
Background Analysis for B^0 Lifetime and Oscillation Frequency Measurement using Hadronic $B^0 \rightarrow D^{(*)-} \pi^+$ Channels at Belle II

Caspar Schmitt



Munich 2021

**Hintergrund-Analyse zur Messung
der B^0 Lebenszeit und
Oszillationsfrequenz
mit hadronischen $B^0 \rightarrow D^{(*)-} \pi^+$
Zerfallskanälen bei Belle II**

Caspar Schmitt

Master Thesis
at the Faculty of Physics
of the Ludwig-Maximilians-Universität
Munich

submitted by
Caspar Schmitt
from Munich

Munich, 13 September 2021

Supervisor: Prof. Dr. Christian Kiesling

Institute: Max Planck Institute for Physics (Werner Heisenberg Institute)

Abstract

To explain the origin of the observed matter-antimatter asymmetry in the universe, new sources of charge-parity (CP) symmetry violation beyond the standard model of particle physics are required. The oscillation frequency Δm in the neutral B meson system is an important parameter in overconstraining standard model predictions for CP violation in the quark sector.

We aim for a precision measurement of the lifetime τ_{B^0} and oscillation frequency Δm at the Belle II experiment. At the asymmetric electron-positron collider SuperKEKB, $B^0\bar{B}^0$ meson pairs are produced in an entangled coherent quantum state. We reconstruct B mesons in three hadronic signal decay channels $B^0 \rightarrow \pi^+ D^{(*)-}$ and employ flavor tagging algorithms to determine the flavor eigenstate of the accompanying B meson decay. From a time-dependent fit to the mixing asymmetry in the decay time difference of both B mesons, in seven bins of the flavor tag figure-of-merit, we extract the oscillation frequency and lifetime.

From simulation, we analyse the (peaking) background compositions in all three signal channels. We develop multi-dimensional fit strategies to separate signal and backgrounds and to extract the background composition from Belle II data. We fit in the energy difference ΔE of the reconstructed B mesons and the output of a continuum suppression boosted decision tree, trained on simulated event samples. In simulation the fit consistently tells apart signal and backgrounds with deviations of less than 10%. Pseudo-experiment indicate a stable and un-biased fit and the fit result on $176.9 \pm 12.6 \text{ fb}^{-1}$ of Belle II data agrees with the expectations from simulation.

Kurzfassung

Um den Ursprung der im Universum beobachteten Materie-Antimaterie Asymmetrie zu erklären, sind neue CP (Ladungs-Paritäts-Symmetrie) verletzende Quellen, jenseits des Standardmodells der Teilchenphysik, notwendig. Die Oszillationsfrequenz Δm im neutralen B-Meson System ist ein wichtiger Parameter, um die Beschreibung von CP-Verletzung im Quark-Sektor des Standardmodells zu überprüfen.

Wir beabsichtigen eine Präzisionsmessung der Lebenszeit τ_{B^0} und der Oszillationsfrequenz Δm am Belle II Experiment. Am asymmetrischen Elektron-Positron-Collider SuperKEKB werden $B^0\bar{B}^0$ Mesonpaare in verschränkten kohärenten Quantenzuständen produziert. Wir rekonstruieren B Mesonen in drei hadronischen Signalzerfallskanälen $B^0 \rightarrow \pi^+ D^{(*)-}$ und bestimmen den Flavor-Eigenzustand des begleitenden B Mesonerfalls mit flavor-tag Algorithmen. Aus einem zeitabhängigen Fit der Mischungsasymmetrie als Funktion der Zerfallszeitdifferenz beider B Mesonen, in sieben bins der flavor-tag figure-of-merit, kann die Oszillationsfrequenz und Lebenszeit bestimmt werden.

Wir untersuchen die Hintergrundzusammensetzung in allen drei Signalkanälen in simulierten Teilchenkollisionen. Mehrdimensionale Fitstrategien wurden entwickelt, um Signal und Hintergründe zu trennen und die Hintergrundzusammensetzung aus Belle II Daten zu bestimmen. Der Fit wird in der Energiedifferenz ΔE der rekonstruierten B Mesonen und im Output eines boosted decision trees zur Kontinuumsunterdrückung, trainiert auf Simulation, durchgeführt. In simulierten Kollisionen trennt der Fit Signal und Hintergründe konsistent mit Abweichungen von unter 10%. Pseudo-Experimente weisen auf einen stabilen und unvoreingenommenen Fit hin. In $176.9 \pm 12.6 \text{ fb}^{-1}$ Belle II Daten, stimmt das Fit-Ergebnis mit den Erwartungen aus der Simulation überein.

Contents

1	Introduction	7
2	Theory	8
2.1	The Standard Model of Particle Physics	8
2.2	Flavor Physics	9
2.2.1	CP Violation	9
2.2.2	Neutral Meson Mixing	12
2.2.3	Neutral B^0 Meson Mixing	14
3	Experiment	15
3.1	SuperKEKB Collider	16
3.2	Belle II Detector	17
3.3	Data Simulation, Processing and Analysing	19
4	Analysis Strategy for Lifetime and Mixing Measurement	20
5	Sample Processing and Reconstruction	28
5.1	Simulation Samples	28
5.2	Data Samples	28
5.3	Reconstruction	29
5.3.1	Track Reconstruction	29
5.3.2	Signal B meson Reconstruction	29
6	Background Analysis for Lifetime and Mixing Measurement	31
6.1	Decay Categorization Procedure	31
6.2	Decay Categorization in $B^0 \rightarrow \pi^+ D^- (\pi^- \pi^- K^+)$	35
6.3	Decay Categorization in $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$	38
6.4	Decay Categorization in $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- K^+))$	40
7	Fit Strategy to Separate Backgrounds and Signal	41
7.1	Simplified Decay Categorization	42
7.2	Event Variables for Fit	43
7.2.1	Continuum Suppression <i>Boosted Decision Tree</i> output (BDT)	44
7.2.2	Correlations between ΔE and BDT	44
7.2.3	Correlations between Δt and BDT	45
7.3	Extraction of Fit Shapes from Simulation Samples	48
7.3.1	Extraction of Fit Shapes in Continuum Suppression BDT	48
7.3.2	Extraction of Fit Shapes in ΔE	48
7.4	Two-dimensional fits in $(\Delta E, \text{BDT})$	51
7.4.1	Fit in Full Simulation Sample	52
7.4.2	Fit in Full Data Sample	53
7.4.3	Stability of Fit in Pseudo-Experiments	54
7.5	One-dimensional Fit in ΔE	60
7.5.1	Fit in Full Simulation Sample	61
7.5.2	Fit in Full Data Sample	61
7.5.3	Stability of Fit in Pseudo-Experiments	62
8	Lifetime and Mixing Measurement	66
8.1	Time-dependent Fits	66
8.2	Integration of Background Fit in Time-Dependent Fit	66
9	Conclusion and Outlook	69

A	Categorization Tool for Reconstructed Decays	71
B	Correlations between ΔE and M_{bc}	71
C	Extracted Shapes for Two-Dimensional Fit in all Dilution Bins for Same- and Opposite-Flavor Events	74
D	Extracted Shapes for One-Dimensional Fit in all Dilution Bins for Same- and Opposite-Flavor Events	88

1 Introduction

Astrophysical experiments measure a matter-antimatter asymmetry in the universe. In order to generate such an asymmetry from an initially symmetric universe, the so-called charge-parity symmetry has to be broken in particle decays (CP violation). The standard model of particle physics incorporates CP violation, but not to an extent necessary to explain the asymmetry. Our understanding of physics at the most elementary level is hence incomplete and we are searching for new physics processes that violate the CP symmetry to the observed magnitude during the evolution of the early universe.

The study of B mesons is a prime candidate in the search for new sources of CP violation. Mesons are particles composed of elementary quarks, more precisely of one quark and one anti-quark. Neutral mesons can oscillate into their antiparticles, a phenomenon which we call *mixing*. A precise measurement of the neutral B meson lifetime and oscillation frequency is an important result towards searching for new CP-violating processes.

At the *SuperKEKB* collider in Japan, large quantities of B meson pairs are produced. The particles and their decays are observed by the *Belle II* detector. The collider provides kinematically well-known collisions in an exceptionally clean environment. The Belle II detector is optimized to enable high-precision, high-statistics measurements of fundamental particle properties and physics parameters, such as CP-violating parameters.

In this thesis, we aim for a precision measurement of the lifetime and oscillation frequency in the neutral B meson system at the *Belle II* experiment. The thesis is structured as follows.

In Chapter 2, we briefly review our current understanding of particle physics. Special emphasis is placed on the discussion of *flavor physics* and we give a short historic retrospect of the development of the quark model. We refer to astrophysical experiments which measure a matter-antimatter asymmetry in the universe and discuss the discrete P , C and CP asymmetries and how they are essential for the formation of a matter-antimatter asymmetry. We then review the present status of CP-measurements in the quark sector and how we can search for new sources of CP violation.

We subsequently derive the formalism for neutral meson mixing from first principles and apply it to the neutral B meson system. We then see that the oscillation frequency can be extracted from a *time-dependent measurement of the mixing asymmetry*.

In Chapter 3, we discuss the production of B mesons at the *SuperKEKB collider* and the detection principles of the Belle II detector. Furthermore, we present the general workflow of an analysis at Belle II.

Chapter 4 presents our analysis strategy in detail. We summarize the previously introduced concepts and explain how we employ them for a precision measurement of the B meson lifetime and oscillation frequency. The explanations focus on the treatment of backgrounds and in particular on so-called *peaking backgrounds*. We review previous measurements and explain how our analysis strategy is expected to perform better in background suppression, resulting in decreased measurement errors on the lifetime and oscillation frequency.

Chapter 5 discusses the reconstruction of detector responses into physics processes. Thereafter a detailed analysis of (peaking) backgrounds is presented in Chapter 6, which was partly omitted in previous analyses. This background analysis serves to understand the contributing backgrounds and is important to assess their impact on the measurement.

Based on the findings from the detailed background analysis, we develop a multi-dimensional

fit strategy to separate backgrounds and signal (Chapter 7). The fit performance and stability is assessed in simulation and in so-called pseudo-experiments. Thereafter, we apply the fits to real Belle II data. We compare the fit performance with a one-dimensional fit model, such as employed in previous analyses.

In Chapter 8 the full analysis is implemented, which consists of the fit to separate signal and backgrounds and of a time-dependent fit of the mixing asymmetry. Chapter 9 concludes the thesis.

2 Theory

The *Standard Model of particle physics* (SM) describes our understanding of the elementary building blocks of matter and how they interact on the microscopic level. This chapter briefly describes how we came to our current understanding of particle physics and why it is incomplete. Subsequently, the B^0 meson mixing behavior is described. We here focus only on the physics that is relevant for the measurement of the B^0 lifetime and mixing, and in no way aim to give a complete description of the SM.

2.1 The Standard Model of Particle Physics

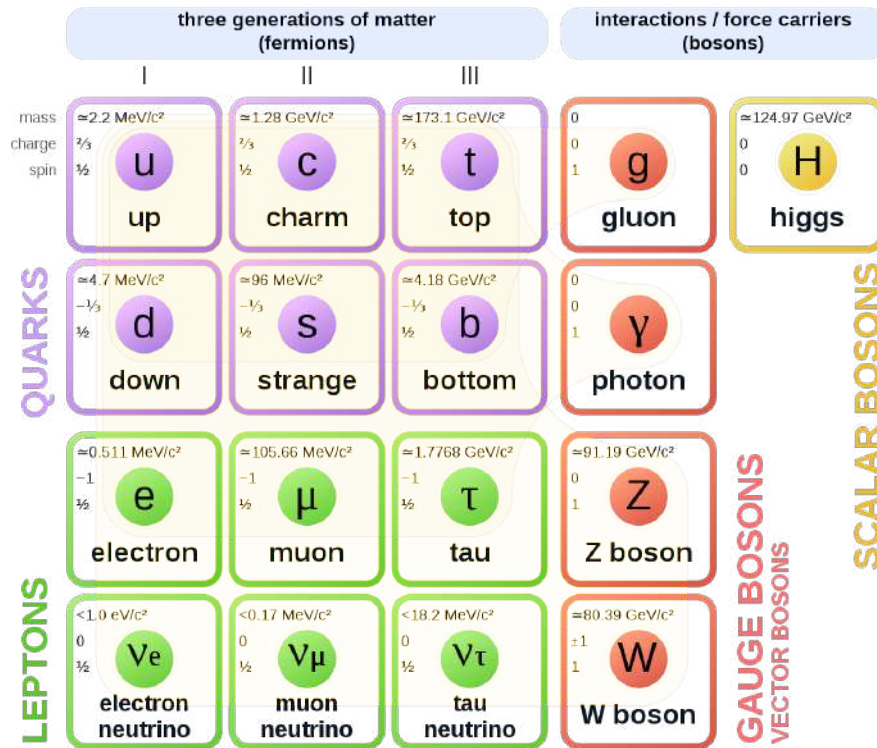


Figure 1: Particle content of the Standard Model of particle physics [1].

Figure 1 shows the elementary particles in the SM, which are all pointlike. All matter is built up of spin 1/2 fermions. They consist of 6 flavors of leptons (green in Figure 1) and of 6 flavors of quarks (purple). We categorize them by their electric charge q into charged and neutral leptons, as well as in up- ($q = 2/3$) and down-type ($q = -1/3$) quarks. Furthermore, we distinguish three generations of fermions which differ by their masses, but have the same spin and charge.

The fermions interact through the following fundamental forces of the universe: the electromagnetic and weak forces (unified as electro-weak force) and the strong force. In the context of quantum field theories (QFTs), the forces are mediated by the exchange of spin 1 gauge bosons (red in Figure 1) and of the spin 0 Higgs boson. The gauge bosons arise in interaction terms, when imposing the observed $U(1)_Y \times SU(2)_L \times SU(3)_C$ symmetries on the physical predictions [2]. By the *Noether theorem*, each continuous symmetry is associated with a conserved charge [3]. The gauge bosons couple to the charge specific for the force they mediate. All charged fermions interact via the photon and interact weakly via the W , Z and Higgs bosons. The quarks in addition interact strongly via the gluons.

The gauge bosons and fermions obtain their masses through interacting with the scalar field associated with the spin 0 Higgs boson. The latter is the final particle of the SM, which was discovered most recently in 2012 at the Large Hadron Collider (LHC) [4][5].

2.2 Flavor Physics

In this thesis, we focus on how the fermions differ in their interaction properties, a field called *flavor physics*. This section follows the explanations in [2] and [6].

Historically, the up- and down-quark were first known as constituents of neutrons and protons in atomic nuclei. The discovery of strange particles involving strange-quarks lead to a number of surprising observations. Initially, due to their unusually long lifetimes, strangeness was introduced as flavor quantum number. Today we know that strangeness is conserved in strong decays and can only change in weak decays. To account for the suppression of flavor-changing decays, Cabibbo suggested that the up-quark couples to a superposition of the down-type quarks d and s

$$d \cdot V_{ud} + s \cdot V_{us} = d \cdot \cos \theta_c + s \cdot \sin \theta_c, \quad (1)$$

where the weak couplings V are expressed in terms of the Cabibbo mixing angle θ_c . He was able to determine $\theta_c \approx 13^\circ$ from the relative branching fractions of leptonic decays of kaons $K^+ = |u\bar{s}\rangle$ and pions $\pi^+ = |u\bar{d}\rangle$ [7]. However this could not account for the strong suppression of flavor changing neutral currents (FCNCs) in decays of strange particles, for example in

$$\frac{\Gamma(K^+ \rightarrow \pi^+ l^- l^+)}{\Gamma(K^+ \rightarrow \pi^0 l^+ \nu_l)} \approx 10^{-6}. \quad (2)$$

The problem was resolved by the postulation of the charm-quark. Glashow, Iliopoulos and Maiani (GIM) proposed a fourth quark which couples to the down-quark with same magnitude but opposite sign as the strange-quark couples to the up-quark [8]. The weak and strong eigenstates are hence related by

$$\begin{bmatrix} d \\ s \end{bmatrix}_{\text{weak}} = \begin{pmatrix} V_{ud} & V_{us} \\ V_{cd} & V_{cs} \end{pmatrix} \begin{bmatrix} d \\ s \end{bmatrix}_{\text{strong}} = \begin{pmatrix} \cos \theta_c & \sin \theta_c \\ -\sin \theta_c & \cos \theta_c \end{pmatrix} \begin{bmatrix} d \\ s \end{bmatrix}_{\text{strong}}. \quad (3)$$

In the so-called *GIM mechanism*, the combined processes result in a strong suppression of FCNCs. The observed remaining FCNCs indicate the non-degeneracy of the masses $m_c \neq m_s$.

2.2.1 CP Violation

Another unexpected observation was made in decays of strange particles: in the so-called $\theta - \tau$ puzzle, two particles of the same mass decayed to 2 and 3 pions respectively. These are fundamentally different modes, since they are of opposite symmetries under so-called *parity transformation* P . P acts on any wavefunction ψ as $P\psi(t, \vec{x}) = P\psi(t, -\vec{x})$ and hence

$$\begin{aligned} P|\pi\pi\pi\rangle &= P(\pi)^3 \cdot (-1)^l |\pi\pi\pi\rangle = -(-1)^l |\pi\pi\pi\rangle \\ P|\pi\pi\rangle &= P(\pi)^2 \cdot (-1)^l |\pi\pi\rangle = +(-1)^l |\pi\pi\rangle, \end{aligned} \quad (4)$$

where l is the total angular momentum of the decaying particle and $P(\pi) = -1$ [2]. According to Lee and Yang, θ and τ were the same particle (today known as K^+) that weakly decayed to both 2 and 3 pions, thus violating parity in the weak force [9]. Shortly afterwards, Wu confirmed their idea experimentally in β decays [10].

Parity P was only one of the discrete symmetries, that the laws of physics were believed to exhibit. *Charge conjugation* C inverts all additive quantum numbers and thus interchanges particles with anti-particles. Although C and P were maximally violated, nature was expected to be symmetric under the combined CP transformation. The observation of a small degree of CP symmetry violation (CPV) in kaon decays [11], motivated Kobayashi and Maskawa (KM) to extend GIM's four-quark mixing matrix by another generation of quarks [12]. Analogously to strangeness, the flavor quantum numbers Charm, Bottom and Top (although named differently at the time) were introduced, that are not conserved in weak decays. As will be shown in the following, the postulation of a third quark generation makes it possible to introduce CPV in weak quark decays.

Typically, the 3×3 CKM mixing matrix is parametrized by three mixing angles $\theta_{12}, \theta_{13}, \theta_{23}$ and one complex phase δ

$$\begin{bmatrix} d \\ s \\ b \end{bmatrix}_{\text{weak}} = V_{\text{CKM}} \begin{bmatrix} d \\ s \\ b \end{bmatrix}_{\text{strong}}$$

with
$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix},$$

where $c_{ij} \equiv \cos \theta_{ij}$ and $s_{ij} \equiv \sin \theta_{ij}$. In any QFT, we require hermitian Lagrangians $\mathcal{L} = \mathcal{L}^\dagger$, and thus hermitian conjugate terms in \mathcal{L} , in order to make measurable predictions. In the SM, the Lagrangian and its CP conjugate for a weak decay of an up-type quark $q \rightarrow q'W^+$ are given by

$$\begin{aligned} \mathcal{L} &\propto V_{qq'} \bar{q}_L \gamma^\mu q'_L W_\mu^+ + V_{qq'}^* \bar{q}'_L \gamma^\mu q_L W_\mu^- \\ \mathcal{L} &\xrightarrow{CP} V_{qq'} \bar{q}'_L \gamma^\mu q_L W_\mu^- + V_{qq'}^* \bar{q}_L \gamma^\mu q'_L W_\mu^+, \end{aligned} \quad (5)$$

where $V_{qq'}$ are the CKM mixing matrix elements. From Equation 5, it is clear that the complex phase δ in the CKM matrix induces CPV. While the complex phase in the CKM matrix is irreducible, one can parametrize the 2×2 GIM mixing matrix such that it includes no complex phase and hence induces no CPV. Thus by extending the quark-model to three generations, KM introduced a mechanism for CPV.

Usually, the CKM matrix is expressed in terms of the Wolfenstein parametrization [13], to reveal its hierarchic structure $V_{ud} \gg V_{us} \gg V_{ub}$.

$$V_{\text{CKM}} = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \quad \text{with} \quad \begin{cases} \lambda \equiv s_{12} \\ A \equiv s_{23}/\lambda^2 \\ A\lambda^3(\rho - i\eta) \equiv s_{13}e^{-i\delta} \end{cases} \quad (6)$$

The *GIM mechanism* is implemented in the 6-quark model by the unitarity of the CKM matrix $V_{\text{CKM}} V_{\text{CKM}}^\dagger = \mathbb{1}$. The unitarity condition yields six independent relations for the off-diagonal zeros of the unit matrix, e.g. the following relation involving quarks of the third generation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \quad (7)$$

which can be visualised as closing triangles in the complex plane (Figure 2). In these so-called *unitarity triangles*, non-trivial angles $\phi \neq 0, \pi$ indicate CPV. While all triangles have the same

area, their sides are of different lengths. Inserting the Wolfenstein parametrization in Equation 7, we find that for this particular triangle all sides are of same magnitude $\mathcal{O}(\lambda^3)$. This results in large angles and hence large CPV in systems, where the b-quarks are involved.

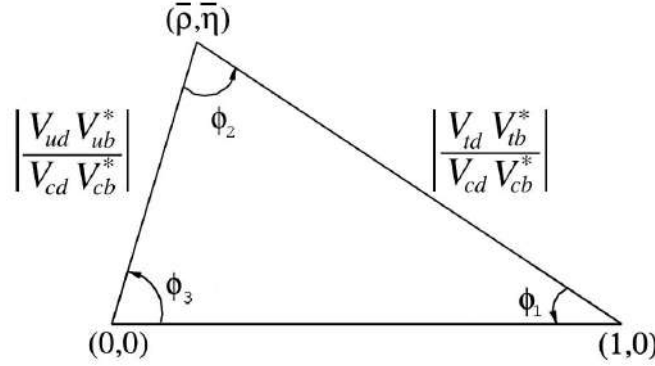


Figure 2: Unitarity triangle corresponding to Equation 7.

$$(\bar{\rho}, \bar{\eta}) = (\text{Re } z, \text{Im } z) \text{ with } z \equiv -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \quad [14].$$

After the discovery of the third-generation bottom-quark at Fermilab [15], a measurement of CP-violating effects in the B mesons was left to confirm the KM mechanism. With the observation of long B meson lifetimes and substantial mixing rates in neutral B^0 mesons [16], two necessary conditions for large CPV were met in the B meson system. This is due to the fact that the *GIM suppression* of FCNCs in B^0 mixing is weakened by the large top-quark mass. The CKM elements in Equation 7 correspond to B meson decays and the large angles in the corresponding unitarity triangle reflect large CPV. Consequently, a great effort was made to construct the first generation of high-luminosity B factories *KEKB* and *PEP* with their detectors, which eventually measured [17][18]

$$\sin(\phi_1) = \sin\left(\arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right]\right) = 0.699 \pm 0.017 \quad [19]. \quad (8)$$

Although the SM can explain most observations, we know that it is not complete. Among others, it does not account for the baryon-antibaryon (= matter-antimatter) asymmetry in the universe. Satellite-based experiments measure no primary cosmic anti-proton or anti-helium fluxes and non-sufficient γ fluxes from annihilation [20][21]. From anisotropies in the cosmic microwave background (CMB) and measurements of light element abundances [22][23], we know that the universe exhibits a baryon asymmetry

$$\eta_B \equiv \frac{n_B - n_{\bar{B}}}{n_{\text{CMB}}} \approx 6 \cdot 10^{-10} \neq 0, \quad [24] \quad (9)$$

where n denote the number densities of (anti-) baryons and CMB-photons. According to Sakharov, three conditions have to be met to generate a non-zero baryon asymmetry from an initially symmetric universe [25]. First, the universe must be in thermal non-equilibrium. To illustrate the two remaining conditions, consider the example of a particle X decaying to only two final states. The decay rates and their CP conjugates are given by

$$\begin{aligned} \Gamma(X \rightarrow f_1) &= r \xrightarrow{CP} \Gamma(\bar{X} \rightarrow \bar{f}_1) = \bar{r} \\ \Gamma(X \rightarrow f_2) &= 1 - r \xrightarrow{CP} \Gamma(\bar{X} \rightarrow \bar{f}_2) = 1 - \bar{r} \end{aligned} \quad \implies \Delta N_B = (r - \bar{r}) (N_B^{(1)} - N_B^{(2)}), \quad (10)$$

where $N_B^{(1,2)}$ are the baryon numbers of the two final states and ΔN_B is the total change in baryon number for an equal number of X and \bar{X} decays. Clearly, for a baryon asymmetry $\Delta N_B \neq 0$, we

need both baryon number violating $N_B^{(1)} - N_B^{(2)} \neq 0^1$ and CP violating $r - \bar{r} \neq 0$ decays [6].

In the SM, the source of CPV is given by the complex phase of the CKM matrix [26]. However, the induced degree of baryon asymmetry $\eta_B^{\text{SM}}/\eta_B \approx 10^{-16} \ll 1$ is far from being able to explain the observed asymmetry [27]. Hence, we search for additional sources of CPV in *new physics beyond the SM* (NP). Two complementary approaches are pursued in the search for NP: *direct searches* at high energies to produce new particles (at LHC) and *indirect searches* at the high-intensity frontier. At the second-generation B factory *SuperKEKB* with *Belle II*, deviations from SM parameter predictions are searched in high precision analyses. One approach is to overconstrain CKM-matrix-unitarity by precisely measuring all sides and angles of the unitarity triangles. Figure 3 shows the current experimental limits.

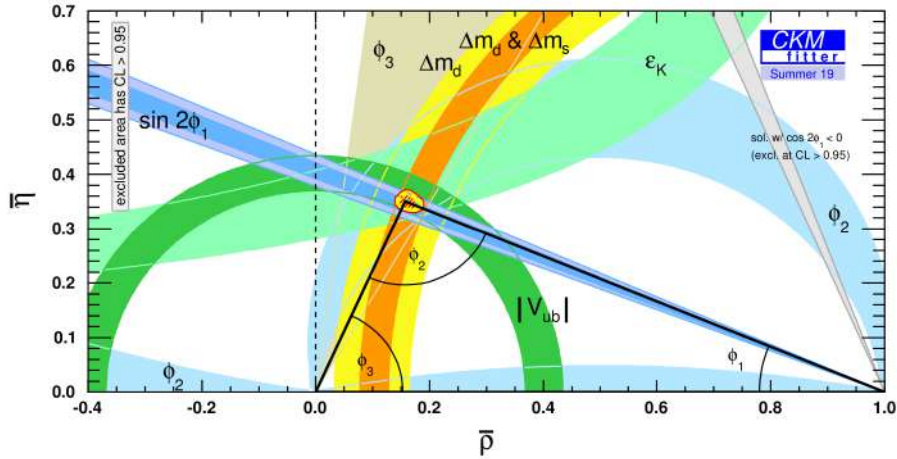


Figure 3: Unitarity triangle with current experimental limits [28].

2.2.2 Neutral Meson Mixing

An important parameter for overconstraining CKM-matrix-unitarity is the mixing parameter Δm (i.e. oscillation frequency) in the neutral B_d^0 meson system. *Mixing* describes the phenomenon of neutral mesons oscillating into their antiparticles and will here be discussed in detail. In this section, we describe the mixing behavior of generic neutral mesons, here denoted P^0 . Thereafter, we specify the results for neutral B^0 meson mixing.

Due to mixing, neutral mesons propagating in time must be treated as superposition of the meson and antimeson states, described by the wavefunctions $|P^0(t)\rangle$ and $|\bar{P}^0(t)\rangle$ respectively. The time-evolution of the neutral meson/anti-meson state is governed by the Schrödinger equation (Equation 11).

$$i \frac{d}{dt} \begin{bmatrix} |P^0(t)\rangle \\ |\bar{P}^0(t)\rangle \end{bmatrix} = \underbrace{\begin{pmatrix} \mathbf{M} & \\ & -i\frac{\mathbf{\Gamma}}{2} \end{pmatrix}}_{=\mathcal{H}_{\text{eff}}} \begin{bmatrix} |P^0(t)\rangle \\ |\bar{P}^0(t)\rangle \end{bmatrix} = \left[\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \right] \begin{bmatrix} |P^0(t)\rangle \\ |\bar{P}^0(t)\rangle \end{bmatrix}, \quad (11)$$

where the effective Hamiltonian \mathcal{H}_{eff} describes neutral meson mixing and in general consists of hermitian matrices \mathbf{M} and $\mathbf{\Gamma}$, describing the mass and decay rate components [6]. In the Schrödinger Picture formulation of quantum mechanics, the time-dependence is absorbed in the meson/anti-meson state and \mathcal{H}_{eff} is time-independent.

¹It is worth noting that baryon number conservation can be violated in the SM [23].

Neutral meson mixing denotes transitions between the flavor eigenstates P^0 and \bar{P}^0 and vice versa, which are governed by weak flavor-changing interactions. While the weak interaction operates on the flavor eigenstates, the mesons propagate in their physical mass eigenstates, which are given by the eigenstates of the Hamiltonian. Non-zero off-diagonal components of \mathcal{H}_{eff} induce mixing, which results in the flavor eigenstates to not coincide with the propagating physical mass eigenstates.

Since CPT (with time reversal $T(t, \vec{x}) = (-t, \vec{x})$) invariance implies $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$, solving Equation 11 for the time evolution amounts to finding the eigenstates of

$$\mathcal{H}_{\text{eff}} = \begin{pmatrix} M_{11} - \frac{i}{2}\Gamma_{11} & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M_{11} - \frac{i}{2}\Gamma_{11} \end{pmatrix} \equiv \begin{pmatrix} a & b \\ c & a \end{pmatrix}. \quad (12)$$

We find the heavy (H) and light (L) mass eigenstates

$$\begin{aligned} |P_{H,L}\rangle &= \sqrt{\frac{b}{c+b}} |P^0\rangle \mp \sqrt{\frac{c}{c+b}} |\bar{P}^0\rangle \\ &\equiv p |P^0\rangle \mp q |\bar{P}^0\rangle, \end{aligned} \quad (13)$$

normalized such that $|p|^2 + |q|^2 = 1$ and with eigenvalues

$$m_{H,L} - \frac{i}{2}\Gamma_{H,L} = a \mp \sqrt{b \cdot c}. \quad (14)$$

The resulting time-evolution for the mass eigenstates is

$$|P_{H,L}(t)\rangle = \exp\left[-i\left(m_{H,L} - \frac{i}{2}\Gamma_{H,L}\right)t\right] \cdot |P_{H,L}\rangle. \quad (15)$$

By the Born rule, the square of the magnitude of the wave function corresponds to the probability density for detecting a particle at a specific point in time and space. Since \mathcal{H}_{eff} is not hermitian, the eigenstates are non-orthogonal with complex eigenvalues. This leads to a non-unitary time-evolution of the eigenstates, which means that probability is not conserved here. This is due to the fact, that both P^0 and \bar{P}^0 eventually decay into other particles and disappear from the (P^0, \bar{P}^0) subspace [6].

From Equations 13 - 15, we can find the time-evolution for the flavor eigenstates

$$\begin{aligned} |P^0(t)\rangle &= \frac{1}{2p} [|P_H(t)\rangle + |P_L(t)\rangle] \\ &= \frac{1}{2} e^{-iat} e^{-i\sqrt{bc}t} \cdot \left[\left(1 + e^{+2i\sqrt{bc}t}\right) |P^0\rangle + \frac{q}{p} \left(1 - e^{+2i\sqrt{bc}t}\right) |\bar{P}^0\rangle \right] \\ &= \left[g_+(t) |P^0\rangle + \frac{q}{p} g_-(t) |\bar{P}^0\rangle \right] \quad \text{and} \\ |\bar{P}^0(t)\rangle &= \frac{1}{2q} [|P_L(t)\rangle - |P_H(t)\rangle] = \frac{p}{q} g_-(t) |P^0\rangle + g_+(t) |\bar{P}^0\rangle, \end{aligned} \quad (16)$$

where the time-dependence is encoded in

$$\begin{aligned} g_{\pm}(t) &\equiv \frac{1}{2} e^{-iat} e^{-i\sqrt{bc}t} \left(1 \pm e^{+2i\sqrt{bc}t}\right) \\ &= \frac{1}{2} e^{-i(m_L - \frac{i}{2}\Gamma_L)t} \left(1 \pm e^{(-i\Delta m + \frac{1}{2}\Delta\Gamma)t}\right). \end{aligned} \quad (17)$$

Therein, we defined the mass and decay width differences

$$\begin{aligned} \Delta m &\equiv m_H - m_L > 0 \\ \Delta\Gamma &\equiv \Gamma_L - \Gamma_H. \end{aligned} \quad (18)$$

In an experiment, we measure the decay rates

$$\Gamma(P(t) \rightarrow f) = \frac{1}{N_P} \frac{d}{dt} N(P(t) \rightarrow f)$$

for decays to some final state f and analogously for the decay rates to the CP-conjugate final state \bar{f} . Up to some common normalization, the decay rates for the flavor eigenstates are given by [29]

$$\begin{aligned} \Gamma(P^0(t) \rightarrow f) &= |\langle f | H | P^0(t) \rangle|^2 = |A_f|^2 |g_+(t) + g_-(t) \lambda_f|^2 \quad \text{and} \\ \Gamma(\bar{P}^0(t) \rightarrow \bar{f}) &= |\bar{A}_{\bar{f}}|^2 \left| g_+(t) + \frac{g_-(t)}{\lambda_{\bar{f}}} \right|^2, \end{aligned} \quad (19)$$

where H is the Hamiltonian governing the decay transition with corresponding decay amplitudes

$$\begin{aligned} A_f &\equiv \langle f | H | P^0 \rangle \quad \text{and} \quad A_{\bar{f}} \equiv \langle \bar{f} | H | P^0 \rangle \\ \bar{A}_f &\equiv \langle f | H | \bar{P}^0 \rangle \quad \text{and} \quad \bar{A}_{\bar{f}} \equiv \langle \bar{f} | H | \bar{P}^0 \rangle, \end{aligned} \quad (20)$$

and we defined the parameters $\lambda_f \equiv \frac{q}{p} \frac{\bar{A}_f}{A_f}$ and $\lambda_{\bar{f}} \equiv \frac{q}{p} \frac{\bar{A}_{\bar{f}}}{A_{\bar{f}}}$. As will be shown in the following, no detailed calculation of the hadronic effects and their uncertainties is necessary for a measurement of the B^0 lifetime and mixing behavior.

2.2.3 Neutral B^0 Meson Mixing

The previous expressions simplify considerably, if we take into account the relative magnitudes of m_L , Γ_L , Δm , $\Delta\Gamma$ for the considered type of meson. In general, mixing can either occur through decays to intermediate meson states or through the exchange of virtual heavy particles. We respectively call these the *long-* and *short-distance contributions* to mixing. Decays in long-distance mixing contribute to a non-zero decay width difference $\Delta\Gamma$. The exchange of virtual heavy particles in short-distance mixing contribute to a non-zero mass difference Δm of the physical mass eigenstates [30].

Long-distance contributions are more difficult to compute than short-distance contributions. This is due to confinement at long distances, i.e. low energies. Then, the degrees of freedom of Quantum Chromodynamics (QCD) are composite hadrons rather than individual quarks. In this regime, QCD behaves non-perturbatively and the uncertainties on the theoretical predictions become significant. In B^0 mixing, Γ_{12} is strongly CKM suppressed and long-distance contributions can be neglected [6]. We therefore approximate

$$\Delta\Gamma \ll \Delta m \implies \Delta\Gamma \approx 0 \implies \Gamma_L \approx \frac{\Gamma_L + \Gamma_H}{2} = \frac{1}{\tau_{B^0}}, \quad (21)$$

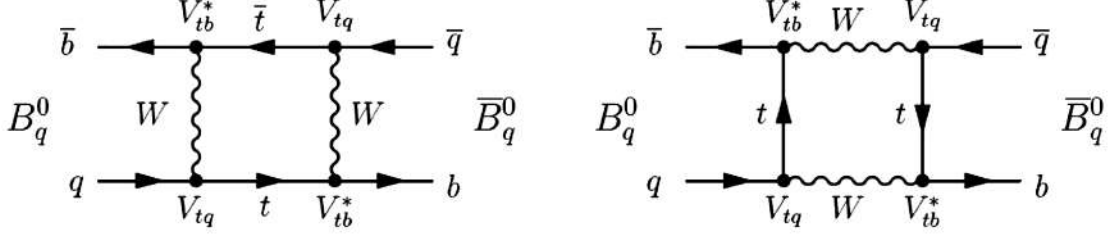
which leads to

$$g_{\pm}(t) \approx \frac{1}{2} e^{-im_L t} e^{-t/(2\tau_{B^0})} (1 \pm e^{-i\Delta m t}). \quad (22)$$

Short-distance mixing proceeds via FCNCs in so-called box-diagrams, for example as depicted in Figure 4. For a particular box-diagram, the larger the mass of the exchanged virtual particle, the larger its contribution [31]. We have

$$m_t > m_W \gg m_{c,u},$$

where m denotes the masses of the involved particles. Additionally, one has to take into account the magnitudes of the corresponding CKM elements in the diagram. Thus for B^0 mixing, the exchange of virtual top quarks vastly dominates, since the top quark is the heaviest quark and $b \rightarrow t W^-$ transitions are not CKM suppressed (see Equation 6).

Figure 4: Dominating box-diagrams in B^0 mixing.

In summary for B^0 mixing, we only account for the contribution in Figure 4. For D^0 and K^0 mixing no such simplification is possible, since mixing via top quarks is CKM suppressed and long-distance mixing via real light mesons dominates.

Plugging Equation 22 in Equation 19, we find

$$\begin{aligned} \Gamma(B^0(t) \rightarrow f) &= |A_f|^2 \cdot \frac{e^{-t/\tau_{B^0}}}{2\tau_{B^0}} \cdot \left[1 + \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2} \cos(\Delta mt) - \frac{2 \operatorname{Im} \lambda_f}{1 + |\lambda_f|^2} \sin(\Delta mt) \right] \\ \Gamma(\bar{B}^0(t) \rightarrow \bar{f}) &= |\bar{A}_{\bar{f}}|^2 \cdot \frac{e^{-t/\tau_{B^0}}}{2\tau_{B^0}} \cdot \left[1 - \frac{1 - |1/\lambda_{\bar{f}}|^2}{1 + |1/\lambda_{\bar{f}}|^2} \cos(\Delta mt) + \frac{2 \operatorname{Im}(1/\lambda_{\bar{f}})}{1 + |1/\lambda_{\bar{f}}|^2} \sin(\Delta mt) \right]. \end{aligned} \quad (23)$$

For a mixing measurement, we look at *flavor-specific* final states, i.e. final states f that are only accessible from B^0 meson decays and not from \bar{B}^0 meson decays. Then, we have $\bar{A}_{\bar{f}} = A_{\bar{f}} = 0$ and hence $\lambda_f = 1/\lambda_{\bar{f}} = 0$. Assuming that there is no CPV in the decays [29], i.e. $|A_f| = |\bar{A}_{\bar{f}}|$, we compute the *mixing asymmetry* to

$$\mathcal{A}(t) \equiv \frac{\Gamma(B^0(t) \rightarrow f) - \Gamma(\bar{B}^0(t) \rightarrow \bar{f})}{\Gamma(B^0(t) \rightarrow f) + \Gamma(\bar{B}^0(t) \rightarrow \bar{f})} = \cos(\Delta mt). \quad (24)$$

Thus, from a time-dependent measurement of the *mixing asymmetry* $\mathcal{A}(t)$, we can extract the mixing parameter Δm .

3 Experiment

To measure the time-dependent mixing asymmetry (Equation 24), we look at decays to *flavor-specific* final states, i.e. final states that are only accessible from either B^0 or \bar{B}^0 decays. As will be shown in Chapter 4, these decays exhibit low branching fractions (see Table 1). Thus, large numbers $\mathcal{O}(10^6)$ of B meson decays, i.e. *high-luminosity colliders*, are needed for a sufficient statistical precision of the mixing asymmetry measurement.

In this high-luminosity environment, the decay rates $\Gamma(B \rightarrow f)$ have to be measured as a function of the decay time t and of the flavor in which the B meson decays (either B^0 or \bar{B}^0). To achieve this, high-resolution detectors with excellent particle identification and vertex resolution are required. According to the principles of relativity, the decay time t_{lab} of a B meson with non-zero velocity $\beta_{\text{lab}} = v/c > 0$ in the laboratory frame is dilated relative to the decay time t_{cms} in the centre-of-mass frame.

$$t_{\text{lab}} = t_{\text{cms}} \cdot \gamma_{\text{lab}} = \frac{t_{\text{cms}}}{\sqrt{1 - \beta_{\text{lab}}^2}} > t_{\text{cms}}, \quad (25)$$

where γ_{lab} denotes the Lorentz factor in the laboratory frame. Naturally in an experiment, larger decay times can be measured with higher relative precision than smaller decay times. Thus the

principle of time dilation is exploited in experiments, by producing the B mesons with non-zero boosts $\beta\gamma > 0$.

Two complementary experimental approaches exist to over-constrain CKM-unitarity and to investigate B meson mixing. At *hadron colliders*, high B meson production cross-sections provide high rates of B meson decays. The B mesons are produced at high energies and carry significant boosts, e.g. $\beta\gamma \sim 100$ at LHC [26]. However, the hadronic nature of the interactions results in large numbers of additional particles besides the B mesons, whose kinematics vary from event to event. In contrast at *lepton colliders*, B mesons can be produced exclusively with constant well-known kinematics. In the following, this chapter describes the experimental setup of the electron-positron collider *SuperKEKB* and the *Belle II* detector and is based on [32] and [33].

3.1 SuperKEKB Collider

The *SuperKEKB* collider near Tsukuba in Japan is an asymmetric electron-positron collider, optimized for the production of B mesons. As explained above, we can exploit relativistic time dilation to increase the B meson decay length in the laboratory frame, and thus improve its measurement precision in the detector. To this end, the $\Upsilon(4S)$ resonance is not produced at rest, but with a boost $\beta\gamma = 0.284$ [33]. This is achieved using asymmetric beam energies,

$$\begin{aligned} E_{\text{beam}}^{e^-} &= 7 \text{ GeV} \\ E_{\text{beam}}^{e^+} &= 4 \text{ GeV}, \end{aligned} \tag{26}$$

which requires separate beam pipes for electrons e^- and positrons e^+ . Figure 5 shows a schematic view of the SuperKEKB collider with its pre-accelerator stages and the *Belle II* detector at the beam interaction point.

Electrons and positrons are collided at a center-of-mass energy \sqrt{s} corresponding to the $\Upsilon(4S)$ resonance rest mass,

$$\sqrt{s} = m_{\Upsilon(4S)}c^2 \approx 10.58 \text{ GeV}. \tag{27}$$

The $\Upsilon(4S)$ resonance decays exclusively to a $B\bar{B}$ meson pair with a branching fraction of above 96% [34]. It thus provides a clean, low-background B^0 meson pair production mode, in which each B meson carries exactly half of the collision energy \sqrt{s} in the center-of-mass frame. This can be used as kinematical constraint to identify B mesons, which we will describe in Chapter 5.

In the e^+e^- collision, the quark-antiquark ($q = u, d, s, c$) pair production cross section is about 3 nb, which is large compared to a production cross section of about 1 nb for the $\Upsilon(4S)$ resonance [32]. To investigate backgrounds from $q\bar{q}$ production, the collision energy can be decreased down to the $\Upsilon(1S)$ resonance rest mass. Runs at the $\Upsilon(5S)$ and $\Upsilon(6S)$ resonance rest mass produce heavier B mesons such as strange B mesons B_s . However, B_s^0 mesons oscillate much faster than B_d^0 mesons and the Belle II resolution is not sufficient to resolve B_s^0 meson mixing (even so they are resolved at LHC due to the large boost $\beta\gamma$).

For the electron beam, electrons are generated by a *radio-frequency gun* and accelerated to 7 GeV in the *injector linac* (linear accelerator), before being injected into the electron ring. To generate positrons a *tungsten target* is irradiated with 3.5 GeV electrons about half-way down the linac, producing brems-photons which are immediately converted into electron-positron pairs in a second close-by target. Since the positrons are produced as secondary particles, their emittance is high, i.e. the positron beam width is large and they exhibit a wide momentum spread. To maximize the luminosity and minimize beam backgrounds, the captured positrons are fed into a 1 GeV *damping ring*. Therein, the positron beam emittance is decreased through synchrotron-radiation. Subsequently, they are reinjected into the linac, accelerated to 4 GeV and injected into

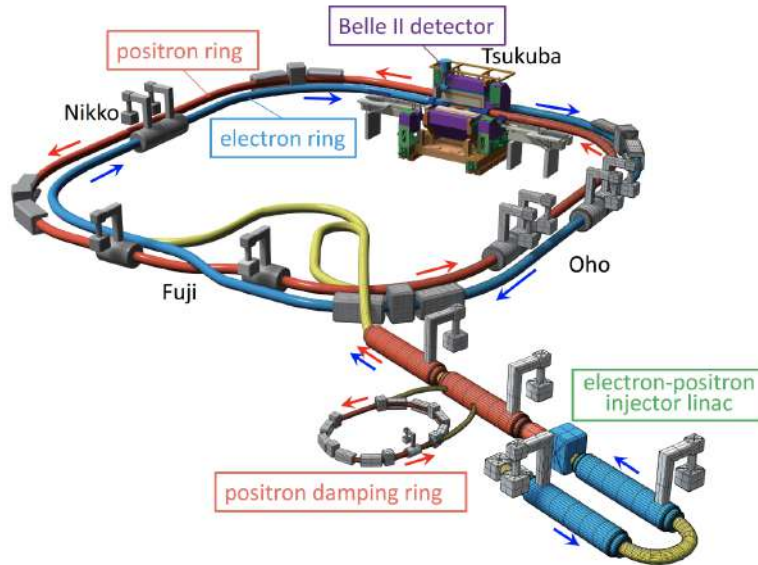


Figure 5: SuperKEKB collider schematic [35].

the positron main ring.

SuperKEKB is the successor of the KEKB collider, which operated in the same tunnel and provided e^+e^- collisions for the first-generation *Belle* detector. To probe physics beyond the SM with high statistics, the second generation collider SuperKEKB is required to deliver an instantaneous luminosity 40 times greater than KEKB [35]. Assuming gaussian beam profiles with transverse widths (σ_x, σ_y) at the interaction point, the instantaneous luminosity of a storage ring is given by

$$\mathcal{L} = \frac{N_b n_e n_{e^-} f}{4\pi\sigma_x \cdot \sigma_y}, \quad (28)$$

where N_b is the number of bunches, f the circulation frequency and n the number of particles per bunch [6]. To increase \mathcal{L} , the number of bunches N_b cannot be increased arbitrarily at constant ring radius, since neighboring bunches eventually interact unwantedly. Instead at SuperKEKB, the beam currents are doubled and in the so-called nano-beam scheme, the beam size orthogonal to the collider-plane is decreased to $\sigma_y \sim 50$ nm [35]. In June 2021, SuperKEKB delivered a record instantaneous luminosity of $3.1 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. In total, the experiment is expected to accumulate an integrated luminosity of $L = \int \mathcal{L} dt \approx 50 \text{ ab}^{-1}$, a 50-fold increase compared to *Belle*.

3.2 Belle II Detector

The *Belle II* detector is installed at the SuperKEKB beam interaction point and records collisions with close to 4π solid angle coverage. It supercedes the *Belle* detector for which an upgrade was mandatory in the new high-luminosity environment and due to the resulting increased backgrounds and event rates. Belle II uses the same solenoid magnet and hence the same iron flux return yoke (so that the outer dimensions are the same), with all detector components upgraded for Belle II. Figure 6 shows schematic views of the Belle II spectrometer. Its forward/backward asymmetry is based on the asymmetric collider.

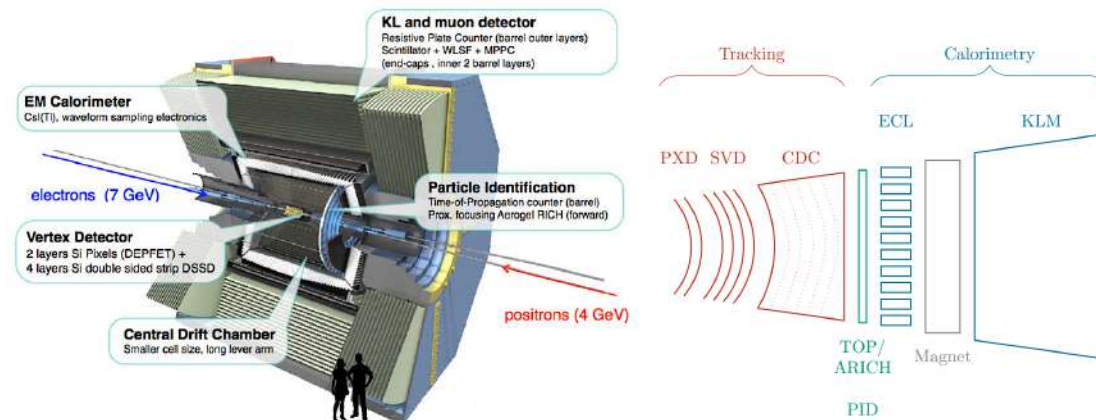


Figure 6: Left: cross section of the Belle II spectrometer [36]. Right: schematic of sub-detector structure from detector center to outer edge [37].

To resolve neutral B^0 meson mixing at *Belle II*, an excellent vertex resolution is required. The **vertex detector** (VXD) is the innermost detector component and consists of six layers of silicon detectors. To keep the detector occupancy at a feasible level, the two innermost VXD layers were realized as pixel detectors (PXD), while the outer layers function as strip detectors. In the PXD, the DEPFET (DEPLETED Field Effect Transistor [32]) technology allows for sufficiently radiation-hard and very thin pixel sensors and thus minimizes multiple scattering in the detector material, which is important for the vertex resolution. The innermost PXD layer is placed at a radius of only 14 mm around the 10 mm beam-pipe, while the outermost layer of the VXD stretches out to 140 mm [32]. The resulting vertex resolution is momentum-dependent and reaches a spatial resolution of up to $\sim 15 \mu\text{m}$ [33]. The VXD also enables to measure tracks of particles that decay before reaching the other detector components. However for the inner layers, the background and hit rates become very large.

The tracking system of Belle II consists of the VXD and the **Central Drift Chamber** (CDC). The CDC is build up of a cylindrical gas volume around the VXD, in which charged particles leave trails of ionized gas. Many so-called field wires stretch through the gas volume and build up a well-known electric field, along which the ionized particles drift and in turn ionize gas molecules. The induced currents are read out in sense wires and the particle's track can be reconstructed. The tracks curve in the magnetic field of 1.5 T, provided by the surrounding **superconducting solenoid magnet**. From this curvature, the particle's momentum is deduced. The CDC also serves to measure the energy loss dE/dx of particles within its gas volume. The energy loss at known momentum is specific for the particle's mass and thus serves as PID for low-momentum tracks that do not reach the outer detectors.

For every particle type, the PID is a combination of information from all subdetectors. Two kinds of Cherenkov PID detectors are dedicated to improve charged hadron identification, e.g. to separate K and π . In the barrel region a **time-of-propagation counter** (TOP) is used, while an **aerogel ring-imaging Cherenkov detector** (ARICH) is used in the forward end-cap. Both detectors make use of the emission of Cherenkov photons by charged particles, propagating at speeds $v > c_n = c/n$ through a medium with refractive index n and speed of light c_n . The photon emission angle θ is given by

$$\cos \theta = \frac{t \cdot c/n}{v \cdot t} = \frac{1}{n\beta}, \quad (29)$$

with time t and hence specific for the particle's velocity $\beta = v/c$. Plugging in the momentum

$|\vec{p}| = m\gamma\beta c$, we find

$$m^2c^2 = |\vec{p}|^2 \left(\frac{1}{\beta^2} - 1 \right) = |\vec{p}|^2 (n^2 \cdot \cos^2 \theta - 1), \quad (30)$$

and we can identify the particle by its invariant mass m , once its momentum $|\vec{p}|$ and Cherenkov angle θ is measured. In the ARICH, Cherenkov photons are produced in aerogel radiators and a ring of photons is projected on photons sensors. Then, θ can be deduced from the radius. In the TOP, quartz bars serve as radiator. The produced photons travel in total reflection to the lateral photon detectors. θ then is a function of the propagation time and the impact position on the detector.

The tracking and PID detectors are surrounded by the **electromagnetic calorimeter** (ECAL). Its main purpose lays in the identification and energy measurement of electrons and photons. The ECAL consists of highly segmented arrays of thallium-doped caesium iodide CsI(Tl) crystal scintillators, which are read out by photodiodes at their rear ends. Particles that interact with the detector material, loose energy in the crystals through forming particle showers. The scintillator material becomes temporarily excited by ionizing radiation of shower particles and relaxates through the emission of scintillation light. The detected scintillation light serves as measure for the energy loss dE/dx of the incident particle. For fully absorbed particles, the ECAL measures the particle's total energy. In the forward region of the ECAL, the new high-luminosity environment may lead to increased pile-up noise and radiation damage. In the future, the ECAL forward end-cap could be upgraded to radiation-hard, un-doped CsI crystals, which provide shorter scintillation times but lower light outputs [32].

The final detector component is the **K_L^0 and muon detector** (KLM), located outside the superconducting solenoid. It serves to identify K_L^0 mesons and μ leptons that interact sparsely with the detector material. The KLM is a sampling calorimeter, alternating between iron plates and active detector elements. While resistive plate chambers (RPC) were used everywhere in Belle, their dead-time in between hits is too large for the increased background rates in the forward-region of Belle II. Instead, Belle II employs scintillator strips in the KLM-endcaps, that are read out by silicon photomultipliers. The iron plates serve as magnetic flux return for the solenoid magnet and significantly add to the calorimeter's material density, in which the particles shower hadronically.

Besides upgrading the detector components, Belle II employs newly developed triggers to provide a broad scope of physics analysis topics despite the increased background levels at SuperKEKB [33].

3.3 Data Simulation, Processing and Analysing

Figure 7 illustrates the general workflow of a physics analysis at Belle II. During *data taking*, SuperKEKB supplies the detector with collisions. For the particles created in the collision (the so-called *event*), the responses of parts of the sub-detector systems are evaluated in real-time. If these satisfy a set of *trigger* conditions, the event is triggered, i.e. all detector components are read out and stored for later evaluation. Triggers are vital to filter out physically interesting collisions and to keep the event rate at a feasible level for the data recording systems.

To compare the recorded data from the detector with our expectations, we need to simulate the detector response to the final state particles. We use *Monte Carlo* methods (MC) to simulate events in two steps. First, the particle interaction from the electron-positron collision is generated according to some physics model, e.g. the SM. This means generating the positions $x^\mu = (t, \vec{x})$ and four-momenta $p^\mu = (E, \vec{p})$ of all particles. Second, we need to simulate how the generated particles propagate through the detector and how exactly each sub-detector responds.

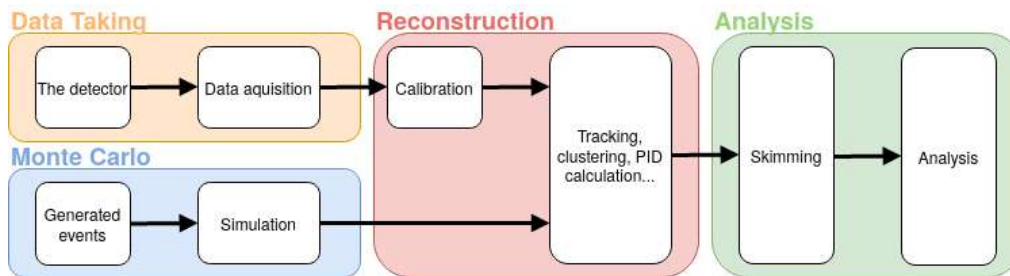


Figure 7: Basic Belle II analysis workflow scheme [38].

In the *reconstruction* of each event, we try to reconstruct the particles and their kinematics, based on the detector responses. Due to detector resolution and background processes, the particle identification can only give a most likely estimate. Aside from detector calibration, the simulated and recorded events are reconstructed in exactly the same manner. Particle types and kinematics are reconstructed from clusters of hits in the calorimeters and from trajectories (also called tracks) in the tracking detectors. Additionally for the simulated events, we know the truly generated particles and can compare them with the corresponding reconstructed particle. This process is called *truth-matching* the particles.

For the measurement of B^0 meson mixing, we need to reconstruct the flavor of the decaying B mesons, i.e. whether it contained a b quark (\bar{B}^0) or \bar{b} quark (B^0) at the time of decay. At B factories such as SuperKEKB, the B mesons are produced exclusively, i.e. without other particles in the event. Therefore after reconstructing a specific B-decay channel, called the *signal B*, all remaining tracks and photons are assigned to the reconstruction of the other B meson, which decays generically. A large fraction of B meson decay channels is approximately flavor specific, i.e. can only be reached either through the decay of a b quark or a \bar{b} quark. Therefore, the flavor of the generically decaying B meson can be inferred from its reconstructed daughter-particles and is to a good approximation independent of the reconstructed signal decay [6]. At Belle II, dedicated multivariate *flavor-tagging algorithms* were developed that compute flavor-tag estimates from kinematic and PID information. These algorithms inclusively extract flavor-specific signatures from a multitude of B meson decay channels and combine them into a final flavor tag. In Belle II MC, the flavor tagging efficiency was estimated at about 37% [33]. For further details on flavor-tagging, we refer to [6],[39] and [40].

When performing an *analysis* of a specific physics process, we need to separate signal from background. Usually, a large portion of background processes can be discarded by so-called selection cuts. We impose limits on selected quantities, such that only events with decays similar to the expected signal remain. Further discrimination can for example be achieved by sophisticated (multi-dimensional) fits or machine-learning techniques. In some cases, it is useful to first *skim* the reconstructed events. Therein, we apply loose selection cuts to produce smaller data sets and thus decrease the computational cost of subsequent analyses.

4 Analysis Strategy for Lifetime and Mixing Measurement

We perform a precision measurement of the lifetime $\tau_{B_d^0}$ and mixing parameter Δm of the neutral B^0 system. This section summarizes the concepts of the measurements and outlines the analysis strategy. Special emphasis is put on the systematic analysis of backgrounds and on developing a fit strategy to separate signal and backgrounds.

In the *Belle II* detector at the SuperKEKB collider, $B^0\bar{B}^0$ meson pairs are produced exclusively from electron-positron annihilation, i.e. without other particles in the event. The B mesons are

produced in the flavor-conserving strong decay of the $\Upsilon(4S)$ resonance and hence are in opposite flavor eigenstates. Subsequently, they propagate in their physical mass eigenstates, which contain both flavors in superposition until one of the B mesons decays in a certain flavor (B or anti-B meson). At this point in time, the flavor of the other B meson is determined as opposite, since the $B^0\bar{B}^0$ pair was produced in an entangled coherent quantum state. Thereafter, the B meson is free to oscillate into the other flavor with a certain probability. Therefore, both B mesons can either be observed in opposite flavor eigenstates (opposite-flavor events), or in the same flavor eigenstate (same-flavor events).

For both cases, we expect time-dependent distributions for the numbers of opposite- and same-flavor events, denoted \tilde{N}_{OF} and \tilde{N}_{SF} respectively.²

$$\tilde{N}_{\text{OF, SF}} = \frac{\exp\left[-|\tilde{\Delta}t|/\tau_{B^0}\right]}{4\tau_{B^0}} \cdot \left[1 \pm \cos\left(\Delta m \tilde{\Delta}t\right)\right], \quad (31)$$

from which we can extract the lifetime τ_{B^0} , once the mixing parameter Δm is measured. Here, we do not measure the absolute decay times $\tilde{t}_1, \tilde{t}_2 \in [0, \infty)$ of both B mesons, but their decay time difference $\tilde{\Delta}t \equiv \tilde{t}_2 - \tilde{t}_1 \in (-\infty, \infty)$.

From a time-dependent measurement of the asymmetry of these events,

$$\tilde{\mathcal{A}}_{\text{mix}}(\tilde{\Delta}t) \equiv \frac{\tilde{N}_{\text{OF}}(\tilde{\Delta}t) - \tilde{N}_{\text{SF}}(\tilde{\Delta}t)}{\tilde{N}_{\text{OF}}(\tilde{\Delta}t) + \tilde{N}_{\text{SF}}(\tilde{\Delta}t)} = \cos\left[\Delta m \cdot \tilde{\Delta}t\right], \quad (32)$$

we can extract the mixing parameter Δm . By considering an asymmetry instead of absolute decay rates, experimental uncertainties cancel out.

From Equation 32, it is clear that for every event we need to determine whether it was opposite- or same-flavor. To distinguish opposite- and same-flavor events, we employ *flavor-tagging*. One B meson, called B_{sig} , is reconstructed in one of three signal channels with specific flavor (Table 1). For the reconstruction of anti-B mesons, all charge signs are reversed and neutral D mesons are replaced with neutral anti-D mesons.³

Table 1: Reconstructed B_{sig}^0 decay channels.

B_{sig}^0 decay channel	branching fraction [34]
$B^0 \rightarrow D^- (\rightarrow K^+ \pi^- \pi^-) \pi^+$	$2.5 \cdot 10^{-3} \cdot 0.09 = 2.3 \cdot 10^{-4}$
$B^0 \rightarrow D^{*-} \left[\rightarrow \bar{D}^0 (\rightarrow K^+ \pi^+ \pi^- \pi^-) \pi^+ \right]$	$2.7 \cdot 10^{-3} \cdot 0.68 \cdot 0.082 = 1.5 \cdot 10^{-4}$
$B^0 \rightarrow D^{*-} \left[\rightarrow \bar{D}^0 (\rightarrow K^+ \pi^-) \pi^- \right] \pi^+$	$2.7 \cdot 10^{-3} \cdot 0.68 \cdot 0.04 = 7.3 \cdot 10^{-5}$

Since apart from the two B mesons no other particles are produced in the event, we assign all particles remaining after the B_{sig}^0 reconstruction to the other B meson, called B_{tag} . In *flavor-tagging*, the flavors of B_{sig} and B_{tag} are inferred from their reconstructed daughter-particles. For B_{sig} decays, the charge sign of the π^\pm produced associated with the $D^{(*)\mp}$ encodes the B_{sig} flavor, when neglecting sub-leading order Feynman diagrams. At sub-leading order the final state is accessible from both flavors (Figure 8), but the doubly CKM-suppressed ($\sim 10^{-4}$) amplitude is negligible for practical purposes.

²Quantities with \sim denote the true values, i.e. measurements from an ideal experiment with perfect flavor-tagging, detector resolution, etc. and no approximations.

³In the following and this entire work, CP-conjugate processes in anti-B meson decays are always implied.

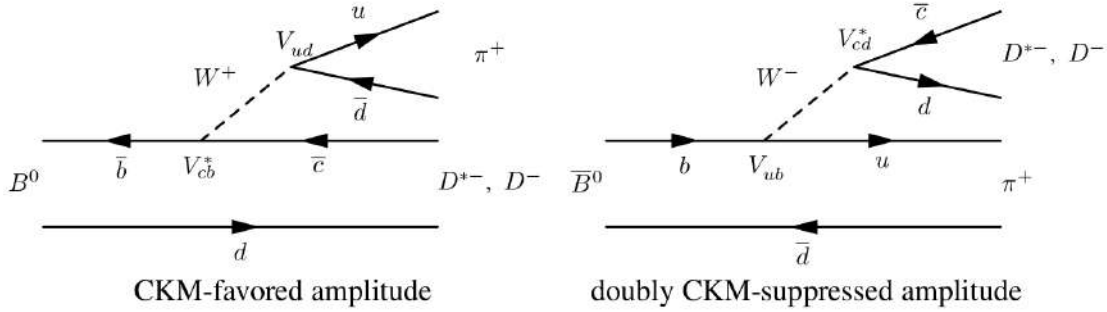


Figure 8: Feynman diagrams for B_{sig} decays, adapted from [39]. Left: leading-order. Right: sub-leading order suppressed by $\sim 10^{-4}$ [34].

The flavor of B_{tag} is inclusively determined in *flavor-tagging algorithms* from a multitude of different decays [40]. We may wrong-tag the B_{tag} decay, i.e. assign the wrong flavor, depending on the considered B_{tag} decay and the reconstruction efficiencies. We here denote the fraction of wrong-tagged decays w . We can express the number of decays N_{B^0} and $N_{\bar{B}^0}$, tagged as B^0 and \bar{B}^0 respectively, in the form

$$\begin{aligned} N_{B^0} &= \epsilon(1-w) \cdot \tilde{N}_{B^0} + \epsilon w \cdot \tilde{N}_{\bar{B}^0} \\ N_{\bar{B}^0} &= \epsilon(1-w) \cdot \tilde{N}_{\bar{B}^0} + \epsilon w \cdot \tilde{N}_{B^0}, \end{aligned}$$

where ϵ denotes the tagging efficiency and \tilde{N}_{B^0} and $\tilde{N}_{\bar{B}^0}$ denote the true numbers of B^0 and \bar{B}^0 mesons respectively⁴. We can now rewrite Equation 32 as

$$\mathcal{A}_{\text{mix}}(\tilde{\Delta}t, w) \equiv \frac{N_{\text{OF}}(\tilde{\Delta}t) - N_{\text{SF}}(\tilde{\Delta}t)}{N_{\text{OF}}(\tilde{\Delta}t) + N_{\text{SF}}(\tilde{\Delta}t)} = (1-2w) \cdot \cos[\Delta m \cdot \tilde{\Delta}t], \quad (33)$$

where N_{OF} and N_{SF} respectively denote the measured numbers of opposite- and same-flavor events. Due to wrong-tags the asymmetry becomes *diluted* by a factor $r \equiv (1-2w)$. By comparing Equations 32 and 33, we find that the statistical errors scale as

$$\sigma_{\tilde{\mathcal{A}}_{\text{mix}}} = \frac{\sigma_{\mathcal{A}_{\text{mix}}}}{r} \propto \frac{1/\sqrt{N_{B^0} + N_{\bar{B}^0}}}{r} \propto \frac{1}{\sqrt{\epsilon \cdot r^2}}. \quad (34)$$

When determining the flavor of B_{tag} with flavor-tagging algorithms, we assign it a value $q \cdot r$. Therein, $q = +1$ (-1) denotes the tagged flavor B^0 (\bar{B}^0) and the dilution factor $r \in [0, 1]$ depends on the tagged flavor-specific signature. For $r = 1$ (0), the flavor tag is certain (undetermined). We now show that the statistical uncertainty (Equation 34) can be decreased, by separating the tagged decays according to their dilutions r .

Suppose we group the tagged decays into dilution bins. In each bin c , we have a fraction of decays ϵ_c with dilution r_c . The total tagging efficiency ϵ and the average dilution r are given by

$$\epsilon = \sum_c \epsilon_c \quad r = \sum_c \frac{\epsilon_c}{\epsilon} \cdot r_c$$

After grouping the tagged decays into dilution bins, the statistical error is given by

$$\frac{1}{\sigma_{\tilde{\mathcal{A}}_{\text{mix}}}} \propto \sqrt{\sum_c \epsilon_c \cdot r_c^2} = \sqrt{\epsilon r^2 + \sum_c \epsilon_c \cdot (r_c - r)^2} > \sqrt{\epsilon r^2}. \quad (35)$$

⁴We here assume that the wrong-tag fraction w and efficiency ϵ are identical for both flavors. This is an approximation since the detector performance may depend on the flavor. In the lifetime and mixing analysis, we account for this as well as for tag-side interference (interfering processes in the B_{tag} decay) in systematic errors.

Hence, the statistical error is decreased through the binning [6]. In the lifetime and mixing analysis, we group the tagged decays into 7 bins according to Table 2.

Table 2: Utilized bins in dilution r .

Bin Number c	Dilution r
0	$0.000 \leq r < 0.100$
1	$0.100 \leq r < 0.250$
2	$0.250 \leq r < 0.500$
3	$0.500 \leq r < 0.625$
4	$0.625 \leq r < 0.750$
5	$0.750 \leq r < 0.875$
6	$0.875 \leq r < 1.000$

The bins are chosen in accordance with the *Belle II* flavor tagging calibration [40]. The wrong-tag fraction in Equation 33 becomes binned: w_c , $c = 0, \dots, 6$.

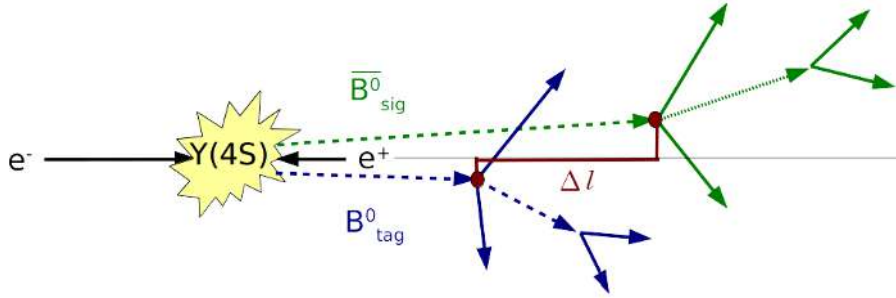


Figure 9: Schematic illustration of the distance Δl between both reconstructed B vertices (adapted from [41]).

Besides separating opposite- and same-flavor events, we need to measure the decay time difference Δt of both B mesons (see Equation 33). It can be inferred from the distance Δl between both reconstructed B decay vertices along the boost direction (Figure 9) and the accelerator boost $\beta\gamma$ of the electron-positron cms-system. By measuring the relative decay length Δl instead of the absolute decay lengths, uncertainties on the beam collision point are avoided. Then one has [6]

$$\tilde{\Delta}t = \frac{\tilde{\Delta}l}{\tilde{\beta}\tilde{\gamma}\tilde{\gamma}_{Bc}} + \frac{\tilde{\beta}_B}{\tilde{\beta}} \cdot \tilde{\Sigma}t \cdot \cos\tilde{\theta}_{\text{cms}} \quad \xrightarrow{\text{approx.}} \quad \Delta t = \frac{\Delta l}{\beta\gamma c} \cdot \frac{m_B c^2}{E_B^{\text{cms}}}, \quad (36)$$

where $\beta_B\gamma_B$, E_B^{cms} and θ_{cms} denote the boost, energy and polar angle of the B_{sig} in the cms-frame. Σt denotes the sum of the decay times from both B mesons. The B mesons carry a small momentum in the cms-frame and thus the true decay time difference $\tilde{\Delta}t$ is dependent on $\tilde{\theta}_{\text{cms}}$. In Δt , we neglect the second term in Equation 36 and approximate the B_{sig} to be at rest in the cms-frame. We call this the *kinematic approximation*.

Figure 10 shows the simulated distributions in Δt (Equation 36) and $\tilde{\Delta}t$ in the signal channel $B^0 [D^-(K^+\pi^-\pi^+)\pi^+]$. As expected the measured distribution Δt is smeared out, due to a finite resolution of the boost $\beta\gamma$ and the vertices, as well as due to the kinematic approximation.

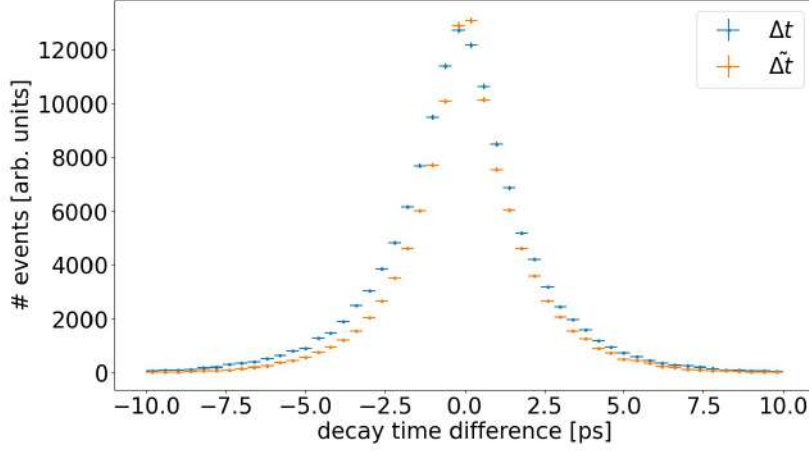


Figure 10: Simulated distributions in measured Δt and true $\tilde{\Delta t}$ for signal channel $B^0 [D^-(K^+\pi^-\pi^-\pi^+)]$.

In this analysis, we use a *resolution function* \mathcal{R} to account for all detector effects and for the kinematic approximation. Since these are of similar magnitude as the average Δt , the resolution function has to be studied in great detail from simulation [6]. The resolution function was developed in [42] and is presented as part of the full analysis in Chapter 8. By convoluting with the resolution function in each of the seven dilution bins $c = 0, \dots, 6$, we relate the measured to the true distributions.

$$\begin{aligned} \mathcal{A}_{\text{mix}}^c(\Delta t) &\equiv \frac{N_{\text{OF}}^c(\Delta t) - N_{\text{SF}}^c(\Delta t)}{N_{\text{OF}}^c(\Delta t) + N_{\text{SF}}^c(\Delta t)} = (1 - 2w_c) \cdot \cos[\Delta m \cdot \Delta t] \\ &\text{with } N^c(\Delta t) = (N^c(\tilde{\Delta t}) * \mathcal{R})(\Delta t) \\ &\text{and } c = 0, \dots, 6 \end{aligned} \quad (37)$$

By simultaneously fitting Equation 37 on the measured asymmetries in all dilution bins, we can in principle extract the wrong-tag fractions w_c and the mixing parameter Δm .

However, one has to take into account systematic uncertainties from background decays that are falsely reconstructed as signal and distort the measured values. Figure 11 shows the measured distribution for reconstructed B_{sig}^0 decays in the so-called beam-constrained mass M_{bc} from an earlier *Belle* analysis [43]. We define M_{bc} and the so-called energy difference ΔE as

$$\begin{aligned} M_{bc} &\equiv \sqrt{(E_{\text{beam}}^{\text{cms}})^2 - (\vec{p}_B^{\text{cms}})^2} \quad \text{and} \\ \Delta E &\equiv E_B^{\text{cms}} - E_{\text{beam}}^{\text{cms}}, \end{aligned} \quad (38)$$

where $E_{\text{beam}}^{\text{cms}}$, E_B^{cms} and \vec{p}_B^{cms} respectively are the beam energy and the reconstructed B meson energy and momentum in the cms-frame. For reasons that will be explained in Chapter 5, the distributions in M_{bc} and ΔE are convenient for separating backgrounds and reconstructed signal decays.

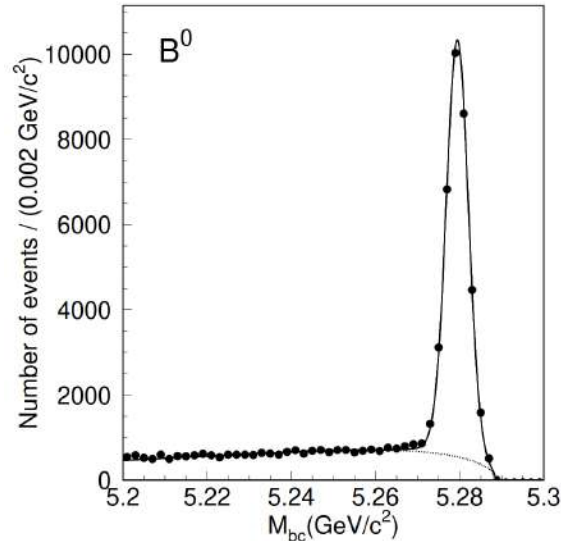


Figure 11: Dots: Distribution of reconstructed B_{sig}^0 decays in the beam-constrained mass M_{bc} [43]. Dashed: Fitted background component. Solid: Fitted signal+background component.

In Figure 11, the distribution of reconstructed B_{sig}^0 decays (dots) consists of true signal decays and backgrounds falsely reconstructed as signal. In the real data, these background decays cannot be distinguished from signal event-by-event. However, they can be estimated statistically in fits. To separate out the true signal decays, a fit in M_{bc} is performed in [43]. From simulation, the expected shapes in M_{bc} are extracted for backgrounds and signal. These shapes are then fitted onto the measured distribution in M_{bc} . Figure 11 shows the fitted background (dashed) and signal+background (solid) components. In this way, the number of true signal decays can be estimated and the mixing asymmetry (Equation 37) is then computed only from these.

The fit strategy in [43] thus amounts to a two-dimensional fit in $(M_{bc}, \Delta t)$. This fit strategy was also implemented in an early *Belle II* analysis [44]. The shape of background decays in Δt is extracted from reconstructed decays with $M_{bc} < 5.27 \text{ GeV}/c^2$. In this so-called *sideband* in M_{bc} , we expect only background decays (see Figure 11). The extracted background and signal shapes in Δt and M_{bc} are then simultaneously fitted onto the measured distributions in the *signal region* $M_{bc} \geq 5.27 \text{ GeV}/c^2$. In [43], the signal shape in Δt was extracted from simulation and it was assumed that the background shape is identical in the side-band and the signal region. To scrutinize this assumption we plot the background distributions in Δt , in both the sideband and signal region in M_{bc} (left side of Figure 12). We see that the assumption from [43] is an approximation. A more accurate background shape for Δt in the signal region can be extracted from the side-band in the energy difference $|\Delta E| > 0.05 \text{ GeV}$ (right side of Figure 12). Therefore in this analysis we fit in ΔE instead of in M_{bc} and we expect lower biases in the time-dependent fit.

For any fit, the extracted number of true signal decays is afflicted with uncertainties. For the fit in Figure 11, an uncertainty arises from a particular kind of background. *Peaking backgrounds* consist of B meson decays with missing or mis-identified tracks, that are falsely reconstructed as signal decays. Since they also peak at $M_{bc} \approx 5.28 \text{ GeV}/c^2$ like the signal decays (therefore the name *peaking backgrounds*), they cannot be identified as background from a fit in M_{bc} . In [43] and [44], peaking backgrounds were not evaluated in detail and resulted in a systematic uncertainty on the lifetime τ_{B^0} and mixing parameter Δm .

For this lifetime and mixing analysis, we perform an extensive analysis of the (peaking) backgrounds in simulation in all three signal channels (Chapter 6). This analysis serves to understand

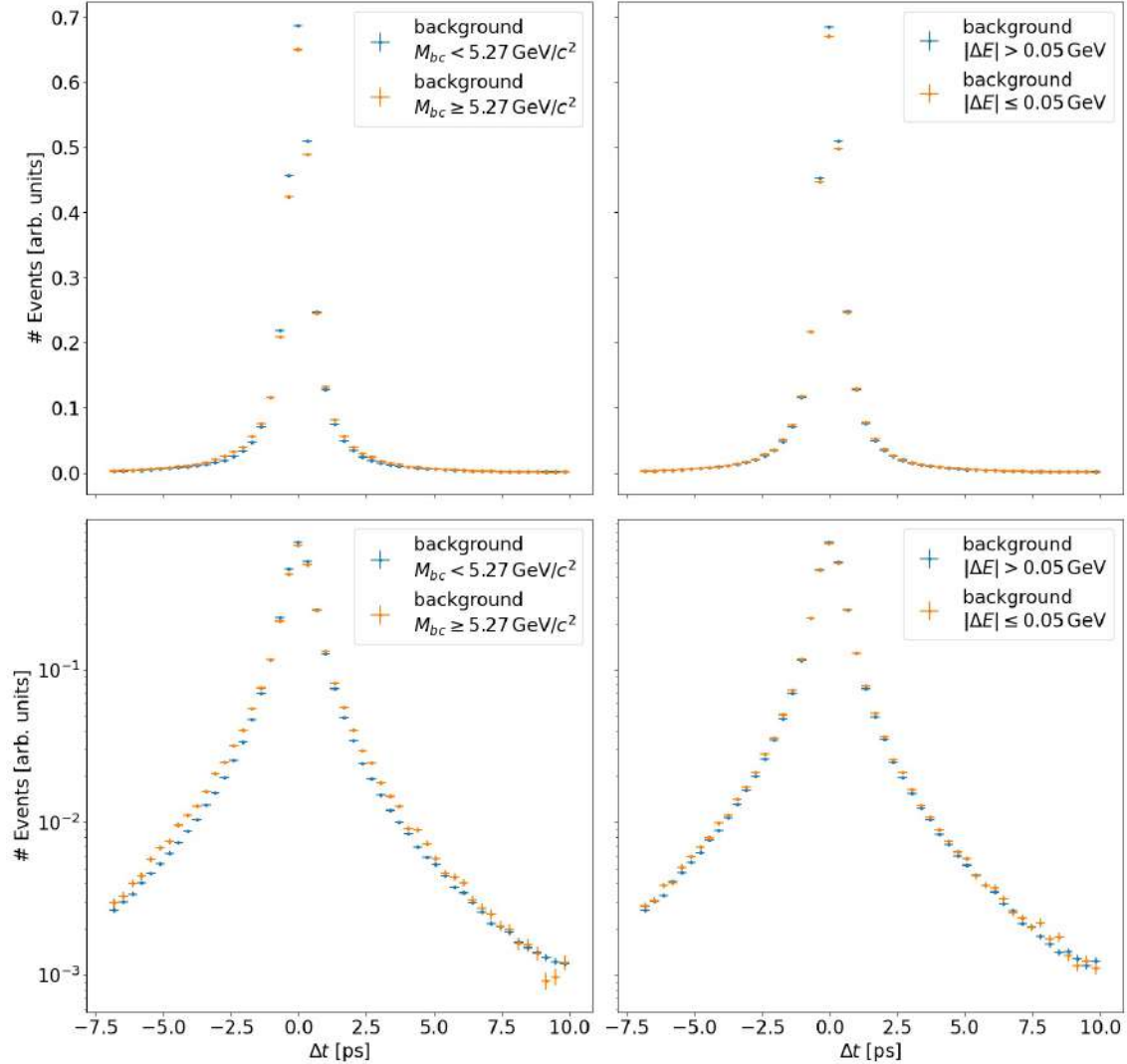


Figure 12: Distributions of reconstructed background decays in simulation in all three signal channels combined. Right: Side-band and signal region in ΔE . Left: Side-band and signal region in M_{bc} . Top: Linear scale. Bottom: Logarithmic scale.

the contributing backgrounds and is important to assess their impact on the measurement. In Chapter 7, we present a fit strategy that aims to decrease the systematic uncertainties on τ_{B^0} and Δm by also fitting the peaking backgrounds.

Several fit strategies were tested to separate signal and backgrounds. As in [43], we studied the backgrounds and signal in simulation and extracted the shapes of their distributions in selected variables from simulation. We then fitted these shapes on the measured distributions, in order to separate out the true signal decays from which the mixing asymmetry is computed. The fits were first validated in simulation and then applied on *Belle II* data. Best results were obtained for a two-dimensional fit in the event variables ΔE and the output of a *continuum suppression boosted decision tree*, here denoted BDT.⁵ These event variables and the detailed fit strategy will

⁵Fits including Fox-Wolfram moments were tested, but proved inferior to the CS-BDT and are omitted here. The same applies to fits of ΔE in 2 bins of M_{bc} . M_{bc} and ΔE cannot be simultaneously fitted without accounting for their correlation in signal (see Section 7.2.2 and Appendix B).

be presented in Chapter 7. In the following we explain why the two-dimensional fit is expected to perform better than a one-dimensional fit in only ΔE .

For the lifetime and mixing measurement, we combine the two-dimensional fit in $(\Delta E, \text{BDT})$ with the time-dependent fit in Δt (Equation 37). In this analysis, we also take into account the uncertainty on Δt , which we denote $\sigma_{\Delta t}$ in the following. This amounts to a four-dimensional fit function in each dilution bin:

$$N(\Delta t, \sigma_{\Delta t}, \Delta E, \text{BDT}) = N_{\text{sig}} \cdot G_{\text{sig}}(\Delta t, \sigma_{\Delta t}) \cdot H_{\text{sig}}(\Delta E, \text{BDT}) + \sum_{\text{bkg } i} N_{\text{bkg}}^i \cdot G_{\text{bkg}}^i(\Delta t, \sigma_{\Delta t}) \cdot H_{\text{bkg}}^i(\Delta E, \text{BDT}), \quad (39)$$

where the sum runs over all backgrounds, and G and H respectively are the distributions in the fit variables $(\Delta t, \sigma_{\Delta t})$ and $(\Delta E, \text{BDT})$. N_{sig} and N_{bkg}^i respectively denote the fitted numbers of signal and background decays.

For comparison, a simpler three-dimensional fit strategy in $(\Delta E, \Delta t, \sigma_{\Delta t})$ is also adopted. Here, the backgrounds and signal are separated in a one-dimensional fit in ΔE . We present this fit strategy in Section 7.5. It amounts to a three-dimensional fit function in each dilution bin:

$$N(\Delta t, \sigma_{\Delta t}, \Delta E) = N_{\text{sig}} \cdot G_{\text{sig}}(\Delta t, \sigma_{\Delta t}) \cdot H_{\text{sig}}(\Delta E) + N_{\text{bkg}} \cdot G_{\text{bkg}}(\Delta t, \sigma_{\Delta t}) \cdot H_{\text{bkg}}(\Delta E), \quad (40)$$

where different (peaking) backgrounds cannot be separated from each other, and are modelled collectively in the second term. This assumes, that the background distribution G_{bkg} in the decay time difference is independent of ΔE . Figure 13 shows the simulated G_{bkg} for different regions in ΔE .

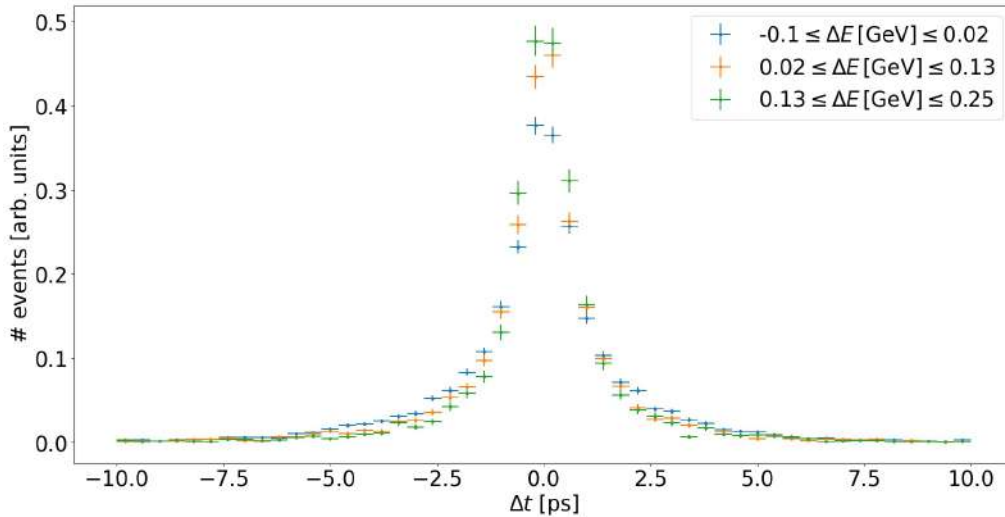


Figure 13: Simulated normalized distributions $G_{\text{bkg}}(\Delta t)$ in bins of event variable ΔE . All signal channels combined.

Clearly, the assumption of a constant $G_{\text{bkg}}(\Delta t)$ distribution is not supported by the simulation. This behavior can be explained in the following way: Different backgrounds exhibit different distributions $G_{\text{bkg}}^i(\Delta t)$. Since the background composition changes with ΔE , so does $G_{\text{bkg}}(\Delta t)$. Our two-dimensional fit strategy in $(\Delta E, \text{BDT})$ is able to separate different backgrounds and extracts the background composition from data. Hence, we expect it to perform better than the one-dimensional fit strategy in ΔE , resulting in lower systematic uncertainties on the measured lifetime τ_{B^0} and mixing parameter Δm . We quantify the improvement by comparing the results

from both fit strategies.

The analysis presented here builds on a previous analysis [44] and employs, besides the detailed background analysis and new fit strategy, an improved resolution function model (see [42]) and uses all available statistics of $176.9 \pm 12.6 \text{ fb}^{-1}$ recorded on $\Upsilon(4S)$ resonance until summer 2021.

5 Sample Processing and Reconstruction

The analysis of backgrounds is performed with simulated event samples. This section introduces the simulation samples and the samples from real data. We explain the reconstruction which is identical for simulation and real data samples.

5.1 Simulation Samples

The production cross-section from e^+e^- collisions at the $\Upsilon(4S)$ resonance does not only consist of B meson pairs but also receives large contributions from non-resonant $q\bar{q}$, $q = (u, d, s, c)$ and lepton pair production. We generate simulation samples for each of the components listed in Table 3. The combined set of these samples simulates 1 ab^{-1} of data at the $\Upsilon(4S)$ resonance. The numbers of events for each contribution depend on the respective cross-sections and are also indicated in Table 3.

Table 3: Generated generic simulation samples for background analysis.

Generic Simulation Sample	Luminosity [fb^{-1}]	Number of Events in all Signal Channels Combined
$B^0\bar{B}^0$	1000	212400 ± 460
B^+B^-	1000	67410 ± 260
$u\bar{u}$	1000	361200 ± 600
$d\bar{d}$	1000	84100 ± 290
$c\bar{c}$	1000	505880 ± 710
$s\bar{s}$	1000	78960 ± 280
$\tau^+\tau^-$	1000	4046 ± 64

Additionally, large samples of $2 \cdot 10^6$ signal-only events are generated for each signal channel. These samples are used in the following to estimate the total signal selection and reconstruction efficiency.

5.2 Data Samples

At the time of writing, the Belle II data samples listed in Table 4 are available for analysis.⁶ The applied reconstruction and selection cuts are identical to the ones applied on simulation samples.

Table 4: Available Belle II data samples for this thesis.

B^0 decay channel	Available luminosity in Belle II data [fb^{-1}]
$B^0 [D^-(K^+\pi^-\pi^-\pi^+)]$	176.9 ± 12.9
$B^0 [D^{*-}[D^0(K^+\pi^+\pi^-\pi^-\pi^-\pi^+)]\pi^+]$	<i>Processing in progress</i> ⁶
$B^0 [D^{*-}[D^0(K^+\pi^-\pi^-\pi^+)]\pi^+]$	176.9 ± 12.9

⁶We do not perform the processing and reconstruction of the data samples ourselves. In one channel, the process was delayed due to malfunctions in the reconstruction of the event shape variables.

5.3 Reconstruction

In simulation, the responses of all sub-detector systems are imitated as accurately as possible. The resulting simulated event samples are then reconstructed identically to the real data samples.

We reconstruct a total of three hadronic decay channels $B^0 \rightarrow D^{(*)-} \pi_{\text{fast}}^+$ with different final states in the subsequent $D^{(*)-}$ decay. Two of them involve a $D^{*-} \rightarrow \bar{D}^0 \pi_{\text{slow}}^-$ transition, in which the resulting pion carries only a small momentum due to the small mass difference between the D^* and D mesons. This fact proves to be most useful in background suppression, as will be shown in Section 6.3.

In a first step of the reconstruction, tracks are reconstructed which are later combined to form the various mesons in the decay chain.

5.3.1 Track Reconstruction

Initially, tracks are selected according to the requirements listed in Table 5.

Table 5: Impact parameter and other requirements for track reconstruction.

Track	dr [cm]	dz [cm]	θ in CDC acceptance	particle ID	p_{cms} [GeV]/ c
π^\pm	< 0.5	$ dz < 3$	✓		
K^\pm	< 0.5	$ dz < 3$	✓	KaonID > 0.01	
π_{fast}^\pm	< 0.5	$ dz < 3$	✓	PionID > 0.01 or KaonID < 0.95	
π_{slow}^\pm					< 0.3

where dz and dr denote the longitudinal and radial impact parameters relative to the interaction point (IP). The IP denotes the collision point of the beams and is measured using di-muon $e^+e^- \rightarrow \mu^+\mu^-$ events [33]. The above cuts are intended to reject beam background tracks that do not originate from an electron-positron annihilation at the IP.

The angle θ denotes the polar angle of the track, which is required to be within the Central Drift Chamber (CDC) acceptance $17^\circ < \theta < 150^\circ$ [38], except for low-momentum pions π_{slow}^\pm . For the latter, a loose cut $p_{\text{cms}} < 0.3 \text{ GeV}/c$ on the momentum in the center-of-mass-frame (cms) is imposed instead.

PionID and KaonID denote the respective particle identification probabilities for a π or K , using the available detector information. For the pions π_{fast}^+ generated in association with the $D^{(*)-}$, additional cuts on the particle identification are required, to reject misidentified kaons. Still, a background $B^0 \rightarrow D^{(*)-} K^+$ remains, which is treated separately from other backgrounds (see section 6.1).

5.3.2 Signal B meson Reconstruction

After having selected tracks, they are combined to form charm mesons and eventually B mesons.

To combine reconstructed tracks to form $D^{(*)\pm}$ mesons, we impose selection cuts on their reconstructed invariant masses m and the free energy in the decay $Q \equiv [m(D^*) - m(D) - m(\pi)]c^2$ (Table 6).

Table 6: Charm meson reconstruction selection cuts.

D decay channel	m [GeV/c^2]	Q [GeV/c^2]
$D^+ \rightarrow K^- \pi^+ \pi^+$	$1.844 < m < 1.894$	
$D^0 \rightarrow K^- \pi^+$	$1.84 < m < 1.89$	
$D^0 \rightarrow K^- \pi^+ \pi^+ \pi^-$		
$D^{*+} \rightarrow D^0 \pi_{\text{slow}}^+$	$1.990 < m < 2.040$	< 0.02

Again, we make use of the fact that very little free energy $Q \approx 5.9 \text{ MeV}/c^2$ [34] is available in the D^{*+} decays, to impose an additional selection cut $Q < 0.02 \text{ GeV}/c^2$ to suppress backgrounds.

Lastly, neutral B mesons are reconstructed from particles and tracks (Table 7),

Table 7: B meson reconstruction selection cuts and efficiencies.

B^0 decay channel	M_{bc} [GeV/c^2]	ΔE [GeV]	ϵ [%]
$B^0 [D^-(K^+ \pi^- \pi^-) \pi^+]$	$5.27 \leq M_{bc} < 5.3$	$-0.1 < \Delta E < 0.25$	41.6
$B^0 [D^{*-} [D^0 (K^+ \pi^+ \pi^- \pi^-) \pi^-] \pi^+]$	$5.27 \leq M_{bc} < 5.3$	$-0.1 < \Delta E < 0.25$	26.4
$B^0 [D^{*-} [D^0 (K^+ \pi^-) \pi^-] \pi^+]$	$5.27 \leq M_{bc} < 5.3$	$-0.1 < \Delta E < 0.25$	37.4

where M_{bc} and ΔE respectively denote the beam-energy-constrained mass and energy difference in the decay:

$$\begin{aligned}
 M_{bc} &\equiv \sqrt{(E_{\text{beam}}^{\text{cms}})^2 - (\vec{p}_B^{\text{cms}})^2} \\
 \Delta E &\equiv E_B^{\text{cms}} - E_{\text{beam}}^{\text{cms}}
 \end{aligned}
 \tag{41}$$

For correctly reconstructed signal decays, each B meson carries half of the collision energy in the cms-frame, i.e. $E_{\text{beam}}^{\text{cms}}$. Hence, their distributions in M_{bc} and ΔE are expected to peak at the true $m_B \approx 5.28 \text{ GeV}/c^2$ mass and at 0 GeV respectively. We use these kinematical constraints to place selection cuts on the reconstructed B mesons in all channels (Table 7). The beam-constrained mass M_{bc} provides a better signal/background discrimination than the reconstructed invariant B meson mass $m_B = \sqrt{E_B^2 - \vec{p}_B^2}$. This is because the beam-energy resolution is smaller than the resolution of E_B , composed from energy measurements of all B meson daughter particles.

The last column in Table 7 denotes the total reconstruction and selection efficiencies ϵ , estimated from the large signal-only simulation samples. Among others, they depend on the number of reconstructed tracks, individual track reconstruction efficiencies and the imposed selection cuts. Here, we only assess the combined effect by counting the correctly reconstructed signal decays.

When reconstructing the meson decay vertices, we let the individual track parameters vary within their errors, in order to converge on the exact standard model masses of the charm mesons. This is called *mass-constraining* the charm mesons and improves the ΔE and M_{bc} resolutions. Figure 14 illustrates this effect of *mass-constraining* on the distribution of simulated signal-only events in ΔE (fitted as described in section 7). The width of the distribution decreases significantly for mass-constrained decays.

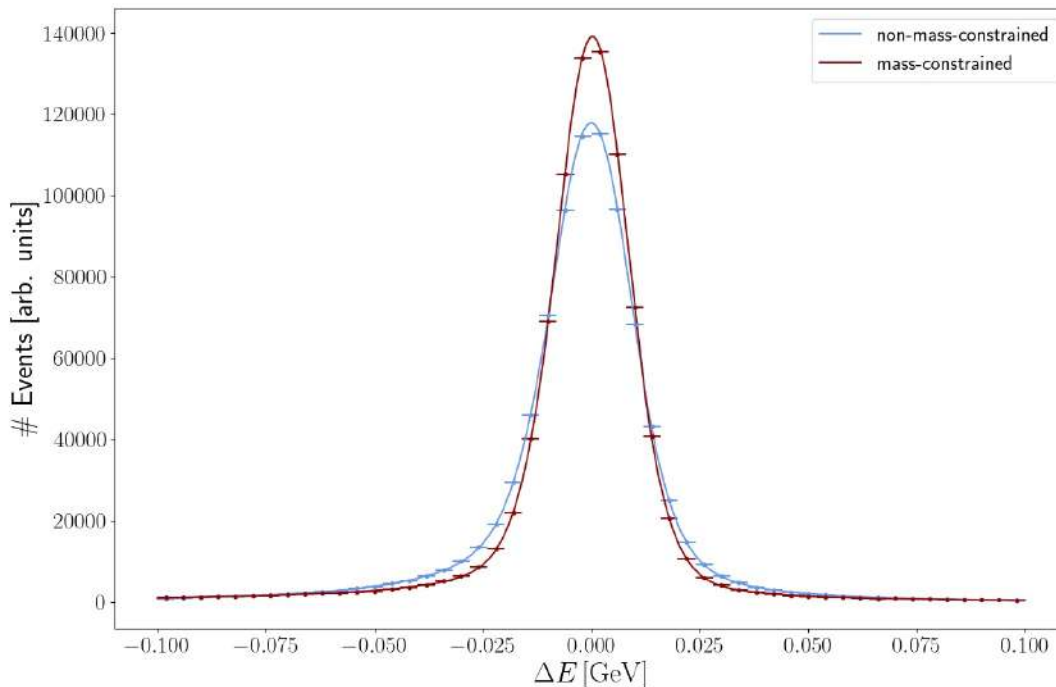


Figure 14: Normalized ΔE distributions for simulated signal-only events in $B^0 [D^-(K^+\pi^-\pi^-)\pi^+]$ with and without *mass-constraining*.

6 Background Analysis for Lifetime and Mixing Measurement

This section covers the detailed background analysis for the lifetime and mixing measurement. The fits to separate backgrounds and signal are described in Chapter 7.

In previous analyses at *Belle* and *Belle II*, so-called *peaking backgrounds* were not evaluated in detail [43][44]. In this background analysis, different types of (peaking) backgrounds are identified and categorized. We first explain the categorization and then apply it on each of the signal channels separately.

6.1 Decay Categorization Procedure

A tool for an event-by-event analysis of the reconstructed generic simulation samples was developed. We access generator-level event variables to compare the reconstructed decay channel to the truly generated decay channel. This tool is used to verify the reconstruction process and to identify *peaking backgrounds*. For the interested reader, its detailed workings are deferred to Appendix A.

Using this tool and several additional event variables (see below), we categorize every reconstructed decay into one of multiple categories. Figure 15 shows the categorization procedure.

We analyse the generic simulation samples (Table 3) and first discriminate between generated events with and without the signal decay. We employ the event variable `reconstructMCdecay`⁷, which returns a binary value indicating whether or not the signal decay was present in the event.

⁷Provided within the `basf2` software framework.

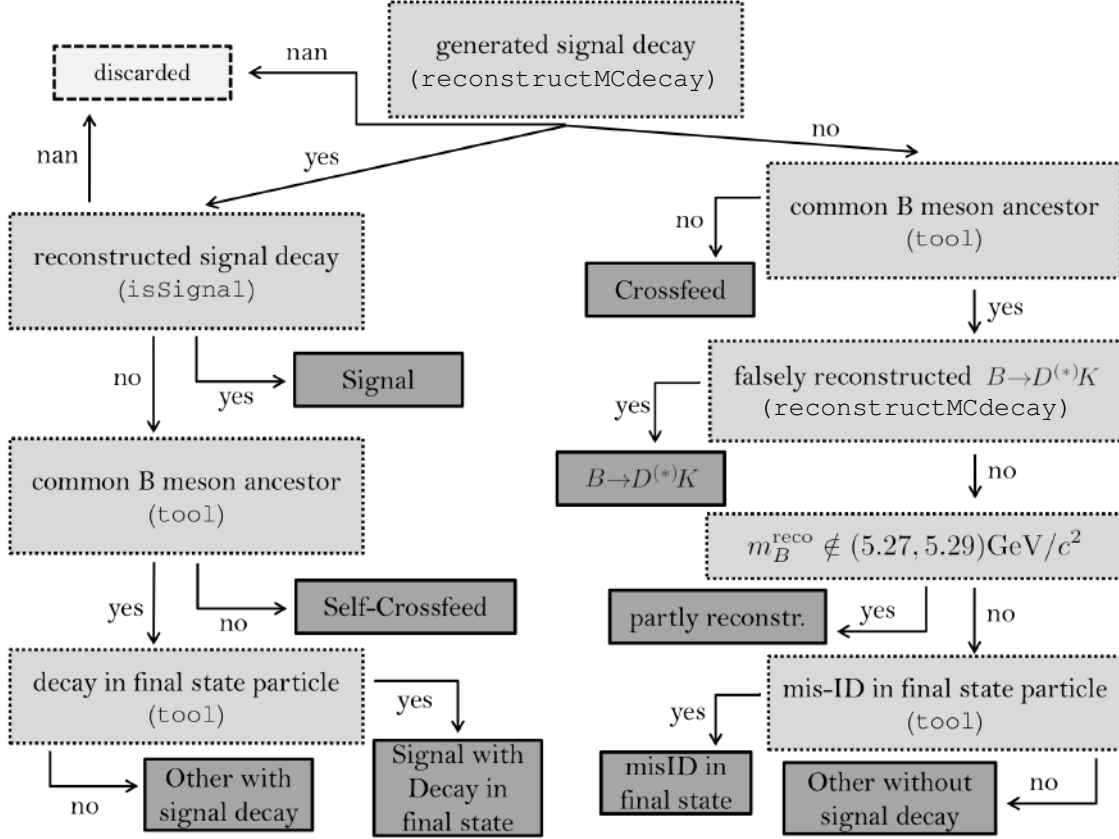


Figure 15: Categorization of reconstructed B meson decays into decay categories for $B^0\bar{B}^0$ and B^+B^- samples.

In the following, events generated with signal decays are described first. Afterwards, events generated with generic non-signal B decays are described.

In events that were **generated with signal decays** (left-hand-side of Figure 15) the signal decay should ideally be reconstructed correctly and efficiently. However in practice, the reconstruction cannot achieve a 100% efficiency. Besides the signal tracks, there are additional tracks in the event. These stem from the accompanying B_{tag} decay or from particles that interact with the detector material, generating secondary tracks. We can reconstruct the decays in the categories listed in Table 8.

Table 8: Reconstructed decay categories for events generated with signal decays.

Decay category	Description
Signal	Correctly reconstructed signal decays.
Signal with decay in final state	Generated signal decays with <i>decays in flight</i> $K \rightarrow \pi(\pi), \pi \rightarrow \mu(\nu)$ or $\mu \rightarrow e(\nu\bar{\nu})$. Reconstructed as signal decay, respectively with mis-identified π, μ, e in final state.
Self-Crossfeed	One B meson decays in signal channel. Reconstructed as signal decay with mix-up of tracks from both B meson-daughters of $\Upsilon(4S)$ decay.
Others with signal decay	Remaining un-categorized decays with generated signal decay, reconstructed as signal decay.

We use the event variable `isSignal`⁷ to extract correctly reconstructed signal decays. `isSignal` requires that all generated final state particles are correctly reconstructed. It does not account for signal decays with subsequent *decays in flight* of the final state particles. These decays are categorized as *signal with decay in final state* and specified in Table 8. The particles in brackets escape the reconstruction, while the remaining are reconstructed as final state particles.

In a negligible number of decays ($\sim 1\%$), the event variables `reconstructMCdecay` and `isSignal` fail to classify the reconstructed decay. This can happen in case of failed truth-matching and we discard these events (see Figure 15).

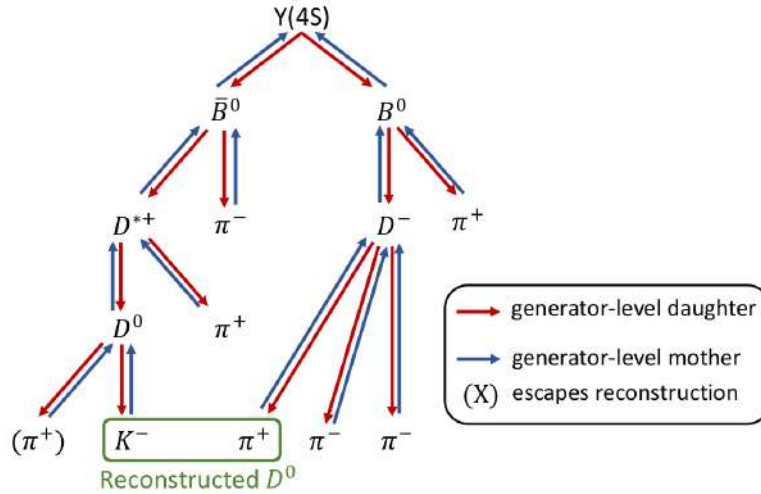


Figure 16: Example of a generator-level decay-chain with track mix-up in reconstruction (green).

After the reconstruction, we identify mix-up of tracks from both B mesons using the developed tool. Particles with the same mother-particle are grouped together, until the full generator-level decay-chain is reassembled. An example for the generator-level decay-chain in the described channels is given in Figure 16. If there is no common identical B meson ancestor for all reconstructed B_{sig} daughter-particles, there has been a mix-up of tracks in the reconstruction, such as in Figure 16. Here a π^+ track from the B_{tag}^0 decay was falsely assigned to the \bar{B}_{sig}^0 reconstruction.

Within the developed tool, this procedure remains functional even when two particles of the same type decay. This can happen, as one of the B mesons may oscillate into its anti-particle and a reliable discrimination between both same-flavored B mesons is necessary.

For future analyses, this method of track mix-up detection was incorporated into a new event variable `isBBCrossfeed` within the Belle II software framework.⁸ It can simply be called in the reconstruction and returns 1 (0) for (no) track mix-up in the reconstruction of a given B meson. Furthermore, it returns `NaN` for failed truth-matching or if it was applied on a particle other than a B meson.

If there is track mix-up in the reconstruction and a signal decay in the event, we call the decay *self-crossfeed* (Table 8). Reconstructed decays with track mix-up and without signal decay in the event are called *crossfeed* (Table 9).

⁸See `basf2-software/analysis/variables/src/MCTruthVariables.cc`. At the time of writing, `isBBCrossfeed` was not yet released collaboration-wide, but is expected to be within release 6 of `basf2`.

We now come to events that were **generated with generic non-signal B decays** (right-hand-side of Figure 15). These decays should ideally be discarded by the selection cuts. However, they may mimic the signal decay, pass the selection requirements and may be falsely reconstructed as signal decays. We can falsely reconstruct these decays as signal decays in the following ways:

Table 9: Reconstructed decay categories for events generated with generic non-signal decays.

Decay category	Description
B to $D^{(*)}K$	Generated $B^0 \rightarrow D^{(*)-}K^+$ decays reconstructed as signal decay with mis-identification of K^+ as π^+ in reconstruction.
Partly Reconstructed	Generated non-signal decays reconstructed as signal decay, where not all tracks of the generated decay are reconstructed, i.e. missing four-momentum in reconstruction.
Mis-identified final state	Generated non-signal decays reconstructed as signal decay, where the particle type of a final state particle is mis-identified. This does not include $B^0 \rightarrow D^{(*)-}K^+$ decays.
Crossfeed	Neither of both B mesons decay in signal channel. Reconstructed as signal decay with mix-up of tracks from both B meson-daughters of $\Upsilon(4S)$ decay.
Others without signal decay	Remaining un-categorized decays without generated signal decay, reconstructed as signal decay.

We explicitly separate out falsely reconstructed $B^0 \rightarrow D^{(*)-}K^+$ decays, since they represent a significant fraction of the *peaking backgrounds*. In the reconstruction of these decays, the PID from the detector responses fails and identifies the kaon as pion.

Furthermore, we identify non-signal decays *partly reconstructed* as signal decays through the invariant B meson mass m_B :

$$m_B = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i p_i^x\right)^2 - \left(\sum_i p_i^y\right)^2 - \left(\sum_i p_i^z\right)^2} \quad \text{with } i \in \text{reconstr. tracks}$$

where i runs over all reconstructed B_{sig} daughter-particles and $E, p^{(x,y,z)}$ are their generated energies and momenta. For partly reconstructed tracks, not all daughter-particles of the generated decay are reconstructed and $m_B < 5.28$ GeV.

In the following, the previously described decay categorization is applied to all three signal channels. The continuum backgrounds $\tau^+\tau^-$ and $q\bar{q}$ are added from their separate simulation samples.

6.2 Decay Categorization in $B^0 \rightarrow \pi^+ D^-(\pi^- \pi^- K^+)$

Figure 17 shows the simulated distributions of reconstructed decays in the energy-difference ΔE and in the beam-energy-constrained mass M_{bc} (see the definitions in Equation 41) in the signal channel $B^0 \rightarrow \pi^+ D^-(\pi^- \pi^- K^+)$. Table 10 shows the numbers of reconstructed decays categorized in each decay category.

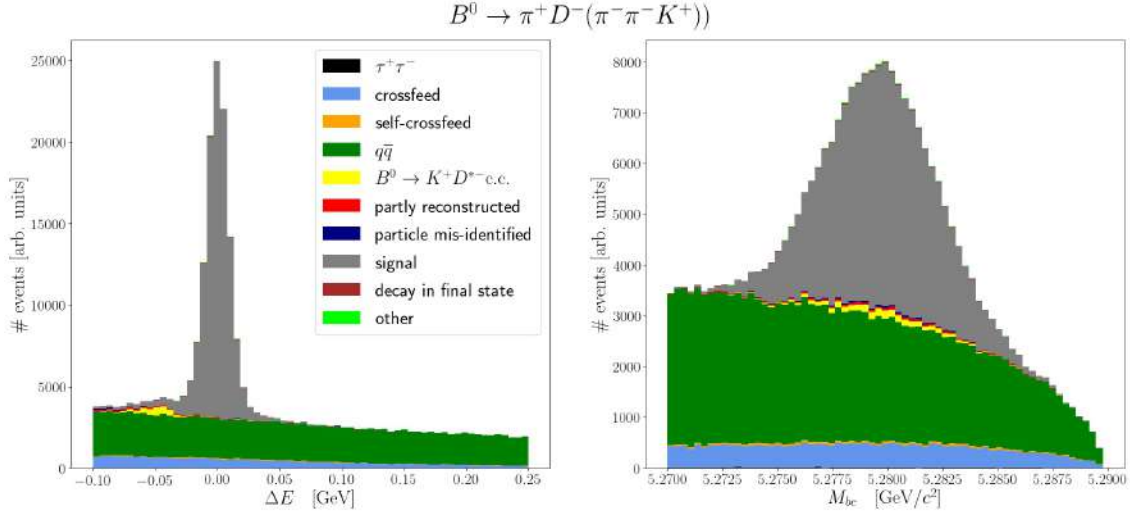


Figure 17: Reconstructed simulated decays and decay categories. Left: Energy difference ΔE . Right: beam-energy-constrained mass M_{bc} .

Table 10: Abundances of decay categories in the $B^0 \rightarrow \pi^+ D^-(\pi^- \pi^- K^+)$ signal channel reconstruction in 1 ab^{-1} of simulation.

Decay Category	Number of Reconstructed Decays in 1 ab^{-1} of Simulation
$\tau^+ \tau^-$	773 ± 28
Crossfeed	25160 ± 160
Self-Crossfeed	2329 ± 48
$q\bar{q}$	145980 ± 380
$B^0 \rightarrow K^+ D^{(*)-} + \text{h.c.}$	2913 ± 54
Partly reconstructed	1472 ± 38
Particle mis-identified	764 ± 28
Signal	100130 ± 320
Decay in final state	459 ± 21
Other with(out) signal decay	399 ± 20

As expected, the signal contribution peaks at $\Delta E \approx 0 \text{ GeV}$ and $M_{bc} \approx 5.28 \text{ GeV}$. Due to imperfect detector resolution and beam energy measurements, the distributions have a finite width. The range $\Delta E > 0.05 \text{ GeV}$ is called *sideband* in ΔE and contains a negligible amount of signal decays. Still, we do not discard decays in the sideband, in order to retain some information about the continuum backgrounds (see section 7.4 and Chapter 8).

The number of $\tau^+ \tau^-$ events that are falsely reconstructed as signal is small. This is expected from the selection cuts and since we require 4 charged tracks in the reconstruction (the τ branching fractions to 4 or more charged particles are $\leq \mathcal{O}(10^{-3})$ [34]).

The contributions from crossfeed and self-crossfeed consist of decays with a mix-up of tracks from both B meson decays in the reconstruction of B_{sig} . A vast number of different decays can be falsely reconstructed as signal decay, when mixing up tracks from both B meson decays. We can list only the most abundant decays with crossfeed in the reconstruction in Table 11. As can be seen in Table 11, also charged B mesons decays can be falsely reconstructed as B^0 signal decay, when assigning tracks from the B_{tag} decay to it.

Table 11: Selected decays reconstructed as $B^0 \rightarrow \pi^+ D^- (\pi^- \pi^- K^+)$ with crossfeed in reconstruction.

Decay Reconstructed as B_{sig} decay	Particles from B_{tag} decay assigned to B_{sig} Reconstruction	Abundance
$B^+ \rightarrow \pi^+ \bar{D}^0 (\pi^- K^+)$	$B^- \rightarrow \pi^- X$	33%
$B^+ \rightarrow \pi^+ \bar{D}^{*0} (X \bar{D}^0 (\pi^- K^+)), \quad X = \pi_{\text{slow}}^0$	$B^- \rightarrow \pi^- X$	13%
$B^0 \rightarrow \pi^+ D^{*-} (X \bar{D}^0 (\pi^- K^+)), \quad X = \pi_{\text{slow}}^-$	$\bar{B}^0 \rightarrow \pi^- X$	3%
...		

where X are particles that escape reconstruction and a π^- from the accompanying B_{tag} decay is falsely assigned to the signal decay. Especially low-momentum pions from D^* decays frequently escape detection and reconstruction.

Decays with self-crossfeed in the reconstruction consist of correctly reconstructed signal decays, except for a missing D^- daughter-particle that is falsely assigned from the accompanying B_{tag} decay to the B_{sig} reconstruction.

A large fraction of falsely reconstructed decays consists of the $q\bar{q}$ continuum background. Quark anti-quark production at the cms-energy $\sqrt{s} = m_{\Upsilon(4S)} c^2$ is about three times more frequent than the $\Upsilon(4S)$ resonance formation itself. We only detect the resulting hadronized particles for which the reconstructed energies and masses stretch over the whole range of ΔE and M_{bc} .

Falsely reconstructed $B^0 \rightarrow D^{(*)-} K^+$ decays represent a significant fraction of the *peaking backgrounds*. The K^+ track can be mis-identified as π^+ , i.e. as a particle with smaller rest mass. Since the momentum $|\vec{p}|$ of the kaon is precisely determined by its gyroradius in the solenoid field, the energy of its track must be underestimated to meet the lower pion mass. For the shifted $\Delta E'$, we find

$$\Delta E - \Delta E' = E_B - E'_B = E_\pi - E_K = \sqrt{m_\pi^2 + \vec{p}^2} - \sqrt{m_K^2 + \vec{p}^2} \approx \frac{m_\pi^2 - m_K^2}{2|\vec{p}|},$$

where the square roots were approximated in Taylor series up to first order in $m^2/\vec{p}^2 \ll 1$. For typically measured momenta $|\vec{p}| \approx 2.5 \text{ GeV}/c$ in the laboratory frame, we find approximately

$$\Delta E - \Delta E' \approx -0.045 \text{ GeV},$$

in agreement with the simulation (Figure 17). The beam-energy-constrained mass depends only on the reconstructed momentum \vec{p}_B and is hence unaffected.

Partly reconstructed decays miss some tracks and hence four-momentum in the reconstruction. Therefore they also peak at $\Delta E < 0 \text{ GeV}$ and the shape in M_{bc} is smeared out. Some common non-signal decays partly reconstructed as signal involve D^* decays where low-momentum pions escape reconstruction, incompletely reconstructed ρ resonance decays and semileptonic decays of charm mesons with undetected neutrinos (Table 12).

In the leftmost column in Table 12 the truly generated decay is shown, which is falsely reconstructed as signal decay. Here, the most common partly reconstructed decay is the decay $B^+ \rightarrow \pi^+ \bar{D}^0 (\pi^- K^+ \pi^0)$ of a charged B meson. In the reconstruction, the charged particles are correctly identified. The subsequent π^0 decay involves an electron, which is mis-identified as π^- in the reconstruction. The other daughter particles of the π^0 escape the reconstruction (X in Table 12). The reconstructed final state is hence $(\pi^+, \pi^-, \pi^-, K^+)$, from which the signal decay is reconstructed. From the kinematics of these four particles in the final state, the B meson mass is reconstructed. Since the generated B meson was not fully reconstructed, $m_B^{\text{reco}} < 5.28 \text{ GeV}/c^2$ and the decay is categorized as *partly reconstructed*.

Table 12: Selected decays that are partly reconstructed as $B^0 \rightarrow \pi^+ D^- (\pi^- \pi^- K^+)$.

Decay Partly Reconstructed as B_{sig}^0 Decay	Missing Particle in Reconstruction	Abundance
$B^+ \rightarrow \pi^+ \bar{D}^0 (\pi^- K^+ \pi^0 (X e^-))$	$X = \gamma, e^+$	20%
$B^0 \rightarrow \pi^+ D^{*-} (X D^- (\pi^- \pi^- K^+))$	$X = \pi_{\text{slow}}^0$	14%
$B^0 \rightarrow \rho^+ (\pi^+ X) D^- (\pi^- \pi^- K^+)$	$X = \pi^0$	13%
$B^0 \rightarrow \pi^+ D^- (\pi^- K^+ l^- X)$	$X = \bar{\nu}$	13%
$B^+ \rightarrow X \bar{D}^{*0} (\pi^+ D^- (\pi^- \pi^- K^+))$	$X = \pi_{\text{fast}}^+$	10%
$B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (l^- X K^+))$	$X = \bar{\nu}$	8%
...		

Decays reconstructed as signal decays, with mis-identified final states in the reconstruction, mostly involve kaon/pion misidentifications and less often lepton/pion misidentifications. Some common decays are listed in Table 13.

Table 13: Selected decays reconstructed as $B^0 \rightarrow \pi^+ D^- (\pi^- \pi^- K^+)$ with mis-identified final states.

Decay with Final State Mis-identified as B_{sig}^0	Abundance
$B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (K^- K^+))$	54%
$B^0 \rightarrow \pi^- D_s^+ (\pi^+ K^+ K^-)$	14%
$B^0 \rightarrow J/\psi(1S) (l^+ l^-) \bar{K}^{*0} (K^- \pi^+)$	4%
...	

The number of decays categorized into *other with(out) signal decay* are small (see Table 10). More than 95% of these are prompt decays $B^0 \rightarrow \pi^- \pi^+ \pi^- K^+$, some of them with short-lived (strongly decaying) resonances, such as ρ^0 and K^{*0} . Since these decays do not involve a charm meson decay, their reconstructed D meson mass distribution does not peak at $m_D \approx 1.87 \text{ GeV}/c^2$, as can be seen from the distributions in Figure 18. Instead, the distribution is homogeneous in m_D and an additional selection cut cannot be imposed without decreasing the signal reconstruction efficiency.

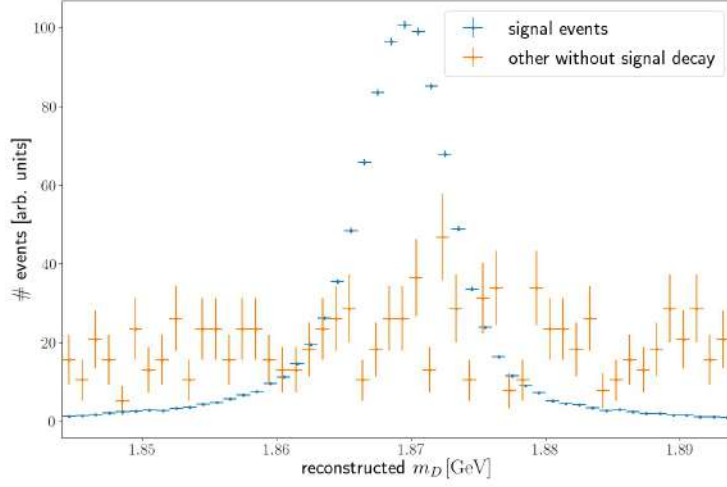


Figure 18: Normalized distributions in simulation. Reconstructed D meson mass for signal and prompt decays.

6.3 Decay Categorization in $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$

Figure 19 shows the simulated distributions of reconstructed decays in the energy-difference ΔE and in the beam-energy-constrained mass in the signal channel $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$. Table 14 shows the numbers of reconstructed decays categorized in each decay category.

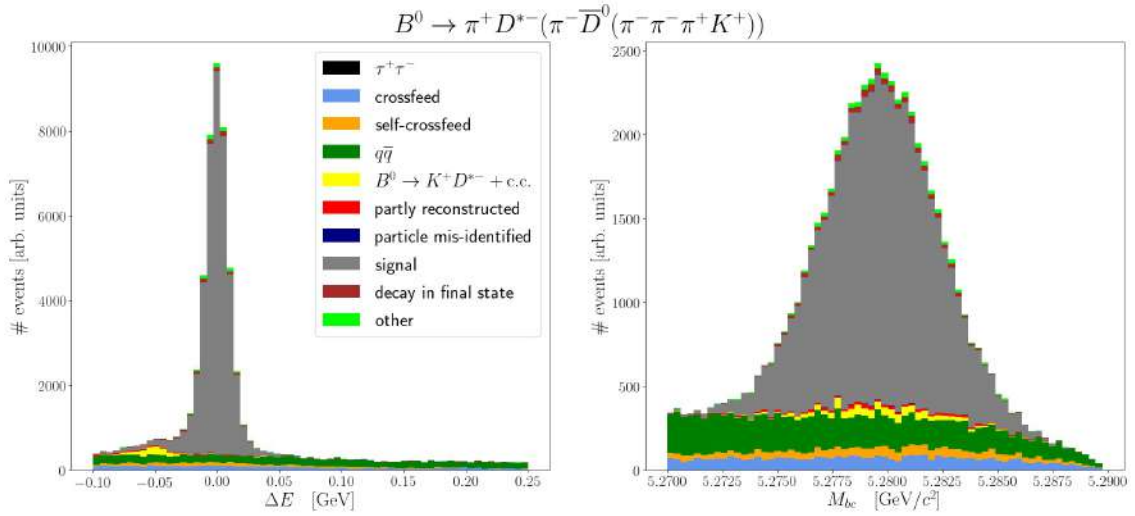


Figure 19: Reconstructed simulated decays and decay categories. Left: Energy difference ΔE . Right: beam-energy-constrained mass M_{bc} .

Table 14: Abundances of decay categories in the $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$ signal channel reconstruction in 1 ab^{-1} of simulation.

Decay Category	Number of Reconstructed Decays in 1 ab^{-1} of Simulation
$\tau^+ \tau^-$	2 ± 1
Crossfeed	4134 ± 64
Self-Crossfeed	2630 ± 51
$q\bar{q}$	11140 ± 110
$B^0 \rightarrow K^+ D^{(*)-} + \text{h.c.}$	1182 ± 34
Partly reconstructed	545 ± 23
Particle mis-identified	25 ± 5
Signal	40780 ± 200
Decay in final state	866 ± 29
Other with(out) signal decay	725 ± 27

Compared to the previously described signal channel $B^0 \rightarrow \pi^+ D^- (\pi^- \pi^- K^+)$, this signal channel exhibits significantly fewer decays at the same luminosity. This depends on the branching fractions of the decays (see Table 1) and the number of reconstructed tracks. The track reconstruction efficiency depends on the momentum and every additional reconstructed track reduces the overall reconstruction efficiency. Low-momentum π_{slow} have an estimated track reconstruction efficiency of $\epsilon \sim 75\%$ and for the other charged particles $\epsilon \sim 95\%$ [45].

From the decreased number of background decays, the strong background suppression through the additional selection cuts on the D^* decay is immediately evident. A very small number of continuum $\tau^+ \tau^-$ events is falsely reconstructed as signal. As before, decays with crossfeed in the reconstruction involve low-momentum pions from D^* decays and \bar{D}^0 daughter-particles that are falsely assigned from the B_{tag} decay (Table 15) to the B_{sig} reconstruction.

Table 15: Selected decays reconstructed as $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$ with crossfeed in reconstruction.

Decay Reconstructed as B_{sig} decay	Particles from B_{tag} decay assigned to B_{sig} Reconstruction	Abundance
$B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^+ K^-))$	$\bar{B}^0 \rightarrow \pi^+ \pi^- X$	37%
$B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^+ K^- \pi^- X)), X = \pi^+$	$\bar{B}^0 \rightarrow \pi^+ X$	
$B^+ \rightarrow \pi^+ D^{*0} (X \bar{D}^0 (\pi^- \pi^- \pi^+ K^+)), X = \pi_{\text{slow}}^0$	$B^- \rightarrow \pi^- X$	19%
$B^+ \rightarrow \pi^+ \bar{D}^0 (\pi^- \pi^- \pi^+ K^+)$	$B^- \rightarrow \pi^- X$	14%
...		

Falsely reconstructed $B^0 \rightarrow K^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$ decays are the main contribution to the *peaking backgrounds*. Since they involve a $D^* \rightarrow D\pi$ decay, their reconstruction is not suppressed by the additional selection cuts. For the majority of reconstructed decays with self-crossfeed, the D^* decay is correctly reconstructed and a \bar{D}^0 daughter-particle is falsely assigned from the B_{tag} decay to the B_{sig} reconstruction.

Decays partly reconstructed as signal decays are shown in Table 16. Again, neutral pions π^0 are only partly reconstructed and electrons misidentified as π^- .

Table 16: Selected decays that are partly reconstructed as $B^0 \rightarrow \pi^+ D^{*-}(\pi^- \bar{D}^0(\pi^- \pi^- \pi^+ K^+))$

Decay Partly Reconstructed as B_{sig} Decay	Missing Particle in B_{sig} Reconstruction	Abundance
$B^+ \rightarrow \pi^+ D^{*0}(\pi^0(Xe^-)\bar{D}^0(\pi^- \pi^- \pi^+ K^+))$	$X = \gamma, e^+$	28%
$B^0 \rightarrow \pi^+ D^{*-}(\pi^- \bar{D}^0(\pi^- K^+ \pi^0(Xe^+e^-)))$	$X = \gamma$	14%
$B^0 \rightarrow \rho^+(\pi^+ X)D^{*-}(\pi^- \bar{D}^0(\pi^- \pi^- \pi^+ K^+))$	$X = \pi^0$	6%
...		

The number of decays with kaon/pion and lepton/pion misidentifications in the reconstructed final state is small (see Table 14). Of the decays categorized as *other with(out) signal decays*, 25% are decays of the type $D^0 \rightarrow \pi^+ K^- K_S^0(\pi^+ \pi^-)$. The weakly decaying K_S^0 resonance has a non-negligible lifetime of 89.5 ps and can broaden the distribution in the decay time difference Δt . An additional selection cut on the reconstructed mass of any two pions is impractical, as it severely cuts into the signal reconstruction efficiency.

6.4 Decay Categorization in $B^0 \rightarrow \pi^+ D^{*-}(\pi^- \bar{D}^0(\pi^- K^+))$

Figure 20 shows the simulated distributions of reconstructed decays in the energy-difference ΔE and in the beam-energy-constrained mass in the signal channel $B^0 \rightarrow \pi^+ D^{*-}(\pi^- \bar{D}^0(\pi^- K^+))$. Table 17 shows the numbers of reconstructed decays categorized in each decay category.

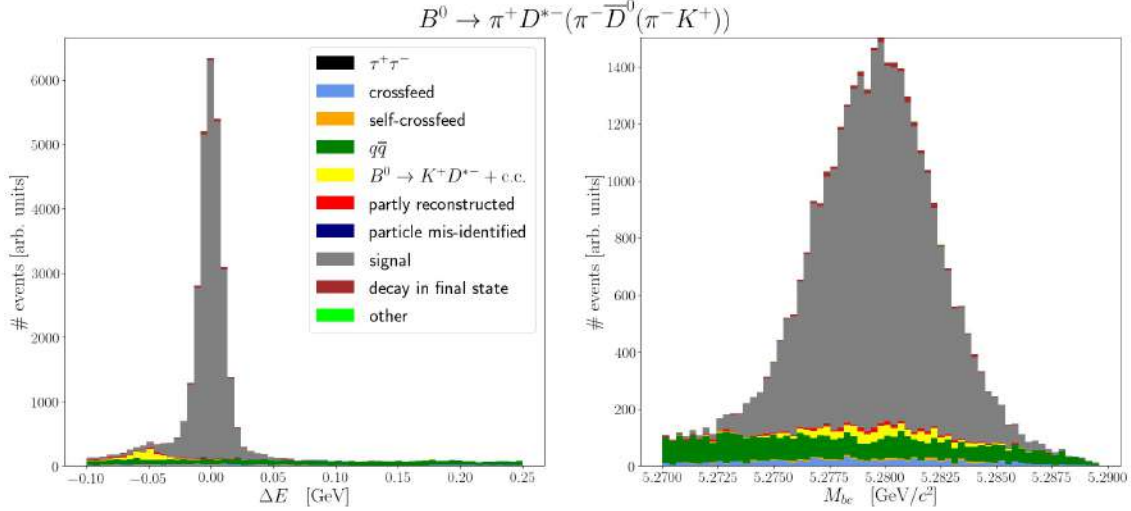


Figure 20: Reconstructed simulated decays and decay categories. Left: Energy difference ΔE . Right: beam-energy-constrained mass M_{bc} .

Table 17: Abundances of decay categories in the $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- K^+))$ signal channel reconstruction in 1 ab^{-1} of simulation.

Decay Category	Number of Reconstructed Decays in 1 ab^{-1} of Simulation
$\tau^+ \tau^-$	4 ± 2
Crossfeed	878 ± 30
Self-Crossfeed	199 ± 14
$q\bar{q}$	4392 ± 66
$B^0 \rightarrow K^+ D^{(*)-} + \text{h.c.}$	814 ± 29
Partly reconstructed	301 ± 17
Particle mis-identified	7 ± 3
Signal	27660 ± 170
Decay in final state	338 ± 18
Other with(out) signal decay	12 ± 3

Compared to the previously described signal channel $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$, here the number of signal decays is lower due to a lower branching fraction (see Table 1). This channel exhibits low (self-) crossfeed background contaminations, compared to the previously discussed signal channels (Table 10 and 14). This is expected, since the \bar{D}^0 decays to only 2 charged tracks. The lower the number of reconstructed tracks, the lower the probability for mix-up of any tracks in the reconstruction.

In the few remaining reconstructed decays with self-crossfeed, low-momentum pions in the D^* decays are falsely assigned from the accompanying B_{tag} decay to the B_{sig} decay. Due to the additional selection cuts, these decays are strongly suppressed. In reconstructed decays with crossfeed, there is an additional component of $B^+ \rightarrow \pi^+ \bar{D}^0 (\pi^- K^+)$ decays with an additional π^- track from the B_{tag} decay assigned to the B_{sig} decay.

There is no contamination from decays of K_S^0 resonances, since we require only one charged pion in the \bar{D}^0 decay. Decays *partly reconstructed* as signal decays are analogous to the ones presented for the $B^0 \rightarrow \pi^+ D^{*-} (\pi^- \bar{D}^0 (\pi^- \pi^- \pi^+ K^+))$ signal channel (Table 16).

7 Fit Strategy to Separate Backgrounds and Signal

The previous section described a detailed background analysis, performed in generic simulation samples. Of course, in data no generator-level event information is available as in the simulation samples. Hence, to separate the backgrounds and signal, we perform a fit to the measured distributions of selected event variables. As explained in Chapter 4, we extract the shapes of the distributions of backgrounds and signal from simulation. These shapes are then fitted onto the measured distributions to separate out true signal decays. The fits are first validated in simulation samples and pseudo-experiments. The final lifetime and mixing measurement is then performed with *Belle II* data.

From the background analysis in Chapter 6, we know the detailed background contributions and how they are reconstructed. While the background analysis was aiming to understand the backgrounds in as much detail as possible, the fit does not aim to separate all categories described in Section 6.1. The complexity of the fit increases with the number of backgrounds to separate. For the lifetime and mixing measurement, we only need to separate backgrounds that differ in their behavior in Δt (see Chapter 4). Hence for the fit we apply a simplified decay categorization, compared to the one in Section 6.1, which copes with fewer categories.

7.1 Simplified Decay Categorization

The simplified decay categorization can be obtained by combining categories in Section 6.1 or simply by the procedure shown in Figure 21. Since all described channels are sensitive to the same lifetime τ_{B^0} and mixing parameter Δm , we combine the samples of all reconstructed channels to obtain the best possible statistics in the individual distributions to be fitted. This is especially useful, since data-taking at *Belle II* is still in an early phase, with $176.9 \pm 12.6 \text{ fb}^{-1}$ collected on $\Upsilon(4S)$ resonance until summer 2021. We separate out signal decays and falsely reconstructed

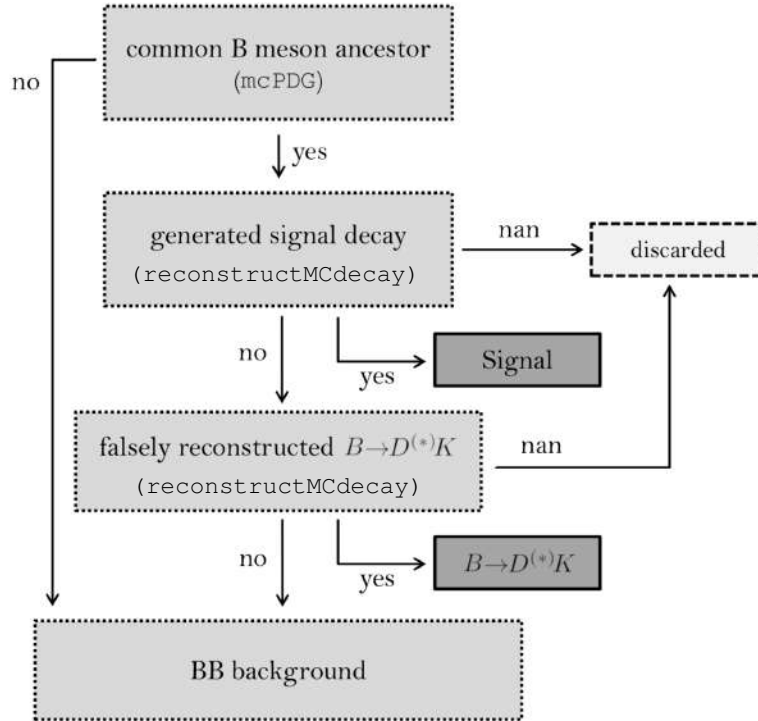


Figure 21: Simplified decay categorization for $B^0\bar{B}^0$ and B^+B^- samples.

$B \rightarrow D^{(*)}K$ decays as before and call all remaining decays *BB backgrounds*. Since *other signal decays* and *signal decays with decays in the final state* (Section 6.1) constitute only a negligible fraction of the decays and their distributions in the fit variables provide no distinctive features, we do not separate them from signal.

$\tau^+\tau^-$ and $q\bar{q}$ backgrounds are combined as *continuum backgrounds*. As can be seen in Figure 22, they exhibit a significantly narrower distribution in Δt than signal and non-signal B meson decays. This is expected, since the *continuum background* tracks originate directly from the vertex at the beam interaction point.

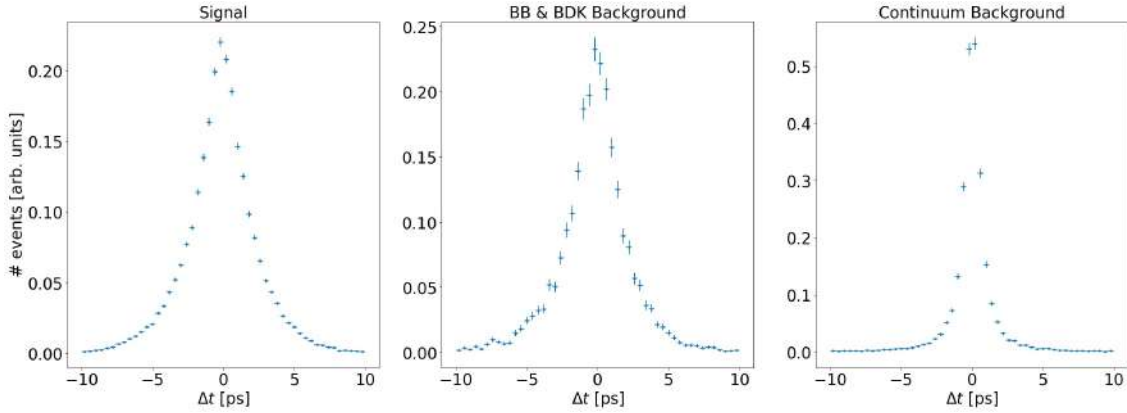


Figure 22: Simulated normalized distributions in decay time difference Δt . All signal channels combined.

Figure 23 illustrates the composition of the total simulation sample. The inner pie chart features the simplified decay categories, while the outer pie chart shows the components in the detailed background analysis.

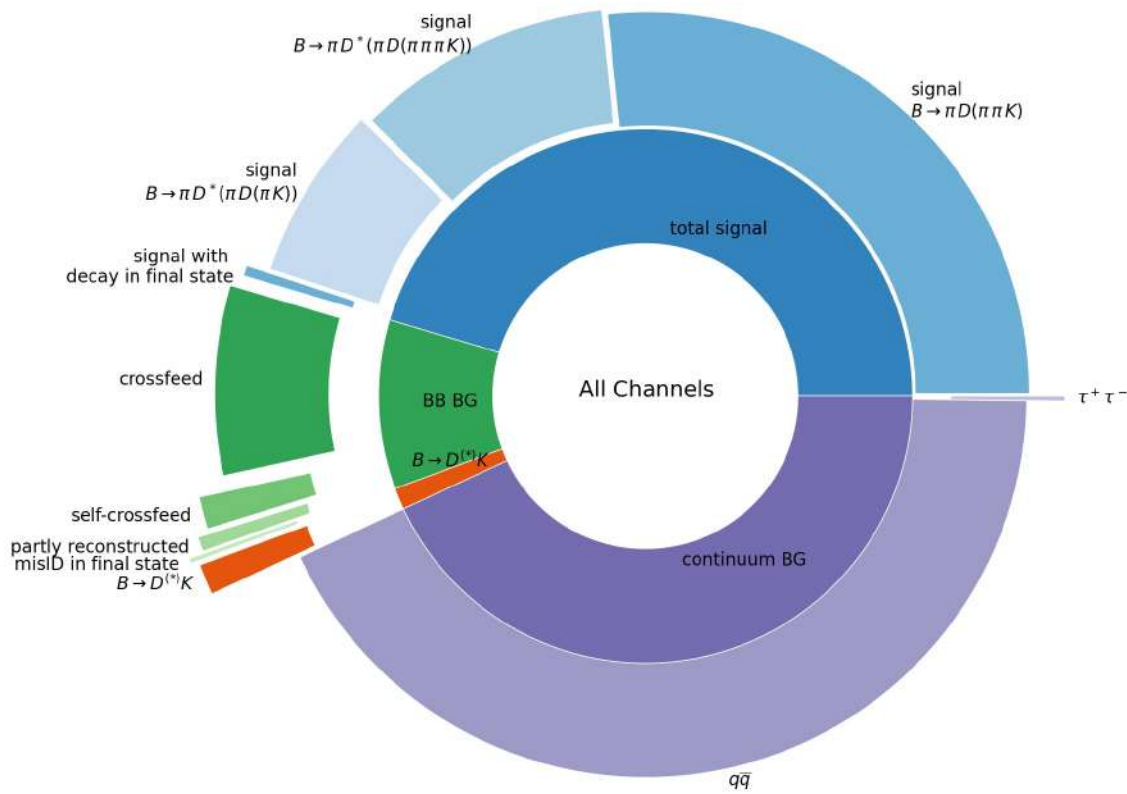


Figure 23: Composition of full simulation sample of all channels combined.

7.2 Event Variables for Fit

To separate backgrounds and signal in data, we perform fits in ΔE and the output of a continuum suppression *Boosted Decision Tree*, here denoted BDT. The latter was developed in [46] and is

briefly introduced in this section. Furthermore, the correlations between ΔE , BDT and Δt are assessed.

7.2.1 Continuum Suppression Boosted Decision Tree output (BDT)

As explained above, quark-antiquark ($q = u, d, s, c$) pairs are produced with a cross section about three times larger than the $B\bar{B}$ cross section at the $\Upsilon(4S)$ resonance and constitute the main background source in the signal reconstruction. A strong suppression of continuum backgrounds can be achieved by requirements on the so-called event shape variables. Since the hadronized quark-antiquark pairs carry much smaller masses than the hadronized B meson pairs, the former carry much larger momenta. This results in a back-to-back topology of the hadronized quarks, with small transverse momenta of the individual tracks (see Figure 24). In contrast to this, B mesons are almost at rest in the cms-frame and their decay products are therefore isotropically distributed.

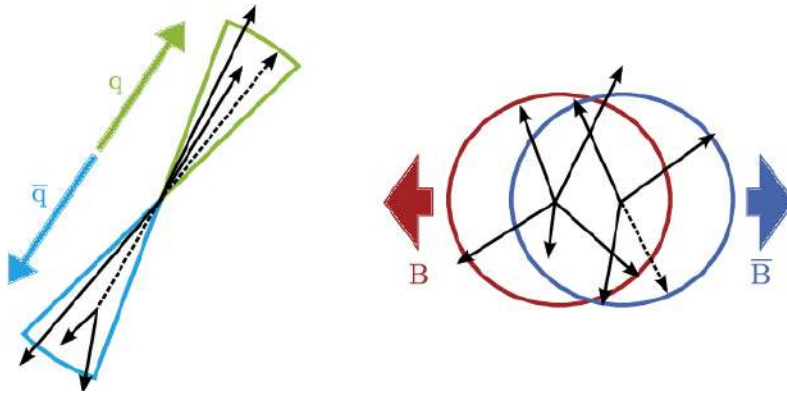


Figure 24: Event shape for continuum $q\bar{q}$ jets and B meson pair decay [47].

A continuum suppression boosted decision tree (CS-BDT) is applied, in order to separate out continuum background decays with high reliability. This CS-BDT uses event-shape variables as discrimination variables. For every decay, the CS-BDT returns an output classifier BDT, estimating the probability for a continuum background decay reconstructed as signal. An output of 0 indicates the highest probability for a continuum background decay, while an output of 1 indicates the lowest probability.

Figure 25 shows the distributions of decays in ΔE and BDT for the full simulation sample combining all channels. The strong separation power of the CS-BDT between continuum background decays and other decays is clearly illustrated. We impose a loose selection cut $\text{BDT} > 0.05$ on all channels, which removes 48.7% of continuum background decays, but retains 98.6% of the signal decays. We cannot remove all continuum background decays with a single cut on BDT, without severely cutting into the signal efficiency. This cut merely serves as loose pre-selection of events, before applying the developed fit strategies.

7.2.2 Correlations between ΔE and BDT

When fitting simultaneously in multiple variables (we simultaneously fit the one-dimensional distributions in ΔE and BDT), we need to make sure that they are independent, i.e. not correlated. This is because each of the two one-dimensional distributions has no knowledge about the other and they are implicitly assumed "independent" of each other.

Figure 26 illustrates the correlations between ΔE and BDT in the total simulation sample, separately for signal and backgrounds. The left column shows the normalized distributions in ΔE ,

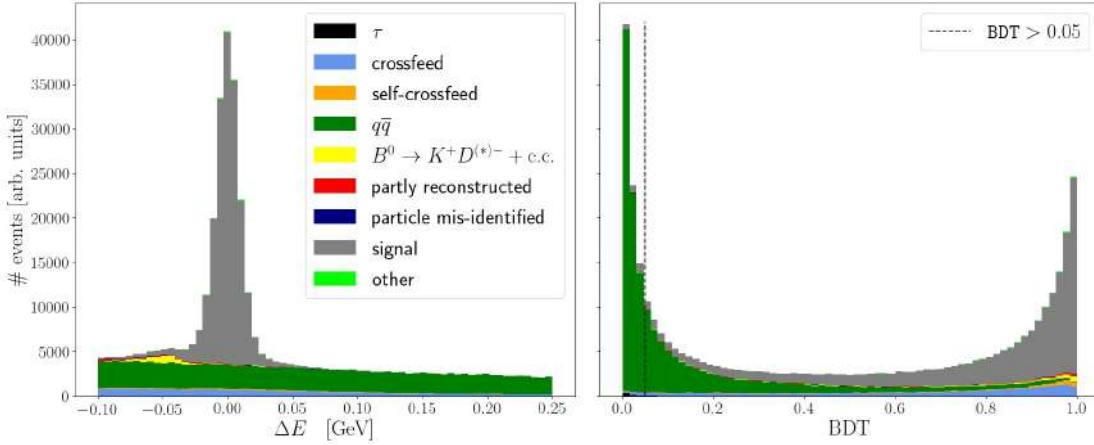


Figure 25: Distributions of decays in ΔE and continuum suppression BDT output for full simulation sample combining all channels.

in three different regions of BDT. The close agreement of the shapes shows their independence of BDT. The right column of Figure 26 shows the corresponding density plots in the $(\text{BDT}, \Delta E)$ plane.

To quantitatively determine the linear (in-) dependence of two variables X and Y , one computes their correlation coefficient

$$\rho_{X,Y} \equiv \frac{\text{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])]}{\sigma_X \cdot \sigma_Y},$$

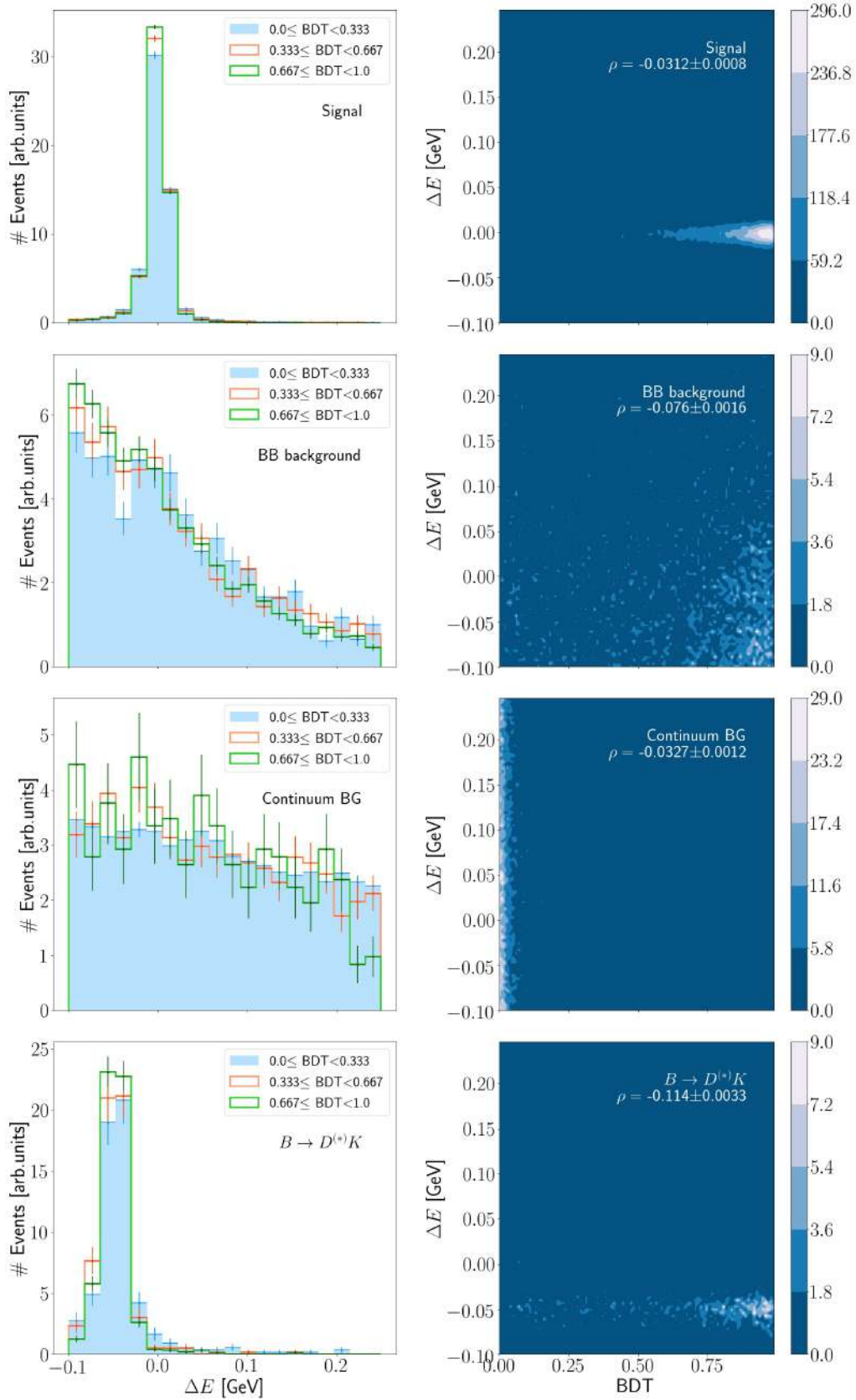
i.e. the normalized covariance, with expectation value \mathbb{E} . If increasing values in X yield increasing (or decreasing) values in Y , the variables are linearly dependent and their correlation is non-zero. In Figure 26, we indicate the correlations $\rho_{\Delta E, \text{BDT}}$ for signal and backgrounds. They are small, compared to a maximal correlation of 1, and confirm the validity of simultaneous fits in ΔE and BDT.

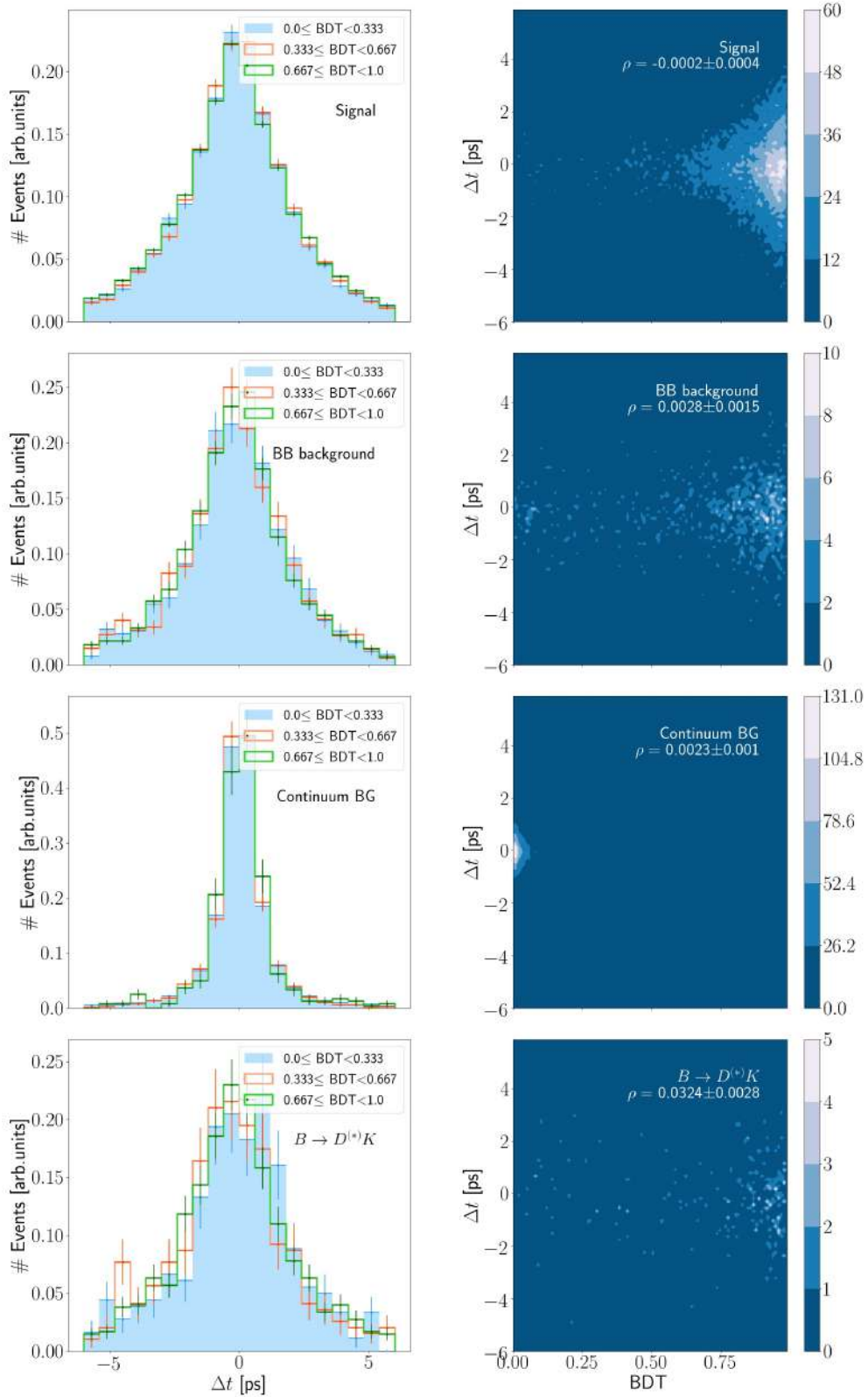
We *bootstrap* the simulation samples to estimate the confidence intervals for the correlation coefficients ρ . In bootstrapping we sample from the original samples, i.e. randomly choose pairs of data points from the original samples. We sample *with replacement*, such that a particular data point can be chosen multiple times. In this way, we use the original samples as estimator for the underlying distributions that are unknown. Multiple samplings of a data point are understood as multiple independent samplings from this distribution [6]. We sample repeatedly and generate 200 samples of the same size as the original sample. For each of the generated samples, we compute the correlation coefficient ρ and estimate the confidence intervals from the distribution in ρ .

7.2.3 Correlations between Δt and BDT

As introduced in Chapter 4, the fit strategy for the lifetime and mixing measurement consists in the two-dimensional fit in $(\text{BDT}, \Delta E)$ and the time-dependent fit. Again, we assume that the variables are independent in our fit. Therefore, we also need to assess the correlations between Δt and $(\text{BDT}, \Delta E)$. Figure 27 shows the correlations between Δt and BDT in the total simulation sample, analogously to Figure 26. The correlations ρ are at most at the percent level (Figure 27), approving the assumption of independent distributions in Δt and BDT.

As mentioned in Chapter 4, we do not fit in $(\Delta E, M_{bc})$ due to larger correlations in the signal. The corresponding plots are deferred to Appendix B.

Figure 26: Correlations between ΔE and BDT in total simulation sample combining all channels.

Figure 27: Correlations between Δt and BDT in total simulation sample combining all channels.

7.3 Extraction of Fit Shapes from Simulation Samples

For the fits in ΔE and BDT, the shapes for backgrounds and signal are extracted from simulation samples. The extraction of fit shapes is performed separately for same- and opposite-flavor events and separately in the 7 dilution bins.

7.3.1 Extraction of Fit Shapes in Continuum Suppression BDT

Instead of directly fitting BDT, we fit its logarithmic transform LTBDT,

$$\text{LTBDT}(\text{BDT}) \equiv \ln \left[\frac{\text{BDT} - \text{BDT}_{\min}}{\text{BDT}_{\max} - \text{BDT}} \right] = \ln \left[\frac{\text{BDT} - 0.05}{1 - \text{BDT}} \right],$$

where BDT_{\min} and BDT_{\max} denote the minimal and maximal BDT output values respectively. LTBDT provides a shape that is preferably fitted with a Johnson S_U probability density function⁹

$$J(y; s, t) = \frac{t}{\sqrt{2\pi(1+y^2)}\sigma} \cdot \exp \left[-\frac{1}{2} (s + t \cdot \sinh^{-1} y)^2 \right] \quad \text{with} \quad y = \frac{x - \mu}{\sigma},$$

with real scalars $s, t > 0$ called skewness and tailweight respectively. The probability density function is shifted and scaled by the real scalar parameters $\mu, \sigma > 0$.

For the distributions in LTBDT, we perform *unbinned maximum-likelihood fits* of Johnson S_U probability density functions $N \cdot J(\text{LTBDT}; s, t, \mu, \sigma)$ with normalization N , corresponding to the number of reconstructed decays. The continuum backgrounds exhibit a different shape in LTBDT, which we use to separate them from signal. We model continuum backgrounds by a superimposition of a Johnson S_U and a normal probability density function ϕ :

$$f_{\text{cont.}}(x; a, s, t, \mu, \sigma, \mu', \sigma') = a \cdot \phi(x; \mu', \sigma') + (1 - a) \cdot J(x; s, t, \mu, \sigma)$$

$$\text{with} \quad \phi(x; \mu', \sigma') = \frac{1}{\sqrt{2\pi}\sigma'} \cdot \exp \left[-\frac{1}{2} \left(\frac{x - \mu'}{\sigma'} \right)^2 \right]$$

where $a \in [0, 1]$ denotes the relative normalization of the two distributions. The left column of Figure 28 shows the individual simulated distributions in LTBDT for all decay categories and the shapes extracted from it. The distributions are for flavor-tag bin 0¹⁰, for opposite-flavor events. Below each shape, the pull distribution $(\text{data}(x_i) - \text{fit}(x_i))/\sigma_{\text{data}}(x_i)$ is shown as a measure of the fit quality.

7.3.2 Extraction of Fit Shapes in ΔE

The right column of Figure 28 shows the extracted shapes in ΔE for all categories. Since the different backgrounds and signal exhibit distinct shapes in ΔE , we employ various fit models. Again, we perform *unbinned maximum-likelihood fits* to extract the shapes.

The BB background is modelled by a linear combination of (first kind) Chebyshev polynomials up to third order:

$$C(x; c_0, c_1, c_2, c_3) = \sum_{i=0}^3 c_i \cdot T_i(x)$$

with real scalars c_i and polynomials $T_i(x)$ that are recursively defined as [49]

$$T_{i+1}(x) \equiv 2x \cdot T_i(x) - T_{i-1}(x) \quad \text{with} \quad T_0 = 1, T_1 = x.$$

⁹The Johnson probability density function was originally introduced as transformation of the normal distribution ϕ . $J(y) \propto \phi(t \cdot g(y) + s)$ with some suitable function g . The support S of the probability density function depends on the choice of g , in this case unbounded support $-\infty < y < \infty$, i.e. S_U [48].

¹⁰The shapes for the other dilution bins can be found in Appendix C.

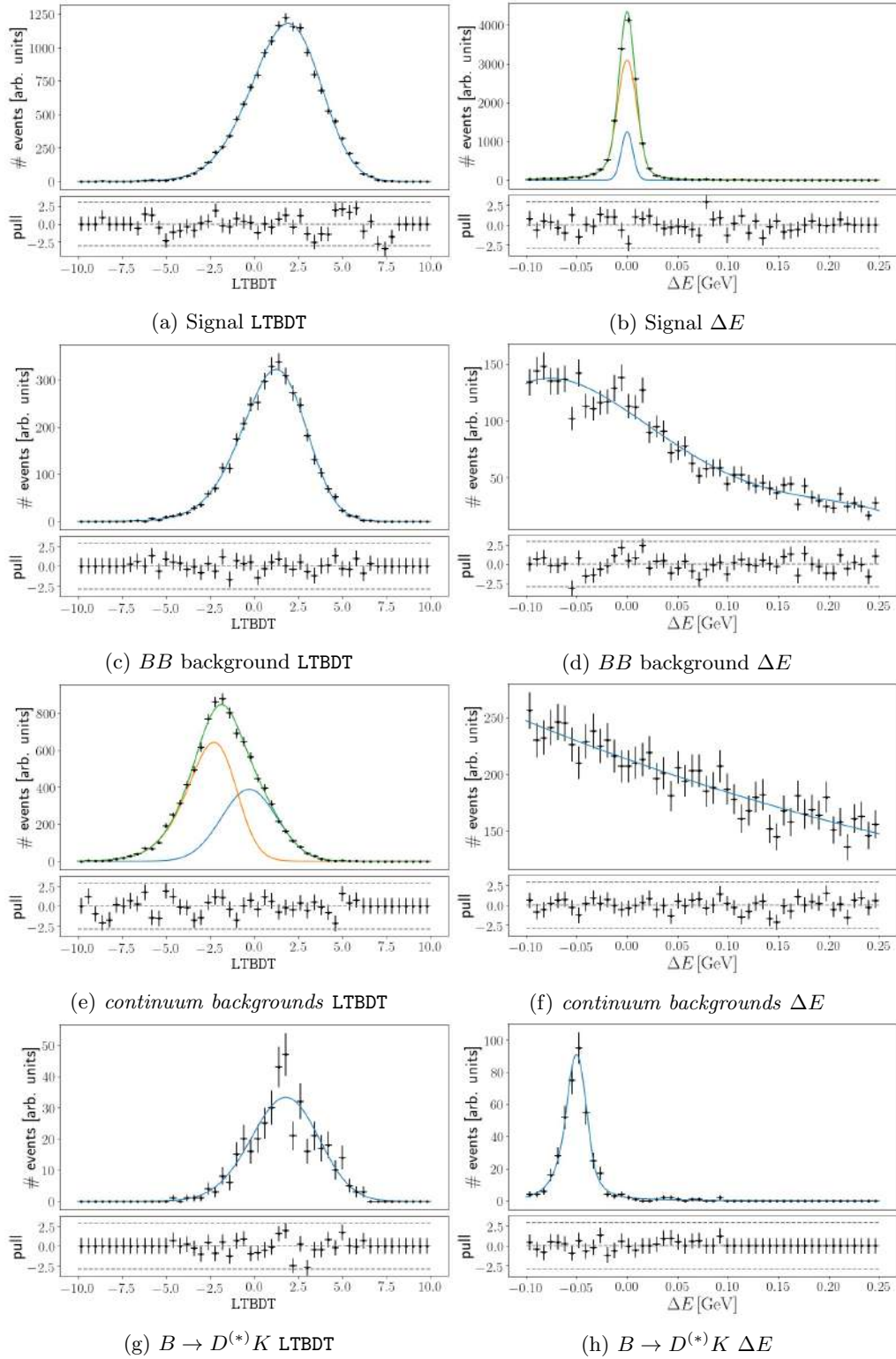


Figure 28: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Opposite-flavor events, bin 0 in $q \cdot r$.

The falsely reconstructed $B \rightarrow D^{(*)}K$ shape is modelled by a Crystal Ball shape probability density function [49]

$$CB(x; \mu, \sigma, \alpha, n) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] & \text{for } \frac{x-\mu}{\sigma} \geq -\alpha \\ A \cdot \left(B - \frac{x-\mu}{\sigma}\right)^{-n} & \text{for } \frac{x-\mu}{\sigma} < -\alpha \end{cases}$$

$$\text{with } A = \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left[-\frac{|\alpha|^2}{2}\right]$$

$$\text{and } B = \frac{n}{|\alpha|} - |\alpha|$$

which has a gaussian core with a powerlaw tail to its left. The parameter α indicates where to make the transition between core and tail.

The signal shape is modelled by a linear combination of a normal distribution and a double sided Crystal Ball probability density function [49]

$$f_{\text{sig}}(x; a, \mu, \sigma, \sigma', \alpha_L, n_L, \alpha_R, n_R) = a \cdot \phi(x; \mu, \sigma') + (1-a) \cdot DCB(x; \mu, \sigma, \alpha_L, n_L, \alpha_R, n_R).$$

The two distributions share the same mean μ and $a \in [0, 1]$ denotes their relative normalization. The double crystal ball shape DCB also has a gaussian core, but with powerlaw tails to the left and right. These are parametrized separately by (α_L, n_L) and (α_R, n_R) :

$$DCB(x; \mu, \sigma, \alpha_L, n_L, \alpha_R, n_R) = \begin{cases} A_L \cdot \left(B_L - \frac{x-\mu}{\sigma}\right)^{-n} & \text{for } \frac{x-\mu}{\sigma} < -\alpha_L \\ \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] & \text{for } -\alpha_L \leq \frac{x-\mu}{\sigma} \leq \alpha_R \\ A_R \cdot \left(B_R - \frac{x-\mu}{\sigma}\right)^{-n} & \text{for } \frac{x-\mu}{\sigma} > \alpha_R \end{cases}$$

with $A_{L,R}$ and $B_{L,R}$ defined analogously to the one-sided crystal ball distribution.

Finally, the *continuum backgrounds* are modelled by an exponential distribution

$$E(x; \lambda) \propto \exp[\lambda \cdot x]$$

where the real scalar parameter λ dictates the slope of the distribution.

A number of different fit strategies were tested in ΔE , LTBDT and other variables.⁵ Best results were obtained for a simultaneous fit in ΔE and LTBDT, which is used in the lifetime and mixing measurement. We here only describe this two-dimensional fit. In section 7.5, we compare it to a conventional one-dimensional fit in ΔE .

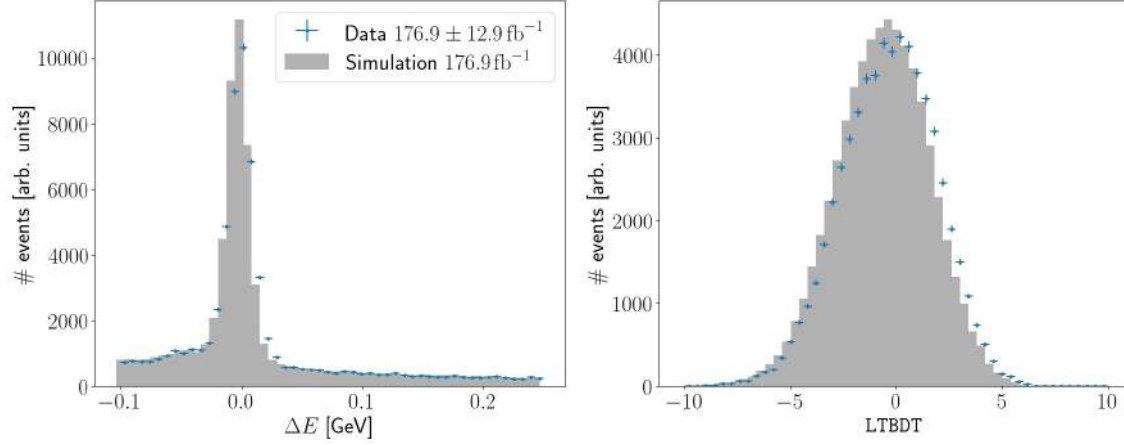


Figure 29: Distributions in ΔE (left) and LTBDT (right) for data and simulation scaled to the same luminosity.

7.4 Two-dimensional fits in $(\Delta E, \text{BDT})$

As described in the previous section, we perform unbinned maximum-likelihood fits to extract the shapes for each background and signal in ΔE and LTBDT. For each background and signal, we multiply the extracted one-dimensional distributions for ΔE and LTBDT to obtain a two-dimensional probability density function

$$\text{pdf}(\Delta E, \text{LTBDT}) = \text{pdf}(\Delta E) \cdot \text{pdf}(\text{LTBDT}).$$

We individually normalize each pdf over the fit range and sum the two-dimensional probability density functions, to obtain the full two-dimensional fit function f :

$$f(\Delta E, \text{LTBDT}, N_1, \dots, N_n) = \sum_{i=1}^n N_i \cdot \text{pdf}_i(\Delta E, \text{LTBDT}) \quad \text{with } i \in \text{decay categories}$$

where i runs over signal and all backgrounds. In the two-dimensional maximum likelihood fit to separate backgrounds and signal, we fix this fit function but let the normalizations N_i for signal and backgrounds float free.

The agreement between simulation and data is not perfect (Figure 29). Typically, an inferior energy resolution in data leads to slightly broader shapes in ΔE . Also, imperfect beam-energy measurements lead to slight shifts of the distributions in ΔE . To account for these differences, we include additional free parameters (Table 18) in the fit, that allow the extracted fit shapes to vary slightly.

Separating out the small $B \rightarrow D^{(*)}K$ component is challenging for the fit. Therefore, we fix the relative normalization of $B \rightarrow D^{(*)}K$ to the signal normalization from simulation. This is reasonable, since we expect the relative branching fractions to be constant over all three signal channels.

To reduce the reliance on the simulation, we let the extracted shapes as variable as possible in the fit. We free the slope λ of the extracted exponential shape for *continuum backgrounds* in ΔE in each $q \cdot r$ bin. The fit is able to determine λ from the *sideband* $\Delta E > 0.05$ GeV. We also introduce free parameters in the shapes in LTBDT, to account for the differences in Figure 29. The full set of free parameters is summarized in Table 18. Therein \checkmark (\times) means that we do (not) use independent free parameters for each $q \cdot r$ bin or for same- and opposite-flavor events.

Table 18: Free parameters in two-dimensional fit.

Shape	Free Parameters	independent free params. for each $q \cdot r$ bin	independent free params. for opposite & same-flavor
Signal and $B \rightarrow D^{(*)}K$	normalization N_{sig}	✓	✓
continuum background	normalization $N_{\text{cont.}}$	✓	✓
BB background	normalization N_{BB}	✓	✓
Signal and $B \rightarrow D^{(*)}K$ in ΔE	shift μ_1 , scale σ_1	✗	✗
continuum background in ΔE	slope λ	✓	✗
continuum background in LTBDT	shift μ_2 , scale σ_2	✗	✗
other shapes in LTBDT	shift μ_3 , scale σ_3	✗	✗

7.4.1 Fit in Full Simulation Sample

Figure 30 shows the fit result in the full simulation sample of 1 ab^{-1} . For each flavor tag bin (rows) and opposite- and same-flavor events (columns), the two-dimensional fit function f is projected out on ΔE or LTBDT

$$f(\text{LTBDT}) = \int_{\Delta E_{\text{min}}}^{\Delta E_{\text{max}}} f(\Delta E, \text{LTBDT}) d\Delta E$$

and analogously for $f(\Delta E)$. In Figure 30, the projected one-dimensional fit functions are compared to the fitted simulation samples (black dots). The fitted signal and background components are shown in color, along with the pulls of their normalizations. Furthermore, the pull of the free slope λ , for the *continuum background* in ΔE , is shown. As indicated in Table 18, λ is shared between opposite- and same-flavor events for every flavor tag bin. Table 19 shows the fit results for the remaining free parameters.

Table 19: Fit result for two-dimensional fit in full simulation sample.

Parameter	Fitted Value	True Value	Pull
μ_1 [GeV]	$(3.0 \pm 2.7)10^{-5}$	0.0	1.1
σ_1	0.990 ± 0.003	1.0	-3.3
μ_2	0.014 ± 0.021	0.0	0.67
σ_2	0.9963 ± 0.0049	1.0	-0.76
μ_3	-0.132 ± 0.021	0.0	-6.3
σ_3	0.9895 ± 0.0014	1.0	-7.5

Although the pulls in μ_3, σ_3 significantly deviate from the expectation ("true value" in Table 19), we accept this performance since the physically relevant parameters are the normalizations indicated in Figure 30. Their maximal pull is below 2.5, which corresponds to a deviation of less than 10%.

For comparison, we also attempt the fit with common fit shapes for all bins, extracted from the full cumulative simulation sample. The result in Figure 31 shows large deviations between fit

and fitted simulation samples in LTBDT. We conclude that the shapes vary significantly over the bins, and that common fit shapes are not adequate.

7.4.2 Fit in Full Data Sample

To confirm the simulation/data agreement, we repeat the two-dimensional fit *with individual fit shapes in each bin* in the full available data sample of $176.9 \pm 12.6 \text{ fb}^{-1}$ (Table 4). Since data is available only for two of the three signal channels, we fit shapes that were extracted from simulation in only these two channels. In simulation, the fit performance in two and three signal channels is equal. Figure 32 shows the projected fits. Table 20 shows the fit results in all free parameters and compares them to the expected values from simulation samples of the same luminosity.

For the numbers N of signal and background decays, the uncertainties on the expected values from simulation result from the uncertainty on the collected luminosity $176.9 \pm 12.6 \text{ fb}^{-1}$ (see [50]). For λ the uncertainties on the expected values result from the uncertainties on the extracted shapes from simulation. As can be seen in Table 20, the fitted values in data agree well with the expectations from simulation. The increased value in σ_1 and the shift in μ_1 are expected from an inferior energy resolution and imperfect beam energy measurement in data (see Figure 29).

Table 20: Fit result for two-dimensional fit in full data sample.

Parameter	Fitted Value	Expected Value from Simulation	Pull
μ_1 [GeV]	$(-27.4 \pm 8.3) \cdot 10^{-5}$	0.0	-3.3
σ_1	1.098 ± 0.009	1.0	1.1
μ_2	$(-3.4 \pm 8.5) \cdot 10^{-2}$	0.0	-0.4
σ_2	0.99 ± 0.02	1.0	-0.5
μ_3	$(2.7 \pm 6.0) \cdot 10^{-2}$	0.0	0.5
σ_3	0.995 ± 0.004	1.0	1.3
bin 0, λ [GeV $^{-1}$]	-1.09 ± 0.33	-1.54 ± 0.10	1.3
bin 1, λ [GeV $^{-1}$]	-1.99 ± 0.35	-1.55 ± 0.10	-1.2
bin 2, λ [GeV $^{-1}$]	-2.23 ± 0.36	-1.61 ± 0.11	-1.6
bin 3, λ [GeV $^{-1}$]	-1.44 ± 0.46	-1.56 ± 0.13	0.25
bin 4, λ [GeV $^{-1}$]	-2.14 ± 0.44	-1.42 ± 0.14	-1.6
bin 5, λ [GeV $^{-1}$]	-2.34 ± 0.70	-1.32 ± 0.20	-1.4
bin 6, λ [GeV $^{-1}$]	-2.2 ± 2.3	-1.5 ± 0.48	-0.3
OF, bin 0, N_{sig}	1852 ± 49	1890 ± 138	-0.3
OF, bin 1, N_{sig}	2104 ± 52	2109 ± 154	-0.03
OF, bin 2, N_{sig}	2340 ± 54	2307 ± 168	0.2
OF, bin 3, N_{sig}	1793 ± 48	1782 ± 130	0.08
OF, bin 4, N_{sig}	1634 ± 45	1604 ± 117	0.2
OF, bin 5, N_{sig}	1496 ± 43	1526 ± 111	-0.2
OF, bin 6, N_{sig}	1885 ± 46	1937 ± 141	-0.4
SF, bin 0, N_{sig}	1871 ± 49	1812 ± 132	0.4
SF, bin 1, N_{sig}	1803 ± 48	1773 ± 129	0.2
SF, bin 2, N_{sig}	1606 ± 45	1573 ± 115	0.3
SF, bin 3, N_{sig}	976 ± 35	1047 ± 76	-0.9
SF, bin 4, N_{sig}	775 ± 31	805 ± 59	-0.5
SF, bin 5, N_{sig}	634 ± 27	623 ± 45	0.2
SF, bin 6, N_{sig}	574 ± 26	549 ± 40	0.5
OF, bin 0, $N_{\text{cont.}}$	742 ± 56	877 ± 62	-1.6
OF, bin 1, $N_{\text{cont.}}$	809 ± 58	839 ± 60	-0.4
OF, bin 2, $N_{\text{cont.}}$	827 ± 58	820 ± 58	0.09

Continued on next page

Table 20 – continued from previous page

Parameter	Fitted Value	Expected Value from Simulation	Pull
OF, bin 3, $N_{\text{cont.}}$	553 ± 50	646 ± 46	-1.4
OF, bin 4, $N_{\text{cont.}}$	568 ± 42	575 ± 41	-0.1
OF, bin 5, $N_{\text{cont.}}$	280 ± 32	312 ± 22	-0.8
OF, bin 6, $N_{\text{cont.}}$	46 ± 15	50 ± 4	-0.3
SF, bin 0, $N_{\text{cont.}}$	892 ± 57	854 ± 61	0.5
SF, bin 1, $N_{\text{cont.}}$	730 ± 56	787 ± 56	-0.7
SF, bin 2, $N_{\text{cont.}}$	560 ± 50	628 ± 45	-1.0
SF, bin 3, $N_{\text{cont.}}$	388 ± 35	370 ± 26	0.4
SF, bin 4, $N_{\text{cont.}}$	252 ± 25	275 ± 20	-0.7
SF, bin 5, $N_{\text{cont.}}$	94 ± 19	137 ± 10	-2.0
SF, bin 6, $N_{\text{cont.}}$	19 ± 8	27 ± 2	-1.0
OF, bin 0, N_{BB}	532 ± 62	486 ± 35	0.6
OF, bin 1, N_{BB}	511 ± 63	515 ± 37	-0.05
OF, bin 2, N_{BB}	504 ± 63	602 ± 43	-1.3
OF, bin 3, N_{BB}	564 ± 57	506 ± 36	0.9
OF, bin 4, N_{BB}	292 ± 46	440 ± 31	-2.7
OF, bin 5, N_{BB}	245 ± 38	307 ± 22	-1.4
OF, bin 6, N_{BB}	219 ± 27	227 ± 16	-0.3
SF, bin 0, N_{BB}	361 ± 59	474 ± 34	-1.7
SF, bin 1, N_{BB}	425 ± 62	453 ± 32	-0.4
SF, bin 2, N_{BB}	374 ± 55	377 ± 27	-0.05
SF, bin 3, N_{BB}	166 ± 38	221 ± 16	-1.3
SF, bin 4, N_{BB}	134 ± 28	164 ± 12	-1.0
SF, bin 5, N_{BB}	107 ± 23	98 ± 7	0.4
SF, bin 6, N_{BB}	15 ± 15	49 ± 4	-2.2

7.4.3 Stability of Fit in Pseudo-Experiments

To validate the fit method, a set of pseudo-experiments is performed. By sampling from the extracted shapes in ΔE and LTBDT, we generate 1000 pseudo simulation samples with poisson-fluctuated normalizations. Their individual size corresponds to the expected number of decays in 150 fb^{-1} of data. For each pseudo sample, we perform the full two-dimensional fit and compare the results to the original shapes. Figures 33 and 34 show the resulting pull distributions in all fit parameters.

When assuming likelihood functions that are symmetric in the fit parameters, maximum-likelihood fits find their average values. According to the *Central Limit Theorem*, these average values are normal distributed, in the limit of large statistics. When computing their pull, we expect standard normal distributions with mean $\mu = 0$ and width $\sigma = 1$. If the fit is free of bias, it will on average estimate the correct fit parameters and the mean of their pulls will be zero. If the fit on average correctly estimates the uncertainties on the fit parameters, the pulls will have a variance $\sigma^2 = 1$.

In Figures 33 and 34, we fit a normal distribution on all pulls and extract their mean μ and standard deviation σ . Their values are as expected within their errors, indicating that the fit is stable and not significantly biased.

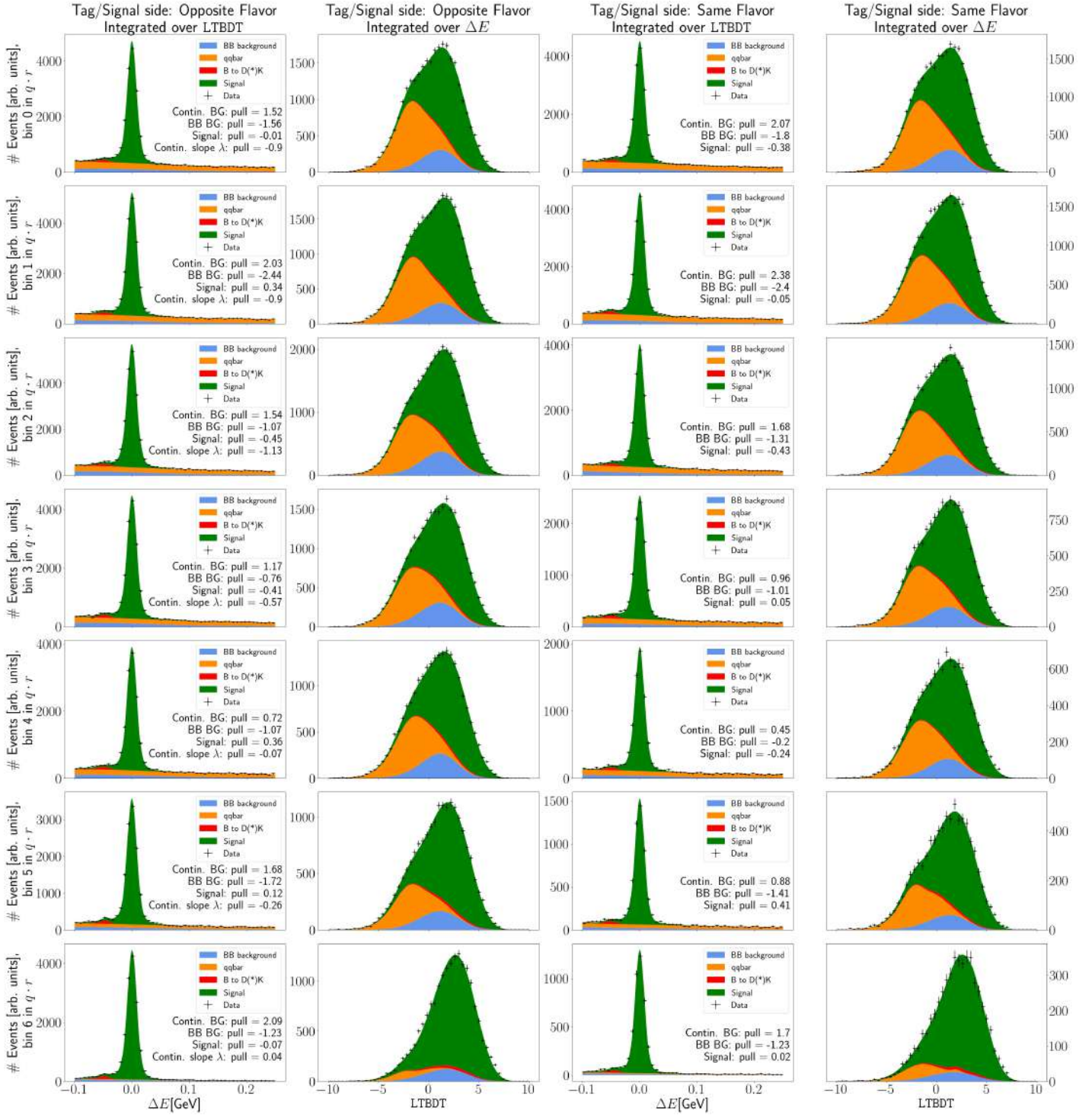


Figure 30: Full two-dimensional fit to total simulation sample of 1 ab^{-1} . Projected fit results in LTBDT and ΔE , separately for 7 bins in $q \cdot r$ and opposite- and same-flavor events. *Individual fit shapes for each bin.*

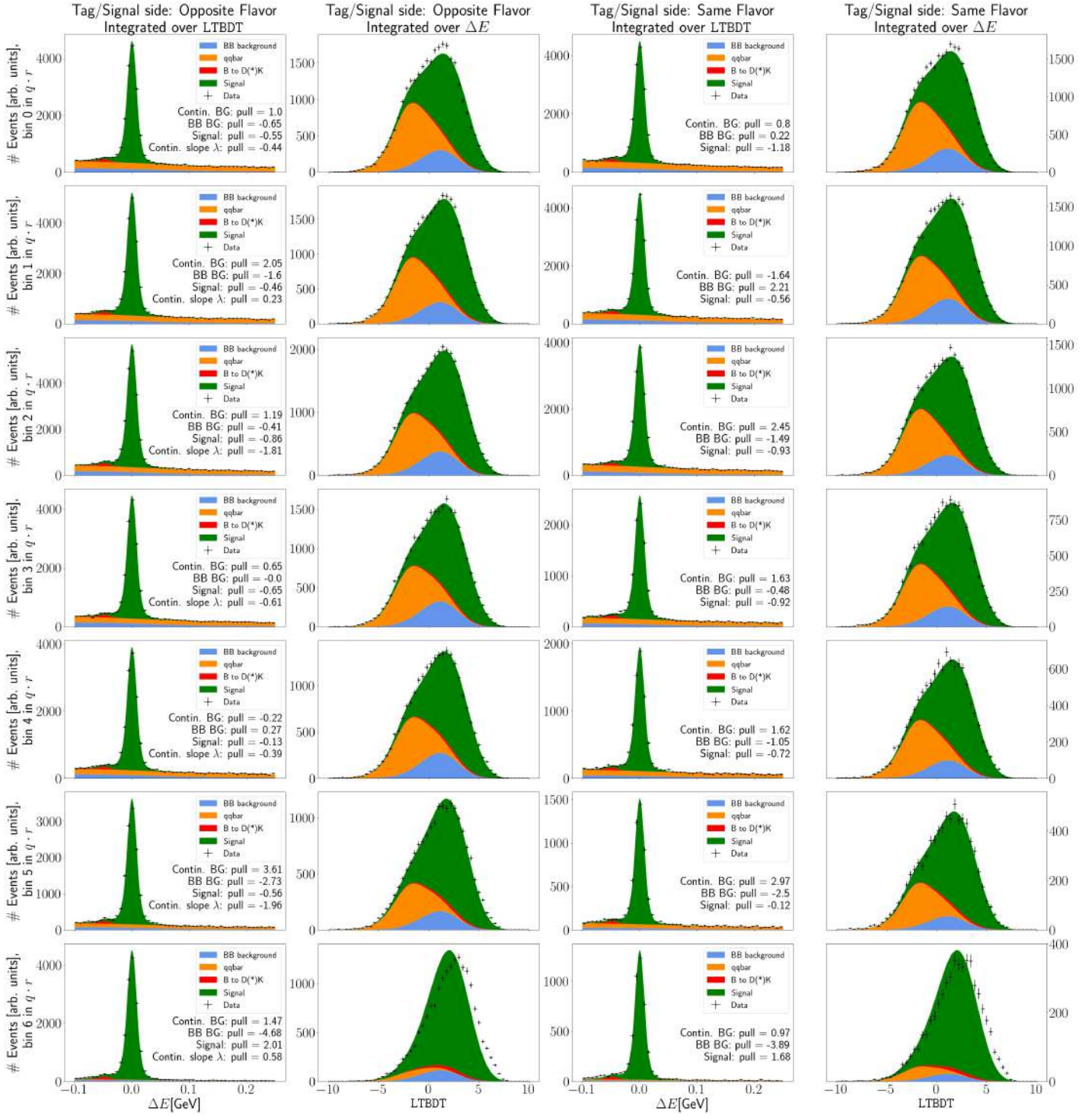


Figure 31: Full two-dimensional fit to total simulation sample of 1 ab^{-1} . Projected fit results in LTBDT and ΔE , separately for 7 bins in $q \cdot r$ and opposite- and same-flavor events. *Common fit shapes over all bins.*

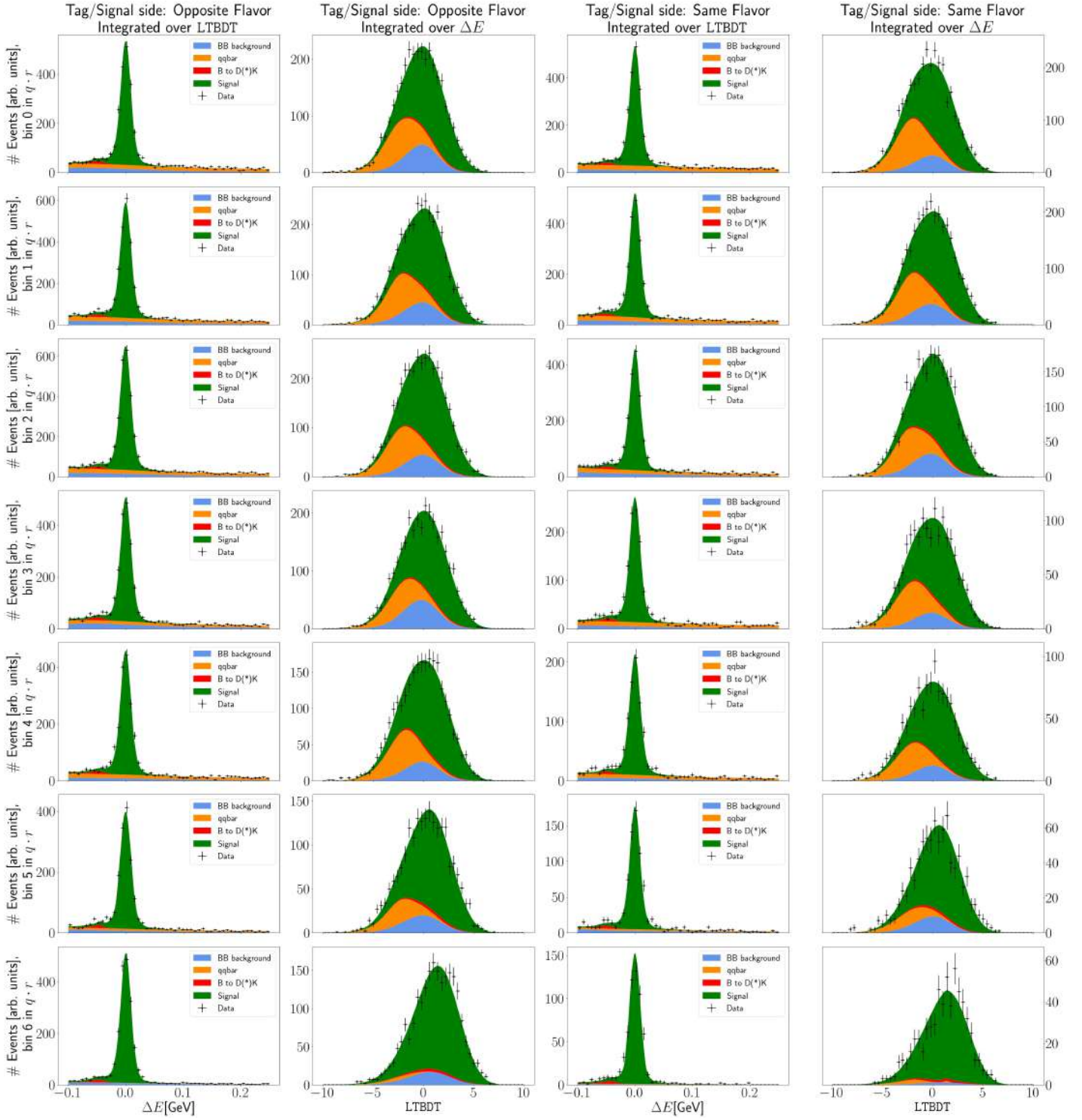


Figure 32: Full two-dimensional fit to total data sample. Projected fit results in LTBDDT and ΔE , separately for 7 bins in $q \cdot r$ and opposite- and same-flavor events. *Individual fit shapes for each bin.*

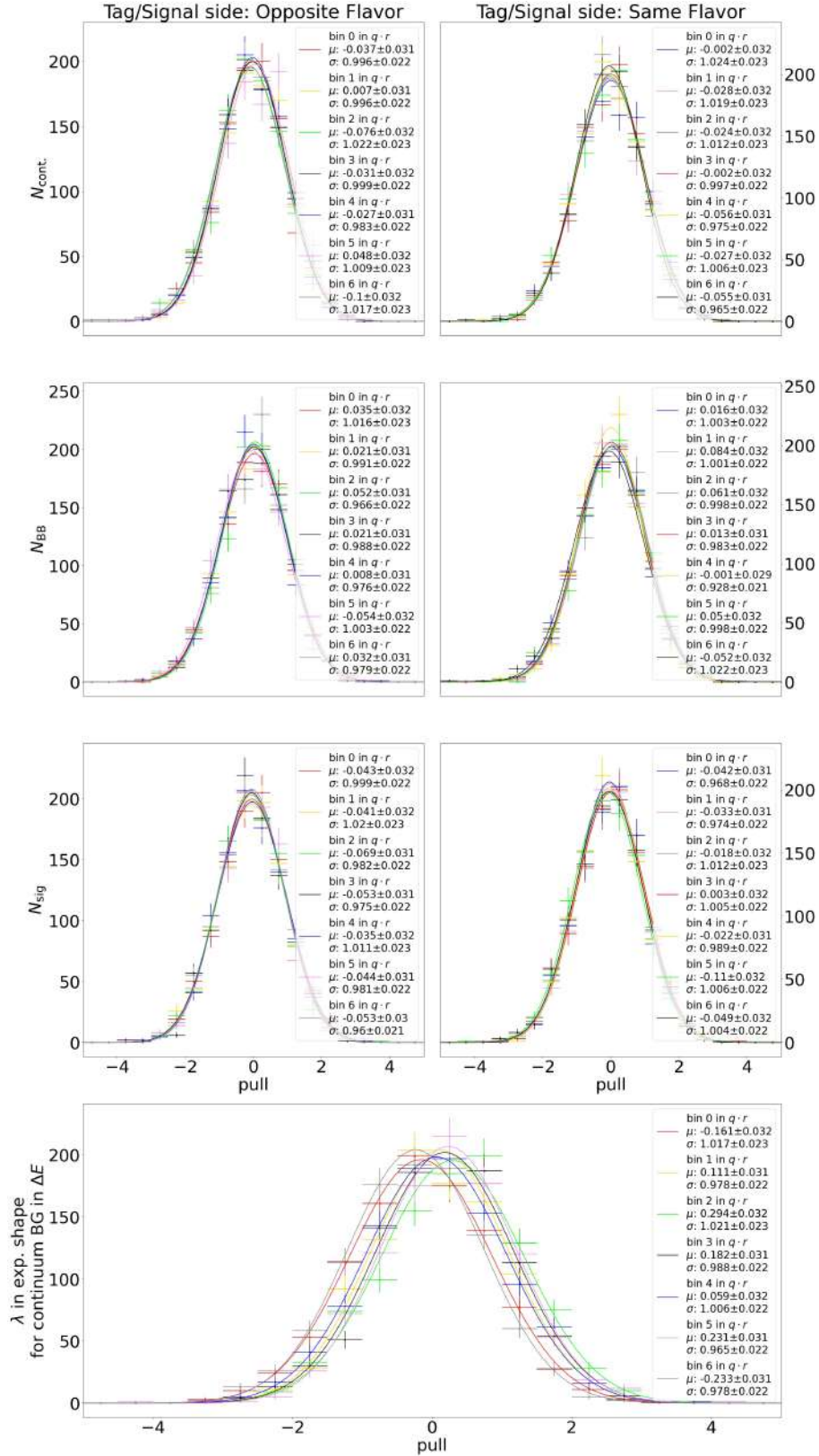


Figure 33: Toy-study of two-dimensional fit in 1000 fits of 150fb^{-1} each (Part 1 of 2). *Individual fit shapes in each bin.* For variable definitions see Table 18.

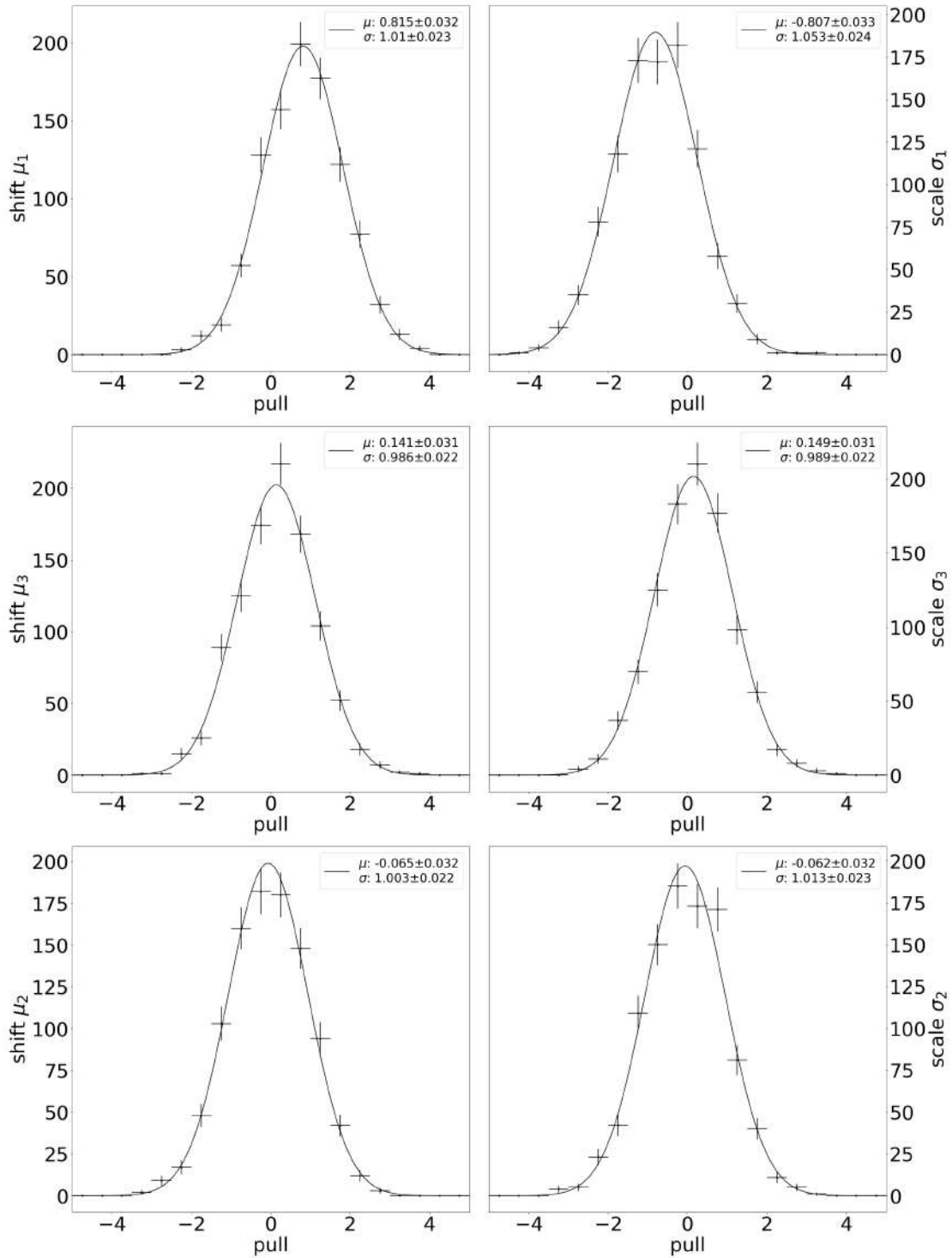


Figure 34: Toy-study of two-dimensional fit in 1000 fits of 150fb^{-1} each. Continued (Part 2 of 2). Individual fit shapes in each bin. For variable definitions see Table 18.

7.5 One-dimensional Fit in ΔE

The previously described two-dimensional fit separates backgrounds with different behaviors in Δt . In this section, we compare it with a one-dimensional fit strategy in ΔE . As explained in Chapter 4, different background contributions cannot be separated when only fitting in ΔE . By comparing the results for the one- and two-dimensional fits, we aim to quantify the improvement on the fitted lifetime and mixing parameter uncertainties.

Since the distributions of *continuum backgrounds* and *BB* backgrounds are very similar in ΔE and there is no additional observable to distinguish them, we cannot separate them in the one-dimensional fit. Instead, we model continuum and *BB* backgrounds together by a single exponential. Again, the shape is extracted from simulation samples, separately for opposite- and same-flavor events and in the 7 dilution bins. Figure 35 shows the extracted shape for opposite-flavor events in flavor-tag bin 0¹¹, along with the pull distribution.

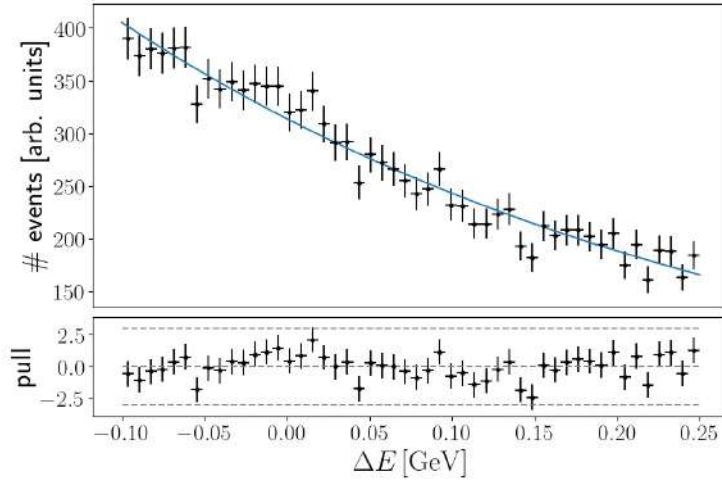


Figure 35: Simulated distributions in ΔE for combined backgrounds, except $B \rightarrow D^{(*)}K$, with extracted exponential shape and pull. Opposite-flavor events, bin 0 in $q \cdot r$.

As for the two-dimensional fit, we fix the relative normalizations of the $B \rightarrow D^{(*)}K$ component to signal from simulation. Again, we introduce a free shift μ_1 and a free scale parameter σ_1 . To reduce the reliance on simulation, we free the slopes of the exponential background shape in every bin. Table 21 summarizes all free parameters in the one-dimensional fit.

Table 21: Free parameters in one-dimensional fit.

Shape	Free Parameters	individual free params. for each $q \cdot r$ bin	individual free params. for opp. & same-flavor
Signal and $B \rightarrow D^{(*)}K$	normalization N_{sig}	✓	✓
continuum and BB background	normalization $N_{\text{cont. \& BB}}$	✓	✓
Signal and $B \rightarrow D^{(*)}K$ in ΔE	shift μ_1 , scale σ_1	✗	✗
continuum and BB background in ΔE	slope λ	✓	✗

¹¹The shapes for the other dilution bins can be found in Appendix D.

7.5.1 Fit in Full Simulation Sample

A fit on the full simulation sample of 1 ab^{-1} yields the results in Figure 36 and Table 22.

Table 22: Fit result for one-dimensional fit in full simulation sample.

Parameter	Fitted Value	True Value	Pull
μ_1 [GeV]	$(-0.4 \pm 2.8)10^{-5}$	0.0	-0.1
σ_1	1.0060 ± 0.0031	1.0	1.9

The pulls are all within expectations and the fit correctly tells apart signal, falsely reconstructed $B \rightarrow D^{(*)}K$ and other backgrounds. However, as discussed before, a separation between *continuum backgrounds* and BB backgrounds is not possible. Hence when using this fit strategy, these backgrounds have to be treated collectively in the full lifetime and mixing measurement. As motivated in Chapter 4, this is not desirable due to their different behavior in Δt and may lead to larger uncertainties.

7.5.2 Fit in Full Data Sample

To confirm the simulation/data agreement, we repeat the one-dimensional fit in the full available Belle II data sample of $176.9 \pm 12.6 \text{ fb}^{-1}$ (Table 4). Again, the shape extraction from simulation and the fit in data is performed in only two signal channels. The fit results are shown in Figure 37 and in Table 23.

Table 23: Fit result for one-dimensional fit in full data sample.

Parameter	Fitted Value	Expected Value from Simulation	Pull
μ_1 [GeV]	$(-28.0 \pm 8.5) \cdot 10^{-5}$	0.0	-3.3
σ_1	1.12 ± 0.01	1.0	12.0
bin 0, λ [GeV $^{-1}$]	-2.38 ± 0.23	-2.91 ± 0.08	2.2
bin 1, λ [GeV $^{-1}$]	-3.28 ± 0.23	-3.00 ± 0.08	-1.1
bin 2, λ [GeV $^{-1}$]	-3.41 ± 0.24	-3.17 ± 0.09	-0.9
bin 3, λ [GeV $^{-1}$]	-3.06 ± 0.28	-3.15 ± 0.10	0.3
bin 4, λ [GeV $^{-1}$]	-3.13 ± 0.34	-3.10 ± 0.11	-0.1
bin 5, λ [GeV $^{-1}$]	-4.04 ± 0.44	-3.25 ± 0.15	-1.7
bin 6, λ [GeV $^{-1}$]	-4.35 ± 0.81	5.10 ± 0.24	0.9
OF, bin 0, N_{sig}	1887 ± 49	1890 ± 138	-0.02
OF, bin 1, N_{sig}	2122 ± 52	2109 ± 154	0.08
OF, bin 2, N_{sig}	2369 ± 55	2307 ± 168	0.4
OF, bin 3, N_{sig}	1823 ± 49	1782 ± 130	0.4
OF, bin 4, N_{sig}	1657 ± 46	1604 ± 117	0.4
OF, bin 5, N_{sig}	1493 ± 43	1526 ± 111	-0.3
OF, bin 6, N_{sig}	1899 ± 47	1937 ± 141	-0.3
SF, bin 0, N_{sig}	1990 ± 49	1812 ± 132	1.3
SF, bin 1, N_{sig}	1828 ± 49	1773 ± 129	0.3
SF, bin 2, N_{sig}	1620 ± 46	1573 ± 115	0.4
SF, bin 3, N_{sig}	987 ± 35	1047 ± 76	-0.7
SF, bin 4, N_{sig}	798 ± 32	805 ± 59	-0.1
SF, bin 5, N_{sig}	637 ± 28	623 ± 45	0.3
SF, bin 6, N_{sig}	577 ± 25	549 ± 40	0.6
OF, bin 0, $N_{\text{cont.}\&\text{BB}}$	1239 ± 43	1362 ± 99	-1.1
OF, bin 1, $N_{\text{cont.}\&\text{BB}}$	1301 ± 44	1354 ± 99	-0.5

Continued on next page

Table 23 – continued from previous page

Parameter	Fitted Value	Expected Value from Simulation	Pull
OF, bin 2, $N_{\text{cont.}\&\text{BB}}$	1301 ± 45	1422 ± 104	-1.1
OF, bin 3, $N_{\text{cont.}\&\text{BB}}$	1086 ± 41	1152 ± 84	-0.7
OF, bin 4, $N_{\text{cont.}\&\text{BB}}$	837 ± 37	1015 ± 74	-2.2
OF, bin 5, $N_{\text{cont.}\&\text{BB}}$	528 ± 31	619 ± 45	-1.7
OF, bin 6, $N_{\text{cont.}\&\text{BB}}$	251 ± 24	278 ± 20	-0.9
SF, bin 0, $N_{\text{cont.}\&\text{BB}}$	1222 ± 43	1328 ± 97	-1.0
SF, bin 1, $N_{\text{cont.}\&\text{BB}}$	1130 ± 42	1240 ± 90	-1.1
SF, bin 2, $N_{\text{cont.}\&\text{BB}}$	919 ± 38	1005 ± 73	-1.0
SF, bin 3, $N_{\text{cont.}\&\text{BB}}$	543 ± 29	591 ± 43	-0.9
SF, bin 4, $N_{\text{cont.}\&\text{BB}}$	362 ± 24	439 ± 32	-1.9
SF, bin 5, $N_{\text{cont.}\&\text{BB}}$	199 ± 23	235 ± 17	-1.4
SF, bin 6, $N_{\text{cont.}\&\text{BB}}$	30 ± 10	76 ± 6	-3.9

The uncertainties on the expected values from simulation are analogous to Table 20. As can be seen in Table 23, the fitted values in data agree well with the expectations from simulation. Again, the increased value in σ_1 and the shift in μ_1 are expected from an inferior energy resolution and imperfect beam energy measurement in data (see Figure 29).

7.5.3 Stability of Fit in Pseudo-Experiments

As for the two-dimensional fit, we assess the stability of the fit in 1000 pseudo-experiments. Again, we expect standard normal distributions in the pulls of all fit parameters. Figure 38 shows the pull distributions. Analogously to Figures 33 and 34, we fit normal distributions to extract the mean μ and width σ of the pull distributions. The result agrees well with the expectation of standard normal distributions, indicating that the fit is stable and not significantly biased.

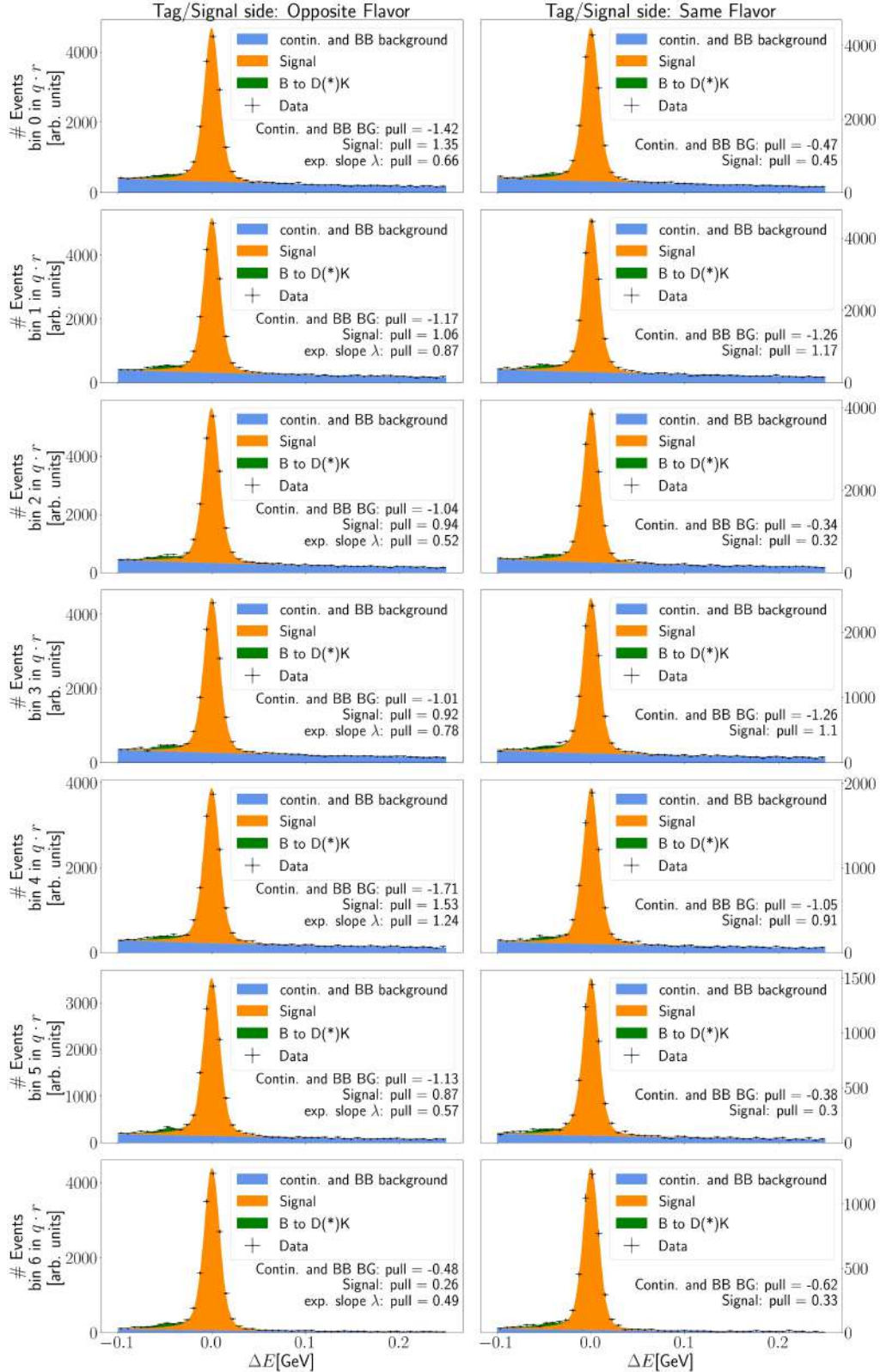


Figure 36: Full one-dimensional fit to total simulation sample of 1 ab^{-1} . Fit results in ΔE , separately for 7 bins in $q \cdot r$ and opposite- and same-flavor events. *Individual fit shapes for each bin.* For variable definitions see Table 21.

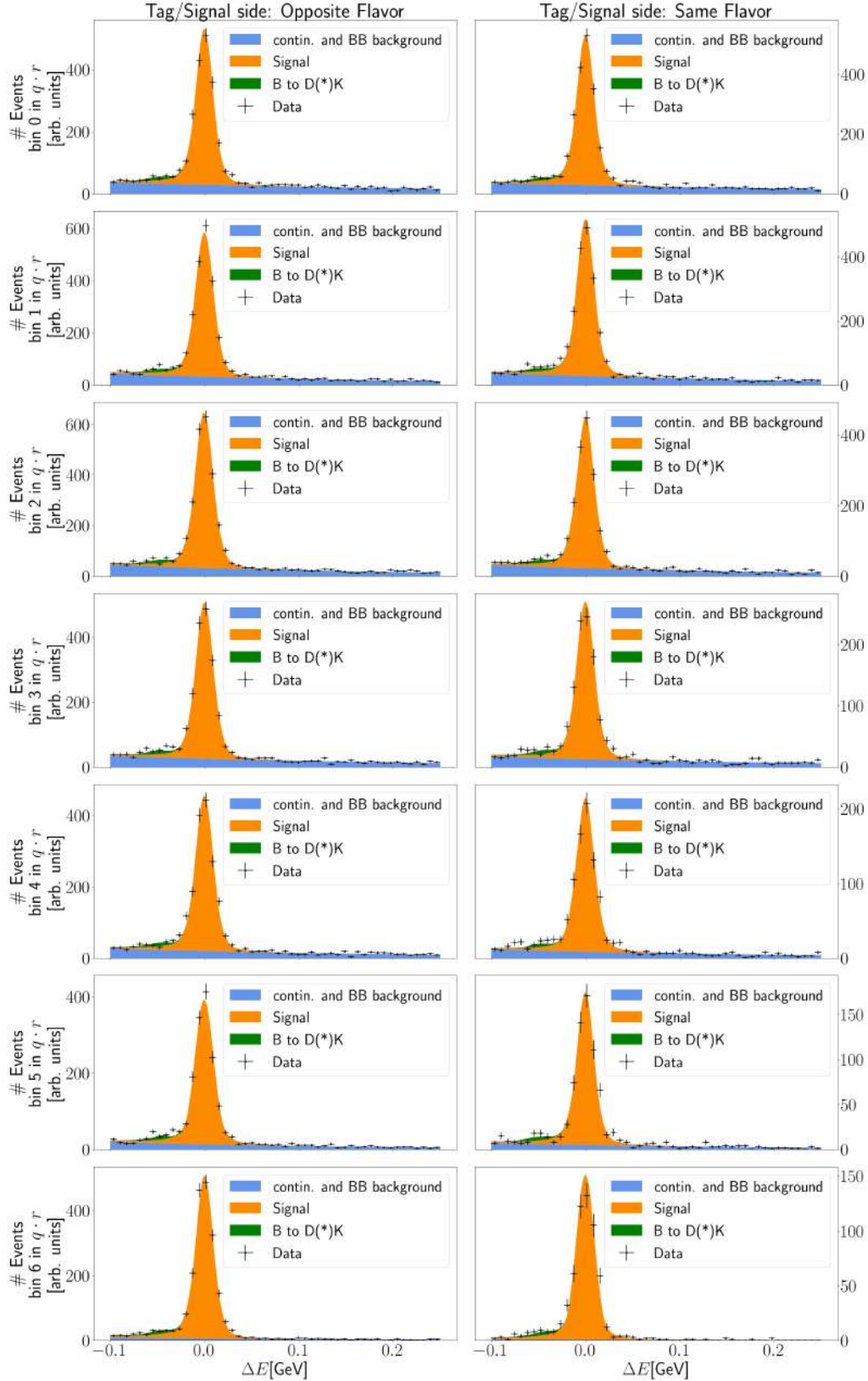


Figure 37: Full one-dimensional fit to total data sample. Fit results in ΔE , separately for 7 bins in $q \cdot r$ and opposite- and same-flavor events. *Individual fit shapes for each bin.* For variable definitions see Table 21.

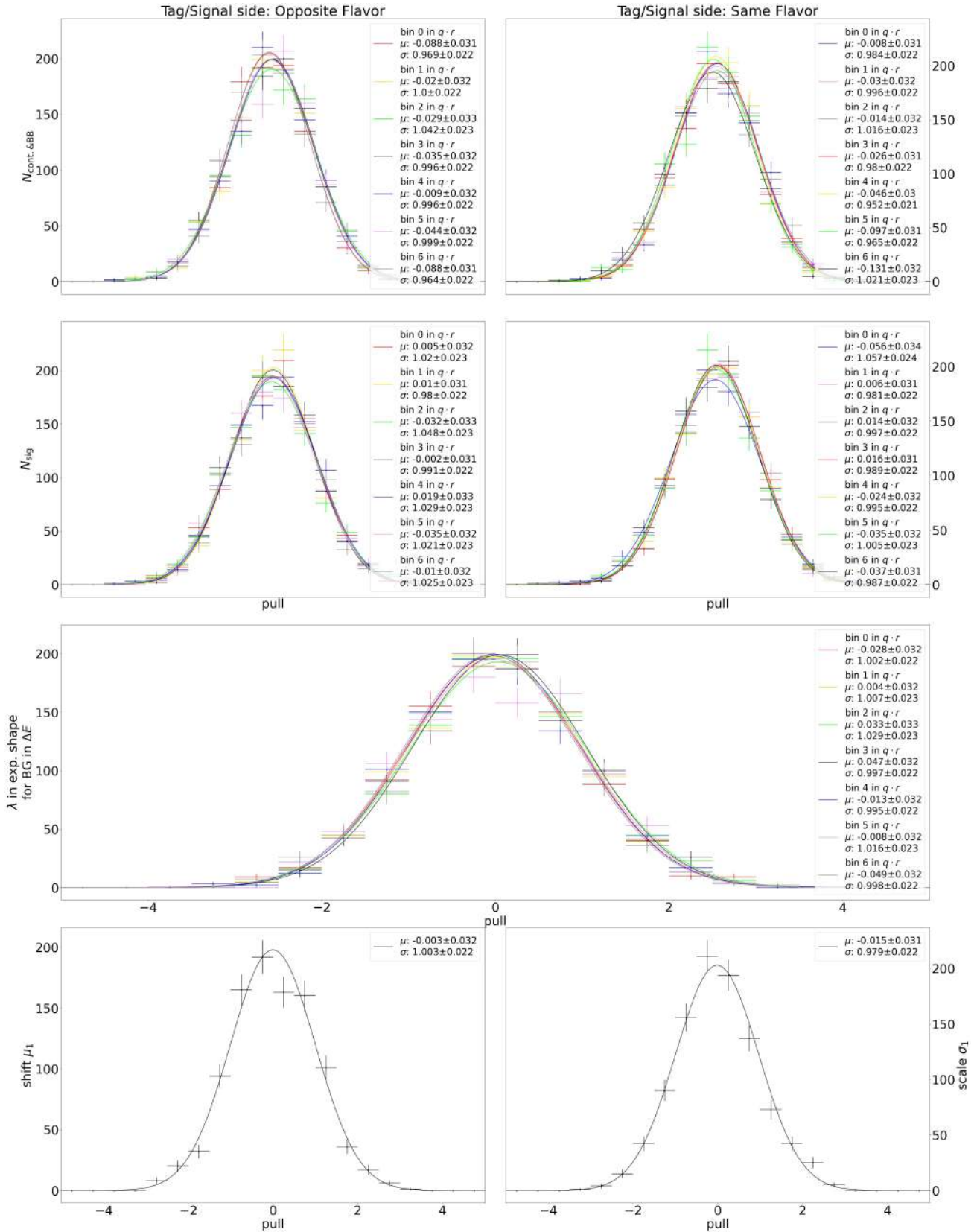


Figure 38: Toy-study of one-dimensional fit in 1000 fits of 150fb^{-1} each. *Individual fit shapes for each bin.* For variable definitions see Table 21.

8 Lifetime and Mixing Measurement

In this Chapter we describe the time-dependent fits and how the previously described fits in ΔE and BDT are implemented in the lifetime and mixing analysis.

8.1 Time-dependent Fits

As explained in Chapter 4, the measured distributions in the decay time difference Δt of the B mesons are altered by the detector resolution and the kinematic approximation. We employ a *resolution function* \mathcal{R} to relate the measured and true distributions in Δt . For this analysis, a resolution function with a double gaussian core and exponential tails was developed from simulation (Equation 42).

$$\begin{aligned} \mathcal{R}(\Delta t, \sigma_{\Delta t}) = & (1 - f_{\text{tail}}) \cdot \phi(\Delta t; \mu_{\text{main}} \cdot \sigma_{\Delta t}, s_{\text{main}} \cdot \sigma_{\Delta t}) \\ & + f_{\text{tail}} \cdot (1 - f_{\text{exp}}) \cdot \phi(\Delta t; \mu_{\text{tail}} \cdot \sigma_{\Delta t}, s_{\text{tail}} \cdot \sigma_{\Delta t}) \\ & + f_{\text{tail}} \cdot f_{\text{exp}} \cdot \phi(\Delta t; \mu_{\text{tail}} \cdot \sigma_{\Delta t}, s_{\text{tail}} \cdot \sigma_{\Delta t}) \\ & * ((1 - f_{\text{R}}) \cdot \exp_L[\Delta t/c \cdot \sigma_{\Delta t}] + f_{\text{R}} \cdot \exp_R[-\Delta t/c \cdot \sigma_{\Delta t}]), \end{aligned} \quad (42)$$

where ϕ denotes the normal distribution with mean $\mu \cdot \sigma_{\Delta t}$ and width $s \cdot \sigma_{\Delta t}$. Therein $\sigma_{\Delta t}$ is the uncertainty on the decay time difference Δt and f_{tail} denotes the fraction of reconstructed events in the tails of the distribution. The tails originate from the decays of secondary particles in the B meson decays and consist of a gaussian outlier component and exponential tails. f_{exp} (f_{R}) denotes the number of reconstructed events in the (right-sided) exponential tail. In Equation 42, parameters in blue are free in the fit and parameters in magenta are fixed from simulation [42].

8.2 Integration of Background Fit in Time-Dependent Fit

The full fit strategy for the lifetime and mixing analysis consists in simultaneously fitting in $(\Delta t, \sigma_{\Delta t})$ and in $(\Delta E, \text{BDT})$. The fit variables $(\Delta E, \text{BDT})$ serve to separate signal and backgrounds, for each of which the resolution function $\mathcal{R}(\Delta t, \sigma_{\Delta t})$ fits the time-dependance. As motivated in Chapter 4, the background shapes in $(\Delta t, \sigma_{\Delta t})$ are extracted from the side band in ΔE , i.e. in $|\Delta E| > 0.05 \text{ GeV}$. Together with the signal shapes, they are then fitted on $(\Delta t, \sigma_{\Delta t})$ in the signal region in ΔE , i.e. in $|\Delta E| \leq 0.05 \text{ GeV}$.

Figure 39 shows an exemplary fit on simulation in dilution bin 6. The fitted samples (dots) are overlaid with the projections of the fitted shapes (solid) in all four fit variables. Fitted BB background contributions are shown in orange, continuum background contributions in blue and the remaining contribution is signal.¹² The pull distributions in Figure 39 indicate a good fit quality in all variables.

By simultaneously fitting on the distributions in each of the 7 dilution bins and for opposite- and same-flavor events, we can extract the wrong-tag fractions w , the lifetime τ_{B^0} and the oscillation frequency Δm (see Equation 37). Table 24 summarizes all free parameters in the fit.

¹²For simplicity, the $B \rightarrow D^{(*)}K$ component is not shown in this preliminary fit.

Table 24: Free fit parameters in full fit strategy.

Fit Parameter	Number of Free Parameters	Explanation
τ_{B^0}	1	Neutral B meson lifetime
Δm	1	Neutral B meson oscillation frequency
N_{sig}	14 (7 dilution bins and OF,SF)	Number of reconstr. signal decays
N_{BB}	14 (7 dilution bins and OF,SF)	Number of reconstr. BB BG decays
$N_{\text{cont.}}$	14 (7 dilution bins and OF,SF)	Number of reconstr. continuum BG decays
λ	7 (7 dilution bins)	Free parameter for continuum shape in ΔE
μ_1, σ_1	2	Free shift and scale in ΔE shapes
$\mu_2, \sigma_2, \mu_3, \sigma_3$	4	Free shift and scale in BDT shapes
w_{sig}	7 (7 dilution bins)	Wrong-tag fraction for signal decays
w_{BB}	7 (7 dilution bins)	Wrong-tag fraction for BB BG decays
$w_{\text{cont.}}$	7 (7 dilution bins)	Wrong-tag fraction for continuum BG decays
$\mu_{\text{main}}^{\text{sig}}, s_{\text{main}}^{\text{sig}}$ $\mu_{\text{tail}}^{\text{sig}}, s_{\text{tail}}^{\text{sig}}, f_{\text{tail}}^{\text{sig}}$	5	Resolution function pars. for signal decays
$\mu_{\text{main}}^{\text{BB}}, s_{\text{main}}^{\text{BB}}$ $\mu_{\text{tail}}^{\text{BB}}, s_{\text{tail}}^{\text{BB}}, f_{\text{tail}}^{\text{BB}}$	5	Resolution function pars. for BB BG decays
$\mu_{\text{main}}^{\text{cont.}}, s_{\text{main}}^{\text{cont.}}$ $\mu_{\text{tail}}^{\text{cont.}}, s_{\text{tail}}^{\text{cont.}}, f_{\text{tail}}^{\text{cont.}}$	5	Resolution func. pars. for continuum decays

We aim to compare the fit results with the one- and two-dimensional fit strategies in ΔE and BDT. This serves to assess the impact of extracting the background composition from the data on the lifetime and the oscillation frequency, which is only possible in the two-dimensional fit. When using the one-dimensional fit in ΔE , all fit parameters in BDT are omitted and BB backgrounds are treated collectively with the continuum backgrounds.

At the time of writing, a malfunction was discovered in the flavor tagger algorithms provided within the `basf2` software framework. In events with multiple B meson candidates, a flavor-tag is computed only for one candidate and is assigned to all candidates. In simulation the average candidate multiplicity is 1.1 and this malfunction induces a significant bias on the fitted value of Δm . This is why, at the time of writing, we cannot yet provide a precision measurement of the lifetime τ_{B^0} and oscillation frequency Δm .

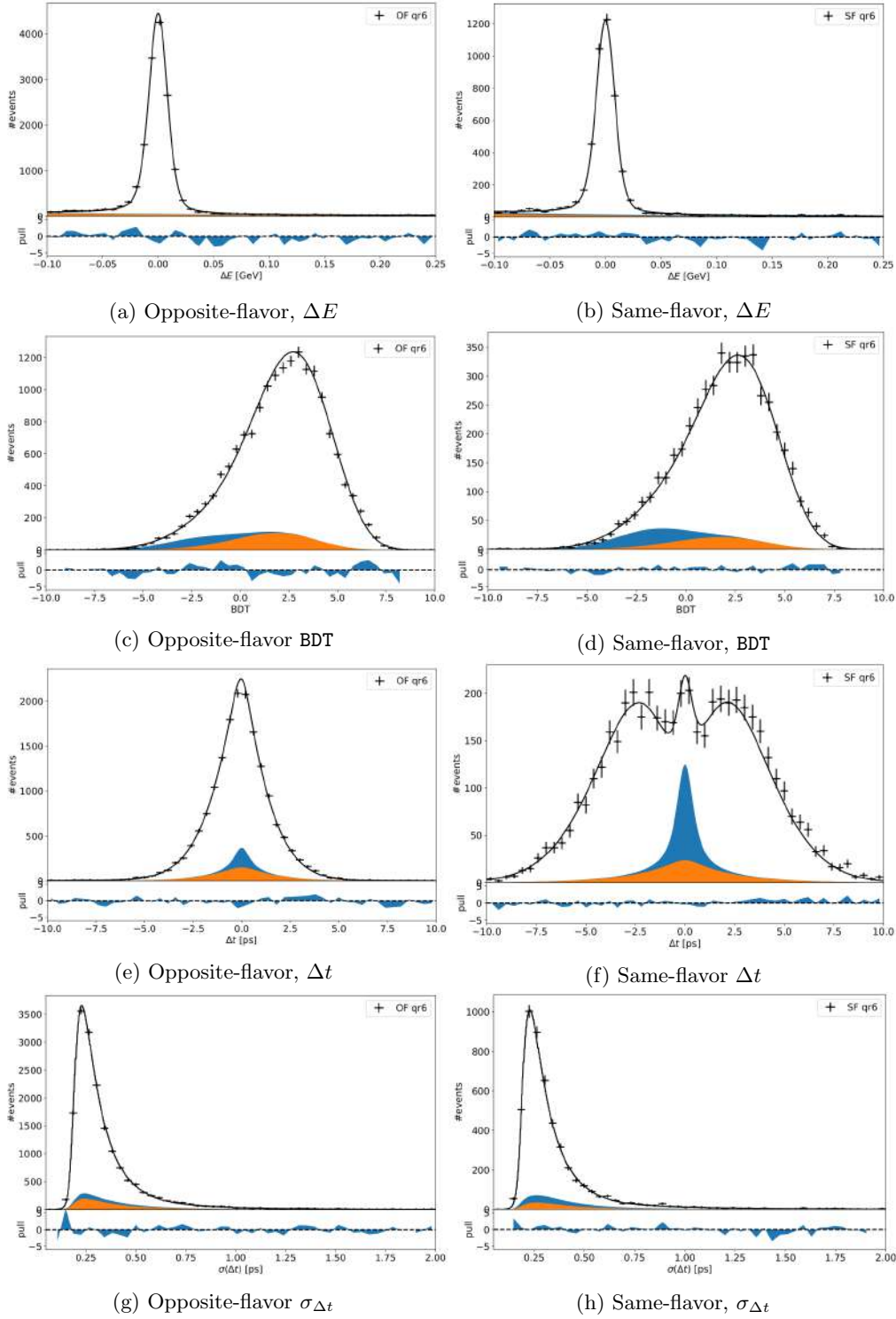


Figure 39: Full fit on simulation in dilution bin 6. Left: Opposite-flavor events. Right: Same-flavor events. From top to bottom: Distributions and fit projections in ΔE , LTBDT, Δt , $\sigma_{\Delta t}$. Fitted BB background contribution in orange, continuum background in blue and the remaining contribution is signal.

9 Conclusion and Outlook

The universe exhibits a matter-antimatter asymmetry, whose origin cannot be explained within the standard model of particle physics. To do so, new sources of charge-parity (CP) symmetry violation in particle decays are a key requirement. A precise measurement of the oscillation frequency in the neutral B meson system is an important probe towards overconstraining standard model predictions for CP violation in the quark sector.

In this thesis we aim for a precision measurement of the lifetime τ_{B^0} and oscillation frequency Δm in the neutral B meson system at the Belle II experiment.

In the Belle II detector at the asymmetric SuperKEKB collider, $B^0\bar{B}^0$ meson pairs are produced from decays of the $\Upsilon(4S)$ resonance. In this analysis we make use of the well-known event kinematics, to identify B meson candidates and to reconstruct the decay time difference Δt of both B mesons. We reconstruct B mesons in three hadronic signal decay channels $B^0 \rightarrow \pi^+ D^{(*)-}$ and employ flavor tagging algorithms to determine the flavor eigenstate of the accompanying B meson decay. This allows us to measure the mixing asymmetry as a function of the decay time difference of both B mesons. From a time-dependent fit to the mixing asymmetry in seven bins of the flavor tag figure-of-merit, we extract the oscillation frequency Δm .

Peaking background decays are decays that mimic signal decays and alter the measured value for Δm and τ_{B^0} . In earlier analyses, peaking background decays were not evaluated in detail and resulted in a systematic uncertainty on the measurement. In this thesis, (peaking) background contributions are analysed in detail from simulation in all three signal decay channels. Based on the findings from the background analysis, we develop a two-dimensional fit strategy to separate signal and background decays and to extract the background composition in Belle II data. We perform unbinned maximum likelihood fits in the energy difference ΔE of the reconstructed signal B meson and in the output of a continuum suppression boosted decision tree (BDT) that was trained on simulated event samples. In simulation, the fit consistently tells apart signal and three kinds of backgrounds with deviations of less than 10%. Pseudo-experiment studies indicate a stable and non-biased fit. In $176.9 \pm 12.6 \text{ fb}^{-1}$ of Belle II data, the fit results agree with the expectations from simulation. For comparison, we also implement and test a one-dimensional fit strategy in ΔE .

Our full fit strategy for the lifetime and oscillation frequency measurement amounts to a four-dimensional fit in ΔE , BDT, Δt and in the uncertainty on the decay time difference $\sigma_{\Delta t}$. Tests in simulation indicate a good fit quality in all fit variables and a promising signal and background separation. However at the time of writing, a malfunction in the flavor tagger algorithm induces significant biases on the fitted value for the oscillation frequency Δm .

Outlook

Once the malfunction in the flavor tagger is corrected, we hope to demonstrate that our fit strategy performs better than previous analyses and hope to provide the most accurate measurement yet of τ_{B^0} and Δm at Belle II. Furthermore we aim to show that the two-dimensional fit strategy in ΔE and BDT is superior compared to the one-dimensional fit in ΔE . In contrast to the two-dimensional fit, the one-dimensional fit cannot extract the background composition from data. Due to this, we expect smaller systematic uncertainties on the measured values with the two-dimensional fit strategy. By comparing the fit results for the two strategies, we aim to quantify the improvement.

Acknowledgement

I would like to thank everybody who has made this work possible and has supported me throughout my master studies in Munich.

First and foremost, I would like to thank Prof. Dr. Christian Kiesling and Dr. Hans-Günther Moser for welcoming me in the Belle II group at MPP and for supporting my research with valuable discussions.

A great thank you goes to Dr. Thibaud Humair for his supervision and for taking the time to answer my countless questions. I am looking forward to continue working on our analysis! I would also like to express my gratitude to the members of the TDCPV group at Belle II and to the members of the MPP for their help.

I thank Christian, Hans-Günther and Thibaud for reviewing and suggesting improvements for this thesis.

Finally, I would like to thank my parents and Jackie for their support over the last 2 years.

A Categorization Tool for Reconstructed Decays

In the generation of simulation samples, each generated particle is equipped with its PDG-identifier `genParticlePDG` denoting the particle type, and a unique identification code `genParticleID`. We use the latter to distinguish between particles of the same type.

For every reconstructed final state particle, we extract both `genParticlePDG` and `genParticleID` for the corresponding truth-matched particle. Subsequently, we also extract both variables for all of their ancestor-particles up to five generations down the generator decay-chain. Typically in the described channels, the mother-particles are descendingly identified as charm mesons, then B mesons and finally as the $\Upsilon(4S)$ resonance (Equation 43).

$$\begin{array}{l} \text{truth-matched particle of} \\ \text{reconstructed final state } \pi, K \end{array} \xrightarrow[\text{mother}]{\text{generator}} \bar{D}^0 \xrightarrow[\text{mother}]{\text{generator}} D^* \xrightarrow[\text{mother}]{\text{generator}} B^0 \xrightarrow[\text{mother}]{\text{generator}} \Upsilon(4S) \quad (43)$$

For falsely reconstructed decays, they differ from this. After having identified the ancestors of the reconstructed final state particles, we reassemble them to form the true generated decay-chain. Particles with the same mother-particle are grouped together as daughter particles. This remains functional even when multiple particles of the same type decay, since the mother-particle is uniquely identified through `genParticleID`.

In the simulation sample generation, the assigned value of the unique `genParticleID` ascends from the primary $\Upsilon(4S)$ resonance to the final state particles. Typically, the $\Upsilon(4S)$ resonance gets assigned a value of 0, and its two daughter B mesons values of 1 and 2 respectively. Depending on the number of generated particles in the event, increasing values are assigned until the maximal values are assigned to the final state particles. Therefore to reassemble the decay-chain, a hierarchical approach is applied, starting from the highest `genParticleID` and grouping daughter-particles in descending order.

In some cases, particles of different generations in the generator decay-chain are direct joint daughters of a mother-particle. For instance, in the decay of a D^0 meson in Equation 44.

$$\left. \begin{array}{l} \mu^+ \xrightarrow[\text{mother}]{\text{generator}} \pi^+ \\ \phantom{\xrightarrow[\text{mother}]{\text{generator}}} K^- \end{array} \right\} \xrightarrow[\text{mother}]{\text{generator}} D^0 \quad (44)$$

The π^+ and K^- are of different generations in the generator decay-chain, but must be matched as joint daughters of the D^0 . Hence it is important, to first assign the μ^+ , which has the highest `genParticleID`, to its mother-pion. Only afterwards, we can group the D^0 daughter-particles.

B Correlations between ΔE and M_{bc}

Figure 40 shows the correlations between ΔE and M_{bc} in the total simulation sample, analogously to Figures 26 and 27. Here the correlations in signal are larger, with about 14% compared to 3% in Figure 26. This may be ascribed to the common dependence on the measured beam energy (see Equation 41). In any case, the signal distributions in ΔE and M_{bc} cannot simply be assumed independent. For a precision lifetime and mixing measurement, a combined fit in $(\Delta E, M_{bc}, \Delta t)$ would thus require to take the correlations into account. This is why we instead adopted a fit strategy in $(\Delta E, \text{BDT}, \Delta t)$, where the correlations are negligible (Figures 26 and 27).

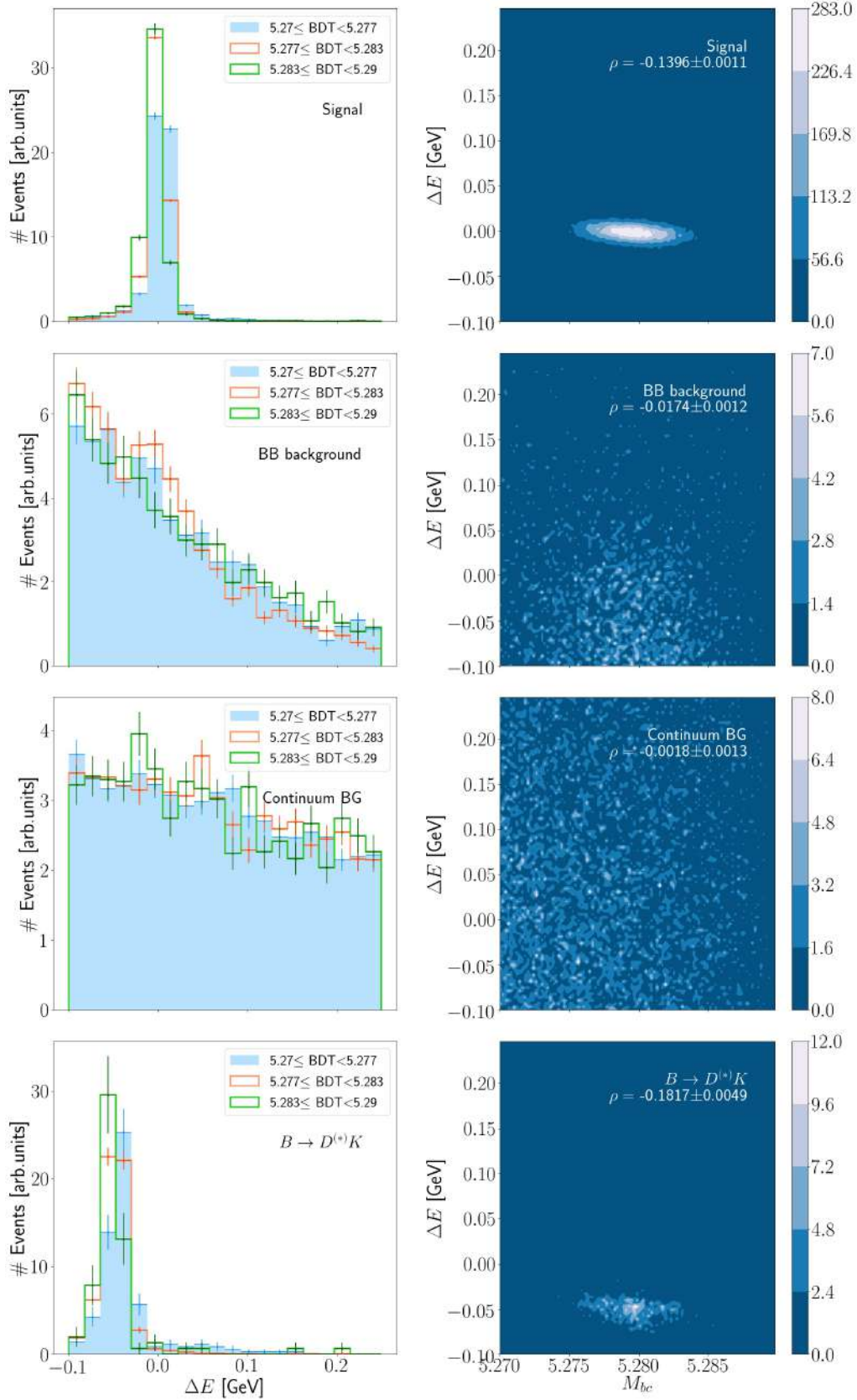


Figure 40: Correlations between ΔE and M_{bc} in total simulation sample combining all channels.

C Extracted Shapes for Two-Dimensional Fit in all Dilution Bins for Same- and Opposite-Flavor Events

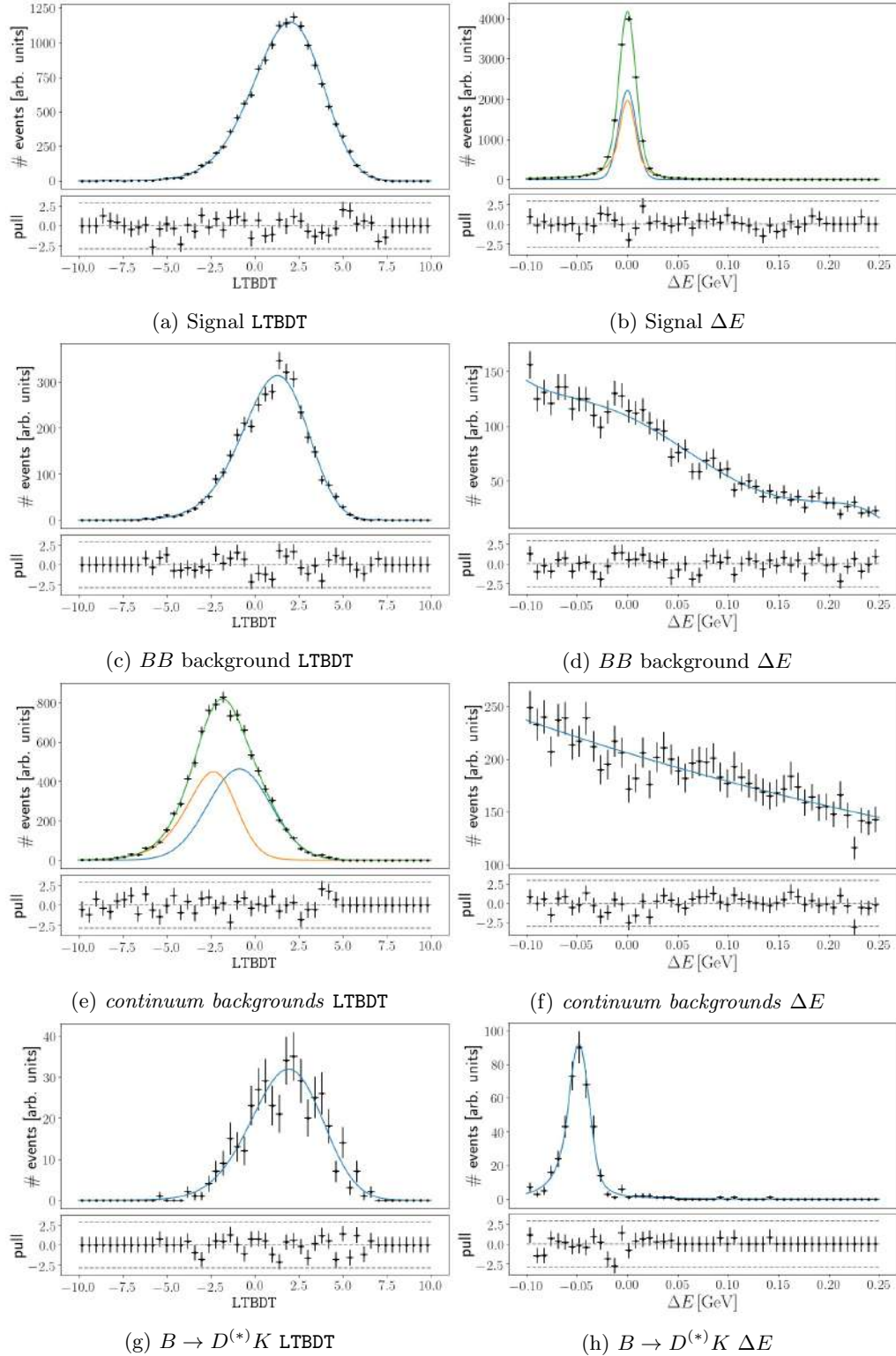


Figure 41: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Same-flavor events, bin 0 in $q \cdot r$.

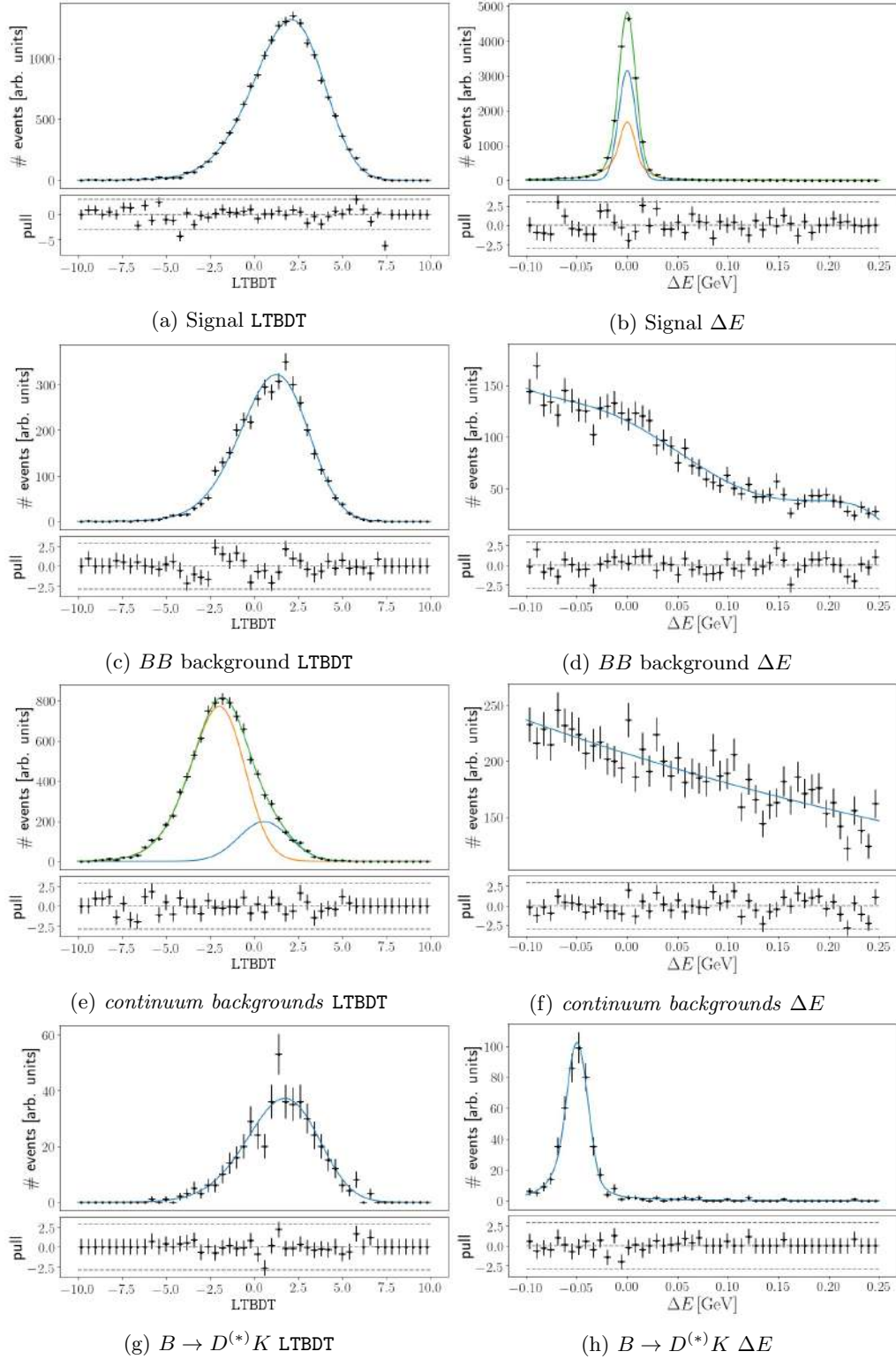


Figure 42: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Opposite-flavor events, bin 1 in $q \cdot r$.

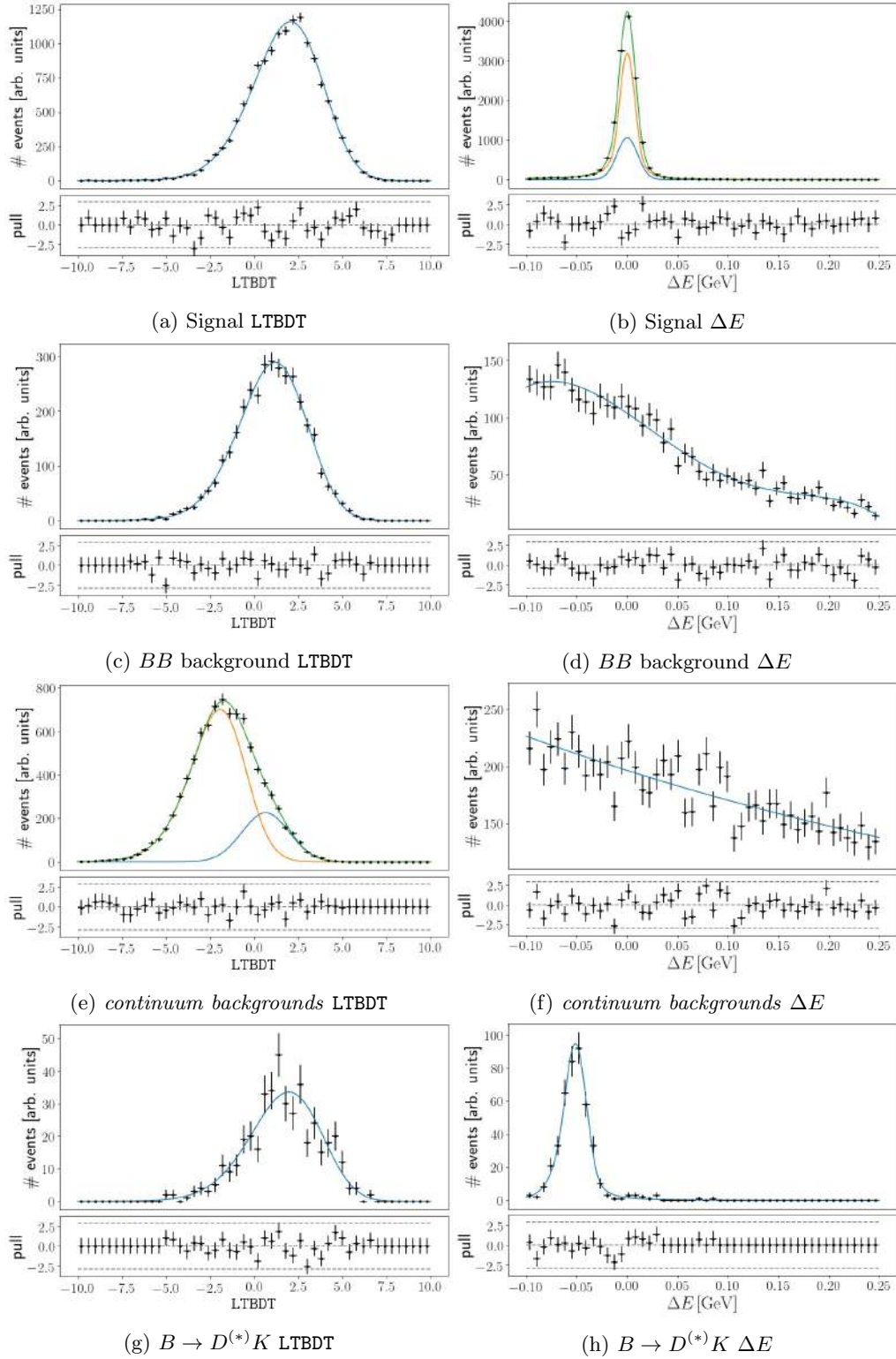


Figure 43: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Same-flavor events, bin 1 in $q \cdot r$.

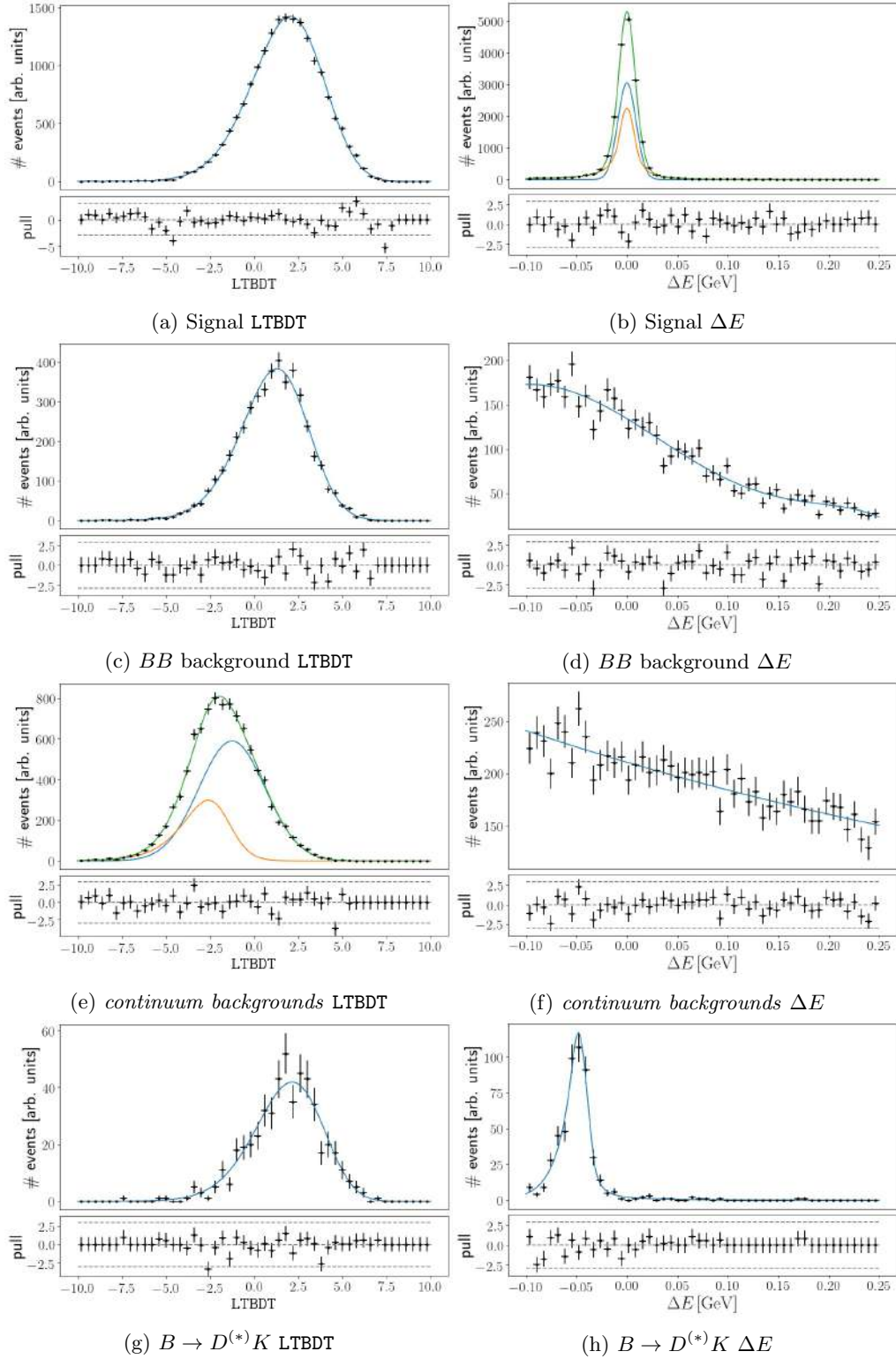


Figure 44: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Opposite-flavor events, bin 2 in $q \cdot r$.

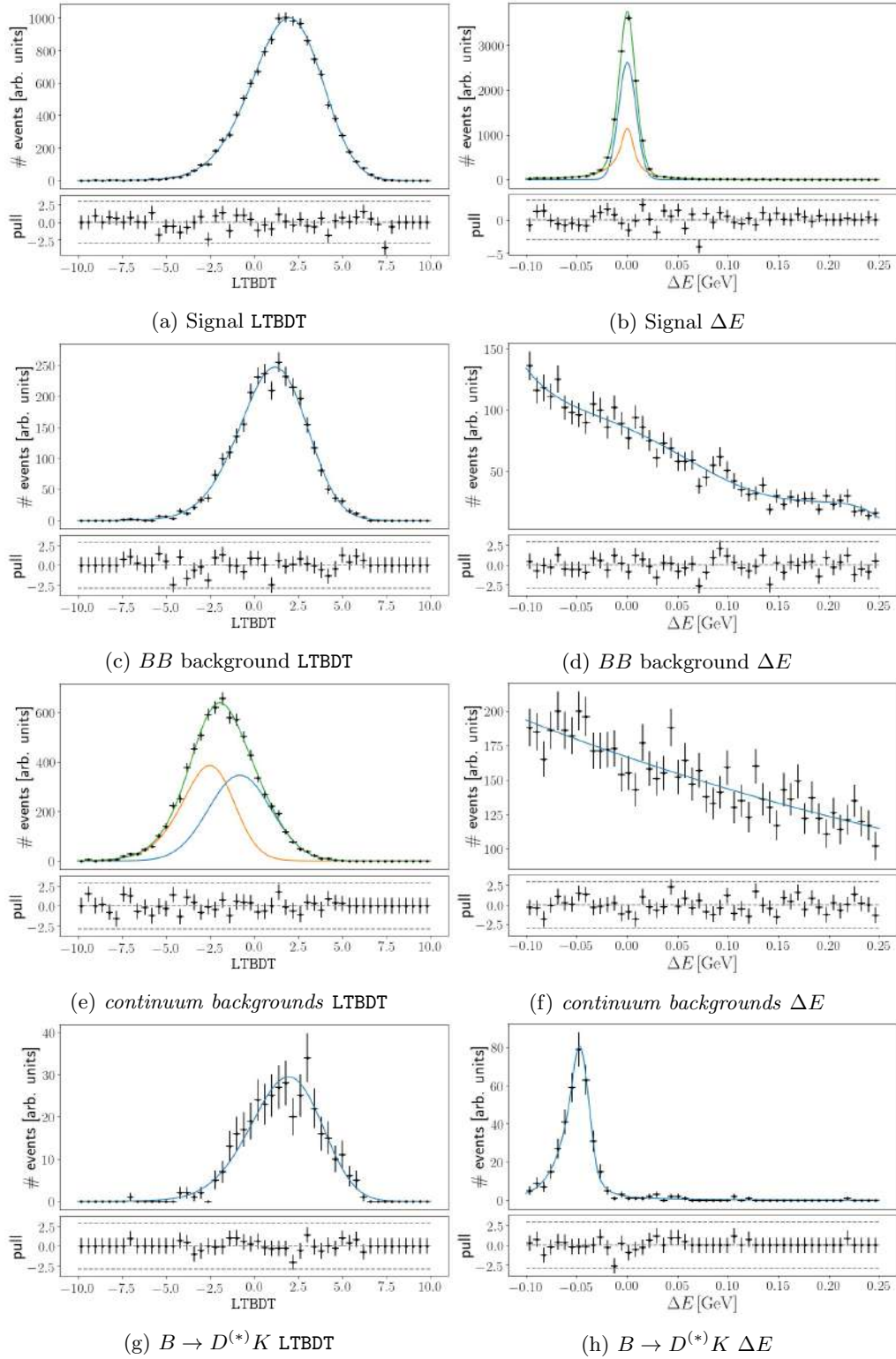


Figure 45: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Same-flavor events, bin 2 in $q \cdot r$.

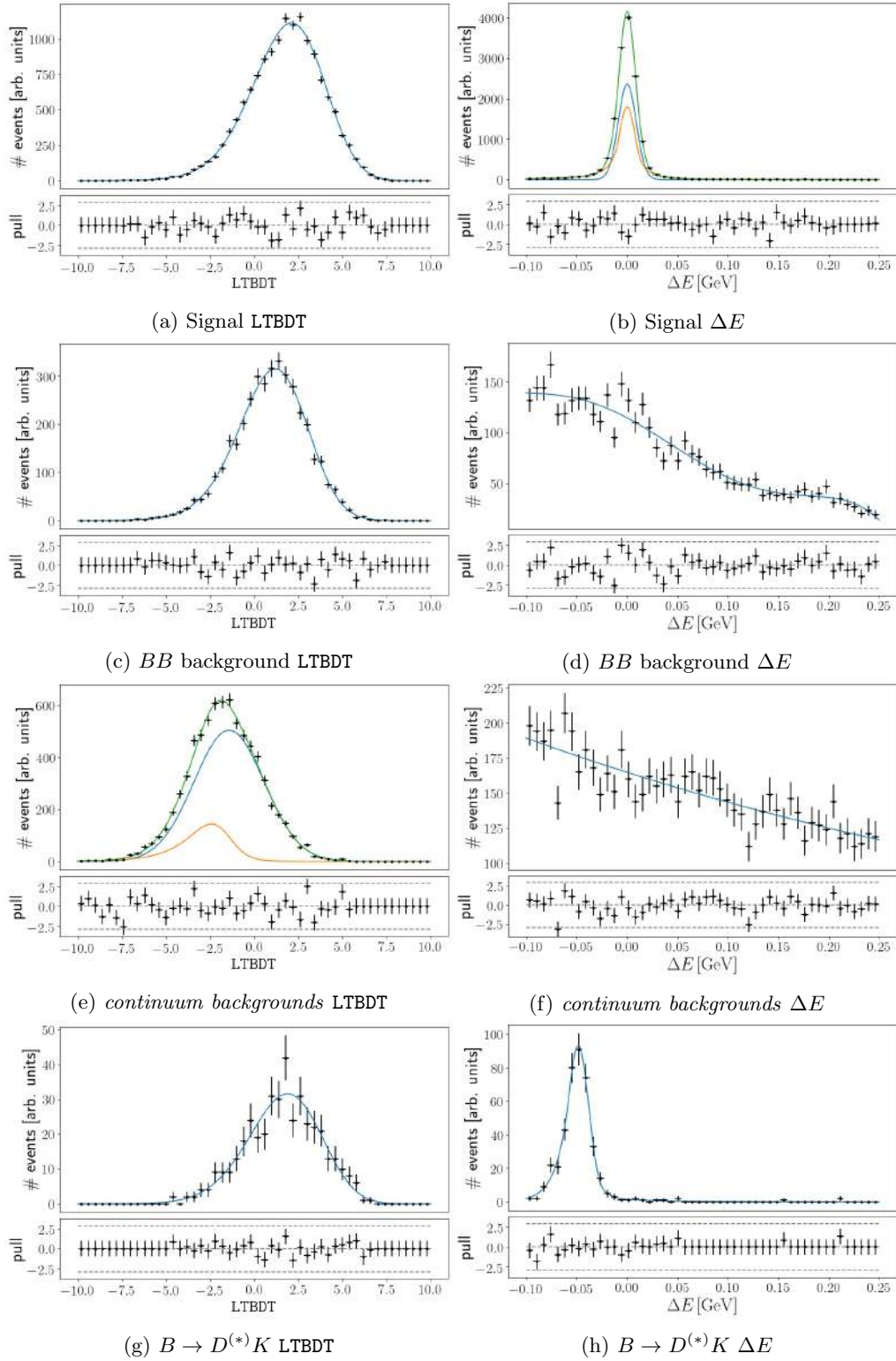


Figure 46: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Opposite-flavor events, bin 3 in $q \cdot r$.

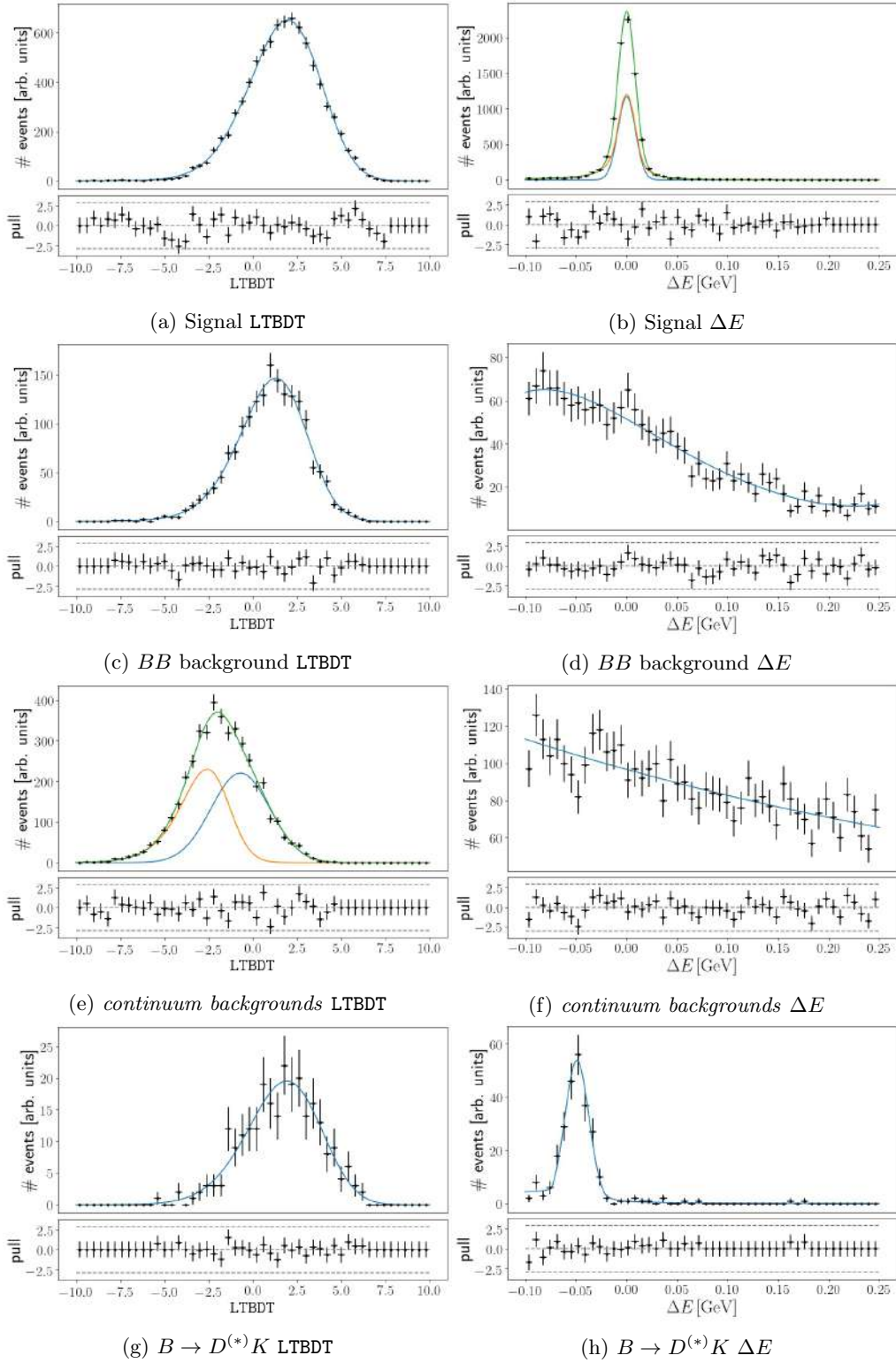


Figure 47: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Same-flavor events, bin 3 in $q \cdot r$.

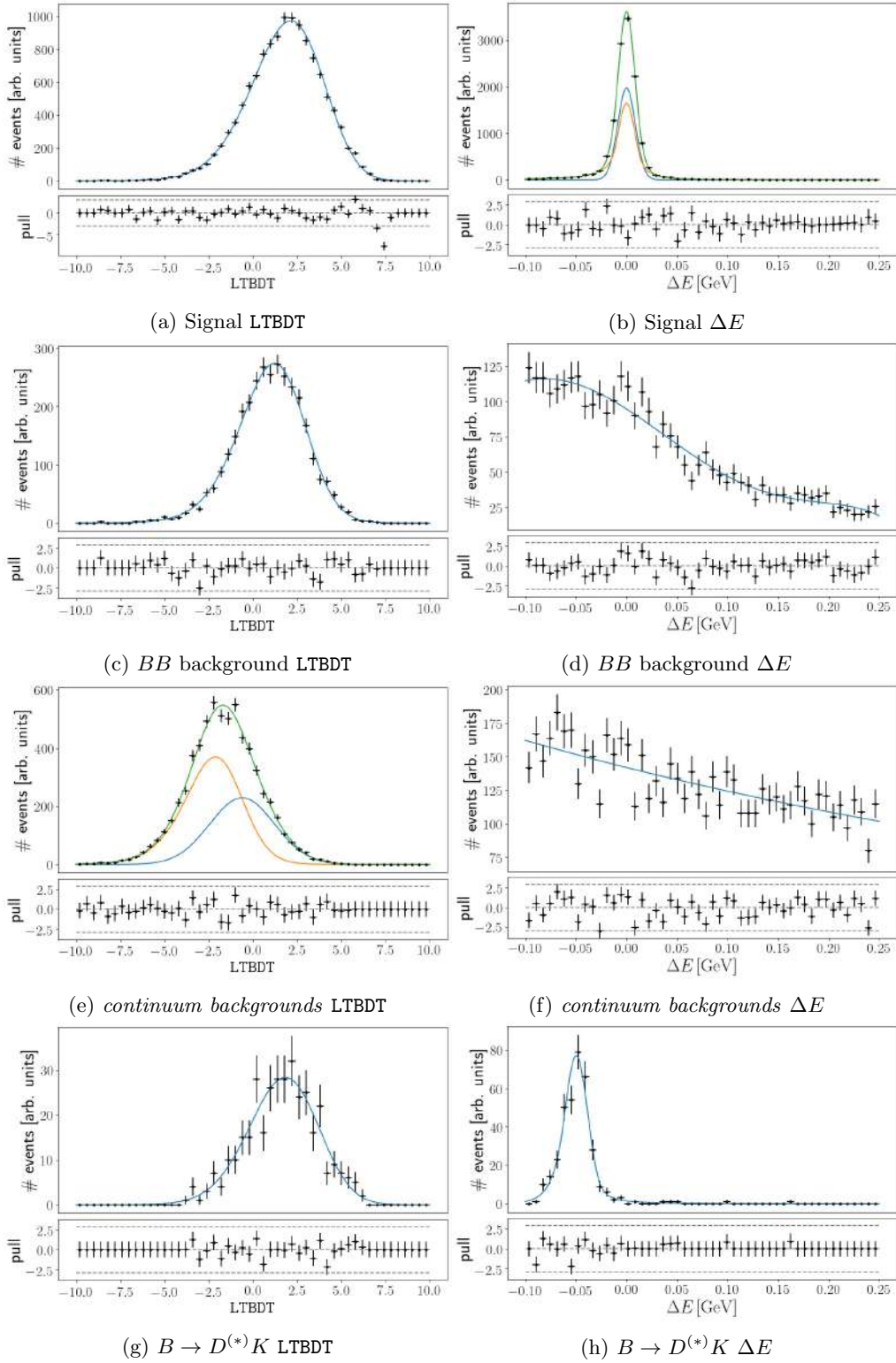


Figure 48: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Opposite-flavor events, bin 4 in $q \cdot r$.

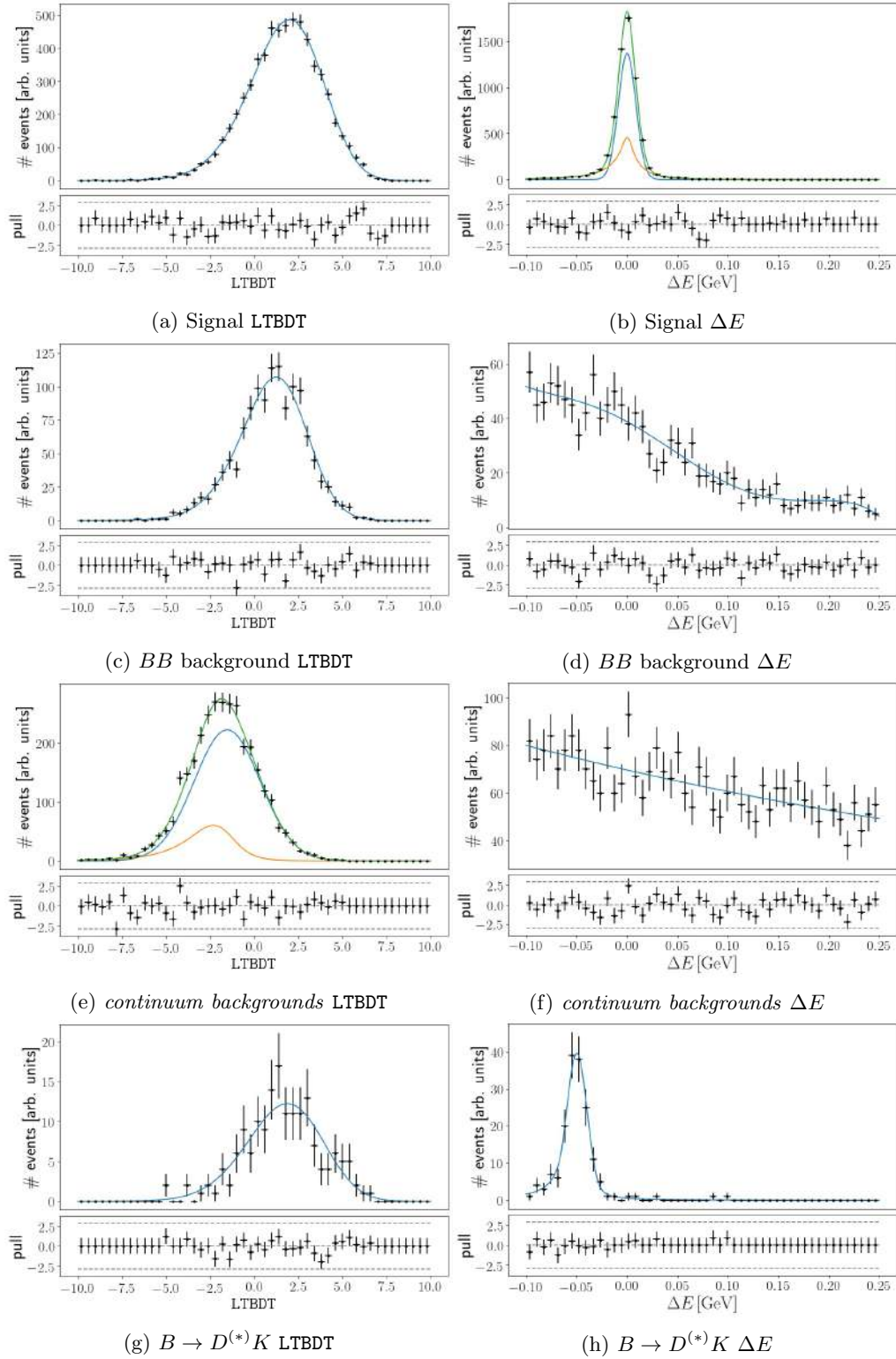


Figure 49: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Same-flavor events, bin 4 in $q \cdot r$.

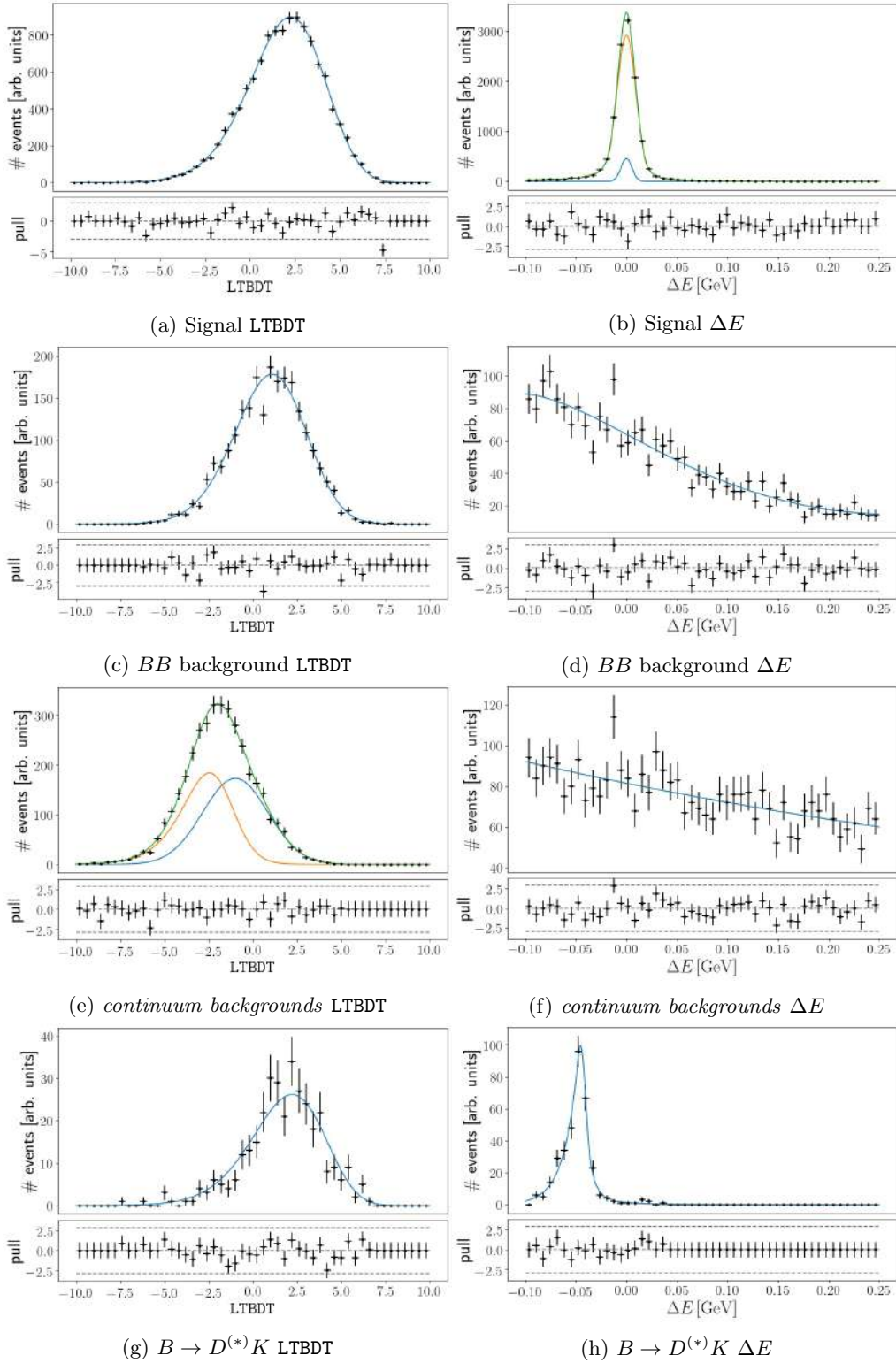


Figure 50: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Opposite-flavor events, bin 5 in $q \cdot r$.

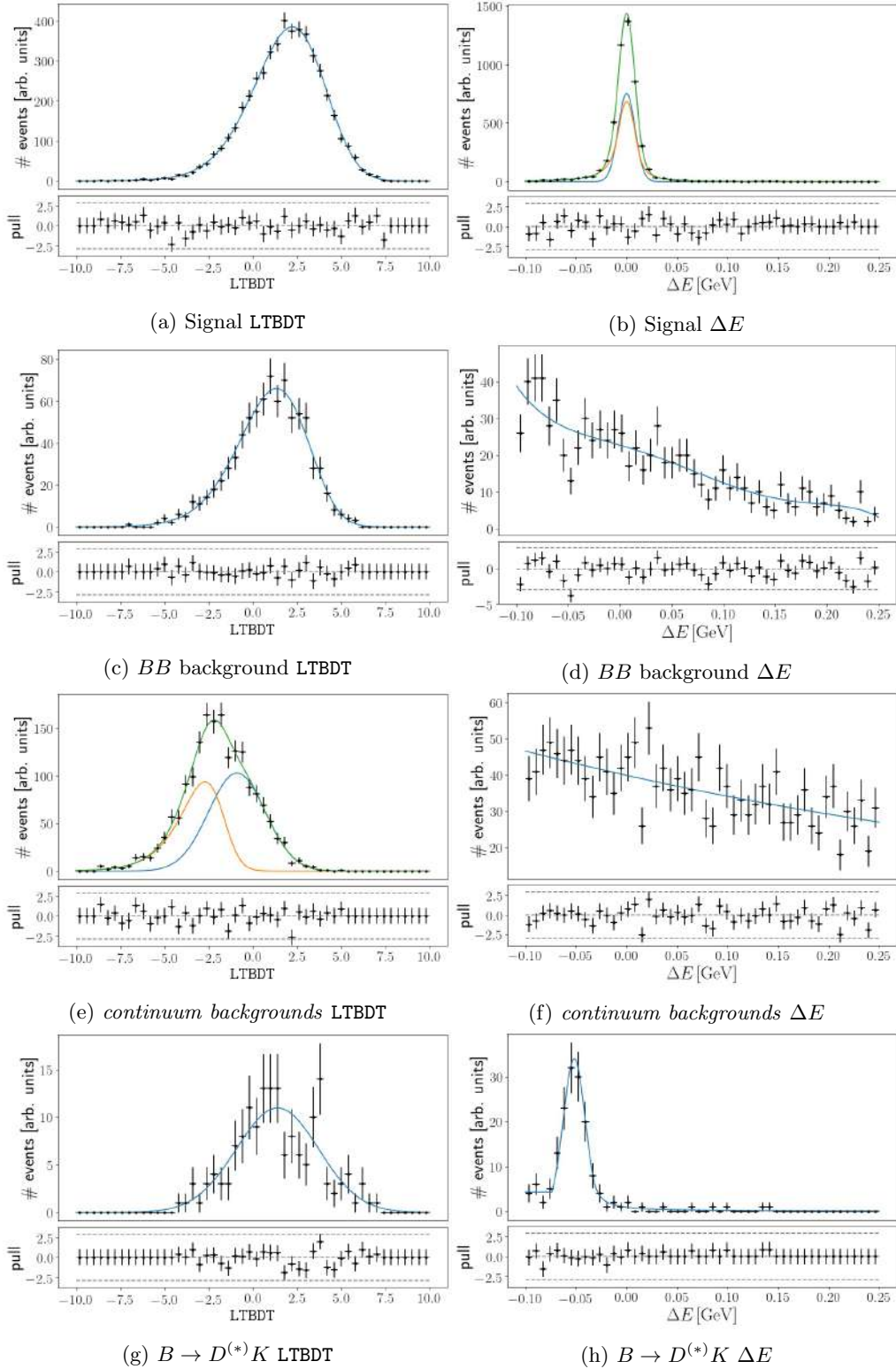


Figure 51: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Same-flavor events, bin 5 in $q \cdot r$.

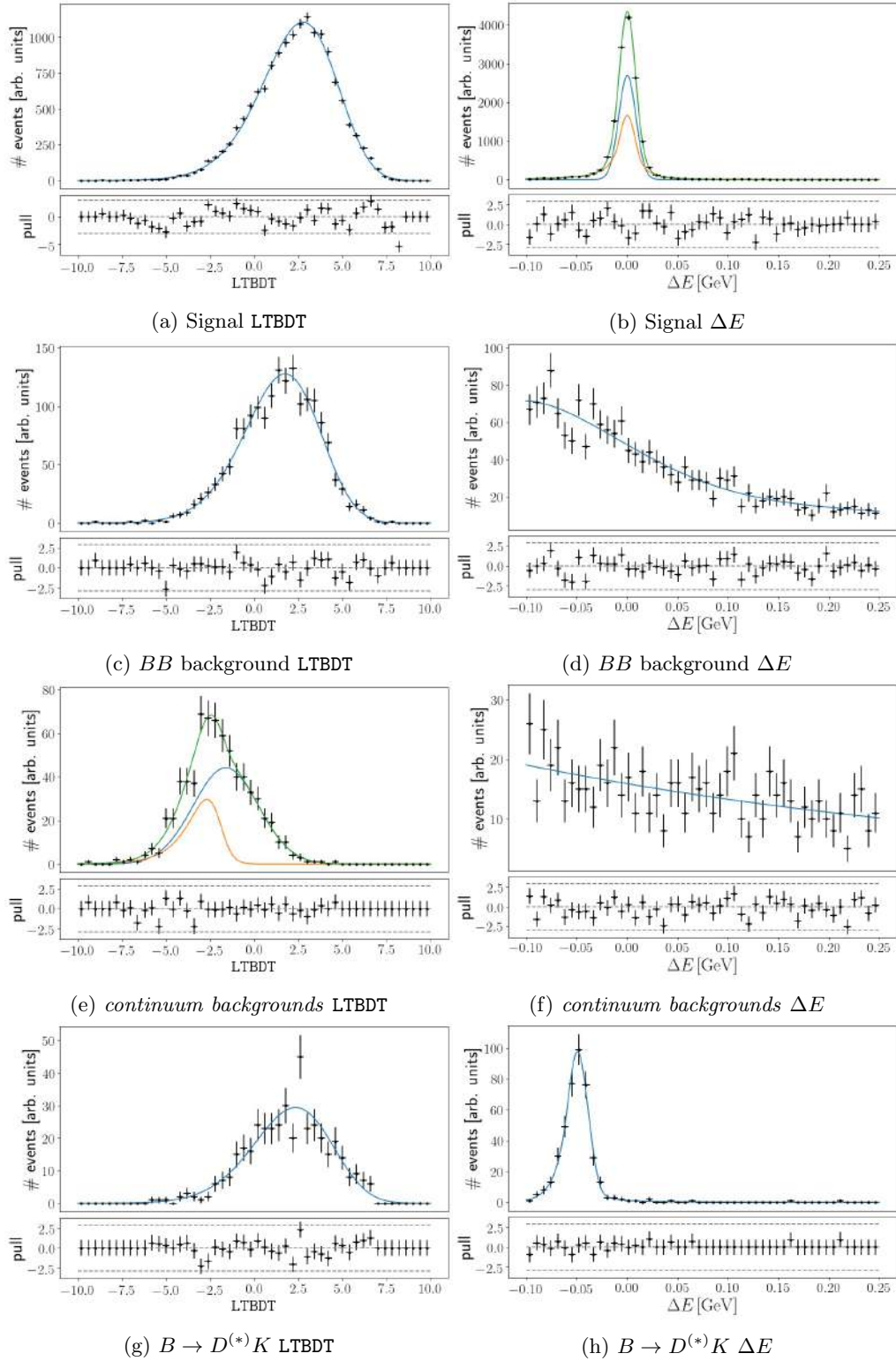


Figure 52: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Opposite-flavor events, bin 6 in $q \cdot r$.

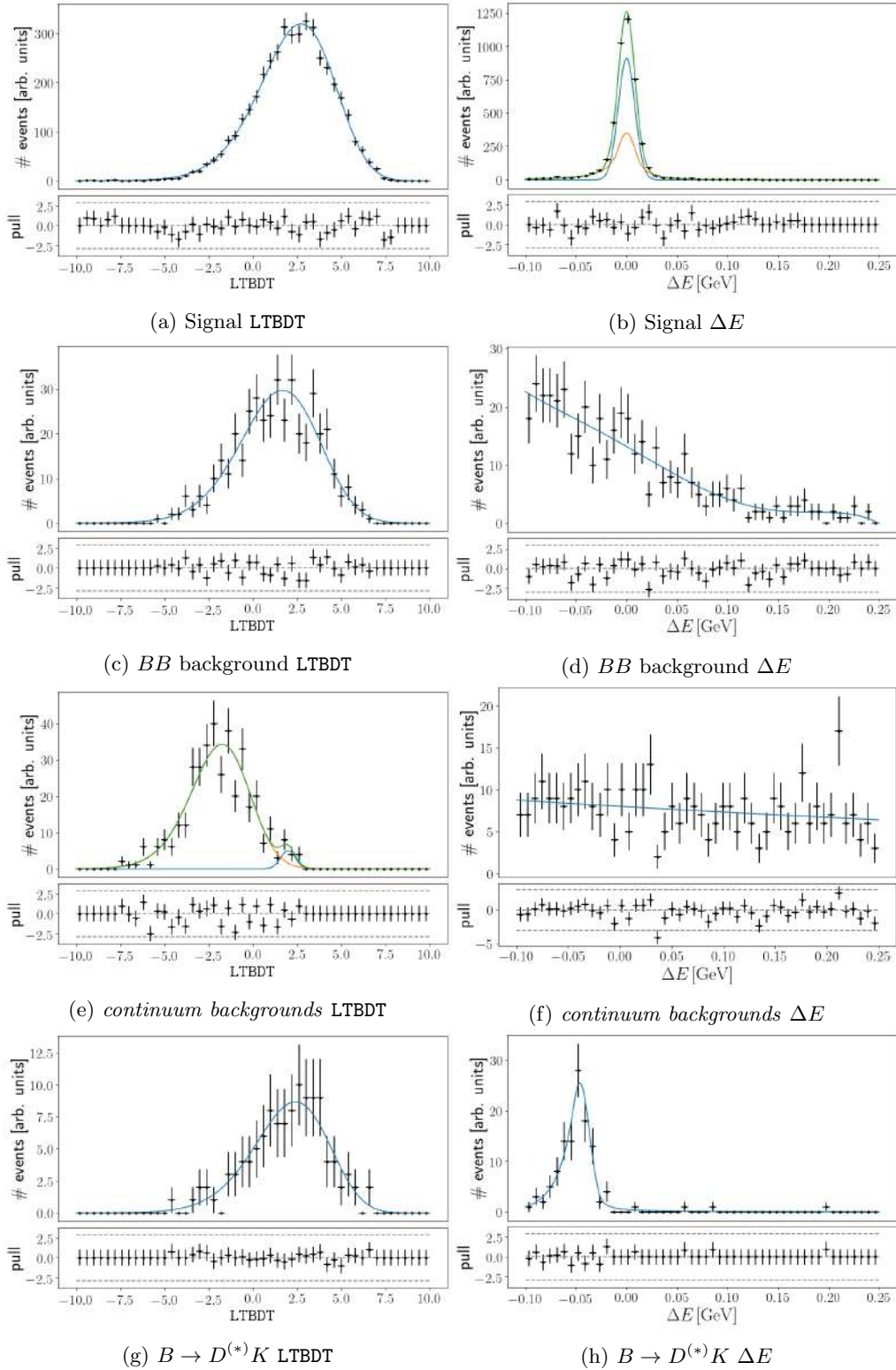


Figure 53: Simulated distributions in LTBDT (left) and ΔE (right) with extracted shapes and pull distributions. Same-flavor events, bin 6 in $q \cdot r$.

D Extracted Shapes for One-Dimensional Fit in all Dilution Bins for Same- and Opposite-Flavor Events

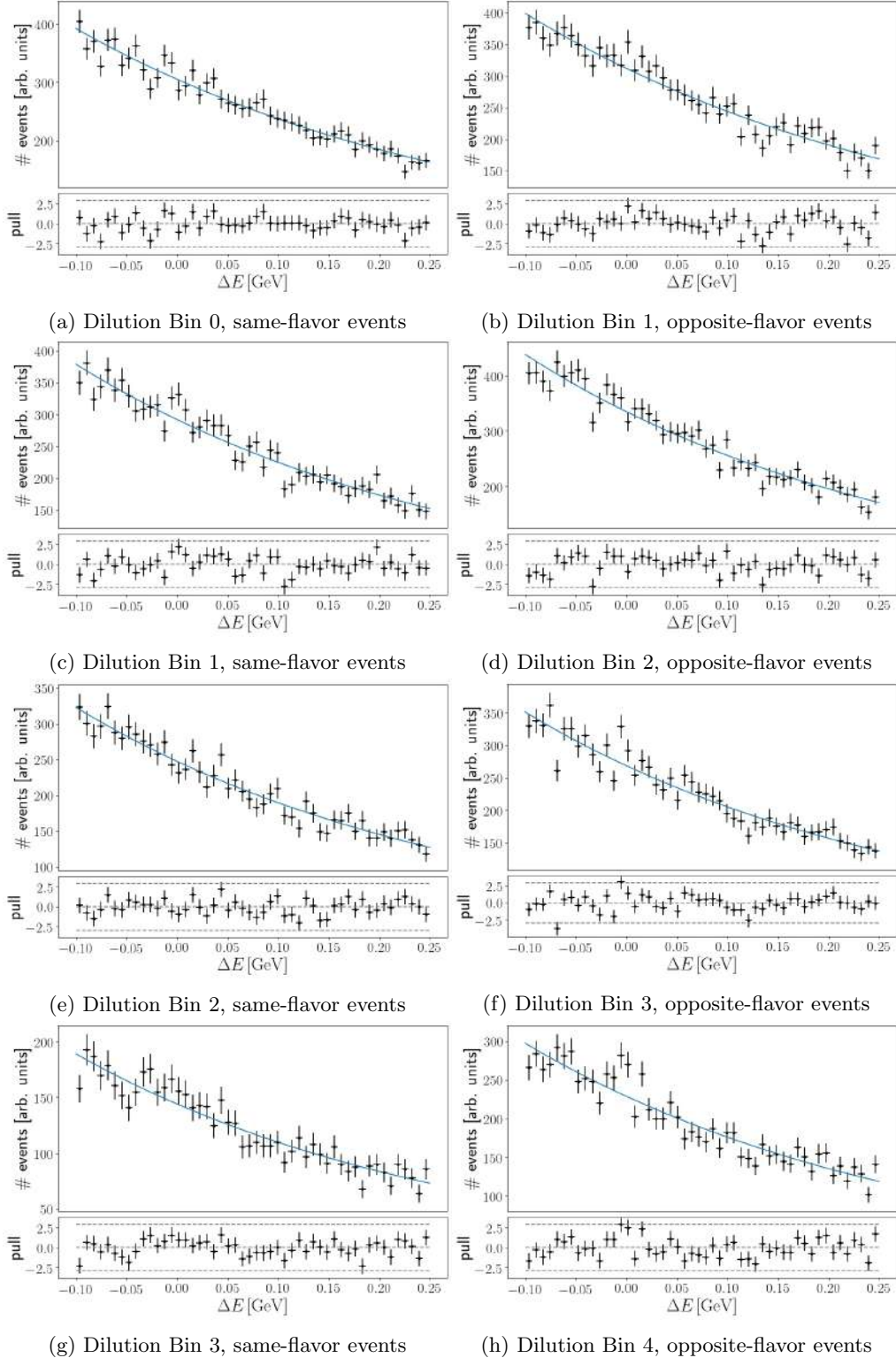


Figure 54: Simulated distributions in ΔE with extracted shapes and pull distributions.

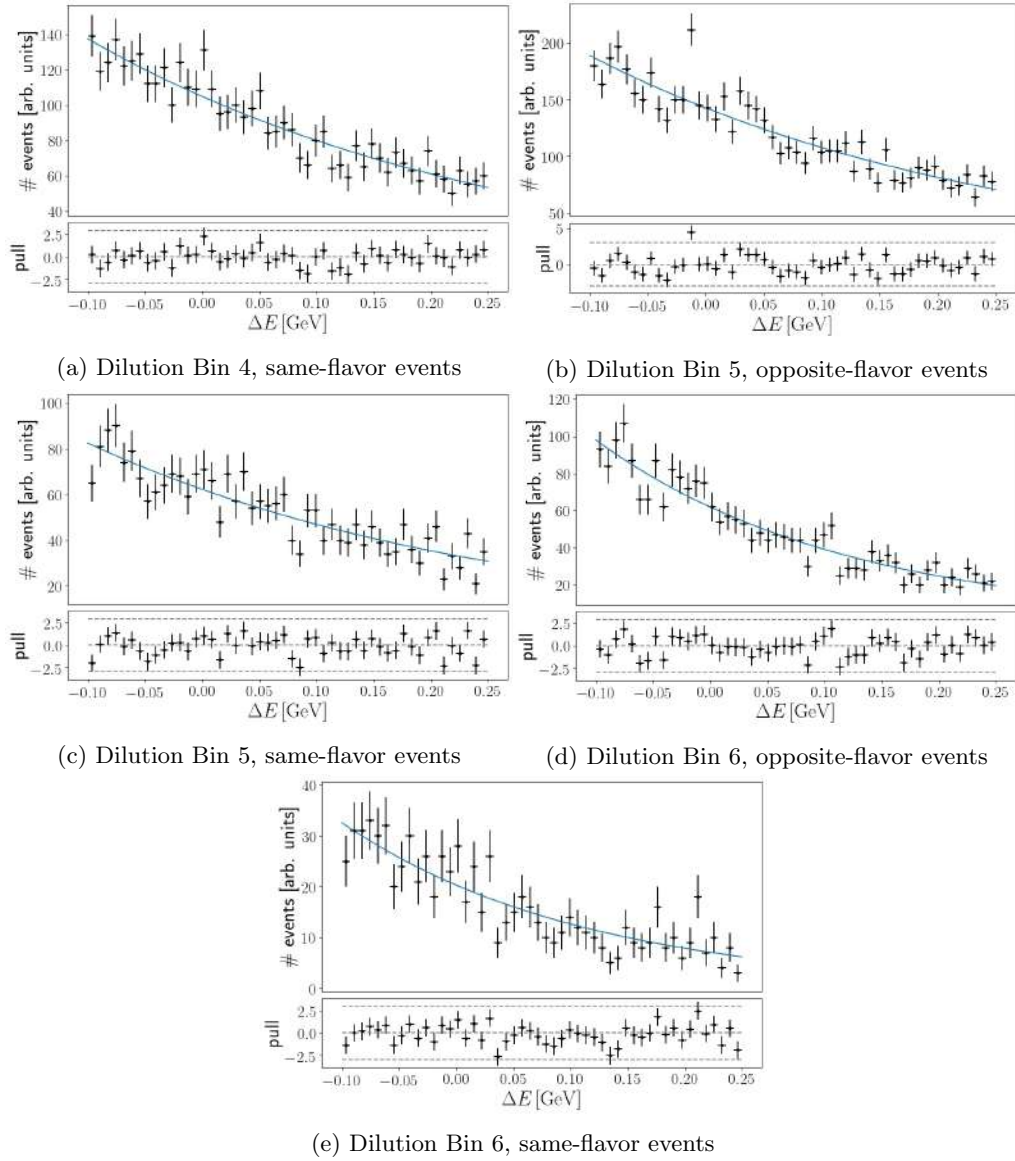


Figure 55: Simulated distributions in ΔE with extracted shapes and pull distributions.

References

- [1] Pictorial Source: Standard model of particle physics. <https://commons.wikimedia.org>.
- [2] Mark Thomson. *Modern particle physics*. Cambridge University Press, 2013.
- [3] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.
- [4] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A.A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, and et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, Sep 2012.
- [5] S. Chatrchyan et. al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(1):30–61, 2012.
- [6] Adrian Bevan, Bostjan Golob, Thomas Mannel, Soeren Prell, and Bruce Yabsley. *The physics of the B factories*. Springer Nature, 2017.
- [7] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Physical Review Letters*, 10(12):531, 1963.
- [8] Sheldon L Glashow, Jean Iliopoulos, and Luciano Maiani. Weak interactions with lepton-hadron symmetry. *Physical review D*, 2(7):1285, 1970.
- [9] Tsung-Dao Lee and Chen-Ning Yang. Question of parity conservation in weak interactions. *Physical Review*, 104(1):254, 1956.
- [10] Chien-Shiung Wu, Ernest Ambler, Raymond W Hayward, Dale D Hoppes, and Ralph Percy Hudson. Experimental test of parity conservation in beta decay. *Physical review*, 105(4):1413, 1957.
- [11] James H Christenson, Jeremiah W Cronin, Val L Fitch, and René Turlay. Evidence for the 2π decay of the K^0 meson. *Physical Review Letters*, 13(4):138, 1964.
- [12] Makoto Kobayashi and Toshihide Maskawa. Cp-violation in the renormalizable theory of weak interaction. *Progress of theoretical physics*, 49(2):652–657, 1973.
- [13] Lincoln Wolfenstein. Parametrization of the kobayashi-maskawa matrix. *Physical Review Letters*, 51(21):1945, 1983.
- [14] Konrad Kleinknecht and Burkhard Renk. Unitarity triangle from cp invariant quantities. *Physics Letters B*, 639(6):612–615, 2006.
- [15] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.
- [16] H. Albrecht et al. Observation of B^0 - anti- B^0 Mixing. *Phys. Lett. B*, 192:245–252, 1987.
- [17] Bernard Aubert, D Boutigny, J-M Gaillard, A Hicheur, Y Karyotakis, JP Lees, P Robbe, V Tisserand, A Palano, GP Chen, et al. Observation of cp violation in the B^0 meson system. *Physical review letters*, 87(9):091801, 2001.
- [18] Kazuo Abe, R Abe, I Adachi, Byoung Sup Ahn, H Aihara, M Akatsu, G Alimonti, K Asai, M Asai, Y Asano, et al. Observation of large cp violation in the neutral b meson system. *Physical Review Letters*, 87(9):091802, 2001.
- [19] World averages
<https://hflav-eos.web.cern.ch/hflav-eos/triangle/pdg2021/index.shtml#sin2b>.

- [20] O Adriani, GC Barbarino, GA Bazilevskaya, R Bellotti, M Boezio, EA Bogomolov, M Bongi, V Bonvicini, S Bottai, A Bruno, et al. The pamel mission: Heraldng a new era in precision cosmic ray physics. *Physics Reports*, 544(4):323–370, 2014.
- [21] Martin Harwit. What is the evidence on antigalaxies? *The Astrophysical Journal*, 347:688–691, 1989.
- [22] Andrew Liddle. *An introduction to modern cosmology*. John Wiley & Sons, 2015.
- [23] Cosimo Bambi and Alexandre D Dolgov. *Introduction to particle cosmology: The standard model of cosmology and its open problems*. Springer, 2015.
- [24] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and antimatter in the universe. *New Journal of Physics*, 14(9):095012, 2012.
- [25] Andrei D Sakharov. Violation of cp-invariance, c-asymmetry, and baryon asymmetry of the universe. In *In The Intermissions... Collected Works on Research into the Essentials of Theoretical Physics in Russian Federal Nuclear Center, Arzamas-16*, pages 84–87. World Scientific, 1998.
- [26] T Gershon and VV Gligorov. Cp violation in the b system. *Reports on Progress in Physics*, 80(4):046201, 2017.
- [27] Renata Jora. Baryon asymmetry in the standard model revisited. *arXiv preprint arXiv:1806.00597*, 2018.
- [28] Ckm fitter group. http://ckmfitter.in2p3.fr/www/html/ckm_results.html.
- [29] K Anikeev, D Atwood, F Azfar, S Bailey, CW Bauer, William Bell, G Bodwin, E Braaten, G Burdman, JN Butler, et al. B physics at the tevatron: Run ii and beyond. *arXiv preprint hep-ph/0201071*, 2002.
- [30] U. Uwer. Advanced Particle Physics: IX. Flavor Oscillation and CP Violation. https://www.physi.uni-heidelberg.de/~uwer/lectures/ParticlePhysics/Vorlesung_SS07/Lect-9a.pdf.
- [31] Ulrich Nierste. Three lectures on meson mixing and ckm phenomenology. *arXiv preprint arXiv:0904.1869*, 2009.
- [32] T. Abe et al. Belle ii technical design report, 2010.
- [33] E Kou, P Urquijo, W Altmannshofer, F Beaujean, G Bell, M Beneke, I I Bigi, F Bishara, M Blanke, C Bobeth, and et al. The belle ii physics book. *Progress of Theoretical and Experimental Physics*, 2019(12), Dec 2019.
- [34] Particle Data Group. <https://pdg.lbl.gov/>.
- [35] Kazunori Akai, Kazuro Furukawa, Haruyo Koiso, et al. Superkekb collider. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 907:188–199, 2018.
- [36] Dmitry Matvienko. The belle ii experiment: status and physics program. In *EPJ Web of Conferences*, volume 191, page 02010. EDP Sciences, 2018.
- [37] Justin Skorupa. Characterization of the Particle Identification of the Belle II Detector with Bhabha Radiation and First Measurement of $\sin \phi_1$ and Δm_d Using Belle II data. Master’s thesis, TUM Munich, 2021.
- [38] Belle ii software documentation online book. <https://software.belle2.org/development/sphinx/index.html>.

- [39] Fernando Abudinén (PhD thesis). Development of a B0 flavor tagger and performance study of a novel time-dependent CP analysis of the decay $B^0 \rightarrow \pi^0 \pi^0$ at Belle II. <https://docs.belle2.org/record/1215?ln=en>.
- [40] Fernando Abudinén, Takeo Higuchi, and Colm Murphy. First flavor tagging calibration using 2019 belle ii data. Jul 2020.
- [41] Thomas Kuhr. Experimental Flavor Physics LMU Munich. <https://www.flavor.physik.uni-muenchen.de>.
- [42] Thibaud Humair. Resolution function for time-dependent fits. <https://indico.belle2.org/event/3922/contributions/22963/attachments/11268/17222/pres.pdf>.
- [43] K Abe, I Adachi, H Aihara, M Akatsu, Y Asano, T Aushev, T Aziz, S Bahinipati, AM Bakich, Y Ban, et al. Improved measurement of cp-violation parameters $\sin 2 \phi_1$ and $-\lambda$, b meson lifetimes, and $b^0 \rightarrow b^0$ mixing parameter δ m d. *Physical Review D*, 71(7):072003, 2005.
- [44] Vladimir Chekelian, Thibaud Humair, Jakub Kandra, Justin Skorupa, and Radek Žlebčik. Prompt $b^0 \rightarrow j/\psi k_s$ analysis. Jun 2020.
- [45] Valerio Bertacchi, Tadeas Bilka, Nils Braun, Giulia Casarosa, Luigi Corona, Sam Cunliffe, Filippo Dattola, Gaetano De Marino, Michael De Nuccio, Giacomo De Pietro, and et al. Track finding at belle ii. *Computer Physics Communications*, 259:107610, Feb 2021.
- [46] Jakub Kandra. Background suppression using boosted decision trees. https://indico.belle2.org/event/3922/contributions/22962/attachments/11273/17227/Lifetime-07_06.pdf.
- [47] D. Weyland. Continuum suppression with deep learning techniques for the belle ii experiment. 2017.
- [48] Selim Gündüz and Mustafa Korkmaz. A new unit distribution based on the unbounded johnson distribution rule: The unit johnson s_u distribution. *Pakistan Journal of Statistics and Operation Research*, 16, 09 2020.
- [49] zfit. zfit documentation. https://zfit.readthedocs.io/en/latest/user_api_generated/pdf/zfit.pdf.Chebyshev.html.
- [50] Belle ii data production states. <https://confluence.desy.de/display/BI/Data+Production+Status>.

Erklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig verfasst zu haben und keine als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

München, 13. September 2021

Caspar Schmitt