$	$m(\pi^0)$ with $\text{E}(\gamma_{\text{ECL}}) > 20\,\text{MeV}$
$\cos\Theta_{\gamma\nu}$	$m(\eta)$ with $\text{E}(\gamma_{\text{ECL}}) > 300\,\text{MeV}$
$m(\eta)$ with $\text{E}(\gamma_{\text{ECL}}) > 300\,\text{MeV}$	$m(\pi^0)$ with ECL cuts scaled by 0.6
$m(\pi^0)$ with $\text{E}(\gamma_{\text{ECL}}) > 20\,\text{MeV}$	$m(\eta)$ with $\text{E}(\gamma_{\text{ECL}}) > 100\,\text{MeV}$
$m(\pi^0)$ with $\text{E}(\gamma_{\text{ECL}}) > 60\,\text{MeV}$	$m(\pi^0)$ with ECL cuts scaled by 1.2
$m(\pi^0)$ with ECL cuts scaled by 0.6	
$m(\eta)$ with $\text{E}(\gamma_{\text{ECL}}) > 100\,\text{MeV}$	

4.4 Known data-MC discrepancies

Two corrections are applied to the MC samples to correct for known data-MC differences of the B_{tag} and lepton ID efficiency.

4.4.1 B_{tag} efficiency correction

The reconstruction efficiency of the hadronic tag shows discrepancies between MC and data which are studied in [31]. Here, the following charmed semi-leptonic decay channels with large branching fractions are analyzed in fully reconstructed $B\bar{B}$ events:

- $B^+ \to D^0(K^-\pi^+)\ell\nu_\ell$
- $B^+ \to D^0(K^-\pi^+\pi^0)\ell\nu_\ell$
- $B^+ \to D^0(K^-\pi^+\pi^+\pi^-)\ell\nu_\ell$
- $B^+ \to D^{*0}(D^0(K^-\pi^+)\pi^0)\ell\nu_\ell$
- $B^+ \to D^{*0}(D^0(K^-\pi^+)\gamma)\ell\nu_\ell.$
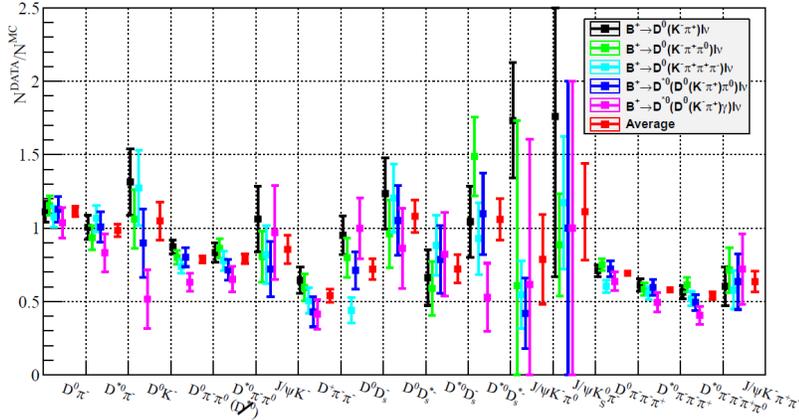
Figure 4.22: Data-MC tag-side efficiency difference plotted for all tag-side decay modes and several charmed signal-side decay channels [31].

The data-MC yield differences are measured in fits to m^2_{miss} distributions of the signal-side decay where the results are shown in Fig. 4.22 grouped by the tag-side decay modes. From the plot it is evident that tag-side decay modes exhibiting the largest discrepancy contain two or more charged or neutral pions. This effect dominates the data-MC differences, since it includes many of the high-statistics D^0 channels. Also modes with a charged D_s meson seem to have a higher efficiency on MC, on the other hand, modes with excited charged D_s mesons show no difference.

An average over all decay modes shows that yields on MC are overestimated by approximately one third. The author of the study provides a function to correct for these differences which computes the MC weight on an event basis depending on the tag-side decay channel and the B_{tag} network output. The event weights are shown in Fig. 4.23 for signal and background MC.
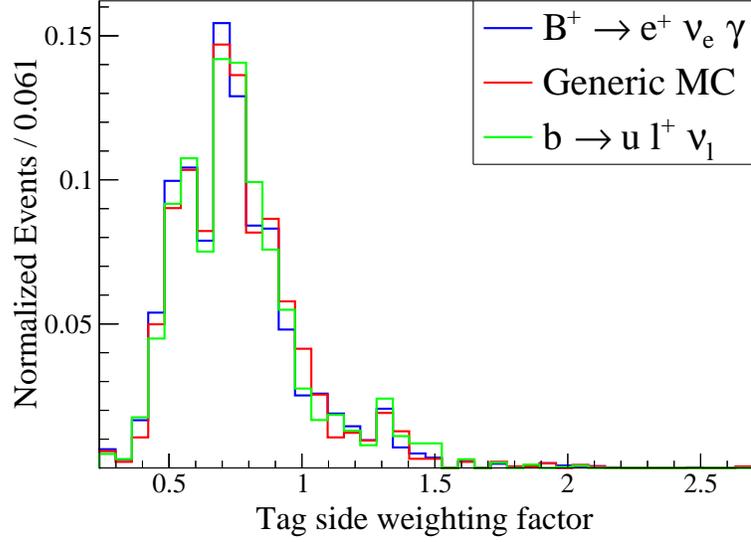
Figure 4.23: Event weights correcting for data-MC tag-side efficiency differences which are determined in [31]. The weights depend on the tag-side decay mode and the B_{tag} network output.

It is observed, that only events with a correctly reconstructed B_{tag} exhibit a data-MC discrepancy. This can be seen in Fig. 4.25 where the M_{bc} mass is plotted after the signal selection from section 4.2 in the sideband of $m_{\mathrm{miss}}^2 > 0.3 \mathrm{GeV}^2/c^4$. A difference between data and MC is only observed for the mass peak at 5.28 GeV/c^2 which contains events where the tag-side is correctly reconstructed. A correct B_{tag} is defined according to section 3.5. The discrepancy is larger for the muon channel, and the agreement between data and MC improves after the corrective weights are applied to the MC.

4.4.2 Lepton ID efficiency correction

Also the Belle lepton ID variables, described in section 3.1, exhibit systematic data-MC differences. The difference is investigated in [32] for four-lepton processes

(a) Electron channel

(b) Muon channel
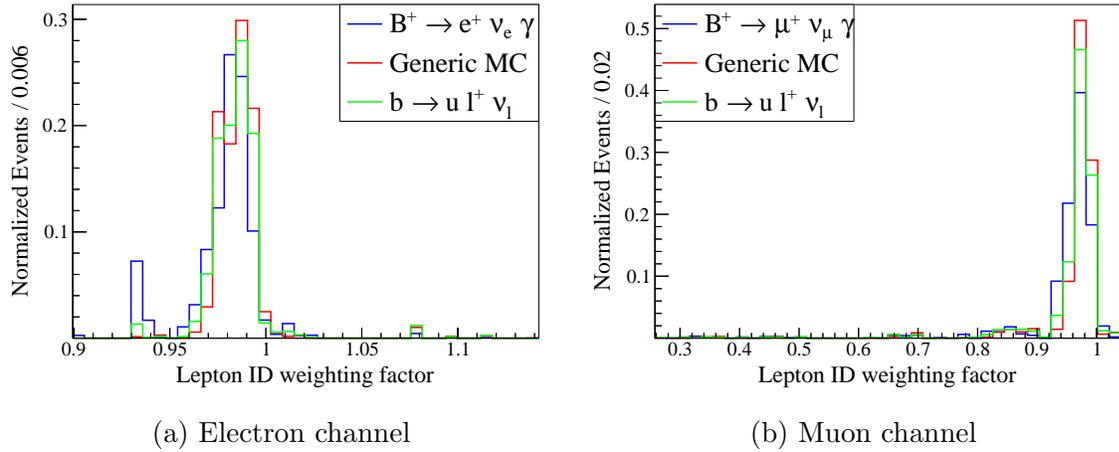
Figure 4.24: Event weights correcting the data-MC efficiency difference in the lepton ID.

$e^+e^- \to e^+e^-\ell^+\ell^-$. The systematic deviation of the likelihoods depends on the polar angle and the momentum of the particle, as indicated in Fig. 3.1. For that reason, the systematic errors are studied in seven bins of the polar angle and ten $500\,\mathrm{MeV}/c$ wide bins of the momentum. Additionally, inclusive $B \to XJ/\psi(\to \ell^+\ell^-)$ decays are used to study the systematic deviation in a B decay environment with a higher track multiplicity. The observed difference to the four-lepton process is added to the systematic error. The discrepancy is determined for different selections of the lepton ID variable and separately for the SVD1 and SVD2 setups, which were two different SVD detectors used during the operation of the experiment. The weights which correct the discrepancy for the MC samples used in the analysis are shown in Fig. 4.24. The average yield on MC is few percent higher than on data.

m^2_{miss} distribution after correction

The effect on the m^2_{miss} distribution after applying both corrections described above is shown in Fig. 4.26. An overall yield reduction is observed with minor changes in the shape of the distributions. The slight shape change at the signal peak in Fig. 4.26b and 4.26d is due to the higher B_{tag} candidate quality in this region which results in a smaller weight than in regions with larger portions of wrong B_{tag} candidates.
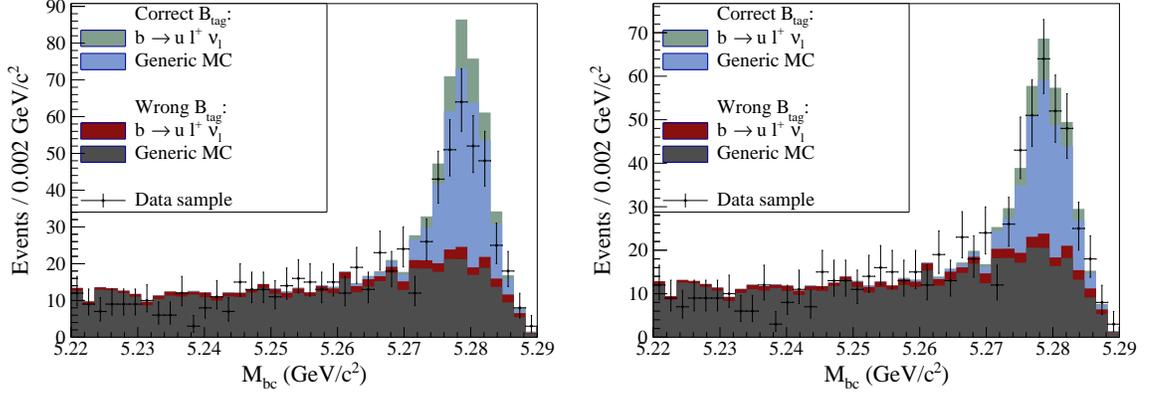
(a) Muon channel without corrective weights.

(b) Muon channel with corrective weights.
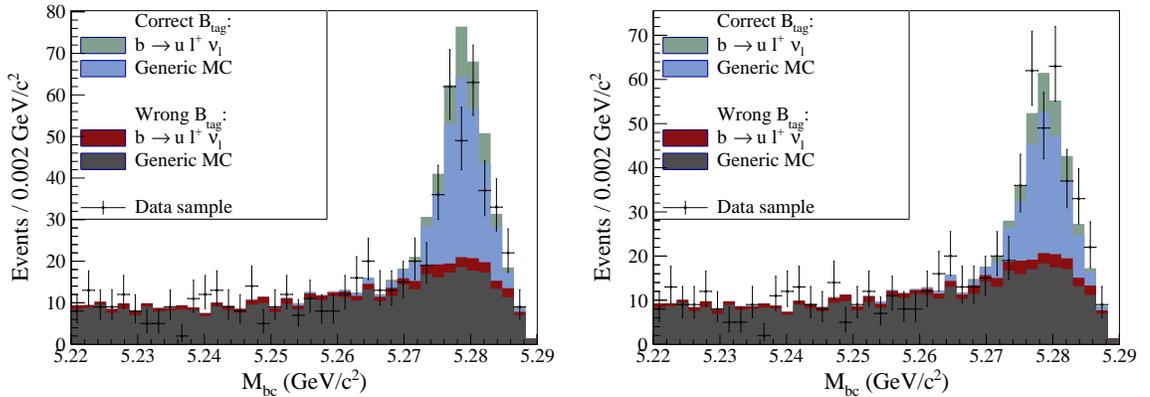
(c) Electron channel without corrective weights.

(d) Electron channel with corrective weights.

Figure 4.25: The M_{bc} mass is plotted in the sideband of $m^2_{miss} > 0.3 \text{GeV}^2/c^4$ to show the improvement in the data-MC agreement after applying the efficiency correction. The distributions show samples with the selection from section 4.2, excluding the requirement on the M_{bc} mass. For the muon channel, the excess in MC at the mass peak (a) vanishes after correction (b). The excess for the electron channel is weaker (c) and after correction a good agreement is observed (d).
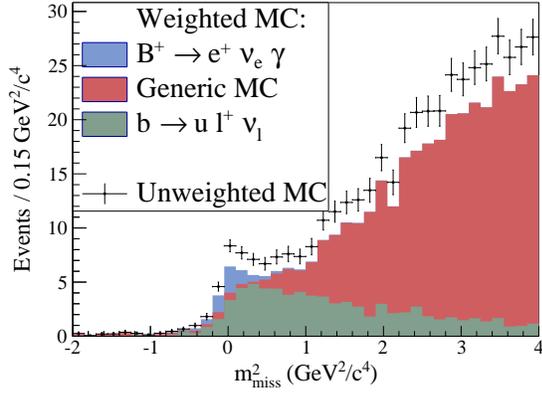
(a) Electron channel expected events.
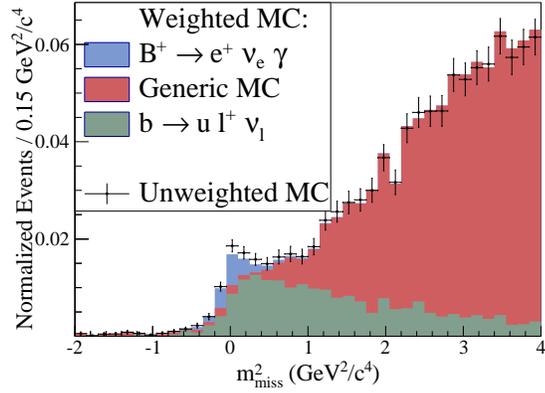
(b) Electron channel normalized.
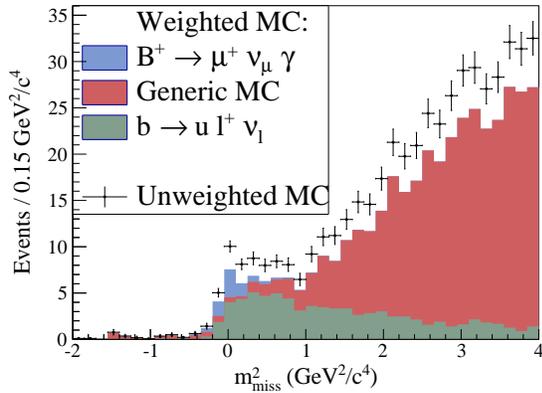
(c) Muon channel expected events.
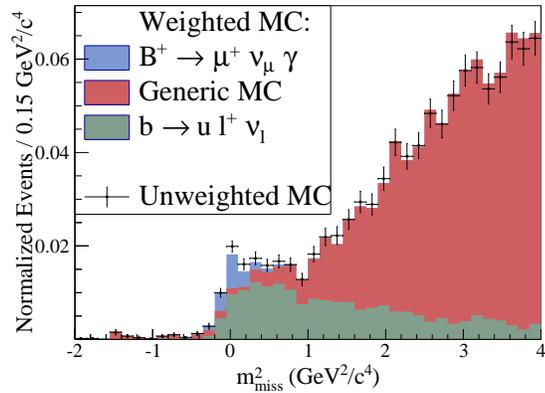
(d) Muon channel normalized.

Figure 4.26: Effect of the weighting on m^2_{miss} shown after all selections from section 4.2. The histogram shows the weighted MC samples; the unweighted MC samples are overlayed as data points. The changes in the expected luminosity on data is displayed in (a) and (c). Shape comparisons are given in (b) and (d) with normalized distributions. An overall reduction in efficiency is observed with little changes in the shape.

5   Signal extraction

The signal yield is measured in an unbinned extended maximum likelihood fit to m_{miss}^2. The fit is performed in bins of the NN output (n_{out}) defined in section 4.3, and the number of bins is optimized for signal significance. This approach is compared to an alternative setup where a cut on n_{out} is determined to maximize the signal significance in a one-dimensional fit of m_{miss}^2. It is shown, that the binned fit has a higher significance than the cut based approach. All fits are constructed with the RooFit package [33]].

The fit model consists of three components: the $B^+ \to \ell^+ \nu_\ell \gamma$ signal and two background PDFs. Each PDF is determined on MC and fixed for the fit on data. The yield of the signal and one background component are the only free parameters determined on the data sample with the yield of a second background component being fixed. To prevent a fit bias the free yields can also assume negative values.

The $b \to u \ell^+ \nu_\ell$ MC contains well known exclusive decays where the up-quark forms a meson of the type $X_u = \pi^0, \eta, \omega, \rho^0, \eta', \pi^+,$ and ρ^+. These are the dominant backgrounds, and their yields are fixed according to the latest PDG values [34] in the fit on data. The π^0 and η modes cover over 90% of the events of this contribution. This component is hereinafter referred to as the $B \to X_u \ell^+ \nu_\ell$ component.

A second background component contains the generic MC and the remaining $b \to u \ell^+ \nu_\ell$ processes not included in the fixed $B \to X_u \ell^+ \nu_\ell$ sample. The latter sample consists of decays proceeding via intermediate resonances of f_{0-2}, a_{0-2}, h_1, b_1^+, and a_2^+; and other previously accounted processes which are summarized in an inclusive model. The contribution is hereinafter referred to as "fitted background." As can be seen in Fig. 5.1, the shapes of the remaining $b \to u \ell^+ \nu_\ell$ MC and the generic MC align well enough for them to be subsumed into a single PDF.

5.1   Significance and limit determination

The significance of the signal is defined as

$$\sigma_{\mathrm{sig}} = \sqrt{-2\ln\left(\frac{\mathcal{L}_b}{\mathcal{L}_{(s+b)}}\right)}, \tag{8}$$

where $\mathcal{L}_b$ and $\mathcal{L}_{(s+b)}$ are the maximum likelihood value of the background and signal plus background hypothesis, respectively.

This uses Wilks' theorem which states that the log-likelihood ratio is proportional to a χ^2 distribution with its degrees of freedom being equal to the difference of free parameters in numerator and denominator. With this proportionality it is possible

(a) Electron channel.   (b) Muon channel.
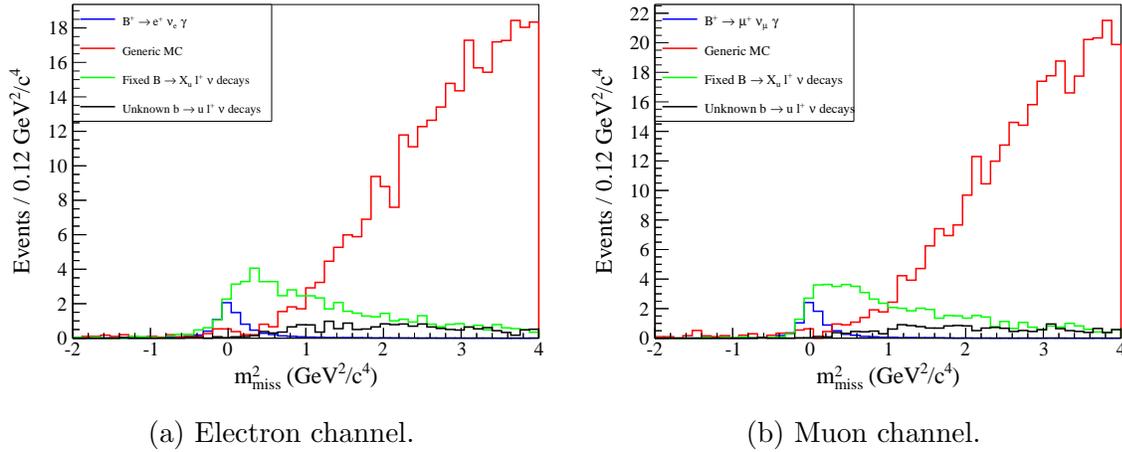
Figure 5.1: m^2_{miss} distribution where the $b \to u\ell^+\nu_\ell$ MC is split into its well known exclusive $B \to X_u\ell^+\nu_\ell$ decays and its remaining contributions. The latter component aligns well with the shape of the generic MC. Hence, in the fit they are summarized as the "fitted background" component.

to quantify the significance of the signal in multiples of the standard deviation of a Gaussian distribution.

The maximum likelihood value for null and signal hypothesis is obtained from the profile likelihood function where the fitted background yield is treated as a nuisance parameter. The likelihood for the signal hypothesis is identified with the maximum value of the likelihood function, and the likelihood for the background-only hypothesis is obtained at the position where the signal yield equals zero.

In case no significant measurement can be made, the upper limit is also determined from the profile likelihood. The one-sided confidence interval is computed by integrating the likelihood to determine the 90% quantile. Since negative signal yields are allowed in the fit, a Bayesian prior is introduced assuming the signal yield to be positive. This translates into integrating the likelihood only for positive signal yields which ensures a positive limit on the branching fraction. The bias introduced by this confidence interval construction is examined in section 5.4.2.

5.2 Inclusion of systematic errors

Systematic errors are included in the significance and the upper limit determination by convolving the likelihood function with a Gaussian whose width is equal to the systematic error. Only additive systematic errors are used in the convolution. These

affect the signal but not the background yield. The total systematic error also includes multiplicative systematic errors which affect all yields. The total error is used when determining the branching fraction since the total yield is needed for its calculation.

The systematic errors are asymmetric. The downward error is taken as the error on the significance, because it corresponds to the amount of events by which the signal is potentially overestimated. After smearing the likelihood distribution with a Gaussian, the maximum likelihood value of the signal hypothesis decreases and the likelihood for the null hypothesis increases. Therefore the difference between signal and null hypothesis reduces and with it the signal significance. The effect of the convolution on the likelihood function can be seen in Fig. 5.2a.
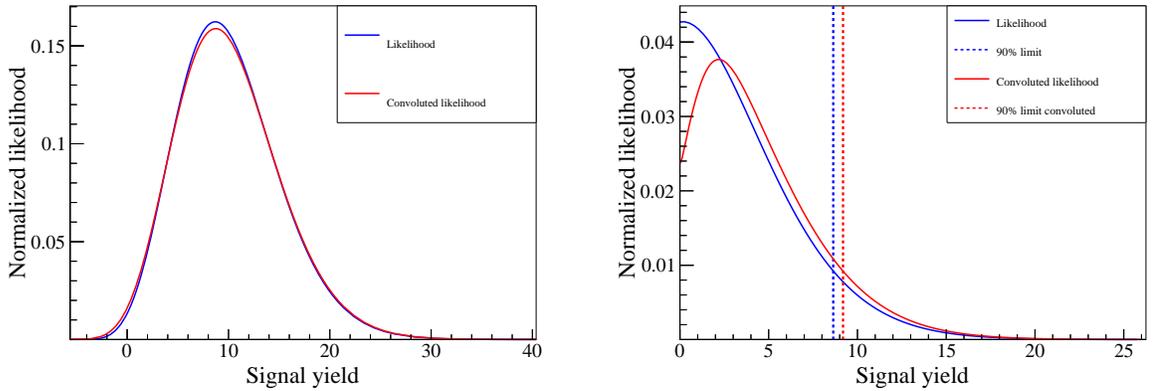
(a) The significance decrease after convolution, since the maximum likelihood decreases and the value of the null hypothesis at zero increases.

(b) The 90% quantile moves to higher values after convolution.

Figure 5.2: Likelihood functions before and after convolution for two different toy MC samples.

For the upper limit the amount of events by which the signal is underestimated is of interest. This corresponds to the upward systematic error which is used here in the convolution. After convolution, the likelihood function becomes broader and the 90% quantile moves to higher values, which can be seen in Fig. 5.2b. In the limit determination, only the region for positive signal yields is convoluted. This leads to a distortion with the curve bending downwards at values towards zero, as seen in the plot. The effect could be avoided if the likelihood function would be convoluted in the whole range including negative values. However, this leads in rare cases to a reduction of the upper limit after convolution when the fit has a very negative

maximum likelihood value with only a small tail of the distribution reaching into the region above zero. In these cases, parts of the distribution shift from negative values to values just above zero which results in a steeper function in the positive region. Therefore only the positive region of the likelihood is convoluted.

5.3 Fit shapes

The significance of the signal is maximal for the fit of the m^2_{miss} in six bins of n_{out}, as determined section 5.1. Therefore the PDFs for all components are shown in six bins of n_{out} and the same function is used for all bins.

The correlation between m^2_{miss} and n_{out} is checked with a tool [29] described in section 3.6. Significant correlations are found for the signal and $b \rightarrow u\ell^+\nu_\ell$ MC, and a lower correlation for the generic MC. It is concluded that the parameters of the functions must be determined in each n_{out} bin separately since significant shape differences among the bins are to be expected.

The fit shapes for the secondary analysis are given in appendix A.4.

Signal PDF

The signal has an asymmetric tail to positive values in m^2_{miss} and the width of the peak increases for more background-like network outputs. A Crystal Ball function together with a Gaussian is used to describe the signal shape, where the former describes the tail to positive mass values and the latter accounts for the smaller tail on the left-hand side of the peak. The Crystal Ball function [35] is a modified Gaussian function which is given by

$$
f(x; \alpha, n, \bar{x}, \sigma) = N \cdot \begin{cases} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right), & \text{for } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot \left(B - \frac{x-\bar{x}}{\sigma}\right)^{-n}, & \text{for } \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases}
$$

$$
A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right),
$$

$$
B = \frac{n}{|\alpha|} - |\alpha|.
$$

Leaving all parameters free over-parameterizes the fit. This leads to large correlations in the error matrix and thus to a problematic evaluation of the systematic errors originating from fit shape uncertainties. This is prevented by fixing the parameters n and α of the Crystal Ball function and by fitting the Gaussian and Crystal Ball function with a common mean. The PDFs are shown in Fig. 5.3.

61

Fitted background PDF

The fitted background component is modeled with a modified exponential function having a second order polynomial in the exponent. The function is given by

$$f(x; x_0, \alpha, \beta) = e^{\alpha(x-x_0)^2 + \beta(x-x_0)}. \tag{9}$$

The fit results are shown in Fig. 5.4. The shapes do not change much among the bins since the correlation between m_{miss}^2 and n_{out} is small for this fit component.

Fixed $B \to X_u \ell^+ \nu_\ell$ PDF

No analytical function could be found to describe the fixed $B \to X_u \ell^+ \nu_\ell$ contribution in all bins of n_{out}. Therefore a Gaussian kernel estimator algorithm is used where each data point is approximated by a Gaussian PDF [36]. This procedure gives a smooth shape for badly parameterizable distributions. The algorithm is provided with a parameter quantifying the width of the Gaussian kernels, where the shape follows the MC sample more closely if the kernel width become smaller.

The shapes of the high statistics samples of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$ are fitted separately from the other $B \to X_u \ell^+ \nu_\ell$ decays with lower statistics. The kernel width is chosen to be smaller for the high statistics samples so that the algorithm follows the MC more closely. The two shapes are combined to obtain the $B \to X_u \ell^+ \nu_\ell$ sum PDF which is shown in Fig. 5.5. The separate shapes are displayed in appendix A.1.

5.4   Model of the binned fit

The extended likelihood function for the binned fit is given by

$$\ln \mathcal{L} = \sum_{j=1}^{N_{\text{tot}}} \ln\left\{ \sum_i^{N_c} N_i \mathcal{P}_i(m_{\text{miss}}^2, n_{\text{out}}) \right\} - \sum_i^{N_c} N_i,$$

where N_{tot} is the total number of events in the data set, N_c denotes the number of components in the fit, N_i is the number of events for the ith component, and $\mathcal{P}_i$ represents the PDF for that component as a function of m_{miss}^2 and n_{out}. The determination of the PDFs for the three fit components is described in the previous section.

The shapes as well as the relative normalizations among the bins are fixed for the fit on data, and only the total yield is a free parameter. The PDF for the ith fit

component in the bin of n_{out} is given by

$$\mathcal{P}_i(m_{\mathrm{miss}}^2, n_{\mathrm{out}}) = f_i^{n_{\mathrm{out}}} \mathcal{P}_i^{n_{\mathrm{out}}}(m_{\mathrm{miss}}^2),$$

where $f_i^{n_{\mathrm{out}}}$ denotes the fixed fraction of N_i events in the bin and $\mathcal{P}_i^{n_{\mathrm{out}}}$ is the PDF in that bin.

By design, each bin contains the same number of expected signal events and the number of n_{out} bins is chosen to maximize the expected significance of the signal. The bin boundaries are shown in Fig. 4.20. The two signal channels $B^+ \to e^+ \nu_e \gamma$ and $B^+ \to \mu^+ \nu_\mu \gamma$ are measured in separate fits. Additionally, a simultaneous fit to both channels is performed to measure the $B^+ \to \ell^+ \nu_\ell \gamma$ branching fraction. Lepton universality is assumed for the latter measurement, where the signal branching fractions of the two channels are fixed to the same value. The fits to the separate channels have two yields as free parameters and the combined fit model has three free yields. All yields are unconstrained which means that negative values are allowed in the fit. The models for the fit in six bins are shown in Fig. 5.6.

5.4.1 Significance optimization

The expected signal significance is determined from an average over many toy MC experiments. The toy samples are generated from PDFs which are obtained by fitting MC. The significance is measured for bin counts ranging from one to eight to find the optimal number of n_{out} bins. For each bin count, the functions described in section 5.3 are refitted to obtain the PDFs for the toy MC study.

A plot showing the fit significance against the number of n_{out} bins is given in Fig. 5.7 for the nominal analysis and in Fig. 5.8 for the secondary analysis. The size of the error bars is determined by 1000 toy experiments performed for each data point. The significance makes a jump when increasing the number of bins from one to two, and after that a slower increase in significance is found, where a saturation is observed for larger bin numbers. According to these plots, it is decided that both signal channels are fitted in six bins of n_{out} for the nominal and secondary analysis.

The significances for the separate and simultaneous fit in six bins are summarized in table 5.1 together with the results for the upper limits. The expected upper limit is determined in toy MC studies generated without any signal. The table also shows values in parentheses which contain the systematic errors determined in section 7. The measurement is dominated by the statistical error and the results do not deteriorate significantly with the inclusion of systematic errors.

5.4.2   Bias check

A toy study is performed to check if the fit in six bins of n_{out} reproduces the yields used for toy generation. This is done in extended pure toy MC studies which generate toy samples from the PDF shapes and fluctuate the input yields of the components according to a Poisson distribution. Each toy sample is fitted with the binned fit model from section 5.4 in an extended maximum likelihood fit to extract the signal and fitted background yield. For toy MC generated with a signal branching fraction of 5×10^{-6}, the results are shown in Fig. 5.9, 5.10, and 5.11 for the simultaneous fit, the electron channel, and the muon channel, respectively.

The plots show: the measured yield; the error of the yield, which is obtained from the points where the negative log likelihood function increases by $1/2$ of its minimum value; and the pull distribution, which subtracts the measured yield from the true yield, divided by the measured error. Since the statistical error is asymmetric, the pull value is divided by the negative error if the mean is negative and vice versa. If the parameter and its error are estimated correctly from the likelihood function, the pull distribution has the shape of a standard normal distribution with a mean of zero and a standard deviation of one. To obtain a quantitative measure for the possible fit bias, the pull distribution is fitted with a Gaussian to extract the mean and the standard deviation.

In conclusion, the average fitted signal yields of the binned fit reproduce the input signal yield used for toy generation, and no relevant deviation is found for a signal branching fraction of 5×10^{-6}. The pull distributions for the separate channels show asymmetric features where the means are a little below zero with widths larger than one. The asymmetry shows only in the negative half of the pull distribution where the negative fit error of the parameter is used in the division. It seems that this downward error is underestimated, which causes negative pull values to deviate to larger values.

Furthermore, the estimates for the background yields show no bias for any of the fits. The toy MC results for the fitted background yields of the simultaneous fit can be found in appendix A.3. No fit bias is found for the secondary analysis whose results are shown in appendix A.4.

Linearity test

A linearity test is performed, where the dependence between the input signal branching fraction and the fit bias is examined. Here, pure toy MC studies are performed for different signal branching fractions. For each toy study the mean and the standard deviation is obtained from a fit to the pull distribution. The results are shown in

Fig. 5.12, where for each data point 1000 toy experiments are performed. For unbiased fits the mean and the standard deviation agree with the horizontal lines drawn in the plot.

At low signal branching fractions an increasing bias is observed for all fit models where the pull mean deviates to lower values and the standard deviation to higher values. The measurement underestimates the true signal yield for low branching fractions which lowers the signal significance, since it shifts the likelihood function to lower values thus reducing the value of the likelihood ratio of signal to background hypothesis. The downward bias of the fit is not taken into account in the systematic error of the significance, because only upward errors, which lead to an increase in significance, are of interest for this quantity, as described in section 5.2. The observed bias is small compared to the statistical error, and it leads to a conservative estimate of the signal yield for low branching fractions.

The bias hints to problems in the fit model for low event numbers in the signal region. For background depleted n_{out} bins which experience an additional downward fluctuation of the background, a negatively measured signal yield can cause the overall PDF of the fit model to become negative in the signal region, which is shown in Fig. 5.14. This leads to an undefined behavior in the likelihood function and the fitting package RooFit interprets negative PDF values as zero. This entails, that decreasing the signal component in the signal-like n_{out} bins has a smaller effect on the likelihood function. Still, the signal yield cannot assume arbitrarily negative values since a decrease has an effect on the likelihood value due to the background-like bins in which the overall PDF stays positive. For the toy MC study without signal, approximately half of the toy experiments have a PDF which assumes negative values of varying extent in the signal region of the signal-like n_{out} bins. This behavior is also observed for the electron channel and the simultaneous fit.

This difficulty is only present for low yields and therefore the bias disappears for larger signal branching fractions. The secondary analysis has more events in the signal region of the signal-like n_{out} bins and accordingly a lower bias is observed in the linearity test in Fig. 5.13.

Upper limit

The confidence interval for the upper limit is determined by the 90% quantile of the signal likelihood function where only positive signal yields are taken into account. This construct excludes unphysical values for the branching fraction but it also leads to an over-coverage of the confidence interval.

This is confirmed in a toy MC study in which the percentage of confidence intervals

containing the true value is determined. The results of the check in Fig. 5.15 plot the fraction of intervals not containing the true value against the input branching fraction used for toy generation. For unbiased results, 10% of the true values should be outside the 90% confidence interval which is indicated by a horizontal line. For low branching fractions the confidence intervals are too large, containing the true value in all of the cases. With increasing branching fractions the bias decreases since the likelihood function moves into the positive region thus being integrated more completely. However, this happens for branching fractions for which a significant measurement is expected.

This bias counteracts the bias described in the previous section which also appears for low branching fractions. The bias introduced by this Bayesian construction is stronger which results in a harmless situation where the upper limit is overestimated and thus conservative. For that reason, the bias is not included as a systematic error. As described in section 5.2, only downward errors are taken into account in the upper limit.

The same consideration apply to the secondary analysis, where the results shown in Fig. 5.16 display a similar over-coverage of the confidence interval.

MC studies sampled from MC

As of now, pure toy MC studies have been used to identify fit bias. To check if the PDFs describe the data properly, toy MC is generated directly from the MC samples from which events are randomly selected to obtain samples representing data. The number of toys which can be generated in this way is statistically constrained by the luminosity of the MC. The generic $B\bar{B}$ and the $b \to u\ell^+\nu_\ell$ MC only contain ten times the events of data, where the remaining high luminosity samples for the signal and the dominant backgrounds contain about 100 times the amount of data. A hybrid approach is taken to generate 100 toy experiments. The high luminosity MC is sampled and PDFs are used for the fitted background component to generate the toy MC.

The results are shown in appendix A.5 for both the nominal and secondary analysis. No significant bias is found for the binned fits with a signal branching fraction of 5×10^{-6}. The signal yield is found to be slightly underestimated for the nominal analysis in the fit of the separate channels.

5.5 Significance of a cut-based approach

As a cross check to the binned fit, a cut-based approach is implemented. The significance is optimized by incrementally increasing the cut value on n_{out} with a toy

MC study being carried out for each incremental selection. The functions given in section 5.3 are refitted for each cut to obtain the PDFs.

The result of this procedure is shown in Fig. 5.17 and 5.18 for the nominal and secondary analysis, respectively. An improvement in significance can be observed when applying a selection on the network. A suitable threshold is found with a value of 0.25 for both channels of the nominal and secondary analysis. The significances obtained with this procedure are smaller than for the binned fit procedure and therefore the cut-based procedure is not used for the measurement on data. Differences between systematic errors of the binned fit and the cut-based approach do not change the picture. The significances of the binned fit which include systematic errors are of the same size as the significances of the cut-based approach without systematic errors.

The best significances for the cut-based approach are listed in table 5.1. No fit bias is observed for this procedure which is confirmed in appendix A.2 for toy MC studies in which the n_out cut maximizes the significance.

Table 5.1: Expected signal yields and significances which are obtained from toy studies for $\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) = 5 \times 10^{-6}$. Upper limits are obtained from toy studies without signal. The values in parentheses contain systematic errors, and for the yield the first error is statistical and the second error systematic. For comparison, the lower significances of the cut-based approach are also listed.

Nominal analysis with $E_\gamma^\text{sig} > 1\,\text{GeV}$				
Mode	Yield	Significance (σ)	$\mathcal{B}$ limit (10^{-6})	Cut-based significance (σ)
$B^+ \to e^+\nu_e\gamma$	$8.0 \pm 4.5\,^{+1.0}_{-1.3}$	2.4 (2.1)	$< 7.0\,(< 7.5)$	2.0
$B^+ \to \mu^+\nu_\mu\gamma$	$8.7 \pm 4.6\,^{+1.0}_{-1.5}$	2.5 (2.2)	$< 6.5\,(< 6.9)$	2.1
$B^+ \to \ell^+\nu_\ell\gamma$	$16.5 \pm 6.5\,^{+1.6}_{-2.2}$	3.6 (2.9)	$< 4.4\,(< 4.8)$	-

Secondary analysis with $E_\gamma^\text{sig} > 400\,\text{MeV}$				
Mode	Yield	Significance (σ)	$\mathcal{B}$ limit (10^{-6})	Cut-based significance (σ)
$B^+ \to e^+\nu_e\gamma$	$12.4 \pm 6.2\,^{+1.8}_{-2.3}$	2.4 (2.1)	$< 6.5\,(< 6.8)$	2.1
$B^+ \to \mu^+\nu_\mu\gamma$	$11.9 \pm 6.0\,^{+1.7}_{-2.1}$	2.5 (2.2)	$< 6.0\,(< 6.2)$	2.3
$B^+ \to \ell^+\nu_\ell\gamma$	$24.9 \pm 8.7\,^{+3.0}_{-3.5}$	3.4 (2.9)	$< 4.1\,(< 4.3)$	-
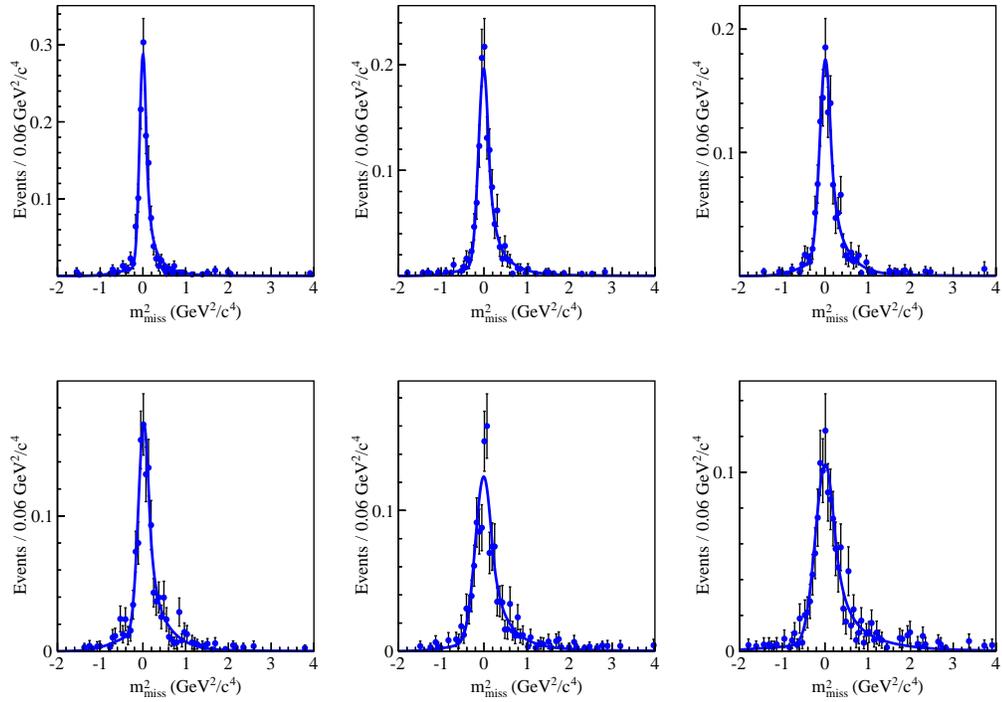
(a) Electron channel.
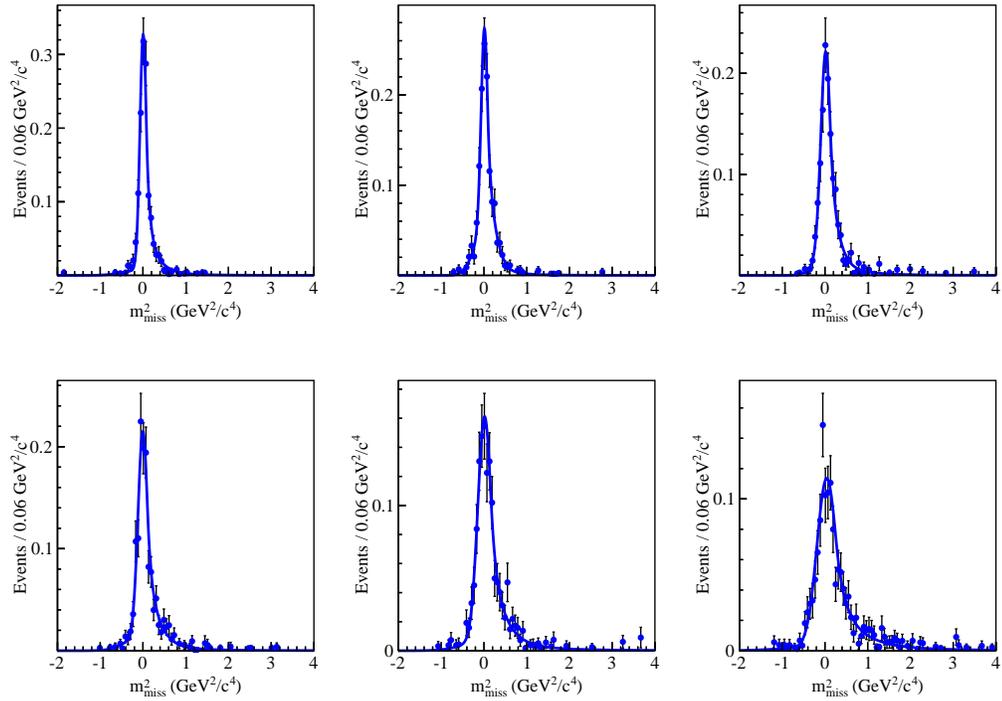
(b) Muon channel.

Figure 5.3: Signal m^2_{miss} PDF fitted in six bins of n_{out}. The data points show the signal MC.
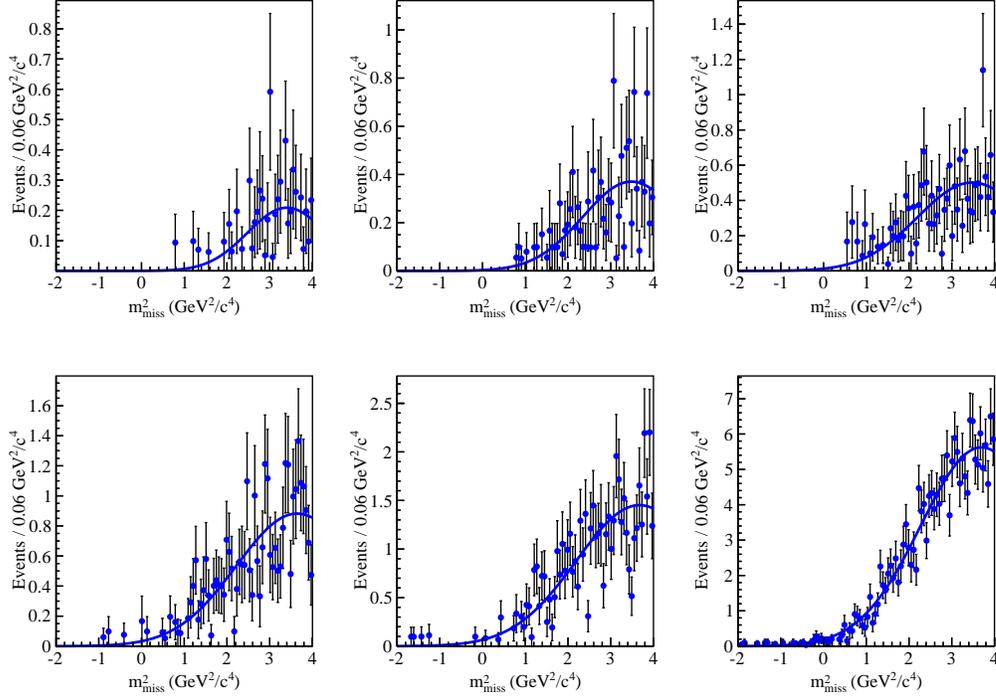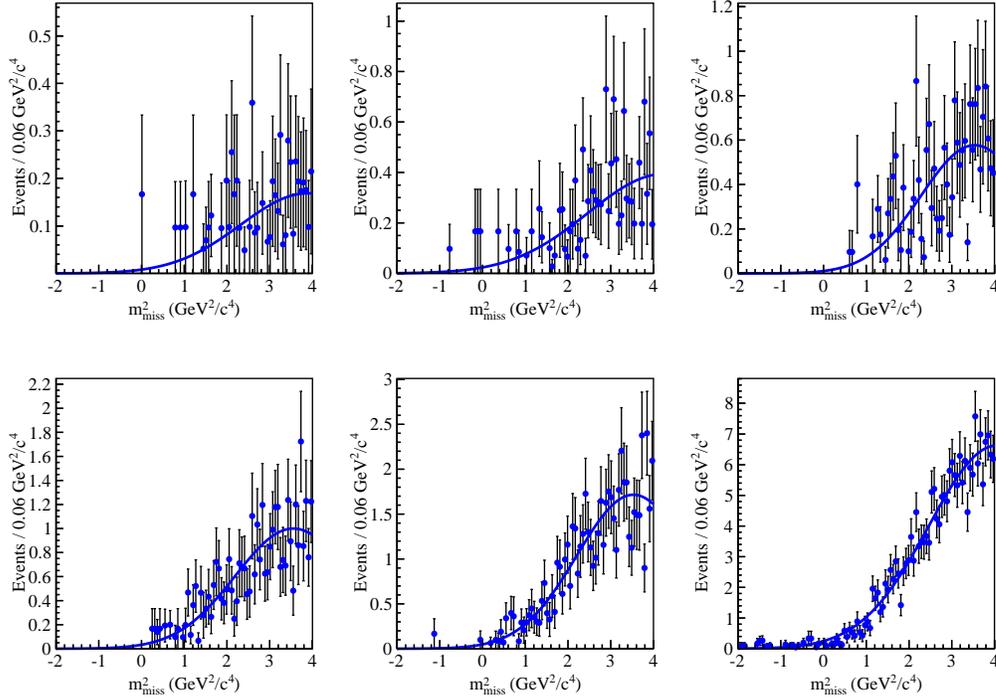
(a) Electron channel.

(b) Muon channel.

Figure 5.4: Fitted background m^2_{miss} PDF determined in six bins of n_{out}. The data points show the generic MC together with badly known decays of the $b \to u\ell^+\nu_\ell$ MC.
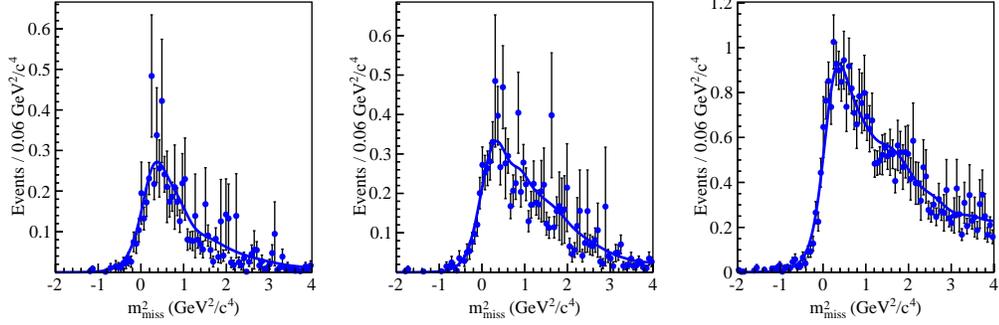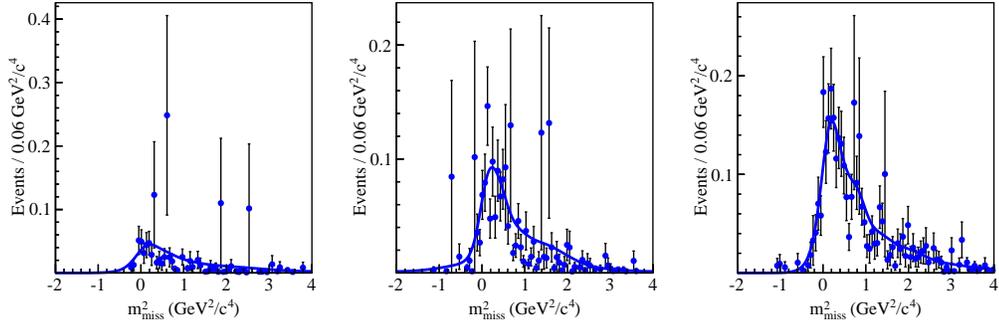
(a) Electron channel.
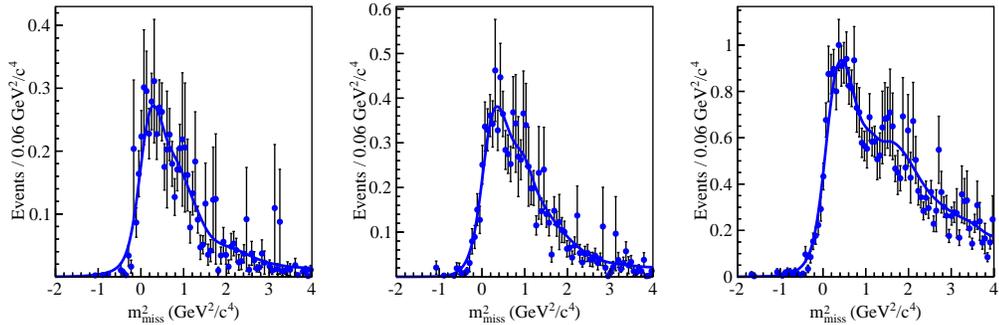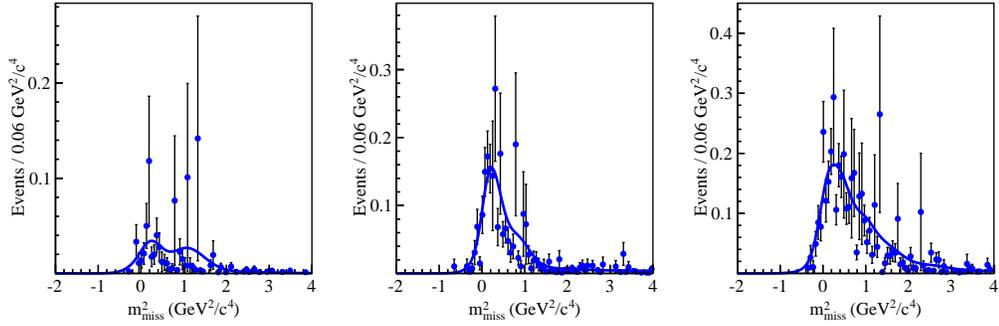
(b) Muon channel.

Figure 5.5: Combined $B \to X_u \ell^+ \nu_\ell$ PDF in m^2_{miss} determined in six bins of n_{out}. The data points show well known decays of the $b \to u \ell^+ \nu_\ell$ MC together with the high luminosity samples of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$. The differing luminosities lead to non-uniform error bars.
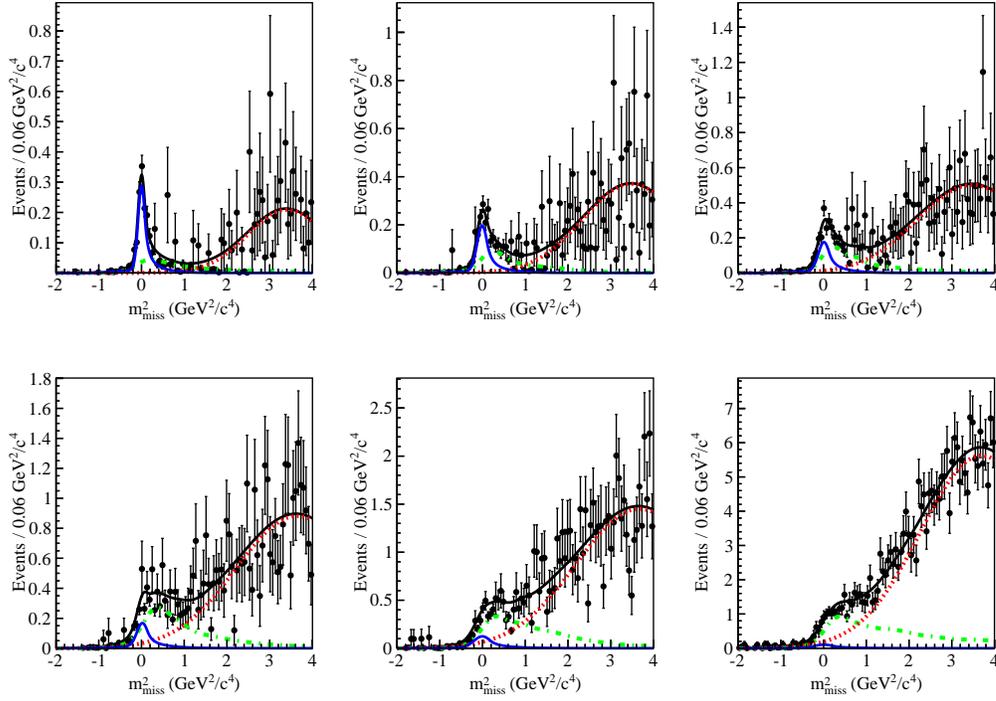
(a) Electron channel.
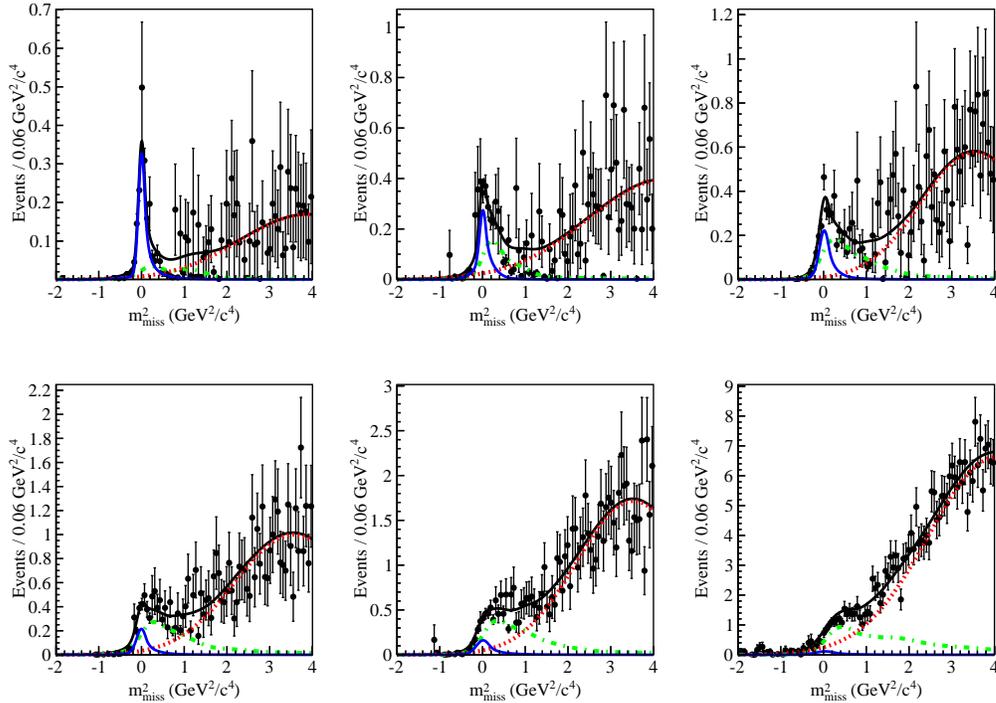
(b) Muon channel.

Figure 5.6: Fit model of m^2_{miss} in six bins of n_{out} in MC (points with error bars), where the MC samples are weighted to the data luminosity. Accordingly, the error bars on the data points represent the luminosity of the MC samples and not the expected error on data. Shown are the PDFs for: the signal (blue), fixed $B \to X_u \ell^+ \nu_\ell$ backgrounds (green), fitted backgrounds (red), and the sum (black). The signal is normalized to a branching fraction of 5×10^{-6}. The most signal-like bin is found in the upper left plot. Proceeding from left to right, the distributions become increasingly background-like and the most background-like bin is shown in the lower right panel.
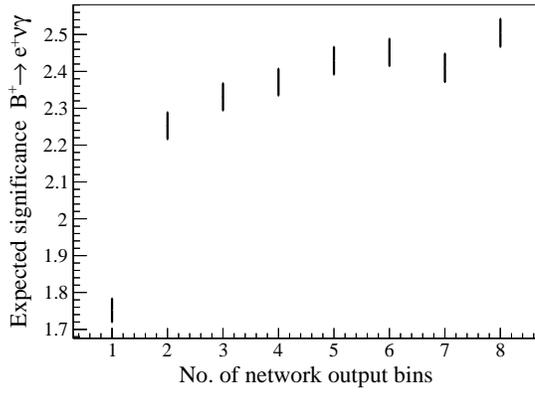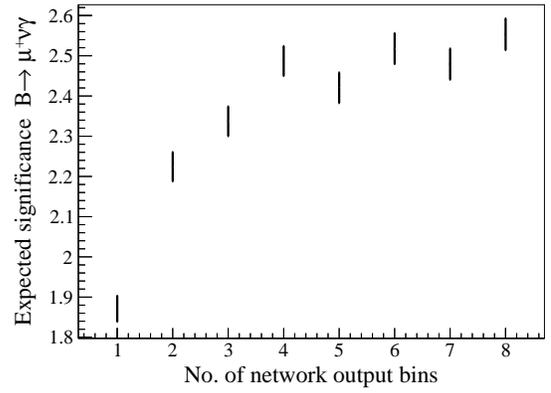
(a) Electron channel.

(b) Muon channel.

Figure 5.7: Signal significance for fits performed with different numbers of n_{out} bins for the nominal analysis.
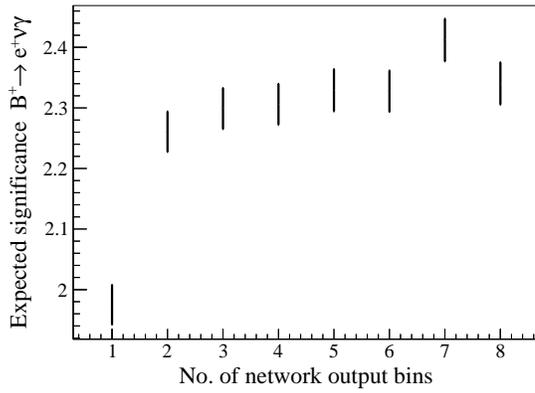
(a) Electron channel.
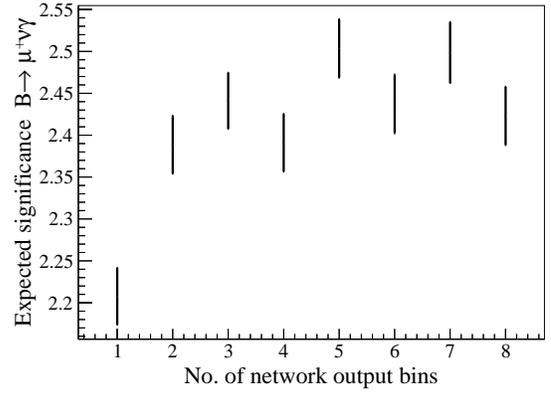
(b) Muon channel.

Figure 5.8: Signal significance for fits performed with different numbers of n_{out} bins for the secondary analysis.
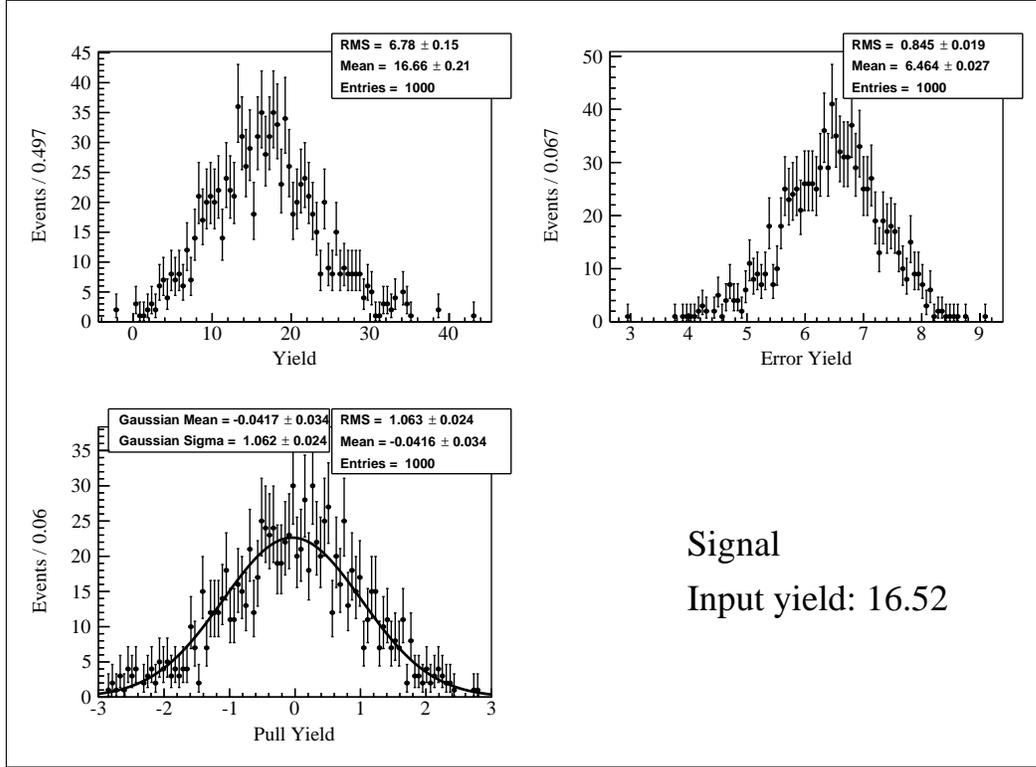
Figure 5.9: Toy study result for the joint signal yield of the simultaneous fit to both signal channels performed with a branching fraction of 5×10^{-6}. The plot shows: the measured yields in the upper left panel, the statistical errors on the measured yields in the upper right panel, the pull distribution computed from the difference between measured and true yield divided by the statistical error in the lower left panel, and true yield used for toy generation in the lower right panel. A Gaussian is fitted to the pull distribution to determine its deviation from a unit Gaussian, where mean and standard deviation of the fit are displayed in the upper left corner of the panel. Each histogram has its number of entries, the mean, and the root mean square displayed in the upper right corner of the panel.
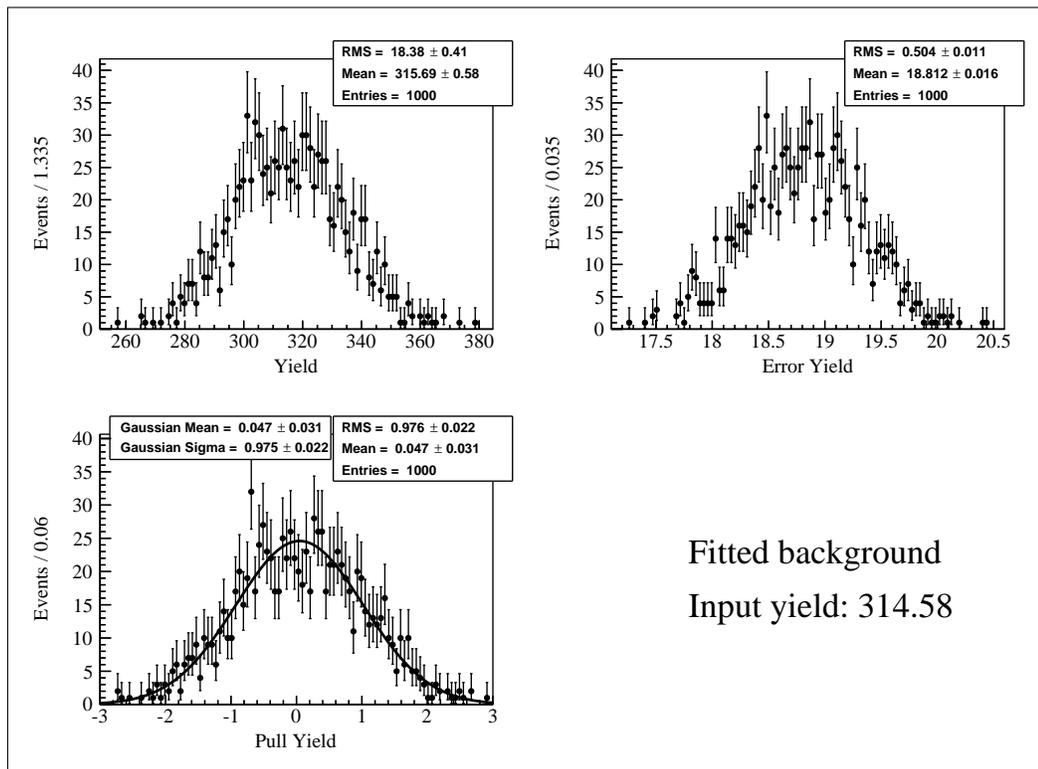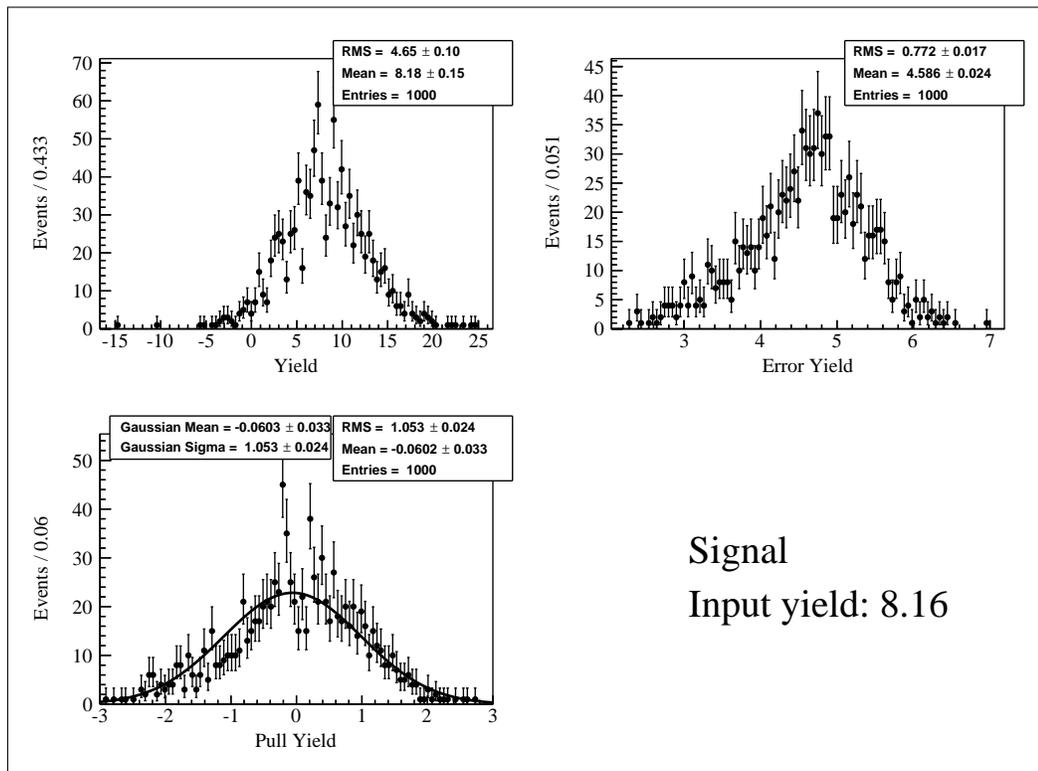
Figure 5.10: Toy study result for the electron channel with a signal branching fraction of 5×10^{-6}. See Fig. 5.9 for a description of the plot.
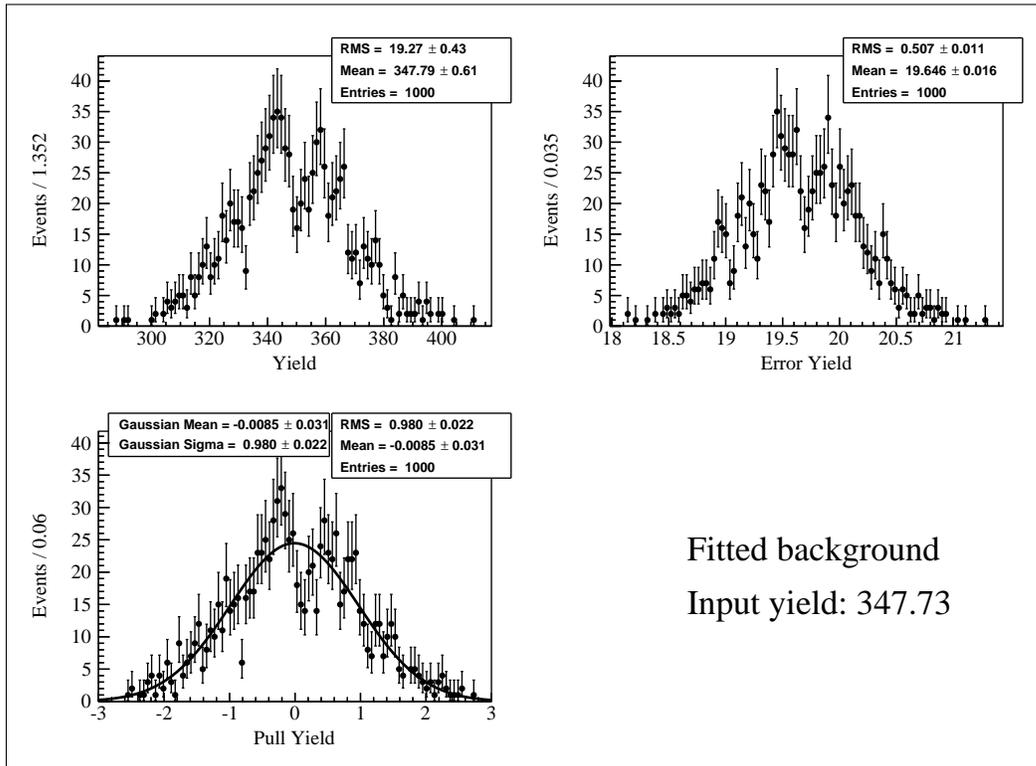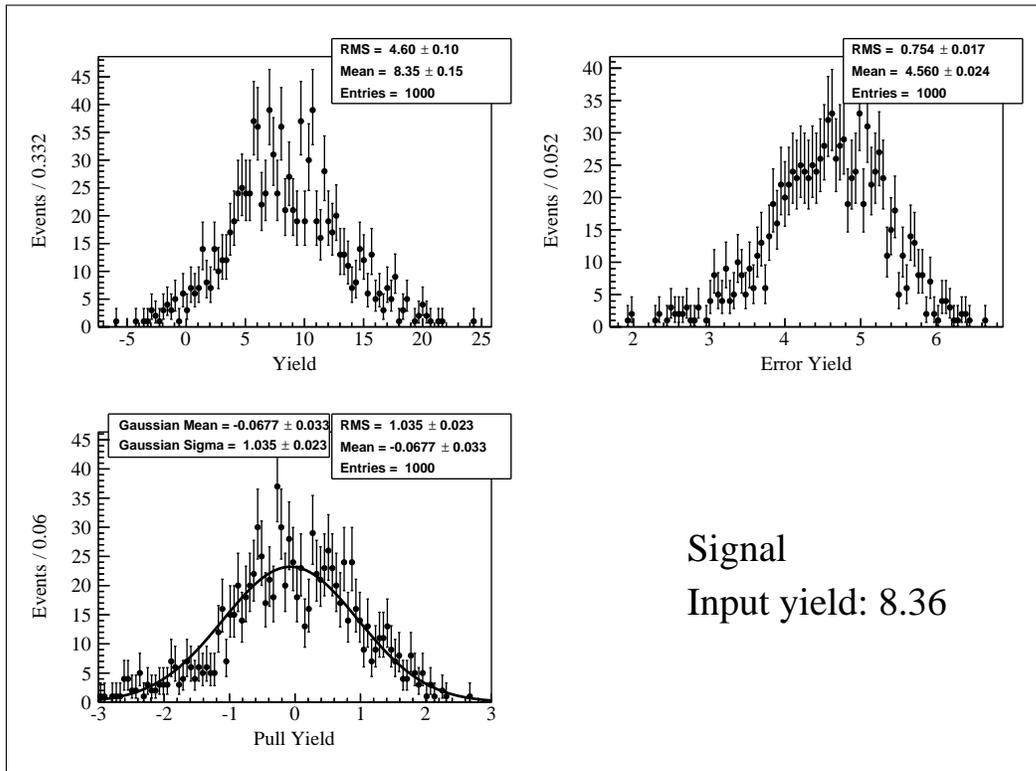
Figure 5.11: Toy study result for the muon channel with a signal branching fraction of 5×10^{-6}. See Fig. 5.9 for a description of the plot.
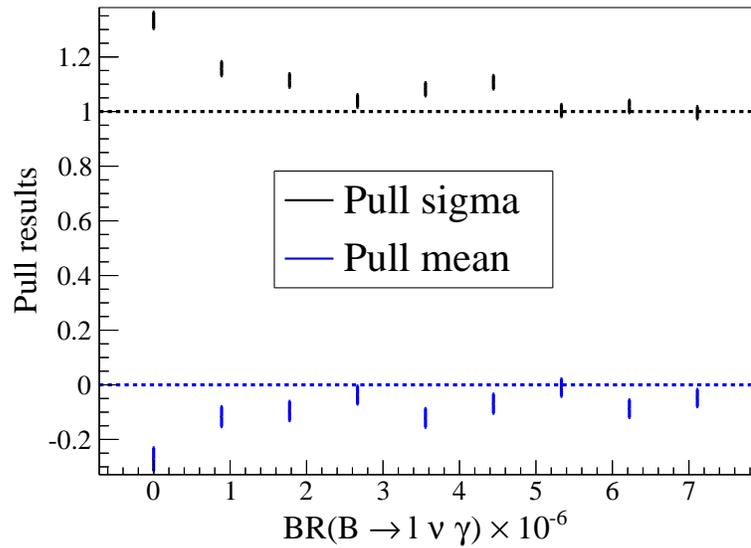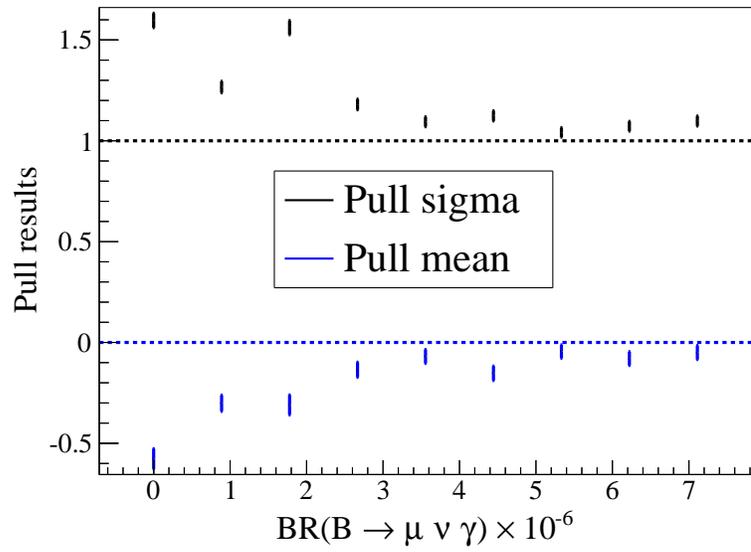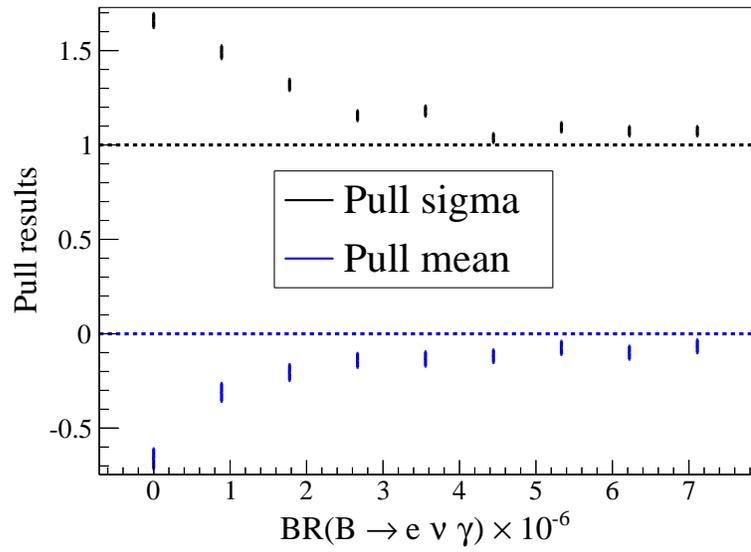
Figure 5.12: Linearity test for the nominal analysis showing the mean and width of the pull distribution for different signal branching fractions. The upper series shows the width and the mean of the pull distribution. Unbiased results are in agreement with the respective horizontal lines.
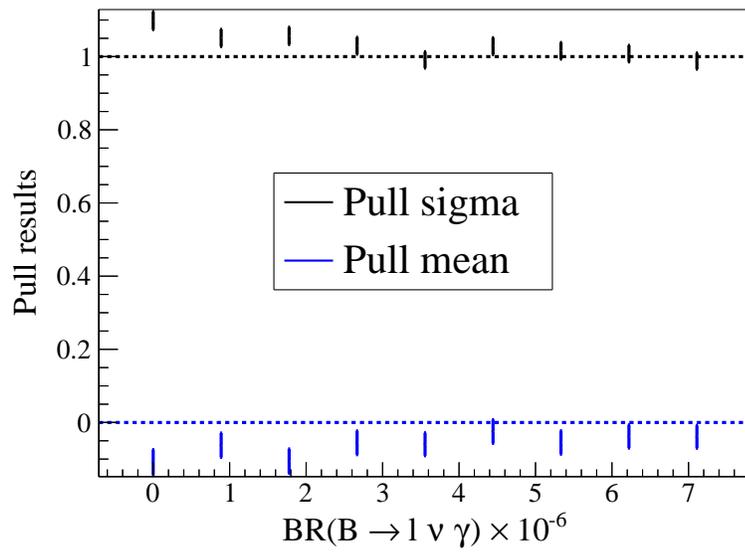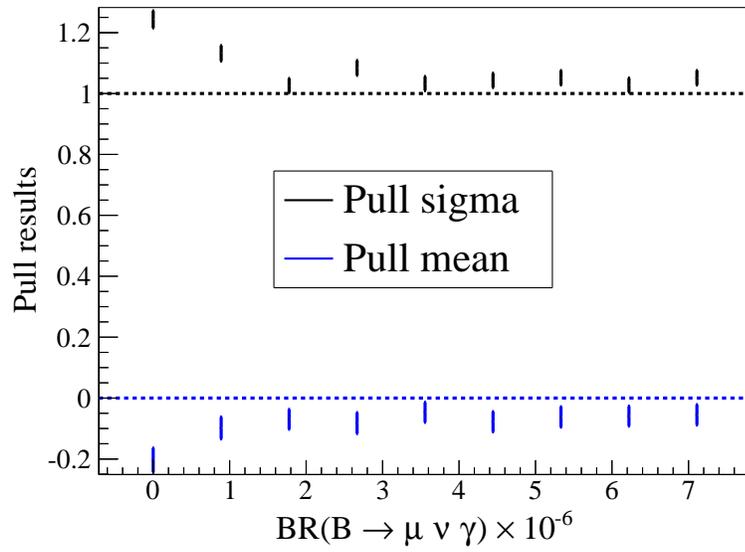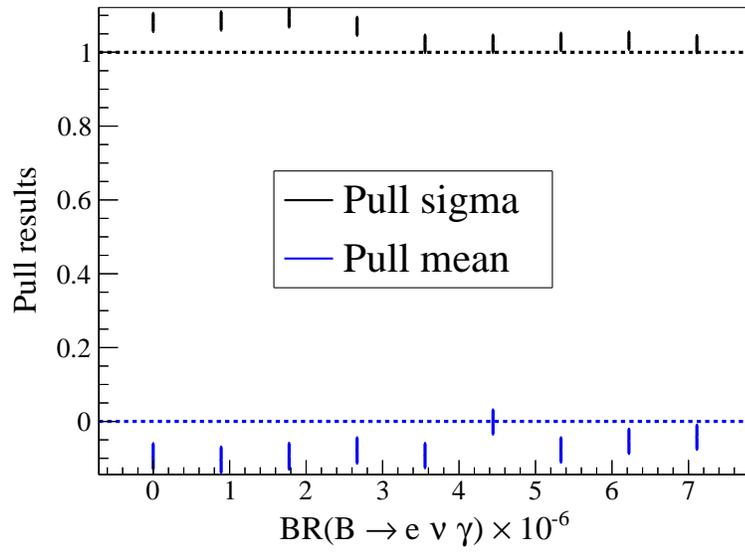
Figure 5.13: Linearity test for the secondary analysis. See description of Fig. 5.12.
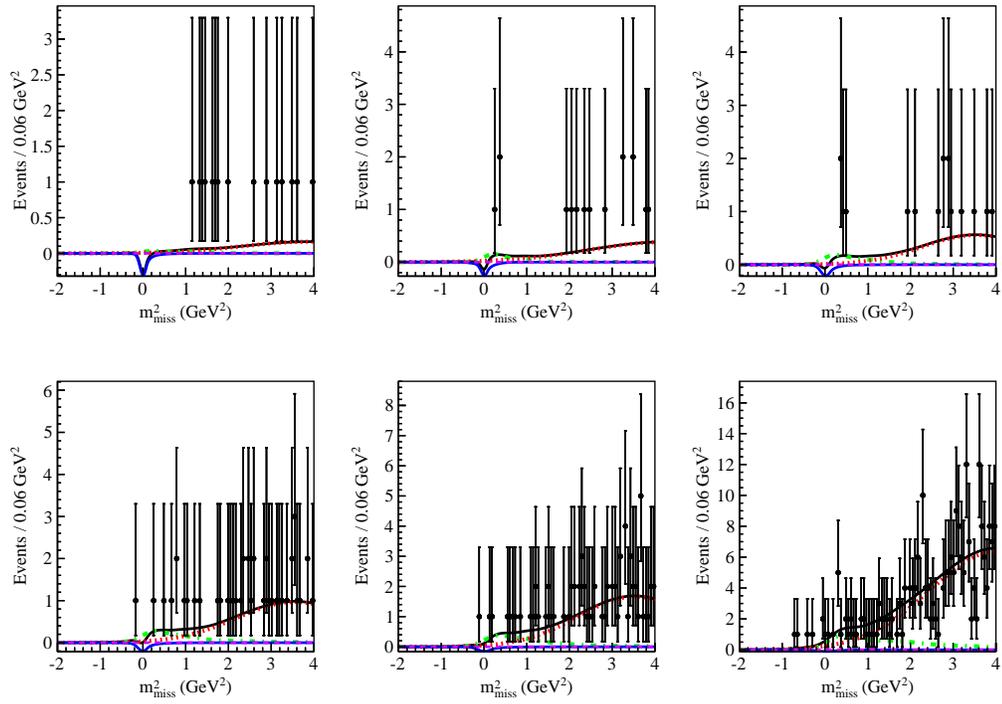
Figure 5.14: Fit of a toy sample generated without signal. A negative signal yield is fitted which is large enough to cause the total PDF to be negative in the signal-like n_{out} bins.
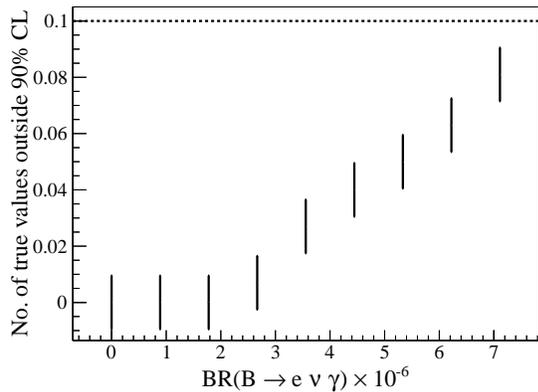
(a) Electron channel.
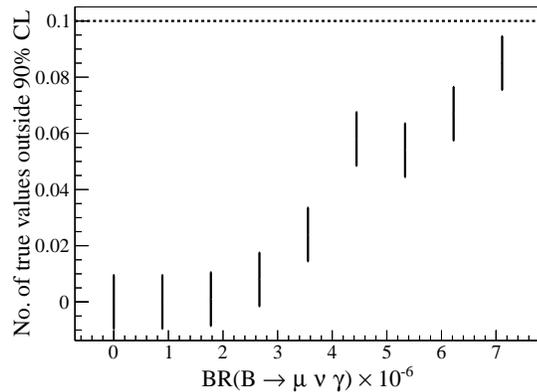
(b) Muon channel.
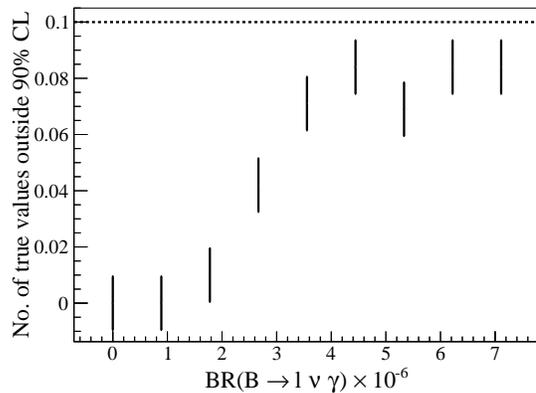
(c) Simultaneous fit to both channels.

Figure 5.15: Confidence interval coverage test in toy MC studies for the nominal analysis. The plots show the fraction of measurements for which the true signal yield is outside the 90% confidence interval for different branching fractions. The horizontal line indicates unbiased values.
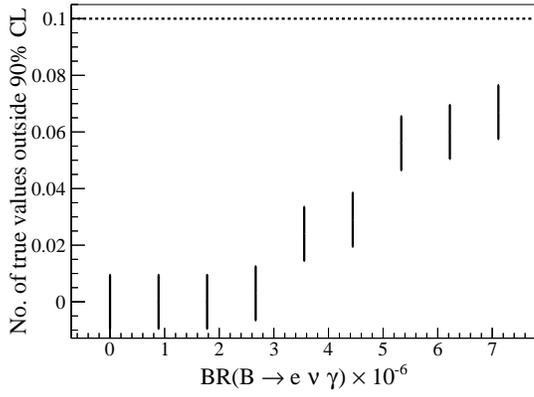
(a) Electron channel.
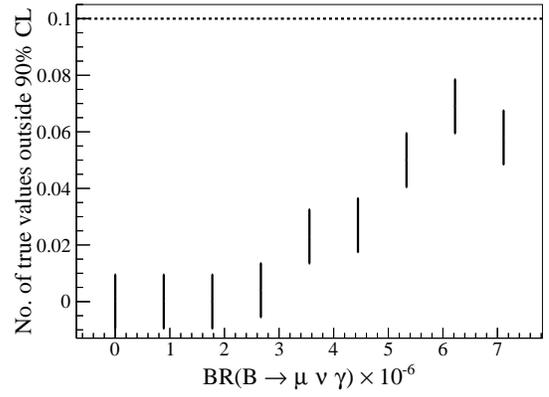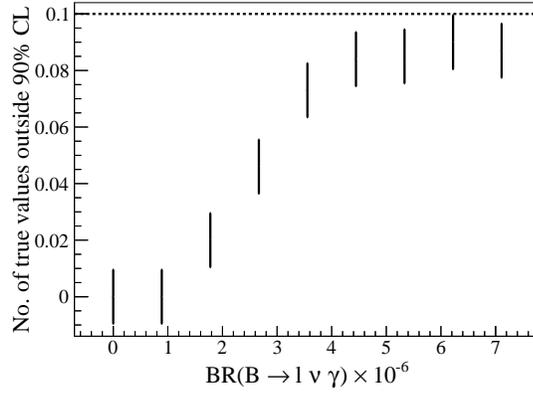
(b) Muon channel.

(c) Simultaneous fit to both channels.

Figure 5.16: Confidence interval coverage test in toy MC studies for the secondary analysis. See description of Fig. 5.15.
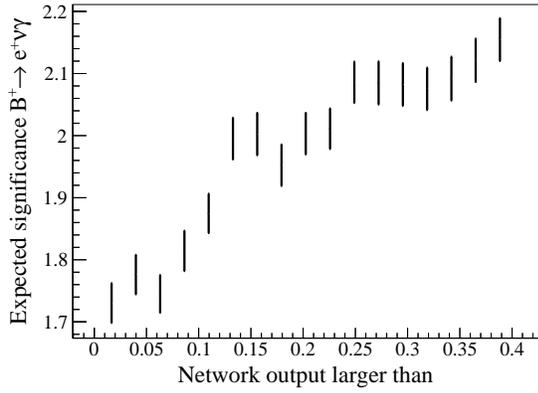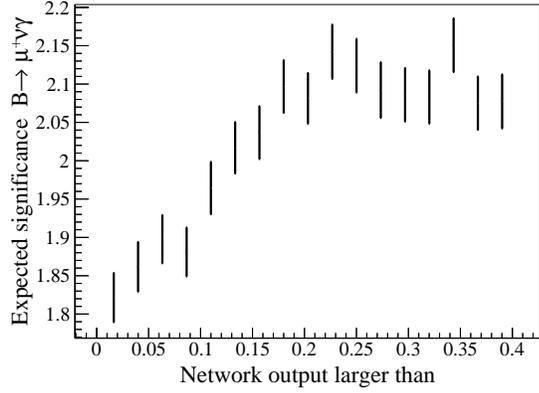
(a) Electron channel.   (b) Muon channel.

Figure 5.17: Signal significance of the cut-based procedure for incrementally increasing cuts on n_{out} for the nominal analysis.
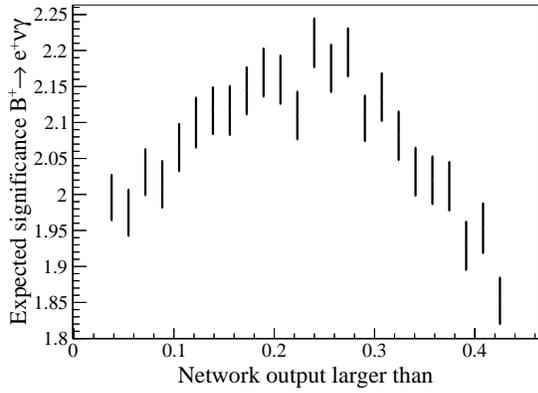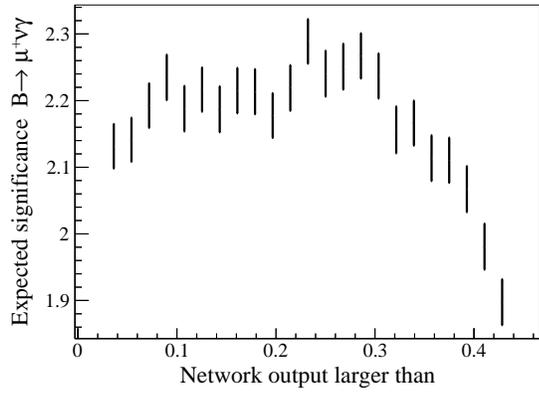
(a) Electron channel.   (b) Muon channel.

Figure 5.18: Signal significance of the cut-based procedure for incrementally increasing cuts on n_{out} for the secondary analysis.

6 Sideband consistency check

6.1 Sideband distributions

The data and MC distributions of several variables are compared in sideband distributions for the nominal analysis to look for deviations between the samples. The variables m^2_{miss} and n_{out} used in the binned fit in section 5.4 as well as the input variables used in the network training in section 4.3 are shown.

The signal region is defined by $m^2_{\text{miss}} < 0.3$ GeV$^2/c^4$, $n_{\text{out}} > 0.25$, and $M_{\text{bc}} > 5.27$ GeV$/c^2$. The m^2_{miss} variable is plotted in the sidebands $M_{\text{bc}} > 5.27$ GeV$/c^2$ and $n_{\text{out}} < 0.25$; and also $M_{\text{bc}} < 5.27$ GeV$/c^2$ and $n_{\text{out}} > 0.25$. The sidebands are chosen to resemble the background composition of the signal region.

The remaining variables are shown in two other sidebands which are defined as $m^2_{\text{miss}} \in (1.0, 4.0)$ GeV$^2/c^4$ and $m^2_{\text{miss}} \in (0.3, 1.0)$ GeV$^2/c^4$, where both selections also satisfy the condition $M_{\text{bc}} > 5.27$ GeV$/c^2$. These sidebands check the distributions for the two background fit components defined in section 5.3. The first sideband, which is dominated by $B\bar{B}$ events from the generic MC, tests the consistency of the fitted background component (fitted background SB). The second sideband is enriched with events from the fixed $B \to X_u \ell^+ \nu_\ell$ component ($B \to X_u \ell^+ \nu_\ell$ enriched SB), where still about half of the background events originate from $B\bar{B}$ decays of the generic MC. A pure $B \to X_u \ell^+ \nu_\ell$ sideband cannot be obtained since most of the background in the sideband originates from $B\bar{B}$ generic MC.

In the following, all sidebands plots are shown in a row where plots for the fit variables are given in section 6.1.1 and the network input variables are shown in section 6.1.2.

6.1.1 Fit variables
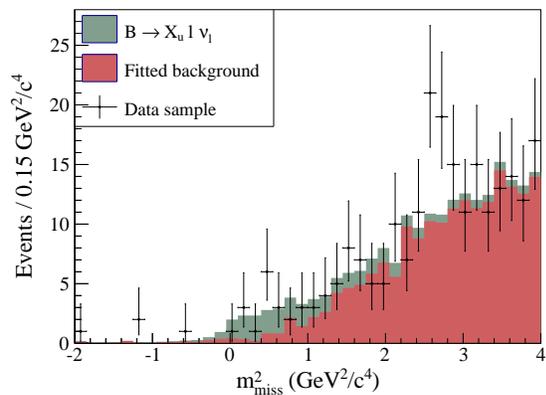
(a) Electron channel in the sideband satisfying $M_{bc} > 5.27$ GeV/c^2 and $n_{out} < 0.25$.
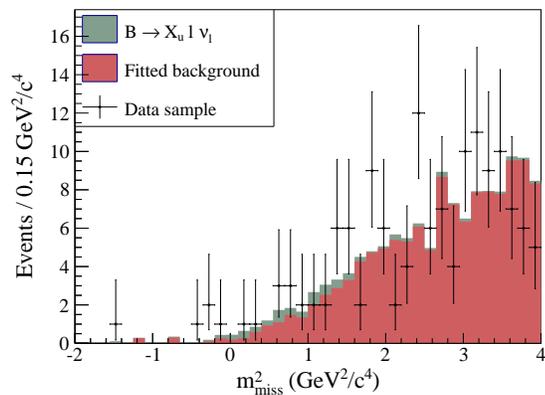
(b) Electron channel in the sideband satisfying $M_{bc} < 5.27$ GeV/c^2 and $n_{out} > 0.25$.
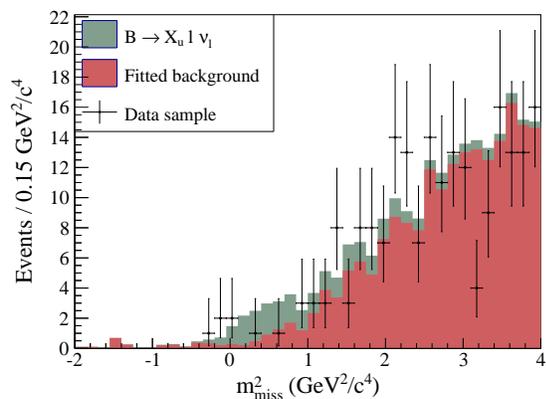
(c) Muon channel in the sideband satisfying $M_{bc} > 5.27$ GeV/c^2 and $n_{out} < 0.25$.
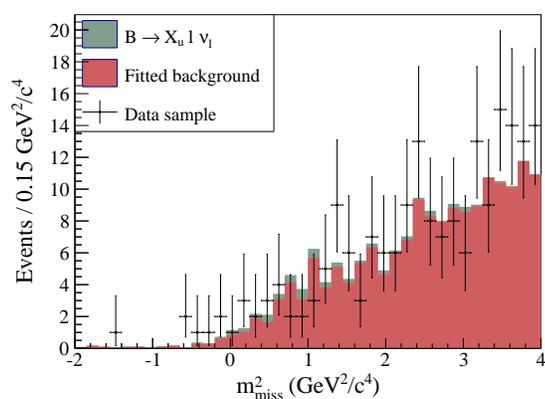
(d) Muon channel in the sideband satisfying $M_{bc} < 5.27$ GeV/c^2 and $n_{out} > 0.25$.

Figure 6.1: Sideband of the m^2_{miss} distribution.
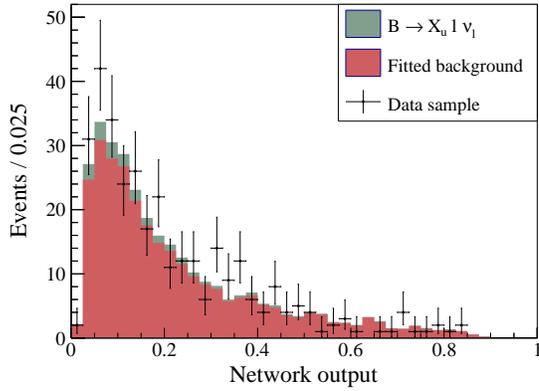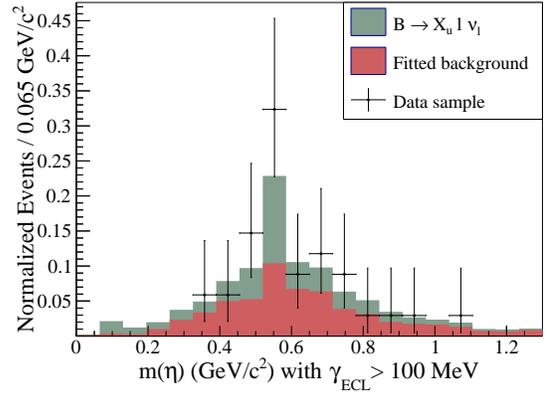
(a) Elec. channel: fitted background SB. (b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
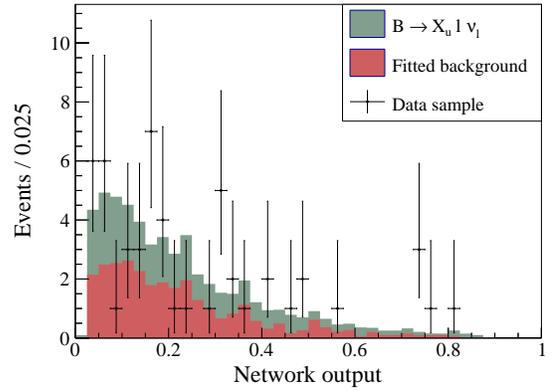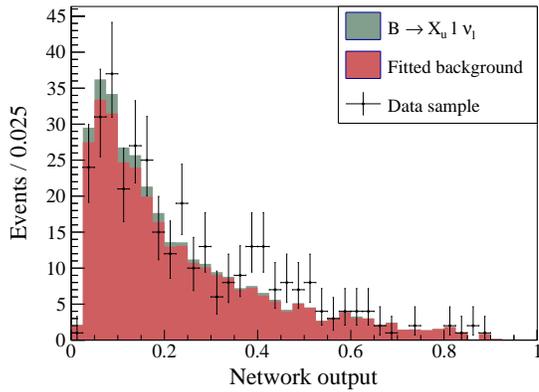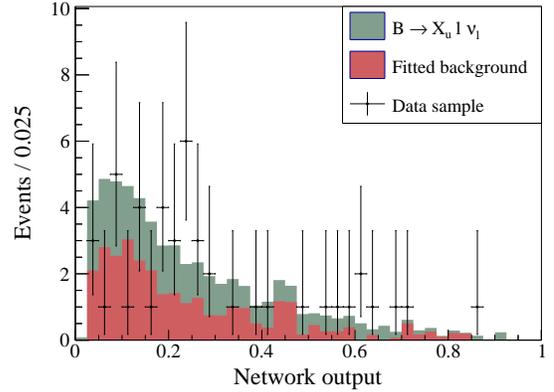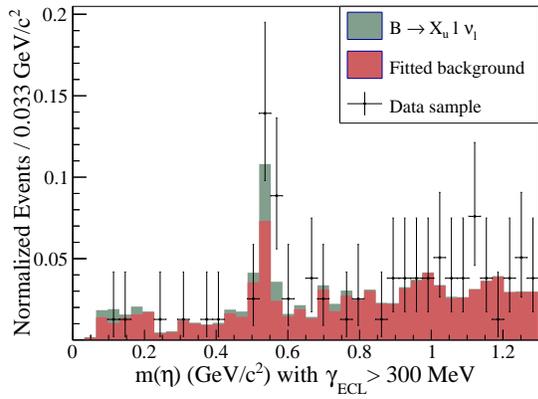
(c) Muon channel: fitted background SB. (d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.2: Sideband of the n_{out} distribution which is defined in section 4.3.

6.1.2 Network training variables
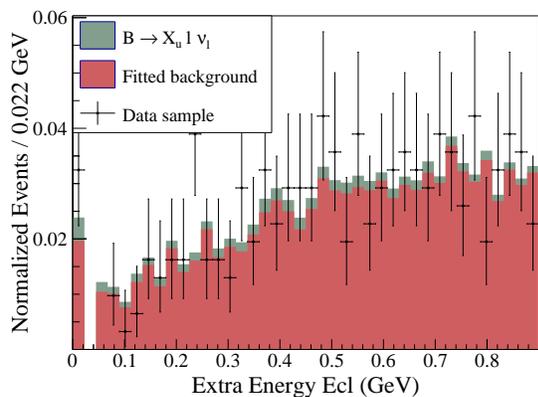
(a) Elec. channel: fitted background SB.
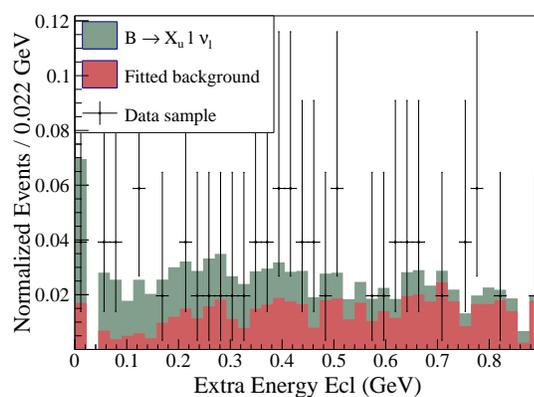
(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
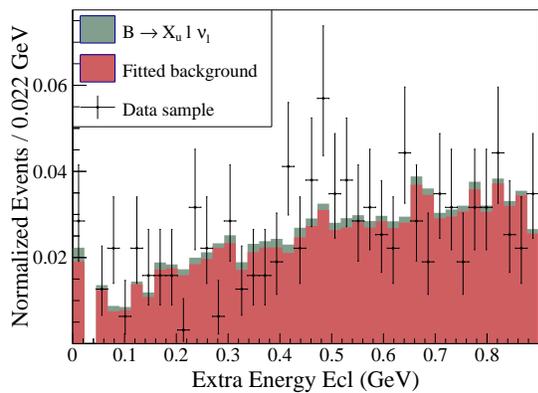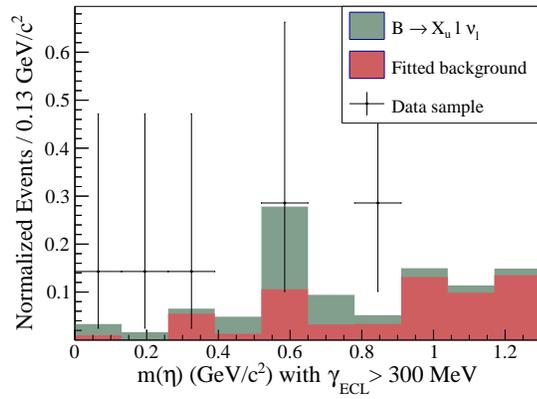
(c) Muon channel: fitted background SB.
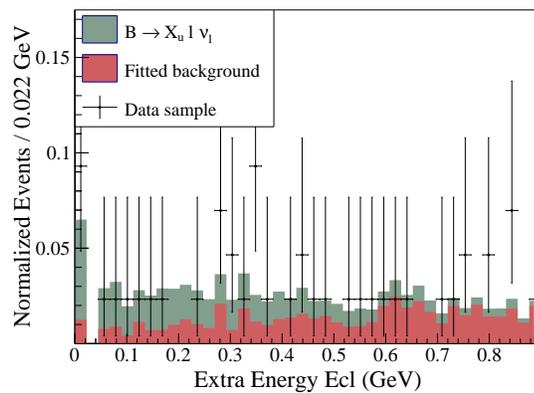
(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.3: Sideband of the extra energy in the ECL.
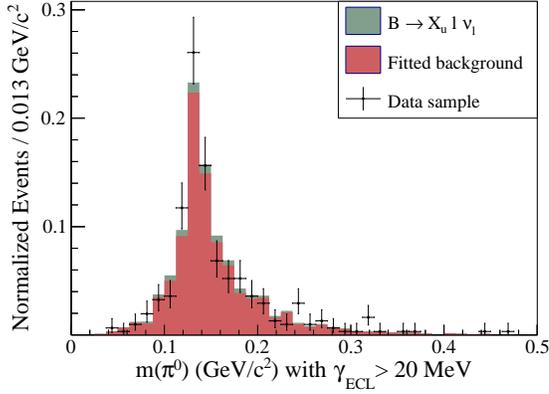
(a) Elec. channel: fitted background SB.
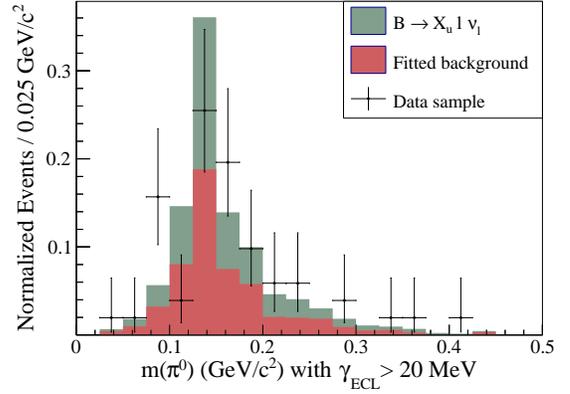
(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
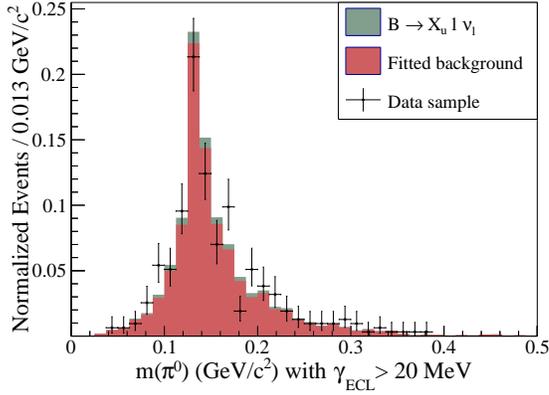
(c) Muon channel: fitted background SB.
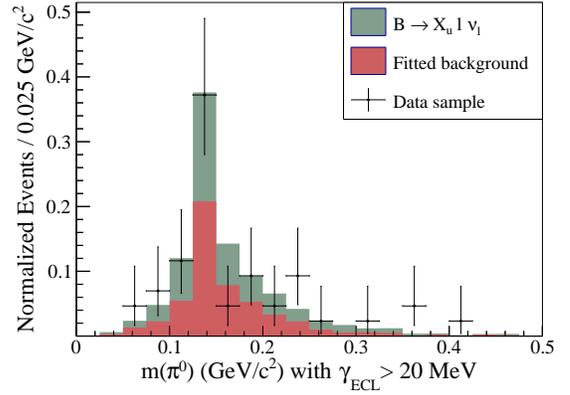
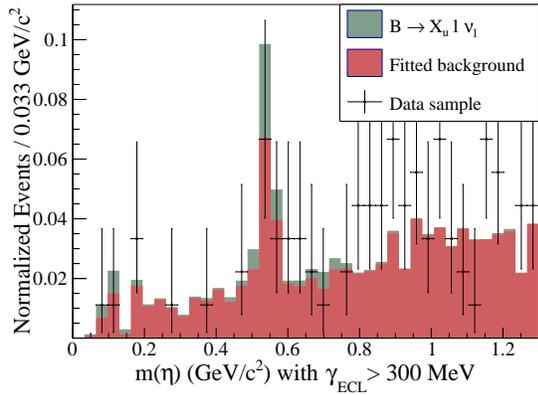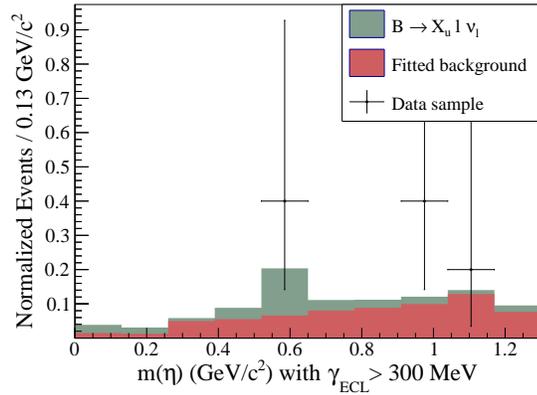(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.4: Sideband of $m(\pi^0)$ computed with $E(\gamma_{ECL}) > 20\,\text{MeV}$.
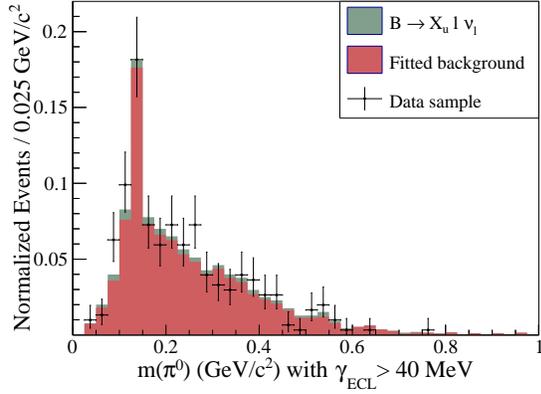
(a) Elec. channel: fitted background SB.
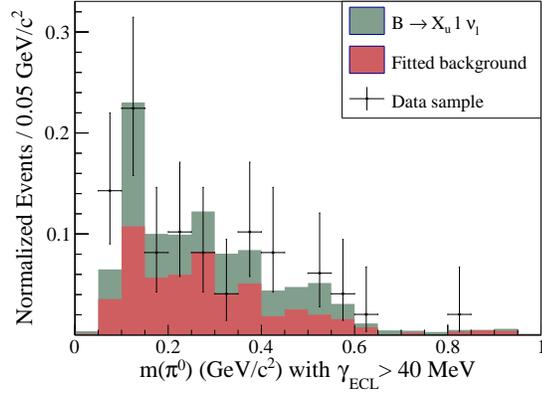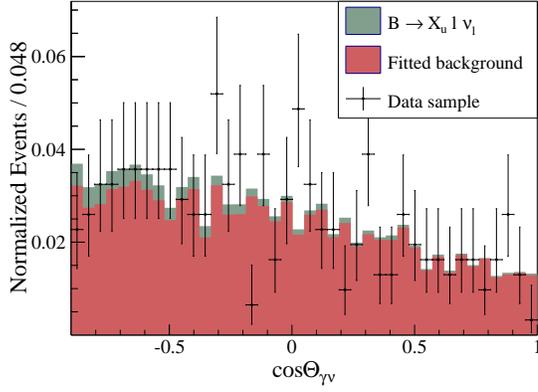
(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
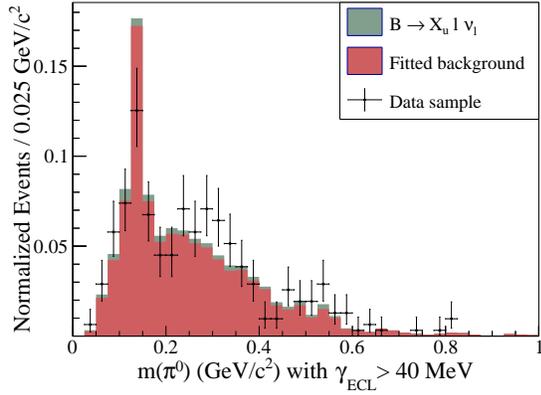
(c) Muon channel: fitted background SB.
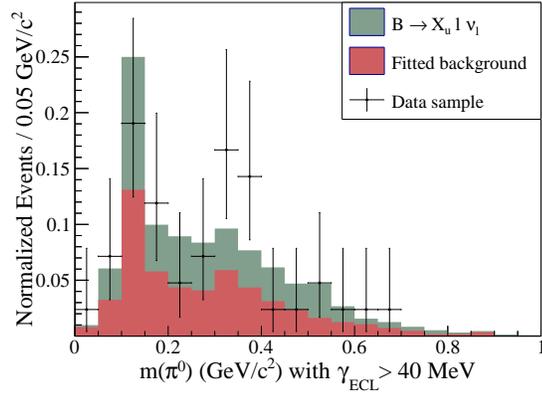
(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.5: Sideband of m(π^0) computed with E(γ_{ECL}) > 40 MeV.
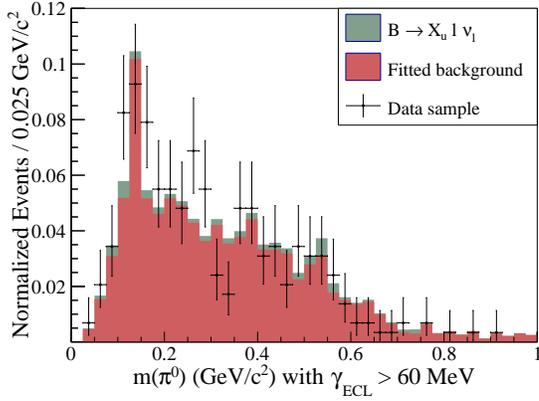
(a) Elec. channel: fitted background SB.
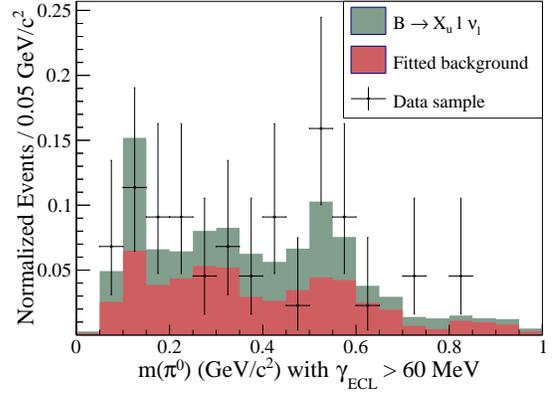
(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
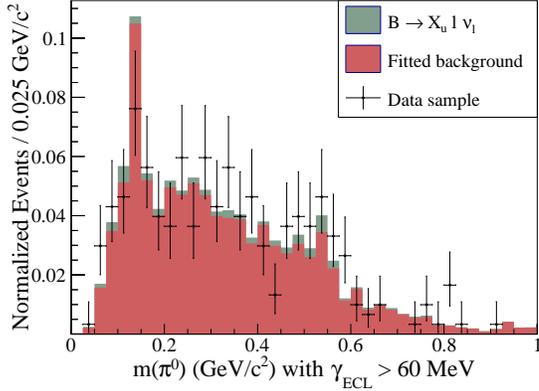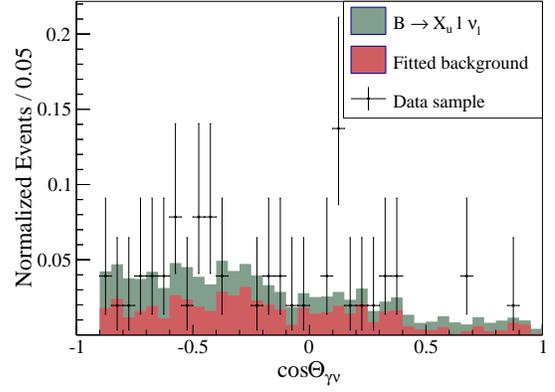
(c) Muon channel: fitted background SB.
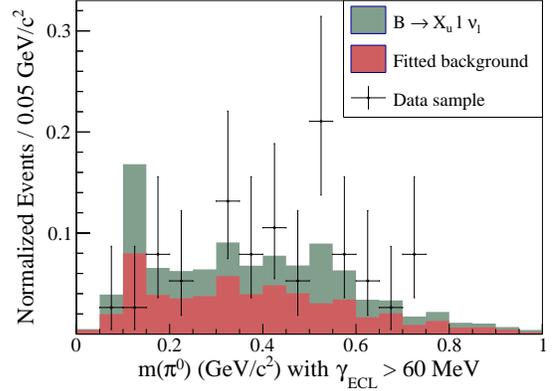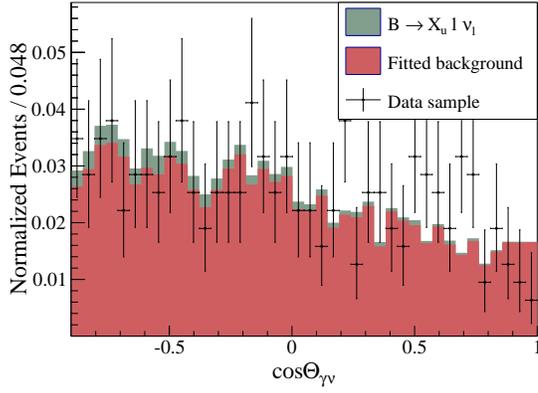
(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

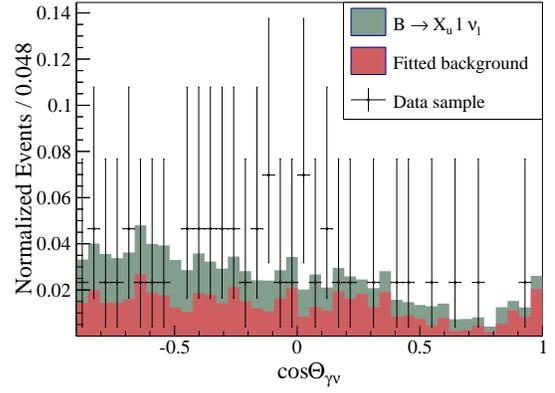Figure 6.6: Sideband of $m(\pi^0)$ computed with $E(\gamma_{\mathrm{ECL}}) > 60\,\mathrm{MeV}$.
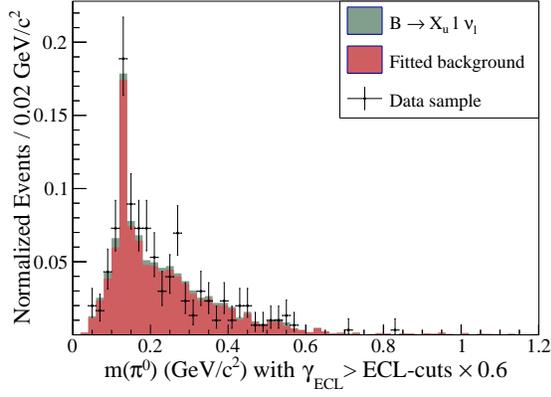
(a) Elec. channel: fitted background SB.
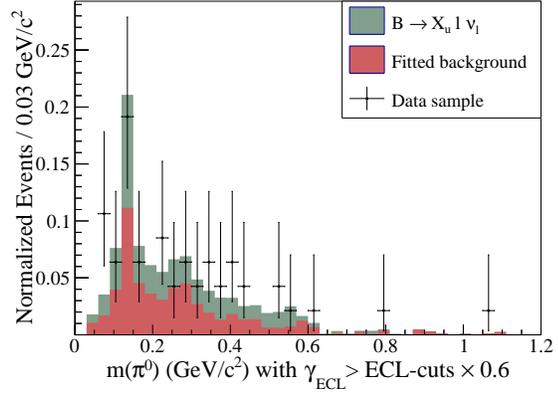
(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
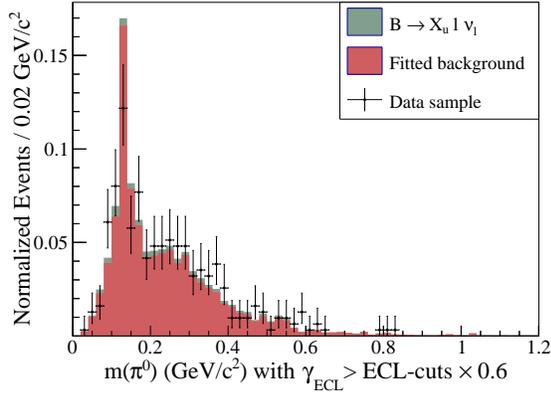
(c) Muon channel: fitted background SB.
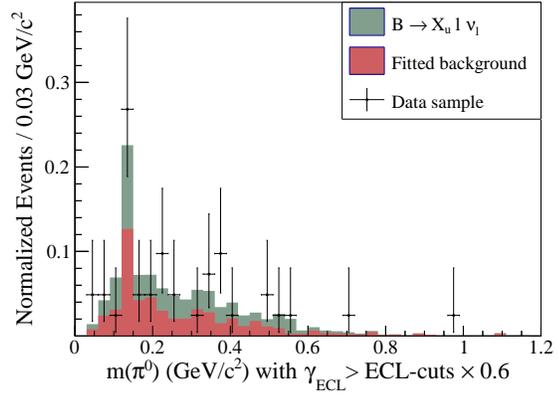
(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.7: Sideband of $m(\pi^0)$ computed with ECL photons satisfying the ECL energy cuts scaled by 0.6.
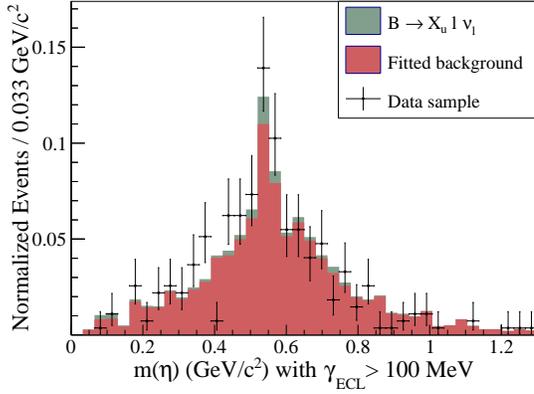
(a) Elec. channel: fitted background SB.   (b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
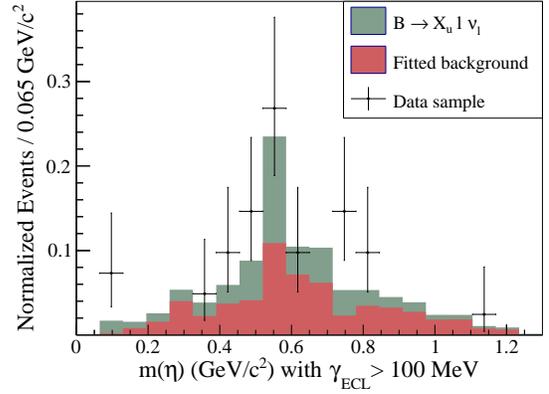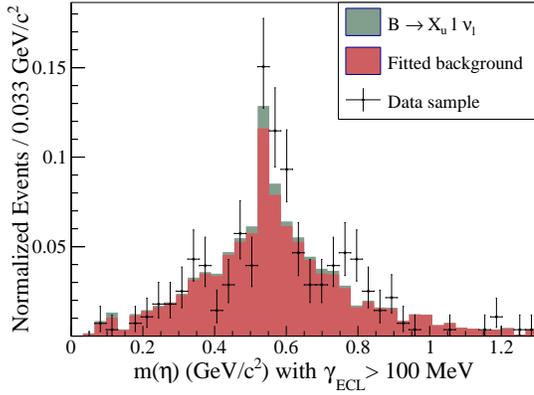
(c) Muon channel: fitted background SB.   (d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.8: Sideband of m(η) computed with E(γ_{ECL}) > 100 MeV.

(a) Elec. channel: fitted background SB.

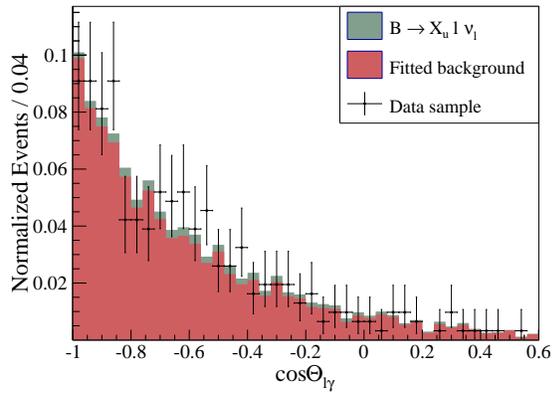(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
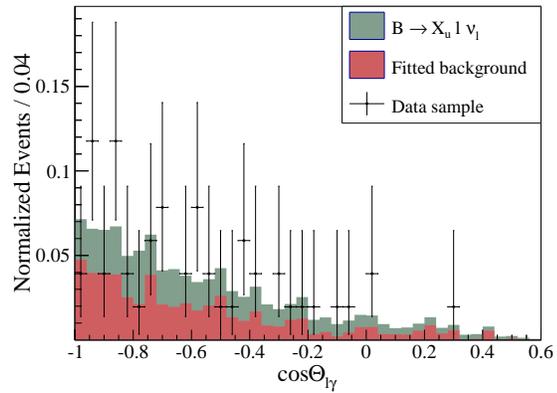
(c) Muon channel: fitted background SB.

(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.9: Sideband of m(η) computed with E(γ_{ECL}) > 300 MeV.

(a) Elec. channel: fitted background SB.

(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

(c) Muon channel: fitted background SB.

(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.
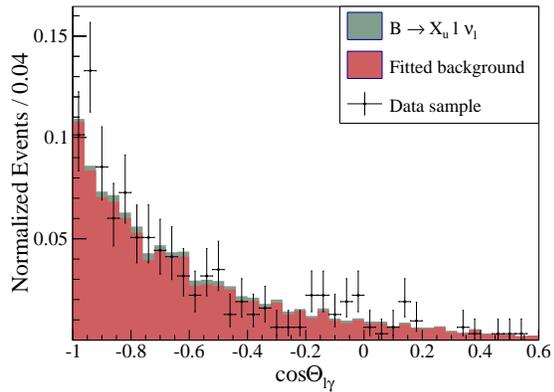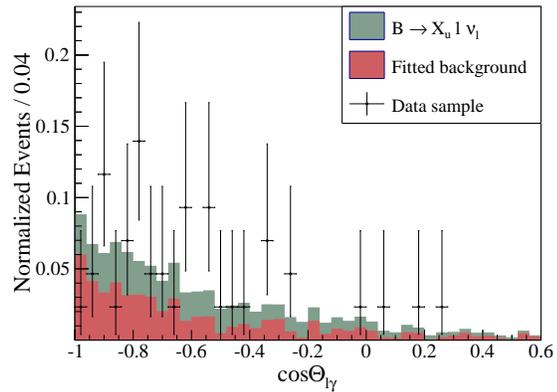
Figure 6.10: Sideband of the $\cos\Theta_{\gamma\nu}$ angle.

(a) Elec. channel: fitted background SB.

(b) Elec. channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

(c) Muon channel: fitted background SB.

(d) Muon channel: $B \to X_u \ell^+ \nu_\ell$ enriched SB.

Figure 6.11: Sideband of the $\cos \Theta_{\gamma \ell}$ angle.

6.2 Training of MC against data

An additional check for the agreement of data and MC is obtained by using a neural network which is trained to separate MC from data. If any differences in the shapes or the correlation structure of the input variables is found between data and MC, the neural network will be able to separate the samples.

The MC sample representing the data sample in composition and luminosity is obtained by sampling the generic and $b \to u\ell^+\nu_\ell$ MC. A binary classification is trained with NeuroBayes in the sideband of $m^2_{\mathrm{miss}} > 0.3 \ \mathrm{GeV}^2/c^4$. The training is identical to the training in section 4.3, and the same set variables are used to train simultaneously in the electron and muon channel.

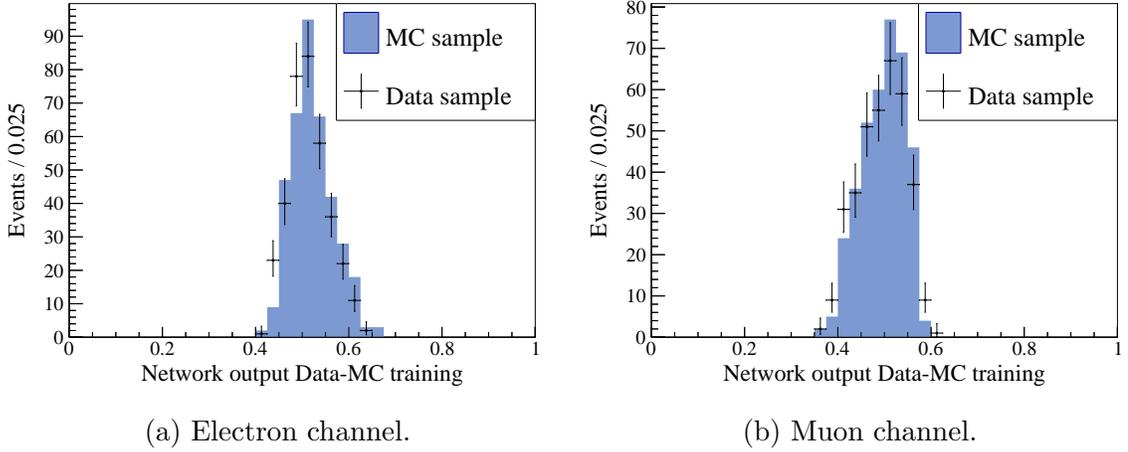(a) Electron channel.   (b) Muon channel.

Figure 6.12: Network output obtained by training data against MC.

The network output obtained from the training is plotted separately for the electron and muon channel in Fig. 6.12. It shows a weak separation of the samples from which a good data-MC agreement for the background is deduced. The most significant variables in the training are given by

- $m(\pi^0)$ with $E(\gamma_{\mathrm{ECL}}) > 20 \, \mathrm{MeV}$ (Fig. 6.4)

- $m(\eta)$ with $E(\gamma_{\mathrm{ECL}}) > 300 \, \mathrm{MeV}$ (Fig. 6.9)

- Extra Energy in the ECL (Fig. 6.3)

- $m(\pi^0)$ with $E(\gamma_{\mathrm{ECL}}) > 60 \, \mathrm{MeV}$ (Fig. 6.6)

- $m(\pi^0)$ with $E(\gamma_{\mathrm{ECL}}) > 40 \, \mathrm{MeV}$ (Fig. 6.5).

These variables show no apparent deviation between data and MC.

A quantitative error on the signal shape in the network output is determined with the $B^0 \rightarrow K^*\gamma(K^* \rightarrow K^+\pi^-)$ control channel in section 7.3.
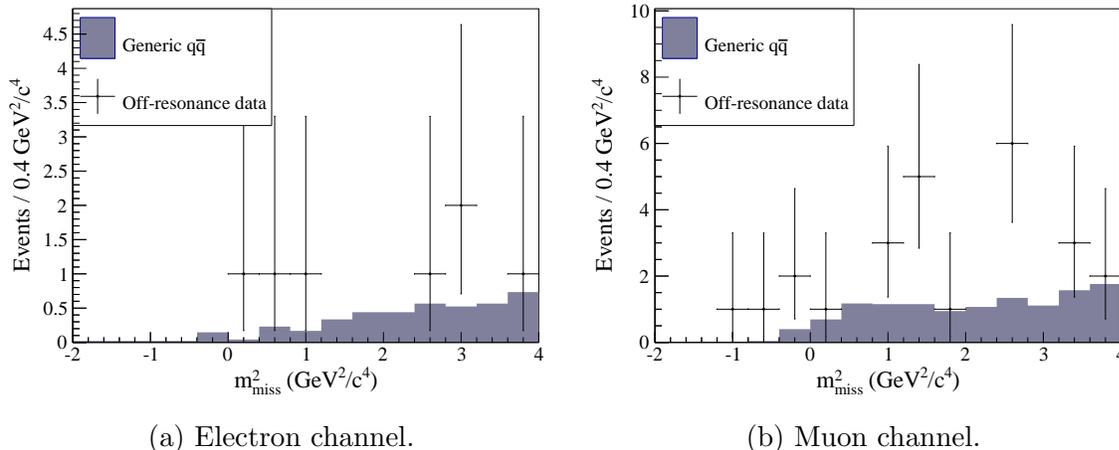

(a) Electron channel. (b) Muon channel.

Figure 6.13: m^2_{miss} distribution for $q\bar{q}$ off-resonance data and generic $q\bar{q}$ MC for the signal selection without the $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$ requirement. The MC sample is scaled to the luminosity of the off-resonance data.

6.3   Off-resonance data

An off-resonance data sample has been recorded $60\,\mathrm{MeV}$ below the $\Upsilon(4S)$ resonance with the Belle experiment. The sample has a luminosity of $79.37\,\mathrm{fb}^1$ which corresponds to 11.3% of the size of the on-resonance data. This sample is used in $B\bar{B}$ analyses to determine the shape of the $q\bar{q}$ continuum background which is identical for on- and off-resonance data. This background consists of two-jet QCD processes whose modeling uncertainty is larger than for $B\bar{B}$ events.

Discrepancies for the continuum background between data and MC are observed in sideband distributions in section 4.2.1. The m^2_{miss} distribution in Fig. 4.11a has an excess in data around zero. The off-resonance data is used to check for any data-MC discrepancy in the signal region. The distribution of generic $q\bar{q}$ MC and off-resonance data is shown in Fig. 6.13, where the MC sample is scaled to the luminosity of the off-resonance data. The signal selection is applied to the samples, with the exception of the $M_{\mathrm{bc}} > 5.27\,\mathrm{GeV}/c^2$ requirement. A selection on M_{bc} cannot be performed due to the lower CMS energy in the off-resonance sample which causes the hadronically reconstructed fake B_{tag} candidates to be lower in mass. The statistics of

95

the off-resonance sample is low, but the distributions show no significant discrepancy, especially in the signal region below $m_{\mathrm{miss}}^2 < 1\mathrm{GeV}^2/c^4$. Therefore no large deviation is to be expected of the continuum background in the on-resonance data.

7 Systematic error estimation

7.1 Determination of systematic errors

The systematic errors are estimated in toy MC studies. To obtain the deviation for the given systematic, an alternative model is determined by a fit to a re-weighted MC or by varying parameters of the existing model by their errors. The alternative and default model are then fitted to the same toy sample. Performing many toy experiments, the differences between the signal yields have a Gaussian shape whose mean is identified as the systematic error.

The systematic errors are summed in quadrature to obtain the total error for the separate fits of the electron and muon channel. The systematic error for the simultaneous fit is calculated from the errors of the separate channels where uncorrelated errors are being summed in quadrature and correlated errors are summed linearly. This assumes correlation coefficients of one for all correlated errors.

All systematic deviations determined in this section are given table 7.1, and the errors for the secondary analysis are estimated analogously to the nominal analysis.

7.2 Meson veto systematics

An important systematic error to be checked, originates from the n_{out} variable which is used to bin m_{miss}^2. As described in section 4.3, the network contains: several meson masses calculated with different energy thresholds, the extra energy in the ECL, and the angles between the signal particle candidates. The main purpose of the network is to identify the dominant backgrounds of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$ decays. Only the systematic error originating from the meson masses in the neural network is determined in this section. Two scenarios can be distinguished in these variables, which lead to different systematic uncertainties.

In the first case, the meson mass is calculated from two photons, the signal photon candidate and a remaining photon in the ECL, which both originate from the π^0 or η meson. Here, the Belle experiment prescribes a reconstruction efficiency error of 2% for high momentum π^0 and η particles in two-photon decays. This error changes the height of the peaks in the invariant meson mass spectra by a negligible amount. Furthermore, no large discrepancies can be identified in the sideband distributions of the invariant meson masses in section 6.1.1. Especially in the $B \to X_u \ell^+ \nu_\ell$ enriched sideband, the data-MC comparison of the peak height at the nominal meson masses is a good indicator for the reconstruction efficiency difference.

In the second case, two random photons are combined to compute the meson mass where either one or both of the photons do not originate from a meson decay.

This happens, most importantly, for $B^+ \to \ell^+ \nu_\ell \gamma$ signal events where the signal photon is combined with photonic ECL background. Also in many $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$ events only one photon from the π^0 or η meson is used in the calculation of the meson mass. For these backgrounds, nearly all of the signal photon candidates originate from the meson where in 40% of the cases the second photon is not reconstructed in the ECL. The most important mechanism for photon loss is the merging of the two daughter photons from a high momentum meson into one ECL cluster. Here, the decaying system is strongly boosted which results in a small separation angle between the daughter particles in the laboratory frame. Photons can also be converted via $\gamma \to e^+ e^-$ pair production in the detector material which is traversed before reaching the ECL. In most cases, remaining photons not originating from a meson are beam background or photons missed by the full reconstruction of the B_{tag}.

The systematic error introduced by using the meson masses in the network training is estimated with the $B^+ \to K^*(892)\gamma$ control channel. This decay is chosen to obtain a high-energetic signal photon from the two-body B decay which is then combined with the remaining photons in the ECL to calculate the meson masses. This control channel is a similar environment for the calculation of the meson masses as it is found in the semi-leptonic B decays of the signal and dominant background processes.

The data-MC agreement for the remaining variables in the network training are checked in a different way. The extra energy in the ECL and the angles between the signal particle candidates show no deviations in the sideband distributions.

For the extra energy in the ECL, a large part of the photonic background originates from beam background which is not simulated but measured by the experiment. Furthermore, systematic efficiency differences of this variable are checked in [31] for hadronically tagged exclusive $B \to X_u \ell^+ \nu_\ell$ decays. In that study, the data-MC efficiency ratio for three different extra energy in the ECL variables is compared. The three variables differ in their minimal cluster energies used to build the sum of the energy in the ECL. One of the variables is the same as in this analysis, and the remaining ones use thresholds of $50\,\text{MeV}$ and $150\,\text{MeV}$ for all parts of the ECL. In conclusion, no significant differences in the efficiency ratios is found among the extra energy in the ECL variables.

The angles between the signal particles are only included in the training for the nominal analysis for which the theoretical model is valid. The model dependence of the angular distribution is checked by comparing leading-order and next-to-leading-order calculations for the signal model in section 7.4.

7.3 $B^+ \to K^*(892)\gamma$ control channel

The $B^+ \to K^*(892)\gamma$ control channel is used to check systematic deviations in the meson mass variables which are input in the network training in section 4.3. The event selection for the $B^+ \to K^*(892)\gamma$ decay is taken from [37]. Signal MC is generated in exclusive $B^+ \to K^*(892)\gamma$ decays with 1×10^6 events and the background estimation includes only the generic $B\bar{B}$ and non-resonant $q\bar{q}$ decays with six times the size of the data sample.

7.3.1 Primary particle selection

Charged pions and kaons are selected with the kaon ID variable described in section 3.1. Particles with a kaon ID value larger and smaller than 0.1 are assumed to be kaons and pions, respectively.

The signal photon originates from a B meson two-body decay, and therefore the signal photon candidates are selected to be high-energetic by requiring energies above $1.4\,\mathrm{GeV}$. Additionally, the cluster shape of the photon is constrained with the $\frac{E_9}{E_{25}}$ variable denoting the ratio of the energy deposition in a 3×3 over 5×5 tile array in the ECL. Values above 0.95 are selected for this variable, which means that only narrow showers are permitted for which 95% of the shower energy is deposited in the inner part of the cluster area. This servers as a photon identification.

The signal photon candidate must have a low probability originating from a π^0 or η meson, which is ensured by pion and eta probability variables [30]. These are required to be below 0.75 and 0.65, respectively (The performance of these variables is tested in the network training of the $B^+ \to \ell^+\nu_\ell\gamma$ analysis where no increase in significance is observed).

7.3.2 Reconstructed particle selection

The K^* meson is reconstructed by combining a K^+ with a π^- candidate, and the mass is required to be within $75\,\mathrm{MeV}/c^2$ of the nominal K^* mass of approximately $896\,\mathrm{MeV}/c^2$ [34]. The best K^* candidate in an event is defined to have the lowest χ^2 value which is obtained in a vertex fit of its two daughter particles K^+ and π^-.

After selecting the best K^* candidate, it is combined with the signal photon candidates to form B^0 candidates. For these it is required that $\Delta E = E(B^0) - E_{\mathrm{Beam}}$ lies within the energy window of $\Delta E \in (-0.4, 0.2)\,\mathrm{GeV}$. The B^0 candidate closest to the nominal mass of about $5.28\,\mathrm{GeV}/c^2$ is selected as the best candidate.

A final requirement is imposed on the helicity angle of the K^* decay. This is the angle between the K^* candidate and its positively charged daughter, where the

K^* momentum is boosted into the rest frame of the B^0 candidate and the daughter momentum into the K^* rest frame. The absolute value of the cosine of this angle is required to be below 0.8.

After selection, only $q\bar{q}$ continuum background is left which can be seen in the distribution of the M_{bc} mass in Fig 7.1. The plot shows the signal and generic MC weighted according to their expected yields on data, where the signal sample is weighted to the world average value of the $B^+ \to K^*(892)\gamma$ branching fraction of $(4.33 \pm 0.15) \times 10^{-5}$. It is evident that the yield of the continuum background is underestimated in MC while the shape is approximately reproduced.
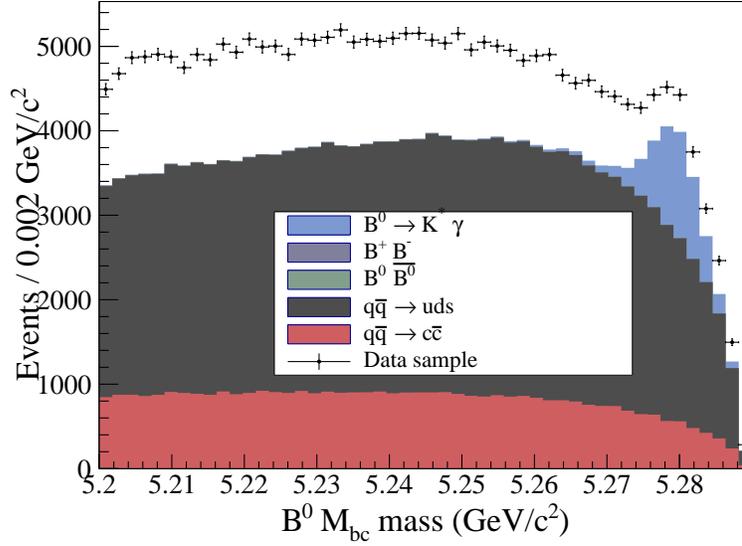
Figure 7.1: M_{bc} mass distribution of the B^0 meson after the $K^*\gamma$ signal selection.

A continuum suppression is used to reduce the remaining background, where signal MC is trained against continuum events of the generic MC with a network containing event shape variables described in the section 3.3. The training is performed with the NeuroBayes package. The resulting network output in Fig. 7.2 is required to be larger than 0.6. The M_{bc} mass distribution after the network selection is shown in Fig. 7.3. The continuum background is still underestimated on MC, but the yield of the signal, which peaks at an M_{bc} mass of around $5.28\,\mathrm{GeV}/c^2$, is in agreement with the world average value. The resonant $B\bar{B}$ background is negligible at this point and not taken into account in the remaining analysis.
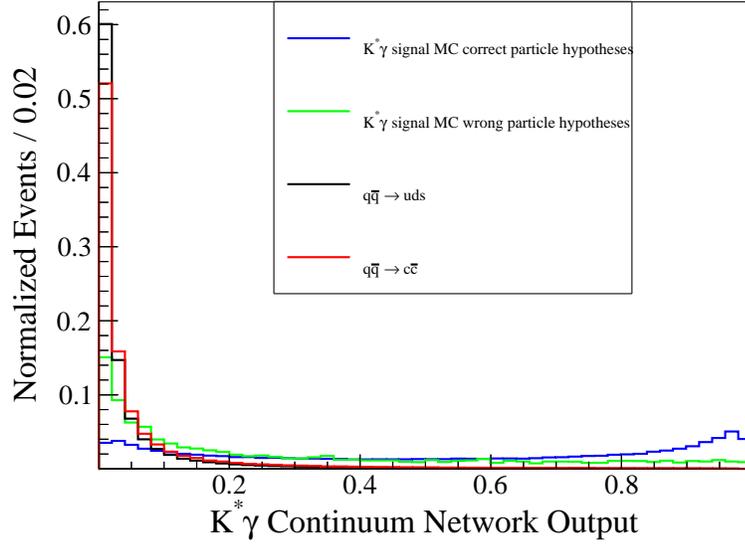
Figure 7.2: Network output for the $B^+ \to K^*(892)\gamma$ continuum suppression.
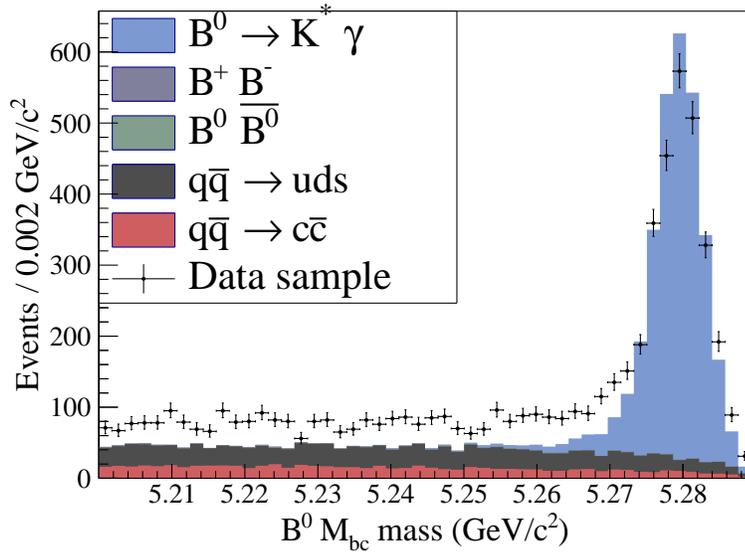
Figure 7.3: M_{bc} mass distribution of the B^0 after requiring the network output of the continuum suppression to be above 0.6.

7.3.3 Fitting and yield extraction

The M_{bc} mass is fitted with a Crystal Ball function for the signal, which is already used for the $B^+ \to \ell^+ \nu_\ell \gamma$ decay in section 5.3, and an Argus function for the continuum background. The Argus function is defined as [38]

$$f(x; \chi, c) = \frac{\chi^3}{\sqrt{2\pi}\, \Psi(\chi)} \frac{x}{c^2} \sqrt{1 - \frac{x^2}{c^2}} \exp\left\{ -\frac{1}{2}\chi^2\left(1 - \frac{x^2}{c^2}\right)\right\},$$

$$\Psi(\chi) = \Phi(\chi) - \chi\phi(\chi) - \frac{1}{2},$$

where χ specifies the end-point of the function, and parameter c is the curvature of the shape. The PDF ϕ is a unit Gaussian with Φ being its cumulative distribution.

In Fig. 7.4, the functions are fitted on data and plotted on the MC in order to compare the shapes between data and MC. The shape of the continuum background agrees well between simulation and data. A difference is observed for the signal where a better M_{bc} mass resolution is found on MC.
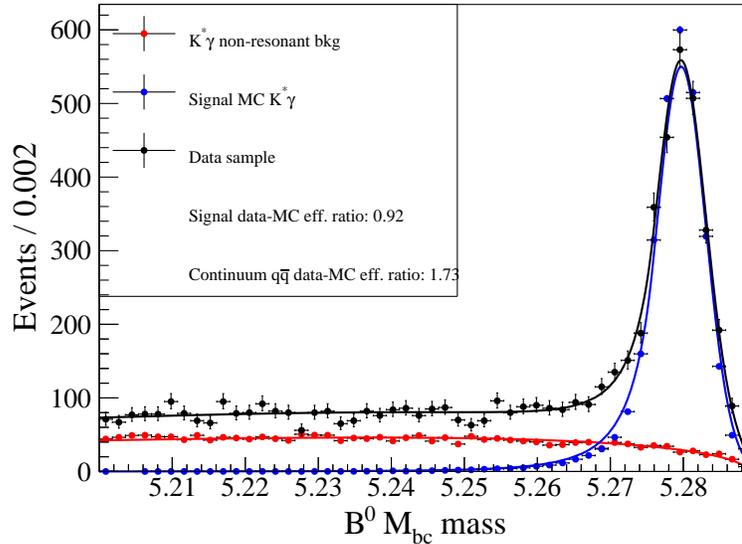
Figure 7.4: Signal and background shape comparison between data and MC for the M_{bc} mass. The PDFs are fitted on data and plotted on the MC samples. The background shape agrees well, and a broader signal peak is observed on data than on MC.

With the fit results, the yield ratio of data and MC is determined. A ratio of 0.93

and 1.6 is found for signal and continuum processes, respectively. The branching fraction of $B^+ \to K^*(892)\gamma$ is computed as

$$\mathcal{B}(B^+ \to K^*(892)\gamma) = \frac{N_{\text{sig}}}{N_{B\bar{B}}\epsilon_{\text{sig}}} = (4.0 \pm 0.15) \times 10^{-5},$$

where the error is purely statistical, with N_{sig} and ϵ_{sig} being the measured signal yield and the signal selection efficiency obtained from signal MC, respectively. The result is in agreement with the world average value stated above, especially if systematic errors would be taken into account.

In the following analysis, the MC samples are normalized to the yields measured from the fits on data in the M_{bc} mass.

7.3.4   Meson veto network crosscheck

A pure signal sample from data has to be obtained in order to perform a shape comparison of the network output between data and the signal MC. Instead of applying a simple selection to the M_{bc} mass, a better result is achieved by using the sPlot background subtraction [39]. This technique uses the signal and background PDFs of the M_{bc} mass to obtain event weights, which are then used for a set of variables uncorrelated to the M_{bc} mass. By applying the weights, data samples corresponding to pure signal and pure background samples can be constructed. The weights are computed from the PDF values and the covariance between the PDF components. The computation is given by

$$w_{\text{sPlot}} = \frac{\sum_{i=1}^{n_{\text{sp}}} \mathbf{V}_{ni} f_i(y_e)}{\sum_{k=1}^{n_{\text{sp}}} N_k f_k(y_e)}, \mathbf{V}_{ni}^{-1} = \sum_{\text{events}} \frac{f_i(y_e) f_n(y_e)}{(\sum_k N_k f_k(y_e))^2},$$

here, f_i denotes the PDF for species i, $\mathbf{V}_{ni}$ is the covariance between species i and n in the discriminating variable, and N_k is the estimated number of events for species i. In this case, the species are given by the signal and the continuum background component.

A network is trained on the MC samples of the $B^+ \to \ell^+ \nu_\ell \gamma$ analysis with the same setup as described in section 4.3 with the difference that only the meson masses are used. The remaining variables used in the training of the $B^+ \to \ell^+ \nu_\ell \gamma$ analysis cannot be computed in the control channel. The extra energy in the ECL is different for the control channel since it is not fully reconstructed, and also the angles between the signal children candidates cannot be included in the training.

The network output is computed for the control channel and the comparison of the output is shown for the background subtracted data sample and signal MC in Fig. 7.5 where a good agreement is observed.
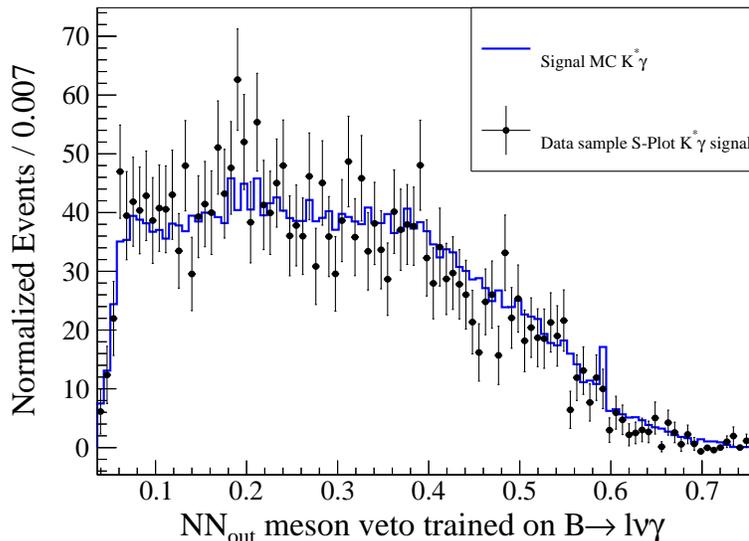
103

Figure 7.5: Network output distribution for sPlot subtracted data and $B^+ \to K^*(892)\gamma$ signal MC. The training is performed on the $B^+ \to \ell^+\nu_\ell\gamma$ MC samples where only with meson masses are used.

To determine the systematic error of the network in the binned fit of the $B^+ \to \ell^+\nu_\ell\gamma$ analysis, data-MC double ratios are calculated in bins of the network output of the control channel. This quantifies the data-MC yield deviations among the bins where the ratios are computed as

$$r = \frac{N_i^{\mathrm{MC}}/N_{\mathrm{sum}}^{\mathrm{MC}}}{N_i^{\mathrm{Data}}/N_{\mathrm{sum}}^{\mathrm{Data}}},$$

where N_i is the event count in bin i and N_{sum} the total number of events. The result of the calculation is plotted in Fig. 7.6, where the bin boundaries are determined to contain an equal amount signal events in the $B^+ \to \ell^+\nu_\ell\gamma$ signal MC. The data sample has a higher yield for low network outputs and accordingly a lower one for high output values. The values for the deviations among the bins are extracted with a linear fit to the distribution. The deviation in a given bin is determined by the value of the fit function at the central position in the bin.

The network output deviations obtained in the control channel are used to re-weight the MC samples of the $B^+ \to \ell^+\nu_\ell\gamma$ analysis. The weights are scaled so that the total expected yield of each fit component from section 5.4 is preserved. The error is then estimated in a toy MC study in which the toy samples are fitted with
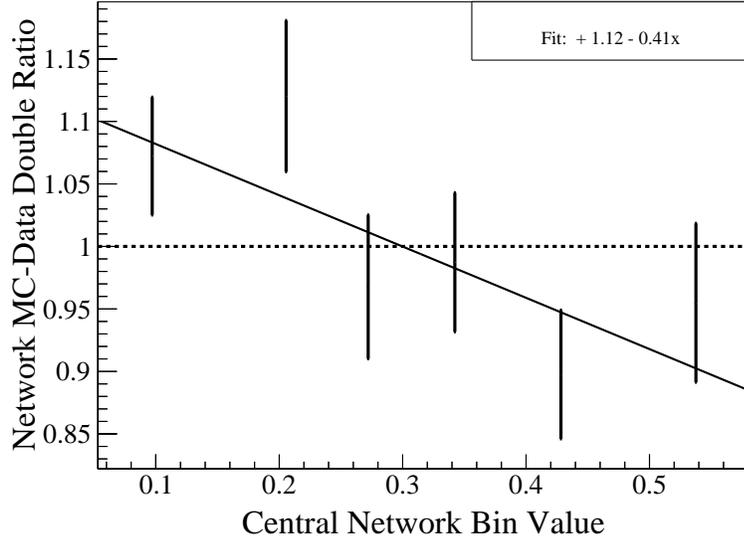
Figure 7.6: Double ratio in bins of the network output calculated with the $B^+ \to K^*(892)\gamma$ signal MC and the sPlot subtracted data sample. A linear fit is used to obtain the deviations in the individual network output bins.

the default fit model and the model obtained from the re-weighted MC samples. The differences between the yields obtained for the two fit models is taken as the systematic error which is listed in table 7.1.

The $B^+ \to \ell^+\nu_\ell\gamma$ analysis and control channel differ since the control channel is not fully reconstructed. This means that particles from the other B meson also enter the calculation of the meson masses and the beam background becomes less important. However, the beam background is assumed to be well described since it is measured at the experiment and added to the MC.

7.4   Remaining systematic errors

PDF shape variation

The PDFs defined in section 5.3 introduce a systematic error since they are determined on MC with limited statistics. To estimate the error, the analytic PDF shapes used for the signal and the "fitted background" fit components are varied by their fit errors obtained on MC. The fit determines asymmetric error bounds in a maximum likelihood fit, where the errors are defined by the value at which the negative logarithmic

Table 7.1: Systematic uncertainties of the signal yield grouped by error-types for the nominal and secondary analysis. Deviations are given in signal yields.

Analysis	Nominal		Secondary	
Decay mode	$B^+ \to \mu^+\nu_\mu\gamma$	$B^+ \to e^+\nu_e\gamma$	$B^+ \to \mu^+\nu_\mu\gamma$	$B^+ \to e^+\nu_e\gamma$
Fit shapes	$^{+0.75}_{-1.34}$	$^{+0.64}_{-1.06}$	$^{+1.18}_{-1.69}$	$^{+1.27}_{-1.9}$
Meson veto network	± 0.58	± 0.66	± 0.91	± 1.08
Fixed $B \to X_u\ell^+\nu_\ell$ yield	± 0.18	± 0.24	± 0.31	± 0.41
$B^+ \to \ell^+\nu_\ell\gamma$ model	-0.01	-0.05	$+0.8$	$+0.4$
Additive error	$^{+0.97}_{-1.47}$	$^{+0.95}_{-1.27}$	$^{+1.72}_{-2.10}$	$^{+1.76}_{-2.26}$
Lepton ID	± 0.42	± 0.18	± 0.62	± 0.27
Tag-side efficiency	± 0.35	± 0.34	± 0.52	± 0.51
Tag-side NN	± 0.13	± 0.40	± 0.19	± 0.48
Tracking efficiency	-0.01	-0.01	-0.02	-0.02
$N_{B\bar{B}}$	± 0.11	± 0.11	± 0.17	± 0.17
Multiplicative Error	± 0.57	± 0.55	± 0.85	± 0.77
Combined error	$^{+1.12}_{-1.58}$	$^{+1.10}_{-1.39}$	$^{+1.92}_{-2.27}$	$^{+1.92}_{-2.39}$

Decay mode	$B^+ \to \ell^+\nu_\ell\gamma$	$B^+ \to \ell^+\nu_\ell\gamma$
Additive error	$^{+1.64}_{-2.15}$	$^{+2.98}_{-3.52}$
Multiplicative error	± 0.99	± 1.44
Combined error	$^{+1.92}_{-2.37}$	$^{+3.31}_{-3.80}$

likelihood function increases by 1/2 of its minimum value to either side.

The PDFs are fitted in each bin of the network output separately, and therefore also each parameter of the PDF is varied in each bin separately. Due to negligible correlations among the parameters, the total error is calculated by adding the individual errors in quadrature.

The PDF for the $B \to X_u\ell^+\nu_\ell$ fitting component is obtained by a kernel estimator algorithm [36] which takes one parameter to determine the size of the Gaussian kernels. This influences the rigidity of the resulting PDF shape. The systematic error

for this component is determined by varying the kernel size to achieve an appreciable change in the shape of the PDF.

The errors for the individual components of the fit model and their sum are given in Fig. 7.7. The fitted background component dominates the error due to the low luminosity of the generic MC sample.
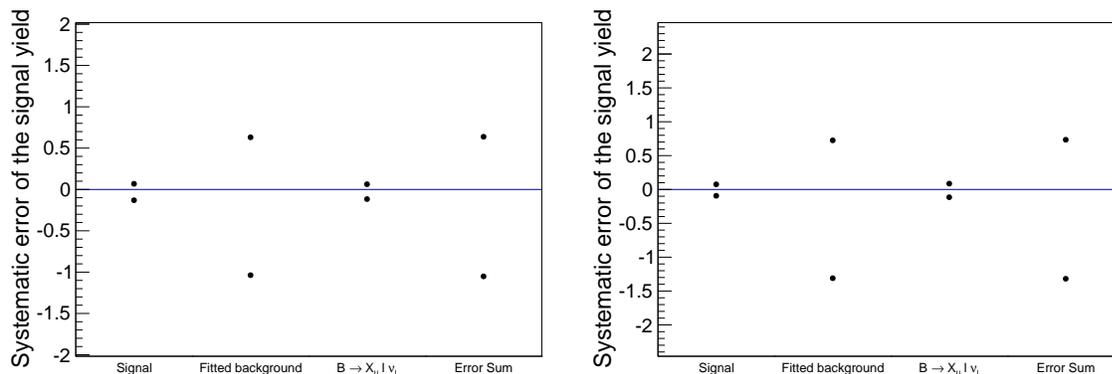
(a) Electron channel.

(b) Muon channel.

Figure 7.7: Systematic error of the PDF shape uncertainties separated for each fit component.

Fixed $B \to X_u \ell^+ \nu_\ell$ branching fraction

The yield of the $B \to X_u \ell^+ \nu_\ell$ component is fixed in the fit to data to the world average values. This component consists of the well known exclusive decays with relatively small errors on their branching fractions. To estimate the impact on the signal yield, the fixed branching fractions are varied by one standard deviation of the PDG errors [34] for each decay channel separately. The result is shown in Fig. 7.8.

The error that an individual decay channel contributes to the total systematic error depends on the number of the background events as well as the relative error of the branching fraction. This leaves the largest backgrounds of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$ decays as the processes contributing the most to the systematic error where η decay mode has a larger contribution due to the higher uncertainty on the branching fraction.

Uncertainties regarding the form factors of the $B \to X_u \ell^+ \nu_\ell$ decays are implicitly contained in the world average errors. Additionally, the Belle analysis [31] finds only small efficiency differences between data and MC for two competing form factor models

107

of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and several other exclusive $B \to X_u \ell^+ \nu_\ell$ decays. It uses comparable selection criteria and the same hadronic full reconstruction as this analysis, and it measures exclusive branching fractions for $B \to X_u \ell^+ \nu_\ell$ with fits to the missing mass distributions in bins of q^2. Here, q^2 denotes the squared four-momentum of the leptonic $\ell^+ \nu_\ell$ system. Complementary to that, a flat reconstruction efficiency is observed for $B \to X_u \ell^+ \nu_\ell$ decays in q^2. Therefore, different form factor models, which change the shape of the differential decay width $d\Gamma/dq^2$, will have a negligible impact on the selection efficiency as well as the shape of the $B \to X_u \ell^+ \nu_\ell$ contribution in the missing mass distribution.
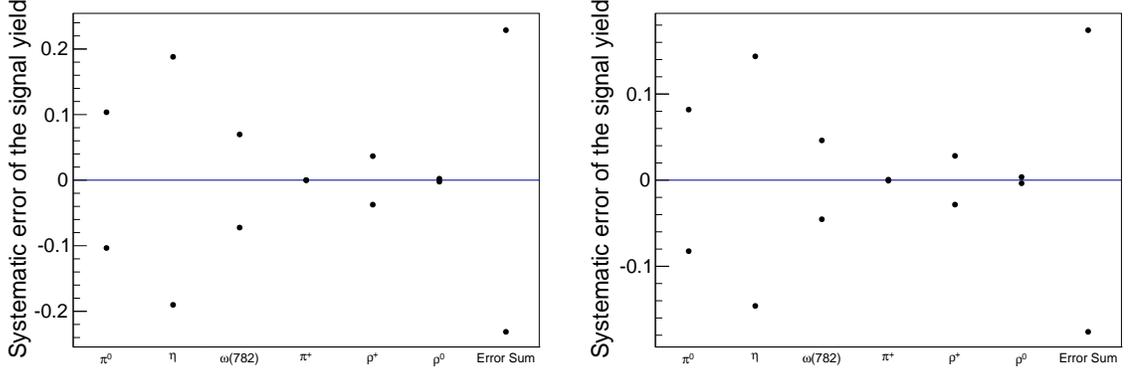
(a) Electron channel.   (b) Muon channel.

Figure 7.8: Systematic error obtained by varying the branching fraction of the $B \to X_u \ell^+ \nu_\ell$ fit component.

Signal model comparison

To obtain an estimate of the model uncertainty for signal decay, the latest NLO calculation [9] is compared the an older LO calculation [11]. The missing mass shapes for the two models are given in Fig. 7.9. The differences in the distributions are small, and also the angular distributions between the signal particles are similar. The difference in the signal yield is -0.05 and -0.01 for the electron and muon channel, respectively.

Tag-side efficiency

The error of the tag-side efficiency is determined in [31] to be 4.2%. This error is taken from the study which is used to obtain the efficiency correction for the B_{tag} as

(a) Electron channel.

(b) Muon channel.
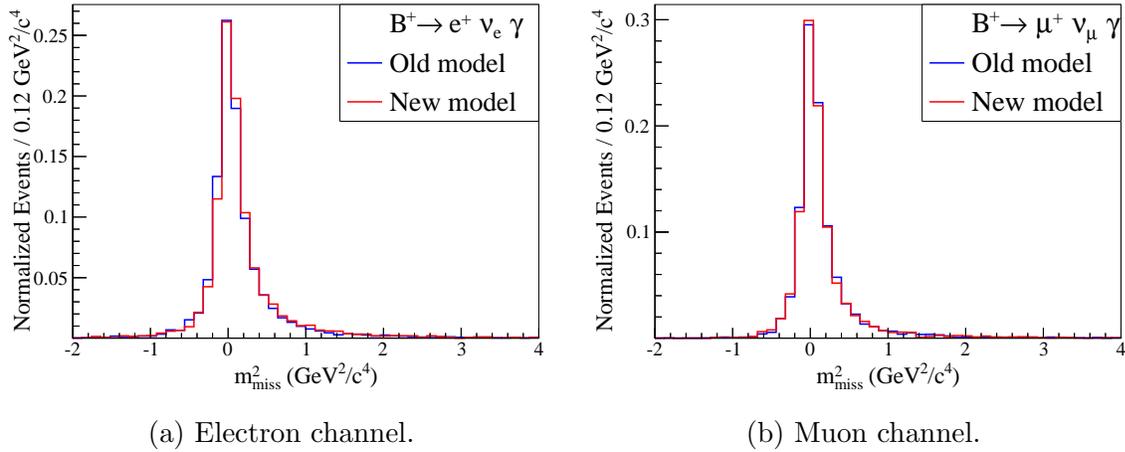
Figure 7.9: m^2_{miss} shape comparison for the LO and NLO calculations.

described in section 4.4.

Tag-side continuum network

The B_{tag} network output includes continuum suppression variables. It is shown in Fig. 4.16 in sideband distributions where a data-MC good agreement is observed. The systematic deviation introduced by the continuum suppression is estimated by comparing the data-MC efficiency ratio for the selection used in the analysis in the sideband of $m^2_{\mathrm{miss}} > 0.3\mathrm{GeV}^2/c^4$. For that, the signal selection of section 4.2 is applied with the exception of the selection on the B_{tag} network output. The resulting errors are -0.4 and -0.13 for the electron and muon channel, respectively.

Tracking systematic

The tracking efficiency is determined by [40] with high tracks of high transverse momentum from partially reconstructed D^* decays. It applies to tracks with a momentum of at least $200\,\mathrm{MeV}$ which is given for this analysis. The tracking error amounts to a negligible error of -0.13%. The error is introduced by the charged track of the signal side lepton. The tracking efficiencies from the tag-side tracks are accounted for by the error on the tag-side efficiency mentioned before.

109

Belle PID systematic

The Belle lepton ID shows a data-MC discrepancy which is described in section 4.4. The error is determined in [32] in $\gamma\gamma \to \ell^+\ell^-$ processes, and it depends on the selection of the ID variable. For the selection of this analysis an efficiency difference of 2.2% and 5.0% is cited for the electron and muon ID, respectively.

Number of $B\bar{B}$ pairs

The error on the total number of $B\bar{B}$ pairs is provided by Belle to be $(771.581 \pm 10.566) \times 10^6$.

8 Measurement on data

The results for the measurement on data are listed in table 8.1 with the systematic errors included. No significant result is found for any of the decay channels. The fits to the separate decay channels are shown in Fig. 8.1 and 8.2 for the nominal analysis, with signal photon energies of $E_\gamma^{\text{sig}} > 1\,\text{GeV}$, and for the secondary analysis, with signal photon energies of $E_\gamma^{\text{sig}} > 400\,\text{MeV}$, respectively. To better illustrate the fit results obtained from the fit in bins of n_{out}, the unbinned m_{miss}^2 distribution is shown in Fig. 8.3.

The signal yield in the muon channel is close to zero, where for the electron channel few events are fitted which are not significant. This observation results in a better limit for the muon than for the electron channel. Also the limit for $B^+ \to \ell^+ \nu_\ell \gamma$ determined in the simultaneous fit benefits from the strong limit of the muon channel.

All PDFs show a good agreement with the data sample. Additionally, the yield of the fitted background component on data is in agreement with its MC prediction, as shown in table 8.2. Only for background of the $B^+ \to e^+ \nu_e \gamma$ channel in the secondary analysis, the MC prediction is more than two standard deviations below the yield on data. However, this is not a significant discrepancy when looking at all background yields combined. Also for the network outputs no discrepancy between data and MC is found in the signal region, as shown in Fig. 8.4 where the vertical lines indicate the binning used for the fit of m_{miss}^2.

The likelihood functions for the two separate fits are shown in Fig. 8.5. The likelihoods are plotted before and after convolution with their respective 90% quantiles which define the upper limits on the yields. From the upper limit on the signal yield, the limit on the branching fraction is computed.

8.1 Limit determination for λ_B

With the upper limit on the partial branching fraction $\Delta\mathcal{B}$, a limit on the λ_B parameter can be determined. The partial branching fraction is connected to the differential decay width $d\Gamma/dE_\gamma$ as

$$\Delta\mathcal{B} = \frac{\tau_{B_d}}{\hbar} \int\limits_{\min(E_\gamma^{\text{sig}})}^{m_B c^2/2} dE_\gamma \frac{d\Gamma}{dE_\gamma} = \frac{\tau_{B_d}}{\hbar} \int\limits_{\min(x_\gamma)}^{1} \frac{m_B}{2} dx_\gamma \frac{d\Gamma}{dx_\gamma},$$

with $x_\gamma = 2E_\gamma/(m_B c^2)$. The differential decay width is integrated over the phase space of the photon energy used in the analysis. The lower integration bound is given by $\min(E_\gamma^{\text{sig}}) = 1\,\text{GeV}$ and $\min(E_\gamma^{\text{sig}}) = 400\,\text{MeV}$ for the nominal and secondary

111

analysis, respectively, and the integration goes up to the maximal possible energy of half the B meson mass. The integral is reparametrized by substituting E_γ with x_γ in the following expressions.

The differential decay width for $B^+ \to \ell^+ \nu_\ell \gamma$ given in Eq. 4 can be rewritten as

$$\frac{d\Gamma}{dE_\gamma} = \frac{\alpha_{em} G_F^2 |V_{ub}|^2}{48\pi^2} m_B^4 (1 - x_\gamma) x_\gamma^3 \left[F_A^2 + F_V^2 \right],$$

$$F_V(E_\gamma) = \frac{f_B}{x_\gamma} \left(\frac{1}{\lambda_B(\mu)} \underbrace{Q_u R(E_\gamma, \mu)}_{t_1} + \underbrace{\frac{c}{m_B} + \frac{Q_u}{x_\gamma m_B} + \frac{Q_b}{m_b}}_{t_2} \right),$$

$$F_A(E_\gamma) = \frac{f_B}{x_\gamma} \left(\frac{1}{\lambda_B(\mu)} Q_u R(E_\gamma, \mu) + \underbrace{\frac{c}{m_B} - \frac{Q_u}{x_\gamma m_B} + \frac{Q_b}{m_b} + \frac{Q_\ell}{m_B}}_{t_3} \right),$$

where t_{1-3} are the terms within the form factors. These terms are used to solve the equations for λ_B. The result for λ_B is obtained as

$$0 = \lambda_B^2 \left(1 - \underbrace{\frac{\text{const}}{\Delta\mathcal{B}} \int_{\min(x_\gamma)}^{1} dx_\gamma (1 - x_\gamma) x_\gamma (t_2^2 + t_3^2)}_{I_1} \right)$$

$$-2\lambda_B \underbrace{\frac{\text{const}}{\Delta\mathcal{B}} \int_{\min(x_\gamma)}^{1} dx_\gamma (1 - x_\gamma) x_\gamma (t_1 t_2 + t_2 t_3)}_{I_2} - \underbrace{\frac{\text{const}}{\Delta\mathcal{B}} \int_{\min(x_\gamma)}^{1} dx_\gamma (1 - x_\gamma) x_\gamma 2 t_1^2}_{I_3}$$

$$\Rightarrow \lambda_B = \frac{I_2 + \sqrt{I_2^2 + I_3 - I_1 I_3}}{1 - I_1},$$

$$\text{const} = \frac{\alpha_{em} G_F^2 |V_{ub}|^2}{48\pi^2} \frac{m_B^5}{2} \frac{\tau_{B_d}}{\hbar} f_B.$$

The integrals containing the term t_1 have to be integrated numerically due to the complicated form of the radiative corrections $R(E_\gamma, \mu)$. The results obtained with the function for λ_B are cross checked with a plot provided in [9]. Additionally, the calculation is performed using the double differential decay width from Eq. 3 and a two-dimensional numerical integration. It is found that all results agree.

The value of λ_B depends on a number of input parameters whose errors influence the precision with which it can be determined. The parameter values given in table 8.3

Table 8.1: Measured signal yields on data for the nominal and secondary analysis where the first error is statistical and the second error is systematic. The significances and the 90% limits contain systematic errors. The BaBar result is directly comparable with the nominal analysis. Under the assumption of a negligible signal contribution for $E_\gamma^{\rm sig} < 400\,{\rm MeV}$, the BaBar result can also be compared to the secondary analysis.

Nominal analysis ($E_\gamma^{\rm sig} > 1\,{\rm GeV}$)			
Decay mode	$B^+ \to e^+ \nu_e \gamma$	$B^+ \to \mu^+ \nu_\mu \gamma$	$B^+ \to \ell^+ \nu_\ell \gamma$
Signal yield	$6.1^{+4.9\,+1.0}_{-3.9\,-1.3}$	$0.9^{+3.6\,+1.0}_{-2.6\,-1.5}$	$6.6^{+5.7\,+1.6}_{-4.7\,-2.2}$
Significance (σ)	1.7	0.4	1.4
$\Delta\mathcal{B}$ central value (10^{-6})	$3.8^{+3.0\,+0.7}_{-2.4\,-0.9}$	$0.6^{+2.1\,+0.7}_{-1.5\,-1.1}$	$2.0^{+1.7\,+0.6}_{-1.4\,-0.7}$
$\Delta\mathcal{B}$ limit (10^{-6})	< 6.1	< 3.4	< 3.5
BaBar $\Delta\mathcal{B}$ limit (10^{-6})	-	-	< 14

Secondary analysis ($E_\gamma^{\rm sig} > 400\,{\rm MeV}$)			
Decay mode	$B^+ \to e^+ \nu_e \gamma$	$B^+ \to \mu^+ \nu_\mu \gamma$	$B^+ \to \ell^+ \nu_\ell \gamma$
Signal yield	$11.9^{+7.0\,+1.8}_{-6.0\,-2.3}$	$-0.1^{+5.2\,+1.7}_{-4.1\,-2.1}$	$11.3^{+8.4\,+3.0}_{-7.4\,-3.5}$
Significance (σ)	2.0	-	1.4
$\Delta\mathcal{B}$ central value (10^{-6})	$4.9^{+2.9\,+0.8}_{-2.5\,-1.0}$	-	$2.3^{+1.7\,+0.7}_{-1.5\,-0.8}$
$\Delta\mathcal{B}$ limit (10^{-6})	< 9.3	< 4.3	< 5.1
BaBar $\mathcal{B}$ limit (PDG values) (10^{-6})	< 17	< 26	< 15.6

Table 8.2: Yield of the fitted background component on data compared to the MC prediction, where the errors are purely statistical.

	Nominal analysis ($E_\gamma^{\rm sig} > 1\,{\rm GeV}$)		Secondary analysis ($E_\gamma^{\rm sig} > 400\,{\rm MeV}$)	
Decay mode	$B^+ \to e^+ \nu_e \gamma$	$B^+ \to \mu^+ \nu_\mu \gamma$	$B^+ \to e^+ \nu_e \gamma$	$B^+ \to \mu^+ \nu_\mu \gamma$
Measured yield	336^{+20}_{-19}	352^{+20}_{-19}	739^{+29}_{-28}	759^{+29}_{-28}
MC prediction	315 ± 4.2	348 ± 4.5	668 ± 6.1	714 ± 6.4

are taken from Ref. [9] with the exception of the parameter c which is taken from Ref. [10]. This latter parameter has the largest uncertainty. With these proposed parameters the central value of λ_B is determined. The parameters are changed by their errors to determine minimal and maximal possible values for λ_B which is the defined as the error range.

Using the measured values for $\Delta\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma)$ from table 8.1, the limit $\lambda_B > 238\,\text{MeV}$ is obtained within the error range $\lambda_B > (172, 410)\,\text{MeV}$ at 90% confidence level for the nominal analysis. A limit on the λ_B can only be derived in a clean way with this partial branching fraction since the differential decay width $d\Gamma/dE_\gamma$ is only valid for that photon energy range. Nevertheless, similar values are obtained for the secondary analysis with $\lambda_B > 234\,\text{MeV}$ within a range of $\lambda_B > (175, 399)\,\text{MeV}$.

Table 8.3: Parameter values from Ref. [9] used in the form factors where the value for c is taken from Ref. [10]. The values for σ_1 and σ_1 are given for the partial phase space of $E_\gamma^{\text{sig}} > 1\,\text{GeV}$.

Parameter	Value	Parameter	Value
G_F	$1.16637 \times 10^5\,\text{GeV}^2$	τ_{B_d}	$1.64\,\text{ps}$
α_{em}	$1/129$	f_B	$195 \pm 10\,\text{MeV}$
$\Lambda_{\overline{\text{MS}}}^{n_f=4}$	$289.9\,\text{MeV}$	m_b	$4.8 \pm 0.1\text{GeV}$
m_B	$5279\,\text{MeV}$	Λ_{QCD}	$217\,\text{MeV}$
$|V_{ub}|_{\text{incl.}}$	4.27×103	$\sigma_1(1\text{GeV})$	1.5 ± 1
$|V_{ub}|_{\text{excl.}}$	3.38×103	$\sigma_2(1\text{GeV})$	3 ± 2
c	$-1.11^{+0.95}_{-0.27}$	ξ_0	$-0.5^{+0.04}_{-0.12}$

8.2 Comparison with the BaBar measurement

The current PDG limits [34] for $B^+ \to \ell^+\nu_\ell\gamma$ decays are based on the previous best measurement by BaBar [7] which is carried out on a sample of $465 \times 10^6 B\bar{B}$ events with a hadronic tag. In that analysis, it is claimed that a model-independent measurement of the total $B^+ \to \ell^+\nu_\ell\gamma$ branching fraction is measured. To achieve this, neither the signal photon energy nor the angles between the signal particles are used in the event selection. This, however, neglects effects due to the acceptance and reconstruction efficiency of the detector which depends on the angle and the energy of a photon as well as the other signal particles. Furthermore, many variables used in the selection

depend directly on the signal decay model. This includes a π^0 mass veto, which computed with the signal photon candidate, and the missing mass variable, which is calculated similarly to section 4.1. Therefore, the claim of model-independence can be viewed as contentious.

Making similar assumptions for the secondary analysis, the downward slope of the photon energy spectrum below $1\,\mathrm{GeV}$ in Fig. 2.7 can be taken as an indication that only few signal events are to be expected below $400\,\mathrm{MeV}$. Thus it can be concluded that the BaBar limit of the total branching fraction can be compared to the limit of the partial branching fraction obtained from the secondary analysis for $E_\gamma^{\mathrm{sig}} > 400\,\mathrm{MeV}$. Also the secondary analysis reduces its model-dependency by not using the angles between the signal particles in the training of the network, as described in section 4.3. This leads to the conclusion that the current PDG of the total branching fractions are significantly improved by the limits of the secondary analysis.

The BaBar analysis provides a partial branching fraction with $E_\gamma^{\mathrm{sig}} > 1\,\mathrm{GeV}$ only for the combined $B^+ \to \ell^+ \nu_\ell \gamma$ decay mode. Also here, a significant improvement is observed comparing this result to the nominal analysis. Accordingly, using this BaBar limit with the result for λ_B calculated above, a limit of $\lambda_B > 129\,\mathrm{MeV}$ within the bounds $\lambda_B > (95, 216)\,\mathrm{MeV}$ is deduced at 90% confidence level. This limit is significantly lower than the limit given in the BaBar paper of $\lambda_B > 300\,\mathrm{MeV}$. There, form factors from the old calculation [11] are used to determine this limit. These increase the value since the calculation misses NLO terms as well as radiative corrections given for the LO term in [9] which decrease the predicted branching fraction by more than 30%. Additionally, the limit is falsely computed with the better measurement of the total branching fraction described above.

(a) Electron channel.

(b) Muon channel.
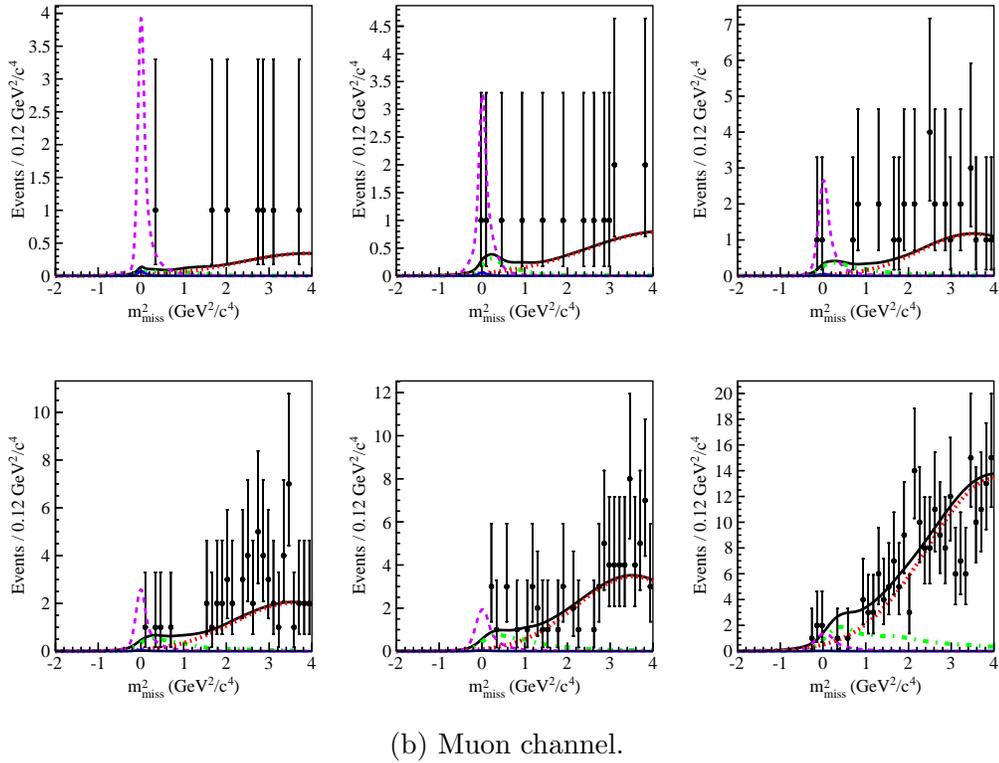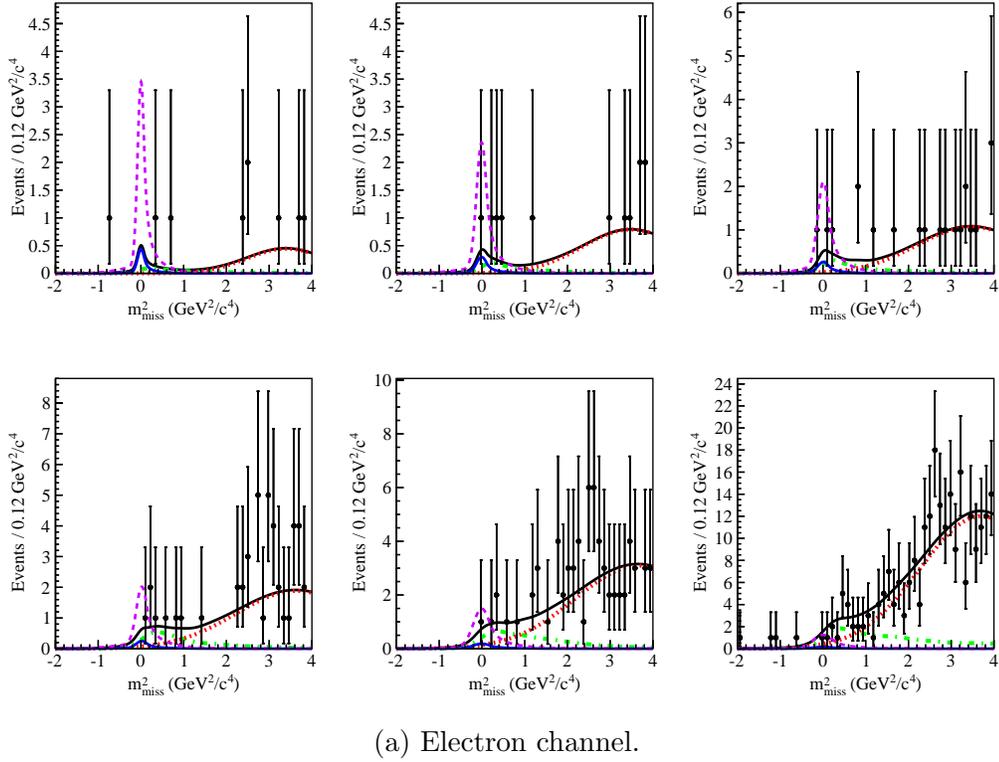
Figure 8.1: Measurement of the nominal analysis with the fit of the m_{miss}^2 in six bins of n_{out} on data (points with error bars). Shown are the PDFs for: the signal (blue), enhanced signal (dashed violet), fixed $B \to X_u \ell^+ \nu_\ell$ backgrounds (green), fitted backgrounds (red), and the sum (black). The enhanced signal function, which has the same normalization for each bin, corresponds to a branching fraction of 30×10^{-6}. The most signal-like bin is found in the upper left plot. Proceeding from left to right, the distributions become increasingly more background-like and the most background-like bin is shown in the lower right panel.
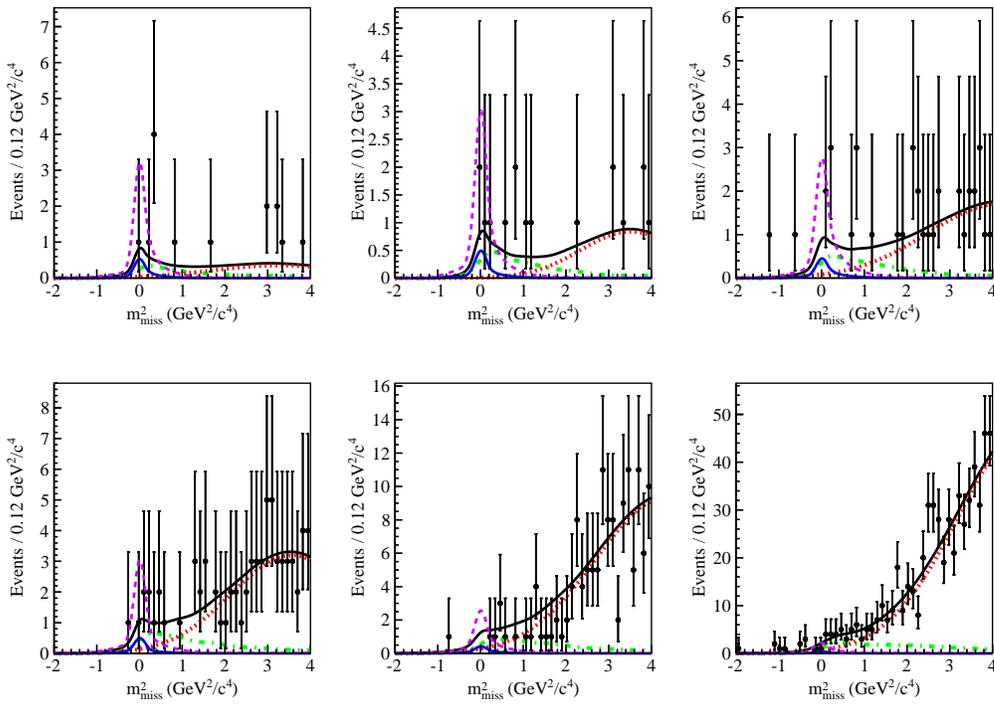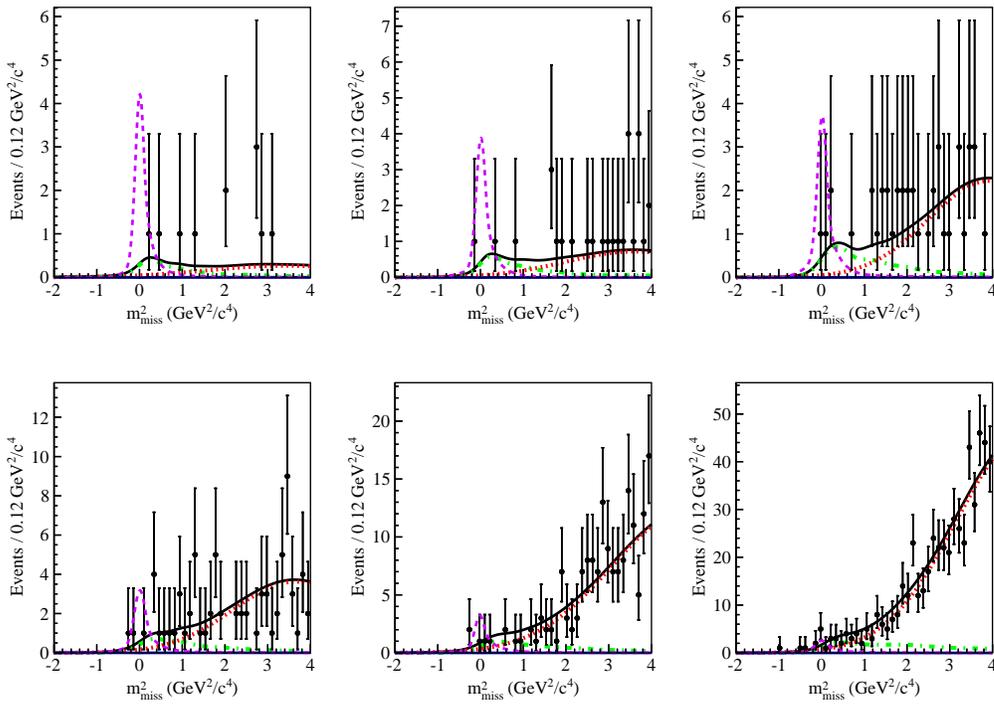
(a) Electron channel.

(b) Muon channel.

Figure 8.2: Measurement for the secondary analysis with the fit of the m^2_{miss} in six bins of n_{out} on data. See description Fig. 8.1
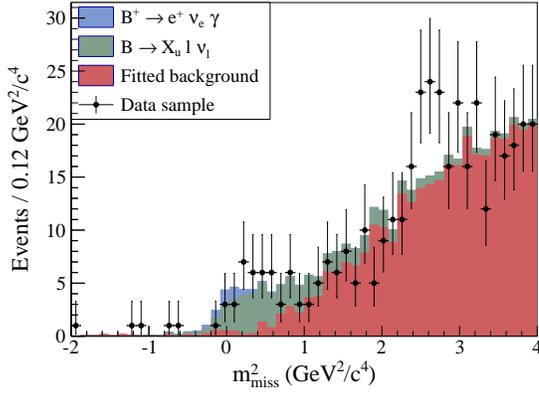
(a) Electron channel of the nominal analysis.   (b) Muon channel of the nominal analysis.
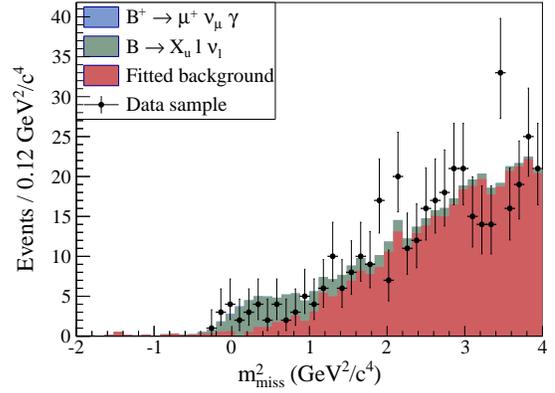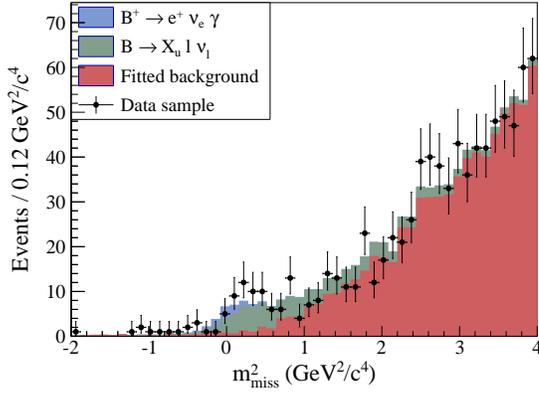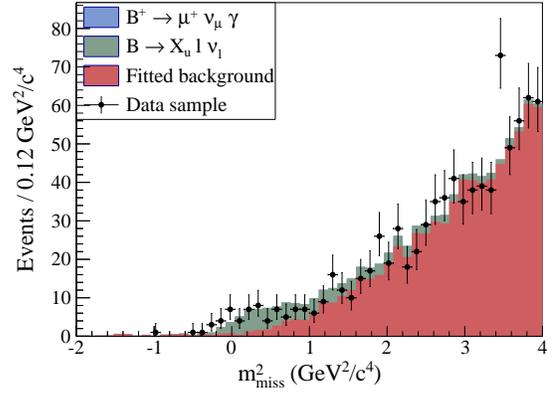
(c) Electron channel of the secondary analysis. (d) Muon channel of the secondary analysis.

Figure 8.3: Measurement on data illustrated in the unbinned m^2_{miss} distribution where the enhanced signal corresponds to a branching fraction of 30×10^{-6}.
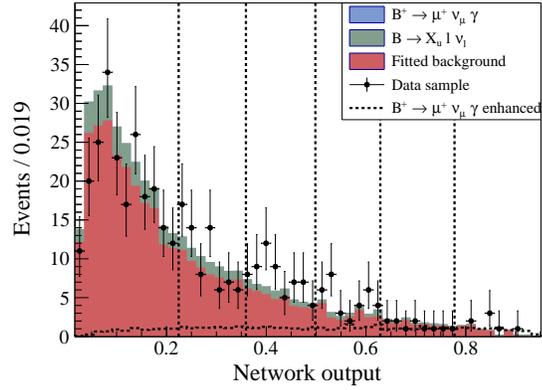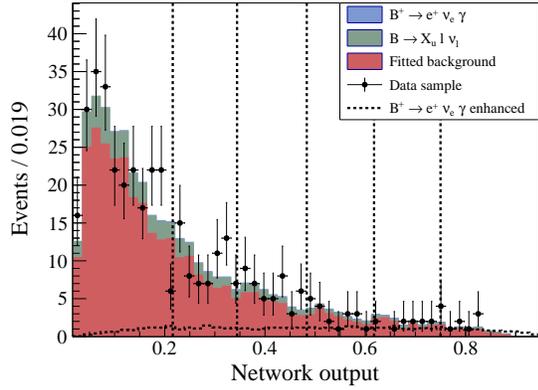
(a) Electron channel of the nominal analysis.

(b) Muon channel of the nominal analysis.
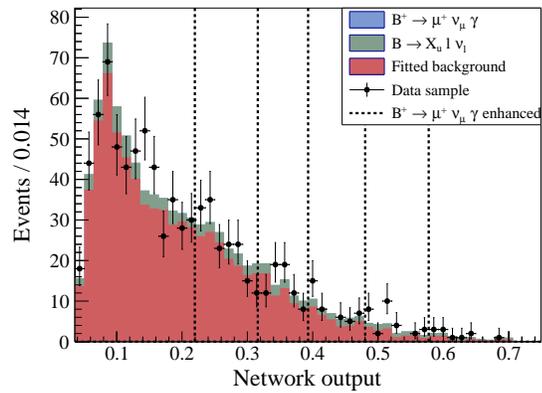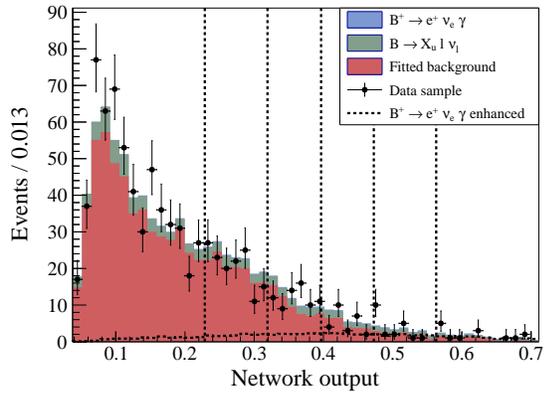
(c) Electron channel of the secondary analysis. (d) Muon channel of the secondary analysis.

Figure 8.4: Network outputs used to bin m^2_{miss} where the bin boundaries are indicated by the vertical lines. The normalizations of the MC distributions are taken from the fit results in m^2_{miss} and the enhanced signal corresponds to a branching fraction of 30×10^{-6}.
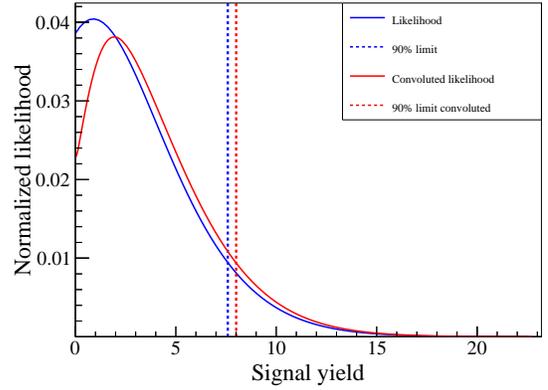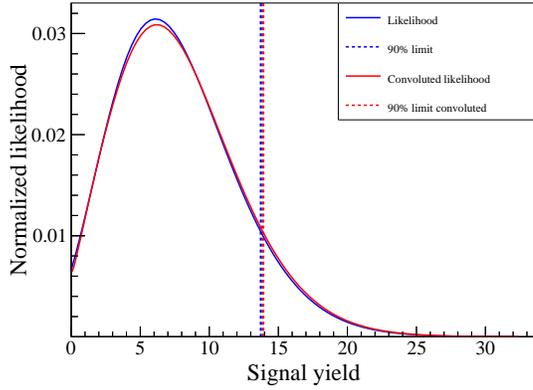
(a) Electron channel of the nominal analysis. (b) Muon channel of the nominal analysis.
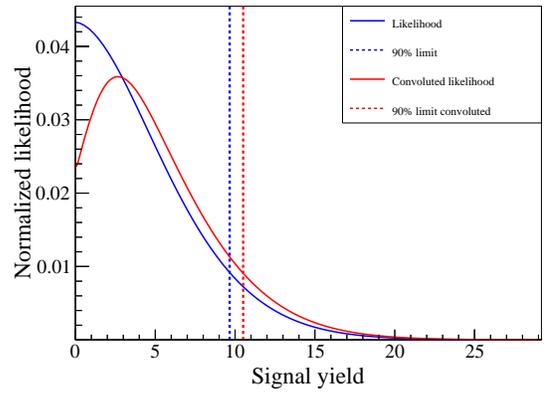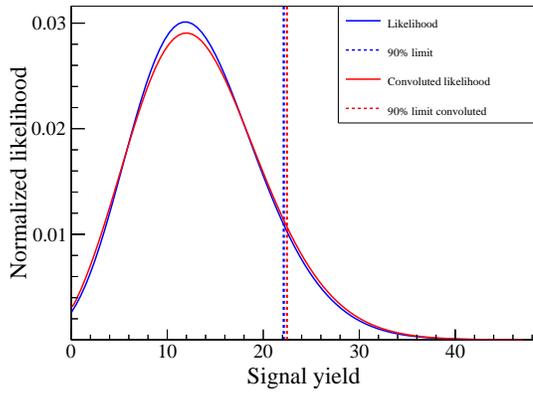
(c) Electron channel of the secondary analysis. (d) Muon channel of the secondary analysis.

Figure 8.5: Likelihood function obtained from the measurement on the data sample in the profile of the signal yield. The likelihoods are shown before and after convolution together with their respective limits.

9 Conclusion

This work presents the measurement of $B^+ \to \ell^+\nu_\ell\gamma$ in the decay modes where ℓ^+ is an electron or a muon. Additionally, the two decay modes are measured in a combined fit to obtain the $B^+ \to \ell^+\nu_\ell\gamma$ branching fraction. The results of the analysis have been published in Phys. Rev. D **91**, 112009 (2015) [41].

The analysis uses the full Belle data set of $772 \times 10^6 B\bar{B}$ pairs. No significant signal is found in any of the channels where the limits at 90% confidence level read

$$\Delta\mathcal{B}(B^+ \to e^+\nu_e\gamma) < 6.1 \times 10^{-6},$$
$$\Delta\mathcal{B}(B^+ \to \mu^+\nu_\mu\gamma) < 3.4 \times 10^{-6},$$
$$\Delta\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) < 3.5 \times 10^{-6}.$$

These are the limits for the partial branching fraction of signal photon energies above $1\,\text{GeV}$. This value is chosen since the signal model is only valid for these photon energies. The best previous limit of the partial branching fraction of $\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) < 14 \times 10^{-6}$, which was reported by the BaBar collaboration on a data set of comparable size, is significantly improved by this measurement.

The improvement to the BaBar limit is achieved through the application of state of the art analysis techniques to optimize the significance of the result. These include neural network estimators for a multi-variate signal and background separation and a quasi two-dimensional fit model to extract the signal. The fit uses the variable with the best signal-to-noise ratio with the neural network output as the second fit dimension.

From the limit on $B^+ \to \ell^+\nu_\ell\gamma$, the limit on the hadronic parameter λ_B is calculated to be $\lambda_B > 238\,\text{MeV}$ at 90% confidence level. The calculation depends on several input parameters which are varied by their errors to obtain the error range of $\lambda_B > (172, 410)\,\text{MeV}$. The large error is dominated by soft higher-order terms in the theoretical model of the $B^+ \to \ell^+\nu_\ell\gamma$ decay whose unknown contribution has to be varied within a large range.

A secondary analysis is performed to obtain the partial branching fraction for signal photon energies above $400\,\text{MeV}$, and it is performed to minimize the signal model dependence. Consistent results are obtained in this analysis at 90% confidence level with

$$\mathcal{B}(B^+ \to e^+\nu_e\gamma) < 9.3 \times 10^{-6},$$
$$\mathcal{B}(B^+ \to \mu^+\nu_\mu\gamma) < 4.3 \times 10^{-6},$$
$$\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) < 5.1 \times 10^{-6}.$$

The current PDG values are provided by the BaBar collaboration for the total branching fraction. Under the assumption that a negligible amount of signal events is found below a photon energy of $400\,\text{MeV}$, the PDG values can be compared with the secondary analysis. The old limits are $\mathcal{B}(B^+ \to e^+\nu_e\gamma) < 17 \times 10^{-6}$, $\mathcal{B}(B^+ \to \mu^+\nu_\mu\gamma) < 26 \times 10^{-6}$, and $\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) < 15.6 \times 10^{-6}$. Also here, all of the values are significantly improved. With the $B^+ \to \ell^+\nu_\ell\gamma$ limit of the secondary analysis, a consistent limit of $\lambda_B > 234\,\text{MeV}$ is calculated which varies within a range of $\lambda_B > (175, 399)\,\text{MeV}$.

The parameter λ_B is used in the QCD factorization which calculates charmless hadronic B meson decays into two mesons. The theoretical error due to the poor knowledge of the parameter varies by decay channel and it is particularly high in color-suppressed modes of $B \to \pi\pi, \pi\rho$, and $\rho\rho$. Measurements of these decays favor values of about $\lambda_B \sim 200\,\text{MeV}$. On the other hand, values of $\lambda_B \sim (350 - 500)\,\text{MeV}$ are obtained in non-perturbative calculations with light-cone sum rules. Although the limit on λ_B measured here seems to be larger than the favored value for color-suppressed decays, it has a large error. Therefore the tension can not yet be resolved.
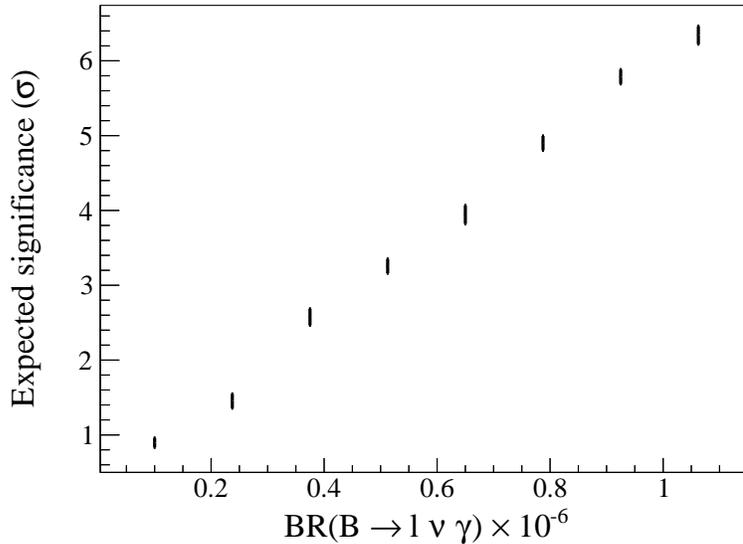
Figure 9.1: Expected significance of the $B^+ \to \ell^+\nu_\ell\gamma$ decay for the full Belle II data set of $50\,\text{ab}^{-1}$.

The Belle II experiment is scheduled to acquire a data set that will be 50 times larger than that of its predecessor. Furthermore, several improvements in the reconstruction of the data are being implemented for Belle II which will additionally

enhance the efficiency of an analysis. Since the branching fraction of the $B^+ \to \ell^+\nu_\ell\gamma$ decay is expected to be of the order of 10^{-6}, a significant signal is expected in the Belle II data. The expected significance for different $B^+ \to \ell^+\nu_\ell\gamma$ branching fractions is estimated in Fig. 9.1 for the full Belle II data set. For that, the analysis presented in this thesis is performed on simulated Belle I data which is increased 50 times in size. A discovery with a signal significance of 5σ is expected for a branching fraction of $\mathcal{B}(B^+ \to \ell^+\nu_\ell\gamma) = 1 \times 10^{-6}$

A   Appendix

A.1 $B \to X_u \ell^+ \nu_\ell$ PDFs separated by components

Two separate PDFs are added to obtain the fixed $B \to X_u \ell^+ \nu_\ell$ PDF where the fit is performed with a Gaussian kernel estimator algorithm [36]. The high luminosity samples of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$ are fitted in one shape with narrow kernels that follow the MC sample more closely. The result for the nominal analysis is shown in Fig. 1.1 and 1.2 for the electron and muon channel, respectively.

The remaining low statistics channels of the $b \to u \ell^+ \nu_\ell$ MC fitted with broader kernels are shown in Fig. 1.3a and 1.3b for electron and muon channel, respectively.
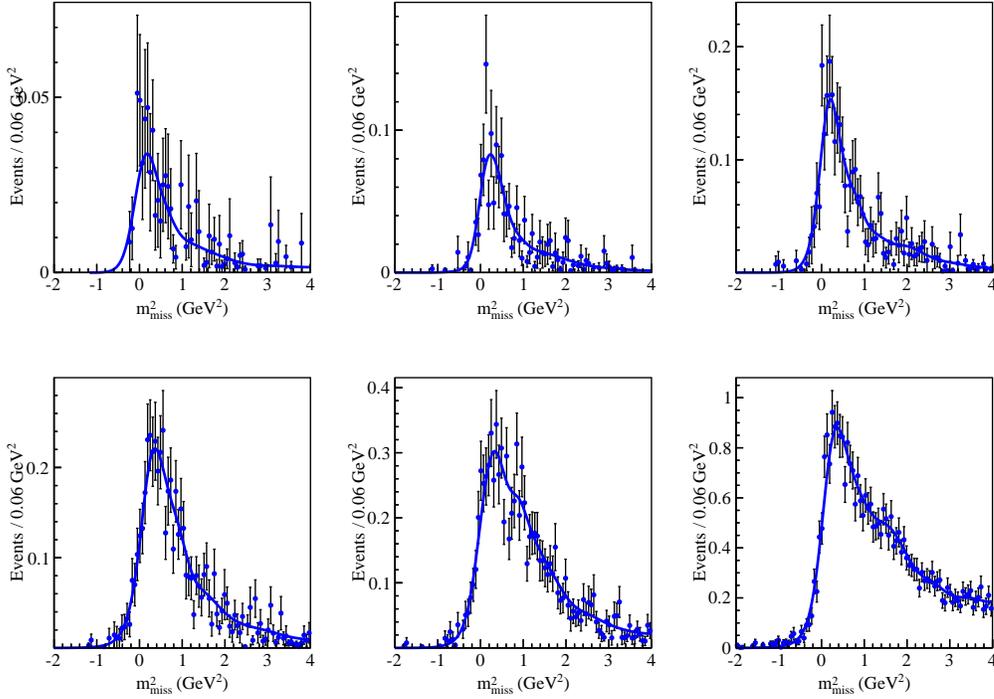
Figure 1.1: m^2_{miss} distribution for the high luminosity samples of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$ MC fitted in bins of the n_{out}. Results show the electron channel of the nominal analysis.
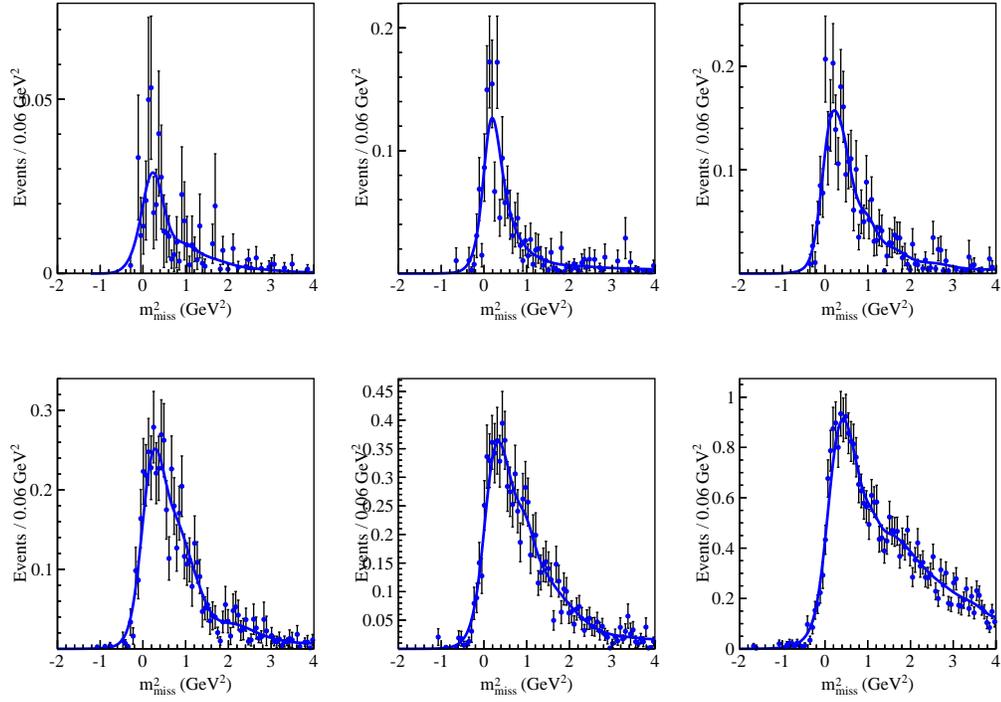
Figure 1.2: m^2_{miss} distribution for the high luminosity samples of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$ MC fitted in bins of the n_{out}. Results show the muon channel of the nominal analysis.
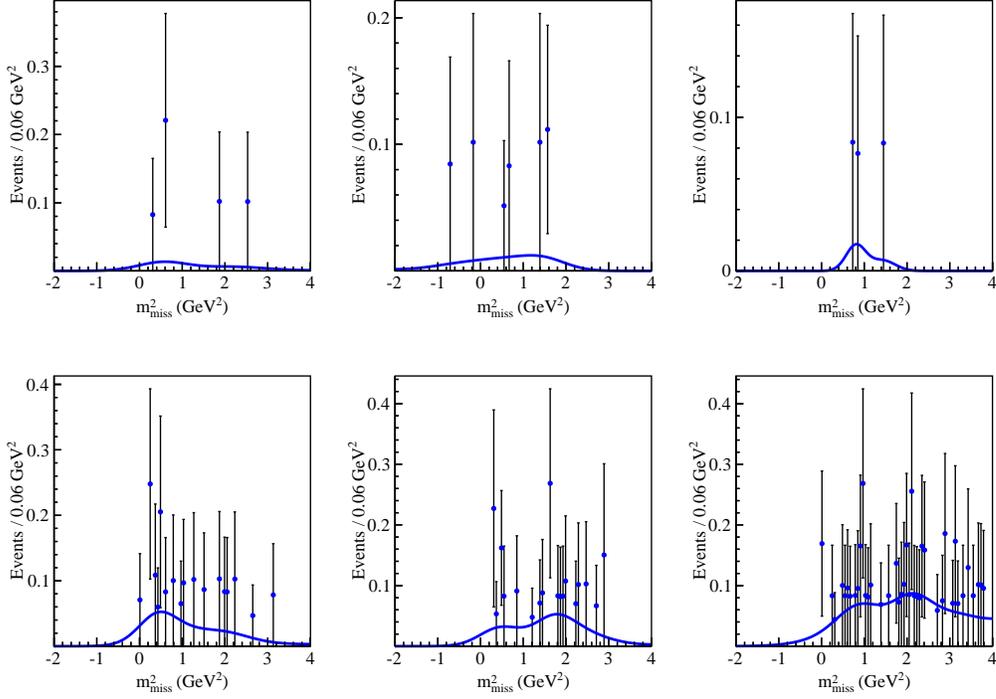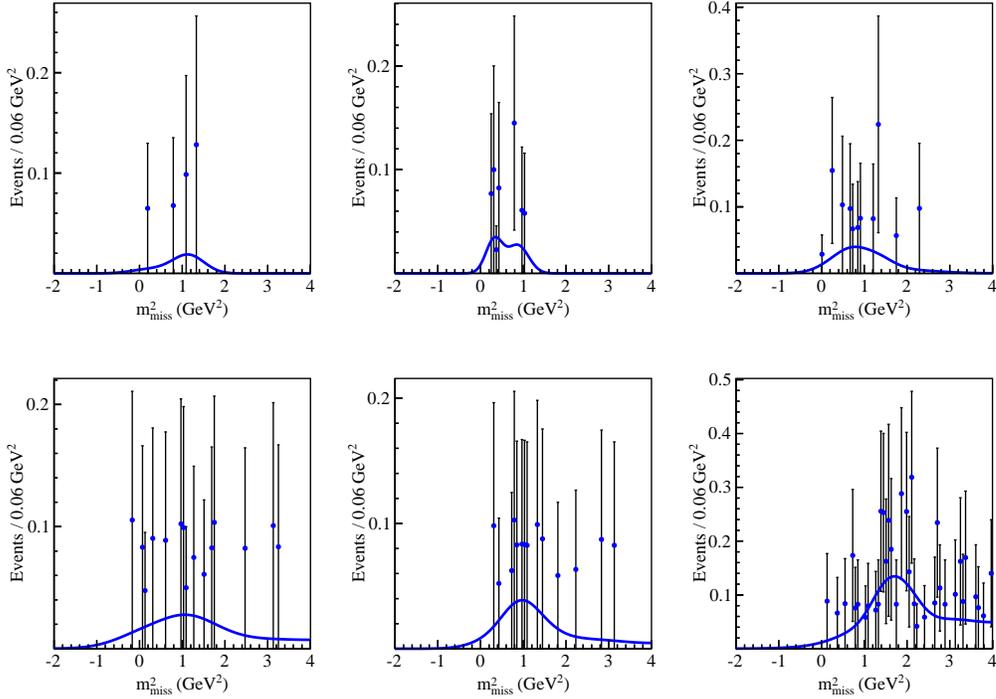
(a) Electron channel.

(b) Muon channel.

Figure 1.3: m^2_{miss} distribution for the low statics channels of the $b \to u\ell^+\nu_\ell$ MC in bins of n_{out} for the nominal analysis.

A.2 Fit bias check for the cut-optimized result
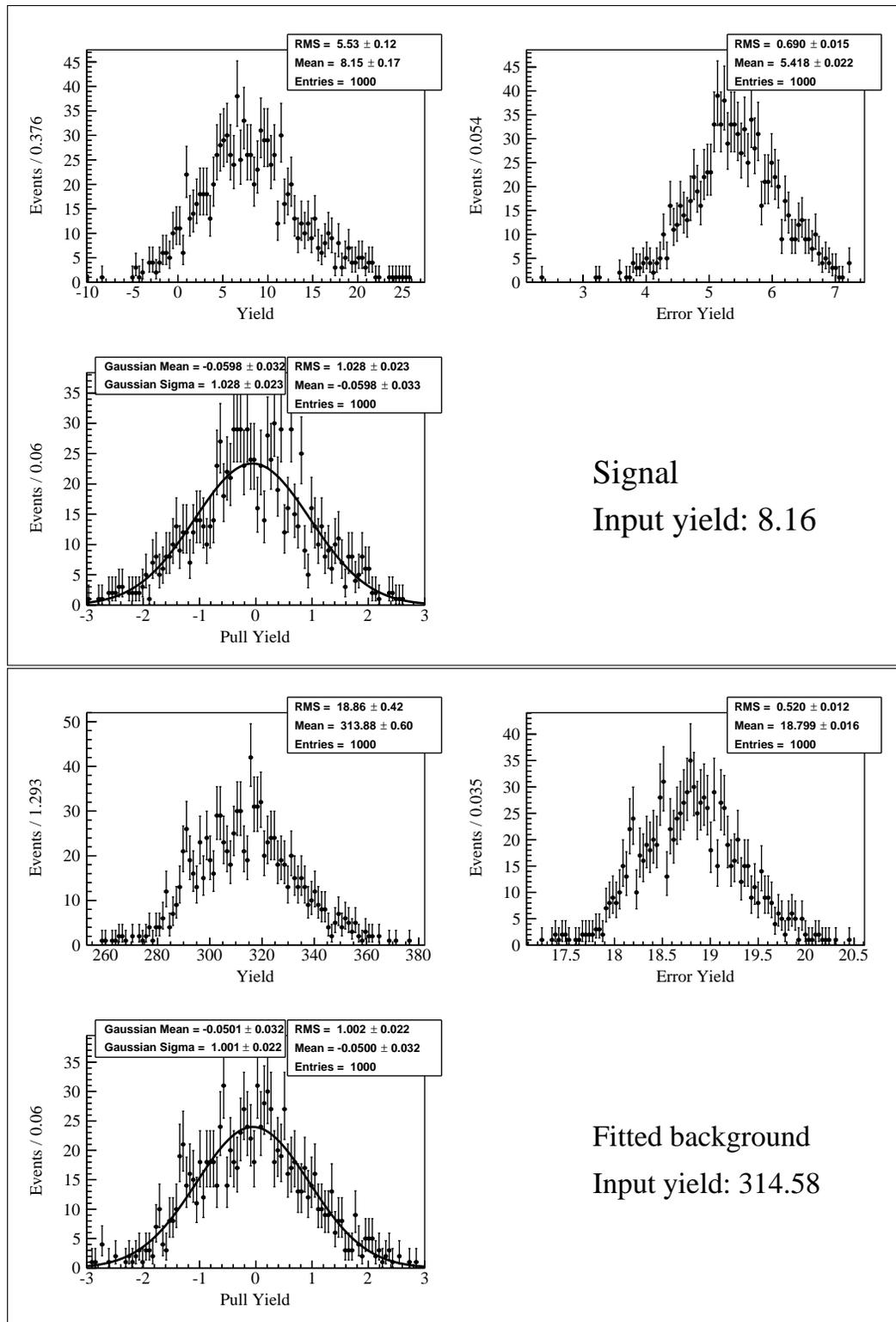
Figure 1.4: Toy study result for the optimal n_{out} selection of the cut-based procedure. Results show the electron channel of the nominal analysis.
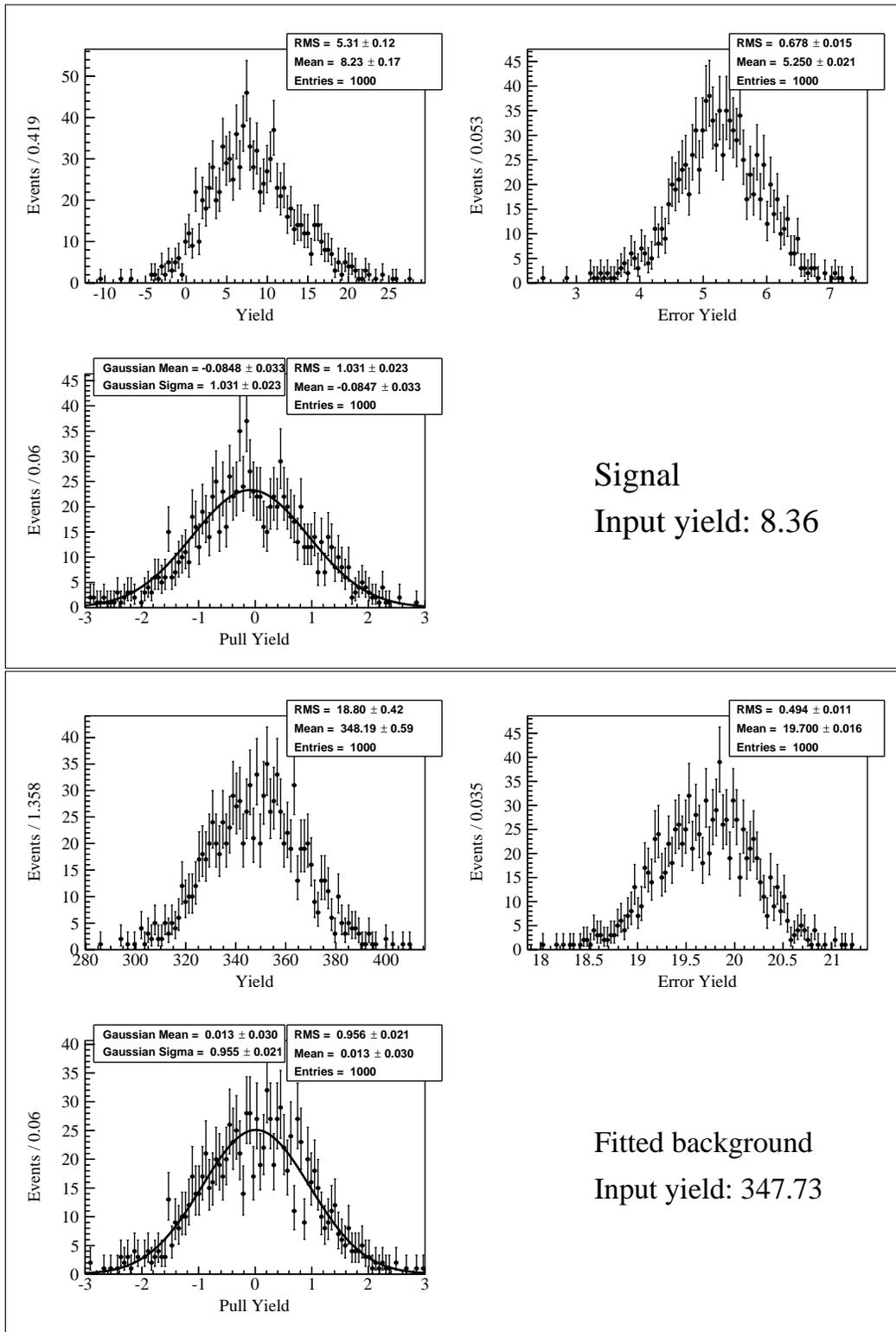
Figure 1.5: Toy study result for the optimal n_{out} selection of the cut-based procedure. Results show the muon channel of the nominal analysis.
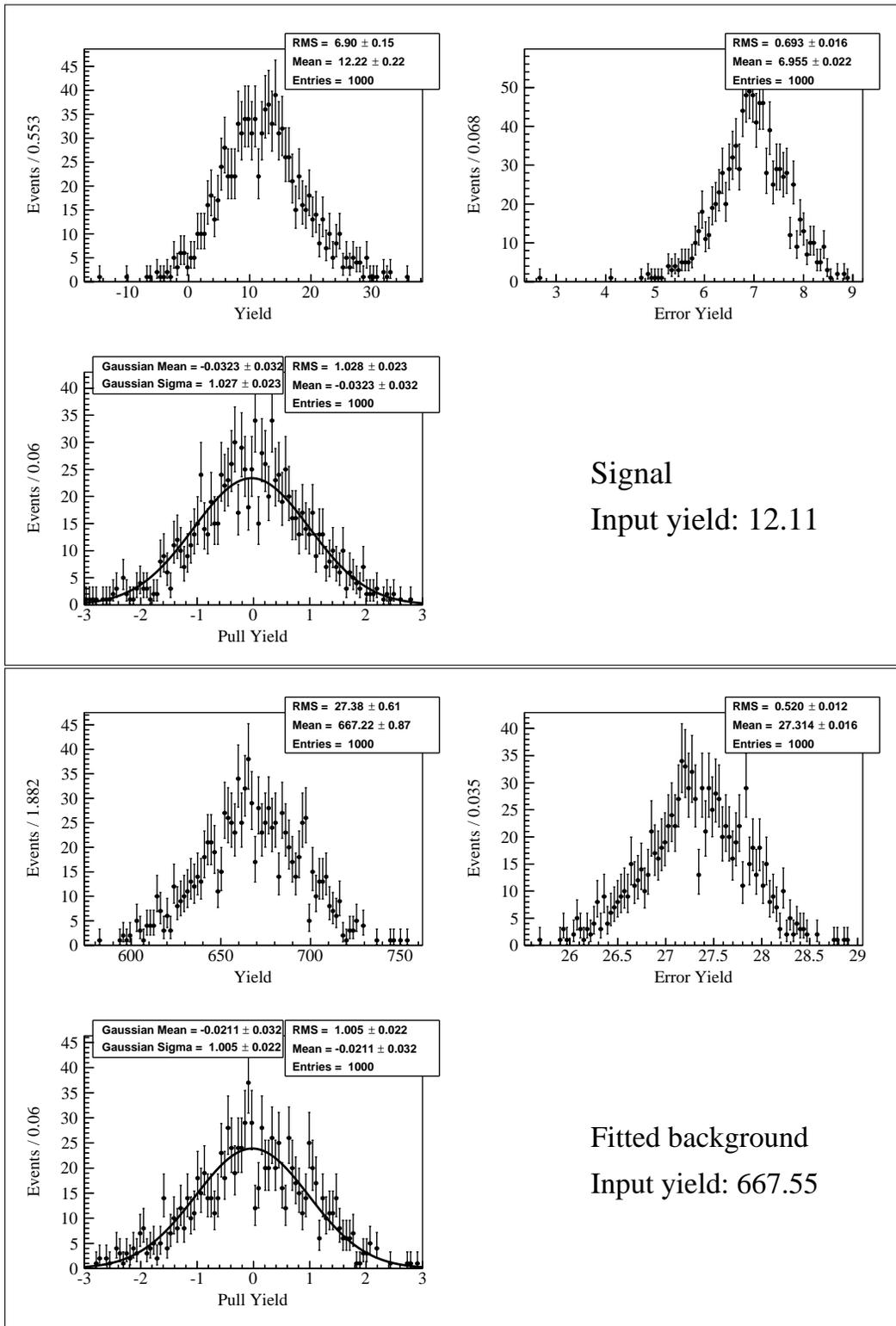
Figure 1.6: Toy study result for the optimal n_{out} selection of the cut-based procedure. Results show the electron channel of the secondary analysis.
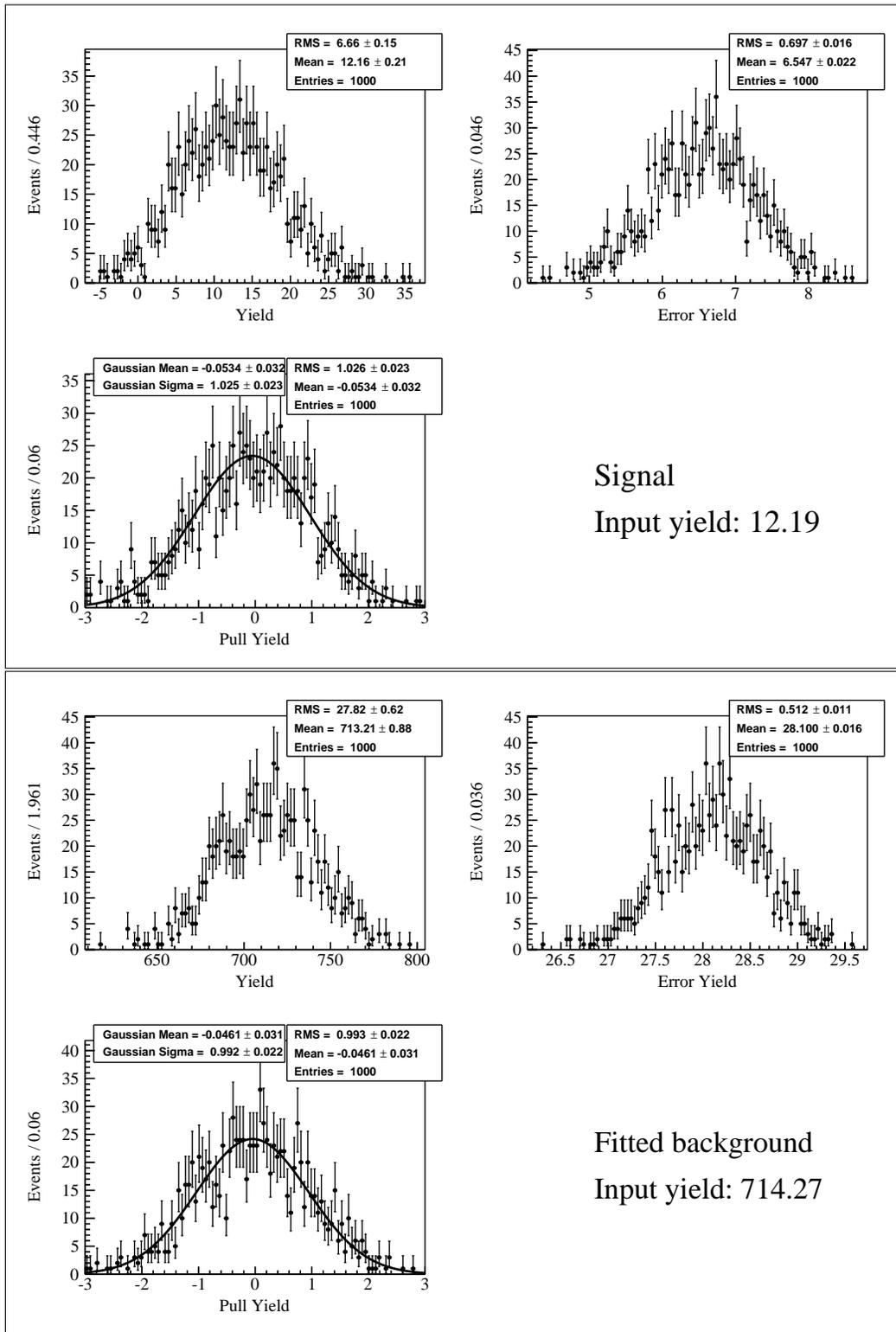
Figure 1.7: Toy study result for the optimal n_{out} selection of the cut-based procedure. Results show the muon channel of the secondary analysis.

A.3 Toy study result for the simultaneous fit
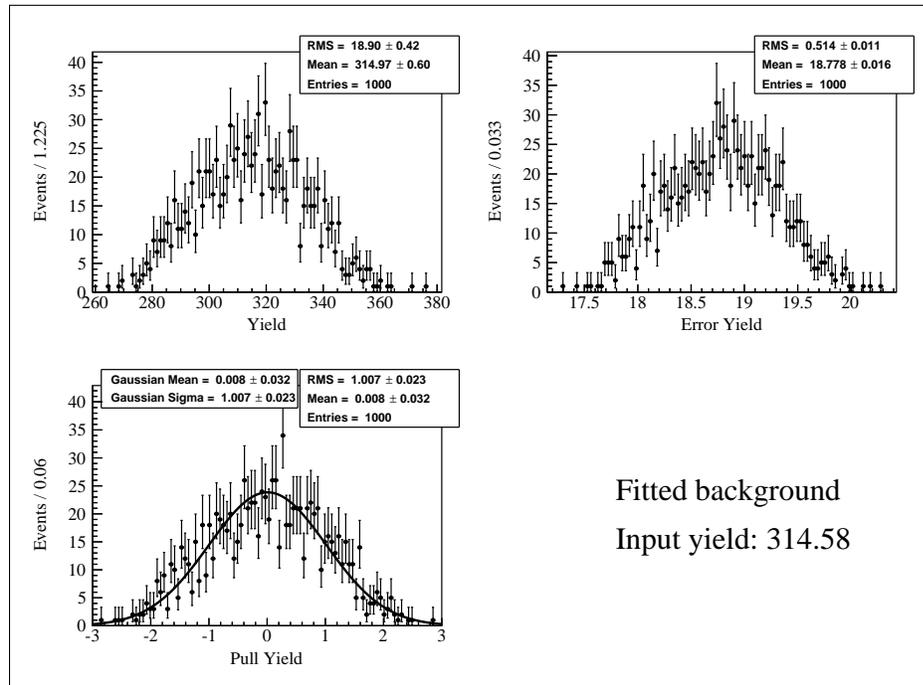
(a) Electron channel.
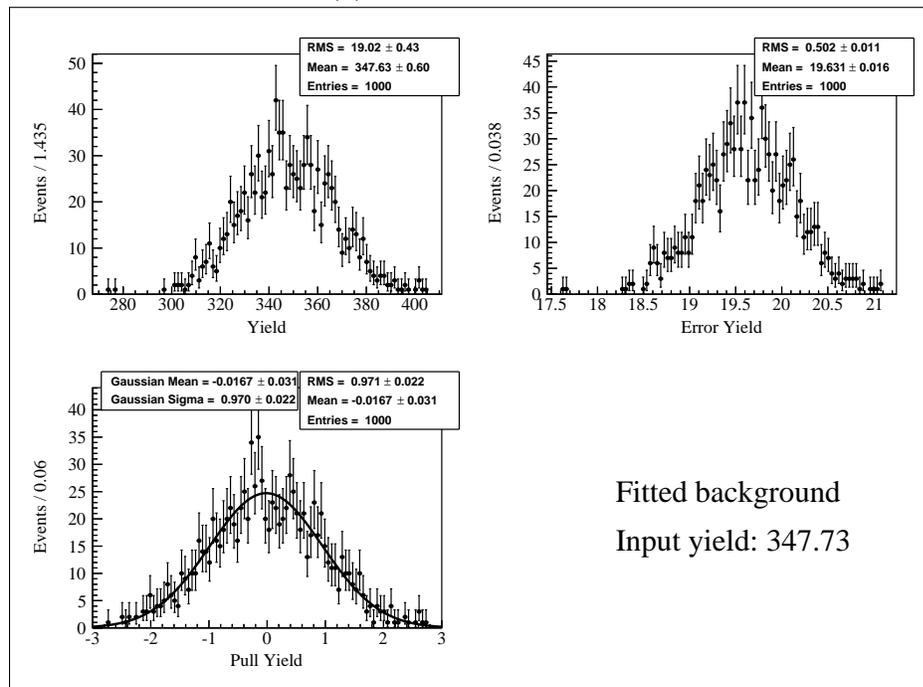
(b) Muon channel.

Figure 1.8: Toy study result of two fitted background yields for the binned fit in the simultaneous fit to both channels.

A.4   Fit results and consistency checks for secondary analysis

Several results of the secondary analysis are gathered in this section and described in the order in which they appear in the analysis.

The check for over-training of the neural network from section 4.3 is shown in Fig 1.9. As for the nominal analysis, a good agreement between training and independent samples is observed.

The binned fit described in section 5.4 is carried out in the same way as for the nominal analysis. The same functions are used to determine the PDFs for the fit components in six bins of the network output. The results are shown for: the signal in Fig. 1.10, the fitted background component Fig. 1.11, the fixed $B \to X_u \ell^+ \nu_\ell$ component in Fig. 1.12, and the combined model Fig. 1.13.

Results for the pure toy MC study for the binned fit of m^2_{miss} are shown in Fig 1.14 and 1.15 for the electron and muon channel, respectively.

(a) Electron channel signal MC. (b) Muon channel signal MC.

(c) Electron channel $b \to u\ell^+\nu_\ell$ MC. (d) Muon channel $b \to u\ell^+\nu_\ell$ MC.
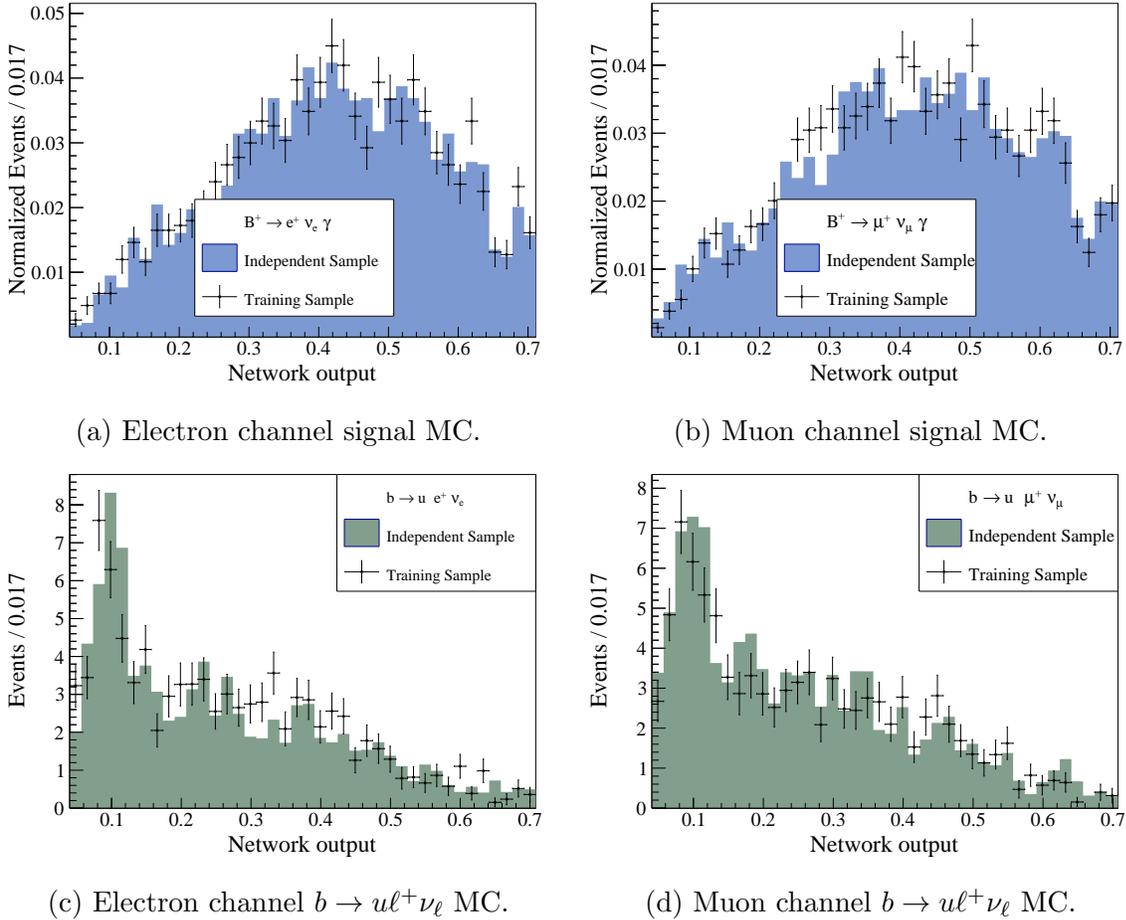
Figure 1.9: Secondary analysis n_{out} comparison between the training and independent sample for the training components of the signal MC and $b \to u\ell^+\nu_\ell$ MC.

133
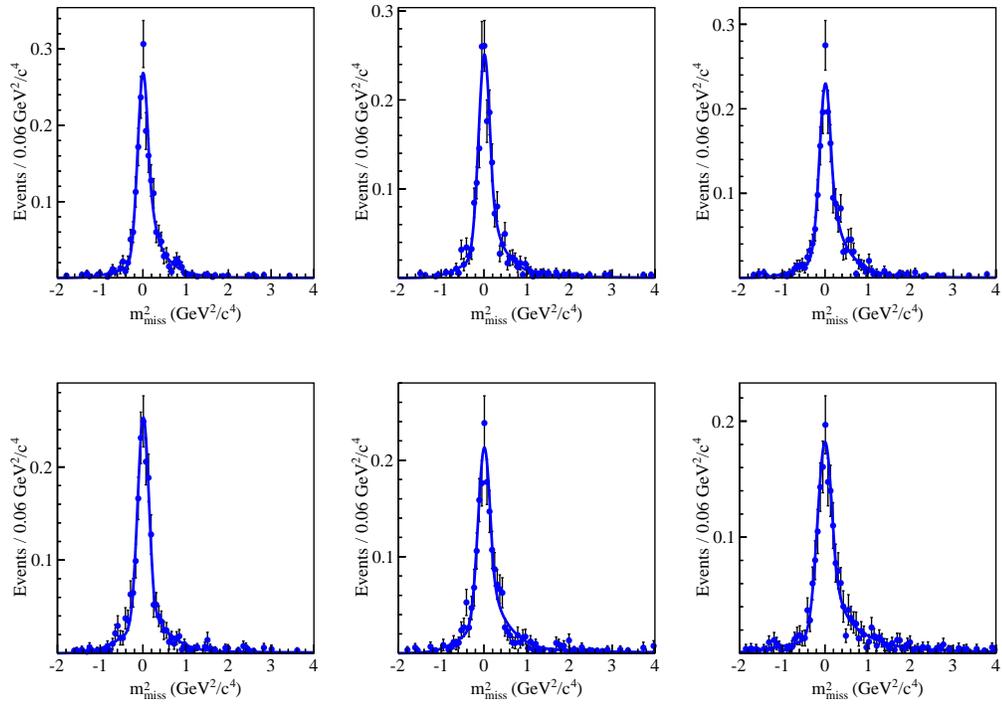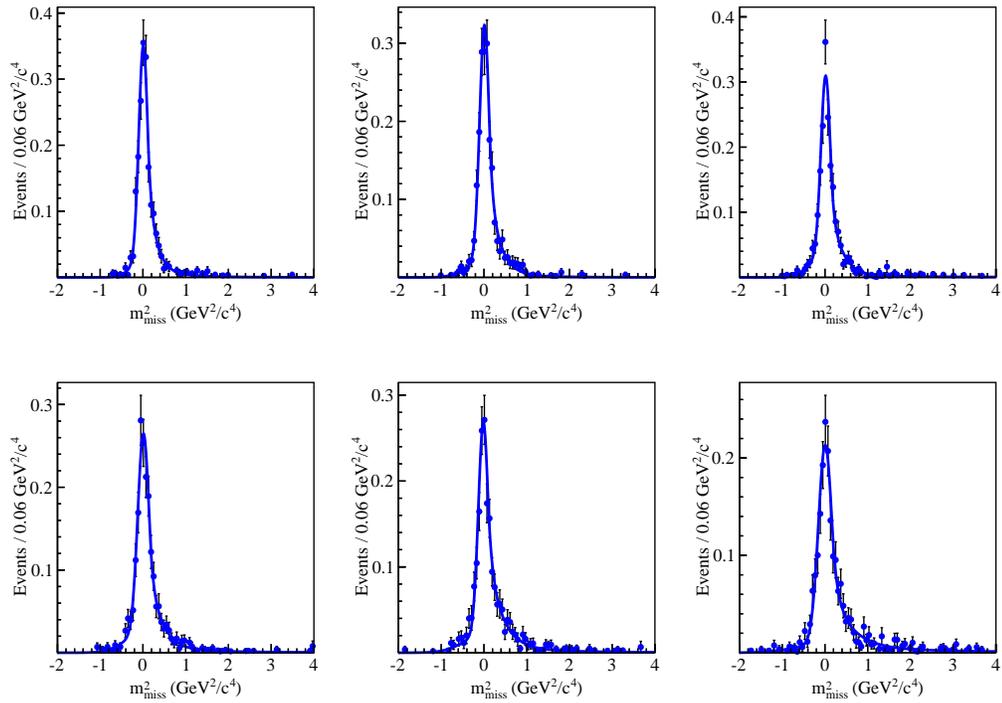
(a) Electron channel.

(b) Muon channel.

Figure 1.10: Signal PDF of the m^2_{miss} fit in six bins of n_{out}. The data points show the signal MC.
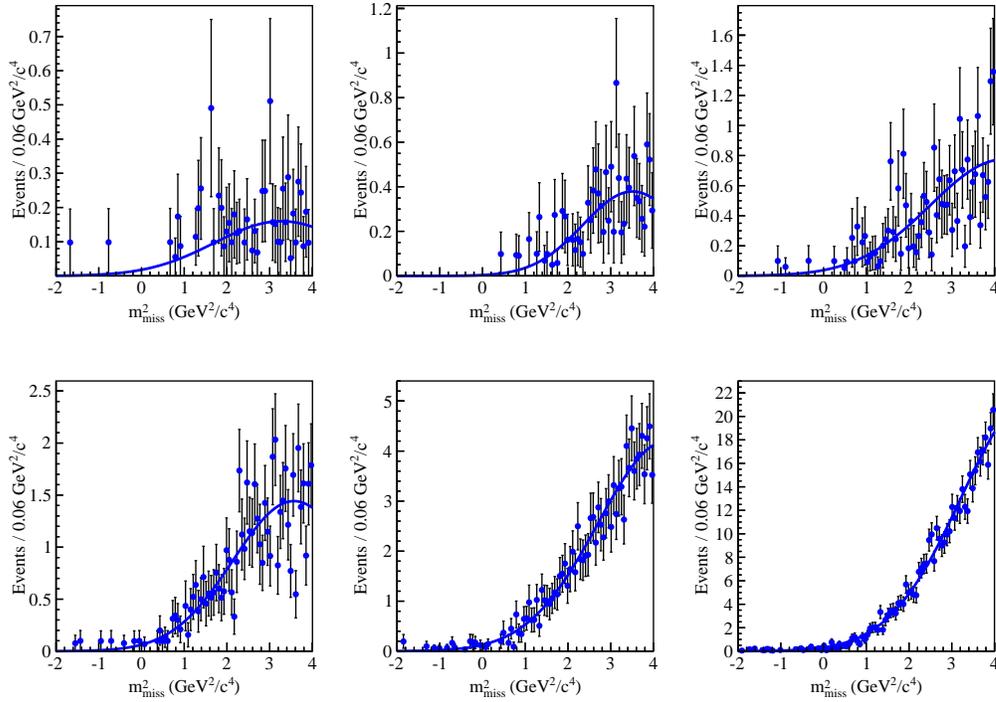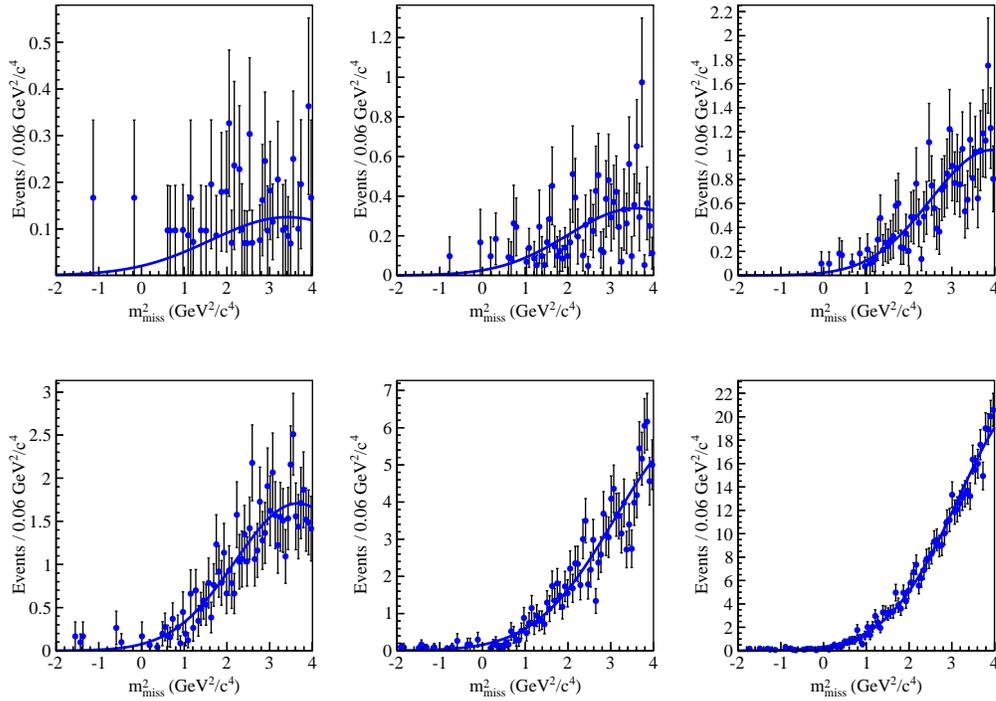
(a) Electron channel.

(b) Muon channel.

Figure 1.11: Fitted background PDF of the m_{miss}^2 fit in six bins of n_{out}. The data points show the generic MC together with badly known decays of the $b \to u\ell^+\nu_\ell$ MC.
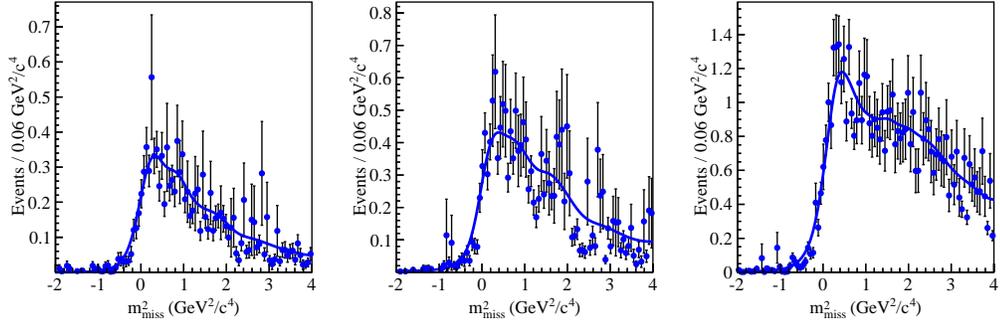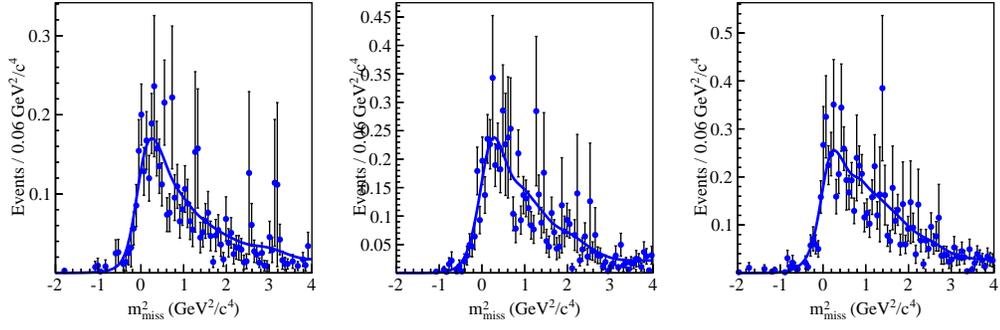
(a) Electron channel.
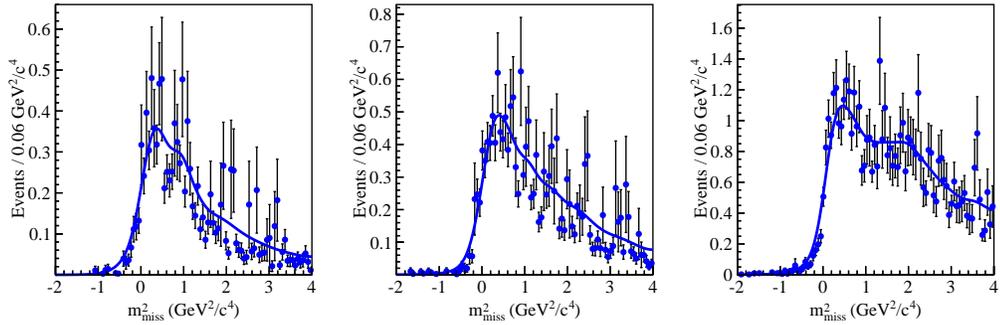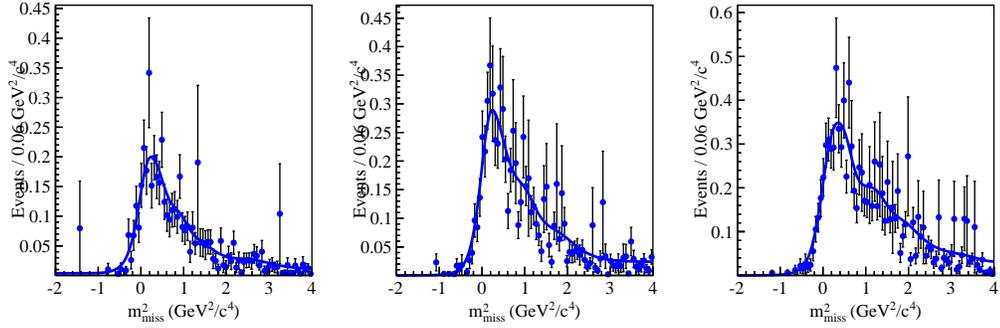
(b) Muon channel.

Figure 1.12: Combined $B \to X_u \ell^+ \nu_\ell$ PDF of the m^2_{miss} fit in six bins of n_{out}. The data points show well known decays of the $b \to u \ell^+ \nu_\ell$ MC together with the high luminosity samples of $B^+ \to \pi^0 \ell^+ \nu_\ell$ and $B^+ \to \eta \ell^+ \nu_\ell$. The differing luminosities lead to non-uniform error bars.
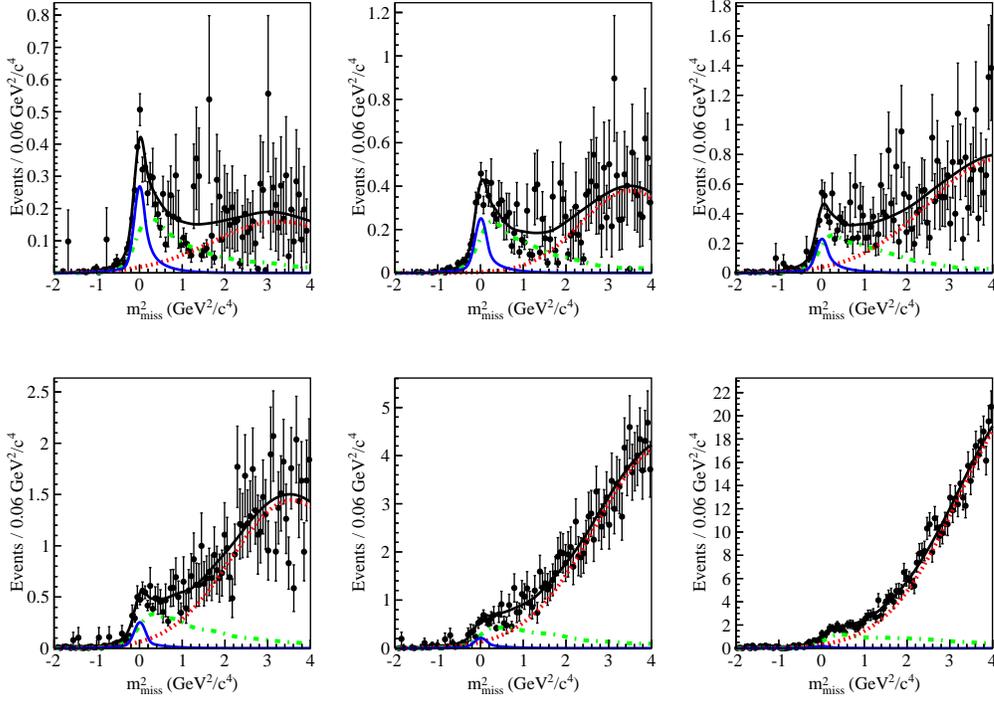
(a) Electron channel.
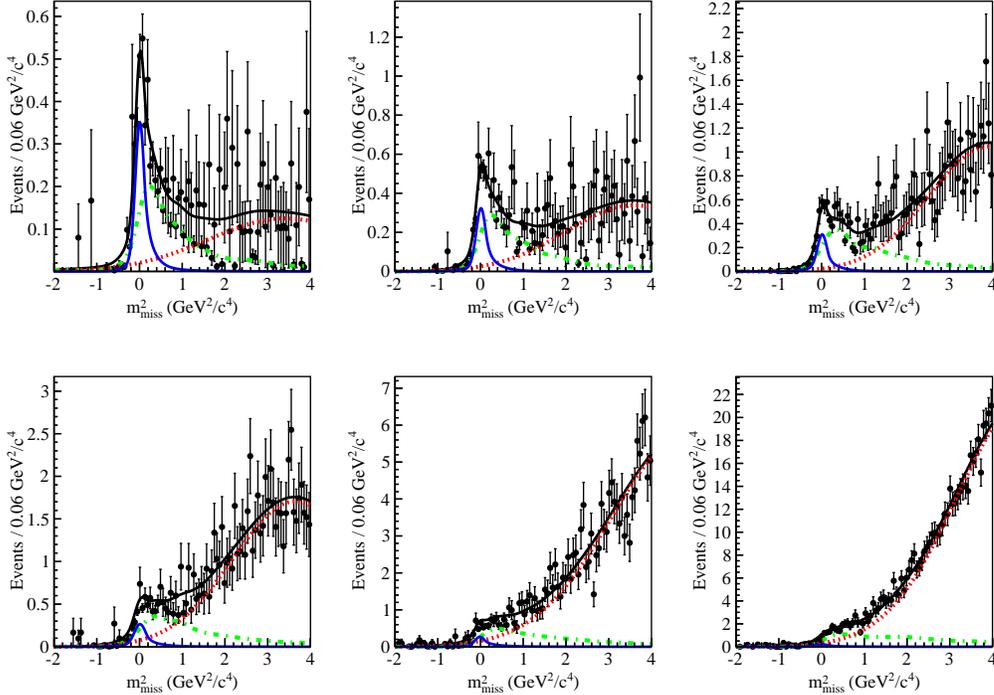
(b) Muon channel.

Figure 1.13: Fit of m^2_{miss} in six bins of n_{out} on MC (points with error bars) for secondary analysis where the samples are weighted to data luminosity. The PDFs show the: signal (blue), fixed $B \to X_u \ell^+ \nu_\ell$ backgrounds (green), fitted backgrounds (red), and the sum (black). The error bars on the data points represent the luminosity of the MC samples and not the expected error on data. The signal is normalized to 5×10^{-6}. The most signal-like bin is found in the upper left plot. The most background-like bin is shown in the lower right panel.
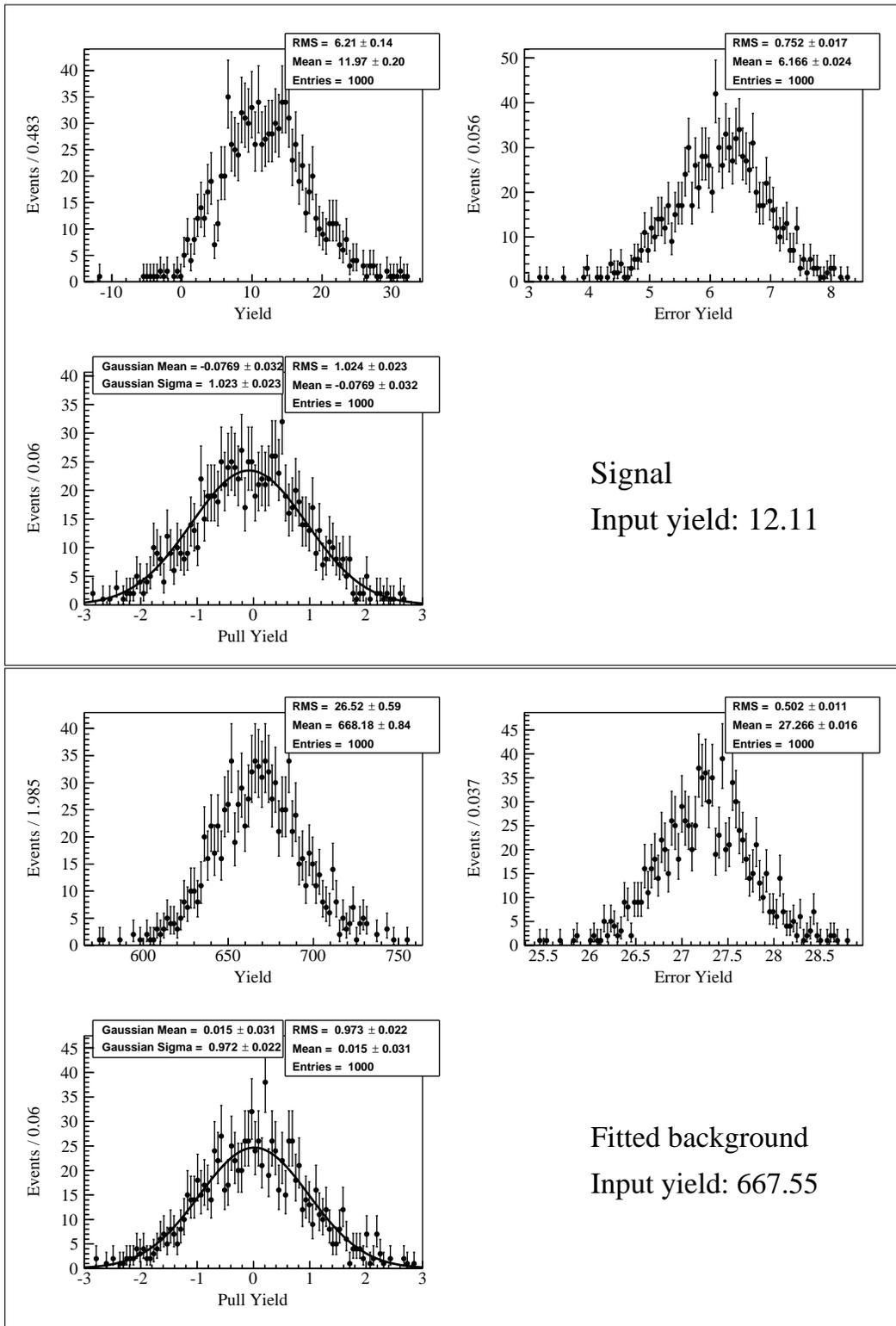
Figure 1.14: Toy study result for the fit in six bins of n_{out} for the electron channel
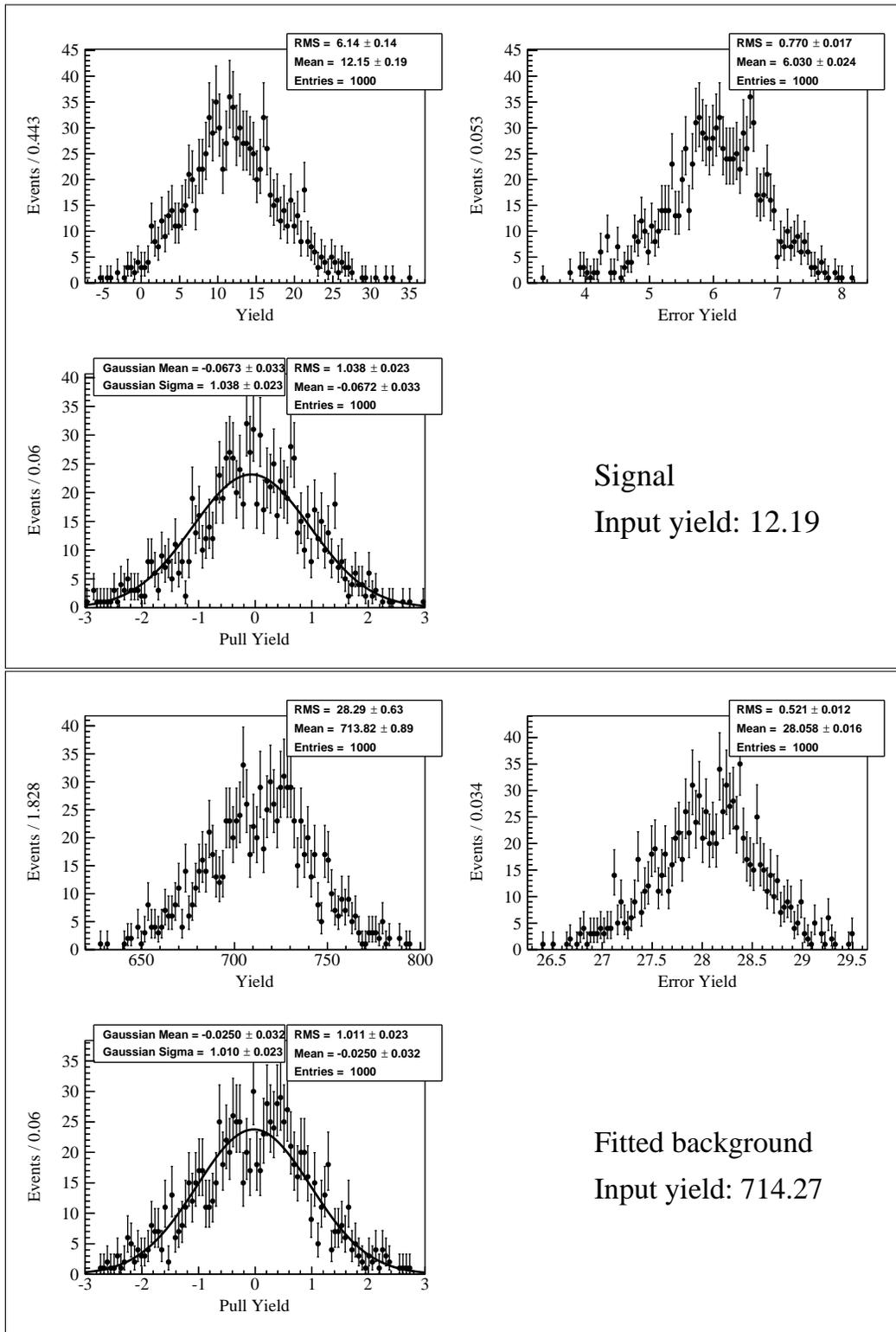
Figure 1.15: Toy study result for the fit in six bins of n_{out} for the muon channel

A.5   Toy MC study sampled from MC
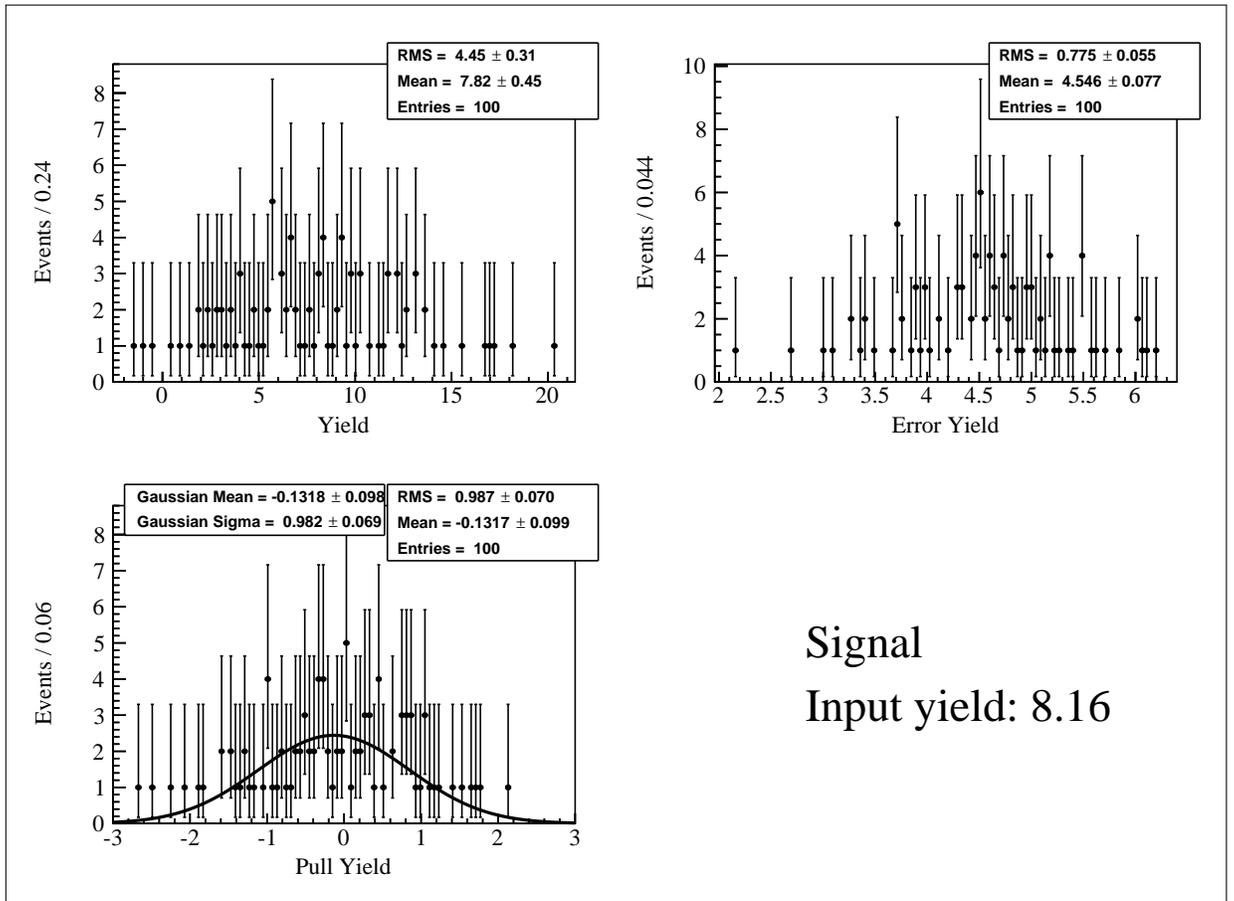
Figure 1.16: Toy study sampled from MC for the m^2_{miss} fit in six bins for the electron channel of the nominal analysis.
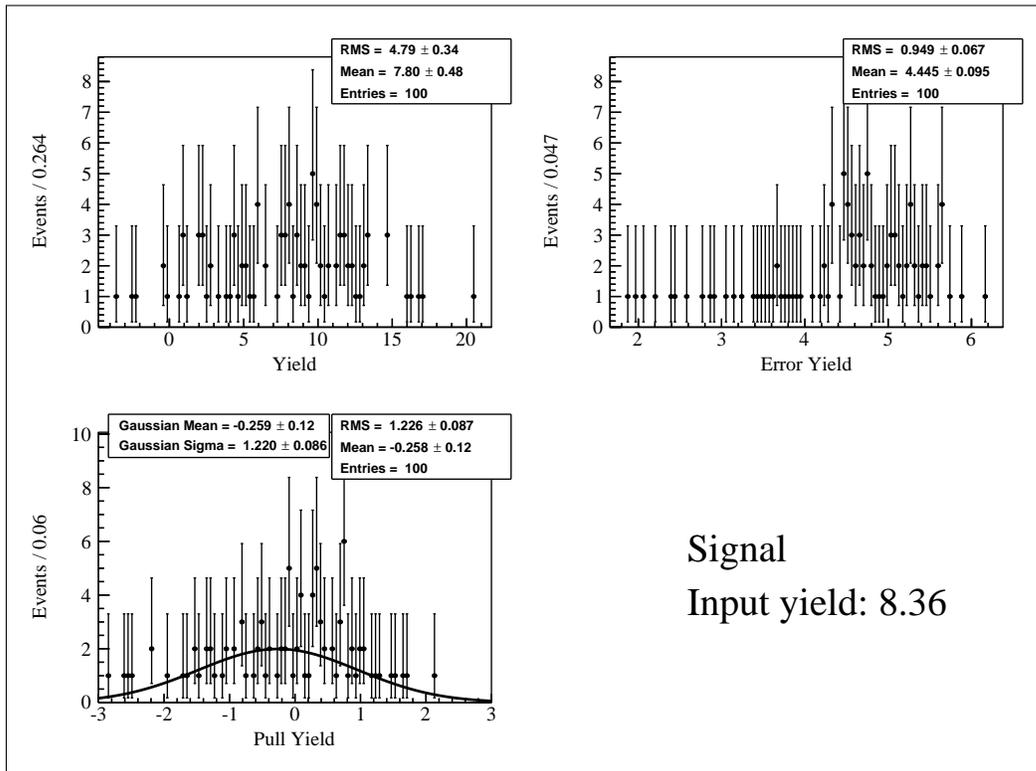
(a) Muon channel.
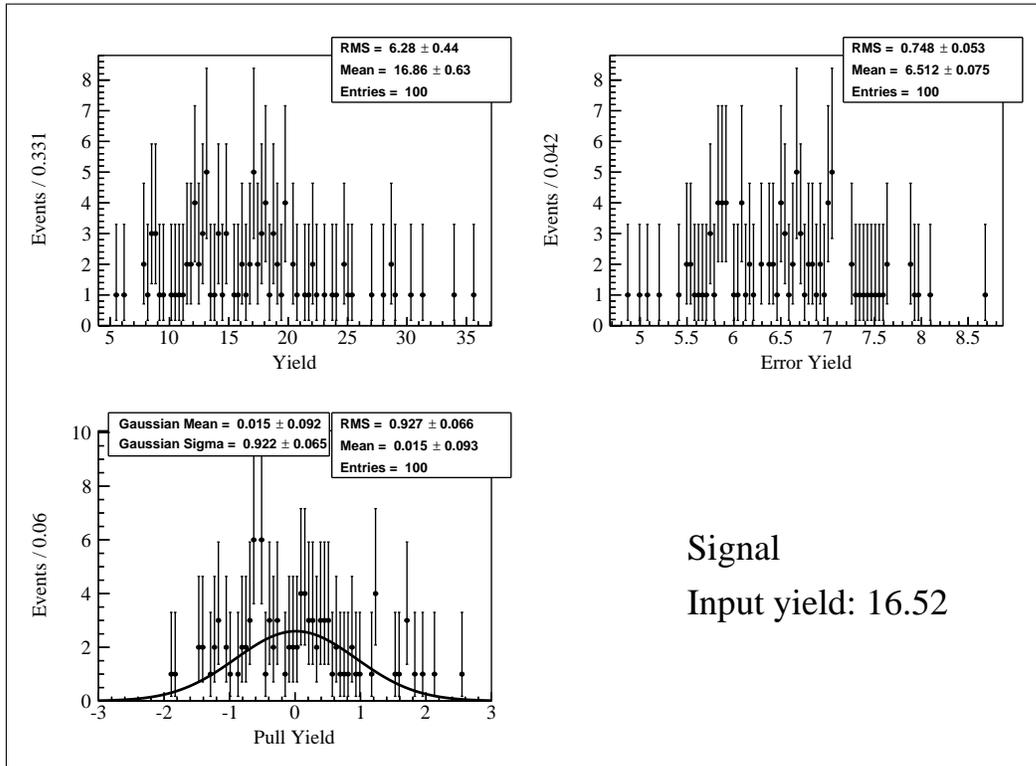
(b) Simultaneous fit.

Figure 1.17: Toy study sampled from MC for the m^2_{miss} fit in six bins for the nominal analysis.

A.6 Data-MC comparison of the input variables for network

The network variables used for the training in section 4.3 are plotted in the signal region for the nominal analysis. No discrepancy is found for any of the variables which agrees with the observation for network output shown in section 8.

(a) Electron channel.
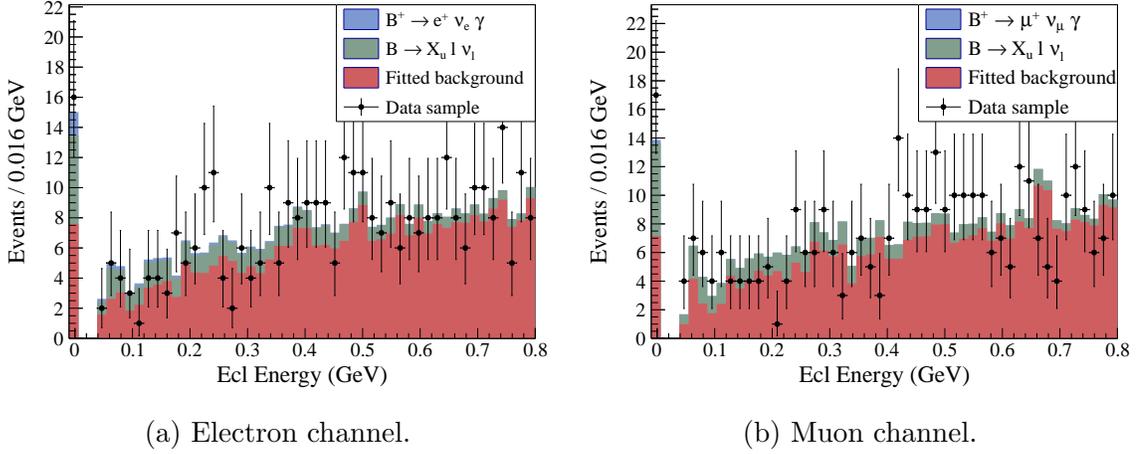
(b) Muon channel.

Figure 1.18: Extra energy in the ECL in the signal region.

(a) Electron channel.
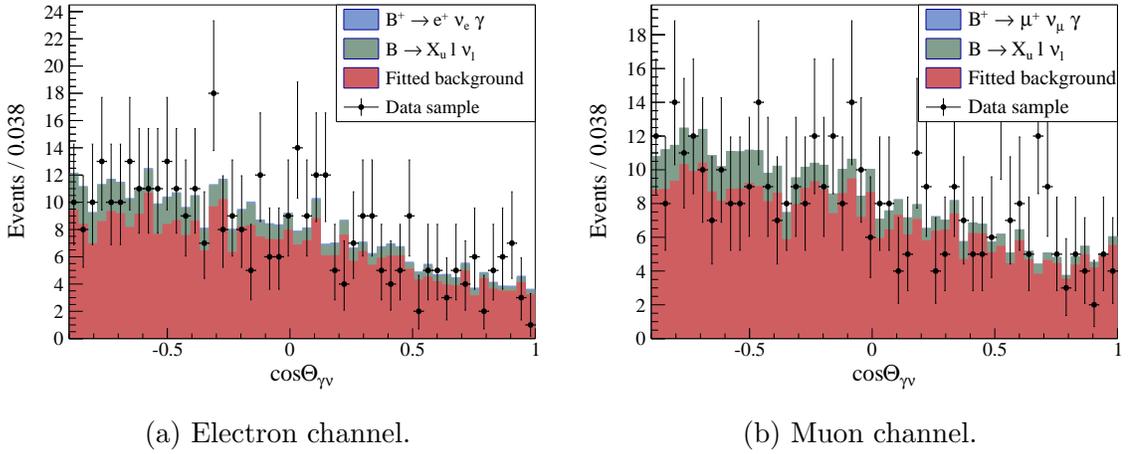
(b) Muon channel.

Figure 1.19: $\cos\Theta_{\gamma\nu}$ angle in the signal region.
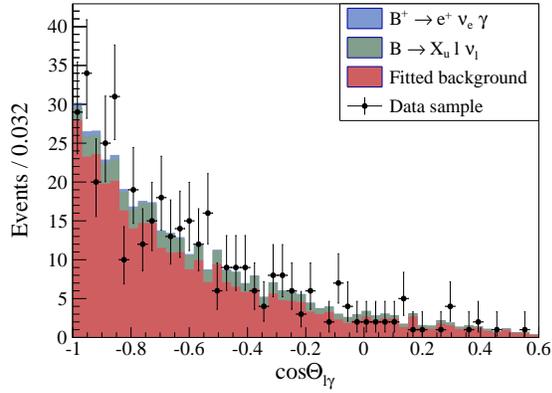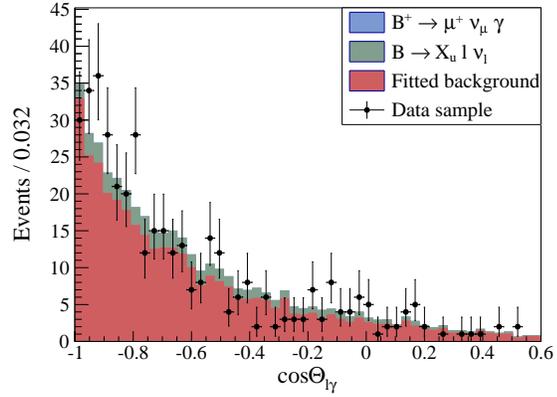
(a) Electron channel.

(b) Muon channel.

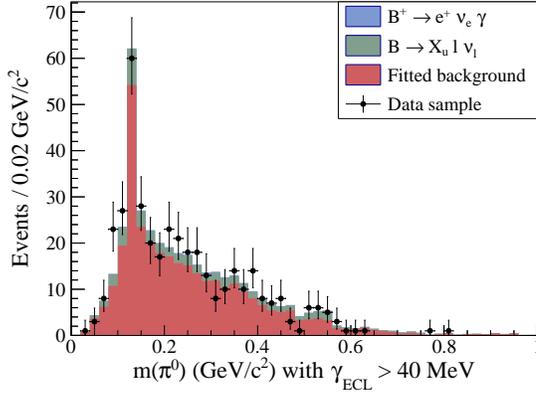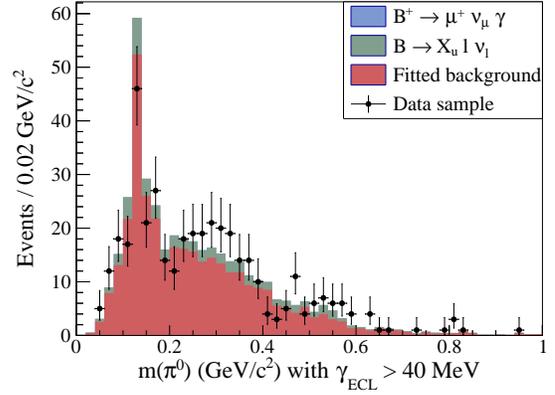Figure 1.20: $\cos\Theta_{\gamma\ell}$ angle in the signal region.
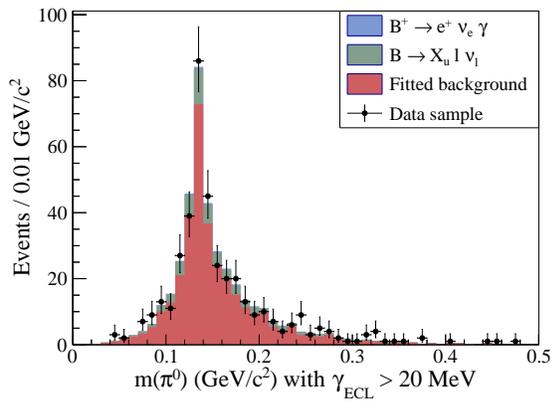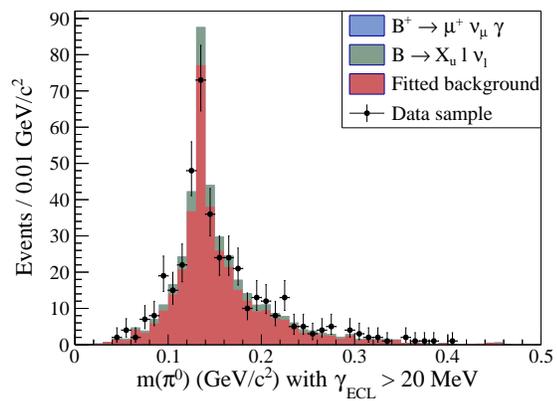
(a) Electron channel.

(b) Muon channel.

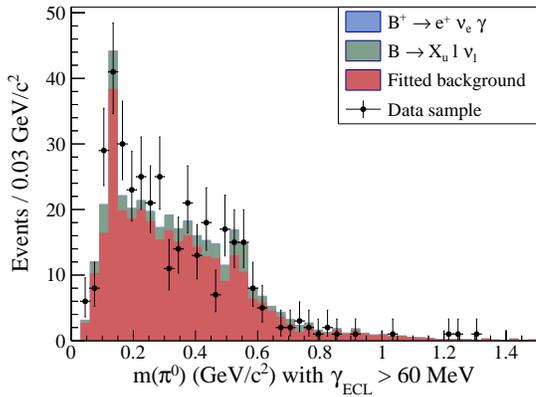Figure 1.21: $m(\pi^0)$ computed with ECL photons above $20\,\mathrm{MeV}$ in the signal region.
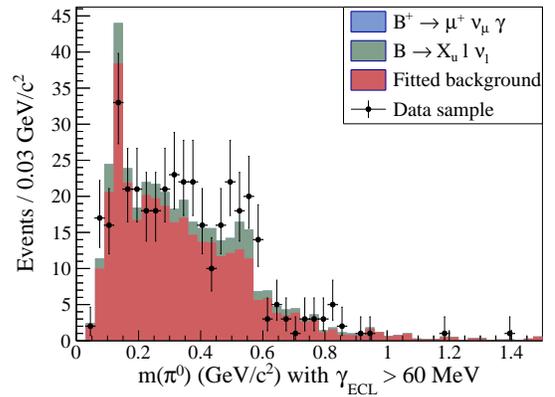
(a) Electron channel.

(b) Muon channel.

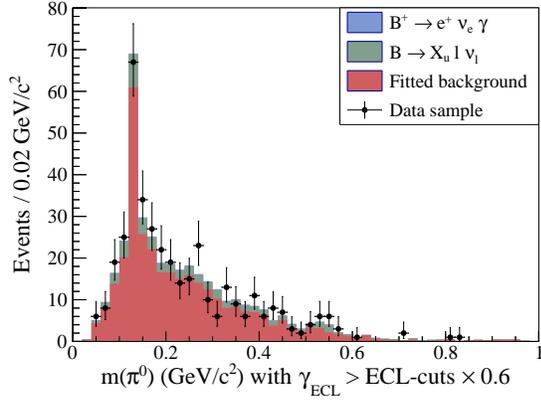Figure 1.22: $m(\pi^0)$ computed with ECL photons above $40\,\text{MeV}$ in the signal region.
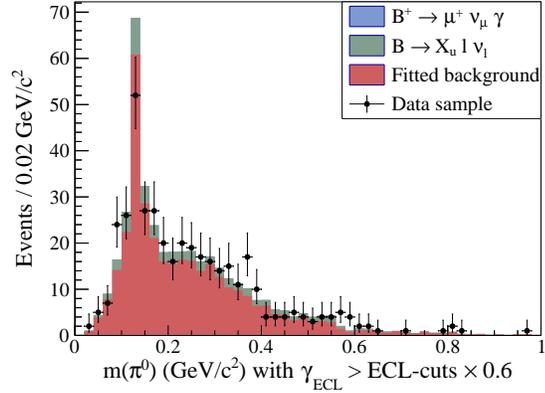

(a) Electron channel.

(b) Muon channel.

Figure 1.23: $m(\pi^0)$ computed with ECL photons above $60\,\text{MeV}$ in the signal region.
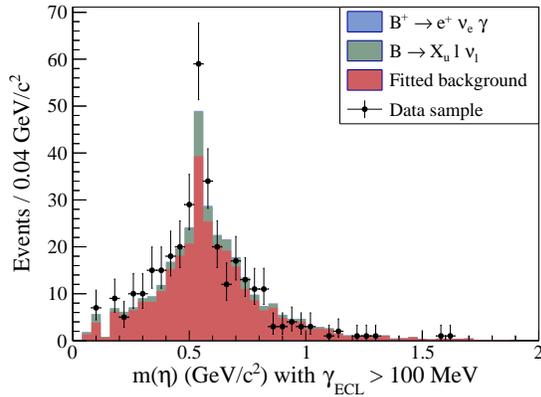
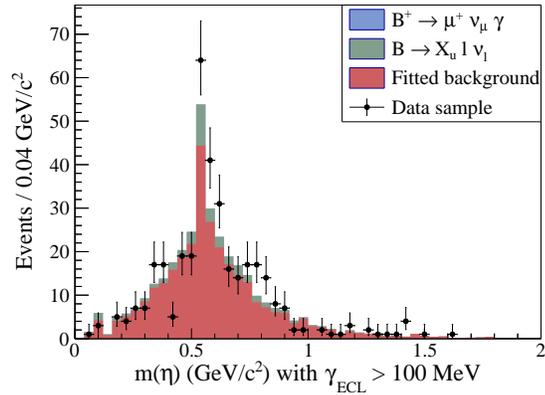(a) Electron channel.

(b) Muon channel.

Figure 1.24: m(π^0) with ECL cuts scaled by 0.6 in the signal region.
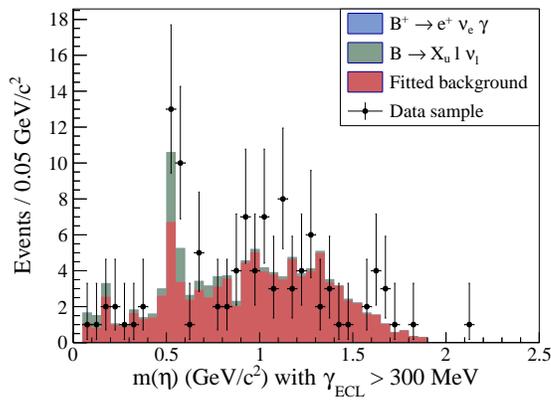

(a) Electron channel.

(b) Muon channel.
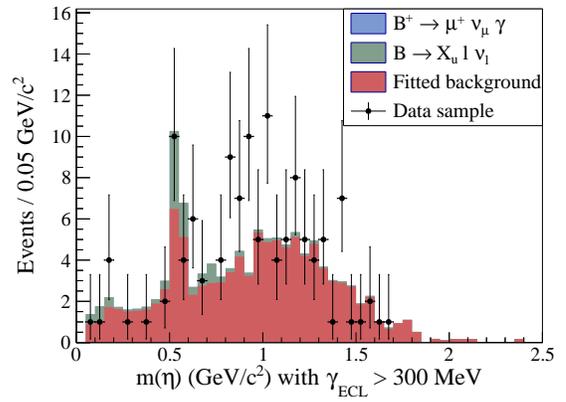
Figure 1.25: m(η) computed with ECL photons above $100\,\mathrm{MeV}$ in the signal region.

145

(a) Electron channel.

(b) Muon channel.

Figure 1.26: m(η) computed with ECL photons above $300\,\text{MeV}$ in the signal region.

References

[1] M. Beneke and G. Buchalla and M. Neubert and C.T. Sachrajda, "QCD factorization for exclusive non-leptonic B-meson decays: general arguments and the case of heavylight final states," *Nuclear Physics B* **591** no. 12, (2000) 313 – 418.

[2] T. Feldmann, "Non-Leptonic Heavy Meson Decays - Theory Status," `arXiv:1408.0300 [hep-ph]`.

[3] G. Bell and V. Pilipp, "$B^- \to \pi^-$, π^0/ρ^-, ρ^0 to NNLO in QCD factorization," *Phys.Rev.* **D80** (2009) 054024, `arXiv:0907.1016 [hep-ph]`.

[4] G. Bell, "Charmless hadronic B decays: theory status," CKM Conference 2014.

[5] M. Sevior *et al.* (Belle Collaboration), "Updated measurement of the Branching Ratio and A_{CP} for $B^0 \to \pi^0\pi^0$ and improved constraint on ϕ_2." Belle paper draft (Belle Note 1291), 2015.

[6] J.P. Lees *et al.* (BaBar Collaboration), "Measurement of CP Asymmetries and Branching Fractions in Charmless Two-Body B-Meson Decays to Pions and Kaons," *Phys.Rev.* **D87** no. 5, (2013) 052009, `arXiv:1206.3525 [hep-ex]`.

[7] B. Aubert *et al.* (BaBar Collaboration), "Model-independent search for the decay $B^+ \to \ell^+\nu_l\gamma$," *Phys. Rev. D* **80** (Dec, 2009) 111105.

[8] J. D. Bjorken, "Topics in B-physics," *Nuclear Physics B - Proceedings Supplements* **11** no. 0, (1989) 325 – 341.

[9] M .Beneke and J. Rohrwild, "B meson distribution amplitude from $B \to \gamma\ell\nu$," *Eur. Phys. J. C* **71** no. 12, (2011) 1818.

[10] V.M. Braun and A. Khodjamirian, "Soft contribution to $B \to \gamma\ell\nu_\ell$ and the B-meson distribution amplitude," *Physics Letters B* **718** no. 3, (2013) 1014 – 1019.

[11] G. P. Korchemsky and D. Pirjol and T. M. Yan, "Radiative leptonic decays of B mesons in QCD," *Phys. Rev. D* **61** (May, 2000) 114510.

[12] A. Abashian *et al.* (Belle Collaboration), "The Belle detector," *Nuclear instruments & methods in physics research A* **479** no. 1, (2002) 117 – 232.

[13] J. Brodzicka *et al.*, "Achievements of KEKB," *Progress of Theoretical and Experimental Physics* **04** no. 04D001, (2012) .

[14] S. Kurokawa and E. Kikutani, "Overview of the KEKB accelerators," *Nuclear instruments & methods in physics research A* **499** no. 1, (2003) 1 – 7.

[15] T. Abe *et al.*, "Achievements of KEKB," *Progress of Theoretical and Experimental Physics* **03** no. 03A001, (2013) .

[16] K Hanagaki and H Kakuno and H Ikeda and T Iijima and T Tsukamoto, "Electron identification in Belle," *Nuclear instruments & methods in physics research A* **485** no. 3, (2002) 490 – 503.

[17] A.J. Bevan and others (BaBar and Belle Collaboration), "The Physics of the B Factories," *Eur.Phys.J.* **C74** no. 11, (2014) 3026.

[18] A Abashian *et al.*, "Muon identification in the Belle experiment at KEKB," *Nuclear instruments & methods in physics research A* **491** no. 12, (2002) 69 – 82.

[19] David J. Lange, "The EvtGen particle decay simulation package," *Nuclear instruments & methods in physics research A* **462** no. 12, (2001) 152 – 155.

[20] R. Brun *et al.*, "GEANT 3," Report DD/EE/84-1, CERN, 1984.

[21] Heavy Flavor Averaging Group (HFAG), 2014.
http://www.slac.stanford.edu/xorg/hfag/semi/index.html.

[22] G. C. Fox and S. Wolfram, "Observables for the analysis of event shapes in e^+e^- annihilation and other processes," *Phys. Rev. Lett.* **41** (Dec, 1978) 1581–1585.

[23] S. H. Lee *et al.* (Belle Collaboration), "Evidence for $B^0 \rightarrow \pi^0\pi^0$," *Phys. Rev. Lett.* **91** (Dec, 2003) 261801.

[24] D. M. Asner *et al.* (CLEO Collaboration), "Search for exclusive charmless hadronic B decays," *Phys. Rev. D* **53** (Feb, 1996) 1039–1050.

[25] M. Prim, "Neural network based B^0 flavor tagging at the Belle experiment," diploma thesis, Karlsruhe Institute of Technology (KIT), Institut für experiementelle Kernphysik (EKP), 2009.

[26] M. Feindt *et al.*, "A hierarchical NeuroBayes-based algorithm for full reconstruction of B mesons at B factories," *Nuclear instruments & methods in physics research A* **654** no. 1, (2011) 432 – 440.

[27] M. Feindt and U. Kerzel, "The NeuroBayes neural network package," *Nuclear instruments & methods in physics research A* **559** no. 1, (2006) 190 – 194.

[28] A. Zupanc, "Truth matching," Belle Analysis School, 2011.

[29] M. Prim and M. Feindt, "An algorithm for quantifying dependence in multivariate data sets," *Nuclear instruments & methods in physics research A* **698** (2013) 84 – 89. Preprint: http://arxiv.org/abs/1207.0981.

[30] P. Koppenburg, "An improved π^0 and η veto." Belle Note 665, 2004.

[31] A. Sibidanov *et al.* (Belle Collaboration), "Study of exclusive $B \to X_u \ell \nu$ decays and extraction of $|V_{ub}|$ using full reconstruction tagging at the Belle experiment," *Phys. Rev. D* **88** (Aug, 2013) 032005.

[32] L. Hinz, "Lepton ID correction and systematic error." Belle Note 954, 2006.

[33] W. Verkerke and D. P. Kirkby, "The RooFit toolkit for data modeling," *eConf* **C0303241** (2003) MOLT007, `arXiv:physics/0306116 [physics]`.

[34] Particle Data Group K.A. Olive *et al.*, "Review of Particle Physics," *Chin.Phys.* **C38** (2014) 090001.

[35] T. Skwarnicki report/Ph.D. thesis, Institute for Nuclear Physics, Krakow, 1986.

[36] K. S. Cranmer, "Kernel estimation in high-energy physics," *Comput.Phys.Commun.* **136** (2001) 198–207.

[37] D. Dutta and B. Bhuyan and K. Trabelsi, "Search for the decay $B_s \to \gamma\gamma$." Belle Note 1250v2, 2013.

[38] H. Albrecht *et al.* (ARGUS Collaboration), "Search for hadronic $b \to u$ decays," *Physics Letters B* **241** no. 2, (1990) 278 – 282.

[39] M. Pivk and F.R. Le Diberder, ": A statistical tool to unfold data distributions," *Nuclear instruments & methods in physics research A* **555** no. 12, (2005) 356 – 369.

[40] B. Bhuya, "High P_T tracking efficiency using partially reconstructed D^* decays." Belle Note 1165, 2010.

[41] A. Heller *et al.* (Belle Collaboration), "Search for $B^+ \to \ell^+ \nu_\ell \gamma$ decays with hadronic tagging using the full Belle data sample," *Phys. Rev. D* **91** (Jun, 2015) 112009, `arXiv:1504.05831 [hep-ex]`.

Danksagung

Diese Arbeit wäre ohne die Hilfe zahlreicher Personen nicht möglich gewesen.

An erster Stelle danke ich Prof. Michael Feindt, der mich in seine Arbeitsgruppe aufgenommen und mich nach Kräften unterstützt hat. Sein Einsatz und die gute Arbeitsgruppe die er aufgebaut hat, waren maßgeblich für die erfolgreiche Ausführung dieser Arbeit.

Prof. Thomas Müller danke ich sehr für die Übernahme des Korreferats.

Auch möchte ich der B-Gruppe danken, welche mit breiter Fachkenntnis immer behilflich war. Im Besonderen danke ich hierbei den Post-Docs: Pablo Goldenzweig, Martin Heck, Thomas Kuhr und Anže Zupanc. Jeder von Ihnen hat wichtige Beiträge zu dieser Arbeit geleistet.

Im weiteren danke ich allen Admins und im besonderen den Core-Admins, ohne deren Einsatz die komplexe IT-Infrastruktur des Instituts nicht zu betreiben wäre, welche essenziell in der Hochenergiephysik ist.

Ich danke meiner Familie für die immerwährende Unterstützung.