

Dalitz analysis of
 $B^- \rightarrow D^+ \pi^- \pi^-$

Simon Andreas Alexander Kohl

A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE

INSTITUTE FOR EXPERIMENTAL NUCLEAR PHYSICS,
DEPARTMENT OF PHYSICS,
KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT)

Advisor: Prof. Dr. M. Feindt
Co-Advisor: Prof. Dr. G. Quast

JUNE 17, 2016

IEKP-KA/2016-9

Contents

1. Introduction	1
1.1. Analysis of $B^- \rightarrow D^+ \pi^- \pi^-$	1
1.2. Excited D -mesons	2
1.3. Current World Averages and Expected Signal Yield	4
2. The Belle Experiment	5
2.1. KEKB Accelerator	5
2.2. The Belle Detector	6
3. Dalitz Plot Analysis Technique	9
3.1. Dalitz Plot Analysis Technique	9
3.1.1. Kinematic Boundaries	10
3.1.2. Amplitude Description	11
3.1.3. Isobar Formalism	11
3.1.4. Branching Ratios from Dalitz Plot Fits	15
3.2. Choice of Dalitz Plot Parametrization	15
3.2.1. Square Dalitz Plot Representation	16
4. Statistical Analysis Techniques	19
4.1. Maximum Likelihood Estimation	19
4.1.1. Introduction	19
4.1.2. Calculation of Errors	20
4.2. Kernel Density Estimation	20
4.3. Statistical Hypothesis Testing	21
4.4. Multivariate Classifiers	23
4.4.1. Artificial Neural Networks	23
4.4.2. Boosted Decision Trees	24
5. Belle II Analysis Software Framework	27
6. Blind Analysis using Monte Carlo Data Samples	29
6.1. Belle Generic Monte Carlo	30
6.2. Continuum Monte Carlo	31
6.3. Evtgen Generic Dalitz Model for Signal Monte Carlo	32
6.4. Detector Simulation	33
7. Reconstruction and Selection	35
7.1. Reconstruction Channel	35
7.2. Final-State Particles	35
7.3. D -Meson Reconstruction	36
7.4. Vertex Reconstruction	36
7.5. B -Meson Reconstruction	37

7.6. Continuum Suppression	37
7.6.1. Multivariate Classifiers for Continuum Suppression	41
7.7. B -Meson Selection	42
7.8. Best Candidate Selection	43
7.9. Enforcing the Dalitz Plot boundaries	46
8. Dalitz Fit Model	49
8.1. Fit Model	49
8.2. Fit Framework	50
9. Background Modelling	51
9.1. Continuum Background	52
9.2. Combinatoric Background	52
9.3. Misreconstructed Signal	54
10. Detector Response	57
10.1. Signal Resolution and Migration	57
10.2. Efficiency	58
11. Fit Validation using individual Resonances	61
12. Fit Validation by Ensemble Tests	67
12.1. Ensemble Test Results	68
13. Outlook	73
14. Conclusion	77
Appendix A. basf2 Steering File	83
Appendix B. Continuum Suppression Variables	87
Appendix C. Fit to coherent Evtgen Model	91
Appendix D. Results of Ensemble Tests	93

*Stands at the sea,
wonders at wondering: I
a universe of atoms
an atom in the universe.*

— Richard Feynman,
The Value of Science

1. Introduction

1.1. Analysis of $B^- \rightarrow D^+ \pi^- \pi^-$

Although the standard model of particle physics was given further credence by the discovery of the Higgs boson in 2012, there is a range of major phenomena, which it fails to explain, e.g. the large excess of gravitating mass over visible mass in our universe, calling for the existence of so-called dark matter. For this reason many particle physics analyses of the past years were focused on particle production and decay channels that promise sensitivity to physics beyond the standard model (SM). An essential part of such analyses is the understanding and knowledge of standard model processes to very high precision, e.g. in order to detect potentially small yet significant deviations from qualifying SM processes. Instead of exploring new territory, this is precisely where the present analyses aims at making a contribution.

Analyzing the decay $B^- \rightarrow D^+ \pi^- \pi^-$ ¹, which accounts for around 1‰ of all B decays, offers to measure the inclusive branching ratio of one of the most frequent charmed hadronic decays in the sector of B physics. Beyond that, the three-body decay shows a rich resonant structure, allowing to study two different orbitally excited charmed mesons, the $D_0^*(2400)^0$ and the $D_2^*(2460)^0$, via a dedicated experimental technique: The Dalitz plot. The technique permits to simultaneously measure the exclusive branching fractions of $B^- \rightarrow D_J^0 \pi^-$ as well as the quantum numbers, masses, widths and relative phases of the resonances. This is interesting for a number of reasons: A precise understanding of excited D -mesons can reduce uncertainties in semileptonic decays and thus the Cabbibo-Kobayashi-Maskawa matrix element $|V_{cb}|$ [6]. Semileptonic decays such as $B^+ \rightarrow \tau^+ \nu_\tau$ [26] or $B^+ \rightarrow D^{(*)} \tau^+ \nu_\tau$ [22], which are sensitive to new physics, also directly profit from an improved estimation of D_J background. Furthermore, in providing experimental specifications to the spectroscopy of the D , the analysis of $B^- \rightarrow D^+ \pi^- \pi^-$ can aid in validating heavy quark effective theories [15].

The decay $B^- \rightarrow D^+ \pi^- \pi^-$ has first been analyzed by the Belle collaboration in 2004 [3] and later, in 2009 [6] and also in 2010 [24] by the BaBar collaboration. At the time the first Belle analysis was conducted, less than a 10th of the total data collected at the $\Upsilon(4S)$ resonance by the Belle experiment was at the experimenter's disposal. The full Belle dataset is therefore expected to have in store a significant improvement in statistical precision over the previous analyses. Thus, given the increase in available data today and the persisting unsatisfactory

¹Throughout this work charge-conjugate decays are included unless explicitly stated otherwise.

knowledge of the $D_0^*(2400)^0$ properties (including relatively large uncertainties on the world averages of its mass and width), the endeavor of analyzing the proposed decay channel by the Belle collaboration should be duly legitimized.

The present work lays the ground for the Dalitz plot analysis of $B^- \rightarrow D^+\pi^-\pi^-$ that is intended to use the full Belle dataset of around 772 million $B\bar{B}$ pairs by performing a blind analysis based on simulated data. Moreover, by utilizing the software framework developed to be used in the successor experiment Belle II and by extending it to fit the required use-cases, a road for future Dalitz plot analyses is paved. The phenomenology of excited D -mesons of vital interest in the analysis of $B^- \rightarrow D^+\pi^-\pi^-$ is described in Sec. 1.2. After an introduction of the Belle experiment and detector in Chapter 2, Chapters 3 and 4 elaborate upon the employed experimental techniques such as the Dalitz plot formalism. While Chapters 5 and 6 lay out the settings of how the blind-analysis using simulated data is performed, the subsequent Chapters 7 to 10 are concerned with the reconstruction and selection of B candidates, the formulation of the fit model as well as the treatment of backgrounds and detector effects. Finally Chapters 11 and 12 are dedicated to the validation of the Dalitz plot fit by validating the modelling of detector effects and by performing toy studies aimed at assessing the fit's unbiasedness. The last chapter (Chap. 13) presents the course of action to be undertaken in a prospective fit to recorded data.

1.2. Excited D -mesons

Besides a non-resonant contribution, the decay of $B^- \rightarrow D^+\pi^-\pi^-$ has been observed to proceed via intermediate D -meson excited states. Those states constitute P-wave excitations (i.e. with an intrinsic orbital momentum of $L = 1$) of quark-antiquark systems containing one charmed and one light quark (u). The larger the mass difference between the quarks, the more accurately the system can be described in analogy to the hydrogen model, which is why the light quark spin s_q is predicted to couple with the orbital angular momentum L to form a good quantum number, the total angular momentum j [20]. The P-wave D -states thus constitute two degenerate $j = 3/2$ states, the 2^+ and the 1^+ states, and two degenerate $j = 1/2$ states, the 0^+ and the other 1^+ (expressed in terms of the spin-parity J^P). These excited states are referred to as D_2^* , D_1 , D_0^* and D_1' , respectively. Their mass spectrum, as known today, is illustrated in Fig. 1.1, which also specifies the allowed $D_J \rightarrow D\pi$ (and $D_J \rightarrow D^*\pi$) transitions. The allowed decays are constrained by the conservation of parity (P) and angular momentum (J) in strong decays: The $j = 1/2$ states are expected to decay exclusively through S-waves (meaning there is zero angular momentum between the decay daughters) and hence constitute broad resonant states, while the $j = 3/2$ states are predicted to decay through D-waves (an angular momentum of $L = 2$ between the decay daughters), which are accordingly narrow in nature.

Because this analysis exclusively reconstructs charged $J = 0$ D -mesons (in combination with two pions) as final states, only contributions from the $D_0^*(2400)^0$ and $D_2^*(2460)^0$ are expected. Two tree-level Feynman diagrams contribute to each resonant decay as depicted in Fig. 1.2. The properties of the broad $D_0^*(2400)^0$ are hard to measure, which is why compared to the $D_2^*(2460)^0$, its mass and width are known only to a poor precision, as can be seen from Tab. 1.1, which lists the current world averages. The Dalitz plot technique, which is introduced in detail in Chap. 3, offers to study the dynamics of $B^- \rightarrow (D_J^0 \rightarrow D^+\pi^-)\pi^-$ and thus the J^P values of the contributing states. To anticipate a strength of the technique, note that the distinction of resonant states is readily conducted by characteristic structures that the different angular momenta of the D excitations bring about in the plot.

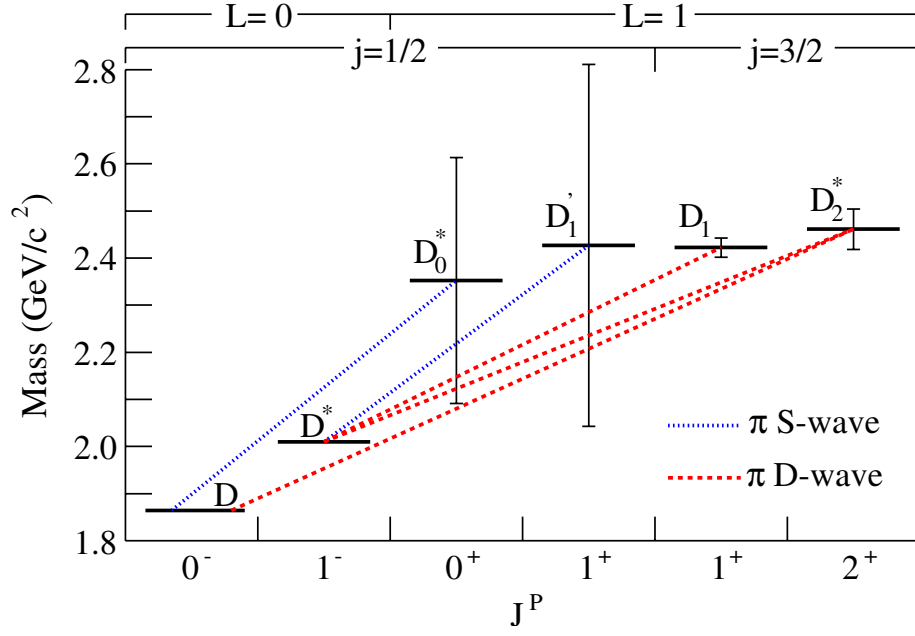


Figure 1.1.: Mass spectrum for $c\bar{u}$ states. The vertical bars show the corresponding widths and the dashed lines indicate the dominant pion transitions [6].

<i>resonance</i>	J^P	<i>mass</i> (MeV/ c^2)	<i>width</i> (MeV/ c^2)	<i>partial wave</i>
$D_0^*(2400)^0$	0^+	2318 ± 29	267 ± 40	S
$D_2^*(2460)^0$	2^+	2462.6 ± 0.6	49 ± 1.3	D

Table 1.1.: Properties of the expected P-wave resonances [31].

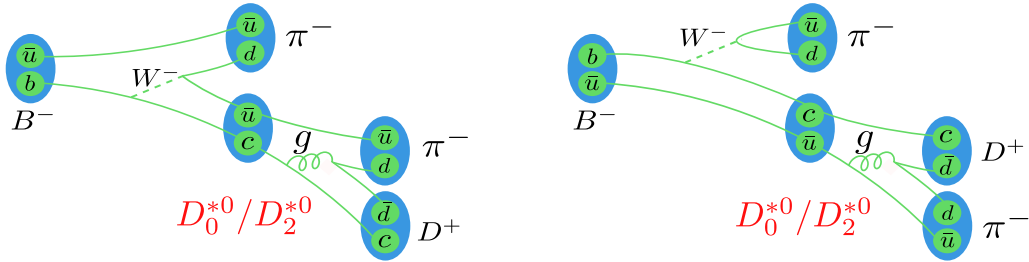


Figure 1.2.: The contributing tree-level Feynman diagrams of the strong decay $B^- \rightarrow (D_0^{*0}/D_2^{*0} \rightarrow D^+ \pi^-) \pi^-$, where the resonances are labeled in red.

1.3. Current World Averages and Expected Signal Yield

The current world averages of the branching ratios (\mathcal{B} s) of the inclusive decay of $B^- \rightarrow D^+ \pi^- \pi^-$ as well as the exclusive branching fractions of two expected resonant contributions are given in Tab. 1.2 below. On the basis of the total number of recorded $B\bar{B}$ events and the branching fraction of the channel that the D is reconstructed in ($D^+ \rightarrow K^+ \pi^- \pi^-$, see Sec. 7.1) as well as the average reconstruction efficiency $\langle\epsilon\rangle$ (see Chap. 10), the expected number of signal events can be computed via

$$N_{signal} = \langle\epsilon\rangle \cdot N_{B\bar{B}} \cdot \mathcal{B}_{B \rightarrow D\pi\pi} \cdot \mathcal{B}_{D \rightarrow K\pi\pi},$$

which, using $\langle\epsilon\rangle = 0.0699$, $N_{B\bar{B}} = 771581000$, $\mathcal{B}_{B \rightarrow D\pi\pi} = 0.00107$ and $\mathcal{B}_{D \rightarrow K\pi\pi} = 0.0913$, comes out to be $N_{signal} \approx 5269$ signal events. On recorded data deviations from this signal yield are expected since the average efficiency is dependent on the physical signal model. If the latter were known, $\langle\epsilon\rangle$ could be calculated via Eq. 3.13.

<i>channel</i>	$\mathcal{B} (\times 10^{-3})$	<i>error</i> ($\times 10^{-3}$)
inclusive	1.07	0.05
via $D_0^*(2400)^0$	0.64	0.14
via $D_2^*(2460)^0$	0.35	0.04

Table 1.2.: The measured inclusive and exclusive B^- branching fractions of decays via different resonances in $B^- \rightarrow D^+ \pi^- \pi^-$ [31].

2. The Belle Experiment

The following chapter introduces the facilities and their technical set-up with which the particle collision experiments, at the bottom of the present work, were conducted. After briefly elaborating upon the KEKB accelerator located in Tsukuba, Japan, the heart of the Belle experiment, the detector device, is featured layer by layer ¹.

2.1. KEKB Accelerator

KEKB is a circular e^+e^- collider, whose design goals were specifically set to meet the demands of decay-time-dependent measurements in B physics (in order to test CP violation in B mesons): An instantaneous luminosity in excess of $10^{33} \text{ cm}^{-2} \text{ s}^{-2}$ as well as a boost of the center of mass frame relative to the laboratory frame sufficient to observe the time evolution of B decays. In order to achieve the latter, the two beams are required to exhibit a pronounced asymmetry in energy necessitating the usage of two different beam pipes. These are referred to as HER (high energy ring) and LER (low energy ring) and are, amongst other KEKB design details, depicted in Fig. 2.1.

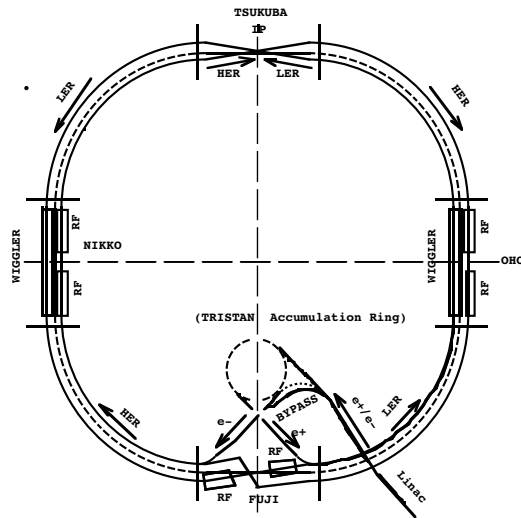


Figure 2.1.: Schematic layout of KEKB [27].

¹Large parts of this chapter's content follow Chap. 1 in [7] and [14]

At the intersection of the rings (the interaction point, IP in Fig. 2.1), the beams collided with a center of mass energy exactly equaling the $\Upsilon(4S)$ ' (an excited $b\bar{b}$ state) rest mass [25]:

$$E_{CMS} = 2\sqrt{E_{HER}E_{LER}} = 10.58 \text{ GeV} = m(\Upsilon(4S)), \quad (2.1)$$

where $E_{HER} = 8 \text{ GeV}$ and $E_{LER} = 3.5 \text{ GeV}$, respectively. Because the $\Upsilon(4S)$ decays to nothing but entangled $B\bar{B}$ pairs close to 100% of the times, the KEKB was an abundant source of B mesons in a clean, low-background environment, accordingly referred to as a ' B -Factory'. KEKB ran from 1998 to 2010, a time span during which it collected the aforementioned dataset of around 772 Mio. $B\bar{B}$ events at the $\Upsilon(4S)$ resonance. Since its shutdown in 2010, KEKB is being upgraded to SuperKEKB with an intended 40-fold increase in instantaneous luminosity accompanied by an improved design of the detector for the successor experiment Belle II.

2.2. The Belle Detector

The interaction point of KEKB was surrounded by the Belle detector, a multi-purpose magnetic spectrometer covering a solid angle of 4π . The schematic longitudinal cross section of the Belle detector is shown in Fig. 2.2. The full detector was composed of the forward and backward endcaps as well as the centrally located barrel, which houses several, to an extent layered, subdetectors as well as a superconducting solenoid with a magnetic field of 1.5 T.

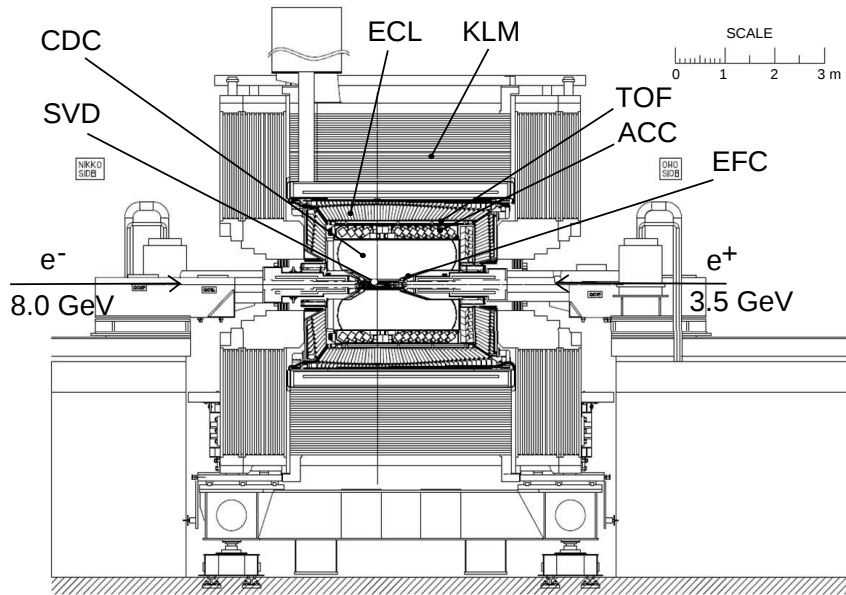


Figure 2.2.: Schematic view of the longitudinal cross section of the Belle detector [7].

Outside of the innermost component, the cylindrical beryllium beam pipe, a *Silicon Vertex Detector* (SVD) was located and used to detect B decay vertices as well as charged tracks. Like the SVD, the connected *Central Drift Chamber* (CDC) supplied hits originating from charged tracks. Additionally, the CDC provided dE/dx measurements, which combined with the signals from an array of *Aerogel Čerenkov Counters* (ACC) and from a set of *Time-of-Flight Scintillation Counters* (TOF), allowed for particle identification of charged tracks. Electromagnetic showers, and thus their energies, were detected in the *Electromagnetic Calorimeter* (ECL) composed of an array of CsI(Tl) crystals. An *Extreme Forward Calorimeter* (EFC) made from

BGO crystals was situated close to the IP and served to increase the angular coverage of the detector. The outermost detector component, the KLM, was an assembly of resistive plate counters aimed at detecting K_L^0 s and μ^\pm s and was placed outside of the iron yoke, which was used to return the magnetic flux. The following descriptions are committed to providing more details on the respective subdetector parts as we transverse the detector outwards starting from the beam-pipe.

SVD With three layers of AC coupled double-sided silicon-strip detectors, the Silicon Vertex Detector first installed, called the SVD1, mirrored a conservative design choice. Its polar angle coverage of $23^\circ < \theta < 140^\circ$ did not match the detector's full coverage of $17^\circ < \theta < 150^\circ$ and the limited radiation hardness of the chips disallowed more aggressive operation of KEKB, which is why after a period of three years, SVD1 was replaced by SVD2, exhibiting 4 layers of double-sided silicon-strip detectors and a full angular acceptance. The silicon strips were aligned perpendicular and parallel to the beam pipe which in combination allows for a measurement of charged tracks in the $r\phi$ and z direction.

CDC The Central Drift Chamber reconstructs charged particle tracks, measures their hit coordinates in the detector volume and enables reconstruction of their momenta. Moreover, the energy loss dE/dx in its gas volume, provides information for particle identification. In order to provide optimal momentum resolution while retaining a good energy loss resolution a compromise in the nucleon number of the gases had to be found which was reflected in the final choice of the gas mixture of 50% He and 50% C_2H_6 . To increase forward acceptance (in the direction of the e^- beam) the CDC is positioned asymmetric in the z direction. Its inner radius lied at 80, its outer radius at 880 mm with respect to the beam-pipe. The chamber had 50 cylindrical layers, each containing between three and six either axial or small-angle stereo layers as well as three cathode strip layers. In total there were 8400 drift cells with a maximum drift distance between 8 and 10 mm.

TOF The Time-of-Flight system consisted of a barrel of 128 plastic scintillator counters and could distinguish between kaons and pions for tracks with momenta below 1.2 GeV. The time resolution of the system was 100 ps, which enables to measure the time the charged particles need to transverse the distance of ~ 1.2 m from the interaction point [21]. Each of the 64 TOF modules consisted of two TOF counters and was augmented by one thin trigger scintillation counter. In total a coverage of $33^\circ < \theta < 121^\circ$ was reached. Using a momentum measurement from the SVD and CDC and the helical distance traveled (L), the mass m of a particle can be deduced via a TOF measurement according to [32] by

$$m = \frac{p}{c} \sqrt{\left(\frac{cT}{L}\right)^2 - 1}. \quad (2.2)$$

ACC The Aerogel Čerenkov Counter provided separation between kaons and pions in a momentum range of $1.2 \text{ GeV} < p < 1.2 \text{ GeV}$ and was thus complimentary to the aforementioned TOF. Charged particles that pass a medium with a velocity larger than the speed of light in the medium emit a cone of Čerenkov light, whose opening angle is proportional to the particle velocity. For a given material's refractive index n , this occurs if a particle with mass m satisfies

$$n > \frac{1}{\beta} = \sqrt{1 + \left(\frac{m}{p}\right)^2}. \quad (2.3)$$

The material of the ACC was therefore chosen such that pions with $p > 1.2 \text{ GeV}$ emitted Čerenkov light, whereas heavier kaons did not, making the ACC a threshold counter [32]. The ACC modules covered a polar angle of $17^\circ < \theta < 127^\circ$, with refractive indices varying between 1.01 and 1.03.

- ECL** The Electromagnetic Calorimeter measured the energy deposits and positions of electrons and photons, both of which undergo a cascade of Bremsstrahlung and pair production in alternation. The ECL consisted of a barrel section and two end caps of segmented arrays of CsI(Tl) crystals, which, aside from a small tilt, pointed towards the interaction point. The scintillation light produced as the electromagnetic cascade dies down inside the crystals was detected with silicon photodiodes. The measurements in the ECL provided the energy to momentum ratio, the main parameter used to distinguish hadrons from electrons and photons and thus contributed to the particle identification. The polar angle coverage amounted to $17^\circ < \theta < 150^\circ$, which was extended by the EFC, described below.
- EFC** By covering a polar angle of $6.4^\circ < \theta < 11.5^\circ$ in the forward direction and an angle of $163.3^\circ < \theta < 171.2^\circ$ in the backward direction, the Extreme Forward Calorimeter's main purpose was to improve the experimental sensitivity for the electron and photon detection [32]. The material choice accommodated for the increased levels of radiation in the proximity of the interaction region in that a more suitable material, bismuth germanium oxide (BGO), was used for detection of scintillation light.
- Magnet** Magnetic fields penetrating the other subdetectors were used to force charged particles on circular paths allowing to measure their momenta and to determine their charge. The superconducting solenoid magnet provided a magnetic field of 1.5 T (parallel to the beam-pipe) in a cylindrical volume of 3.4 m in diameter and 4.4 m in length. The coil included all other subdetectors introduced above and was in turn surrounded by an iron structure that served as the return path of magnetic flux and as absorber material for the KLM described below. The superconducting coil was made of a niobium-titanium-copper (NbTi/Cu) alloy and was cooled by a liquid helium cryostat to achieve superconducting temperatures. The magnet had a nominal current of 4400 A and stored an energy of 35 MJ [32].
- KLM** The KLM detection system was designed to identify K_L s and muons and owing to their persistence it is the outermost part of the detector. The KLM consisted of alternating layers of charged particle detectors (glass-resistive plate counters, 15 layers in the barrel region, 14 in the end-caps) and 4.7 cm thick iron plates (14 layers in both the barrel and end-caps). In addition to the 0.8 interaction lengths the ECL provided, the iron plates delivered a total of 3.9 interaction lengths of material for a particle traveling normal to the detector planes. K_L s that interact in the iron or ECL produced a shower of ionizing particles, which indicated the direction of the incoming K_L s. As compared to charged hadrons, muons travel further and with smaller deflections inside the KLM, which allows to discriminate between the two.

3. Dalitz Plot Analysis Technique

3.1. Dalitz Plot Analysis Technique

The Dalitz Plot analysis technique is an approach that enables the study of three-body resonant decays and was first devised by Prof. Richard H. Dalitz whilst investigating the nature of the so-called Tau-Theta Puzzle in 1953. The technique exploits the fact that the dynamics of three-body decays can be parametrized by a suitable choice of only two intrinsic degrees of freedom, allowing for both a convenient visualization of the decay topology as well as for a quantitative extraction of the decay characteristics. The three final-states of a three-body decay are characterized by their four-momenta and thus in terms of 12 parameters as hinted upon in Fig. 3.1. Fortunately a set of constraints exists and comes to the rescue: four energy-momentum constraints, three constraints on the final states' masses and rotational invariance (alleviating the three degrees of freedom of the Euler angles) help bring down the overall count of degrees of freedom to only two independent observable quantities.

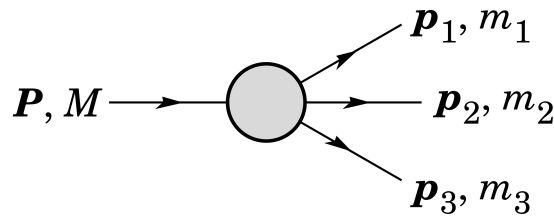


Figure 3.1.: Definitions of variables for three-body decays [31].

The Dalitz Plot itself is a scatter plot of those two remaining degrees of freedom whose density distribution in this plot is a manifestation of the decay dynamics. The two free parameters can be chosen as either the kinetic energies of two of the final-state particles (which was R. H. Dalitz' approach) or as the squares of the invariant masses of two pairs of final-state particles. The latter approach is generally more suitable for relativistic decays [8] and therefore also adopted in this work.

Weak nonleptonic decays of D and B mesons are expected to proceed primarily through resonant two-body decays [31]. Those resonances in turn undergo a two-body decay, thus in succession constituting a three-body decay process with the production of three final-state particles. For three-body decays of a spin-0 particle to all pseudo-scalar final-states, such as

D or $B \rightarrow abc$, the differential decay rate [31] is given as

$$d\Gamma = \frac{1}{(2\pi)^3 32M^3} |\mathcal{A}|^2 dm_{ab}^2 dm_{bc}^2, \quad (3.1)$$

where \mathcal{A} represents the total decay amplitude, M the mass of the mother particle and m_{ab} and m_{bc} are the invariant mass combinations of the daughters a , b and c , respectively. The prefactor contains all kinematic factors, while $|\mathcal{A}|^2$ parametrizes the dynamics and thus gives the probability density according to which the invariant mass combinations are distributed. All structures and variations in the Dalitz plot (the scatter plot of m_{ab}^2 and m_{bc}^2) population are due to dynamic as opposed to kinematic effects.

Further descriptions of the details of the Dalitz plot analysis technique follow closely the respective chapter in the book *The Physics of the B Factories* [8], which was jointly written by the BaBar and Belle collaborations.

3.1.1. Kinematic Boundaries

The squared invariant masses of pairs ij of final-state particles are linearly related through a global constraint arising from Lorentz invariance:

$$m_{ab}^2 + m_{bc}^2 + m_{ac}^2 = M^2 + m_a^2 + m_b^2 + m_c^2. \quad (3.2)$$

The right hand side of Eq. 3.2 constitutes the sum of the rest masses of the involved particles and is therefore constant. As suggested above, the entire dynamics is a unique function of just two of the three possible invariant mass combinations, yielding three different possible definitions of the Dalitz plot, all containing the same information. Fig. 3.2 shows a generic example of a Dalitz plot of one such choice. The almond shaped region therein corresponds to the kinematically allowed phasespace of the daughters as defined through Eq. 3.2.

Each of the squared invariant mass combinations can be expressed in terms of either one of the other squared invariant masses, as exemplified below for the extremal values of m_{bc}^2 (for a given m_{ab}^2):

$$\begin{aligned} (m_{bc}^2)_{max} &= (E_b^* + E_c^*)^2 - (p_b^* - p_c^*)^2, \\ (m_{bc}^2)_{min} &= (E_b^* + E_c^*)^2 - (p_b^* + p_c^*)^2, \end{aligned} \quad (3.3)$$

where

$$E_b^* = \frac{m_{ab}^2 - m_a^2 + m_b^2}{2m_{ab}}, \quad E_c^* = \frac{M^2 - m_{ab}^2 - m_c^2}{2m_{ab}} \quad (3.4)$$

are the energies of particles b and c in the ab rest frame and

$$p_b^* = \sqrt{E_b^{*2} - m_b^2}, \quad p_c^* = \sqrt{E_c^{*2} - m_c^2} \quad (3.5)$$

are the corresponding four-momenta magnitudes. As can be gathered from Eqs. 3.3 the position of a three-body decay in the Dalitz plot reflects the kinematics of the final-states. For a given m_{ab}^2 , the range of m_{bc}^2 is determined by its values when p_b^* is parallel or antiparallel to p_c^* [31]. The 'global' maximum value of a given squared invariant mass pair occurs when the third daughter particle is produced at rest (as viewed in the frame of the decaying particle), while the two other daughters display anti-parallel three-momenta, thus decaying 'back to back'. In contrast, a minimum squared invariant mass pair value is adopted when the third daughter carries its maximally allowed three-momentum while the other two daughters both show anti-parallel three-momenta with respect to the former. Different scenarios of this type with the three-momenta configurations depicted in light grey are demonstrated in Fig. 3.2.

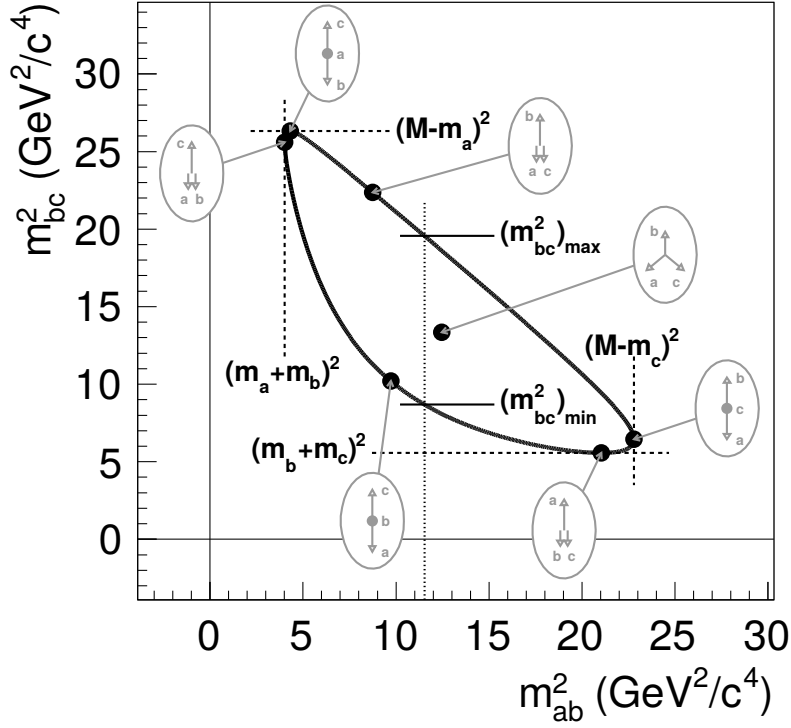


Figure 3.2.: Kinematic boundaries of the three-body decay phase space and illustration of various kinematic configurations of the final-state particles for characteristic Dalitz plot points [8].

3.1.2. Amplitude Description

The decay amplitude $\mathcal{A}(m_{ab}^2, m_{bc}^2)$ of a three-body decay of spin-zero particles to pseudoscalar final-states is commonly modeled as the coherent sum of two-body amplitudes (subscript r) and a ‘nonresonant’ contribution (subscript NR):

$$\mathcal{A}(\mathbf{m}) = \sum_r a_r e^{i\phi_r} \mathcal{A}_r(\mathbf{m}) + a_{NR} e^{i\phi_{NR}} \mathcal{A}_{NR}(\mathbf{m}). \quad (3.6)$$

The parameters a_r (a_{NR}) and ϕ_r (ϕ_{NR}) are the amplitudes and phases of the component r (NR) and $\mathbf{m} \equiv (m_{ab}^2, m_{bc}^2)$ is the corresponding point in the Dalitz plane. The phases are not absolute in their magnitude, instead they are relative quantities, thus leaving the freedom of defining one phase to be zero which implies that the remaining finite phases are defined and measured relative to the former.

3.1.3. Isobar Formalism

The isobar formalism describes the amplitude \mathcal{A} for the process $R \rightarrow rc$, $r \rightarrow ab$ where R is a D or B meson, r is an intermediate resonance and a , b , c are pseudoscalar final-states. The model obtained its name because it was first used for intermediate resonances which are isobars of a particular nuclear state such as resonances occurring in antinucleon-nucleon interactions. The amplitude, as obtained through the isobar formalism, takes the form

$$\mathcal{A}_r = B_L^R \times B_L^r \times Z_r \times T_r. \quad (3.7)$$

$Z_r \times T_r$ corresponds to the resonance propagator, where Z_r describes the angular distribution of the final-states a and b , and T_r is the dynamical function of the resonance r , also referred to as its ‘line shape’. B_L^R and B_L^r in turn denote the barrier factors for the production of rc

and of ab and constitute weighting factors for the amplitudes that account for a resonance spin-dependent production suppression, as further detailed below. The subscripts L refer to the orbital angular momentum between r and c .

The simplest and also most common approach to describe the dynamical function T_r is using a relativistic Breit-Wigner (BW) parametrization with mass-dependent width. In cases where the threshold of another channel via which the considered resonance can decay is accessible or when the resonances have a sizeable width, the relativistic BW might not be the parametrization of choice. In this analysis, however, its applicability is assumed and thus T_r is taken as:

$$T_r(m_{ab}) \propto \frac{1}{m_r^2 - m_{ab}^2 - im_r \Gamma_{ab}(m_{ab})}. \quad (3.8)$$

A standard formulation for the mass-dependent width reads:

$$\Gamma_{ab}(m_{ab}) = \Gamma^r \left(\frac{q}{q_r} \right)^{2L+1} \left(\frac{m_r}{m_{ab}} \right) B_L(q_{ab}, q_r)^2, \quad (3.9)$$

where L , Γ^r and $B_L(q_{ab}, q_r)$ are the angular momentum of the decay products a and b , the partial width and the Blatt-Weisskopf barrier factor for the decay of resonance r . q_{ab} is the three-momentum magnitude of the daughter particles in the center-of-mass frame of a and b and q_r is the momentum the decay products a and b would have in the rest-frame of a resonance with mass m_r , where m_r corresponds to the pole mass.

In strong decays the maximum angular momentum L is limited by the relative momentum q_r of the decay particles in the resonance center of mass frame [17]. Decaying resonances with an impact parameter (meson radius) R of order 1 fm ($\sim 1.2 \text{ GeV}^{-1}$) have trouble raising sufficient angular momentum to conserve the spin of the resonance. The Blatt-Weisskopf barrier factors $B_L(q_{ab}, q_r)$ weight the amplitudes to factor in this spin-dependent effect. For the cases $L = 0, 1, 2$ they are given in Table 3.1, where the effective radius R commonly takes values between 1 and 5 GeV^{-1} and is chosen empirically. This analysis adopts $R = 1.6 \text{ GeV}^{-1}$ and in doing so follows [3], [6], [2].

L	$B_L(q_{ab}, q_r)$
0	1
1	$\sqrt{\frac{1 + R^2 q_r^2}{1 + R^2 q_{ab}^2}}$
2	$\sqrt{\frac{(R^2 q_r^2 - 3)^2 + 9R^2 q_r^2}{(R^2 q_{ab}^2 - 3)^2 + 9R^2 q_{ab}^2}}$

Table 3.1.: The Blatt-Weisskopf factors for different values of the resonance spin L .

The last factor remaining to be defined in Eq. 3.7 is the angular dependence Z_r . It is described using either the helicity formalism or the Zemach tensor formalism. In the latter, transversality is enforced and because all the final-states in the analysis at hand are spin 0 particles, it is the

description of choice in this analysis. The resonance spin-dependent distributions are given in Table 3.2. The angle θ is the angle between the momenta of particles a/b and c as viewed in the resonance' rest frame and it is referred to as the helicity angle.

$J \rightarrow L \oplus l$	angular distribution
$0 \rightarrow 0 \oplus 0$	uniform
$0 \rightarrow 1 \oplus 1$	$\propto \cos\theta$
$0 \rightarrow 2 \oplus 2$	$\propto \cos^2\theta - 1/3$

Table 3.2.: The angular distribution for the decay process $R \rightarrow rc, r \rightarrow ab$. Here θ is the angle between the particles a and c or b and c in the rest frame of the resonance r .

The coherent sum of three-body decay amplitudes as described by the isobar formalism serves as the model according to which the signal candidates are distributed in the Dalitz plot. This model explains all structures, i.e. high density regions, in the scatter plot. From the descriptions above it is apparent that the Dalitz plot technique allows to not only extract the masses and widths of resonant contributions, but also their spin. Sticking with the convention of having the resonance decay to the daughters a and b while c corresponds to the bachelor particle, it is clear that the dynamical function T_r takes effect in the Dalitz plot projections to m_{ab} while the distribution of the helicity angle, that represents a handle on the resonance' spin, is manifested in the projection to m_{bc} .

This is further illustrated in Fig. 3.3 where different intermediate two-body states appear in the shown examples of Dalitz plots. Fig. 3.3(a) shows the flat distribution of a phase-space three-body decay, i.e. without intermediate resonance, which is clearly distinguishable from Figs. 3.3(b-d) featuring bands of uniform density corresponding to scalar (spin 0) resonances. The angular distributions of vector (spin 1) and tensor (spin 2) resonances demonstrate characteristic non-uniformity of the event density along the resonance bands (Figs. 3.3 (e,f)). Moreover, due to the coherent addition of amplitudes the region of overlapping resonances is sensitive to phase difference as is illustrated in Figs. 3.3(g,h).

The isobar formalism neglects final-state interactions and is known to lead to unitarity violation whenever the pole masses of the resonances are close to one another measured in terms of their respective widths [12]. Alternatives such as the so-called K-matrix formalism [11] exist but their complexity outweighs their merits for use in the analysis at hand.

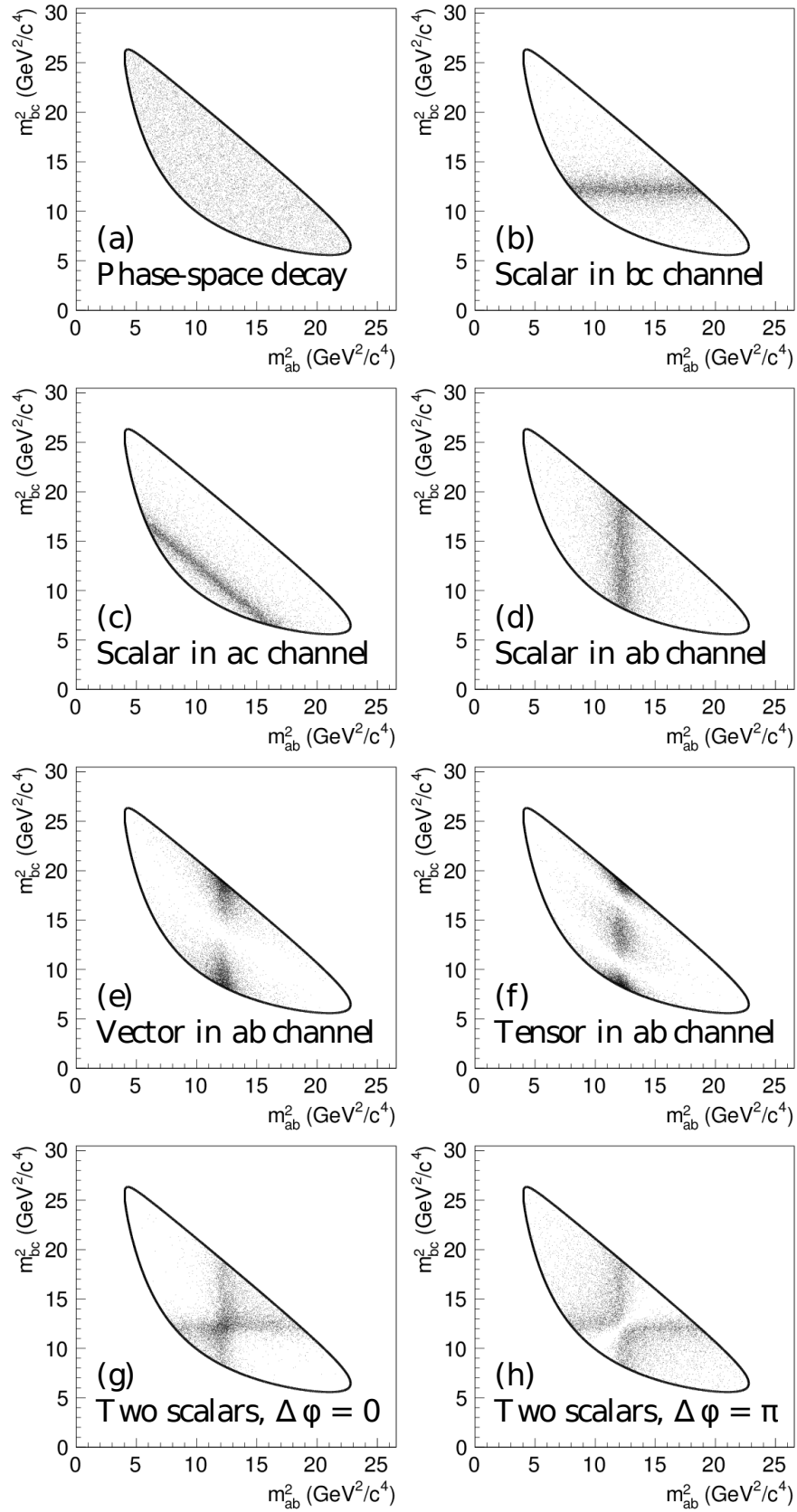


Figure 3.3.: Example Dalitz plots with (a) phase-space decay, (b-d) one scalar resonance appearing in various decay channels, (e, f) vector and tensor resonances, and (g, h) the interference of two scalar resonances with different values of the relative phase $\Delta\phi$ [8].

3.1.4. Branching Ratios from Dalitz Plot Fits

For the sake of comparability of fit results, it is desirable to formulate observables that are largely independent of the choice of normalization, phase convention, and amplitude formalism. 'Fit fractions' are commonly used to provide a means to estimate the branching fractions in such a convention independent manner. The fit fraction of the j^{th} decay channel FF_j is defined as the integral over the Dalitz plane (DP , here m_{ab} vs. m_{bc}) of the single amplitude j squared and normalized by the corresponding integral of the square of the coherent sum of all amplitudes, or

$$FF_j = \frac{\int_{DP} |a_j e^{i\phi_j} \mathcal{A}_j|^2 dm_{ab}^2 dm_{bc}^2}{\int_{DP} \left| \sum_j a_j e^{i\phi_j} \mathcal{A}_j \right|^2 dm_{ab}^2 dm_{bc}^2}. \quad (3.10)$$

It is important to note that the fit fractions do not have to add up to unity. This is due to interference effects, destructive or constructive in nature. To quantify such interference effects more explicitly, interference fit fractions can be defined as

$$FF_{jk} = \frac{\int_{DP} |a_j a_k e^{i(\phi_j - \phi_k)} \mathcal{A}_j \mathcal{A}_k^*|^2 dm_{ab}^2 dm_{bc}^2}{\int_{DP} \left| \sum_j a_j e^{i\phi_j} \mathcal{A}_j \right|^2 dm_{ab}^2 dm_{bc}^2}. \quad (3.11)$$

The exclusive branching fractions are then obtained as:

$$\mathcal{B}_j = \frac{FF_j \cdot N_{sig}}{\bar{\epsilon} \cdot \epsilon_{PID} \cdot N_{B\bar{B}} \cdot \mathcal{B}_{x \rightarrow final}}, \quad (3.12)$$

where N_{sig} is the signal yield and ϵ_{PID} is the K^\pm/π^\pm selection efficiency as obtained from a channel specific particle identification (PID) calibration. The branching fraction $\mathcal{B}_{x \rightarrow final}$ appears in case any one of the three daughters is unstable and reconstructed in form of some channel $x \rightarrow final$. As is introduced in Chapter 1, the considered decay channel requires the reconstruction of a charged D -meson in the channel $D^+ \rightarrow K^- \pi^+ \pi^+$ such that the world average $\mathcal{B}_{D \rightarrow K\pi\pi}$ is employed. Finally, $\bar{\epsilon}$ is the reconstruction efficiency (see Chap. 10) averaged with the total amplitude as in:

$$\bar{\epsilon} = \frac{\int_{DP} \epsilon(m_{ab}^2, m_{bc}^2) \left| \sum_j a_j e^{i\phi_j} \mathcal{A}_j \right|^2 dm_{ab}^2 dm_{bc}^2}{\int_{DP} \left| \sum_j a_j e^{i\phi_j} \mathcal{A}_j \right|^2 dm_{ab}^2 dm_{bc}^2}. \quad (3.13)$$

3.2. Choice of Dalitz Plot Parametrization

In the case of three-body decay modes that produce two identical final-state particles, as for example in the decay $B \rightarrow D\pi_1\pi_2$, a statistically more advantageous parametrization of the Dalitz plot can be chosen. Because the Dalitz plot is symmetric with respect to the first semiaxis, the Dalitz plot can, figuratively speaking, essentially be folded along this line. Thus, instead of spanning the Dalitz plot using $m(D\pi_1)^2$ vs. $m(D\pi_2)^2$, which might appear as the natural choice, the minimal invariant mass pair $m(D\pi)_{min}^2$ is plotted against the maximal invariant mass pair $m(D\pi)_{max}^2$, where the latter is defined as:

$$m(D\pi)_{max}^2 = \begin{cases} m(D\pi_1)^2, & \text{if } m(D\pi_1)^2 > m(D\pi_2)^2 \\ m(D\pi_2)^2, & \text{otherwise,} \end{cases} \quad (3.14)$$

with the respective implications for $m(D\pi)_{min}^2$. This choice represents an effective mirroring of events north of the first semiaxis down onto the region south of the axis (in relation to the Dalitz plot as defined using $m(D\pi_1)^2$ vs. $m(D\pi_2)^2$). The advantage being an effective doubling of the statistics since the allowed Dalitz plot region is cut in half.

3.2.1. Square Dalitz Plot Representation

As hinted upon earlier, the choice of the two kinematic variables that span the Dalitz plot is not unique. The Dalitz plot parametrization in terms of m_{ab}^2 and m_{bc}^2 for example (referred to as the 'regular' representation hereafter) was found unfit for decays producing light meson daughters. This is because daughters with large three-momenta form candidates at the boundaries of the regular Dalitz plot representation, thus impeding the study of interference effects in these regions [8]. To overcome this drawback, the square Dalitz plot representation was introduced, which ensues from a transformation of the invariant mass pairs squared that maps the regular Dalitz plot into a rectangle:

$$dm_{ab}^2 dm_{bc}^2 \longrightarrow |det J| dm' d\theta', \quad (3.15)$$

with the new coordinates being

$$m' \equiv \frac{1}{\pi} \arccos \left(2 \frac{m_{ac} - m_{ac}^{min}}{m_{ac}^{max} - m_{ac}^{min}} - 1 \right), \quad (3.16)$$

$$\theta' \equiv \frac{1}{\pi} \theta_{ac}, \quad (3.17)$$

where $m_{ac}^{max} = M - m_b$ and $m_{ac}^{min} = m_a + m_c$ are the kinematic limits of m_{ac} (as described in Sec. 3.1.1), θ_{ac} is the helicity angle of the ac daughter combination and J is the Jacobian matrix of the transformation. Both new variables generally range between 0 and 1. However in the analysis at hand, if one transforms the size-ordered $m(D\pi)_{min}^2$ and $m(D\pi)_{max}^2$, the θ' variable is confined between 0 and 0.5. This is because the ordering for a given kinematic configuration of the daughter four-momenta, described by Eq. 3.14, is implicitly determined by the angle of the three-momenta of π_1 and π_2 , i.e. by the helicity angle between the daughter pions. As an effect of the ordering, all helicity angles larger than $\pi/2$ are mirrored as according to $\theta'_{ac} = \pi - \theta_{ac}$.

The transformation to the square Dalitz plot is visualized in Fig. 3.4, where the transformation of a sample, uniformly distributed in the regular representation, is depicted.

Although the present analysis does not entail light mesonic resonances, which would necessitate the square Dalitz plot representation for an increased resolution of the interference region, the square representation has another key advantage in specific applications in that the obtained rectangular kinematic boundary facilitates the modelling of non-parametric probability density functions. Therefore the square representation is employed for the description of the expected background components (as is further detailed in Chap. 9).

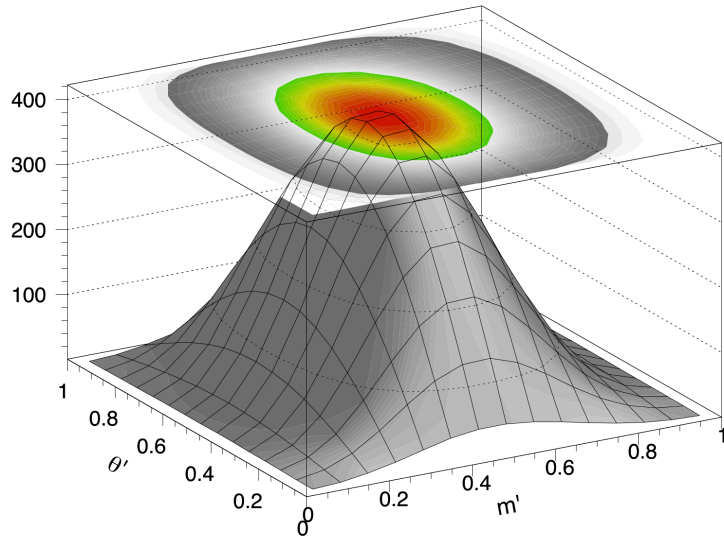


Figure 3.4.: An illustration of the square Dalitz plot representation: The transformed distribution of events uniformly distributed in the regular Dalitz plot.[8].

4. Statistical Analysis Techniques

This chapter presents techniques that are essential to the analysis of the collider data in the work at hand. Because those techniques can be loosely attributed to the field of statistics, their introductions are collected below.

4.1. Maximum Likelihood Estimation

4.1.1. Introduction

In order to estimate the parameters that govern a probability density of interest, the maximum likelihood estimation (MLE), an approach with many desirable statistical properties [9], is often the method of choice. If the probability density $f(x|a)$ at the bottom of a data sample x_1, \dots, x_n is known, the method allows to infer an estimate \hat{a} of the parameter (or the vector of parameters) a from the set of measurements. In the MLE scheme a joint likelihood function is constructed from the individual likelihoods of the data points:

$$\mathcal{L}(a) = \prod_i^n f(x_i|a), \quad (4.1)$$

which, for a given choice of the parameter(s) a , specifies the probability to find the encountered data sample. As according to the maximum likelihood principle, the best estimate \hat{a} is obtained for the value of a that maximizes \mathcal{L} :

$$\hat{a} = \operatorname{argmax}_a \mathcal{L} \leftrightarrow \frac{\partial \mathcal{L}}{\partial a_k} = 0 \quad \forall k, \quad (4.2)$$

whereby it is important to ensure that the probability density $f(x|a)$ is normalized to unity for all values of a . For reasons of practicality the negative logarithmic value of the likelihood function $F(a) = -\sum_i^n \ln f(x_i|a)$ is minimized in most applications, which, with the logarithm being a monotonous function, is admissible:

$$\hat{a} = \operatorname{argmin}_a F(a). \quad (4.3)$$

The task of minimizing the negative log-likelihood is taken care of by the likelihood fitting package `Laura++` [30] in this analysis. The model defined in Eq. 4.1 is extended in the analysis at hand, which is why the performed fit is coined an *extended maximum likelihood fit* (further detailed in Sec. 8.1).

4.1.2. Calculation of Errors

A convenient property of the MLE method is that statistical errors of the parameter(s) a can be readily inferred from the variance of the likelihood function. In the asymptotic case of $n \rightarrow \infty$ the likelihood function \mathcal{L} approaches a (multivariate) Gaussian centered at \hat{a} . By Taylor expanding the negative Log-Likelihood around its minimum \hat{a} and then comparing the Likelihood to a Gaussian, an estimate of the covariance matrix (valid in this limit) can be obtained [9]:

$$F(a) = F(\hat{a}) + \frac{1}{2} \sum_{i,k} \frac{\partial^2 F}{\partial a_i \partial a_k} (a_i - \hat{a}_i)(a_k - \hat{a}_k) + \dots \quad (4.4)$$

$$\rightarrow \mathcal{L}(a) \cong \text{const} \cdot \exp\left(-\frac{1}{2} \sum_{i,k} \frac{\partial^2 F}{\partial a_i \partial a_k} (a_i - \hat{a}_i)(a_k - \hat{a}_k)\right), \quad (4.5)$$

from which the covariance matrix V of the parameter vector a can be inferred:

$$V = \left(\frac{\partial^2 F}{\partial a_i \partial a_k} \Big|_{\hat{a}} \right)^{-1}. \quad (4.6)$$

Multiple σ confidence intervals for the individual parameters a (a_i) can be obtained by simply varying $F(a)$ around its minimum at $a = \hat{a} \pm r \cdot \sigma$:

$$F(\hat{a} \pm r \cdot \sigma) = F(\hat{a}) + \frac{1}{2} r^2. \quad (4.7)$$

Using Laura++, the (1σ) statistical errors of the floating fit parameters are calculated in this fashion.

4.2. Kernel Density Estimation

When accounting for different components in a maximum likelihood fit, their probability densities (PDFs) have to be provided in form of a fit model. As is elaborated upon in more detail in Sec. 9, the PDFs of the background components of the fits in this analysis are a priori unknown. One way to estimate unknown probability densities (PDFs) in a non-parametric way is the so-called kernel density estimation (KDE). The technique allows to approximate the underlying PDF f from a finite data sample (x_1, \dots, x_n) , by assigning kernel functions to each data point. The kernels K are normalized non-negative functions with mean zero and the kernel density estimator for f is given as:

$$\hat{f}(x) = \frac{1}{n} \sum_i^n K_i(x - x_i, \sigma_i^2), \quad (4.8)$$

where each Kernel receives a weight of $1/n$ and is assigned a variance σ_i^2 (also referred to as a bandwidth in this use-case). As opposed to histogramming the data, this technique has the advantage of providing a smooth estimate of the underlying PDF, which is illustrated in Fig. 4.1 below.

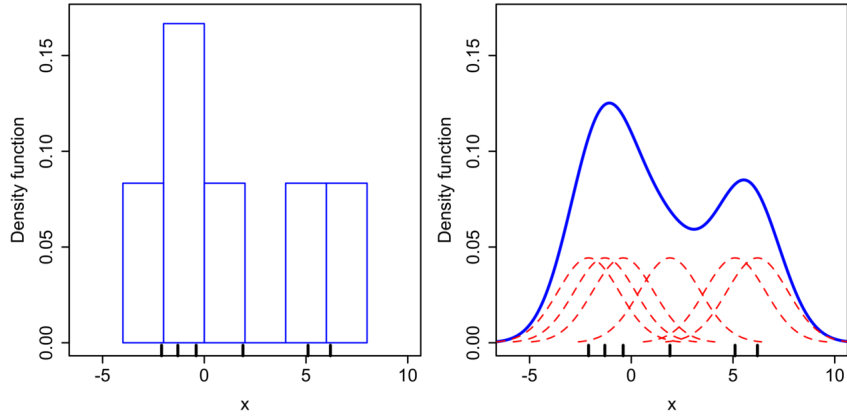


Figure 4.1.: Comparison of the construction of a histogram (left) and a kernel density estimate (right) using the same data [37].

For the KDE application in this analysis Gaussian kernels as implemented in the RooFit toolkit [35] are employed. The kernel bandwidths need to be chosen so as to neither under- nor over-smooth the distribution. For this reason, an adaptive approach is taken in which the results from different bandwidth KDE results are blended together (see Sec. 9).

4.3. Statistical Hypothesis Testing

At a number of occasions in this analysis statistical tests are conducted to classify data as according to different hypotheses. Concepts of such statistical tests are for example applied when evaluating the performance of multivariate classifiers (see Sec. 7.6) or when assessing the goodness of fits to individual resonances (see Chap. 11), and are therefore briefly introduced below.

The data samples under scrutiny in statistical tests are usually samples of random variables following a distribution that is a priori unknown (and often the very goal of the hypothesis test). Practically speaking, a random sample $X = X_1, \dots, X_n$ may be drawn from a distribution $f(\mu, \sigma^2)$ from which the true mean value μ and true variance σ^2 are unknown but can be tested for using different hypotheses.

Depending on what hypothesis the test aims to check for (e.g. $\mu = \mu_0$) a suitable quantity, referred to as an estimator or a test-statistic, needs to be constructed. The test statistic, say $T(X)$, is a function of the random sample. By construction its distribution given a certain hypothesis H_i , g_{H_i} , is known and its distribution thus encodes the hypothesis. The hypothesized distribution of T can then be utilized to choose a critical value z that separates regions in T for which the hypothesis is either rejected or not rejected. z is constructed so that only in α parts of the cases, the correct hypothesis is rejected. So if, for example, we intend to test for the null hypothesis H_0 that $\mu = \mu_0$ versus the alternative hypothesis H_1 that $\mu = \mu_1 > \mu_0$, the critical value is found using G_{H_0} , the cumulative distribution of T under H_0 :

$$G_{H_0}(T \geq z) < \alpha. \quad (4.9)$$

If $T(X) < z$, H_0 is not rejected with significance level α and if $T(X) \geq z$, H_0 is rejected with significance level α . Usually, instead of quoting $T(X)$, the probability to find the encountered $T(X)$ given the validity of H_0 is stated, which is referred to as the p-value p . For the chosen

example, p would be found as:

$$p = \int_{T(X)}^{\infty} g_{H_0}(T') dT'. \quad (4.10)$$

Using this hypothesis testing scheme, different kinds of test outcomes and errors can occur, which are presented in Tab. 4.1 below.

true hypothesis	test result	
	H_0	H_1
H_0	correct	false negative error
H_1	false positive error	correct

Table 4.1.: Possible outcomes of hypothesis tests.

The fraction of the data which is falsely discarded as not being in line with the null-hypothesis is referred to as the false negative rate. By definition it is equivalent to the chosen significance level α and thus amounts to:

$$\alpha = \int_z^{\infty} g_{H_0}(T') dT'. \quad (4.11)$$

The fraction of data falsely accepted as being in line with the null-hypothesis is referred to as the false positive rate and is obtained through:

$$\beta = \int_{-\infty}^z g_{H_1}(T') dT'. \quad (4.12)$$

The above test scenario as well as the related error rates are depicted in Fig. 4.2, where the distributions for two different hypotheses over a test-statistic are illustrated.

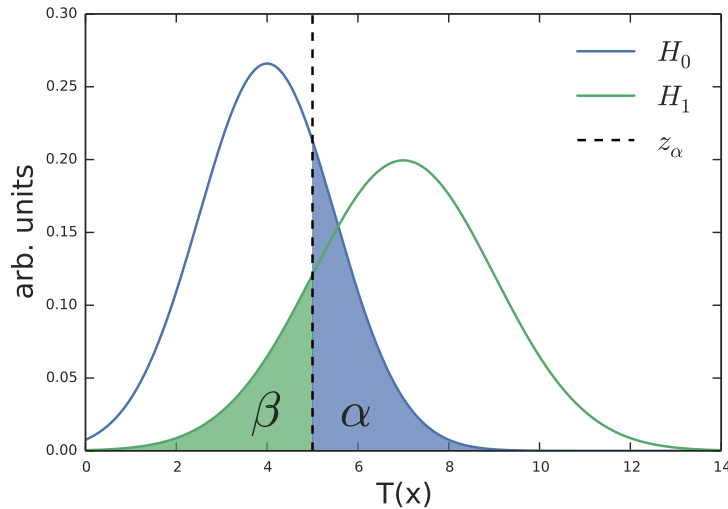


Figure 4.2.: The distributions of two different hypothesis H_0 and H_1 as a function of the test statistic $T(X)$ are shown. z_α , the critical value with significance level α , is illustrated as a dashed line and defines the boundaries from which the probabilities of making a true positive error (shaded green area corresponding to β) and of making a false negative error (shaded blue area corresponding to α) ensue.

Multivariate classification techniques, such as the ones introduced below, methodologically deviate quite some from the classical hypothesis test described above in that they ‘learn’ to combine a multidimensional input from which they then infer a classification. Nonetheless, using an appropriate transformation, their classification can again be interpreted as a test-statistic. For this reason, their performance can be evaluated on labeled data using the concepts described above. To assess the performance of trained multivariate classifiers in binary classification problems it is common practice to construct what is called the *receiver operating characteristic* (ROC). The ROC is a graph of the dependency of the true positive rate ($1 - \alpha$) on the false positive rate (β) as the cut on the test-statistic (z_α) is varied. The integrated curve, referred to as the *area under the receiver operating characteristic* (AUC), is used to compare the performance of different classifiers, an example of which is shown in Fig. 7.4.

4.4. Multivariate Classifiers

Suppressing specific types of background can be viewed as a classical binary classification problem. Such tasks are commonly approached using multivariate classifiers. All classifiers used in this work have to be trained on explanatory features using annotated data sets, which in part are required to contain quite substantial amounts of labeled examples. Fortunately, high-energy physics does not suffer from a shortage of labeled training data, quite contrary, arbitrary amounts can be generated with Monte Carlo event generators (see Chap. 6).

During this so-called supervised training of the classifiers, internal parameters are adjusted so as to minimize a given loss-function, which parametrizes the learning task. In the learning process, there generally is a risk of overfitting, which means that the classifier could learn and memorize statistical fluctuations of the training data. Obviously, this is highly undesirable, given that the task is to generalize well enough so as to be able to reliably classify previously ‘unseen’ data, i.e. events from independent data sets. The classifiers’ ability to generalize well can be enforced through different regularization techniques. These either reflect specific choices of the classifiers’ internal architecture or are imposed in the form of particular learning strategies. Both are tuned by hand-regulated parameters, so-called hyperparameters.

In the following, two different classifiers popular in high-energy physics, a commercial version of an artificial neural network as well as a boosted decision tree are introduced as they are applied and evaluated in terms of their performance in the suppression of a specific background (see Sec. 7.6). In order to probe the trained classifiers on data, they hadn’t come across before, all performance related measures, such as the AUC, are assessed on an independent test set equally large as the employed training set. Both classifiers are found to exhibit a strong out of the box ability to generalize.

4.4.1. Artificial Neural Networks

Artificial neural networks are loosely inspired and motivated by biological neurons, as found in nervous systems such as ours. In these systems, each neuron receives impulses through synaptical connections and, once a certain activation threshold is reached, in turn transmits a signal response to other neurons. The basic element of an artificial neural network can be pictured as exhibiting several input nodes, which each take in one element of a given feature vector \vec{x} . Each basic artificial neuron computes the scalar product of the feature vector and a weight vector \vec{w} , which is then transformed using a nonlinear activation function σ , that often maps the scalar product to a range between either 0 and 1 or -1 and 1. This scheme is further illustrated in Fig. 4.3.

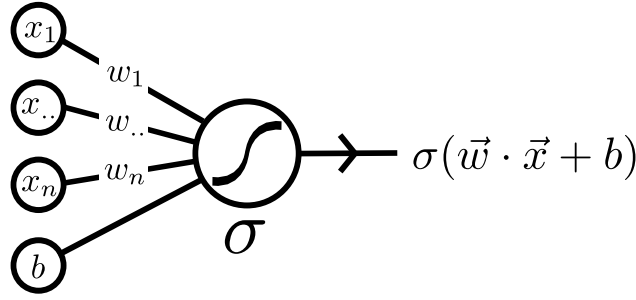


Figure 4.3.: Schematic illustration of a single artificial neuron.

During the learning process, the weights of the weight vector are adjusted so as to span a hyperplane that constitutes the separation between different classes. The weight vector can be viewed as the normal vector to the hyperplane and, as according to the *Hessian normal form* for planes, the resulting scalar product gives the distance d of a given input vector to the plane:

$$d = \vec{w} \cdot \vec{x} + b, \quad (4.13)$$

where a bias term b is added to allow for a shift of the planes' distance to the origin. The non-linearity σ thus effectively gives a continuous response as to whether the input lies below (negative d) or above (positive d) the learned separating plane.

Many such neurons can be combined to form what is called a multilayer perceptron. More sophisticated methods exist, but are not covered here. Each node in every layer of such a network is connected to all outputs of the previous layer, which is again implemented via weights (w_{ij}). This is described as a feed-forward network, in which the responses to the input vector are propagated through the network from its first layer to the last, where its classification response is evaluated using a loss-function L . The learning is performed using the backpropagation algorithm, which constitutes successive updates of the weights as according to their influence on the loss-function:

$$w'_{ij} = w_{ij} - \eta \frac{\partial L}{\partial w_{ij}}, \quad (4.14)$$

where the parameter η is the learning-rate, a key hyperparameter for a stable and fast convergence to a near-optimal minimum of L . There are many other hyperparameters such as different methods to optimally modify η , the number of hidden layers, the number of neurons in each such layer, the choice of activation function and the choice of loss-function.

This analysis makes use of the commercial package NeuroBayes [18], which is based on an artificial neural network and includes intensive preprocessing of the input variables as well as advanced weight-update strategies.

4.4.2. Boosted Decision Trees

Boosted decision trees (BDTs) assign a given candidate a signal probability (\mathcal{O}_{BDT}) according to the region it is part of in the provided variable space. The boundaries (cuts) of those disjunct regions are established in the supervised training phase of the BDT. During this phase, the (binary) class labels for each candidate are known and the cuts defining each such region are determined so as to maximize the separation between the classes, as quantified by a separation

measure like the entropy difference Q :

$$Q(p) = p \log p + (1 - p) \log(1 - p), \quad (4.15)$$

where p is the fraction of correctly classified events at each split. The cut itself is determined by maximizing the sum of the Q values of the daughter nodes weighted by the number of events.

Like the name already suggests, BDTs can be represented by a tree structure which is depicted in Fig. 4.4. Each node therein stands for a binary decision executed in the form of a cut on one of the variables in the given parameter space. This analysis employs a speed-optimized version of a *stochastic gradient boosted decision tree* as described in [19]. Several different regularization measures ensure that the BDT performs well on independent data: The depth of the tree (the number of consecutive cuts on different variables) is limited to three, the training is performed stochastically in that the samples to train on are drawn randomly and most importantly an ensemble of weakly classifying trees is assembled to form a strongly separating classifier, a ‘forest’ of trees, figuratively speaking [25].

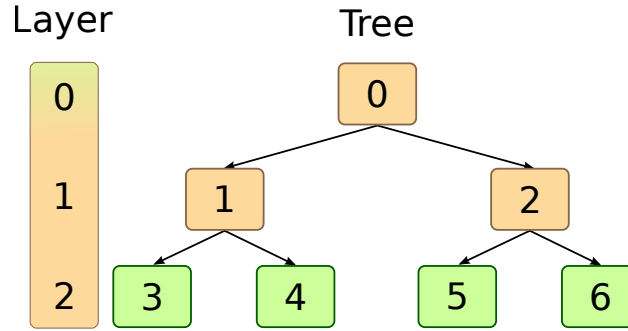


Figure 4.4.: An exemplified tree classifier with a layer depth of three as taken from [25].

During the training a boosting algorithm is used to weight individual candidates according to how well they are classified. The training consists of a succession of trainings of individual trees, where the average error is calculated for each tree. During the training of the successive tree, misclassified events receive a higher weight as a function of this average error of the previous tree and thus regions that are hard to classify gain in importance. In that manner weak learners are added consecutively to the ensemble and eventually build a well-regularized classifier, whose classification is given in terms of a performance-weighted sum of the individual trees' classification.

5. Belle II Analysis Software Framework

Although the analysis at hand utilizes data recorded by the Belle detector, the reconstruction of signal candidates from the data is performed using the *Belle II Analysis Software Framework*, in short *basf2*. This choice was made as Belle analyses usually ran highly customized code, which was often both implemented inefficiently and inconvenient for others to reproduce. Using the functionalities provided through *basf2*, reconstruction procedures common across analyses do not have to be reimplemented from scratch. In the present case there is the additional benefit that the provided modules can be tested in the context of a realistic use-case for the very first time.

The Belle II software framework is implemented in C++ and can be steered using Python3. It is written in a modular fashion, where each module is a small processing block built on top of a range of available libraries, which are shared between the modules. Each such block operates on data event-by-event and performs self-contained tasks. The modules communicate with one another through a common *DataStore*, from and to which they can read and write data-objects. This concept is illustrated in Fig. 5.1.

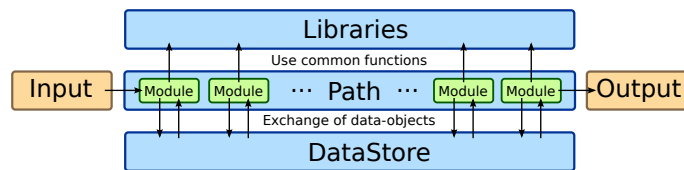


Figure 5.1.: The individual *basf2* modules are lined up in a path and communicate through the common *DataStore* [25].

In order to be able to process Belle data in *basf2*, the Belle *mdst*-files have to be converted to meet the expected format including for example serialized data-objects (amongst others *Track*, *ECLCluster*, *KLMCluster* and *MCParticle* data-objects). This functionality is available through the *b2bii* conversion module. Once the data is made available in this fashion in the *DataStore*, the ‘online’ reconstruction, a term further specified and used in Chap. 7, can proceed according to the path of modules defined in the steering-file. As part of this work, new functionalities were added to *basf2* (to and in the form of modules), such as the introduction of the *variablesToExtraInfo* module. In order to show how this module can be used in Dalitz analyses as well as to enable full reproducibility of the online reconstruction of this analysis, the employed steering-file including further annotations is given in Appendix A.

6. Blind Analysis using Monte Carlo Data Samples

In particle physics it is of great importance to eliminate any forms of unintentional biases introduced by the experimenter, which is a source of error that easily creeps in when analysis procedures are tweaked on real data. To circumvent this human error, it is good scientific standard in the community to perform what is called a 'blind analyses'. A blind analysis is a measurement such as that of a branching fraction or upper limit that is performed without looking at the signal box of recorded data until most or all analysis criteria are finalized [8]. After the analysis procedure is optimized and fixed on simulated data, a so-called 'box-opening' can take place, at which point real recorded data is subjected to the analysis for the first and only time.

The upcoming Chapters 7, 9 and 10 all describe the analysis steps prepared on the basis of simulated data. These steps entail the reconstruction of the considered decay channel subject to a so-called (cut-based) 'online' selection, an 'offline' selection (based on multivariate classification) as well as a set of preparatory steps for ingredients vital to the eventual fit of the amplitudes in the Dalitz plane. This basic work-flow is pictorially represented in Fig. 6.1.

It is clear that the approach of a blind analysis is only justified if the used simulated data is essentially indistinguishable from recorded data. Fortunately a range of very sophisticated event generators exist that model the final-states of e^+e^- collisions to a mostly trust-worthy degree of accuracy. The generators depend on theoretical models of interactions to calculate the four-vectors of the final-state particles and have to be able to describe both signal and background decays of e^+e^- -events. Once produced, the four-vectors are passed by the software framework to the detector simulation where they are tracked in the detector, taking into account the interaction between the particles and the different materials, specific to the Belle detector and its geometry, in order to eventually compute the electronic signals which mimic the detector response [8].

The event generators heavy usage of random number generators to draw statistically independent samples is reminiscent of the underlying randomness that is also at the core of gambling. This is why the simulated data is most often referred to as *Monte Carlo* data (MC). The different Monte Carlo samples used for the blind analysis at hand are described in the following.

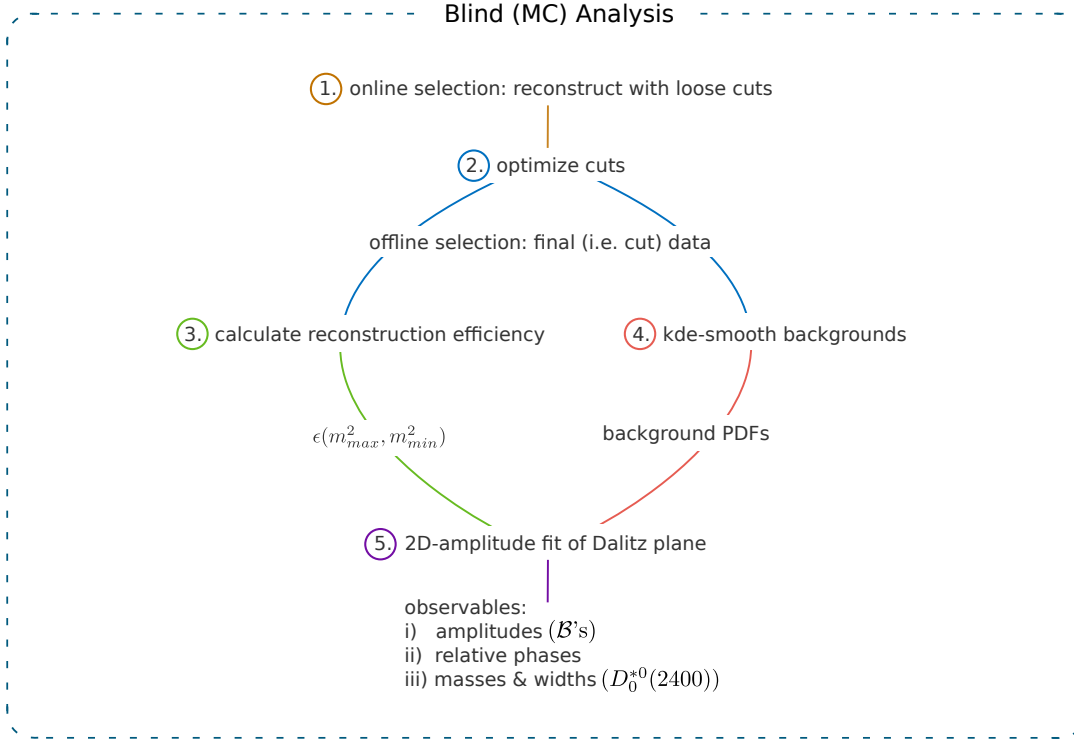


Figure 6.1.: A pictorial representation of the work-flow of the analysis at hand. In accordance with the principles of a blind analysis all analysis criteria are determined and fixed on Monte Carlo data before a prospective look at recorded data.

6.1. Belle Generic Monte Carlo

The software `Evtgen` [29] is used to simulate *generic* Monte Carlo data. The term refers to the most frequent B -meson decay channels proceeding through $b \rightarrow c$ transitions at the quark level. `Evtgen` uses decay amplitudes instead of probabilities for each node in the decay tree in order to simulate the entire decay chain [8] and reads in the *PDG* world averages of 2010 for the respective branching ratios. Ten generic Monte Carlo datasets, referred to as streams, are available. They are further split into 'charged' and 'mixed' samples, which is Belle-speak for B^+B^- and $B^0\bar{B}^0$ decays respectively. Each of the streams corresponds to the total integrated luminosity of Belle which matches around $772 \times 10^6 \Upsilon(4S) \rightarrow B\bar{B}$ pairs. The generic Monte Carlo also contains the considered signal decay channels. The employed `Evtgen` model for the generic MC however does not model the interferences of the resonant sub-channels, which is why the signal channels are deselected using the MC truth information and then replaced by MC events obtained through a more accurate MC model, see Sec. 6.3 below. Moreover, it is apparent that the charged Belle generic MC wrongfully includes both the inclusive branching fraction as well as the resonant exclusive branching fractions (as gathered from the employed decay-file 'DECAY.DEC', lines 3557 ff.) in $B^- \rightarrow D^+\pi^-\pi^-$, thus severely overestimating the combined yield, which is alarming considering analyses were the decay constitutes a background channel. The deselected generic MC is one source of background that comes about either through simply mimicing signal events or by lending particles to the signal event during the particle combination in the reconstruction, both of which is clearly undesirable. Fig. 6.2 displays Dalitz plots of both charged and mixed backgrounds. As introduced in Sec. 3.2 the Dalitz plots are cut in half and thus display a sharp and straight boundary in the left upper corner of the plots.

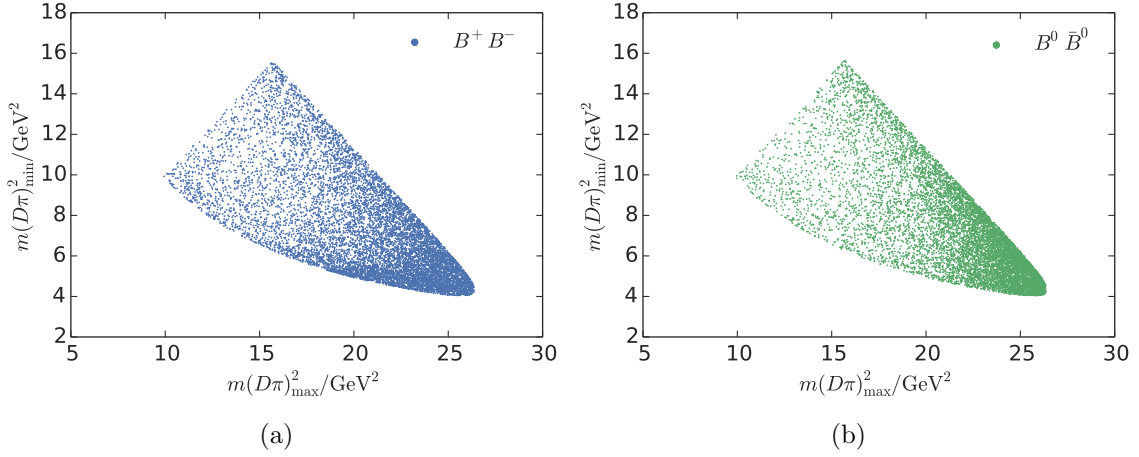


Figure 6.2.: Charged and mixed background samples from generic MC are shown in the left and right plot, respectively. In the charged sample the signal component is deselected. The figures depict equally sized sub-samples of reconstructed candidates of stream 1 generic MC.

6.2. Continuum Monte Carlo

Another background source are continuum events from $e^+e^- \rightarrow q\bar{q}$ ($q \in \{u, d, s, c\}$) processes, further detailed in Sec. 7.6. The strong hadronization process of the ensuing shower of quarks requires dedicated generators, which in this case are Pythia and Jetset [34]. The generation is done separately for $q = c$ and $q \in \{u, d, s\}$ processes, ingeniously referred to as *charm* and *uds* MC samples. Again, there are ten streams of MC for the two distinguished types, each containing the appropriate amounts of continuum background as expected to be found within the entity of recorded Belle data. Fig. 6.3 illustrates Dalitz plots of both charm and uds background components.

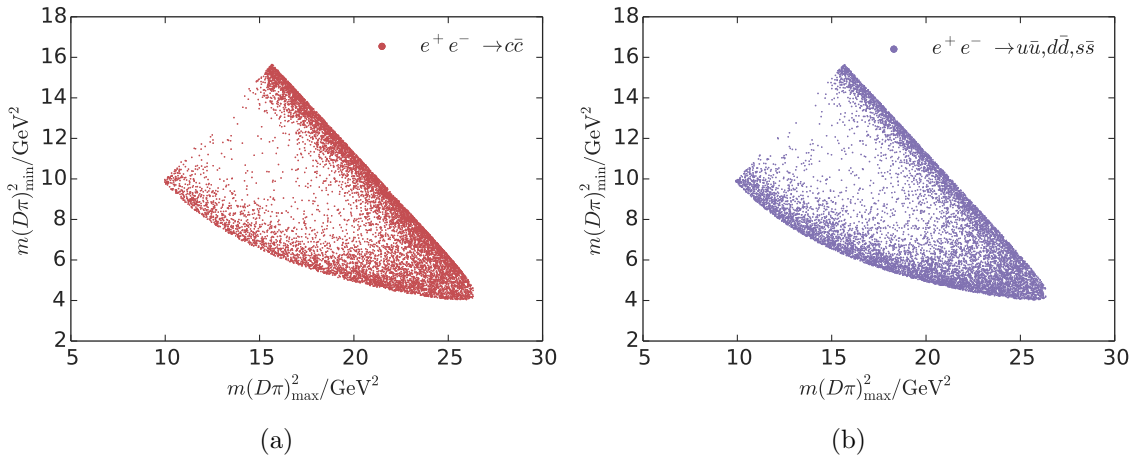


Figure 6.3.: Charm and uds background samples from continuum MC are shown in the left and right plot, respectively. The figures depict equally sized sub-samples of reconstructed candidates of stream 1 continuum MC.

6.3. Evtgen Generic Dalitz Model for Signal Monte Carlo

As hinted upon earlier, the Evtgen model used for the generation of generic MC is severely impaired by its failure to account for interference effects. Besides the necessity of a physically sound model it is also apparent from the expected signal yield in recorded data (see Chapter 1) that larger numbers of signal MC are required in certain preparatory steps heavily relying on signal statistics than the ones available in all generic MC streams combined. One such step hungry for statistics is the determination of the reconstruction efficiency where 50 Mio. non-resonant $B \rightarrow D\pi\pi$ -decays are used (see Chap. 2).

For these reasons signal MC is generated separately using Evtgen's *EvtGenericDalitz* model which is in principal also able to simulate signal events based on the coherent sum of resonant sub-channel amplitudes as given in Eq. 3.6. Signal MC samples are mostly used to determine detector effects quantified in the form of the mentioned reconstruction efficiency. As is common in many Dalitz analyses, the phase-space mode is used for this purpose [31]. Thus for this application as well as for testing the fit's performance on single resonances, a coherent addition of amplitudes is redundant.

In general, it would be desirable to assess the fit's performance using the EvtGenericDalitz model for coherent signal samples. Unfortunately, the EvtGenericDalitz model is not free of limitations either as in this model's implementation the individual amplitudes $\mathcal{A}_{i, Evtgen}$ are not normalized, such that their branching-fractions in relation to one another are off. An attempt to rectify this deficiency can be made by re-weighting the amplitudes by the ratio of the maximal Evtgen amplitude and the maximum of a correctly normalized amplitude \mathcal{A}_{LAURA} . This is carried out as sketched in

$$\mathcal{A}_{i, Evtgen} \rightarrow \sqrt{\mathcal{B}_{PDG, i} \cdot \frac{\mathcal{A}_{max, LAURA}^2}{\mathcal{A}_{max, Evtgen}^2}} \times \mathcal{A}_{i, Evtgen}. \quad (6.1)$$

Here it is assumed that the respective maximal values occur at the same location in the Dalitz plane and that the two amplitude models are identical otherwise. Along with this approximate normalization the factor $\mathcal{B}_{PDG, i}$ (the branching ratio of component i as according to [31]) seeks to ensure the correct ratio of the individual components i .

While EvtGenericDalitz model's shortcomings do not affect the separate generation of individual components, tests with meaningful implications on the fit's reliability in case of the coherent signal model are not conducted using this model. Instead, coherent signal samples are generated using the employed fit framework, which is further detailed in Chapter 12. The above approximation in the coherent EvtGenericDalitz model is thus harmless and the only appearance it makes is in Appendix C, appearing as an aside in the context of fits to individual resonances (see Chap. 11).

Signal MC samples modeled individually for single resonances with the EvtGenericDalitz model are shown in Fig. 6.4. In accordance with the descriptions of the Isobar model (Sec. 3.1.3) the non-resonant decay is distributed uniformly across the Dalitz plot whereas the decay of the scalar resonance (D_0^{*0}) displays a uniform event density along its resonance band. The decay of the tensor resonance (D_2^{*0}) in turn exhibits its characteristic lobe-shaped non-uniformity along its resonance band.

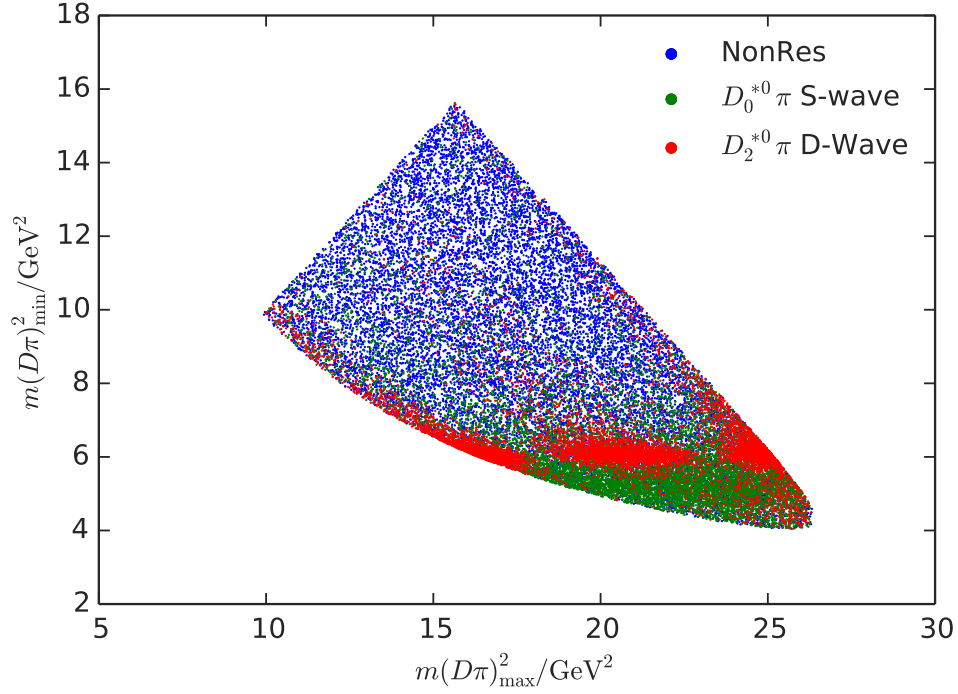


Figure 6.4.: Signal MC of individual resonant components generated using the EvtGenericDalitz model. The blue component shows the non-resonant decay, the green component shows the S-wave contribution and the red component depicts the D-wave contribution. The three components are shown in the correct proportion.

6.4. Detector Simulation

Each of the produced MC samples is further processed by a detector simulation, that models all detector effects and generates the signatures of hits and clusters in the different detector modules that are left behind by the particles. For this purpose the simulation package Geant4 [4] is utilized. Finally we also add beam background caused by intrabeam scattering or interactions with the beam gas. Beam background cannot be simulated easily and is thus recorded with the detector at random points in time in between e^+e^- collisions and added to the simulation [21].

7. Reconstruction and Selection

This chapter discusses the reconstruction of charged B -meson candidates and the developed selection requirements. The reconstruction is structured hierarchically in that it combines the charged final-state tracks, as reconstructed in the Belle detector (see Chap. 2), to form B candidates. In each intermediate step therein, selection criteria have to be met, which consist of cuts on kinematic variables and cuts on the results of multivariate classifiers. The procedure is loosely divided into an ‘online’ and an ‘offline’ selection, with the former referring to such requirements directly enforced when processing the raw MC data (or real data once applied after the prospective unblinding) and the latter denoting requirements that are imposed afterwards (once the respective requirements are optimized). This approach allows for the desired degree of flexibility during the process of streamlining the selection strategy. The criteria are optimized on Monte Carlo simulated data, which is further detailed below.

7.1. Reconstruction Channel

For this study the exclusive decay of $B^- \rightarrow D^+ \pi^- \pi^-$ and the charge conjugate combination was selected. The D -meson was reconstructed in the mode $D^+ \rightarrow K^- \pi^+ \pi^+$ since it constitutes the most frequent hadronic D -decay mode [31]. In doing so, this analysis also follows [3], where it was found that the signal-to-noise ratio for other D decay modes in Belle data is much lower.

7.2. Final-State Particles

As part of the online selection, the charged tracks of the final-state particles are subjected to initial requirements. Those include cuts on the ‘impact parameters’, the absolute distance of closest approach to the interaction point, in both the transverse (dr) and longitudinal (dz) direction with respect to the laboratory frame. Moreover the responses from the Aerogel Čerenkov Counter (ACC) and the Time-of-Flight Counter (TOF) is combined with dE/dx measurements in the Central Drift Chamber (CDC) to provide separation of *kaons* and *pions* in the form of a likelihood ratio (for a description of the detector parts see Chap. 2):

$$PID_{\pi/K} = \frac{\mathcal{L}(\pi)}{\mathcal{L}(\pi) + \mathcal{L}(K)}, \quad (7.1)$$

$$PID_{K/\pi} = \frac{\mathcal{L}(K)}{\mathcal{L}(\pi) + \mathcal{L}(K)} = 1 - PID_{\pi/K},$$

where PID stands for particle identification. The respective PID cuts are adopted from the first $B^- \rightarrow D^+ \pi^- \pi^-$ analysis by A. Kuzmin et al. of the Belle collaboration [3]. The track-finding algorithms that fit the final-state trajectory to the combined hits, registered in the different detector modules, further provide a p value of the χ^2 -probability of the fit result, p_{track} , which is cut upon loosely. The employed final-state preselection requirements are given in Table 7.1.

<i>variable</i>	π^\pm	K^\pm
PID	$PID_{\pi/K} > 0.2$	$PID_{K/\pi} > 0.6$
$ dr $	$< 0.2 \text{ cm}$	$< 0.2 \text{ cm}$
$ dz $	$< 4.0 \text{ cm}$	$< 4.0 \text{ cm}$
p_{track}	> 0.001	> 0.001

Table 7.1.: Final-state particles preselection cuts.

7.3. D -Meson Reconstruction

The final-state particles that meet the above defined criteria (Table 7.1) are subsequently combined to form D^+ candidates from $K^- \pi^+ \pi^+$ combinations. In the online selection, the D candidates' invariant masses are required to lie in the interval $[1.84 \text{ MeV}/c^2, 1.89 \text{ MeV}/c^2]$, which represents a generous padding around the nominal D^+ mass of $1.869 \text{ GeV}/c^2$. The invariant mass distribution of the resulting D candidates is shown in Fig. 7.6e. Besides combining the final-state candidates to fall within this mass window, they are subjected to a mass-constrained vertex fit which is required to perform above a certain goodness of fit threshold, as described in Section 7.4.

7.4. Vertex Reconstruction

It is not only desirable to find the daughter particle combinations in sufficiently good agreement with the nominal mass of their direct mother, but it is also very important to ensure that those particles originate from the same decay vertex. This is checked by performing a kinematic fit which returns the most probable decay vertex for a given set of daughter tracks using the vertex reconstruction toolkit Rave [36]. The fit is conducted by minimizing a χ^2 term that takes into account the track parameters and their error matrices. Here again the p value of the χ^2 -probability of the fit result can be consulted to decide whether or not the daughters are believed to originate from a common mother.

In this analysis, two successive vertex fits are performed per candidate: A vertex fit aiming to reconstruct the D -decay vertex followed by a vertex fit intended to reconstruct the B -decay vertex. In both cases an additional contribution is added to the 'bare' χ^2 -term. This supplementary constraint favors the daughter particles' combined invariant mass to comply with the nominal mass of their direct mother. Each time such a 'mass-constrained' vertex fit is conducted, all kinematic variables of the involved daughters are updated. This approach is chosen so as to assure minimal migration of B candidates, i.e. the candidates' invariant mass combinations $m(D\pi)_{min}^2$ and $m(D\pi)_{max}^2$, away from their position of origin within the

Dalitz plot. While the kinematic variables that are eventually used in the fit to the Dalitz plot are purposely modified in this manner, it is important to not alter the distributions of the variables that are used in the candidate selection procedures. Consequently copies of the variables before the vertex fits are carried out are retained for this intent. Furthermore such candidates for which the vertex fits fail are rejected by exclusively accepting vertex fits with p-value (p_{track}) larger or equal to zero.

7.5. *B-Meson Reconstruction*

B candidates can be identified by their center of mass (c.m.) energy difference ΔE and the beam-constrained mass M_{bc} as defined below:

$$\begin{aligned}\Delta E &= \sum_i E_i - E_b = \sum_i \sqrt{m_i^2 + \vec{p}_i^2} - E_b, \\ M_{bc} &= \sqrt{E_b^2 - \left(\sum_i \vec{p}_i\right)^2},\end{aligned}\tag{7.2}$$

where $E_b = \sqrt{s}/2$ is half the beam energy in the $\Upsilon(4S)$ c.m. rest frame and \vec{p}_i , m_i and E_i are the three-momenta, rest-masses and energies of the B -meson candidates' daughters (indexed by subscripts i). Here one key design feature of B-factories such as KEKB is exploited in that nearly all $\Upsilon(4S)$ -resonances decay into B -meson pairs, each of which is transferred exactly half the beam energy.

Both variables defined in Eq. 7.2 can be utilized to span what's colloquially referred to as the 'signal box'. During the online selection this signal box is chosen such that B candidates comply with $|\Delta E| < 0.2 \text{ GeV}$ and $M_{bc} > 5.27 \text{ GeV}/c^2$. Both variables enter the offline B -meson selection as input to the multivariate classifier employed for this task (see Sec. 7.7). Fig. 7.6a and Fig. 7.6c show the respective distributions for stream 1 MC and the corresponding amount of signal MC as encountered after the online selection. At this point it should be acknowledged once more that the ΔE and M_{bc} variables are taken from before the B -mass-constrained vertex fit is carried out.

7.6. Continuum Suppression

Aside from kinematic variables such as ΔE and M_{bc} there is one more very effective handle on background rejection to work with: The topological difference of continuum events and $b\bar{b}$ events, which manifests itself in distinct event shapes, i.e. angular distributions. The latter are accessible through a large number of respectively chosen variables and hence allow for good separation between the two categories.

The term continuum background refers to B candidates from $e^+e^- \rightarrow q\bar{q}$ ($q \in \{u, d, s, c\}$) events, which is the dominant background in this analysis prior to applying the suppression technique described in the following. Those $q\bar{q}$ events entail two initial quarks with large three-momenta, which move back-to-back to one another and, in the process of strong fragmentation, produce jets of light hadrons [8]. In contrast, both B -mesons produced in the desired $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ events are found almost at rest in the $\Upsilon(4S)$ c.m. frame since the $B\bar{B}$ threshold is barely below the $\Upsilon(4S)$ rest mass. The angular distribution of the B decay products, with the B -meson being a spin 0 particle, is consequently isotropic. The event

structure of the two distinct scenarios is illustrated in Fig. 7.1. As the fundamental idea of the Dalitz plot technique is to exploit the kinematic structure of the considered candidates, it is clear that the jet-like behavior of continuum is also accessible in Dalitz plots by the bare eye. Figs. 6.3a and 6.3b display an elevated event density of continuum events towards the Dalitz plot boundaries, which is expected following the arguments in Sec. 3.1.1. Isotropic events such as in generic MC samples on the other hand are more evenly distributed across the Dalitz plane as is evident in Figs. 6.2a and 6.2b.

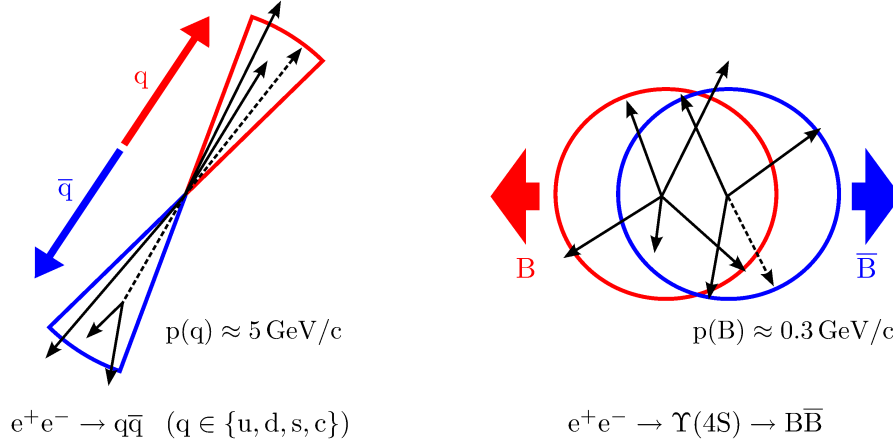


Figure 7.1.: The left-hand side depicts the event shape of a typical continuum event whereas the right-hand side is an illustration of a typical $B\bar{B}$ event. Taken from Reference [33].

As mentioned above, there is an abundance of variables, that provide separation based on the angular distributions of the decay particles. Below find a brief description of the variables employed in the analysis at hand, which, after comparison to an artificial neural network, are further combined in a boosted decision tree classifier (BDT) for an even more powerful discrimination.

Thrust Variables

A variable providing good separation is the angle θ_T between the thrust axis of the decay products of the B (evaluated in the $\Upsilon(4S)$ rest frame) and the thrust axis of the remaining particles of the event, also referred to as the ‘rest of event’ (or ROE). For a set of N particles the thrust axis is defined as the unit vector \mathbf{T} with respect to which the projection of all particles’ three-momenta \mathbf{p}_i is maximal:

$$\mathbf{T} = \underset{\mathbf{T}}{\operatorname{argmax}} \left(\sum_i^N \mathbf{T} \cdot \mathbf{p}_i \right). \quad (7.3)$$

Several related variables are useful. $|\cos\theta_T|$ for example is uniformly distributed for isotropic events as is the case for $B\bar{B}$ processes in contrast to $q\bar{q}$ processes, which display a pronounced peak ascending towards $|\cos\theta_T| = 1$, as can be seen from Fig. 7.2a. Similarly $|\cos\theta_{Tz}|$, where θ_{Tz} is the angle of the thrust axis of the B daughters and the beam pipe (Fig. 7.2b), the magnitude of the thrust axis of the B daughters (B -thrust, as in Fig. 7.2c), as well as the magnitude of the thrust axis of the ROE (ROE -thrust, as in Fig. 7.2d) are employed for discrimination. The thrust axis magnitude is defined as

$$T = \frac{\sum_i^N \mathbf{T} \cdot \mathbf{p}_i}{\sum_i^N |\mathbf{p}_i|}. \quad (7.4)$$

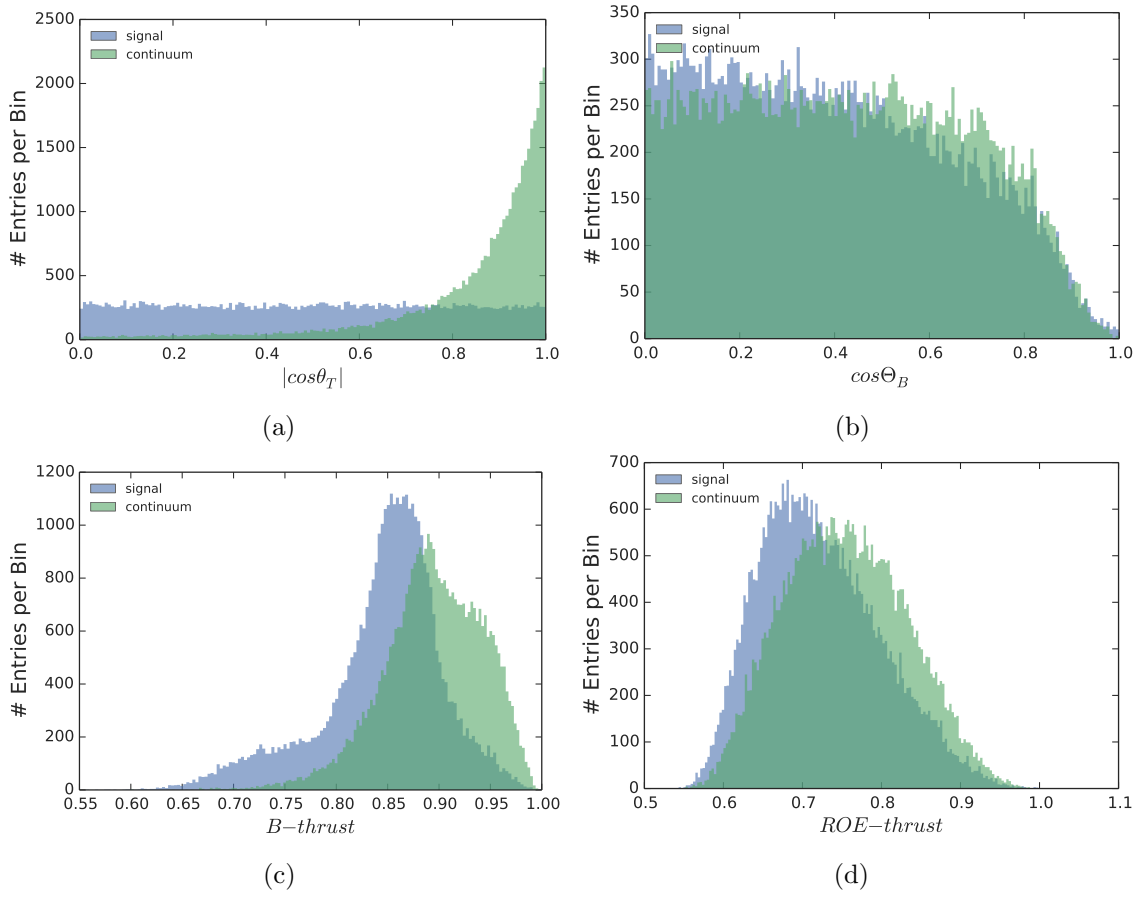


Figure 7.2.: The distribution of different thrust related variables used to suppress continuum background.

Fox-Wolfram- and Modified Fox-Wolfram moments

The k^{th} Fox-Wolfram moment H_k and the k^{th} normalized Fox-Wolfram moment R_k for a set of N particles are given by

$$H_k = \sum_{i,j}^N \frac{|\mathbf{p}_i||\mathbf{p}_j|}{E_{tot}} P_k(\cos\theta_{ij}) \quad \text{and} \quad R_k = \frac{H_k}{H_0}, \quad (7.5)$$

where the summation is over the particle indices i and j , \mathbf{p}_i denotes the i^{th} particle three-momentum and $P_k(\theta_{ij})$ is the k^{th} Legendre polynomial as a function of the angle θ_{ij} between particle pair $\{ij\}$. The moments provide a parametrization of the energy and momentum flow in an event and were first developed by G.C. Fox and S. Wolfram in 1978. Strongly collimated jets display R_k values close to zero (one) for odd (even) values of k [8].

To further enhance the discrimination power for continuum events, the Belle collaboration introduced two Fisher discriminants that are made up of linear combinations of observables similar to the H_k of Eq. 7.5 which are referred to as the ‘Super Fox-Wolfram Moments’ (SFW) and ‘Kakuno Super Fox-Wolfram Moments’ (KSFw) respectively. The modified moments are different in that here the double sum is not carried out over the entire set of particles of the event at hand but instead one summation parameter runs over particles associated with the B candidate (denoted as s) and the other over such associated with the ROE (denoted as o). Additionally, particles of the set o are further distinguished in that charged, neutral and undetected particles (associated with missing mass) are used separately in different moments, labeled accordingly by c , n and v superscripts. Instead of combining the moments in the mentioned dedicated Fisher discriminants they are used separately as input variables to multivariate classifiers as described in Section 4.4.2. A total of 16 KSFw variables is available in the Belle II Framework (their distributions as used in the classifier training are shown in Fig. B.1 and Fig. B.2; for more intricate descriptions and definitions refer to Ref. [8]):

- E_T : scalar sum of transverse energy of all particles of the event
- mm^2 : missing mass squared
- $H_i^{so,c}$: using charged tracks of ROE , $i \in \{0, 2, 4\}$
- $H_i^{so,n}$: using neutral tracks of ROE , $i \in \{0, 2, 4\}$
- $H_i^{so,v}$: using missing momentum of ROE , $i \in \{0, 2, 4\}$
- H_i^{oo} : double sum over ROE tracks only, $i \in \{0, 1, 2, 3, 4\}$

CLEO-Cones

The CLEO collaboration came up with another discriminant aiding continuum suppression: A Fisher discriminant combining nine variables, each quantifying the momentum flow around the thrust axis of the B candidate, binned in cones of 10° around the thrust axis. Just as before, instead of making use of the discriminant itself, this analysis employs the variables as input to a dedicated classifier. Their distributions, as entered in the classifier training, are displayed in Fig. B.3.

Flavor Tag Variables

B -flavor tagging refers to techniques aiming at deducing the quark flavor of the other B (also referred to as B_{tag}), i.e. the B meson not decaying through the signal channel. Flavor tagging algorithms operate on all final-state particles associated with the ROE and classify the ROE

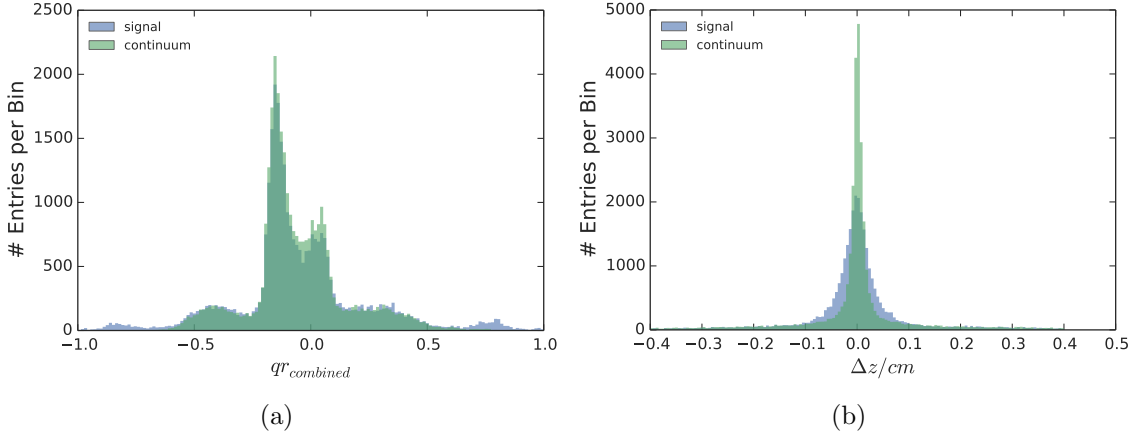


Figure 7.3.: Left: $q_{r_combined}$, the distribution of the flavor tag classifier output. The distribution is expected to be symmetric, why this is not the case here is unclear and beyond the scope of this work. Right: the distribution of the decay vertex distance Δz between the presumed signal B and B_{tag} .

as according to a test statistic called $q_{r_combined}$, ranging between the limits -1 and 1 , which correspond to either \bar{b} or b quark content of B_{tag} . Although flavor tagging is a concept crucial to analyses targeted at time-dependent CP -violation, $q_{r_combined}$ can be a good variable in the context of continuum suppression also. The reason for this is that the flavor tag classifier should be unable to decide what flavor to assign to the ROE particle set, constructed from a continuum event, as of course there never was any b quark content in the first place, which is reflected in $q_{r_combined}$ being confined to values closer to zero for continuum candidates.

Aside from $q_{r_combined}$, one can also use other variables typically used in time-dependent analyses, such as Δz , the distance between the vertices of the signal B and B_{tag} . Similarly, one expects separation arising from the fact that continuum events should be assigned vertex distances that tend to be smaller compared to those of true $B\bar{B}$ events. The two variables' distributions are shown in Fig. 7.3.

7.6.1. Multivariate Classifiers for Continuum Suppression

The binary classification of candidates to either stem from $q\bar{q}$ - or $B\bar{B}$ -events is a task well suited for multivariate classifiers. Two different approaches, artificial neural networks (ANNs) and boosted decision trees (BDTs) are considered and compared in terms of their receiver-operating characteristic (see Sec. 4.3 for a definition). Both classifier types are fed different sets of variables using 65000 coherent signal MC candidates and 300000 stream 1 continuum MC candidates. The training target is the MC truth of whether the event is a continuum or a $B\bar{B}$ -event. The samples are split in two, with equal numbers of candidates used for the classifier training and testing respectively. For the BDT 400 individual trees are assembled with the number of tree layers restricted to three. The ANN is set up using a single hidden layer of width $n + 1$, where n is the number of input features. Each feature undergoes sophisticated preprocessing that helps to handle outliers, smoothes statistical fluctuations and transforms the variables in a way that allows for a fast convergence to a near-optimal minimum of the loss-function.

Both, the ANN and the BDT, are trained using a basic feature set consisting of all KSF Moments, the Cleo-Cones and the thrust variables, therefore referred to as the 'basic' variable

set. Additionally, two further sets are considered: One consisting of the basic set plus Δz and one further adding $qr_{combined}$. The resulting different receiver-operating characteristics as evaluated on the test data set are shown in Fig. 7.4. As can be gathered from the areas under the curves (AUCs, given in the legend of the plot), the two classifiers show a strong out-of-the-box performance, i.e. neither of the two requires an extensive hyperparameter search to show an acceptable separation power. For each variable set the employed BDT outperforms the commercial ANN by roughly one percent in AUC. This finding should not be mistaken for a general statement, as more intensive hyperparameter tweaking holds the potential of changing the situation. As the achieved performance is satisfying, a potentially lengthy parameter search is omitted. The BDT using the basic variable set alongside Δz is chosen as the final classifier employed for the continuum suppression in this analysis. This results in an AUC of 0.939, a barely weaker performance as compared to the corresponding BDT with the additional flavor-tagging variable $qr_{combined}$, which is disfavored due to its insignificant improvement over the former variable set.

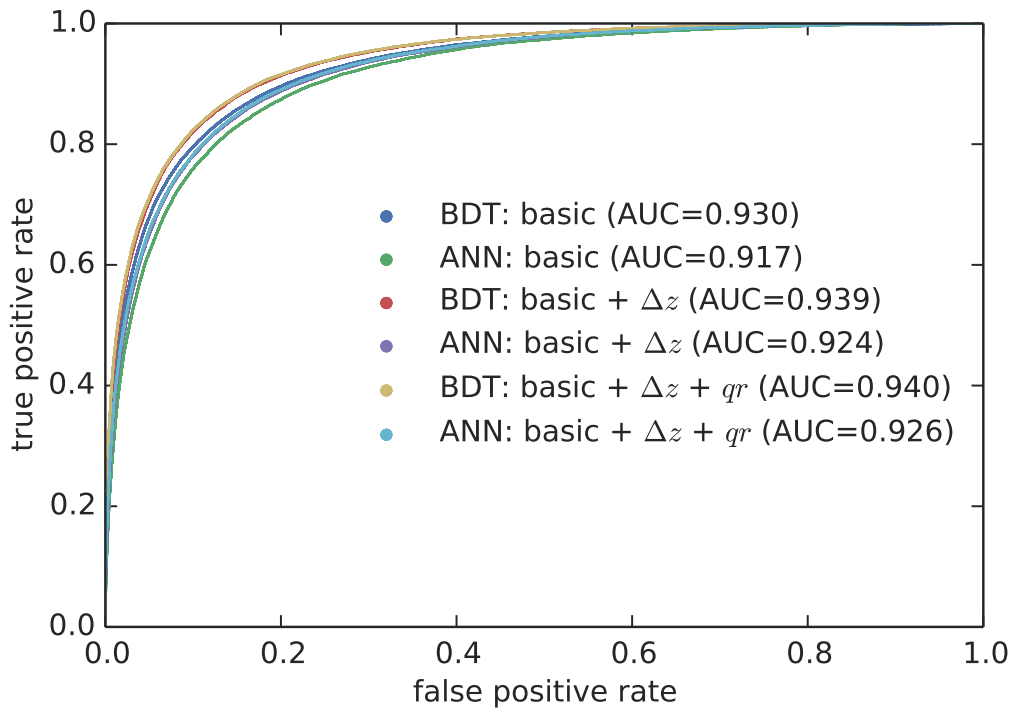


Figure 7.4.: The ROC curves of the boosted decision trees (BDTs) and artificial neural networks (ANNs) evaluated on the test data set, using different input variable sets. The corresponding areas under the shown curves (AUCs) are given in brackets.

The BDT's classification of all (charm, uds, charged and mixed) stream 1 MC backgrounds and the appropriate amount of coherent signal is depicted in Fig. 7.6g, where a good separation is clearly visible.

7.7. B -Meson Selection

Many particle physics analyses in the past and present, including the Dalitz analysis of the channel at hand ([28]), have employed orthogonal cuts on kinematic variables for the purpose of candidate selection. While there are certain merits to this straight-forward approach, e.g. simple interpretability of the selection requirements, other strategies are able to capture correlations much better. As this constitutes a significant advantage, we opt for a deviation from

the simple cut-based selection and instead, just as in the case of the continuum suppression, choose boosted-decision trees for the purpose of an offline B selection.

In principle all variables described previously and used for the loose preselection during the reconstruction can be used as an input to a multivariate classifier aiming to select true B candidates. However, as an aside, in order to avoid biasing, it can be important not to include specific variables in the selection, in case they are intended to be utilized for separate analysis specific procedures. Similar Dalitz analyses, such as [21] and [24], for instance had to extract the signal yield N_s (a quantity explained in more detail in Sec. 8.1) from a fit to the ΔE distribution. This was necessary in order to be able to constrain N_s in the Dalitz fit thereby guaranteeing the fit's stability. In this analysis however no variables are exempt and therefore the following set of variables is used as an input to the BDT: The beam-constrained mass (M_{bc}), the deviation from the expected B energy (ΔE), the reconstructed D mass ($M(D)$), the continuum suppression classification (\mathcal{O}_{CS}) as well as the p -value of the mass-constrained vertex fit of the B candidates ($p_{vertex}(B)$). The p -value of the mass-constrained vertex fit of the D candidates ($p_{vertex}(D)$) has also been considered and was found to bring about negligible gains and is therefore not utilized.

The BDT is trained on the Monte Carlo truth for whether the reconstructed B is a correct signal candidate or not. The training employs all experiments of stream 1 of charm, uds, mixed and charged Belle MC totaling around 480000 reconstructed background candidates as well as around 60000 reconstructed signal MC candidates. Again half of those candidates are retained to be used for testing. The optimal cut value on the BDT output, coined \mathcal{O}_{select} , is determined based upon the so-called figure of merit (FOM), which aims at optimizing the signal to background ratio:

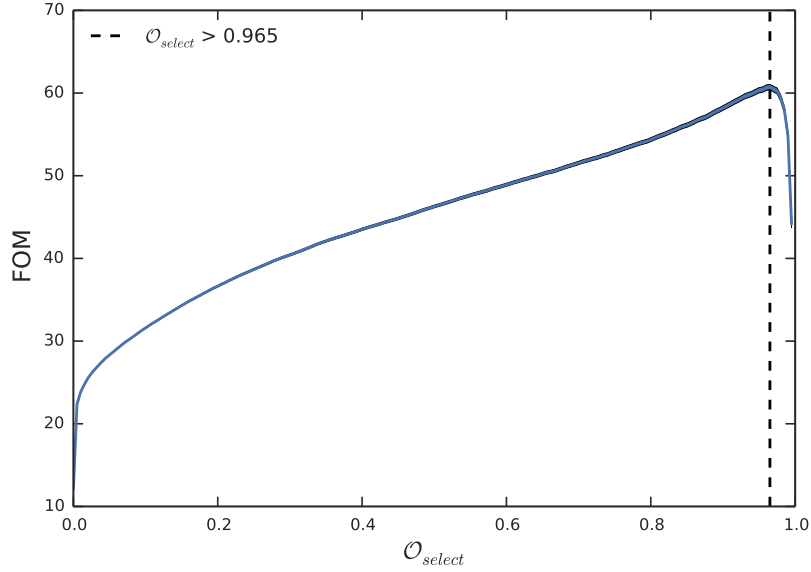
$$FOM = \frac{S}{\sqrt{S+B}}, \quad (7.6)$$

where S denotes the number of signal candidates retained after the application of the cut on \mathcal{O}_{select} and B refers to the respective number of background candidates. The optimal FOM is determined using an entire stream of signal deselected Belle MC backgrounds as found after the online selection, mixed with the corresponding amount of signal MC candidates. The FOM as a function of the classifier output \mathcal{O}_{select} is depicted in Fig. 7.5a, where its maximal, and thus optimal, value of 0.965 is marked by a dashed vertical line. The distributions of \mathcal{O}_{select} for individual components, i.e. for signal, for the combination of all backgrounds as well as for continuum background only, is illustrated in Fig. 7.5b. The expected amounts of individual components after the application of the optimal cut, so after finalizing the offline selection, as determined from stream 1 MC and the corresponding expected amount of signal candidates, are given in Tab. 7.2 below.

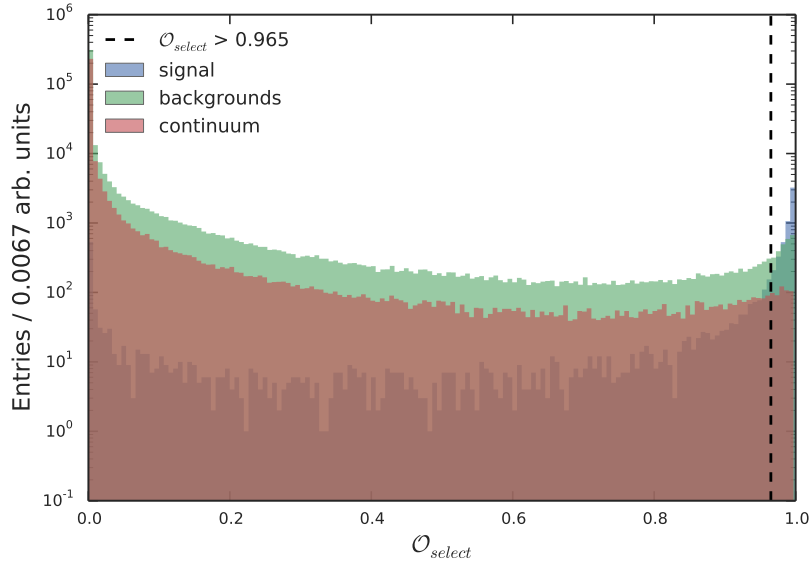
The distributions of the variables used in the classification for the B selection are shown in Fig. 7.6, each illustrated for both the signal component as well as for all backgrounds combined. The subfigures in the left depict the respective distributions as encountered after the online selection, whereas the subfigures on the right show the corresponding distributions as found after the finalized offline selection.

7.8. Best Candidate Selection

As the reconstruction of a B candidate boils down to combining daughter candidates in accordance with the above mentioned online and offline selection requirements, the possibility of finding multiple candidates for one true signal event is not ruled out or countered as of yet. Neglecting the small chance of both B 's of a B^+B^- -event decaying in the considered signal



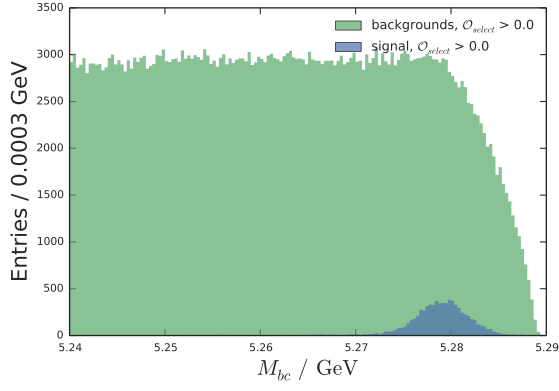
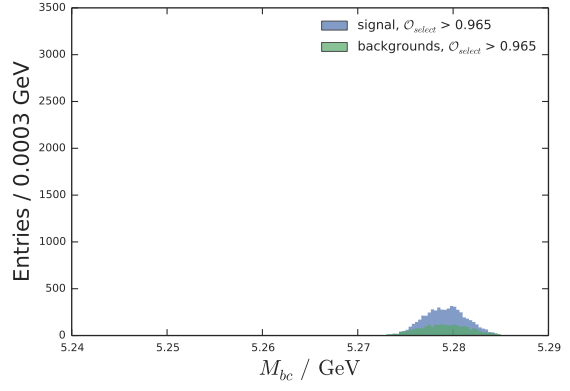
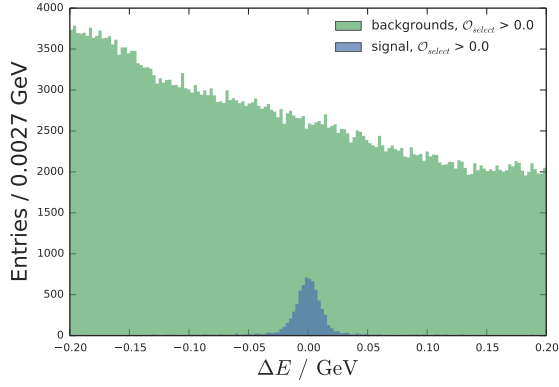
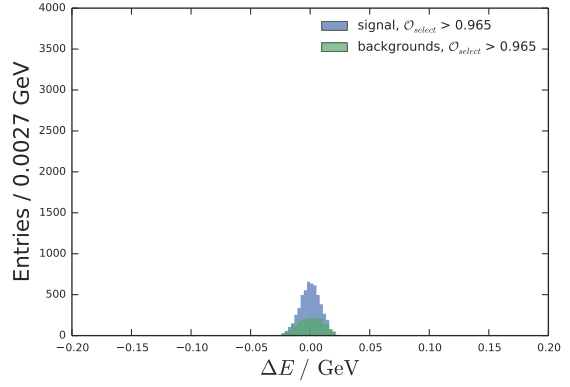
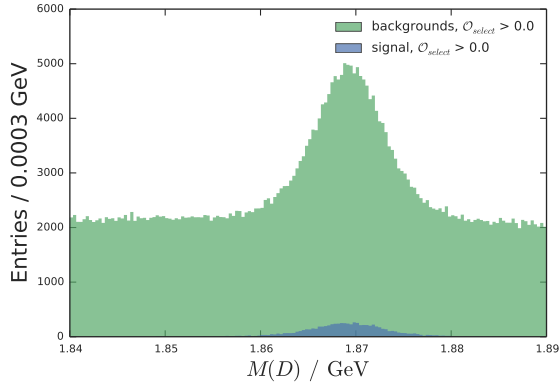
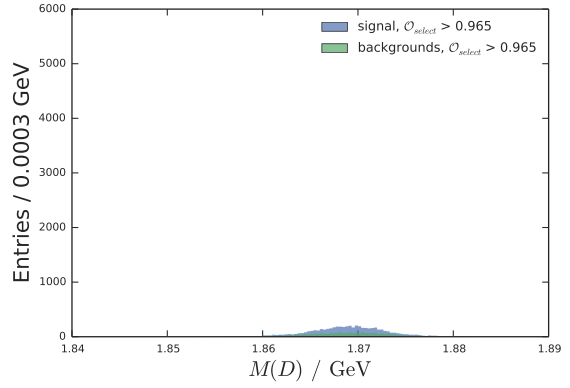
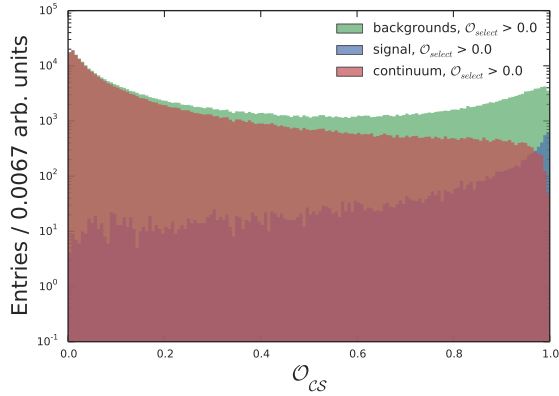
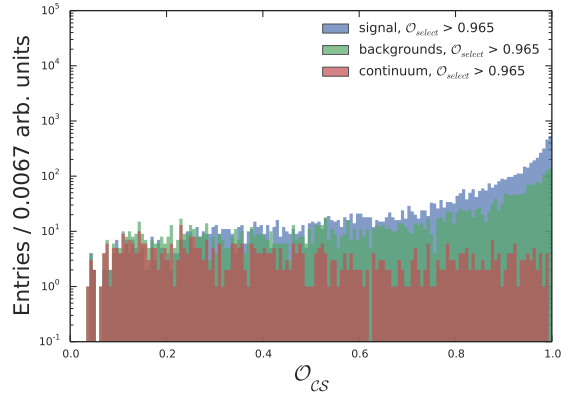
(a) The FOM as a function of the classifier output \mathcal{O}_{select} . The dashed line corresponds to the maximal, and thus optimal, FOM value of 0.965.



(b) The classifier output used for the B selection, shown using a logarithmic y-scale. The blue component illustrates signal candidates, the red component refers to the candidates reconstructed from continuum background only whereas the green component marks the candidates from the continuum and the generic backgrounds. The dashed line shows the optimized cut on \mathcal{O}_{select} .

Figure 7.5.: Multivariate classification of B candidates.

channel, there can only be one correct B -candidate. In principle several variables provide good separation to choose the correctly reconstructed candidate from a set of multiples reconstructed from one event. This analysis chooses $\Delta M_{bc} = |M_{bc} - M(B)_{PDG}|$, the absolute deviation of M_{bc} from the nominal B -mass, to discard misreconstructed multiples. The best candidate selection procedure is evaluated using a sample of approximately 65000 reconstructed signal MC candidates. The multiplicity, i.e. the average number of reconstructed candidates per correct candidate (after the reconstruction and selection described above), is found to amount to 1.02. In events with more than one candidate, the ΔM_{bc} best candidate selection selects the correct candidate in 75.6% of the cases.


 (a) M_{bc} after the online selection.

 (b) M_{bc} after the offline selection.

 (c) ΔE after the online selection.

 (d) ΔE after the offline selection.

 (e) $M(D)$ after the online selection.

 (f) $M(D)$ after the offline selection.

 (g) \mathcal{O}_{CS} after the online selection.

 (h) \mathcal{O}_{CS} after the offline selection.

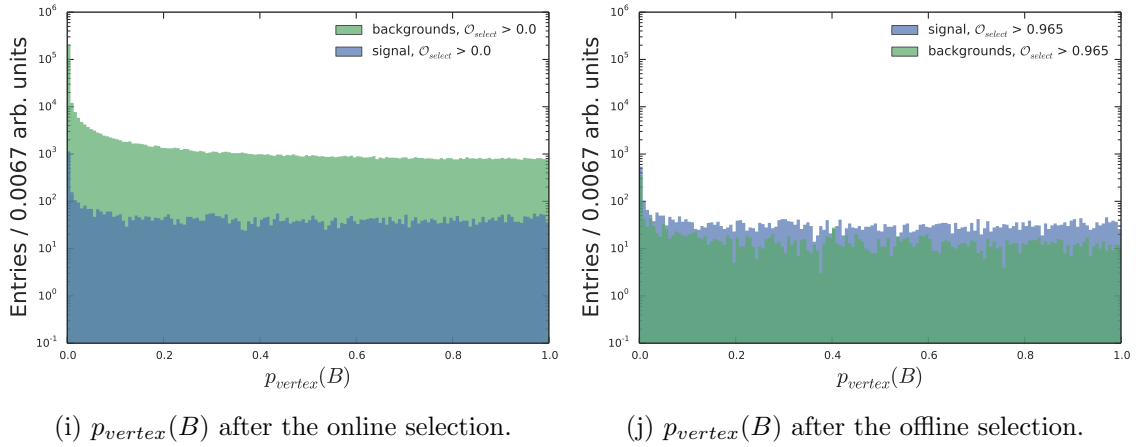


Figure 7.6.: Left: The ‘raw’ distributions of the variables used in the offline B selection (before cutting on \mathcal{O}_{select}). Right: The corresponding distributions after the finalized offline selection, i.e. after imposing the optimal cut of $\mathcal{O}_{select} = 0.965$. The figures show stream 1 Belle MC plus the corresponding number of signal MC candidates.

<i>component</i>	<i>no. candidates</i>
<i>signal</i>	5269
<i>charged</i>	433
<i>mixed</i>	1546
<i>continuum</i>	506

Table 7.2.: The number of B candidates after the finalized offline selection of individual components of stream 1 MC and the corresponding number of expected signal candidates.

7.9. Enforcing the Dalitz Plot boundaries

Even after completion of the best candidate selection and after performing the B -mass-constrained vertex fit, candidates from the EvtGenericDalitz model (see Sec. 6.3) as well as from all streams of the different types of background are reconstructed outside of the allowed Dalitz plot boundary as given in Eq. 3.2 (and as determined through Laura++, the employed fit framework, see Chap. 8). For these candidates reconstructed outside of the physical range, the likelihoods for signal or background respectively are zero. But since the employed fit method factors in the total number of events (see Eq. 8.3) they could potentially harm the result. For this reason only such candidates that satisfy the physical bounds are used for the different MC samples.

Candidates reconstructed outside of the physical bound have undergone migration due to detector, combinatoric or misreconstruction effects (see Chapter 9) and can be handled as mentioned above. More worrisome are such signal MC events that are generated outside of the allowed boundaries but migrate into the allowed bounds. No signal events should ever be generated outside of the DP boundary. But as this apparent violation of the physical bounds

by the generator does not occur very often, we don't view it as a general shortcoming and rectify the effect by neglecting candidates that are generated outside of the physical bounds. This is also accounted for in the efficiency determination, as is described in Chapter 10.

8. Dalitz Fit Model

8.1. Fit Model

The adopted fit model to the data sample that complies with all selection criteria is the joint likelihood density of the sample with an additional constraint on the fitted total number of events. This is referred to as an *unbinned extended maximum likelihood* (EML). The constraint is implemented in the form of an additional multiplicative term that assumes a Poissonian distribution of the total number of events. The individual likelihood \mathcal{L} for each candidate is comprised of two main components: The signal PDF S weighted with the corresponding efficiency ϵ as well as the sum of background PDF templates B_i . Both of these components are normalized across the Dalitz plot, as referred to by the integration area denoted by DP . The number of signal events N_s , also referred to as the signal yield, as well as the n_b different numbers of background events $N_{b,i}$ are left floating. The likelihood components for signal and background are then weighted as according to the ratio of the frequencies of the individual components so as to assure that the global probability of finding each candidate in the Dalitz plot equals unity. The likelihood \mathcal{L} for each candidate, as given by its Dalitz plot coordinates ($x \equiv m(D\pi)_{max}^2, y \equiv m(D\pi)_{min}^2$), is a function of the numbers of the individual components $N = \{N_s, N_{b,1}, \dots, N_{b,n_b}\}$ and the free parameters θ of the coherent signal model S , which is introduced in Eq. 3.6. The parameters θ that can be left floating are the masses and widths of the individual resonance amplitudes \mathcal{A}_i as well as the relative phases $\Delta\alpha_i$.

$$\begin{aligned} \mathcal{L}(x, y|\theta, N) = & \frac{N_s}{N_s + \sum_i^{n_b} N_{b,i}} \frac{S(x, y|\theta)\epsilon(x, y)}{\int_{DP} S(x', y'|\theta)\epsilon(x', y')dx'dy'} \\ & + \sum_j^{n_b} \frac{N_{b,j}}{N_s + \sum_i^{n_b} N_{b,i}} \frac{B_j(x, y)}{\int_{DP} B_j(x', y')dx'dy'}, \end{aligned} \quad (8.1)$$

$$S(x, y|\theta) = |\mathcal{A}_{NonRes} + e^{i\Delta\alpha_1}\mathcal{A}_{D^{*0}} + e^{i\Delta\alpha_2}\mathcal{A}_{D_2^{*0}}|^2, \quad (8.2)$$

$$\mathcal{L}'(data|\theta, N) = \frac{(N_s + \sum_i^{n_b} N_{b,i})^{N_{tot}}}{N_{tot}!} e^{-(N_s + \sum_i^{n_b} N_{b,i})} \prod_i^{N_{tot}} \mathcal{L}(x_i, y_i|\theta, N). \quad (8.3)$$

The total likelihood \mathcal{L}' simplifies when taking the negative logarithm:

$$\begin{aligned}
-\log \mathcal{L}' &\sim -N_{tot} \cdot \log \left(N_s + \sum_j^{n_b} N_{b,j} \right) + \left(N_s + \sum_j^{n_b} N_{b,j} \right) - \sum_i^{N_{tot}} \log \left(\left(N_s + \sum_j^{n_b} N_{b,j} \right) \cdot \mathcal{L} \right) \\
&\quad + N_{tot} \cdot \log \left(N_s + \sum_j^{n_b} N_{b,j} \right) \\
&= - \sum_i^{N_{tot}} \log \left(N_s \frac{S(x, y | \theta) \epsilon(x, y)}{\int_{DP} S(x', y' | \theta) \epsilon(x', y') dx' dy'} + \sum_i^{n_b} N_{b,i} \frac{B_i(x, y)}{\int_{DP} B_i(x', y') dx' dy'} \right) \\
&\quad + \left(N_s + \sum_j^{n_b} N_{b,j} \right),
\end{aligned}$$

where constant, and thus redundant, terms are dropped for brevity (as indicated by the ' \sim ' symbol). The form above shows the negative log-likelihood term as implemented and minimized in the fit framework introduced below.

The signal model is internally parametrized in terms of real and imaginary coefficients rather than in the amplitude and phase convention as given in Eq. 8.2, which was found to yield more stable fit results by earlier works [21].

Note that the distributions and even the contributing channels, and thus the signal PDF S is *a priori* unknown in Dalitz analyses, such that often different models involving more exotic contributions are tried competitively on data (e.g. see [16]).

8.2. Fit Framework

The Dalitz fits in this work, i.e. the minimizations of the negative log-likelihood as given above, are performed using the software package Laura++ developed by Thomas Latham *et al* [30]. Laura++ is a framework dedicated to Dalitz plot analyses and has been employed in many recent works such as [5] by the BaBar collaboration or [1] by the LHCb collaboration. The framework makes use of classes from the data analysis package ROOT [10], e.g. the fitting routine Minuit [23]. Some of its main features, of importance to this analysis, are outlined below:

- Laura++ is an extended maximum likelihood fitter
- Laura++ can form an isobar model for a scalar parent particle decaying to three scalar daughters
- both amplitude/phase and real/imaginary coefficients are supported
- different background templates can be fed in as 2D-histograms
- varying reconstruction efficiencies can be handled by providing 2D-histograms
- both the regular and the squared Dalitz Plot representation are supported
- Laura++ enables the generation of data samples allowing for ensemble tests
- although Laura++ does not provide a direct handle to the fitted *PDF* shapes, samples (also referred to as toys) can be drawn from the fit result and directly compared to the input data

9. Background Modelling

Even after a careful signal selection procedure, background candidates from a range of different sources fulfill the installed requirements. The sources relevant to this analysis comprise both continuum and generic background. The processes giving rise to continuum background do not proceed via the production of a B meson and are introduced in Section 7.6. Generic background on the other hand does stem from B decays, proceeding, however, in channels different to the considered signal channel, which thus result in different final-state particles. Generic background can mimic signal by randomly fulfilling the selection criteria and due to the ensuing false recombination of particles it is coined *combinatoric* background. A typical example of a potential error of this type is the misidentification of a kaon for a pion or vice versa.

Both continuum and combinatoric background are studied using five streams of Belle MC. For the entire Belle dataset of around 772×10^6 $B\bar{B}$ events we expect around 400 combinatoric background candidates from charged $B\bar{B}$ events, 1500 combinatoric background candidates from mixed $B\bar{B}$ events and around 500 continuum candidates from $q\bar{q}$ events. Those candidates are accounted for in the fit by dedicating different template probability densities B_i to the distinct background sources (see Eq. 8.1).

The template probability densities are directly deduced from the reconstructed candidates of the respective backgrounds and fed into the fit in form of a two dimensional histogram. In order to even out statistical fluctuations in the reconstructed distributions, a kernel density estimation (*KDE*) is performed before the bins of the respective histogram are filled. The technique assigns each reconstructed point in the Dalitz plot a Gaussian kernel as is introduced in Sec. 4.2. The regular Dalitz plot representation with its curved kinematic boundary is not ideal for this task given that it is desirable to prevent bins from crossing the physical boundary. Because straight boundaries facilitate the task of keeping the assigned kernel densities within bounds, the square Dalitz plot representation, parametrized in terms of the transformed kinematic variables θ' and m' , (see Sec. 3.2.1) is employed.

As the KDE serves to smooth out fluctuations, the kernel width needs to take on a reasonable large value. But because the different backgrounds all exhibit a more or less pronounced drop-off in population towards the edges of the allowed range in the θ' -projection, large kernel widths result in 'leakage' outside of the bounds which is accompanied by an overestimation of entries close to the edge. In order to achieve both aims, a reasonable smooth background template

and its adherence to kinematic bounds, a hybrid model, merging the histograms resulting from two different choices of kernel widths, is adopted. This approach follows the work of Manuel Heider *et al.* ([21]).

The first histogram, denoted K_{high} , uses a kernel width of 1.2 in the m' direction and a kernel width of 0.85 in the θ' direction. The second histogram, denoted K_{low} , also employs a kernel width of 1.2 in the m' direction, but uses a significantly lower width of 0.2 in the θ' direction. The kernel widths employed and stated in the above are found to work well for all background types. The two histograms K_{high} and K_{low} , containing the results of the KDE smoothed probability densities, are constructed using a binning of 200×200 and are then combined into a single histogram K_{merged} . This resulting histogram is formed using purely K_{high} entries in its central region but is made up of a linear merger of K_{low} and K_{high} towards the boundaries of θ' .

The merger was performed through a weighted sum of the respective i^{th} histogram entries $K_{low,i}$ and $K_{high,i}$. The regions in which the merger is performed are those for which $\theta' \in [0.0, 0.04]$ and $\theta' \in [0.46, 0.5]$, respectively. The final histogram bins $K_{merged,i}$ then come out to be:

$$K_{merged,i} = \begin{cases} \frac{\theta'}{0.04} \cdot K_{high,i} + \frac{0.04 - \theta'}{0.04} \cdot K_{low,i}, & \text{if } \theta' \in [0.0, 0.04], \\ K_{high,i}, & \text{if } \theta' \in (0.04, 0.46), \\ \frac{0.5 - \theta'}{0.04} \cdot K_{high,i} + \frac{\theta' - 0.46}{0.04} \cdot K_{low,i}, & \text{if } \theta' \in [0.46, 0.5]. \end{cases}$$

Bins beyond the kinematic boundary ($\theta' > 0.5$) that still receive contributions are set to contain zero entries.

9.1. Continuum Background

As described above, the high kernel width of 0.85 fails to describe the underlying probability density towards the edges of the θ' projection, which can be confirmed by a consultation of Fig. 9.1a. This is successfully corrected for by combining the former KDE result with the one obtained using a lower kernel width, as depicted in red in Fig. 9.1a. The projection onto m' reveals essentially congruent line shapes for the unmerged and the merged KDE results with a slight deviation of the high and the low kernel width results apparent at the most frequent m' values (this can be gathered from Fig. 9.1b). This artifact can likely be attributed to the fact that bin entries in regions with $\theta' < 0.0$ and $\theta' > 0.5$ are clipped off, thus slightly underestimating the high kernel width contribution. Figure 9.2a illustrates the point of departure, i.e. the bare candidates from five streams of Belle MC, while Figure 9.2a shows the final continuum probability density as employed in the Dalitz fit.

9.2. Combinatoric Background

Combinatoric background stems from a false recombination of particles from B decays, which can be studied and modeled using charged and mixed Belle MC. The two categories are modeled in terms of two different background templates. This aims at allowing for a potential relative difference in the frequencies of charged and mixed combinatoric background as found in the generic MC samples and in recorded data.

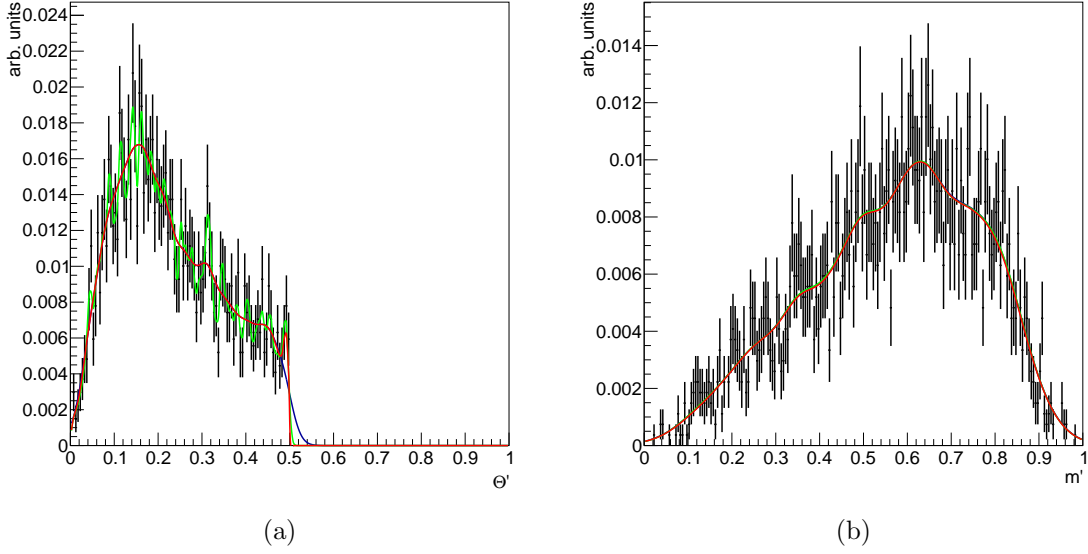


Figure 9.1.: The result of the KDE smoothing of five streams of Belle **continuum** MC in the two projections to θ' and m' . The binned candidates are depicted in black, also shown are the corresponding Poissonian errors. The green line shape shows the result of the high kernel width KDE, the blue line shape depicts the result of the low kernel width KDE result and the red line shape represents the merged results of the two.

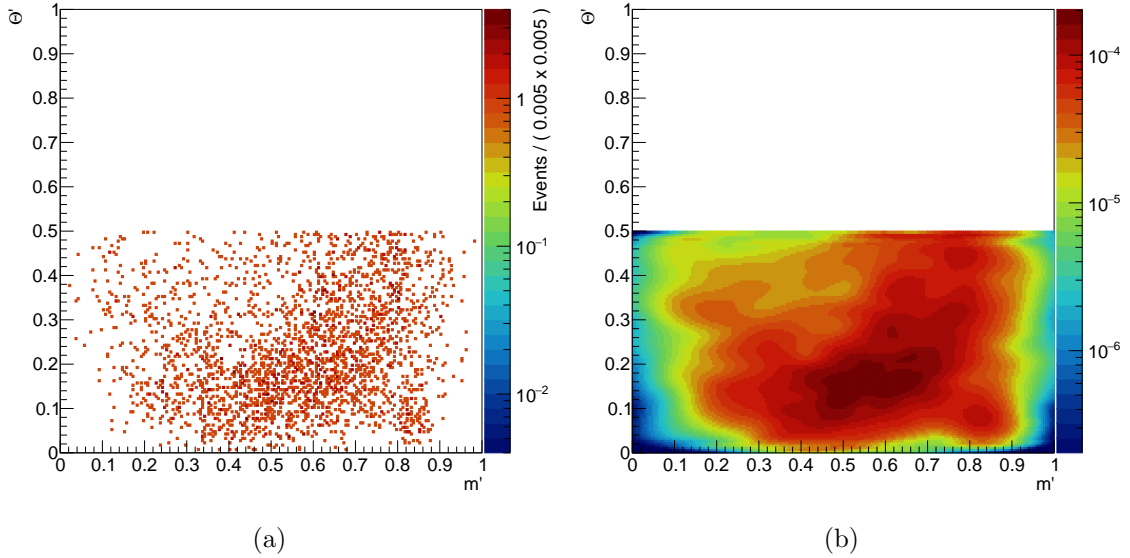


Figure 9.2.: Left: The candidates of five streams of Belle **continuum** MC are depicted in the square DP representation. Right: The corresponding continuum background template for the Dalitz plot after merging the two distinct KDE results is shown.

Again, the charged background component natively contains signal candidates. Using generator information, often referred to as the *MC Truth*, those signal candidates are deselected in the signal box of the utilized MC samples.

Figures 9.3 and 9.4 and Figures 9.5 and 9.6 show the individual KDE results in the respective projections to θ' and m' as well as the final background templates for both charged and mixed combinatoric background. Again five streams of Belle MC were used in the process.

9.3. Misreconstructed Signal

Another background source often important in Dalitz Plot analyses is *misreconstructed signal*. This refers to an interchange of true signal particles during the course of the reconstruction, resulting in a false 'placement' of the particles in the signal decay chain or an only partially correct reconstruction of the signal decay where one or more particle from the other B are falsely combined with true signal particles. To be more graphic, this could occur in the present analysis, when a daughter pion of the D meson would be mistaken for a direct daughter pion of the B meson; or when any kinematically fitting daughter pion of the other B would be mistaken for a daughter pion of the signal B meson. Both error types would yield a potentially large migration of the candidate away from its true position in the Dalitz plot.

However, in contrast to other Dalitz analyses, e.g. such involving neutral pions [21], this analysis does not find a significant component of misreconstructed signal: Determined from the reconstruction of a sample of 50×10^6 generated non-resonant signal MC events, only about 0.4% of the reconstructed candidates are found to be misreconstructed signal. The background due to misreconstructed signal can thus be safely neglected, which is in line with Kuzmin et al's analysis of the signal decay at hand [3]. This component is further discussed in the context of detector effect, where Fig. 10.1 also depicts the misreconstructed signal component.

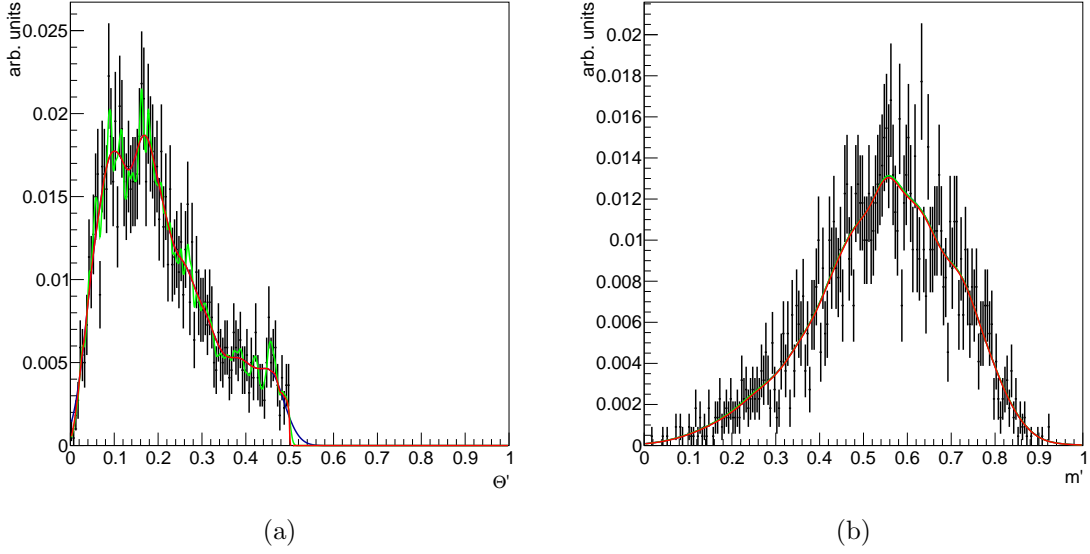


Figure 9.3.: The result of the KDE smoothing of five streams of Belle **charged** MC in the two projections to θ' and m' . The binned candidates are depicted in black, also shown are the corresponding Poissonian errors. The green line shape shows the result of the high kernel width KDE, the blue line shape depicts the result of the low kernel width KDE result and the red line shape represents the merged results of the two.

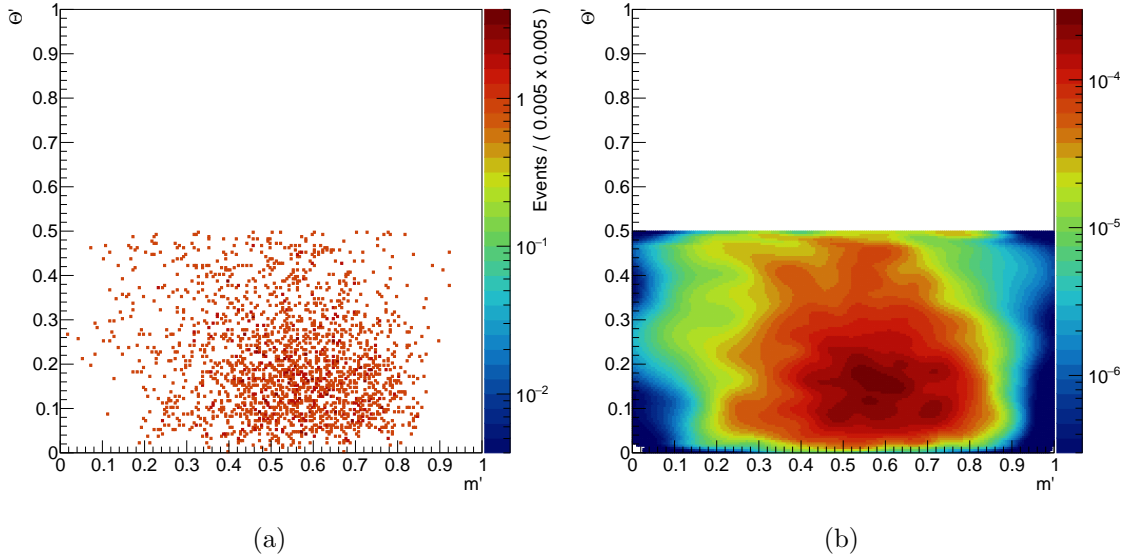


Figure 9.4.: Left: The candidates of five streams of Belle **charged** MC are depicted in the square DP representation. Right: The corresponding continuum background template for the Dalitz plot after merging the two distinct KDE results is shown.

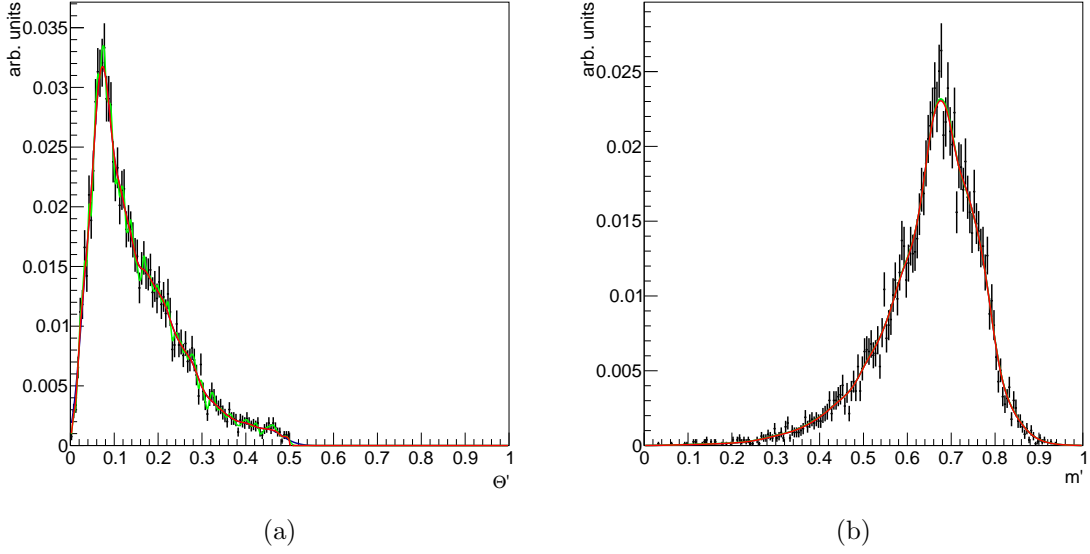


Figure 9.5.: The result of the KDE smoothing of five streams of Belle **mixed** MC in the two projections to θ' and m' . The binned candidates are depicted in black, also shown are the corresponding Poissonian errors. The green line shape shows the result of the high kernel width KDE, the blue line shape depicts the result of the low kernel width KDE result and the red line shape represents the merged results of the two.

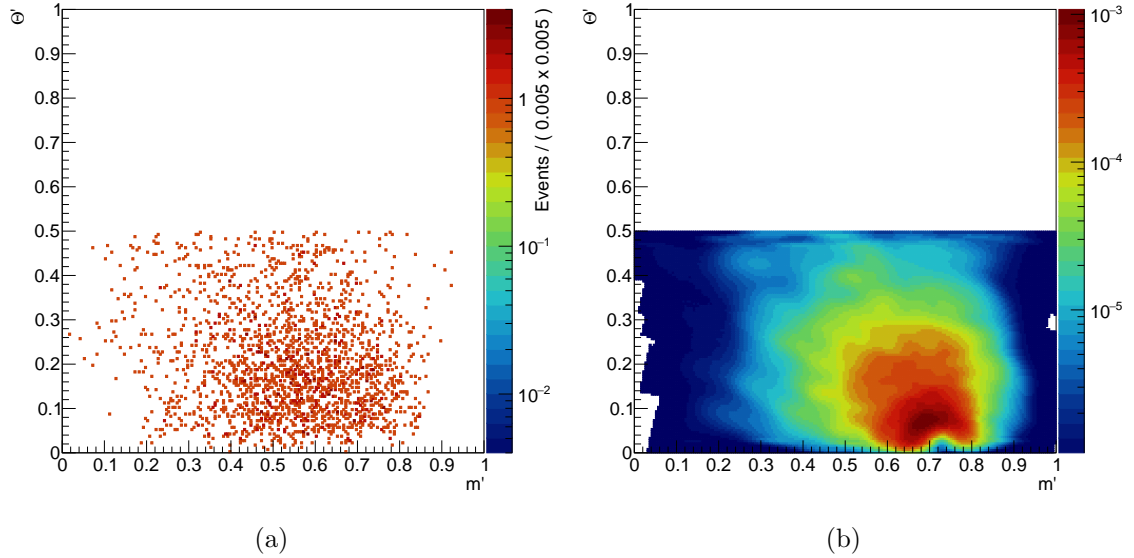


Figure 9.6.: Left: The candidates of five streams of Belle **mixed** MC are depicted in the square DP representation. Right: The corresponding continuum background template for the Dalitz plot after merging the two distinct KDE results is shown.

10. Detector Response

The detector acceptance can vary across the Dalitz plot, which is due to the fact that the detector's behavior differs for distinct momenta of the same particles. Therefore the detector response needs to be modeled and accounted for, such that the correct number of candidates can be uncovered. In particle physics analyses, the detector response is often factored in by 'unfolding' the data. The approach taken here can be viewed as a simple form of applying the reverse operation to the signal model (we 'fold' the signal model, if you will). More specifically, this means performing the correction by parametrizing the detector acceptance in the form of a two dimensional histogram ϵ , spanned by the regular Dalitz plane representation, according to which we modify the signal likelihood for each candidate (as is introduced in Eq. 8.1).

10.1. Signal Resolution and Migration

Before detailing the handling of the detector's acceptance, an additional effect is considered in the following. Namely, the migration of candidates in the Dalitz plot away from their true position. This shift of candidates in the Dalitz plane can be caused by the misreconstruction of signal, an artifact introduced in Sec. 9.3. In case of correctly reconstructed signal, this shift corresponds to the detector resolution, which can be quantified in terms of the absolute migration distance $d_{migration}$, defined as the difference between the candidates' generated position and its reconstructed position (denoted by the subscript *gen* and *rec*, respectively):

$$d_{migration} = \sqrt{\left(m(D\pi)_{max,gen} - m(D\pi)_{max,rec}\right)^2 + \left(m(D\pi)_{min,gen} - m(D\pi)_{min,rec}\right)^2}. \quad (10.1)$$

On a sample of 50×10^6 phasespace signal MC events, the distribution of the migration distance of all reconstructed events is depicted in Fig. 10.1. The 1- σ -percentile of correctly reconstructed signal is found at around 3.9 MeV, a value that can be regarded as insignificant compared to the lowest resonance' width, the width of the D_2^* , which is expected to take on a value at around $\Gamma(D_2^*) = 49 \text{ MeV}$ (and is further depicted as a dashed line in Fig. 10.1). The 1- σ -percentile estimate for the signal resolution seems in line with the findings of the previous Belle Dalitz plot analysis of the channel at hand. As described in the corresponding collaboration internal note ([28], referred to as 'Belle Note'), A. Kuzmin et al. found a resolution of 4 MeV. This value was deemed negligible in the sense that the results of fits, when convolving the signal model with the resolution function and when leaving the signal model

untouched, yield insignificant differences. This analysis therefore refrains from convolving the signal model (Eq. 8.2) with a resolution function, which is also legitimized by corresponding statements in the *Physics of the B-Factories* [8]. As can be understood from Fig. 10.1, the migration distance does not seem to peak at 0.0, the reason for which remains unclear.

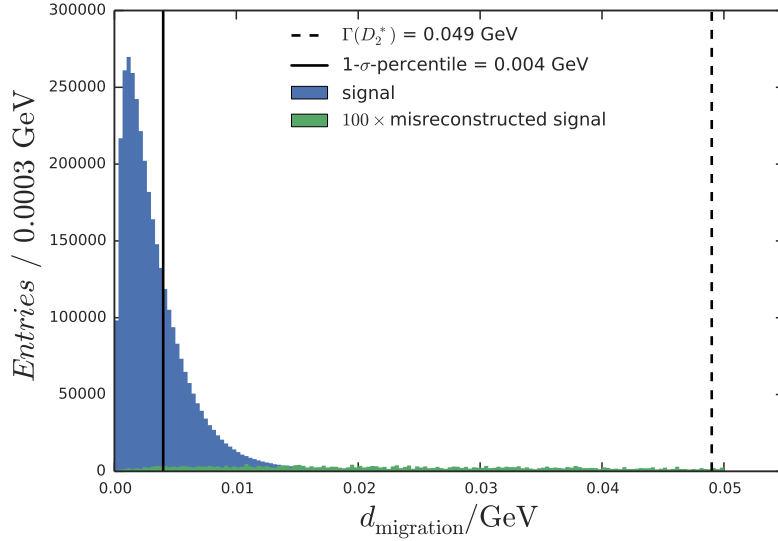


Figure 10.1.: The migration distance $d_{\text{migration}}$ from a sample of 50×10^6 phasespace signal MC events. Correctly reconstructed signal events are shown in blue, whereas misreconstructed candidates, enhanced by a factor of 100, are shown in green. The solid black line shows the $1\text{-}\sigma$ -percentile of correctly reconstructed signal and for comparison the dashed black line denotes the width of the most narrow resonance, the D_2^* , that is contributing.

As already mentioned in Sec. 9.3, many Dalitz plot analyses need to handle significant self-crossfeed due to misreconstructed signal. This behavior is especially prevalent in decays containing neutral pions. The present analysis however seems to find itself in the comfortable position of only dealing with the occurrence of very little misreconstructed signal. On the phasespace MC sample considered in this context, only about 0.4% of all candidates are misreconstructed. The migration distance of the latter, enhanced by a factor of one hundred, is depicted in Fig. 10.1 and found to be uniformly distributed. Because of the apparent insignificance of misreconstructed signal, there is no need for elaborate but systematics-prone correction mechanisms, as, for example, [21] had to employ.

10.2. Efficiency

The detector's efficiency ϵ is defined as the probability of a given event to pass the selection as a function of the event's true position in the Dalitz plane which is determined using a large sample of phasespace signal Monte Carlo. This approach is common ([31]) as the latter is uniformly distributed in the Dalitz plane. For the task of producing the employed 50×10^6 signal events, we use *Evtgen* and then apply *gsim*, a software package based on *Geant4* that models the detector effects, as is described in more detail in Chap. 6. The events are then reconstructed and subjected to the full set of selection requirements after which the efficiency can essentially be determined by calculating the bin wise ratio of reconstructed signal candidates and generated candidates.

There is however a range of subtleties to attend to. For one, due to the migration effects described above, it is obvious that not all candidates are reconstructed in their bin of origin. Yet the efficiency in a given bin i should describe the probability of a candidate, generated in bin i , to be reconstructed at all, regardless of which bin j it really ends up being reconstructed in. For this reason N_i^{reco} , the number of reconstructed candidates originating in bin i , is determined irrespective to its reconstructed position j . Therefore the efficiency histogram results from dividing the latter by N_i^{gen} , the number of candidates generated in bin i :

$$\epsilon_i(m(D\pi)_{max}^2, m(D\pi)_{max}^2) = \frac{N_i^{reco}}{N_i^{gen}} = \frac{\sum_j N_{ij}^{reco}}{N_i^{gen}}. \quad (10.2)$$

As hinted upon earlier, the regular Dalitz plot representation is chosen to determine the efficiency histogram. This is due to the fact that while the phasespace signal candidates are distributed uniformly in the regular representation, the transformation into the square representation renders them highly non-uniformly distributed. As shown by others (see [21]), this leads to relatively large regions located towards the Dalitz plot edges that contain only very little candidates, thus resulting in quasi-zero efficiency bins.

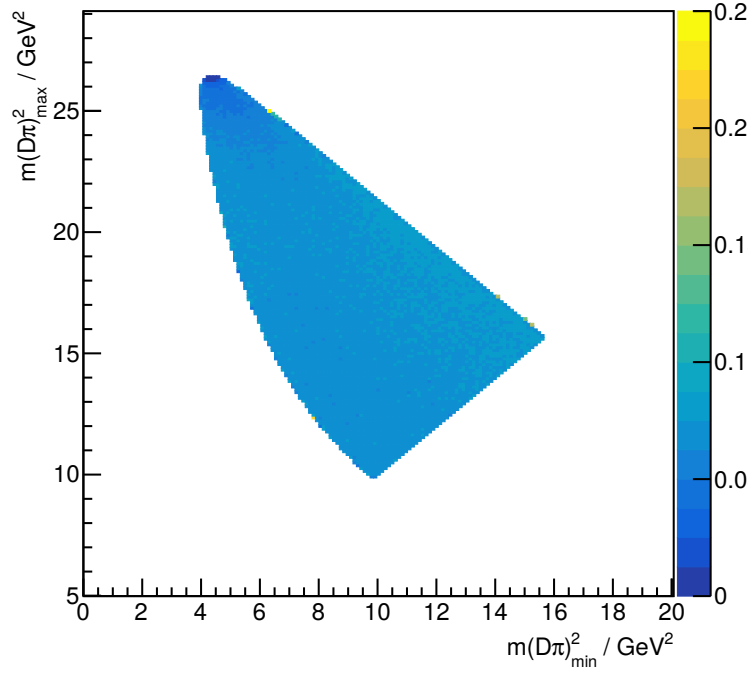
The regular Dalitz plot representation with its mostly flat distribution for non-resonant signal is therefore more suitable, but presents other challenges: The curved kinematic boundary crosses bins, leaving some of the N_i^{gen} , especially at the DP boundary, underpopulated. In the efficiency this can present itself in the form of quasi-empty bins or very large bin populations, depending on how the effect interferes with N_i^{gen} in relation to N_i^{reco} . This of course is an artificial effect due only to the binning, which thus needs to be corrected for. One obvious way out, that is also adopted here, is averaging over neighboring bins. To do so, a threshold for a bin to be considered under-populated needs to be defined. In this regard we consider such bins low, that exhibit a bin content that is 20 Poissonian standard-deviations below the average of all generated entries N^{gen} . If a bin i falls short of the threshold, its contents N_i^{reco} and N_i^{gen} are updated with the average of the respective values of all its direct neighbors j , that rank above the threshold (coined ‘good’ bins). The schematic for this modified efficiency is then given as

$$\epsilon'_i = \frac{N_i^{reco} + \sum_j N_{j, good}^{reco}}{N_j^{gen} + \sum_j N_{j, good}^{gen}}. \quad (10.3)$$

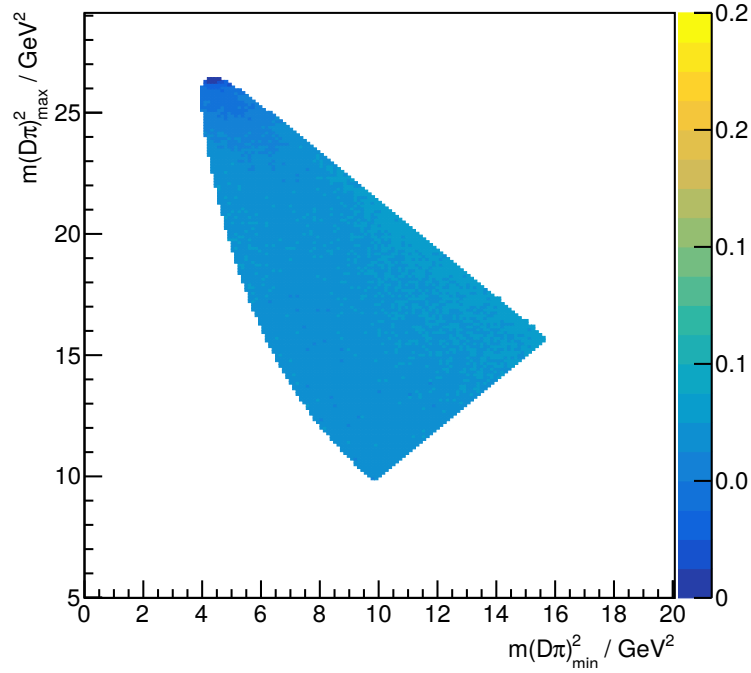
The efficiency before the averaging over unaffected neighbor bins is shown in Fig. 10.2a. Here both under- and overpopulated bins located at the Dalitz plot boundary are clearly visible. Fig. 10.2b on the other hand depicts the modified efficiency (ϵ' , as in Eq. 10.3), where the statistical fluctuations due to the binning are highly (if not fully) reduced. A binning of 250×250 corresponding to individual bin sizes of about $0.114 \text{ GeV}^2 \times 0.114 \text{ GeV}^2$ is found to work well.

There is one more thing to pay attention to and that is the fact that Evtgen's EvtGenericDalitz model wrongfully generates events outside of the kinematic boundary (see Sec. 7.9). Although the portion of reconstructed candidates that were generated out of bounds is low ($\mathcal{O}(1\%)$ evaluated on the sample of 50×10^6 phasespace signal MC), it still needs to be accounted for. For this purpose, before calculating the efficiency histogram, all candidates generated outside of the Dalitz plot bounds and reconstructed within are taken out of the sample used in the determination of N^{reco} and so are all those candidates generally generated outside of the limits used in the determination of N^{gen} .

The overall average of the efficiency calculated in above fashion and utilized to determine the expected signal yield (see Eq. 1.3), comes out to be $\langle \epsilon \rangle = 0.0699$.



(a) The efficiency histogram ϵ before the averaging scheme, where the occurrence of under- and overpopulated bins is visible particularly at the upper tip and the left and right boundary of the depicted Dalitz plane (bins colored dark blue and yellow/pale green, respectively).



(b) The efficiency histogram ϵ' , i.e. after the averaging scheme, with a reduction of statistical fluctuations due to the binning.

Figure 10.2.: The construction of the efficiency histogram, which encodes the detector response for the signal component.

11. Fit Validation using individual Resonances

In order to test whether the constructed efficiency histogram (see Chap. 10) is able to capture the detector effects well, fits to truth-matched signal MC samples which have been generated using a full detector simulation have to be performed. For this task, as described before, events are generated with Evtgen and then passed on to Geant4, which models the detector effects. As is argued in Sec. 6.3, the coherent Dalitz signal model as implemented in Evtgen is unfortunately unfit for this exercise, as the resonances' relative contributions, as generated, are not known. Individual resonances however are unaffected by this shortcoming and are thus utilized to probe the modelling of detector effects. Signal MC samples of 5×10^5 events are generated for both the D_0^{*0} and the D_2^{*0} component using the parameters given in Tab. 11.1 below. Note that the meson radii, R_B and R_D , used in the Blatt-Weisskopf barrier factors (see Sec. 3.1.3) for the D_0^{*0} component are irrelevant as the Blatt-Weisskopf barrier factors are constant in that case. Attention should be drawn to R_B , the meson radius of the B meson, used for the D_2^{*0} component: It is set to a nonphysical value of 0.0 GeV^{-1} (the natural choice would be for R_B to match R_D). This setting is necessary as we found the Evtgen (v12r0) model and the Laura++ (v3r0) model for D_2^{*0} incompatible when assigning non-zero values to R_B (see Fig. 11.3 below), resulting in poor fit results due to this non-conformity at the generator level. Besides a broader Breit-Wigner shape, Laura++' amplitude also nonphysically rises towards the upper boundary of the Dalitz plot (see Fig. 11.3a), giving rise to the suspicion that the amplitude is mismodelled in the ML fitter rather than in the MC generator. However, for the purpose of evaluating the efficiency treatment, resorting to a nonphysical barrier radius value of zero is admissible as it does not alter the test's validity.

In the Dalitz fits to the two resonant components the same parameters as employed for the MC generation are chosen and fixed. The amplitude of the resonance addressed in the fit is fixed at 1.0 and its phase at 0.0, while the amplitude and phase of the non-resonant component and the resonance not-present in the MC sample are left floating. The fits make use of the extended maximum likelihood scheme, hence the signal yield is left floating. The performance of the fits is evaluated in terms of the agreement between the fitted and the input signal shapes as projected into the variables $m(D\pi)_{min}$ and $m(D\pi)_{max}$. The projections of the Dalitz fits to D_0^{*0} and D_2^{*0} are shown in Figs. C.1 and 11.2. As mentioned in Sec. 8.2, Laura++ does not allow for an extraction of a parametrized fitted distribution, instead samples from the fitted distribution can be drawn and compared to the input distribution, both of which are shown in the respective projections. The reconstructed signal MC samples both contain small fractions

<i>resonance</i>	<i>mass (GeV)</i>	<i>width (MeV)</i>	T_r	R_B (GeV ⁻¹)	R_D (GeV ⁻¹)
$D_0^{*0}(2400)$	2.3180	267	Rel. B.-W.	-	-
$D_2^{*0}(2462)$	2.4626	49	Rel. B.-W.	0.0	1.6

Table 11.1.: Parameters used for the generation of individual resonance MC samples with Evtgen’s EvtGenericDalitz model. In both cases relativistic Breit-Wigner functions are employed for the line shape T_r (see Sec. 3.1.3). R_B denotes the meson radius for the B and R_D refers to the meson-radius of the resonant D as utilized in the respective Blatt-Weisskopf barrier factors. The choice of $R_D = 1.6$ GeV⁻¹ follows previous analyses (also see Sec. 3.1.3).

of misreconstructed signal (see Sec. 10.1), also illustrated in the plots. The deviation of the fit result from the input (quantified in terms of Poissonian standard deviations) is illustrated in the form of a pull distribution that is shown below the respective fit projections. In this case the pull is defined for every bin i as

$$pull_i = \frac{N_{i,fitted} - N_{i,generated}}{\sqrt{N_{i,fitted}}},$$

where $N_{i,fitted}$ refers to the number of fitted candidates and $N_{i,generated}$ denotes the number of input candidates in bin i .

The floated signal components’ amplitudes and phases are found compatible with zero in both fits. Due to its apparent insignificance, the misreconstructed signal component is not addressed with a dedicated background template or more advanced approaches (see Sec. 10.1). This is justified by the success of the fits without such correction mechanisms, which is elaborated upon below.

Commonly the reduced χ^2 , or its p-value, p_{χ^2} , for that matter, is utilized to conduct a statistical test for whether a sample is consistent with the shape of a distribution that it is considered to be drawn from. In this case however, as the line shape of the fitted distribution is unavailable, a χ^2 -test is not the appropriate choice. Nonetheless, for the sake of completeness the respective reduced χ^2 and its p-value are given in boxes overlaying Figs. C.1 and 11.2 and as expected the reduced χ^2 comes out high (a successful test for conformity should exhibit a reduced χ^2 value of around unity), also reflected in the comparatively large deviations from zero in the pull distributions. In lieu of the χ^2 -test we conduct a two-sided two-sample Kolmogorov-Smirnov-test (KS-test), which, from the difference between the cumulative distributions of the samples, constructs a p-value, p_{KS} , to test for the null hypothesis that both samples come from a population with the same distribution.

The p-values p_{KS} are given in the respective boxes overlaying Figs. C.1 and 11.2. In all cases the null hypothesis cannot be rejected (with the lowest value of $p_{KS} = 0.044$ obtained in the $m(D\pi)_{min}$ projection of the fit to D_2^{*0}), which is taken as confirmation for what seems apparent by plain eye: The fits to individual signal components (using above settings and thus consistent models in both Evtgen and Laura++) are successful, thus confirming the validity of the constructed efficiency histogram. For the sake of completeness the fit to a coherent signal model as simulated with Evtgen is included in Appendix C. Here, the Evtgen amplitudes are

normalized using Eq. 6.1, therefore only approximately reflecting the relative contributions as known from the current world averages. A thus anticipated mismatch of the model shapes is visible particularly in the region of the D_2^* pole mass. Here too $R_B = 0.0 \text{ GeV}^{-1}$ was fixed in order to make Laura++' model fit with Evtgen as closely as possible.

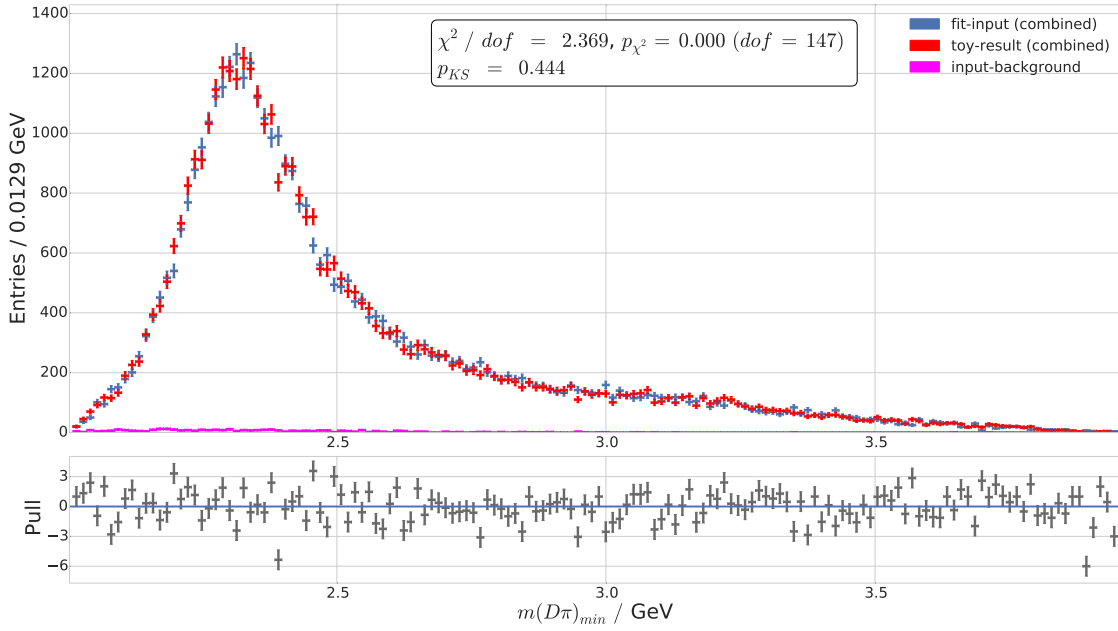
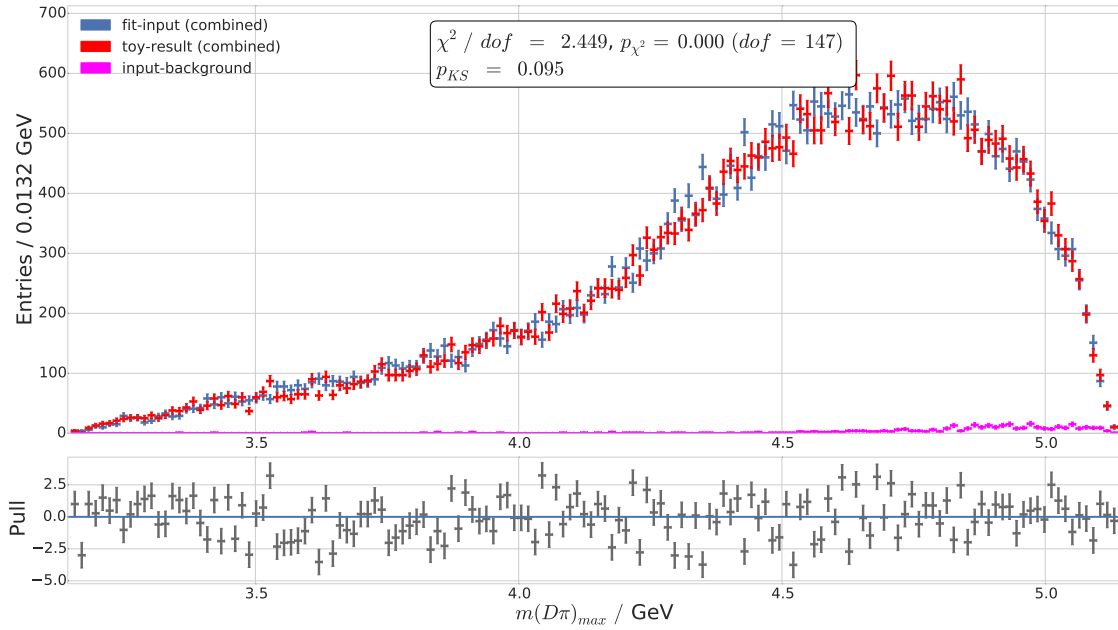
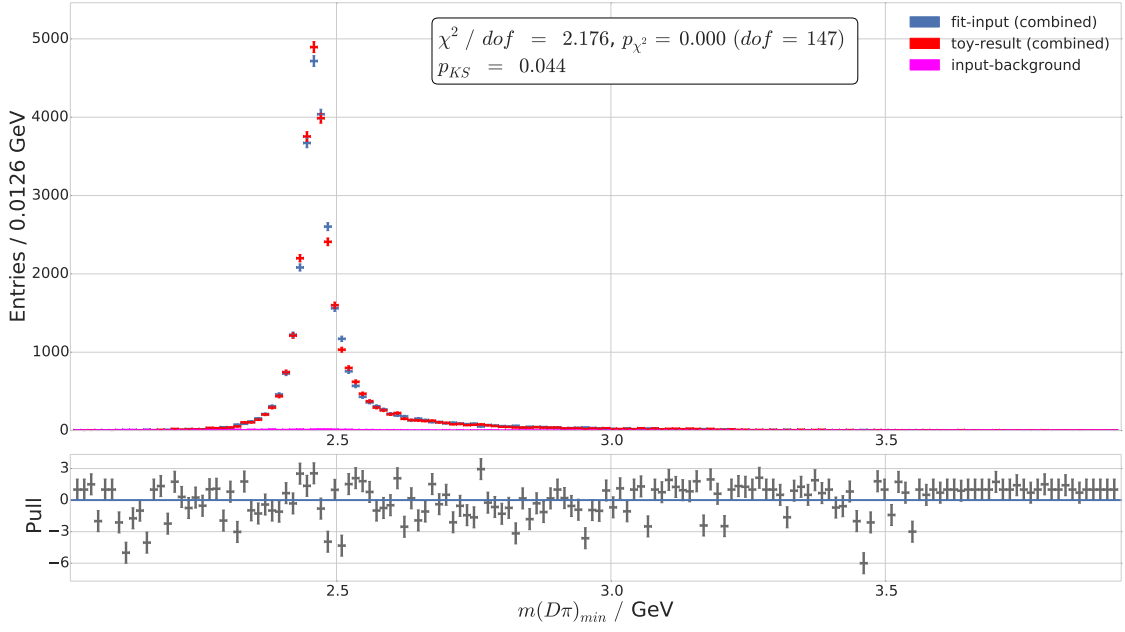
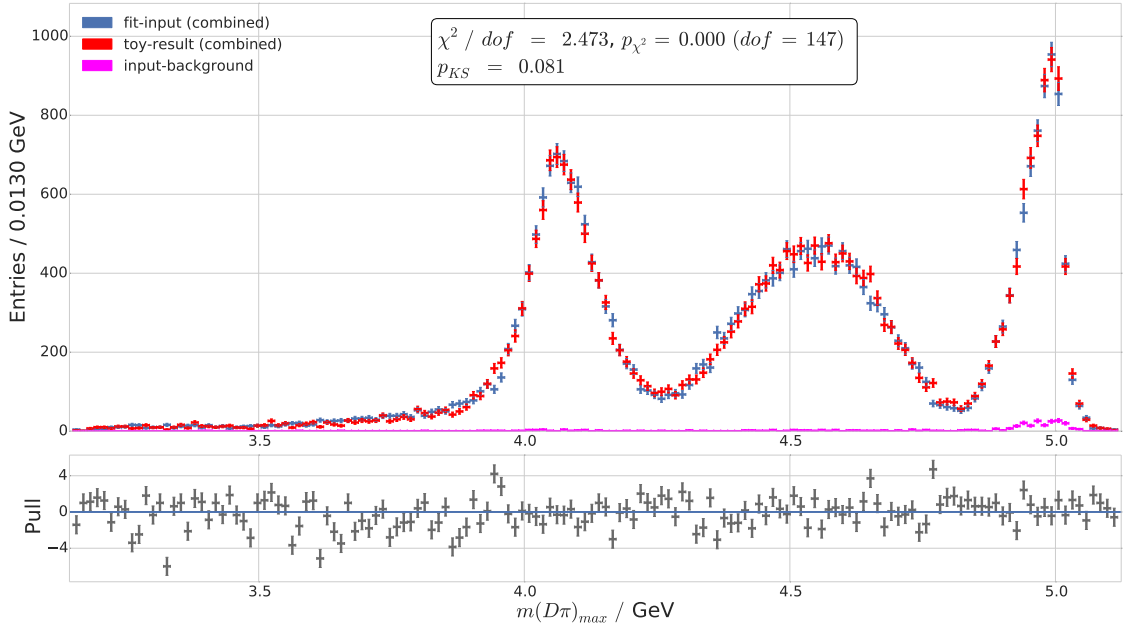

 (a) Fit to D_0^{*0} signal MC in the $m(D\pi)_{min}$ -projection.

 (b) Fit to D_0^{*0} signal MC in the $m(D\pi)_{max}$ -projection.

Figure 11.1.: Fit to D_0^{*0} signal MC projected into the kinematic variables $m(D\pi)_{min}$ and $m(D\pi)_{max}$. The upper panels show the **fit input**, the **fit result** in the form of toys, as well as **misreconstructed signal** candidates. The lower panels depict the pull distributions. The boxes overlaying the figures give the respective reduced χ^2 , its p-value p_{χ^2} as well as the p-value of a Kolmogorov-Smirnov-test, p_{KS} .



(a) Fit to D_2^{*0} signal MC in the $m(D\pi)_{min}$ -projection.



(b) Fit to D_2^{*0} signal MC in the $m(D\pi)_{max}$ -projection.

Figure 11.2.: Fit to D_2^{*0} signal MC projected into the kinematic variables $m(D\pi)_{min}$ and $m(D\pi)_{max}$. The upper panels show the **fit input**, the **fit result** in the form of toys, as well as **misreconstructed signal** candidates. The lower panels depict the pull distributions. The boxes overlaying the figures give the respective reduced χ^2 , its p-value p_{χ^2} as well as the p-value of a Kolmogorov-Smirnov-test, p_{KS} .

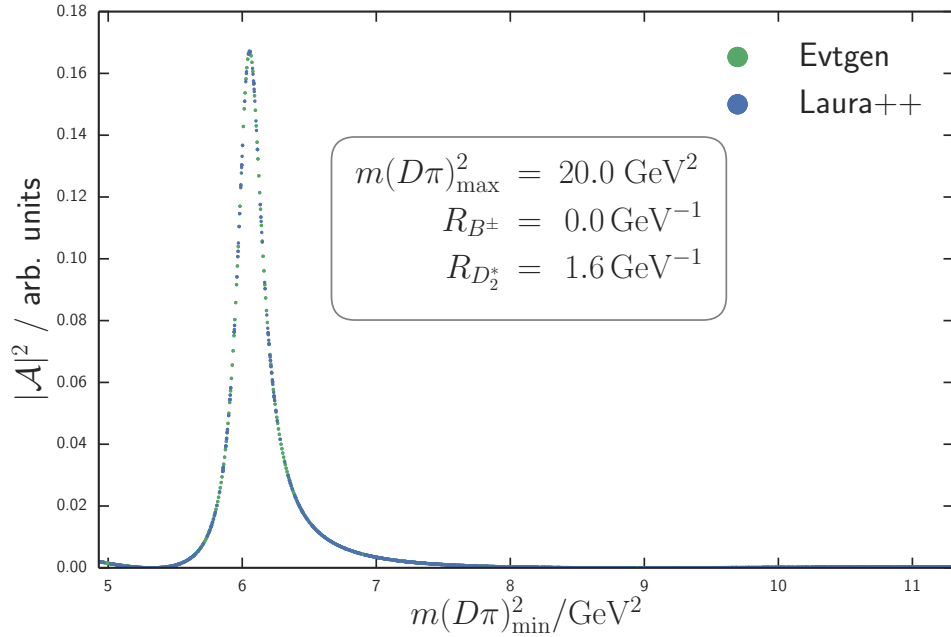
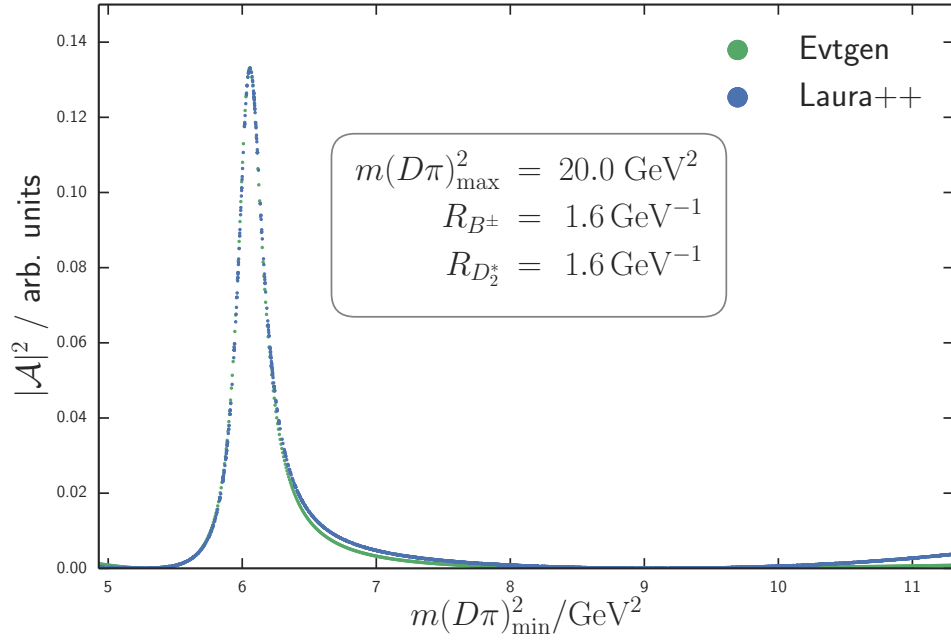


Figure 11.3.: Generator level comparison of the signal model (the amplitude squared, $|\mathcal{A}|^2$) as implemented in **Evtgen** (v12r0) and **Laura++** (v3r0). Both subfigures show $|\mathcal{A}|^2$ sampled at fixed $m(D\pi)_{\max}^2$ and plotted as a function of $m(D\pi)_{\min}^2$ within its physical range. In Fig. 11.3a a meson radius of $R_B = 1.6 \text{ GeV}^{-1}$ is used and a discrepancy between the two models is clearly visible. Remarkably **Laura++**' model exhibits a broader relativistic Breit-Wigner and its amplitude squared values rise for values $m(D\pi)_{\min}^2 \gtrsim 9 \text{ GeV}^2$. For Fig. 11.3b on the other hand a (nonphysical) meson radius of $R_B = 0.0 \text{ GeV}^{-1}$ is employed, as a consequence of which the two models appear to coincide.

12. Fit Validation by Ensemble Tests

In order to examine the fit's ability to correctly capture different input values, so-called ensemble tests (also known as toy studies) are performed. The underlying idea of the tests is to simulate different, physically realistic, scenarios for which checks for potential fit biases are conducted. Each ensemble test constitutes a series of fits to a MC sample for which the signal component is drawn from a 'random model'. The signal components of the random models consist of combinations of amplitudes and phases that are randomly drawn from ranges within the current world averages, as given by the Particle Data Group [31].

Unfortunately, as described in Sec. 6.3, we do not have a MC generator at our disposal, which perfectly satisfies our demands for coherent signal MC, that could then be passed to a full detector simulation. The fit framework Laura++ however can be used to generate MC samples for both the signal component (drawn from the coherent isobar model, using user specified amplitudes and phases, and weighted with the detector efficiency) and the different background components (drawn from the KDE smoothed background histograms, see Chap. 9). Each such random model is chosen to contain the expected amounts of signal and background of the entire Belle dataset.

As mentioned above, the random models are chosen to be coherent and are thus cast in the following form

$$\mathcal{A}_{tot} = \alpha e^{i\phi_\alpha} \mathcal{A}_{NonRes} + \beta e^{i\phi_\beta} \mathcal{A}_{D_0^*} + \gamma \mathcal{A}_{D_2^*},$$

where \mathcal{A}_i ($i \in \{NonRes, D_0^*, D_2^*\}$) are the amplitudes of the contributing channels, α , β and γ are their relative weights, and ϕ_α and ϕ_β are the phases relative to the D_2^* amplitude $\mathcal{A}_{D_2^*}$, which is also described in Eq. 3.6. Because Laura++ individually normalizes the amplitudes \mathcal{A}_i , α , β and γ can be drawn from Gaussian distributions \mathcal{N} , centered at the relative weights $(\alpha_{PDG}, \beta_{PDG}, \gamma_{PDG})$ and varying with the respective errors $(\sigma_{\alpha,PDG}, \sigma_{\beta,PDG}, \sigma_{\gamma,PDG})$. The relative weights can be obtained from the PDG branching ratios \mathcal{B}_i , since $\mathcal{A}_i^2 \sim \mathcal{B}_i$, and are normalized to the inclusive fraction \mathcal{B}_{inc} . This is exemplified below for the random weight β of the D_0^* -amplitude:

$$\begin{aligned}\beta &\sim \mathcal{N}(\beta_{PDG}, \sigma_{\beta, PDG}), \\ \beta_{PDG} &= \sqrt{\frac{\mathcal{B}_{D_0^*, PDG}}{\mathcal{B}_{inc, PDG}}}, \\ \sigma_{\beta, PDG}^2 &= \sum_{i=inc, D_0^*} \left(\frac{\partial \beta_{PDG}}{\partial \mathcal{B}_{i, PDG}} \right)^2 (\delta \mathcal{B}_{i, PDG})^2,\end{aligned}$$

where \mathcal{B}_i are the respective PDG branching ratios and $\delta \mathcal{B}_{i, PDG}$ the combined errors as given by the PDG. Because the relative phases ϕ_β and ϕ_γ are not known, they are sampled from a uniform distribution, \mathcal{U} as in

$$\phi_\beta / \phi_\gamma \sim \mathcal{U}(-\pi, \pi).$$

Note that by construction \mathcal{A}_{tot} is not normalized, which is ok, as the accept-reject method for the generation of samples only requires relative weights to reproduce a signal sample distributed according to $|\mathcal{A}_{tot}|^2$, as intended.

For each such random model 300 MC samples are drawn and fitted. Different fit settings are employed, i.e. different variables are kept floating or fixed, in order to evaluate the performance of distinct settings. The fit results for the (floating) parameters $p_{i, fit}$ are compared to the input values $p_{i, true}$ and the distribution of their pulls should follow a normal distribution. The pull is defined as

$$pull_i = \frac{p_{i, fit} - p_{i, true}}{\sigma_{i, fit}},$$

where $\sigma_{i, fit}$ is the parabolic error of the fit for parameter i . Unbiased parameters with correctly estimated statistical errors should therefore be centered at zero and exhibit a 1- σ statistical error in accordance with unity [13].

12.1. Ensemble Test Results

There generally is a risk that the fit performance with respect to the central values of the pulls of individual parameters is affected by the number of free parameters in the fit. In order to check for this, we gradually add floating parameters, that are of interest to this analysis, and compare the effect this has on the mean and the width of the pull distributions. Contrasting the pull means with their respective errors sheds light on whether the pulls exhibit a bias, that eventually needs to be corrected for. Again, note that aside from excluding $J = 0$ charmed resonances other than the D_0^* and D_2^* (see Sec. 1.2), the presence of other contributing channels can not *a priori* be excluded in Dalitz analyses, such that often different models involving for instance more exotic contributions are tried competitively on data (e.g. see [16], where models including virtual B and virtual D mesons were evaluated on data). In this light, the following ensemble tests are performed under the assumption that besides the confirmed D_0^* and D_2^* there is only a third contribution, which is the non-resonant channel. Since the D_2^* 's exclusive branching fraction as well as its mass and width are well known, we fix the D_2^* 's properties and float the properties of the D_0^* and the non-resonant component, for which there does not even exist a world average value.

Tab. 12.1 - 12.4 show the results of fits to 300 different samples of a (randomly chosen but fixed) random model, where the set of floating parameters differs for each ensemble test shown.

The floating parameters are given in the column *fit parameter*. Note that although the ‘amplitude/phase’ formulation is employed to randomly draw the settings for the random models, the generation as well as the fit to the random models makes use of the ‘real/imaginary-part’ convention, e.g. Laura++ simply uses $\beta e^{i\phi_\beta} = X_{D_0^{*0}} + iY_{D_0^{*0}}$. Commonly the amplitude and phase of the channel with the largest contribution is fixed. But because the amplitude of the D_0^{*0} has additional floating parameters, we choose to fix the real part of the amplitude of the second largest contribution (D_2^*), $X_{D_2^*}$, at its random value the imaginary part $Y_{D_2^*}$ to zero. N_i refers to the number of candidates associated with event type i . The column *#fits* reports the number of successful fits, here all fits (300 of 300) are successful.

Tab. 12.1 presents the result for the simplest case, where only the real and imaginary parts of the D_0^{*0} and the non-resonant amplitude as well as the background yields are left floating (thus representing a minimal set of free parameters). The results depicted in Tab. 12.2 correspond to a test in which the signal yield N_{signal} is added as a free parameter. It is visible that the means and widths of the other floating parameters are left largely unaffected. As can be concluded from Tab. 12.3, leaving the mass and width of the D_0^{*0} floating (while again fixing N_{signal}) seems to introduce a slight bias in the non-resonant amplitude’s real part X_{NonRes} . The test results for the case where all three, the width and mass of the D_0^{*0} as well as the signal yield, are added to the minimal set of free fit parameters are given in Tab. 12.4. From the latter, we can conclude that adding N_{signal} , together with $m(D_0^{*0})$, $\Gamma(D_0^{*0})$, to the minimal set of free parameters does not deteriorate the picture as compared to the case when $m(D_0^{*0})$, $\Gamma(D_0^{*0})$ are left floating, but potential biases persist. This is further investigated below.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i>	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	1.000	-0.0042 ± 0.064	1 ± 0.059
$Y_{D_0^{*0}}$	300	0.317	0.13 ± 0.059	0.92 ± 0.054
X_{NonRes}	300	0.643	-0.093 ± 0.054	0.86 ± 0.048
Y_{NonRes}	300	0.364	0.054 ± 0.056	0.87 ± 0.056
$N_{charged}$	300	433	0.1 ± 0.063	0.97 ± 0.054
N_{mixed}	300	1546	-0.059 ± 0.056	0.89 ± 0.049
$N_{continuum}$	300	506	-0.13 ± 0.06	0.96 ± 0.045

Table 12.1.: Results of ensemble tests for a coherent PDG model with a minimal set of free parameters, i.e. with fixed mass and width of the D_0^{*0} and fixed signal yield.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i>	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	1.000	0.053 ± 0.069	1 ± 0.071
$Y_{D_0^{*0}}$	300	0.317	0.12 ± 0.063	0.94 ± 0.062
X_{NonRes}	300	0.643	-0.042 ± 0.052	0.85 ± 0.05
Y_{NonRes}	300	0.364	-0.0066 ± 0.057	0.91 ± 0.058
N_{signal}	300	5279	-0.058 ± 0.042	0.69 ± 0.033
$N_{charged}$	300	433	0.098 ± 0.064	0.98 ± 0.057
N_{mixed}	300	1546	-0.035 ± 0.053	0.86 ± 0.047
$N_{continuum}$	300	506	-0.1 ± 0.06	0.97 ± 0.046

Table 12.2.: Results of ensemble tests for a coherent PDG model with fixed mass and width of the D_0^{*0} but floating signal yield.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i>	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	1.000	-0.015 ± 0.07	1.1 ± 0.071
$Y_{D_0^{*0}}$	300	0.317	0.18 ± 0.059	0.9 ± 0.054
X_{NonRes}	300	0.643	-0.12 ± 0.055	0.85 ± 0.048
Y_{NonRes}	300	0.364	0.091 ± 0.06	0.95 ± 0.058
$m(D_0^{*0})$	300	2.32	0.0026 ± 0.056	0.86 ± 0.049
$\Gamma(D_0^{*0})$	300	0.267	-0.092 ± 0.058	0.86 ± 0.059
$N_{charged}$	300	433	0.083 ± 0.062	0.95 ± 0.054
N_{mixed}	300	1546	0.039 ± 0.053	0.85 ± 0.046
$N_{continuum}$	300	506	-0.14 ± 0.061	0.95 ± 0.047

Table 12.3.: Results of ensemble tests for a coherent PDG model with floating mass and width of the D_0^{*0} but fixed signal yield. $m(D_0^{*0})$ and $\Gamma(D_0^{*0})$ values are given in GeV.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i>	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	1.000	-0.05 ± 0.069	1.1 ± 0.059
$Y_{D_0^{*0}}$	300	0.317	-0.019 ± 0.063	0.97 ± 0.056
X_{NonRes}	300	0.502	-0.15 ± 0.069	1 ± 0.07
Y_{NonRes}	299	0.243	-0.03 ± 0.06	0.93 ± 0.063
$m(D_0^{*0})$	300	2.32	0.041 ± 0.062	0.98 ± 0.056
$\Gamma(D_0^{*0})$	300	0.267	-0.046 ± 0.062	0.95 ± 0.054
N_{signal}	300	5279	-0.057 ± 0.048	0.79 ± 0.044
$N_{charged}$	300	433	0.01 ± 0.067	0.98 ± 0.048
N_{mixed}	300	15 463	0.081 ± 0.053	0.85 ± 0.045
$N_{continuum}$	300	506	-0.03 ± 0.063	1 ± 0.055

Table 12.4.: Results of ensemble tests for a coherent PDG model with floating mass and width of the D_0^{*0} as well as a floating signal yield N_{signal} . $m(D_0^{*0})$ and $\Gamma(D_0^{*0})$ values are given in GeV. The minimal set of floating parameters (as in Tab. 12.1) is unaffected by the additional degrees of freedom.

The pull distributions corresponding to Tab. 12.4 are depicted in Fig. D.1. Moreover Appendix D lists the results of fits to a number of further toy models in which the same set of parameters as in Tab. 12.4 is left floating (Tab. D.1-D.6). These results seem to suggest that particularly $m(D_0^{*0})$ and X_{NonRes} , are prone to exhibit significant bias (while their standard deviations remain in accordance with unity). There are two potential ways to deal with this effect. Either one finds a way to alleviate the tendency towards fit biases, e.g. by constraining the signal yield, or one counters this deficit by incorporating it as a systematic effect:

- Although the fits vulnerability to biasedness does not necessarily seem to be connected to floating the signal yield, one could try to increase the fit's stability through a reduction of the total number of free parameters by determining N_{signal} from a fit to the ΔE -distribution (and fix it to the value obtained in this manner). This has been done in a similar analysis by M. Heider et al. [21], but was refrained from in this work, as in comparison substantially less parameters are left floating, which was expected to yield an unbiased fit. A disadvantage of having to extract N_{signal} from ΔE ensues from the fact that the latter variable thus becomes unavailable to a multivariate B selection, resulting in increased background yields. Moreover an additional systematic error is introduced.

- Instead of fixing N_{signal} , the biases could be treated by dedicated approaches. As can be argued from the ensemble tests, the unbiasedness of the fit parameters is sensitive to the underlying true values of the random models, even when varied within the ranges known from world averages. A possibility to automatically account for the potential bias is the construction of frequentist confidence intervals (e.g. by a Neyman construction [31]) using toy studies. In this case such a study would entail scanning a two dimensional space (i.e. for X_{NonRes} and $m(D_0^{*0})$) and its feasibility would have to be investigated. A possibility to conservatively factor in this uncertainty would be to generate and fit a number of random models in the vicinity of the values of the amplitudes and phases found in the prospective fit to data. In this toy-study the values could be varied around their fitted values within a range taking into account the statistical fit error. The largest deviation encountered could then serve as a systematic error on the parameter in question.

Judging from Tab. 12.4 and Tab. D.1-D.6, the floating imaginary parts appear quite stable while the pull of N_{mixed} and $N_{continuum}$ tend to exhibit non-negligible biases. This however is not of large concern, as the two background yields are not targets of the aspired measurement.

In conclusion, the ensemble tests reveal that the Dalitz plot analysis of $B^- \rightarrow D^+ \pi^- \pi^-$ using the full Belle data set will allow to simultaneously measure the inclusive branching fraction, the individual branching fractions of the resonant components (the D_2^* component implicitly), the non-resonant component, as well as the mass and width of the D_0^{*0} component, when potential biases are either reduced or factored in as a systematic effect.

13. Outlook

This chapter seeks to lay out how the results from a prospective fit to data would be utilized for a measurement of the inclusive branching fraction as well as the exclusive branching fractions of the non-resonant component and of the two charmed resonances. Moreover the persisting shortcomings of the current MC generation and the fit framework are restated.

To illustrate the anticipated steps, a form of mock analysis is sketched out in the following, in which a model employing the current PDG world averages is used. As the coherent Evtgen model is unable to capture correctly the relative contributions of the channels, Laura++ is used for both the generation and the fit of a coherent signal model (similar to the ensemble tests) in which the relative contributions are taken from current world averages (see Tab. 1.2). In the generation the expected signal and background yields (see Tab. 7.2) are utilized and the signal model is weighted according to the provided reconstruction efficiency. Moreover the parent Blatt-Weisskopf barrier factor radius is set to $R_B = 0.0 \text{ GeV}^{-1}$ in order to prevent a nonphysical rise in the D_2^* amplitude as described in Chap. 11. In order to show that the fit recaptures the PDG inclusive branching ratio, the chosen number of signal events differs from the calculated signal yield in Eq. 1.3. This is because the expected signal yield is dependent on the signal model which enters the game in the form of the average efficiency $\langle \epsilon \rangle$. In this case the average efficiency is therefore taken so as to equal $\bar{\epsilon}$ (see Eq. 3.13) of the chosen PDG model.

The projections of the Dalitz fit are shown in Fig. 13.1. Exceptionally high p -values for both the χ^2 - and the KS-test are apparent, these stellar goodness of fit results are however not to be taken at face value as Laura++' MC generation can only approximate the effects the detector exerts. The fit results together with the respective statistical errors are given in Tab. 13.1 below. From the fit results the inclusive branching fraction and the corresponding statistical error come out as:

$$\mathcal{B}_{B \rightarrow D\pi\pi} = \frac{N_{\text{signal}}}{\bar{\epsilon} \cdot N_{B\bar{B}} \cdot \mathcal{B}_{D \rightarrow K\pi\pi}} = (1.0833 \pm 0.0240) \times 10^{-3}, \quad (13.1)$$

thus recapturing the current PDG world average and promising the potential to almost half the statistical error of the previous Belle analysis [15]. In the latter equation we are however still lacking the K^\pm/π^\pm selection efficiency (ϵ_{PID}) for particle identification, which as of now is not yet available in the b2bii conversion in basf2 (see Chap. 5). Using the fit fractions

(as defined in Eq. 3.10) and the signal averaged reconstruction efficiency $\bar{\epsilon}$ as obtained from Laura++, the exclusive branching fractions can be calculated as shown below. Note that the fit fractions do not sum up to unity, which is caused by destructive interference between the non-resonant and the D_0^* amplitude. This can be understood from the provided interference fit fractions (for a definition see Eq. 3.11), where only the interference fit fraction $FF_{D_0^{*0}, NonRes}$ takes on a sizeable (and negative) value. The corresponding exclusive branching fractions and their statistical errors come out as follows:

$$\mathcal{B}_{B \rightarrow (D_0^* \rightarrow K\pi\pi)\pi\pi} = \frac{FF_{D_0^*} \cdot N_{signal}}{\bar{\epsilon} \cdot N_{B\bar{B}} \cdot \mathcal{B}_{D \rightarrow K\pi\pi}} = (8.3072 \pm 0.5248) \times 10^{-4},$$

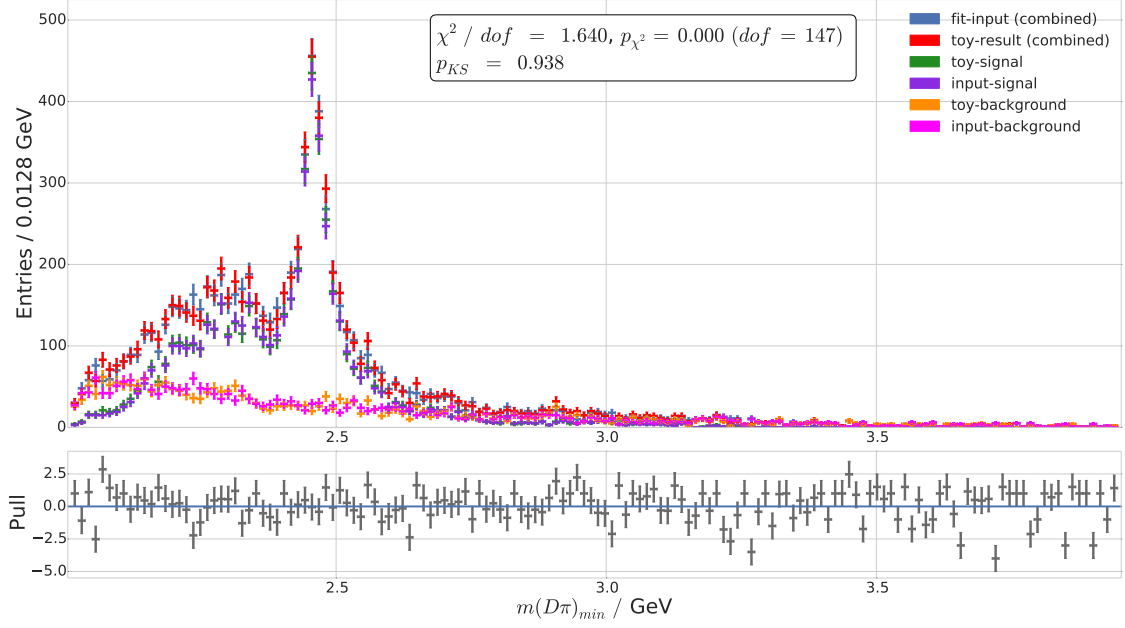
$$\mathcal{B}_{B \rightarrow (D_2^* \rightarrow K\pi\pi)\pi\pi} = \frac{FF_{D_2^*} \cdot N_{signal}}{\bar{\epsilon} \cdot N_{B\bar{B}} \cdot \mathcal{B}_{D \rightarrow K\pi\pi}} = (4.3249 \pm 0.2727) \times 10^{-4},$$

$$\mathcal{B}_{B \rightarrow D\pi\pi, NonRes} = \frac{FF_{NonRes} \cdot N_{signal}}{\bar{\epsilon} \cdot N_{B\bar{B}} \cdot \mathcal{B}_{D \rightarrow K\pi\pi}} = (1.2321 \pm 0.0780) \times 10^{-4}.$$

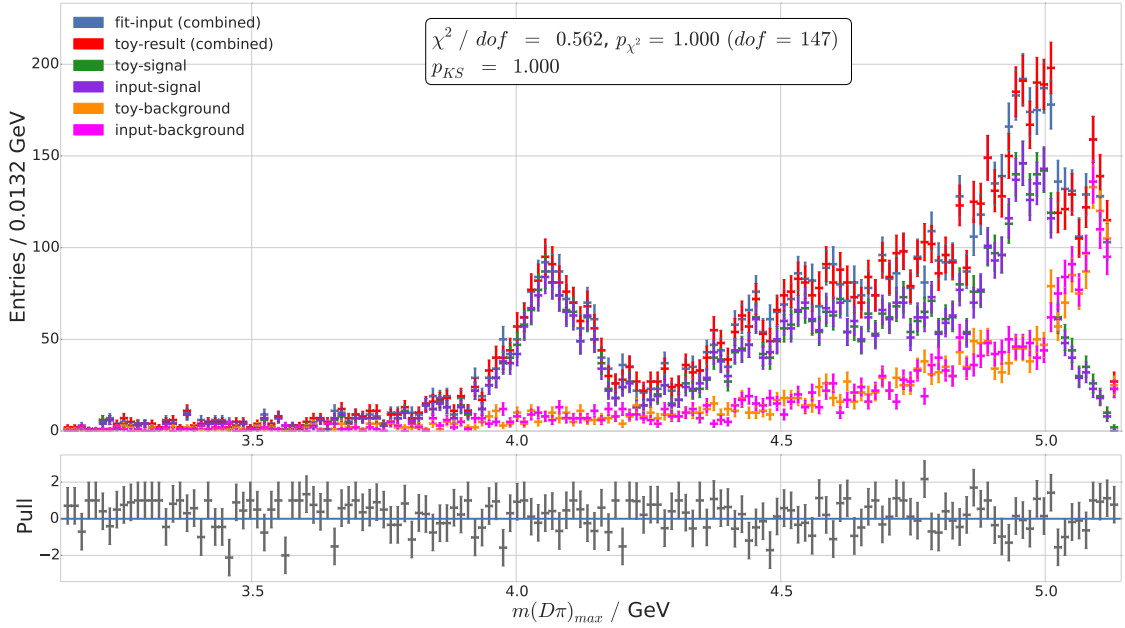
Although the fit fraction results recapture their input values within their statistical error, the exclusive branching fractions deviate significantly from the world averages (see Tab. 1.2), with which the fitted MC sample was modelled. This is however expected as the true interference between the channels, amongst others ensuing from the phase differences between the decay channels, which are set to zero in the present case, is unknown. On real data we expect to recapture the physical interplay of the real and imaginary parts of the amplitudes that governed the previous measurements of the exclusive fractions. The other observables of interest, the mass and width of the D_0^* , $m(D_0^{*0})$ and $\Gamma(D_0^{*0})$, are of course readily available as they are floating parameters of the fit. Although Tab. 13.1 only states the statistical errors, the potential to help lower the uncertainties of the respective world averages (Tab. 1.1) is apparent.

Systematic errors in this analysis are mostly attributed to fixed assumptions. The uncertainties of the following sources of systematic uncertainties should be evaluated as discussed below and added in quadrature:

- ϵ As opposed to the background histograms, the reconstruction efficiency ϵ is unsmoothed, which is why it introduces errors due to statistical fluctuations in the bins. Laura++ allows to fluctuate the bin contents according to Poisson distributions, which allows to extract the associated uncertainty in the form of a toy study.
- ϵ_{PID} The K^\pm/π^\pm selection efficiency usually exhibits a systematic shift due to differences between the simulated and the recorded data, which, in ϵ_{PID} , is corrected for through a channel specific calibration that however introduces a calibration uncertainty.
- $N_{B\bar{B}}$ The total number of $B\bar{B}$ events is only known to a certain precision, which needs to be propagated through in the branching ratios.
- D_2^{*0} The parameters $m(D_2^{*0})$ and $\Gamma(D_2^{*0})$ are fixed to their PDG world averages. By varying both parameters within their PDG errors, the corresponding systematic errors can be evaluated.
- R_B/R_D The meson radii used in the Blatt-Weisskopf barrier factors are unknown and generally assumed to lie within $1\text{-}5 \text{ GeV}^{-1}$ [31]. The radii should be varied within these limits to evaluate the associated uncertainty.
- fit bias** Finally, following the discussion in Chap. 12, it remains to be resolved whether the observed fit bias can and should be mitigated by fixing the signal yield in the fit or whether instead the bias should be evaluated in a toy study and treated as a systematic error.



(a) Fit to coherent signal plus background Laura++ MC in the $m(D\pi)_{min}$ -projection.



(b) Fit to coherent signal plus background Laura++ MC in the $m(D\pi)_{max}$ -projection.

Figure 13.1.: Fit to coherent signal plus background Laura++ MC projected into the kinematic variables $m(D\pi)_{min}$ and $m(D\pi)_{max}$. The upper panels show the **combined fit input** and the **combined fit result**. Moreover the signal and background components which are combined in the former are differentiated: The **toy signal** and the **input signal** as well as the **input background** and the **toy background** candidates. The lower panels depict the pull distributions. The boxes overlaying the figures give the respective reduced χ^2 , its p-value p_{χ^2} as well as the p-value of a Kolmogorov-Smirnov-test, p_{KS} .

<i>fit parameter</i>	<i>true value</i>	<i>result</i>	<i>error</i>
$X_{D_0^{*0}}$	0.773	0.793	0.027
$X_{D_2^{*0}}$	0.572	fixed	fixed
X_{NonRes}	0.273	0.305	0.034
$Y_{D_0^{*0}}$	0.0	-0.001	0.057
$Y_{D_2^{*0}}$	0.0	fixed	fixed
Y_{NonRes}	0.0	-0.020	0.064
$m(D_0^{*0})$	2318	2318.31	6.22
$\Gamma(D_0^{*0})$	267	281.86	15.22
N_{signal}	4870	4921.14	109.14
$N_{charged}$	433	389.65	140.95
N_{mixed}	1546	1457.20	78.79
$N_{continuum}$	506	587.02	85.83
$FF_{D_0^{*0}}$	0.7352	0.7668	0.0454
$FF_{D_2^{*0}}$	0.402	0.3992	0.0236
FF_{NonRes}	0.0919	0.1137	0.0067
$FF_{D_0^{*0}, D_2^{*0}}$	0.0049	0.0050	-
$FF_{D_0^{*0}, NonRes}$	-0.2340	-0.2847	-
$FF_{D_2^{*0}, NonRes}$	$1.4 \cdot 10^{-5}$	$1.3 \cdot 10^{-5}$	-
$\bar{\epsilon}$	-	0.0644	-

Table 13.1.: Results of a fit for a coherent PDG model generated with **Laura++** with all desired parameters left floating. The parameters above the horizontal separation are direct observables in the fit, the ones below are each a function of the direct observables.

As can be gathered from above, the work at hand is able to perform satisfyingly on MC data. This is true within the scope of technical hindrances ranging from major challenges such as the unclear validity of **Laura++**' signal model in the case of finite R_B meson radii to smaller obstructions such as Evtgens unnormalized signal model or the current unavailability of the ϵ_{PID} calibration in b2bii. For these reasons the present analysis is not yet in the position to undergo an immediate unblinding.

14. Conclusion

The present work performs a Dalitz analysis of the decay $B^- \rightarrow D^+ \pi^- \pi^-$ on Monte Carlo samples corresponding to the total integrated luminosity recorded by the Belle experiment. The decay channel exhibits a rich resonant substructure, providing an excellent handle to measure the properties of charmed P-wave excitations as well as their exclusive branching fractions. A precise understanding of excited D -mesons is highly desirable as it reduces uncertainties in semileptonic decays, which are sensitive to new physics.

While the Belle collaboration has already performed a Dalitz analysis of the channel at hand, using only about a 10^{th} of the data available today, we show that revisiting the channel holds the potential of increasing the precision of our knowledge of the mass and width of the $D_0^*(2400)^0$ resonance. The present work prepares a prospective Dalitz fit on recorded data and in doing so explores different analysis specific procedures. Compared to Belle's previous analysis a number of improvements is achieved. For instance, an improved suppression of continuum background is expected and by accounting for correlations between selection criteria we also project an enhanced selection of B candidates.

Although the present work aims at utilizing Belle data, it was conducted utilizing the software framework of the successor experiment Belle II. Besides evaluating the software, new functionality, required in the context of Dalitz analyses, was added in the course of the work, thus paving the way for future Dalitz analyses to come.

A number of shortcomings in third-party software libraries was discovered and poses hindrances, which will have to be overcome before a prospective unblinding can take place. For example the Monte Carlo generator Evtgen does not correctly handle the relative normalizations of the sub-channels, therefore preventing both the evaluation of the reconstruction efficiency and the execution of ensemble tests using a coherent MC sample including a sophisticated modelling of detector effects. More severely even, the maximum likelihood fitter Laura++ is found to mismodel the relativistic Breit-Wigner amplitude when using finite meson radii, thus questioning its applicability in its current state. Due to the limited time allocated to this work, these flaws persist.

Working around these hindrances, toy studies, so-called ensemble tests, are used to test Dalitz fits to signal models randomly varied within the errors from world average expectations. Those

tests indicate that depending on the (randomly) chosen input parameters the Dalitz fit's unbiasedness is not always guaranteed. Here, further investigations need to be undertaken in order to clarify whether steps to eliminate the biases can be put in place or whether it has to be addressed systematically.

To reiterate and conclude, the present work establishes a basis from which a prospective Dalitz fit to the full set of recorded Belle data can start off and tie in with. Besides a projected increase in the precision of the $D_0^*(2400)^0$ ' mass and width, the targeted channel of the present MC study, $B^- \rightarrow D^+ \pi^- \pi^-$, could potentially prove a worthy accomplice in the quest for discovering physics beyond the standard model.

Bibliography

- [1] Roel Aaij, B Adeva, M Adinolfi, A Affolder, Ziad Ajaltouni, S Akar, J Albrecht, F Alessio, M Alexander, S Ali, et al. Observation of overlapping spin-1 and spin-3 $\bar{D}^0 K^-$ resonances at mass 2.86 GeV/c². *Physical review letters*, 113(16):162001, 2014.
- [2] Roel Aaij, B Adeva, M Adinolfi, A Affolder, Ziad Ajaltouni, S Akar, J Albrecht, F Alessio, M Alexander, S Ali, et al. Dalitz plot analysis of $B^0 \rightarrow \bar{D}^0 \pi^+ \pi^-$ decays. *Physical Review D*, 92(3):032002, 2015.
- [3] Kazuo Abe, T Abe, I Adachi, H Aihara, M Akatsu, Y Asano, T Aso, T Aushev, AM Bakich, Y Ban, et al. Study of $B^- \rightarrow D^{*0} \pi^- (D^{*0} \rightarrow D^{(*)+} \pi^-)$ decays. *Physical Review D*, 69(11):112002, 2004.
- [4] Sea Agostinelli, John Allison, K al Amako, J Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.
- [5] Bernard Aubert, M Bona, Y Karyotakis, JP Lees, V Poireau, E Prencipe, X Prudent, V Tisserand, J Garra Tico, E Grauges, et al. Evidence for direct CP violation from dalitz-plot analysis of $B^\pm \rightarrow K^\pm \pi^\mp \pi^\pm$. *Physical Review D*, 78(1):012004, 2008.
- [6] Bernard Aubert, M Bona, Y Karyotakis, JP Lees, V Poireau, E Prencipe, X Prudent, V Tisserand, J Garra Tico, E Grauges, et al. Dalitz plot analysis of $B^- \rightarrow D^+ \pi^- \pi^-$. *Physical Review D*, 79(11):112004, 2009.
- [7] A.J. Bevan et al. The Physics of the B Factories. *Eur.Phys.J.*, C74(11):3026, 2014.
- [8] AJ Bevan, Boštjan Golob, Th Mannel, S Prell, BD Yabsley, H Aihara, F Anulli, N Arnaud, T Aushev, M Beneke, et al. The physics of the B factories. *The European Physical Journal C*, 74(11):1–928, 2014.
- [9] Volker Blobel and Erich Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Springer-Verlag, 2013.
- [10] Rene Brun and Fons Rademakers. ROOT - An Object Oriented Data Analysis Framework. *Proceedings AIHENP'96 Workshop, Lausanne, Nucl. Inst. & Meth.*, A389:81–86, 1997. <http://root.cern.ch>.
- [11] SU Chung, J Brose, R Hackmann, E Klempt, S Spanier, and C Strassburger. Partial wave analysis in K-matrix formalism. *Annalen der Physik*, 507(5):404–430, 1995.
- [12] Pablo del Amo Sánchez. *Time-dependent Dalitz-Plot Analysis of the Charmless Decay $B_0 \rightarrow K_0$* . PhD thesis, University of Birmingham, 2007.
- [13] Luc Demortier and Louis Lyons. Everything you always wanted to know about pulls. 2002.
- [14] A. Abashian et al. The belle detector. *Nucl. Instrum. Meth.*, A79:117–232, 2002.

-
- [15] Abe et al. Study of $B^- \rightarrow D^{*0}\pi^- (D^{*0} \rightarrow D^{(*)+}\pi^-)$ decays. *Phys. Rev. D*, 69:112002, Jun 2004.
 - [16] Aubert et al. Dalitz plot analysis of $B^- \rightarrow D^+\pi^-\pi^-$. *Phys. Rev. D*, 79:112004, Jun 2009.
 - [17] K. Nakamura et al. (Particle Data Group),. *Journal of Physics*, G37, 075021, 2010 and 2011 partial update for the 2012 edition.
 - [18] M Feindt and U Kerzel. The NeuroBayes neural network package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(1):190–194, 2006.
 - [19] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
 - [20] Stephen Godfrey and Richard Kokoski. Properties of P -wave mesons with one heavy quark. *Phys. Rev. D*, 43:1679–1687, Mar 1991.
 - [21] Manuel Heider. *Dalitz analysis of $B^- \rightarrow D^0\pi^-\pi^0$ and $B^0 \rightarrow D^-\pi^+\pi^0$ at the Belle experiment*. PhD thesis, Karlsruhe Institute of Technology, 2015.
 - [22] Matthias Huschle. *Measurement of the branching ratio of $B^+ \rightarrow D^{(*)}\tau^+\nu_\tau$ relative to $B \rightarrow D^{(*)}l\nu_l$ with hadronic tagging at Belle*. PhD thesis, KIT, 2015.
 - [23] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput.Phys.Commun.*, 10:343–367, 1975.
 - [24] TM Karbach et al. Dalitz plot analysis of $B^- \rightarrow D^+\pi^-\pi^-$. *arXiv preprint arXiv:1001.3345*, 2010.
 - [25] Thomas Keck. The Full Event Interpretation for Belle II. Master’s thesis, Karlsruhe Institute of Technology, 2014.
 - [26] Bastian Kronenbitter. *Measurement of the branching ratio of $B^+ \rightarrow \tau^+\nu_\tau$ decays with the semileptonic tagging method*. PhD thesis, KIT, 2015.
 - [27] S. Kurokawa and E. Kikutani. Overview of the KEKB accelerators. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 499(1):1 – 7, 2003.
 - [28] A. Kuzmin. Study of $B^- \rightarrow D^{*0}\pi^- (D^{*0} \rightarrow D^{(*)+}\pi^-)$ processes. *Internal Note: Belle Note*, 552, Nov. 2002.
 - [29] David J Lange. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152–155, 2001.
 - [30] Thomas Latham. The Laura++ Dalitz plot fitter. *arXiv preprint arXiv:1603.00752*, 2016.
 - [31] K. A. Olive et al. Review of Particle Physics. *Chin. Phys.*, C38:090001, 2014.
 - [32] M. Prim et al. Angular analysis of $B^0 \rightarrow \phi K^*$ decays and search for CP violation at Belle. *Phys.Rev.*, D88(7):072004, 2013.
 - [33] Markus Röhrken. *Time-Dependent CP Violation Measurements in Neutral B Meson to Double-Charm Decays at the Japanese Belle Experiment*. PhD thesis, Karlsruhe Institute of Technology, 2012.
 - [34] Torbjörn Sjöstrand. High-energy-physics event generation with pythia 5.7 and jetset 7.4. *Computer Physics Communications*, 82(1):74–89, 1994.

- [35] Wouter Verkerke. The roofit toolkit for data modeling. June 2003. arXiv:physics/0306116v1.
- [36] Wolfgang Waltenberger. Rave—a detector-independent toolkit to reconstruct vertices. *Nuclear Science, IEEE Transactions on*, 58(2):434–444, 2011.
- [37] Wikipedia. Kernel density estimation — Wikipedia, the free encyclopedia, 2015. [Online; accessed 18-May-2016].

A. basf2 Steering File

The steering file below shows the online reconstruction of the channel $B^- \rightarrow D^+ \pi^- \pi^-$ and the charge conjugate thereof step by step. First, all identifiers from the basf2 and modularAnalysis Python modules as well as functionality to convert the Belle data to the Belle II format (functions from the b2biiConversion module) are loaded into the current namespace. Modules can either be registered directly using the register_module function or by using the convenience functions provided by modularAnalysis.

```
from basf2 import *
from modularAnalysis import *
from b2biiConversion import convertBelleMdstToBelleIIMdst, setupB2BIIDatabase
```

First a local database that stores the Belle beam parameters is set up after which, using the Belle detector parameters, the provided Belle MC input *mdst*-file (referred to as *infile* below) is converted on the fly and made available in the DataStore. By default the analysis path implicitly created below is named *analysis_main*.

```
setupB2BIIDatabase(True)
convertBelleMdstToBelleIIMdst(infile)
```

For each Track in the DataStore hypotheses for so-called charged final-state particles are created. Below, Tracks in line with either the kaon or the pion hypothesis fulfilling specific selection requirements (as elaborated upon in Sec. 7.2) are selected and assigned to ParticleLists.

```
pions_plus = ('pi+:good', 'chiProb > 0.001 and atcPIDBelle(2,3) > 0.2 and abs(dz) < 4.0 and abs(dr) < 0.2')
kaons = ('K-:good', 'chiProb > 0.001 and atcPIDBelle(3,2) > 0.6 and abs(dz) < 4.0 and abs(dr) < 0.2')
fillParticleLists([kaons, pions_plus])
```

Candidates from the two ParticleLists that fulfill a chosen criterion for their combined invariant mass are teamed up to form D -meson candidates. As is true for all other modules, reconstructDecay('D+->K- pi+ pi+') creates two ParticleLists. One for $D^+ \rightarrow K^- \pi^+ \pi^+$ and one for the charge conjugate decay. All convenience functions automatically operate on the charge conjugate ParticleList as well.

```
reconstructDecay('D+:Kpipi -> K-:good pi+:good pi+:good', '1.84 < M < 1.89')
```

Before performing a mass-constrained vertex fit, kinematic variables of interest for an offline selection are copied to ExtraInfo variables in order to prevent them from being modified by the vertex fit:

```
extraInfo_Variables_D = {'M':'D_Mass', 'deltaM':'deltaMD', 'abs_deltaM':'abs_deltaMD'}
variablesToExtraInfo('D+:Kpipi', extraInfo_Variables_D)
```

This is followed by an actual vertex fit using the Tracks of the D 's daughters, which allows to discriminate against unsuited candidates by rejecting failed fits (by dropping candidates with a negative p -value). `fitVertex()` takes into account the Belle detector geometry and performs a vertex fit using the RAVE package [36]. The momenta, their covariance matrix and the vertex p -value of the D 's daughters are updated in this step:

```
fitVertex('D+:Kpipi', 0.0, fitter='rave', fit_type='massvertex', daughtersUpdate=True)
```

In order to preserve the p -value of the χ^2 mass-constrained vertex fit of the D from being updated after the B 's vertex fit its value is also saved in an ExtraInfo:

```
extraInfo_Var_D = {'chiProb':'chiProb_D_init'}
variablesToExtraInfo('D+:Kpipi', extraInfo_Var_D)
```

D candidates recombined with two pions which fulfill the kinematic preselection given below are used to form B candidates:

```
reconstructDecay('B-:Dpipi -> D+:Kpipi pi-:good pi-:good', '5.24 < Mbc < 5.29 and abs(deltaE) < 0.2')
```

After this step the best candidate selection is performed, where the custom variable `dMbc` computes the quantity $M_{bc} - M(B)_{PDG}$ (see Sec. 7.8) and the candidates within each event are ranked in the order of ascending magnitudes of `dMbc`.

```
rankByLowest('B-:Dpipi', variable='abs(dMbc)', outputVariable='bestCandRank')
```

Subsequently, kinematic variables of the B candidates relevant to the offline selection are again saved to ExtraInfo variables. This is followed by a mass-constrained vertex fit of the B candidates, similar to the one for the D s above.

```
extraInfo_Variables_B = {'Mbc':'Mbc', 'M':'M', 'deltaE':'deltaE', 'abs_deltaE':'abs_deltaE'}
variablesToExtraInfo('B-:Dpipi', extraInfo_Variables_B)
fitVertex('B-:Dpipi', 0.0, fitter='rave', fit_type='massvertex', daughtersUpdate=True)
```

Afterwards a number of modules which compute additional variables of interest are called through convenience functions. The following calls compute the MC truth for the B candidates and either directly prepare or compute variables used for continuum suppression:

```
matchMCTruth('B-:Dpipi')
buildRestOfEvent('B-:Dpipi')
buildContinuumSuppression('B-:Dpipi')
TagV('B-:Dpipi', 'breco')
```

To enable the multivariate offline selection of candidates, described in Sec. 7.7, the BDT for the continuum suppression needs to be initialized by reading in the trained weights from the specified `workingDirectory`. The module that applies the BDT for the multivariate offline selection is registered in the path subsequently (here too the location of the pretrained weights needs to be specified).

```
applyTMVAMethod('B-:Dpipi', method='FastBDT', signalClass=1, signalFraction='-1.',
  expertOutputName='FastBDT_out', workingDirectory=train_dir, prefix='all2',
  transformToProbability=False)
applyTMVAMethod('B-:Dpipi', method='FastBDT', signalClass=1, signalFraction='-1.',
  expertOutputName='mva_selection', workingDirectory=mva_dir, prefix='mva2',
  transformToProbability=False)
```

Once all modules directly involved in the selection and reconstruction are added to the path, the only remaining thing to be specified is the list of variables to be written to a *root*-file. To provide more descriptive or abbreviated variable names, aliases are configured. The functions used therein which compute variables of the B candidates' daughters utilize the given daughter indices: $D = 0, \pi_1 = 1, \pi_2 = 2$.

```
from variables import variables as v
v.addAlias("MDpi1", "daughterInvariantMass(0, 1)")
v.addAlias("MDpi2", "daughterInvariantMass(0, 2)")
v.addAlias("Mpipi", "daughterInvariantMass(1, 2)")
v.addAlias("dMbc", "formula(Mbc-5.27926)")
v.addAlias("Mbc_init", "extraInfo(Mbc)")
v.addAlias("D_Mass", "daughter(0, M)")
v.addAlias("D_Mass_init", "daughter(0, extraInfo(D_Mass))")
v.addAlias("deltaMD_init", "daughter(0, extraInfo(deltaMD))")
v.addAlias("abs_deltaMD_init", "daughter(0, extraInfo(abs_deltaMD))")
v.addAlias("bestCandRank", "extraInfo(bestCandRank)")
v.addAlias("bestCandRank_init", "extraInfo(bestCandRank)")
v.addAlias("M_init", "extraInfo(M)")
v.addAlias("deltaE_init", "extraInfo(deltaE)")
v.addAlias("deltaMD", "daughter(0, deltaM)")
v.addAlias("chiProb_D", "daughter(0, chiProb)")
v.addAlias("chiProb_D_init", "daughter(0, extraInfo(chiProb_D_init))")
v.addAlias("FBDT_out", "extraInfo(FastBDT_out)")
v.addAlias("mva_selection", "extraInfo(mva_selection)")
v.addAlias("abs_deltaE_init", "extraInfo(abs_deltaE)")
```

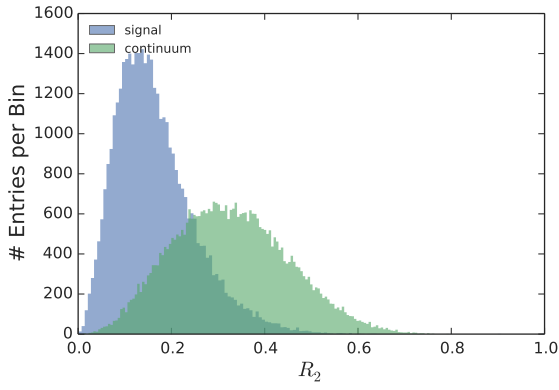
After defining the aliases above, the specified lists of variables are written to a *root*-file (whose location and name of the data tree is specified through *treename* and *filename*, respectively):

```
variables = ['dMbc', 'bestCandRank', 'bestCandRank_init', 'M', 'M_init', 'isSignal',
            'chiProb', 'chiProb_D', 'chiProb_D_init', 'Mbc', 'Mbc_init', 'deltaE', 'deltaE_init',
            'abs_deltaE', 'abs_deltaE_init', 'TrueMDpiMax', 'TrueMDpiMin', "mcErrors",
            "MDpi1", "MDpi2", "Mpipi", 'expNum', 'runNum', 'evtNum', 'isContinuumEvent',
            'D_Mass', 'D_Mass_init', 'deltaMD', 'deltaMD_init', 'abs_deltaMD_init',
            'FBDT_out', 'mva_selection']
CS_variables = ['R2', 'cosTBT0', 'KSFVariables(hso02)', 'KSFVariables(hso12)', 'cosTBz',
               'CleoCone(9)', 'thrustBm', 'thrust0m', 'KSFVariables(et)', 'KSFVariables(mm2)',
               'KSFVariables(hso00)', 'KSFVariables(hso04)', 'KSFVariables(hso10)',
               'KSFVariables(hso14)', 'KSFVariables(hso20)', 'KSFVariables(hso22)',
               'KSFVariables(hso24)', 'KSFVariables(hso0)', 'KSFVariables(hso1)',
               'KSFVariables(hso2)', 'KSFVariables(hso3)', 'KSFVariables(hso4)',
               'CleoCone(1)', 'CleoCone(2)', 'CleoCone(3)', 'CleoCone(4)',
               'CleoCone(5)', 'CleoCone(6)', 'CleoCone(7)', 'CleoCone(8)']
variablesToNTuple('B-Dpipi', variables=variables + CS_variables, treename=treename,
                 filename=outfile)
```

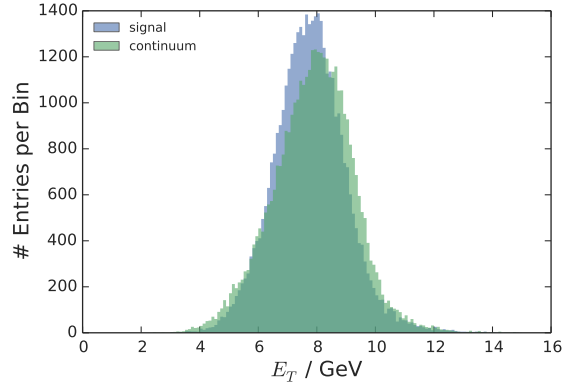
The code execution on steering file level ends with the following command, that triggers the execution of the entire path:

```
process(analysis_main)
```

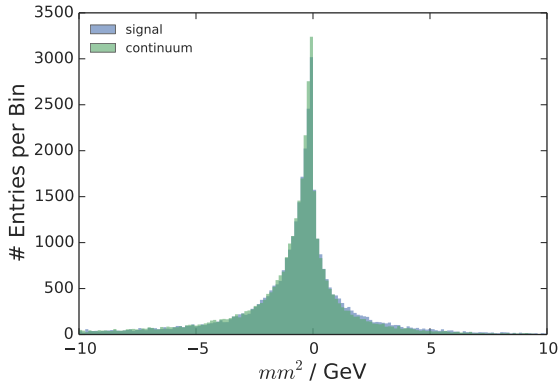

B. Continuum Suppression Variables



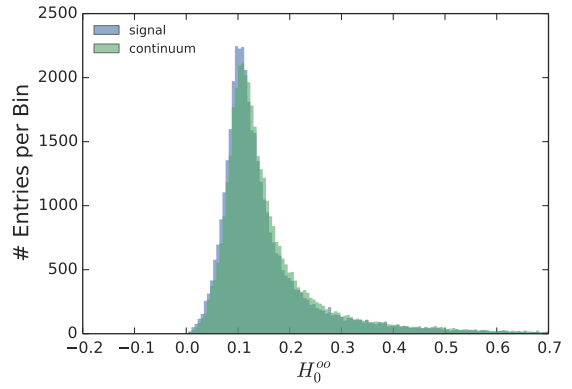
(a)



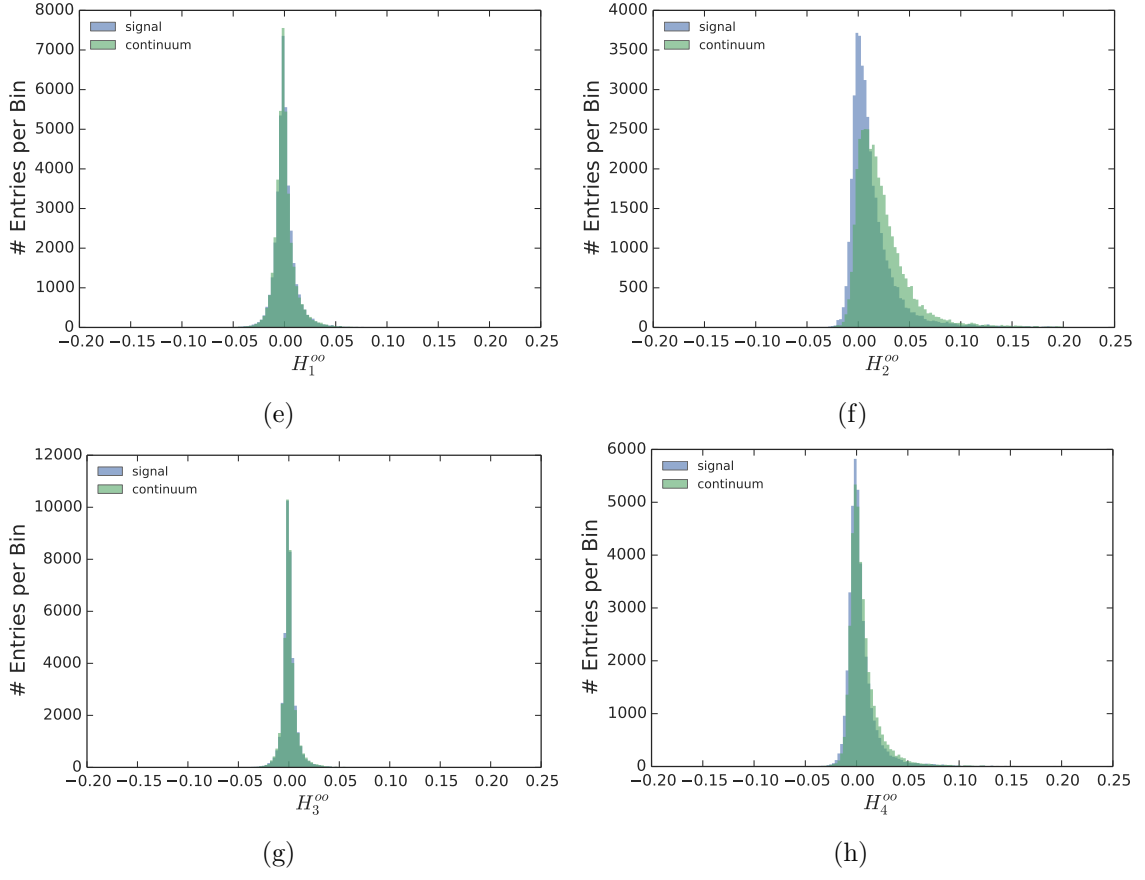
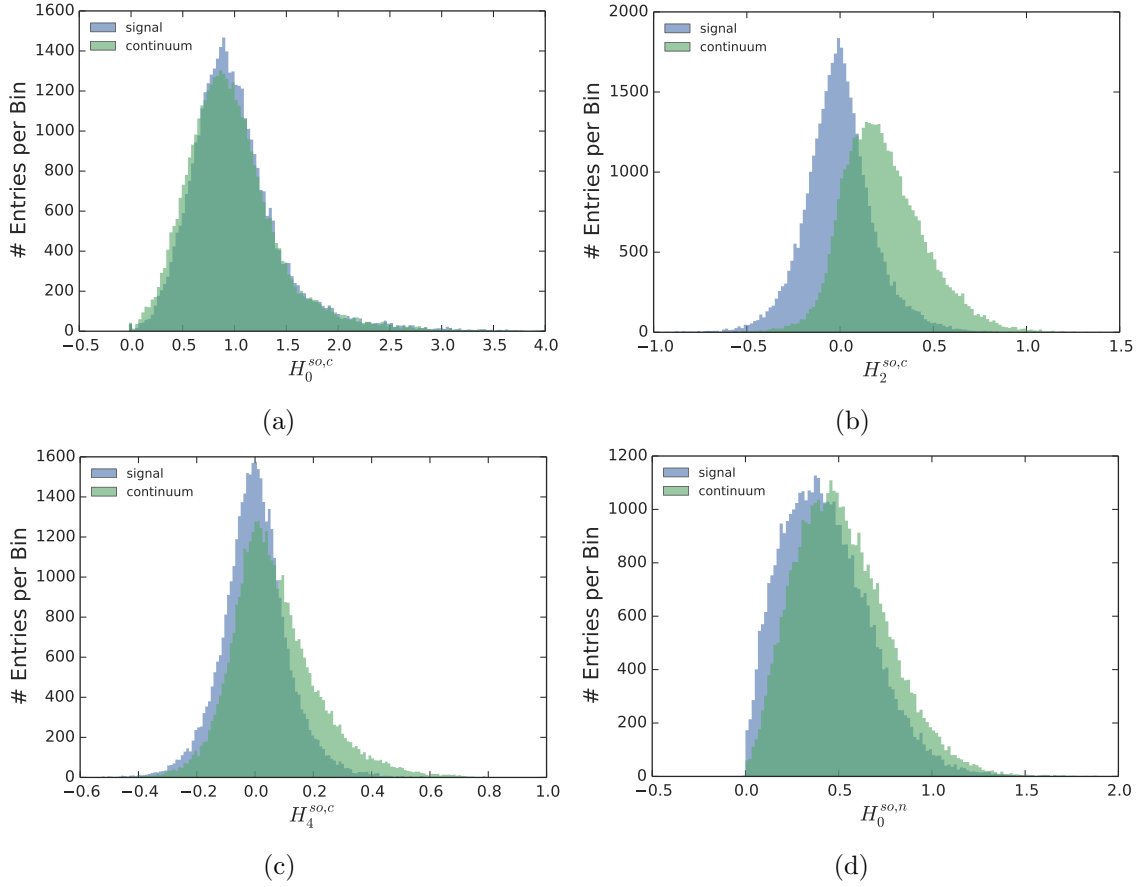
(b)



(c)



(d)


 Figure B.1.: The distributions of $R2$, E_T , mm^2 and the H_i^{oo} .


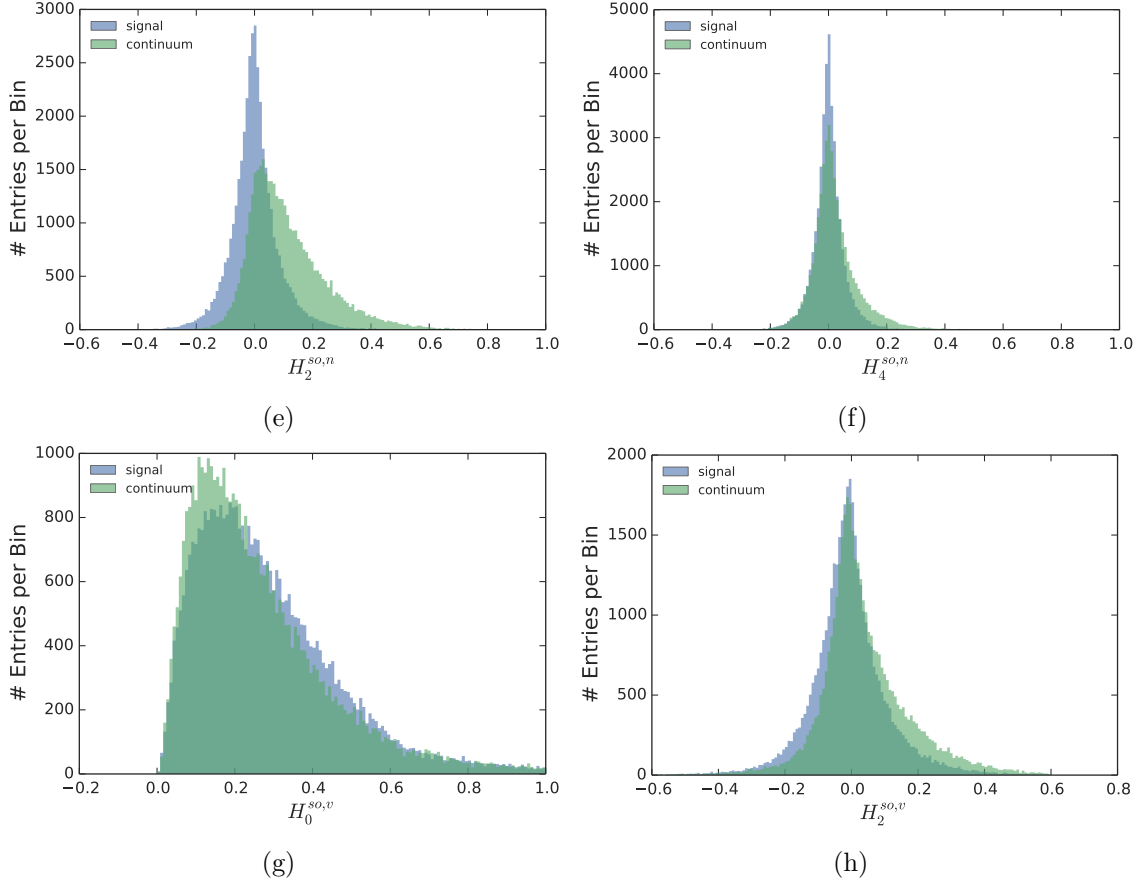
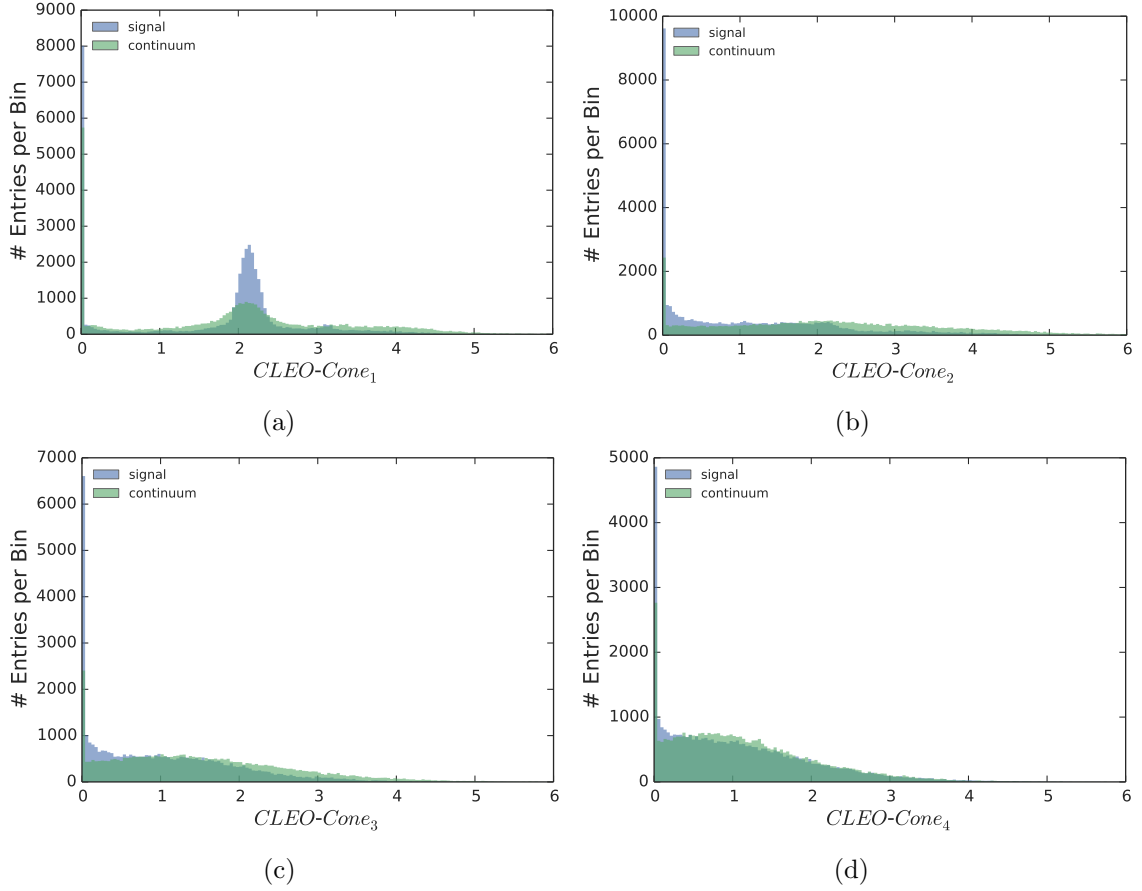


Figure B.2.: The distributions of the $H_i^{so,c}$, $H_i^{so,n}$ and the $H_i^{so,v}$.



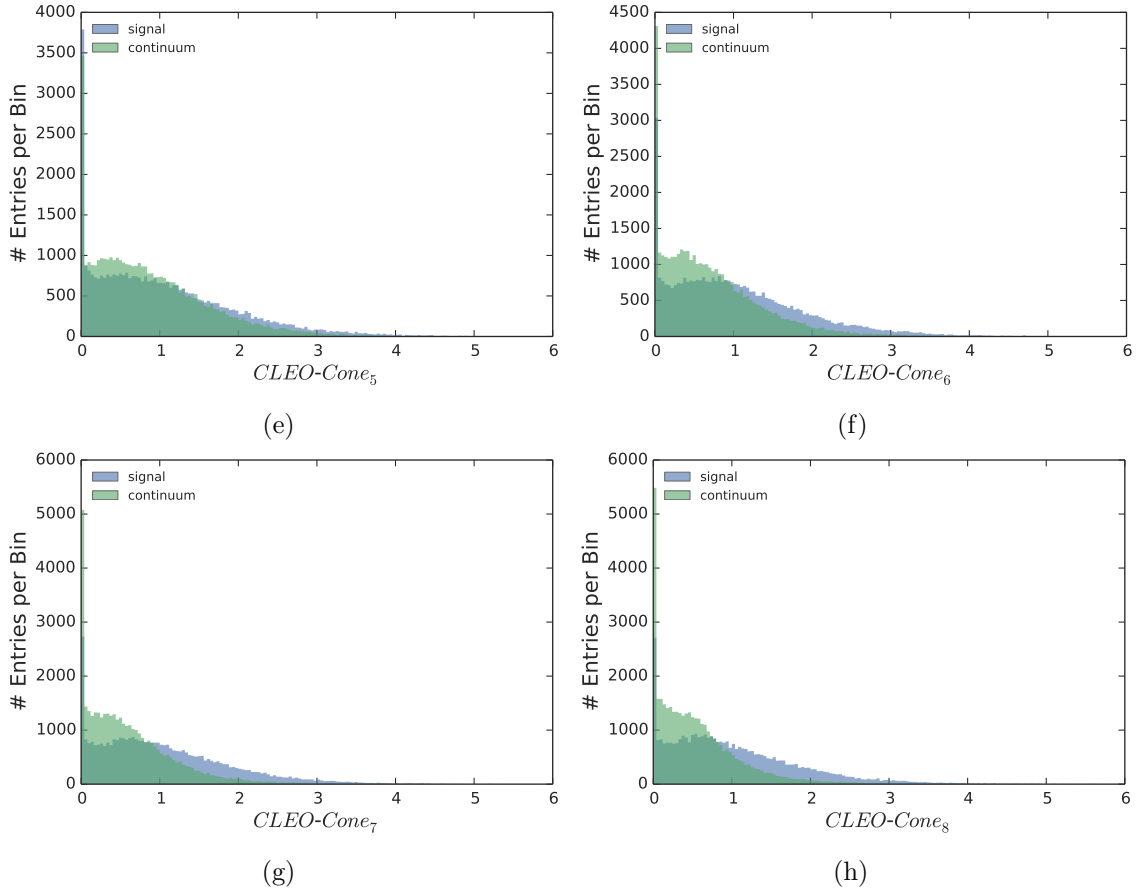


Figure B.3.: The distributions of the CLEO-Cones.

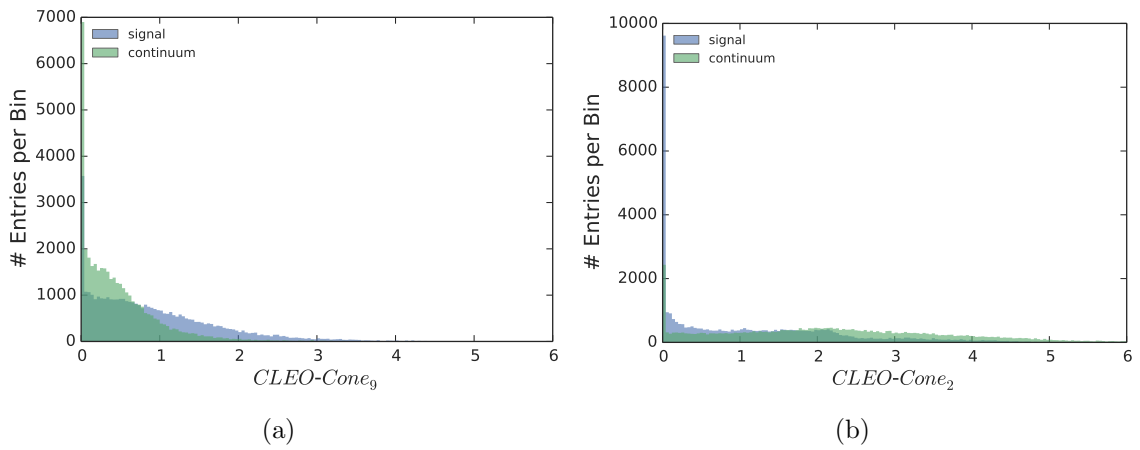


Figure B.4.: The distributions of the CLEO-Cones.

C. Fit to coherent Evtgen Model

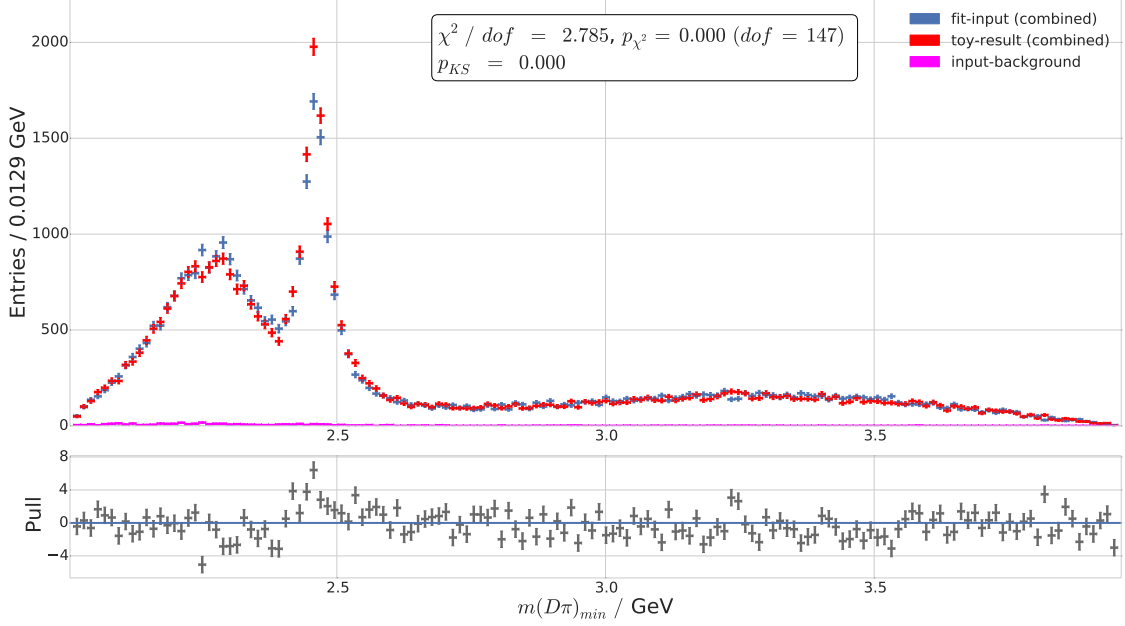
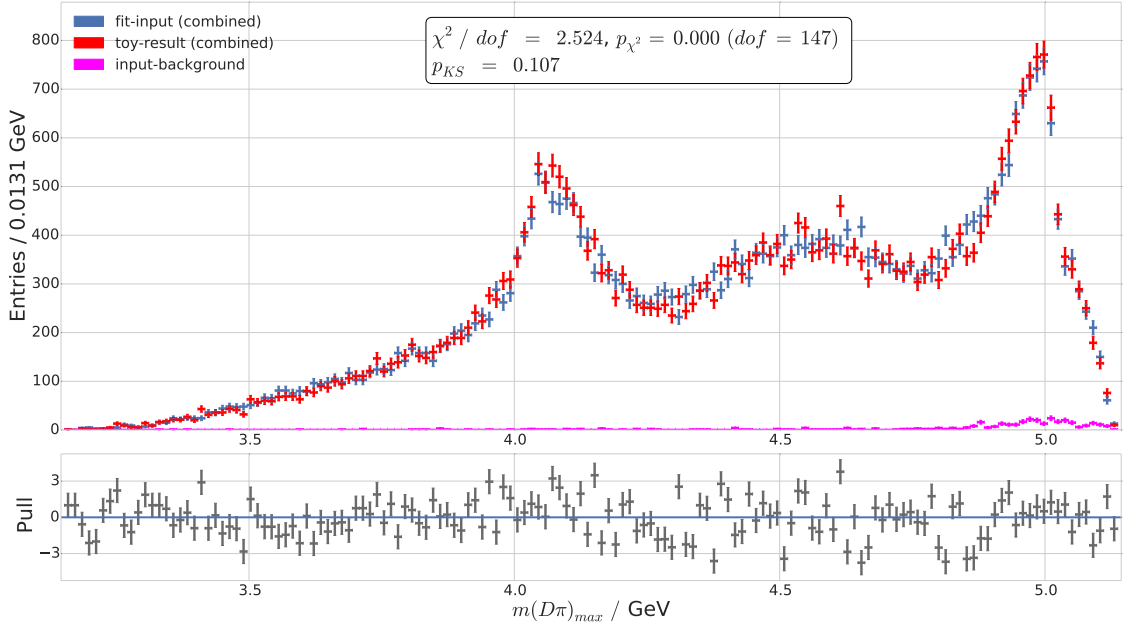
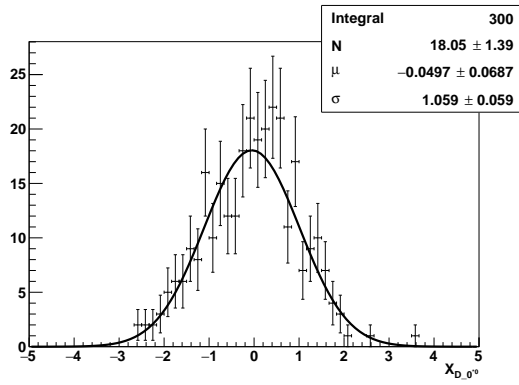
(a) Fit to D_0^{*0} signal MC in the $m(D\pi)_{min}$ -projection.(b) Fit to D_0^{*0} signal MC in the $m(D\pi)_{max}$ -projection.

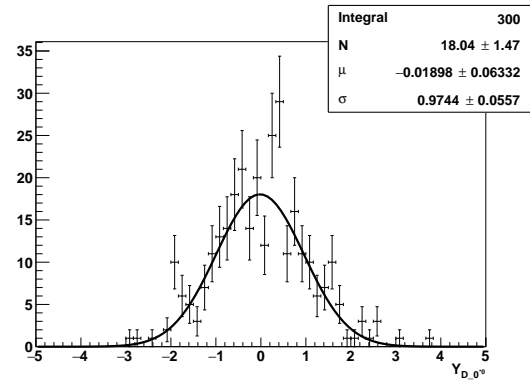
Figure C.1.: Fit to coherent signal MC projected into the kinematic variables $m(D\pi)_{min}$ and $m(D\pi)_{max}$. Note that in *Evtgen* the individual amplitudes were approximately normalized using Eq. 6.1 and that in *Laura++*'s fit model R_B was fixed to 0.0, as described in Chap. 11. Here all amplitudes are fixed at their PDG values and the phases are fixed at 0.0, only the signal yield is left floating. The upper panels show the fit input, the fit result in the form of toys, as well as misreconstructed signal candidates. The lower panels depict the pull distributions. The boxes overlaying the figures give the respective reduced χ^2 , its p-value p_{χ^2} as well as the p-value of a Kolmogorov-Smirnov-test, p_{KS} .

D. Results of Ensemble Tests

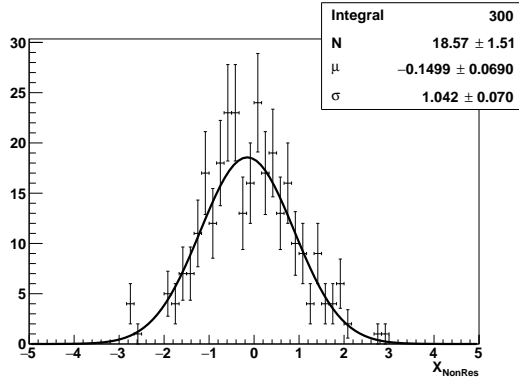
Pull distributions of a fit to a random model leaving all parameters of interest floating



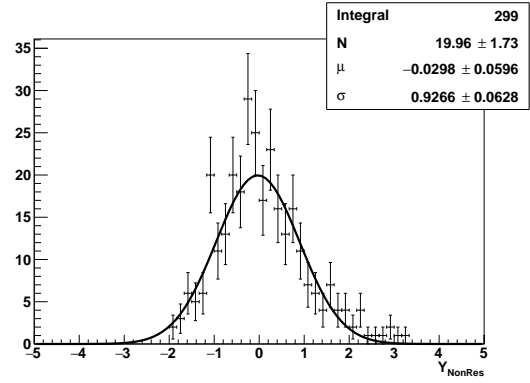
(a)



(b)



(c)



(d)

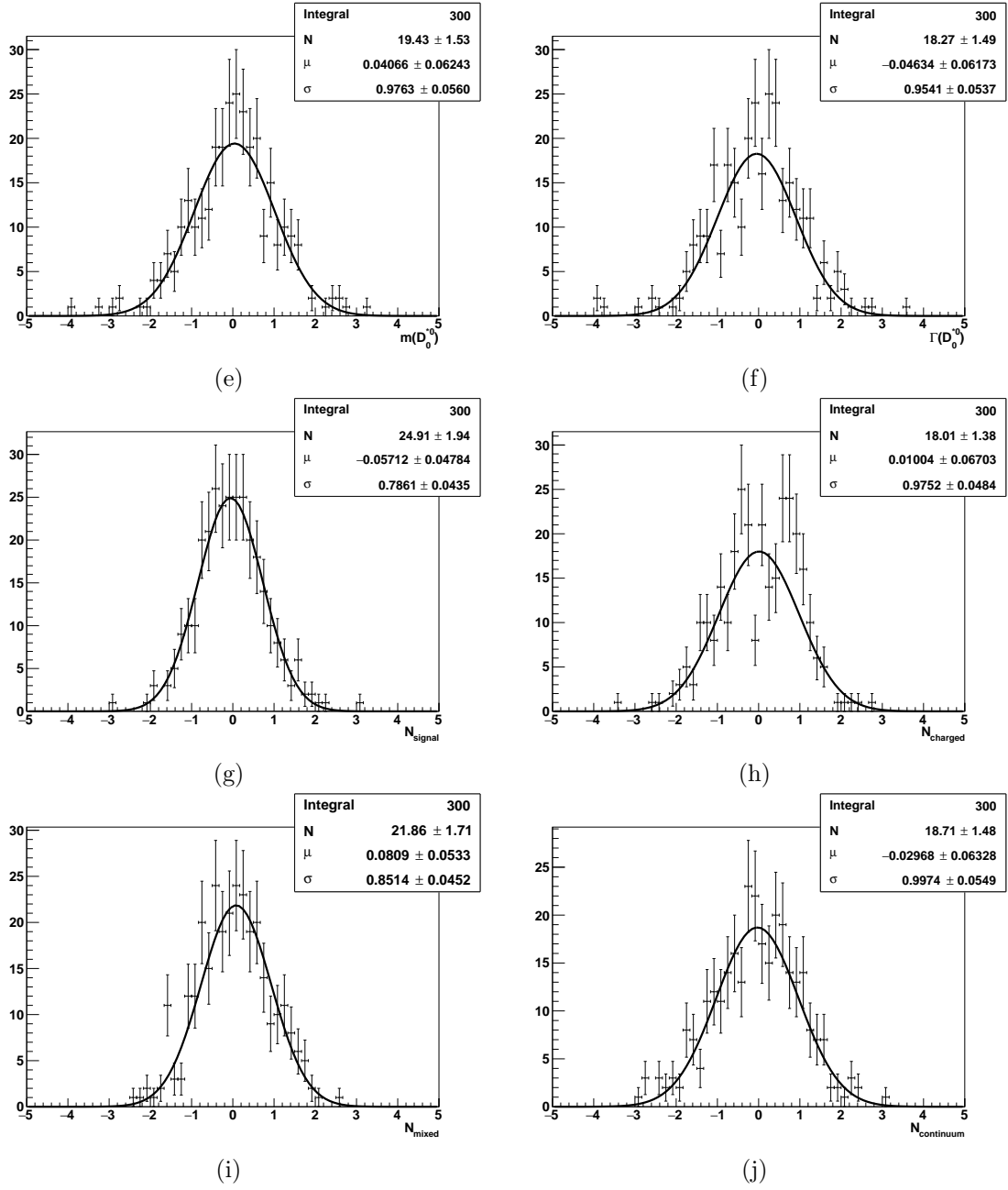


Figure D.1.: Illustration of the pull distributions of the floating parameters in the fit to a PDG random model, whose results are described in Tab. 12.4.

Fit results of different random models leaving all parameters of interest floating

<i>fit parameter</i>	<i># fits</i>	<i>true value</i> ¹	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	-0.599	-0.0044 ± 0.072	1.1 ± 0.069
$Y_{D_0^{*0}}$	300	0.89	0.037 ± 0.061	0.93 ± 0.052
X_{NonRes}	300	-0.724	0.093 ± 0.069	1.1 ± 0.06
Y_{NonRes}	300	0.26	-0.039 ± 0.061	0.98 ± 0.048
$m(D_0^{*0})$	300	2.32	0.094 ± 0.065	1 ± 0.062
$\Gamma(D_0^{*0})$	300	0.267	-0.082 ± 0.063	0.97 ± 0.052
N_{signal}	300	5279	-0.081 ± 0.055	0.86 ± 0.054
$N_{charged}$	300	433	-0.014 ± 0.064	1 ± 0.064
N_{mixed}	300	1546	0.017 ± 0.058	0.91 ± 0.062
$N_{continuum}$	300	506	-0.11 ± 0.066	1 ± 0.064

Table D.1.: Results of ensemble tests for a random model leaving all parameters of interest floating.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i> ¹	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	0.854	0.00011 ± 0.068	1 ± 0.053
$Y_{D_0^{*0}}$	300	0.247	-0.094 ± 0.059	0.96 ± 0.052
X_{NonRes}	300	0.32	-0.13 ± 0.053	0.87 ± 0.047
Y_{NonRes}	300	-0.219	0.034 ± 0.06	0.97 ± 0.055
$m(D_0^{*0})$	300	2.32	0.048 ± 0.059	0.92 ± 0.046
$\Gamma(D_0^{*0})$	300	0.267	-0.11 ± 0.057	0.9 ± 0.054
N_{signal}	300	5279	0.0073 ± 0.046	0.76 ± 0.038
$N_{charged}$	300	433	0.012 ± 0.058	0.91 ± 0.044
N_{mixed}	300	1546	0.087 ± 0.055	0.87 ± 0.049
$N_{continuum}$	300	506	-0.037 ± 0.062	1 ± 0.053

Table D.2.: Results of ensemble tests for another random model leaving all parameters of interest floating.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i> ¹	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	0.576	-0.086 ± 0.067	1 ± 0.059
$Y_{D_0^{*0}}$	300	0.566	0.14 ± 0.07	1 ± 0.058
X_{NonRes}	300	-0.105	-0.028 ± 0.061	0.97 ± 0.066
Y_{NonRes}	300	0.0227	0.099 ± 0.063	1 ± 0.058
$m(D_0^{*0})$	300	2.32	0.17 ± 0.06	0.96 ± 0.049
$\Gamma(D_0^{*0})$	300	0.267	0.0011 ± 0.063	1 ± 0.06
N_{signal}	300	5279	0.055 ± 0.051	0.83 ± 0.043
$N_{charged}$	300	433	0.025 ± 0.057	0.92 ± 0.046
N_{mixed}	300	1546	0.14 ± 0.053	0.84 ± 0.044
$N_{continuum}$	300	506	-0.14 ± 0.059	0.93 ± 0.045

Table D.3.: Results of ensemble tests for another random model leaving all parameters of interest floating.

¹ $m(D_0^{*0})$ and $\Gamma(D_0^{*0})$ values are given in GeV.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i> ¹	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	0.754	-0.098 ± 0.061	0.98 ± 0.05
$Y_{D_0^{*0}}$	300	-0.312	-0.054 ± 0.061	0.98 ± 0.048
X_{NonRes}	300	0.299	-0.11 ± 0.063	1 ± 0.051
Y_{NonRes}	300	0.696	-0.011 ± 0.064	1 ± 0.07
$m(D_0^{*0})$	300	2.32	0.034 ± 0.065	1 ± 0.062
$\Gamma(D_0^{*0})$	300	0.267	-0.078 ± 0.065	1.1 ± 0.057
N_{signal}	300	5279	-0.03 ± 0.058	0.92 ± 0.05
$N_{charged}$	300	433	0.017 ± 0.064	1 ± 0.061
N_{mixed}	300	1546	0.047 ± 0.056	0.83 ± 0.057
$N_{continuum}$	300	506	-0.1 ± 0.065	1 ± 0.059

Table D.4.: Results of ensemble tests for another random model leaving all parameters of interest floating.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i> ¹	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	300	-0.671	0.098 ± 0.063	0.95 ± 0.062
$Y_{D_0^{*0}}$	300	-0.25	0.023 ± 0.064	1 ± 0.054
X_{NonRes}	300	-0.0313	0.032 ± 0.064	0.96 ± 0.056
Y_{NonRes}	300	-0.975	0.03 ± 0.065	1 ± 0.055
$m(D_0^{*0})$	300	2.32	0.12 ± 0.06	0.97 ± 0.048
$\Gamma(D_0^{*0})$	300	0.267	-0.061 ± 0.067	1 ± 0.062
N_{signal}	300	5279	-0.035 ± 0.053	0.84 ± 0.043
$N_{charged}$	300	433	0.0027 ± 0.065	1 ± 0.056
N_{mixed}	300	1546	-0.0059 ± 0.056	0.89 ± 0.048
$N_{continuum}$	300	506	0.044 ± 0.066	1 ± 0.06

Table D.5.: Results of ensemble tests for another random model leaving all parameters of interest floating.

<i>fit parameter</i>	<i># fits</i>	<i>true value</i> ¹	<i>pull mean</i>	<i>pull width</i>
$X_{D_0^{*0}}$	299	0.622	-0.098 ± 0.06	0.96 ± 0.053
$Y_{D_0^{*0}}$	299	-0.559	0.028 ± 0.056	0.85 ± 0.043
X_{NonRes}	299	-0.207	-0.048 ± 0.06	0.9 ± 0.052
Y_{NonRes}	299	1.42	0.033 ± 0.061	0.97 ± 0.052
$m(D_0^{*0})$	298	2.32	-0.006 ± 0.074	1.1 ± 0.084
$\Gamma(D_0^{*0})$	298	0.267	-0.11 ± 0.062	0.96 ± 0.064
N_{signal}	299	5279	-0.016 ± 0.063	0.96 ± 0.057
$N_{charged}$	299	433	0.055 ± 0.067	1.1 ± 0.057
N_{mixed}	299	1546	0.019 ± 0.054	0.88 ± 0.052
$N_{continuum}$	299	506	-0.17 ± 0.061	0.97 ± 0.058

Table D.6.: Results of ensemble tests for another random model leaving all parameters of interest floating. In this case one of the 300 fits failed.

¹ $m(D_0^{*0})$ and $\Gamma(D_0^{*0})$ values are given in GeV.

Danksagung

Ich danke Herrn Prof. Dr. Michael Feindt herzlich für die Aufnahme in seiner Gruppe am Institut für experimentelle Kernphysik und das damit in mich gesteckte Vertrauen. Die Gespräche mit ihm, fachspezifischer Natur oder anderweitigen Bezuges, waren stets eine große Inspiration für mich.

Herrn Prof. Dr. Günter Quast möchte ich für die bereitwillige Übernahme des Korreferats danken.

Außerordentlicher Dank gilt Dr. Pablo Goldenzweig und Dr. Martin Heck, die stets Zeit fanden um mir mit ihrer Expertise beratend zur Seite zu stehen und mir wesentliche Hilfestellungen geleistet haben.

Der gesamten *B*-Gruppe gebührt mein Dank für die durchweg sehr angenehme und auch humorvolle Arbeitsatmosphäre, dank welcher ich mich hier stets sehr wohl gefühlt habe. Im Besonderen erwähnen möchte ich Dr. Manuel Heider, der mich geduldig und kompetent in das Thema meiner Arbeit eingeführt hat und ohne den die Arbeit sicherlich weit weniger erfolgreich verlaufen wäre. Auch mein Zimmerkollege Johannes Grygier war häufig Adressat von Fragen bezüglich Hindernissen aller Art, für deren Beantwortung ich meinen Dank aussprechen möchte.

Meinen Eltern, Andreas und Claudia Kohl, möchte ich meinen Dank für ihre bedingungslose Unterstützung während meiner Studienzeit aussprechen.

Eigenständigkeitserklärung

Hiermit versichere ich die vorliegende Arbeit eigenständig verfasst und nur die angegebenen Hilfsmittel verwendet zu haben.

Simon Kohl

Karlsruhe, den 17.06.2016.