Full Reconstruction And Search For Charged Higgs Effects In Semi-Tauonic B Decays

Daniel Zander

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
von der Fakultät für Physik des
KIT – Karlsruher Institut für Technologie

genehmigte

Dissertation

von

Dipl.-Phys. Daniel Zander
aus Tübingen

Tag der mündlichen Prüfung: 8.2.2013

Referent: Prof. Dr. Michael Feindt
Korreferent: Prof. Dr. Thomas Müller

Contents

1. Introduction 6

2. Theory 8
   2.1. The Particles of the Standard Model . . . 8
      2.1.1. Matter and Antimatter . . . 8
      2.1.2. Leptons . . . 9
      2.1.3. Quarks . . . 9
   2.2. The Interactions of the Standard Model . . . 9
      2.2.1. The Strong Interaction . . . 10
      2.2.2. The Electromagnetic Interaction . . . 13
      2.2.3. The Weak Interaction . . . 13
      2.2.4. Electroweak Unification . . . 17
   2.3. The Higgs Mechanism . . . 19
      2.3.1. Charged Higgs . . . 19

3. Data Analysis 23
   3.1. Probability . . . 23
      3.1.1. Quantum Mechanical Probability . . . 23
      3.1.2. Probability In Physics In General . . . 24
      3.1.3. Bayes' Theorem . . . 24
      3.1.4. Distributions . . . 25
      3.1.5. Probability Density Functions . . . 25
      3.1.6. Parameter Estimation . . . 27
   3.2. Classification . . . 30
      3.2.1. Classification in Particle Physics . . . 30
      3.2.2. Monovariate Classification . . . 31
      3.2.3. Multivariate Classification . . . 32
   3.3. NeuroBayes . . . 34
      3.3.1. Theory . . . 34
      3.3.2. The Neural Network . . . 37
      3.3.3. Training . . . 39
      3.3.4. Performance . . . 41
   3.4. Interpretation Of The NeuroBayes Output . . . 43

4. The Experimental Setup 47
   4.1. Particle Accelerators . . . 47
   4.2. The KEKB Accelerator . . . 48

4.3. The Belle Detector . . . 51
4.3.1. Interaction Region . . . 51
4.3.2. Silicon Vertex Detector . . . 52
4.3.3. Central Drift Chamber . . . 53
4.3.4. Aerogel Cherenkov Counter System . . . 54
4.3.5. Time Of Flight Counters . . . 54
4.3.6. Electromagnetic Calorimeter . . . 55
4.3.7. K_L and Muon Detection System . . . 56
4.3.8. Detector Solenoid and Iron Yoke . . . 57
4.3.9. Trigger System . . . 57
4.3.10. Overview and Summary . . . 58

5. The Full Reconstruction **61**
5.1. The Principle Of The Full Reconstruction . . . 61
5.1.1. Missing Momentum And Additional Particles . . . 62
5.1.2. Inclusive Searches . . . 63
5.1.3. Broken China Plates . . . 64
5.2. Efficiency and Purity . . . 64
5.2.1. Channels . . . 65
5.2.2. Hierarchical System . . . 65
5.2.3. Intermediate Cuts . . . 69
5.2.4. Probabilities . . . 70
5.2.5. The Product Of Children Probabilities . . . 71
5.2.6. Cut Determination . . . 71
5.3. The B_{tag} Collection . . . 74
5.3.1. Continuum Suppression . . . 74
5.3.2. Purity And Efficiency . . . 75
5.3.3. Reduced Channel Set . . . 77
5.4. Summary . . . 79

6. Signal Side **82**
6.1. Previous Measurements . . . 83
6.1.1. Tagging Methods . . . 83
6.1.2. First Measurement With Inclusive Tagging . . . 83
6.1.3. Charged B Measurement With Inclusive Tagging . . . 83
6.1.4. All Decay Channels With Hadronic Tagging . . . 84
6.1.5. BaBar Measurements . . . 85
6.1.6. Conclusion . . . 86
6.2. Simulated And Real Data . . . 88
6.3. Signal Side Selection . . . 89
6.3.1. Pre-cuts . . . 89
6.3.2. Choice Of B_{tag} . . . 89
6.4. Analysis Samples . . . 91
6.4.1. D Mass Sideband Sample . . . 96

6.4.2. $\Delta M(D^*D)$ Sideband Sample . . . 96
6.5. Correction of Monte Carlo . . . 97
6.5.1. Tag Side Bias . . . 97
6.5.2. D^{**} Components . . . 97
6.5.3. Lepton ID Efficiency . . . 98
6.5.4. Lepton Fake Rates . . . 98
6.5.5. D Meson Widths . . . 100
6.6. The Fitting Variables . . . 102
6.7. NeuroBayes Trainings . . . 105
6.7.1. Correlation . . . 106
6.8. Best Candidate Selection . . . 110
6.9. Cross-Checks for the NeuroBayes Trainings . . . 112
6.9.1. Signal Efficiencies . . . 112
6.9.2. o_{NB}in D mass sideband . . . 113
6.9.3. o_{NB}in $\Delta M(D^*, D)$ sideband . . . 113
6.9.4. Figure Of Merit . . . 113
6.10. Composition . . . 116
6.11. Normalisation Check . . . 121
6.11.1. The Fit Components . . . 121
6.11.2. Intermediate Branching Fractions . . . 122
6.11.3. Results . . . 124
6.11.4. Conclusion . . . 131
6.12. D^{**} Backgrounds . . . 132
6.12.1. Origin Of The D^{**} Backgrounds . . . 132
6.12.2. D^{**} Enriched Sample . . . 132
6.12.3. Composition And Fitting Of The D^{**} Enriched Sample . . . 132
6.12.4. Data-Monte Carlo Comparison . . . 139
6.13. Fit Strategies . . . 145
6.13.1. Fit Components . . . 145
6.13.2. Factors . . . 149
6.13.3. Shapes . . . 149
6.13.4. Illustration . . . 155
6.14. Expected Yields . . . 156
6.15. Toy Study . . . 157
6.15.1. $B^0 \to D^- \ell^+ \nu_\ell$ Sample . . . 158
6.15.2. $B^0 \to D^{*-} \ell^+ \nu_\ell$ Sample . . . 162
6.15.3. $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ Sample . . . 165
6.15.4. $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ Sample . . . 169
6.15.5. Conclusion . . . 172
6.16. Corrected Expected Yields . . . 173
6.17. Expected Statistical Uncertainty . . . 176
6.18. Determination Of Cut Values On Generic Monte Carlo . . . 177
6.19. Simultaneous Signal Extraction Fit . . . 187

6.20. Results 194
6.21. Systematic Uncertainties 206
6.21.1. D^{**} backgrounds 206
6.21.2. Factors For Simultaneous Fitting 207
6.21.3. PDF Shapes 207
6.21.4. Summary 209

7. Conclusion **213**

A. Appendix **215**
A.1. A Brief Introduction To Feynman Diagrams 215
A.1.1. Phase Space 215
A.1.2. Matrix Element 216
A.2. D Meson Masses 217
A.3. Signal Side Composition 225
A.4. Signal-Background Curves 229
A.5. NeuroBayes Trainings of the Full Reconstruction 233

1. Introduction

The theoretical foundation of particle physics is the *Standard Model*. This theory is based on the principles of quantum mechanics and special relativity. Despite its quite compact form, it has so far incorporated an immense amount of experimental data. The Standard Model allows the extremely precise prediction of certain values, that are also experimentally very well accessible. By passing these highly accurate tests, it has become the most precisely-measured and well-tested model to date. Since its formulation in the 1970s, there has not yet been any conclusive evidence capable of disproving the Standard Model.

There are, however, theoretical extensions to this highly successful model. These extensions are motivated by a number of phenomena. The most prominent one is the existence of so-called Dark Matter. Indicated by its gravitational effects on the spatial dimensions of galaxies, the Standard Model cannot provide a suitable candidate for the nature of this Dark Matter. Theoretical considerations, either concerned with quantum corrections of certain properties or arising from symmetry considerations, give further motivation for the search for extensions of the Standard Model.

Experimental particle physics is devoted to find evidence to support, constrain or disprove the Standard Model itself and in particular its extensions, also known as *New Physics*. There are two different approaches in the search for new physics in high energy particle physics:

With a large enough energy density, the direct production of new particles, postulated by extensions of the Standard Model becomes possible. The largest energy densities are produced in the collisions of high-energetic protons at the *Large Hadron Collider* at CERN in Geneva, Switzerland.

The second method to probe for effects of new physics are indirect searches for its effects in high-precision measurements. Measurements of this kind are performed at electron-positron colliders, also known as B factories, that operate at much lower energies than the Large Hadron Collider. There have been two B factories: The PEP-II collider at the SLAC National Accelerator Laboratory in Stanford, USA and the KEKB collider at the KEK High Energy Accelerator Research Organisation in Tsukuba, Japan. While the energy at B factories is far from sufficient to produce particles postulated by new physics models, the effects of new physics are nonetheless predicted to be visible in high precision measurements. The long-lasting success of the Standard Model is, however, a strong indication that the new physics effects in these indirect searches are

most likely very small, subtle effects. The very simple argument is that strong effects would have been long discovered. One of these subtle effects is postulated to occur in the decays $B \to D\tau\nu_\tau$ and $B \to D^*\tau\nu_\tau$.

The analysis described in this thesis is devoted to the search for effects of new physics in the decays $B \to D\tau\nu_\tau$ and $B \to D^*\tau\nu_\tau$. It has been performed within the Belle Collaboration at the KEKB B factory. A crucial tool for this analysis and indeed for a large number of other analyses within the Belle Collaboration is the Full Reconstruction. This tool allows the measurement of *missing momentum* in an event, providing extermely useful and often crucial information for many analyses. The efficiency of the Full Reconstruction is directly proportional to the size of the available data set for these analyses. Thus, a large part of this thesis is devoted to the redesign and improvement of this well-known full reconstruction algorithm.

2. Theory

The Standard Model of Particle Physics comprises the description of all known particles and their interactions amongst each other. It is based on the principles of quantum mechanics and special relativity. As mentioned in the introduction, this thesis is devoted to testing certain extensions of the Standard Model. Thus, the Standard Model and the tested extensions are described in the following.

2.1. The Particles of the Standard Model

All of the objects in the Standard Model can be categorised by the *spin* they carry. The spin is an internal quantum mechanical property of any particle. The mathematical treatment of spin is similar to the treatment of the macroscopic phenomenon of spin angular momentum. But this is as far as the analogy goes. [1] Suffice it to say that the spin is an internal property of particles and is measured in units of the reduced Planck constant $\hbar$ [2], where

$$\hbar = \frac{h}{2\pi} \approx 1.0546 \times 10^{-34}\text{Js} \,. \tag{2.1}$$

Particles that carry half-odd-integer spins ($1/2\hbar$, $3/2\hbar$, ...) are called *fermions*, particles that carry integer spins ($0\hbar$, $1\hbar$, ...) are called *bosons*. [3]

Only the lightest particles within the Standard Model are long-lived, or in other words *stable*. Generally, heavier particles decay very quickly into lighter particles, unless that decay is forbidden by a conservation law. In a manner of speaking, the stable particles are only stable, because there is nothing lighter for them to decay into or a conservation law forbidding it. One exemplary decay that is of special interest for this thesis it that of a relatively heavy b quark into the lighter c quark.

2.1.1. Matter and Antimatter

Quantum Mechanics does not only allow "normal" particles, but also anti-particles [4]. All leptons and quarks (see below) have corresponding anti-particles. Particle and anti-particle have the same mass and differ by the sign of their internal quantum numbers.

[1] The spin angular momentum of a macroscopic object can be interpreted as the sum total of the angular momenta of all the mass points on their orbit around the axis of rotation. But as the fundamental particles in the Standard Model have no sub-structure and appear as true point particles, this picture cannot hold. [1]

2.1.2. Leptons

The six known leptons (e, ν_e, μ, ν_μ, τ, ν_τ) form three different families. Each of these families consists of a charged, massive lepton and a very light uncharged corresponding neutrino. Table 2.1 gives an overview of the three lepton families. Both the charged leptons and the corresponding neutrinos carry a lepton number (N_e, N_μ and N_τ respectively) that has to be conserved in all processes [2]. Thus, the decay

$$\tau^- \rightarrow e^- \tag{2.2}$$

is forbidden by lepton number conservation and the decay of a τ lepton into an electron proceeds as follows:

$$\tau^- \rightarrow e^- \nu_\tau \overline{\nu}_e \,. \tag{2.3}$$

This decay and the fact that it proceeds emitting two neutrinos will be of special interest for the analysis presented in this thesis.

Family	Name	Charge	N_e	N_μ	N_τ	Mass
I	e	-1	1	0	0	$0.511 \pm 1.3 \cdot 10^{-8}$ MeV/c^2
	ν_e	0	1	0	0	< 2 eV/c^2
II	μ	-1	0	1	0	$105.7 \pm 0.4 \cdot 10^{-5}$ MeV/c^2
	ν_μ	0	0	1	0	< 2 eV/c^2
III	τ	-1	0	0	1	1776.8 ± 0.17 MeV/c^2
	ν_τ	0	0	0	1	< 2 eV/c^2

Table 2.1.: Overview of the leptons of the Standard Model [2].

2.1.3. Quarks

The six known quarks (d, u, s, c, b, t) also form three families, in analogy to the leptons. They carry non-integer electric charge ($\pm 1/3e$ and $\pm 2/3e$) and also differ greatly in their masses. Table 2.2 gives a systematic overview of the 6 quarks.

2.2. The Interactions of the Standard Model

Currently, there are only four different types of interactions known:

[2] with the possible exception of neutrino osciallations

Family	Name	Charge	Mass
I	up (u)	$+\frac{2}{3}$	$2.55^{+0.75}_{1.05}$ MeV/c^2
	down (d)	$-\frac{1}{3}$	$5.04^{+0.96}_{-1.54}$ MeV/c^2
II	charm (c)	$+\frac{2}{3}$	$1.27^{+0.07}_{-0.11}$ GeV/c^2
	strange (s)	$-\frac{1}{3}$	105^{+25}_{-35} MeV/c^2
III	top (t)	$+\frac{2}{3}$	$171.3 \pm 1.3 \pm 1.2$ GeV/c^2
	bottom (b)	$-\frac{1}{3}$	$4.20^{+0.17}_{-0.07}$ GeV/c^2

Table 2.2.: Overview of the quarks of the Standard Model [2].

Gravitation is a macroscopically long ranged interaction between masses, described by General Relativity. It is not part of the Standard Model. Its effects on the microscopic processes of Particle Physics are completely negligible.

Electromagnetism causes the attraction of two unequally signed electric charges and the repulsion of equally signed electric charges. Its classical formulation are the famous Maxwell equations.

The Weak Interaction manifests itself in the decay of unstable particles. It was first observed in the nuclear beta decay.

The Strong Interaction causes the attraction between quarks and nucleons.

We will see in chapter 2.2.4, that the electromagnetic and the weak interaction are merely two different low energy manifestations of a unified interaction, which we call the electroweak interaction.

2.2.1. The Strong Interaction

Quarks do not occur isolated in nature, but only in bound states of two or more quarks. This phenomenon is called *confinement*. From the observation of the energy spectra of bound two-quark systems (e.g. "Charmonium" $c\bar{c}$), the form of the potential between two quarks can be derived [5] as

$$V_S = -\frac{4}{3}\frac{\alpha_S}{r} + kr\,. \tag{2.4}$$

The first term ($\sim 1/r$) is clearly dominant at small ranges and proportional to the Coulomb potential in electrodynamics: $V(r \to 0) \sim 1/r$. At large distances, however, the second term ($\sim r$) dominates clearly and gives an explanation for the above mentioned confinement: $V(r \to \infty) \to \infty$. Figure 2.1 visualises equation 2.4. This quickly rising potential between two quarks has an interesting effect: With increasing distance

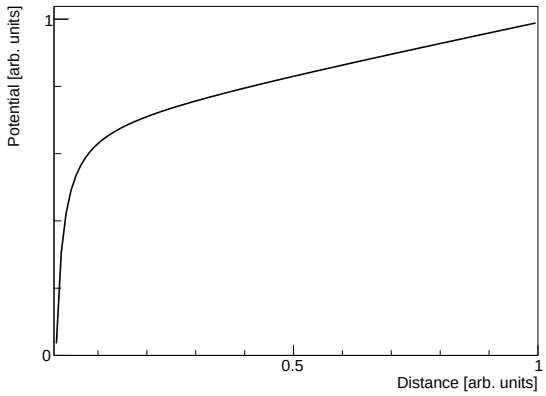


Figure 2.1.: Visualisation of the two-quark potential given in equation 2.4.

between two quarks, at about 10^{-15} m, it becomes energetically preferable to create a new quark-anti-quark pair instead of increasing the distance further. This process is known as hadronisation.

The theory that describes the behaviour of quarks under the strong interaction is called Quantum Chromodynamics (QCD). It is a gauge theory of the SU(3) gauge group, that defines a charge with three possible types. This strong charge, an internal quantum number, is called *colour*. It is a requirement of Quantum Chromodynamics that only colour-neutral states are allowed. The chosen colours *red*, *green* and *blue* (r,g,b) and their corresponding *anti-colours* ($\bar{r}$,$\bar{g}$,$\bar{b}$) can be combined to colour-neutral states in the following ways:

$$r + \bar{r} = \text{neutral} \qquad g + \bar{g} = \text{neutral} \qquad b + \bar{b} = \text{neutral}$$
$$r + g + b = \text{neutral} \qquad \bar{r} + \bar{g} + \bar{b} = \text{neutral}\,.$$

The concept of assigning a "colour" to quarks should not be taken literally, of course. It is an internal quantum number, but its name provides a useful analogy to explain the fact that quarks occur in bound states of two or three quarks. Experimental evidence for the internal quantum number colour comes from the existence of the Δ^{++} particle, that is measured as a bound state of three up-quarks in the spatially symmetric ground state with angular momentum $l = 0$ with the spins coupled to $J = \frac{3}{2}$, which can figuratively be described as "all three spins parallel". So the wave function of this particle is symmetric in flavour, space and spin, which seems to be a violation of the Pauli principle, which states that multiple fermions cannot exist in the same quantum state. This issue is resolved by an additional quantum number, the colour. Quantum Chromodynamics also allows other possibilities for colourless states, such as four-quark or five-quark

bound states.

$$(r + \bar{r}) + (g + \bar{g}) = \text{neutral}$$
$$(r + g + b) + (r + \bar{r}) = \text{neutral}$$

There has not been yet irrevocable evidence for the existence of such exotic hadrons.

The colour must be conserved in all processes of Quantum Chromodynamics. The fact that Quantum Chromodynamics is a non-abelian theory results in the self-interactions of the gauge field. This means, in effect, that the mediators of the strong force carry the strong charge "colour" themselves. The mediators of the strong interaction are the *gluons* (g). Gluons are massless spin-1 particles that couple to colour and carry colour themselves. The resulting self-coupling explains the short range of the strong interaction, despite the masslessness of the gluons. An exemplary process in Quantum Chromodynamics is schematically shown as a Feynman diagram in figure 2.2. A quark

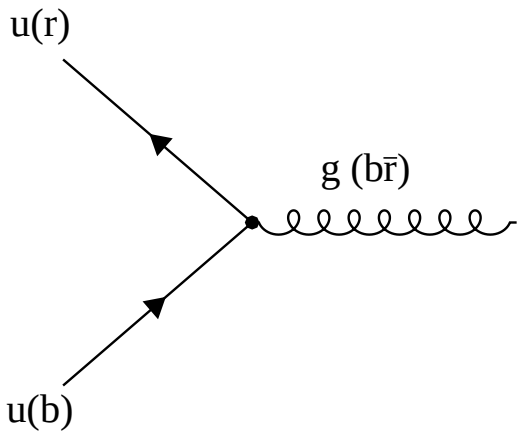


Figure 2.2.: Example for the strong interaction

carrying the colour blue is converted into a quark carrying the colour red. As colour has to be conserved, the gluon must carry the difference, which, in this case, is blue and anti-red. All gluons carry a certain colour and a different anti-colour. As there are three different colours in Quantum Chromodynamics, one would expect nine possible gluons: $r\bar{r}$, $r\bar{b}$, $r\bar{g}$, $b\bar{r}$, $b\bar{b}$, $b\bar{g}$, $g\bar{r}$, $g\bar{b}$, $g\bar{g}$. However, as Quantum Chromodynamics is based on the SU(3) symmetry, the possible colour combinations form a colour octet and a colour singlet. The above mentioned confinement of quarks can be rephrased in terms of the SU(3) symmetry: All naturally occurring free particles must be colour singlets. The hypothetical ninth gluon would be a colour singlet and occur as a free particle and a long-ranged mediator of the strong force between two other colour singlets, e.g. proton and neutron. This does not occur in nature and we can therefore exclude the existence of a gluon singlet [3].

2.2.2. The Electromagnetic Interaction

The electromagnetic interaction is one of the pillars of classical physics. Quantum Electrodynamics (QED) describes the effects of the electromagnetic interaction at the microscopic level. All electrically charged particles interact via the exchange of the electromagnetic exchange boson, the photon γ. QED describes the annihilation and production of charged particles. The Feynman diagram in figure 2.3 shows the coupling of charged particles to the photon. The annihilation of electron and position is

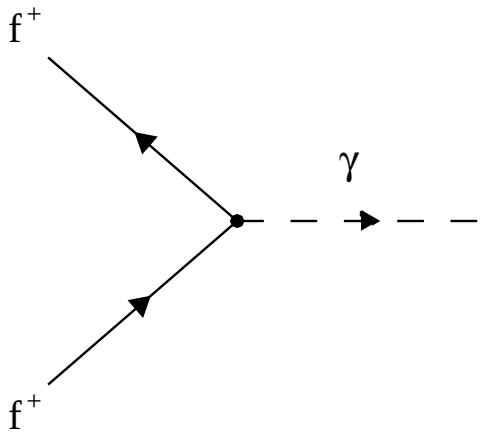


Figure 2.3.: Example for the electromagnetic interaction.

an example of the electromagnetic interaction which is of special interest. This annihilation and the subsequent production of the $\Upsilon(4S)$ resonance is the basis for the entire physics program at B factories and for the Full Reconstructionin particular (see chapters 4.2 and 5.1).

2.2.3. The Weak Interaction

The only interaction in which all known fermions participate is the weak interaction. Its effects are small compared to the strong and electromagnetic interaction and often are obfuscated by them. For instance, the annihilation of electron and positron depicted in figure 2.3 and the subsequent production of the $\Upsilon(4S)$ resonance can also occur via the weak interaction.
There are no known bound states that are formed due to the weak interaction, as there are in all other fundamental interactions. This makes its description less intuitive, as we cannot rely on analogies from atomic physics. The weak interaction manifests itself in the decay of leptons and quarks. Two exemplary weak processes are schematically shown as a Feynman diagrams in figures 2.4 and 2.5. In figure 2.4, we can see a down-type quark that is converted into an up-type quark in a weak process. Figure 2.6 shows the weak interaction of a massive lepton and its corresponding neutrino. The weak interaction is the only Standard Modelinteraction in which neutrinos participate.

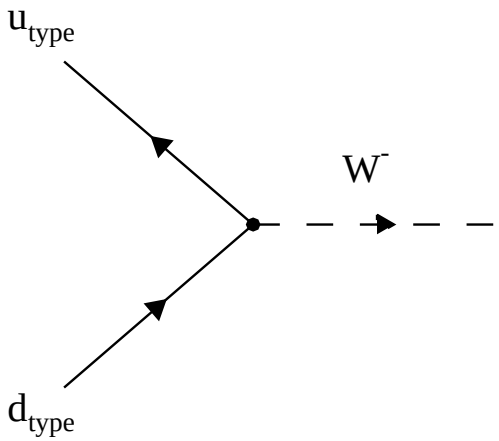


Figure 2.4.: Example for the weak interaction.

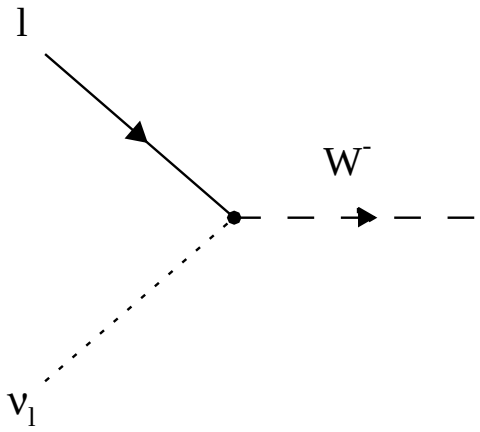


Figure 2.5.: Example for the weak interaction.

First experimental evidence for the weak interaction was found in the β decay of neutrons in atomic nuclei, where a neutron decays into a proton, an electron and a neutrino, depicted in figure 2.6. It can be easily seen, that the β decay comprises exactly the two processes that were already shown in figures 2.4 and 2.5. In contrast to all other interactions, the mediators of the weak interaction (called $W^{\pm}$ and Z^0) are substantially massive particles. Their masses of ~ 80 GeV/c^2 and ~ 91 GeV/c^2 are similar to the mass of heavy atomic nuclei. The above mentioned relative weakness of the weak interaction can be attributed to the large masses of the exchange particles. Their properties are summarised in table 2.3.

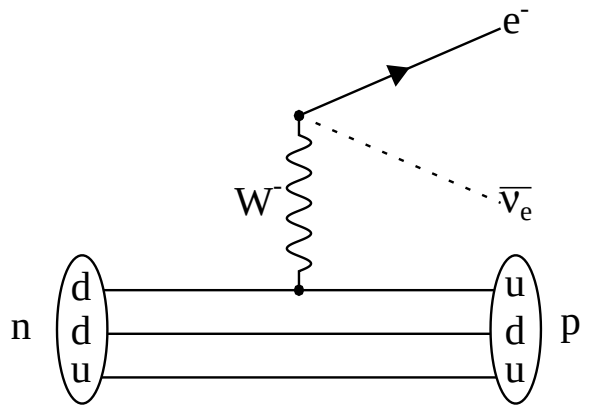


Figure 2.6.: The β decay.

Name	Charge	Spin	Mass
$W^{\pm}$	± 1	1	80.399 ± 0.023 GeV/c^2
Z^0	0	1	91.1876 ± 0.0021 GeV/c^2

Table 2.3.: Overview of the weak exchange bosons [2].

Parity Violation

A unique property of the weak interaction is the fact, that it is not invariant under the parity transformation. Any coupling mediated by a spin-1 particle can have either vectorial or axial-vectorial characteristics and their fraction is described by the two factors c_A and c_V [5]. An interaction is invariant under parity transformation ("conserves parity"), if either $c_A = 0$ or $c_V = 0$ and it violates parity if $c_A \neq 0$ and $c_V \neq 0$. The violation is maximal if $|c_A| = |c_V|$. From the analysis of the properties of the decay $\mu \rightarrow e\nu_e\nu_\mu$, one can extract $c_A \approx -c_V$ and $|c_A|/|c_V| = 0.985 \pm 0.024$ [6]. This means that, within the experimental errors, parity is maximally violated in weak interactions. For illustration of this phenomenon, the concept of "handedness" is instructive: "Handedness" is defined in such a way, that a particle is right-handed if its spin and momentum point into the same direction and it is left-handed, if its spin and momentum point in opposite directions. A consequence of the maximal violation of parity is that only left-handed fermions and right-handed anti-fermions interact weakly. For massive particles, that always travel slower than the speed of light in a vacuum ($v < c$), handedness depends on the frame of reference, as there can always be a frame of reference where the momentum of the particle is reversed. So right-handed fermions and left-handed anti-fermions do interact weakly, but the interaction is suppressed. For neutrinos, that travel at the speed of light ($v = c$), the situation is different: While left-handed neutrinos and right-handed anti-neutrinos interact weakly, right-handed neutrinos and left-handed anti-neutrinos

do not. There is no frame of reference in which their handedness was reversed. Without a "correctly-handed component" they participate in none of the known interactions.

The CKM Matrix

In the weak interaction of leptons, only lepton transitions within the same generation are allowed, see figure 2.5. Quarks, on the other hand, do not exhibit this behaviour, and there are transitions beyond the own generation. This difference can be explained by the CKM mechanism as follows: The weak interaction does not couple to the pairs of quark mass-eigenstates

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix},$$

but to a set of "skewed" generations:

$$\begin{pmatrix} u \\ d' \end{pmatrix} \begin{pmatrix} c \\ s' \end{pmatrix} \begin{pmatrix} t \\ b' \end{pmatrix},$$

where d', s' and b' are linear combinations of the mass-eigenstates of quarks, given by

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \underbrace{\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}}_{\text{CKM Matrix}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.5}$$

There is no evidence that would favour skewing the down-type quarks rather than the up-type quarks, so our choice is arbitrary but follows convention. The matrix that describes these linear combinations is called the "CKM Matrix". The current best experimental values for the magnitues of the matrix elements [2] are summarized in the following equation:

$$V_{\text{CKM}} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} V_{cb} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix}. \tag{2.6}$$

The diagonal elements of the matrix correspond to quark transitions within the same generation. Their values very close to 1 show that this transition is by far the most frequent. The squares of the off-diagonal elements give the relative frequency of transitions beyond the quark-generation. Thus, figure 2.4 also describes $b \to u$ or $t \to d$ transitions. The change of generation without the change of electric charge would require Flavour Changing Neutral Currents, which are highly suppressed in the Standard Model by the GIM mechanism [7].

Unitarity Triangle

The CKM matrix is a unitary matrix and can be parameterised by three amplitudes and one phase [8]. The unitarity constraint leads to six orthogonality relations, that can be depicted as six triangles in the complex plane [9]. The most popular one arises from the condition

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0\,. \tag{2.7}$$

We can now divide this equation by the best known summand $V_{cd}V_{cb}^*$ and obtain a triangle with the base stretching from (0,0) to (1,0) as depicted in figure 2.7. The over-

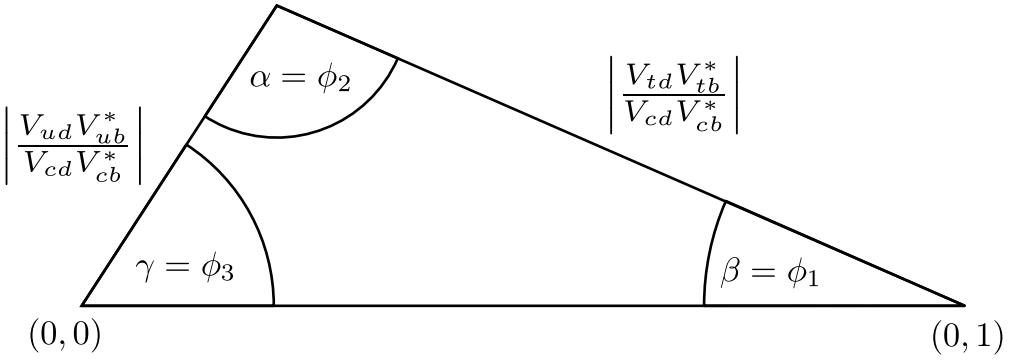


Figure 2.7.: A unitarity triangle.

constraining of the parameters of this triangle is one of the main pursuits of flavour physics today. It allows to test the theoretical predictions by the CKM mechanism and numerous measurements are designed for this purpose.

2.2.4. Electroweak Unification

The W bosons couple to all left-handed fermions with the same strength. Only the CKM mechanism leads to minor corrections for the quark coupling. These expectations could be confirmed in the analysis of the branching fractions for the W decay. If the Z boson mediated the weak interaction in the same way as the W, we would expect the same branching ratios in the decay to pairs of charged leptons and pairs of neutrinos. Highly precise measurements at all 4 LEP experiments show that the branching ratio of $Z^0 \to \nu_{e,\mu,\tau}\bar{\nu}_{e,\mu,\tau}$ is much larger than the branching ratio $Z^0 \to (e^-, \mu^-, \tau^-)(e^+, \mu^+, \tau^+)$ (e.g. [10]). The Z boson can therefore not be treated as an "uncharged W", but it mediates a more complicated interaction, that seems to depend on the electric charge.

To explain this phenomenon, a new quantum number is introduced: The weak isospin T. The families of left-handed (L) quarks and leptons are identified as weak isospin doublets ($T_3 = \pm 1/2$), the right-handed (R) fermions are identified as weak isospin singlets ($T_3 = 0$) [5, 11], as shown in table 2.4. Members of the doublets can transform

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

Table 2.4.: Fermions and their weak isospin [5].

into their doublet-counterpart by W exchange. The weak isospin structure for exchange bosons is given in table 2.5 and gives rise to a weak-isospin triplet and a weak-isosping singlet. The two $W^\pm$ bosons are already known from the "purely weak" interaction.

Boson multiplet	T_3	T
$\begin{pmatrix} W^+ \\ W^0 \\ W^- \end{pmatrix}$	$+1$	
	0	1
	-1	
B^0	0	0

Table 2.5.: Bosons and their weak isospin.

Both the W^0 and the B^0 couple to fermions without changing their weak isospin. The physically-known vector-bosons with that property are the γ and the Z^0. These two vector-bosons can be formulated as orthogonal linear combinations of the W^0 and the B^0:

$$\left|\gamma\right\rangle = \cos\theta_W \left|B^0\right\rangle + \sin\theta_W \left|W^0\right\rangle \tag{2.8}$$

$$\left|Z^0\right\rangle = -\sin\theta_W \left|B^0\right\rangle + \cos\theta_W \left|W^0\right\rangle , \tag{2.9}$$

where θ_W is the electroweak mixing angle. θ_W also gives a relation between the weak and the electromagnetic charge g:

$$e = g \cdot \sin\theta_W . \tag{2.10}$$

The electroweak mixing angle was measured as [2]

$$\sin^2\theta_W \approx 0.231 . \tag{2.11}$$

The electroweak mixing renders the interaction more complicated: The charged $W^{\pm}$ bosons couple to all fermions with the same strenght. The couplings to the Z^0 boson, on the other hand, depend also on the electrical charge of the fermion:

$$g_Z(f) = \frac{g}{\cos\theta_W} \cdot (T_3 - z_f \sin^2\theta_W) \,, \tag{2.12}$$

where z_f is the electric charge of the fermion in units of the elementary charge e.

2.3. The Higgs Mechanism

The fact that the W and Z bosons are massive particles poses a central issue in the theory of electroweak physics. Theoretical consideration would prefer all electroweak exchange particles to have the same - "symmetric" - zero mass. Looking at table 2.3, we see that the masses of the W and Z bosons are far from zero, while the photon is indeed massless. One possible explanation for this electroweak symmetry breaking is the Higgs mechanism.

The Higgs mechanism adds an additional scalar field ϕ to the theory, whose minimum does not lie at $\phi = 0$ [12], resulting in a doublet of complex scalars $\phi = (\phi^+, \phi^0)$. The four free scalar fields correspond to four bosons: Three Goldstone bosons are created, that appear as the longitudinal components of the - now massive - W^+, W^- and Z^0 bosons. The fourth free field that has not bee "absorbed" by the massive electroweak exchange bosons manifests itself as the Higgs Boson. It is worth noting that the Higgs mechanism is not only capable of giving mass to the electroweak interaction particles, but to all fermions as well via the Yukawa interaction of scalar fields and Dirac fields.

Indirect and direct searches for this boson have been going on for decades. In July 2012, an excess of events in the search for the Higgs boson have been simultaneously reported by the ATLAS [13] and CMS [14] collaborations at the Large Hadron Collider at CERN. The both excesses were found at masses of $M_H = 126.5\ \mathrm{GeV}/c^2$ (ATLAS) and $M_H = 125.5\ \mathrm{GeV}/c^2$ (CMS). While further investigation is needed and ongoing, these excesses seem to be promising candidates for the direct experimental discovery of the Higgs boson.

2.3.1. Charged Higgs

Although the Standard Model has been a very successful model, there are still areas of uncertainty left by it. A prominent example is the nature of dark matter, which is not explained by the Standard Model. Also, the masses of quarks and leptons or the magnitude of the CKM matrix elements are measured, but not *understood* [15]. Several extensions of the Standard Model postulate the existence of charged Higgs bosons in addition to the neutral Higgs particle postulated by the simplest Higgs mechanism. The simplest extensions of the Standard Model postulate one additional Higgs field doublet

and are realized e.g. in the *Minimal Supersymmetric Standard Model* (MSSM) [16]. In certain types of extended Higgs models (also in the MSSM at tree level) up-type quarks get their mass solely from one of the two doublets, while down-type quarks get their mass from the other doublet [17]. The existence of these charged Higgs bosons could alter the structure of the weak interaction [18]. Thus, it could have measurable impact on the branching ratios of several B meson decays, already accessible to experimental tests. The decays $B \to D^{(*)}\tau\nu$ are promising candidates for indirect evidence of extended Higgs model contributions.

The interaction of charged Higgs bosons with fermions is proportional to the fermion masses m_b and m_τ (see eq. 2.14 and [19]), thus the largest eventual contribution of the charged Higgs interaction at a B factory should be visible in decays of B mesons with τ leptons in the final state. This choice of the largest possible involved mass leads us to the decay $B \to \tau\nu$, which has been analysed by the Belle Collaboration [20], and to the decay $B \to D^*\tau\nu$, which is the subject of this analysis. The decay is depicted in figure 2.8.

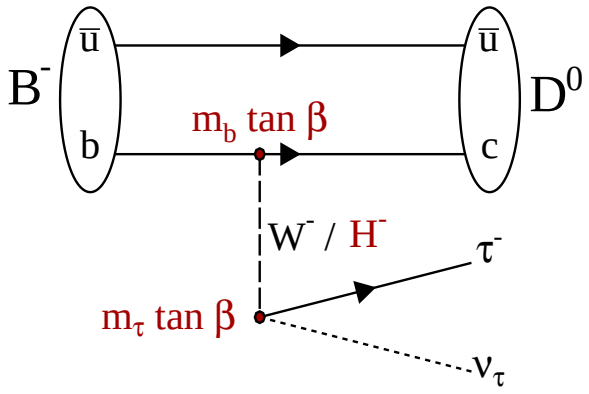


Figure 2.8.: Feynman graph for the decay $B \to D^*\tau\nu_\tau$. The hypothetical decay via charged Higgs is highlightes in red (coloured version).

The effective Hamiltonian (see section A.1) describing the decays $B \to D^{(*)}\tau\nu$ and $B \to \tau\nu$ is

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} V_{qb} \Big\{ [\bar{q}\gamma^\mu(1-\gamma_5)b][\bar{\tau}\gamma_\mu(1-\gamma_5)\nu_\tau] \tag{2.13}$$

$$-\frac{\bar{m}_b m_\tau}{m_B^2}\bar{q}[g_S + g_P\gamma_5]b[\bar{\tau}(1-\gamma_5)\nu_\tau]\Big\} . \tag{2.14}$$

The first part of the effective Hamiltonian (equation 2.13) contains the weak current standard model interaction. The second part (equation 2.14) describes the charged Higgs interaction. The coupling constant g_p only affects the decay $B \to \tau\nu$, whereas

the decay $B \to D^{(*)}\tau\nu$ is sensitive only to the coupling constant g_S. In the MSSM, the coupling constants take the following form:

$$g_S = g_P = \frac{m_B^2}{M_{H^+}^2} \frac{\tan^2\beta}{(1+\tilde{\epsilon}_0 \tan\beta)(1+\epsilon_\tau \tan\beta)} , \tag{2.15}$$

where $\tilde{\epsilon}_0$ and ϵ_τ are loop correction factors that are 0 in type-II 2 Higgs doublet models. In these models, the effects of the charged Higgs H^+ are thus entirely parameterised by the mass of the H^+ and the ratio of the two vacuum expectation values for the up- and down-type doublets v_u and v_d:

$$\tan\beta = \frac{v_u}{v_d} . \tag{2.16}$$

The effects of the charged Higgs contribution should manifest themselves in a change of the branching ratios of the decays $B \to D^*\tau\nu$. Many hadronic and Standard Model parametric uncertainties must be considered in the measurement of these branching ratios. It is therefore beneficial to define the ratios in equations 2.17 to 2.20. In these ratios, most of the above mentioned uncertainties cancel, while the sensitivity to effects of charged Higgs contributions is still retained. It is the goal of the presented analysis to measure the following ratios:

$$R^0 = \frac{\mathcal{B}(B^0 \to D^- \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \to D^- \ell^+ \nu_\ell)} \tag{2.17}$$

$$R^+ = \frac{\mathcal{B}(B^+ \to \overline{D}^0 \tau^+ \nu_\tau)}{\mathcal{B}(B^+ \to \overline{D}^0 \ell^+ \nu_\ell)} \tag{2.18}$$

$$R^{*0} = \frac{\mathcal{B}(B^0 \to D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu_\ell)} \tag{2.19}$$

$$R^{*+} = \frac{\mathcal{B}(B^+ \to \overline{D}^{*0} \tau^+ \nu_\tau)}{\mathcal{B}(B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell)}. \tag{2.20}$$

The Standard Model predictions are 0.297 ± 0.017 for R^0 and R^+ and 0.252 ± 0.003 for R^{*0} and R^{*+} [21]. As indicated by the negative sign of the supersymmetric part of the effective Hamiltonian (see equation 2.14), a negative interference is expected from the charged Higgs contribution. Thus, the ratio R should decrease down to the point where the supersymmetric contribution dominates over the standard model expectation, as shown in figure 2.8. A significant deviation of the ratios R and R^* from the Standard Model expectation would indicate a physical process beyond the Standard Model.

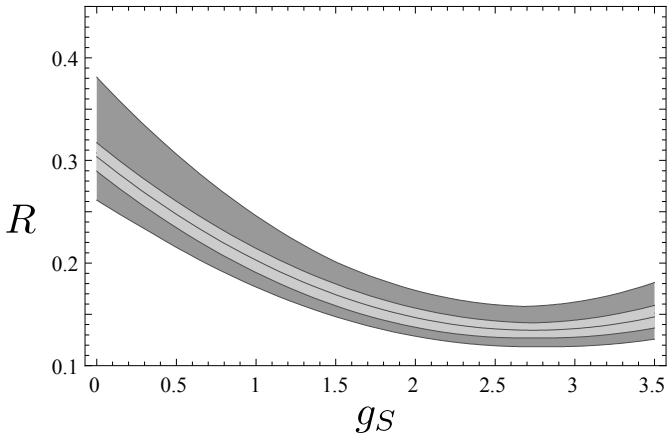


Figure 2.9.: Expected behaviour for R^0 and R^+ depending on g_S [22].

3. Data Analysis

3.1. Probability

When we look at classical physical processes, like a pendulum or the orbits of planets, the outcome of a measurement is exactly predictable. There is no fundamental uncertainty; a defined cause has a defined effect. This is the concept which we call *determinism*. The other extreme would be a process whose outcome is completely non-predictable, like the exact time of a radioactive decay of a single atom or the outcome of a lottery. We call processes of this type *coincidence*.

Most real systems are, however, a mixture of determinism and pure coincidence [23]. The expectation, that a certain experiment should give, under identical circumstances, identical results is an idealisation even in classical physics. The coincidental element in this classical case is the fact that the circumstances of two experiments can never be exactly identical. For many measurements, these deviations are tiny compared to the precision of the measurement. In such cases, we may safely neglect the coincidental element of the measurement and our prediction of the outcome of a measurement (say the position of Jupiter) appears within the inherent precision of the measurement exact.

3.1.1. Quantum Mechanical Probability

We find a totally different situation in the field of quantum mechanics. An experiment, even if the circumstances could be chosen to be identical, would never yield exactly predictable results. And this behaviour would not be due to our current lack of experimental finesse, but it is an essential property of quantum mechanics. It is thereby an unavoidable part of particles physics, ultimately due to Heisenberg's uncertainty principle. As a consequence of the coincidental nature of the processes in particle physics, n measurements will usually yield n different results, that may, when plotted in a histogram, cluster in the vicinity of a certain value. At other times the results will be distributed over a certain range uniformly or according to a certain pattern. Keeping these arguments in mind, it is evident that it is impossible to measure any quantity in the field of particle physics exactly and we are limited to make statements about the *probability* of the outcome of a measurement. We define the probability $P(A)$ for the outcome A of an experiment with n independent measurements and k positive outcomes as:

$$P(A) = \lim_{n\to\infty} \frac{k}{n} \,. \tag{3.1}$$

A typical example would be the probability of an atom to decay within a certain time window (say within 10 seconds).

3.1.2. Probability In Physics In General

While all processes in nature are ultimately governed by quantum mechanics, the effects are often negligibly small. Nonetheless, even macroscopic measurements will never yield the exact same outcome, due to the finite resolution of any measuring apparatus. Think of measuring the exact distance between the physics building at KIT and San-Go-Kan at KEK to a precision of micrometers. Repeated measurements will yield slightly different results. The results will cluster according to Gaussian distribution around the true value [24]

The definition in equation 3.1 yields one practical problem, which is that the infinite number of repetitions of the experiment can never be achieved. It can, however, be shown that the error on the probability can be reduced as far as desired by increasing the number of repetitions.[1]

3.1.3. Bayes' Theorem

An important property of probabilities is the ability to make combinations. Let the probability of event A be $P(A)$ and the probability of event B be $P(B)$. Then the probability that both events A and B occur is

$$P(A \text{ and } B) = P(B|A) \cdot P(A) \tag{3.2}$$

where $P(B|A)$ is the conditional probability that B occurs, given that A has occurred. As A and B are not specified any further we can also write in general

$$P(A \text{ and } B) = P(B|A) \cdot P(A) = P(A|B) \cdot P(B) \tag{3.3}$$

which leads to Bayes' Theorem:

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)} \,. \tag{3.4}$$

In the Bayesian interpretation, this theorem relates an initial degree of belief $P(A)$ ("prior") with additional infomation $P(B)$ ("evidence") and the likelihood for the additional information $P(B|A)$ with the updated degree of belief encompassing additional the information $P(A|B)$ ("posterior"). Bayes' Theorem will be used extensively in this thesis, in particular in the Full Reconstruction algorithm, described in detail in chapter 5.

[1] Additionally, we must assume a perfect experiment with no systematic uncertainties. Systematic uncertainties will not be reduced by increasing the number of repetitions.

3.1.4. Distributions

As explained in section 3.1, repeated measurements do in general not yield one single fixed and predictable value, but rather a distribution of values. Figure 3.1 shows a histogram of multiple independent mass measurements of a certain type of particle. These repeated measurements result in a distribution, typical for particle physics. We

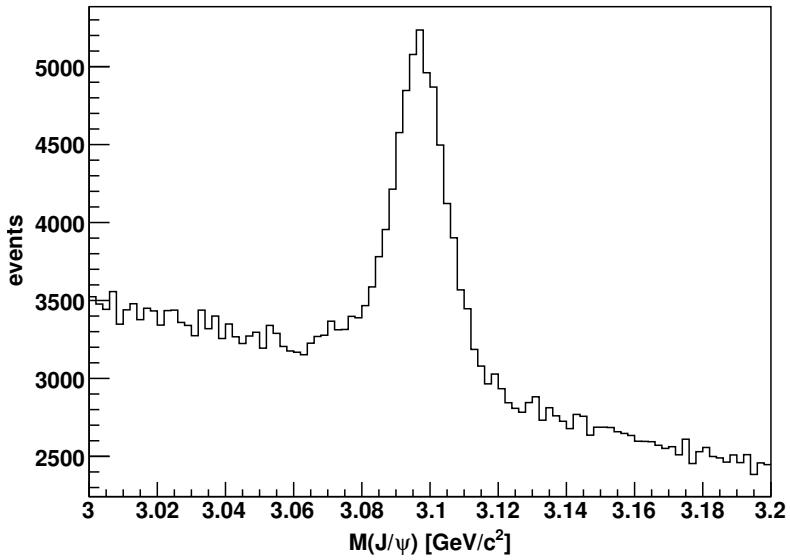


Figure 3.1.: A typical mass distribution.

can intuitively see two component in the histogram: A peaking structure and a downward sloped flat component. The peak can be identified as mass measurements of the particle we are interested in (J/ψ in this example). We call this structure *signal*. The downward sloped component can be identified as detector noise or mass measurements of other particles whose masses were close or identical to the mass of the signal component, but were not signals in reality. We call this component the *background*. The mass measurement of this particular type of particle merely serves as an example. In this case, the relatively wide distribution is caused by the finite resolution of the detector. Nonetheless, a clear signal can emerge, given enough repetitions of the measurement.

3.1.5. Probability Density Functions

When we measure a continuous quantity like mass, it is in principle impossible to measure one exactly predicted value (say $m = 3.096$ GeV/c^2). The probability for any such prediction is exactly 0. We can, however, predict that a measurement will yield a value within a certain range. In general, the probability for such a prediction is larger than 0. Therefore, we introduce the *probability density function (pdf)* $f(x)$, a property that characterises the distribution of a random variable. It denotes the probability

$P(a \leq x \leq b)$ that a measurement of the quantity x lies within the limits a and b by

$$P(a \leq x \leq b) = \int_a^b f(x)\mathrm{dx} \,. \tag{3.5}$$

A pdf has to be non-negative and normalized to 1:

$$f(x) > 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)dx = 1 \,. \tag{3.6}$$

The application of formula 3.5 can be shown in the exemplary plot in figure 3.1: The probability that the outcome of a mass measurement will lie within the peaking structure is higher than within an equally wide range outside the peak. Note that the probability density function $f(x)$ is not a probability, but rather the quantity $f(x)\Delta x$ will become the probability for x to lie within x and Δx as $\Delta x \to 0$.

Expectation Value And Variance

An important parameter that characterizes a pdf is the *expectation value* E of a function $h(x)$ with respect to the pdf $f(x)$. E is defined as

$$E[h] = \int_{-\infty}^{+\infty} h(x)f(x)dx \tag{3.7}$$

with the important special case of the expectation value of "the function" x:

$$E[x] = \langle x \rangle = \int_{-\infty}^{+\infty} xf(x)dx \,, \tag{3.8}$$

also known as the first algebraic moment. In general, the expectation value of the function x^n is called n^{th} algebraic moment.

The *variance* $V(x)$ is another important aspect in the characterization of a pdf and is defined as

$$V[x] = E[(x - \langle x \rangle)^2] = \int_{-\infty}^{+\infty} (x - \langle x \rangle)^2 f(x)dx = \sigma^2 \,. \tag{3.9}$$

It characterizes the width of a distribution and is also known as the second central moment. In general, the expectation value of the function $(x - \langle x \rangle)^n$ is called n^{th} central moment.

There are numerous other momenta or function of momenta that are used to characterize probability density functions. It can be shown, that any probability density function can be defined unambiguously by all of its momenta.

3.1.6. Parameter Estimation

Any measurement in particle physics is subject to uncertainty and measurement error. They are introduced by the above mentioned fundamental uncertainties of the quantum physical measurement process, and also by the unavoidable measurement error due to an imperfect measuring apparatus. It is therefore important to utilize the optimal technique to extract a relatively precise result from imperfect and imprecise measurements and also to determine the error of that result. If we once again look at figure 3.1, we intuitively see what the true value of the measured quantity will most likely be. Judging from the width of the signal peak, we may even estimate the error in our determination of the value. So from a relatively broad spectrum, we can extract a quite precise result and its uncertainty. For proper scientific purposes, the described intuitive approach is, of course, insufficient and an analytical procedure is required to obtain unbiased results. We evaluate the quality of such a parameter estimation procedure by the following four criteria:

Consistency For a large number of measurements n, the result $\hat{a}$ should converge towards the real value a_0 :

$$\lim_{n\to\infty} \hat{a} = a_0 \tag{3.10}$$

Truth to expectation The expectation value $E[\hat{a}]$ should be the true value a_0.

$$E[\hat{a}] = a_0 \tag{3.11}$$

Effectiveness The variance of the result $\hat{a}$ should be as small as possible.

Robustness Data outliers or wrongful modelling should influence the result $\hat{a}$ as little as possible.

One of the most commonly used procedures for parameter estimation is the *maximum likelihood method*.

The Maximum Likelihood Method

Let us assume, that the underlying probability density function $f(x|a)$ of a measured distribution of variable x is known. x may stand for a single variable or a vector of variables with n measured values: x_1, x_2 $\ldots x_n$. a stands for one or more unknown parameters of the pdf. Given this underlying probability density function $f(x|a)$, we can then form the *negative log likelihood function* $F(a)$, defined as

$$F(a) = -\sum_{i=1}^{n} \ln f(x_i|a)\,. \tag{3.12}$$

According to the maximum likelihood principle, the best estimator $\hat{a}$ is the one, that minimizes $F(a)$[2]. It can be proven, that in the limit $n \to \infty$ this method is consistent, true to expectation and effective. A problem of this method is, however, that the a-priori knowledge of the underlying probability density function $f(x|a)$ is necessary.

An instructive example is the parameter estimation using a Gaussian distribution. The Gaussian distribution is defined as:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} . \tag{3.13}$$

Its shape is determined by the two parameters σ and μ. It can be shown that

$$\mu = \hat{a} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and} \tag{3.14}$$

$$\sigma(\hat{a}) = \frac{\sigma}{\sqrt{n}} \tag{3.15}$$

A very fundamental consequence of equation 3.15 is that the variance or "error" of

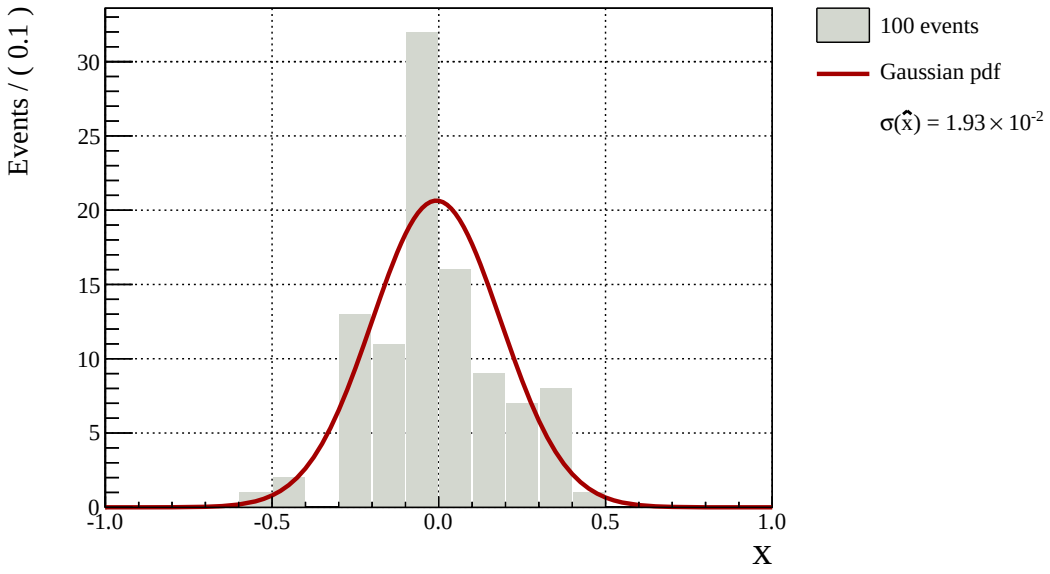


Figure 3.2.: 100 events randomly generated according to a Gaussian pdf. A Gaussian function was fitted to the distribution.

an estimator decreases with increasing number of measurements. Figures 3.2 and 3.3 visualise this phenomenon. While for the low number of measurements in figure 3.2, the exact shape of the Gaussian is not readily apparent to the eye, for the high number

[2]Or, equivalently maximizes $\prod f(x|a)$. Hence the name *maximum* likelihood method.

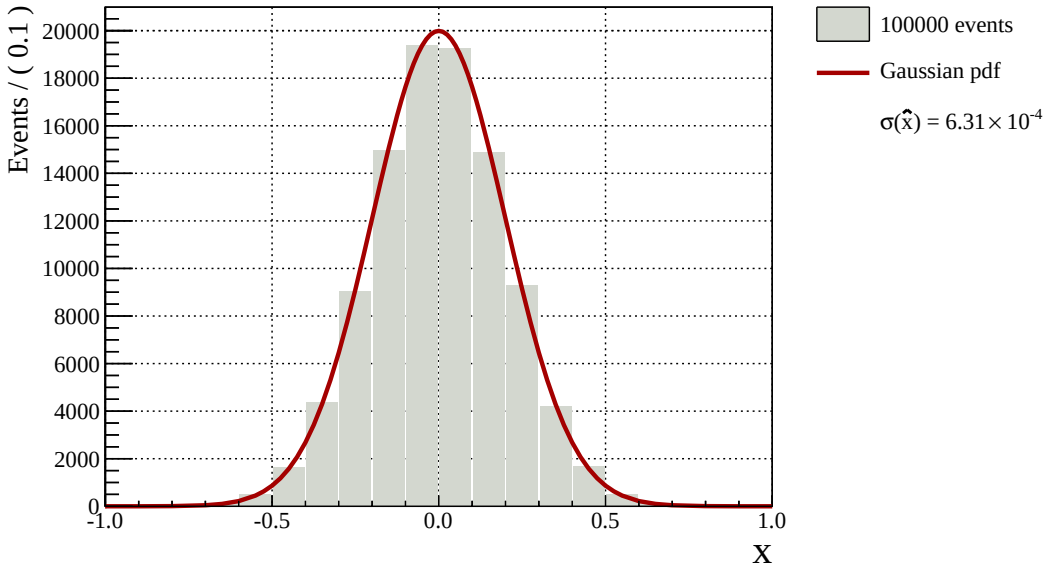


Figure 3.3.: 100000 events randomly generated according to a Gaussian pdf. A Gaussian function was fitted to the distribution.

of measurements in figure 3.3 the shape is much more clearly defined.

The values for $\sigma(\hat{x})$ in the figures quantify the above statement in agreement with equation 3.15: The pdf in figure 3.3 contains a factor 1000 more events and the mean is determined by a factor of $\sqrt{1000} \approx 32$ more precisely. While usually parameter estimation procedures in physics research are much more involved, the above example illustrates an important point: The more measurements there are in a sample, the more precisely the parameter estimation can be done, which is the main motivation for the improvement of the Full Reconstruction that will be described in chapter 5 of this thesis.

3.2. Classification

A major task of data analysis techniques is the classification of an observation, based on one or more properties of the observation. A very important example is the classification, whether a particle was reconstructed correctly or not. The mass of the reconstructed *candidate* is an intuitive and often useful property for this type of classification. Figure 3.1 shows the mass measurements of a number of candidates. Let us define a narrow mass region around $m = 3.096\,\mathrm{GeV}/c^2$ so that the clear peak in figure 3.1 is overed by the region, say $[3.07, 3.11]\,\mathrm{GeV}/c^2$. If a candidate lies within that region, we call it "correct", otherwise "incorrect".
In general, such a simple classification is not the optimal procedure for a given problem, as it does not use any further information that we might have about the measurement. Furthermore, even the information that was used is not used in an optimal way and a binary decision is made, where a probabilistic statement could have been given, e.g. based on the proximity to the maximum value. Many statistical problems require a classification into more than just two categories, but usually these multi-class problems can be reduced to many consecutive binary decisions.
The quality of a classification procedure depends on the amount of information that is available up to the decision-making moment, and also on the quality of the algorithm that is used for the decision making.

3.2.1. Classification in Particle Physics

During a typical high energy physics analysis, one is often presented with a collection of reconstructed particles, called *candidates*, some of which are actually the particle one is looking for, which we call the *signal component*, and some of which are not the particles one is interested in, which we call *background*. In order to obtain results from an analysis, it is essential to separate signal from background. For all selection methods, a compromise has to be found between good *efficiency* and good *purity*, which we define as:

$$\text{efficiency} = \frac{\text{N(selected signals)}}{\text{N(signals in the sample)}}, \qquad \text{purity} = \frac{\text{N(selected signals)}}{\text{N(selected candidates)}}\,. \tag{3.16}$$

Increasing the efficiency by applying a less restrictive classification usually results in more background events passing that classification and thereby lowering the purity. On the other hand, applying a more restrictive classification will improve the purity by rejecting more background but will in turn, of course, lower the efficiency by rejecting a certain portion of the signal component. Figure 3.4 shows an example of purity plotted over efficiency for two different classification procedures. The curve comprises all possible pairings of purity and efficiency for a given classification procedure with varying restrictiveness. Plots of this type can be used as a quality characteristic for the used procedure. Better classification methods are able to reach further into the top right hand corner of the graph, having simultaneously better purity and efficiency than a less powerful classification procedure.

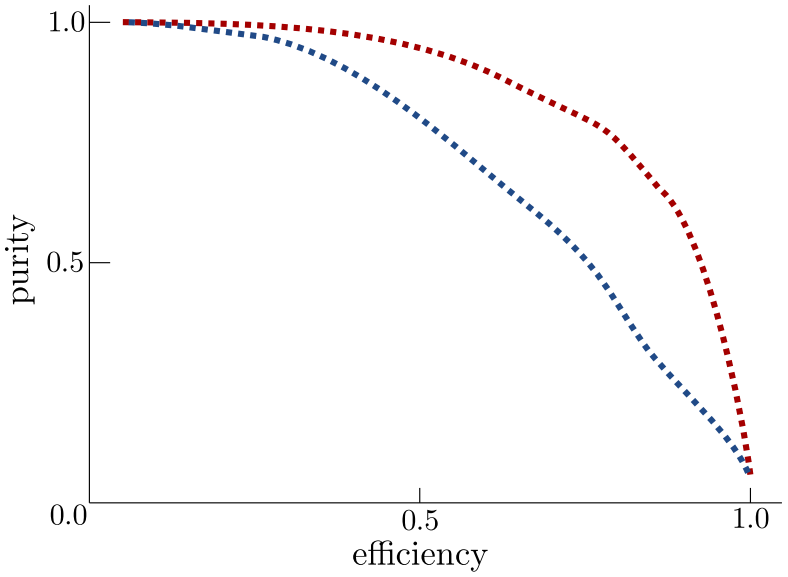


Figure 3.4.: An exemplary purity-efficiency-plot. The dots represent purity-efficiency values of two different classification procedures. The red dots, reaching better purity-efficiency values, stem from an algorithm with higher classification power than the blue dots.

3.2.2. Monovariate Classification

The easiest way to separate signal from background is to apply a binary decision on one single property of a candidate. It is therefore called a *monovariate classification*. If the property in question of a candidate lies above or below a certain requirement, it is considered signal or background respectively. As this procedure resembles cutting away certain areas of the sample, these binary decisions are usually called "cuts". In particle physics, typical properties on which cuts are performed are mass, momentum or impact parameter. In fact, the classification procedure described in the beginning of section 3.2 is a fairly typical monovariate classification. This procedure has the advantage that it is extremely simple and is well-suited for trivial classification tasks. It can be beneficial to perform the more trivial part of a classification (should there be one) by applying cuts. With the trivial decision out of the way, more sophisticated methods are better able to learn the subtle aspects of the remaining sample. Apart from that point, monovariate classification has in practice very limited potential. Usually, multiple cuts on several properties of the candidates have to be performed and it is very difficult to take correlations amongst these properties into account. Neglecting these correlations can result in relatively low efficiencies or purities. Figure 3.5 gives a visualisation of the method of consecutive cuts. Let us assume, that the two dimensions x and y represent correlated properties of a candidate. The task is to select as many signal candidates (represented by the red dots for clarity), and leave out as many background candidates (blue dots) as possible. We may choose either a relatively pure or a relatively efficient

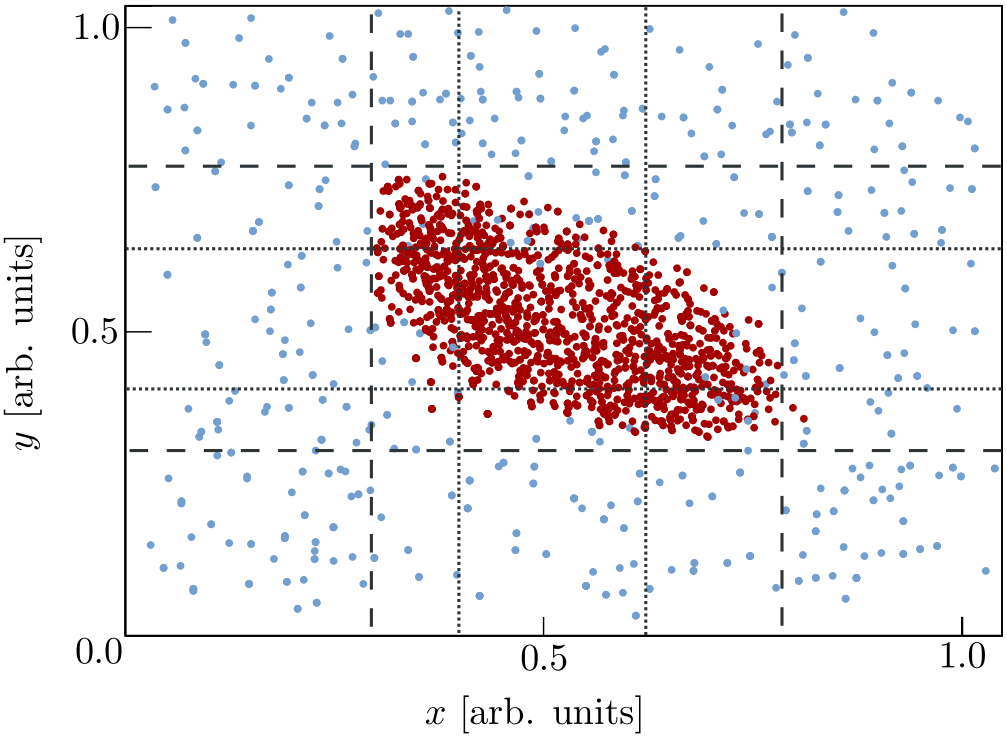


Figure 3.5.: Visualisation of sequential cuts. More effecient selection: dashed lines. Purer selection: dotted lines. Red: signal, blue: background.

selection, represented by the dashed and dotted lines. We can see, that it is unavoidable to either cut away a lot of the signal to achieve a good purity (dotted lines) or to leave a lot of the background in to achieve a good efficiency (dashed lines).

Variable Transformation

A better solution to this problem would be to tranform the variables x and y. The transformation would correpond to a rotation of the coordinate system. A tilt of the coordinate system of $\sim 30°$ in figure 3.5 would allow for an even better selection.

3.2.3. Multivariate Classification

In order to improve the simple selection shown in figure 3.5 and even the variable transformation, it would be desirable to not be forced to cut in straight lines. This could in principle be achieved by varying the cut window width of one variable, depending on the other and thus define an ellipse that wraps around the signal component. While this might seem feasable for two dimensions, it will quickly become a futile task, when there are more than two properties to be used for the discrimination procedure.

A more sophisticated approach for discrimination problems with multiple variables is *multivariate classification*. It is a generic term for a great number of procedures that attempt to map a set of input variables $x_1, \ldots x_d$ to one single output variable y [25]. The goal of these procedures is to combine the discrimination power of many input variables, taking into account correlations amongst the input variables. The result of a multivariate analysis is ideally one very powerful discriminator, suited for a simple cut. In general, there can be more than one output variable. In most cases in physics research, and also in this thesis, however, only one output variable for the classification of signal and background is used.

To perform the desired mapping, a mathematical function is used that contains a number of adjustable parameters, that are commonly referred to as *weights*. We can write this function as

$$y = y(\vec{x}, \vec{w}) \,,$$

where $\vec{x}$ is the vector of input variables and $\vec{w}$ is the vector of weights. In general, the exact form of the mapping function is not known and has to be found using a data set of examples, where the correct classification is known. This data set is called a *training data set*. In particle physics, usually simulated data is used. For other applications, historical data can also be a source for the training data. The process of adjusting the weights of the mapping function is called a *training* and has to be done for all multivariate classification methods.

Multivariate techniques are in principle capable of taking correlations amongst the input variables into account. The application of these techniques can, however, be rather involved. Simplified models can deliver quite good results when correlations between the different variables are small. An important example of an advanced multivariate technique is a *Neural Network*. Neural Networks were used prominently in this thesis, as a major part of the NeuroBayes package. A detailed description of NeuroBayes, including the inner workings of a Neural Network, is given in chapter 3.3.

3.3. NeuroBayes

The NeuroBayes package is a very powerful multivariate classification method. One of its core components is a Neural Network that maps a vector of input variables $\vec{x}$ to an ouput variable o_{NB}. A sophisticated algorithm transforms the raw, unprocessed input variables into suitable input variables for the Neural Network [26]. The Neural Network then learns the actual classification for the training data set and produces the output. The NeuroBayes package is capable of operating in two different modes: classification-mode or density-mode.

- The classification mode is the most frequently used mode in particle physics and also the only mode used in this thesis. For a classification task, which is usually the decision if a candidate is signal or background, the network's task is to map the input variables to a single output variable, while taking into account the correlations amongst the input variables. The produced output is a measure of the signal probability for the candidate. Given the prior of a candidate to be a signal, the output can be transformed easily into a real probability in the Bayesian sense (see equation 3.4). A single output node is used for this measure of probability.
- In density mode, there is not one, but many output nodes, which enable the network to estimate a probability density function, from which one can obtain estimates for absolute values rather than just a probability.

As the NeuroBayes package, and its classification mode in particular, is used extensively in the analysis described in this thesis, it will be explained in more detail in the following.

3.3.1. Theory

Let us assume that we have a random variable t distributed according to its probability density function (pdf) $f(t)$ (see section 3.1.4). In particle physics, our knowledge of the pdf will usually come from simulated data, known as *Monte Carlo simulations*. [3] Given the knowledge of this shape, it is easy to say what outcome a measurement will have *on average*. NeuroBayes, however, will aim to produce a better estimate for *each single event* taking into account a set of measured values $\vec{x_i}$. This means that NeuroBayes aims to find the conditional probability density of variable t, given the measured input values $\vec{x_i}$: $f(t|\vec{x_i})$. If there should be no additional information in the measured values $\vec{x_i}$ for the current classification problem, then our best estimate is the pdf: $f(t|\vec{x_i}) = f(t)$. If there is, however, additional information in the input vector $\vec{x_i}$, one should be able to obtain a better estimator for a given event with the conditional probability density $f(t|\vec{x_i})$ than with the pdf $f(t)$ alone.

[3] If the signal can be seen clearly in a distribution from real measured data, the pdf can be also estimated from a large number of measurements. The pdf of the measurement shown in figure 3.1 could in principle be estimated easily.

Preprocessing

The training of a Neural Network is a crucial step in the application of NeuropBayes. While it would technically be possible to perform such a training with the raw input variables, a sophisticated preprocessing procedure can increase the training speed and the robustness of the classification. As shown in the example in figure 3.6, all the

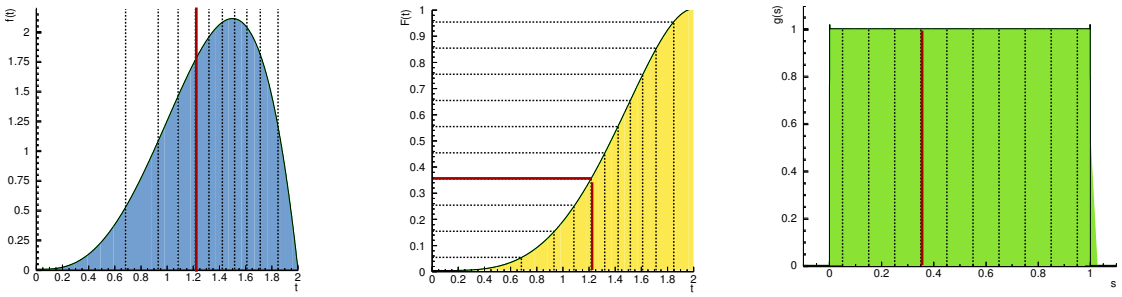


Figure 3.6.: Variable transformation during preprocessing.

distributions $f(t)$ of the input variables (left plot in figure 3.6)) are transformed to other distributions $g(s)$ during preprocessing by a non-linear variable transformation $F : t \rightarrow s$ (middle plot in figure 3.6), so that the resulting distribution (signal and background) is flat and within the interval 0 and 1 (right plot in figure 3.6). This avoids that outliers in the training data set have the power to completely dominate the Neural Network training.[4] It also avoid numerical problems, should the input variables have extermely small or large numerical values.

Flattening the Distributions

As an example, we use the mass distribution of D meson candidates, that have a nominal mass of ~ 1.865 GeV/c^2. The flattened distribution can be seen in figure 3.7. The structure of the original mass distribution can still be seen in this distribution with many signal events and relatively few background events around the nominal mass.

Normal-Distributed Purity

The flattening of the distribution is followed by obtaining the purity in each bin of the new, flattened distribution. The resulting purity distribution lies within the interval 0 and 1. Optionally, a spline fit of the resulting purity distribution is performed, as shown in figure 3.8. The obtained purity distribution is then again transformed, optionally according to the performed fit, to have a mean of 0 and a width of 1. The result of this final transformation can be seen in figure 3.9. These transformations inprove the learning speed of the Neural Network and avoid neuron saturation.

[4] These outliers are usually unreliable data anyway.

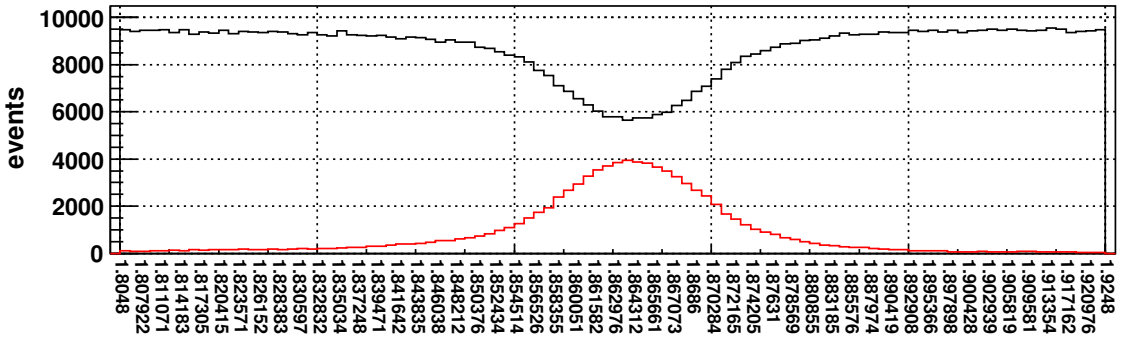


Figure 3.7.: Flattened distribution, red: signal, black: background.

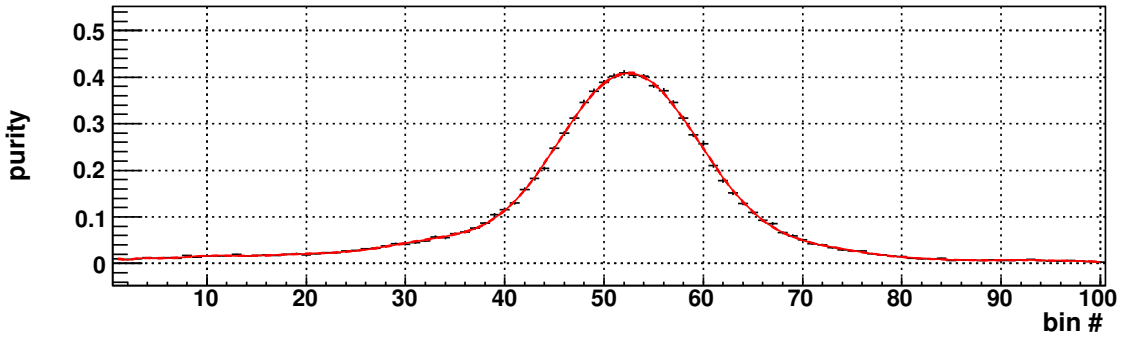


Figure 3.8.: Purity distribution, red: spline fit, black: measurement.

Decorrelation Of Input Variables

As a final step before the training of a Neural Network, the input variables are decorrelated from each other. For that purpose, the known classification of the training sample (the "target") is added to the variable set. In the next step, the correlation matrix of the preprocessed input variables and the target is calculated. This matrix is then diagonalised by iterative Jacobian rotations. The rotated input vectors are finally normalized by their eigenvalues. This results in an input dataset that is rotated in such away that all correlations are to the target.

Separation Power Of Each Variable

To estimate the separation power of this one variable compared to the whole neural network, a purity-efficiency plot is created for each input variable, which can be seen in figure 3.10, where now the black line indicates the purity-efficiency curve for the current variable alone, while the red curve is for all of the input variables combined.

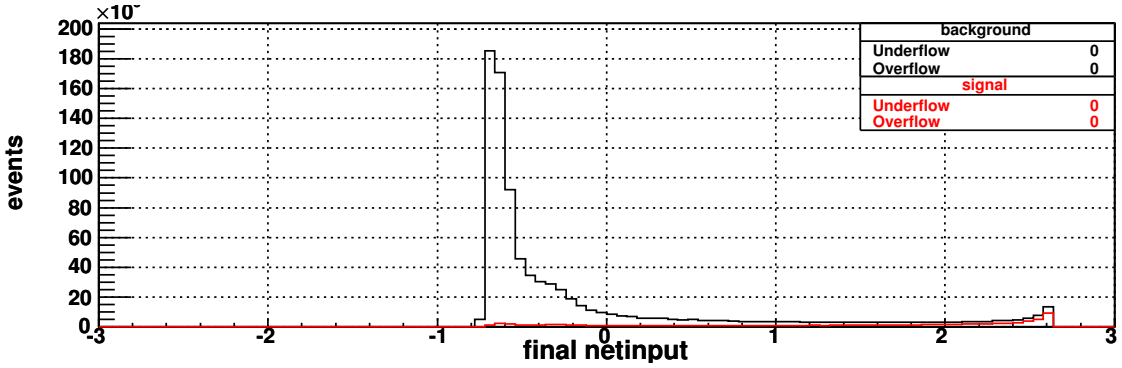


Figure 3.9.: Final distribution, red: signal, black: background.

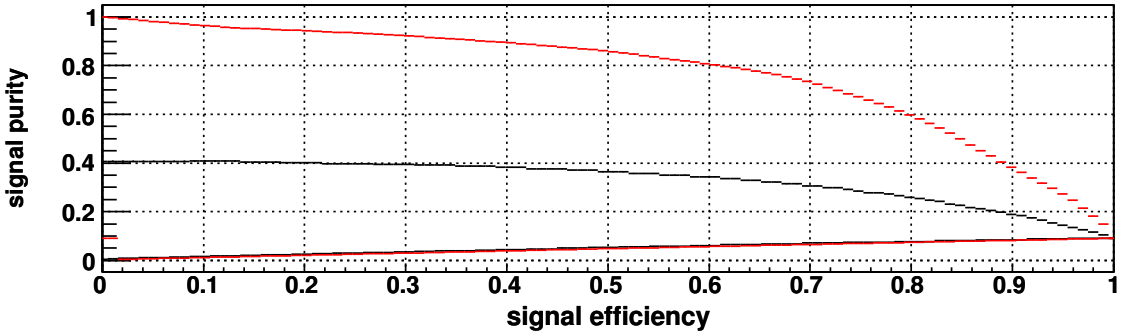


Figure 3.10.: Purity efficiency plot, red: current variable only, black: all variable combined.

3.3.2. The Neural Network

The neural networks utilised in the NeuroBayes package share their common topology of a three-layered feed forward neural network. Figure 3.11 shows an exemplary neural network with 4 input nodes ($i_1 \dots i_4$), 3 hidden nodes ($h_1 \dots h_3$) and one output node (o). It can be shown, that any neural network can be simplified to a three layer network, if there are enough nodes in the hidden layer. Neural networks are based on the working principles of neurons in nature. The input nodes are connected to the nodes in the hidden layer, who produce, when activated, an output to the output node. The output y of the nodes in the hidden layer can be expressed as

$$y = S\left(\sum_i w_i x_i\right) \tag{3.17}$$

where x_i are the different inputs of the nodes with their respective weights w_i that can assume any real value. The widths of the connecting lines between the nodes in figure 3.11 represent their individual weights. The function $S(x)$ is called the activation function and is chosen to be a sigmoid function. This function has the virtue to map

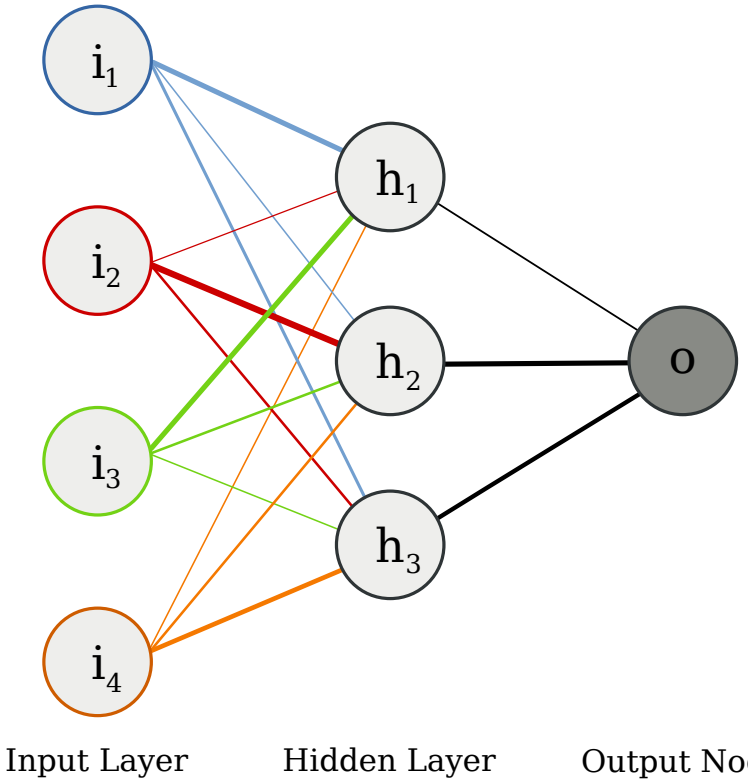


Figure 3.11.: Topology of a three-layered feed forward neural network.

the whole range of real numbers to the interval between −1 and 1:

$$S: \quad]-\infty,\infty[\quad \longrightarrow \quad]-1,1[\tag{3.18}$$

The sigmoid function is defined by

$$S(x) = \frac{1}{1+e^{-cx}} \tag{3.19}$$

and shown in figure 3.12, where we can see that this function has

- two approximately constant regions at relatively large and small values of x,
- a linear region around $x = 0$
- an exponential region at negative values near zero and
- a logarithmic region at positive values near zero.

Where exactly these regions are depends on the choice of the parameter c, which can also be seen in figure 3.12. If we now look at the whole three layered network, we can

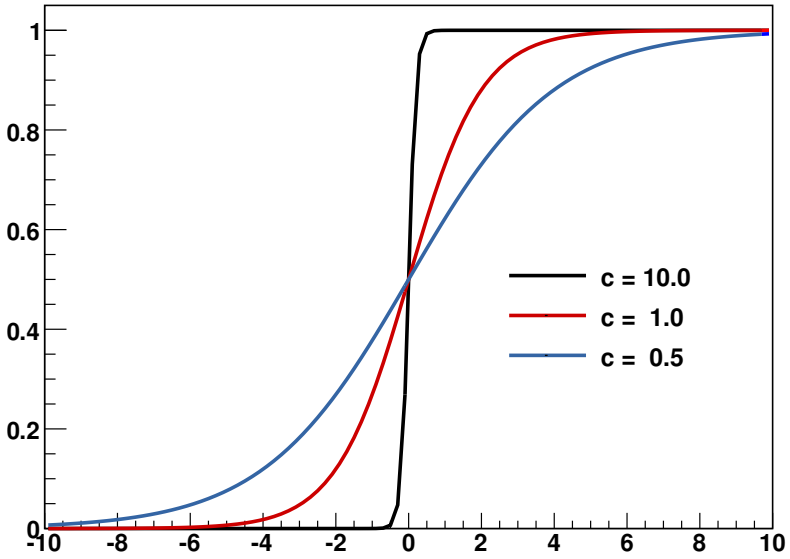


Figure 3.12.: The sigmoid function for different values of c.

express the output of the node in the output layer o using equation 3.17 twice as:

$$o = S\left(\sum_l w_l^{2\to 3} \cdot S\left(\sum_k w_{kl}^{1\to 2} \cdot x_k\right)\right) \tag{3.20}$$

where $w_{kl}^{1\to 2}$ is the weight of node k of the input layer to node l in the hidden layer and $w_l^{2\to 3}$ is the weight of node l in the hidden layer to the the single output node.

3.3.3. Training

In order to utilise a neural network, first a training has to be performed, the results of which are then applied to characterise the data in question. To perform the necessary training we need a data-set, for which the true category of an input event ("signal" or "background") and thus the desired value of the output node is known. Generally, this can be historical data, but in particle physics, Monte Carlo simulations are often used. The neural network is then trained with a sample of signal and another sample of background events. The target for the neural network is known for each event in the training samples and can be either $t = 1$ for signal or $t = -1$ for background. The process of a neural network training is, after setting the weights described in formula 3.17 initially randomly, the readjustment of these weights in order to describe the known target values as good as possible. In other words, the neural network training is essentially a fit with the weights as free parameters. To measure the quality of this fit, a loss function is used, which has to be minimised for all events i in the sample

simultaneously. This can either be a quadratic (χ^2) or an entropy (E_D) loss function:

$$\chi^2 = w\frac{1}{2}\sum_i (t_i - o_i)^2 \,, \tag{3.21}$$

$$E_D = w\sum_i \log\left(\frac{1}{2}(1 + t_i \cdot o_i + \epsilon)\right), \tag{3.22}$$

where t_i are the known targets (+1 or −1) and o_i are the corresponding network outputs. ϵ is a regularisation constant which helps to avoid numerical problems with an untrained network. It is reduced to 0 after a few iterations. w is a weight parameter, introduced to minimise overfitting and to maintain generalisation ability.

For classification problems, the entropic loss function is generally used. It has the advantage, that its value increases towards infinity for completely wrong classification ($o_i = +1$ while $t_i = -1$ or $o_i = -1$ while $t_i = +1$). Consequently, this error is automatically strongly punished and the distinction is quickly learned by the neural network. The two above mentioned loss functions can only be evaluated after a sufficiently large subsample of the available data has been used. In order to achieve an improvement in the prediction power of the network, it is common to perform multiple iterations on the dataset.

Over-Training

This training process by itself contains one common problem of neural networks, however, which is over-training. When the signal (or background) fraction in the training is not large enough and there are many nodes in the input layer, the neural network might be able to learn the signal (or background) events by heart. This means that a few very specific pattern might be able to trigger an extreme network output value. While this value would be true for the one corresponding event, the input pattern leading to the value has in general not got a general discrimination power.
To avoid this problem, the concept of decaying weights is used, which means that the weights of all nodes w_i are moved towards 0 after a fixed interval, which can either be one iteration or a fixed number of events. In a figurative way, one could say that the neural network is programmed to forget part of what it has learned to avoid over-training. Additionally the training is not performed on the whole available data-set, but a small fraction of signal and background events are left out of the training and used for validation of the neural network. Over-training would be indicated by exceptionally bad performance on such a validation sample.

Signal And Background Fraction

While it would be possible, and it sometimes indeed is practical, to train the neural network with the same signal to background ratio as expected on data, it is at other times more practical and sometimes even only possible using a different ratio. If, for

example, the desired signal is very rare in the training data, a training would not learn to distinguish the few signal events from the millions of background events. The network would rather try to learn something from statistical fluctuations that surpass the signal by far and therefore also dominate the loss function. Therefore, a training with a higher signal fraction is the only way, in which such rare signals can be detected. This would, however mean, that the output of the NeuroBayes training could no longer be interpreted as a probability in the Bayesian sense. It can, however be transformed to regain that property (see section 3.4).

3.3.4. Performance

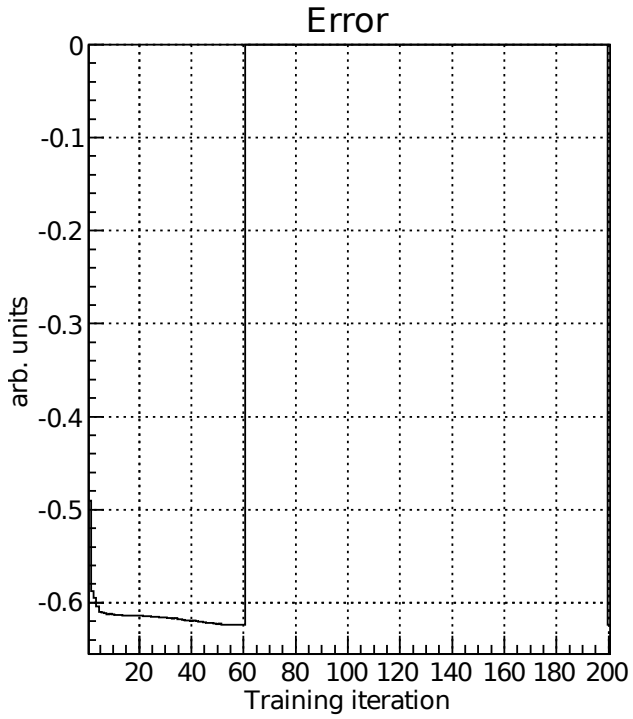


Figure 3.13.: The error function during a NeuroBayes training. The training was stopped approximately after the 60th iteration.

After a performed neural network training, there are several plots and figures that provide information about the quality of a network training. The error plots, shown in figure 3.13, show the value of the loss function during the individual training iterations. During a normal training, the value of the error function should decrease towards a minimal value, which then indicates that a local minimum of the loss function is found. As it would not bring much improvement to do further iterations at this point, NeuroBayes can be set to stop the training here. This was in fact done in the

exemplary plot in figure 3.13 around the 60th iteration.
The quality of the achieved classification can be described by the ability of the network to separate signal and background of a sample. This ability is shown in figure 3.14. We

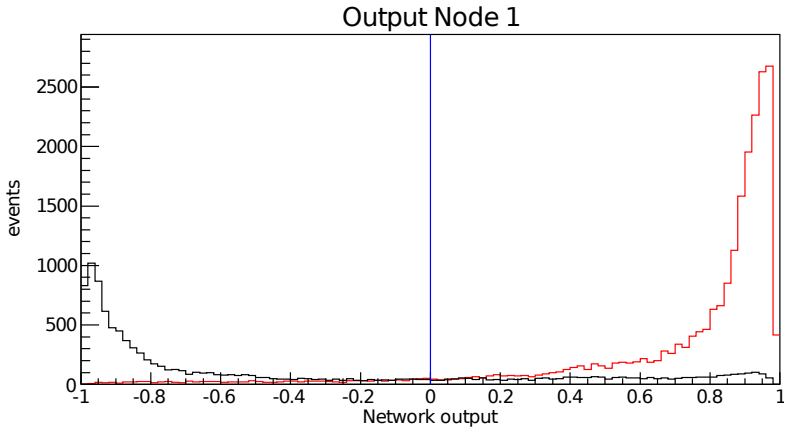


Figure 3.14.: The distribution of the NeuroBayes neural network output, red: signal, black: background.

can see a small background peak (black line) at low NeuroBayes outputs and a large signal peak (red line) at high NeuroBayes outputs. The good separation power of this training is immediately visible.

The separation, as good as it may look on any plot like in figure 3.14 can never reach 100% signal purity in any realistic situation and, as mentioned in section 3.2, we always have to deal with a compromise between purity and efficiency. The purity-efficiency plot in figure 3.15, automatically generated by NeuroBayes also gives valuable information about the quality of the neural network training. The upper curve shows a usual purity-efficiency curve, that is obtained by calling the candidates above a certain cut "signal" and those below that cut "background". The lower curve is obtained doing just the opposite: Calling the candidates below a certain cut "signal" and those above the cut "background".

Finally, there is one plot that summarises the importance of the given input variables. It is a colour-coded correlation matrix of the input variables and the target before decorrelation. It assigns each variable a row and a column and shows the correlation amongst the variables and the correlation of each variable to the target. An exemplary correlation matrix is shown in figure 3.16. A strong (anti-)correlation to column no. 1, which is the known truth during a training, indicates the importance of a variable. If this variable is, however, strongly correlated to another variable, only one of these is in fact important for the neural network, as they carry similar information.

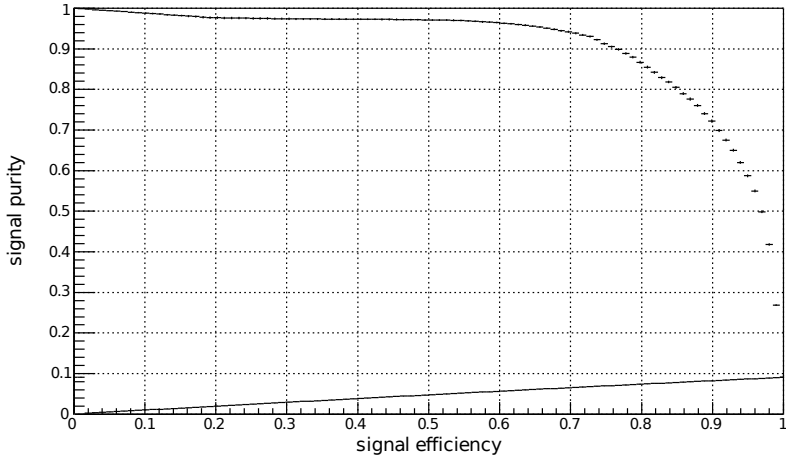


Figure 3.15.: The purity-efficiency plot, generated by NeuroBayes.

3.4. Interpretation Of The NeuroBayes Output

The NeuroBayes algorithm is constructed in such a way that the output of a training, rescaled to lie within the interval $[0,1]$, is the actual posterior signal probability of a candidate. The assumption that is made for the above statement is that the signal to background ratio used for the training is the same as one would expect in the real application. As mentioned in chapter 3.3.3, it is sometimes beneficial or even necessary to artificially change the signal to background fraction for the training in order to improve the separation power. The artificial change requires a transformation to be applied to the NeuroBayes output, in order to still be interpretable as the actual posterior signal probability of a candidate.
For the calculation of this correction [27, 28], we start with two instances of Bayes' Theorem for two alternative theories S and B for given data D:

$$P(S|D) = \frac{P(D|S) \cdot P(S)}{P(D)} , \tag{3.23}$$

$$P(B|D) = \frac{P(D|B) \cdot P(B)}{P(D)} . \tag{3.24}$$

We relate them to each other [29] and define Bayes' factor Λ and the prior O:

$$\frac{P(S|D)}{P(B|D)} = \underbrace{\frac{P(D|S)}{P(D|B)}}_{=\Lambda(D|S)} \cdot \underbrace{\frac{P(S)}{P(B)}}_{=O(S)} \tag{3.25}$$

The posterior odds are then given by

$$O(S|D) = O(S) \cdot \Lambda(D|S) . \tag{3.26}$$

We can express the Bayes' factor by using the output of a NeuroBayes training o_t using data set t as data. Also, we can assign the two alternative theories S and B to correspond to the signal hypothesis (S) and background hypothesis (B):

$$\Lambda(D|S) = \frac{P(o_t|S)}{P(o_t|B)} \tag{3.27}$$

Here, $P(o_t|S)$ is the likelihood for a certain NeuroBayes output given the candidate in question was indeed a signal. Accordingly, $P(o_t|B)$ is the likelihood for a certain NeuroBayes output given the candidate in question was background. The conditional probabaility for a candidate to be signal or background can also be expressed in terms of a given NeuroBayes output o_t. Keeping in mind that the probabilities for the signal hypothesis and the background hypothesis must add up to exactly = 1, we get

$$o_t = P_t(S|o_t) \qquad \text{signal} \tag{3.28}$$

$$(1 - o_t) = P_t(B|o_t) \qquad \text{background} \tag{3.29}$$

We can apply Bayes' theorem in a similar manner to equation 3.25:

$$\frac{P_t(S|o_t)}{P_t(B|o_t)} = \underbrace{\frac{P(o_t|S)}{P(o_t|B)}}_{=\Lambda(o_t|S)} \cdot \frac{P_t(S)}{P_t(B)} \tag{3.30}$$

and insert equations 3.28 and 3.29 into equation 3.30:

$$\Lambda(o_t|S) = \frac{o_t}{1 - o_t} \cdot \frac{P_t(B)}{P_t(S)} \tag{3.31}$$

We thus obtained the factor $\Lambda(o_t|S)$, independent of the signal to background ratio used in the training dataset. When we wish to apply the NeuroBayes training on a data set with a different signal to background ratio than in the training, we have the means to correct it in such a way, that it is the actual posterior signal probability of a candidate. We use equation 3.25 once more, now with a prediction dataset p in mind:

$$\underbrace{\frac{P_p(S|o_p)}{P_p(B|o_p)}}_{=\frac{o_p}{1-o_p}, \text{ see 3.28, 3.29}} = \Lambda(o_t|S) \cdot \frac{P_p(S)}{P_p(B)} \tag{3.32}$$

We can now solve equation 3.32 for o_p and thus gain the desired correction:

$$o_p = \left[1 + \left(\frac{1}{o_t} - 1\right) \frac{P_p(B)}{P_p(S)} \frac{P_t(S)}{P_t(B)}\right]^{-1} \tag{3.33}$$

Equation 3.33 allows us to re-interpret the output of a NeuroBayes expert as the actual posterior signal probability of a candidate, regardless whether the signal to background

ratio that was used in the training sample is the same as on the prediction dataset or not. The information needed for this re-interpretation is the signal to background ratio in the training dataset and in the prediction dataset. This probabilistic interpretation of the NeuroBayes output will be used prominently in the Full Reconstruction algorithm, described in detail in chapter 5.

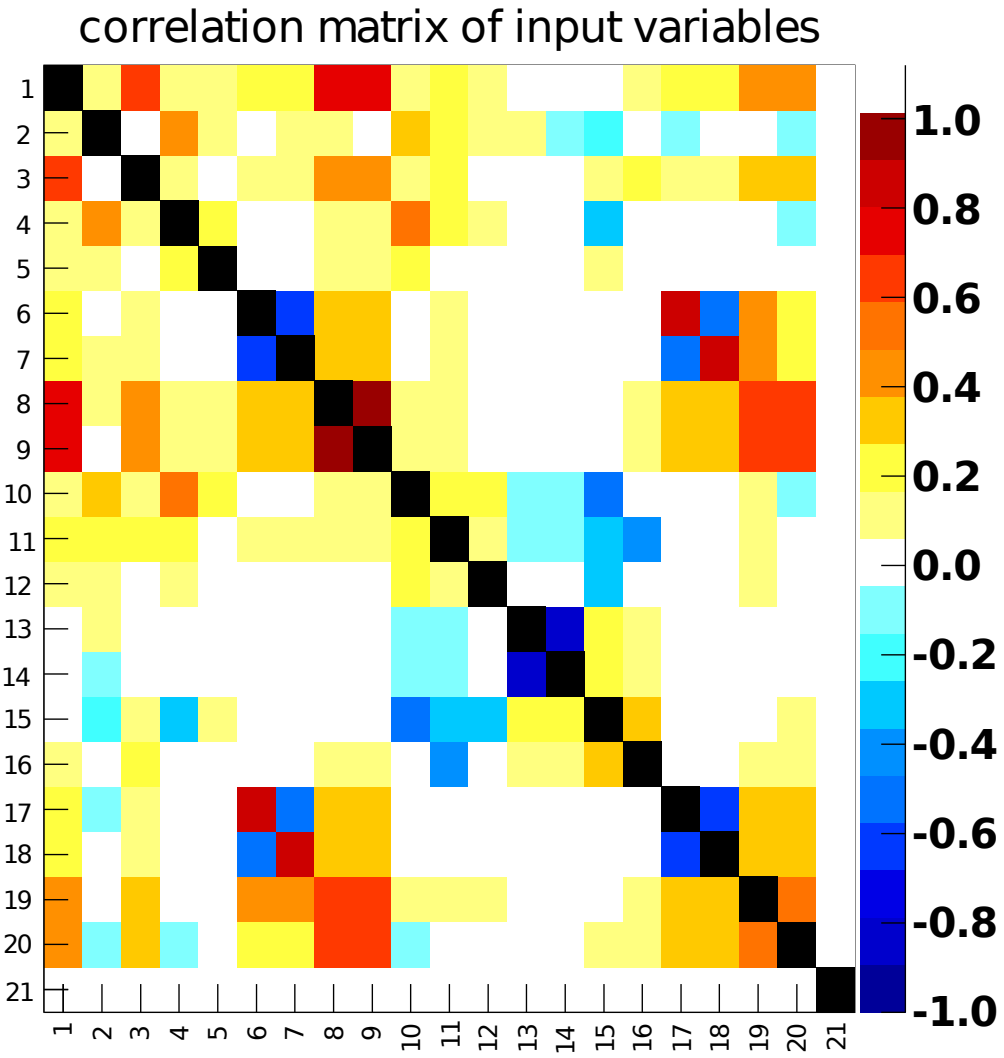


Figure 3.16.: Correlation matrix of a NeuroBayes training.

4. The Experimental Setup

4.1. Particle Accelerators

The vast majority of the matter that surrounds us consists only of electrons, protons and neutrons. First evidence that there could exist other particles was found in the analysis of cosmic particles. Neutrinos, muons, pions and kaons were discovered in these experiments. These newly found particle did not come from another place in the universe, where there are other forms of matter, but they were *produced*. The energy densities required to produce the heavier quarks (c, b and t) occur in nature in cosmic showers. These originate from interstellar high-energy particles that collide with atoms in earth's atmosphere. Under controlled circumstances, the heavier quarks (c, b and t) can only be produced in modern particle accelerators. Although the several machines in existence differ largely in their details, they share the common principle of accelerating stable, charged particles using electric fields and bringing the accelerated particles to collision. The collision can either occur with a fixed target or with other accelerated particles of a second beam moving in the opposite direction. An important difference between fixed-target and two-beam collisions is the achievable centre of mass energy $\sqrt{s}$ [30]. The historically first experiments were conducted with fixed targets, an experimental setup where the accelerated particles with energy E_1 and momentum p_1 are directed onto a fixed target with energy m_2 and no momentum. The achievable energy in the centre of mass frame $\sqrt{s}$ for this setup is

$$s = 2E_1 m_2 + m_1^2 + m_2^2 \,, \tag{4.1}$$

$$\sqrt{s} \sim \sqrt{E_1} \,. \tag{4.2}$$

In more recent two-beam colliders, two relativistic particles with energies E_1, E_2 are colliding head-on with momenta p_1, p_2. In this case, the centre of mass energy $\sqrt{s}$ can be calculated as

$$s = 2(E_1 E_2 + p_1 p_2) + 2M^2 \approx 4E_1 E_2 \,, \tag{4.3}$$

$$\sqrt{s} \sim E \,, \tag{4.4}$$

where $E^2 = E_1 \cdot E_2$. We can see that the available energy for the creation of new particles grows with $\sqrt{E}$ for fixed-target machines, but with E for two-beam colliders.
Modern colliders can be categorised by the types of particles that are accelerated:

Lepton colliders have the advantage to utilise particles that have no substructure. Usually, electrons are brought to collision with positrons. Thus, the available centre of mass energy in a collision can be chosen precisely. Additionally, in electron-positron collisions, the initial quantum-mechanical state is known exactly. The

largest drawback of electron-positron colliders is the huge amount of synchrotron radiation they produce. This type of radiation is emitted by any charged particle under acceleration. The consequence of this is energy loss ΔE for the particles in a collider. The energy loss for circular acceleration can be expressed as [1]

$$\Delta E = \frac{4\pi}{3}\left(\frac{e^2\beta^3\gamma^4}{\rho}\right), \tag{4.5}$$

where ρ is the bending radius, β the velocity of the charged particle and $\gamma = (1 - \beta^2)^{-1/2}$. With $\beta \sim |\vec{p}|/m$, it becomes clear that the lighter the accelerated particle is, the larger its energy loss by synchrotron radiation is:

$$\Delta E \sim \frac{1}{m^4} \tag{4.6}$$

Hadron colliders usually accelerate protons or protons and anti-protons and bring them to collision. As the energy loss due to synchrotron radiation is strongly mass dependent, according to equation 4.6 we can easily see that this energy loss is a factor of $(m_p/m_e)^4 \approx 10^{13}$ smaller for protons than for electrons. Therefore much higher energies can be reached with hadron colliders. The disadvantage of these colliders is the substructure of the accelerated particles. Only one constituent of each hadron plays an active role in the collision, and this constituent carries only a small fraction of the entire energy of the hadron. As this fraction varies strongly, a precise choice of energy is impossible.

Hadron colliders, like the LHC and the Tevatron, yield enough energy to push the boundaries of achievable energies and thereby the potential to discover new particles further. Lepton colliders such as LEP, PEP-II or KEK B can achieve a much higher precisions in the collision energy. They are therefore more suited to perform high-precision measurements that also allow the search for new physics and the indirect search for new fundamental particles.

4.2. The KEKB Accelerator

The KEKB accelerator is an asymmetric electron-positron collider [31, 32, 33] located at Tsukuba, Japan. Its construction began in 1994 in the underground tunnels that were formerly occupied by the TRISTAN collider. Figure 4.1 shows a schematic drawing of the KEKB accelerator. KEKB consists of two rings, 3.016 km in circumference, constructed side by side: a low energy ring (LER) that contains a positron current of approximately 1600 mA at an energy of $E_+ = 3.5$ GeV and a high energy ring (HER) that contains an electron current of approximately 1200 mA at an energy of $E_- = 8.0$ GeV. The two beams yield a centre of mass energy $\sqrt{s}$ of 10.58 GeV, which is approximately the energy of the $\Upsilon(4S)$ resonance:

$$\sqrt{s} = \sqrt{4E_+E_-} \approx 10.58 \text{ GeV} = E_{\Upsilon(4S)}\,. \tag{4.7}$$

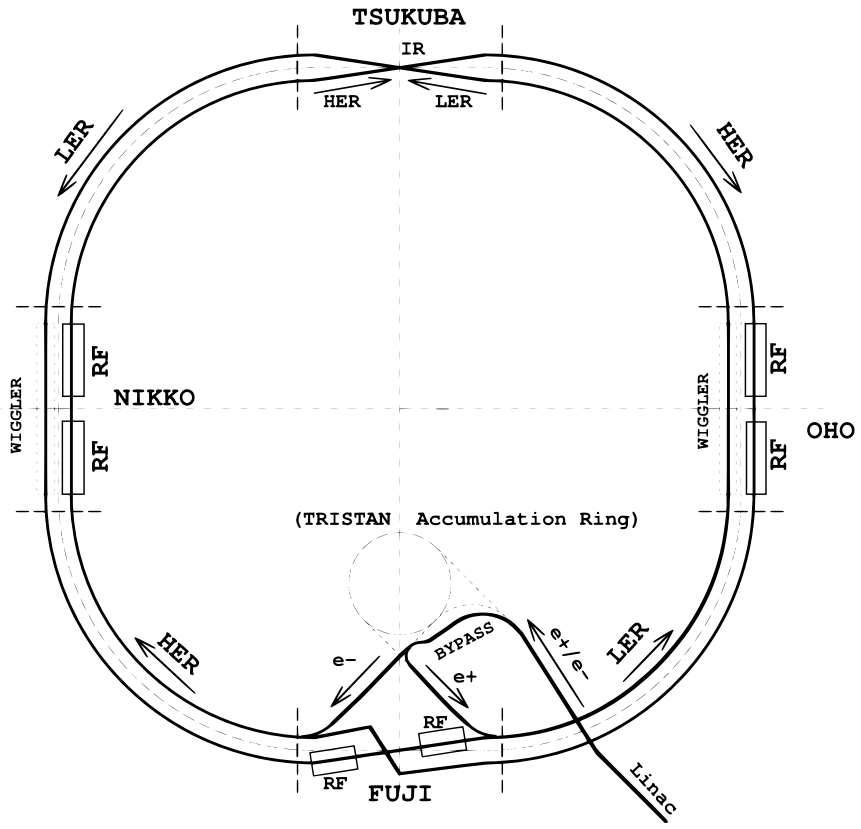


Figure 4.1.: The KEKB accelerator.

The asymmetry between the two rings can be expressed by the resulting Lorentz-boost parameter $\beta\gamma$ of the $\Upsilon(4S)$ resonance:

$$\beta\gamma = \frac{E_- - E_+}{\sqrt{s}} = 0.42\,. \tag{4.8}$$

The need for this asymmetry is derived from physics considerations. One of the main goals of the experiments is to measure CP violations in the decay of B mesons. Due to the relatively small lifetime of the B mesons, a large Lorentz-boost is beneficial for distinguishing the B decay vertices.

Although we speak of a "particle beam", the shape of the particles in a collider is not a constant flow, but rather thousands of particle bunches, a few centimetres in length with equal spacing of approximately 2m between them. This topology is due to the mechanism of acceleration by high frequency electromagnetic fields that self-actingly produces this bunch structure.

The production rate for new particles in collisions is determined not only by the achieved

energy, which is only an indicator, whether a certain type of particle can be produced at all with the available energy, but by the *luminosity* $\mathcal{L}$ of the collider:

$$\text{production rate} = \text{luminosity} \cdot \text{cross section} \tag{4.9}$$

The luminosity of a collider depends only on the properties of the beam and is given by

$$\mathcal{L} = \frac{N_1 N_2 f}{4\pi\sigma_x\sigma_y}\,, \tag{4.10}$$

where $N_{1,2}$ is the number of bunches in the HER and LER ring respectively, σ_x and σ_y are the spatial dimensions of the bunches and f is the collision rate of these bunches.

The KEKB accelerator operates using trains of 1582 bunches, separated by gaps of 270 buckets (empty bunches) which are needed for a safe beam abort procedure. The bunch crossing frequency is at $f \approx 508.9$ MHz and bunches are crossed at an angle of 22 mrad in the horizontal plane. The KEKB has only got one interaction region, which lies in the Tsukuba area, where the Belle detector was located. The KEKB accelerator was designed for a luminosity of $1 \cdot 10^{-34}$ cm^{-2}s^{-1}, but has reached a record luminosity of $2.1083 \cdot 10^{-34}$ cm^{-2}s^{-1} in June 2009, towards the end of its runtime. It was shut down on June 30th, 2010 and will be upgraded to Super KEKB, with decreased beam size for an even higher luminosity at the $\Upsilon(4S)$ resonance.

4.3. The Belle Detector

The Belle detector was a multi purpose particle detector designed to perform high precision measurements of asymmetric electron-positron collisions at the $\Upsilon(4S)$ resonance [34] and to carry out quantitative studies of rare B meson decays with very small branching fractions [35]. Figure 4.2 shows a schematic side view of the detector. The Belle detector consists of several, specialized sub-detectors, that shall be described

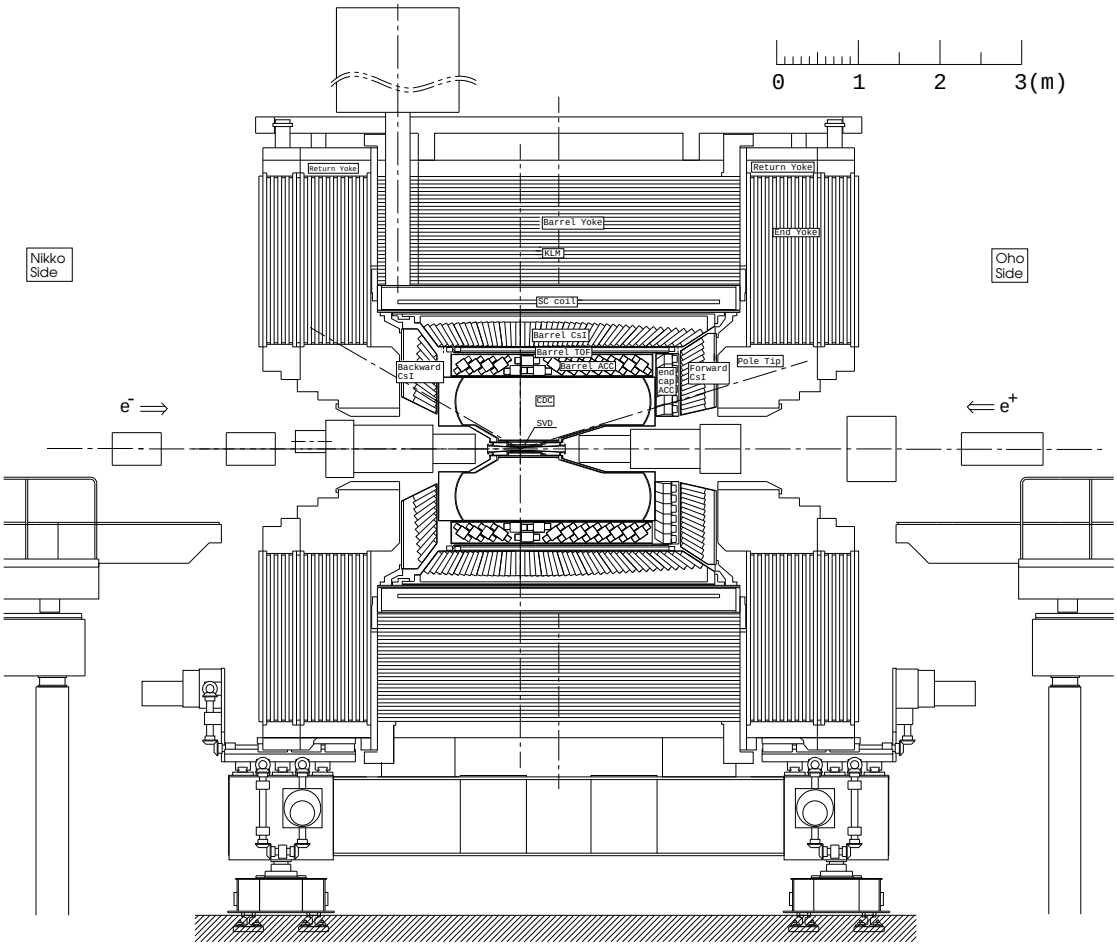


Figure 4.2.: Side view of the Belle detector. [35]

in the following. As a lot of components of the detector have roughly cylindrical shape, it is convenient to describe the geometry using cylinder coordinates. Figure 4.3 gives an overview of the used coordinates.

4.3.1. Interaction Region

The entire Belle detector is build in several layers around the interaction region, where the e^+e^- collisions occur. The magnetic field of the Belle detector is aligned with the

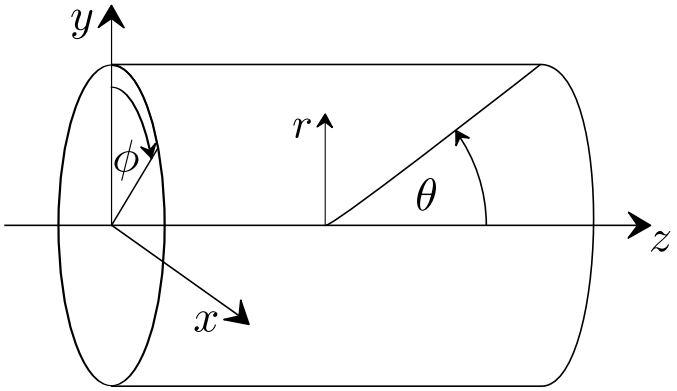


Figure 4.3.: Coordinates used in the description of the Belle detector. [35]

direction of the low energetic e^+ beam. The high energetic e^- beam intersects the low energetic beam at an angle of 22mrad in the interaction point. This finite angle avoids parasitic collisions and makes it unnecessary to use separation bend magnets, allowing for higher luminosity.
After the collision products were produced in the interaction region, they have to cross the walls of the beam line before they reach the next layers of the detector. Coulomb scattering during this process is a limiting factor for the vertex position resolution. Another important factor for the resolution is the spatial proximity of the first layer of the Vertex detector to the interaction point. Thus, it is important to have a beam pipe with low diameter and thin walls. Two Beryllium walls of 0.5mm thickness were used, separated by a layer of Helium of 2.5mm thickness. The Helium layer is necessary as a gas cooling system to avoid damage to the beam pipe and detector induced by a few hundred watts of beam induced heating. The beam pipe is covered by a 20μm thick gold sheet to reduce certain backgrounds. The inner edge of this compound wall is located at a radius of r = 20mm from the interaction point. A schematic view of the interaction region and the beam line can be found in figure 4.4. In the process of the update of the Silicon Vertex Detector (see section 4.3.2), the radius of the beam pipe was reduced from 20 mm to 15 mm [36].

4.3.2. Silicon Vertex Detector

As one of the main goal of the Belle detector is to measure time-dependent CP-asymmetries, precise measurement of the position of the interaction vertex is essential. The Silicon Vertex Detector (SVD) was designed for that purpose. Additionally, the decay vertices of relatively long-lived particles such as D mesons can be measured by the SVD improving the overall reconstruction quality. The SVD also delivers an important input for the tracking algorithm and can therefore also help to achieve a good tracking quality.
The original Silicon Vertex Detector was replaced multiple times by almost identical spares due to radiation damage. An updated construction was installed in September

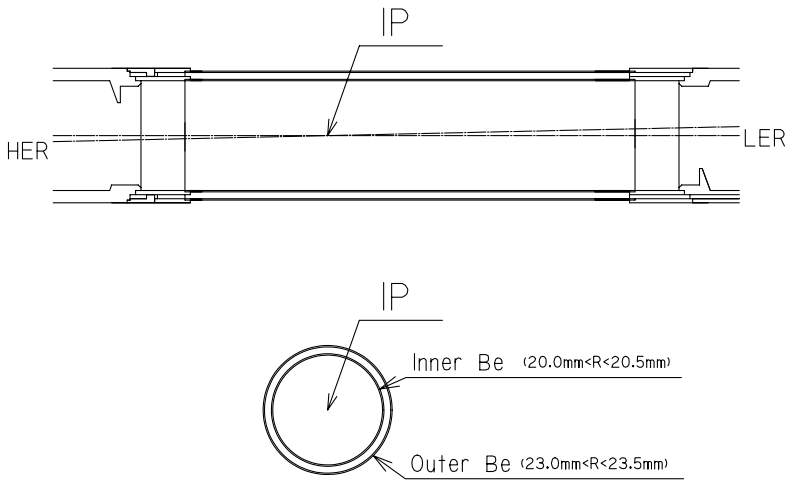


Figure 4.4.: Schematic view of the interaction region and the beam line. [35]

2003 and remained in usage until the end of the runtime. It was named Silicon Vertex Detector 2 (SVD2) and as the bulk of the Belle data was taken with the SVD2 in use, only the SVD2 shall be described in further detail.
The measurement devices utilized by the SVD2 are two types of double-sided silicon strip detectors (DSSD), differing in their spatial dimensions. The DSSD are arranged in four layers of, with a range of radii from $r = 20.0$mm for the innermost layer to $r = 88.0$mm for the outermost layer [37, 38, 39]. The four SVD2 layers cover a polar angle of $17° < \theta < 150°$. The SVD2 delivers an intrinsic spatial resolution of $12.0 \pm 0.4\mu$m in the ϕ plane and $22.3 \pm 0.8\mu$m in the z plane.

4.3.3. Central Drift Chamber

The majority of all measurements that can be conducted at a modern multi-purpose particle detector as the Belle detector depend greatly on good quality tracking of the trajectories of charged particles. Within a magnetic field, charged particles are deflected and the curvature of their trajectories is inversely proportional to their momenta. The trajectories of charged particles are determined by measuring the ionisation trail they leave when transversing gases. These measurements are conducted using the central drift chamber (CDC). Additionally, important information for the particle identification can be extracted from the energy loss which can be measured by the weakening of the ionisation signal in the CDC.
The CDC comprises 8400 drift cells, distributed amongst 50 cylindrical layers, organized into 11 super-layers. The CDC stretches from an inner radius of $r = 103.5$mm to an outer radius of $r = 874$mm. A schematic depiction of the CDC can be found in figure 4.5. In order to achieve a good compromise between as little multiple scattering as possible to achieve a good momentum resolution and good energy loss measurements, the CDC is filled with a mixture of Helium and Ethane to equal parts. The CDC de-

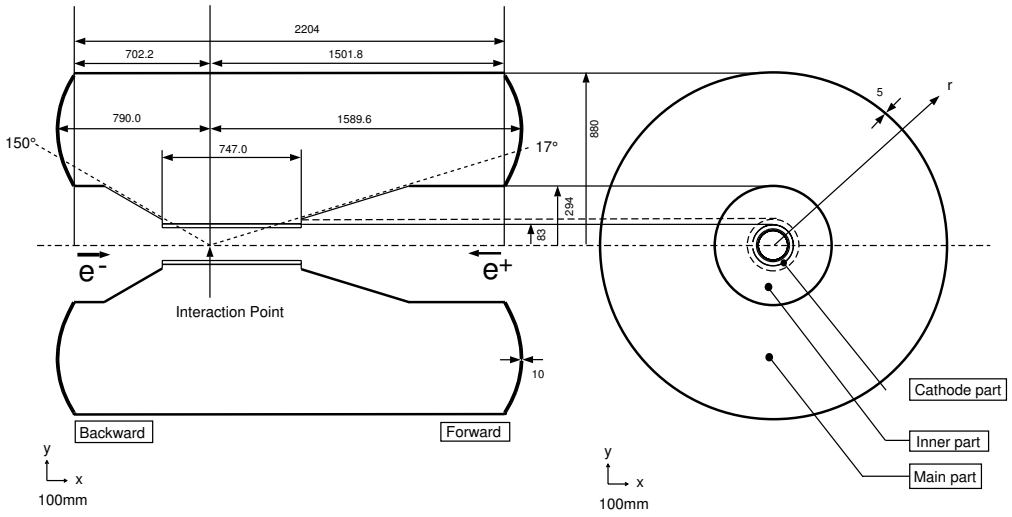


Figure 4.5.: Schematic view of the central drift chamber. [35]

livers an overall spatial resolution of 130μm. The resolution for transverse momentum p_t, measured in muon pair production processes is 1.64 ± 0.04 % in the p_t range from $4 - 5.2$ GeV/c. The energy loss dE/dx resolution was measured in K_S decays and was found to be 7.8% in the momentum range $0.4 - 0.6$ GeV/c.

4.3.4. Aerogel Cherenkov Counter System

Situated just outside of the CDC, the aerogel Cherenkov counter system (ACC) plays a key role for particle identification at the Belle experiment together with energy loss information of the CDC (see section 4.3.3) and time-of-flight measurements (see section 4.3.5). The ACC is composed of 960 silica aerogel counter modules with refractive indices between 1.01 and 1.03. The ACC structure has an angular coverage of $\theta = 17° - 34°$ for the forward endcap and $\theta = 34° - 127°$ for the barrel ACC. There is no backward endcap module.
The Cherenkov radiation is amplified by photon multipliers. The descriminating variable that is used for particle identifiaction is the pulse height caused by different types of particles passing through the aerogel. The individual cells were calibrated using $\mu^+\mu^-$ events. The opening angle of the Cherenkov cone is not measured. An exemplary pulse height spectrum distinguishing kaons and electrons is shown in figure 4.6.

4.3.5. Time Of Flight Counters

At a distance of $r = 120$ cm from the interaction point, the 192 time of flight counters (TOF) are located. They are composed of scintillation counters with directly attached photomultipliers. They cover angles of $\theta = 34° - 120°$ and have an intristic time resolution of 80 ps. Time of flight measurements can be performed for transverse particle momenta from $p_t \approx 0.28$ GeV/c, which is the minimum momentum required to reach

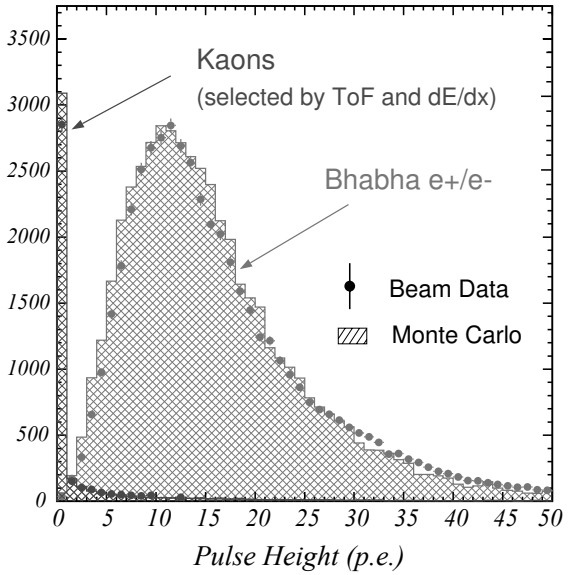


Figure 4.6.: Exemplary pulse height spectrum of the ACC separating kaons from electrons. [35]

the TOF at all to $p_t \approx 1.2$ GeV/c, when the resolution of the TOF does not suffice anymore. Several additional contributions (e.g. beam bunch length) are added in quadrature and result in a resolution of ~ 100 ps. The time of flight measurements are another important variable for particle identification at the Belle detector. Figure 4.7 shows an exemplary time of flight spectrum separating pions from protons.

4.3.6. Electromagnetic Calorimeter

The electromagnetic caloriemter (ECL) is designed to perform precise measurements of electromagnetic clusters of both relatively low and relatively high energies (< 0.5 GeV – 4 GeV) [40]. The separation of individual clusters is another important aspect for certain physics analyses, thus a good spatial resolution, provided by fine granularity of the caloriemter cells, is desirable. The ECL of the Belle detector is composed of CsI(Tl) crystals that are arranged in a finely segemented array.
The ECL is arranged asymmetrically around the interaction point. The barrel region of the ECL streches from $z = -1.0$m to $z = 2.0$m relative to the interaction point and covers an angle of $\theta = 32.2° - 128.7°$. It contains the bulk of the CsI(Tl) crystals that are tilted by $\sim 1.3°$ in order to reduce photon loss through the gaps between the crystals. The forward and backward endcap calorimeters are attached to the barrel region and cover angles of $\theta = 12.4° - 31.4°$ and $\theta = 130.7° - 155.1°$ respectively. An overview of the three parts of the ECL is given in figure 4.8. The choice of the crystal size is governed by two considerations: The smaller the crystals, the better the resolution. But with smaller crystals the energy resolution deteriorates due to the gaps and inactive material between them and additional channels have to be read out. For different parts of

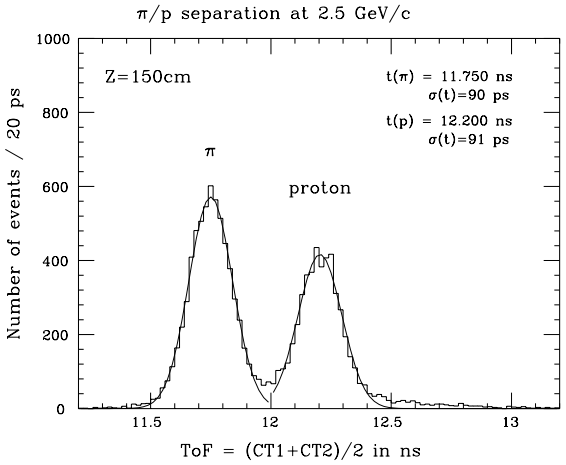


Figure 4.7.: Exemplary time of flight spectrum of the TOF separating pions from protons. [35]

the ECL, different sizes of the crystals have been chosen. The front faces of the crystals are squares with edge lenght ranging from 44.5mm to 70.8mm. The back faces of the crystals have edge lenghts ranging from 54mm to 82mm and the crystals are 300mm long.
Calibrated with Bhabha and $\gamma\gamma$ events, the ECL delivers an energy resolution of 1.7 % in the barrel region, 1.74 % in the forward endcap and 2.85 % in the backward endcap.

4.3.7. K_L and Muon Detection System

The purpose of the K_L and muon detection system (KLM) is to separate muons from other charged particles by shielding them off and to detect the relatively long-lived K_L mesons. The KLM consists of 14 iron layers, each 47mm thick, separating 15 detector layers [41]. The detector layers are composed of glass-electrode-resistive plate counters, that detect the passage of charged particles by ionisation. The separation of muons against other charged particles is based on their range and deflection through the 4.7 radiation lenghts the KLM construction provides. The KLM is divided into three parts: A barrel region and two end caps. Together, these three components cover an angle of $\theta = 20° - 155°$. The overall layout of the KLM is shown in figure 4.2. The glass-electrode-resistive plate counters are of rectangular shape and vary in size from 3.3m^2 to 5.9m^2.
The average number of K_L clusters is 0.5 per event, which is in agreement with simulation. Muons with energies larger than 0.5 GeV can reach the KLM. A comparison of the simulated to the measured range of particle tracks allows to assign a muon likelihood. The muon detection efficiency depending on the muon momentum for a likelihood cut of 0.66 is shown in figure 4.9.

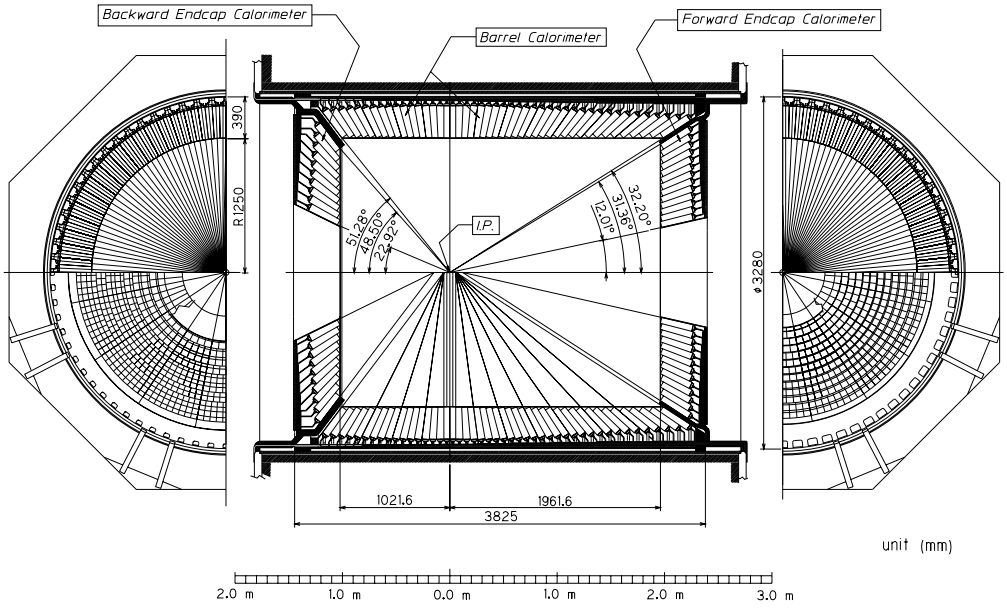


Figure 4.8.: Schematic overview of the electromagnetic calorimeter. [35]

4.3.8. Detector Solenoid and Iron Yoke

The inner sections (SVD, CDC, ACC, TOF, ECL) of the Belle detector are in a constant magnetic field of 1.5 T. The field is cylindrical, 4.4 m long and 3.4 m in diameter. As mentioned in section 4.3.3, this is essential to measure the momenta of charged particles. The magnetic field is produced by a NbTi/Cu superconductor, storing an energy of 35 MJ when fully charged. The iron support structure for all Belle detector components serves, together with the iron absorber plates of the KLM detector, as flux return paths. A contour plot of the measured field strength inside the tracking volume can be found in figure 4.10.

4.3.9. Trigger System

It would be very difficult and unpractical to record every single collision that occurs at the interaction region of the Belle detector. Many processes are of small or no interest at all. Apart from processes that are physically interesting, Bhabha and $\gamma\gamma$ events are useful to calibrate the detector and measure luminosity. Given that estimates of the beam-related backgrounds are difficult, a flexible trigger system is required that can keep the background level low, while still giving good efficiency for physically interesting events. The Belle trigger system [42] consists of a Level-1 hardware trigger and a Level-3 software trigger. The Level-1 trigger combines up to 48 signals from track triggers and energy triggers and forms the trigger signal in fixed processing time of 350 ns. An overview of possible inputs to the Level-1 trigger system is given in figure 4.11.

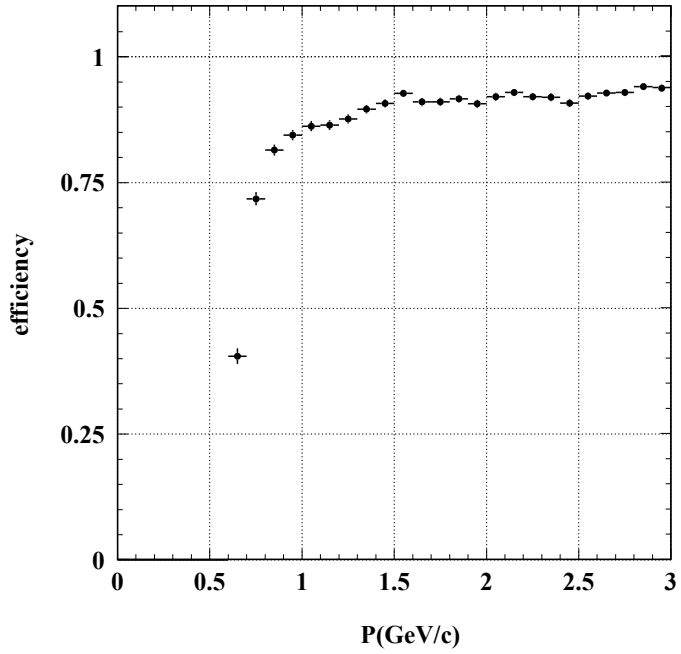


Figure 4.9.: Muon detection efficiency vs. muon momentum in the KLM. [35]

4.3.10. Overview and Summary

To summarize this chapter, let us see how different types of particles interact with different subdetectors and how they can be distinguished. Figure 4.12 gives a graphical representation of the interactions.

An important group are the charged particles: e^-, μ^-, π^-, K^- and p^- are relatively long-lived particles, that travel macroscopic distances through the detector. Due to their electrical charge, they leave ionisation trails when passing through the silicon strips of the SVD and the gas in the CDC. When they reach the ECL, they leave electromagnetic showers in the crystals. Out of this group, only muons travel through the iron absorper plates of the KLM and leave a signal there. The other charged particles are usually absorbed there and can be distinguished by energy loss measurements in the CDC and inforamtion from the ACC and TOF subdetectors.

There is only one neutral particle, that can be detected directly, the photon (γ). γ particles are directly detected by the electromagnetic showers they leave in the ECL. They can be distinguished from charged particles that also leave showers there by the fact that there is no matching track in the CDC to their showers. There is a number of neutral particles that can be detected indirectly. The π^0 decays into two γ particles, that can be detected and matched to reconstruct the original π^0. K_S mesons frequently decay into two charged pions within the CDC and leave a v-shaped structure of two charged tracks that do not originate from near the interaction point. Photons that con-

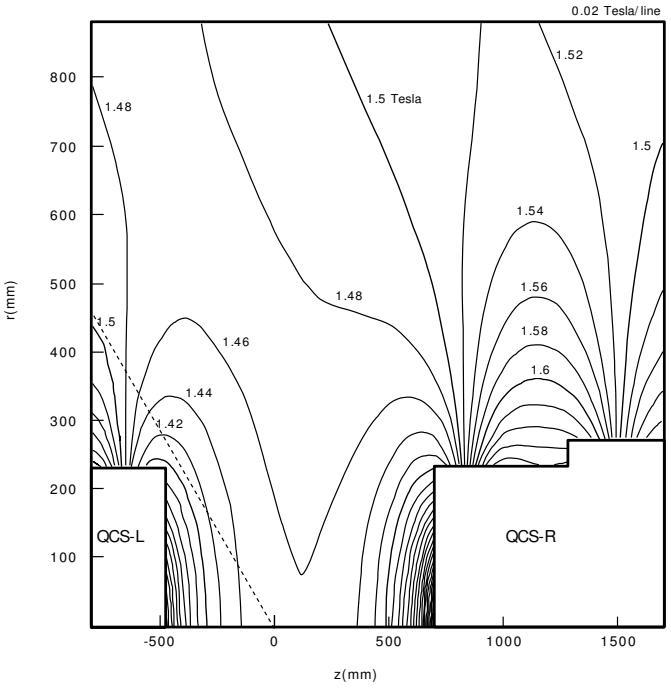


Figure 4.10.: Contour plot of the mesured field strength in the Belle detector. [35]

vert into an e^+e^- pair can be detected in the same way. K_L mesons are long-lived and are usually the only particles that reach the KLM, where they decay into two or more charged particles.

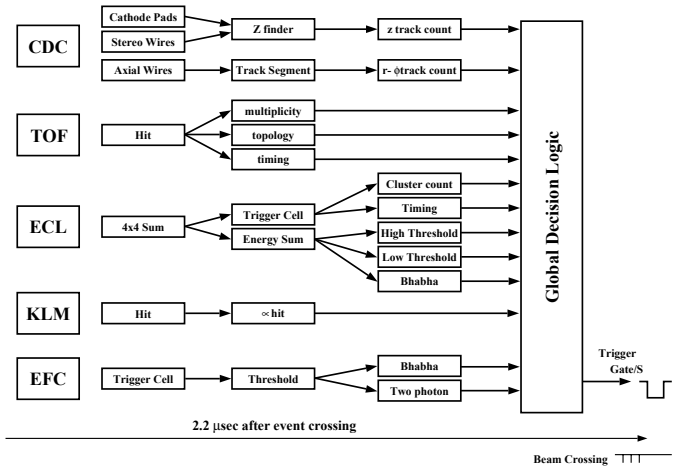


Figure 4.11.: Overview of the Level-1 trigger system of the Belle detector. [35]

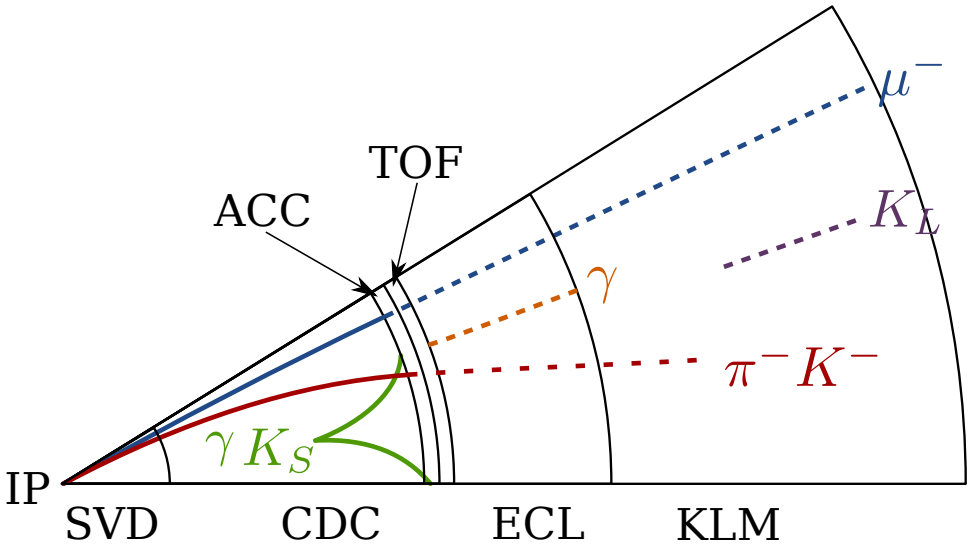


Figure 4.12.: Detector response to different kinds of stable particles. Detector components are not drawn to scale.

5. The Full Reconstruction

Many of the B meson decays studied at B-factories are semi-leptonic or even purely leptonic decays, i.e. the B meson decays into hadrons and leptons or into leptons only. Typical examples are found in equations 5.1 and 5.2.

$$B \to D^{(*)} \ell \nu_\ell \qquad \text{semi-leptonic} \tag{5.1}$$

$$B \to \tau(\to \ell \bar{\nu}_\ell \nu_\tau)\nu_\tau \qquad \text{purely leptonic} \tag{5.2}$$

Another example would be the decay $B \to D^{(*)}\tau\nu_\tau$, which is the main focus of this thesis. One of the common features of these decay modes is the fact that they contain neutrinos. There is no known detector that would be able to identify the only weakly interacting neutrinos with the rate and the accuracy that would be necessary for useful event reconstruction. Furthermore, shielding the $7 \times 10^9 \text{cm}^{-2}\text{s}^{-1}$ neutrinos from the sun reaching earth, would pose another huge problem. That means that it is currently impossible to measure the momentum of neutrinos at collider experiments directly in a satisfactory manner.

5.1. The Principle Of The Full Reconstruction

There is, however, a way to measure the momentum of neutrinos indirectly. This indirect measurement basically relies on the momentum conservation in two-body decays. To understand this technique, we must remember the fact that KEKB operates at the $\Upsilon(4S)$ resonance, as described in equation 4.7. In the cross-section spectrum of e^+e^- collisions, shown in figure 5.1, the $\Upsilon(4S)$ resonance is the first resonance, where the creation of a B meson pair is energetically possible. The $\Upsilon(4S)$ resonance produced at B-factories decays almost always into a pair of B mesons and *no additional particles*:

$$\Upsilon(4S) \to B\bar{B} \quad > 96\% \tag{5.3}$$

$$\Upsilon(4S) \to B^+B^- = (51.6 \pm 0.6)\% \tag{5.4}$$

$$\Upsilon(4S) \to B^0\bar{B}^0 = (48.4 \pm 0.6)\% \tag{5.5}$$

As the decay $\Upsilon(4S) \to B\bar{B}$ is a two-body decay, the principle of conservation of momentum yields a very simple relation:

$$p_{\text{CMS}}(B) + p_{\text{CMS}}(\bar{B}) + p_{\text{CMS}}(\text{Beam}) = 0 \tag{5.6}$$

Let us imagine that one of these B mesons, which we shall call the *signal side* B meson B_{sig}, decays semi-leptonically and the other B meson, which we shall call the *tag side*

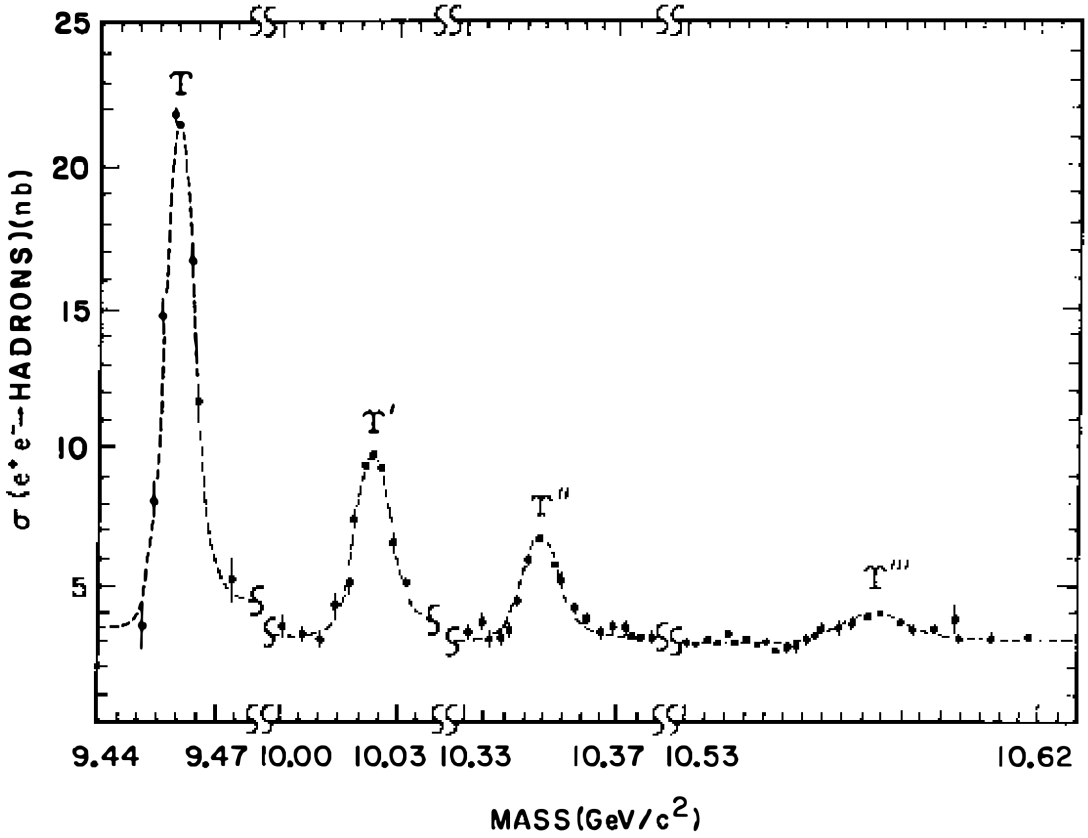


Figure 5.1.: The Υ resonances in the cross-section spectrum of e^+e^- collisions [43]. The $\Upsilon(4S)$ resonance is written as Υ''' in this plot.

B meson B_{tag}, decays purely hadronically. If this is the case, we are, in principle, able to reconstruct the purely hadronically decaying tag side B meson B_{tag} completely, or "fully". This is not possible for the signal side, we will always miss the momentum carried away by the neutrinos. But using equation 5.6, with the precise knowledge of the 4-momentum of the beam, we can derive the true 4-momentum of B_{sig}, compare it with the reconstructed 4-momentum of B_{sig} and thereby measure the missing momentum:

$$p_{\text{CMS}}(B_{\text{tag}}) + p_{\text{CMS}}(B_{\text{sig, rec}}) + p_{\text{CMS}}(\text{missing}) + p_{\text{CMS}}(\text{Beam}) = 0 \tag{5.7}$$

It can be easily seen from equation 5.7, that the precise knowledge of the beam momentum is another crucial element for the Full Reconstruction algorithm.

5.1.1. Missing Momentum And Additional Particles

The ability to make this missing momentum measurable is one of the main features of the Full Reconstruction. Another very important piece of information can be deduced

from the fact that in $\Upsilon(4S)$ decays, there are no particles in addition to the two B mesons (see equations 5.4 and 5.5). It is therefore possible to measure the number and 4-momenta of additional tracks and electromagnetic clusters, that were included neither in the tag-side nor in the signal-side reconstruction. This knowledge can greatly help to understand signal-side decays or yield much cleaner signal samples by requiring that there shouldn't be any tracks or electromagnetic clusters left.
An exemplary schematic for a fully reconstructed event with a signal side can be found in figure 5.2. It should be noted, that the Full Reconstruction is not an analysis in itself,

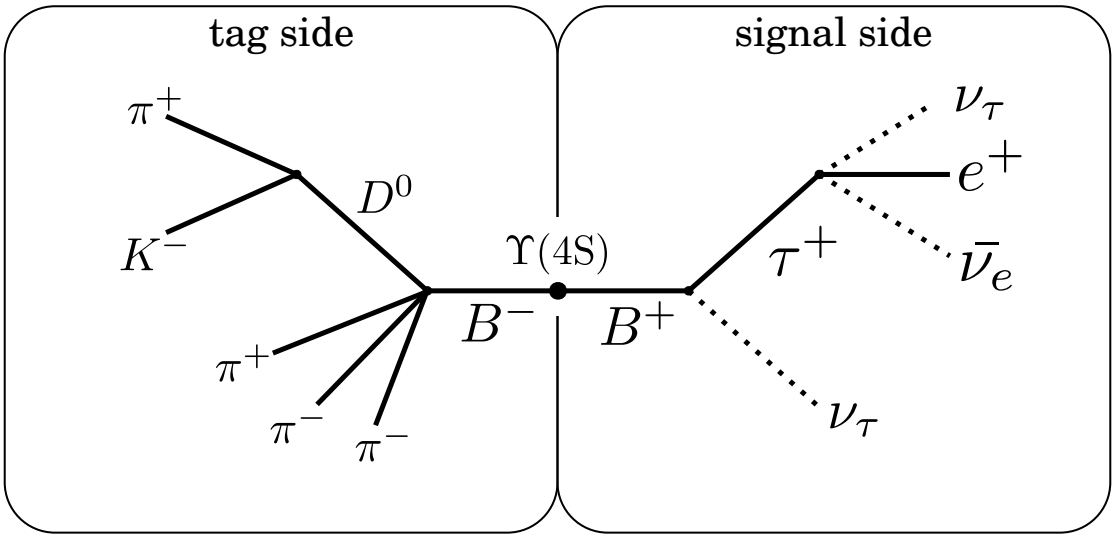


Figure 5.2.: A fully reconstructed event with $B^- \to D^0(\to K^-\pi^+)\pi^+\pi^-\pi^-$ on the tag side and the purely leptonic decay $B^+ \to \tau+(\to e^+\nu_\tau\bar{\nu}_e)\nu_\tau$ on the signal side

but rather a tool that makes many other analyses at all possible. It takes advantage of the properties of the $\Upsilon(4S)$ resonance and the very precisely known beam energy and is therefore only applicable at lepton colliders.

5.1.2. Inclusive Searches

The missing momentum on the signal side can not only originate from missing neutrinos. The Full Reconstruction allows to measure the missing momentum of every particle combination that can be missed during reconstruction. This is a very useful property for inclusive searches. Consider for example an inclusive search for $c\bar{c}$ resonances:

$$B^+ \to X_{c\bar{c}}K^+ \tag{5.8}$$

By reconstructing one K^+ particle and then measuring the missing momentum of the rest of the signal side, $X_{c\bar{c}}$ resonances can be found in the spectrum of missing mass. This approach makes the explicit reconstruction of the particles of interest unnecessary. It is therefore a useful approach in the search of yet unknown states.

5.1.3. Broken China Plates

A simple analogy can be made to illustrate the principle of the Full Reconstruction: Let us imagine two identical china plates that were dropped to the floor. If one is successful to piece one of the two plates together completely, one can be sure that the entire pile of shards that remains must come from the other plate. If we imagine further, that one or more shards were taken away from the pile, we can exactly determine the mass of these shards by measuring how much lighter the remaining pile of shards is than the pieced together china plate.

5.2. Efficiency and Purity

Many analyses at B factories, predominantly those concerned with (semi-)leptonic B meson decays, depend on a tag side. If we recall figure 5.2, we can directly see, that the more often we are successful in fully reconstructing a tag side, the larger the available data sample for the signal side analysis is. The number of fully reconstructed tag side events depends basically on two factors:

1. The covered branching ratio of the B meson and
2. the reconstruction and selection efficiency.

Both of these factors have direct influence on the number of tag side B mesons, or the *efficiency*. Keeping in mind the need for as many tag-side candidates as possible, or the highest possible efficiency, it seems, at first glance, a reasonably good idea to implement as many purely hadronic B and D meson decays as possible. As the D mesons and especially the B mesons decay into a rich spectrum of final states, it seems to be beneficial to include as many of these decay modes as possible into the reconstruction algorithm. At the same time, one might think about not introducing any pre-cuts at all, since it is only possible to loose efficiency by applying cuts. This approach was indeed followed at first and very soon serious problems arose: Using every possible combination of tracks and electromagnetic clusters present in an event for the combination of a B meson candidate corresponds to picking k items out of n in a specific order leading to $n!/(n-k)!$ possible candidates. This means for the exemplary typical tag-side decay

$$B^+ \to \bar{D}^0 \pi^+ \pi^+ \pi^- \tag{5.9}$$

$$D^0 \to K^- \pi^+ \tag{5.10}$$

five tracks have to be picked from a typical event track multiplicity of ten. This leads to

$$\frac{10!}{5!} = 30240 \tag{5.11}$$

possible combinations in this channel alone, where at most one can be true. The above argument runs a bit short, as the tracks are already divided up by their charge and even

very wide mass windows for D and B mesons don't allow for completely non-sensible combinations etc.... It becomes nonetheless clear, that combinatorics pose a very serious problem. Apart from all technical problems, this is important above all for the signal to background ratio of the tag side sample, as it would introduce an enormous amount of background. Furthermore, we cannot neglect the technical difficulties this approach would bring with it. The necessary computing time to perform the reconstruction of the decay chains and the necessary storage capacity to save the results are not available. Even higher requirements would be expected for the classification of this enormous tag side sample. It is therefore crucial for the success of the Full Reconstruction tool to find tag side channels, that cover a fair amount of the B and D meson branching ratio and pose still manageable combinatorial problems. At the same time, efficient pre-cuts are absolutely necessary to simultaneously avoid giving away too much of the tag-side signal on which many analyses directly depend, while reducing the background to manageable levels.

5.2.1. Channels

With inspiration taken from the predecessor of this Full Reconstruction [44, 45] and the Particle Data Group [2], B and D meson channels were chosen for the Full Reconstruction that cover a fair portion of the respective total branching ratios. Despite their relatively low branching ratio, J/Ψ mesons were included, as they provide a very clean signal. Additionally, experience with runtime and trial and error played a significant role in the selection of decay channels. The B meson decay channels are listed in table 5.1, the D meson decay channels are listed in table 5.2, the excited D^* meson decay channels are listed in table 5.3 and the $D_S^{(*)}$ and J/Ψ meson decay channels are listed in table 5.4, all of them with their respective branching ratios.

5.2.2. Hierarchical System

Tables 5.1 through 5.4 give a clean overview of the decay channels used for the Full Reconstruction. One has, however, to keep in mind that one single line in table 5.1 does not only correspond to one exclusive decay channel, but rather to a large number. This comes from the fact that D mesons also decay in numerous different channels. For example the line

$$B^+ \to \bar{D}^0 \pi^+$$

Channel			$\mathcal{BR}$	Channel			$\mathcal{BR}$
B^+	$\to$	$\bar{D}^0\pi^+$	0.484%	B^0	$\to$	$D^-\pi^+$	0.268%
B^+	$\to$	$\bar{D}^0\pi^+\pi^0$	1.340%	B^0	$\to$	$D^-\pi^+\pi^0$	0.760%
B^+	$\to$	$\bar{D}^0\pi^+\pi^+\pi^-$	1.100%	B^0	$\to$	$D^-\pi^+\pi^+\pi^-$	0.800%
B^+	$\to$	$D_S^+\bar{D}^0$	1.000%	B^0	$\to$	$\bar{D}^0\pi^0$	0.026%
B^+	$\to$	$\bar{D}^{0*}\pi^+$	0.519%	B^0	$\to$	$D_S^+D^-$	0.720%
B^+	$\to$	$\bar{D}^{0*}\pi^+\pi^0$	0.980%	B^0	$\to$	$D^{*-}\pi^+$	0.276%
B^+	$\to$	$\bar{D}^{0*}\pi^+\pi^+\pi^-$	1.030%	B^0	$\to$	$D^{*-}\pi^+\pi^0$	1.500%
B^+	$\to$	$\bar{D}^{0*}\pi^+\pi^+\pi^-\pi^0$	1.800%	B^0	$\to$	$D^{*-}\pi^+\pi^+\pi^-$	0.700%
B^+	$\to$	$D_S^{+*}\bar{D}^0$	0.760%	B^0	$\to$	$D^{*-}\pi^+\pi^+\pi^-\pi^0$	1.760%
B^+	$\to$	$D_S^+\bar{D}^{0*}$	0.820%	B^0	$\to$	$D_S^{+*}D^-$	0.740%
B^+	$\to$	$D_S^{+*}\bar{D}^{0*}$	1.710%	B^0	$\to$	$D_S^+D^{*-}$	0.800%
B^+	$\to$	$\bar{D}^0K^+$	0.037%	B^0	$\to$	$D_S^{+*}D^{*-}$	1.770%
B^+	$\to$	$D^-\pi^+\pi^+$	0.107%	B^0	$\to$	$J/\psi K_S^0$	0.087%
B^+	$\to$	$J/\psi K^+$	0.101%	B^0	$\to$	$J/\psi K^+\pi^-$	0.120%
B^+	$\to$	$J/\psi K^+\pi^+\pi^-$	0.107%	B^0	$\to$	$J/\psi K_S^0\pi^+\pi^-$	0.100%
B^+	$\to$	$J/\psi K^+\pi^0$	0.047%				
B^+	$\to$	$J/\psi K_S^0\pi^+$	0.094%				
Sum			12.0%	Sum			10.4%

Table 5.1.: B meson decay modes used in the Full Reconstruction

Channel			$\mathcal{BR}$	Channel			$\mathcal{BR}$
D^0	$\rightarrow$	$K^-\pi^+$	3.89%	D^+	$\rightarrow$	$K^-\pi^+\pi^+$	9.40%
D^0	$\rightarrow$	$K^-\pi^+\pi^+\pi^-$	8.09%	D^+	$\rightarrow$	$K^0_S\pi^+$	1.49%
D^0	$\rightarrow$	$K^-\pi^+\pi^0$	6.90%	D^+	$\rightarrow$	$K^0_S\pi^+\pi^0$	6.90%
D^0	$\rightarrow$	$\pi^+\pi^-$	0.14%	D^+	$\rightarrow$	$K^-\pi^+\pi^+\pi^0$	6.08%
D^0	$\rightarrow$	$\pi^+\pi^-\pi^0$	1.44%	D^+	$\rightarrow$	$K^0_S\pi^+\pi^+\pi^-$	3.10%
D^0	$\rightarrow$	$K^0_S\pi^0$	1.22%	D^+	$\rightarrow$	$K^+K^-\pi^+$	0.98%
D^0	$\rightarrow$	$K^0_S\pi^+\pi^-$	2.94%	D^+	$\rightarrow$	$K^+K^-\pi^+\pi^0$	1.50%
D^0	$\rightarrow$	$K^0_S\pi^+\pi^-\pi^0$	5.40%				
D^0	$\rightarrow$	K^+K^-	0.39%				
D^0	$\rightarrow$	$K^+K^-K^0_S$	0.47%				
Sum			37.9%	Sum			29.4%

Table 5.2.: D meson decay modes used in the Full Reconstruction

Channel			$\mathcal{BR}$	Channel			$\mathcal{BR}$
D^{0*}	$\rightarrow$	$D^0\pi^0$	61.9%	D^{+*}	$\rightarrow$	$D^0\pi^+$	67.70%
D^{0*}	$\rightarrow$	$D^0\gamma$	38.10%	D^{+*}	$\rightarrow$	$D^+\pi^0$	30.70%
Sum			100.0%	Sum			98.4%

Table 5.3.: D^* meson decay modes used in the Full Reconstruction

Channel			$\mathcal{BR}$	Channel			$\mathcal{BR}$
D^+_S	$\rightarrow$	$K^+K^0_S$	1.49%	D^{+*}_S	$\rightarrow$	$D^+_S\gamma$	94.20%
D^+_S	$\rightarrow$	$K^+\pi^+\pi^-$	0.69%				
D^+_S	$\rightarrow$	$K^+K^-\pi^+$	5.50%				
D^+_S	$\rightarrow$	$K^+K^-\pi^+\pi^0$	5.60%				
D^+_S	$\rightarrow$	$K^+K^0_S\pi^+\pi^-$	0.96%				
D^+_S	$\rightarrow$	$K^-K^0_S\pi^+\pi^+$	1.64%				
D^+_S	$\rightarrow$	$K^+K^-\pi^+\pi^+\pi^-$	0.88%	J/ψ	$\rightarrow$	e^-e^+	5.94%
D^+_S	$\rightarrow$	$\pi^+\pi^+\pi^-$	1.10%	J/ψ	$\rightarrow$	$\mu^-\mu^+$	5.93%
Sum			17.9%	Sum			11.9%

Table 5.4.: $D^{(*)}_S$ and J/Ψ meson decay modes used in the Full Reconstruction

corresponds to 10 exclusive decay modes:

$$
\begin{aligned}
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^-\pi^+ \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^-\pi^+\pi^+\pi^- \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^-\pi^+\pi^0 \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to \pi^+\pi^- \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to \pi^+\pi^-\pi^0 \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^0_S\pi^0 \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^0_S\pi^+\pi^- \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^0_S\pi^+\pi^-\pi^0 \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^+K^- \\
B^+ &\to \bar{D}^0\pi^+ \quad \text{with} \quad D^0 \to K^+K^-K^0_S
\end{aligned}
$$

When we look at table 5.1 again with this in mind, it can be calculated that the Full Reconstruction comprises the reconstruction of 1104 different exclusive decay channels. In terms of coding and code maintenance, it would be a futile task to implement all those decay channels one by one. This is why a *hierarchical* reconstruction framework was developed. It enables us to follow the naturally given approach to reconstruct a collection of unstable particles, say D^0 mesons, and then use the entire collection for further reconstruction. A visualisation of this framework can be found in figure 5.3.

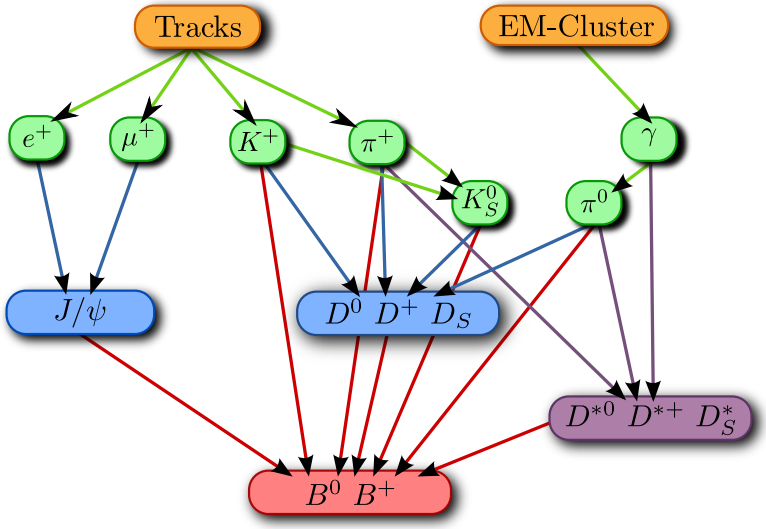


Figure 5.3.: The hierarchical reconstruction framework developed for the Full Reconstruction

Four Stages

The hierarchical system, pictured in figure 5.3 comprises four stages: In the first stage, tracks measured by the Belle-Detector are divided into e^+, μ^+, K^+ and π^+ and electromagnetic clusters are skimmed for good photons γ. Then, K_S^0 and π^0 mesons are reconstructed using K^+ and π^+ and γ respectively. In the next stage, D and J/Ψ mesons are reconstructed using the particles from the stage before according to the decay channels listed in tables 5.2 and 5.4. After that, excited D mesons are reconstructed using all available D mesons and all of the remaining π^+, π^0 and γ particles, according to tables 5.3 and 5.4. In the last stage, all available excited D mesons, all D and J/Ψ mesons and all remaining e^+, μ^+, K^+, π^+, K_S^0 and π^0 are combined into the final B mesons, of course according to table 5.1.

The hierarchical approach has us thus allowed to implement the reconstruction of B mesons in a very convenient and natural manner, minimizing the potential for human error. But this was not only advantageous, as described in section 5.2.3.

5.2.3. Intermediate Cuts

When we try to maximise the efficiency of the Full Reconstruction as much as possible, we have to question the need for cuts during the reconstruction process. It is a trivial fact, that one can only loose signal by applying even the softest cut. It became, however, evident very quickly that it is unavoidable to include several cuts in the reconstruction framework. Otherwise, there would be no chance to obtain a tag-side sample with a reasonable fraction of correctly reconstructed tags and it would take minutes to process one single event, instead of a feasible time in the order of $10^{-2}s$.

As mentioned above, it is of paramount importance to have the most efficient reconstruction possible, while not loosing the signal fading in enormous amounts of background. At the same time we must take care not to be overwhelmed by excessive need for computing time and storage capacity. The relatively large number of decay channels for B and D mesons listed in tables 5.1 to 5.4 made the need for one single solution for this problem that worked for all decay channels equally well clear. Tweaking different selection criteria for all of the decay modes one-by-one seemed not only a lengthy task, prone to human error. It would also be a procedure that would lack well-defined criteria for the selection purity and efficiency. Let us clarify this with an example: Our goal is to achieve the maximum amount of signal, while keeping background levels and computing requirements manageable. Considering that, how can one compare the quality of a candidate in the two decay channels

$$B^+ \to \bar{D}^0\pi^+ \quad \text{and}$$
$$B^+ \to \bar{D}^0\pi^+\pi^+\pi^- \ ?$$

Let us furthermore consider only two decay modes of the D meson:

$$D^0 \to K^-\pi^+ \quad \text{and}$$
$$D^0 \to K^0_S\pi^+\pi^-\pi^0 \,.$$

Considering the relatively *clean*[1] decay channel $D^0 \to K^-\pi^+$, we might consider quite loose cuts. This approach appears especially compelling when combining the resulting D meson with the clean B decay mode $B^+ \to \bar{D}^0\pi^+$. If we, on the other hand, combine the D meson with the much less clean B decay channel $B^+ \to \bar{D}^0\pi^+\pi^+\pi^-$ stronger cuts are much more appropriate. So for clean B decay modes, we would select loose D meson cuts; for less clean B decay modes, we would select stronger D meson cuts. A large list of cuts for the $D^0 \to K^-\pi^+$ decay channel would be necessary. And a very different, but equally large list for the $D^0 \to K^0_S\pi^+\pi^-\pi^0$ would have to be determined.

Implementing different cuts for a D meson, depending on the B decay mode that will eventually be used, seems to be the solution for this issue. This would be a truly sisyphean undertaking. The 1104 exclusive decay channels mentioned earlier to illustrate the benefit of the hierarchical system would in this case have to be treated really individually, thus making the hierarchical system superfluous. At first glance, it seems impossible to combine the need for individually adjusted selection criteria with the simple and clear, hierarchical system, where all particles of the same type are to be treated equally. Abandoning the hierarchical system was not an option, considering the huge amount of cut determination and coding, that would be required for the exclusive reconstruction, not to mention the still missing well-defined criteria for efficiency and purity and the large amount of human errors, that would come with this approach.

5.2.4. Probabilities

The above described predicament was solved by calculating the signal *probability* for each decay channel. This probability has the benefit to be comparable between different decay channels of the same particle, say

$$D^0 \to K^-\pi^+ \quad \text{and}$$
$$D^0 \to K^0_S\pi^+\pi^-\pi^0 \,.$$

This comparability is based on the signal probability of the candidates and is therefore independent of the cleanliness of the channel. Thus, we can now use D mesons of different decay channels for the reconstruction of B mesons and do not need to explicitly distinguish the decay modes of the D mesons. The task of calculating the probability was performed by the multivariate analysis framework NeuroBayes.

While we limited ourselves in the given example to the reconstruction of B and D decay modes, the very same procedure was applied for all stages. The final stage particles in

[1] A decay channel is typically considered "clean" when it has a relatively large signal component and low background levels.

the first stage are assigned a signal probability, as well as all unstable particles in the higher stages.

NeuroBayes trainings were performed for all particles of stage 1. Typical variables used in these trainings include information coming from the various sub-detectors and also their combination into the standard Belle Particle ID [46, 47, 48], energy, total momentum and shower shape parameters, where applicable. Complete lists and brief descriptions of the variables can be found in tables A.1 to A.7 in appendix A.5. The output of these trainings was then recalculated to represent the actual signal probability as described in section 3.4. This recalculation was performed to enable a fair comparison between the decay channels on the same stage.

The variables used in the trainings of the higher stages ($D^{(*)}_{(S)}$, J/Ψ and B mesons) were quite similar for all channels. This is due to the fact that they all are unstable particles and the quality of the decay products and the kinametics of the decay are the two main pieces of information in such cases. The variable with the largest discrimination power was almost always o_{prod}, described in more detail in section 5.2.5. Typically, vertex fit information provided useful information as well. A complete list of all variables can be found in table A.8 in appendix A.5.

5.2.5. The Product Of Children Probabilities

Let us stay with the exemplary B meson decays a bit further. While we have now the signal probability of each of the children at our disposal, the question remains how to determine ideal cuts on these different probabilities. A first step is the combination of all the signal probabilities of all children in a product:

$$o_{\text{prod}} = \prod_{i=0}^{N} (o_i) \,, \tag{5.12}$$

where N is the number of children and o_i is the NeuroBayes output of child i. For each B decay mode, o_{prod} is a very powerful variable on which an individual cut, depending on the B decay mode can be performed. This means that we perform a cut on all D meson decays simultaneously and at a later time when further information (vertex fits, decay mode of the B meson, additional final state particles) is available.
The variable o_{prod} is one of the central elements of this Full Reconstruction algorithm. It allows for very efficient intermediate cuts at all stages of the hierarchical system. This was an essential requirement for an efficient Full Reconstruction.

5.2.6. Cut Determination

Still unanswered is the question, how the exact values of the cuts on o_{prod} were determined. For simplicity, let us consider the D^0 mesons alone for the moment. We only

considered D^0 mesons that decay into final state particles in our choices of decay channels. For each of these final state particles, we know the signal probability from the corresponding NeuroBayes training. We are therefore able to construct o_{prod}. The goal of our selection process is to select as much signal N_{sig} as possible *summed over all* D^0 *decay modes* for a given total number of accepted events $N_{\text{all}} = N_{\text{sig}} + N_{\text{bg}}$. In mathematical terms, we wish to maximise the following expression for a given total number of accepted events N_{all}:

$$\frac{\sum_f (N_{\text{sig},f})}{\sum_f (N_{\text{sig},f} + N_{\text{bg},f})}\,, \tag{5.13}$$

where f denotes the final state in which the D^0 meson decays. There is no obvious criterion for the number of accepted events N_{all}, apart from the needed processing time and storage space. But let us first discuss the optimisation strategy.

The following strategy for the above mentioned optimisation was chosen: On generic Monte Carlo simulation, several possible cut values over the whole range of o_{prod} were tested for all D^0 meson channels individually. Each cut yielded a certain number of signal and background events. All of those pairs of numbers were plotted as points on a plane. This resulted in the curves shown in figure 5.4. The slope of the tan-

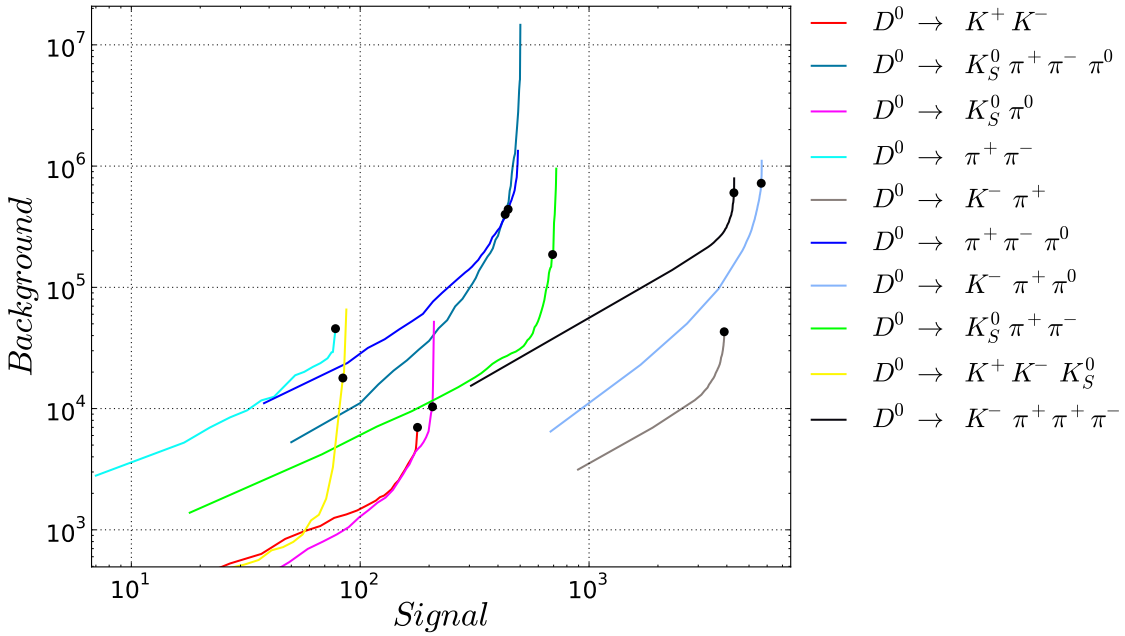


Figure 5.4.: Signal and background for different cuts on o_{prod} for all D^0 decay channels in the Full Reconstruction

gent at a fixed point of one of the curves gives the number of additional background

candidates that will be accepted for one additional signal candidate. Very clean channels (e.g. $D^0 \to K^-\pi^+$) have a relatively flat slope over their whole range. Other less clean channels (e.g. $D^0 \to K^0_S K^+ K^-$) have much steeper slopes for low cut values of o_{prod}. Choosing the same slope for all channels gives us cut values that accept the same amount of background events for one additional signal event. This statement is true for all D^0 modes that are considered in this Full Reconstruction. The cuts are therefore optimised with respect to the efficiency and purity integrated over all D^0 decay channels. Increasing or decreasing one single cut value would lead to a non-optimal signal to background ratio with respect to the overall number of accepted events N_{all}.

Let us now return to the still missing criterion for the number of accepted events N_{all}. This factor is proportionally linked to the numerical value of the chosen slope in the above described optimisation.The steeper the slope is, the more candidates per event are accepted and the more computing time is needed to process the Belle data set and several Monte Carlo sets. The flatter the slope is, the less computing time is needed. Ultimately, experience with the CPU requirements of previous full reconstruction attempts guided the decision. An average processing time of less than 0.1s per event was chosen and the obtained numerical values of the slope can be found in table 5.5.
The above described procedure was not only performed for D^0 mesons, but for all $D_{(S)}$

Particle	N_{CPE}	Slope
D^+	6.88	7087
D^0	8.62	5404
D_S^+	4.17	22443
D^{*+}	1.62	8223
D^{*0}	5.24	8223
D_S^{*+}	1.88	8223
B^+	0.93	31934
B^0	0.52	31729

Table 5.5.: Slopes and numbers of candidates per event N_{CPE} that were chosen in the Signal-Background curves method for intermediate cuts

mesons in the Full Reconstruction. It was also used for the intermediate cuts in the reconstruction of $D^*_{(S)}$ and finally B mesons.
Let us remind ourselves that the cuts were performed on NeuroBayes trainings. Furthermore, the product of the NeuroBayes outputs of the children o_{prod} were used in these trainings. This way, information on the tracks and electromagnetic clusters found its way into the NeuroBayes trainings for $D^*_{(S)}$ and B mesons, along with vertex fit information for the D mesons etc. In terms of the hierarchical system, described in sec-

tion 5.2.2, both the reconstruction and the flow of information is organised in a hierarchical manner. This hierarchical flow of information is one of the very strong features of the Full Reconstruction framework.

The chosen values for all slopes can be found in table 5.5 and the signal-background curves can be found in figures A.15 to A.20 in appendix A.4.

5.3. The B_{tag} Collection

After the reconstruction of the four stages, we have obtained the desired collection of B_{tag}mesons. Although there is no following reconstruction step that would require any quality criterion, NeuroBayes trainings were performed for these B mesons as well. This has proven very beneficial, as the purity and efficiency of this final sample can now be chosen freely. This gives different types of analyses with different needs the ability to utilise the Full Reconstruction with optimal purity and efficiency. These final trainings are described in more detail in the following.

5.3.1. Continuum Suppression

A large background contribution in the B_{tag}sample provided by the Full Reconstruction stems from events, where no $\Upsilon(4S)$ resonance has been produced. This class of events is called *continuum*. For many analyses, it is beneficial to suppress this type of background. There are many variables that can help to suppress this sort of background. However, many of these variables rely on analysis of the spatial event shape in the detector, which generally tends to be more jet-like for continuum events. For certain analyses, the usage of such variables might introduce difficulties. Therefore, a dedicated B_{tag}network training was performed with all continuum suppression variables left out.

For certain analyses, the usage of continuum suppression is less problematic. Therefore, additional trainings were performed that included continuum suppression variables.

Fox-Wolfram Moments

Most variables that were used in the NeuroBayes trainings to suppress the continuum background were the Fox-Wolfram moments [49, 50] H_l defined as:

$$H_l = \sum_i \sum_j |p_i| \cdot |p_j| \cdot P_l \cos\theta_{ij} \ , \tag{5.14}$$

where i and j are final state particles in an event, p_i and p_j their momenta, θ_{ij} is the angle between their momenta and P_l is the Legendre polynomial of l^{th} order.

The First Continuum Suppression

The first continuum suppression network training contains the second reduced Fox-Wolfram moment

$$R_2 \equiv \frac{H_2}{H_0} \tag{5.15}$$

as well as the thrust angle $\cos\theta_{\text{thrust}}$. For the definition of the thrust angle, all final state particles that belong to the B_{tag}candidate in question were combined into one set. The thrust axis $\vec{n}_{\text{tag}}$ of this set is defined by maximising

$$\frac{\sum_i \vec{n}\cdot\vec{p_i}}{\sum_i |\vec{p_i}|}\,. \tag{5.16}$$

In a second step, all particles in the event that do not belong to the B_{tag}candidate are combined into a second set and a second thrust axis $\vec{n}_{\text{rest}}$ is determined. The angle between the two thrust axes is the thrust angle:

$$\cos\theta_{\text{thrust}} = \cos\sphericalangle(\vec{n}_{\text{tag}}, \vec{n}_{\text{rest}}) \tag{5.17}$$

The Second Continuum Suppression

Additionally to the first continuum suppression network, as second network was trained, which included more variables than the first network. Mainly, the Super-Fox-Wolfram moments [51] were used. They provide more information than the Fox-Wolfram moments by exploiting the distinction between tag side and signal side. The summation of all final state particles, as given in equation 5.14 is split into the particles that were used in the reconstruction of the B_{tag}candidate, those that were not used, and mixed terms. This splitting yields 18 different terms, all of which were used in the second continuum suppression network. As this approach brings information about the signal side into the continuum suppression, it can lead to difficulties for certain analyses. This was the reason to provide three methods of evaluating the B_{tag}quality: Without continuum suppression, including only event shape variables and including Super-Fox-Wolfram moments. The user is then free to choose, which of the three NeuroBayes experts to use for the selection of purity and efficiency.

5.3.2. Purity And Efficiency

As explained in chapter 3.2.1, a good and intuitive way to compare different classification methods are purity-efficiency plots. They give a complete overview of the discrimination power of a classification algorithm and also allow thorough comparisons between two different algorithms. The basis of comparison for the Full Reconstruction algorithm presented in this chapter is the Full Reconstruction algorithm that was in usage by the Belle collaboration for many years. This predecessor of the new Full

Reconstruction used a cut-based algorithm and presented a fixed point in the purity-efficiency plane as a result. It is described in more detail in references [44, 45].

For the purity-efficiency comparison, B_{tag}signal yields have to be extracted from the data. They were extracted by fits to the beam-constrained mass M_{bc}, defined as

$$M_{\text{bc}} \equiv \sqrt{E_{\text{beam}}^2 - \sum_{FS} p_{FS}^2}\,, \tag{5.18}$$

where E_{beam} is the measured beam energy and FS runs over the final state particles used in the B_{tag}reconstruction. Multiple fits to this variable were performed for different cut values of o_{NB}. It should be stressed that the following comparisons were all done on real data. Three exemplary plots that allow a quick comparison to the predecessor are shown in figures 5.5, 5.6 and 5.7. From these plots alone, it is evident that

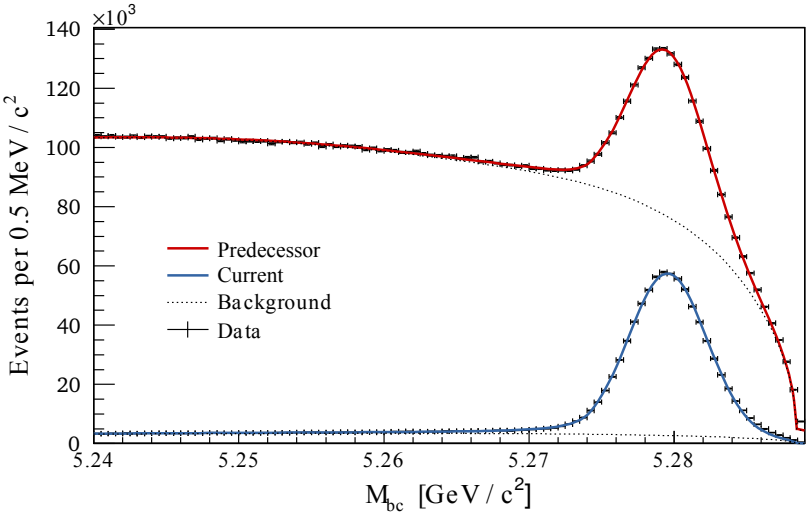


Figure 5.5.: $M_{\text{bc}}$for the B^0_{tag} sample provided by the Full Reconstruction and its predecessor. The cut on $o_{\text{NB}}$was chosen to achieve the same signal efficiency for both samples.

...

- For the same signal efficiency, the purity of the current Full Reconstruction algorithm is substantially higher than that of the predecessor (figure 5.5).
- For the same background level, the signal fraction is much higher than for the predecessor (figure 5.6).
- For the same purity, the signal yield is approximately by a factor 2 larger than for the predecessor (figure 5.7).

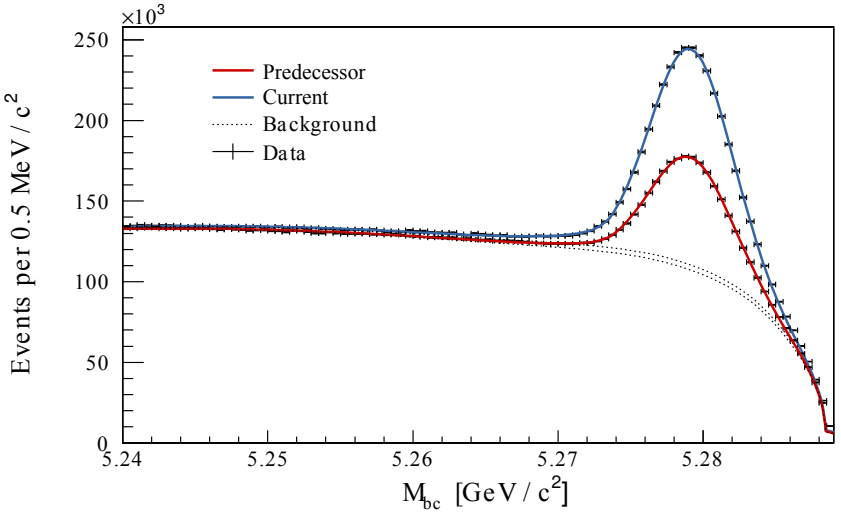


Figure 5.6.: $M_{\mathrm{bc}}$for the B^+_{tag} sample provided by the Full Reconstruction and its predecessor. The cut on $o_{\mathrm{NB}}$was chosen to achieve the same background level for both samples.

The three described plots are taken at random points on the purity efficiency curve, and were chosen only for an illustrative comparison to the predecessor.

The purities and efficiencies for many different cuts were extracted from similar fits and composed into purity-efficiency plots, which give a more complete picture. Figures 5.8 and 5.9 show the performance of all three B_{tag}classification methods of the new Full Reconstruction and the purity-efficiency working point of the predecessor. It can be easily seen from these figures, that the new algorithm supersedes its predecessor both in purity and efficiency simultaneously. If a purity similar to the predecessor is chosen, the new Full Reconstruction delivers roughly double the efficiency. If, on the other hand, a similar efficiency to the predecessor is chosen, the purity of the sample rises from below 25% to over 80%. The point for maximum reconstruction efficiency yields an overall efficiency of 0.18% for B^0 mesons and 0.28% for B^+ mesons. These efficiencies correspond to 1.4 and 2.1 million correctly reconstructed B_{tag}mesons, respectively, given the complete Belle data-set of 771×10^6 $B\bar{B}$ events.
It can also be seen, that the network outputs including continuum suppression variables deliver better performance than the one without. Our expectation that the more information is used in the classification, the better the result will be, was confirmed.

5.3.3. Reduced Channel Set

When faced with the comparison between the current Full Reconstruction and its predecessor, the question arises how much of the improvement is due to the new decay channels and how much due to the improved algorithm. This question cannot be

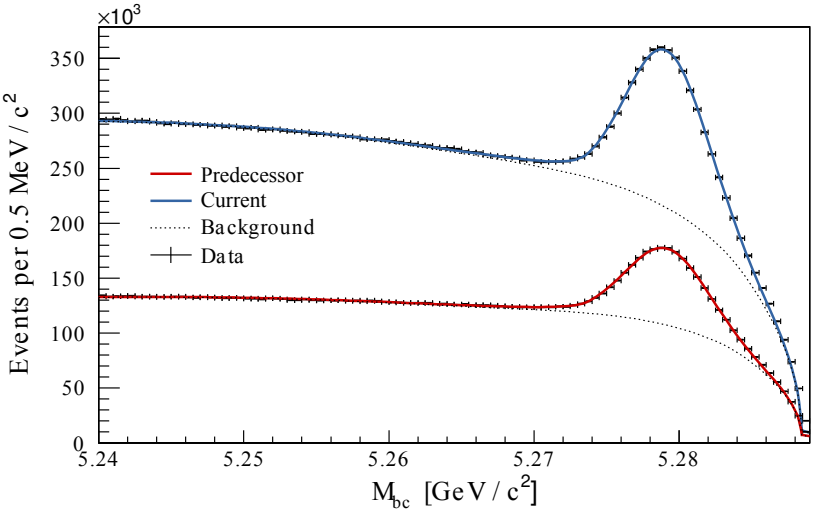


Figure 5.7.: $M_{\mathrm{bc}}$for the B^+_{tag} sample provided by the Full Reconstruction and its predecessor. The cut on $o_{\mathrm{NB}}$was chosen to achieve the same purity for both samples.

clearly answered, as the two aspects are connected: Without the improved algorithm, the addition of new channels would not have been possible. Nonetheless, this issue was investigated. For testing purposes, the Full Reconstruction described in this thesis was stripped of all newly added channels. The sample that resulted from this channel reduction was then analysed and compared to a sample of the predecessor. Both samples were obtained from the very same data set. The signal and background yields were obtained from fits to M_{bc}. The fit results are shown in figure 5.10. Table 5.6 summarises the results. It is evident, that a large fraction of the improvement was indeed achieved by the application of the more sophisticated algorithm. When performing this com-

Sample	Predecessor [10^3]	Current [10^3]	Factor
B^0 signal	48.4	71.6	1.48
B^0 background	135	135	1.00
B^+ signal	64.2	102	1.59
B^+ background	177	176	1.01

Table 5.6.: Extracted yields from the reduced channel fits. The cut on $o_{\mathrm{NB}}$was chosen in such a way that equal background levels could be achieved.

parison on a channel-by-channel basis, it became evident that the largest improvement could be achieved in channels with two or more light mesons in the final state, where

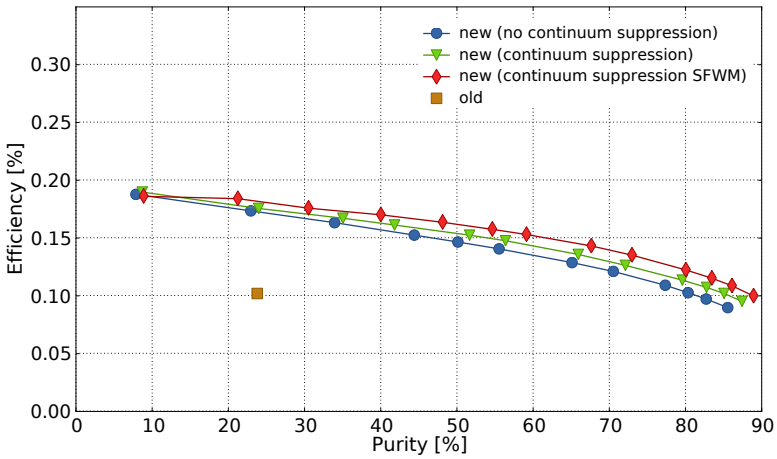


Figure 5.8.: The purity-efficiency curves for the B^0_{tag} sample provided by the Full Reconstruction.

this new Full Reconstruction algorithm imposes much weaker phase space-limits.

5.4. Summary

The Full Reconstruction algorithm is an essential analysis tool for a great number of analyses. It provides a tag side and is therefore also known as the *hadronic tag*. Using a hierarchical reconstruction algorithm combined with the advanced data analysis toolkit NeuroBayes allowed for a hierarchical flow of information. Thus, the signal probability for each candidate on each stage could be calculated and used as a very powerful discrimination variable. The postponing of the cut decision to later stages led to a very efficient selection. With a maximum reconstruction efficiency of 0.18% for B^0 mesons and 0.28% for B^+ mesons, this improved algorithm, combined with the addition of more decay channels, yielded an efficiency improvement over its predecessor of roughly a factor 2. This means, in effect, a doubling to the available data-set for many analyses overnight after 10 years of data-taking at the Belle detector.

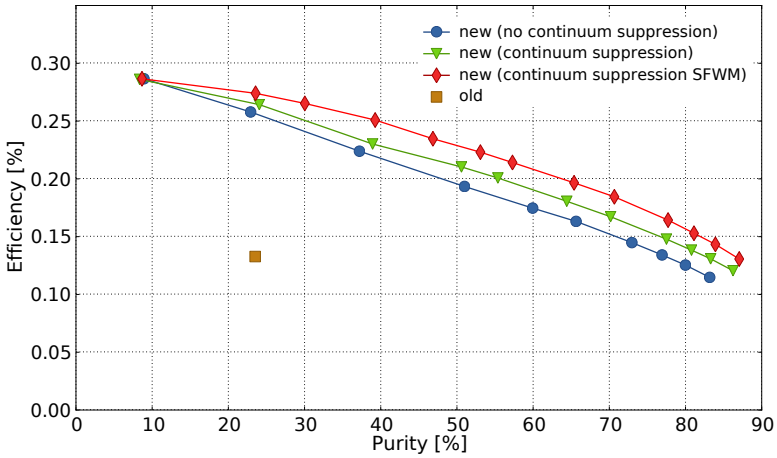
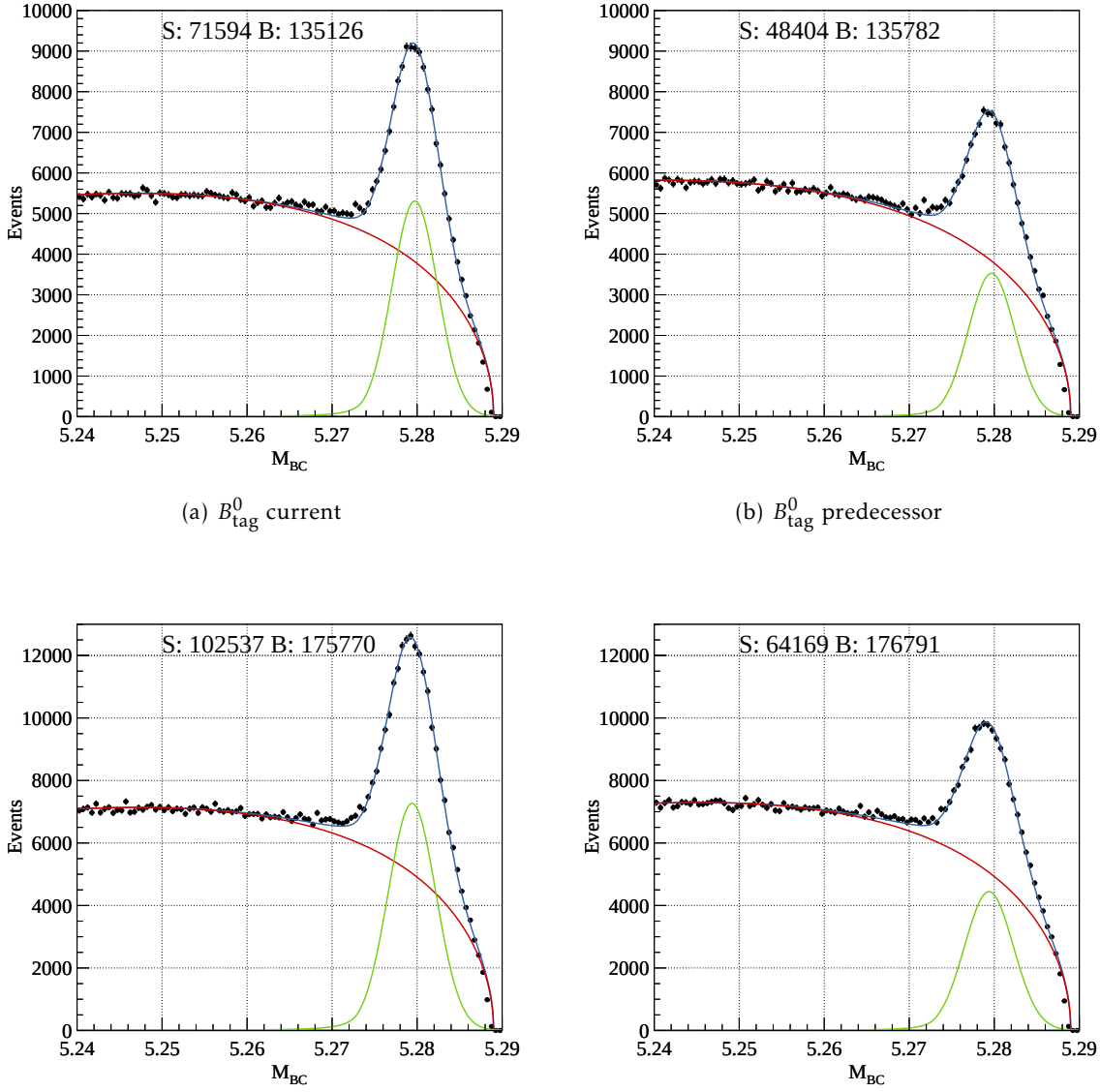


Figure 5.9.: The purity-efficiency curves for the B^+_{tag} sample provided by the Full Reconstruction.



(a) B^0_{tag} current

(b) B^0_{tag} predecessor

(c) B^+_{tag} current

(d) B^+_{tag} predecessor

Figure 5.10.: A comparison of the B_{tag}signal yields of the current Full Reconstruction and its predecessor. Only the decay channels that were present in both algorithms are included. The cut on $o_{\text{NB}}$was chosen in such a way that equal background levels could be achieved.

6. Signal Side

The goal of the presented analysis is to measure the ratios R^0, R^+, R^{*0} and R^{*+}, which are defined as

$$R^0 = \frac{\mathcal{B}(B^0 \to D^- \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \to D^- \ell^+ \nu_\ell)} \tag{6.1}$$

$$R^+ = \frac{\mathcal{B}(B^+ \to \overline{D}^0 \tau^+ \nu_\tau)}{\mathcal{B}(B^+ \to \overline{D}^0 \ell^+ \nu_\ell)} \tag{6.2}$$

$$R^{*0} = \frac{\mathcal{B}(B^0 \to D^{*-} \tau^+ \nu_\tau)}{\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu_\ell)} \tag{6.3}$$

$$R^{*+} = \frac{\mathcal{B}(B^+ \to \overline{D}^{*0} \tau^+ \nu_\tau)}{\mathcal{B}(B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell)}, \tag{6.4}$$

where $\ell = e, \mu$. Note that isospin symmetry suggests that $R^0 = R^+$ and $R^{*0} = R^{*+}$. The measurement of these ratios rather than branching fractions has the advantage that several systematic effects cancel. We call $B \to D^{(*)} \tau \nu_\tau$ signal decays the *signal modes* and the $B \to D^{(*)} \ell \nu_\ell$ decays the *normalisation modes*. The decays of the τ mesons that are considered are those into lighter leptons ℓ. These decay channels cover a relatively large branching fraction and are also experimentally comparably well accessible. This means that by reconstructing $B \to D^{(*)} \ell \nu_\ell$ decays, we implicitly also reconstruct $B \to D^{(*)} \tau \nu_\tau$ decays. It is therefore crucial for this analysis to distinguish not only the τ and ℓ signal channels from various backgrounds but also from each other. This introduces a fundamental problem, however: When we reconstruct the τ lepton from its decays into lighter leptons, the decay $B \to D^{(*)} \ell \nu_\ell$ is reconstructed from the same final state particles as the decay $B \to D^{(*)} \tau \nu_\tau$, as shown in equations 6.5 and 6.6.

$$B \to D \tau (\to \ell \nu \nu) \nu \tag{6.5}$$

$$B \to D \ell \nu \tag{6.6}$$

The only difference between the two decay channels is the number of neutrinos in the decay. These can, however, not be reconstructed with current particle detectors with even remotely sufficient efficiency. It is therefore necessary to measure the momentum of the neutrino system in an indirect manner. As already explained in great detail in chapter 5, the Full Reconstruction enables us to do exactly that.

6.1. Previous Measurements

Before we proceed with the decsription of the analysis, let us review the previous measurements done by the Belle and BaBar collaborations. Let us fist focus on the Belle measurements. There have been numerous efforts to measure the branching fractions $\mathcal{B}(B \to D^{(*)}\tau\nu_\tau)$ and the ratios R and R^*. They fall into two categories, distinguished by the different tagging methods.

6.1.1. Tagging Methods

The hadronic tagging, also known as the Full Reconstruction, endeavours to reconstruct as many purely hadronic $b \to c$ decays as possible. The efficiency of this reconstruction procedure is typically $\sim (0.1 - 0.2)\%$. Only after the hadronic tag is available, the signal side is reconstructed. This is also the tagging method that is used for the analysis presented in this thesis.

The inclusive tagging method follows a different approach. The visible final state particles of the signal side are reconstructed and combined first. All remaining final state particles in the event are then combined to form a tag side. This approach typically yields larger efficiency than the hadronic tagging, but also brings higher background levels with it.

6.1.2. First Measurement With Inclusive Tagging

The first measurement of the channel at the Belle collaboration [52] was performed in 2007 on a data set corresponding to $535\times10^6 \bar{B}B$ pairs. The only decay channel that was reconstructed was $B^0 \to D^{*-}\tau^+\nu_\tau$. The τ leptons were reconstructed in the channels $\tau^+ \to e^+\nu_e\bar{\nu}_\tau$ and $\tau^+ \to \pi^+\bar{\nu}_\tau$. The reconstruction of the D^* mesons was done in the channel $D^{*-} \to \bar{D}^0\pi^-$ and the D meson decay channel was $D^0 \to K^-\pi^+$. Additionally, the channel $D^0 \to K^-\pi^+\pi^0$ was used in the electron mode. The signal extraction was performed by a fit to the M_{bc}distribution of the inclusive B_{tag}. Figure 6.1 shows the signal extraction fit and the results are summarised in table 6.1. Taking the systematic uncertainty of 18.5% into account, this measurement yields a signal significance of 5.2σ. Thus, this measurement constitutes the first observation of an exclusive semitauonic B decay.

6.1.3. Charged B Measurement With Inclusive Tagging

With a larger data set of $657 \times 10^6 \bar{B}B$ pairs, another analysis using the inclusive tagging method was performed in 2010 by the Belle collaboration [53]. The B mesons were reconstructed in the channels $B^+ \to \bar{D}^0\tau^+\nu_\tau$ and $B^+ \to \bar{D}^{*0}\tau^+\nu_\tau$. The D^* mesons were reconstructed in the decay $D^{*0} \to D^0\pi^0$ with the D meson channels $D^0 \to K^-\pi^+$ and $D^0 \to K^-\pi^+\pi^0$. The τ mesons were reconstructed in leptonic and hadronic decays: $\tau^+ \to e^+\nu_e\bar{\nu}_\tau$, $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$ and $\tau^+ \to \pi^+\bar{\nu}_\tau$. The signal was extracted by a 2-

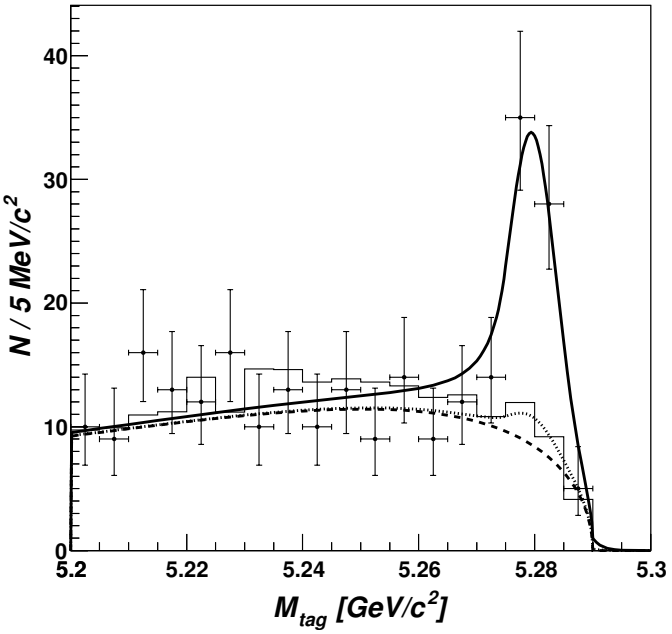


Figure 6.1.: Summary of the results of the first measurement with inclusive tagging [52].

dimensional fit to the $M_{\rm bc}$distribution of the $B_{\rm tag}$and the momentum of the signal side D meson in the $\Upsilon(4S)$ frame. Figure 6.2 shows the signal extraction fit and the results are summarised in table 6.1. This measurement had systematic uncertainties of 13.9% in the $B^+ \to \bar{D}^{*0}\tau^+\nu_\tau$ channel and 15.2% in the $B^+ \to \bar{D}^0\tau^+\nu_\tau$ channel. Taking these uncertainties into account, this measurement yields a signal significance of 8.1σ for the D^* channel and 3.5σ for the D channel. Thus, it constitutes the first evidence for the decay $B^+ \to \bar{D}^0\tau^+\nu_\tau$.

6.1.4. All Decay Channels With Hadronic Tagging

The only Belle measurement of $B \to D^{(*)}\tau\nu_\tau$ using hadronic tagging was done in 2009 on the dataset of $657 \times 10^6 \bar{B}B$ pairs [54]. The B meson decay channels were $B^0 \to D^-\tau^+\nu_\tau$, $B^0 \to D^{*-}\tau^+\nu_\tau$, $B^+ \to \overline{D}^0\tau^+\nu_\tau$ and $B^+ \to \overline{D}^{*0}\tau^+\nu_\tau$. The D^* mesons were reconstructed in the decay channels $D^{*0} \to (D^0\pi^0), (D^0\gamma)$ and $D^{*+} \to (D^+\pi^0), (D^0\pi^+)$. The used D meson decay channels were $D^0 \to (K^-\pi^+), (K^-\pi^+\pi^0), (K^-\pi^+\pi^+\pi^-), (K^-\pi^+\pi^+\pi^-\pi^0)$, $(K^0_S\pi^0), (K^0_S\pi^+\pi^-)$ and $(K^0_S\pi^+\pi^-\pi^0)$ as well as $D^- \to (K^+\pi^-\pi^-), (K^+\pi^-\pi^-\pi^0)$ and $(K^0_S\pi^-)$. The τ leptons were reconstructed only leptonically: $\tau^+ \to e^+\nu_e\bar{\nu}_\tau$ and $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$. For the first time, a normalisation sample was reconstructed in a Belle analysis as well. It consisted of the decays $B^0 \to D^-\ell^+\nu_\ell$, $B^0 \to D^{*-}\ell^+\nu_\ell$, $B^+ \to \overline{D}^0\ell^+\nu_\ell$ and $B^0 \to D^{*-}\ell^+\nu_\ell$. These were used for the calculation of R and R^*. The signal was ex-

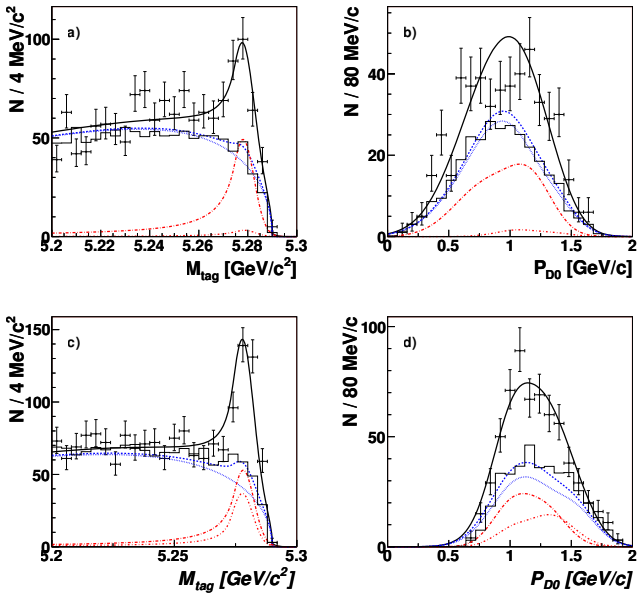


Figure 6.2.: Signal extraction fit of the charged B measurement with inclusive tagging [53].

tracted by 2-dimensional fits to M^2_{miss} and $E^{\text{ECL}}_{\text{extra}}$ defined as

$$M^2_{\text{miss}} = \left[p(\text{Beam}) - \left(p(B_{\text{tag}}) + p(D^{(*)}) + p(\ell)\right)\right]^2 \text{ and} \tag{6.7}$$

$$E^{\text{ECL}}_{\text{extra}} = \sum E_{\text{Calor.}} - \left(\sum E_{\text{tag}} + \sum E_{\text{signal}}\right) \tag{6.8}$$

Projections of the fits can be found in figure 6.3, the results are also summarised in table 6.1.

6.1.5. BaBar Measurements

The BaBar collaboration has made similar efforts as the Belle collaboration to measure the branching fractions $\mathcal{B}(B \to D^{(*)}\tau\nu_\tau)$ and the ratios R and R^*. In 2007 a first measurement was performed [55], using a data sample of approximately 232×10^6 $B\bar{B}$ pairs. The extracted values for $R^{(*)}$ are

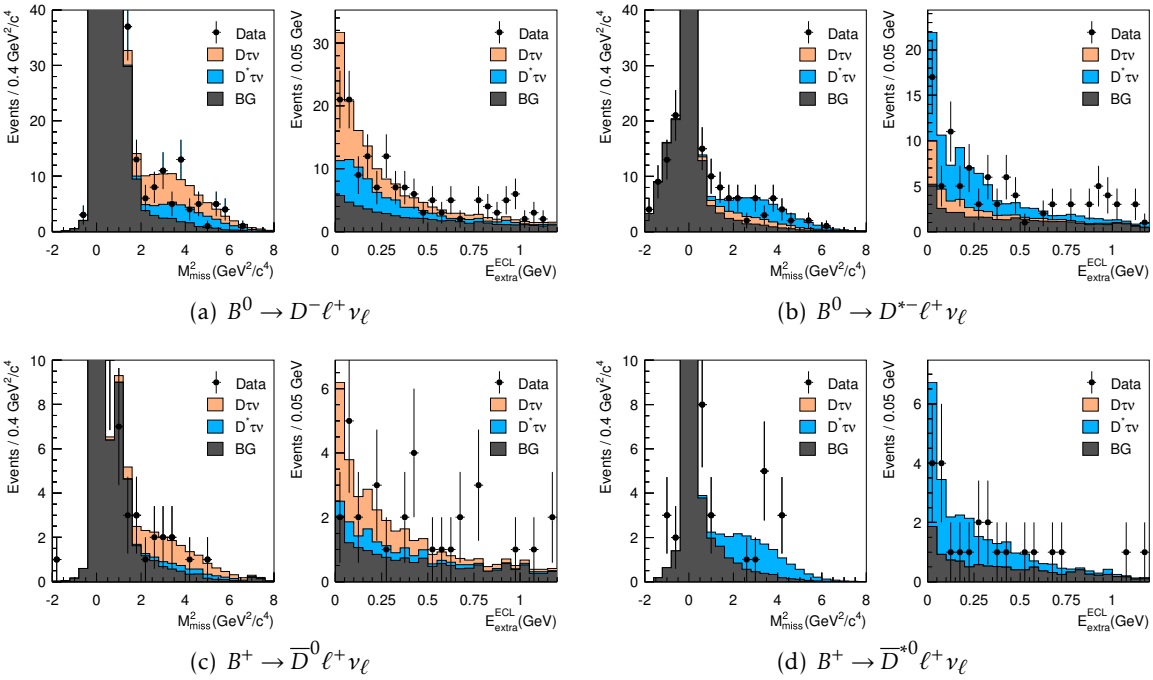
$$R = 0.416 \pm 0.117 \pm 0.052 \quad \text{and} \tag{6.9}$$

$$R^* = 0.297 \pm 0.056 \pm 0.018\,. \tag{6.10}$$

The analysis was updated in 2012 [56] with and increased data sample of approximately 471×10^6 $B\bar{B}$ pairs and and improved tagging method. The extracted values for $R^{(*)}$ are

$$R = 0.440 \pm 0.058 \pm 0.042 \quad \text{and} \tag{6.11}$$

$$R^* = 0.332 \pm 0.024 \pm 0.018\,. \tag{6.12}$$



(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.3.: Projections of the signal extraction fits [54].

Combined, these two values deviate from the Standard Model by 3.4σ and exclude the type II 2 Higgs Doublet Model with a 99.8% confidence level.

6.1.6. Conclusion

Since 2007, there have been numerous measurements by the Belle collaboration focusing on $B \to D^{(*)}\tau\nu_\tau$. Both hadronic and inclusive tagging methods have been used successfully. Table 6.1 summarises the results. Recent measurements by the BaBar collaboration indicate a deviation from the Standard Model and rule out the type II 2 Higgs Doublet Model at the same time. The analysis in this thesis endeavours to test the observed tension.

Tagging	Measurement	Result
Inclusive	$\mathcal{B}(B^0 \to D^{*-}\tau^+\nu_\tau)$	$(2.02^{+0.40}_{-0.37} \pm 0.37)\%$
Inclusive	$\mathcal{B}(B^+ \to \overline{D}^0\tau^+\nu_\tau)$	$(0.77 \pm 0.22 \pm 0.12)\%$
Inclusive	$\mathcal{B}(B^+ \to \overline{D}^{*0}\tau^+\nu_\tau)$	$(2.12^{+0.28}_{-0.27} \pm 0.29)\%$
Hadronic	$R(D^0)$, $R(D^+)$	$0.70^{+0.19+0.11}_{-0.18-0.9}$, $0.48^{+0.22+0.06}_{-0.19-0.05}$
Hadronic	$R(D^{*0})$, $R(D^{*+})$	$0.47^{+0.11+0.06}_{-0.10-0.07}$, $0.48^{+0.14+0.06}_{-0.12-0.04}$
Hadronic	$\mathcal{B}(B^+ \to \overline{D}^0\tau^+\nu_\tau)$	$(1.51^{+0.41+0.24}_{-0.39-0.19} \pm 0.15)\%$
Hadronic	$\mathcal{B}(B^+ \to \overline{D}^{*0}\tau^+\nu_\tau)$	$(3.04^{+0.69+0.40}_{-0.66-0.47} \pm 0.22)\%$
Hadronic	$\mathcal{B}(B^0 \to D^-\tau^+\nu_\tau)$	$(1.01^{+0.46+0.13}_{-0.41-0.11} \pm 0.10)\%$
Hadronic	$\mathcal{B}(B^0 \to D^{*-}\tau^+\nu_\tau)$	$(2.56^{+0.75+0.31}_{-0.66-0.22} \pm 0.10)\%$

Table 6.1.: Summary of previous Belle measurements [52, 53, 54].

6.2. Simulated And Real Data

In many collaborations in particle physics, it is good practice to develop an analysis not on the real data that was measured by the detector. Instead, a sample of simulated data, usually called "Monte Carlo" is used. For the Belle collaboration, the accurate simulation of the decays of the $B\bar{B}$ meson pairs from the $\Upsilon(4S)$ resonance is of great importance. Another important point is the correct estimation of the various background processes that are present on real data. Therefore, $e^+e^- \to (c\bar{c}), (s\bar{s}), (u\bar{u}), (d\bar{d})$ transitions are also simulated. The set of simulated data containing all those processes is called "generic Monte Carlo".

A rather large computational effort is required to generate simulated data that reflects the real data in a satisfactory manner. The first step is to simulate the mere physical processes. Correct simulation of masses, widths, lifetimes, branching ratios and angular distributions are amongst the most important requirements. This simulation is performed by two different packages: The *EvtGen* package [57] and the PYTHIA software [58] are used for the simulation of the decays of the B mesons coming from $\Upsilon(4S) \to B\bar{B}$ and of the e^+e^- annihilation and production of lighter quark-antiquark pairs (u, d, s, c)

The second step in the production of Monte Carlo data is to simulate the detector response to each simulated physics event. This requires a detailed description of the geometry and the materials of the detector. This description is then used by the GEANT software [59], which simulates the detector response.

For the development of certain analyses it is desirable to have as much simulated data as possible. But the production of Monte Carlo data is time consuming, the simulation of the detector response especially so.

The equivalent of the amount of real data that was recorded by the Belle detector is called one *stream*. 6 streams of generic Monte Carlo were produced. Additionally, 4 streams were produced that contained only $B\bar{B}$ events. While they do not represent real data with the same accuracy as the complete generic Monte Carlo samples, they are still useful for certain aspects of some analyses. Signal shapes or $B\bar{B}$ background can still be modeled using these samples.

6.3. Signal Side Selection

For the measurement of R and R^*, we need to reconstruct the *normalization modes* $B \to D^{(*)}\ell\nu_\ell$ as well as the *signal modes* $B \to D^{(*)}\tau\nu_\tau$. The signal side decays were selected in the modes listed in table 6.2. The neutrinos can, of course, not be selected and are just mentioned for completeness. As the τ leptons are reconstructed from their decays into e and μ only, the reconstruction of the $B \to D\tau\nu_\tau$ and $B \to D^*\tau\nu_\tau$ decay channels is done implicitly. The decay channels of the D and D^* mesons are listed in table 6.3.

$B^0 \to D^- e^+ \nu_e$	$B^+ \to \overline{D}^0 e^+ \nu_e$
$B^0 \to D^{*-} e^+ \nu_e$	$B^+ \to \overline{D}^{*0} e^+ \nu_e$
$B^0 \to D^- \mu^+ \nu_\mu$	$B^+ \to \overline{D}^0 \mu^+ \nu_\mu$
$B^0 \to D^{*-} \mu^+ \nu_\mu$	$B^+ \to \overline{D}^{*0} \mu^+ \nu_\mu$

Table 6.2.: B meson decay channels used for the signal side reconstruction.

$D^+ \to K^-\pi^+\pi^+$	$D^0 \to K^-\pi^+$
$D^+ \to K^0_S\pi^+$	$D^0 \to K^-\pi^+\pi^+\pi^-$
$D^+ \to K^0_S\pi^+\pi^0$	$D^0 \to K^-\pi^+\pi^0$
$D^+ \to K^-\pi^+\pi^+\pi^0$	$D^0 \to K^0_S\pi^0$
$D^+ \to K^0_S\pi^+\pi^+\pi^-$	$D^0 \to K^0_S\pi^+\pi^-$
$D^{*+} \to D^0\pi^+$	$D^{*0} \to D^0\pi^0$
$D^{*+} \to D^+\pi^0$	$D^{*0} \to D^0\gamma$

Table 6.3.: D and D^* meson decay channels used for the signal side reconstruction.

6.3.1. Pre-cuts

A small set of relatively soft pre-cuts was applied during the reconstruction in order to keep the background levels manageable. These cuts are listed in table 6.4.

6.3.2. Choice Of B_{tag}

In every event, the reconstruction of each D meson listed in table 6.3 was attempted. All of the candidates were checked for overlaps with the B_{tag} collection. Overlaps would mean that a track or an electromagnetic cluster that was already used in the B_{tag} reconstruction was used again for the signal side reconstruction. If at least one B_{tag} was found, that had no overlap with the D meson candidate, the candidate was kept. It was

Cuts
Allowed combinations: $(B^+_{\text{tag}}B^-_{\text{sig}})$, $(B^-_{\text{tag}}B^+_{\text{sig}})$, $(B^0_{\text{tag}}B^0_{\text{sig}})$, $(\bar{B}^0_{\text{tag}}B^0_{\text{sig}})$, $(B^0_{\text{tag}}\bar{B}^0_{\text{sig}})$ and $(\bar{B}^0_{\text{tag}}\bar{B}^0_{\text{sig}})$
Signal Side final state particles consistent with at least one tag side
$M_{bc,\text{tag}} > 5.22$ GeV/c^2 (as dictated by Full Reconstruction)
-0.15 GeV $< \Delta E_{\text{tag}} < 0.10$ GeV (as dictated by Full Reconstruction)
$e^\pm$ ID > 0.4, $\mu^\pm$ ID > 0.4, $\pi^\pm$ ID > 0.1 and $K^\pm$ ID > 0.1
goodKs== 1
$dr < 2.0$cm and $dz < 4.0$cm for charged tracks.
$E(\gamma) > 50$ MeV
$E(\gamma$ from $\pi^0) > 50$ MeV
110 MeV/$c^2 < M(\pi^0) < 150$ MeV/c^2
$\lvert\vec{p}^*(D^{0/+})\rvert < 3.0$ GeV/c
$M(D)$ within a 160 MeV/c^2 window of the nominal mass except for the modes below
1.70 GeV/$c^2 < M(D) <$ 2.10 GeV/c^2 for $D^+ \to K^0_S\pi^+\pi^0$ and $D^0 \to K^-\pi^+\pi^0$
1.60 GeV/$c^2 < M(D) <$ 2.10 GeV/c^2 for $D^0 \to K^0_S\pi^0$
$0.125 < \Delta M(D^*, D) < 0.160$ for $D^{*+} \to D^0\pi^+$ and $D^{*+} \to D^+\pi^0$
$0.122 < \Delta M(D^*, D) < 0.162$ for $D^{*0} \to D^0\pi^0$
$0.109 < \Delta M(D^*, D) < 0.175$ for $D^{*0} \to D^0\gamma$
Not more than 1 remaining track with $dr < 2.0$cm and $dz < 4.0$cm

Table 6.4.: Reconstruction cuts

then combined with each e and μ candidate that passed the cuts in table 6.4 and was not already used for any B_{tag} or on the signal side. The resulting signal side candidates were again checked for overlaps with B_{tag}, as it can occur that the D meson is compatible with a B_{tag} with which the used ℓ candidate is not. While in principle, only one final check for overlaps would be sufficient, it has proven a great advantage in program execution speed to perform the check for overlaps also for D mesons and lepton tracks. If there were multiple B_{tag} candidates compatible with the signal side B meson, a choice had to be made which B_{tag} candidate should be used. The B_{tag} candidate with the largest number of charged tracks without overlap was chosen. In a true signal event no additional charged tracks can occur, so this method of choosing is the most efficient one. Choosing a B_{tag} candidate with less charged tracks would lead to remaining tracks in the event. Such a candidate would not pass the later selection critera, described in section 6.4. If there should be two candidates with the same (largest possible) number

of tracks, the one with the better $o_{\mathrm{NB}}$was chosen.

6.4. Analysis Samples

After the reconstruction finished, the data set was cleaned up further with several soft pre-cuts listed in tables 6.5 and 6.6. More finely grained cuts on the D meson mass and the mass difference between the D^* mesons and their D mesons were applied, the exact cut values are listed in table 6.7. In order to obtain meaningful cuts on $M(D)$ and $\Delta M(D^*D)$, fits to these distributions were performed for each D and D^* meson decay channel individually. The fit results can be found in table 6.8 and are the basis for the cuts listed in table 6.7. Additionally, a cut on the momentum transfer squared q^2, defined as

$$q^2 = \left(p(B_{\mathrm{tag}}) - p(D^*_{\mathrm{sig}})\right)^2 \tag{6.13}$$

was performed. The purpose of this cut is to reduce the otherwise overwhelming signal contribution of the $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ normalisation modes. The requirement for no remaining π^0 particles applies to π^0 candidates within the mass range $[0.1178, 0.1502]\,\mathrm{GeV}/c^2$ and with photon energies $E(\gamma) > 50$ MeV in the barrel region, $E(\gamma) > 100$ MeV in the forward endcap or $E(\gamma) > 150$ MeV in the backward endcap. The cuts on $o_{\mathrm{NB,tag\ side}}$ were determined by requiring 90% efficiency of each individual cut for the combined τ signal of the channels $B \to D\tau\nu_\tau$ and $B \to D^*\tau\nu_\tau$.

Cuts
Allowed combinations: $(B^+_{\text{tag}}B^-_{\text{sig}})$, $(B^-_{\text{tag}}B^+_{\text{sig}})$, $(B^0_{\text{tag}}B^0_{\text{sig}})$, $(\bar{B}^0_{\text{tag}}B^0_{\text{sig}})$, $(B^0_{\text{tag}}\bar{B}^0_{\text{sig}})$ and $(\bar{B}^0_{\text{tag}}\bar{B}^0_{\text{sig}})$
Signal Side final state particles consistent with at least one tag side
Tag Side channel dependent $o_{\text{NB,tag side}}$ cut, see table 6.6
$M_{bc,\text{tag}} > 5.27$ GeV/c^2
-0.15 GeV $< \Delta E_{\text{tag}} < 0.10$ GeV
$e^\pm$ ID > 0.6
$\mu^\pm$ ID > 0.9
$\pi^\pm$ ID > 0.1
$K^\pm$ ID > 0.1
eID for $\pi^\pm$ and $K^\pm < 0.9$
μID for $\pi^\pm$ and $K^\pm < 0.9$
goodKs== 1
$dr < 2.0$cm and $dz < 4.0$cm for charged tracks.
$E(\gamma) > 50$ MeV
$E(\gamma$ from $\pi^0) > 50$ MeV
$M(\pi^0)\ \epsilon\ (110\ldots150)$ MeV/c^2
$p^*(\pi^0) > 200$ MeV/c^2 except for slow π^0 from D^*
D mass cuts and D^*D mass difference cuts, see table 6.7
$\lvert\vec{p^*}(D^{0/+})\rvert < 3.0$ GeV/c
$q^2 > 4.0$ GeV$^2/c^4$
No remaining tracks with $dr < 2.0$cm and $dz < 4.0$cm
No remaining π^0 particles

Table 6.5.: Analysis cuts

Channel	Cut	Channel	Cut
$B^- \to D^{*0}\pi^-$	-3.72	$B^0 \to D^{*+}\pi^-$	-2.40
$B^- \to D^{*0}\pi^-\pi^0$	-5.40	$B^0 \to D^{*+}\pi^-\pi^0$	-5.28
$B^- \to D^{*0}\pi^-\pi^-\pi^+$	-6.12	$B^0 \to D^{*+}\pi^-\pi^+\pi^-$	-4.92
$B^- \to D^0\pi^-$	-3.60	$B^0 \to D^+\pi^-$	-3.24
$B^- \to D^0\pi^-\pi^0$	-5.04	$B^0 \to D^+\pi^-\pi^0$	-4.80
$B^- \to D^0\pi^-\pi^-\pi^+$	-6.12	$B^0 \to D^+\pi^-\pi^+\pi^-$	-6.24
$B^- \to D^{*0}D_S^{*-}$	-5.76	$B^0 \to D^{*+}D_S^{*-}$	-6.12
$B^- \to D^{*0}D_S^-$	-6.36	$B^0 \to D^{*+}D_S^-$	-4.68
$B^- \to D^0D_S^{*-}$	-6.48	$B^0 \to D^+D_S^{*-}$	-4.80
$B^- \to D^0D_S^-$	-7.32	$B^0 \to D^+D_S^-$	-6.24
$B^- \to J/\psi K^-$	-0.96	$B^0 \to J/\psi K_s^0$	-0.48
$B^- \to J/\psi K^-\pi^+\pi^-$	-2.88	$B^0 \to J/\psi K^-\pi^+$	-0.60
$B^- \to D^0K^-$	-5.40	$B^0 \to J/\psi K_s^0\pi^+\pi^-$	-4.32
$B^- \to D^+\pi^-\pi^-$	-5.52	$B^0 \to D^0\pi^0$	-5.40
$B^- \to D^{*0}\pi^-\pi^-\pi^+\pi^0$	-7.20	$B^0 \to D^{*+}\pi^-\pi^-\pi^+\pi^0$	-6.24
$B^- \to J/\psi K^-\pi^0$	-5.76		
$B^- \to J/\psi K_S^0\pi^-$	-1.56		

Table 6.6.: Analysis cuts on $\log(o_{\text{NB,tag side}})$

Channel	sample	lower SB	Signal region	upper SB
			[GeV/c^2]	
$D^+ \to K^0_S\pi^+$	MC	[1.816, 1.838]	[1.848, 1.892]	[1.902, 1.924]
$D^+ \to K^0_S\pi^+$	Data	[1.811, 1.835]	[1.845, 1.894]	[1.904, 1.929]
$D^+ \to K^0_S\pi^+\pi^0$	MC	[1.740, 1.780]	[1.830, 1.910]	[1.920, 1.960]
$D^+ \to K^0_S\pi^+\pi^0$	Data	[1.721, 1.770]	[1.820, 1.919]	[1.929, 1.978]
$D^+ \to K^-\pi^+\pi^+$	MC	[1.821, 1.841]	[1.850, 1.889]	[1.899, 1.918]
$D^+ \to K^-\pi^+\pi^+$	Data	[1.815, 1.838]	[1.847, 1.892]	[1.902, 1.924]
$D^+ \to K^-\pi^+\pi^+\pi^0$	MC	[1.801, 1.830]	[1.840, 1.899]	[1.909, 1.938]
$D^+ \to K^-\pi^+\pi^+\pi^0$	Data	[1.801, 1.831]	[1.841, 1.899]	[1.909, 1.938]
$D^+ \to K^0_S\pi^+\pi^+\pi^-$	MC	[1.823, 1.841]	[1.851, 1.888]	[1.898, 1.917]
$D^+ \to K^0_S\pi^+\pi^+\pi^-$	Data	[1.820, 1.840]	[1.850, 1.890]	[1.900, 1.920]
$D^0 \to K^0_S\pi^0$	MC	[1.625, 1.695]	[1.795, 1.935]	[1.945, 2.014]
$D^0 \to K^0_S\pi^0$	Data	[1.616, 1.691]	[1.791, 1.939]	[1.949, 2.023]
$D^0 \to K^-\pi^+$	MC	[1.814, 1.834]	[1.844, 1.885]	[1.895, 1.916]
$D^0 \to K^-\pi^+$	Data	[1.806, 1.830]	[1.840, 1.889]	[1.899, 1.924]
$D^0 \to K^-\pi^+\pi^0$	MC	[1.711, 1.758]	[1.818, 1.912]	[1.922, 1.969]
$D^0 \to K^-\pi^+\pi^0$	Data	[1.687, 1.746]	[1.806, 1.924]	[1.934, 1.992]
$D^0 \to K^0_S\pi^+\pi^-$	MC	[1.812, 1.833]	[1.843, 1.886]	[1.896, 1.918]
$D^0 \to K^0_S\pi^+\pi^-$	Data	[1.805, 1.830]	[1.840, 1.890]	[1.900, 1.924]
$D^0 \to K^-\pi^+\pi^+\pi^-$	MC	[1.814, 1.835]	[1.845, 1.885]	[1.895, 1.915]
$D^0 \to K^-\pi^+\pi^+\pi^-$	Data	[1.807, 1.831]	[1.841, 1.889]	[1.899, 1.923]
$D^{*+} \to D^+\pi^0$	MC	-	[0.139, 0.143]	[0.145, 0.149]
$D^{*+} \to D^+\pi^0$	Data	-	[0.138, 0.143]	[0.145, 0.149]
$D^{*+} \to D^0\pi^+$	MC	-	[0.143, 0.148]	[0.150, 0.155]
$D^{*+} \to D^0\pi^+$	Data	-	[0.142, 0.148]	[0.150, 0.156]
$D^{*0} \to D^0\gamma$	MC	[0.114, 0.127]	[0.129, 0.155]	[0.157, 0.170]
$D^{*0} \to D^0\gamma$	Data	[0.114, 0.127]	[0.129, 0.155]	[0.157, 0.171]
$D^{*0} \to D^0\pi^0$	MC	-	[0.140, 0.144]	[0.147, 0.152]
$D^{*0} \to D^0\pi^0$	Data	-	[0.140, 0.145]	[0.148, 0.153]

Table 6.7.: Analysis cuts for D meson masses and D^*D mass differences. The cut width correspond to 3 to 4 times the fitted widths, depending on the channel.

	width [MeV/c^2]		
Channel	MC	Data	ratio
$M(D^+ \to K^0_S \pi^+)$	5.51	6.14	1.11
$M(D^+ \to K^0_S \pi^+ \pi^0)$	10.01	12.30	1.23
$M(D^+ \to K^- \pi^+ \pi^+)$	4.78	5.53	1.16
$M(D^+ \to K^- \pi^+ \pi^+ \pi^0)$	7.36	7.27	0.99
$M(D^+ \to K^0_S \pi^+ \pi^+ \pi^-)$	4.64	5.00	1.08
$M(D^0 \to K^0_S \pi^0)$	17.45	18.58	1.06
$M(D^0 \to K^- \pi^+)$	5.12	6.14	1.20
$M(D^0 \to K^- \pi^+ \pi^0)$	15.62	19.58	1.20
$M(D^0 \to K^0_S \pi^+ \pi^-)$	5.37	6.20	1.15
$M(D^0 \to K^- \pi^+ \pi^+ \pi^-)$	5.08	5.96	1.17
ΔM (D, $D^{*+} \to D^+ \pi^0$)	2.05	2.21	1.08
ΔM (D, $D^{*+} \to D^0 \pi^+$)	1.31	1.48	1.13
ΔM (D, $D^{*0} \to D^0 \gamma$)	13.09	13.27	1.01
ΔM (D, $D^{*0} \to D^0 \pi^0$)	2.15	2.61	1.21

Table 6.8.: Results of the fits to D meson mass spectra and D^*D mass difference spectra

6.4.1. D Mass Sideband Sample

Apart from the signal sample, an additional sample was produced, also with the cuts described in table 6.5, but with D meson mass cuts that excluded the D meson signal region. Two D mass regions were selected for this sample, one above the mass window for the signal region, and the other one below. The two regions were chosen to have a certain distance from the signal window and to have a combined width equal to the width of signal mass region. This choice of sideband widths ensures that the composition of backgrounds in the signal box is correctly reflected in the sidebands. The exact values for each channel are shown in table 6.7. The sideband regions are also marked in figures A.4 to A.10 in appendix A.2.

6.4.2. $\Delta M(D^*D)$ Sideband Sample

Additionally, a $\Delta M(D^*D)$ sideband sample was produced. It had to fulfil the cuts described in table 6.5 and the D meson mass signal region cuts in table 6.7. The $\Delta M(D^*D)$ sideband cuts from the same table were then applied. The sideband regions are also marked in figures A.4 to A.10 in appendix A.2.

6.5. Correction of Monte Carlo

Simulated data is used extensively in many areas of high energy physics as it is crucial in the developement of a great number of analyses. It is often the only method of determining the composition of an analysed sample, reconstruction efficiencies of the individual components and detector acceptance. The Belle collaboration has endeavoured to create a simulated data set that describes the real data taken by the Belle detector as precisely as possible (see section 6.2). While this *generic Monte Carlo* sample delivers a quite accurate description of real data for many cases, there are also some known imperfections. In order to better simulate the real data, several correction can be applied to the generic Monte Carlo samples.

6.5.1. Tag Side Bias

The developement of the Full Reconstruction, and especially the NeuroBayes Experts that are used, relied heavily on the usage of Monte Carlo samples. During the tests of the Full Reconstruction, it has become apparent, that the branching fractions of certain B meson decays were not modelled correctly. In fact, for some high multiplicity B meson decays, the Monte Carlo delivers as much as double the signal that is observed on real data. A set of correction factors for each single tag side B meson decay channel has to be determined. By comparing the fit yields obtained in a $B \to D^* \ell \nu$ signal side analysis on generic Monte Carlo and data, such a set of correction factors can be obtained [60].
The original distribution for B_{tag} M_{bc} together with the reweighted distributions, accounting for the tag side bias, can be seen in figure 6.4. We expect the effects of the tag side bias to cancel in the measurement of the ratios R and R^*. This correction is still aplied, as it improves the estimates of expected signal and background events.

6.5.2. D^{**} Components

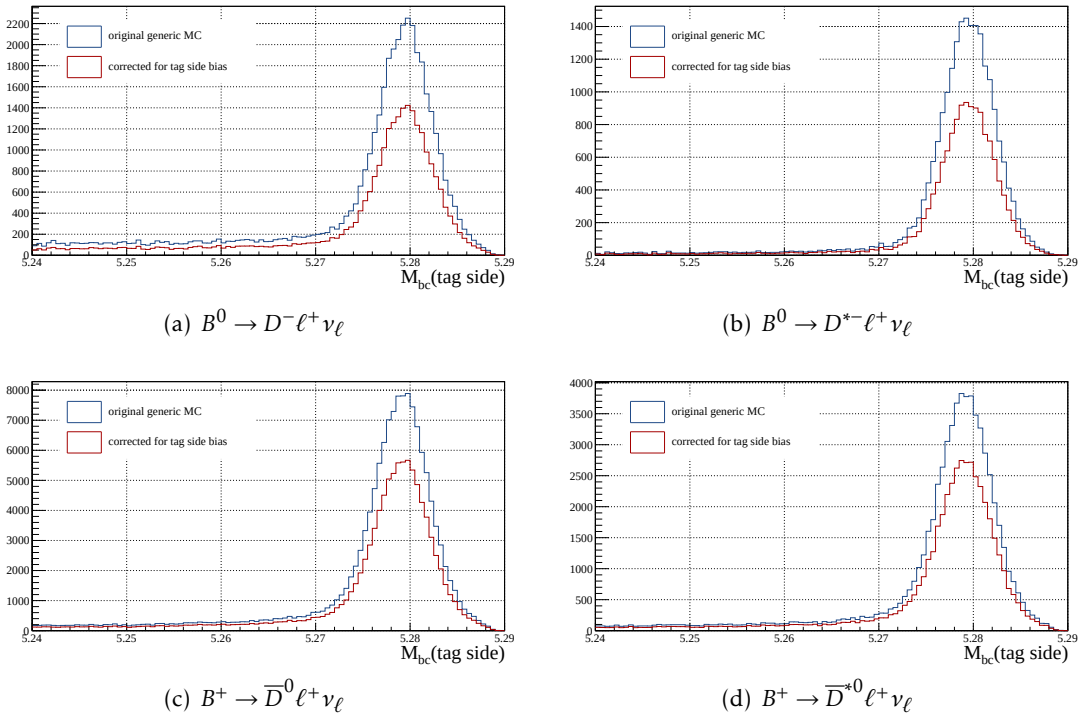
In the standard Belle generic Monte Carlo, the semileptonic $B \to D^{**} \ell \nu_\ell$ decays are generated accoring to the ISGW2 model [61, 62]. More sophisticated models have appeared after the generation of the generic Monte Carlo [63], called LLSW [64, 65]. Event weights to change the ISGW2 shapes into LLSW shapes were calculated [66]. Three different weights have been calculated, corresponding to the two extreme values in the allowed parameter space and the average:

$$\eta = -0.45 \qquad \hat{\tau}_1 = -1.2 \qquad \hat{\zeta}_1 = -1.6 \tag{6.14}$$

$$\eta = -0.15 \qquad \hat{\tau}_1 = -1.2 \qquad \hat{\zeta}_1 = -1.6 \tag{6.15}$$

$$\eta = +0.15 \qquad \hat{\tau}_1 = -1.2 \qquad \hat{\zeta}_1 = -1.6 \tag{6.16}$$

The reweighting was done in 22 bins in the q^2 range of $0 \ldots 11\ \text{GeV}^2/c^4$. The original M^2_{miss}distributions for the D^{**} background together with the reweighted distributions can be seen in figure 6.5.



(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.4.: Effect of the tag side bias corection on $B_{\text{tag}} M_{\text{bc}}$.

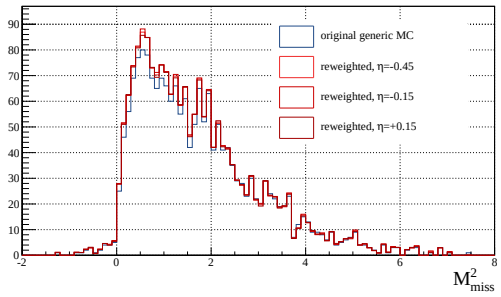
6.5.3. Lepton ID Efficiency

There are known discrepancies between data and Monte Carlo in the lepton ID efficiency. The necessary corrections were calcualted using $\gamma\gamma \to \ell^+\ell^-$ processes [67] as functions of polar angle θ and momentum p. To account for influences of a hadronic environment to the lepton ID efficiency, an inclusive $B \to XJ/\psi(\to \ell^+\ell^-)$ sample was reconstructed and used for comparison.

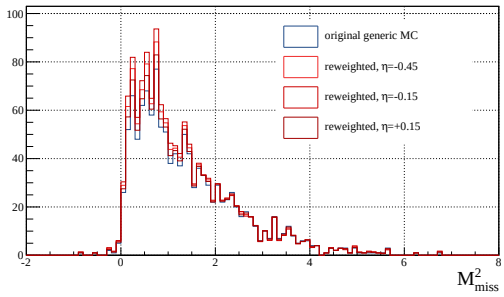
Using a standardised tool to calculate the correction [68], tables 6.9 and 6.10 were produced.

6.5.4. Lepton Fake Rates

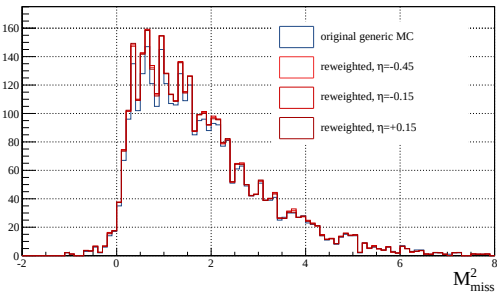
The cuts on particle ID (see table 6.5) result in quite clean samples of charged final state particles. These samples, though very clean, can never be perfectly clean. Some real K^+ mesons, wrongfully identified as μ^+ leptons, end up in the μ^+ sample and vice versa. Such fakes occur for all types of particles. For the K^+ and π^+ mesons used for the reconstruction of the D meson, this sort of background is part of the wrong $D^{(*)}$ background component (see section 6.10). The lepton samples (e^+ and μ^+) contain incorrectly identified particles as well. The lepton fake rates can differ on Monte Carlo



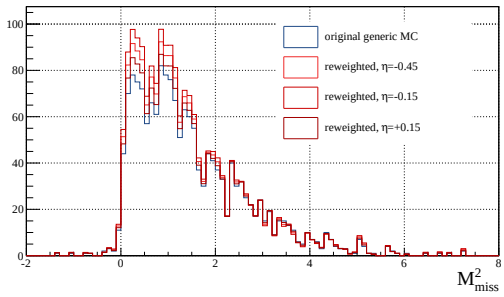
(a) $B^0 \to D^- \ell^+ \nu_\ell$



(b) $B^0 \to D^{*-} \ell^+ \nu_\ell$



(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$



(d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.5.: Effect of the D^{**} shape correction on M^2_{miss}.

	exp	lepton signal	tau signal	ratio
eID	7 - 27	0.9866 ± 0.0213	0.9629 ± 0.0309	1.0246
eID	31 - 65	0.9825 ± 0.0165	0.9681 ± 0.0243	1.0149
μID	7 - 27	0.9661 ± 0.0252	0.9492 ± 0.0253	1.0178
μID	31 - 39, 45a	0.9554 ± 0.0193	0.9795 ± 0.0182	0.9754
μID	41 - 49	0.9503 ± 0.0197	0.9740 ± 0.0180	0.9757
μID	51 - 65	0.9794 ± 0.0194	0.9906 ± 0.0195	0.9887

Table 6.9.: Lepton Efficiency corrections for B^0

simulation and on real data. To evaluate the differences in the fake rates, a dedicated D^* sample was analysed [69, 70]. This sample can deliver extremely clean π^+ and K^+ samples. By counting, how many of these real π^+ and K^+ particles pass the electron-ID or muon-ID cut, an estimate of the real fake rates on data can be made.
Events that contained such a fake lepton were reweighted according to the expectations from the above procedure. The correction factors were determined in 8 bins of θ ($0 \ldots 151°$) and 11 bins in the momentum of the particle in the lab frame p_{lab} ($0 \ldots 4.0$ GeV/c).

	exp	lepton signal	tau signal	ratio
eID	7 - 27	0.9863 ± 0.0213	0.9641 ± 0.0321	1.0231
eID	31 - 65	0.9824 ± 0.0165	0.9690 ± 0.0230	1.0138
μID	7 - 27	0.9657 ± 0.0253	0.9480 ± 0.0249	1.0187
μID	31 - 39, 45a	0.9539 ± 0.0197	0.9692 ± 0.0206	0.9842
μID	41 - 49	0.9467 ± 0.0198	0.9669 ± 0.0205	0.9790
μID	51 - 65	0.9768 ± 0.0195	0.9917 ± 0.0207	0.9849

Table 6.10.: Lepton Efficiency corrections for B^+

6.5.5. D Meson Widths

It has been observed by many analyses before, that the resolution of the D meson mass is better on simulated data than on real data. Imperfect modelling of the magnetic field, the number of hits left by a charged particle and other detector responses lead to that imperfection. While it is possible to correct for that effect by smearing of tracks, an alternative approach was chosen. The D meson signal in the $M(D)$ distribution and the D^* signal in the $\Delta M(D^*, D)$ distribution are clearly visible for all decay modes and can thus be fitted on generic Monte Carlo and on data. The results of these fits are summarized in table 6.8.

With the knowledge of the widths on data and Monte Carlo, we can then construct a transformed D mass variable:

$$M(D)_{\text{trafo}} = \frac{M(D) - M(D)_{\text{WA}}}{\sigma_{\text{data,MC}}} \tag{6.17}$$

$$M(D)_{\text{trafo}} = \frac{M(D) - \text{middle of SB}}{\sigma_{\text{data,MC}}} \tag{6.18}$$

Equation 6.17 is valid for the signal region of the D meson mass, while equation 6.18 is valid for the sideband regions. $M(D)_{\text{WA}}$ is the current world average value of the D meson mass, "middle of SB" is the middle of the upper or lower sideband and $\sigma_{\text{data,MC}}$ is the fitted width on Monte Carlo or data respectively.

An analogous transformation was performed for the mass difference $\Delta M(D^*, D)$:

$$\Delta M(D^*, D)_{\text{trafo}} = \frac{\Delta M(D^*, D) - \Delta M(D^*, D)_{\text{WA}}}{\sigma_{\text{data,MC}}} \tag{6.19}$$

$$\Delta M(D^*, D)_{\text{trafo}} = \frac{\Delta M(D^*, D) - \text{middle of SB}}{\sigma_{\text{data,MC}}} \tag{6.20}$$

Again, equation 6.19 is valid for the signal region of the D^*, D meson mass difference, while equation 6.20 is valid for the sideband regions. $\Delta M(D^*, D)_{\text{WA}}$ is the current world average value of the D^*, D meson mass difference, "middle of SB" is the middle of the

upper or lower sideband and $\sigma_{\text{data,MC}}$ is the fitted width on Monte Carlo or data respectively.

These two tranformed variables look the same on data and Monte Carlo. They were therefore better-suited for further usage in the NeuroBayes trainings described in section 6.7.

6.6. The Fitting Variables

The ultimate goal of this analysis is to measure the ratios R and R^*

$$R = \frac{\mathcal{B}(B \to D\tau\nu_\tau)}{\mathcal{B}(B \to D\ell\nu_\ell)} \quad \text{and} \quad R^* = \frac{\mathcal{B}(B \to D^*\tau\nu_\tau)}{\mathcal{B}(B \to D^*\ell\nu_\ell)}\,, \tag{6.21}$$

already defined in more detail in equations 6.1 to 6.4. It is therefore crucial to distinguish the τ signal from the ℓ signal and also to distinguish both the τ and the ℓ signal from various backgrounds. As the τ leptons are reconstructed in the channels

$$\tau^- \to e^-\bar{\nu}_e\nu_\tau \quad \text{and} \tag{6.22}$$

$$\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau\,, \tag{6.23}$$

the decay $B \to D^{(*)}\tau(\to \ell\nu_\ell\nu_\tau)\nu_\tau$ differs from the decay $B \to D^{(*)}\ell\nu_\ell$ only in the number of undetectable neutrinos. The one decay channel is therefore in itself not distinguishable from the other. However, with the knowledge of the 4-momentum of the tag-side meson $p(B_{\text{tag}})$ that is provided by the Full Reconstruction and the 4-momentum of the beam $p(\text{Beam})$, we can calculate the 4-momentum of the missing neutrino system. This allows us to distinguish the lepton signal from the τ signal.

Missing Mass Squared

Several observables can be calculated using the information about the missing momentum of the neutrino system. A very intuitive example is the *Missing Mass Squared* (M^2_{miss}), defined as

$$M^2_{\text{miss}} = \Big[p(\text{Beam}) - \Big(p(B_{\text{tag}}) + p(D^{(*)}) + p(\ell)\Big)\Big]^2\,. \tag{6.24}$$

For the electron and muon signal, there is only one neutrino involved in the decay. As the neutrino has a mass very close to 0, we therefore expect a peaking component around $M^2_{\text{miss}} = 0\ \text{GeV}^2/c^4$. In the τ signal decay, the three neutrino system constitutes a much larger missing mass, and we therefore expect to see a much broader structure at $M^2_{\text{miss}} > 1\ \text{GeV}^2/c^4$[1] This variable is very useful to distinguish the lepton and τ signal. It has, however, only very limited power to distinguish the τ signal from a number of backgrounds. It is therefore desirable to construct a variable, that is better suited to separate the τ signal from backgrounds. An exemplary distribution of $M^2_{\text{miss}}$is shown in figure 6.6. For a more detailed description of the M^2_{miss}-shape of the fit components, see chapter 6.10.

Other Variables and Correlations to M^2_{miss}

It is possible to construct many more variables that can separate τ and ℓ signal from each other. Examples of such variables are

[1] Further detailed explanations can be found in section 6.10.

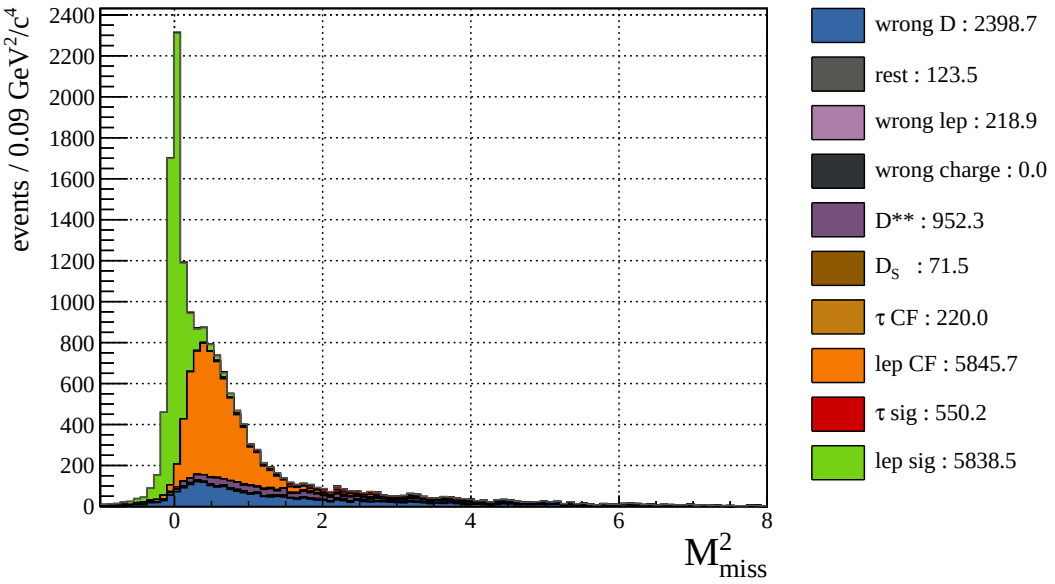


Figure 6.6.: M^2_{miss}distribution for $B^0 \to D^- e^+ \nu_e$. For a detailed description of the components, see section 6.10.

- $q^2 = (p_{\text{miss}} + p_\ell)^2$
- $P_\ell^{\text{in } B_{sig} \text{ frame}}$

Those variables are constructed using information provided by the Full Reconstruction. The primary information that comes from the Full Reconstruction is the momentum of the $B_{\text{tag}}$and a list of final state particles that were used in the reconstruction of the B_{tag}. The powerful variables use exactly this information. While the variables differ in the details of their construction, they all use similar information. It is therefore not surprising that there are large correlations amongst them. A fit in a second dimension to such a variable would therefore bring not much additional information, but larger correlations between the two fitting variables. These variables were therefore not used in the analysis.

An exception is the additional energy in the electromagnetic calorimeter $E^{\text{ECL}}_{\text{extra}}$:

$$E^{\text{ECL}}_{\text{extra}} = \sum E_{\text{Calor.}} - \left(\sum E_{\text{tag}} + \sum E_{\text{signal}}\right) \tag{6.25}$$

The exact implementation is done in the same way as in reference [71]. While M^2_{miss}uses information about the momentum of the neutrino system, $E^{\text{ECL}}_{\text{extra}}$ uses information about the remaining particles in the detector. The correlation between the two variables is therefore small, while the information that they bring is complementary. Using $E^{\text{ECL}}_{\text{extra}}$ as a second dimension in a fit is therefore reasonable. There are, however, more variables that have discriminative power to distinguish signal from various sources of

background. These additional variables do not use Full Reconstruction information and have therefore little correlation to M^2_{miss}. In order to increase the discriminative power that a second dimension would bring compared to a fit only to M^2_{miss}, many variables were combined in a NeuroBayes training. It is described in detail in chapter 6.7.

6.7. NeuroBayes Trainings

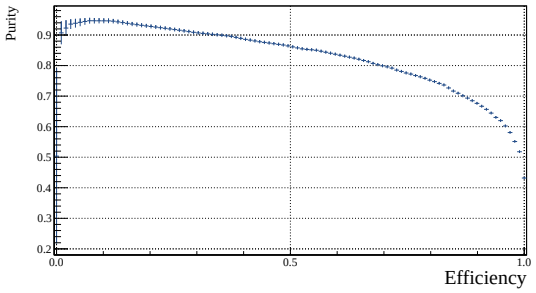
Four separate trainings were performed for the four different channels: $B^0 \to D^- \ell^+ \nu_\ell$, $B^0 \to D^{*-} \ell^+ \nu_\ell$, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ and $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$. The categories for the trainings were the folllowing:

Target $B \to D^{(*)} \tau \nu_\tau$ signal

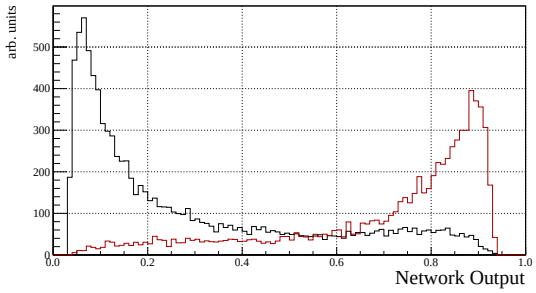
Background Wrong D meson background, wrong D^* meson background, D^{**} background, wrong lepton background, wrong charge cross-feed, D_s background and remaining backgrounds. (see section 6.10)

Not included $B \to D^{(*)} \ell \nu_\ell$ signal, $B \to D^* \ell \nu_\ell$ cross-feed, $B \to D^* \tau \nu_\tau$ cross-feed

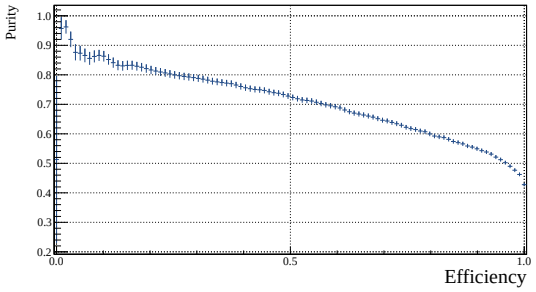
Each of the above trainings contained the same 8 variables. They are listed, together with their significance in table 6.11. The significance is defined as the correlation of the variable to the target multiplied by the square-root of the sample size [72]. Other variables are not taken into account. The network output was recalculated in such a way that it lies within the interval $[0,1]$. The purity efficiency plots and the network outputs of the trainings can be found in figures 6.7 and 6.8.



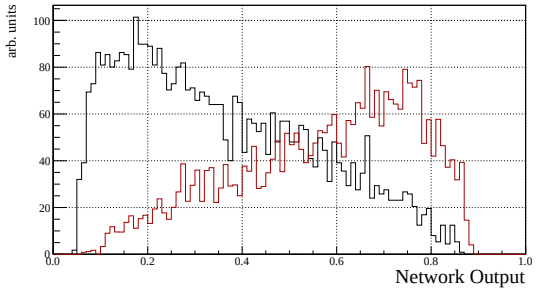
(a) $B^0 \to D^- e^+ \nu_e$



(b) $B^0 \to D^- e^+ \nu_e$

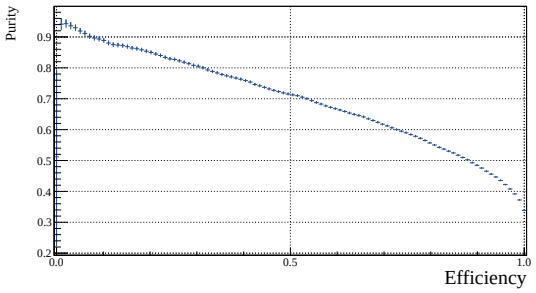


(c) $B^0 \to D^{*-} e^+ \nu_e$

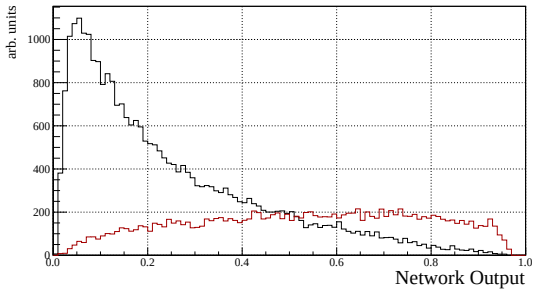


(d) $B^0 \to D^{*-} e^+ \nu_e$

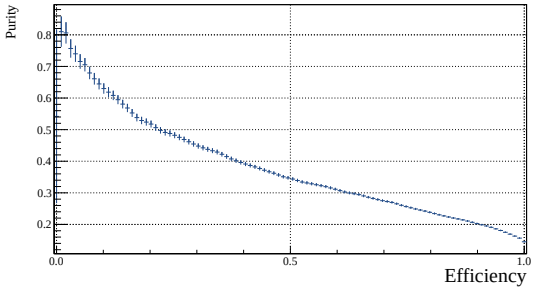
Figure 6.7.: Purity efficiency plots and network output plots for channels $B^0 \to D^- e^+ \nu_e$ and $B^0 \to D^{*-} e^+ \nu_e$.



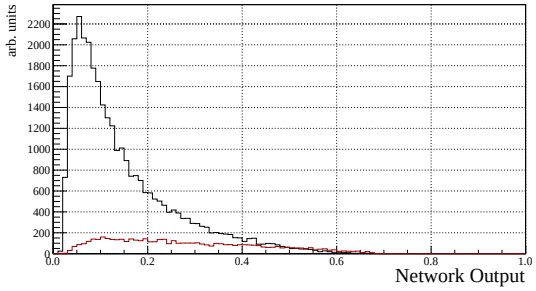
(a) $B^+ \to \overline{D}^0 e^+ \nu_e$



(b) $B^+ \to \overline{D}^0 e^+ \nu_e$



(c) $B^+ \to \overline{D}^{*0} e^+ \nu_e$



(d) $B^+ \to \overline{D}^{*0} e^+ \nu_e$

Figure 6.8.: Purity efficiency plots and network output plots for channels $B^+ \to \overline{D}^0 e^+ \nu_e$ and $B^+ \to \overline{D}^{*0} e^+ \nu_e$.

While it would in principle be possible to use the output of the trainings o_{NB}directly in the fit, there is some benefit in transforming it. The tranformation is given by

$$o_{\mathrm{NB,\,trafo.}} = \log \frac{o_{\mathrm{NB}} - o_{\mathrm{min}}}{o_{\mathrm{max}} - o_{\mathrm{NB}}} \,, \tag{6.26}$$

where o_{min} and o_{max} are the minimum and maximum network output respectively. It has the effect of turning the network output into a smooth distribution, the fitting of which generally poses less technical issues. Although there is no obvious mathematical reason for it, $o_{\mathrm{NB,\,trafo.}}$often shows Gaussian-like shape and can indeed be fitted by a bifurcated Gaussian function. It should be noted, this transformation only provides the smoother shape of the distribution and does not improve the discrimination power of the NeuroBayes trainings.

6.7.1. Correlation

In order to check the correlations between the first fitting variable $M^2_{\mathrm{miss}}$and the output of the NeuroBayes trainings (o_{NB}), a correlation analysis was performed using the correlation analysis tool CAT [73]. The results of this analysis are summarised in table 6.12. The low absolute values of Pearson's [74] and Spearman's [75] correlation coeffcients

(both range from −1 to +1) indicate an only small linear correlation of o_{NB} and M^2_{miss}. There are, however, non-linear forms of correlation that are not quantified by these two types of correlation coefficient. A hypothesis test, described in reference [73] will test the two variables for any type of correlation, linear or non-linear. The significance of the hypothesis of the two variables being correlated is given under the column *Sign* in table 6.12. Especially for the case of no cut in o_{NB}(i.e. $o_{\mathrm{NB}} > 0.0$), there seem to be large correlations. One must, however, keep two important facts about the analyzed sample in mind: A cut on o_{NB} will definitely be performed at a later stage during the analysis, so the large correlation for the completely uncut sample is of no big importance. Secondly, this analysis was performed using 5 streams of generic Monte Carlo. So the sample is 5 times larger than the expected sample on real data. To correctly estimate the expected significance for correlations on real data, one has to divide the significance figures by $\sqrt{5} \approx 2.24$. This leads to realistic cut values in the region $o_{\mathrm{NB}} > 0.1$ or $o_{\mathrm{NB}} > 0.2$[2], depending on the channel.

[2]Reminder: o_{NB}ranges from 0 to 1.

Variable	Sig [σ]			
	D^+	D^{*+}	D^0	D^{*0}
Hash identifying the type and decay channel of the D meson	84.16	33.51	92.22	25.38
Hash identifying the type and decay channel of the D^* meson	0.00	4.17	0.00	34.09
Relative deviation of the D meson mass from the world average value (in sidebands: from middle of sideband), in units of fitted D mass width. See section 6.5.5.	30.18	10.24	42.62	10.39
Relative deviation of the D^*, D mass difference from the world average value (in sidebands: from middle of sideband), in units of fitted D^*, D mass difference width. See section 6.5.5.	0.00	14.01	0.00	24.92
Number of photons from remaining π^0 candidates. These π^0 candidates must lie within the mass region $[0.110, 0.150]\,\mathrm{GeV}/c^2$ and have photon energies $E(\gamma) > 50\,\mathrm{MeV}$, independent of the detector region (compare different requirements in section 6.4)	16.29	8.31	34.23	20.43
Remaining energy in the ECL after removal of the clusters used for the tag side or signal side	52.16	28.14	81.77	52.62
Cosine of the angle between the momentum of the D meson and the line through the D meson decay vertex and the interaction point	8.57	0.35	1.72	3.18
$o_{\mathrm{NB}}$of the tag side B meson	6.17	4.22	15.68	4.58
Hash identifying the type and decay channel of the tag side B meson	18.45	14.64	29.17	24.21

Table 6.11.: Variables in the Signal Side NeuroBayes training

	$o_{\mathrm{NB}} > 0.0$			$o_{\mathrm{NB}} > 0.1$			$o_{\mathrm{NB}} > 0.2$			$o_{\mathrm{NB}} > 0.3$		
Component	Spea	Pear	Sign	Spea	Pear	Sign	Spea	Pear	Sign	Spea	Pear	Sign
$B^0 \to D^- \ell^+ \nu_\ell$ sample												
ℓ signal	-0.10	-0.07	+9.0	-0.08	-0.06	+9.0	-0.06	-0.04	+7.1	-0.07	-0.03	+5.4
τ signal	+0.00	+0.01	+0.9	+0.00	+0.01	+0.9	+0.02	+0.02	+1.0	+0.01	+0.01	+1.1
ℓ cross-feed	-0.07	-0.05	+3.9	-0.08	-0.05	+3.4	-0.06	-0.04	+2.9	-0.06	-0.04	+2.4
τ cross-feed	-0.09	-0.11	+1.6	-0.13	-0.14	+1.9	-0.12	-0.14	+2.1	-0.15	-0.15	+1.8
wrong D meson	-0.12	-0.13	+9.0	-0.06	-0.05	+1.9	-0.04	-0.04	+1.0	-0.03	-0.03	+0.6
wrong ℓ	-0.07	-0.08	+1.2	-0.12	-0.11	+1.0	-0.09	-0.11	+1.3	-0.10	-0.11	+0.9
D_s background	-0.22	-0.13	+0.5	-0.08	-0.06	+1.2	-0.15	-0.08	+0.3	-0.09	-0.04	+0.1
other backgrounds	+0.01	+0.03	+0.9	+0.05	+0.07	+0.5	+0.13	+0.18	+1.0	+0.09	+0.14	+0.4
D mass sidebands	-0.14	-0.14	+9.0	-0.06	-0.06	+3.4	-0.06	-0.06	+3.4	-0.04	-0.04	+2.0
$B^0 \to D^{*-} \ell^+ \nu_\ell$ sample												
ℓ signal	-0.09	-0.05	+3.7	-0.08	-0.04	+3.7	-0.06	-0.04	+3.2	-0.05	-0.04	+3.8
τ signal	-0.07	-0.07	+1.8	-0.07	-0.07	+1.8	-0.08	-0.08	+1.9	-0.06	-0.06	+1.5
wrong D^* meson	-0.30	-0.25	+9.0	-0.22	-0.18	+9.0	-0.13	-0.14	+6.6	-0.09	-0.08	+3.1
wrong ℓ	-0.16	-0.15	+1.5	-0.18	-0.16	+1.4	-0.20	-0.20	+1.6	-0.13	-0.13	+0.9
D_s background	+0.01	+0.06	+0.5	-0.05	+0.06	+0.5	+0.08	+0.07	+0.4	+0.06	+0.04	+0.3
other backgrounds	-0.08	-0.09	+1.6	-0.06	-0.10	+0.9	+0.14	+0.04	+0.5	+0.05	-0.09	+0.7
$\Delta M(D^*, D)$ sidebands	-0.28	-0.29	+9.0	-0.17	-0.20	+7.9	-0.06	-0.11	+3.2	+0.01	-0.08	+1.7
$B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample												
ℓ signal	-0.07	-0.05	+9.0	-0.07	-0.04	+9.0	-0.05	-0.02	+8.2	-0.05	-0.03	+6.4
τ signal	-0.01	-0.01	+0.6	-0.02	-0.01	+1.9	-0.00	+0.00	+1.3	-0.01	-0.00	+0.9
ℓ cross-feed	-0.09	-0.06	+9.0	-0.08	-0.06	+6.9	-0.07	-0.05	+4.5	-0.08	-0.05	+4.5
τ cross-feed	+0.02	+0.02	+1.6	+0.02	+0.02	+1.7	+0.00	+0.01	+2.0	+0.00	+0.01	+1.3
wrong D meson	-0.14	-0.13	+9.0	-0.07	-0.07	+3.2	-0.06	-0.07	+3.5	-0.08	-0.07	+3.0
wrong ℓ	-0.08	-0.10	+4.2	-0.07	-0.09	+3.2	-0.06	-0.07	+1.3	-0.04	-0.05	+0.8
D_s background	-0.18	-0.17	+1.7	-0.19	-0.15	+2.2	-0.12	-0.11	+0.6	-0.15	-0.13	+0.7
other backgrounds	+0.11	+0.09	+9.0	+0.12	+0.12	+9.0	+0.13	+0.15	+7.4	+0.14	+0.16	+6.6
D mass sidebands	-0.15	-0.16	+9.0	-0.11	-0.12	+9.0	-0.10	-0.11	+6.5	-0.10	-0.10	+4.6
$B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ sample												
ℓ signal	-0.06	-0.03	+9.0	-0.05	-0.03	+6.6	-0.04	-0.03	+3.7	-0.03	-0.02	+3.0
τ signal	+0.00	+0.00	+0.2	+0.00	+0.01	+0.6	-0.01	-0.01	+0.4	-0.03	-0.04	+1.0
wrong D^* meson	-0.20	-0.21	+9.0	-0.09	-0.12	+9.0	-0.06	-0.09	+7.7	-0.03	-0.04	+2.8
wrong ℓ	-0.07	-0.07	+1.6	-0.04	-0.07	+1.5	-0.21	-0.21	+2.1	-0.09	-0.11	+1.4
D_s background	-0.02	-0.01	+0.4	+0.03	+0.02	+0.7	+0.02	+0.07	+0.8	-0.05	+0.03	+0.3
other backgrounds	-0.15	-0.18	+1.5	-0.17	-0.24	+2.2	+0.09	+0.01	+0.1	+0.11	+0.19	+0.2
$\Delta M(D^*, D)$ sidebands	-0.19	-0.19	+9.0	-0.07	-0.10	+9.0	-0.04	-0.06	+3.0	-0.04	-0.06	+3.3

Table 6.12.: Correlation coefficients for the correlation of M^2_{miss} vs. o_{NB}. *Spea*: Spearman's correlation coefficient, *Pear*: Pearman's correlation coefficient, *Sign*: Significance of the correlation. The significance is cut off at 9.0 by CAT [73].

6.8. Best Candidate Selection

In this analysis, we deal with 4 signal samples: $B^0 \to D^- \ell^+ \nu_\ell$, $B^0 \to D^{*-} \ell^+ \nu_\ell$, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ and $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$. We can ultimately only accept one candidate per event. It is therefore necessary to choose the best B candidate in an event, for which there are many possible criteria. Three criteria were investigated for this analysis:

1. We choose the best candidate at random.
2. We choose the candidate with the lowest value for $E^{\mathrm{ECL}}_{\mathrm{extra}}$.
3. We define two qualities:

$$M(D)_{\mathrm{trafo}} = \frac{M(D)_{\mathrm{measured}} - M(D)_{\mathrm{WA}}}{\sigma(M(D))_{\mathrm{fitted}}} \tag{6.27}$$

$$\Delta M(D^*D)_{\mathrm{trafo}} = \frac{\Delta M(D^*D)_{\mathrm{measured}} - \Delta M(D^*D)_{\mathrm{WA}}}{\sigma(\Delta M(D^*D))_{\mathrm{fitted}}} \tag{6.28}$$

   We then choose the candidate with the lowest value for $M(D)_{\mathrm{trafo}}$ in $B \to D\ell\nu_\ell$ channels and the candidate with the lowest value for $\Delta M(D^*D)_{\mathrm{trafo}}$ in $B \to D^*\ell\nu_\ell$ channels.

To evaluate the quality of these three criterions, a sample was produced for each selection with special NeuroBayes trainings. These special trainings are based on the trainings described in section 6.7, but they do not contain the variable that was used for the best candidate selection:

- The selection using $M(D)_{\mathrm{trafo}}$ and $\Delta M(D^*D)_{\mathrm{trafo}}$ was performed on a sample, where the two variables were left out of the NeuroBayes training.
- The selection using $E^{\mathrm{ECL}}_{\mathrm{extra}}$was performed on a sample, where this variable was left out of the NeuroBayes training.
- The random selection was performed on a sample that contained all variables listed in table 6.11.

Before the best candidate selection was performed, the cuts on o_{NB}described in section 6.18 were applied. As a reference for comparison, a sample where no best candidate selection and no cut on $o_{\mathrm{NB}}$was performed at all was also produced. Table 6.13 summarises and compares the results of the best candidate selection.
Column "all" shows the maximal achievable number of τ and ℓ signal. No selection was performed there. Columns "$E^{\mathrm{ECL}}_{\mathrm{extra}}$" and "$M(D)$ / ΔM" show the achieved signal yields for the best candidate selection according to the corresponding criterion. Column "random" finally shows the signal yields for the completely random best candidate selection, but with all the information in the NeuroBayes training. As this step in the selection procedure chooses one candidate per event, the number of *candidates* for

	all		$E^{\mathrm{ECL}}_{\mathrm{extra}}$		$M(D)$ / ΔM		random	
Channel	ℓ sig	τ sig	ℓ sig	τ sig	ℓ sig	τ sig	ℓ sig	τ sig
$B^0 \to D^- \ell^+ \nu_\ell$	9857	891	8732	816	8749	800	9000	844
$B^0 \to D^{*-} \ell^+ \nu_\ell$	15754	685	15341	670	15176	658	15239	660
$B^+ \to \overline{D}^0 \ell^+ \nu_\ell$	19258	1716	16155	1428	15400	1368	15871	1418
$B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$	27083	1284	22695	1077	20294	974	17393	816
All Signals added	71952	4576	62923	3991	59619	3800	57503	3738

Table 6.13.: Best candidate selection evaluated on 10 streams of generic Monte Carlo.

the "$E^{\mathrm{ECL}}_{\mathrm{extra}}$", "$M(D)$ / ΔM" and the "random" sample are fixed[3]. Therefore, a larger τ and ℓ signal yield implies a lower background yield.
We can conclude from table 6.13 that each method has its advantages. The random selection performs quite well for most channels (except for $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$). This is, of course, due to the fact that the omission of $E^{\mathrm{ECL}}_{\mathrm{extra}}$or $M(D)$ / ΔM from the NeuroBayes trainings lessens their quality. So the necessary cuts remove more signal in the $E^{\mathrm{ECL}}_{\mathrm{extra}}$and $M(D)$ / ΔM methods.

The random best canidate selection does not use any variables that are even slightly correlated to the NeuroBayes trainings. So, no dditional systematic error is introduced by this method. The random best canidate selection was chosen for the signal extraction.

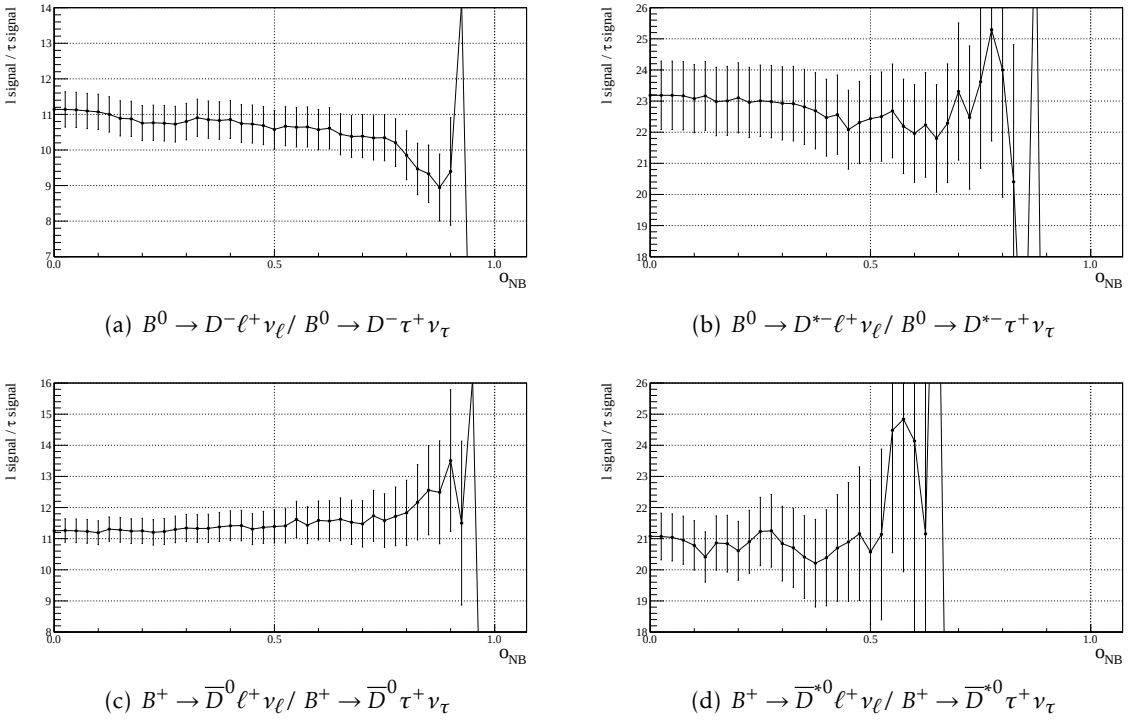
[3]They are slightly different from each other, as the NeuroBayes trainings are not identical

6.9. Cross-Checks for the NeuroBayes Trainings

After the best candidate selection, several cross-checks were performed to ascertain the quality of the NeuroBayes trainings.

6.9.1. Signal Efficiencies

As a first cross-check, the ratio of $B \to D^{(*)}\ell\nu_\ell$ signal to $B \to D^{(*)}\tau\nu_\tau$ signal, depending on the cut on the NeuroBayes expertise, was studied using Monte Carlo information. Let us recall that those two signal components occupy quite different regions of M^2_{miss}. When plotting the ratio $B \to D^{(*)}\ell\nu_\ell / B \to D^{(*)}\tau\nu_\tau$, against o_{NB}, a small slope would indicate a low correlation of o_{NB} to $M^2_{\text{miss}}$and the inability of the NeuroBayes expert to distinguish between $B \to D^{(*)}\ell\nu_\ell$ signal and $B \to D^{(*)}\tau\nu_\tau$ signal. The resulting plots are shown in figure 6.9. From figure 6.9 we can conclude that there is no strong correlation



(a) $B^0 \to D^-\ell^+\nu_\ell / B^0 \to D^-\tau^+\nu_\tau$

(b) $B^0 \to D^{*-}\ell^+\nu_\ell / B^0 \to D^{*-}\tau^+\nu_\tau$

(c) $B^+ \to \overline{D}^0\ell^+\nu_\ell / B^+ \to \overline{D}^0\tau^+\nu_\tau$

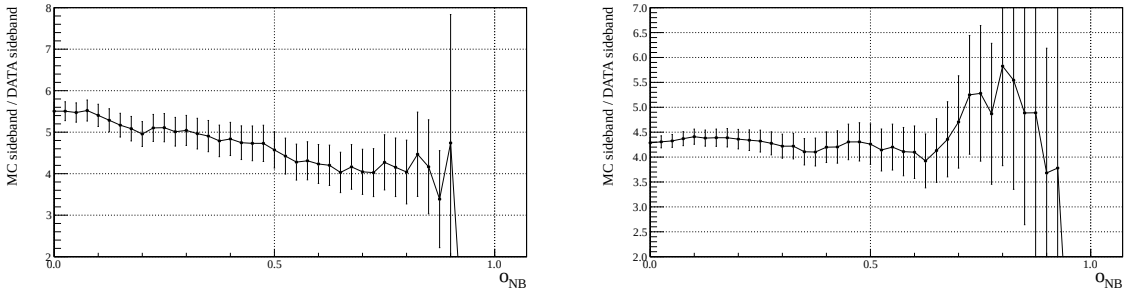
(d) $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell / B^+ \to \overline{D}^{*0}\tau^+\nu_\tau$

Figure 6.9.: Ratios of $B \to D^{(*)}\ell\nu_\ell$ signal to $B \to D^{(*)}\tau\nu_\tau$ signal. When either $B \to D^{(*)}\ell\nu_\ell$ signal or $B \to D^{(*)}\tau\nu_\tau$ signal was 0, then the ratio was set to 0.

between $M^2_{\text{miss}}$and o_{NB}.

6.9.2. o_{NB} in D mass sideband

As a cross-check involving real data, the number of events in the D mass sidebands (see section 6.4.1) was determined in the generic Monte Carlo simulation and on real data for different cuts on o_{NB}. A ratio of "wrong D in MC-sideband" to "wrong D in data-sideband" was formed. The resulting plots are shown in figure 6.10. The NeuroBayes



(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$

Figure 6.10.: Ratios of "wrong D in MC-sideband" to "wrong D in data-sideband". When either the "wrong D in MC-sideband" component or the "wrong D in data-sideband" component was 0, then the ratio was set to 0.

expert has similar effects on data and Monte Carlo, visible as *small* slopes (close to 0) in figure 6.10.

6.9.3. o_{NB} in $\Delta M(D^*, D)$ sideband

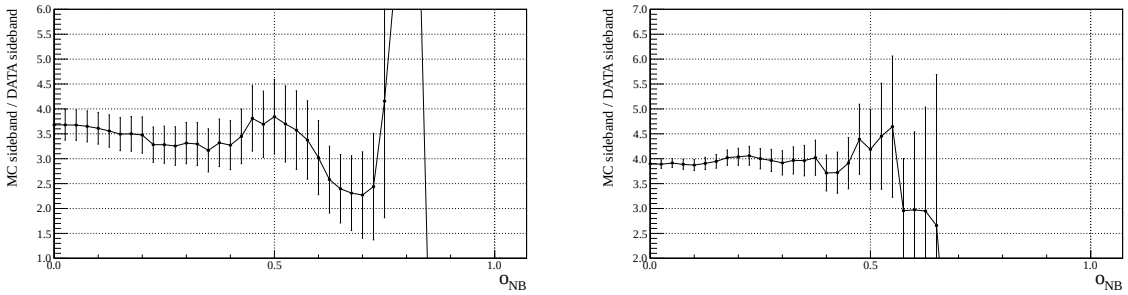
As a similar cross-check as above, the number of events in the $\Delta M(D^*, D)$ sidebands (see section 6.4.2) was determined in the generic Monte Carlo simulation and on real data for different cuts on o_{NB}. Again, the ratio of "wrong D in MC-sideband" to "wrong D in data-sideband" was formed. The resulting plots are shown in figure 6.11. Again, we can observe similar effects of NeuroBayes expert on data and Monte Carlo.

6.9.4. Figure Of Merit

In order to get a feeling for different possible cut values for o_{NB}, a figure of merit was calculated. It is defined as

$$\mathrm{FOM} = \frac{\#\text{ of } \tau \text{ signal events}}{\sqrt{\#\text{ of } \tau \text{ signal events} + \#\text{ of background events}}} \tag{6.29}$$

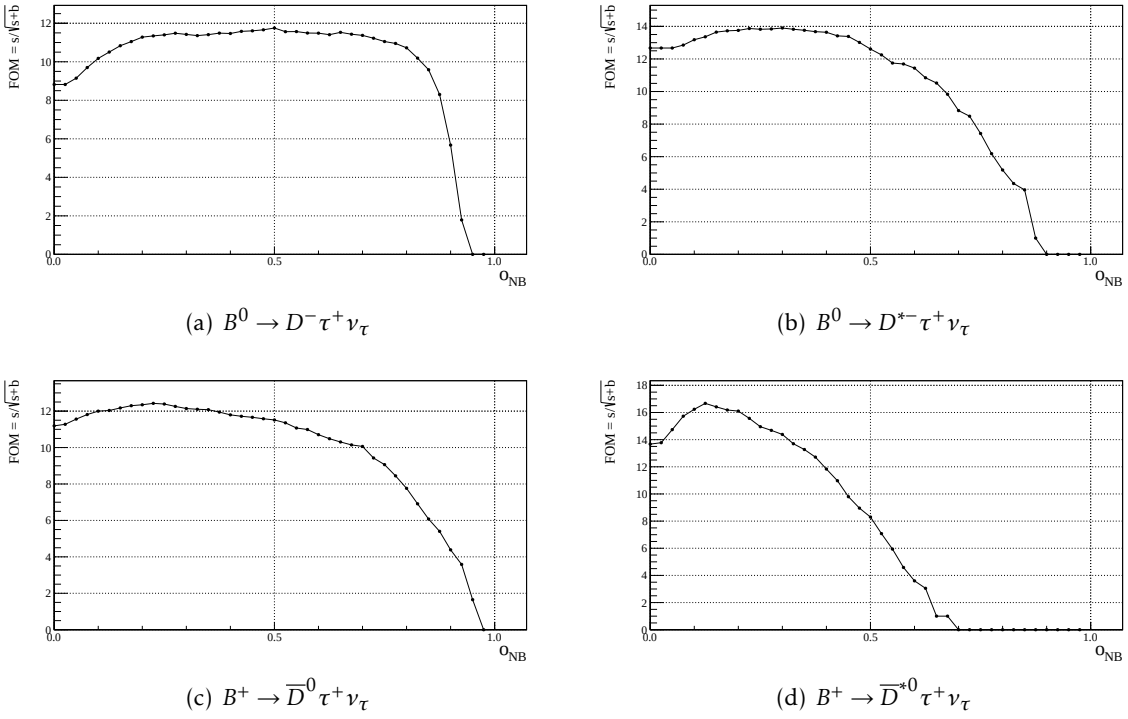
The calculation was performed on 5 streams of generic Monte Carlo in the region $M^2_{\mathrm{miss}} > 1.5\,\mathrm{GeV}^2/c^4$. The results are shown in figure 6.12. These results should not be used to determine the final NeuroBayes cut, as this check lacks information about the 2-dimensional fitting including M^2_{miss}. They indicate, however, that for most networks the exact choice of the cut value is not critical. This is evident by the plateau



(a) $B^0 \to D^{*-}\ell^+\nu_\ell$ (b) $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$

Figure 6.11.: Ratios of "wrong D in MC-sideband" to "wrong D in data-sideband". When either the "wrong D in MC-sideband"component or the "wrong D in data-sideband" component was 0, then the ratio was set to 0.

that the figure of merit shows. Only in the channel $B^+ \to \overline{D}^{*0}\tau^+\nu_\tau$ it seems important to select the cut value with special care.



(a) $B^0 \to D^- \tau^+ \nu_\tau$

(b) $B^0 \to D^{*-} \tau^+ \nu_\tau$

(c) $B^+ \to \overline{D}^0 \tau^+ \nu_\tau$

(d) $B^+ \to \overline{D}^{*0} \tau^+ \nu_\tau$

Figure 6.12.: Figure of merit calculation.

6.10. Composition

Using large amounts of simulated data described in section 6.2, the expected composition of the selected sample was analysed. As the available streams of complete generic Monte Carlo was used, the signal decays $B \to D^{(*)}\tau\nu_\tau$ are also included. The analysis cuts listed in tables 6.5, 6.6 and 6.7 were applied. A very instructive variable that can be constructed is the missing mass squared M^2_{miss} in an event. It can only be calculated with a tag side. It is defined as

$$M^2_{\text{miss}} = \left[p(\text{Beam}) - \left(p(B_{\text{tag}}) + p(D^{(*)}) + p(\ell)\right)\right]^2 . \tag{6.30}$$

and is described in more detail in section 6.6. For improved clarity on the plots, the sample was divided into two regions: The "low M^2_{miss} region" ($M^2_{\text{miss}} < 1.5\ \text{GeV}^2/c^4$) contains a very dominant and narrow component, which would, in a common plot, overshadow the small, but important components in the "high M^2_{miss} region" ($M^2_{\text{miss}} > 1.5\ \text{GeV}^2/c^4$). The different components in the sample are described in the following.

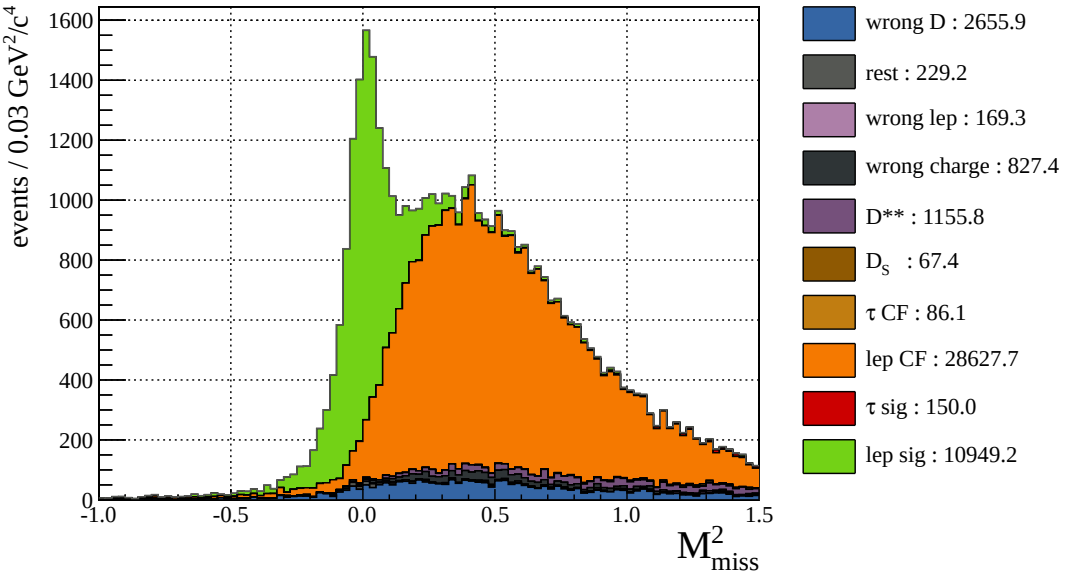


Figure 6.13.: The composition of the analysis sample of the decay $B^+ \to \overline{D}^0 e^+ \nu_e$, using five times the expected amount of data. Shown is the M^2_{miss} distribution for $M^2_{\text{miss}} < 1.5\ \text{GeV}^2/c^4$.

Lepton signal One of the most prominent components is the lepton signal component. It comprises a correctly reconstructed D or D^* meson, as well as a correctly matched light lepton (e or μ) originating from the B decay. They both form a B meson with no missing particles apart from one neutrino. This component can be seen as a narrowly peaking structure around $M^2_{\text{miss}} = 0.0\ \text{GeV}^2/c^4$ in the M^2_{miss}

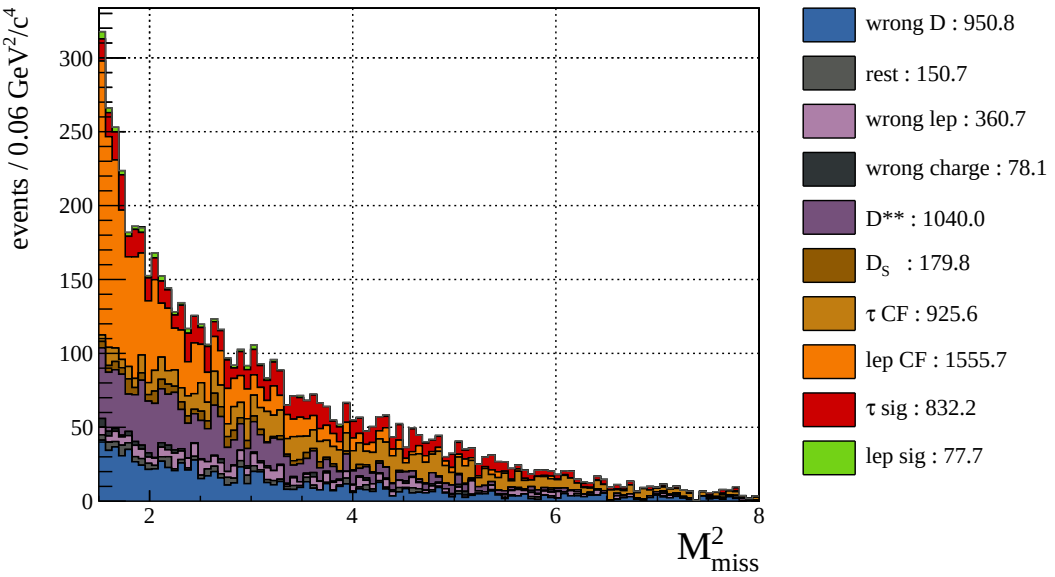


Figure 6.14.: The composition of the analysis sample of the decay $B^+ \to \overline{D}^0 e^+ \nu_e$, using five times the expected amount of data. Shown is the M^2_{miss} distribution for $M^2_{\text{miss}} > 1.5\ \text{GeV}^2/c^4$.

spectrum. It peaks in the M_{bc} spectrum of the B_{tag} meson, in the D meson mass spectrum and, where applicable, in the $\Delta M(D, D^*)$ spectrum. In all plots within this thesis it is coloured green.

τ signal This is the far less prominent of the two signal components. A correctly reconstructed D or D^* meson is required, as well as a correct lepton originating from the τ decay. They both form a B meson with no missing particles apart from three neutrinos. This component can be seen as a broad structure in the M^2_{miss} spectrum above $M^2_{\text{miss}} = 1\ \text{GeV}^2/c^4$. It also peaks in the M_{bc} spectrum of the B_{tag} meson, in the D meson mass spectrum and, where applicable, in the $\Delta M(D, D^*)$ spectrum. In all plots within this thesis it is coloured red.

Lepton cross feed It frequently happens that in reality a $B \to D^* \ell \nu_\ell$ decay occurred, but the decay $B \to D \ell \nu_\ell$ was reconstructed, missing a slow pion or gamma from the D^* decay. By virtue of the Full Reconstruction this missing 4-momentum can be measured, just as the missing 4-momentum of the neutrino system. Slow pions or gammas are relatively difficult to reconstruct directly, but the Full Reconstruction yields an indirect measurement. It can therefore happen that the cross feed component in the channel $B \to D \ell \nu_\ell$ is larger than the signal component in $B \to D^* \ell \nu_\ell$. The cross feed component resembles real $B \to D \ell \nu_\ell$ signal in many ways: A correct D meson is combined with a correctly identified lepton to form a $B \to D \ell \nu_\ell$ candidate. This component therefore peaks at the correct D mass and

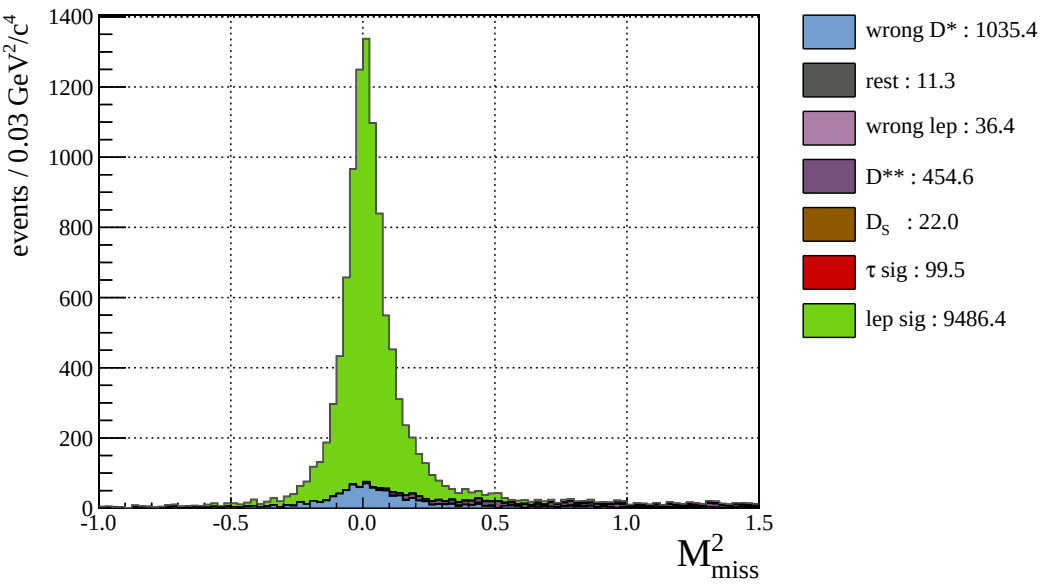


Figure 6.15.: The composition of the analysis sample of the decay $B^0 \to D^{*-} e^+ \nu_e$, using five times the expected amount of data. Shown is the M^2_{miss} distribution for $M^2_{\text{miss}} < 1.5\ \text{GeV}^2/c^4$.

B_{tag} M_{bc} values. It also shows a peaking structure in the M^2_{miss} spectrum, but with a much broader peak than the lepton signal and shifted towards higher values of M^2_{miss}. With the explanation above, it is clear that this component can only occur in the $B \to D\ell\nu_\ell$ samples. In all plots within this thesis it is coloured orange.

Wrong charge cross feed It can also easily happen, that in a $B^0 \to D^{*-}\ell^+\nu_\ell$ decay with $D^{*+} \to D^0\pi^+$ the charged pion is missed. It peaks in the same region as the "Lepton cross feed", described in the previously. This component only occurs as a small background contribution in the $B^+ \to \overline{D}^0\ell^+\nu_\ell$ sample and not at all in the $B^0 \to D^-\ell^+\nu_\ell$ sample. In all plots within this thesis it is coloured black.

τ cross feed In a very similar way to the lepton cross feed, an actual $B \to D^*\tau\nu_\tau$ decay can be misreconstructed as the decay $B \to D\tau\nu_\tau$, missing a slow pion or gamma from the D^* decay. As it was already discussed for the lepton cross feed, this component peaks at the correct D mass and B_{tag} M_{bc} values. In the M^2_{miss} spectrum, the relative shift and widening of the cross feed component with respect to the already very broad $B \to D\tau\nu_\tau$ signal structure is not noticeable. It can also only occur in the $B \to D\ell\nu_\ell$ samples. In all plots within this thesis it is coloured light brown.

Wrong D meson The most common misreconstruction is that of the D meson. This can occur when one or more tag side tracks or clusters are used for the signal side D meson reconstruction. Also, the tracks or electromagnetic clusters belonging

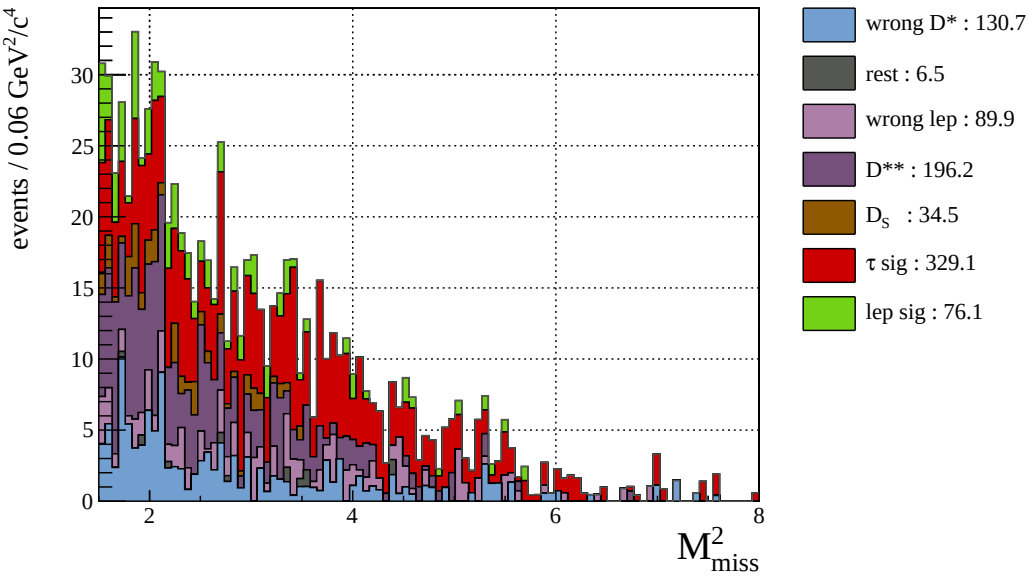


Figure 6.16.: The composition of the analysis sample of the decay $B^0 \to D^{*-} e^+ \nu_e$, using five times the expected amount of data. Shown is the M^2_{miss} distribution for $M^2_{\text{miss}} > 1.5\ \text{GeV}^2/c^4$.

to the D meson could simply be missed by the detector or detector noise could wrongly be interpreted as a track or cluster. Finally, actual signal side lepton tracks may have been wrongly used in the D meson reconstruction. As this background originates from events with correct as well as from events with wrong B_{tag} , there is a peaking and a non-peaking component in the B_{tag} M_{bc} spectrum. The wrong D component is flat in the D mass spectrum and shows up as a very broad structure in the entire M^2_{miss} spectrum In all plots within this thesis it is coloured blue.

Wrong D^* meson For the $B \to D^* \ell \nu_\ell$ and $B \to D^* \tau \nu_\tau$ samples, there is another major source of misreconstruction. Even with a correctly reconstructed D meson, the reconstruction of the D^* meson can still go wrong. Again, usage of one or more final state particles from the tag side or detector effects are conceivable causes. A non-peaking component in the $\Delta M(D, D^*)$ spectrum is caused by this background and it shows a very broad structure in the entire M^2_{miss} spectrum , similar to the wrong D component. By definition, this component also includes wrongly reconstructed D mesons. In the B_{tag} M_{bc} distribution there is a peaking and a non peaking part, as explained above. In all plots within this thesis this component is coloured light blue.

D^{**} component It is not only possible to miss a slow pion or gamma from a D^* decay, but also from higher excited states of the D meson. The states $D_0^*(2400)^0$,

$D_1^*(2420)^0$, $D_2^*(2460)^0$ and $D_2^*(2460)^\pm$ are summarized as D^{**} mesons and can pollute the $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ samples in a similar way as the lepton cross-feed. The lepton and τ components of this background have practically indistingushable shapes. Therefore they are treated in the same way. With correct D and D^* mesons (by definition), this component peaks in the $M(D)$ and in the $\Delta M(D, D^*)$ spectra. As it differs from a real signal decay only by the missing 4-momentum of the slow pions, the D^{**} component peaks in the M_{bc} distribution of the B_{tag}. With a large system of missing particles, it is visible as a very broad structure in the entire M^2_{miss} spectrum above $M^2_{\text{miss}} > 0.0\,\text{GeV}^2/c^4$. In all plots within this thesis this component is coloured purple.

Wrong lepton In the reconstruction of the decay $B \to D\ell\nu_\ell$, not only the D meson can contribute to the background, but also the lepton. The lepton candidate can be a wrongly identified K or π, either from the signal side decay $B \to DK/\pi$ or from the tag side. The case of a charged track taken from the signal side D meson would result in a wrong D meson and is already covered in the category "Wrong D meson". The argument also holds for D^* mesons. The wrong lepton component is visible as a very broad structure in the entire M^2_{miss} spectrum above $M^2_{\text{miss}} > 0.0\,\text{GeV}^2/c^4$ and it peaks in the $M(D)$ and in the $\Delta M(D, D^*)$ spectra. In all plots within this thesis this component is coloured light purple.

D_s decay The helicity suppressed decay $B \to D^{(*)}D_s$ with $D_s \to \tau\nu$ has the very same particles in the final state as the τ signal decay and can therefore fake a τ signal decay. It also shows very similar structures as the actual τ signal decay. For the decays of D_s to lighter leptons, the helicity suppression is stronger and their contribution is so small that it was not found in the Monte Carlo samples. In all plots within this thesis this component is coloured dark brown.

Remaining backgrounds The remaining background events were so few, that further investigation as to their origin seems pointless. No clear shape could be established due to lack of statistics. In all plots within this thesis this component is coloured dark grey.

6.11. Normalisation Check

For the development of the fitting procedure to extract R (see equations 6.1 to 6.4), a preliminary test was performed. The τ signal region in the M^2_{miss}spectrum was blinded by requiring $M^2_{\text{miss}} < 1.5\,\text{GeV}^2/c^4$. On this blinded sample, a simultaneous 2-dimensional fit to $M^2_{\text{miss}}$and $o_{\text{NB, trafo.}}$on the $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ samples was performed. The aim was to extract the ratios

$$X^0 = X(B^0) \equiv \frac{\mathcal{B}(B^0 \to D^{*-}\ell^+\nu_\ell)}{\mathcal{B}(B^0 \to D^-\ell^+\nu_\ell)} \quad \text{and} \tag{6.31}$$

$$X^+ = X(B^+) \equiv \frac{\mathcal{B}(B^+ \to \overline{D}^{*0}\ell^+\nu_\ell)}{\mathcal{B}(B^+ \to \overline{D}^{0}\ell^+\nu_\ell)}\,. \tag{6.32}$$

The individual components of the fits are described in the following.

6.11.1. The Fit Components

For each component that could be identified in the $B \to D\ell\nu_\ell$ and the $B \to D^*\ell\nu_\ell$ samples (see section 6.10) a PDF was determined and included in the fit. The shapes of all components in M^2_{miss}were determined on generic Monte Carlo samples as smoothed kernel estimation functions [76]. The shapes in the $o_{\text{NB, trafo.}}$dimension were parametrised by bifurcated Gaussian PDFs. With 5 streams of generic Monte Carlo available, four streams were used for determining the shapes of the components and the remaining stream was used for the actual fitting procedure. The following components are included in the fit:

Lepton signal This is the main signal component. The normalisation is a free parameter of the fit.

Lepton cross-feed This is a major source of background in the $B \to D\ell\nu_\ell$ sample. The normalisation is a free parameter of the fit.

τ signal The contribution of τ signal in the region $M^2_{\text{miss}} < 1.5\,\text{GeV}^2/c^4$ is negligible. The normalisation was fixed to standard model expectations of the generic Monte Carlo.

τ cross-feed The contribution of τ cross-feed in the region $M^2_{\text{miss}} < 1.5\,\text{GeV}^2/c^4$ is negligible. The normalisation was fixed to standard model expectations of the generic Monte Carlo.

wrong D background This background component can be easily separated from the signal component by the fit to $o_{\text{NB, trafo.}}$. The normalisation is a free parameter of the fit.

wrong D^* background Similar to the wrong D background, this component is quite well separable from the signal using $o_{\text{NB, trafo.}}$. Therefore, its yield is also a free parameter of the fit.

D^{**} background The normalisation of this background component was determined before the actual fit. An additional fit was performed to the D^{**} enriched sample. The following components (see section 6.12.2) were distinguished in this additional fit:

Signal The normalisation is a free parameter of the additional fit.

Random π^0 The normalisation is a free parameter of the additional fit.[4]

Combinatorial background The normalisation is a free parameter of the additional fit.

On the 4 streams of generic Monte Carlo (subscript 4), a factor g_4 was determined using Monte Carlo information:

$$g_4 = \frac{D^{**}\text{ bg yield}_4}{\text{signal yield in } D^{**}\text{ enriched sample}_4} \tag{6.33}$$

Before the actual fit on the remaining stream started, the yield of the D^{**} background was determined by the additional fit to the remaining stream (subscript 1). Thus, an estimate for the D^{**} background in the actual fit to the remaining stream was obtained:

$$D^{**}\text{ bg yield}_1 = g_4 \cdot \text{signal yield in } D^{**}\text{ enriched sample}_1 \,. \tag{6.34}$$

The normalisation of this component was fixed in the fit.

D_S background and wrong lepton background These background components have much smaller yields than the statistical uncertainty of the lepton signal or the lepton cross-feed. They have therefore little effect on the result. The yields were fixed to the standard model expectations of the generic Monte Carlo.

remaining backgrounds This background component has a much smaller yield than the statistical uncertainty of the lepton signal or the lepton cross-feed. It has therefore little effect on the result. The normalisation is a free parameter of the fit.

6.11.2. Intermediate Branching Fractions

In order to immediately get a proper estimate on the fit precision, it is beneficial to actually fit the variable X, rather than the $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ signal yields separately. Technically, one of the two signal yields is expressed in terms of the other signal yield, the relative efficiency and X. An interesting issue involving intermediate branching fractions occurs when fitting $X(B^0)$: For the decays $B^0 \to D^-\ell^+\nu_\ell$ and $B^0 \to D^{*-}\ell^+\nu_\ell$ the ratio X^0 is calculated as follows:

$$X^0 = \frac{\mathcal{B}(B^0 \to D^{*-}\ell^+\nu_\ell)}{\mathcal{B}(B^0 \to D^-\ell^+\nu_\ell)} \,. \tag{6.35}$$

[4] The two components *Random π^0 to signal* and *Random π^0 to CF* (see section 6.12.2) were not distinguished in this fit.

The number of generated and reconstructed events are connected by the efficiency

$$\text{Rec} = \text{Gen} \cdot \epsilon \ . \tag{6.36}$$

We can express the number of reconstructed $B^0 \to D^- \ell^+ \nu_\ell$ events as

$$\text{Rec}_{B^0 \to D^- \ell^+ \nu_\ell} = N_{B\bar{B}} \cdot \mathcal{B}(B^0 \to D^- \ell^+ \nu_\ell) \cdot \epsilon_{B^0 \to D^- \ell^+ \nu_\ell} \cdot \sum_{FS} \mathcal{B}(D^+ \to FS) \epsilon_{D^+ \to FS} \tag{6.37}$$

and the number of generated $B^0 \to D^{*-} \ell^+ \nu_\ell$ events as

$$\text{Rec}_{B^0 \to D^{*-} \ell^+ \nu_\ell} = N_{B\bar{B}} \cdot \mathcal{B}(B^0 \to D^{*-} \ell^+ \nu_\ell) \cdot \epsilon_{B^0 \to D^{*-} \ell^+ \nu_\ell} \cdot (A + B) \ , \qquad \text{where} \tag{6.38}$$

$$A = \mathcal{B}(D^{*+} \to D^0 \pi^+) \cdot \sum_{FS'} \left[\mathcal{B}(D^0 \to FS') \epsilon_{D^0 \to FS'} \right] \tag{6.39}$$

$$B = \mathcal{B}(D^{*+} \to D^+ \pi^0) \cdot \sum_{FS''} \left[\mathcal{B}(D^+ \to FS'') \epsilon_{D^+ \to FS''} \right] \ . \tag{6.40}$$

Solving equations 6.37 and 6.38 for the branching ratios $\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu_\ell)$ and $\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu_\ell)$, we can rewrite equation 6.35 using 6.36 as follows:

$$X^0 = \frac{\mathcal{B}(B^0 \to D^{*-} \ell^+ \nu_\ell) \cdot \epsilon_{B^0 \to D^{*-} \ell^+ \nu_\ell}}{\mathcal{B}(B^0 \to D^- \ell^+ \nu_\ell) \cdot \epsilon_{B^0 \to D^- \ell^+ \nu_\ell}} \cdot \frac{\sum_{FS} \mathcal{B}(D^+ \to FS) \epsilon_{D^+ \to FS}}{A(\text{eq } 6.39) + B(\text{eq } 6.40)} \tag{6.41}$$

For clarity, let us limit ourselves for the moment to the factors in the second term, which relevant for this argument:

$$X^0 \sim \frac{\sum_{FS} \mathcal{B}(D^+ \to FS) \epsilon_{D^+ \to FS}}{\sum_{FS'} [\mathcal{B}(D^0 \to FS') \epsilon_{D^0 \to FS'}] + \sum_{FS''} [\mathcal{B}(D^+ \to FS'') \epsilon_{D^+ \to FS''}]} \tag{6.42}$$

The D^+ and D^0 branching fractions in equation 6.42 differ on the generic Monte Carlo from the current world averages. These values enter in the calculation of X. Thus, different values for these branching fractions had to be used in the following fits, depending on whether Monte Carlo or real data was analysed.
When we consider the decays $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ and $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$, equation 6.42 simplifies to

$$X^+ \sim \frac{\sum_{FS} \mathcal{B}(D^0 \to FS) \epsilon_{D^0 \to FS}}{\sum_{FS'} \mathcal{B}(D^0 \to FS') \epsilon_{D^0 \to FS'}} \ . \tag{6.43}$$

This is due to the fact, that D^{*0} mesons only decay into D^0 mesons. The effect of different D meson branching ratios on generic Monte Carlo and on data thus cancels in the case of $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ and $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ decays.

Fit on	Shapes and ratios determined on
Generic Monte Carlo, stream 0	Generic Monte Carlo, streams 1,2,3,4
Generic Monte Carlo, stream 1	Generic Monte Carlo, streams 0,2,3,4
Generic Monte Carlo, stream 2	Generic Monte Carlo, streams 0,1,3,4
Generic Monte Carlo, stream 3	Generic Monte Carlo, streams 0,1,2,4
Generic Monte Carlo, stream 4	Generic Monte Carlo, streams 0,1,2,3
Data	Generic Monte Carlo, streams 0,1,2,3,4

Table 6.14.: Samples used for the fits in the normalisation check

6.11.3. Results

To extract the ratios X^0 and X^+, 2-dimensional unbinned maximum likelihood fits were performed. A detailed description of the components can be found in section 6.11.1. A listing of the data samples that were used for the fit and the determination of the shapes can be found in table 6.14. The fits were performed for no cut on o_{NB} and for realistic cut values for the extraction of R and R^*: $o_{\mathrm{NB}}(X^{0/+}) > 0.2$. The results of the fits are summarised in a graphical form in figures 6.17 and 6.18. The expected values for the $B \to D\ell\nu_\ell$ and $B \to D^*\ell\nu_\ell$ signals are indicated by the dotted line with the uncertainty in yellow. The expectation and the uncertainty were determined on 5 streams of generic Monte Carlo. It should be noted at this point that the error bars on the fit results include statistical uncertainties only. The extracted values are therefore easily compatible with the expectation. For the fit to real data, the Monte Carlo Expectation in known to be incorrect. Assuming isospin invariance, the current expectation [77] (see also [78, 79, 80, 81]) is

$$X^{0/+}_{\mathrm{expect.}} = 2.32 \pm 0.11 \; . \tag{6.44}$$

The fits to data for $o_{\mathrm{NB}}(X) > 0.0$ and $o_{\mathrm{NB}}(X^*) > 0.0$ yielded:

$$X^0 = 2.54 \pm 0.09 \quad \text{and} \quad X^+ = 2.21 \pm 0.07 \; . \tag{6.45}$$

The fits to data for $o_{\mathrm{NB}}(X) > 0.2$ and $o_{\mathrm{NB}}(X^*) > 0.2$ yielded:

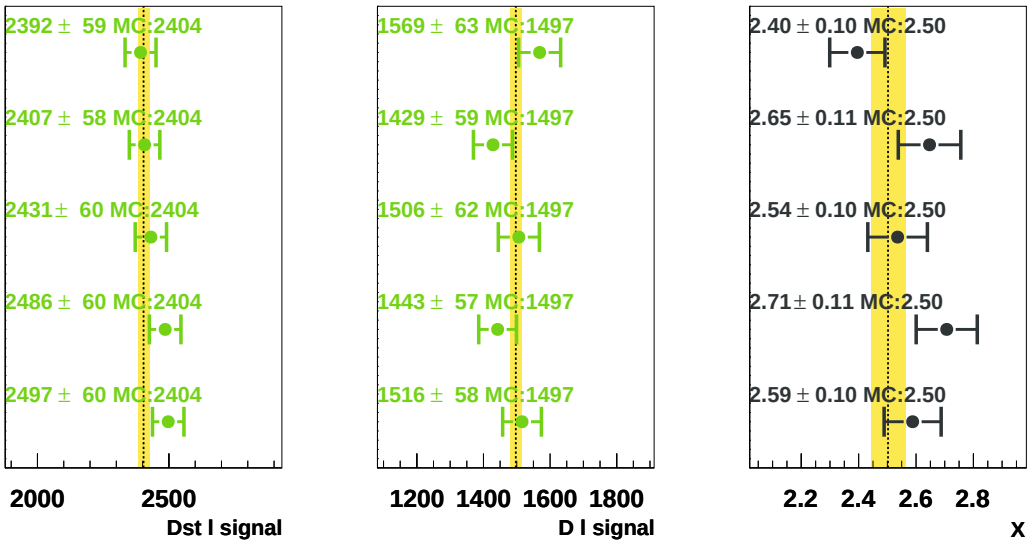
$$X^0 = 2.49 \pm 0.12 \quad \text{and} \quad X^+ = 2.09 \pm 0.09 \; . \tag{6.46}$$

All results treated separately are compatible with the expectation. The combinations of both values for X are

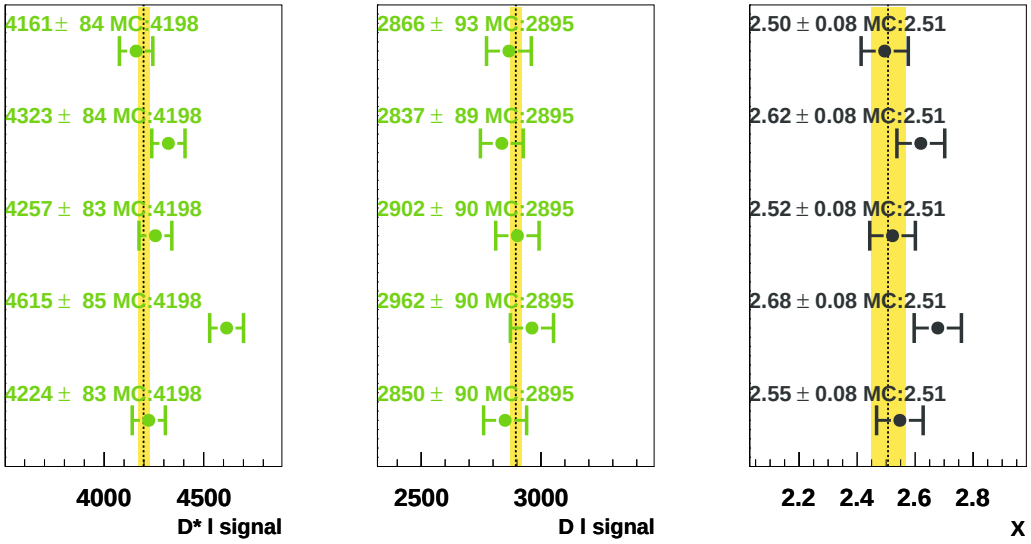
$$X_{\mathrm{comb.},1} = 2.334 \pm 0.055 \qquad \text{for } o_{\mathrm{NB}}(X^{0/+}) > 0.0 \tag{6.47}$$

$$X_{\mathrm{comb.},2} = 2.234 \pm 0.072 \qquad \text{for } o_{\mathrm{NB}}(X^{0/+}) > 0.2 \; . \tag{6.48}$$

Both of these values are also well compatible with the expectation. Projections of the fits can be found in figures 6.19 to 6.22.

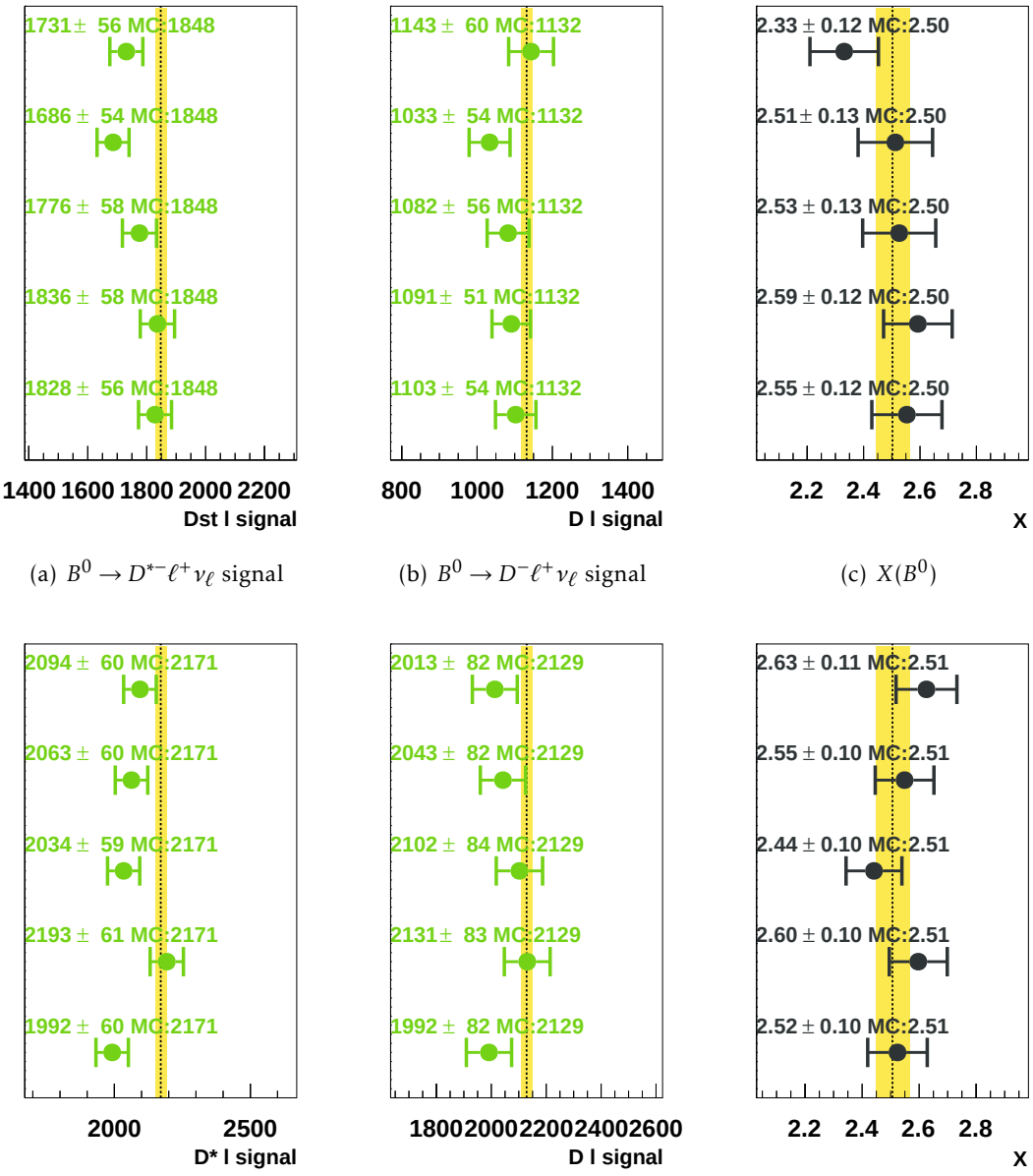


(a) $B^0 \to D^{*-}\ell^+\nu_\ell$ signal (b) $B^0 \to D^-\ell^+\nu_\ell$ signal (c) $X(B^0)$



(d) $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ signal (e) $B^+ \to \overline{D}^0\ell^+\nu_\ell$ signal (f) $X(B^+)$

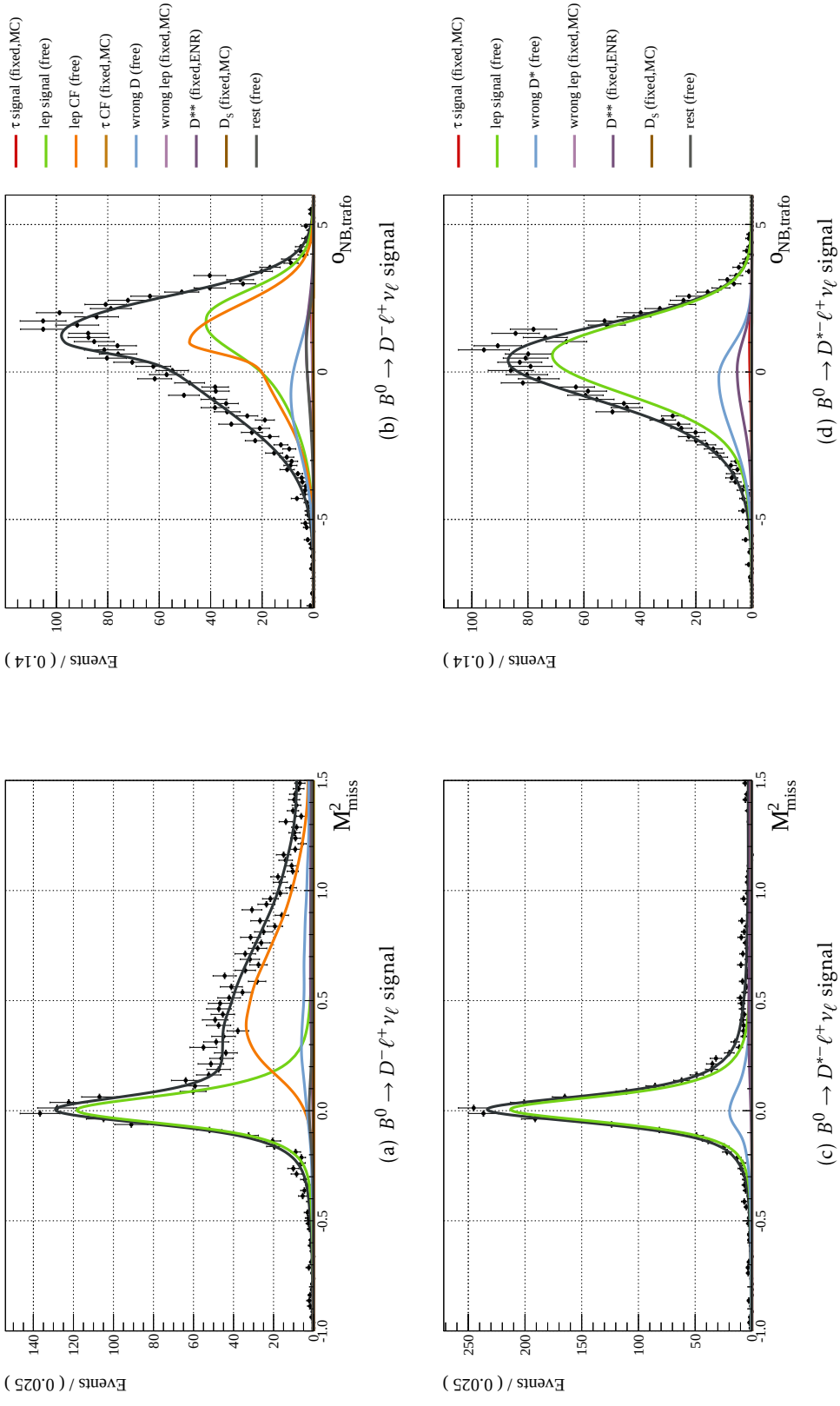
Figure 6.17.: Fit results of the normalisation check for $o_{\mathrm{NB}}(X) > 0.0$ and $o_{\mathrm{NB}}(X^*) > 0.0$. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4. The expectation for X on generic Monte Carlo is $X = 2.50$. The known uncertainty on data was used for the Monte Carlo expectation.



(a) $B^0 \to D^{*-}\ell^+\nu_\ell$ signal (b) $B^0 \to D^-\ell^+\nu_\ell$ signal (c) $X(B^0)$

(d) $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ signal (e) $B^+ \to \overline{D}^0\ell^+\nu_\ell$ signal (f) $X(B^+)$

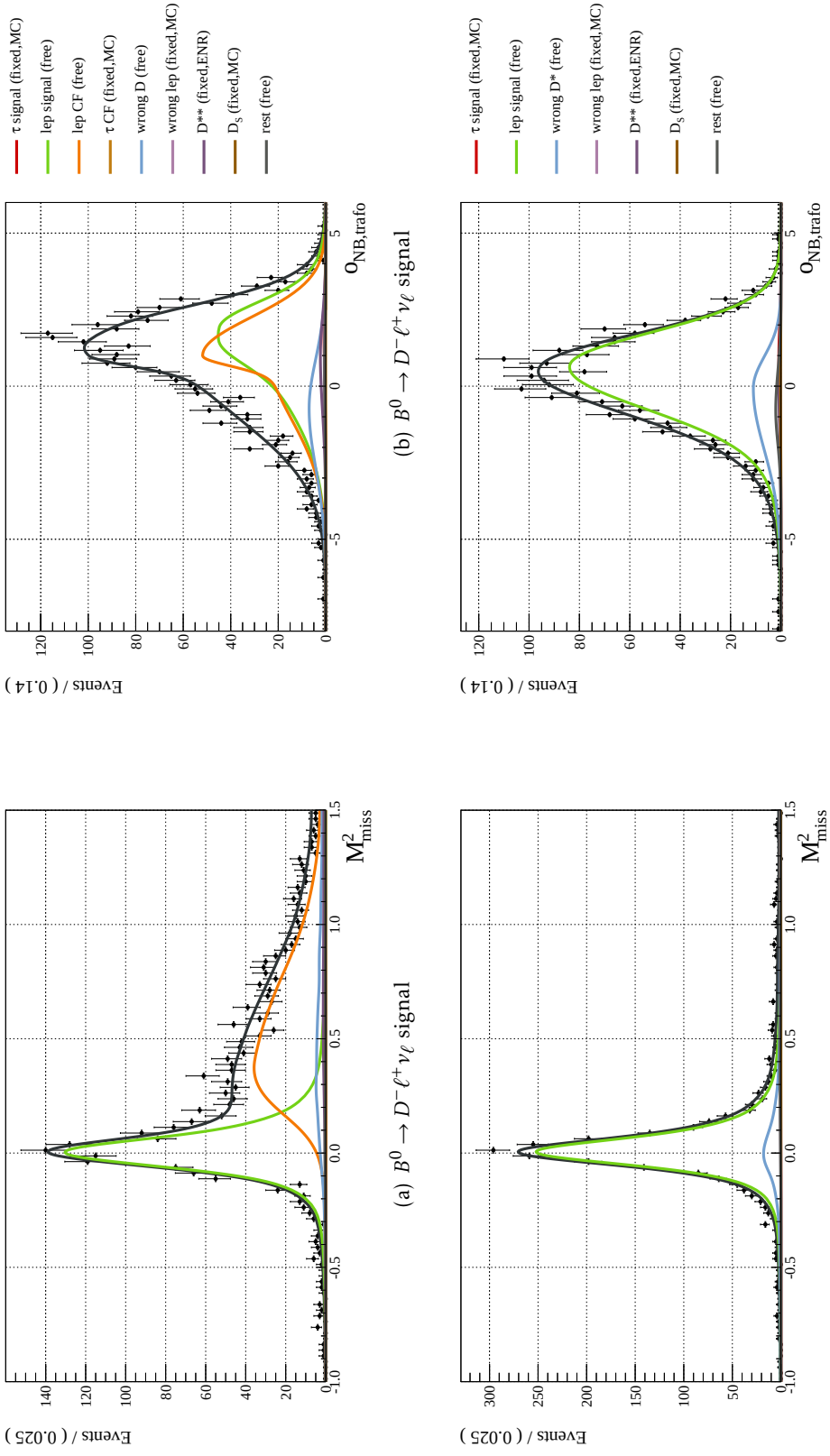
Figure 6.18.: Fit results of the normalisation check for $o_{\mathrm{NB}}(X) > 0.2$ and $o_{\mathrm{NB}}(X^*) > 0.2$. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4. The expectation for X on generic Monte Carlo is $X = 2.50$. The known uncertainty on data was used for the Monte Carlo expectation.



(a) $B^0 \to D^- \ell^+ \nu_\ell$ signal (b) $B^0 \to D^- \ell^+ \nu_\ell$ signal

(c) $B^0 \to D^{*-} \ell^+ \nu_\ell$ signal (d) $B^0 \to D^{*-} \ell^+ \nu_\ell$ signal

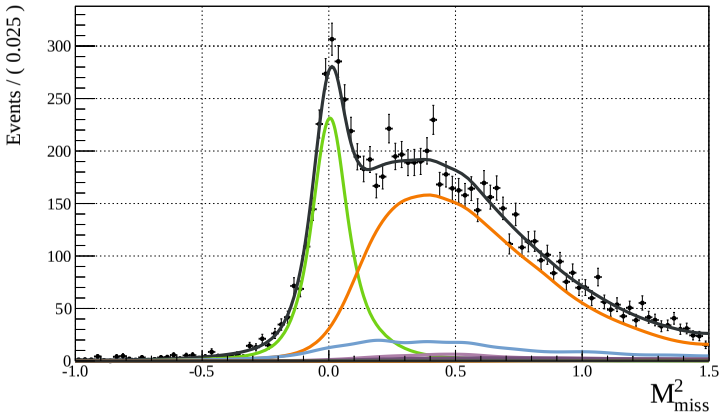
Figure 6.19.: Fit projections of the normalisation check. B^0 sample on generic Monte Carlo with $o_{\mathrm{NB}}(X) > 0.2$ and $o_{\mathrm{NB}}(X^*) > 0.2$.



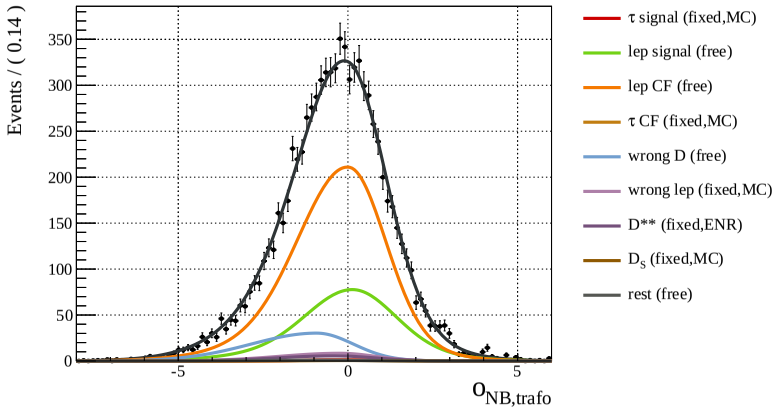
(a) $B^0 \to D^- \ell^+ \nu_\ell$ signal (b) $B^0 \to D^- \ell^+ \nu_\ell$ signal

(c) $B^0 \to D^{*-} \ell^+ \nu_\ell$ signal (d) $B^0 \to D^{*-} \ell^+ \nu_\ell$ signal

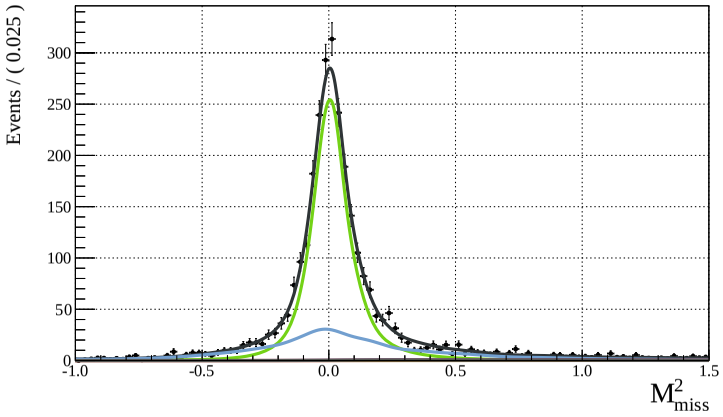
Figure 6.20.: Fit projections of the normalisation check. B^0 sample on data with $o_{\mathrm{NB}}(X) > 0.2$ and $o_{\mathrm{NB}}(X^*) > 0.2$.



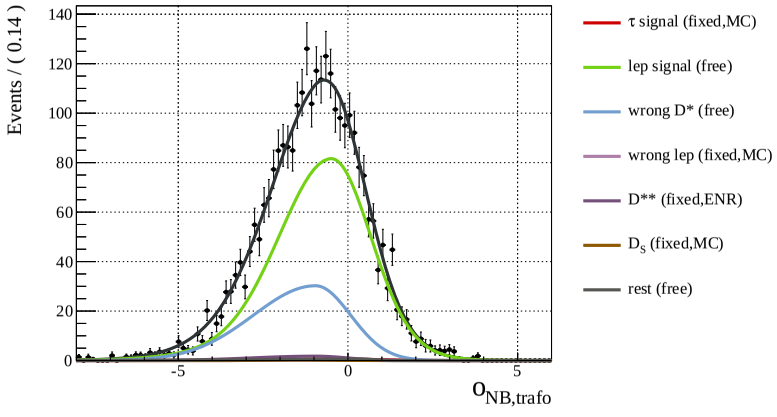
(a) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ signal



(b) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ signal



(c) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ signal



(d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ signal

Figure 6.21.: Fit projections of the normalisation check. B^0 sample on generic Monte Carlo with $o_{\mathrm{NB}}(X) > 0.2$ and $o_{\mathrm{NB}}(X^*) > 0.2$.

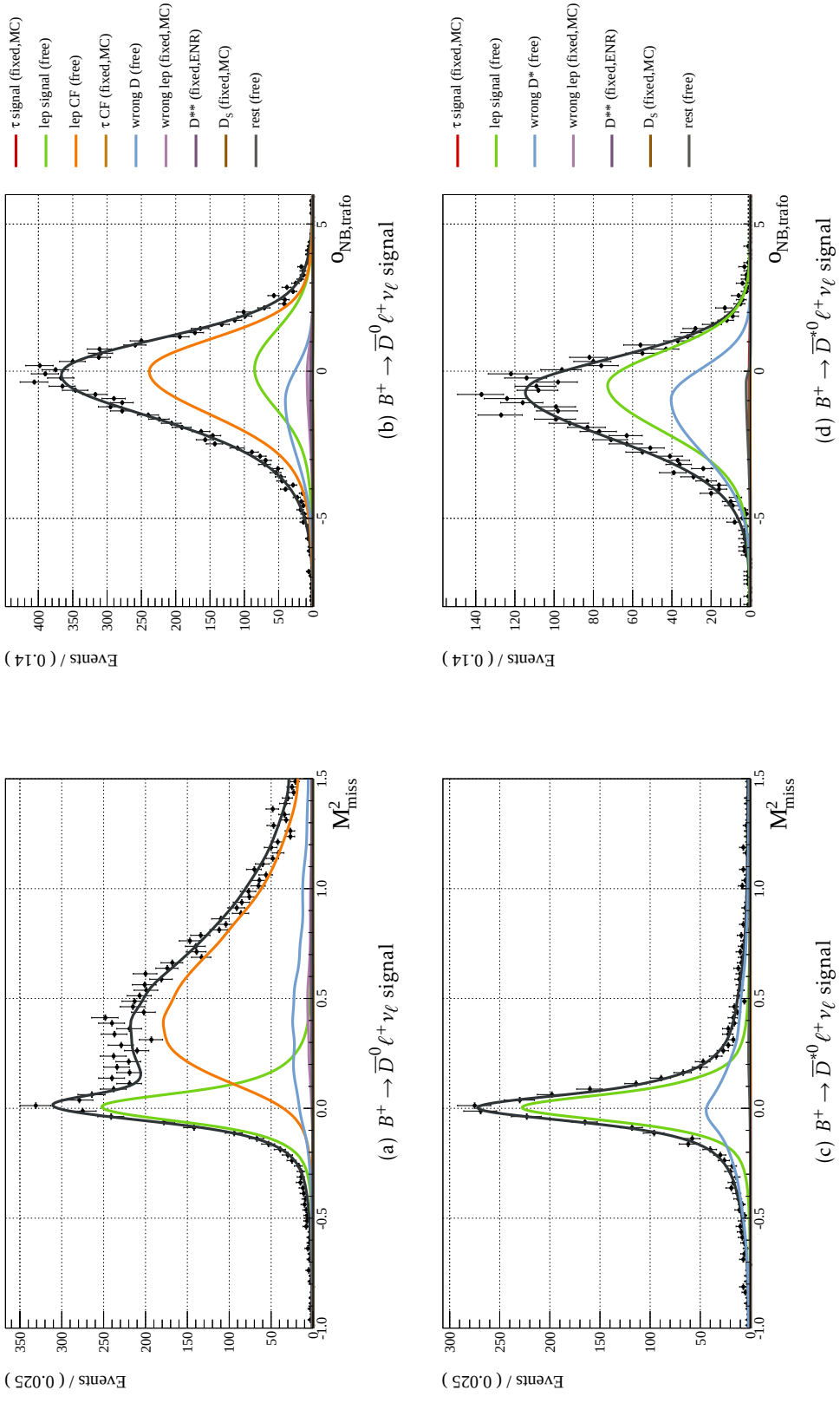


Figure 6.22.: Fit projections of the normalisation check. B^+ sample on data with $o_{\mathrm{NB}}(X) > 0.2$ and $o_{\mathrm{NB}}(X^*) > 0.2$.

6.11.4. Conclusion

As the sample is entirely dominated by the lepton signal, lepton cross-feed and the wrong $D^{(*)}$ background, only the modelling of these components could be checked. Other components of the samples have got a smaller yield than the statistical uncertainty of the three dominant components. It can be concluded from this check, that the modelling of the lepton signal, lepton cross-feed and the wrong $D^{(*)}$ background works satisfactorily and the expected results could be reproduced on data.
While the study on fitting X was helpful to confirm part of the fitting procedure for R on real data, a very important aspect could not be clarified by this procedure. The D^{**} background component had very little impact on the fitting of $X^{0/+}$ and will be investigated in further detail in section 6.12.

6.12. D^{**} Backgrounds

The correct estimation of background coming from $B \to D^{**}\ell\nu$ events is crucial for the success of this analysis. Those events cluster towards large values of M^2_{miss}, the same region where the $B \to D^{(*)}\tau\nu_\tau$ signal is expected. This background can therefore have substantial influence on the extraction of the τ signal.

6.12.1. Origin Of The D^{**} Backgrounds

Based on generic Monte Carlo simulation, the contributions of all D^{**} decay modes included in the generic Monte Carlo were estimated. Tables 6.15 and 6.16 summarise the results. From these tables, it is evident, that the majority of the D^{**} signal ends up as background in the signal sample by missing one neutral π^0 during reconstruction. Only a small fraction of the D^{**} signal ($< 10\%$) has a missing charged track in the final state.

6.12.2. D^{**} Enriched Sample

In order to estimate the contribution of the D^{**} backgrounds, a D^{**} enriched sample was reconstructed for each signal decay channel. This was achieved by explicitly adding a π^0 to each signal mode. Thus, there are 4 different D^{**} enriched samples:

- $B^0 \to D^-\ell^+\nu_\ell\pi^0$
- $B^0 \to D^{*-}\ell^+\nu_\ell\pi^0$
- $B^+ \to \overline{D}^0\ell^+\nu_\ell\pi^0$
- $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell\pi^0$

Each of these channels is used for the D^{**} background estimation in the respective signal channel without the added π^0. The justification for this procedure was given in section 6.12.1, specifically by tables 6.15 and 6.16. In decays with unexcited D mesons (i.e. $B^0 \to D^-e^+\nu_e$, $B^0 \to D^-\mu^+\nu_\mu$, $B^+ \to \overline{D}^0e^+\nu_e$ and $B^+ \to \overline{D}^0\mu^+\nu_\mu$) the D^{**} enriched sample contains by construction a certain amount of D^* signal. The expected mass differences $\Delta M(D^{**}, D)$ are well above 300 MeV/c^2. In choosing $\Delta M(D^{**}, D) > 200$ MeV/c^2, all of the D^* signal is suppressed while the bulk of the D^{**} signal is retained.

6.12.3. Composition And Fitting Of The D^{**} Enriched Sample

The D^{**} enriched samples consist of four distinct components:

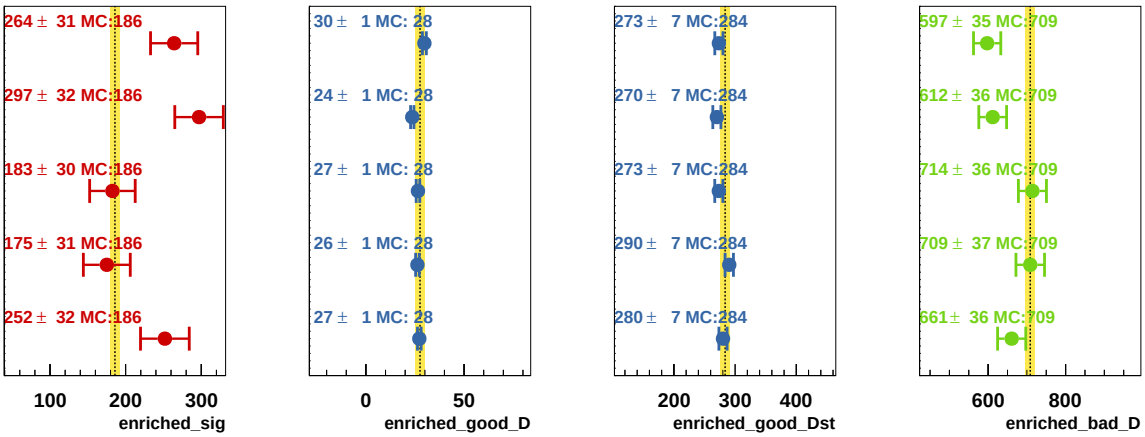
Signal The D^{**} signal component consists of real $B \to D^{**}\ell\nu$ events and peaks around $M^2_{\text{miss}} = 0.0$ GeV$^2/c^4$. This component is is used to estimate D^{**} background component in the signal sample (e.g. $B^0 \to D^-e^+\nu_e$). For that purpose, the two components are fitted simultaneously.

Random π^0 to $B \to D\ell\nu_\ell$ signal It can easily happen, that a correct $B \to D\ell\nu_\ell$ candidate was reconstructed and a random π^0 was added. By erroneously increasing the correct mass of the candidate, the *missing* mass is reduced below $M^2_{\text{miss}} = 0.0\ \text{GeV}^2/c^4$. To suppress this background, $M^2_{\text{miss}}$was calculated without the additional π^0. This results in a distribution, where the $B \to D\ell\nu_\ell$ signal is clearly visible as a sharp peak around $M^2_{\text{miss, no }\pi^0} = 0.0\ \text{GeV}^2/c^4$. Cutting away the sharp signal peak by requiring $M^2_{\text{miss, no }\pi^0} > 0.2\ \text{GeV}^2/c^4$ reduces this background. This component was fitted simultaneously with the $B \to D\ell\nu_\ell$ signal in the actual signal fit. The two yields are connected via a factor, determined on generic Monte Carlo. It is negligibly small in the $B \to D^*\ell\nu_\ell\pi^0$ samples and was not included in the corresponding PDFs.

Random π^0 to $B \to D^*\ell\nu_\ell$ signal In a similar way as for the $B \to D\ell\nu_\ell$ channel, a random π^0 can be attached to a correct $B \to D^*\ell\nu_\ell$ signal. With the same argument as above, it can also be reduced by a cut to M^2_{miss}, calculated without the extra π^0. This component can, however also occur in the $B \to D\ell\nu_\ell\pi^0$ channel. In that case, a random π^0 was added to cross-feed, a component where actually a π^0 was missed. Being cross-feed, this component does not peak around $M^2_{\text{miss, no }\pi^0} = 0.0\ \text{GeV}^2/c^4$, so it is not cut away by requiring $M^2_{\text{miss, no }\pi^0} > 0.2\ \text{GeV}^2/c^4$. But having roughly the correct missing mass (missing a π^0 and then adding one again), it peaks around $M^2_{\text{miss}} = 0.0\ \text{GeV}^2/c^4$. The peak is broader than the actual D^{**} signal peak and has to be separated by the fit. This component was fitted simultaneously with the $B \to D^*\ell\nu_\ell$ signal in the actual signal fit. The two yields are connected via a factor, determined on generic Monte Carlo.

Combinatorial background The normalisation is a free parameter of the fit.

To estimate the $B \to D^{**}\ell\nu$ background contribution in the signal sample, the signal component in the D^{**} enriched sample is used. It is extracted by a 2-dimensional fit to $M^2_{\text{miss}}$and $o_{\text{NB, trafo.}}$. For each of the four components described above, a smooth kernel estimation PDF was constructed in the M^2_{miss}dimension. Bifurcated Gaussian PDFs were used to model the distribution in $o_{\text{NB, trafo.}}$. In an analogous way as it was done for the normalisation check, the shapes were determined on an independent sample from the fitting sample. Table 6.17 gives an overview of the samples that were used. The results of the fits to generic Monte Carlo are summarised in figures 6.23 to 6.26.



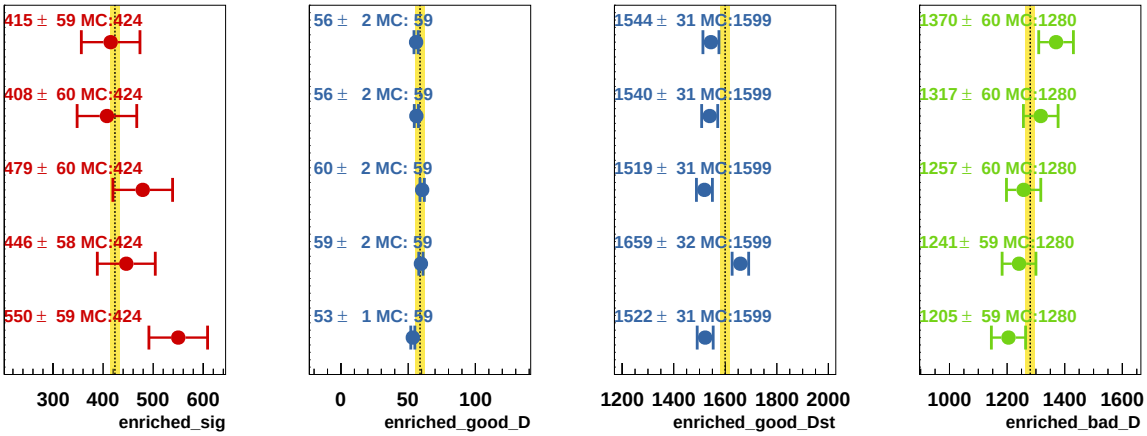
(a) D^{**} Signal (simultaneous with D^{**} BG in $B^0 \to D^- \ell^+ \nu_\ell$ sample)

(b) Random π^0 to $B^0 \to D^- \ell^+ \nu_\ell$ signal (simultaneous with $B^0 \to D^- \ell^+ \nu_\ell$ signal)

(c) Random π^0 to $B^0 \to D^{*-} \ell^+ \nu_\ell$ signal (simultaneous with $B^0 \to D^{*-} \ell^+ \nu_\ell$ signal)

(d) Combinatorial BG (free)

Figure 6.23.: Summary of the results of the fits to the D^{**} enriched samples ($B^0 \to D^- \ell^+ \nu_\ell \pi^0$) on generic Monte Carlo. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4.



(a) D^{**} Signal (simultaneous with D^{**} BG in $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample)

(b) Random π^0 to $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ signal (simultaneous with $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ signal)

(c) Random π^0 to $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ signal (simultaneous with $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ signal)

(d) Combinatorial BG (free)

Figure 6.24.: Summary of the results of the fits to the D^{**} enriched samples ($B^+ \to \overline{D}^0 \ell^+ \nu_\ell \pi^0$) on generic Monte Carlo. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4.

Percentage	Events	Channel
	$B^0 \to D^- e^+ \nu_e$	
67.13%	876	$B^0 \to D_0^{*+}(\to D^+\pi^0)e$
17.24%	225	$B^0 \to D_2^{*+}(\to D^+\pi^0)e$
6.67%	87	$B^+ \to D_0^{*0}(\to D^+\pi^-)e$
	$B^0 \to D^- \mu^+ \nu_\mu$	
64.95%	706	$B^0 \to D_0^{*+}(\to D^+\pi^0)\mu$
19.32%	210	$B^0 \to D_2^{*+}(\to D^+\pi^0)\mu$
7.73%	84	$B^+ \to D_0^{*0}(\to D^+\pi^-)\mu$
	$B^0 \to D^{*-} e^+ \nu_e$	
38.68%	345	$B^0 \to D_1^{\prime+}(\to D^{*+}\pi^0)e$
37.56%	335	$B^0 \to D_1^{+}(\to D^{*+}\pi^0)e$
6.28%	56	$B^0 \to D_2^{*+}(\to D^{*+}\pi^0)e$
5.38%	48	$B^+ \to D_1^{\prime 0}(\to D^{*+}\pi^-)e$
4.93%	44	$B^+ \to D_1^{0}(\to D^{*+}\pi^-)e$
	$B^0 \to D^{*-} \mu^+ \nu_\mu$	
44.21%	313	$B^0 \to D_1^{\prime+}(\to D^{*+}\pi^0)\mu$
32.91%	233	$B^0 \to D_1^{+}(\to D^{*+}\pi^0)\mu$
6.07%	43	$B^0 \to D_2^{*+}(\to D^{*+}\pi^0)\mu$
5.93%	42	$B^+ \to D_1^{\prime 0}(\to D^{*+}\pi^-)\mu$
5.37%	38	$B^+ \to D_1^{0}(\to D^{*+}\pi^-)\mu$

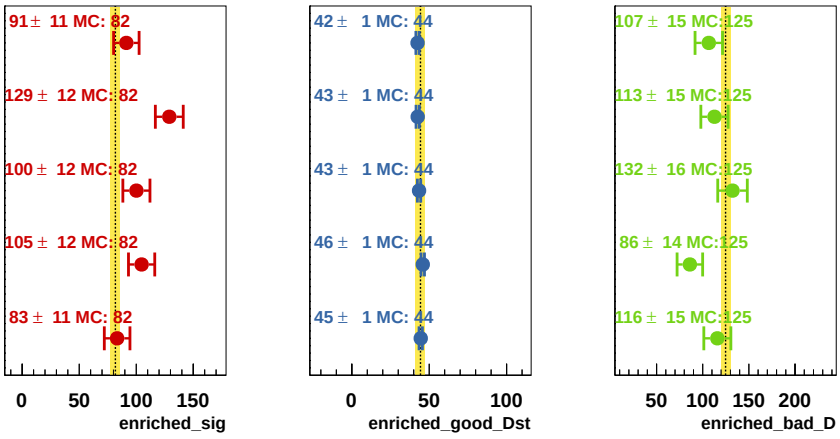
Table 6.15.: Composition of the D^{**} backgrounds in the B^0 sample. For clarity, only the top 90% of the backgrounds are listed.

Percentage	Events	Channel
	$B^+ \to \overline{D}^0 e^+ \nu_e$	
67.90%	1231	$B^+ \to D_0^{*0}(\to D^0\pi^0)e$
18.81%	341	$B^+ \to D_2^{*0}(\to D^0\pi^0)e$
4.30%	78	$B^0 \to D_0^{*+}(\to D^0\pi^-)e$
	$B^+ \to \overline{D}^0 \mu^+ \nu_\mu$	
66.19%	971	$B^+ \to D_0^{*0}(\to D^0)\pi^0\mu$
19.97%	293	$B^+ \to D_2^{*0}(\to D^0\pi^0)\mu$
3.82%	56	$B^0 \to D_0^{*+}(\to D^0\pi^-)\mu$
2.18%	32	$B^+ \to D_0^{*0}(\to D^0\pi^0)\tau$
	$B^+ \to \overline{D}^{*0} e^+ \nu_e$	
39.21%	158	$B^+ \to D_1^0(\to D^{*0}\pi^0)e$
38.71%	156	$B^+ \to D_1^{\prime 0}(\to D^{*0}\pi^0)e$
8.19%	33	$B^+ \to D_2^{*0}(\to D^{*0}\pi^0)e$
2.98%	12	$B^0 \to D_1^{\prime +}(\to D^{*0}\pi^+)e$
1.99%	8	$B^0 \to D_1^+(\to D^{*0}\pi^+)e$
	$B^+ \to \overline{D}^{*0} \mu^+ \nu_\mu$	
43.22%	137	$B^+ \to D_1^{\prime 0}(\to D^{*0}\pi^0)\mu$
40.06%	127	$B^+ \to D_1^0(\to D^{*0}\pi^0)\mu$
5.68%	18	$B^+ \to D_2^{*0}(\to D^{*0}\pi^0)\mu$
2.52%	8	$B^0 \to D_1^+(\to D^{*0}\pi^+)\mu$

Table 6.16.: Composition of the D^{**} backgrounds in the B^+ sample. For clarity, only the top 90% of the backgrounds are listed.

Fit on	Shapes determined on
Generic Monte Carlo, stream 0	Generic Monte Carlo, streams 1,2,3,4
Generic Monte Carlo, stream 1	Generic Monte Carlo, streams 0,2,3,4
Generic Monte Carlo, stream 2	Generic Monte Carlo, streams 0,1,3,4
Generic Monte Carlo, stream 3	Generic Monte Carlo, streams 0,1,2,4
Generic Monte Carlo, stream 4	Generic Monte Carlo, streams 0,1,2,3
Data	Generic Monte Carlo, streams 0,1,2,3,4

Table 6.17.: Samples used for the fits in the D^{**} backgrounds check.

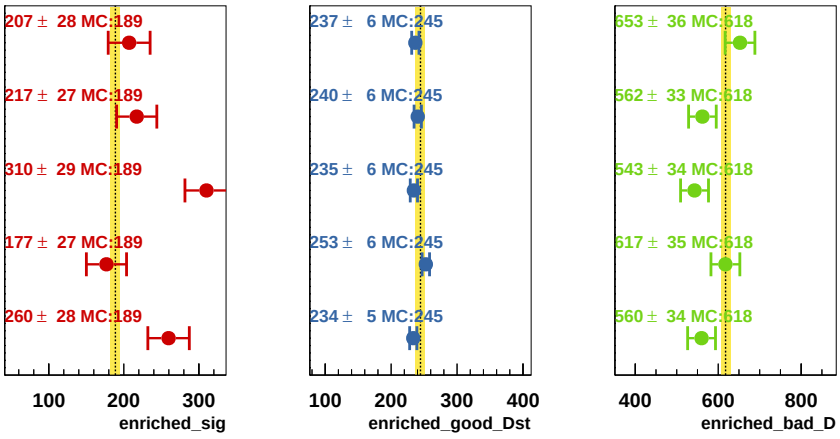


(a) D^{**} Signal (simultaneous with D^{**} BG in $B^0 \to D^{*-}\ell^+\nu_\ell$ sample)

(b) Random π^0 to $B^0 \to D^{*-}\ell^+\nu_\ell$ signal (simultaneous with $B^0 \to D^{*-}\ell^+\nu_\ell$ signal)

(c) Combinatorial BG

Figure 6.25.: Summary of the results of the fits to the D^{**} enriched samples ($B^0 \to D^{*-}\ell^+\nu_\ell\pi^0$) on generic Monte Carlo. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4.



(a) D^{**} Signal (simultaneous with D^{**} BG in $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ sample)

(b) Random π^0 to $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ signal (simultaneous with $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ signal)

(c) Combinatorial BG

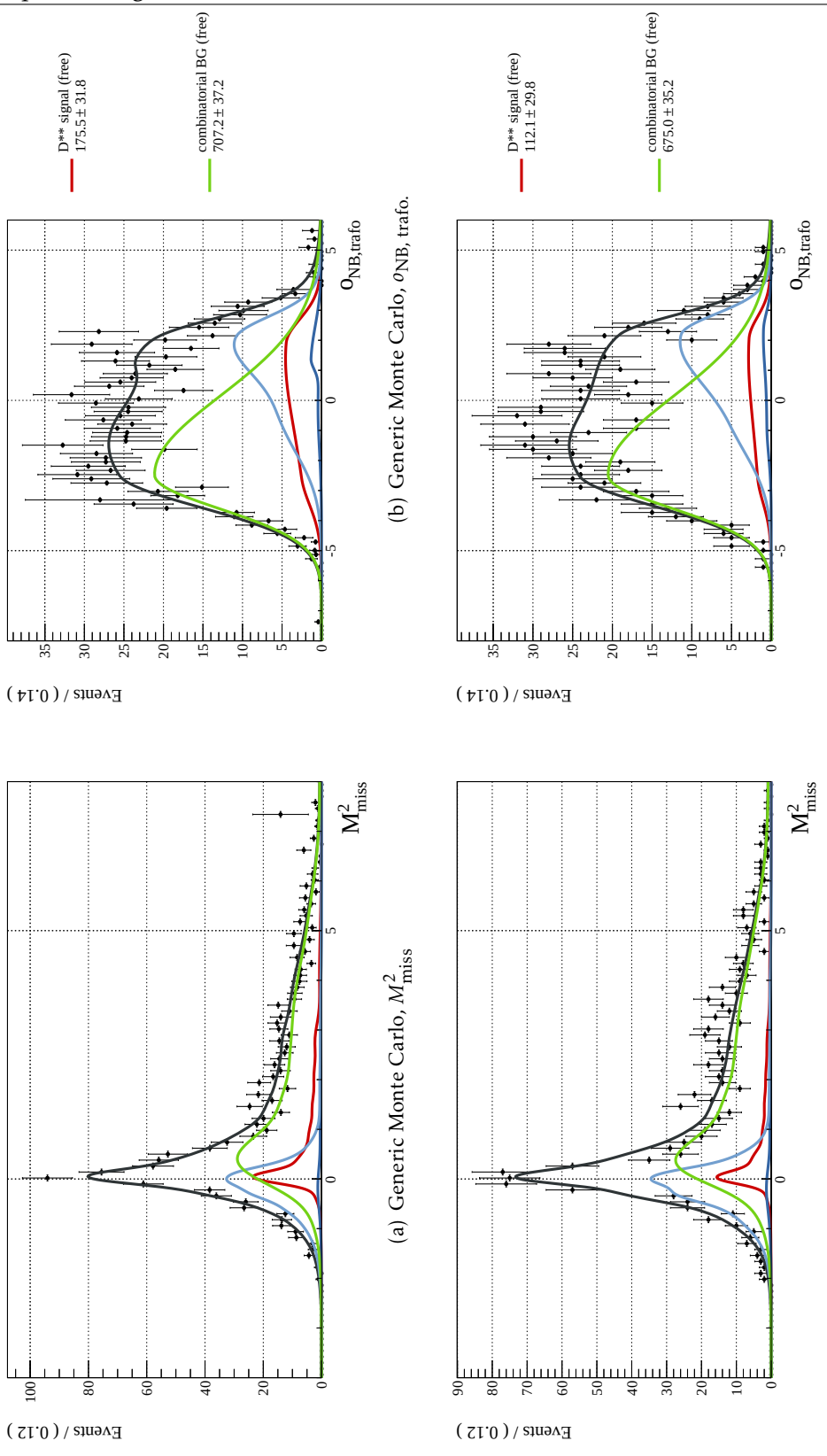
Figure 6.26.: Summary of the results of the fits to the D^{**} enriched samples ($B^+ \to \overline{D}^{*0}\ell^+\nu_\ell\pi^0$) on generic Monte Carlo. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4.

6.12.4. Data-Monte Carlo Comparison

While the above described strategy works reasonably well on generic Monte Carlo, this is not guaranteed on real data. To test this, the above described fits were performed on the D^{**} enriched sample from real data and compared to generic Monte Carlo. It is known, that the branching fractions that are used in the generic Monte Carlo for the D^{**} decay channels are too large [77, 82, 83, 84, 85]. Table 6.18 gives the expected data/Monte Carlo ratios along with the fit yields. A good agreement within the uncertainties could be found. Please note that the fit results on Monte Carlo were obtained by using 5 streams of Monte Carlo. This accounts for the significantly lower uncertainties. The projections of the fits to real data, along with exemplary fit projections on Monte Carlo can be seen in figures 6.27 to 6.30.

Sample	$B^+ \to \overline{D}^0 \ell^+ \nu_\ell \pi^0$	$B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell \pi^0$	$B^0 \to D^- \ell^+ \nu_\ell \pi^0$	$B^0 \to D^{*-} \ell^+ \nu_\ell \pi^0$
		Fitted yields		
Data (fit)	154.1 ± 69.1	76.9 ± 18.4	112.1 ± 29.8	30.4 ± 7.5
MC (fit)	427.6 ± 29.8	157.5 ± 8.9	234.0 ± 14.3	72.3 ± 3.9
Ratio	$\mathbf{0.36 \pm 0.16}$	$\mathbf{0.49 \pm 0.12}$	$\mathbf{0.48 \pm 0.13}$	$\mathbf{0.42 \pm 0.11}$
		Branching fractions		
Data (expected)	$(0.21 \pm 0.03)\%$	$(0.31 \pm 0.03)\%$	$(0.22 \pm 0.03)\%$	$(0.25 \pm 0.03)\%$
MC (used)	0.45%	0.63%	0.42%	0.58%
Ratio	$\mathbf{0.47 \pm 0.07}$	$\mathbf{0.49 \pm 0.05}$	$\mathbf{0.52 \pm 0.07}$	$\mathbf{0.43 \pm 0.05}$

Table 6.18.: D^{**} enriched samples: Expectation and fit results



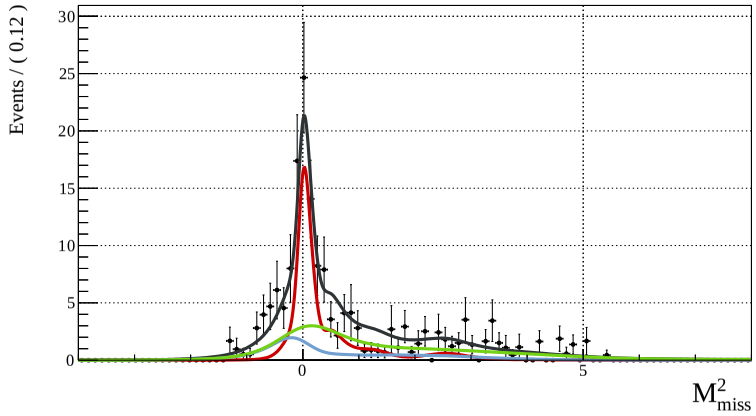
(a) Generic Monte Carlo, M^2_{miss}

(b) Generic Monte Carlo, $o_{\mathrm{NB,\ trafo.}}$

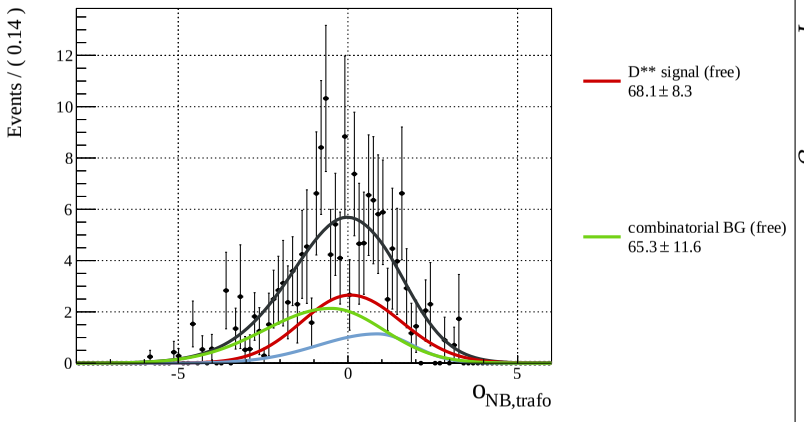
(c) Data, M^2_{miss}

(d) Data, $o_{\mathrm{NB,\ trafo.}}$

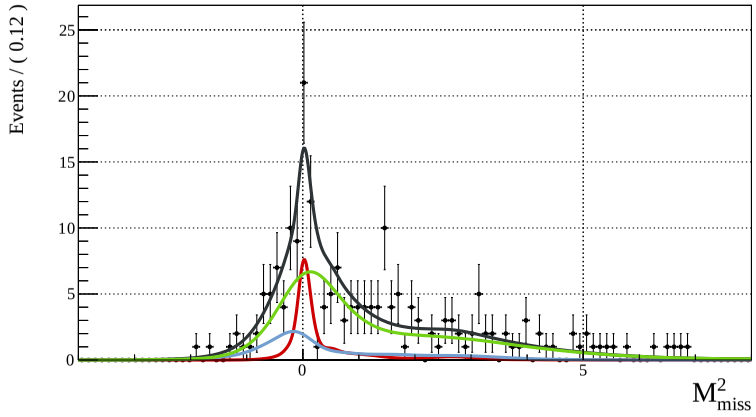
Figure 6.27.: Fit projections of the fits to the D^{**} enriched samples ($B^0 \to D^- \ell^+ \nu_\ell \pi^0$) on generic Monte Carlo and data. $o_{\mathrm{NB}}(D_{\mathrm{enr.}}) > 0.0$ and $o_{\mathrm{NB}}(D^*_{\mathrm{enr.}}) > 0.0$.



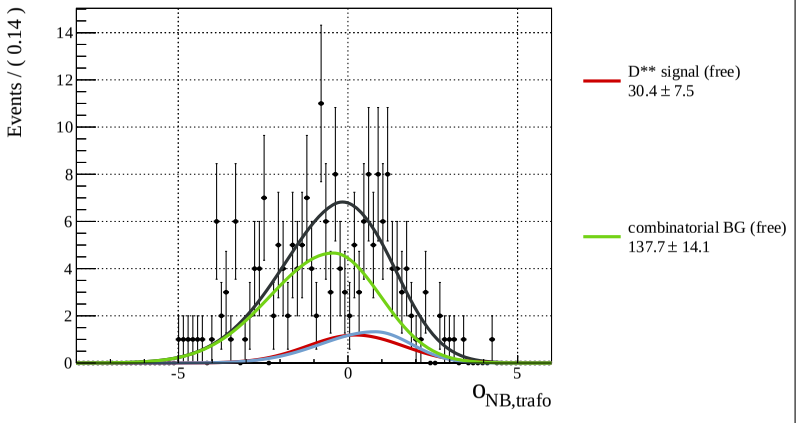
(a) Generic Monte Carlo, M^2_{miss}



(b) Generic Monte Carlo, $o_{\text{NB, trafo.}}$

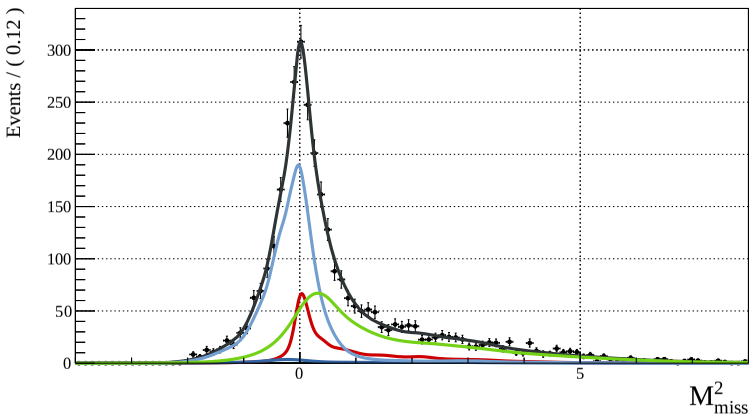


(c) Data, M^2_{miss}

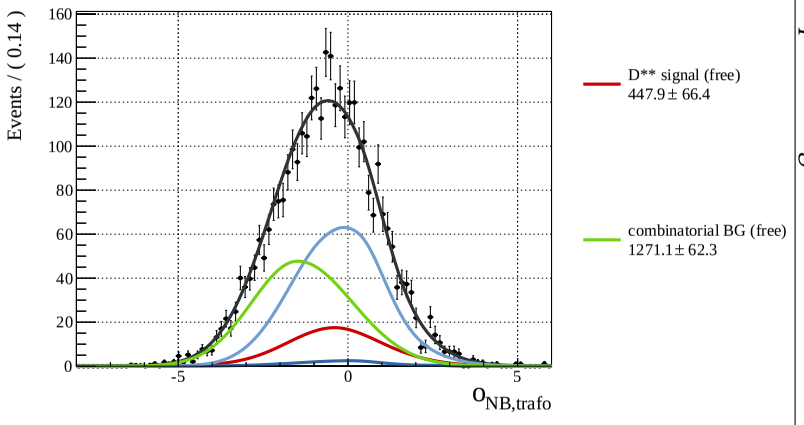


(d) Data, $o_{\text{NB, trafo.}}$

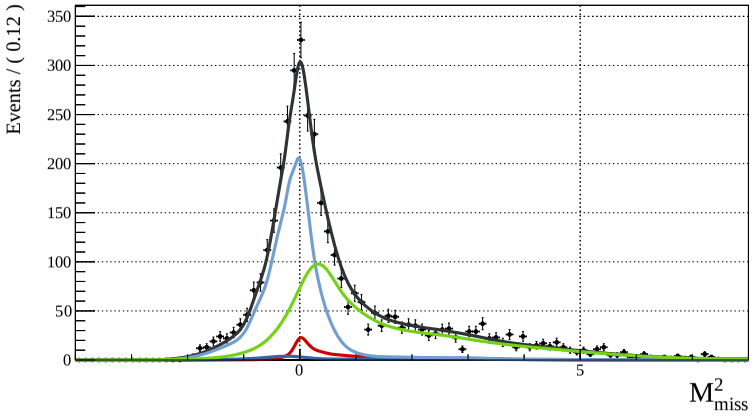
Figure 6.28.: Fit projections of the fits to the D^{**} enriched samples ($B^0 \to D^{*-}\ell^+\nu_\ell\pi^0$) on generic Monte Carlo and data. $o_{\text{NB}}(D_{\text{enr.}}) > 0.0$ and $o_{\text{NB}}(D^*_{\text{enr.}}) > 0.0$.



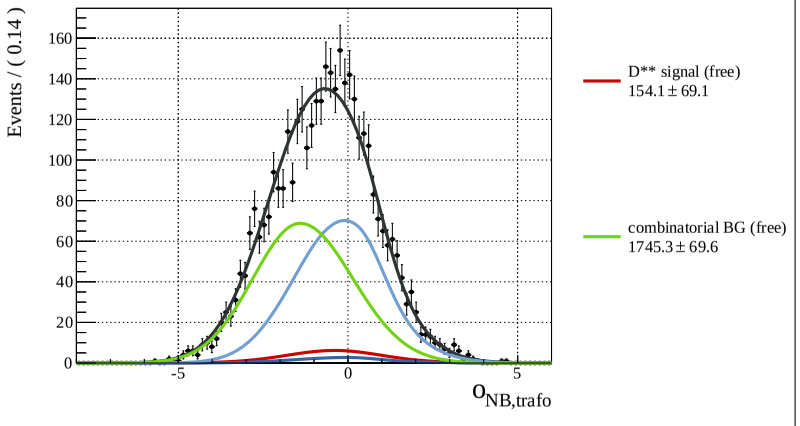
(a) Generic Monte Carlo, M^2_{miss}



(b) Generic Monte Carlo, $o_{\text{NB, trafo.}}$

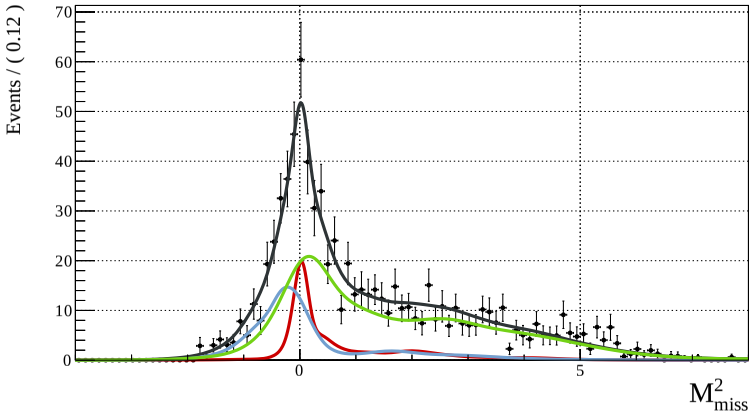


(c) Data, M^2_{miss}

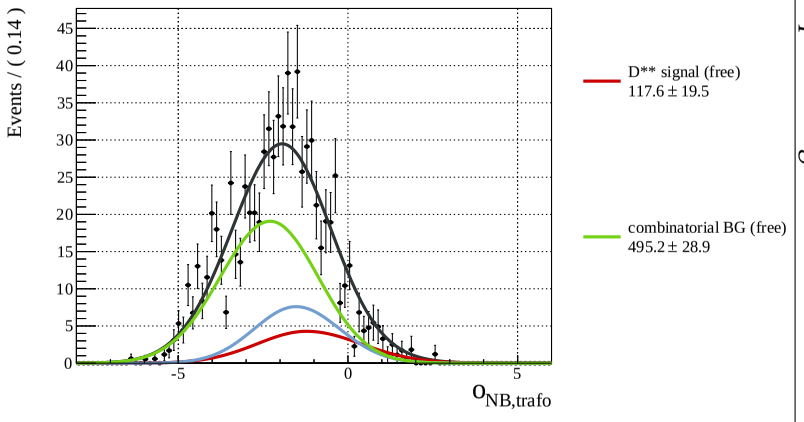


(d) Data, $o_{\text{NB, trafo.}}$

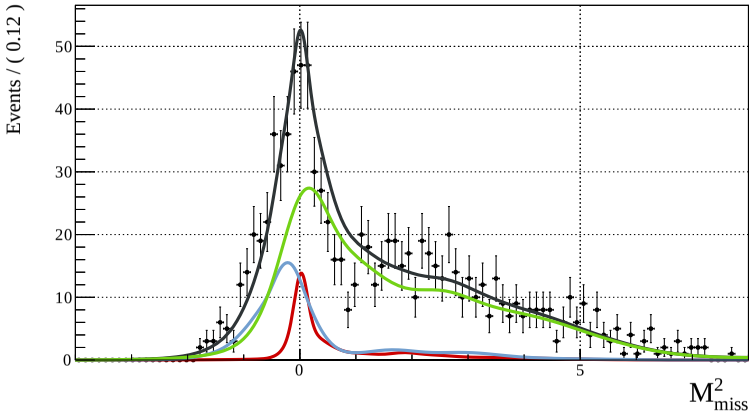
Figure 6.29.: Fit projections of the fits to the D^{**} enriched samples ($B^+ \to \overline{D}^0 \ell^+ \nu_\ell \pi^0$) on generic Monte Carlo and data. $o_{\text{NB}}(D_{\text{enr.}}) > 0.0$ and $o_{\text{NB}}(D^*_{\text{enr.}}) > 0.0$.



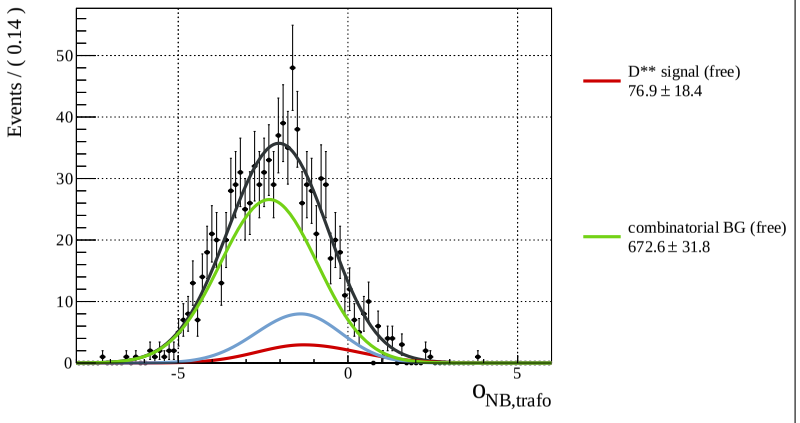
(a) Generic Monte Carlo, M^2_{miss}



(b) Generic Monte Carlo, $o_{\text{NB, trafo.}}$



(c) Data, M^2_{miss}



(d) Data, $o_{\text{NB, trafo.}}$

Figure 6.30.: Fit projections of the fits to the D^{**} enriched samples ($B^+ \to \overline{D}^{*0}\ell^+\nu_\ell\pi^0$) on generic Monte Carlo and data. $o_{\text{NB}}(D_{\text{enr.}}) > 0.0$ and $o_{\text{NB}}(D^*_{\text{enr.}}) > 0.0$.

6.13. Fit Strategies

There are two different approaches to extract R and R^* from the B^0 and B^+ samples.

1. Two independent fits are performed, one to the B^0 sample and another one to the B^+ sample.
2. The B^0 and B^+ samples are fitted simultaneously. The shared parameters are R and R^*.

Both approaches have their advantages, so both were implemented. In the following descriptions, let us focus on the two independent fits. The small technical differences of the two separate fits to the simultaneous fit will be explained in section 6.19. Be aware that the two independent fits that are described in the following are also "simultaneous" fits. The two samples that are simultaneously fitted in that case are the D and the D^* sample. In a manner of speaking, the two independent fits are "simultaneous" fits and the fit to the combined B^0 and B^+ samples is a "double-simultaneous" fit.

6.13.1. Fit Components

The fit strategy for extracting R and R^* is very similar to the one used for extracting X (see section 6.11.1). There are, however, some key differences. Therefore, the entire strategy shall be explained again to avoid confusion.
For each component that could be identified in the $B \to D\ell\nu_\ell$ and the $B \to D^*\ell\nu_\ell$ samples (see section 6.10) a PDF was determined and included in the fit. For the fitting procedure, the shapes of all components were determined on generic Monte Carlo samples unless stated otherwise. The pdfs in $M^2_{\text{miss}}$are determined as smoothed kernel estimation functions [76]. The shapes in the $o_{\text{NB, trafo.}}$dimension (see section 6.7) were parametrised by bifurcated Gaussian pdfs. The following components are included in the fit:

Lepton signal The normalisation is a free parameter of the fit.

Lepton cross-feed This is a major source of background in the $B \to D\ell\nu_\ell$ sample. The normalisation is a free parameter of the fit.

Wrong charge cross-feed This is a small source of background that only occurs in the $B^+ \to \overline{D}^0\ell^+\nu_\ell$ sample. Its yield is fixed to Monte Carlo expectation.

***R*, *R** and τ signal** The main focus of this fitting procedure lies on the extraction of R and R^*, as defined in equations 6.1 to 6.4. Therefore, R and R^* were implemented as free parameters of the fit. They are linked to the lepton and τ signal by the factors f_R and f_R^*. These factors were determined on generic Monte Carlo

as follows:

$$f_R = R_{\mathrm{WA}} \cdot \frac{D\ell\text{signal yield}}{D\tau\text{signal yield}} \tag{6.49}$$

$$f_R^* = R_{\mathrm{WA}}^* \cdot \frac{D^*\ell\text{signal yield}}{D^*\tau\text{signal yield}}\,, \tag{6.50}$$

where $R_{\mathrm{WA}}^{(*)}$ are the current world average values, which are also used on generic Monte Carlo. Therefore, the τ signal yield itself is not a free parameter of the fit, but a function of $R^{(*)}$ and the lepton signal yield. The shape of the τ signal component was determined on especially produced τ signal Monte Carlo.

τ cross-feed This component cannot be reliably separated by the fit, as its shape is very similar to the τ signal. It has to be determined using other information. The lepton and τ cross-feeds are in reality D^* lepton or $D^*\tau$ signal events, that were misreconstructed. The same reconstruction error - missing a slow pion or photon - was made for both components. It is therefore reasonable to assume, that the two cross-feeds and their corresponding signal components are connected in the following way:

$$\frac{\ell\mathrm{CF}}{D^*\ell\text{signal}} = g \cdot \frac{\tau\mathrm{CF}}{D^*\tau\text{signal}}\,. \tag{6.51}$$

The factor g takes possible differences between $B \to D^*\ell\nu_\ell$ and $B \to D^*\tau\nu_\tau$ decays into account, most likely kinematic differences due to the large τ mass. The factor g is determined on generic Monte Carlo and the τ cross-feed is then determined in the following way:

$$\tau\mathrm{CF} = \frac{1}{g} \cdot D^*\tau\text{signal} \cdot \frac{\ell\mathrm{CF}}{D^*\ell\text{signal}}\,. \tag{6.52}$$

The τ cross-feed is therefore not a free parameter of the fit. It is determined by the free parameters ℓCF, $D^*\ell$signal and $D^*\tau$signal, the latter of which is a function of R^*, as explained above.

wrong *D* background The normalisation of this background component was determined before the actual fit by using D mass sideband data. On generic Monte Carlo, a factor f_{SB} was determined using Monte Carlo information::

$$f_{SB} = \frac{\text{wrong } D \text{ bg yield}}{\#\text{ of events in } D \text{ mass SB}} \tag{6.53}$$

Before the actual fit started, the yield of the wrong D background was determined by counting the events found in the D mass sidebands. Thus, an estimate for the wrong D background in the actual fit was obtained:

$$\text{wrong } D \text{ bg yield} = f_{SB} \cdot \#\text{ of events in } D \text{ mass SB} \tag{6.54}$$

The normalisation of this component was then fixed in the fit.

wrong D^* background The normalisation of this background component was determined before the actual fit by using $\Delta M(D^*, D)$ sidebands. Otherwise, the procedure is completely analogous to the wrong D background. The factor connecting the $\Delta M(D^*, D)$ sidebands and the wrong D^* component is f^*_{SB} and defined as

$$f^*_{SB} = \frac{\text{wrong } D^* \text{ bg yield}}{\#\text{ of events in } \Delta M(D^*, D) \text{ SB}} \tag{6.55}$$

D^{**} background The normalisation of this background component was determined simultaneously with the actual signal fit by using the D^{**} enriched sample (see section 6.12.2). For each component that could be identified in the D^{**} enriched sample (see section 6.12.2) a PDF was determined and included in the fit.

Signal This component is fitted simultaneously with the D^{**} background component in the signal sample.

Random π^0 to $B \to D\ell\nu_\ell$ A random π^0 was added to a correctly reconstructed $B \to D\ell\nu_\ell$ decay. This component was fitted simultaneously with the signal in the actual signal fit. The two yields are connected via a factor f_{e1}. It is determined on generic Monte Carlo as follows:

$$\text{Random } \pi^0 \text{ to } D\ell \text{ signal yield} = f_{e1} \cdot\ B \to D\ell\nu_\ell \text{ signal yield} \tag{6.56}$$

Random π^0 to $B \to D^*\ell\nu_\ell$ signal A random π^0 was added to an actual $B \to D^*\ell\nu_\ell$ decay. This can happen in the $B \to D^*\ell\nu_\ell\pi^0$ sample as well as in the $B \to D\ell\nu_\ell\pi^0$ sample ("cross-feed"). These two components were fitted simultaneously with the $B \to D^*\ell\nu_\ell$ signal in the actual signal fit. The two yields are connected to the $B \to D^*\ell\nu_\ell$ signal via two factors f_{e2} and f^*_e, determined on generic Monte Carlo:

$$\text{Random } \pi^0 \text{ to } D^*\ell \text{ cross-feed} = f_{e2} \cdot\ B \to D^*\ell\nu_\ell \text{ signal yield} \tag{6.57}$$

$$\text{Random } \pi^0 \text{ to } D^*\ell \text{ signal yield} = f^*_e \cdot\ B \to D^*\ell\nu_\ell \text{ signal yield} \tag{6.58}$$

Combinatorial background The normalisation is a free parameter of the fit.

With the fitting procedure for the D^{**} signal component ready, the simultaneous fitting of the D^{**} background can be set up: On generic Monte Carlo, a factor $f_D^{(*)}$ was determined:

$$f_D^{(*)} = \frac{D^{**} \text{ bg yield in } B \to D^{(*)}\ell\nu_\ell \text{ sample}}{\text{signal yield in } B \to D^{(*)}\ell\nu_\ell\pi^0 \text{ sample}} \tag{6.59}$$

The normalisation of the D^{**} background component was fitted simultaneously with the signal component in the D^{**} enriched sample, as explained above.

$$D^{**} \text{ bg yield in } B \to D^{(*)}\ell\nu_\ell \text{ sample} = f_D^{(*)} \cdot \text{signal yield in } B \to D^{(*)}\ell\nu_\ell\pi^0 \text{ sample} \tag{6.60}$$

D_S background This background component has a very small yield. The yield was fixed to standard model expectations of the generic Monte Carlo. For the fit on data, the branching ratio was corrected to the most precise measurements.

wrong lepton background This background component has a very small yield. The yield was fixed to MC expectation. For the fit on data, the fake lepton rates for real data were taken into account.

remaining backgrounds This background component has a very small yield and is fixed to MC expectation.

The full PDF can thus be written as a sum of the following 4 components:

$$
\begin{aligned}
\mathcal{P}_D &= Y_{D\ell\text{sig}}\mathcal{P}_{D\ell\text{sig}} + Y_{D\ell\text{CF}}\mathcal{P}_{D\ell\text{CF}} + (R\cdot Y_{D\ell\text{sig}}\cdot f_R)\mathcal{P}_{D\tau\text{sig}} \\
&\quad + \frac{Y_{D\ell\text{CF}}\cdot Y_{D^*\tau\text{sig}}}{Y_{D^*\ell\text{sig}}\cdot g}\mathcal{P}_{Dtcf} \\
&\quad + Y_{\text{wrong}D}\mathcal{P}_{\text{wrong}D} + (Y^{\text{enr}D}_{D^{**}\ell\text{sig}}\cdot f_D)\mathcal{P}_{D^{**}\text{in}D} + Y_{D_S\text{in}D}\mathcal{P}_{D_S\text{in}D} \\
&\quad + Y_{\text{wrong charge in}D}\mathcal{P}_{\text{wrong charge in}D} + Y_{\text{wrong}\ell\text{in}D}\mathcal{P}_{\text{wrong}\ell\text{in}D} + Y_{D\text{rest}}\mathcal{P}_{D\text{rest}} \\
\mathcal{P}_{\text{enr}D} &= Y^{\text{enr}D}_{D^{**}\ell\text{sig}}\mathcal{P}^{\text{enr}D}_{D^{**}\ell\text{sig}} + (f_{e1}\cdot Y_{D\ell\text{sig}})\mathcal{P}^{\text{enr}D}_{D\text{sig+random}\pi^0} \\
&\quad + (f_{e2}\cdot Y_{D^*\ell\text{sig}})\mathcal{P}^{\text{enr}D}_{D^*\text{sig+random}\pi^0} + Y^{\text{enr}D}_{\text{comb}}\mathcal{P}^{\text{enr}D}_{\text{comb}} \\
\mathcal{P}_{D^*} &= Y_{D^*\ell\text{sig}}\mathcal{P}_{D^*\ell\text{sig}} + (R^*\cdot Y_{D^*\ell\text{sig}}\cdot f^*_R)\mathcal{P}_{D^*\tau\text{sig}} \\
&\quad + Y_{\text{wrong}D^*}\mathcal{P}_{\text{wrong}D^*} + (Y^{\text{enr}D^*}_{D^{**}\ell\text{sig}}\cdot f^*_D)\mathcal{P}_{D^{**}\text{in}D^*} + Y_{D_S\text{in}D^*}\mathcal{P}_{D_S\text{in}D^*} \\
&\quad + Y_{\text{wrong}\ell\text{in}D^*}\mathcal{P}_{\text{wrong}\ell\text{in}D^*} + Y_{D^*\text{rest}}\mathcal{P}_{D^*\text{rest}} \\
\mathcal{P}_{\text{enr}D^*} &= Y^{\text{enr}D^*}_{D^{**}\ell\text{sig}}\mathcal{P}^{\text{enr}D^*}_{D^{**}\ell\text{sig}} \\
&\quad + (f^*_e\cdot Y_{D^*\ell\text{sig}})\mathcal{P}^{\text{enr}D^*}_{D^*\text{sig+random}\pi^0} + Y^{\text{enr}D^*}_{\text{comb}}\mathcal{P}^{\text{enr}D^*}_{\text{comb}}
\end{aligned}
$$

where $\mathcal{P}$ stands for a 2-dimensional pdf, consisting of a non-parametric kernel-estimation pdf ($\mathcal{K}$) in one dimension (M^2_{miss}) and a (sum of) bifurcated Gaussians ($\mathcal{G}$) in the other dimenson ($o_{\text{NB, trafo.}}$):

$$\mathcal{P} = \mathcal{K}(M^2_{\text{miss}})\cdot\mathcal{G}(o_{\text{NB, trafo.}})$$

$Y_{D\ell\text{sig}}$, $Y_{D^*\ell\text{sig}}$, $Y_{D\ell\text{CF}}$, R, R^*, $Y^{\text{enr}D}_{D^{**}\ell\text{sig}}$, $Y^{\text{enr}D^*}_{D^{**}\ell\text{sig}}$, $Y^{\text{enr}D}_{\text{comb}}$ and $Y^{\text{enr}D^*}_{\text{comb}}$ are free parameters of the fit, $Y_{\text{wrong charge in}D}$ is fixed to Monte Carlo expectation in the B^+ sample and set to 0 in the B^0 sample. f_R and f^*_R are factors containing intermediate branching ratios and efficiencies for the calculation of R and R^* respectively. g (see equation 6.51), f_D, f^*_D (see equation 6.59), f_{e1}, f_{e2} and f^*_e (see equations 6.56 to 6.58) are determined from Monte Carlo simulation. The numerical values of these factors are given in the following section. Factors that occur multiple times are coloured to help readability.

6.13.2. Factors

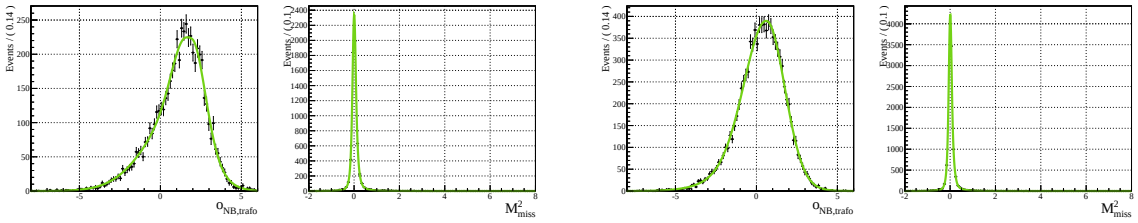
When performing the simultaneous fitting procedure, the yields of certain components are linked using a numerical factor, as descried in detail in section 6.13.1. The numerical values of the used factors are given in table 6.19.

Component		Eqn	Value (B^0)	Value (B^+)	Streams
wrong D bg	f_{SB}	6.53	0.835 ± 0.020	0.748 ± 0.017	5
wrong D^* bg	f^*_{SB}	6.55	1.628 ± 0.074	0.737 ± 0.016	5
D^{**} bg	f_D	6.59	1.044 ± 0.033	1.011 ± 0.031	10
D^{**} bg	f^*_D	6.59	2.323 ± 0.114	0.781 ± 0.047	10
$D\ell\nu$ sig in enr	f_{e1}	6.56	0.022 ± 0.001	0.027 ± 0.002	10
$D^*\ell\nu$ CF in enr	f_{e2}	6.57	0.156 ± 0.003	0.726 ± 0.011	10
$D^*\ell\nu$ sig in enr	f^*_e	6.58	0.016 ± 0.001	0.079 ± 0.003	10
Eff factor for R	f_R	6.49	3.350 ± 0.106	3.381 ± 0.115	10
Eff factor for R^*	f^*_R	6.50	5.719 ± 0.200	5.192 ± 0.232	10
τ CF	g	6.51	1.137 ± 0.065	1.450 ± 0.080	10

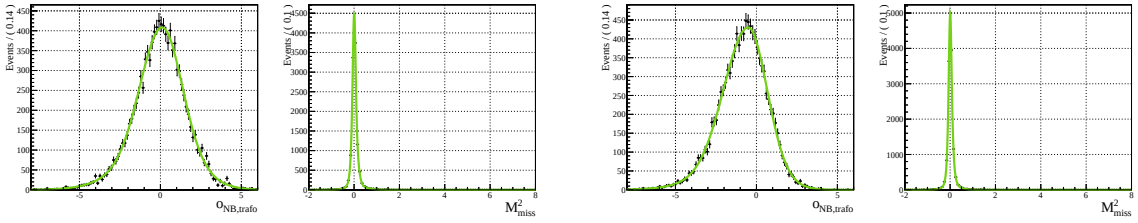
Table 6.19.: Factors used for simultaneous fitting determined on 5 or 10 streams of generic Monte Carlo. Some values could only be calculated using 5 streams of generic Monte Carlo, because the other 5 streams did not contain non-$B\bar{B}$ background.

6.13.3. Shapes

The shapes for the individual components described in section 6.13.1 are shown in figures 6.31 to 6.39. For this illustration, the shapes were determined on all 10 streams of generic Monte Carlo. An exception is the τ signal shape. This was determined on signal Monte Carlo, as the sample size yielded from even 10 streams of generic Monte Carlo was far from the size of the signal Monte Carlo samples.

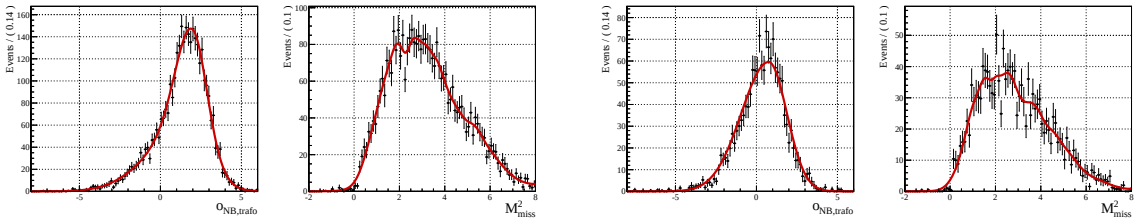


(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

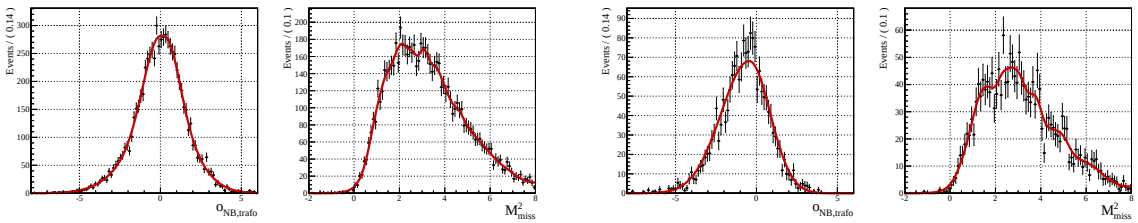


(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.31.: Lepton signal shapes determined on 10 streams of generic Monte Carlo.

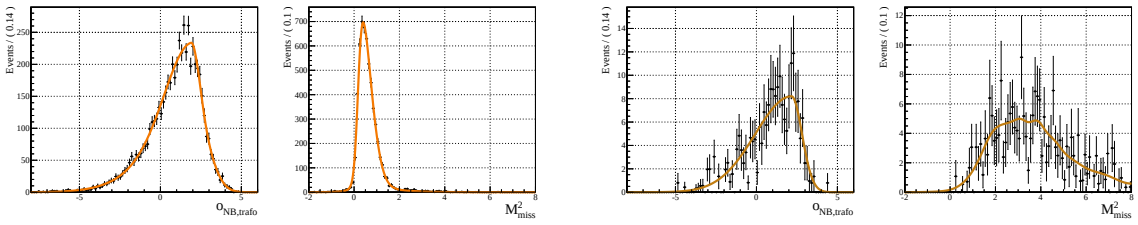


(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$



(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.32.: τ signal shapes determined on signal Monte Carlo.



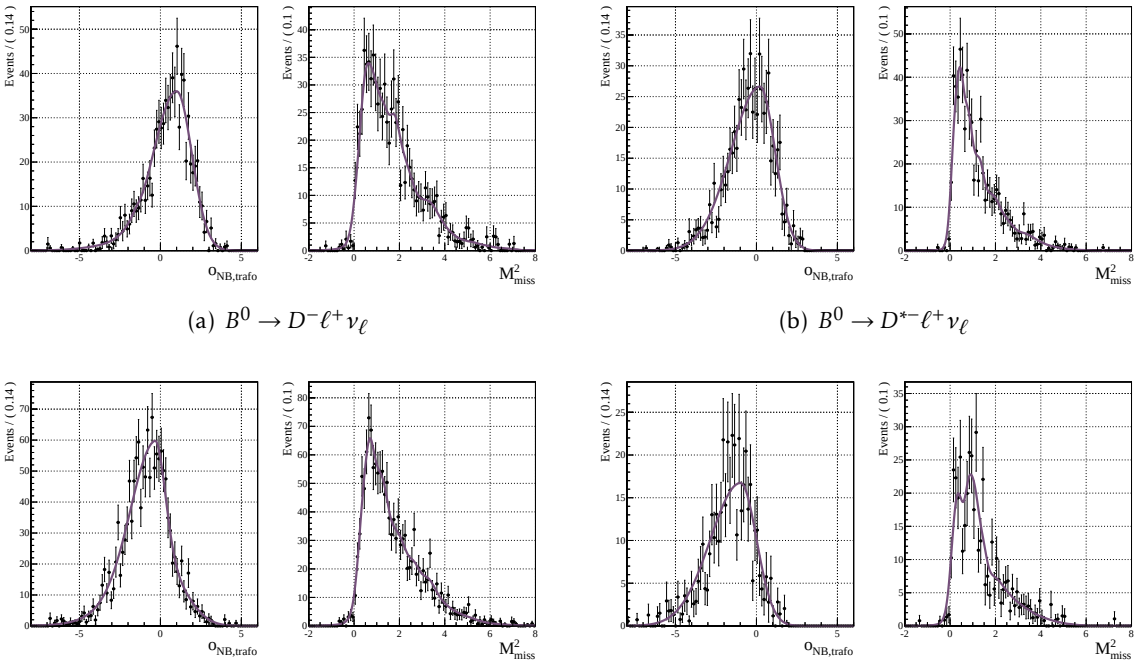
(a) $B^0 \to D^- \ell^+ \nu_\ell$ lepton cross-feed

(b) $B^0 \to D^- \ell^+ \nu_\ell$ τ cross-feed

(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ lepton cross-feed

(d) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ τ cross-feed

Figure 6.33.: Lepton and τ cross-feed shapes determined on 10 streams of generic Monte Carlo.



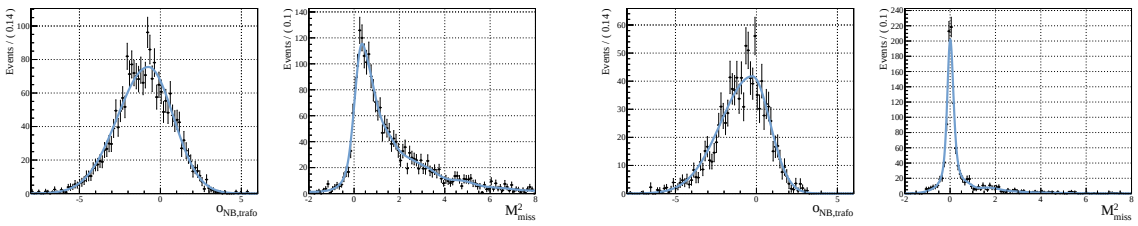
(a) $B^0 \to D^- \ell^+ \nu_\ell$

(b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

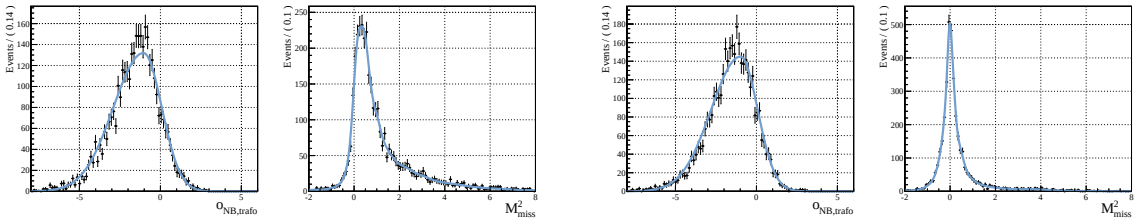
(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$

(d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.34.: D^{**} background shapes determined on 10 streams of generic Monte Carlo.

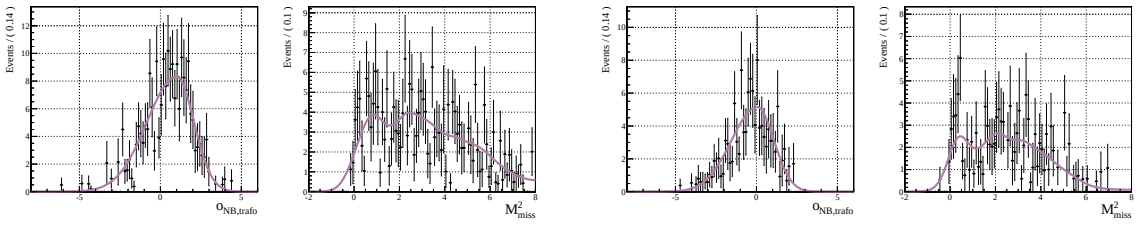


(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

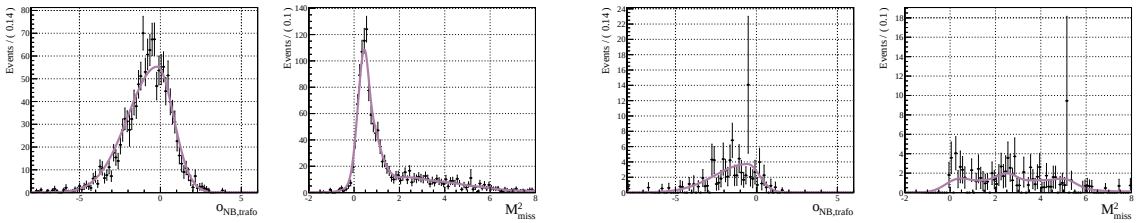


(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.35.: Wrong $D^{(*)}$ background shapes determined on 10 streams of generic Monte Carlo.

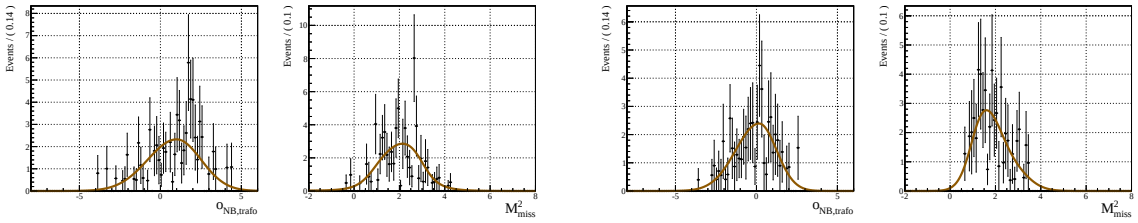


(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$



(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

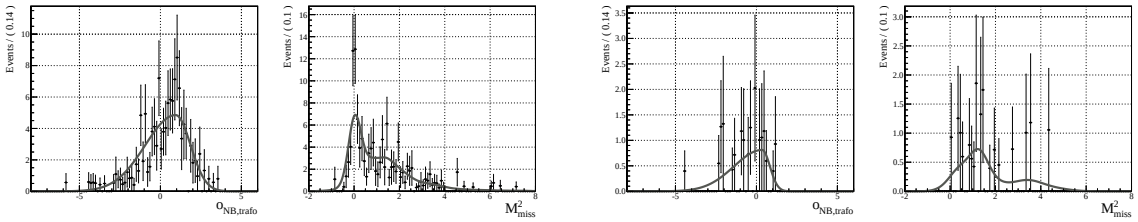
Figure 6.36.: Wrong lepton background shapes determined on 10 streams of generic Monte Carlo.



(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

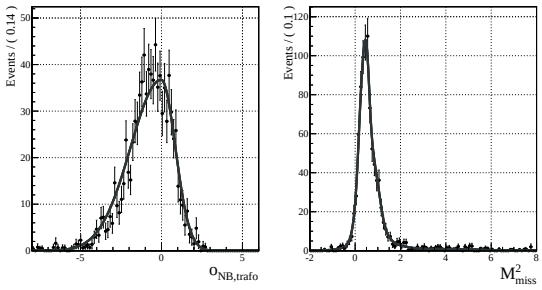
Figure 6.37.: D_S background shapes determined on 10 streams of generic Monte Carlo.



(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

(c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.38.: Remaining background shapes determined on 10 streams of generic Monte Carlo.



(a) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$

Figure 6.39.: Wrong charge cross-feed background shapes determined on 10 streams of generic Monte Carlo.

6.13.4. Illustration

Figure 6.40 tries to illustrate the rather complicated web of simultaneously fitted yields. The words in parentheses in figure 6.40 only try to clarify the *intention* behind the si-

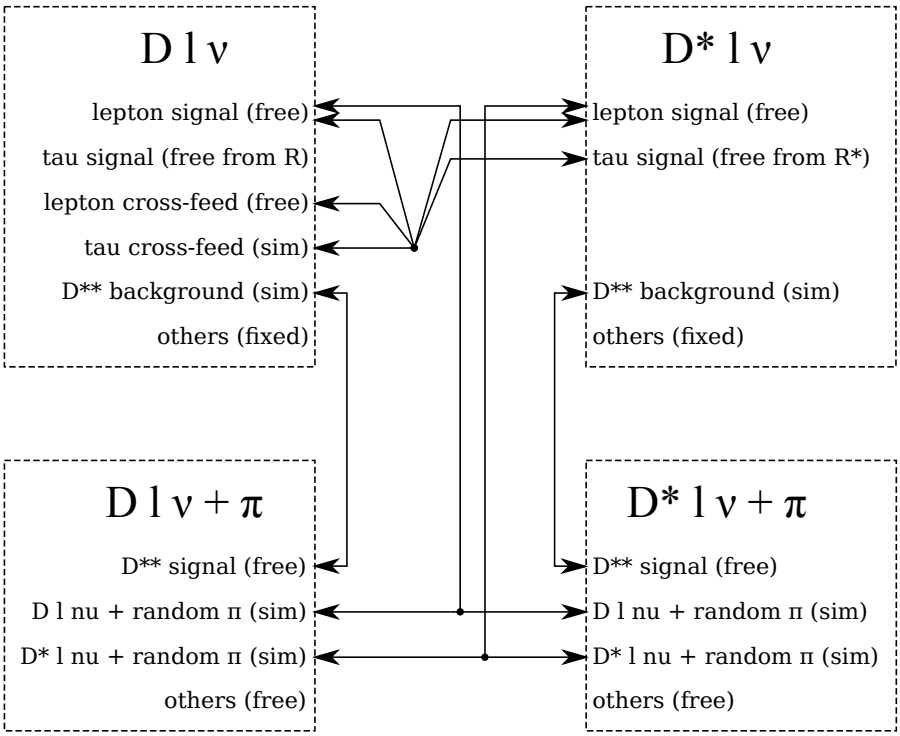


Figure 6.40.: Visualisation of the simultaneous fitting procedure.

multaneous fitting. Take, for instance, the signal component in the sample. It is labelled as "free", whereas the D^{**} background in the $B \to D\ell\nu_\ell$ sample is labelled as "sim" (simultaneously). The message is that the normalisation of the D^{**} background is constrained by the D^{**} signal component in the sample. In reality, this is a simultaneous fit, so in a manner of speaking, both component constrain each other [5].

[5] Usually, the statistically dominant component(s) in the simultaneous fits are the "free" ones

6.14. Expected Yields

Based on the 5 available complete streams of generic Monte Carlo, the expected yields of the components were calculated. They are listed in tables 6.20 and 6.21. These numbers were produced on a sample which was reweighted for tag side bias, the lepton fake rate and the D^{**} background shape.

Component	$B^0 \to D^- \ell^+ \nu_\ell$	$B^+ \to \overline{D}^0 \ell^+ \nu_\ell$
Lepton Signal	1173.93	2306.16
Tau Signal	110.15	205.03
Lep Cross-Fee	1198.43	7184.14
Tau Cross-Fee	45.13	255.02
Wrong D	495.58	820.58
D**	191.60	469.99
Wrong Lepton	44.38	122.50
Charge Cross-Feed	0.00	183.56
D_S	14.26	55.87
Rest	24.69	83.60

Table 6.20.: Yield expectations for the individual components in the $B \to D\ell\nu_\ell$ signal sample with $o_{\mathrm{NB}}(R) > 0.2$. Best candidate selection done at random.

Component	$B^0 \to D^{*-} \ell^+ \nu_\ell$	$B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$
Lepton Signal	1960.41	2557.09
Tau Signal	86.31	117.89
Wrong D*	297.13	1469.66
D**	145.45	149.17
Wrong Lepton	28.12	29.10
D_S	11.70	9.44
Rest	4.78	3.73

Table 6.21.: Yield expectations for the individual components in the $B \to D^*\ell\nu_\ell$ signal sample with $o_{\mathrm{NB}}(R^*) > 0.1$.. Best candidate selection done at random.

6.15. Toy Study

In order to check the fit model, a toy Monte Carlo study has been performed. According to the pdfs described in section 6.13, 2-dimensional datasets were generated. The shapes of the pdfs were determined on 10 streams of generic Monte Carlo. This study was performed with $o_{\mathrm{NB}}(R) > 0.2$ and $o_{\mathrm{NB}}(R^*) > 0.1$ (see section 6.18). The generated yields of the components were sampled from Poisson distributions the means of which are the expected numbers of events according to generic Monte Carlo expectations. Six samples were produced:

- The D signal sample,
- the D^* signal sample,
- the D mass sideband,
- the $\Delta M(D^*, D)$ sideband,
- the $D^{*,*}$ enriched sample from the D signal sample and
- the $D^{*,*}$ enriched sample from the D^* signal sample.

The fitting procedure described in section 6.13 was then applied and the resulting yields and errors for each component saved. A pull distribution, defined as

$$\text{pull} = \frac{\text{Yield} - \text{Expectation}}{\text{Error}} \tag{6.61}$$

was calculated.
Overall, 1000 iterations were performed and the resulting distributions for fit yields, errors and pulls can be found in figures 6.41 to 6.72. Gaussian pdfs were fitted to the distributions and the mean and width of these Gaussians are also noted on the plots. A bias in the fit would manifest itself in the pull distributions: A mean of the fitted Gaussian which would be signifcantly different from zero would indicate a bias.

6.15.1. $B^0 \to D^- \ell^+ \nu_\ell$ Sample

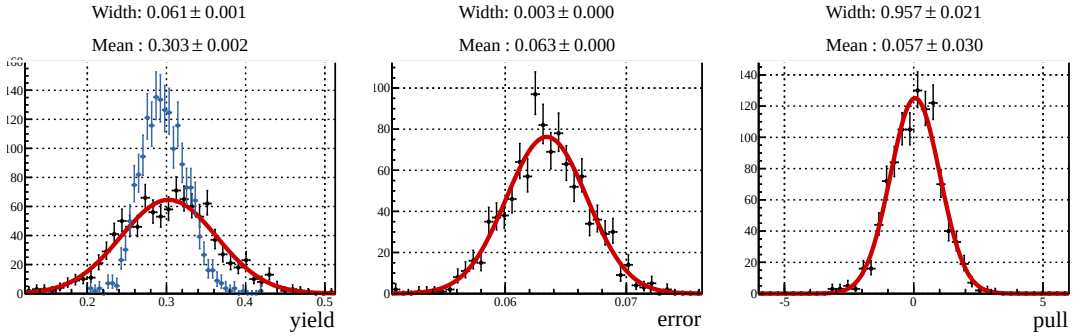


Figure 6.41.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, R. Generated values are coloured blue.

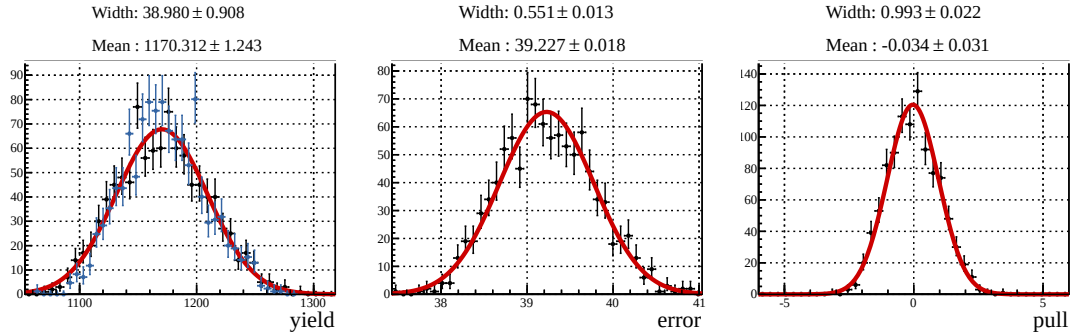


Figure 6.42.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, lepton signal. Generated values are coloured blue.

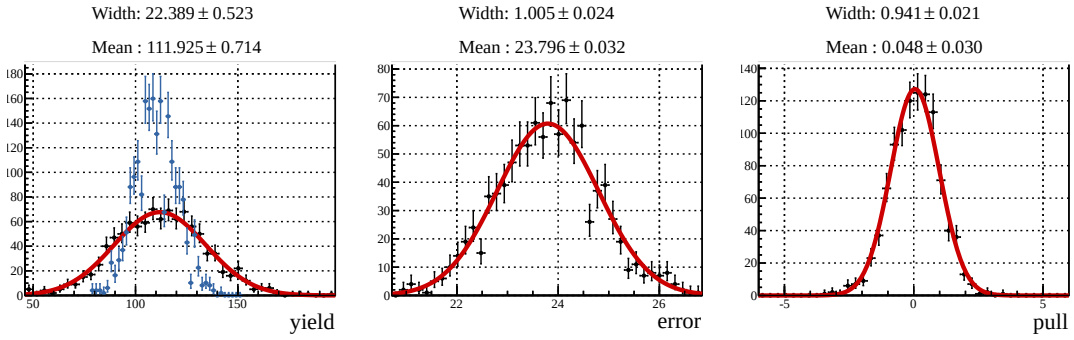


Figure 6.43.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, τ signal. This is not a free parameter of the fit.. Generated values are coloured blue.

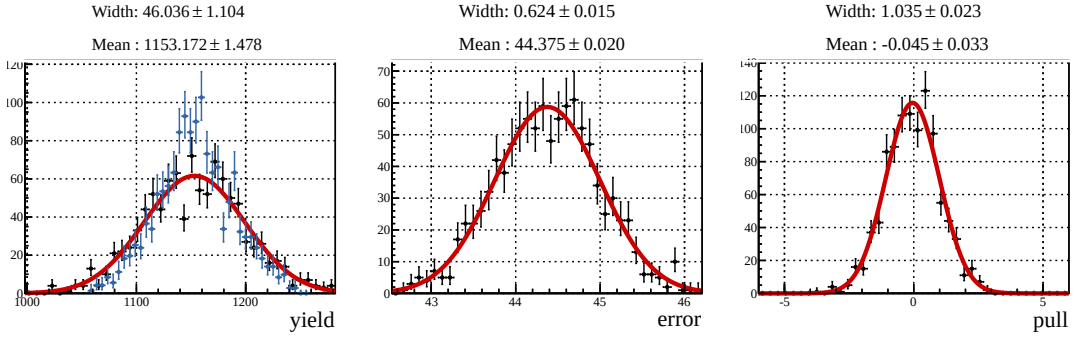


Figure 6.44.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, lepton cross-feed. Generated values are coloured blue.

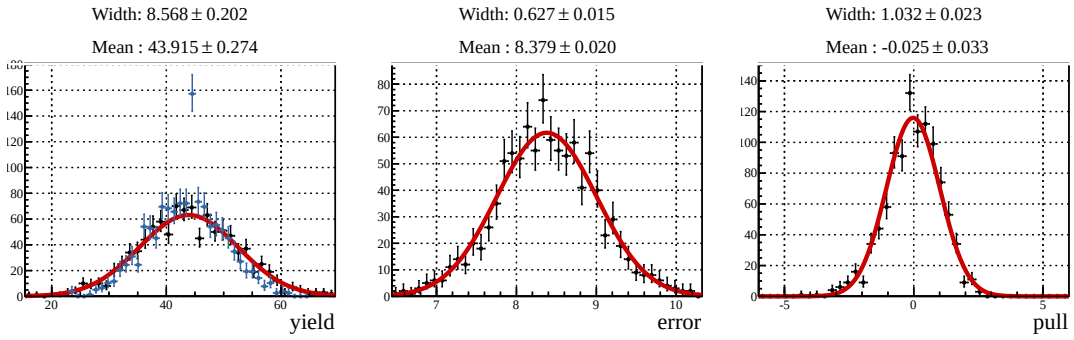


Figure 6.45.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, tau cross-feed. This is not a free parameter of the fit.. Generated values are coloured blue.

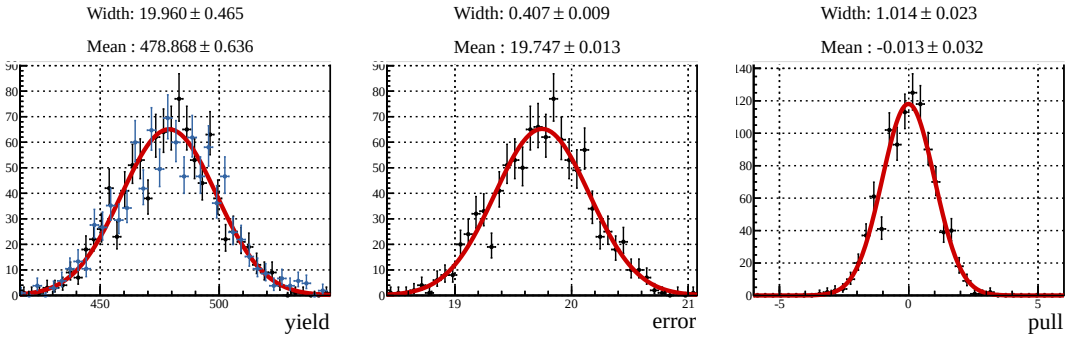


Figure 6.46.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, wrong D background. This is not a free parameter of the fit.. Generated values are coloured blue.

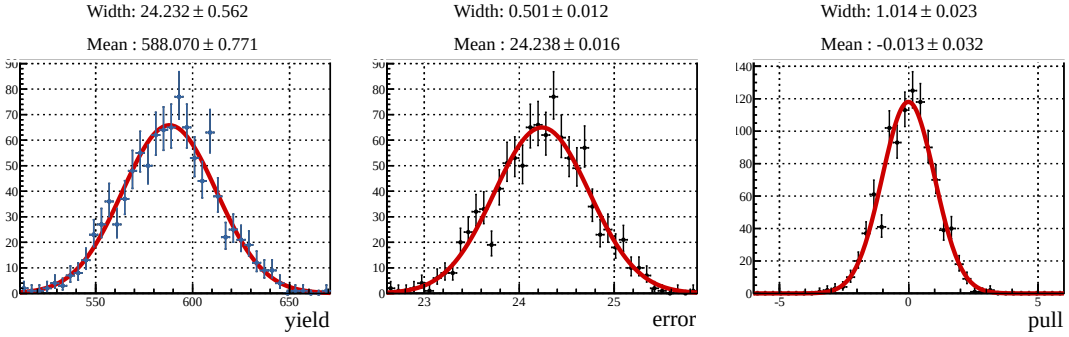


Figure 6.47.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, sideband. Generated values are coloured blue.

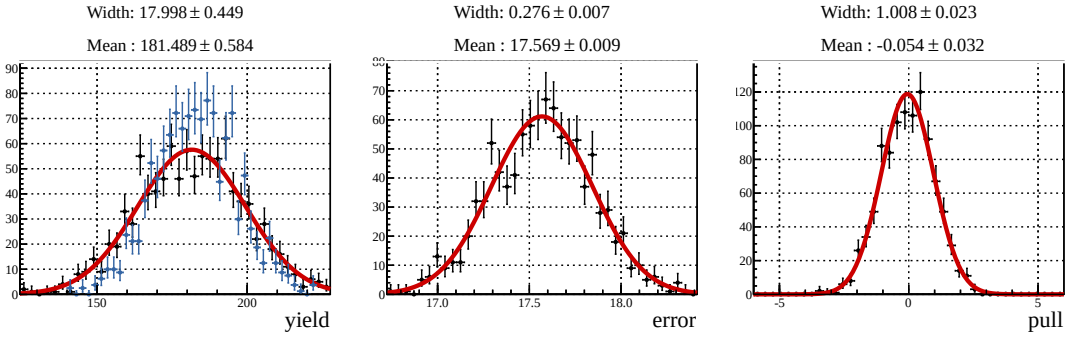


Figure 6.48.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell$ sample, D^{**} background. This is not a free parameter of the fit.. Generated values are coloured blue.

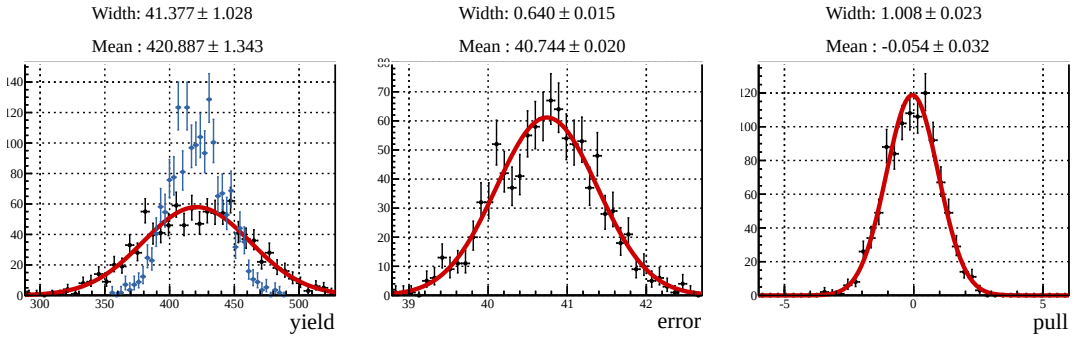


Figure 6.49.: Toy Monte Carlo study, $B^0 \to D^- \ell^+ \nu_\ell \pi^0$ sample, D^{**} signal. Generated values are coloured blue.

6.15.2. $B^0 \to D^{*-}\ell^+\nu_\ell$ Sample

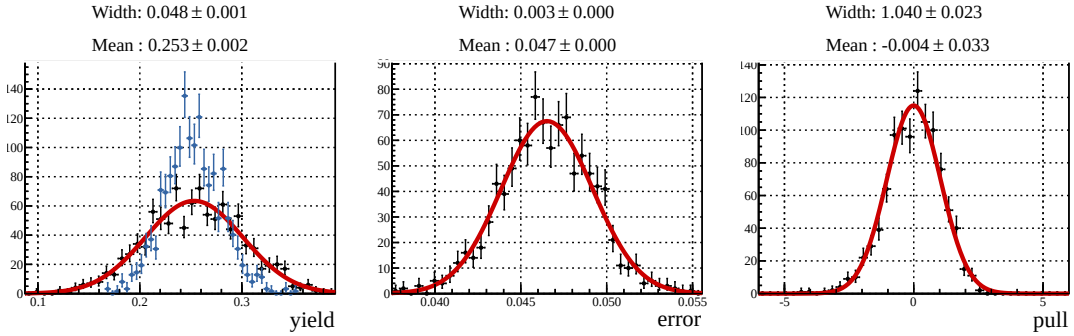


Figure 6.50.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell$ sample, R^*. Generated values are coloured blue.

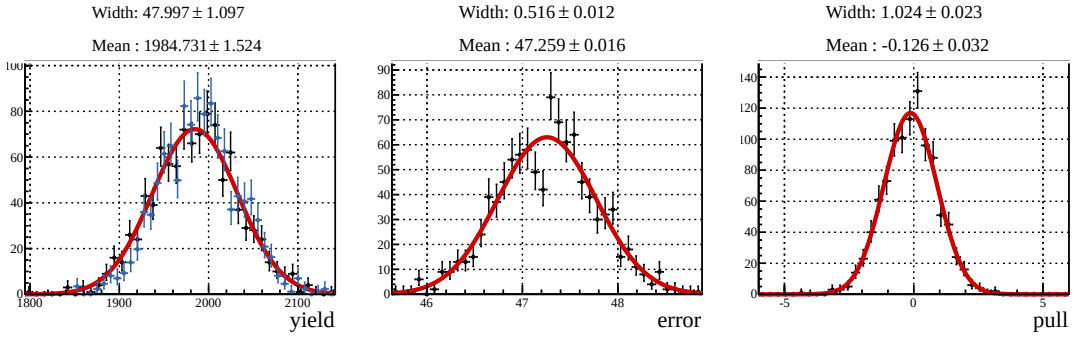


Figure 6.51.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell$ sample, lepton signal. Generated values are coloured blue.

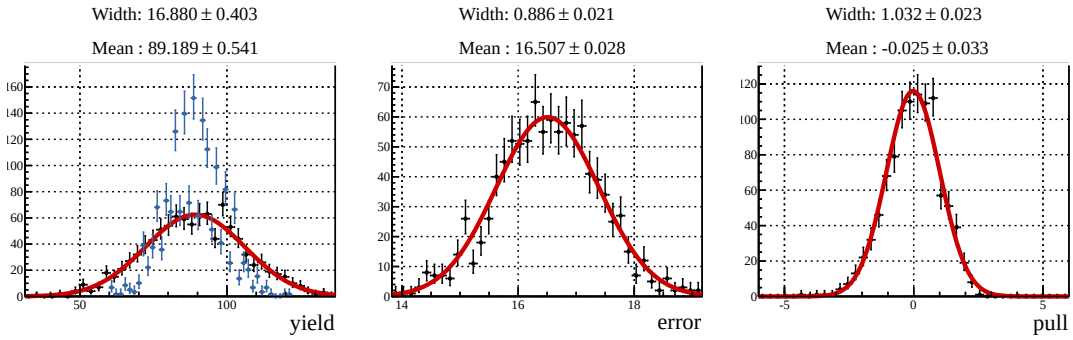


Figure 6.52.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell$ sample, τ signal. This is not a free parameter of the fit.. Generated values are coloured blue.

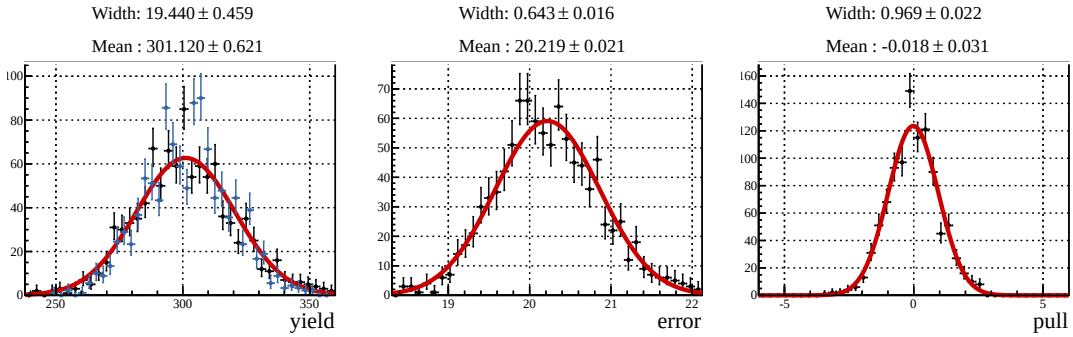


Figure 6.53.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell$ sample, wrong D^* background. This is not a free parameter of the fit.. Generated values are coloured blue.

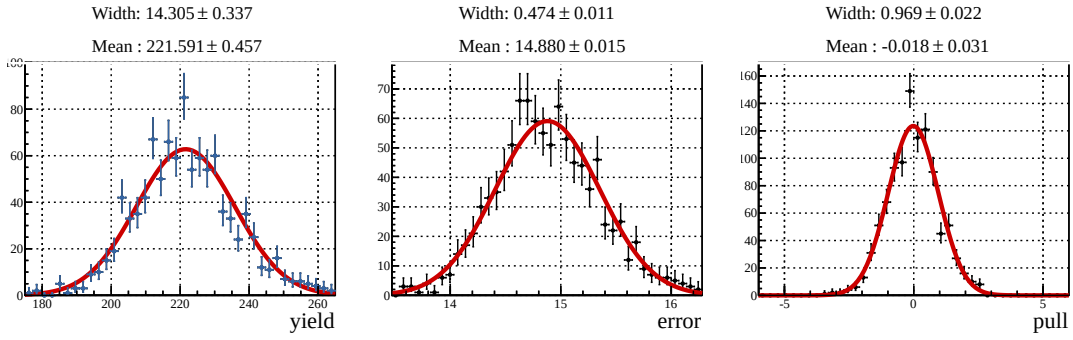


Figure 6.54.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell$ sample, sideband. Generated values are coloured blue.

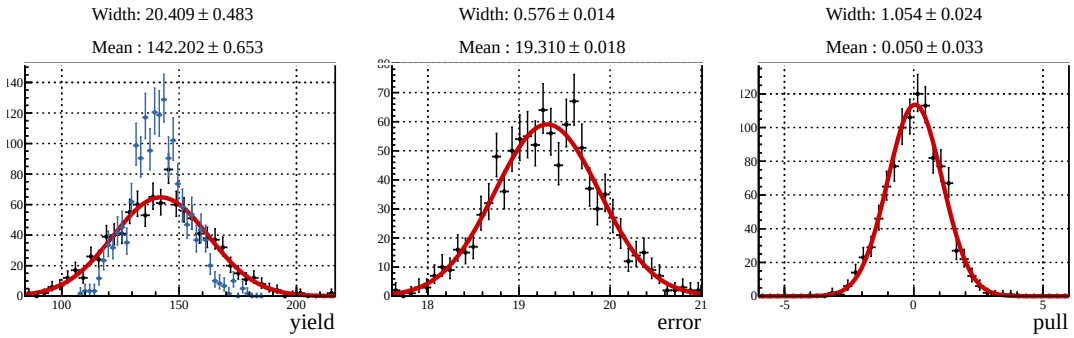


Figure 6.55.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell$ sample, D^{**} background. This is not a free parameter of the fit.. Generated values are coloured blue.

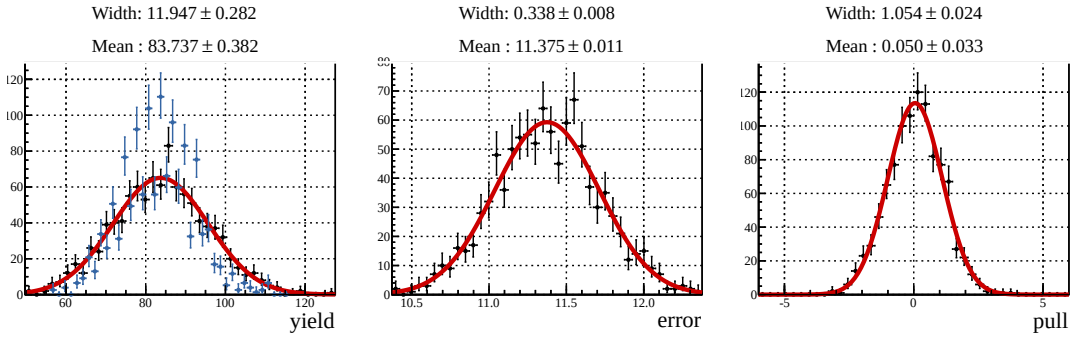


Figure 6.56.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell\pi^0$ sample, D^{**} signal. Generated values are coloured blue.

6.15.3. $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ Sample

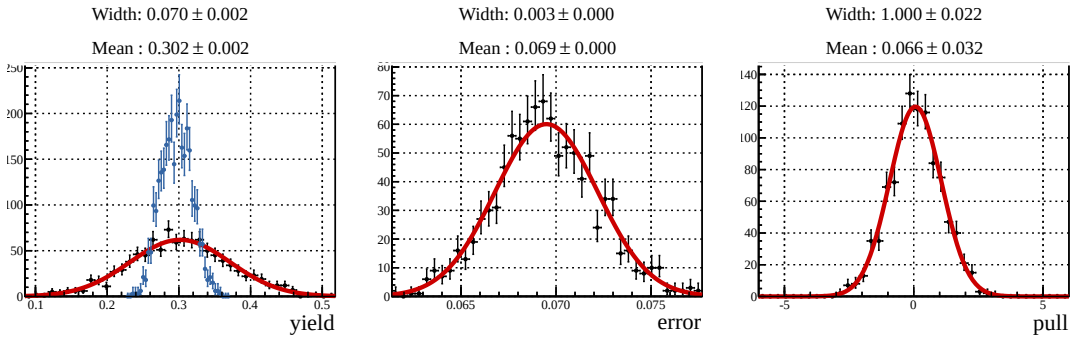


Figure 6.57.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, R. Generated values are coloured blue.

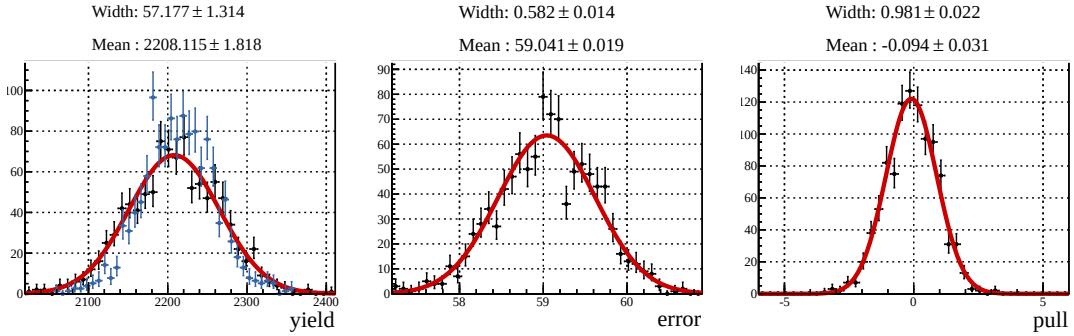


Figure 6.58.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, lepton signal. Generated values are coloured blue.

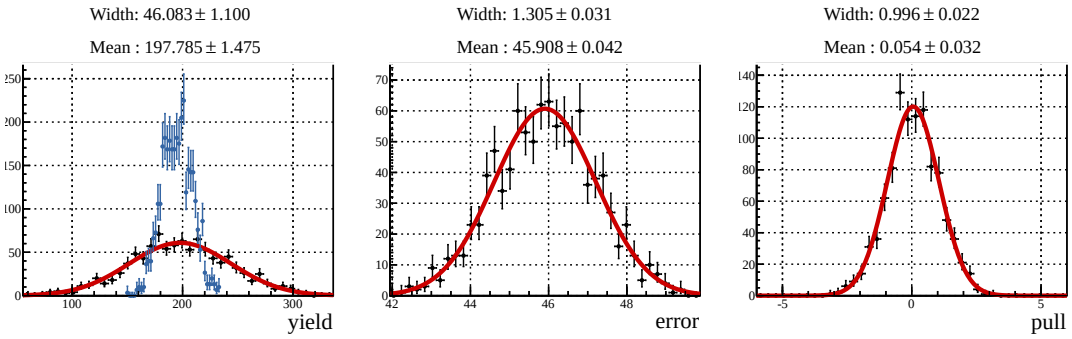


Figure 6.59.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, τ signal. This is not a free parameter of the fit.. Generated values are coloured blue.

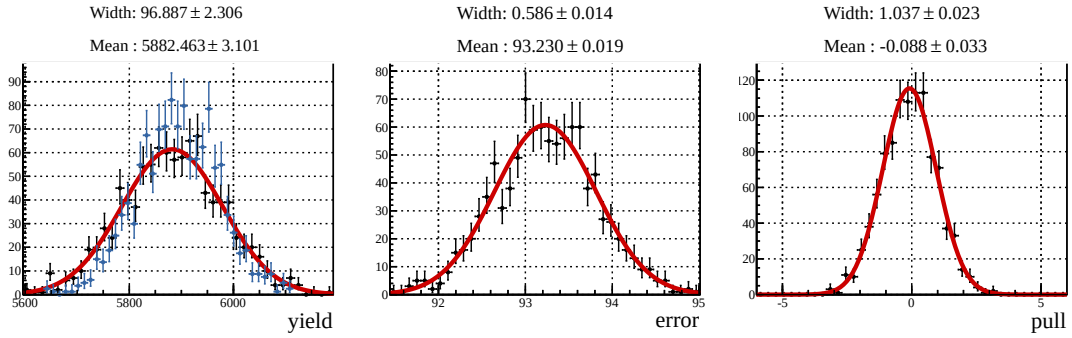


Figure 6.60.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, lepton cross-feed. Generated values are coloured blue.

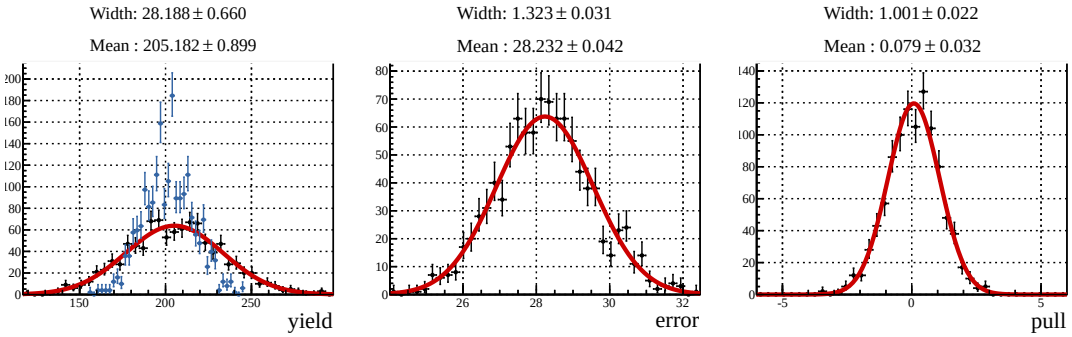


Figure 6.61.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, tau cross-feed. This is not a free parameter of the fit.. Generated values are coloured blue.

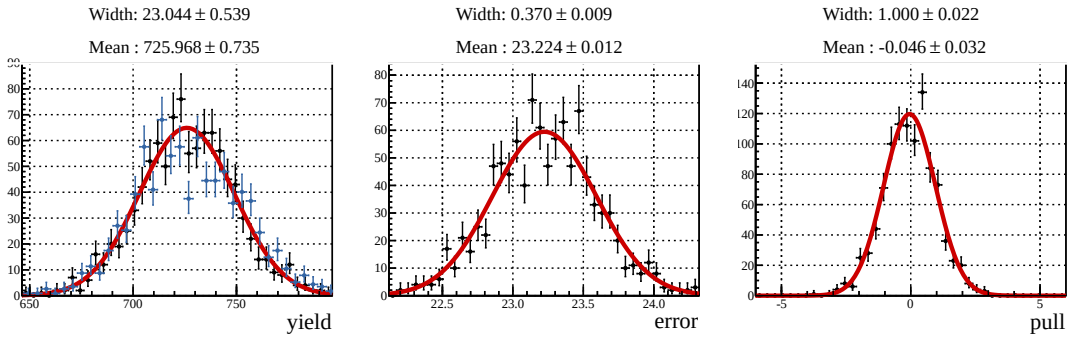


Figure 6.62.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, wrong D background. This is not a free parameter of the fit.. Generated values are coloured blue.

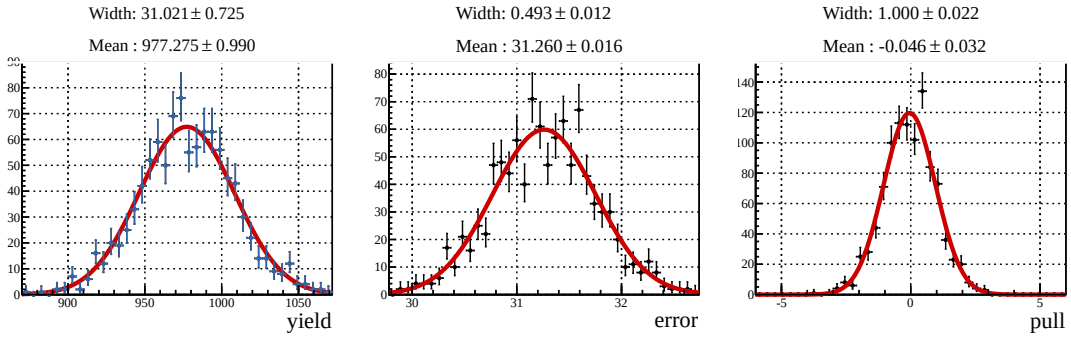


Figure 6.63.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, sideband. Generated values are coloured blue.

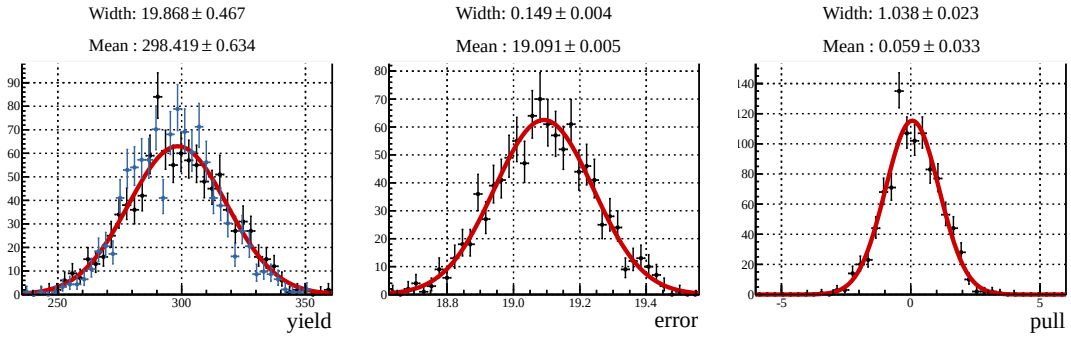


Figure 6.64.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ sample, D^{**} background. This is not a free parameter of the fit.. Generated values are coloured blue.

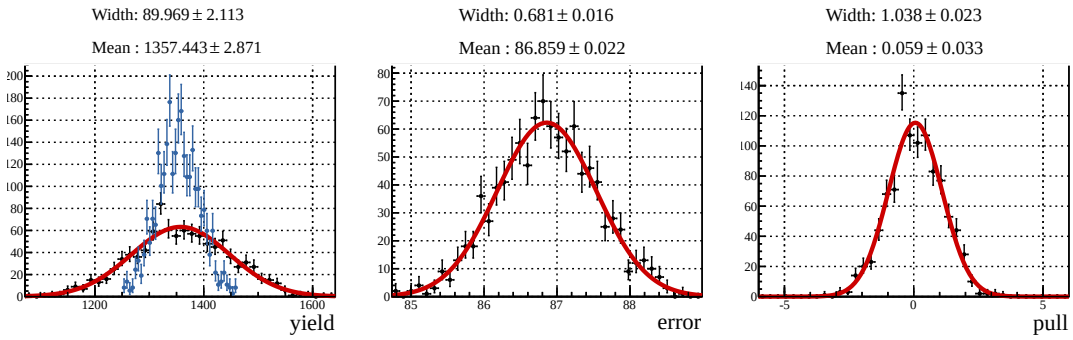


Figure 6.65.: Toy Monte Carlo study, $B^+ \to \overline{D}^0 \ell^+ \nu_\ell \pi^0$ sample, D^{**} signal. Generated values are coloured blue.

6.15.4. $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ Sample

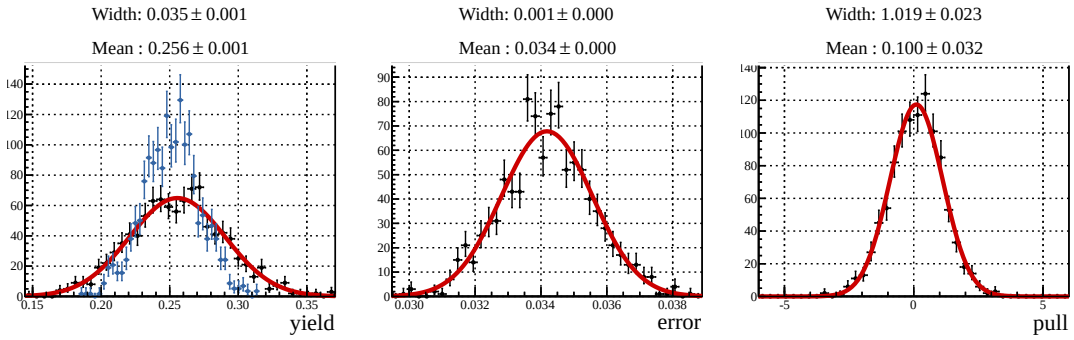


Figure 6.66.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ sample, R^*. Generated values are coloured blue.

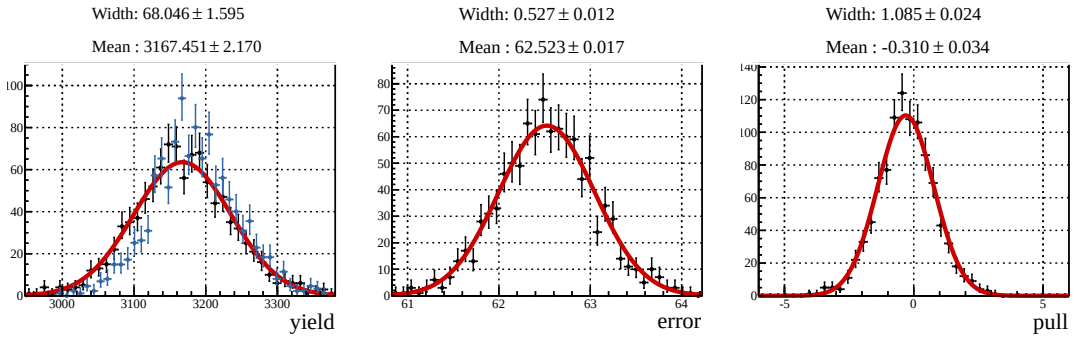


Figure 6.67.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ sample, lepton signal. Generated values are coloured blue.

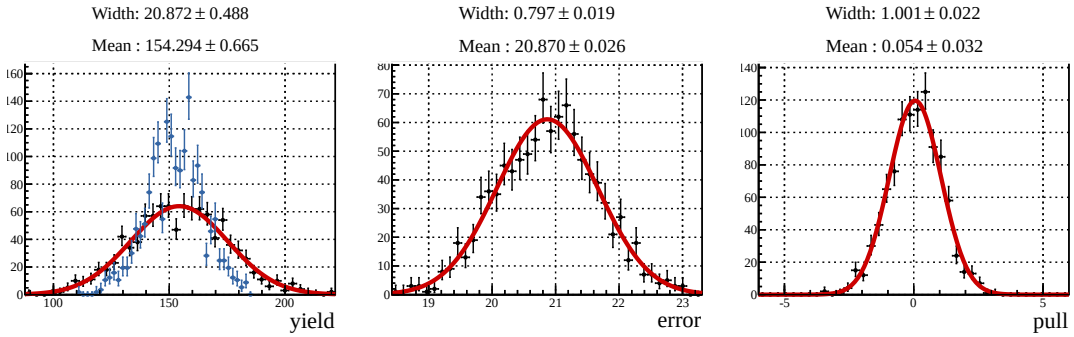


Figure 6.68.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ sample, τ signal. This is not a free parameter of the fit.. Generated values are coloured blue.

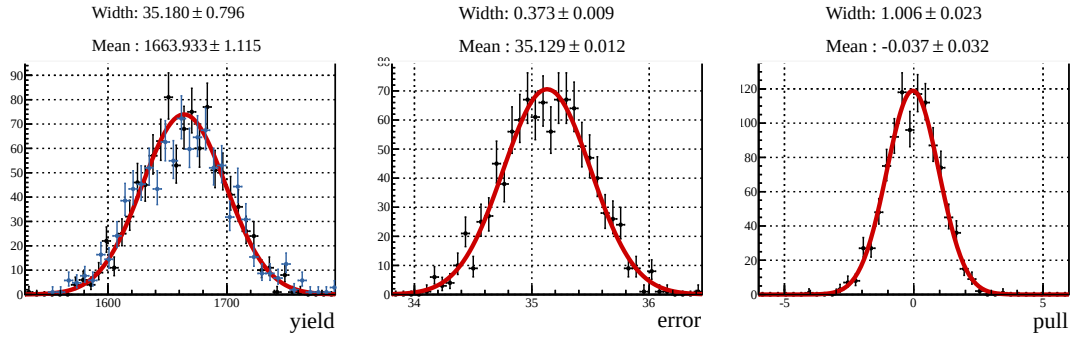


Figure 6.69.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ sample, wrong D^* background. This is not a free parameter of the fit.. Generated values are coloured blue.

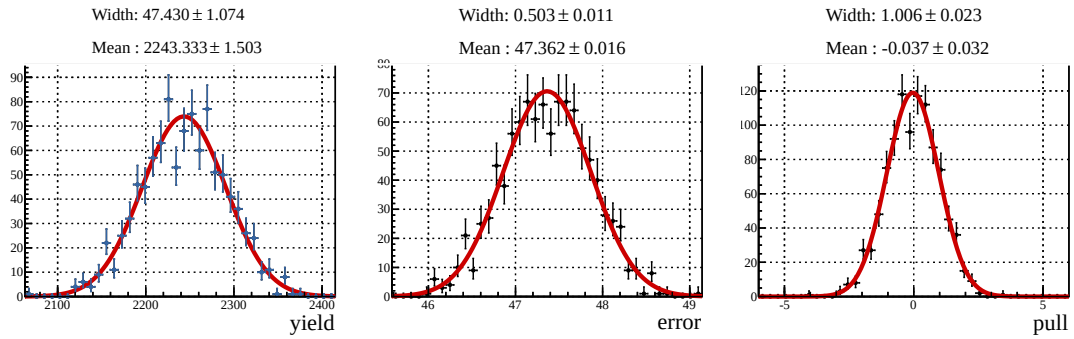


Figure 6.70.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ sample, sideband. Generated values are coloured blue.

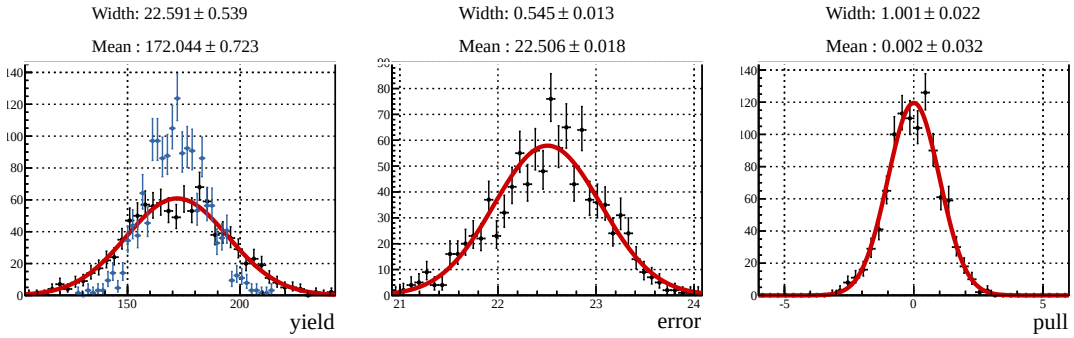


Figure 6.71.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ sample, D^{**} background. This is not a free parameter of the fit.. Generated values are coloured blue.

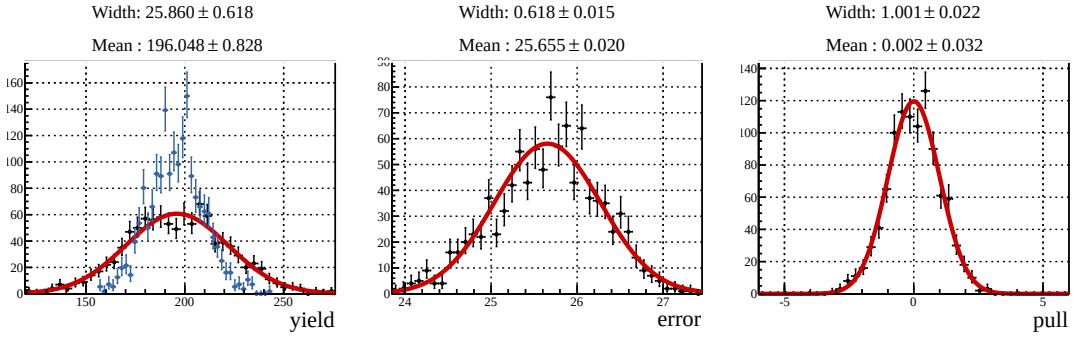


Figure 6.72.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell\pi^0$ sample, D^{**} signal. Generated values are coloured blue.

6.15.5. Conclusion

The pull distributions for R and R^* are all well-compatible with a Gaussian of mean 0.0 and width 1.0 . The lepton signal component can also be extracted correctly in $B \to D\ell\nu_\ell$ channels. The lepton signal seems to have a small bias in the $B \to D^*\ell\nu_\ell$ channels. Let us for instance consider channel $B^+ \to \overline{D}^{*0} e^+ \nu_e$ (figure 6.67). The mean of the pull is shifted to -0.310. From equation 6.61, we see that this means that the extracted yield is on average 31% of the extracted error larger than the generated value. 31% of a mean error of 63 events leads to a bias of ~ 20 events. Compared to an overall expected yields of 3170 events, this bias of 20 events seems negligible. The correct pull distribution for R and R^* (in which the lepton signal enters) support this claim.
τ signal and cross-feed, wrong $D^{(*)}$ and D^{**} backgrounds are not free parameters of the fit and are merely shown here for clarity and completeness. It can be directly seen that wrong $D^{(*)}$ and D^{**} backgrounds are deduced from $M(D)$ and $\Delta M(D^*, D)$ sidebands and the D^{**} enriched sample, respectively. Overall, the toy Monte Carlo study did not reveal any major issues in the fitting procedure.

6.16. Corrected Expected Yields

If we take into account our current best knowledge of what to expect on real data, tables 6.20 and 6.21 in section 6.14 can be recalculated. Our knowledge of what to expect on data stems from the following sources:

1. Fitted yields of the $B \to D^{(*)}\ell\nu_\ell$ samples. See section 6.11.
2. Yields from $M(D)$ and $\Delta M(D^*D)$ sidebands of real data.
3. Fit to the D^{**} enhanced samples on real data. See section 6.12.
4. Previous measurements of R and R^* by the BaBar collaboration [56].

Table 6.22 summarises the corrections for the expected yields. The corrections of

Decay Channel	Fit X		Sideband Data		D^{**} enriched		BaBar	
	MC	Data	MC	Data	MC	Data	MC	Data
$B^0 \to D^- e^+ \nu_e$	1132	1138	570	574	186	112	0.297	0.440
$B^0 \to D^{*-} e^+ \nu_e$	1848	2012	150	216	82	30	0.251	0.332
$B^+ \to \overline{D}^0 e^+ \nu_e$	2129	2228	946	1083	424	154	0.297	0.440
$B^+ \to \overline{D}^{*0} e^+ \nu_e$	2171	1846	1014	1254	189	77	0.251	0.332

Table 6.22.: Data-driven corrections to the expected yields. The first line indicates where our knowledge of data comes from.

the yields are simply the fractions $\frac{Data}{MC}$ obtained from table 6.22. If they are applied, tables 6.23 and 6.24 are obtained. Additional toy Monte Carlo studies were performed with these corrected yields. The results can be seen in figures 6.73 to 6.76.

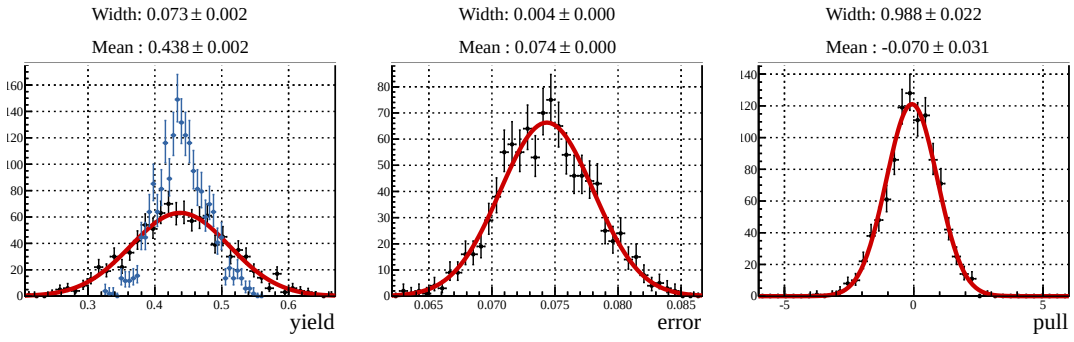


Figure 6.73.: Toy Monte Carlo study, $B^0 \to D^-\ell^+\nu_\ell$ sample, R. Generated values are coloured blue.

Component	$B^0 \to D^- \ell^+ \nu_\ell$	$B^+ \to \overline{D}^0 \ell^+ \nu_\ell$
Lepton Signal	1178.30	2318.06
Tau Signal	163.88	304.54
Lep Cross-Fee	1275.12	5143.80
Tau Cross-Fee	63.37	227.85
Wrong D	550.03	835.13
D**	114.79	159.50
Wrong Lepton	43.78	106.00
Charge Cross-Feed	0.00	181.63
D_S	14.30	49.45
Rest	24.69	75.99

Table 6.23.: Yield expectations for the individual components in the $B \to D\ell\nu_\ell$ signal sample with $o_{\mathrm{NB}}(R) > 0.2$. This table is corrected, taking our current best knowledge of the data into account.

Component	$B^0 \to D^{*-} \ell^+ \nu_\ell$	$B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$
Lepton Signal	2174.91	2715.59
Tau Signal	128.39	171.42
Wrong D*	428.18	2133.79
D**	53.36	69.48
Wrong Lepton	28.81	31.26
D_S	11.91	12.84
Rest	4.54	4.39

Table 6.24.: Yield expectations for the individual components in the $B \to D^*\ell\nu_\ell$ signal sample with $o_{\mathrm{NB}}(R^*) > 0.1$. This table is corrected, taking our current best knowledge of the data into account.

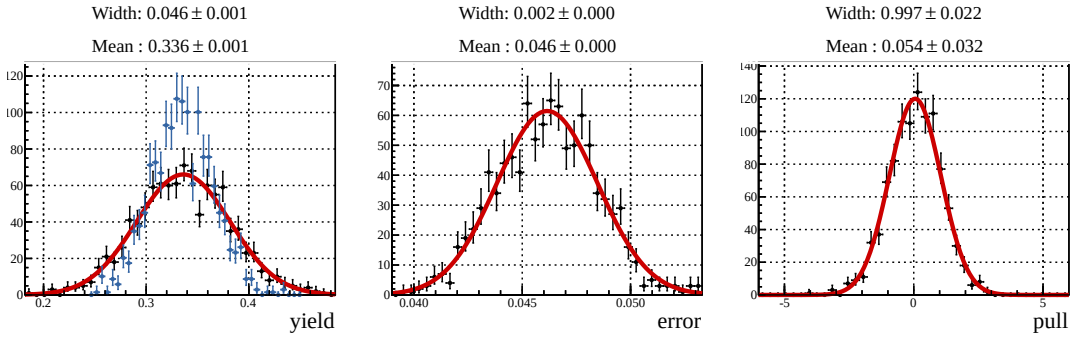


Figure 6.74.: Toy Monte Carlo study, $B^0 \to D^{*-}\ell^+\nu_\ell$ sample, R^*. Generated values are coloured blue.

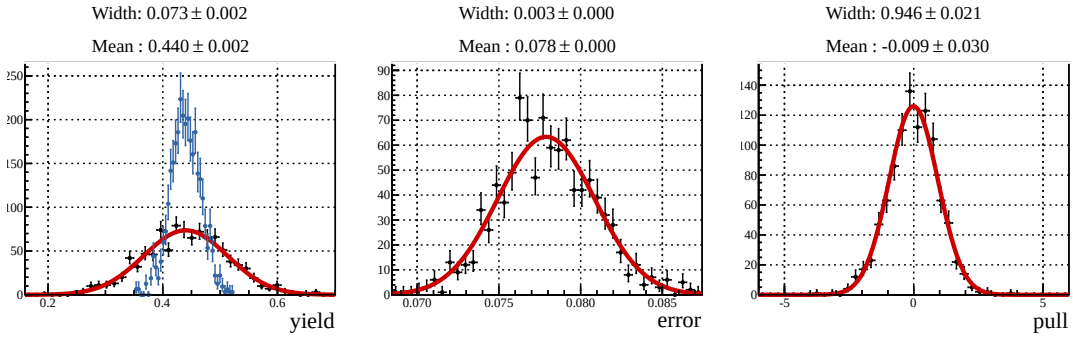


Figure 6.75.: Toy Monte Carlo study, $B^+ \to \overline{D}^0\ell^+\nu_\ell$ sample, R. Generated values are coloured blue.

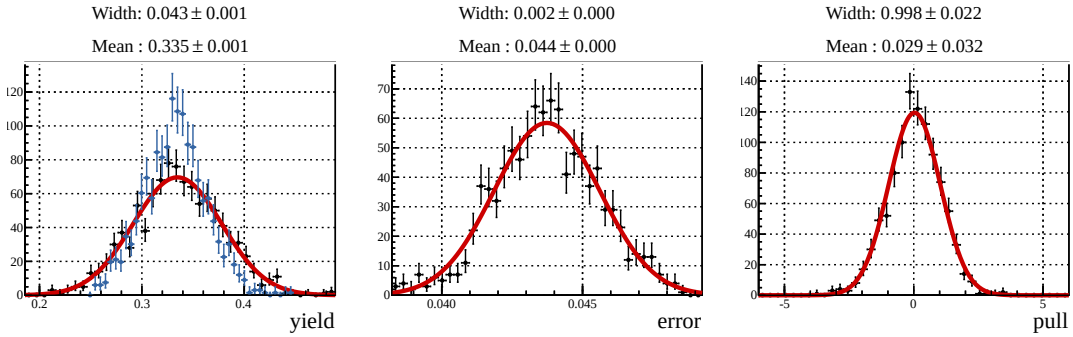


Figure 6.76.: Toy Monte Carlo study, $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$ sample, R^*. Generated values are coloured blue.

6.17. Expected Statistical Uncertainty

An important property of this analysis is the expected relative error on the values of R and R^*. The toy Monte Carlo studies can provide such estimates. Table 6.25 shows the expected statistical uncertainty of the measurements of R and R^*. The values are extracted by fitting Gaussian functions to the error distributions of the toy Monte Carlos. Values for the generic Monte Carlo expectation (section 6.14, figures 6.41, 6.50, 6.57 and 6.66) and for the expectation closest to data (section 6.16, figures 6.73, 6.74, 6.75 and 6.76) are given in table 6.25. For comparison, the uncertainties of the most recent measurements by the BaBar collaboration [56] are also included in table 6.25. We

	generic Monte Carlo			closest expectation			BaBar		
Channel	Value	abs	rel	Value	abs	rel	Value	abs	rel
$B^0 \to D^- e^+ \nu_e$	0.294	0.068	23%	0.438	0.074	17%	0.469	0.084	18%
$B^0 \to D^{*-} e^+ \nu_e$	0.252	0.047	19%	0.336	0.046	14%	0.355	0.039	11%
$B^+ \to \overline{D}^0 e^+ \nu_e$	0.293	0.079	27%	0.440	0.078	18%	0.429	0.082	19%
$B^+ \to \overline{D}^{*0} e^+ \nu_e$	0.253	0.036	14%	0.335	0.044	13%	0.322	0.032	10%

Table 6.25.: Expected statistical uncertainties of the measurements of R and R^*. *abs*: absolute error, *rel*: relative error.

can see that the expected relative errors are competitive with those given by the BaBar collaboration.

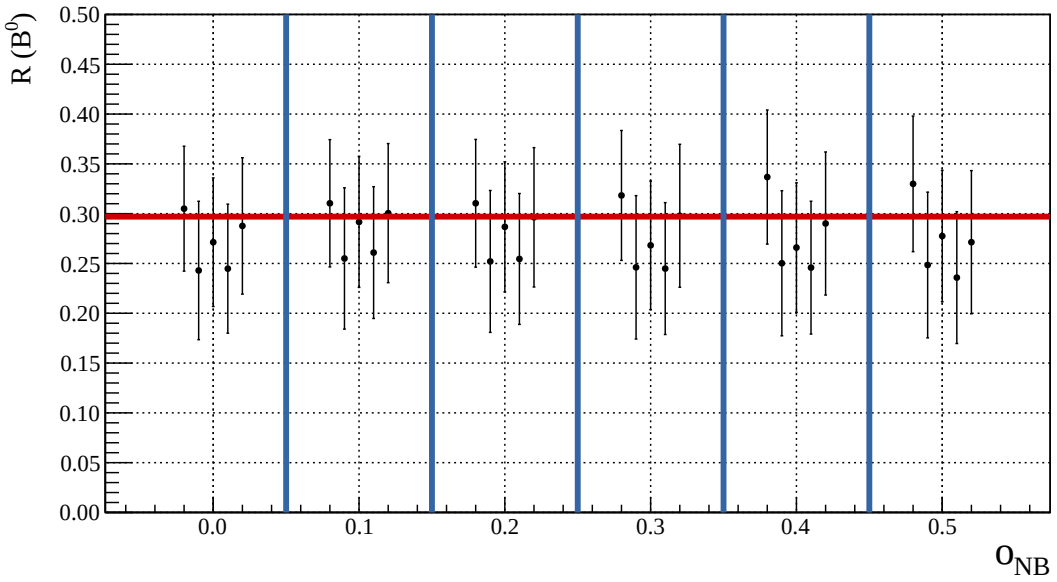
6.18. Determination Of Cut Values On Generic Monte Carlo

The extraction of R and R^* is performed via 2-dimensional extended maximum likelihood fits to $M^2_{\text{miss}}$and $o_{\text{NB, trafo.}}$. While it it possible to cut on o_{NB}before the fitting procedure, we do not expect large benefits, as the variable on which the cut would be performed (o_{NB}) is already included in the fit, albeit in a transformed way. The correlation analysis described in section 6.7.1 suggests however, that cuts of $o_{\text{NB}}(R) > 0.2$ and $o_{\text{NB}}(R^*) > 0.1$ would reduce the correlations between these two variables largely. The final fitting procedure was performed for different cut values on o_{NB} and the resulting values for R and R^* are plotted in figures 6.77 to 6.80. Cut value of $o_{\text{NB}}(R) > 0.2$ and $o_{\text{NB}}(R^*) > 0.1$ seem optimal, as they result in still relatively small statistical errors on the fit result, while the correlation is found to be insignificant. The fits were performed on the 5 complete streams of generic Monte Carlo, the shapes and factors were determined on the 9 streams of generic Monte Carlo that were not used for fitting. Table 6.26 summarises the usage of the samples.

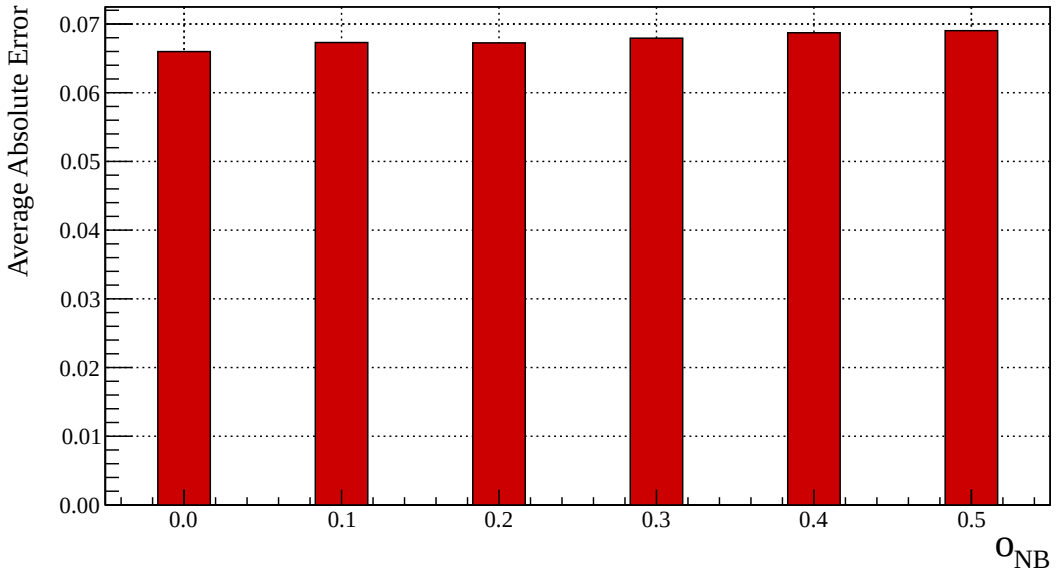
Fit on	Shapes and ratios determined on
Generic Monte Carlo, stream 0	Generic Monte Carlo, streams 1,2,3,4,5,6,7,8,9
Generic Monte Carlo, stream 1	Generic Monte Carlo, streams 0,2,3,4,5,6,7,8,9
Generic Monte Carlo, stream 2	Generic Monte Carlo, streams 0,1,3,4,5,6,7,8,9
Generic Monte Carlo, stream 3	Generic Monte Carlo, streams 0,1,2,4,5,6,7,8,9
Generic Monte Carlo, stream 4	Generic Monte Carlo, streams 0,1,2,3,5,6,7,8,9

Table 6.26.: Samples used for the fits on generic Monte Carlo.

With the cut fixed at $o_{\text{NB}}(R) > 0.2$ and $o_{\text{NB}}(R^*) > 0.1$, the results of the fits to the signal side are summarised in figure 6.81. The resulting values for R and R^* are well-compatible with the expectation, while the uncertainties are consistent with the expectations from toy Monte Carlo studies (see section 6.17). Exemplary fit projections can be found in figures 6.82 to 6.85.

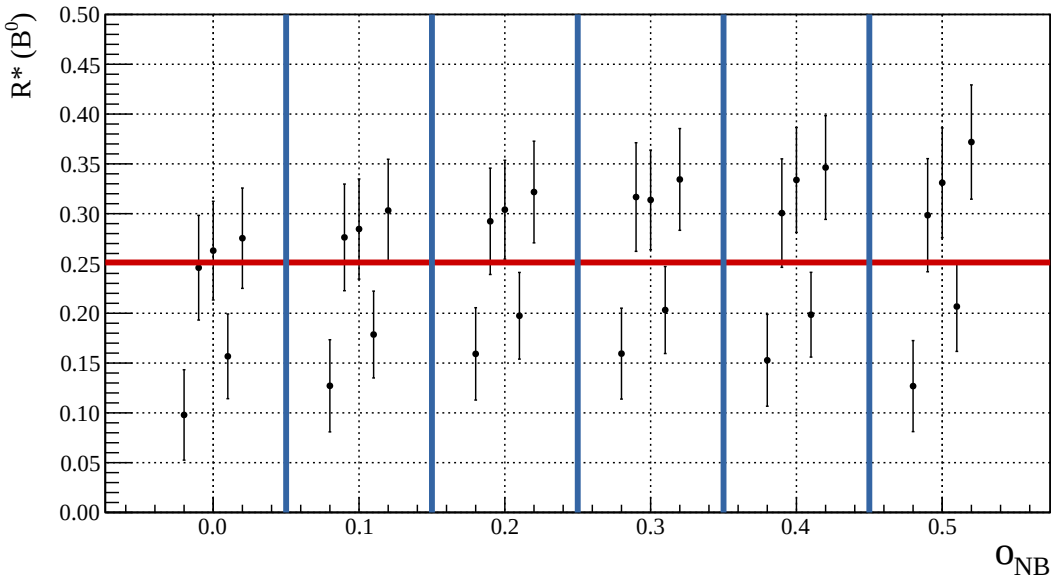


(a) The five results for each cut correspond to the 5 streams of generic Monte Carlo. The red line indicates the expected value of $R = 0.297$.

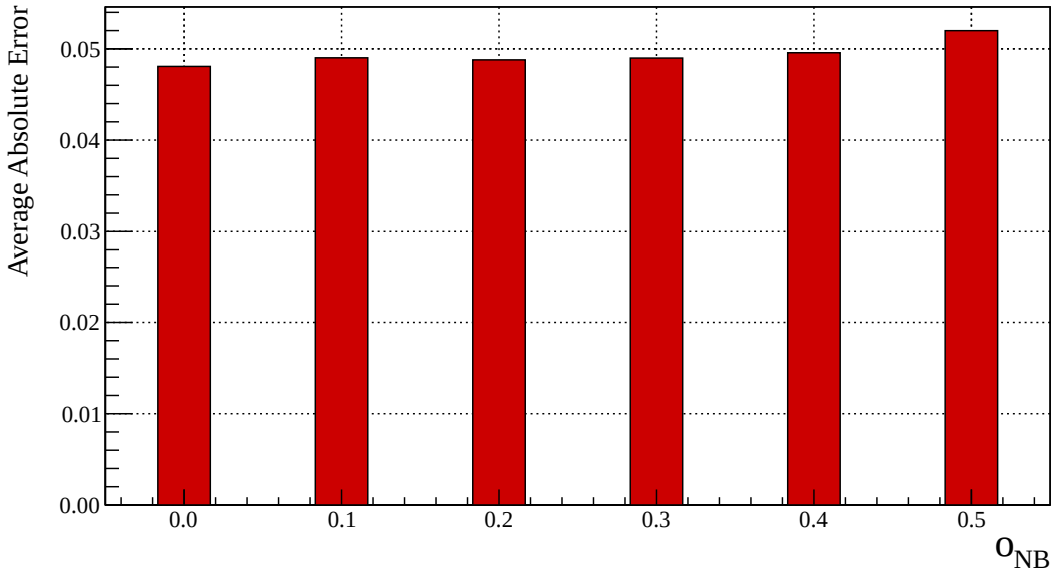


(b) Average absolute error on R

Figure 6.77.: Fit results and average absolute error on R for different cut values of o_{NB} in channel $B^0 \to D^- \ell^+ \nu_\ell$.

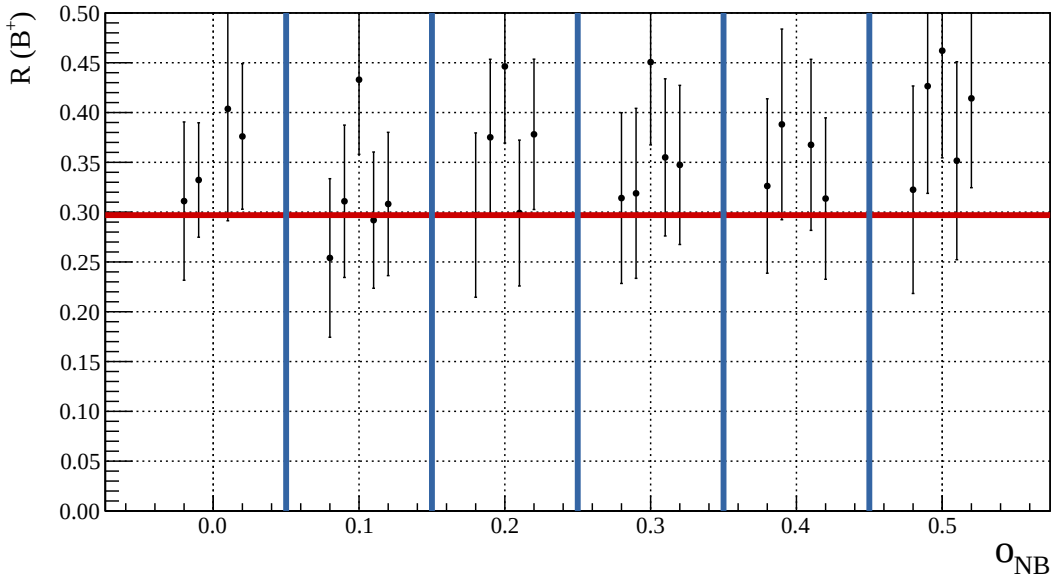


(a) The five results for each cut correspond to the 5 streams of generic Monte Carlo. The red line indicates the expected value of $R = 0.251$.

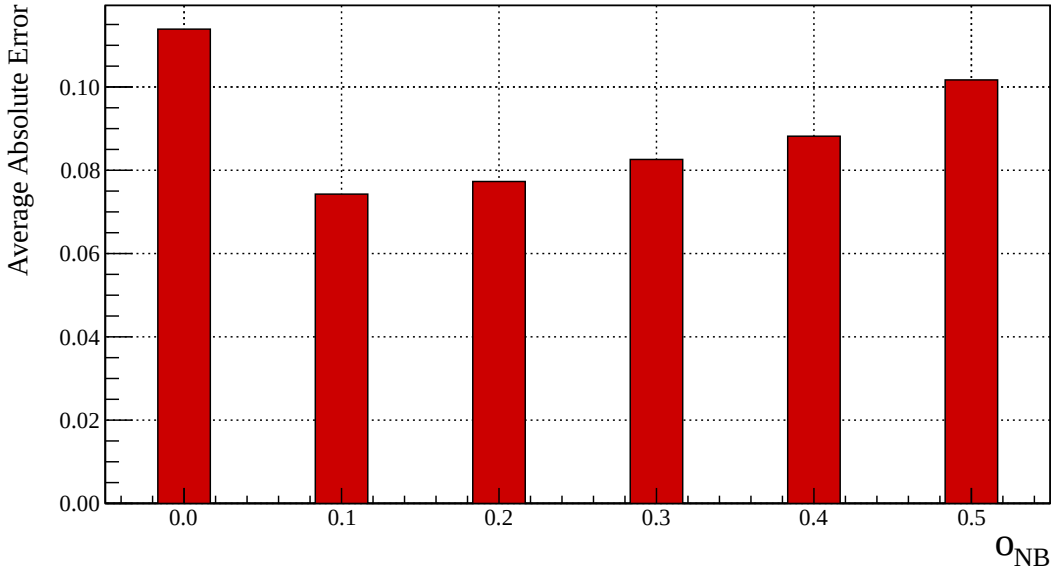


(b) Average absolute error on R^*

Figure 6.78.: Fit results and average absolute error on R for different cut values of o_{NB} in channel $B^0 \to D^{*-}\ell^+\nu_\ell$.

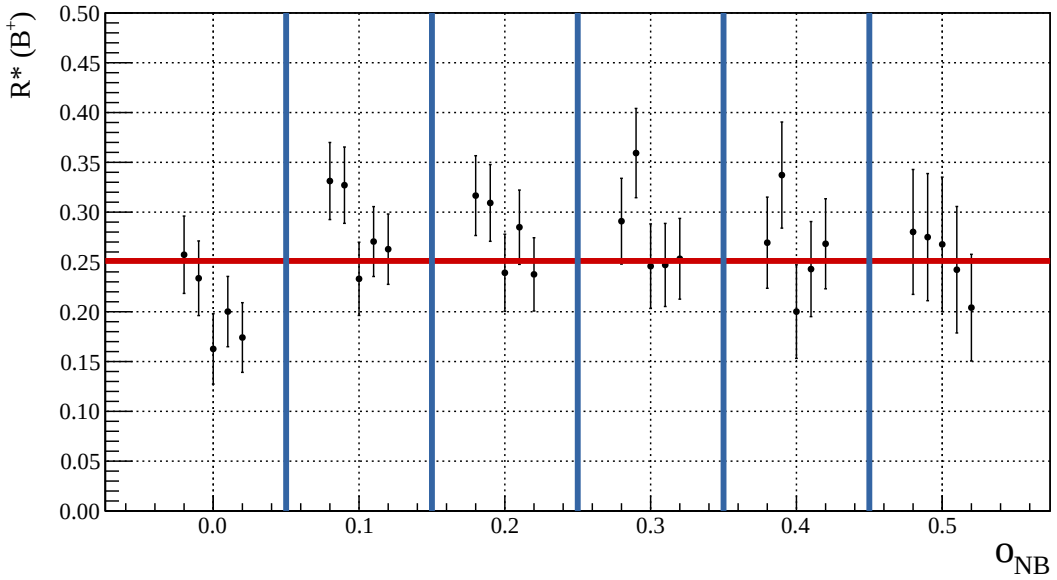


(a) The five results for each cut correspond to the 5 streams of generic Monte Carlo. The red line indicates the expected value of $R = 0.297$.

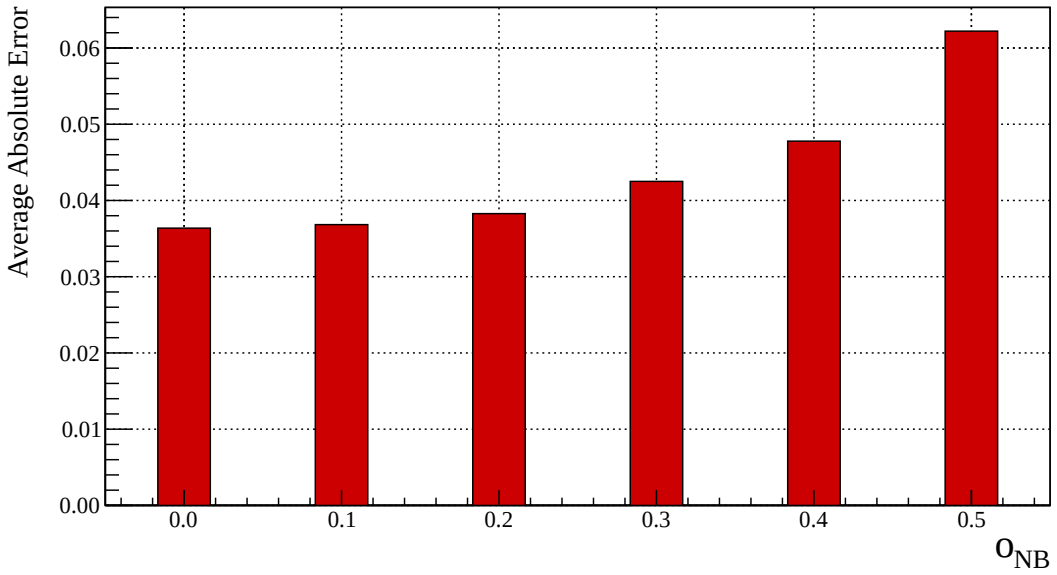


(b) Average absolute error on R

Figure 6.79.: Fit results and average absolute error on R for different cut values of o_{NB} in channel $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$.

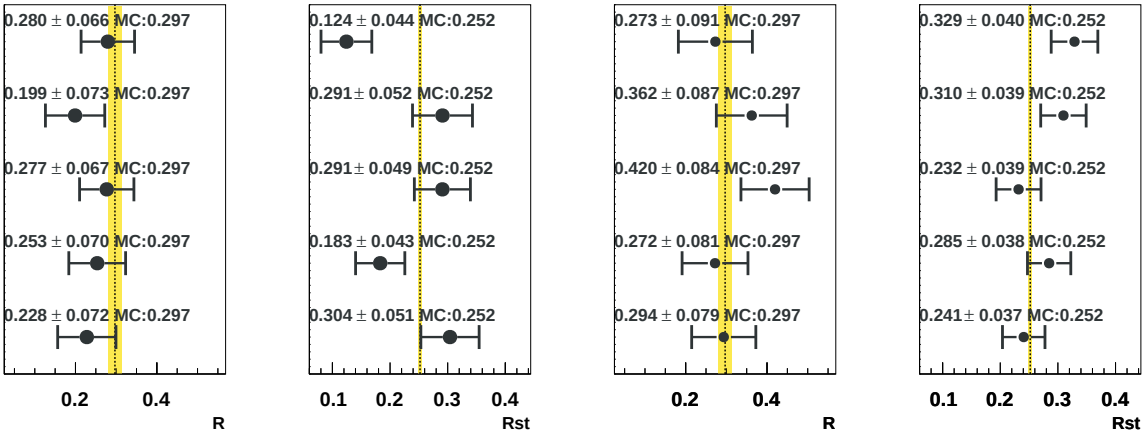


(a) The five results for each cut correspond to the 5 streams of generic Monte Carlo. The red line indicates the expected value of $R = 0.251$.



(b) Average absolute error on R^*

Figure 6.80.: Fit results and average absolute error on R for different cut values of o_{NB} in channel $B^+ \to \overline{D}^{*0}\ell^+\nu_\ell$.



(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$ (c) $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ (d) $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$

Figure 6.81.: Summary of the results of the fits to the signal samples on generic Monte Carlo. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4.

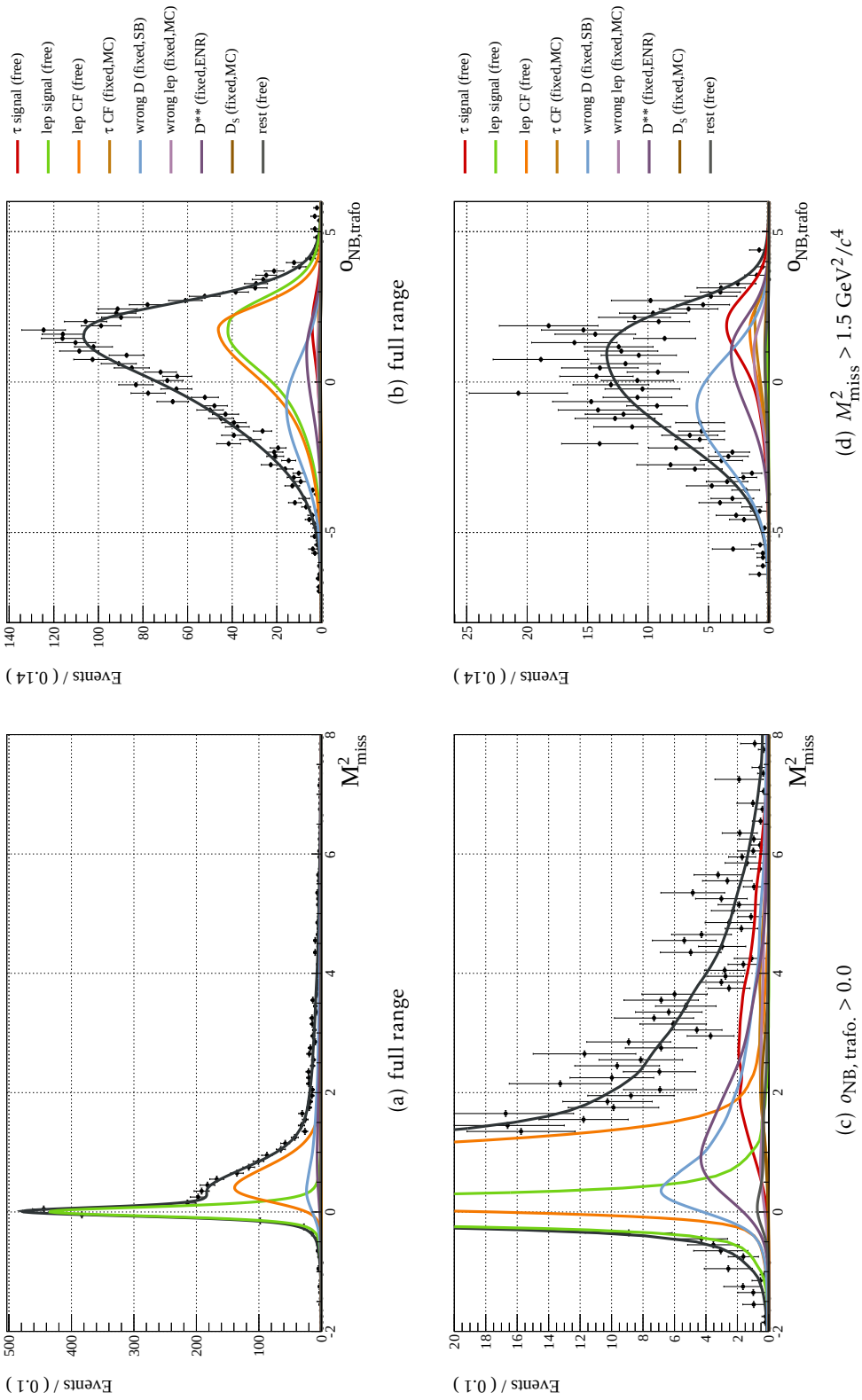
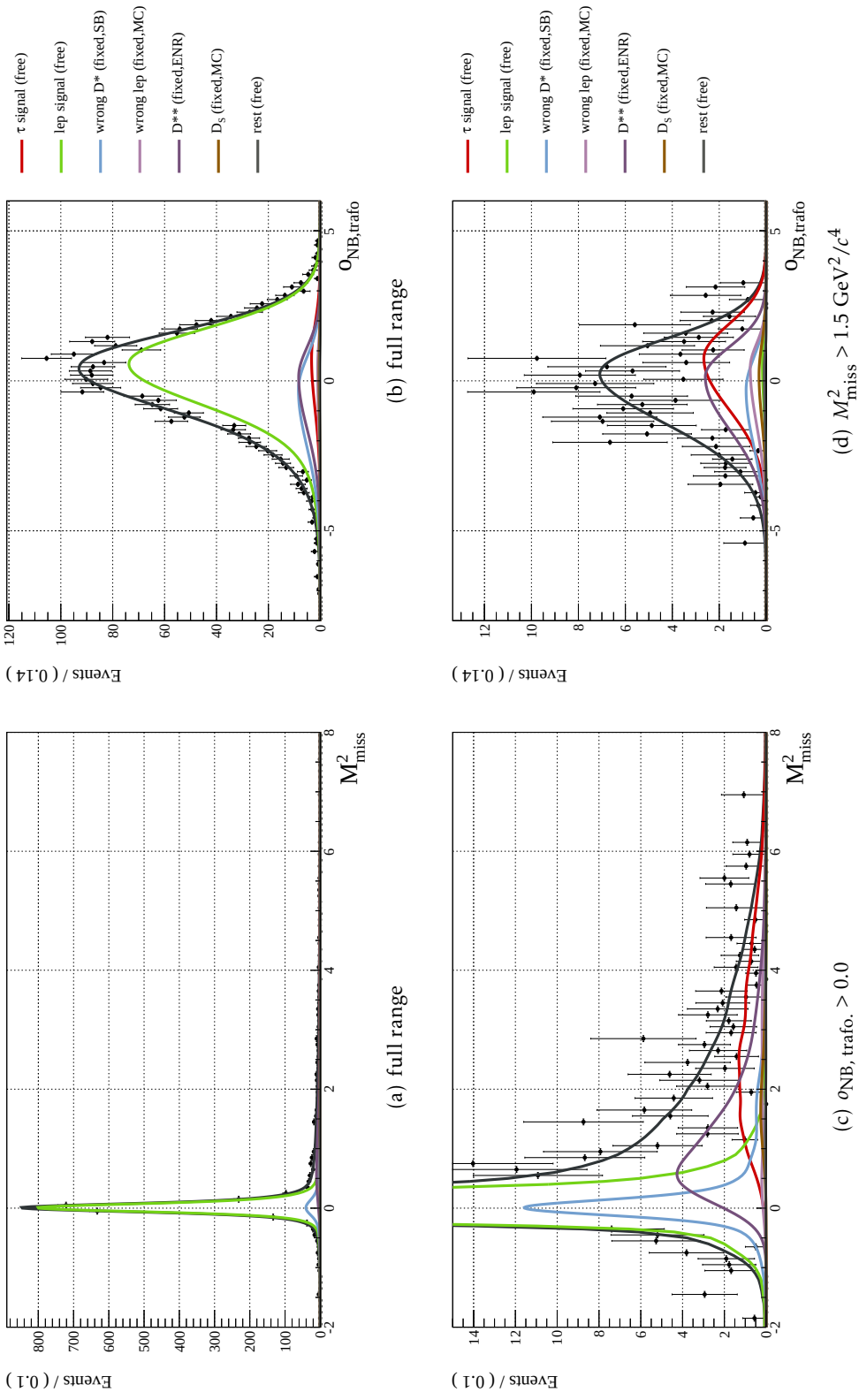


Figure 6.82.: Fit projection of $B^0 \to D^- \ell^+ \nu_\ell$ on generic Monte Carlo.



(a) full range (b) full range

(c) $o_{\text{NB, trafo.}} > 0.0$ (d) $M^2_{\text{miss}} > 1.5\ \text{GeV}^2/c^4$

Figure 6.83.: Fit projection of $B^0 \to D^{*-}\ell^+\nu_\ell$ on generic Monte Carlo.

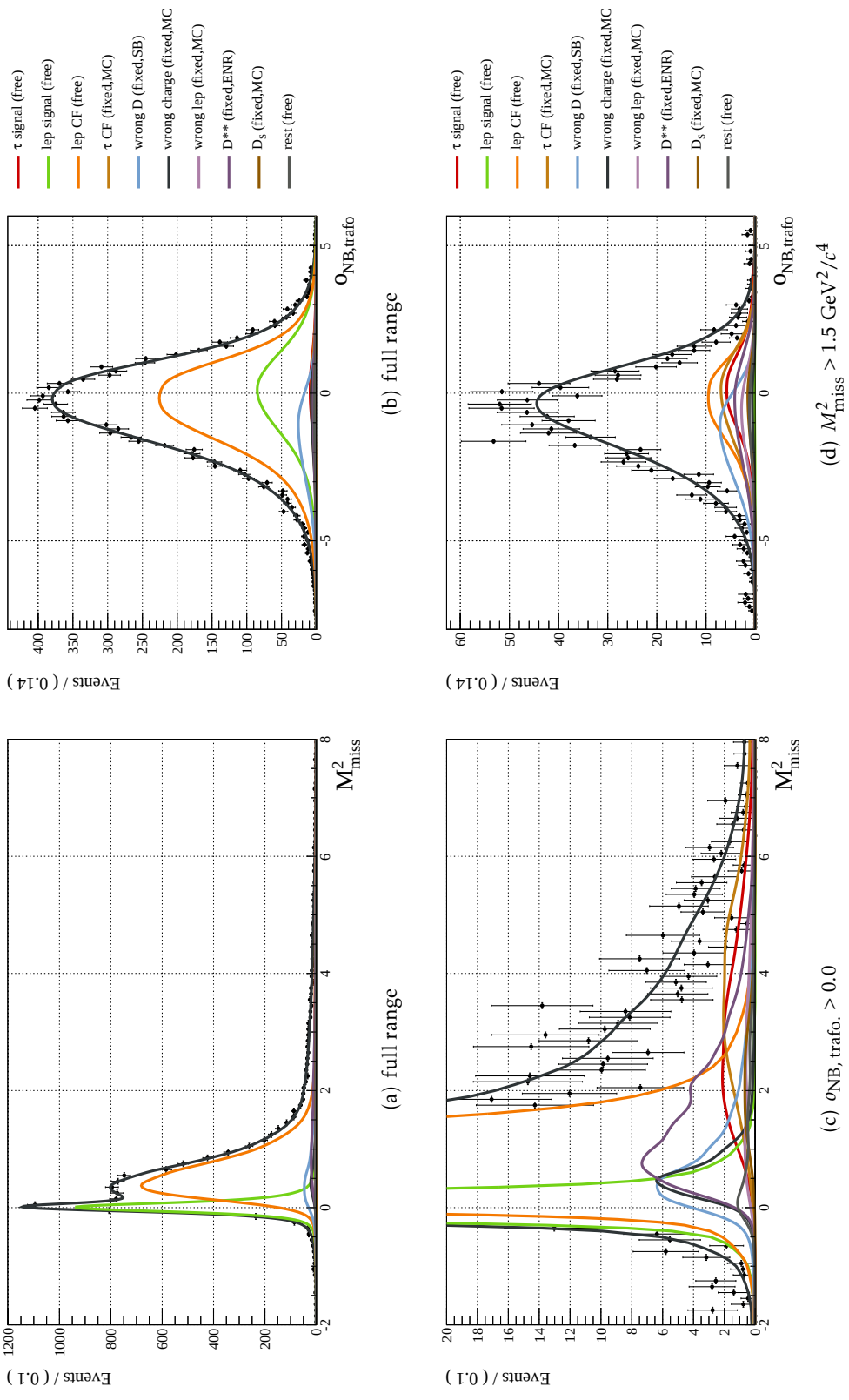


Figure 6.84.: Fit projection of $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ on generic Monte Carlo.

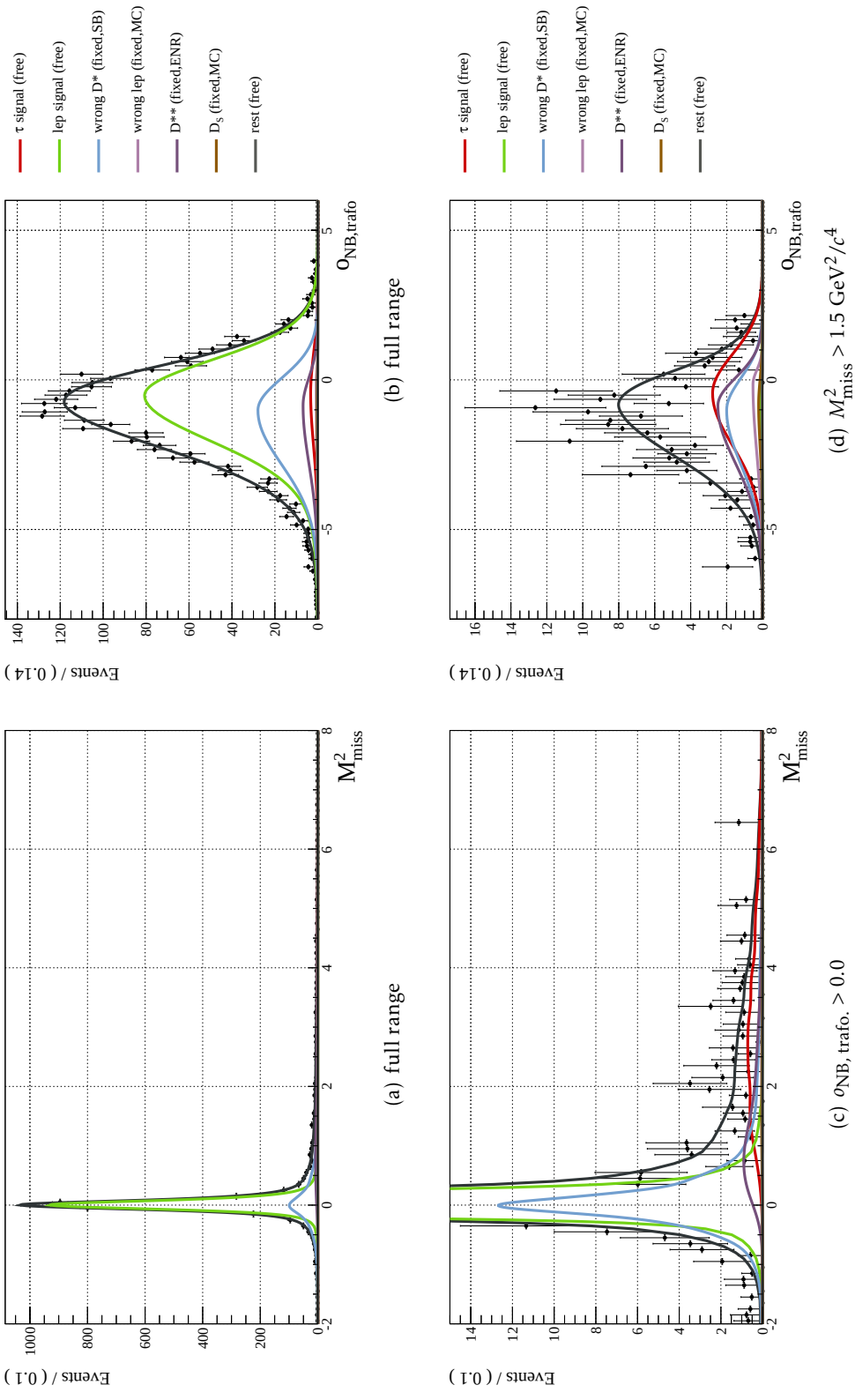


Figure 6.85.: Fit projection of $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ on generic Monte Carlo.

6.19. Simultaneous Signal Extraction Fit

The fits described in section 6.18 are well-suited to extract the $B \to D^{(*)}\tau\nu_\tau$ signal components. There are, however, disadvantages: Two different fits are performed on two different samples: One fit on the B^0 sample, one fit on the B^+ sample. This means, that two different values for R and two different values for R^* are extracted, where in reality we expect only one value each. Additionally, the treatment of the correlation between R and R^* is simplified when there is only one value each. The alternative is to fit both the B^0 and B^+ samples simultaneously, as already mentioned in section 6.13. While we have only considered the two independent fits up to this point, let us now consider the simultaneous fit to the combined B^0 and B^+ sample[6].

The actual fitting procedure is very similar to the two independent fits presented until now. There are two places, where the simultaneous fitting of the B^0 and B^+ samples is utilized:

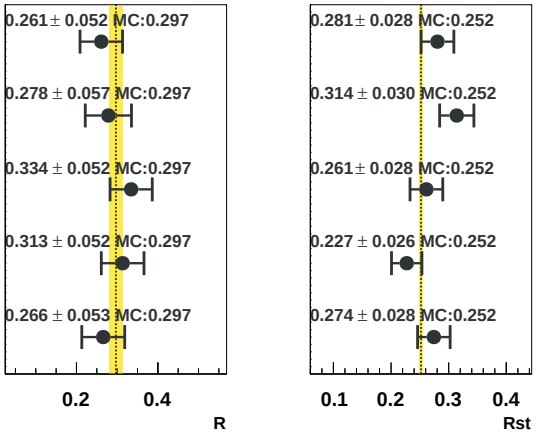
1. There is only one common value fro R and only one common value for R^* in the fit.

2. The wrong charge cross-feed in the $B^+ \to \overline{D}^0\ell^+\nu_\ell$ sample (see section 6.10) can now be constrained by using $B^0 \to D^{*-}\ell^+\nu_\ell$ lepton signal. The two yields are connected by a factor $f_{wc} = 0.091 \pm 0.003$.

Other than those two points, the fitting procedure is identical. The results of this fitting procedure applied to generic Monte Carlo are summarized in figure 6.86. Again, a single stream of generic Monte Carlo was used for fitting, while the shapes and factors (see sections 6.13.3 and 6.13.2) were determined on the remaining streams. See table 6.26 for a detailed description.

Let us compare the results from this simultaneous fitting with the results obtained from fitting two separate samples shown in figure 6.81. We see that the uncertainty of the results in figure 6.86 have the same size that would be obtained by making a weighted mean from the results in figure 6.81. This behaviour is expected from samples that are in good approximation statistically independent. Projections of the individual subsamples can be found in figures 6.87 to 6.90.

[6] Described as "double-simultaneous" fit in section 6.13



(a) $B^0 \to D^- \ell^+ \nu_\ell$ (b) $B^0 \to D^{*-} \ell^+ \nu_\ell$

Figure 6.86.: Summary of the results of the fits to the signal samples on generic Monte Carlo. The lines are from top to bottom: Generic Monte Carlo streams 0,1,2,3 and 4.

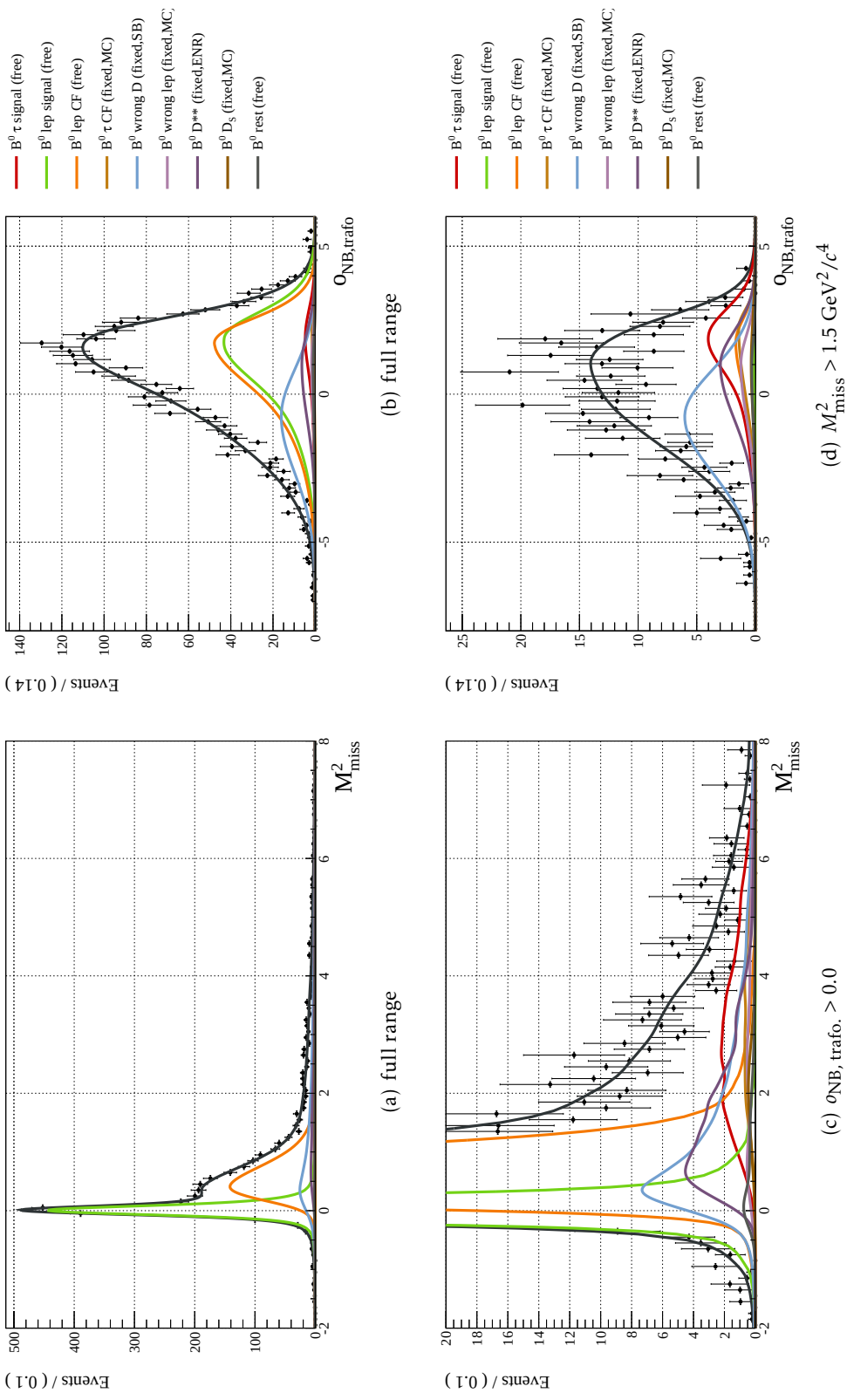


Figure 6.87.: Fit projection of $B^0 \to D^- \ell^+ \nu_\ell$ on generic Monte Carlo.

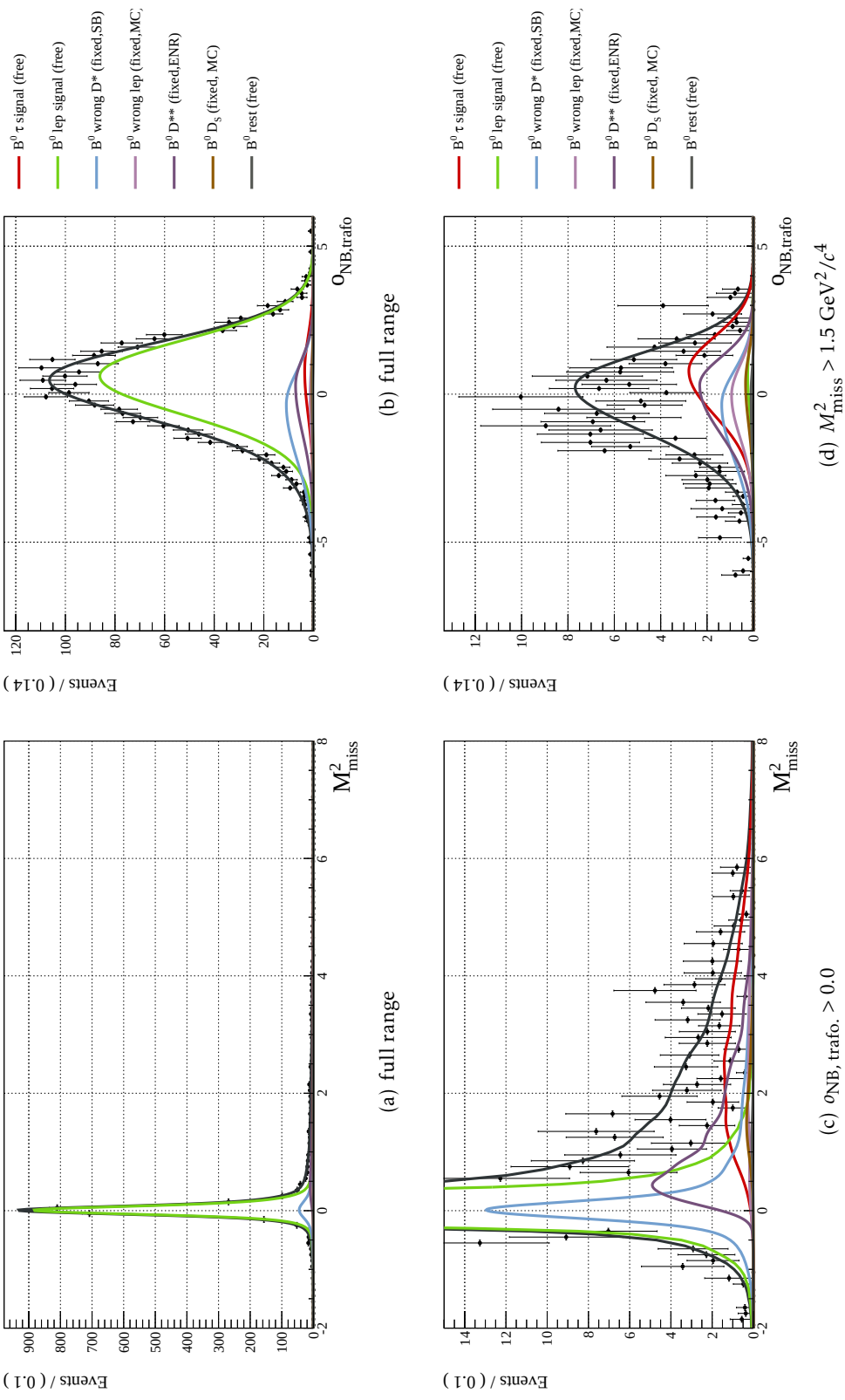


Figure 6.88.: Fit projection of $B^0 \to D^{*-}\ell^+\nu_\ell$ on generic Monte Carlo.

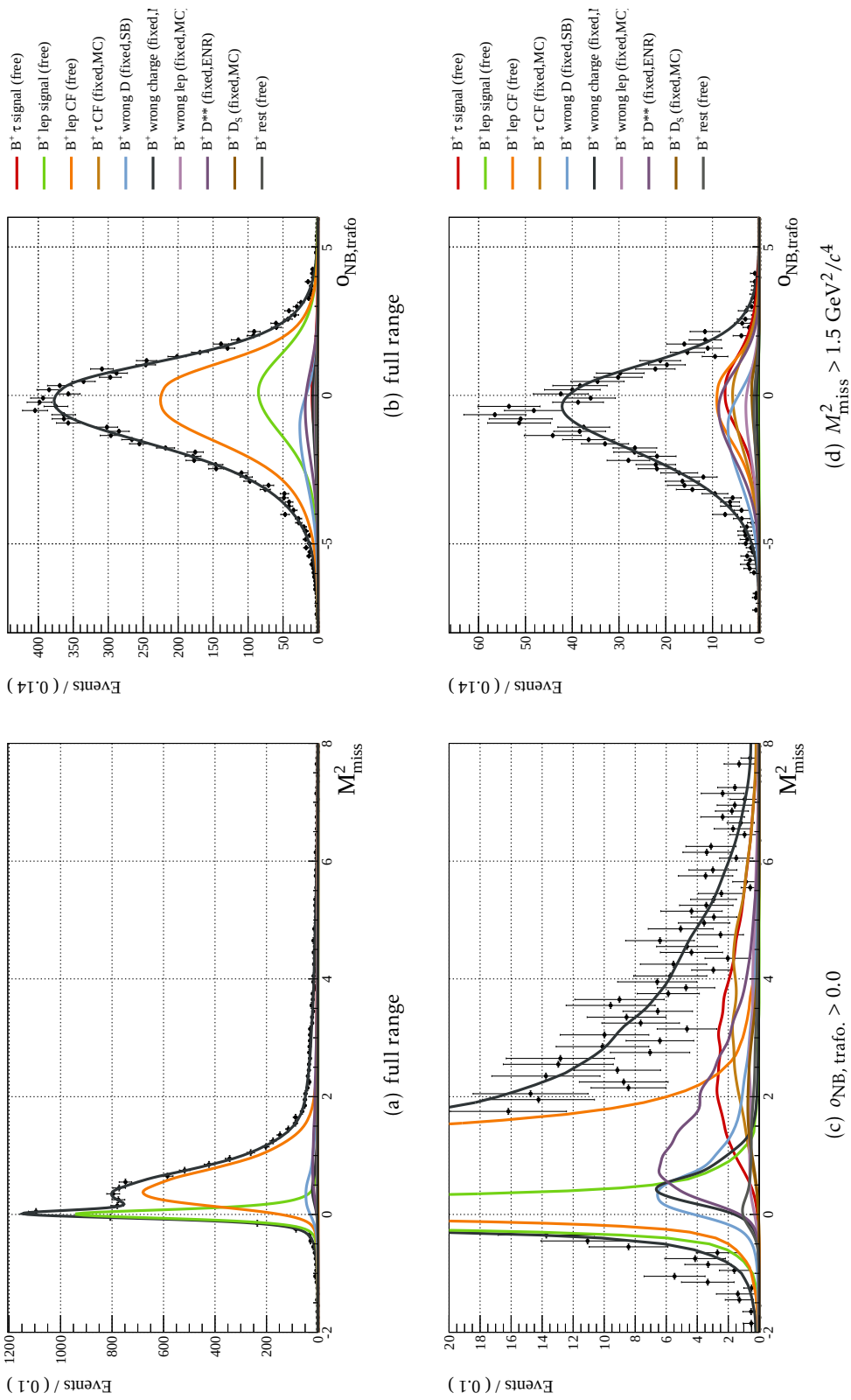


Figure 6.89.: Fit projection of $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ on generic Monte Carlo.

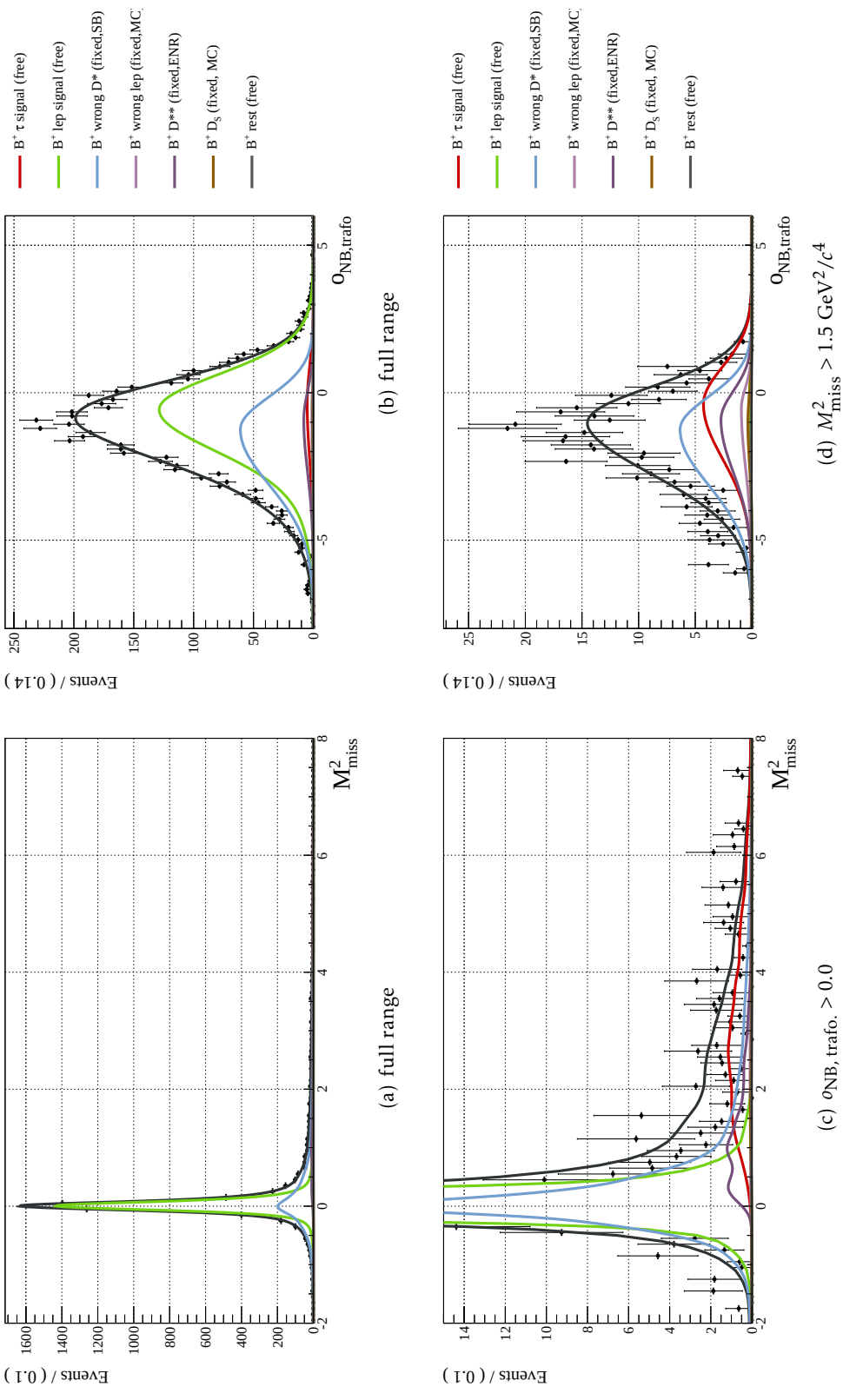


Figure 6.90.: Fit projection of $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ on generic Monte Carlo.

In conclusion, the procedure of extracting R and R^* by a simultaneous fit to the combined B^0 and B^+ sample has produced the expected results. It is the method of choice for extracting R and R^*, while the two independent fits to the B^0 and B^+ samples provide a useful alternative and cross-check.

6.20. Results

As mentioned before the fit to data was performed in two ways.

1. Two independent fits, one to the B^0 sample, the other one to the B^+ sample. This procedure is described in section 6.18.
2. One fit, simultaneously to the B^0 and the B^+ sample. This procedure is described in section 6.19.

Both fits were applied to real data. The free and fixed parameters and the factors connecting certain yields within the simultaneous fitting procedure are described in detail in sections 6.18 and 6.19.

The results of the fits to real data are summarised in table 6.27 and visualised together with the Standard Modelexpectation in the R, R^* plane in figures 6.91 to 6.93.

Channel	$R^{(*)}$	N_τ	N_ℓ	$\mathcal{B}(B \to D^{(*)}\tau\nu_\tau)$
$B^+ \to \overline{D}^0 \ell^+ \nu_\ell$	$0.271 + 0.083\ -0.081$	194.6 ± 59.6	2402.0 ± 63.9	0.61 ± 0.19
$B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$	$0.265 + 0.037\ -0.036$	127.0 ± 17.4	2604.7 ± 46.2	1.51 ± 0.22
$B^0 \to D^- \ell^+ \nu_\ell$	$0.126 + 0.063\ -0.060$	45.0 ± 22.0	1134.7 ± 38.9	0.27 ± 0.14
$B^0 \to D^{*-} \ell^+ \nu_\ell$	$0.267 + 0.047\ -0.044$	98.7 ± 16.7	2106.8 ± 46.5	1.32 ± 0.23
$B \to D \ell \nu_\ell$	$0.199 + 0.051\ -0.050$	$_{(B^+)}\ 143.6 \pm 36.4$ $_{(B^0)}\ 71.1 \pm 18.0$	$_{(B^+)}\ 2409.9 \pm 63.9$ $_{(B^0)}\ 1131.8 \pm 38.8$	
$B \to D^* \ell \nu_\ell$	$0.276 + 0.027\ -0.027$	$_{(B^+)}\ 131.2 \pm 12.8$ $_{(B^0)}\ 102.7 \pm 10.0$	$_{(B^+)}\ 2583.4 \pm 45.4$ $_{(B^0)}\ 2118.3 \pm 46.5$	

Table 6.27.: Final results

The branching ratios in table 6.27 were calculated from the measured factors $R^{(*)}$ and the current world average values for the corresponding $B \to D^{(*)}\ell\nu_\ell$ decays [2]:

$$\mathcal{B}(B \to D^{(*)}\tau\nu_\tau) = R^{(*)} \cdot \mathcal{B}(B \to D^{(*)}\ell\nu_\ell)_{\mathrm{WA}} \tag{6.62}$$

The errors on the branching ratios were calculated by propagating the errors on R^* and the $B \to D^{(*)}\ell\nu_\ell$ branching ratios:

$$\Delta\mathcal{B}(B \to D^{(*)}\tau\nu_\tau) = \sqrt{\left(\frac{\Delta R^{(*)}}{R^{(*)}}\right)^2 + \left(\frac{\Delta\mathcal{B}(B \to D^{(*)}\ell\nu_\ell)_{\mathrm{WA}}}{\mathcal{B}(B \to D^{(*)}\ell\nu_\ell)_{\mathrm{WA}}}\right)^2} \cdot \mathcal{B}(B \to D^{(*)}\tau\nu_\tau) \tag{6.63}$$

As the errors on R^* are asymmetric, the larger of the two values was used.

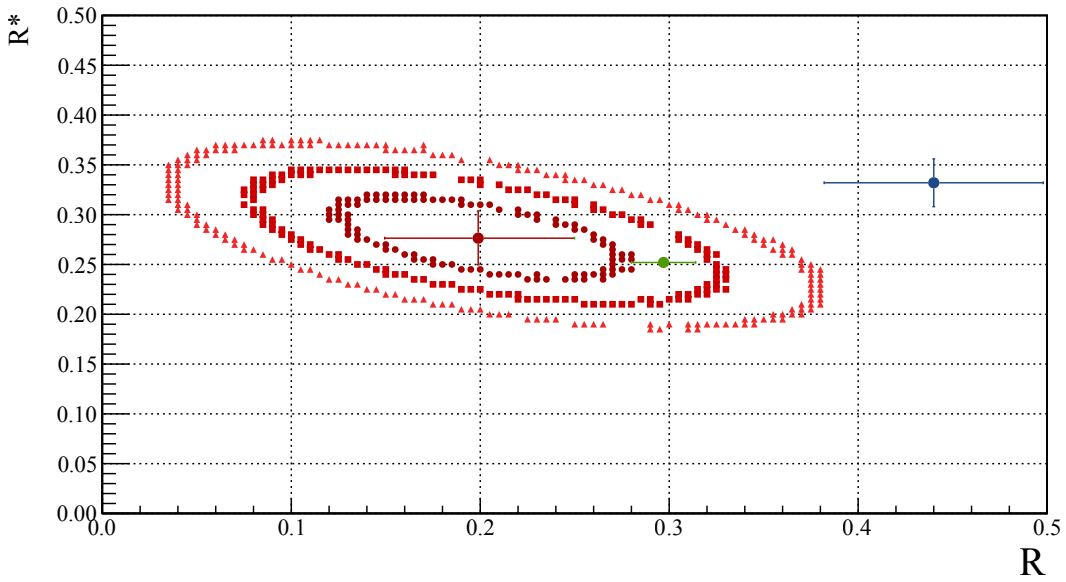


Figure 6.91.: Fit result with 1,2 and 3σ contours and Standard Model expectation in the R, R^* plane. Fit to the combined B^0 and B^+ sample. Fit result: red, BaBar result: blue, Standard Model: green.

The simultaneous fit to the combined B^0 and B^+ sample yields a result that is in excellent agreement with the Standard Model expectation [21] and deviates from the Standard Modelby less than 2σ. It is visualised in figure 6.91. We can not see a significant reduction towards lower values of R, as expected from Minimal Supersymemtic Standard Models, explained in chapter 2.3.1 and in reference [16, 22]. We can also not reproduce the τ signal excess observed by the BaBar collaboration [56], which was in disagrement with the Standard Model and the usual type-II (see chapter 2.3.1) scenario [86, 87].
The two independent fits individually yield similar results: The fit to the B^0 sample, shown in figure 6.92 is also in agreement with the Standard Model, although the value of $R = 0.126^{+0.063}_{-0.060}$ deviates by slightly more than 2σ from it. The extracted value of $R^* = 0.267^{+0.047}_{-0.044}$ is in good agreement with the Standard Model expectation.

The fit to the B^+ sample, shown in figure 6.93 is in good agreement with the Standard Model. The value of $R = 0.271^{+0.083}_{-0.081}$ as well as the value of $R^* = 0.265^{+0.037}_{-0.036}$ is in good agreement with the Standard Model expectation. Projections of the fits can be found in figures 6.94 to 6.101.

The statistical uncertainty of this analysis could be significantly reduced compared to the previous analysis using hadronic tagging (see section 6.1 and reference [54]). They were reduced from ~ 0.2 for R and ~ 0.1 for R^* (see table 6.1 for more details) to ~ 0.08

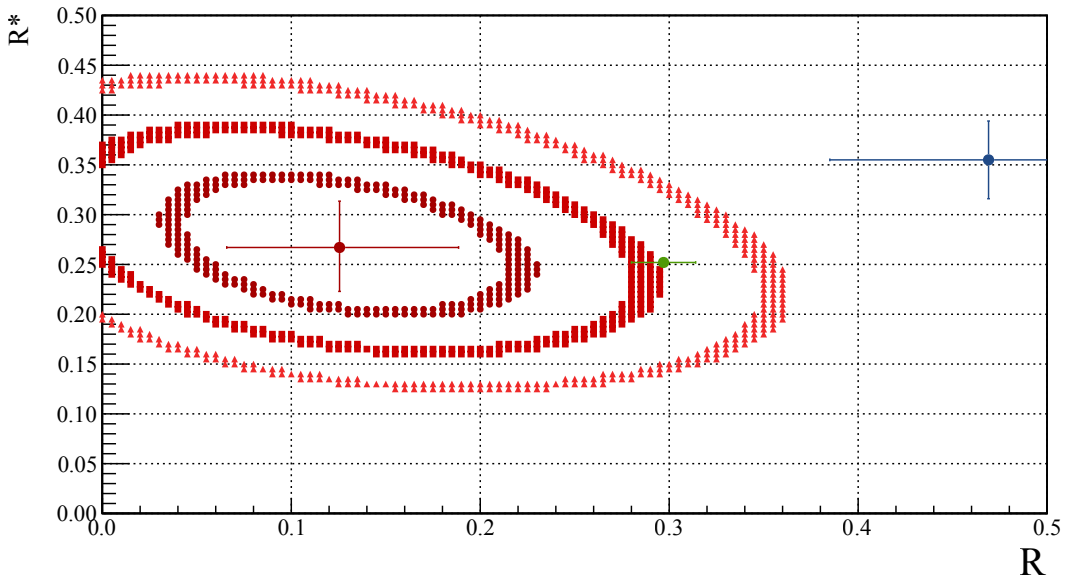


Figure 6.92.: Fit result with 1,2 and 3σ contours and Standard Model expectation in the R, R^* plane. Fit to the B^0 sample. Fit result: red, BaBar result: blue, Standard Model: green.

for R and ~ 0.04 for R^*. This improvement is partly due to the 17.4% larger available data set, but the largest effect comes from the new Full Reconstruction algorithm, which effectively doubled the available dataset with respect to the previous analysis.

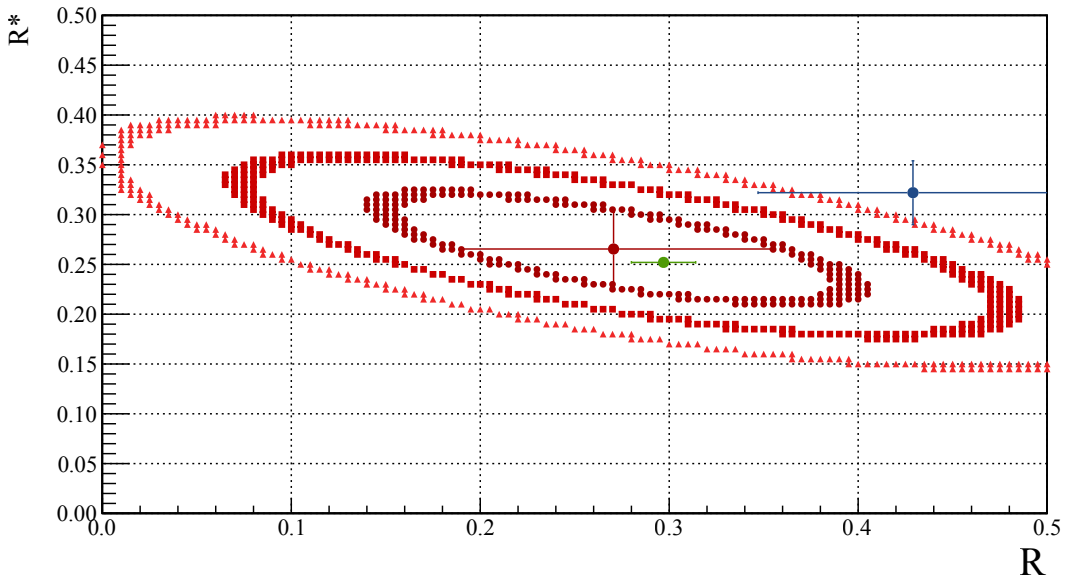


Figure 6.93.: Fit result with 1,2 and 3σ contours and Standard Model expectation in the R, R^* plane. Fit to the B^+ sample. Fit result: red, BaBar result: blue, Standard Model: green.

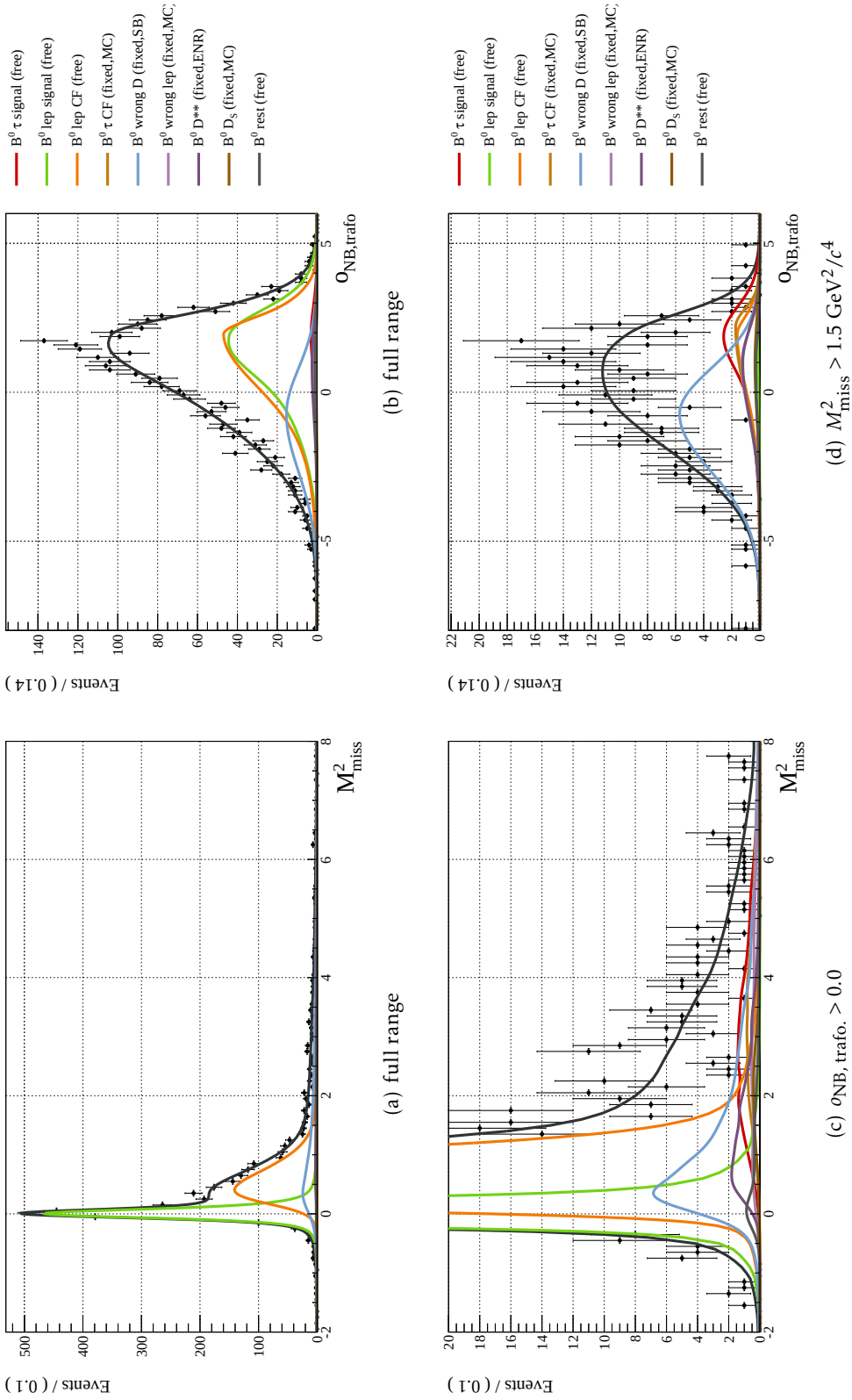


Figure 6.94.: Fit on the real data combined B sample. Projection of the $B^0 \to D^- \ell^+ \nu_\ell$ subsample.

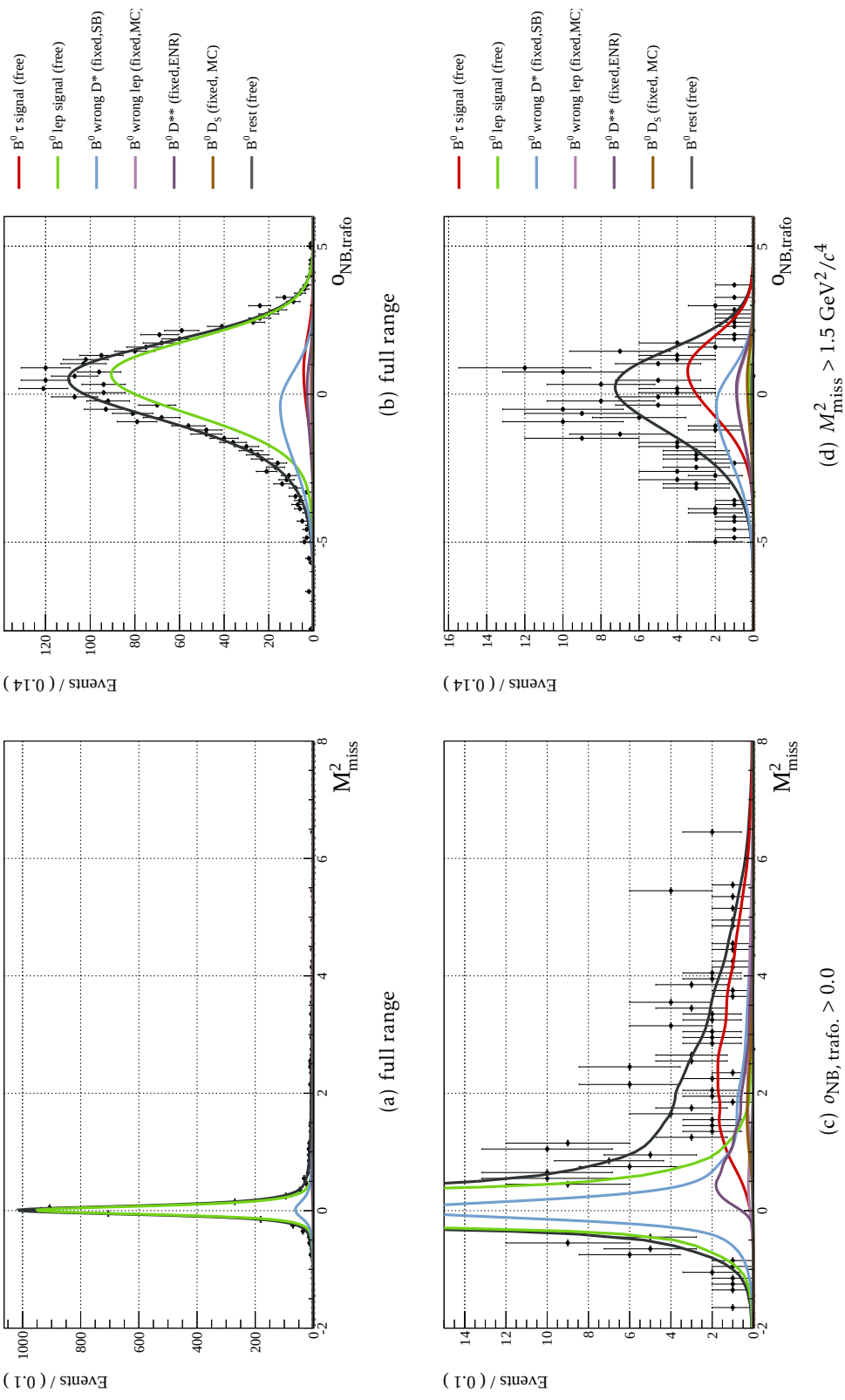


Figure 6.95.: Fit on the real data combined B sample. Projection of the $B^0 \to D^{*-}\ell^+\nu_\ell$ subsample.

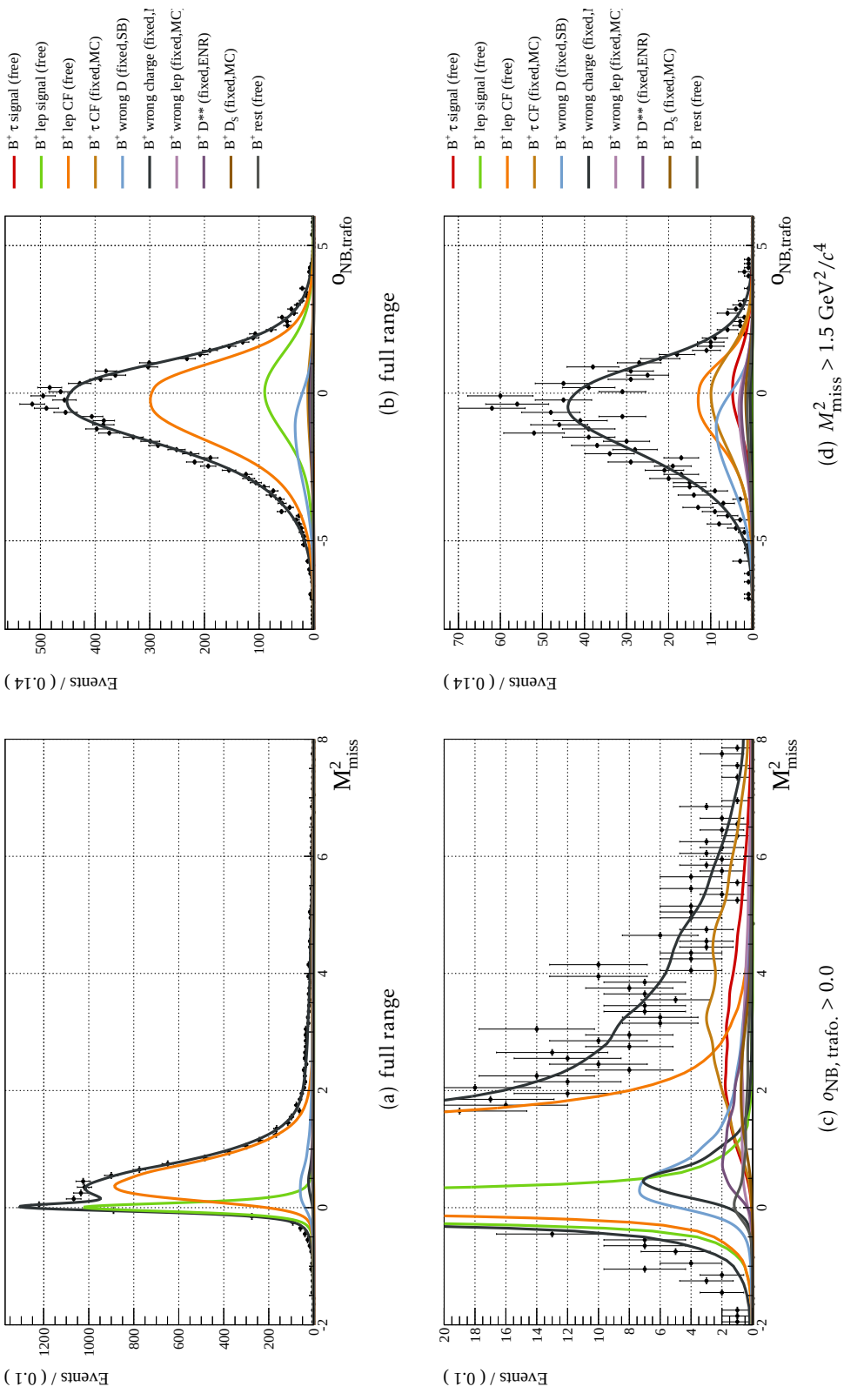


Figure 6.96.: Fit on the real data combined B sample. Projection of the $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ subsample.

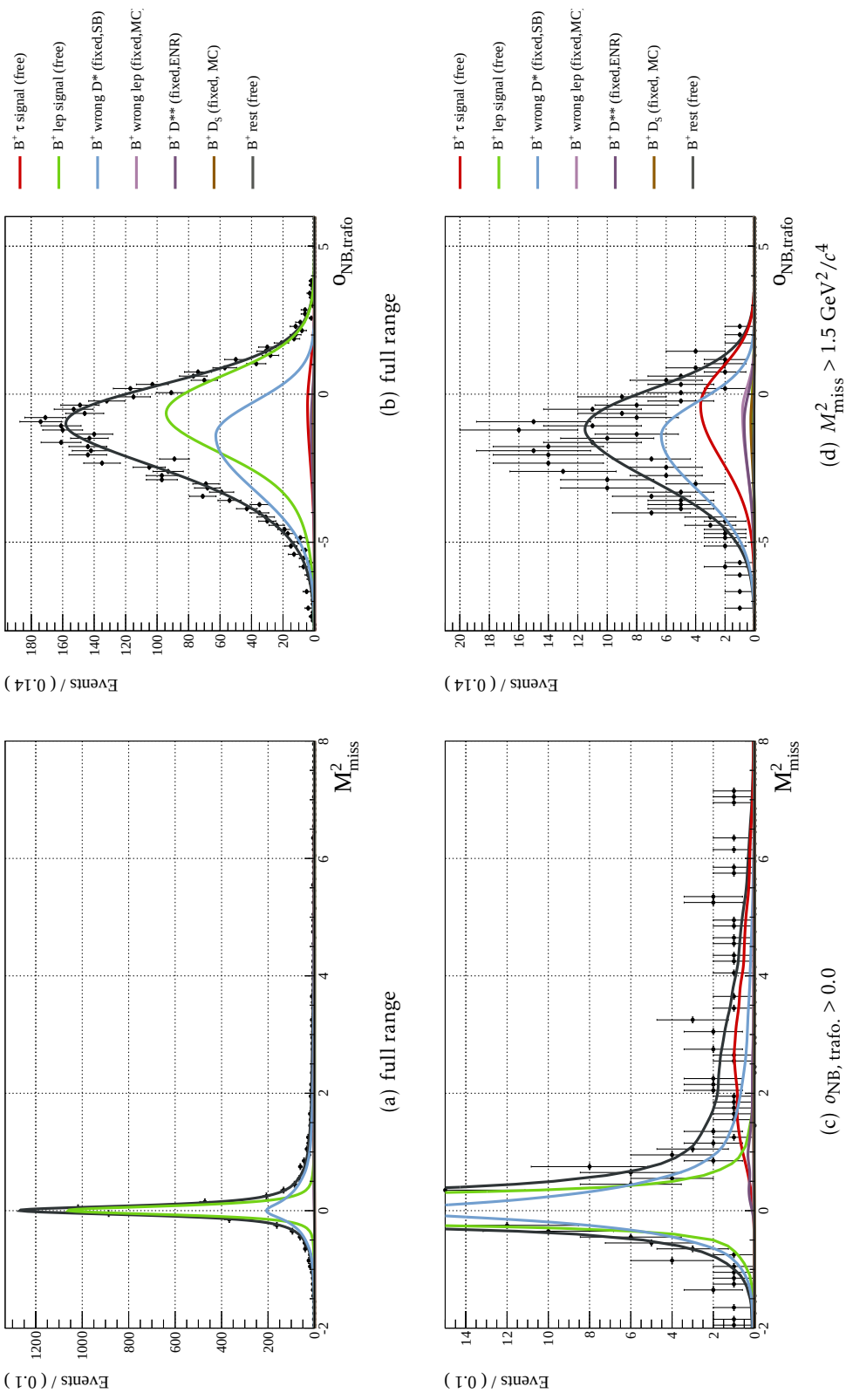


Figure 6.97.: Fit on the real data combined B sample. Projection of the $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ subsample.

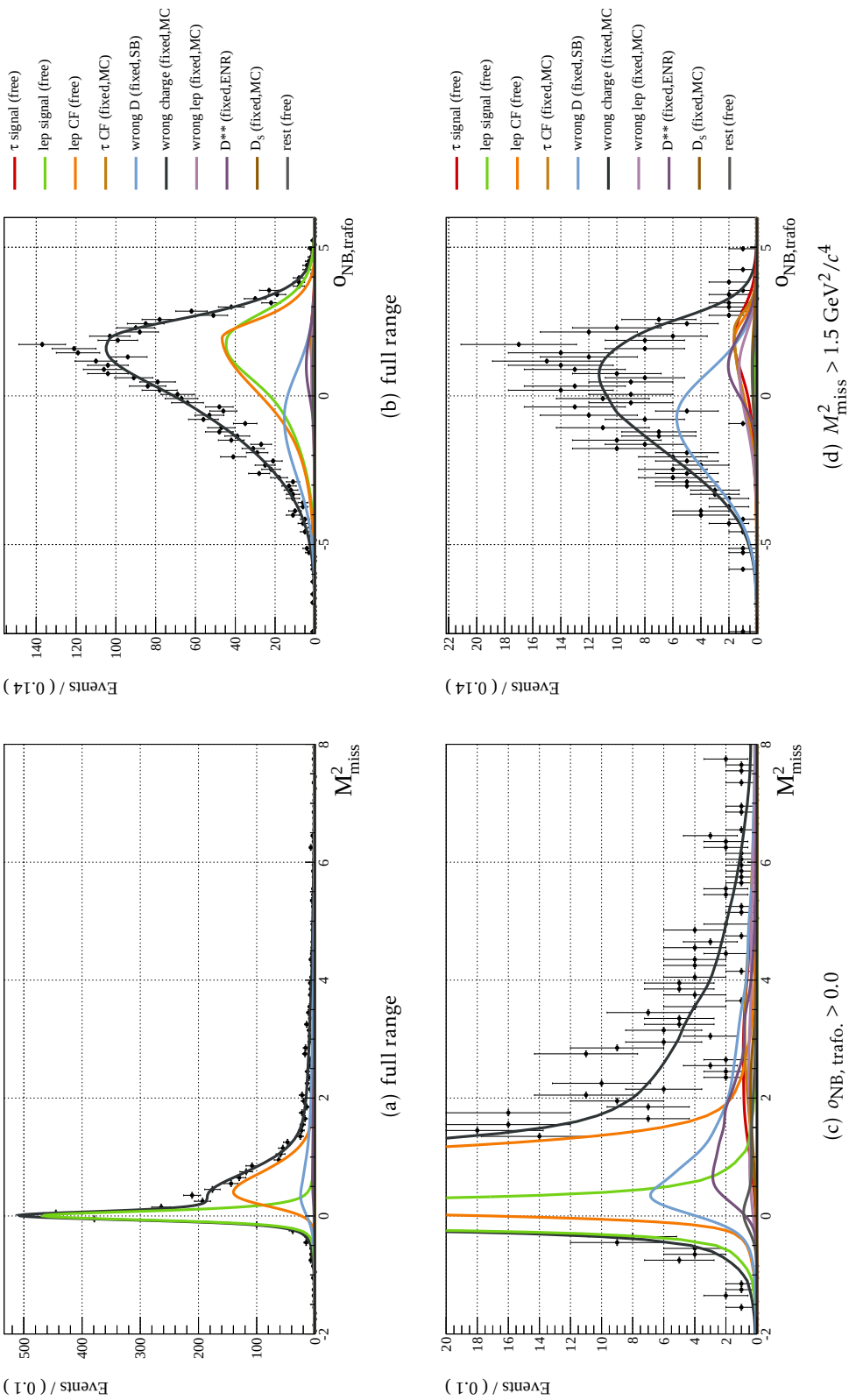


Figure 6.98.: Fit on the real data B^0 sample. Projection of the $B^0 \to D^- \ell^+ \nu_\ell$ subsample.

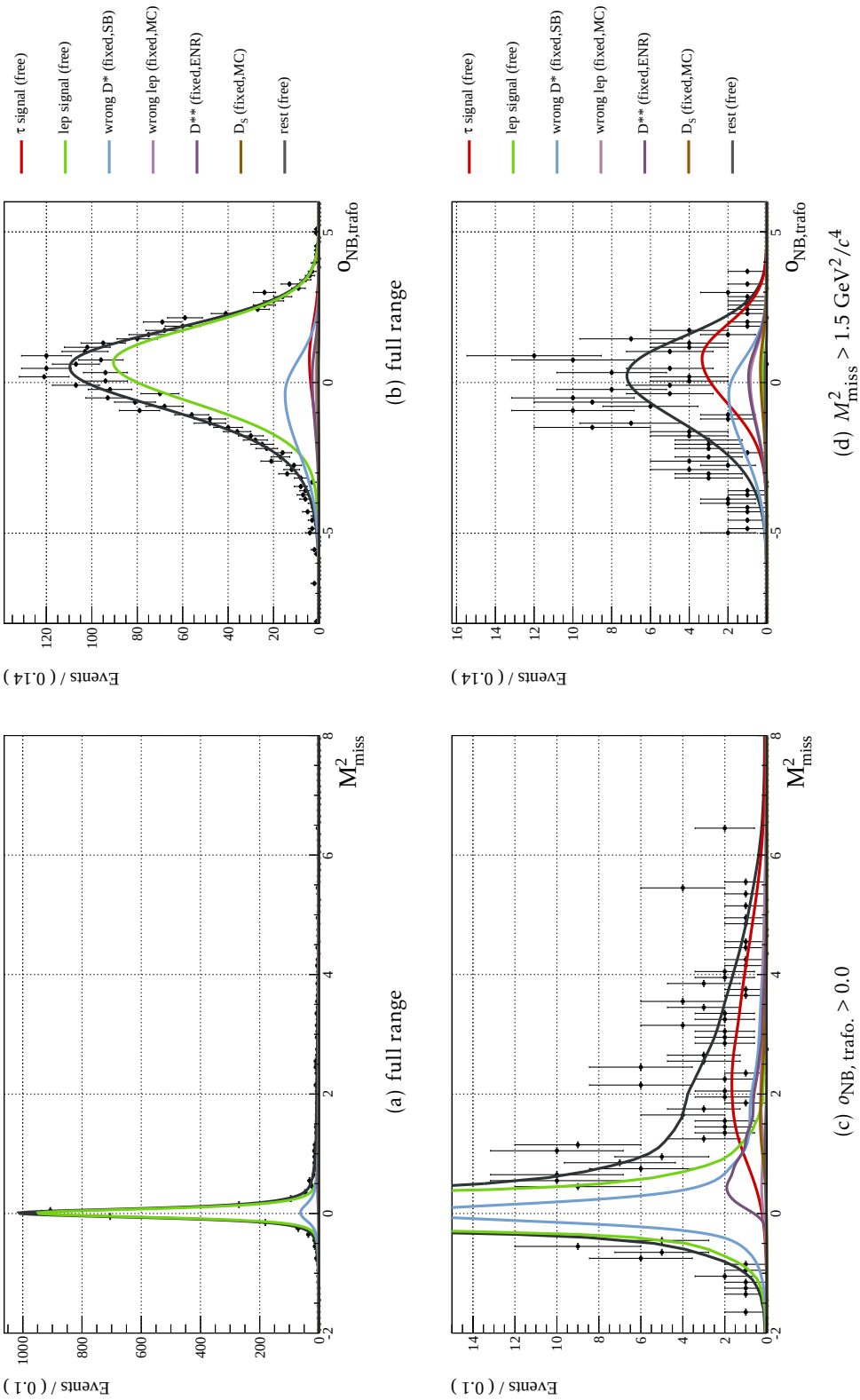


Figure 6.99.: Fit on the real data B^0 sample. Projection of the $B^0 \to D^{*-}\ell^+\nu_\ell$ subsample.

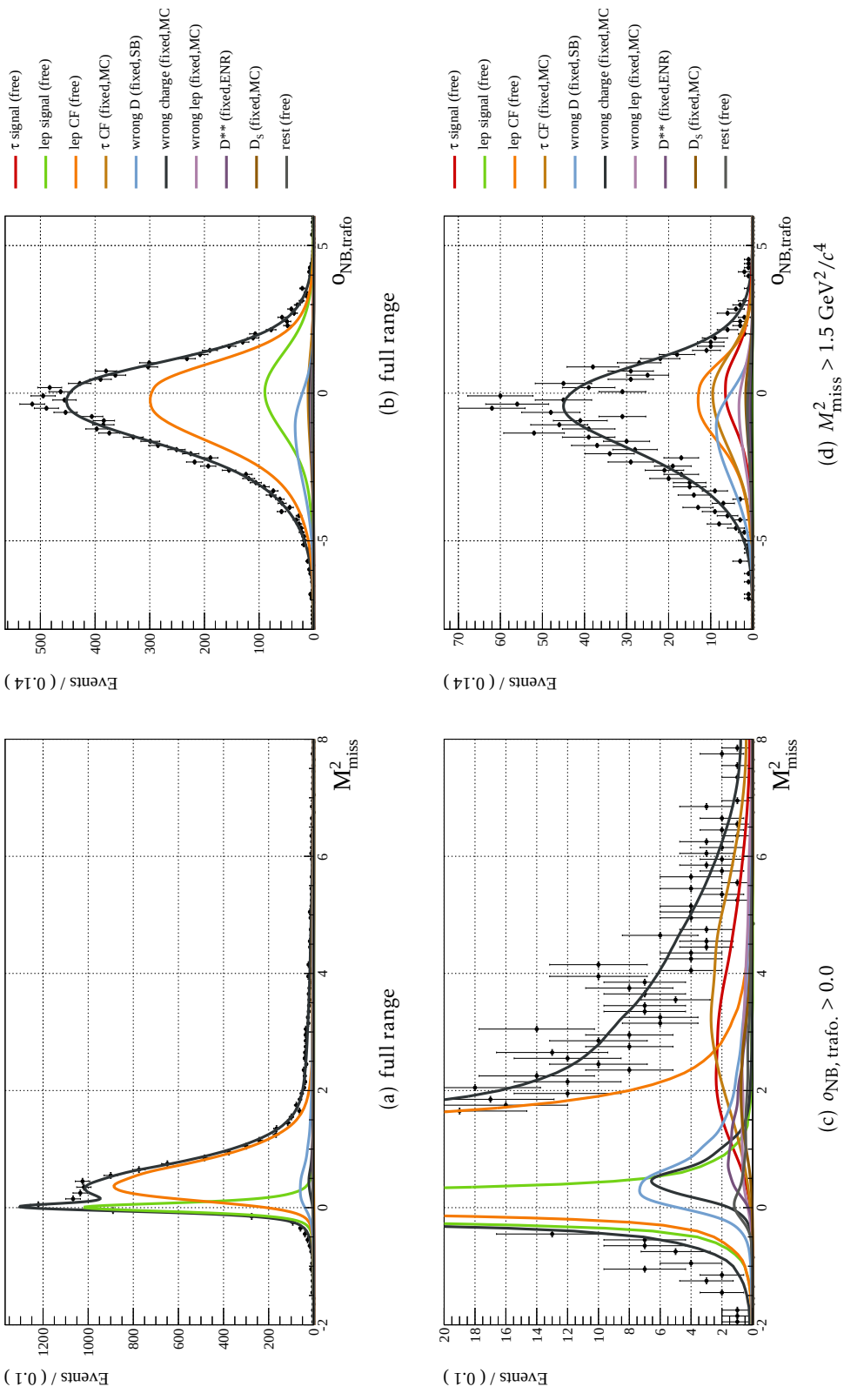


Figure 6.100.: Fit on the real data B^+ sample. Projection of the $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$ subsample.

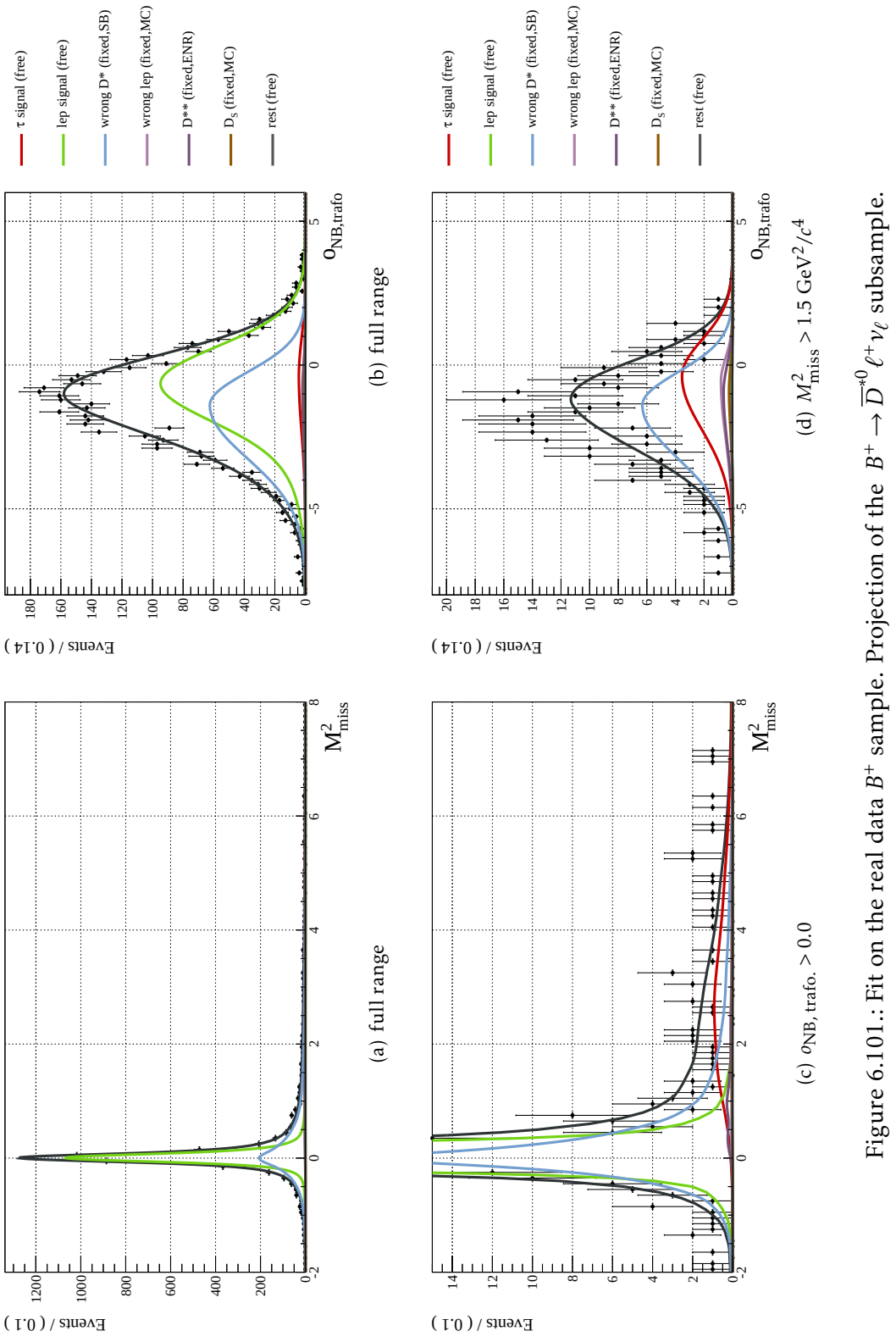


Figure 6.101.: Fit on the real data B^+ sample. Projection of the $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$ subsample.

6.21. Systematic Uncertainties

The measurements of R and R^* are *relative* measurements:

$$R^{(*)} = \frac{B \to D^{(*)} \ell \nu_\ell}{B \to D^{(*)} \tau \nu_\tau} \tag{6.64}$$

We expect therefore many uncertainties, especially those related to detector effects to cancel in the ratio. Such effects include soft pion reconstruction (from the D^* decays), particle identification and final state radiation. Large systematic uncertainties are expected to come from the limited understanding of the D^{**} background.

6.21.1. D^{**} backgrounds

The D^{**} background events have a strong influence on the extracted amount of τ signal. This is because they occupy the same region in the M^2_{miss}spectrum. Several factors contribute to the uncertainty introduced by this background:

- The limited accuracy in the determination of this background by the D^{**} enriched samples. Specifically, the uncertainties of the numerical factors f_D and f_D^* (see table 6.19).
- D^{**} contributions from decays into multi pion final states, which are not simulated on generic Monte Carlo.
- The shape of the known and unknown components of the D^{**} background.
- The possibility of yet undiscovered radial excitations of the D mesons.

All of these uncertainties considered, a conservative error estimation was performed. The two factors f_D and f_D^* were one by one increased or decreased by 50%. This corresponds to changes in the D^{**} background yield by 50%. The effects on the values of R and R^* are listed in table 6.28. As can be seen from table 6.28, the systematic uncer-

Factor		relative change in sample ...					
		B^0		B^+		combined	
f_D	R	+0.047	−0.026	+0.032	−0.017	+0.031	−0.015
	R^*	0.000	0.000	−0.001	0.000	+0.001	−0.001
f_D^*	R	−0.012	+0.004	−0.014	+0.006	−0.014	+0.005
	R^*	+0.026	−0.009	+0.009	−0.003	+0.014	−0.005

Table 6.28.: Relative change in the fit results with varied factors in the D^{**} background estimation

tainty can be quite substantial. Especially the measurement of R in the B^0 sample has a large uncertainty. The uncertainties were added in quadrature and are listed in the summary table 6.32.

6.21.2. Factors For Simultaneous Fitting

All factors that are used in the simultaneous fitting procedure (see table 6.19) were evaluated using generic Monte Carlo. They are only known up to a certain precision. To evaluate the effects of this imprecision to the final result, each factor was varied by its uncertainty. The resulting changes to the extracted values R and R^* are listed in table 6.29. It can be seen from this table, that most factors have little influence on the final result. The largest contribution comes from f_R (see equation 6.49) and g (see equation 6.51). The deviations were added quadratically and are listed in table 6.32.

6.21.3. PDF Shapes

The PDF for all fit componnets have been determined on simulated data and are therefore sensitive to differences between data and Monte Carlo. It is known that the resolution of M^2_{miss}may be not modeled entirely correctly on Monte Carlo. This difference in resolution should have the largest relative effect on the the lepton signal as this is the the narrowest of the components. All of the other components, being much broader, are expected to be less affected by the difference in resolution.

$D^{(*)}$ lepton signal

The M^2_{miss}spectra in the region $-0.45\,\text{GeV}^2/c^4 < M^2_{\text{miss}} < 0.45\,\text{GeV}^2/c^4$ for data and Monte Carlo were analysed. A histogram pdf for the lepton signal component and for the background was taken from 5 streams of the generic Monte Carlo. The signal component was then transformed according to:

$$x = (M^2_{\text{miss}} - a)/s\,, \tag{6.65}$$

where a would indicate a shift of the histogram pdf and s a widening of it. The sum of the tranformed signal histogram pdf and the background histogram pdf was then fitted to the data sample. The results are summarised in table 6.30. These obtained corrections were then applied when fitting data. Only the normalisation data sample ($M^2_{\text{miss}} < 1.5\,\text{GeV}/c^2$) was fitted, as described in section 6.11. The lepton signal pdf was changed to a histogram pdf. The fit was then performed with and without the application of the relative widenings and shifts listed in table 6.30. The results of the fits are listed in table The relative changes are all at most in the order of 0.1%, most are even much smaller. Compared to the relatively large uncertainties brought by the D^{**} background, this component is completely negligible.

		B^0 Sample		B^+ Sample		Combined Sample	
		$+1\sigma$	-1σ	$+1\sigma$	-1σ	$+1\sigma$	-1σ
f_{SB}	R	−0.0020	+0.0020	+0.0011	−0.0011	−0.0007	+0.0007
	R^*	−0.0008	+0.0008	−0.0014	+0.0014	−0.0008	+0.0009
f^*_{SB}	R	+0.0001	−0.0001	+0.0022	−0.0020	+0.0009	−0.0008
	R^*	−0.0018	+0.0018	−0.0023	+0.0021	−0.0018	+0.0018
f_{e1}	R	+0.0011	−0.0011	+0.0003	−0.0003	+0.0008	−0.0008
	R^*	+0.0001	−0.0000	+0.0000	−0.0001	−0.0000	+0.0000
f_{e2}	R	+0.0044	−0.0045	+0.0027	−0.0026	+0.0035	−0.0036
	R^*	+0.0005	−0.0005	+0.0007	−0.0008	+0.0006	−0.0006
f^*_e	R	−0.0007	+0.0006	−0.0017	+0.0018	−0.0010	+0.0010
	R^*	+0.0013	−0.0013	+0.0009	−0.0010	+0.0008	−0.0008
f_R	R	+0.0056	−0.0056	+0.0088	−0.0088	+0.0083	−0.0083
	R^*	+0.0000	+0.0000	+0.0000	+0.0000	−0.0004	+0.0004
f^*_R	R	+0.0000	−0.0000	+0.0001	+0.0001	−0.0003	+0.0003
	R^*	+0.0131	−0.0131	+0.0111	−0.0112	+0.0123	−0.0123
g	R	+0.0095	−0.0110	+0.0169	−0.0176	+0.0135	−0.0145
	R^*	+0.0002	−0.0005	+0.0022	−0.0028	+0.0023	−0.0028
f_{wc}	R					+0.0001	−0.0001
	R^*					0.0000	0.0000

Table 6.29.: Absolute change in the extracted values for R and R^*.

All Other Shapes

While the uncertainty on the lepton signal shape has little impact on the fit result, this might be different for other components. Especially those that are similar to the τ signal shape may have a noticeable influence on the result. To evaluate this behaviour, the shapes of all components were changed and the fit was repeated. As the shapes in $M^2_{\text{miss}}$are determined by Kernel estimation functions (see [76]), the bandwidth of these functions was varied by 50%. The resulting relative changes are summarized in table 6.32. The effect of the varied bandwidth in the kernel estimation is in most cases

Channel	s	a
$B^0 \to D^- e^+ \nu_e$	1.004 ± 0.004	$(0.7 \pm 0.7) \times 10^{-3}$
$B^0 \to D^{*-} e^+ \nu_e$	1.033 ± 0.001	$(3.7 \pm 0.2) \times 10^{-3}$
$B^+ \to \overline{D}^0 e^+ \nu_e$	1.017 ± 0.003	$(-4.8 \pm 0.8) \times 10^{-3}$
$B^+ \to \overline{D}^{*0} e^+ \nu_e$	0.998 ± 0.003	$(1.7 \pm 0.8) \times 10^{-3}$

Table 6.30.: Relative widening s and shift a of the lepton signal

Channel	Change	Lepton Signal	Rel. Change
$B^0 \to D^- \ell^+ \nu_\ell$	original	1609.0	
	transformed	1609.0	0%
$B^0 \to D^{*-} \ell^+ \nu_\ell$	original	2727.7	
	transformed	2727.2	$< 0.1\%$
$B^+ \to \overline{D}^0 \ell^+ \nu_\ell$	original	3384.3	
	transformed	3379.8	0.13%
$B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$	original	4176.7	
	transformed	4175.4	$< 0.1\%$

Table 6.31.: Influence of the relative widening s and shift a of the lepton signal.

smaller than the uncertainties coming from the D^{**} background or the factors used in the simultaneous fitting procedure.

6.21.4. Summary

Table 6.32 lists all of the systematic uncertainties discussed above. With these systematic uncertainties, the result of the simultaneous fit to the combined B^0 and B^+ sample is the following:

$$R = 0.199^{+0.051}_{-0.050}(\text{stat.})^{+0.037}_{-0.027}(\text{syst.}) \tag{6.66}$$

$$R^* = 0.276^{+0.027}_{-0.027}(\text{stat.})^{+0.019}_{-0.017}(\text{syst.}) \tag{6.67}$$

Source		B^0 Sample		B^+ Sample		Combined Sample	
D^{**}	R	+0.047	−0.029	+0.033	−0.022	+0.031	−0.021
	R^*	+0.026	−0.009	+0.009	−0.003	+0.014	−0.005
Factors	R	+0.012	−0.019	+0.019	−0.020	+0.016	−0.017
	R^*	+0.013	−0.013	+0.012	−0.012	+0.013	−0.013
Bandwidth	R	+0.011	−0.000	+0.017	−0.000	+0.013	−0.004
	R^*	+0.000	−0.018	+0.000	−0.016	+0.000	−0.010
Quad. Sum	R	+0.050	−0.035	+0.042	−0.030	+0.037	−0.027
	R^*	+0.029	−0.024	+0.015	−0.020	+0.019	−0.017

Table 6.32.: Absolute systematic uncertainties from various sources.

The result of the two independent fits to the B^0 and the B^+ sample is:

$$R^0 = 0.126^{+0.063}_{-0.060}(\text{stat.})^{+0.050}_{-0.035}(\text{syst.}) \tag{6.68}$$

$$R^{*0} = 0.267^{+0.047}_{-0.044}(\text{stat.})^{+0.029}_{-0.024}(\text{syst.}) \tag{6.69}$$

$$R^{+} = 0.271^{+0.083}_{-0.081}(\text{stat.})^{+0.042}_{-0.030}(\text{syst.}) \tag{6.70}$$

$$R^{*+} = 0.265^{+0.037}_{-0.036}(\text{stat.})^{+0.015}_{-0.020}(\text{syst.}) \tag{6.71}$$

The largest systematic uncertainty comes from the D^{**} background. For each measurement, the statistical uncertainty is dominant over the systematic uncertainty. Assuming that the systematic uncertainties do not change the overall correlation between R and R^*, the contour plots visualising the results have been redone. They can be found in figures 6.102 to 6.104. These contour plots also include 1,2 and 3σ contours for the BaBar results. they are calculated using the errors and correlations given in reference [56].
With the prospect of a vastly larger data sample, provided by the future Super B-factories [88], $B \to D^{(*)}\tau\nu_\tau$ remains an interesting channel to consider for indirect searches for new physics effects [89, 90]. Possible improved understanding of the D^{**} decays might help in the future to further increase the sensitivity of analyses similar to the one presented in this thesis.

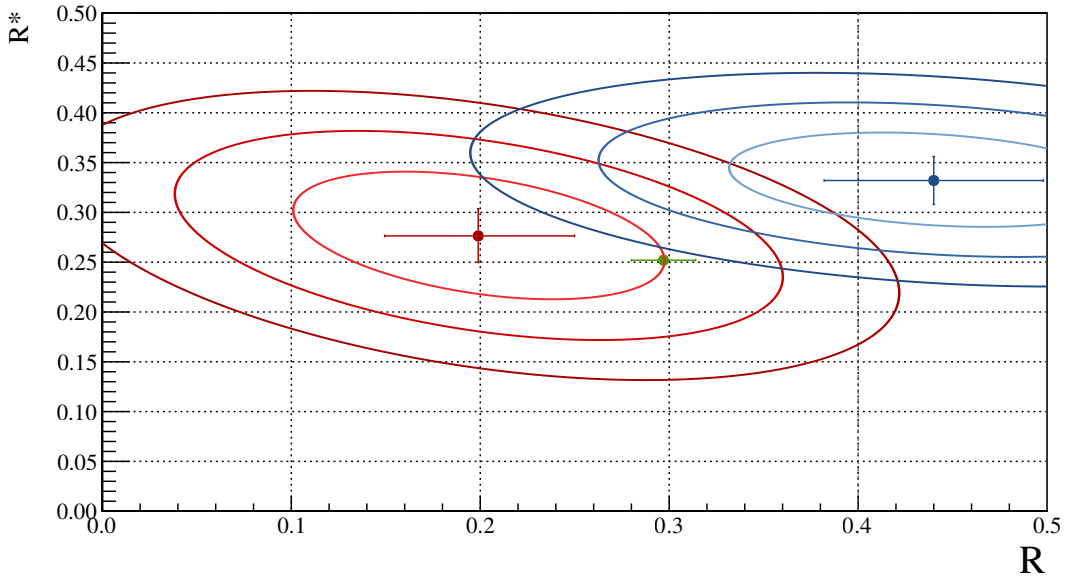


Figure 6.102.: Fit result with 1,2 and 3σ contours and Standard Model expectation in the R, R^* plane. Fit to the combined B^0 and B^+ sample. Fit result: red, BaBar result: blue, Standard Model: green.

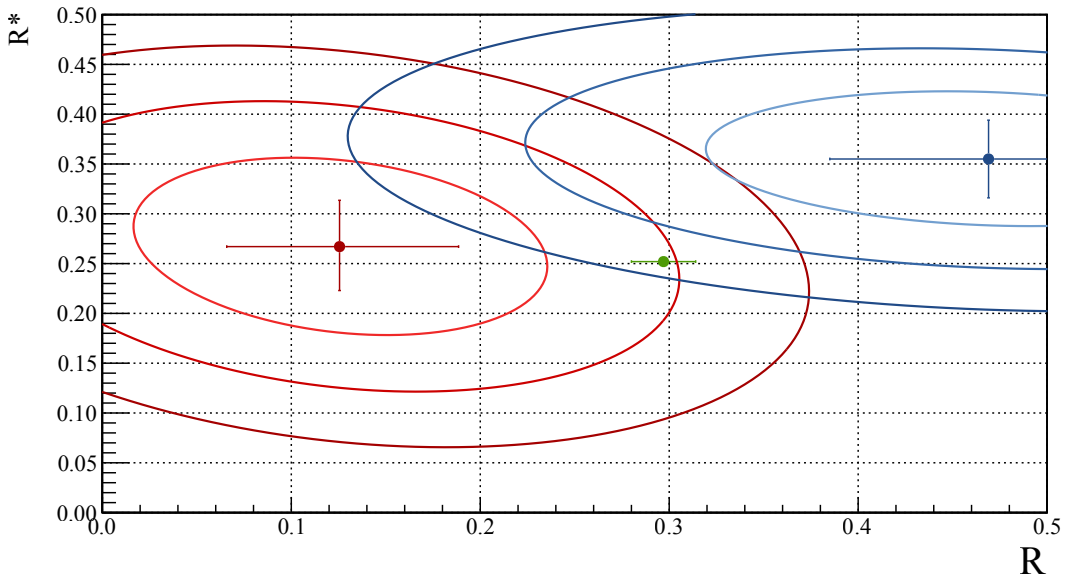


Figure 6.103.: Fit result with 1,2 and 3σ contours and Standard Model expectation in the R, R^* plane. Fit to the B^0 sample. Fit result: red, BaBar result: blue, Standard Model: green.

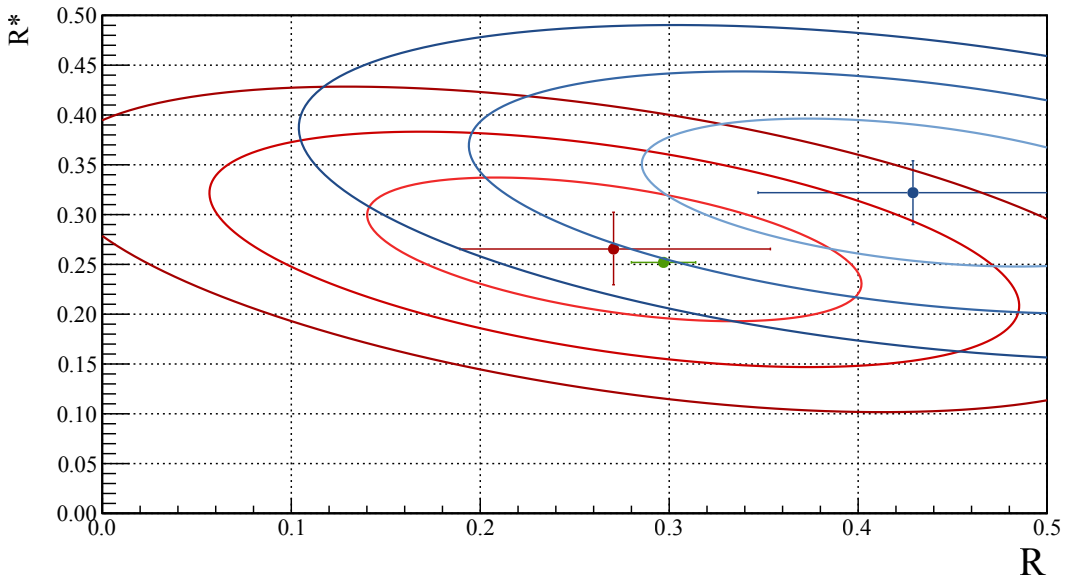


Figure 6.104.: Fit result with 1,2 and 3σ contours and Standard Model expectation in the R, R^* plane. Fit to the B^+ sample. Fit result: red, BaBar result: blue, Standard Model: green.

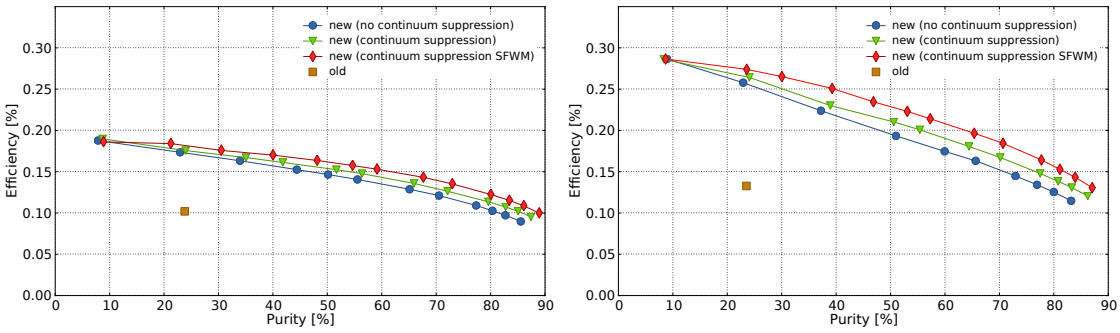
7. Conclusion

The presented thesis consisted of two major parts. Let us start with the second part, for a change. The second part described the analysis of the decay $B \to D^{(*)}\tau\nu_\tau$. The goal of this analysis was the extraction of the values R and R^*, defined as

$$R = \frac{\mathcal{B}(B \to D\tau\nu_\tau)}{\mathcal{B}(B \to D\ell\nu_\ell)} \text{ and}$$
$$R^* = \frac{\mathcal{B}(B \to D^*\tau\nu_\tau)}{\mathcal{B}(B \to D^{(*)}\ell\nu_\ell)} .$$

The τ leptons were reconstructed from their purely leptonic decays. Thus, the final states of $B \to D^{(*)}\tau\nu_\tau$ decays differed from those of $B \to D^{(*)}\ell\nu_\ell$ decays only by two additional, undetectable neutrinos. This is the point, where part one of the presented thesis comes into play.

In part one of this thesis the development of an improved hadronic tagging algorithm, the Full Reconstruction, was described. This algorithm allows to measure missing momentum in a $B\bar{B}$ event indirectly. The efficiency of the Full Reconstructioncould be doubled compared to its predecessor. As shown in figure 7.1, the overall performance in terms of purity and efficiency is largely improved. This improvement could be achieved



(a) B^0 (b) B^+

Figure 7.1.: The purity-efficiency curves for the B_{tag} sample provided by the Full Reconstruction.

in part by the addition of more B meson decay channels to the algorithm, but largely due to the advanced reconstruction procedure, heavily utilising the NeuroBayes package.

With this improved hadronic tagging, the analysis of the decay $B \to D^{(*)}\tau\nu_\tau$ could be performed on the entire Belle dataset of $771 fb^{-1}$ with a hadronic tagging efficiency of $\sim 0.2\%$. The extracted values

$$R = 0.199^{+0.051}_{-0.050}(\text{stat.})^{+0.037}_{-0.027}(\text{syst.}) \text{ and} \tag{7.1}$$

$$R^* = 0.276^{+0.027}_{-0.027}(\text{stat.})^{+0.019}_{-0.017}(\text{syst.}) \tag{7.2}$$

are well compatible with the Standard Model expectations of $R = 0.297 \pm 0.017$ and $R^* = 0.251 \pm 0.003$. The statistical uncertainties could be reduced significantly with respect to the previous analysis[54]. The extracted values are visualised in the contour plot, shown in figure 7.2. Neither the recently found excess by the BaBar collaboration,

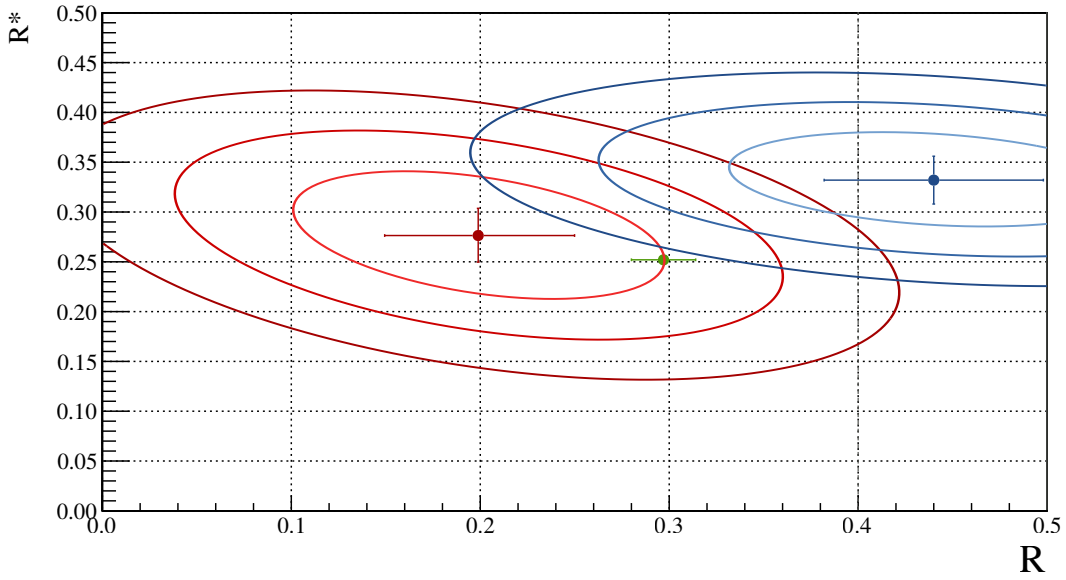


Figure 7.2.: Fit result with 1,2 and 3σ contours and Standard Model expectation in the R, R^* plane. Fit to the combined B^0 and B^+ sample. Fit result: green, BaBar result: blue, Standard Model: black.

nor charged Higgs effects could be observed.

A. Appendix

A.1. A Brief Introduction To Feynman Diagrams

The important experimental probes of particle physics are the measurements of lifetimes and cross sections. The comparison of the measured values to the theoretical predictions enables us to strengthen or disprove the theory on which the calculations were based.
The calculation of lifetimes and cross sections requires two ingredients: [3]

1. The *matrix element* $\mathcal{M}$ for the process (also referred to as the amplitude)
2. The available *phase space* (also referred to as the density of final states)

These two ingredients are used in the calculation of the transition rate of a given process by an equation, often called "Fermi's Golden Rule":

$$\text{transition rate} = \frac{2\pi}{\hbar}|\mathcal{M}|^2 \times (\text{phase space factor}) \tag{A.1}$$

A.1.1. Phase Space

The available *phase space* depends on the masses, energies and momenta of the initial and final state particles. Suppose a particle 1 decays into several particles 2, 3, ...n, then the decay rate for that process is given by

$$d\Gamma \sim |\mathcal{M}|^2 \cdot \underbrace{\left[\frac{d^3\vec{p}_2}{E_2} \cdot \frac{d^3\vec{p}_3}{E_3} \ldots \frac{d^3\vec{p}_n}{E_n}\right] \times \delta^4(p_1 - (p_2 + p_3 \ldots p_n))}_{\text{phase space factor}}, \tag{A.2}$$

where $p_i = (E_i/c, \vec{p_i})$ is the 4 momentum of particle i and the δ function enforces energy and momentum conservation. Figuratively speaking, the more an initial state outweighs a final state in terms of mass, energy and momentum, the larger the phase space factor is, and the more likely a process is to occur. If the initial state only barely outweighs the final state (e.g. in the decay of the neutron $n \to pe\bar{\nu}_e$), the phase space and thereby the decay rate is very small.

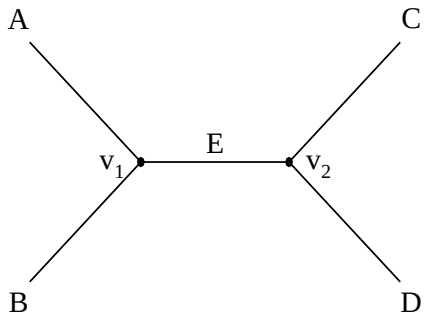


Figure A.1.: A generic Feynman diagram

A.1.2. Matrix Element

The *matrix element* of a process contains all the dynamical information. The calculation of the matrix element can be visualised by "Feynman diagrams". A very generic Feynman diagram is shown in figure A.1. In order to calculate the matrix element for the generic process depicted in figure A.1, we first look at the vertices v_1 and v_2 and notate a factor of

$$-ig$$

for each vertex, where g is the coupling constant of the interaction at the vertex. Please note that during this section, the letter i is reserved for the imaginary unit. The conservation of energy and momentum dictates that at each vertex an additional factor of

$$(2\pi^4)\delta^4(k_1 + k_2 + k_3)$$

comes into the matrix element $\mathcal{M}$, where k_j is the momentum of particle j. The single internal lines give a factor of

$$\frac{i}{q^2 - m^2c^2},$$

where q is the 4-momentum and m the mass of the particle this internal line describes. The necessary integration over internal momenta yields a factor of

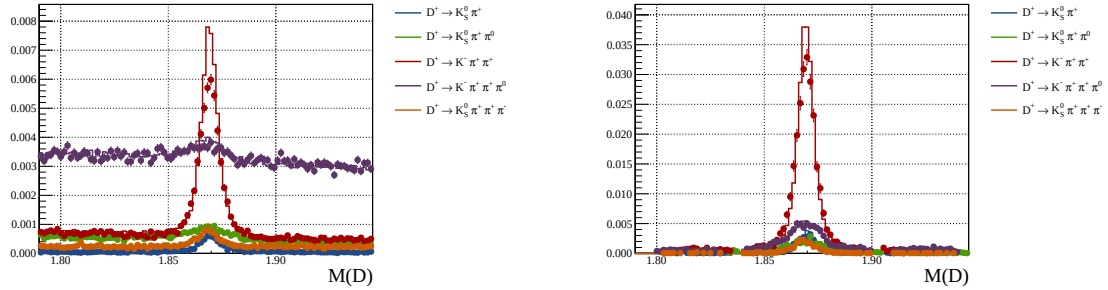
$$\frac{1}{(2\pi)^4} d^4q$$

per internal line, where again q is the 4-momentum of the corresponding particle. Following this procedure, we obtain the matrix element $-i\mathcal{M}$. For the diagram shown in figure A.1, the Matrix element is

$$-i\mathcal{M} = -ig^2 \frac{1}{(p_C - p_D)^2 - m_C^2c^2} (2\pi)^4 \delta^4(p_B + p_D - p_A + p_C)) \,. \tag{A.3}$$

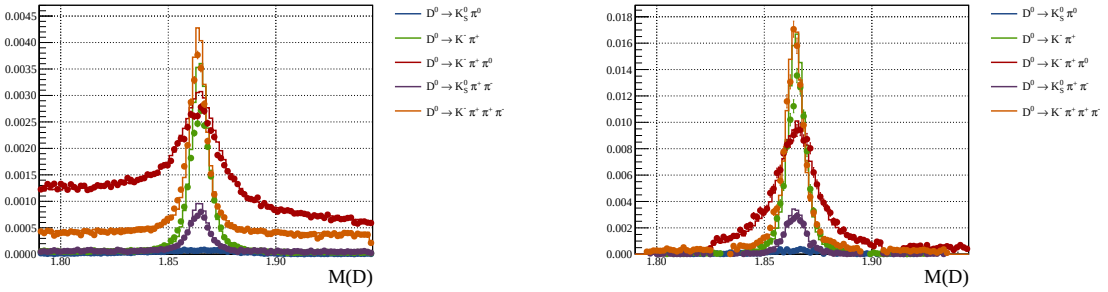
Equation A.3 provides the matrix element $\mathcal{M}$, which can be directly used to calculate the transition rate using Fermi's Golden Rule. The corresponding Feynman Diagram gives a quick and intuitive overview of the physical process.

A.2. D Meson Masses



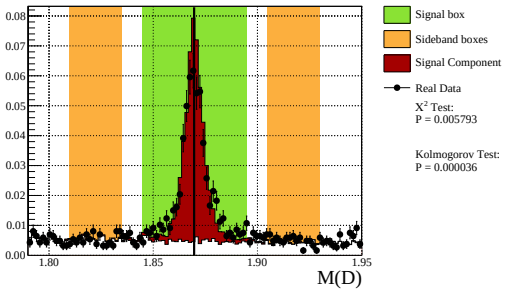
(a) $o_{\mathrm{NB}} > 0.0$ (b) $o_{\mathrm{NB}} > 0.2$

Figure A.2.: D^+ meson masses for all channels used on the Signal Side.

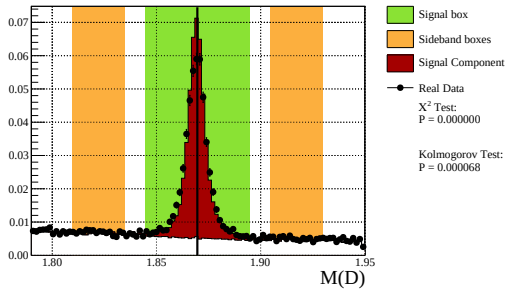


(a) $o_{\mathrm{NB}} > 0.0$ (b) $o_{\mathrm{NB}} > 0.2$

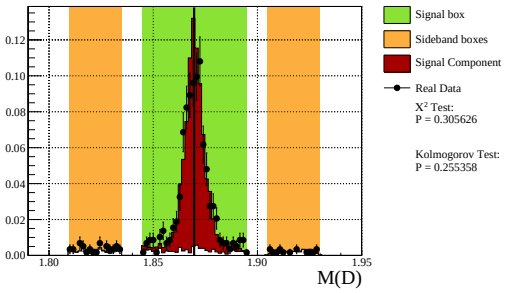
Figure A.3.: D^0 meson masses for all channels used on the Signal Side.



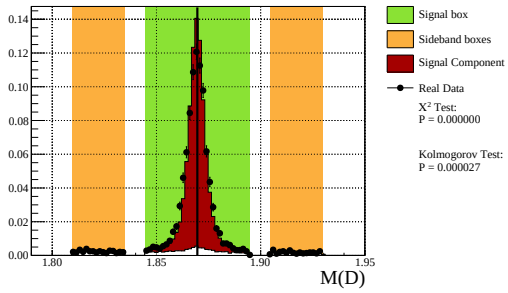
(a) $D^+ \to K^0_S \pi^+$ with $o_{\mathrm{NB}} > 0.0$



(b) $D^+ \to K^- \pi^+ \pi^+$ with $o_{\mathrm{NB}} > 0.0$

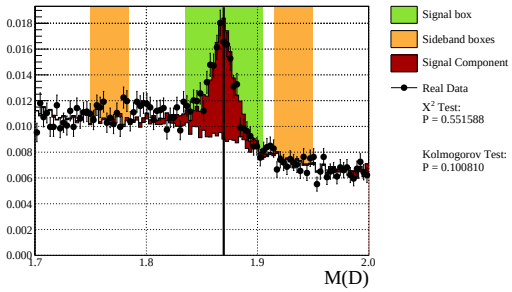


(c) $D^+ \to K^0_S \pi^+$ with $o_{\mathrm{NB}} > 0.2$

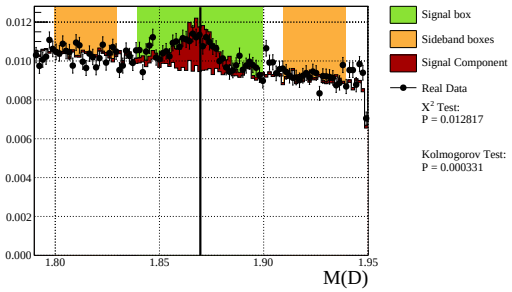


(d) $D^+ \to K^- \pi^+ \pi^+$ with $o_{\mathrm{NB}} > 0.2$

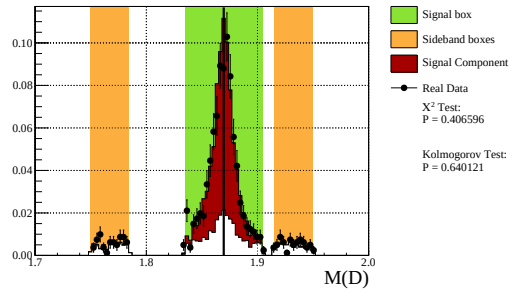
Figure A.4.: D meson mass distributions. Comparison of data (dots with error bars) and generic Monte Carlo simulation (solid line). The signal and the accepted signal and sideband regions are highlighted.



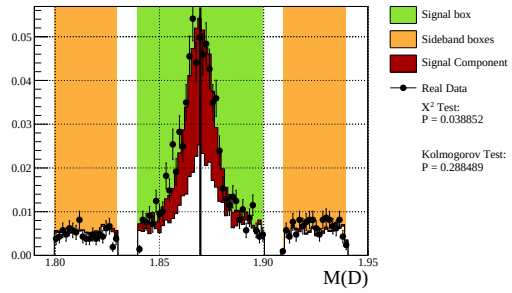
(a) $D^+ \to K^0_S \pi^+ \pi^0$ with $o_{\mathrm{NB}} > 0.0$



(b) $D^+ \to K^- \pi^+ \pi^+ \pi^0$ with $o_{\mathrm{NB}} > 0.0$

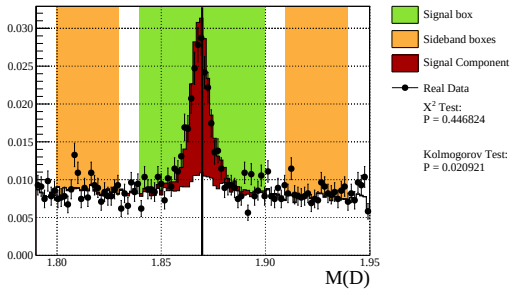


(c) $D^+ \to K^0_S \pi^+ \pi^0$ with $o_{\mathrm{NB}} > 0.2$

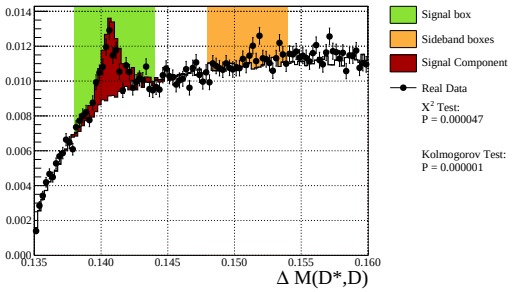


(d) $D^+ \to K^- \pi^+ \pi^+ \pi^0$ with $o_{\mathrm{NB}} > 0.2$

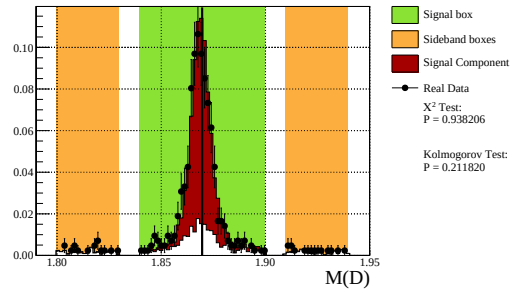
Figure A.5.: D meson mass distributions. Comparison of data (dots with error bars) and generic Monte Carlo simulation (solid line). The signal and the accepted signal and sideband regions are highlighted.



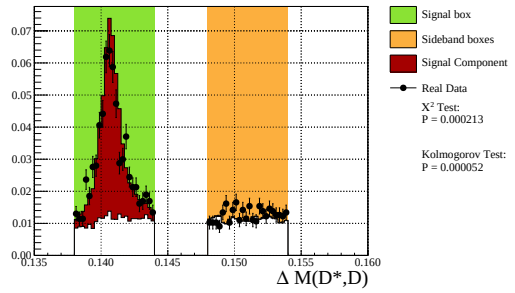
(a) $D^+ \to K^0_S \pi^+ \pi^+ \pi^-$ with $o_{\mathrm{NB}} > 0.0$



(b) $D^{*+} \to D^+ \pi^0$ with $o_{\mathrm{NB}} > 0.0$

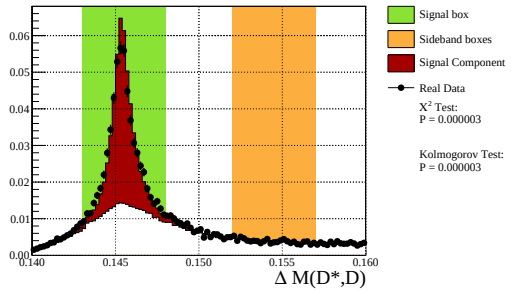


(c) $D^+ \to K^0_S \pi^+ \pi^+ \pi^-$ with $o_{\mathrm{NB}} > 0.2$

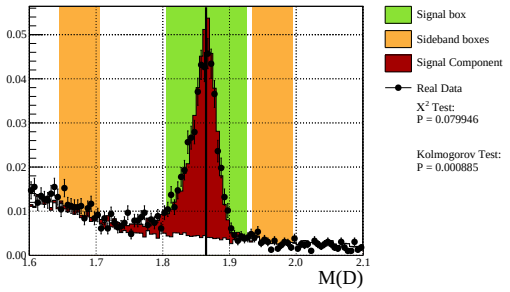


(d) $D^{*+} \to D^+ \pi^0$ with $o_{\mathrm{NB}} > 0.2$

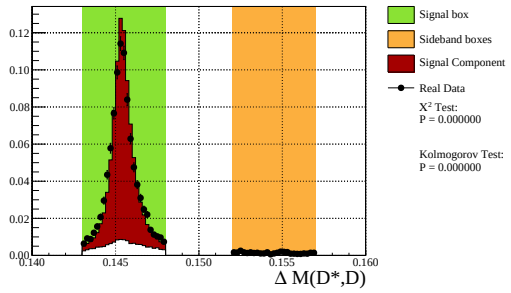
Figure A.6.: D meson mass distributions. Comparison of data (dots with error bars) and generic Monte Carlo simulation (solid line). The signal and the accepted signal and sideband regions are highlighted.



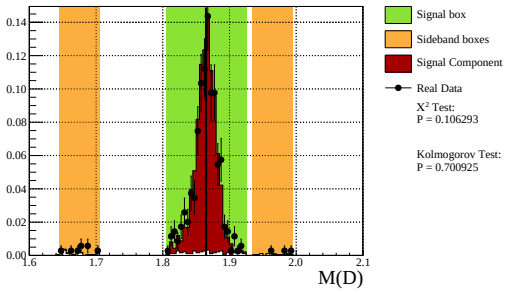
(a) $D^{*+} \to D^0\pi^+$ with $o_{\mathrm{NB}} > 0.0$



(b) $D^0 \to K^0_S\pi^0$ with $o_{\mathrm{NB}} > 0.0$

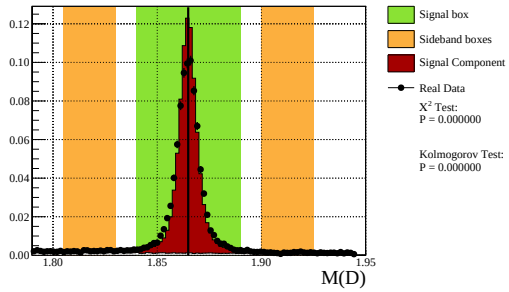


(c) $D^{*+} \to D^0\pi^+$ with $o_{\mathrm{NB}} > 0.2$

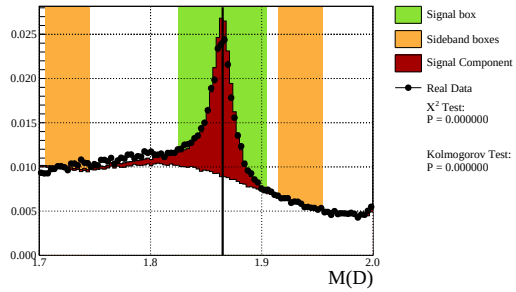


(d) $D^0 \to K^0_S\pi^0$ with $o_{\mathrm{NB}} > 0.2$

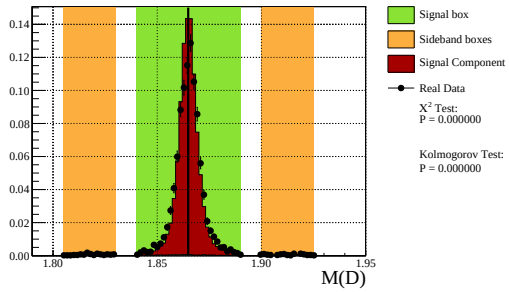
Figure A.7.: D meson mass distributions. Comparison of data (dots with error bars) and generic Monte Carlo simulation (solid line). The signal and the accepted signal and sideband regions are highlighted.



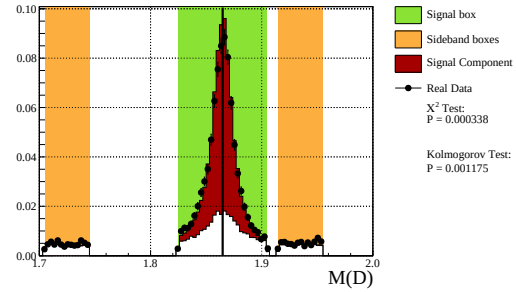
(a) $D^0 \to K^-\pi^+$ with $o_{\mathrm{NB}} > 0.0$



(b) $D^0 \to K^-\pi^+\pi^0$ with $o_{\mathrm{NB}} > 0.0$

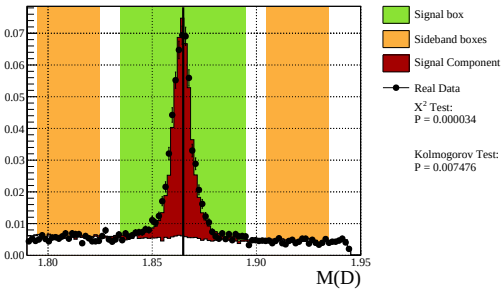


(c) $D^0 \to K^-\pi^+$ with $o_{\mathrm{NB}} > 0.2$

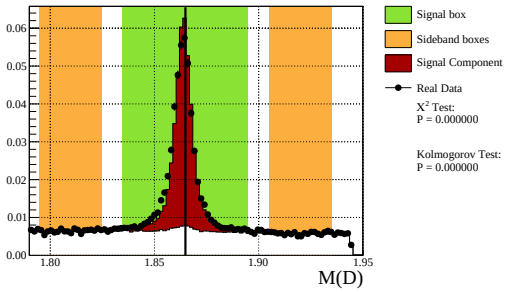


(d) $D^0 \to K^-\pi^+\pi^0$ with $o_{\mathrm{NB}} > 0.2$

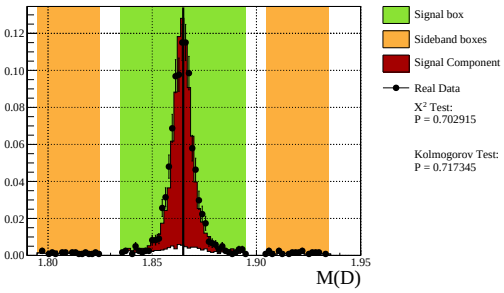
Figure A.8.: *D* meson mass distributions. Comparison of data (dots with error bars) and generic Monte Carlo simulation (solid line). The signal and the accepted signal and sideband regions are highlighted.



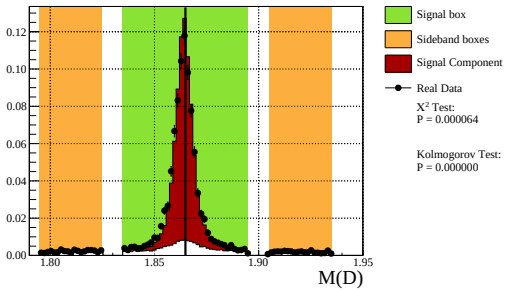
(a) $D^0 \to K^0_S \pi^+ \pi^-$ with $o_{\mathrm{NB}} > 0.0$



(b) $D^0 \to K^- \pi^+ \pi^+ \pi^-$ with $o_{\mathrm{NB}} > 0.0$

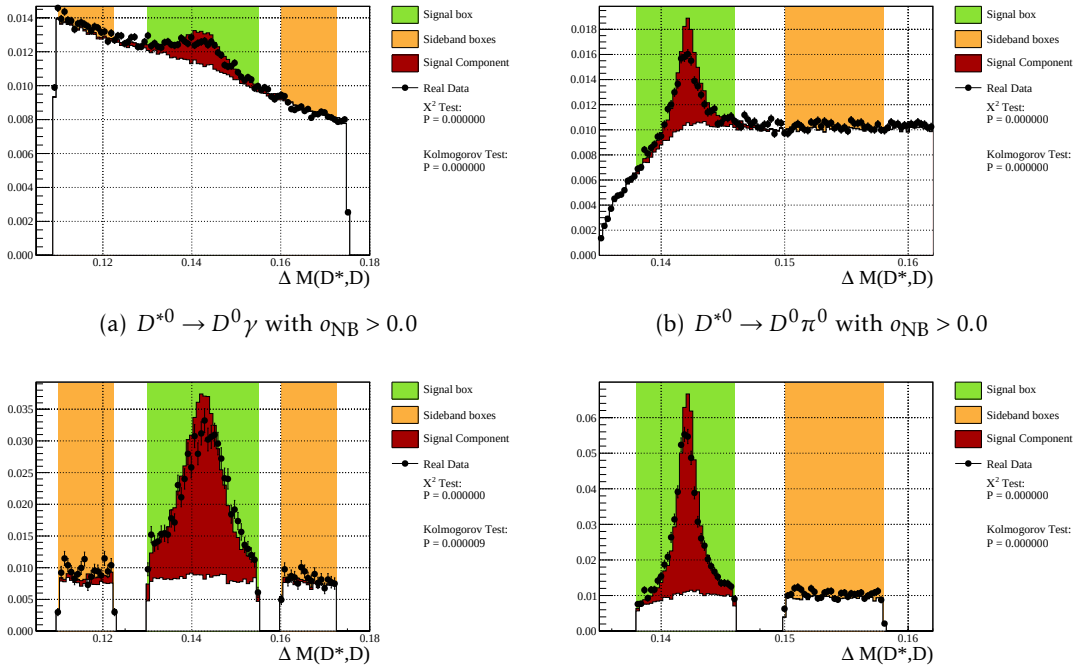


(c) $D^0 \to K^0_S \pi^+ \pi^-$ with $o_{\mathrm{NB}} > 0.2$



(d) $D^0 \to K^- \pi^+ \pi^+ \pi^-$ with $o_{\mathrm{NB}} > 0.2$

Figure A.9.: D meson mass distributions. Comparison of data (dots with error bars) and generic Monte Carlo simulation (solid line). The signal and the accepted signal and sideband regions are highlighted.



(a) $D^{*0} \to D^0\gamma$ with $o_{\mathrm{NB}} > 0.0$

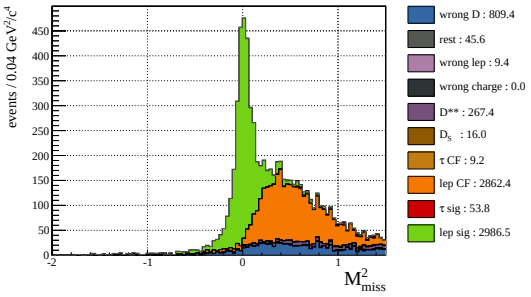
(b) $D^{*0} \to D^0\pi^0$ with $o_{\mathrm{NB}} > 0.0$

(c) $D^{*0} \to D^0\gamma$ with $o_{\mathrm{NB}} > 0.2$

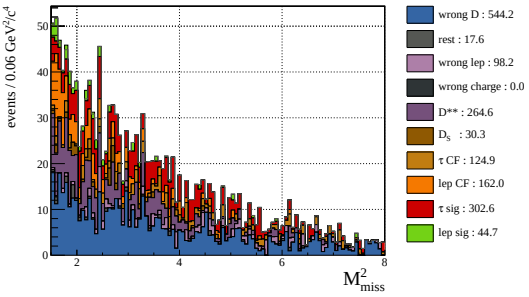
(d) $D^{*0} \to D^0\pi^0$ with $o_{\mathrm{NB}} > 0.2$

Figure A.10.: D meson mass distributions. Comparison of data (dots with error bars) and generic Monte Carlo simulation (solid line). The signal and the accepted signal and sideband regions are highlighted.

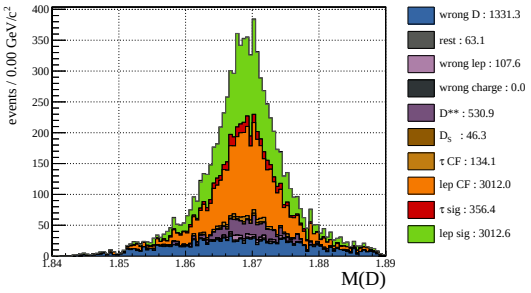
A.3. Signal Side Composition



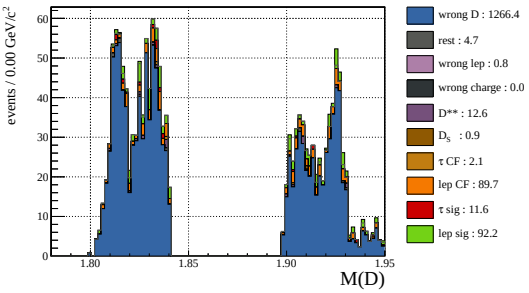
(a) M^2_{miss} in region A



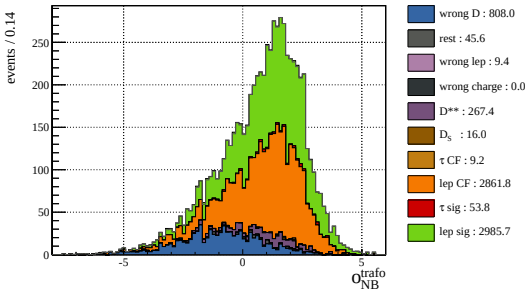
(b) M^2_{miss} in region B



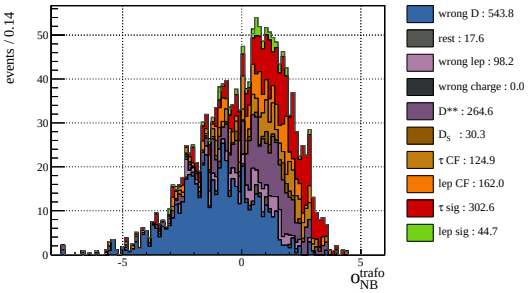
(c) $M(D)$ in signal region



(d) $M(D)$ in sideband region

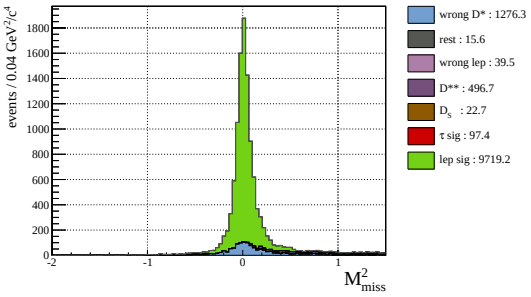


(e) $o_{\text{NB, trafo.}}$ in region A

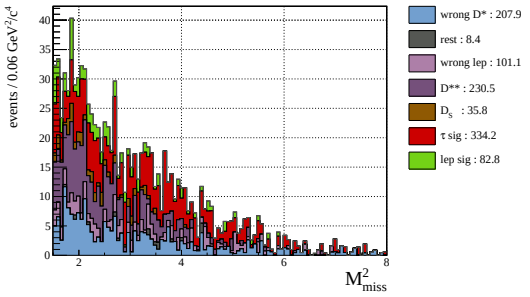


(f) $o_{\text{NB, trafo.}}$ in region B

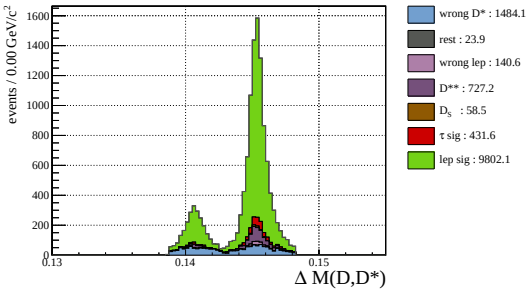
Figure A.11.: Background composition for $B^0 \to D^- \ell^+ \nu_\ell$, evaluated on 5 streams of generic Monte Carlo



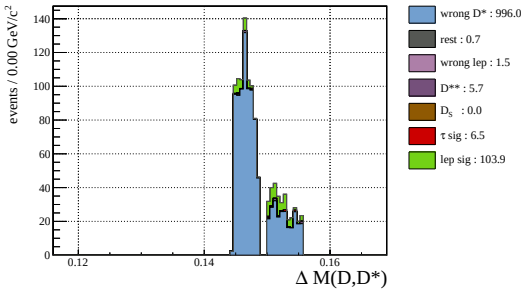
(a) M^2_{miss} in region A



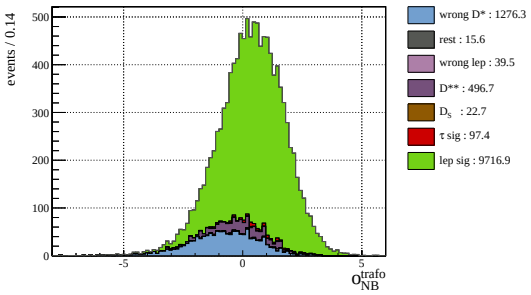
(b) M^2_{miss} in region B



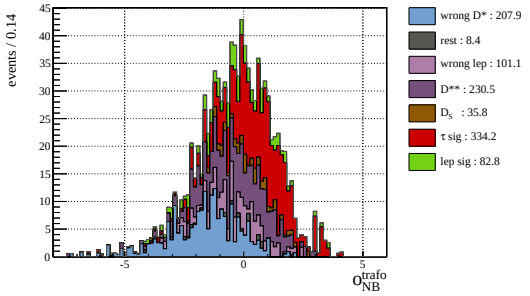
(c) $\Delta M(D^*, D)$ in signal region



(d) $\Delta M(D^*, D)$ in sideband region

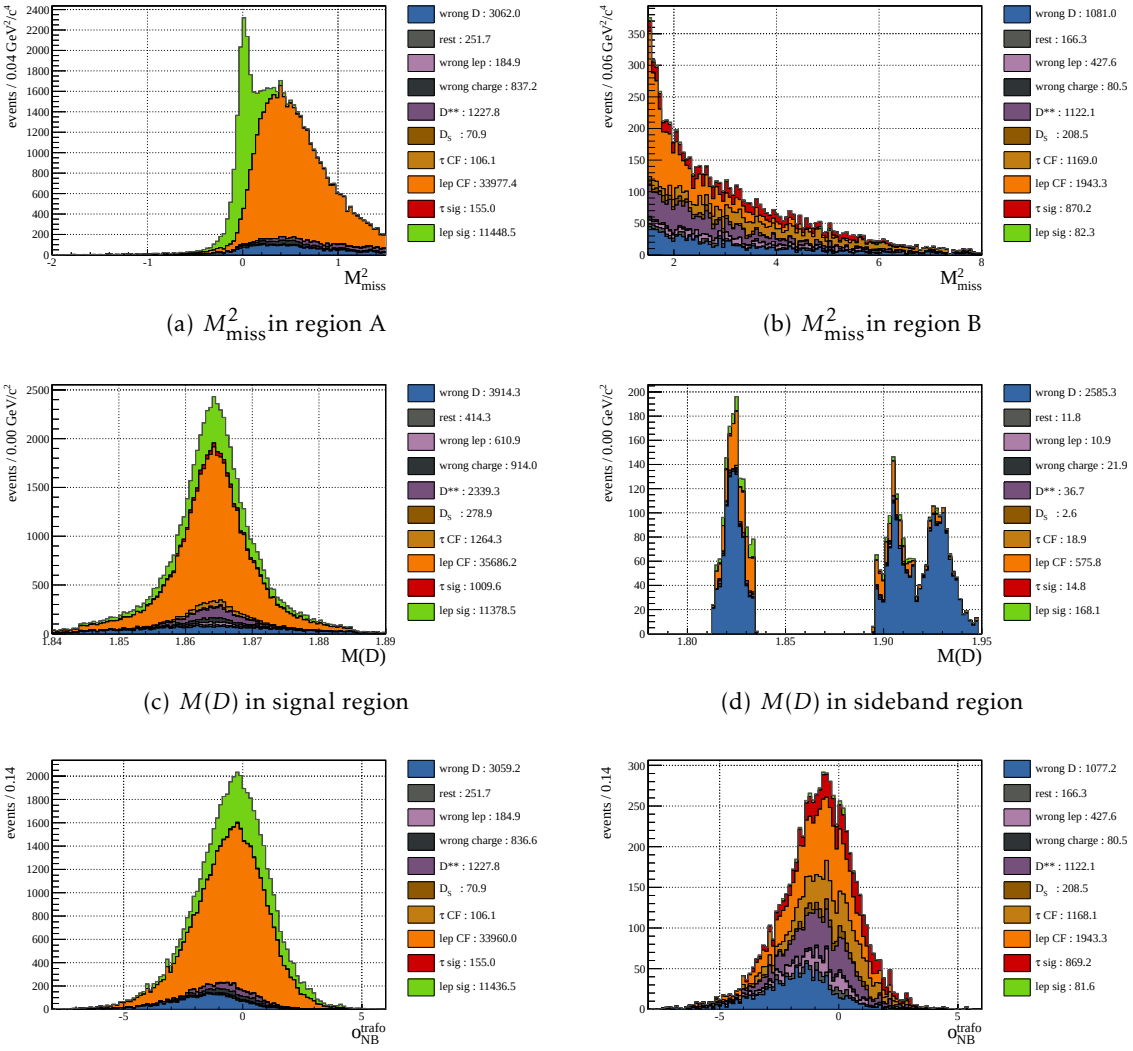


(e) $o_{\text{NB, trafo.}}$ in region A



(f) $o_{\text{NB, trafo.}}$ in region B

Figure A.12.: Background composition for $B^0 \to D^{*-}\ell^+\nu_\ell$, evaluated on 5 streams of generic Monte Carlo



(a) M^2_{miss} in region A

(b) M^2_{miss} in region B

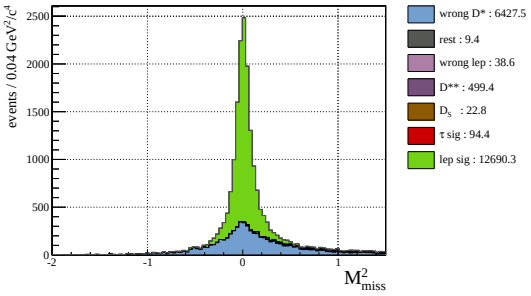
(c) $M(D)$ in signal region

(d) $M(D)$ in sideband region

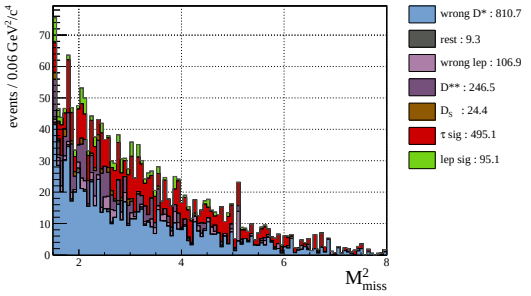
(e) $o_{\text{NB, trafo.}}$ in region A

(f) $o_{\text{NB, trafo.}}$ in region B

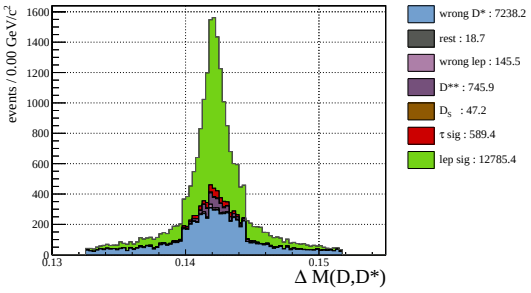
Figure A.13.: Background composition for $B^+ \to \overline{D}^0 \ell^+ \nu_\ell$, evaluated on 5 streams of generic Monte Carlo



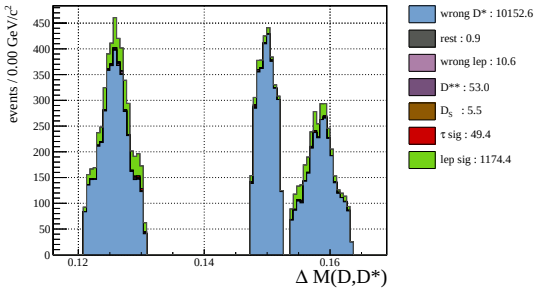
(a) M^2_{miss} in region A



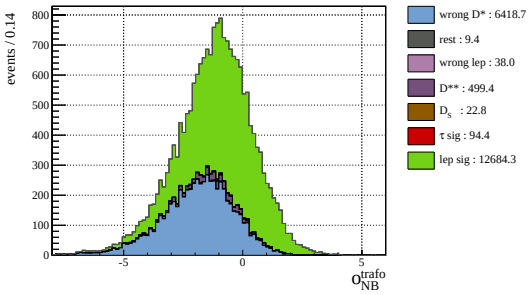
(b) M^2_{miss} in region B



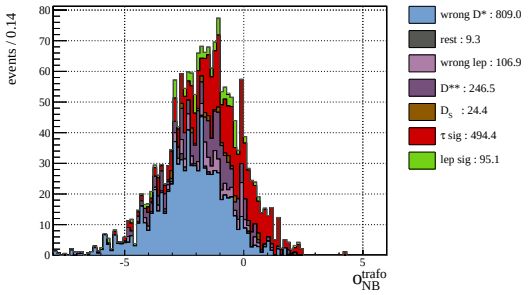
(c) $\Delta M(D^*, D)$ in signal region



(d) $\Delta M(D^*, D)$ in sideband region



(e) $o_{\text{NB, trafo.}}$ in region A



(f) $o_{\text{NB, trafo.}}$ in region B

Figure A.14.: Background composition for $B^+ \to \overline{D}^{*0} \ell^+ \nu_\ell$, evaluated on 5 streams of generic Monte Carlo

A.4. Signal-Background Curves

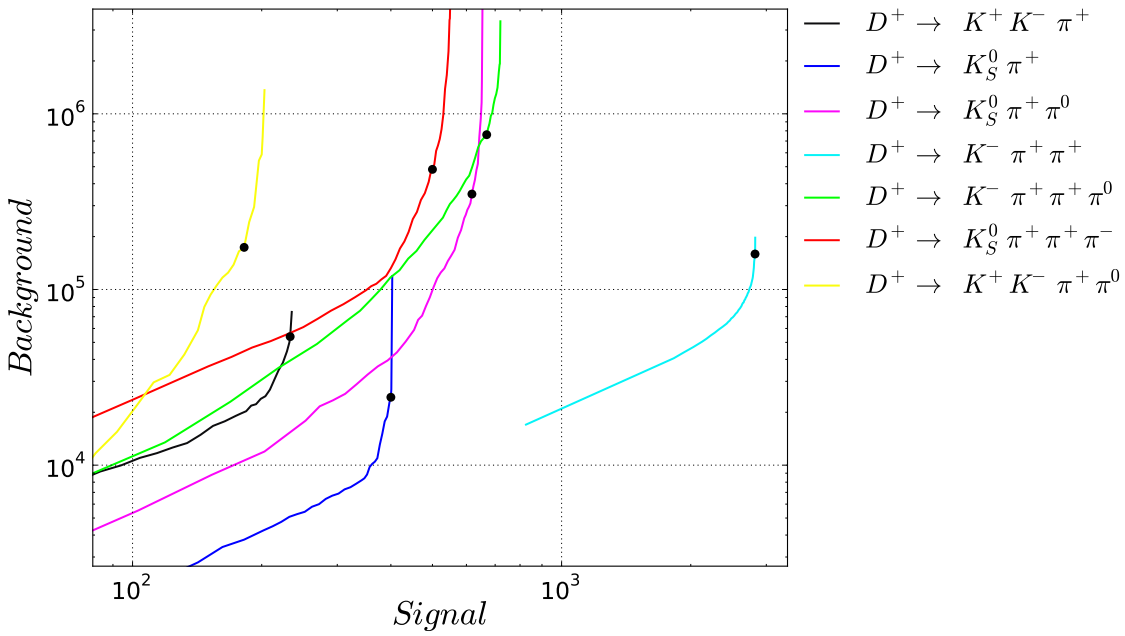


Figure A.15.: Signal and background for different cuts on o_{prod} for all D^+ decay channels in the Full Reconstruction

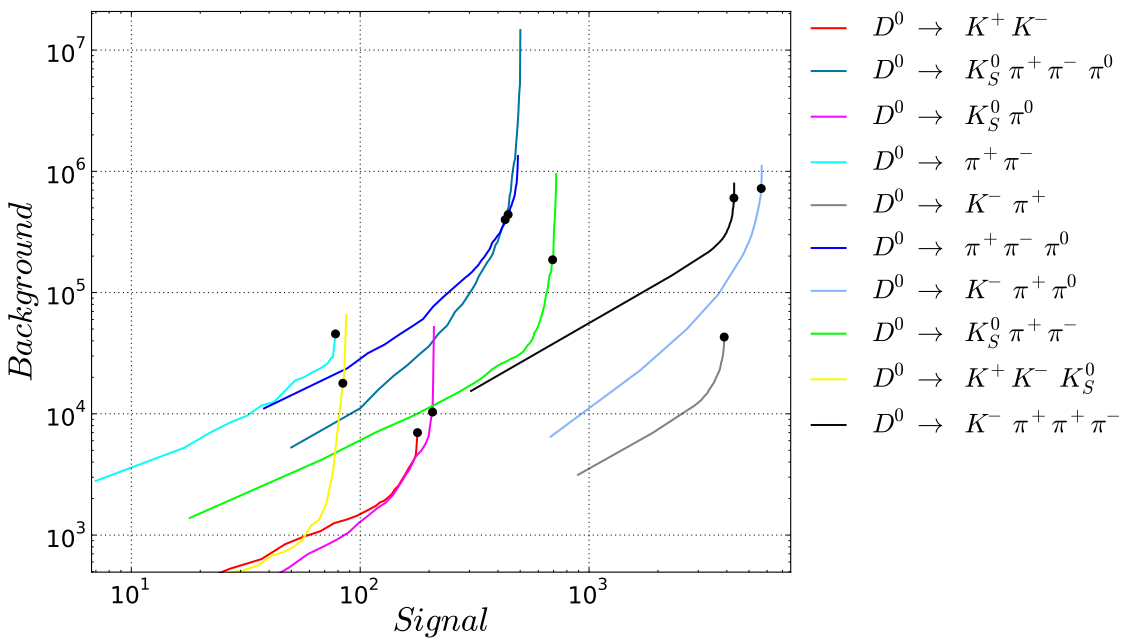


Figure A.16.: Signal and background for different cuts on o_{prod} for all D^0 decay channels in the Full Reconstruction

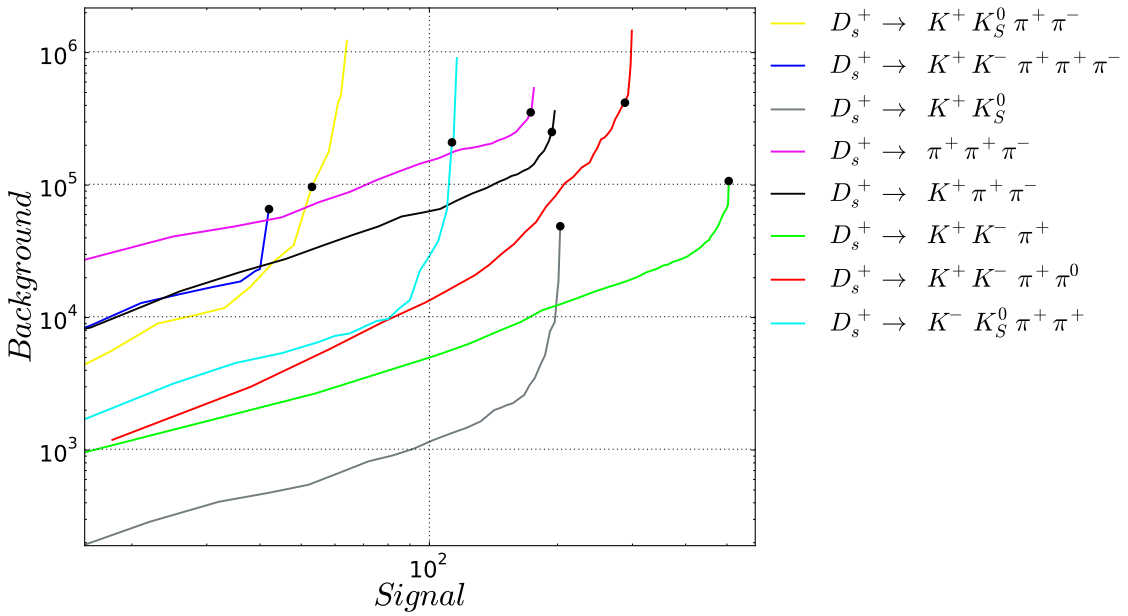


Figure A.17.: Signal and background for different cuts on o_{prod} for all D_S decay channels in the Full Reconstruction

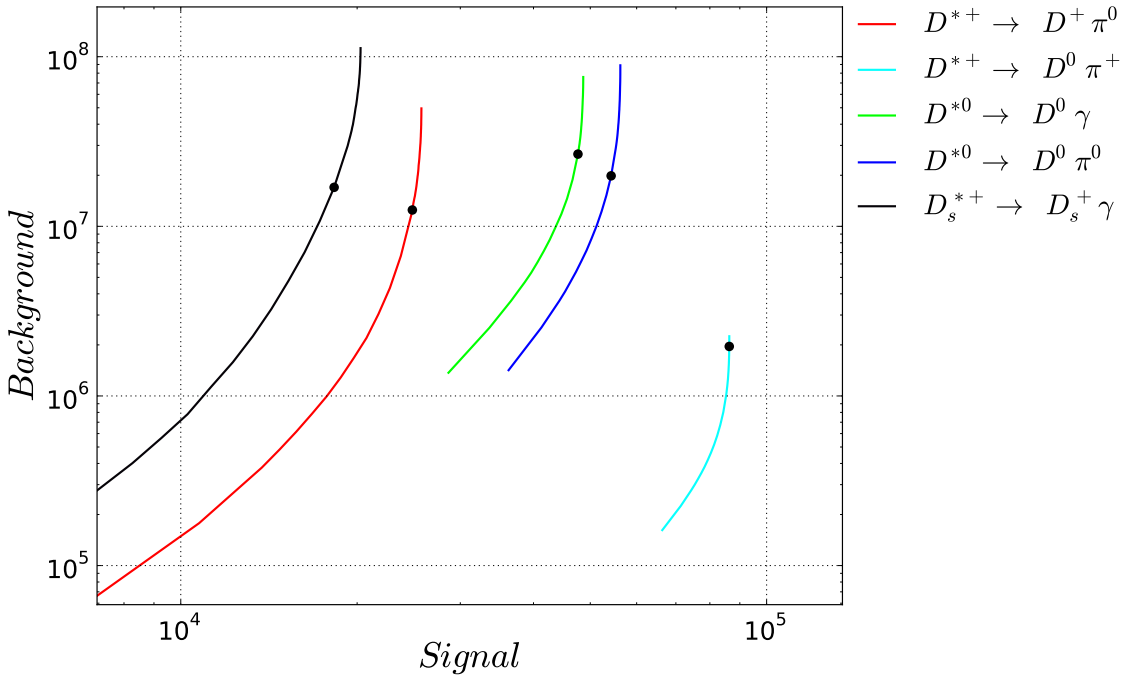


Figure A.18.: Signal and background for different cuts on o_{prod} for all $D^*_{(S)}$ decay channels in the Full Reconstruction

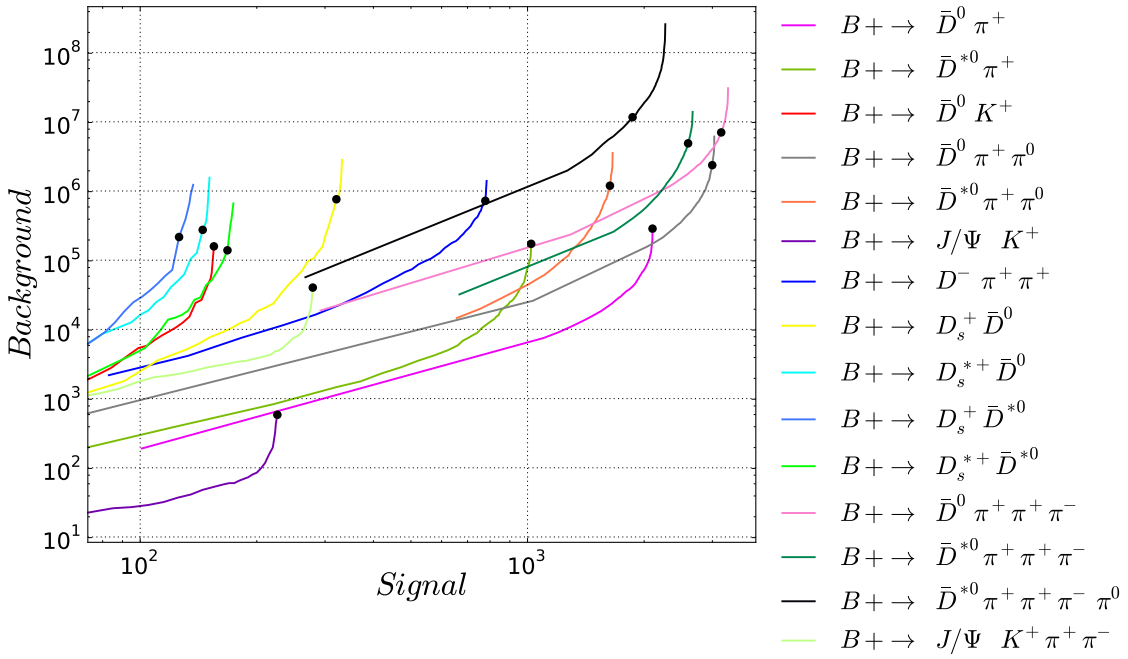


Figure A.19.: Signal and background for different cuts on o_{prod} for all B^+ decay channels in the Full Reconstruction

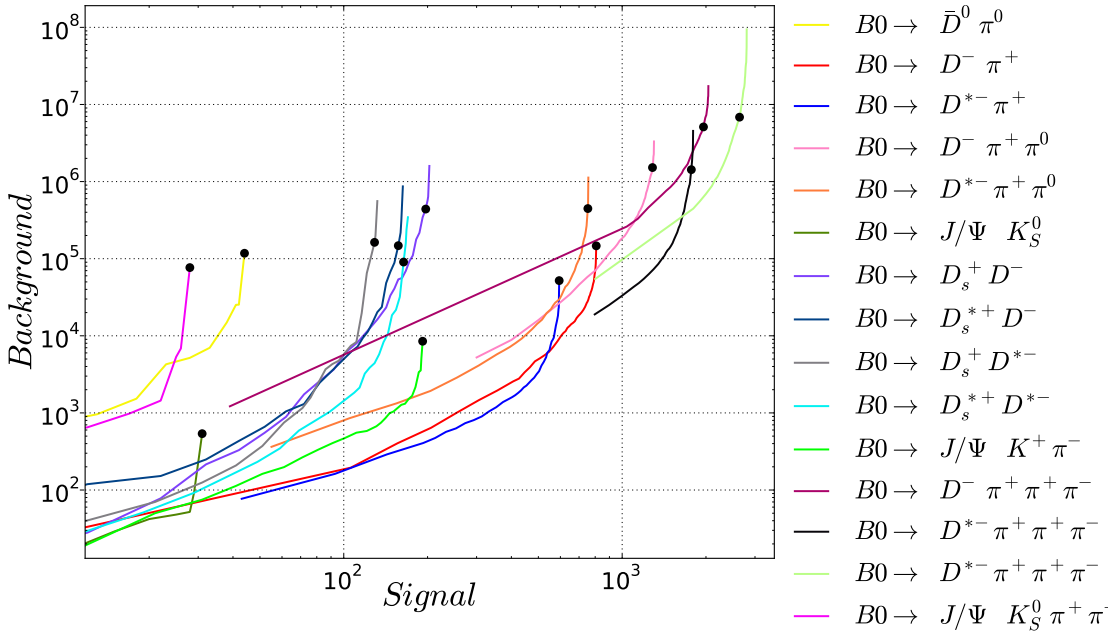


Figure A.20.: Signal and background for different cuts on o_{prod} for all B^0 decay channels in the Full Reconstruction

A.5. NeuroBayes Trainings of the Full Reconstruction

Name	Variable
ATC_PID_binned_23	Particle ID π^+ vs K^+
dEdx_pull	$[(dE/dx)_{\text{measured}} - (dE/dx)_{\text{expected},\pi}]/\sigma(dE/dx)$
PID_eid	Particle ID e^+
PID_muid	Particle ID μ^+
PID_eid_flag	Flag, if Particle ID e^+ exists
ATC_PID_binned_21	Particle ID π^+ vs μ^+
acc_pid	PID information from ACC only
tof_pid	PID information from TOF only
pt	p_t
ATC_PID_binned_24	Particle ID π^+ vs p^+
ATC_PID_binned_20	Particle ID π^+ vs e^+
PID_muid_flag	Flag, if Particle ID μ^+ exists
dEdx_ratio	$(dE/dx)_{\text{measured}}/(dE/dx)_{\text{expected},\mu}$
energy	E
trk_pid	PID information from TRK only
ptot	$\|p\|$

Table A.1.: Variables used in the $\pi^{\pm}$ training

Name	Variable
ATC_PID_binned_32	Particle ID K^+ vs π^+
ATC_PID_binned_34	Particle ID K^+ vs p^+
acc_pid	PID information from ACC only
tof_pid	PID information from TOF only
energy	E
dEdx_pull	$[(dE/dx)_{\text{measured}} - (dE/dx)_{\text{expected},K}]/\sigma(dE/dx)$
PID_eid_flag	Flag, if Particle ID e^+ exists
PID_eid	Particle ID e^+
trk_pid	PID information from TRK only
ATC_PID_binned_30	Particle ID K^+ vs e^+
PID_muid	Particle ID μ^+
pt	p_t
PID_muid_flag	Flag, if Particle ID μ^+ exists
dEdx_ratio	$(dE/dx)_{\text{measured}}/(dE/dx)_{\text{expected},K}$
ATC_PID_binned_31	Particle ID K^+ vs μ^+
ptot	$\|p\|$

Table A.2.: Variables used in the $K^{\pm}$ training

Name	Variable
dPhi	$\Delta\phi$ from K_s finder
mass	m
dr	Δr from K_s finder
Z_dist	$Zdist$ of the two helices
Intersect	Type of intersection of the two helices
PMag	$\|p\|$ from K_s finder
ptot	$\|p\|$
pz	p_z
px	p_x
py	p_y
Chi2	χ^2 of the fit
energy	E
pt	p_t

Table A.3.: Variables used in the K_S^0 training

Name	Variable
EA_seed	ECL AUX energy of seed cell
EA_width	ECL AUX RMS shower width
EA_e9	ECL AUX E9/E25: photon-like & isolation cut
energy	E
dPhi	$\Delta\phi$ from K_s finder (converted photons only)
ECL_en	ECL energy in the cluster
EA_cID	ECL AUX ID of seed cell
EA_nhits	ECL AUX crystals in the cluster
EA_e9unf	ECL AUX E9/E25: photon-like & isolation cut unfolded
Z_dist	Distance in z of the two helices (converted photons only)
dr	Δr from K_s finder (converted photons only)
PMag	$\|p\|$ from K_s finder (converted photons only)
pt	p_t
Intersect	Type of intersection of the two helices (converted photons only)
Ks_goodKs	goodKs Variable from K_s finder (converted photons only)
ptot	$\|p\|$
Chi2	χ^2 of the fit (converted photons only)

Table A.4.: Variables used in the γ training

Name	Variable
PiOChildrenAngle	Angle of the two γ coming from π^0 in the lab frame
PiOMass	m
energy	E
Pi01EA_seed	ECL AUX γ_1 energy of seed cell
Pi00EA_seed	ECL AUX γ_0 energy of seed cell
Pi00EA_cID	ECL AUX γ_0 ID of seed cell
PiOGammaMatch	ECL γ_0 and γ_1 info if clusters match with tracks from CDC
Pi01EA_width	ECL AUX γ_1 RMS shower width
Pi00EA_width	ECL AUX γ_0 RMS shower width
pt	p_t
Pi01EA_cID	ECL AUX γ_1 ID of seed cell
Pi00EA_e9	ECL AUX γ_0 E9/E25: photon-like & isolation cut
Pi01EA_e9	ECL AUX γ_1 E9/E25: photon-like & isolation cut
Pi01EA_e9unf	ECL AUX γ_1 E9/E25: photon-like & isolation cut unfolded
Pi00EA_e9unf	ECL AUX γ_0 E9/E25: photon-like & isolation cut unfolded
Pi01EA_nhits	ECL AUX crystals in γ_1 cluster
Pi00ECL_en	ECL γ_0 energy in the cluster
Pi01ECL_en	ECL γ_1 energy in the cluster
Pi00EA_nhits	ECL AUX crystals in γ_0 cluster
ptot	$\|p\|$

Table A.5.: Variables used in the π^0 training

Name	Variable
PID_eid	Particle ID e^+
ATC_PID_binned_02	Particle ID e^+ vs. π^+ binned
trk_pid	PID information from TRK only
pt	p_t
ATC_PID_binned_01	Particle ID e^+ vs. μ^+ binned
ATC_PID_binned_03	Particle ID e^+ vs. K^+ binned
PID_eid_flag	Flag, if Particle ID e^+ exists
acc_pid	PID information from ACC only
ATC_PID_binned_04	Particle ID e^+ vs. p^+ binned
tof_pid	PID information from TOF only
energy	E
dEdx_pull	$[(dE/dx)_{\text{measured}} - (dE/dx)_{\text{expected},e}]/\sigma(dE/dx)$
dEdx_ratio	$(dE/dx)_{\text{measured}}/(dE/dx)_{\text{expected},e}$
PID_muid	Particle ID μ^+, if it exists
PID_muid_flag	Flag, if Particle ID μ^+ exists
ptot	$\|p\|$

Table A.6.: Variables used in the $e^{\pm}$ training

Name	Variable
PID_muid	Particle ID μ^+
ATC_PID_binned_12	Particle ID μ^+ vs π^+
KLM_likelihood4	KLM Miss likelihood binned
dEdx_pull	$[(dE/dx)_{\text{measured}} - (dE/dx)_{\text{expected},\mu}]/\sigma(dE/dx)$
PID_eid	Particle ID e^+
energy	E
ATC_PID_binned_10	Particle ID μ^+ vs π^+
dEdx_ratio	$(dE/dx)_{\text{measured}}/(dE/dx)_{\text{expected},\mu}$
KLM_likelihood1	KLM Muon likelihood binned
trk_pid	PID information from TRK only
ATC_PID_binned_14	Particle ID μ^+ vs p^+
KLM_likelihood5	KLM Junk likelihood binned
KLM_likelihood2	KLM Pion likelihood binned
pt	p_t
KLM_Chi2	KLM χ^2 of the associated hits
KLM_likelihood3	KLM Kaon likelihood binned
PID_eid_flag	Flag, if Particle ID e^+ exists
tof_pid	PID information from TOF only
acc_pid	PID information from ACC only
KLM_Outcome	KLM Outcome information how far the track came in KLM
ATC_PID_binned_13	Particle ID μ^+ vs K^+
PID_muid_flag	Flag, if Particle ID μ^+ exists
ptot	$\|p\|$

Table A.7.: Variables used in the $\mu^\pm$ training

Name	Variable
sumChldNB	Sum of the NeuroBayes outputs of all of the children
prodChildNB	Product of the NeuroBayes outputs of all of the children
ptot	$\|p\|$
ChN_ptot	$\|p\|$ of child N
ChN_NBout	NeuroBayes output of child N
ChN_PseudoHelAng	Angle between the child momentum in the mother's rest frame and the mother's direction in the $\Upsilon(4S)$ rest frame. For D meons, this really is a good approximation of the helicity angle.
ChN_hash	Decay hash of child N
ChN_Mass	Mass of child N
ChMN_Angle	Angle between children N and M in the CMS
ChMN_InvMassScaled	Invariant mass of children N and M, scaled to the maximum (=1) and minimum (=0) possible theoretical value.
mom_dir_dev	Angle between the momentum of the unstable meson and the line connecting the IP and the fitted decay vertex.
dist_to_IP	Distance of the fitted decay vertex and the IP
sig_dist_to_IP	Significance of the distance of the fitted decay vertex and the IP
deltaE	ΔE
Dstar_D_massdiff	$M(D^*) - M(D)$
D_hash_from_1st_dstar	Decay hash from the first D^* meson
D_hash_from_2nd_dstar	Decay hash from the second D^* meson
CosThetaB	$\cos\theta_B$
ChN_D_dist	Distance of closest approach of a track from the B decay to the fitted vertex of the D meson

Table A.8.: Variables used in the stage 2, 3 and 4 trainings

Bibliography

[1] Donald H. Perkins. *Introduction to High Energy Physics*. Cambridge Univ. Press, Cambridge, New York, Oakleigh, Alarcón, 4. ed., 1. publ. edition, 2000.

[2] K. Nakamura et al. Review of particle physics. *Journal of Physics G*, 37, 075021, 2010.

[3] David J. Griffiths. *Introduction to Elementary Particles*. Textbook Physics. Wiley-VCH, Weinheim, 2004.

[4] P. A. M. Dirac. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A*, 117(778):610–624, 1928.

[5] Bogdan Povh, Klaus Rith, Christoph Scholz, and Frank Zetsche. *Teilchen und Kerne : Eine Einführung in die physikalischen Konzepte*. Springer-Lehrbuch. Springer, Berlin, 7. aufl. edition, 2006.

[6] H. Burkard, F. Corriveau, J. Egger, W. Fetscher, H.-J. Gerber, K.F. Johnson, H. Kaspar, H.J. Mahler, M. Salzmann, and F. Scheck. Muon decay: Measurement of the transverse positron polarization and general analysis. *Physics Letters B*, 160(4–5):343 – 348, 1985.

[7] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970.

[8] Lincoln Wolfenstein. Parametrization of the Kobayashi-Maskawa Matrix. *Phys.Rev.Lett.*, 51:1945, 1983.

[9] Sheldon [Hrsg.] Stone, editor. *b decays*. World Scientific, Singapore [u.a.], 2. ed., rev. edition, 1994.

[10] P. Abreu et al. Cross-sections and leptonic forward backward asymmetries from the Z0 running of LEP. *Eur.Phys.J.*, C16:371–405, 2000.

[11] Gerhard [Mitarb.] Musiol, editor. *Kern- und Elementarteilchenphysik*. VCH, Weinheim, 1988.

[12] John F. Gunion, editor. *The Higgs Hunter's Guide*. Frontiers in physics ; 80. Addison-Wesley, Redwood City, Calif., 1. print. edition, 1996.

[13] The ATLAS collaboration. Observation of an excess of events in the search for the standard model higgs boson with the atlas detector at the lhc. *36th International Conference on High Energy Physics, Melbourne, Australia, 4 - 11 Jul 2012*.

[14] The CMS collaboration. Observation of a new boson with a mass near 125 gev. *CMS-PAS-HIG-12-020*, 2012.

[15] Sheldon Stone. New physics from flavour. 2012.

[16] Ulrich Nierste. Higgs hunting with b decays. *Nuclear Physics B - Proceedings Supplements*, 185(0):195 – 200, 2008. Proceedings of the Second Workshop on Theory, Phenomenology and Experiments in Heavy Flavour Physics.

[17] Stephanie Trine. Charged-Higgs effects in $B \to D\tau\nu$ decays. *arXiv.org:0810.3633*, 2008. 5 pages, parallel talk given at the 34th International Conference on High Energy Physics (ICHEP08), 29 July - 5 August 2008, Philadelphia, USA.

[18] Minoru Tanaka and Ryoutaro Watanabe. New physics in the weak interaction of $\bar{B} \to D^{(*)}\tau\bar{\nu}$. 2012.

[19] Minoru Tanaka and Ryoutaro Watanabe. Tau longitudinal polarization in $\bar{B} \to D\tau\bar{\nu}$ and its role in the search for charged Higgs boson. *arXiv.org:1005.4306*, 2010.

[20] I. Adachi et al. Measurement of $B^- \to \tau^-\bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle. 2012.

[21] Ricciardi, Giulia. Semileptonic B and D decays-latest developments. *arXiv:1209.5650v1 [hep-ph]*, 2012.

[22] Ulrich Nierste, Stephanie Trine, and Susanne Westhoff. Charged-Higgs effects in a new $B \to D\tau\nu_\tau$ differential decay distribution. *arXiv.org:0801.4938*, 2008.

[23] Volker Blobel and Erich Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner-Studienbücher : Physik. Teubner, Stuttgart, 1998.

[24] John R. Taylor. *An introduction to error analysis : the study of uncertainties in physical measurements*. University Science, Sausalito, Calif., 2. ed. edition, 1997.

[25] Christopher M. Bishop. *Neural networks for pattern recognition*. Oxford Univ. Pr., Oxford, repr. edition, 2003.

[26] Michael Feindt. A neural bayesian estimator for conditional probability densities. *arXiv.org:physics/0402093*, 2004.

[27] Sebastian Neubauer. *Search for B → K(*) Neutrino Anti-Neutrino Decays Using a New Probabilistic Full Reconstruction Method*. PhD thesis, 2011.

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

[29] David J. C. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press, Cambridge [u.a.], 2003.

[30] Frank Hinterberger. *Physik der Teilchenbeschleuniger und Ionenoptik*. Springer, Berlin, 2. aufl. edition, 2008.

[31] KEK Report. KEK B B-factory design report. `http://www-acc.kek.jp/kekb/publication/KEKB_design_report/KEKBDesignReport.html`, 1995. [Online; accessed 13-May-2011].

[32] S. Kurokawa. Overview of the KEKB accelerators. *Nucl.Instrum.Meth.*, A499:1–7, 2003.

[33] Tadao Nozaki. The kek b-factory and the belle detector. *Nuclear Physics B - Proceedings Supplements*, 50(1-3):288 – 295, 1996. Proceedings of the Conference on Production and Decay of Hyperons, Charm and Beauty Hadrons.

[34] M. T. Cheng et al. A Study of CP violation in B meson decays: Technical design report. BELLE-TDR-3-95.

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

[36] R. Abe, T. Abe, H. Aihara, Y. Asano, T. Aso, et al. The new beampipe for the Belle experiment. *Nucl.Instrum.Meth.*, A535:558–561, 2004.

[37] Y. Ushiroda. Belle silicon vertex detectors. *Nucl.Instrum.Meth.*, A511:6–10, 2003.

[38] H. Ishino. The belle silicon vertex detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 549(1–3):16 – 19, 2005. Proceedings of the 12th International Workshop on Vertex Detectors.

[39] Z. Natkaniec, H. Aihara, Y. Asano, T. Aso, A. Bakich, et al. Status of the Belle silicon vertex detector. *Nucl.Instrum.Meth.*, A560:1–4, 2006.

[40] K. Miyabayashi. Belle electromagnetic calorimeter. *Nucl.Instrum.Meth.*, A494:298–302, 2002.

[41] J.G. Wang. RPC performance at KLM / BELLE. *Nucl.Instrum.Meth.*, A508:133–136, 2003.

[42] E. Won. Upgrade of the level 1 global trigger in the Belle experiment. *IEEE Trans.Nucl.Sci.*, 55:122–125, 2008.

[43] Paolo Franzini and Juliet Lee-Franzini. Upsilon Resonances. *Ann. Rev. Nucl. Part. Sci.*, 33:1–29, 1983.

[44] P. del Amo Sanchez et al. Evidence for $B^+ \to \tau^+ \nu_\tau$ decays using hadronic B tags. 2010.

[45] I. Adachi et al. Measurement of $B^- \to \tau^- \bar{\nu}_\tau$ with a Hadronic Tagging Method Using the Full Data Sample of Belle. 2012.

[46] K Hanagaki, H Kakuno, H Ikeda, T Iijima, and T Tsukamoto. Electron identification in belle. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 485(3):490 – 503, 2002.

[47] A Abashian, K Abe, K Abe, P.K Behera, F Handa, T Iijima, Y Inoue, H Miyake, T Nagamine, E Nakano, S Narita, L Piilonen, S Schrenk, Y Teramoto, K Trabelsi, J.G Wang, M Yamaga, A Yamaguchi, and Y Yusa. Muon identification in the belle experiment at kekb. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 491(1-2):69 – 82, 2002.

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

[49] Geoffrey C. Fox and Stephen Wolfram. Event Shapes in e+ e- Annihilation. *Nucl.Phys.*, B149:413, 1979.

[50] Geoffrey C. Fox and Stephen Wolfram. Observables for the Analysis of Event Shapes in e+ e- Annihilation and Other Processes. *Phys.Rev.Lett.*, 41:1581, 1978.

[51] K. Abe et al. Evidence for $B^0 \to \pi^0\pi^0$. *Phys.Rev.Lett.*, 91:261801, 2003.

[52] A. Matyja et al. Observation of $B^0 \to D^{*-}\tau^+ nu_\tau$ decay at Belle. *Phys.Rev.Lett.*, 99:191807, 2007.

[53] A. Bozek et al. Observation of $B^+ \to \bar{D}^{*0}\tau^+\nu_\tau$ and Evidence for $B^+ \to \bar{D}^0\tau^+\nu_\tau$ at Belle. *Phys.Rev.*, D82:072005, 2010.

[54] I. Adachi et al. Measurement of $B \to D^{(*)}\tau\nu$ using full reconstruction tags. *arXiv.org:0910.4301*, 2009.

[55] Bernard Aubert et al. Observation of the semileptonic decays $B \to D^*\tau^-\bar{\nu}$ and evidence for $B \to D\tau^-\bar{\nu}$. *Phys.Rev.Lett.*, 100:021801, 2008.

[56] J.P. Lees et al. Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays. *Phys.Rev.Lett.*, 109:101802, 2012.

[57] Anders Ryd, David Lange, Natalia Kuznetsova, Sophie Versille, Marcello Rotondo, et al. EvtGen: A Monte Carlo Generator for B-Physics. 2005.

[58] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *JHEP*, 2006(05):026, 2006.

[59] J. Allison et al. Geant4 developments and applications. *IEEE T. Nucl. Sci.*, 53(1):270 –278, feb. 2006.

[60] The Belle CKM Group. Hadronic tag calibration with charm semileptonic decays `http://belle.kek.jp/secured/wiki/doku.php?id=physics:ckm:htag`, November 2012.

[61] Nathan Isgur, Daryl Scora, Benjamin Grinstein, and Mark B. Wise. Semileptonic B and D Decays in the Quark Model. *Phys.Rev.*, D39:799–818, 1989.

[62] Daryl Scora and Nathan Isgur. Semileptonic meson decays in the quark model: An update. *Phys.Rev.*, D52:2783–2812, 1995.

[63] Adam K. Leibovich, Zoltan Ligeti, Iain W. Stewart, and Mark B. Wise. Semileptonic B decays to excited charmed mesons. *Phys.Rev.*, D57:308–330, 1998.

[64] Adam K. Leibovich, Zoltan Ligeti, Iain W. Stewart, and Mark B. Wise. Semileptonic B decays to excited charmed mesons. *Phys.Rev.*, D57:308–330, 1998.

[65] Adam K. Leibovich, Zoltan Ligeti, Iain W. Stewart, and Mark B. Wise. Model independent results for $B \to D_1(2420)$ lepton anti-neutrino and $B \to D_2^*(2460)$ lepton anti-neutrino at order Lambda(QCD) / m(c,b). *Phys.Rev.Lett.*, 78:3995–3998, 1997.

[66] The Belle CKM Group. Xclnu model uncertainties `http://belle.kek.jp/secured/wiki/doku.php?id=physics:ckm:xcmodel`, November 2012.

[67] L. Hinz. Lepton id efficiency correction and systematic error. Belle Note 954, 2006.

[68] The Belle PID Joint Group. Lepton id `http://belle.kek.jp/group/pid_joint/`, January 2013.

[69] P. Urquijo. Semi-inclusive semileptonic b decays. Belle Note 1251, 2012.

[70] P. Urquijo. Lepton fake rates. private communication, 2012.

[71] T.Iijima K.Hara, Y.Horii. Measurement of $b \to \tau\nu$ with hadronic tagging method. Belle Note 1245, 2012.

[72] phi-t GmbH. The NeuroBayes User's Guide, April 2010.

[73] Michael. Feindt and Michael Prim. An algorithm for quantifying dependence in multivariate data sets. *arXiv.org:1207.0981*, 2012.

[74] Joseph Lee Rodgers and W. Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):pp. 59–66, 1988.

[75] C. Spearman. *The proof and measurement of association between two things*. The American Journal of Psychology, Vol. 15, No. 1 (Jan., 1904), pp. 72-101. University of Illinois Press, Weinheim, 1904.

[76] Kyle S. Cranmer. Kernel estimation in high-energy physics. *Comput.Phys.Commun.*, 136:198–207, 2001.

[77] Heavy Flavor Averaging Group. Updates of Semileptonic Results for End of 2011, 2011.

[78] W. Dungel et al. Measurement of the form factors of the decay $B^0 \to D^{*-}\ell^+\nu$ and determination of the CKM matrix element —Vcb—. *Phys.Rev.*, D82:112007, 2010.

[79] Bernard Aubert et al. Measurements of the Semileptonic Decays anti-B → D l anti-nu and anti-B → D* l anti-nu Using a Global Fit to D X l anti-nu Final States. *Phys.Rev.*, D79:012002, 2009.

[80] K. Abe et al. Measurement of B(anti-B0 → D+ l- anti-nu) and determination of —V(cb)—. *Phys.Lett.*, B526:258–268, 2002.

[81] Bernard Aubert et al. A Measurement of the branching fractions of exclusive $\bar{B} \to D^{(*)}\,(\pi)\,\ell^-\bar{\nu}(\,\ell^)$ decays in events with a fully reconstructed B meson. *Phys.Rev.Lett.*, 100:151802, 2008.

[82] D. Liventsev et al. Study of $B \to D^{**}\ell nu$ with full reconstruction tagging. *Phys.Rev.*, D77:091503, 2008.

[83] Bernard Aubert et al. Measurement of the Branching Fractions of anti-B → D** l- anti-nu(l) Decays in Events Tagged by a Fully Reconstructed B Meson. *Phys.Rev.Lett.*, 101:261802, 2008.

[84] Bernard Aubert et al. Measurement of Semileptonic B Decays into Orbitally-Excited Charmed Mesons. *Phys.Rev.Lett.*, 103:051803, 2009.

[85] J.L. Goity and W. Roberts. Soft pion emission in semileptonic B meson decays. *Phys.Rev.*, D51:3459–3477, 1995.

[86] Alejandro Celis, Martin Jung, Xin-Qiang Li, and Antonio Pich. Sensitivity to charged scalars in $B \to D^{(*)}\tau\nu_\tau$ and $B \to \tau\nu_\tau$ decays. 2012.

[87] Andreas Crivellin, Christoph Greub, and Ahmet Kokulu. Explaining $B \to D\tau\nu$, $B \to D^*\tau\nu$ and $B \to \tau\nu$ in a 2HDM of type III. *Phys.Rev.*, D86:054014, 2012.

[88] Zdenek Dolezal. Super KEKB and Belle II: Status of the KEK Super B Factory. *PoS*, EPS-HEP2009:143, 2009.

[89] T. Browder, M. Ciuchini, T. Gershon, M. Hazumi, T. Hurth, et al. On the Physics Case of a Super Flavour Factory. *JHEP*, 0802:110, 2008.

[90] Thomas E. Browder, Tim Gershon, Dan Pirjol, Amarjit Soni, and Jure Zupan. New Physics at a Super Flavor Factory. *Rev.Mod.Phys.*, 81:1887–1941, 2009.

Danksagung

Bei Herrn Prof. Dr. Michael Feindt möchte ich mich sehr herzlich dafür bedanken, dass ich als Doktorand am EKP an zwei herausfordernden, interessanten Projekten der Belle Kollaboration mitarbeiten durfte.

Herrn Prof. Dr. Thomas Müller danke ich sehr herzlich für die bereitwillige Übernahme des Korreferats.

Ganz besonderer Dank für die tägliche unermüdliche Unterstützung bei meiner Analyse gebührt Dr. Anže Zupanc und Dr. Thomas Kuhr.

Bei meinen Kollegen Dr. Sebastian Neubauer und Fabian Keller bedanke ich mich für jahrelange spannende Zusammenarbeit an der Vollständigen Rekonstruktion.

Für die stets freundschaftliche Atmosphäre bedanke ich mich bei meinen Bürokollegen Dr. Dominik Horn, Manuel Heider und Johannes Grygier. Allen Bachelor- und Masterstudenten, den Diplomanden, Doktoranden, Postdocs und Professoren des EKP danke ich für die sehr angenehme Arbeitsumgebung, und die stets große Hilfsbereitschaft.

Für das Lektorat dieser Arbeit gebührt Manuel Heider, Bastian Kronenbitter, Michael Prim, Simon Wehle, Matthias Huschle, Dr. Aňze Zupanc und Dr. Thomas Kuhr mein Dank.

Weiterhin danke ich dem Administratorenteam des EKP dafür, dass die Computer des EKP stets bestens liefen und ich während meiner Arbeit niemals mit großen technischen Problemen zu kämpfen hatte.

Der Institutssektretärin Frau Bärbel Bäunling gilt mein besonderer Dank für jegliche Organisation und unermüdliche Hilfe bei der Bewältigung der Bürokratie.

Zuletzt möchte ich meiner Familie danken, besonders meinen Eltern Anita und Herbert Zander, meinem Bruder Dr. Frank Zander und meiner Frau Sabina Zander.