

ÉCOLE DOCTORALE DE PHYSIQUE ET CHIMIE-PHYSIQUE

Institut Pluridisciplinaire Hubert Curien (IPHC), UMR 7178

THESIS presented by:

Petros STAVROULAKIS

defended on: 20 November 2025

to obtain the grade of: Doctor of the University of Strasbourg

Discipline: Particle Physics

Measurement of the time-dependent
 CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$
decays in the Belle II experiment

THESIS directed by:

Dr. Christian FINCK
Dr. Isabelle RIPP-BAUDOT

Institut Pluridisciplinaire Hubert Curien
Institut Pluridisciplinaire Hubert Curien

REFEREES:

Dr. Marie-Hélène SCHUNE
Dr. Anton POLUEKTOV

Laboratoire de Physique des 2 Infinis Irène Joliot-Curie
Centre de Physique des Particules de Marseille

OTHER MEMBERS OF THE JURY:

Dr. Iouri BELIKOV
Dr. Thibaud HUMAIR

Institut Pluridisciplinaire Hubert Curien
Deutsches Elektronen-Synchrotron

*Δεῖ τὰς αἰτίας τῶν φυσικῶν φαινομένων δια λογισμοῦ καὶ ἐμπειρίας ζητεῖν,
οὐκ ἐκ φόβου ἢ δεισιδαιμονίας.*

*One must seek the causes of natural phenomena through reasoning and experience,
not through fear or superstition.*

Epicurus — 300 B.C.

Abstract

This thesis presents a study of time-dependent CP violation in radiative B meson decays, focusing on the channel $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$. The analysis uses 365 fb^{-1} of data collected with the Belle II detector at the SuperKEKB asymmetric-energy e^+e^- collider. Such decays provide a sensitive probe of the photon polarisation in $b \rightarrow s\gamma$ transitions, and therefore of possible contributions from physics beyond the Standard Model.

To enable the measurement, key experimental inputs are determined using large control samples. The new Belle II Flavor Tagging algorithm and the decay time resolution function are calibrated *via* a data-driven approach. These results are directly applied in the time-dependent analysis.

The mixing-induced CP asymmetry parameter $S_{K_S^0 \pi^+ \pi^- \gamma}$ is then measured as an *effective* observable, reflecting the admixture of CP eigenstate and non-eigenstate intermediate contributions. In addition, the novel observables S^+ and S^- are measured for the first time, which are defined using complementary regions of the $K_S^0 \pi^+ \pi^- \gamma$ Dalitz plane. These quantities provide access to the real and imaginary parts of $S_{K_S^0 \pi^+ \pi^- \gamma}$ and, together with future amplitude analyses of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$, will help place new constraints on the ratio of Wilson coefficients C'_7/C_7 of the effective theory including possible new physics beyond the Standard Model.

Furthermore, this $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decay channel is utilised as a benchmark to evaluate the physics performance of various designs for the proposed upgrade to the Belle II vertex detector. The improved decay time resolution achievable with such an upgrade is shown to significantly enhance the sensitivity of time-dependent CP asymmetry measurements, underlining the strong physics case for such detector developments.

Résumé

Cette thèse présente une étude de la violation de CP dépendante du temps dans les désintégrations radiatives des mésons B , en se concentrant sur le canal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$. L'analyse utilise 365 fb^{-1} de données collectées avec le détecteur Belle II auprès du collisionneur de faisceaux e^+e^- asymétriques en énergie Super-KEKB. De telles désintégrations constituent une sonde sensible à la polarisation du photon dans les transitions $b \rightarrow s\gamma$, et donc à de possibles contributions de physique au-delà du Modèle Standard.

Pour effectuer cette mesure, des paramètres expérimentaux essentiels sont déterminés à l'aide de grands échantillons de contrôle. Le nouvel algorithme de Flavor Tagging de Belle II ainsi que la fonction de résolution sur le temps de désintégration sont calibrés *via* une approche fondée sur les données. Ces résultats sont appliqués directement dans l'analyse dépendante du temps.

Le paramètre $S_{K_S^0 \pi^+ \pi^- \gamma}$ d'asymétrie de CP induite par le mélange est ensuite mesuré comme une observable *effective*, du fait de contributions intermédiaires qui ne sont pas états propres de CP . De plus, les nouvelles observables S^+ et S^- sont mesurées pour la première fois, définies à partir de deux régions complémentaires du plan de Dalitz $K_S^0 \pi^+ \pi^- \gamma$. Ces grandeurs donnent accès aux parties réelle et imaginaire de $S_{K_S^0 \pi^+ \pi^- \gamma}$ et, combinées à de futures analyses en amplitudes de $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$, permettront de poser de nouvelles contraintes sur les coefficients de Wilson C_7'/C_7 de la théorie effective incluant de possibles effets de nouvelle physique au-delà du Modèle Standard.

En outre, ce canal de désintégration $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ est utilisé comme référence pour évaluer les performances physiques de différentes améliorations proposées pour le détecteur de vertex de Belle II. L'amélioration de la résolution sur le temps de désintégration résultant d'une telle mise à niveau augmente significativement la sensibilité des mesures de l'asymétrie de CP dépendante du temps, ce qui constitue une forte motivation physique pour de tels développements instrumentaux.

Table of Contents

Abstract	i
Résumé	ii
Introduction	1
1 Phenomenological context	3
1.1 The Standard Model of particle physics	3
1.2 Weak interactions between quarks	4
1.3 Quark mixing and the CKM matrix	6
1.4 Mixing and CP violation in the B mesons	9
1.4.1 Mixing of neutral B mesons	9
1.4.2 CP violation in the B mesons	12
1.5 Tree and Penguin transitions	17
1.6 $b \rightarrow s\gamma$ transitions as probes for BSM physics	18
1.6.1 The photon polarisation in $b \rightarrow s\gamma$ transitions	19
1.6.2 Physics observables and current status	22
2 The Belle II experiment	29
2.1 The SuperKEKB accelerator	30
2.2 The Belle II detector	33
2.2.1 The Pixel Detector	36
2.2.2 The Silicon Vertex Detector	37
2.2.3 The Central Drift Chamber	39
2.2.4 Particle Identification (TOP and ARICH)	41
2.2.5 The Electromagnetic Calorimeter	43
2.2.6 The K_L^0 and Muon Detector	44
2.3 Trigger system	45
2.3.1 Level 1 Trigger	46
2.3.2 High Level Trigger	48
2.4 The Belle II Analysis Software Framework	48

TABLE OF CONTENTS

2.5	Simulation	48
2.6	Reconstruction	49
2.6.1	Tracking	49
2.6.2	Charged Particle identification	51
2.6.3	K_S^0 reconstruction	51
2.6.4	Photon reconstruction	51
2.6.5	B meson reconstruction	52
2.6.6	Vertex reconstruction	53
3	VTX Performance Study	55
3.1	The case for the upgrade	55
3.2	The upgraded vertex detector – VTX	58
3.3	VTX performance study	60
3.3.1	Reconstruction efficiency	62
3.3.2	Vertex resolution	64
3.3.3	Sensitivity study	66
3.4	Conclusion and outlooks	66
4	Calibration of the Flavor Tagger and decay time resolution parameters	69
4.1	Flavor Tagging at Belle II	69
4.2	Flavor Tagger and Resolution Function calibration	73
4.2.1	Principle behind the calibration of the GNN-based Flavor Tagger	73
4.2.2	Reconstruction and selection of $B^0 \rightarrow D^{(*)-}\pi^+$ decays and selection criteria	77
4.2.3	Estimation of signal and background yields	79
4.2.4	Interlude: Fit of the Δt residual distribution of simulated signal events	89
4.2.5	Production of the background-subtracted Δt distribution with sWeights	92
4.2.6	Fit of the signal Δt distribution	94
4.3	Conclusion and outlooks	108
5	Measurement of the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0\pi^+\pi^-\gamma$ decays	111
5.1	Overview of the measurement	112
5.2	Signal reconstruction and event selection	114
5.2.1	Signal-side B decay reconstruction	114
5.2.2	Vertex and decay time reconstruction	116
5.2.3	Continuum background suppression	118

TABLE OF CONTENTS

5.2.4	Single candidate selection	120
5.2.5	Selection efficiencies	122
5.2.6	Control mode datasets and selection strategy	124
5.3	Measurement strategy	125
5.3.1	3D fit model	125
5.3.2	Signal modeling	127
5.3.3	Background modeling	135
5.3.4	3D fit to a simulated sample	142
5.4	Validation of the fit strategy	154
5.4.1	Validation using the signal mode	154
5.4.2	Validation using the control mode	193
5.5	Systematic uncertainties	219
5.5.1	Component fit shape parameters	219
5.5.2	Flavor Tagger parameters	219
5.5.3	Δt resolution function parameters	220
5.5.4	B^0 lifetime and $B^0-\bar{B}^0$ oscillation frequency	221
5.5.5	Fixed SCF background fractions	222
5.5.6	CP violation in the $B\bar{B}$ background	223
5.5.7	Effect of biased yields on CP parameters	224
5.5.8	Bootstrapping bias	224
5.5.9	Misalignment	225
5.5.10	Tag-side interference	226
5.5.11	Summary of systematic uncertainties	228
5.6	Results	229
Conclusion and prospects		241
A Phenomenological context		243
A.1	TDCP asymmetry formula derivation	243
A.2	Derivation of exponential decay part of the Δt distribution	246
B Auxiliary PDF distributions		247
B.1	Gaussian function	247
B.2	Bifurcated Gaussian function	247
B.3	Exponential function	247
B.4	Chebyshev polynomials	248
B.5	Double-sided Crystal Ball function	248
B.6	Crystal Ball function	248
B.7	Argus function	249
B.8	Maximum Likelihood Fit	249

TABLE OF CONTENTS

C	Flavor Tagger and decay time resolution calibration	251
C.1	Δt fit on simulated data	251
C.2	Correlation matrix from fit to data	259
D	Time-dependent CP asymmetry measurement	261
D.1	Signal-side B vertex reconstruction	261
D.2	BDT input variable distributions for signal and background events .	262
D.3	Correlations of fit variables	264
D.4	CP fit of SCF and $B\bar{B}$ backgrounds	268
D.5	Dalitz plane plots	270
D.6	Dalitz plane shapes	272
D.7	Ensemble test figures	277
D.8	Component shape comparison between signal and control mode . .	284
D.9	Control mode shape fits	288
D.10	Comparison between toy and generic MC distributions	292
	List of Figures	325
	List of Tables	344

Introduction

In the latter half of the 20th century, a theoretical framework known as the Standard Model (SM) of particle physics was developed in order to explain a number of experimental observations [1–9]. Since its inception, the SM has been extensively tested and has proven to be highly successful at predicting a wide range of experimental results with extraordinary precision [10–12]. This success culminated in the discovery of the Higgs boson by the ATLAS and CMS experiments at CERN in 2012 [13, 14].

Despite its success, the SM has certain notable shortcomings; notably it does not incorporate a description of gravity, and it fails to account for the observed asymmetry between matter and anti-matter in the Universe. Furthermore, it offers no explanation for the origin of the so-called Dark Matter, a form of matter that only interacts gravitationally and whose existence is inferred from various astronomical and cosmological observations [15–18]. Additionally, there exist discrepancies between certain experimental results and the SM predictions [19–21], but no robust conclusions can be made about them.

Before the advent of the B factories, Belle at KEK in Japan and $BABAR$ at SLAC in California, CP violation in the quark sector had not been quantified. The violation of CP symmetry was first observed in the weak decays of neutral kaons in 1964 with the Fitch-Cronin experiment [22]. In the same decade, Andrei Sakharov proposed that the existence of CP violating decay processes is one of the necessary conditions to explain the matter-antimatter asymmetry present in the current state of our Universe [23]. However, the mechanism that allows for CP violation in the SM is found to not be sufficient to explain this asymmetry, which implies the existence of CP violating New Physics (NP) processes.

The absence of *direct* NP signals at high-energy colliders suggests that, if present, NP effects should appear beyond the electroweak scale. Processes such as penguin transitions involving contributions from virtual particles provide an avenue for *indirect* searches at low-energy collider experiments like the B factories. Among them, $b \rightarrow s$ quark transitions are especially sensitive, as they proceed through Flavor Changing Neutral Currents (FCNC), forbidden at tree level in the SM and only possible *via* higher-order loop diagrams. NP contributions could

alter these loops, affecting measurable observables. One such observable is the polarisation of the final state photon in $b \rightarrow s\gamma$ transitions, which is predominantly left-handed in the SM, but can be affected by NP contributions [24, 25]. The photon polarisation can be determined either directly through measurements of angular observables or probed indirectly by measuring the mixing-induced CP asymmetry in radiative B decays proceeding through this transition [26].

In the latter case, the method relies on two considerations: (i) in decays of the form $B \rightarrow f_{CP}\gamma$, where f_{CP} is a CP eigenstate such as $K_S^0\rho^0$, the photon polarisation is correlated with the flavor of the decaying B meson at first order; and (ii) states with different CP quantum numbers do not interfere and therefore do not contribute to the asymmetry. Within the SM, the mixing-induced CP asymmetry is expected to be of order $\mathcal{O}(|m_s/m_b|^2)$, so any observed significant deviation from this prediction would provide a strong indication of NP effects.

In this thesis, the mixing-induced CP asymmetry is determined through a measurement of the time-dependent CP asymmetry in $B^0 \rightarrow K_{res}\gamma \rightarrow K_S^0\pi^+\pi^-\gamma$ decays in data recorded by the Belle II experiment. Here, K_{res} denotes a kaon resonance such as $K_1(1270)$, $K_2^*(1430)$, or $K^*(1680)$. Since not all intermediate states correspond to a CP eigenstate, a full amplitude analysis will be further needed to disentangle the various resonant and non-resonant contributions in order to measure the true CP asymmetry. This is in contrast to decays like $B^0 \rightarrow K_S^0\pi^0\gamma$, where the dominant contribution proceeds through a well-defined CP eigenstate. In the present work, such an analysis is not performed; rather, an *effective* CP asymmetry is measured, reflecting the combined contributions of all accessible intermediate states. In addition, two new observables which can be used to set new constraints on the parameters of NP models, as proposed in Ref. [27], are measured for the first time, establishing a benchmark for future Beyond Standard Model (BSM) searches in processes involving $b \rightarrow s\gamma$ transitions.

This document is organised as follows:

- Chapter 1 outlines SM concepts relevant to the main measurement and introduces the phenomenological motivation.
- Chapter 2 describes the experimental setup used in this work, namely the Belle II detector and the SuperKEKB accelerator.
- Chapter 3 presents a study evaluating the physics performance of a proposed upgrade to the Belle II Vertex Detector.
- Chapter 4 details the calibration of the Flavor Tagger parameters and the decay time resolution model, both of which are used in the time-dependent measurement.
- Chapter 5 presents the measurement of the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0\pi^+\pi^-\gamma$ decays using Belle II data.

Chapter 1

Phenomenological context

Contents

1.1	The Standard Model of particle physics	3
1.2	Weak interactions between quarks	4
1.3	Quark mixing and the CKM matrix	6
1.4	Mixing and CP violation in the B mesons	9
1.4.1	Mixing of neutral B mesons	9
1.4.2	CP violation in the B mesons	12
1.5	Tree and Penguin transitions	17
1.6	$b \rightarrow s\gamma$ transitions as probes for BSM physics	18
1.6.1	The photon polarisation in $b \rightarrow s\gamma$ transitions	19
1.6.2	Physics observables and current status	22

In this chapter, a brief introduction to the fundamental concepts of the Standard Model (SM) is provided, with a focus on the weak interactions between quarks and CP violation in the beauty quark sector. This is followed by an overview of the phenomenology behind $b \rightarrow s\gamma$ quark transitions, highlighting their relevance to the time-dependent CP asymmetry measurement at the core of this thesis research.

1.1 The Standard Model of particle physics

The Standard Model of particle physics is a quantum field theory that describes the interactions between particles with half-integer spin, known as *fermions*, mediated by the exchange of integer-spin particles, known as *bosons*. It is a gauge theory based on the gauge group $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$, which is a combination of the

strong interaction symmetry group $SU(3)_c$ and the electroweak interaction group $SU(2)_L \otimes U(1)_Y$.

The SM contains three generations of fermions, each consisting of a pair of quarks and a pair of leptons. Quarks carry color charge and therefore interact *via* the strong force, whereas leptons do not. All fermions interact *via* the electroweak force, mediated by the photon and the $W^\pm$ and Z bosons. The masses of fermions and the weak bosons are generated *via* the Higgs mechanism, in which electroweak symmetry $SU(2)_L \otimes U(1)_Y$ is spontaneously broken to $U(1)_{em}$.

Despite its success in describing a vast range of phenomena, the SM leaves several important questions unanswered, such as the origin of neutrino masses, the nature of dark matter, the absence of gravity in its framework, and the observed dominance of matter over antimatter in the Universe. These open problems motivate precision studies of the weak interactions and CP violation in the quark sector.

In the framework of the SM, CP violation arises from an irreducible phase in the Lagrangian that describes the weak interactions between electrically charged quarks. This is a direct consequence of the misalignment between the quark flavor eigenstates, which participate in weak interactions, and the quark mass eigenstates, which determine their propagation, as well as a result of the unitary transformation that relates the two.

1.2 Weak interactions between quarks

Among the three fundamental interactions described in the SM, the weak interaction is the only one that violates parity conservation, a fact first established in studies of radioactive β decay [28]. This is encoded in the SM by the $V - A$ structure of the weak interaction, which ensures it can couple only to fermions with *left-handed helicity* and anti-fermions with *right-handed helicity*¹.

The mediators of the weak interaction are the massive $W^\pm$ and Z gauge bosons, which obtain their masses through the Higgs mechanism [5, 6]. This is a consequence of spontaneous symmetry breaking (SSB) of the electroweak gauge group [7]:

$$SU(2)_L \times U(1)_Y \xrightarrow{SSB} U(1)_{em}$$

The quark fields transform as spinors under $SU(2)_L$ and are organised into either doublets or singlets, depending on their helicity. Left-handed quarks are organised

¹A particle's helicity is defined as the projection of its spin onto its direction of motion. For highly relativistic particles, helicity coincides with chirality, which is the property actually distinguished by the weak interaction.

into doublets, such as:

$$Q_L^i = \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix}, \quad (1.1)$$

and right-handed quarks are put into singlets, such as:

$$(u_R^i), (d_R^i), \quad (1.2)$$

where $i = 1, 2, 3$ is the index of the quark family or generation, u and d denote up- and down-type quarks, respectively, and the L or R subscript is used to denote the helicity of the quark spinor, where:

$$\psi_L = P_L \psi, \quad \psi_R = P_R \psi, \quad (1.3)$$

with ψ being the quark spinor. P_L and P_R are known as the projection operators and are defined as:

$$P_L = \frac{1}{2}(1 - \gamma^5), \quad P_R = \frac{1}{2}(1 + \gamma^5), \quad (1.4)$$

where γ^5 is one of the Dirac γ matrices.

In addition to providing masses to the weak gauge bosons, the Higgs field ϕ is also responsible for generating quark masses through Yukawa interactions. The Yukawa Lagrangian for quarks can be written as:

$$\mathcal{L}_{\text{Yukawa}} = -Y_u^{ij} \bar{Q}_L^i \tilde{\phi} u_R^j - Y_d^{ij} \bar{Q}_L^i \phi d_R^j + \text{h.c.}, \quad (1.5)$$

where Y_u and Y_d are 3×3 matrices containing the Yukawa couplings, the indices i and j label the generations, ϕ is the Higgs $SU(2)$ doublet, and $\tilde{\phi} = i\sigma^2 \phi^*$ is its conjugate. After electroweak symmetry breaking, the Higgs acquires a vacuum expectation value $\langle \phi \rangle = v/\sqrt{2}$, giving rise to fermion mass terms $m_q \bar{q}q$ proportional to the Yukawa couplings $Y_{u,d}$.

A distinctive feature of the weak interaction is that so-called *charged current* processes, mediated by the $W^\pm$ bosons, change quark flavor. Taking as an example the $u \rightarrow W^+ d$ quark-level transition, the interaction Lagrangian between the quark fields and the $W^\pm$ bosons can be written as:

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}} \bar{u}_L \gamma^\mu d_L W_\mu^+ + \text{h.c.}, \quad (1.6)$$

where g is the weak coupling constant and W_μ^+ is the weak gauge field.

After electroweak symmetry breaking, however, the quark fields that diagonalize the Yukawa mass matrices (i.e. the mass eigenstates) are not the same as the weak eigenstates appearing above. For the up- and down-type quarks, this misalignment can be expressed through unitary rotations:

$$u_L \rightarrow u'_L = U_u^\dagger u_L, \quad d_L \rightarrow d'_L = U_d^\dagger d_L, \quad (1.7)$$

where U_u and U_d are unitary matrices acting in flavor space.

Importantly, the weak neutral currents, mediated by the Z boson, remain unchanged under such transformations, which implies that so-called Flavor Changing Neutral Current (FCNC) processes do not occur at tree-level in the SM (see Section 1.5). This was predicted in 1970 by the Glashow–Iliopoulos–Maiani (GIM) mechanism [9], which explained the observed suppression of FCNC processes.

On the other hand, the charged current interaction is modified by these rotations and takes the form:

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}} \bar{u}'_L \gamma^\mu V_{CKM} d'_L W_\mu^+ + \text{h.c.} , \quad (1.8)$$

where:

$$V_{CKM} \equiv U_u^\dagger U_d \quad (1.9)$$

is the Cabibbo–Kobayashi–Maskawa (CKM) matrix, which describes the couplings between quarks of different flavors. As a result, the SM predicts the existence of flavor changing charged currents at tree-level. This matrix encodes all quark-flavor-changing transitions in the SM.

1.3 Quark mixing and the CKM matrix

The description of quark mixing in weak interactions was first introduced by Nicola Cabibbo in 1963 [29], when he proposed that the weak interaction eigenstates of the down-type quarks are linear combinations of the mass eigenstates. This idea was extended by Makoto Kobayashi and Toshihide Maskawa in 1973 [30] to three quark generations, leading to the formulation of the CKM mechanism. When considering three generations, an irreducible complex phase naturally arises in the CKM matrix and provides the only source of CP violation in the SM.

The CKM matrix, denoted V_{CKM} , can be parameterised in various ways. Since it is a unitary 3×3 matrix, four independent parameters are sufficient to describe it: three rotation angles and one irreducible complex phase. A widely used choice is the so-called Standard parameterisation, proposed in 1984 by L. L. Chau and W. Y. Keung [31]. It is defined as the product of three rotation matrices and a single complex phase:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} , \quad (1.10)$$

where $c_{ij} = \cos \theta_{ij}$, $s_{ij} = \sin \theta_{ij}$ for $i < j = 1, 2, 3$ are the rotation angles between the quark generations, and δ is the irreducible complex phase.

The hierarchical nature of the CKM matrix elements is now quantified thanks to many experimental results, which have confirmed that the matrix is almost diagonal. This means that interactions between quarks from different generations are suppressed, compared to ones between quarks of the same flavor doublet. To make this hierarchy more transparent, Lincoln Wolfenstein proposed an approximate parameterisation of the CKM matrix in 1983 [32], expressed as an expansion in powers of a small parameter λ :

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & \lambda^3 A(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & \lambda^2 A \\ \lambda^3 A(1 - \rho - i\eta) & -\lambda^2 A & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \quad (1.11)$$

Here A , ρ , and η are parameters of order 1, while η encodes the irreducible complex phase responsible for CP violation in the quark sector. The Wolfenstein parameterisation is useful in that it allows one to visualise the near-diagonal nature of the CKM matrix.

The unitarity of the CKM matrix $V_{\text{CKM}}^\dagger V_{\text{CKM}} = \mathbb{1}$ leads to 9 unitarity relations, of which 3 are of interest for the study of CP violation. They are:

$$V_{ud}V_{us}^* + V_{cd}V_{cs}^* + V_{td}V_{ts}^* = 0 \quad (1.12)$$

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0 \quad (1.13)$$

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0 \quad (1.14)$$

Each of the above equations represents a triangle in the Argand complex plane, called a *unitarity triangle*, where each of the terms can be thought of as a vector in that plane. The first two unitarity triangles are related to CKM matrix elements that govern the neutral $K^0-\bar{K}^0$ and $B_s^0-\bar{B}_s^0$ systems, respectively. In those cases, one of the sides of the unitarity triangle is much shorter than the other two, as can be seen from the power of λ , leading to a triangle which is almost flat. This is why in the leading K and B_s decays CP violation effects are as small as they are. On the other hand, the sides of the unitarity triangle corresponding to Equation 1.14 are all of the same order of magnitude. This leads to relatively large CP asymmetries observed in the $B_d^0-\bar{B}_d^0$ system, making it an ideal laboratory to study CP violation. Another way to write Equation 1.14 is by normalising one of its sides to have unit length. This leads to:

$$\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} + 1 + \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} = 0 \quad (1.15)$$

In this representation, displayed in Figure 1.1, one of the sides is aligned to the real axis and the apex of the triangle can be related to the irreducible phase of the CKM matrix, through the $\bar{\rho}$ and $\bar{\eta}$ parameters. In this representation, the sides of

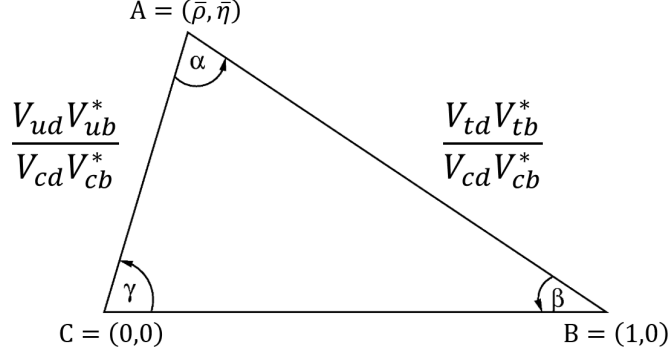


Figure 1.1: Unitarity triangle corresponding to the unitarity relation $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$ of the CKM matrix.

the unitarity triangle are given by:

$$\overline{CB} = 1, \quad (1.16)$$

$$\overline{AB} = \left| \frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{(1 - \bar{\rho})^2 + \bar{\eta}^2} = \frac{1}{\lambda} \left| \frac{V_{td}}{V_{cb}} \right|, \quad (1.17)$$

$$\overline{AC} = \left| \frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right| = \sqrt{\bar{\rho}^2 + \bar{\eta}^2} = \left(1 - \frac{\lambda^2}{2}\right) \frac{1}{\lambda} \left| \frac{V_{ub}}{V_{cb}} \right|, \quad (1.18)$$

where: $\bar{\rho} = \rho(1 - \lambda^2/2)$ and $\bar{\eta} = \eta(1 - \lambda^2/2)$. The angles² are given by:

$$\alpha = \arg \left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*} \right) = \pi - \arctan \left(\frac{\bar{\eta}}{1 - \bar{\rho}} \right) - \arctan \left(\frac{\bar{\eta}}{\bar{\rho}} \right), \quad (1.19)$$

$$\beta = \arg \left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*} \right) = \arctan \left(\frac{\bar{\eta}}{1 - \bar{\rho}} \right), \quad (1.20)$$

$$\gamma = \arg \left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*} \right) = \arctan \left(\frac{\bar{\eta}}{\bar{\rho}} \right). \quad (1.21)$$

An effective approach to test the SM is to overconstrain the parameters of the unitarity triangle by performing as many independent measurements of its sides and angles as possible. So far, the experimental evidence suggests that, within current uncertainties, the SM accurately describes CP violation in the quark sector (see Figure 1.2).

²Here, the angles are denoted as: α, β, γ , but sometimes an alternate notation may be used: ϕ_1, ϕ_2, ϕ_3 , where: $\alpha = \phi_2, \beta = \phi_1, \gamma = \phi_3$.

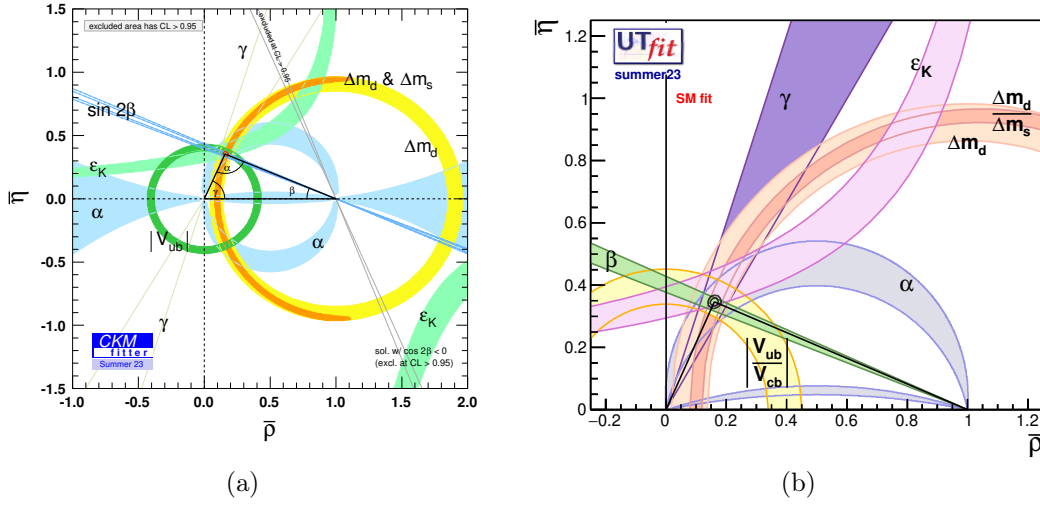


Figure 1.2: Experimental constraints on the sides and angles of the unitarity triangle, (a) by the CKMfitter group [33] and (b) by the UTfit group [34], using the results available in summer 2023.

1.4 Mixing and CP violation in the B mesons

This section describes the quantum mechanical properties of the neutral B meson system and how they relate to CP violation observables. In this system, three distinct types of CP violation can occur, each arising from a different underlying process and allowing the measurement of a different CP violating quantity.

1.4.1 Mixing of neutral B mesons

Neutral mesons and their CP conjugate states are unstable and can decay weakly. Before describing the neutral meson system and the concept of mixing, let us first define the two representations that will be used. First, in order to describe how the neutral meson is produced and how it decays, the flavor basis is used, which consists of two flavor eigenstates, each with a definite quark content (e.g. $B_d^0 = d\bar{b}$ and $\bar{B}_d^0 = b\bar{d}$ in the case of B_d mesons). Second, in order to describe how the neutral mesons propagate in space-time, the mass basis is used, consisting of two mass eigenstates, each with a definite mass and lifetime (e.g. K_S^0 and K_L^0 in the case of kaons). The mass eigenstates are also the eigenstates of the Hamiltonian. Since flavor and mass eigenstates are different from each other, it is possible for flavor eigenstates to get mixed as they propagate. Moreover, because CP symmetry

is violated in the SM, there is an additional misalignment between mass and CP eigenstates.

Neutral meson mixing, which refers to the phenomenon where a neutral meson turns into its antiparticle *via* a two-charged current process, has been observed in kaons [35], in D mesons [36–38], and in B_d and B_s mesons [39–41].

An arbitrary combination of neutral B meson flavor eigenstates has the form:

$$|B(t)\rangle = a(t) |B^0\rangle + b(t) |\bar{B}^0\rangle, \quad (1.22)$$

which obeys the coupled system of Schrödinger equations:

$$\mathcal{H} \begin{pmatrix} a(t) |B^0\rangle \\ b(t) |\bar{B}^0\rangle \end{pmatrix} \equiv (M - i\Gamma) \begin{pmatrix} a(t) |B^0\rangle \\ b(t) |\bar{B}^0\rangle \end{pmatrix} = i \frac{d}{dt} \begin{pmatrix} a(t) |B^0\rangle \\ b(t) |\bar{B}^0\rangle \end{pmatrix}, \quad (1.23)$$

where M and Γ are two 2×2 hermitian matrices known as mass and decay matrices, respectively. The diagonal terms of the mass and decay matrices represent the masses and decay rates of the B^0 and $\bar{B}^0$ flavor eigenstates. Due to the required CPT symmetry of the SM, the masses and decay rates of the flavor eigenstates must be equal, therefore $M_{11} = M_{22} = M$ and $\Gamma_{11} = \Gamma_{22} = \Gamma$. The off-diagonal terms of the decay matrix describe the *interference* between the decays of the B^0 and $\bar{B}^0$ components of the $B(t)$ state, while the off-diagonal terms of the mass matrix account for the *mixing* between the two flavor eigenstates. Figure 1.3 shows the leading-order Feynman diagrams representing the weak transitions between B^0 and $\bar{B}^0$, also known as B^0 – $\bar{B}^0$ mixing.

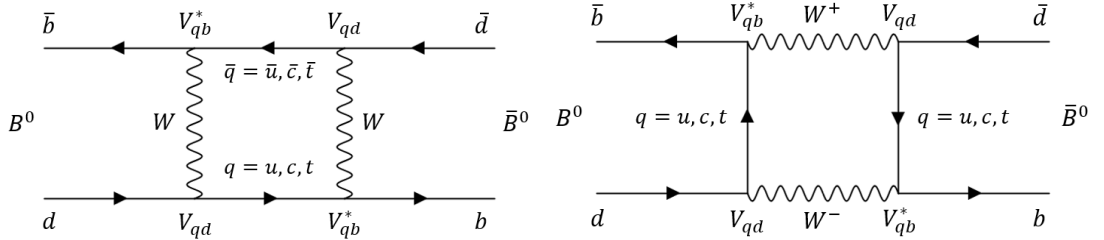


Figure 1.3: Feynman diagrams describing the mixing of B mesons *via* a second-order weak transition. These are also referred to as *box diagrams*.

The presence of the three up-type quarks in the loops of these diagrams should not be too surprising, as all of them couple with a CKM factor proportional to λ^6 . However, the top quark will have the most dominant contribution, since each term is further weighted by the ratio of the quark mass to the W boson mass. Consequently, the contributions from up- and charm-quarks are suppressed and an extra phase of the form $(V_{td}V_{tb}^*)^2 \sim e^{-2\beta}$ appears in the transition amplitude,

where β is one the unitarity triangle angles, defined in Equation 1.20. This means that B^0 mesons that oscillate into $\bar{B}^0$ acquire a relative phase of $\phi_M = 2\beta$ compared to those that do not, which is experimentally accessible in cases where both mesons can decay to a common final state.

Because of the presence of non-zero off-diagonal terms in $\mathcal{H}$, the flavor eigenstates are not the eigenstates of the Hamiltonian. Diagonalising the Hamiltonian of Equation 1.23 involves solving the eigenvalue equation:

$$\begin{pmatrix} M - \frac{i}{2}\Gamma & M_{12} - \frac{i}{2}\Gamma_{12} \\ M_{12}^* - \frac{i}{2}\Gamma_{12}^* & M - \frac{i}{2}\Gamma \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = \lambda \begin{pmatrix} p \\ q \end{pmatrix}, \quad (1.24)$$

where, due to the hermeticity of the mass and decay matrices, $M_{21} = M_{12}^*$ and $\Gamma_{21} = \Gamma_{12}^*$. This leads to the expressions for the mass eigenstates of the neutral B meson system, which are superpositions of the flavor eigenstates:

$$|B_L\rangle = p |B^0\rangle + q |\bar{B}^0\rangle \quad (1.25)$$

$$|B_H\rangle = p |B^0\rangle - q |\bar{B}^0\rangle, \quad (1.26)$$

where the index $L(H)$ denotes the lighter (heavier) mass eigenstate and the complex coefficients p and q must satisfy:

$$\frac{q}{p} = \pm \left(\frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}} \right)^{1/2} \quad \text{and} \quad |p|^2 + |q|^2 = 1. \quad (1.27)$$

The mass and width difference between the physical B_L and B_H states are defined as:

$$\Delta m_d \equiv M_H - M_L, \quad \Delta \Gamma = \Gamma_H - \Gamma_L, \quad (1.28)$$

so that Δm_d is positive by definition.

A unique characteristic of the neutral B mesons is that $\Delta \Gamma_{B_d} \ll \Gamma_{B_d}$, which has some noteworthy consequences with regards to CP violation (see Section 1.4.2). This is a direct result of the small fraction of B decay modes shared between B^0 and $\bar{B}^0$. This situation is in contrast to the neutral kaons, where $\Gamma_S \gg \Gamma_L$ and $\Delta \Gamma_K \sim \Gamma_{K_S}$, leading to the clear distinction between short-lived (K_S^0) and long-lived (K_L^0) states. For B_d mesons, however, the decay widths of B_L and B_H are nearly identical, and so are their lifetimes³. Consequently, the off-diagonal elements of the decay matrix are much smaller than those of the mass matrix, i.e. $|M_{12}| \gg |\Gamma_{12}|$, leading to $|q/p| \sim 1$, an approximation which follows from Equation 1.27.

³Throughout this document τ_{B_d} and τ_{B^0} refer to the same quantity and will be used interchangeably.

Let us now consider a state which starts as a pure B^0 ($\bar{B}^0$) state at time $t = 0$, which we will call $|B_{t=0}^0\rangle$ ($|\bar{B}_{t=0}^0\rangle$). Over time these states will develop an opposite flavor component according to the recipe:

$$|B_{t=0}^0(t)\rangle = g_+(t) |B^0\rangle + \frac{q}{p} g_-(t) |\bar{B}^0\rangle \quad (1.29)$$

$$|\bar{B}_{t=0}^0(t)\rangle = g_+(t) |\bar{B}^0\rangle + \frac{p}{q} g_-(t) |B^0\rangle , \quad (1.30)$$

where:

$$g_+(t) = e^{-\frac{\Gamma}{2}t} e^{-\frac{i}{2}(M_H+M_L)t} \cos\left(\frac{\Delta m_d t}{2}\right) \quad (1.31)$$

$$g_-(t) = e^{-\frac{\Gamma}{2}t} e^{-\frac{i}{2}(M_H+M_L)t} i \sin\left(\frac{\Delta m_d t}{2}\right) , \quad (1.32)$$

which are obtained by first transforming the pure B^0 and $\bar{B}^0$ states in the mass basis, using Equations 1.25 and 1.26, and by applying the Hamiltonian operator on them. Equations 1.29 and 1.30 are then obtained by transforming the states back to the flavor basis.

The mass difference⁴ and lifetime parameters of the neutral B mesons have been precisely measured, with world averages from current experiments [42] given by:

$$\Delta m_d = (0.5069 \pm 0.0019) \text{ ps}^{-1} \text{ and } \tau_{B_d} = (1.517 \pm 0.004) \text{ ps} , \quad (1.33)$$

respectively. Since the values of these parameters are of the same order of magnitude, the mixing amplitude is quite large. As a result, the integrated oscillation probability becomes sizeable:

$$\chi_d = \frac{(\tau_{B_d} \Delta m_d)^2}{2(1 + (\tau_{B_d} \Delta m_d)^2)} = 0.1860 \pm 0.0011 . \quad (1.34)$$

The prominence of mixing in the neutral B meson system can lead to large CP asymmetries, as will be discussed in the next Section.

1.4.2 CP violation in the B mesons

As previously stated, CP violation arises from a single irreducible phase in the SM couplings. There are three distinct ways that CP violation can manifest in the neutral B meson system:

⁴The mass difference Δm_d is commonly referred to as the *oscillation frequency* between B^0 and $\bar{B}^0$, which becomes clear when looking at Equations 1.31 and 1.32.

- **CP violation in decay**, which occurs when the amplitude for a decay has a different magnitude than that of its *CP* conjugate process;
- **CP violation in mixing**, which occurs when the neutral mass eigenstates are not identical to the *CP* eigenstates;
- **CP violation in the interference between mixing and decay**, which occurs when the B^0 and $\bar{B}^0$ mesons can decay into a common final state.

Before describing each of the three types of *CP* violation in more detail, let us first introduce some useful notation. The first is the *transition amplitude* for a $B(\bar{B})$ meson that decays into a final state $f(\bar{f})$, denoted by $A_f(\bar{A}_{\bar{f}})$, which is defined as:

$$A_f = \langle f | \mathcal{H} | B \rangle \quad , \quad \bar{A}_{\bar{f}} = \langle \bar{f} | \mathcal{H} | \bar{B} \rangle \quad . \quad (1.35)$$

The second notation is related to phases that may appear in A_f and $\bar{A}_{\bar{f}}$. Specifically, there are two kinds of phases; *CP-odd*, or *weak* phases, and *CP-even*, or *strong* phases. The former kind is related to the weak interaction part of the Lagrangian. Weak phases can lead to *CP* violation, since the relevant terms for a given process will become complex-conjugated for the *CP* conjugate process, breaking *CP* symmetry. The latter kind is related to the strong interaction part of the Lagrangian, which has *CP* invariant terms and thus cannot lead to *CP* violation.

CP violation in decay

The first kind of *CP* violation arises when there is a non-zero relative phase between the amplitude of a decay process and that of its *CP* conjugate process. The existence of *CP*-odd and *CP*-even phases in a transition amplitude of a process cannot by itself induce *CP* violation, since absolute phases are not measurable. However, if there are multiple amplitudes contributing to a given process, the situation is slightly more complicated.

Consider such a process that is made up of different contributions. The total amplitude is given by the sum of all of them:

$$A_f = \sum_j A_j e^{i\phi_j} e^{i\delta_j} \quad , \quad (1.36)$$

$$\bar{A}_{\bar{f}} = e^{2i\Delta\xi} \sum_j A_j e^{-i\phi_j} e^{i\delta_j} \quad , \quad (1.37)$$

where A_j is the modulus of the contributing amplitude, ϕ_j and δ_j are the *absolute CP-odd* and *CP-even* phases of that given amplitude, respectively, and $\Delta\xi$ is the

relative phase, which is convention-dependent. Taking the modulus of the ratio between the two total amplitudes removes this relative phase and we obtain:

$$\left| \frac{\bar{A}_f}{A_f} \right| = \left| \frac{\sum_j A_j e^{-i\phi_j} e^{i\delta_j}}{\sum_j A_j e^{i\phi_j} e^{i\delta_j}} \right|. \quad (1.38)$$

The only way that CP symmetry is broken in this scenario is if the weak phases ϕ_j are not all equal. Then, it becomes trivial to see that:

$$\left| \frac{\bar{A}_f}{A_f} \right| \neq 1 \implies CP \text{ violation}. \quad (1.39)$$

This type of CP violation is also known as *direct* CP violation and is a result of two or more amplitudes contributing to a process with different phases. The CP asymmetry is defined as:

$$\mathcal{A}_{CP} = \frac{\Gamma(\bar{B} \rightarrow \bar{f}) - \Gamma(B \rightarrow f)}{\Gamma(\bar{B} \rightarrow \bar{f}) + \Gamma(B \rightarrow f)}, \quad (1.40)$$

which is a definition that can be applied to either neutral or charged B mesons.

CP violation in mixing

The second kind of CP violation arises from the fact that neutral B mesons can oscillate into their opposite flavors, which is known as mixing (see Section 1.4.1). More specifically, it is a direct consequence of the misalignment between mass eigenstates and CP eigenstates, which results in an asymmetry between their transition probabilities, $\Gamma(B^0 \rightarrow \bar{B}^0)$ and $\Gamma(\bar{B}^0 \rightarrow B^0)$. These differences can be quantified using the mixing-related quantity q/p , which relates to the off-diagonal elements of the mass and decay matrices, as can be seen from Equation 1.27. Taking the square of the modulus of q/p , we obtain:

$$\left| \frac{q}{p} \right|^2 = \left| \frac{M_{12}^* - \frac{i}{2}\Gamma_{12}^*}{M_{12} - \frac{i}{2}\Gamma_{12}} \right|, \quad (1.41)$$

from which it becomes clear that, if CP was conserved, the mass and CP eigenstates would be identical, i.e. $M_{12} = M_{12}^*$ and $\Gamma_{12} = \Gamma_{12}^*$, and $|q/p| = 1$. In the opposite case, it becomes trivial to see that:

$$\left| \frac{q}{p} \right| \neq 1 \implies CP \text{ violation}. \quad (1.42)$$

This type of CP violation is also known as *indirect* CP violation and was observed for the first time in neutral kaons in 1964 [22]. In the neutral B_d system, however, effects of CP violation due to mixing are expected to be small, $\mathcal{O}(10^{-4})$, and $|q/p| \sim 1$ is a good approximation, with respect to the sensitivity of the past and current B factories.

CP violation in the interference between mixing and decay

The third kind of CP violation, commonly referred to as CP violation in the interference between mixing and decay, was proposed in the 1980s by I. Y. Bigi, A. I. Sanda, A. B. Carter, and others [43, 44]. This effect arises when:

$$\Gamma(B^0 \rightarrow f) \neq \Gamma(B^0 \rightarrow \bar{B}^0 \rightarrow f) , \quad (1.43)$$

which can only occur if the final state f is accessible by both B^0 and $\bar{B}^0$, a condition which is naturally satisfied when f is a CP eigenstate, f_{CP} ⁵. This kind of CP violation is very prominent in the case of neutral B mesons, due to the interference between decay amplitudes with mixing, $A(B^0 \rightarrow \bar{B}^0 \rightarrow f_{CP})$, and ones without, $A(B^0 \rightarrow f_{CP})$.

To measure this kind of CP violation, a similar quantity to that defined in Equation 1.40 can be used. In this case, however, the flavor of the neutral B mesons evolves with time, making its determination more complicated. The experimental setup at B factory experiments such as Belle II facilitates the flavor determination through the production of an $\Upsilon(4S)$ resonance, which decays into a pair of B mesons [45]. The $B\bar{B}$ pair constitutes a quantum-entangled system, meaning that one will have the opposite flavor of the other at any given time. Thus, by determining the flavor of one of the B mesons at time $t = t_1$, the flavor of the other B meson is also known at that same time. The experimental technique of determining the flavor of the B meson which decays into a final state of interest, B_{sig} , by using the decay of the other B meson in the event, B_{tag} , is known as “Flavor Tagging”. More details about how this is performed at Belle II are given in Section 4.1. The quantum entanglement of the $B\bar{B}$ pair is broken when one of the two B mesons decays, after which point the remaining B meson undergoes flavor oscillations until it also decays, at time $t = t_2$. In order to account for these oscillations correctly, precise determination of the time between the two B meson decays, $\Delta t = t_2 - t_1$, is also important.

Considering the above, the CP asymmetry defined in Equation 1.40 can be redefined in a time-dependent context as:

$$\mathcal{A}_{CP}(\Delta t) = \frac{\Gamma(B_{\text{tag}=B^0}(\Delta t) \rightarrow f_{CP}) - \Gamma(B_{\text{tag}=\bar{B}^0}(\Delta t) \rightarrow f_{CP})}{\Gamma(B_{\text{tag}=B^0}(\Delta t) \rightarrow f_{CP}) + \Gamma(B_{\text{tag}=\bar{B}^0}(\Delta t) \rightarrow f_{CP})} . \quad (1.44)$$

Using Equations 1.29–1.32, the time-dependent CP asymmetry can be expressed in terms of parameters that characterize CP violating effects (see Appendix A.1):

$$\mathcal{A}_{CP}(\Delta t) = S \sin(\Delta m_d \Delta t) - C \cos(\Delta m_d \Delta t) , \quad (1.45)$$

⁵It is not necessary, however, for f to be a CP eigenstate for this type of CP violation to occur.

with

$$S = \frac{2\Im(\lambda_{f_{CP}})}{1 + |\lambda_{f_{CP}}|^2}, \quad C = \frac{1 - |\lambda_{f_{CP}}|^2}{1 + |\lambda_{f_{CP}}|^2}. \quad (1.46)$$

The two parameters S and C quantify different sources of CP violation; C corresponds to direct CP violation (a difference in decay amplitudes), while S encodes mixing-induced CP violation, which arises from the interference between mixing and decay amplitudes. Both are expressed in terms of the complex parameter $\lambda_{f_{CP}}$, defined as:

$$\lambda_{f_{CP}} \equiv \eta_{CP} \frac{q}{p} \frac{\bar{A}_{\bar{f}}}{A_{f_{CP}}}, \quad (1.47)$$

where η_{CP} is the CP eigenvalue of the final state, f_{CP} . In this case, $\lambda_{f_{CP}}$ is the complex parameter which carries all the information related to CP violation and it is straightforward to see that:

$$\lambda_{f_{CP}} \neq 1 \implies CP \text{ violation}. \quad (1.48)$$

Note that this condition is satisfied if there is CP violation either in decay ($|\bar{A}_{\bar{f}}/A_f| \neq 1$) or in mixing ($|q/p| \neq 1$). However, it is fulfilled even when $|\bar{A}_{\bar{f}}/A_f| = |q/p| = 1$, since in that case $|\lambda_{f_{CP}}| = 1$ while $\Im(\lambda_{f_{CP}}) \neq 0$. From Equation 1.46, it then follows that $S \neq 0$, giving rise to a non-zero time-dependent CP asymmetry.

Lastly, let us also construct the expression for the time-dependent decay rate which will be used when describing the main measurement in Chapter 5:

$$\mathcal{P}_{CP}(\Delta t; q_{tag}) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \left\{ 1 + q_{tag} [S \sin(\Delta m_d \Delta t) - C \cos(\Delta m_d \Delta t)] \right\}, \quad (1.49)$$

where $q_{tag} = +1(-1)$ when the B_{tag} is determined to be $B^0(\bar{B}^0)$. It is worth noting that the above expression is symmetric under $\Delta t \rightarrow -\Delta t$, meaning that the ordering of the two B decays in the event has no impact on the measurement of the time-dependent CP asymmetry (see Appendix A.2).

1.5 Tree and Penguin transitions

In the SM, weak decays of B mesons can proceed through two main classes of Feynman diagrams: tree-level and loop-level processes. Both types can contribute to the same decay channel, although their relative importance depends on the specific final state under consideration.

Tree-level processes are characterized by the exchange of a virtual W boson between quark fields and are typically the dominant contribution in many B meson decays. A widely studied example is the so-called “golden mode” $B^0 \rightarrow J/\psi K_S^0$, which is dominated by a color-suppressed tree amplitude (see Figure 1.4, left). Tree-level processes are theoretically “cleaner” in the sense that their amplitudes can often be computed with smaller hadronic uncertainties compared to contributions from loop-level processes. This makes them suitable for precision tests of the SM, for instance in measurements of the mixing-induced CP violation introduced in the previous section.

On the other hand, so-called “penguin” diagrams [46] arise at loop-level, where the weak transition is mediated through virtual quarks and a W boson participating in a loop (see Figure 1.4, right). Their contribution is generally suppressed compared to tree amplitudes, both because they occur at higher order in the electroweak interaction and because they often involve additional small CKM factors. Nevertheless, penguin contributions can compete with tree diagrams in certain decay channels and can even dominate in specific cases.

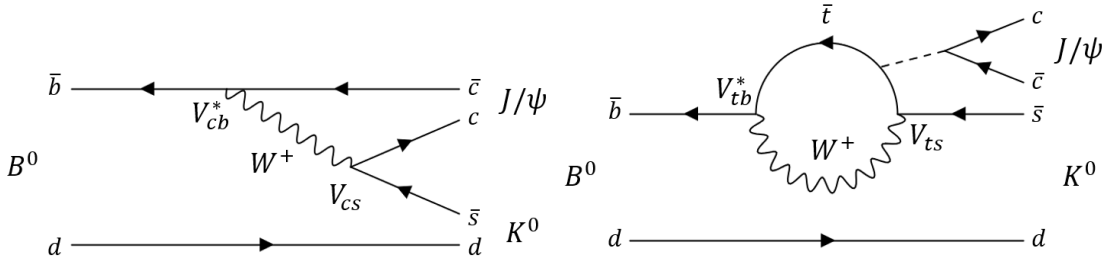


Figure 1.4: Main Feynman diagrams contributing to the $B^0 \rightarrow J/\psi K_S^0$ decay. The left diagram illustrates the tree-level contribution, while the right diagram shows the loop-level (penguin) contribution.

In particular, FCNC processes, such as $b \rightarrow s$ or $b \rightarrow d$ transitions, do not occur at tree-level in the SM, due to the GIM mechanism. Instead, they arise only through penguin loops or box diagrams, making their amplitudes loop-suppressed and CKM-suppressed. This suppression, however, turns out to be an advantage from the point of view of NP searches. Since these decays are predicted to be

rare in the SM, even small NP contributions can induce significant deviations in branching fractions, CP asymmetries, or other observables.

A crucial feature of penguin processes is that they are sensitive to the presence of heavy virtual particles, even if those particles are far above the electroweak scale. Hypothetical heavy particles beyond the SM, such as supersymmetric particles, heavy charged Higgs bosons, or vector-like quarks could enter the loop and alter the effective couplings governing b quark transitions. This makes penguin-dominated decays powerful *indirect* probes of NP effects, complementing *direct* searches at high-energy colliders.

In contrast, looking for NP effects in tree-dominated decays is more challenging. Since the tree amplitude is both theoretically cleaner and typically much larger than any possible penguin contribution, new virtual effects are heavily diluted and difficult to disentangle from the SM expectation. Because of this, tree-dominated decays like $B^0 \rightarrow J/\psi K_S^0$ are best suited for precision tests of CP violation within the SM, while penguin-dominated decays, especially those involving FCNC processes like $b \rightarrow s\gamma$ transitions, are more sensitive to possible contributions from BSM physics.

1.6 $b \rightarrow s\gamma$ transitions as probes for BSM physics

The SM predicts that the photon emitted in $b \rightarrow s\gamma$ ($\bar{b} \rightarrow \bar{s}\gamma$) transitions is predominantly left-handed(right-handed), due to the chiral structure of the weak interaction and the fact that the b quark is much heavier than the s quark (see Section 1.6.1). However, in many NP scenarios, heavy BSM particles can enter in the electroweak loop process with sizeable contributions to the right-handed component of the photon polarisation. Such effects would manifest as deviations from the SM expectation in observables such as branching ratios, CP asymmetries and the photon polarisation itself. To date, no significant deviation from the SM prediction has been observed in either of those observables, which places strong constraints on the parameters of NP models.

The branching ratio of the inclusive $B \rightarrow X_s\gamma$ process has been measured by several experiments with an average of $\mathcal{B}(B \rightarrow X_s\gamma)_{\text{exp}} = (3.49 \pm 0.19) \times 10^{-4}$ [42] and is in agreement with the SM prediction $\mathcal{B}(B \rightarrow X_s\gamma)_{\text{th}} = (3.40 \pm 0.17) \times 10^{-4}$ [47]. However, this by itself is not enough to fully exclude NP contributions to the photon polarisation in $b \rightarrow s\gamma$, as there are some NP models [24, 25] in which the left- and right-handed components are comparable, without leading to an excess in the inclusive branching fraction. Putting constraints on the parameters of those models requires studying the photon polarisation itself in exclusive radiative decay modes.

1.6.1 The photon polarisation in $b \rightarrow s\gamma$ transitions

In the SM, the effective Hamiltonian for the quark-level $b \rightarrow s\gamma$ transition, at leading order in QCD, is given by:

$$\mathcal{H}_{\text{eff}} = -\frac{G_F}{\sqrt{2}} \frac{e}{16\pi^2} V_{tb} V_{ts}^* F_2 \bar{s} \sigma_{\mu\nu} (m_b P_R + m_s P_L) b F^{\mu\nu}, \quad (1.50)$$

where G_F is the Fermi constant, which is related to the weak coupling and the mass of the W boson as $G_F \propto g^2/M_W^2$ at first order of the radiative corrections, V_{tb} and V_{ts} are the leading CKM matrix elements appearing in the loop diagram, F_2 is the Inami-Lim function which accounts for the dominant top quark contribution in the loop [48], P_R and P_L are the projection operators defined in Equation 1.4, $\sigma_{\mu\nu}$ are the commutators of the Dirac γ matrices and $F^{\mu\nu}$ is the electromagnetic field strength tensor.

Because of the dipole structure of the operator in Equation 1.50, only two helicity configurations contribute to the $b \rightarrow s\gamma$ process. The $\sigma_{\mu\nu}$ tensor only couples quarks of opposite helicities, and the mass factors m_b and m_s determine on which quark leg the helicity flip occurs; the m_b term corresponds to $b_R \rightarrow s_L$, while the m_s term corresponds to $b_L \rightarrow s_R$. Angular momentum conservation then fixes the photon helicity, so that the two possible contributions are $b_R \rightarrow s_L \gamma_L$ and $b_L \rightarrow s_R \gamma_R$. Moreover, because the mass of the s quark is much smaller than that of the b quark ($m_s/m_b \sim 0.02$), the term proportional to m_s is suppressed with respect to the one proportional to m_b and that is why in the SM the photon polarisation⁶ is predominantly *left-handed* in $b \rightarrow s\gamma$ transitions, i.e. $P(b_R \rightarrow s_L \gamma_L) \gg P(b_L \rightarrow s_R \gamma_R)$. The Feynman diagram of the quark-level $b \rightarrow s\gamma$ transition is shown in Figure 1.5.

The expression for the effective Hamiltonian corresponding to the $\bar{b} \rightarrow \bar{s}\gamma$ process is obtained by taking the hermitian conjugate of Equation 1.50:

$$\mathcal{H}_{\text{eff}}^\dagger = -\frac{G_F}{\sqrt{2}} \frac{e}{16\pi^2} V_{tb}^* V_{ts} F_2^* \bar{s} \sigma_{\mu\nu} (m_b P_L + m_s P_R) b F^{\mu\nu}. \quad (1.51)$$

Following the same arguments as before, it becomes clear that in $\bar{b} \rightarrow \bar{s}\gamma$ transitions the photon polarisation is predominantly *right-handed*, i.e. $P(\bar{b}_L \rightarrow \bar{s}_R \gamma_R) \gg P(\bar{b}_R \rightarrow \bar{s}_L \gamma_L)$.

In the context of radiative B meson decays proceeding through $b \rightarrow s\gamma$ transitions, such as $B \rightarrow K_{\text{res}} \gamma \rightarrow K \pi \pi \gamma$, it is necessary to account for QCD corrections that modify the decay amplitudes for left- and right-handed photons. The effective Hamiltonian shown in Equation 1.50 captures only the leading contribution, but

⁶The terms photon helicity and photon polarisation are used interchangeably throughout this chapter.

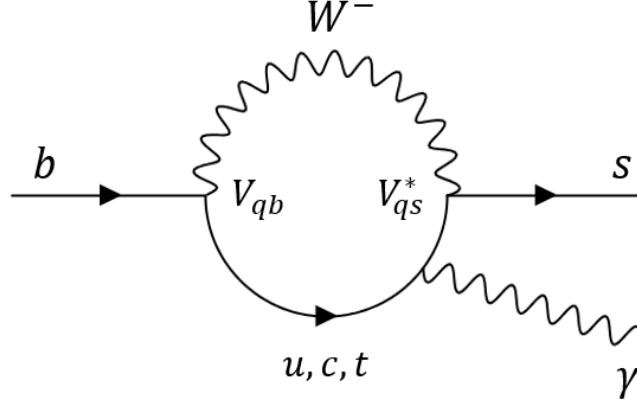


Figure 1.5: Feynman diagram of the $b \rightarrow s\gamma$ quark-level transition in the SM. Since the top quark dominates in the loop, the weak couplings can be taken as $|V_{tb}| \sim 1$ and $|V_{ts}^*| \sim \lambda^2$. The photon helicity is predominantly left-handed in this transition, while in the $\bar{b} \rightarrow \bar{s}\gamma$ transition, it is right-handed.

the full description is obtained within the Operator Product Expansion (OPE). In this framework, the effective Hamiltonian can be written as:

$$\mathcal{H}_{\text{eff}}^{\text{OPE}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_{i=1}^8 \left[C_i(\mu) \mathcal{O}_i(\mu) + C'_i(\mu) \mathcal{O}'_i(\mu) \right], \quad (1.52)$$

where $\mathcal{O}_i$ are local long-distance operators and C_i are short-distance Wilson coefficients related to $b \rightarrow s\gamma_L$ transitions. The primed operators and coefficients are related to the $b \rightarrow s\gamma_R$ transitions. As described earlier, these right-handed terms are suppressed in the SM by a factor m_s/m_b with respect to the left-handed terms.

An intuitive way to understand the structure of $\mathcal{H}_{\text{eff}}^{\text{OPE}}$ is to view it as replacing the underlying short-distance physics with effective point-like interactions, analogous to how the Fermi theory of weak interactions introduces the effective coupling G_F . The local operators $\mathcal{O}_i$ encode these interactions, while the Wilson coefficients C_i act as the corresponding effective couplings. In this picture, $\mathcal{H}_{\text{eff}}^{\text{OPE}}$ is expressed as a sum of effective interaction terms, each given by an operator multiplied by its associated coupling constant.

The local operators and Wilson coefficients depend on the value of the renormalisation scale μ , which is chosen to be of the order of the mass of the decaying hadron, so here $\mu \sim \mathcal{O}(m_b)$. The renormalisation scale is used to separate the physics contributions to the decay amplitude into short-distance contributions at energy scales higher than μ and long-distance ones at scales lower than μ .

Here, the leading contribution to $\mathcal{H}_{\text{eff}}^{\text{OPE}}$ comes from the electromagnetic dipole

operator, $\mathcal{O}_7$, defined as:

$$\mathcal{O}_7 = \frac{e}{16\pi^2} m_b \bar{s}_L \sigma_{\mu\nu} b_R F^{\mu\nu} . \quad (1.53)$$

It is worth noting that, when only considering the contribution from $\mathcal{O}_7$, the expression of the effective Hamiltonian is equivalent to the one in Equation 1.50. The primed operator is given by:

$$\mathcal{O}'_7 = \frac{e}{16\pi^2} m_b \bar{s}_R \sigma_{\mu\nu} b_L F^{\mu\nu} . \quad (1.54)$$

The sub-leading contributions come from the current-current operators $\mathcal{O}_1$ and $\mathcal{O}_2$, which correspond to diagrams with a W boson exchange between the two quark lines, the QCD penguin operators $\mathcal{O}_3$ through $\mathcal{O}_6$, which correspond to diagrams with a gluon exchange between the two quark lines, and the chromomagnetic gluon penguin operator $\mathcal{O}_8$, corresponding to gluon emission from the virtual up-type quark. In the SM, the only contributions from primed operators are those coming from $\mathcal{O}'_7$ and $\mathcal{O}'_8$, but those are negligible since:

$$\frac{C'_7}{C_7} = \frac{C'_8}{C_8} = \frac{m_s}{m_b} \sim 0.02 . \quad (1.55)$$

Thus, at leading order (LO), the subleading contributions from $\mathcal{O}_{i \neq 7}$ and $\mathcal{O}'_{i \neq 7}$ can be absorbed into effective $C_7^{(0)\text{eff}}$ and $C'^{(0)\text{eff}}_7$ coefficients, respectively, and the amplitude for $b \rightarrow s\gamma$ can be written:

$$\mathcal{M}(b \rightarrow s\gamma)_{\text{LO}} = \langle f | \mathcal{H}_{\text{eff}}^{\text{OPE}} | i \rangle = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \left[C_7^{(0)\text{eff}}(\mu) \langle f | \mathcal{O}_7 | i \rangle + C'^{(0)\text{eff}}_7(\mu) \langle f | \mathcal{O}'_7 | i \rangle \right] , \quad (1.56)$$

where i and f denote the hadronic initial and final states, respectively, and $\langle f | \mathcal{O}_7^{(i)} | i \rangle$ are the hadronic matrix elements of $\mathcal{O}_7^{(i)}$ between i and f . The effective coefficient $C_7^{(0)\text{eff}}$ denotes the leading logarithmic expression to C_7^{eff} , while the matrix elements include contributions to the amplitude $\mathcal{M}(b \rightarrow s\gamma)$ from scales lower than μ .

Perturbative QCD corrections to the matrix element part have been computed, in particular for $B \rightarrow K^* \gamma$ decays [49, 50]. These studies indicated that higher-order SM contributions could potentially enhance the right-handed photon polarisation in $b \rightarrow s\gamma$ transitions. In some approaches [51, 52], the right-handed contamination is estimated to be as large as $\sim 10\%$. However, such results are strongly model dependent and not fully in agreement with each other, meaning that further theoretical work is needed to clarify the situation.

A well-motivated class of NP models that can produce sizeable right-handed currents are Left–Right (LR) symmetric extensions of the SM, based on the gauge

group $SU(2)_L \times SU(2)_R \times U(1)_{B-L}$ [53, 54]. These theories introduce a charged gauge boson W_R and a right-handed quark mixing matrix V_R , which is the analogue of the CKM matrix in the SM. Two new effects are relevant for radiative $b \rightarrow s\gamma$ decays: (i) loop diagrams with W_R (and possibly new heavy fermions or scalars) contribute directly to the electromagnetic dipole operators, and (ii) mixing between W_L and W_R can induce C'_7 coefficients that are not suppressed by the light-quark mass ratio m_s/m_b in the same way as in the SM. As a consequence, LR models can generate C'_7 values comparable to C_7 , leading to right-handed photon amplitudes much larger than the SM expectation.

Phenomenological studies have used $b \rightarrow s\gamma$ observables, like the inclusive branching ratio, direct CP asymmetries, and observables sensitive to photon helicity, to constrain the LR parameter space. In many scenarios, these constraints already push W_R to high masses or require small W_L – W_R mixing, but viable regions remain where sizeable C'_7 effects are possible.

A complementary framework in which enhanced right-handed currents can naturally arise is the Minimal Supersymmetric Standard Model (MSSM). Supersymmetry (SUSY) is a proposed extension of the SM in which every fermion has a bosonic superpartner and every boson has a fermionic one. The MSSM is the simplest realisation of this idea, incorporating the minimal particle content and interactions required for a consistent supersymmetric theory of the electroweak and strong interactions [55, 56]. Beyond stabilising the Higgs sector and providing a dark matter candidate, the MSSM also introduces a rich flavor structure that can strongly affect rare B decays.

In radiative $b \rightarrow s\gamma$ transitions, supersymmetric particles such as charginos, gluinos, and charged Higgs bosons can contribute *via* loop diagrams alongside the SM top– W penguin. These additional contributions can substantially modify the Wilson coefficients C_7 and C'_7 . In particular, flavor-violating effects in the squark sector, or left–right mixing among squarks, can enhance the right-handed amplitude, leading to potentially observable deviations in the photon polarisation [57]. Thus, much like LR models, the MSSM provides a concrete scenario in which $b \rightarrow s\gamma$ decays become sensitive probes of new flavor dynamics.

1.6.2 Physics observables and current status

In this Section, the physical observables which are most relevant in probing the photon polarisation in $b \rightarrow s\gamma$ transitions are presented followed by their current experimental status.

Photon polarisation

Direct measurements of the photon polarisation in radiative decays can be quite challenging, as they often require a complicated analysis of angular observables [58, 59]. On this front, the LHCb collaboration has made significant progress.

In particular, the most recent angular analysis of $B^0 \rightarrow K^{*0}(\rightarrow K^-\pi^+)e^+e^-$ decays in the low dilepton invariant mass (q^2) region [60] has set the most stringent constraints to date on the real and imaginary parts of C'_7/C_7 (see Figure 1.6).

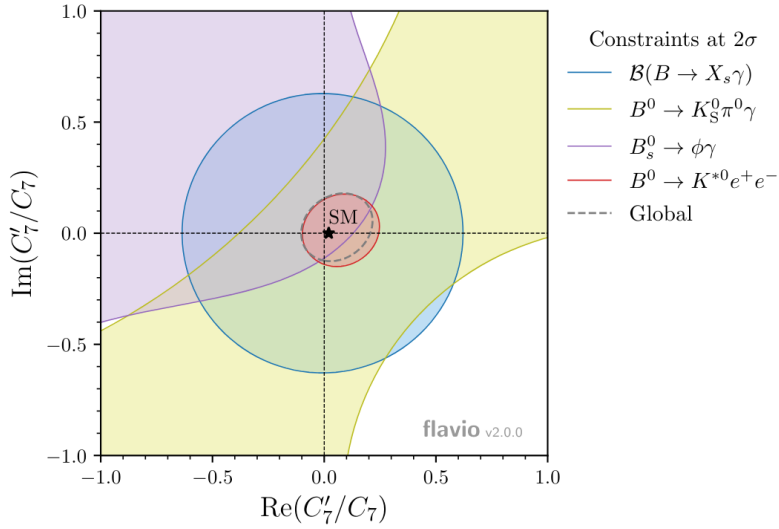


Figure 1.6: Constraints at the two standard deviations level on the real and imaginary parts of the ratio of right- and left-handed Wilson coefficients, C'_7/C_7 . The most stringent constraint comes from the angular analysis of $B^0 \rightarrow K^{*0}e^+e^-$ at very low q^2 . From the LHCb collaboration [60].

In addition, LHCb has measured the photon polarisation for the first time in radiative Λ_b baryon decays [61], which proceed through the same underlying $b \rightarrow s\gamma$ quark-level transition. This measurement is likewise consistent with the SM prediction, placing even more constraints on possible contributions from NP models.

Mixing-induced CP asymmetry

Another method to probe the photon polarisation is to measure the mixing-induced CP asymmetry in neutral B meson decays *via* a time-dependent analysis. This is the strategy followed by the measurement at the core of this thesis.

For radiative decays of neutral B mesons to a hadronic CP eigenstate f_{CP} , where, for simplicity, the flavor of the companion B meson (B_{tag}) is tagged at

time $t = 0$ of its decay, the time-dependent CP asymmetry is given by [26]:

$$\mathcal{A}_{CP}(t) = \frac{\Gamma(\bar{B}^0(t) \rightarrow f_{CP}\gamma) - \Gamma(B^0(t) \rightarrow f_{CP}\gamma)}{\Gamma(\bar{B}^0(t) \rightarrow f_{CP}\gamma) + \Gamma(B^0(t) \rightarrow f_{CP}\gamma)} \quad (1.57)$$

$$\simeq -\xi \sin(2\psi) \sin(\phi_M - \phi_L - \phi_R) \sin(\Delta m_d t) \quad (1.58)$$

$$= S \sin(\Delta m_d t), \quad (1.59)$$

where $\xi = \pm 1$ is the CP eigenvalue of f_{CP} and S is the mixing-induced CP violation parameter. The only difference between this formula for the asymmetry and the one introduced in Section 1.4.2 is that here the total B^0 and $\bar{B}^0$ decay amplitudes are each expressed as the sum of the decay amplitudes involving left- and right-handed photons, such as:

$$\Gamma(B^0(t) \rightarrow f_{CP}\gamma) = |\mathcal{M}_L(t)|^2 + |\mathcal{M}_R(t)|^2, \quad (1.60)$$

$$\Gamma(\bar{B}^0(t) \rightarrow f_{CP}\gamma) = |\bar{\mathcal{M}}_L(t)|^2 + |\bar{\mathcal{M}}_R(t)|^2. \quad (1.61)$$

This is because states with different helicities do not interfere quantum-mechanically and the photon helicity is in principle measurable.

The approximation in Equation 1.58 comes from the fact that the term related to direct CP violation, which is proportional to C , has been neglected. Moreover, in this expression, $\sin(2\psi)$ is used to parameterise the relative amount of left- and right-handed photons, defined as:

$$\sin(2\psi) = \frac{2|\mathcal{M}_L\mathcal{M}_R|}{|\mathcal{M}_L|^2 + |\mathcal{M}_R|^2}, \quad (1.62)$$

while the relative phase ϕ_M is the one related to B^0 – $\bar{B}^0$ mixing (see Section 1.4.1), with $\phi_M = 2\beta$ for B_d^0 mesons and $\phi_M \approx 0$ for B_s^0 mesons. Due to the suppression of the right-handed amplitude in $b \rightarrow s\gamma$, $\sin(2\psi)$ is predicted to be close to zero. The relative phases $\phi_{L,R}$ are the weak phases in the $b \rightarrow s\gamma$ process, which are predicted to be very small in the SM ($\mathcal{O}(\lambda^4)$). They are defined as:

$$\phi_{L,R} = \sin^{-1} \left(\frac{\Im(\mathcal{M}_{L,R})}{|\mathcal{M}_{L,R}|} \right). \quad (1.63)$$

Thus, in the SM, the time-dependent CP asymmetry is expected to be close to zero in radiative B decays of the kind $B^0 \rightarrow f_{CP}\gamma$. Any deviation from this expectation would thus be a clear signal of contributions to the right-handed photon polarisation in $b \rightarrow s\gamma$ from NP processes. This is an indirect method of probing the photon polarisation, as it does not allow to determine the ratio between the left- and right-handed amplitudes $\mathcal{M}_R/\mathcal{M}_L$, and thus relate it to $|C_7'/C_7|$. This

means that, in the case of an experimental result which significantly deviates from the SM prediction of $\mathcal{A}_{CP}(t)$, the relative phases $\phi_{M,L,R}$ would need to be fixed from several NP models to values that best describe the observed value of $\mathcal{A}_{CP}(t)$.

The mixing-induced CP asymmetry has been measured in several radiative B decay modes, but the sensitivities of these measurements are still statistically limited and far from reaching the theoretical precision. The best measurements of this observable come from the B factory experiments, due to the relatively clean experimental environment they provide (see Chapter 2).

In the “golden mode” for studying CP violating effects in radiative B decays, $B^0 \rightarrow K_S^0 \pi^0 \gamma$, both Belle and BABAR have performed measurements of the CP parameter S [62, 63]. They reported $S_{K_S^0 \pi^0 \gamma} = -0.10 \pm 0.31 \pm 0.07$ and $S_{K_S^0 \pi^0 \gamma} = -0.78 \pm 0.59 \pm 0.09$, respectively. In the mass region of the $K^*(892)$, they reported $S_{K^{*}\gamma} = -0.32_{-0.33}^{+0.36} \pm 0.05$ and $S_{K^{*}\gamma} = -0.03 \pm 0.29 \pm 0.03$, respectively. This measurement was also recently performed in Belle II [64], resulting in $S_{K_S^0 \pi^0 \gamma} = 0.04_{-0.44}^{+0.45} \pm 0.10$ and $S_{K^{*}\gamma} = 0.00_{-0.26}^{+0.27} \pm 0.03$, in the $K^*(892)$ mass region. In all these measurements, the measured value of the direct CP violation parameter, C , is consistent with zero, which validates the approximation made in Equation 1.58.

Belle and BABAR have also studied $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays [65, 66]. They reported $S_{K_S^0 \pi^+ \pi^- \gamma} = 0.09 \pm 0.27_{-0.07}^{+0.04}$ and $S_{K_S^0 \pi^+ \pi^- \gamma} = 0.14 \pm 0.25 \pm 0.03$, respectively.

At this point, it is necessary to describe how the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decay proceeds. First, the B meson decays to an intermediate state, which consists of a kaonic resonance K_{res} and a photon. The K_{res} state is very short-lived⁷ and can decay into $K_S^0 \pi^+ \pi^-$ in several different ways. Out of all the different decay paths, only those involving the $K_S^0 \rho^0$ intermediate state, where $\rho^0 \rightarrow \pi^+ \pi^-$, are CP eigenstates and therefore contribute to the mixing-induced CP asymmetry, with the CP parameter $S_{K_S^0 \rho^0 \gamma}$. The rest of the decay paths dilute the measured value of the fundamental CP parameter S and must be separated from the main contribution using a *dilution factor*, in order to obtain $S_{K_S^0 \rho^0 \gamma}$ from any measurement of $S_{K_S^0 \pi^+ \pi^- \gamma}$ including all intermediate contributions to the final state, $K_S^0 \pi^+ \pi^- \gamma$. This is in contrast to $B^0 \rightarrow K_S^0 \pi^0 \gamma$ decays, where the dominant contribution proceeds through a CP eigenstate and no dilution factor is needed at first order. The dilution factor for $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ is defined as:

$$\mathcal{D} = \frac{S_{K_S^0 \pi^+ \pi^- \gamma}}{S_{K_S^0 \rho^0 \gamma}}. \quad (1.64)$$

⁷The lifetime of the K_{res} state is of the order $\mathcal{O}(10^{-24})$ s, which is nowhere near the sensitivity of any current experiment, so it can be considered to decay near-instantly.

The dilution factor can be determined *via* an amplitude analysis to disentangle the contribution of the $K_S^0 \rho^0$ intermediate state leading to the $K_S^0 \pi^+ \pi^-$ final state in the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decay from all other possible intermediate K_{res} paths. Such an analysis is experimentally feasible at Belle II using rather $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays, under the assumption of isospin symmetry, due to the larger branching fraction and higher reconstruction efficiency of the charged mode, though it is beyond the scope of the present thesis. Instead, the analysis presented here measures the *effective* mixing-induced CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays, $S_{\text{eff}} \equiv S_{K_S^0 \pi^+ \pi^- \gamma}$, which also includes contributions from non- CP eigenstates.

The value of the dilution factor in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ has been measured in the past by both Belle and BABAR [65, 66]. Belle reported $\mathcal{D} = 0.83_{-0.03}^{+0.19}$, while BABAR measured $-0.79_{0.17}^{0.19}$. The opposite sign arises from the different conventions used in defining the CP eigenvalues in the two analyses.

Using the above values for the dilution factor, the measurements of S_{eff} by Belle and BABAR [65, 66] can be translated to $S_{K_S^0 \rho^0 \gamma} = 0.11 \pm 0.33_{-0.09}^{+0.05}$ and $S_{K_S^0 \rho^0 \gamma} = -0.18 \pm 0.32_{-0.05}^{+0.06}$, respectively.

Beyond the global measurement of S_{eff} in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays, it is possible to extract additional information from the distribution of events across the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane.

More precisely, the authors of Ref. [27] proposed two new observables, S^+ and S^- , which are sensitive to different combinations of the underlying weak and strong phases and thus provide constraints on the photon helicity that are complementary to the existing ones. They are defined as:

$$S^+ = S^I + S^{\bar{I}}, \quad (1.65)$$

$$S^- = S^I - S^{\bar{I}}, \quad (1.66)$$

where S^I and $S^{\bar{I}}$ are the mixing-induced CP violation parameters measured in the two complementary regions of the Dalitz plane, defined by $m_{K_S^0 \pi^+}^2 < m_{K_S^0 \pi^-}^2$ and $m_{K_S^0 \pi^+}^2 > m_{K_S^0 \pi^-}^2$, respectively (see Figure 1.7).

Measuring S^+ and S^- gives access to the real and imaginary parts of $S_{K_S^0 \pi^+ \pi^- \gamma}$, which allows to place new constraints on the real and imaginary parts of C'_7/C_7 . Figure 1.8 shows projections of these constraints on the C'_7/C_7 complex plane using future expected datasets from Belle II, together with the current and future projected constraints obtained from the angular analysis of $B^0 \rightarrow K^{*0} e^+ e^-$ in the low q^2 region by the LHCb experiment. It can be seen that the projected constraints from S^+ and S^- are expected to be competitive with those obtained from $B^0 \rightarrow K^{*0} e^+ e^-$ at low q^2 .

Note however, that these constraints not only depend on the measured values of S^+ and S^- , but also on the dilution-like factors a and b , which can be determined

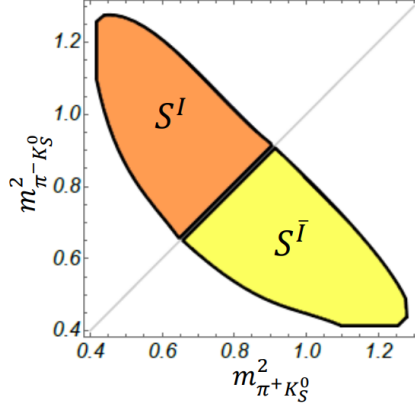


Figure 1.7: Illustration showing the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane. The $m_{K_S^0 \pi^+}^2 = m_{K_S^0 \pi^-}^2$ bisector splits the plane into two halves and the measured S^+ and S^- parameters are linear combinations of the S parameters measured simultaneously in both halves.

from an amplitude analysis of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays. The first measurement of S^+ and S^- in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays is presented in this thesis, acting as the “proof of concept” for such an analysis, together with a measurement of S_{eff} and C , using data recorded by the Belle II experiment. The measurement of S^+ and S^- will provide novel experimental input which, when combined with future amplitude analyses of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$, will allow to further constrain the photon helicity structure in $b \rightarrow s\gamma$ transitions.

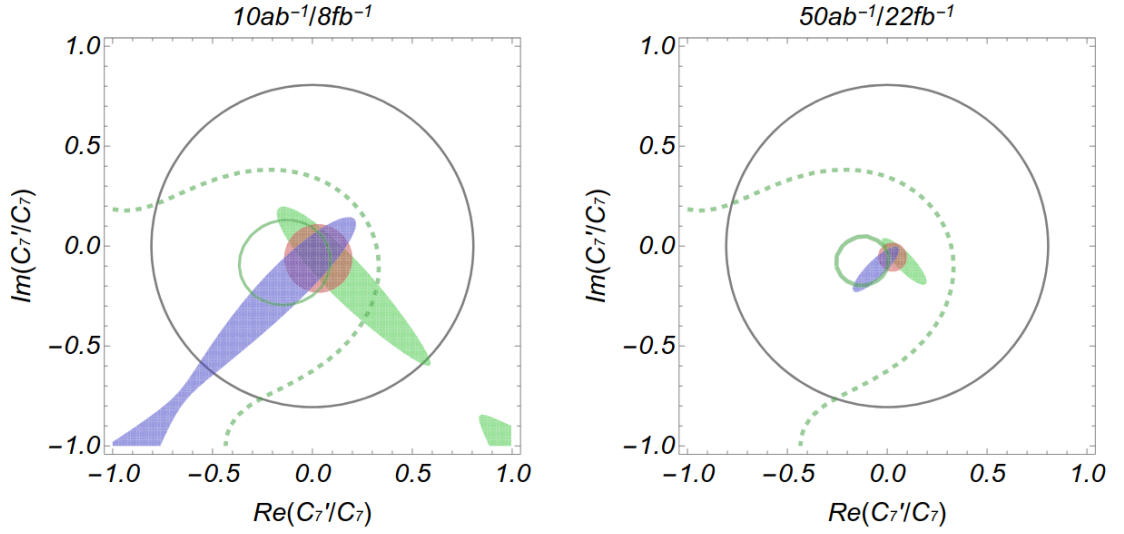


Figure 1.8: Expected constraints on $\Re(C'_7/C_7)$ and $\Im(C'_7/C_7)$ at the three standard deviations level, derived from the measurement of the new proposed observables S^+ and S^- with 10 ab^{-1} and 50 ab^{-1} of Belle II data. The red, green and blue constraints are derived using three different sets of values for $\{S^+, S^-, a \text{ and } b\}$, where a and b are two dilution-like factors. The dashed (solid) green contour is derived from the angular analysis of $B \rightarrow K^* e^+ e^-$ decays at low q^2 using LHCb Run I data [67] (Run II for the left and Run III for the right plot). Taken from [27].

Chapter 2

The Belle II experiment

Contents

2.1	The SuperKEKB accelerator	30
2.2	The Belle II detector	33
2.2.1	The Pixel Detector	36
2.2.2	The Silicon Vertex Detector	37
2.2.3	The Central Drift Chamber	39
2.2.4	Particle Identification (TOP and ARICH)	41
2.2.5	The Electromagnetic Calorimeter	43
2.2.6	The K_L^0 and Muon Detector	44
2.3	Trigger system	45
2.3.1	Level 1 Trigger	46
2.3.2	High Level Trigger	48
2.4	The Belle II Analysis Software Framework	48
2.5	Simulation	48
2.6	Reconstruction	49
2.6.1	Tracking	49
2.6.2	Charged Particle identification	51
2.6.3	K_S^0 reconstruction	51
2.6.4	Photon reconstruction	51
2.6.5	B meson reconstruction	52
2.6.6	Vertex reconstruction	53

This chapter describes the experimental setup which is central to the core measurement presented in this thesis, comprising the SuperKEKB accelerator and the Belle II detector [68, 69].

2.1 The SuperKEKB accelerator

SuperKEKB (see Figure 2.1) is a circular electron-positron ($e^+ e^-$) collider with asymmetric beam energies located at the High Energy Accelerator Research Organization (KEK) in Tsukuba, Japan.

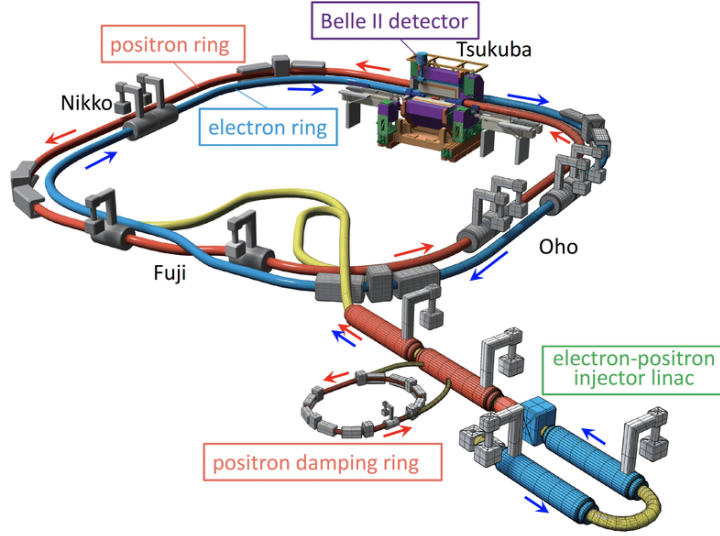


Figure 2.1: Schematic view of the SuperKEKB accelerator. The LINAC is shown as well as the HER and LER rings. Collisions occur at the Interaction Point, at the center of the Belle II detector. Taken from [68].

Electrons are generated in a pre-injector with a linear accelerator (LINAC) and are accelerated to an energy $E_{e^-} = 7.007$ GeV. Positrons are produced by irradiating a tungsten target with electrons produced in the pre-injector and they are subsequently accelerated to an energy $E_{e^+} = 4$ GeV. The beams are then stored in two 3 km storage rings; the High Energy Ring (HER) for electrons and the Low Energy Ring (LER) for positrons. Collisions between electrons and positrons take place at the Interaction Point (IP), found at the center of the Belle II detector (see Figure 2.5). The electrons and positrons collide in so-called “bunches” every 4 ns. The electron and positron beams also collide with a small crossing angle of 83 mrad in order to limit as much as possible the number of beam-beam interactions

as well as to reduce the effective size of the beam spot, in the z -axis. The energy of the collisions at the center-of-mass (C.o.M.) frame is calculated as:

$$\sqrt{s} = \sqrt{\left(\frac{E_e^- + E_e^+}{c^2}\right)^2 - \left(\frac{\mathbf{p}_e^- + \mathbf{p}_e^+}{c}\right)^2} \simeq 10.58 \text{ GeV}/c^2, \quad (2.1)$$

where $\mathbf{p}_e^-$, $\mathbf{p}_e^+$ are the momenta of the electron and positron, respectively. This energy corresponds to the rest mass of the $\Upsilon(4S)$ resonance, which is a bound state consisting of a $b\bar{b}$ quark pair, that decays almost exclusively¹ into pairs of neutral or charged B mesons.

Despite the fact that SuperKEKB operates at a C.o.M. energy that corresponds to the rest mass of the $\Upsilon(4S)$, an $\Upsilon(4S)$ resonance is produced in less than 1% of $e^+ e^-$ collisions. This is because of the relatively low production cross section of the $b\bar{b}$ process, at approximately 1.1 nb, when compared to all other $e^+ e^- \rightarrow f\bar{f}$, where f denotes all fermions. Figure 2.2 shows all the different processes resulting from $e^+ e^-$ collisions with their corresponding cross sections, as they can be found in Table 18, Chapter 4 of Ref. [71].

Cross sections [nb] for $e^+ e^-$ collisions at $\sqrt{s} = 10.58$ GeV

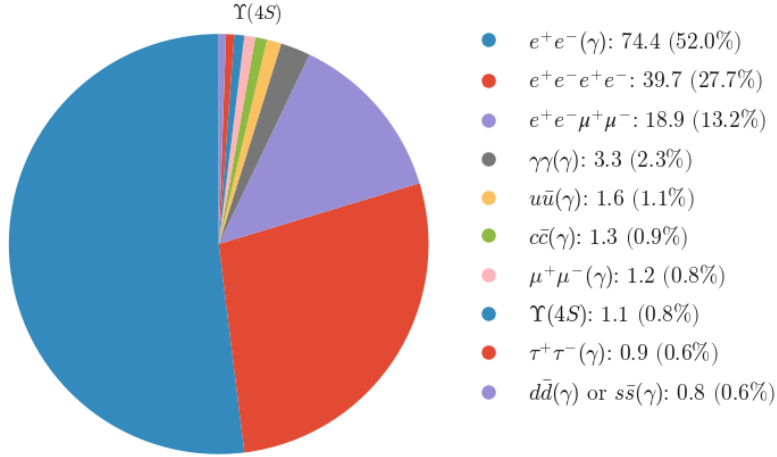


Figure 2.2: Cross sections of $e^+ e^-$ processes observed at $\sqrt{s} = 10.58$ GeV. Cross section numbers taken from [71].

SuperKEKB is the successor to the KEKB accelerator and is designed to achieve an instantaneous luminosity² that is forty times higher than its prede-

¹with branching fraction greater than 96% [70].

²Instantaneous luminosity is a measure of the collision rate per unit area in a particle accelerator, indicating how many particles are likely to interact in a given amount of time.

cessor, with a target peak instantaneous luminosity of $6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$. This will be achieved through the use of the so-called “nano-beam” scheme, which reduces the vertical beam size at the interaction point down to $\sim 50 \text{ nm}$, significantly increasing the probability of collision. The first data-taking period of Belle II took place between 2019–2022, also known as Run 1, in which 365 fb^{-1} of integrated luminosity³ were collected at the $\Upsilon(4S)$ energy, amounting to ~ 388 million $B\bar{B}$ pairs [72]. This dataset is the one analyzed to perform the measurement at the center of this thesis research. Additionally, more collision data were collected at slightly lower energies, called off-resonance, as well as higher energies to produce higher spin resonances, like $\Upsilon(5S)$, bringing the total integrated luminosity of Run 1 up to 428 fb^{-1} , which is comparable to the amount of data collected by the BaBar experiment during its runtime [73]. By the end of data-taking, Belle II aims to collect 50 ab^{-1} of data, which is more than 50 times the amount collected by each of its predecessors, Belle and BaBar. Figure 2.3 shows the timeline of Belle II data-taking during Run 1. The world record for the highest instantaneous luminosity at the time of writing this thesis document was reached by SuperKEKB at $5.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, on December 24th 2024⁴.

There are certain features that make SuperKEKB an ideal environment to study time-dependent CP violation phenomena. The first is a direct consequence of the asymmetry between the electron and positron beams, which results in the $\Upsilon(4S)$ resonance being boosted in the lab frame, with a Lorentz boost factor of: $\beta\gamma = (\mathbf{p}_e^- - \mathbf{p}_e^+)/(\sqrt{s}) \simeq 0.28$. Consequently, the pair of B mesons produced from the $\Upsilon(4S)$ travel a greater distance from the IP ($\sim 100 \mu\text{m}$) before decaying, compared to their flight distance in the case of equal beam energies ($\sim 1 \mu\text{m}$). This displacement is essential, as without it, even the highest-resolution vertex detectors would be unable to resolve the individual decay positions of the two B mesons in an event, using the trajectories of their decay products [71].

The second feature, as explained in Section 1.4.2, is that the two B mesons produced from the $\Upsilon(4S)$ are produced coherently, so there is always one B^0 and one $\bar{B}^0$ in the pair⁵, until one of the two decays. This quantum entanglement allows one to determine the flavor state (either B^0 or $\bar{B}^0$) of one B meson by tagging the flavor of the other B meson in the event at the time of the latter’s decay.

³Integrated luminosity is a measure of the amount of collisions per unit area and is usually measured in inverse femtobarns, where $1 \text{ fb}^{-1} = 10^{39} \text{ cm}^{-2}$, in particle collider experiments. It is a measure of the amount of data collected by the experiment.

⁴The highest instantaneous luminosity reached while the Belle II detector was on is $4.7 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

⁵or one B^+ and one B^- in the case of charged B mesons, but we are focusing on the neutral case here.

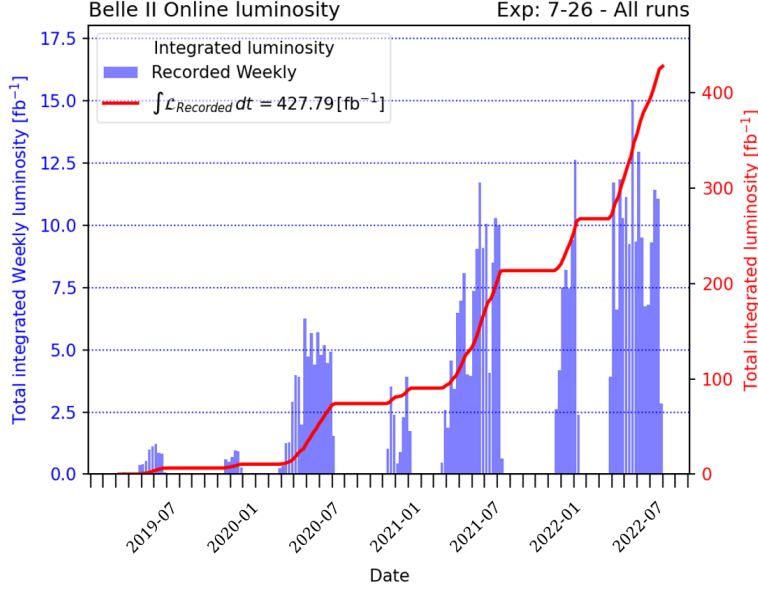


Figure 2.3: Amount of data (in fb^{-1}) collected per week by Belle II during its first physics run period. From the Belle II collaboration.

Both of these aspects are illustrated in Figure 2.4, which shows what a typical $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B^0\bar{B}^0$ event looks like at SuperKEKB. Since the Lorentz boost is well known, the measurable distance between the decay positions of the two B mesons, Δz , can be related to the difference between their decay times, Δt , as: $\Delta z = \beta\gamma c\Delta t$. Moreover, the charge of one of the particles in the decay of the *tag-side* B meson tags its flavor state at the time of its decay, which, due to the quantum entanglement of the $B\bar{B}$ pair, also tags the flavor state of the *CP-side*, or *signal-side* B meson at that same time. More details on how this Flavor Tagging is performed at Belle II are given in Section 4.1.

Reliable Flavor Tagging and precise knowledge of the B mesons' decay positions in each event are both essential in achieving a high experimental precision in measurements of time-dependent CP asymmetry, such as the one at the center of this thesis research. Reaching a high enough performance on both of these fronts is entirely dependent on the instrument used to detect the particles produced in the e^+e^- collisions delivered by SuperKEKB; the Belle II detector.

2.2 The Belle II detector

The Belle II detector (see Figure 2.5) is a state-of-the-art barrel-shaped spectrometer consisting of several detection layers arranged around the beam pipe, each

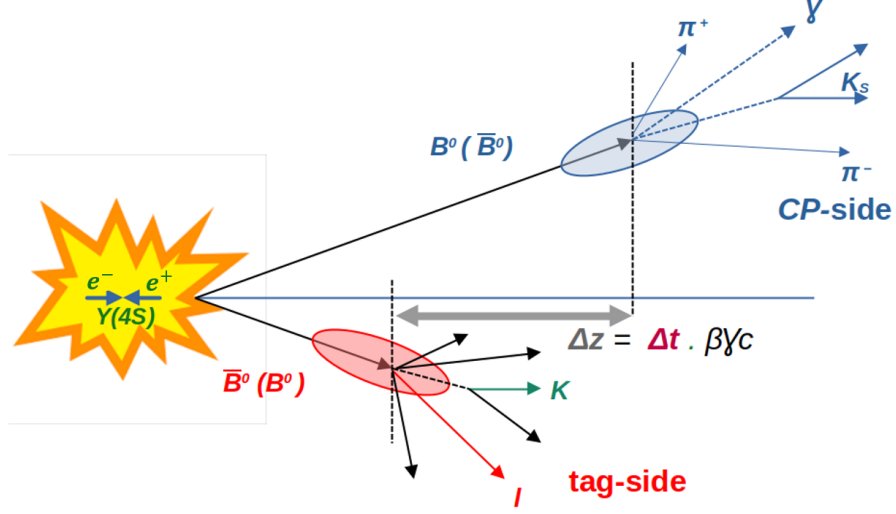


Figure 2.4: Schematic representation of a collision event where an $\Upsilon(4S)$ is produced and decays in a pair of neutral B mesons. The Lorentz boost of the produced $B\bar{B}$ pair ensures they will travel far enough for their individual decay positions to be resolvable. The decay of the *tag-side* B meson tags the flavor state of the *CP-side* B meson at the time of the former's decay.

one being responsible for different aspects of particle detection, such as calorimetry, tracking and particle identification. The sub-detectors are positioned slightly asymmetrically along the beam axis to account for the Lorentz boost, and they provide nearly full angular coverage, which is crucial for reconstructing missing energy from neutrinos. Starting from the one closest to the IP, the Belle II sub-detectors are:

- The PiXel Detector (PXD), consisting of 2 layers of DEPFET silicon pixel sensors⁶. It possesses an excellent spatial resolution, which is essential for the precise measurement of the B mesons' decay positions, a process known as vertex reconstruction. It is also used for track fitting purposes.
- The Silicon Vertex Detector (SVD), consisting of 4 layers of double-sided strip sensors, used for tracking, vertex reconstruction, as well as charged particle identification. The PXD and the SVD comprise the so-called Vertex Detector (VXD) of Belle II.

⁶During the first data-taking period, only one ladder of the second layer was installed; the remaining ladders were added after the end of Run 1 in mid-2022. Since the innermost layer closest to the interaction point was already fully in place, the absence of most of the second layer during Run 1 had only a marginal effect on the vertex resolution.

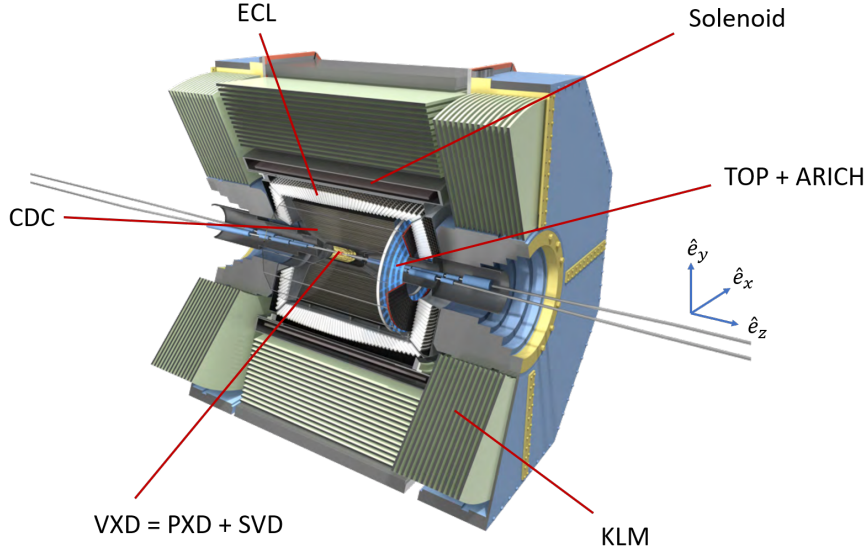


Figure 2.5: Schematic view of the Belle II detector with all sub-detectors denoted. The Belle II coordinate system is centered at the interaction point of the beams. The z -axis is defined to be nearly collinear to the electron beam direction, the x -axis points horizontally outward from the center of the accelerator rings, and the y -axis is vertical, pointing upwards. The polar angle θ is defined over the interval $[-\pi, +\pi]$, with $\theta = 0$ corresponding to the $+z$ direction and $\theta = \pm\pi$ corresponding to the $-z$ direction. The azimuthal angle ϕ spans the $[0, 2\pi]$ range around the xy -plane. Adapted from Belle II collaboration resources.

- The Central Drift Chamber (CDC), the main tracking system of Belle II, consisting of a large number of sense wires suspended in a helium-based gas mixture. It also contributes to particle identification through the measurement of energy loss (dE/dx) of charged particles that pass through it.
- A system of Particle IDentification (PID) detectors, consisting of the Time Of Propagation (TOP) detector in the barrel region and the Aerogel Ring Imaging CHerenkov (ARICH) detector in the forward endcap. They both measure the velocities of charged particles, which, when combined with momentum measurements from the preceding tracking detectors, allow to measure the particles' mass, and thus identify them.
- An Electromagnetic CaLorimeter (ECL), which consists of a large number of CsI(Tl) scintillation crystals. They allow to measure the energy of particles that stop in them, which are mainly photons, electrons and several light hadrons produced in the collision.

- The KLong (K_L^0) and Muon detector (KLM), which makes up the outermost part of the Belle II detector. It consists of an assembly of resistive plate chambers and thick iron plates designed to stop K_L^0 's and muons, as its name suggests.

Moreover, there is a superconductive solenoid magnet that generates a 1.5T magnetic field which is aligned with the beam axis and oriented in the boost direction, bending the trajectories of electrically charged particles in the detector volume. This is crucial in reconstructing the momentum of those particles by determining the curvature of their trajectories, which is the role of the main tracking sub-detectors, SVD and CDC.

In the following Sections, each of the sub-detector systems will be described in more detail.

2.2.1 The Pixel Detector

The rate of parasitic particles produced in $e^+ e^-$ collisions at SuperKEKB is significantly higher than it was at KEKB, due to the use of the nano-beam scheme. Consequently, the Belle setup, which relied solely on silicon strip sensors for vertex detection, would not have been able to handle the increased occupancy⁷ resulting from the higher beam-induced backgrounds⁸ in the Belle II case. The solution to this problem was to use pixel sensors for the innermost layers of the Belle II detector, which have a much higher number of readout channels compared to strip sensors, significantly reducing detector occupancy. Pixel sensors also inherently provide two-dimensional readout, as opposed to strip sensors, which require two layers stacked perpendicularly on top of each other in order to achieve the same two-dimensional hit detection.

The Pixel Detector (PXD) consists of two cylindrical detection layers, labeled L1 and L2, placed at a radius of 14 mm and 22 mm from the axis of the beam (or at 4 mm and 12 mm from the external wall of the beam pipe), respectively, and covering a polar angle range of $17^\circ < \theta < 150^\circ$ (see Figure 2.6). Each detection module consists of a matrix of 768×250 pixel sensors and the modules are glued by pairs to form ladders. The innermost layer is made of 8 ladders, while the outermost contains 12 ladders. Although the PXD is now fully installed, only the first layer was in place during the initial data-taking period, which corresponds to the data sample analysed in this thesis research.

⁷Detector occupancy is defined as the fraction of channels that have registered a hit in each event.

⁸The different types of beam-induced background processes will be explained in Chapter 3.

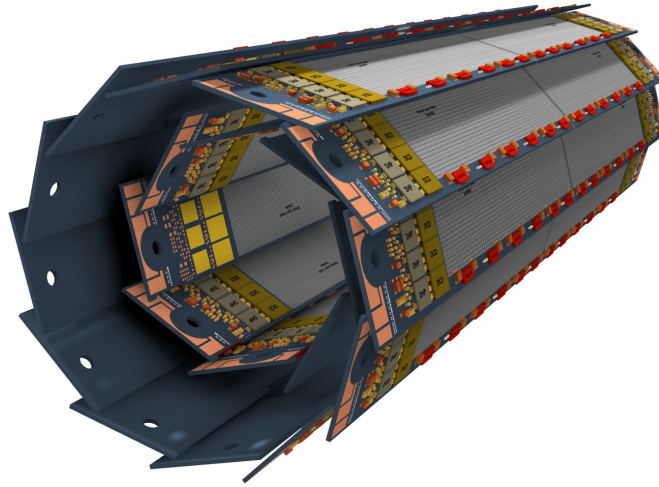


Figure 2.6: Schematic representation of the Pixel Detector with both layers. The gray colored areas correspond to the DEPFET pixel sensors. From the Belle II PXD group.

The sensors used in the PXD are based on the DEPLETED Field Effect Transistor (DEPFET) technology [74], where detection and signal amplification are both handled by a semiconductor detector. A cross section of the sensor is shown in Figure 2.7. The sensor is fully depleted when a high voltage is applied to its p^+ contact. Pairs of electrons and holes are created by incident charged particles and subsequently drift to opposite sides of the sensor, with the electrons accumulating to the so-called “Internal Gate”, which constitutes a potential minimum. The electrons are then amplified to form the signal which is continuously read out every $20\ \mu\text{m}$, with a $100\ \text{ns}$ downtime between the cycles. The most important characteristic of the DEPFET sensor is the small capacitance of its internal gate, which results in very low noise performance, even at room temperatures, making it the ideal choice for the high beam-induced background environment of Belle II.

2.2.2 The Silicon Vertex Detector

After the two layers of the PXD, there are four additional cylindrical detection layers, L3 – L6, that constitute the Silicon Vertex Detector (SVD). It also follows a ladder structure that is similar to the one of the PXD, but there are certain key differences. First and foremost, the ladders consist of Double-sided Silicon Strip Detectors (DSSD), with readout chips (APV25 [75]) that have significantly faster readout compared to the DEPFET pixel sensors of the PXD ($\sim 50\ \text{ns}$ compared to $\sim 20\ \mu\text{m}$). Due to the faster readout, the SVD can reject off-time hits due to beam-induced background more easily compared to the PXD, but it is more

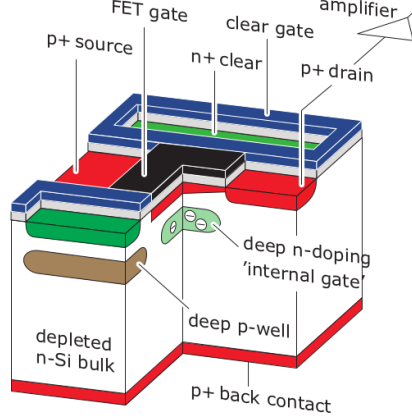


Figure 2.7: Cross section of the DEPFET sensor. Taken from [69].

sensitive to instantaneous noise from high-rate beam-induced particles. This susceptibility is effectively mitigated by placing the SVD ladders several centimeters away from the beam pipe, where the flux of parasitic particles is significantly lower (see Table 2.1). Moreover, layers 4, 5, and 6 have trapezoidal sensors placed in their forward regions, which are slanted in order to increase the acceptance of the detector as much as possible while keeping its coverage in the z -axis the same. The SVD covers a polar angle range of $17^\circ < \theta < 150^\circ$.

Layer	Radius (mm)	Ladders/layer	Sensors/ladder	Material budget ($\%X_0$)
3	39	7	2	0.4
4	80	10	3	0.5
5	104	12	4	0.5
6	135	16	5	0.6

Table 2.1: Specifications of the SVD detector layers. Information taken from [76].

Since the SVD extends out to a cylinder radius almost 14 cm, it is used to reconstruct many of the relatively long-lived K_S^0 particles, which may decay outside the volume of the PXD, using the trajectories of its charged pion daughters ($\pi^\pm$), for about 2/3 of the decays. This of course is of crucial importance in the reconstructing the $K_S^0\pi^+\pi^-\gamma$ final state for the time-dependent CP asymmetry measurement at the core of this thesis research.

Taken together, the PXD and the SVD form the innermost tracker of Belle II, also known as the Vert $\times$ Detector (VXD), that is depicted in Figure 2.8. The spatial resolution of these two sub-detectors is what primarily drives the resolution

of the B decay vertex position measurement, and consequently, the resolution of Δt in all time-dependent analyses at Belle II.

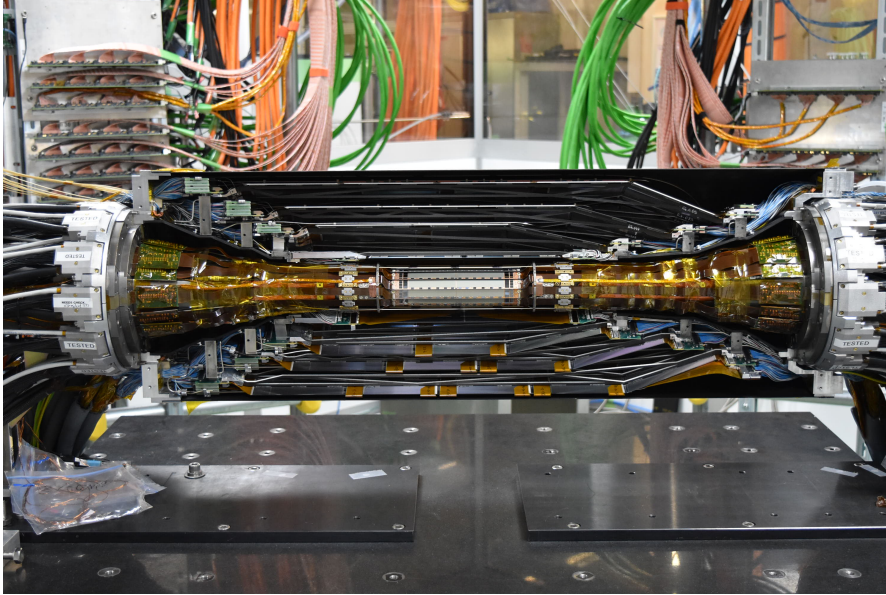


Figure 2.8: Photograph of the fully assembled VXD before installation.

Looking towards the future of Belle II data taking, the vertex detector of Belle II will require a substantial upgrade to cope with the increased background hit rate expected at the nominal SuperKEKB luminosity of $6 \times 10^{35} \text{ cm}^{-2}\text{s}^{-1}$. The most promising of these upgrades is described in more detail in Chapter 3.

2.2.3 The Central Drift Chamber

The Central Drift Chamber (CDC) is primarily responsible for reconstructing the trajectories of charged particles, with their curvature used to determine their momenta. It also contributes to the trigger system by detecting charged particles and provides particle identification information obtained through measurements of energy loss (dE/dx). Since the CDC used in Belle had already shown great performance and reliability throughout its ten year lifespan, the design of the Belle II CDC follows the example of its predecessor. The main difference between them is that the CDC of Belle II has faster readout electronics which are needed to cope with the higher background rate of SuperKEKB compared to KEKB.

The CDC consists of two barrel-shaped frames made of carbon-fiber reinforced plastic and two endplates made out of aluminum. This structure is filled with a helium-ethane gas mixture, which provides adequate drift velocity, low radiation length, good energy loss resolution, good position resolution, as well as low cross

Figure 2.9: Schematic of the CDC structure. Taken from [69].

	Belle	Belle II
Inner cylinder radius (mm)	77	160
Outer cylinder radius (mm)	880	1130
Innermost sense wire radius (mm)	88	168
Outermost sense wire radius (mm)	863	1111.4
No. of layers	50	56
No. of sense wire	8400	14336
Gas mixture	$He - C_2H_6$	$He - C_2H_6$
Sense wire diameter (μm)	30	30

Table 2.2: Specifications of the Belle and Belle II CDCs. Information taken from [69].

2.2.4 Particle Identification (TOP and ARICH)

The Particle Identification (PID) system of Belle II consists of two parts; the Time Of Propagation (TOP) counter in the barrel region and the Aerogel Ring Imaging Cherenkov (ARICH) counter in the forward endcap, both of which are based on the principle of detecting Cherenkov radiation that is emitted from charged particles passing through the detector material with velocity greater than the speed of light in this material. The so-called Cherenkov angle, θ_C is related to the velocity of the incident particle with the formula:

$$\beta = \frac{1}{n \cdot \cos\theta_C}, \quad (2.2)$$

where $\beta = v/c$ and n is the refraction index of the material in which Cherenkov photons are emitted. The measurement of the velocity combined with momentum measurements from the tracking system allows for an estimate of the particle mass.

TOP

The TOP consists of long (~ 2.7 m) bar-shaped quartz radiators that are arranged in 16 modules around the outer wall of the CDC and placed at a radius of 1.24 m. Each module of the TOP has an angular acceptance of $32^\circ < \theta < 120^\circ$, corresponding to the barrel region of the Belle II detector. Cherenkov photons emitted from charged particles that pass through the quartz radiators, go through total internal reflection on the quartz bar walls and are subsequently focused and directed towards micro-channel plate photo multiplier tubes (MCP-PMTs) located at the end of each quartz bar. The PMTs provide high-precision timing measurements, with a resolution of ~ 50 ps, enabling the distinction of different photon path lengths, and thus different Cherenkov angles. As an example, the difference between the arrival time of the kaons and pions is ~ 100 ps, which is easily resolvable by the PMTs. The spatial and temporal distributions of the detected photons are compared to the expected probability distributions for various particle hypotheses (K, π, e, μ, p, d) to evaluate particle identification efficiency. For K/π separation specifically, the TOP achieves an 85% kaon identification efficiency with a 10% pion misidentification rate.

ARICH

The ARICH is located at the forward endcap region of the Belle II detector, covering a polar angle range of $15^\circ < \theta < 30^\circ$. Figure 2.11 depicts the working principle of the ARICH, which, like the TOP, is based on the detection of Cherenkov light emitted by charged particles passing through its radiator material made of silica

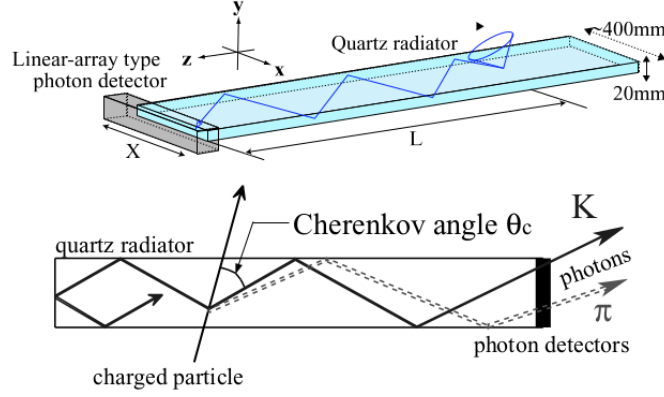


Figure 2.10: Top: Schematic overview of the TOP counter. Bottom: Schematic side-view of TOP counter illustrating how different particle species lead to Cherenkov photons being emitted at different angles. Taken from [69].

aerogel, in this case. Contrary to the TOP, the emitted photons are allowed to propagate freely as a cone of Cherenkov light centered around the particle's trajectory. When this light reaches the photon detector plane, it forms a ring whose radius is characteristic of the particle's velocity. An array of Hybrid Avalanche Photo Detectors, or HAPDs, is used to image the Cherenkov ring pattern, providing precise spatial information used for particle identification. Figure 2.11 depicts this array of photon detectors.

The radiator consists of two layers of aerogel, each 20 mm thick, with refractive indices of 1.055 and 1.065. These values are chosen for two main reasons:

- they fall within the optimal range of 1.04–1.06, enabling reliable separation between kaons and pions with momenta around 4 GeV/c; and
- they cause the Cherenkov cones from the two layers to overlap on the detector plane, which reduces the spread of the ring image due to uncertainties in the photon emission point.

Thanks to its design, the ARICH achieves a kaon identification efficiency of over 90% with a pion misidentification rate below 10% in the forward region for momenta up to around 4 GeV/c.

The high performances of both the TOP and the ARICH make them crucial components of particle identification system of Belle II, particularly for physics channels involving charged kaons in the final state.

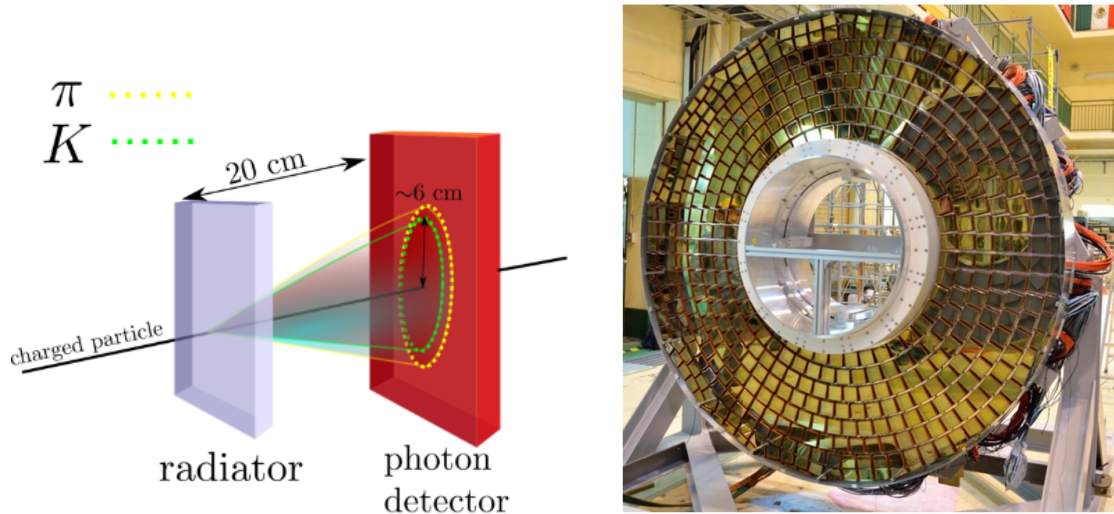


Figure 2.11: Left: Working principle of the ARICH counter. Right: Photo of the ARICH before installation. From the Belle II ARICH group.

2.2.5 The Electromagnetic Calorimeter

The main tasks of the Electromagnetic Calorimeter (ECL) are the detection of photons with high efficiency, the precise determination of their energy and angular coordinates, electron identification, trigger and detection of K_L^0 together with the KLM. It can detect photons with energies as low as 20 MeV and up to 4 GeV. The Belle II calorimeter is almost entirely identical in structure to the calorimeter of Belle, except for its upgraded electronics to cope with the increased rate of soft and hard background photons. It consists of 8736 CsI(Tl) pyramid-shaped scintillation crystals with attached preamplifiers and readout electronics. The 3 m long barrel region consists of 6624 of these CsI crystals and has an inner radius of 1.25 m, while the endcaps consist of 2112 crystals and are located at $z_1 = 2$ m and $z_2 = -1$ m from the IP (see Figure 2.12). The calorimeter covers a polar angle range of $12.4^\circ < \theta < 155.1^\circ$ except for two $\sim 1^\circ$ gaps between the barrel and endcaps.

Each crystal has a $6 \times 6 \text{ cm}^2$ cross section and is 30 cm in length, which is equivalent to about 16 radiation lengths. The crystals are wrapped in a layer of $200 \mu\text{m}$ thick Teflon and covered by a sheet of $25 \mu\text{m}$ thick aluminum and $25 \mu\text{m}$ thick mylar. The scintillation light is read out by two $10 \times 20 \text{ mm}^2$ photodiodes that are glued to the rear surface of each crystal and a preamplifier is attached to each photodiode, providing two independent output signals for each crystal. Using cosmic muons, the average output of output of the crystals was measured to be ~ 5000 photoelectrons per MeV, with a noise level of ~ 200 keV before any

irradiation.

The intrinsic energy resolution of the Belle calorimeter is given by:

$$\frac{\sigma_E}{E} = \sqrt{\left(\frac{0.066\%}{E}\right)^2 + \left(\frac{0.81\%}{\sqrt[4]{E}}\right)^2 + (1.34)^2}, \quad (2.3)$$

where E is measured in GeV. The first term in the formula for the energy resolution is the so-called “noise” term, which dominates at low energies and accounts for instrumental effects such as electronic noise from the readout system and pile-up noise from other low-energy particles created by the accelerator’s high beam currents. The second term is the so-called “stochastic” term, which accounts for the statistical fluctuations in the number of particles produced in the electromagnetic shower. This effect also becomes less significant as the energy of the incident particle increases. Lastly, the constant term represents systematic effects that do not scale with energy, such as non-uniformities in light collection efficiency, shower energy leakage from the finite size of the crystals and calibration uncertainties. This term sets the resolution limit at high energies.

For the Belle II calorimeter, the same effects constrain the energy resolution. However, upgraded readout electronics help mitigate some noise contributions at the lower end of the energy spectrum. The achieved performance is similar to that of the Belle calorimeter, with resolutions of about 2.5% at 100 MeV and 1.7% at 5 GeV [77]. Recent studies using Graph Neural Networks indicate further improvements may be possible, particularly at low energies and under high beam-induced background conditions [78].

2.2.6 The K_L^0 and Muon Detector

Furthest away from the IP and outside of the superconducting solenoid, there is the K_L^0 and muon detector (KLM), which is composed of alternating layers of 4.7 cm thick iron plates and active detector elements. The iron plates serve as return for the magnetic field lines produced by the solenoid, while also providing close to 4 interaction lengths of material in which hadronic showers from K_L^0 mesons can take place. The detection elements in the octagonal barrel region are Resistive Plate Chambers (RPCs), while the ones in the endcaps consist of scintillator strips coupled with silicon photomultipliers (SiPM). The barrel consists of 15 detector layers and 14 iron plates and covers a polar angle range of $55^\circ < \theta < 125^\circ$, while the endcap regions are composed of 14 layers of detectors and 14 iron plates and extend the polar angle range to $20^\circ < \theta < 155^\circ$ (see Figure 2.13).

The RPCs consist of 2 mm glass electrode planes separated by a 2 mm thick gap filled with a 62% HFC-134a (freon 134a), 30% argon and 8% butane-silver gas

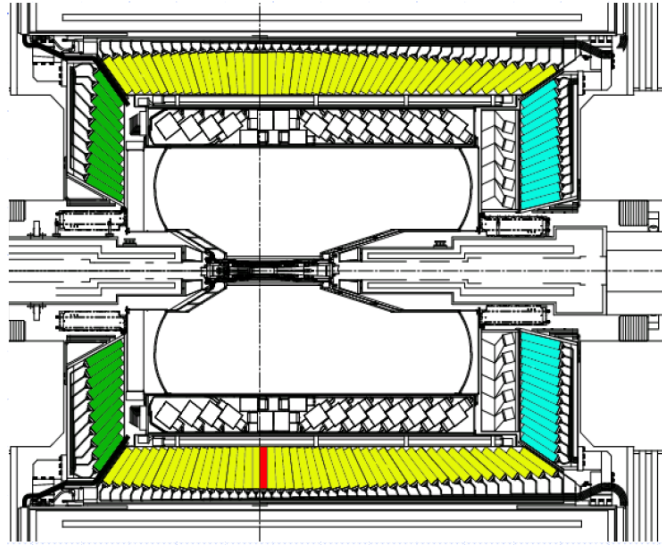


Figure 2.12: Schematic of the ECL showing the arrangement of thallium-doped CsI crystals in the barrel and endcap regions. From the Belle II ECL group.

mixture. High voltage is distributed along the electrodes *via* a thin layer of carbon-doped paint. Electrons generated by particles passing through and ionizing the gas molecules are accelerated by the electric field and collected at metal strips located at the end of the RPCs. Dielectric foil is placed between these strips that acts as a transmission line with a characteristic impedance of 50Ω . Strips of scintillators are used in the endcaps instead of RPCs, due to the long dead time of the latter, which struggles to cope with the increased hit rate caused by the beam-induced background. The scintillator strips have a length of 2.8 m and a cross section of $7 - 10 \text{ mm} \times 40 \text{ mm}$. Fibers glued to the strips carry the scintillation light to the photomultipliers which have a very good time resolution of $\sim 1 \text{ ns}$ and a high output rate.

2.3 Trigger system

The Belle II trigger system is responsible for keeping physics events of interest and discarding background events based on certain criteria. Since triggering is done online, meaning during the actual process of data taking, a maximum average trigger rate of 30 kHz is needed to match the expected collision rate at the nominal SuperKEKB luminosity. Another important requirement is a near 100% detection efficiency for hadronic events coming from $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ and $e^+e^- \rightarrow q\bar{q}$. The trigger must also have a fixed latency of $\sim 5 \mu\text{s}$ and its timing precision must be better than 10 ns. Finally, the separation power between two events must be

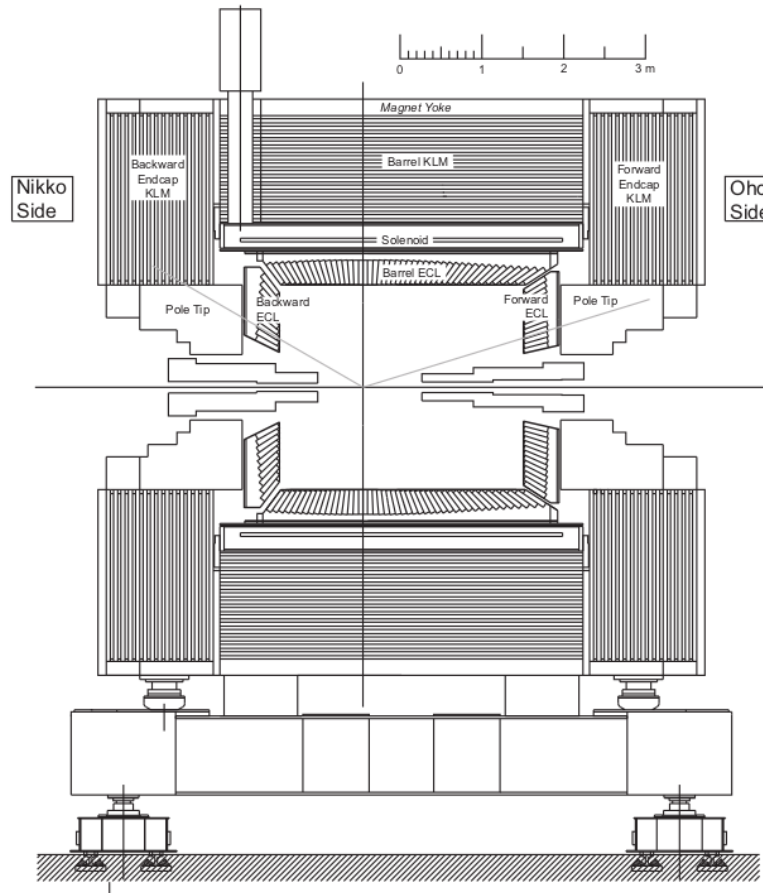


Figure 2.13: Schematic side view of the KLM, located outside the ECL and solenoid. The nominal polar angle acceptance of Belle II is marked by the gray lines that meet at the IP. Taken from [69].

at least 200 ns.

The trigger scheme follows from that of Belle, with all components and connections replaced with new technologies to accommodate the increased event rate of SuperKEKB. It consists of two parts; the Level 1 (L1) Trigger and the High Level Trigger (HLT).

2.3.1 Level 1 Trigger

The L1 Trigger is a hardware-based trigger that uses information from some of the Belle II sub-detectors to make a decision on whether to keep an event or discard it. Information from these so-called sub-triggers is fed to the Global Reconstruction Logic (GRL) of the L1 Trigger, which performs a low-level reconstruction and sends

its output to the Global Decision Logic (GDL), which makes the final decision. Each component of the L1 Trigger utilises a Field Programmable Gate Array (FPGA), which allows the trigger logic to be configured and adjusted, rather than being fixed in hardware.

The L1 trigger scheme is shown in Figure 2.14. The two main sub-triggers are those of the CDC and the ECL, which provide charged track information (momentum, position, charge, multiplicity, etc.) and energy cluster information, respectively. From there, the Barrel PID (BPID) and Endcap PID (EPID) sub-triggers give precise timing information, while the KLM sub-trigger provides muon track information. This information is compiled by the GRL and submitted to the GDL which makes a trigger decision based on the output of the different sub-triggers. The output of the L1 Trigger is then fed to the HLT for a more refined event selection.

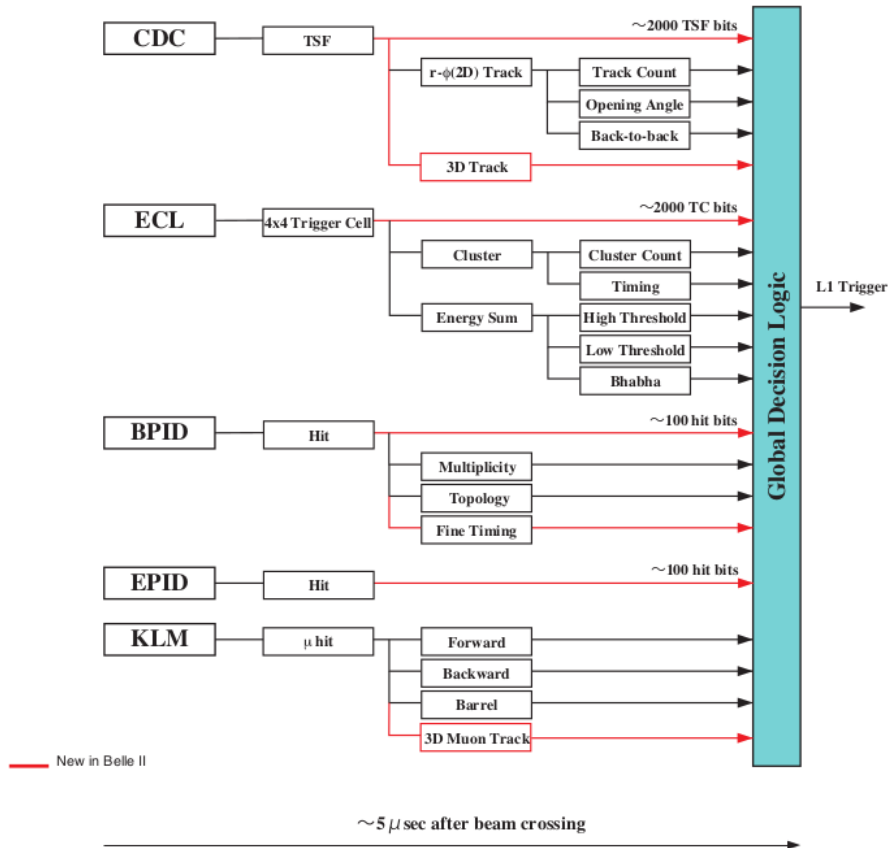


Figure 2.14: Schematic overview of the L1 Trigger. Information from each sub-trigger is fed to the GDL which makes a trigger decision. Taken from [69].

2.3.2 High Level Trigger

The HLT performs a full, real-time reconstruction of each event based on information from all the different sub-systems, except the PXD. The exclusion of the PXD for this step is done to avoid having to process its full output, which is too large to process for each event in the allotted time. Instead, information is gathered from specific regions of interest in the PXD’s readout, which are defined by extrapolating information from the SVD and CDC sub-systems. The event reconstruction is performed based on the same software used in offline reconstruction, described in Section 2.6, to avoid additional systematic uncertainties. Events are then categorised based on their type (hadronic, low multiplicity, cosmic, etc.), for ease of access by the collaboration.

2.4 The Belle II Analysis Software Framework

The Belle II Analysis Software Framework (BASF2) [79] is a set of software tools developed and maintained by the Belle II collaboration for use in data processing, physics analyses, as well as detector simulation studies. It consists of modules written in C++ [80] and interfaced with steering files written in PYTHON [81]. Each steering file comprises a sequence of modules, each responsible for a specific aspect of analysis, such as detector simulation, vertex fitting, or the application of selection criteria, all of which are executed in sequence when the file is run. The various detector configurations and versions of analysis modules are stored in a conditions database and are accessed through so-called “global tags”, while the analysed data are usually stored in ROOT TTREE objects [82].

2.5 Simulation

Analyses at Belle II almost always involve a full Monte Carlo (MC) simulation of the physics process under study, modeled within the specific environment of SuperKEKB and the Belle II detector. The MC simulation used in Belle II is also *run-dependent*, meaning that beam and detector conditions are varied across different runs to more accurately reflect the changing experimental environment, rather than assuming static conditions. Different MC generators implemented within BASF2 are used to simulate the various processes, depending on the underlying physics. For the decays of B and D mesons to exclusive final states, EVTGEN 1.3 is used [83], while PYTHIA 8.2 [84] is used for inclusive meson decays. Continuum events, defined as $e^+e^- \rightarrow q\bar{q}$ with $q = \{u, d, c, s\}$, are simulated using KKMC 4.15 [85] for the initial production, in combination with PYTHIA, which models the subsequent fragmentation and hadronisation processes. KKMC

is also used to simulate the production of τ pairs, while TAUOLA [86] is used to model how the τ leptons decay. For high cross section processes, such as $e^+e^- \rightarrow e^+e^-(\gamma)$ and $e^+e^- \rightarrow \gamma\gamma(\gamma)$, the BABAYAGA [87, 88] generator is used, while for $e^+e^- \rightarrow e^+e^-e^+e^-$ and $e^+e^- \rightarrow e^+e^-\mu^-\mu^+$, the AAFH [89] generator is used. Moreover, initial and final state radiation is simulated using the PHOTOS [90] package. Lastly, detector hits due to beam-induced background events are simulated using SAD [91], while the Belle II detector geometry is simulated with GEANT4 [92], which also models its interaction with the generated particles.

In all simulated data samples, information relating to the generated decay is available and is made possible through the so-called “truth-matching” process, which matches reconstruction-level objects like particle tracks and calorimeter cluster to their generated-level counterparts [93]. This is done by applying requirements that ensure a reconstructed track or calorimeter cluster can be predominantly attributed to a single generated particle, so that only sufficiently pure objects are considered truth-matched. Composite candidates then inherit the match status of their constituents according to defined rules, ensuring consistency throughout the reconstruction hierarchy. Truth-matching allows to validate one’s analysis strategy and extensively investigate certain physics background processes in an effort to better understand and characterise them.

2.6 Reconstruction

Reconstruction is the process through which the large amount of raw data recorded by all the Belle II sub-detectors is transformed into high-level physics objects, such as charged particle tracks, neutral clusters, and identified particle types, that can be used for further analysis. The following sections describe the aspects of Belle II reconstruction that are the most relevant for the measurement at the center of this thesis project, all of which are performed using BASF2.

2.6.1 Tracking

Precise reconstruction of charged particle trajectories is a crucial aspect of most physics analyses in Belle II, and it is a necessary ingredient in momentum and decay vertex reconstruction. Moreover, due to the lower boost in SuperKEKB compared to its predecessor, Belle II track reconstruction needs to be more precise compared to Belle, in order to reach a resolution on the Δt measurement that is comparable. Track reconstruction in Belle II begins with the identification and association of hits in the VXD and CDC produced by charged particles traversing these detectors, which is a process known as *track finding*. This is followed by *track*

fitting, in which a fit to the associated hit positions is performed to determine the particle's trajectory [94].

For the purpose of track finding, two standalone pattern recognition algorithms are used, one for the VXD and one for the CDC, and their results are subsequently combined to form track candidates. The VXD track finding algorithm is based on the Cellular Automaton (CA) model. Track segments are built from pairs of hits in adjacent layers of the SVD, but only if they are deemed compatible by consulting a look-up table, called the sector map, which has been constructed using a large amount of simulated SVD hits. The sector map can vary depending on the particle's momentum. Track segments that pass a set of geometrical requirements are combined into track candidates, and a Hopfield network with a quality indicator is employed to ensure that the resulting track candidates do not overlap. PXD hits are then combined with the track candidates formed solely from SVD hits in order to improve the impact parameter resolution of the track. CDC track finding consists of two parts; a global and a local algorithm. While the local algorithm is based on a CA to identify track segments, similarly to the VXD algorithm, the global algorithm applies a Hough transformation to all the hit positions and searches for intersections in the Legendre space. Both track finding algorithms achieve an excellent track reconstruction efficiency across the momentum spectrum, even for tracks not originating from the IP.

The track candidates from the VXD and CDC are then combined using a Combinatorial Kalman Filter (CKF) and are subsequently fitted to extract the track parameters using the GENFIT2 package [95]. In the Belle II tracking software, five parameters are considered to describe the trajectories of particles that propagate in vacuum and in a constant magnetic field. They are:

- d_0 : the signed distance of the perigee⁹ from the origin in the transverse plane,
- z_0 : the longitudinal signed distance of the perigee from the origin,
- ϕ_0 : the angle between the transverse momentum at the perigee and the x axis,
- $\tan\lambda$: the tangent of the angle between the momentum at the perigee and the transverse plane, and
- ω : the curvature, where the sign corresponds to the charge of the particle.

Due to inhomogeneities in the magnetic field of the solenoid, as well as interactions between the charged particles and the detector material leading to multiple scattering, the trajectories of charged particles in Belle II are not perfect helices. Nevertheless, both effects are accounted for in the track fitting algorithm.

⁹The perigee is the point of closest approach to the origin of the r/ϕ plane.

2.6.2 Charged Particle identification

Efficient charged particle identification is necessary for isolating hadronic final states (i.e. $B^0 \rightarrow D^{(*)-}\pi^+$ and $B^0 \rightarrow D^{(*)-}K^+$), for reducing backgrounds and is also useful for Flavor Tagging. PID information is obtained from the dedicated PID sub-detectors, TOP and ARICH, as well as from dE/dx measurements from CDC and SVD. Moreover, ECL and KLM are the primary sources for electron and muon identification, respectively.

The PID information from each sub-system is analysed independently and is used to calculate a likelihood $\mathcal{L}_\alpha^{det}$ for each charged particle hypothesis, namely: e , μ , π , K , p , d . The individual likelihoods are combined into an overall likelihood for a given hypothesis α :

$$\mathcal{L}_\alpha = \prod_{det} \mathcal{L}_\alpha^{det} . \quad (2.4)$$

The overall likelihood can then be combined into global PID ratios:

$$\text{PID}_\alpha = \frac{\mathcal{L}_\alpha}{\sum_\beta \mathcal{L}_\beta} , \quad (2.5)$$

or binary PID ratios:

$$\text{PID}(\alpha|\beta) = \frac{\mathcal{L}_\alpha}{\mathcal{L}_\alpha + \mathcal{L}_\beta} , \quad (2.6)$$

which can be used in physics analyses directly.

2.6.3 K_S^0 reconstruction

Aside from reconstructing charged particle trajectories, the reconstruction of neutral particles which are relatively long-lived and decay inside the detector volume, like K_S^0 mesons and Λ baryons, is also crucial. These particles, also referred to as V^0 's, usually decay into two charged particles and their decay positions are reconstructed using a dedicated algorithm. Pairs of particle tracks with opposite electric charge are combined and processed by the RAVE vertex reconstruction algorithm [96]. Combinations for which the vertex fit fails are rejected, and selection criteria are applied on the accepted V^0 's, depending on the vertex fit χ^2 and, for vertices located inside the beam pipe, the mass window.

2.6.4 Photon reconstruction

The energy of particles like photons, electrons and light hadrons ($\pi^\pm$, $K^\pm$) is determined using the clusters they form in the ECL through particle showers. A dedicated algorithm is used to reconstruct the cluster energy from energy deposits

in the calorimeter crystals depending on their positions and the amount of energy deposited in them. Each cluster is then matched to a charged particle track if the latter is consistent in time and space with a crystal hit. Otherwise, if the energy of the cluster decreases exponentially with the distance from the incident axis and if the shape of the cluster is cylindrically symmetric in the lateral direction, the cluster is considered to have originated from a photon. In order to separate between photons coming from $\pi^0 \rightarrow \gamma\gamma$ decays and ones coming directly from the B decay, it is typical to use a multivariate analysis, or MVA, classifier.

2.6.5 B meson reconstruction

In cases where the final state of the B decay of interest is known, meaning there are little to no missing particles like neutrinos in the final state, the B meson is reconstructed in each event using the daughter particles which have been reconstructed individually beforehand. Several combinations of the daughter particle candidates are possible, leading to multiple B candidates per event. For each one, its energy and momentum are calculated using the daughter particles' momentum and cluster energy, alongside other significant variables, like the energy difference, ΔE , and the beam-constrained mass, M_{bc} . They are defined as follows:

$$\Delta E \equiv E_B^* - \sqrt{s}/2, \quad (2.7)$$

$$M_{bc} \equiv \sqrt{(\sqrt{s}/2)^2 - p_B^{*2}}. \quad (2.8)$$

ΔE is defined as the difference between half of the beam energy in the C.o.M. frame, $E_{\text{beam}}^* \equiv \sqrt{s}/2$, and the reconstructed B meson energy, in the same frame. Signal events usually peak near 0 GeV in ΔE , while background events either peak at different values or are completely flat in ΔE , depending on the underlying physical process. The resolution in ΔE is typically dominated by the uncertainty of the measured B meson energy, especially when photons are present in the final state. Further details about the various contributions to the ΔE distribution and how they can be modeled will be given in Sections 4.2.3 and 5.3.

M_{bc} is defined similarly to the invariant mass of the B meson, but uses half the beam energy in place of the reconstructed B meson energy in the mass formula. This substitution significantly improves the resolution near the peak of the distribution, at approximately 5.28 GeV/ c^2 , as the beam energy is known with much greater precision than the reconstructed B meson energy. More details about the signal and background contributions to the M_{bc} distribution and how they can be modeled will be given in Section 5.3.

Both of these variables offer strong discriminating power between signal and background and are unique to e^+e^- colliders. Examples of ΔE and M_{bc} distributions for simulated signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ events and continuum background

events are shown in Figures 2.15 and 2.16, respectively. In Belle II analyses, it is typical to either apply selection criteria on these variables or include them in fits with the aim of isolating the signal contribution within a distribution of interest.

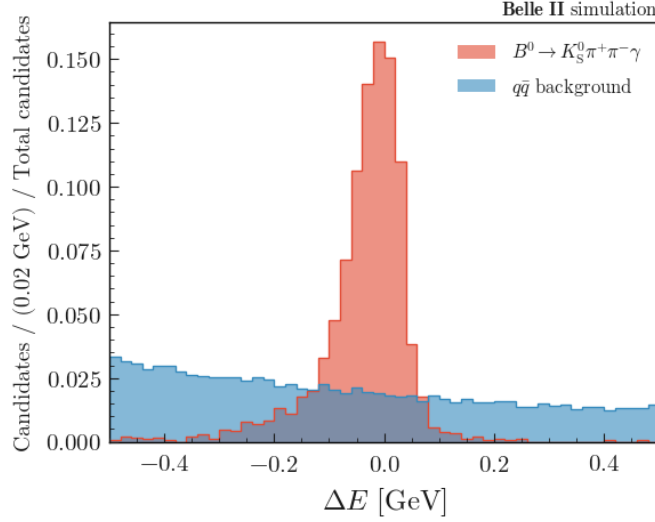


Figure 2.15: ΔE distribution for simulated signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ and continuum background events.

2.6.6 Vertex reconstruction

The decay position, or vertex, of the signal-side B meson (B_{sig}) in each event is determined by performing a kinematic fit of all the charged particle trajectories, V^0 's and calorimeter clusters that have been used to reconstruct it. The TREEFITTER [97] package is used for this purpose, which was originally developed by the BaBar experiment. While it is typical for decay chains to be reconstructed starting with the lowest level objects, such as tracks and clusters, to build higher level objects, like vertices, until the head of the decay is reached, TREEFITTER follows an alternate approach of parameterising and fitting the decay globally. This leads to an improvement in the vertex resolution, especially in decay channels rich in neutral or missing particles, due to the fact that information is shared across the tree. After the B_{sig} vertex fit is performed, the energy and momentum of the B_{sig} are updated, which leads to slight shifts in the values of ΔE and M_{bc} , compared to their values before the vertex fit was performed.

In order to be sensitive to time-dependent CP violating effects, the decay position of the tag-side B meson (B_{tag}) also needs to be determined with good enough precision (see Section 1.4.2). After fitting the B_{sig} decay vertex, a separate fit to

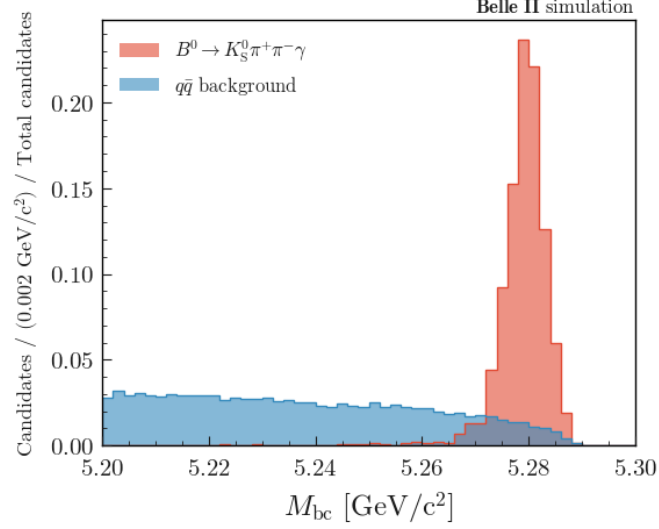


Figure 2.16: M_{bc} distribution for simulated signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ and continuum background events.

determine the B_{tag} decay position is performed using all tracks and clusters in the event that were not used for the B_{sig} vertex fit. The RAVE algorithm [96] is used for this purpose. This is because in this case the decay chain is not reconstructed explicitly and all tracks are instead fitted inclusively with the assumption that they originate from the same point. RAVE employs an iterative fitting approach where the weights assigned to each track are adapted at each iteration based on their compatibility with the vertex hypothesis. This allows to better constrain the B_{tag} vertex position, reducing as much as possible the effect that tertiary tracks coming from D mesons have on the B_{tag} vertex determination. Still, the resolution on the final Δt measurement¹⁰ is dominated by the tag-side B vertex fit in most cases. More details about how the Δt resolution is modeled in a physics analysis context are given in Section 4.2.4.

¹⁰as determined from $\Delta t = \Delta z / \beta \gamma c$, where $\Delta z = z_{\text{sig}} - z_{\text{tag}}$ is the signed difference between the B_{sig} and B_{tag} decay vertex positions in the z direction.

Chapter 3

VTX Performance Study

Contents

3.1	The case for the upgrade	55
3.2	The upgraded vertex detector – VTX	58
3.3	VTX performance study	60
3.3.1	Reconstruction efficiency	62
3.3.2	Vertex resolution	64
3.3.3	Sensitivity study	66
3.4	Conclusion and outlooks	66

This chapter describes a study performed in the context of the optimisation of the design of the VXD detector upgrade. The measurement of the time-dependent CP asymmetry of $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays offers interesting benchmarks to evaluate how the upgrade proposal is performing, in terms of both B vertex and V^0 particle reconstruction. The general questions posed were to evaluate the impact on experimental performance of the number of detection layers, their position, the integration time, the material budget and the detection acceptance. The findings of this study were included in the Physics Performance chapter of the Belle II Detector Upgrades Framework Conceptual Design Report (CDR) [98], and will be influential in deciding which geometry configuration to use for the upgraded vertex detector.

3.1 The case for the upgrade

As discussed in Sections 2.1 and 2.2, the plan for SuperKEKB is to reach ~ 40 times the maximum instantaneous luminosity that KEKB reached and for Belle II

to collect more than 50 times the data that Belle collected. However, this drastic increase in the collision rate does not come without cost, as the rate of parasitic particles produced by the beams will also increase, posing a significant challenge for the sub-detectors, particularly the ones closest to the IP, like the PXD and SVD, as they will be exposed the highest density of incoming particles. The PXD and the CDC in particular have been shown to suffer the most from the high-occupancy environment caused by various kinds of beam-induced background processes. There are two main types of beam-induced backgrounds; those due to beam-beam interactions and those produced by the beam itself while traveling in the accelerator ring. These beam-induced background processes generally produce low momentum particles, that curl in the innermost layers, either directly hitting the detector or producing showers in the material surrounding the interaction region, which causes hits at the outermost detector layers. The following paragraphs discuss the main sources of beam-induced background and their expected evolution from the KEKB to the nominal SuperKEKB conditions.

Synchrotron radiation

Synchrotron radiation¹ emitted from the beam interacts with the material of the beam pipe as well as that of the innermost detectors, PXD and SVD, damaging them in some cases. It is therefore important to characterise this radiation and to take appropriate measures when designing the beam optics, in order to avoid damaging the vertex detectors. Simulation studies show little to no hits due to synchrotron radiation with the current design of the Interaction Region (IR). This is partly due to the 20 μm thick gold layer covering the inner wall of the beam pipe, which helps stop a significant fraction of synchrotron radiation. Moreover, due to the different IR design, this background is expected to be significantly reduced in SuperKEKB, compared to KEKB.

Beam-gas scattering

Bremsstrahlung and Coulomb scattering between the beam particles and the residual gas in the beam pipe alters the momenta of beam particles and causes them to hit the walls of the vacuum chamber and magnets, leading to particle showers. The size of this background depends on the beam current, the pressure of the vacuum in the ring, as well as the strength of the focusing magnets. A background dose with a similar order of magnitude is expected in SuperKEKB compared to KEKB, since the beam currents will be ~ 2 times higher and the vacuum and overall magnetic field strength will be the same as in KEKB.

¹Synchrotron radiation is emitted by charged particles when they are accelerated in a magnetic field.

Touschek scattering

Touschek scattering refers to scattering between the particles in the beam, which changes their momenta so that they hit the walls of the vacuum chamber and magnets, leading to particle showers. The size of this background is proportional to the current of the beam bunches, the number of bunches, and the inverse of the beam size. Due to the nano-beam scheme, the beam size in SuperKEKB is expected to reach $\sim 1/20$ that of KEKB, so Touschek scattering will be a major source of beam-induced background at nominal SuperKEKB luminosity levels. Nevertheless, Touschek scattering is not the main source of beam-induced background in the detector acceptance, which was the case at KEKB.

Radiative Bhabha scattering

The rate of radiative Bhabha events increases with the instantaneous luminosity. This process has a large cross section and is the dominant source of machine-induced background. In these events, photons from the radiative Bhabha process interact with the iron in the magnets, producing a copious amount of neutrons *via* the giant photo-nuclear resonance mechanism. These neutrons constitute the main background source for the KLM. In addition, since the electron and positron energies are decreased in this process, the particles will be significantly bent by the final focusing magnets causing them to hit the magnet walls, leading to the production of electromagnetic showers. However, this is expected to be less relevant in SuperKEKB compared to KEKB, due to the fact that two separate quadrupole magnets are used for incoming and outgoing beams, instead of a shared one. Moreover, the produced particles tend to be produced at low polar angles along the beam pipe, landing outside of the detector acceptance.

Electron-positron pair production *via* two-photon process

Since the pixel detector is placed in close proximity (1.4 cm) to the IP, the background due to very low momentum electron-positron pairs which are produced *via* the two-photon process $e^+e^- \rightarrow e^+e^-e^+e^-$ at the interaction spot of the beams will be significant. For the first layer of the PXD, a maximum hit rate of 670 million tracks per second is expected, corresponding to ~ 14000 e^+e^- pairs being produced per event from this process. This is the dominant source of beam-induced background in the detector acceptance at SuperKEKB by a significant margin.

To maintain the same level of data quality under these demanding conditions, and to ensure the long-term safety and operational integrity of the innermost detector components, a significant upgrade to the vertex detector will be required. This upgrade is essential not only for preserving tracking and vertexing performance in

high-occupancy environments, but also for protecting the most vulnerable components from radiation damage and excessive hit rates as SuperKEKB approaches its target instantaneous luminosity.

3.2 The upgraded vertex detector – VTX

The baseline proposal for the vertex detector upgrade is a replacement of the PXD and SVD sub-detectors, with a fully integrated vertex detector geometry that is based on monolithic CMOS pixel sensors, called VTX [98]. The geometry which was considered at the time of performing this study consisted of 5 layers, compared to the 6 of the current VXD, while covering the same radius. Tables 3.1 and 3.2 list the specifications of the current and upgraded vertex detector geometries which are being considered, respectively. Figures 3.1 shows these geometries as they are simulated by the GEANT4 software in BASF2.

Layer	Radius (mm)	Ladders	Sensors/ladder	$(r - \phi, z)$ pitch (μm)
1	14	8	2	(60, 50)
2	22	12	2	(60, 50)
3	39	7	2	(160, 50)
4	80	10	3	(160, 50)
5	104	12	4	(160, 50)
6	135	16	5	(160, 50)

Table 3.1: Specifications of the ladders and sensors of the VXD (PXD+SVD). Information taken from [69, 76].

Layer	Radius (mm)	Ladders	Sensors/ladder	$(r - \phi, z)$ pitch (μm)
1	14.1	6	4	(35, 35)
2	22.1	10	4	(35, 35)
3a	39.1	17	7	(35, 35)
3b	69.1	30	12	(35, 35)
4	89.5	40	16	(35, 35)
5	140	62	24	(35, 35)

Table 3.2: Specifications of the ladders and sensors of the proposed VTX. The 3a (3b) layer corresponds to the nominal (alternative) VTX geometry. Information taken from [98].

The main benefit of this fully pixelated vertex detector is the significantly smaller pitch of its sensors, resulting in a higher granularity, so it will not suffer from as high an occupancy as the PXD and SVD would. This improvement

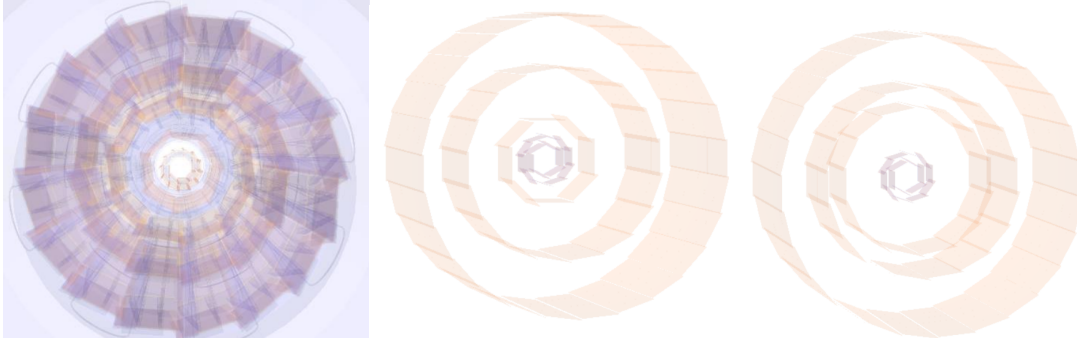


Figure 3.1: Transverse view of vertex detector geometries in simulation. Left is the current VXD, together with cooling and mechanical support structure, center is the nominal 5-layer VTX upgrade, and right is an alternate version of the 5-layer VTX geometry with the third layer placed 30 mm closer to the fourth layer. The diameter of the vertex detector in each schematic is ~ 28 cm.

holds true despite the fact that the SVD features a much faster readout than the proposed VTX. Moreover, when compared to the granularity and integration time of the PXD, the improvement is much more substantial, making it much easier to distinguish between hits coming from beam-induced background processes and ones coming from physical processes of interest. An additional advantage is the integration of functionality into a single sub-detector, rather than maintaining two separate systems, which greatly simplifies coordination and overall system management.

An alternative VTX geometry is also considered with the aim of improving the reconstruction efficiency of K_S^0 mesons, which appear in a significant fraction of B meson decays *via* the $b \rightarrow c \rightarrow s$ quark-level transition, including the decay mode central to this thesis research. In fact, efficiently identifying and reconstructing these K_S^0 particles is one of the main assets of the Belle II experiment, compared to other experiments of its kind. Given their proper lifetime of ~ 90 ps, K_S^0 mesons can travel several centimeters before decaying within the Belle II detector. As a result, the efficiency of reconstructing these particles is expected to improve significantly by shifting the third detection layer closer to the fourth layer of the VTX. Figure 3.2 shows a schematic that demonstrates the expected improvement in K_S^0 reconstruction efficiency which results from this displacement.

Since at least three hits are required to form a track, displacing the third layer of the 5 layer geometry outwards should recover a significant fraction of K_S^0 mesons, in this case reconstructed as $K_S^0 \rightarrow \pi^+\pi^-$, which would otherwise be missed. This is because, in the alternative VTX geometry, a larger fraction of K_S^0 mesons decay before reaching the third layer compared to the nominal one, increasing

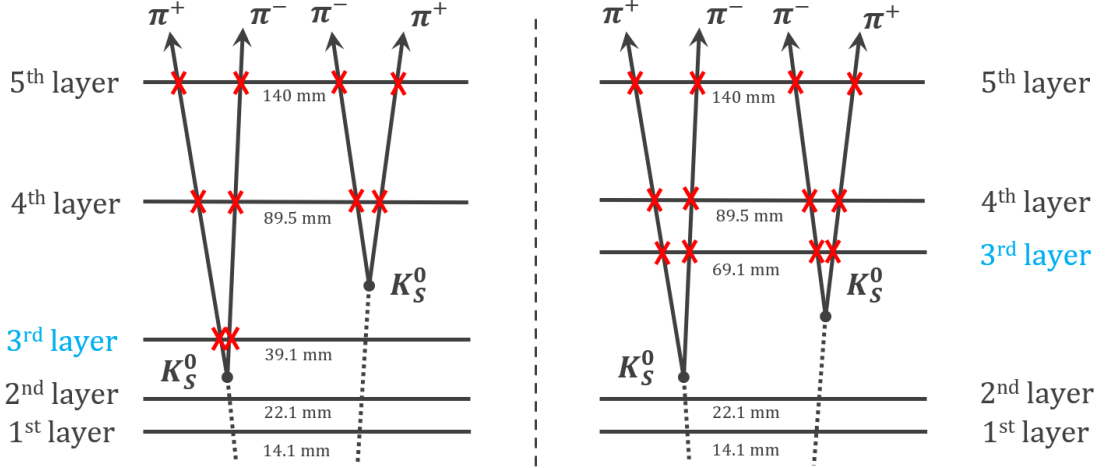


Figure 3.2: Schematic drawing demonstrating the expected improvement in K_S^0 reconstruction efficiency, as a result of displacing the third layer of the VTX. The left drawing represents the nominal proposed VTX geometry and the right drawing represents the alternative one.

the likelihood of correctly reconstructing both charged pion tracks. Part of my work involved the implementation of this alternative VTX geometry into BASF2, so that its performance could be compared to the nominal VTX geometry, as well as the current VXD, both of which were already implemented in the simulation framework.

It is also important to note that both of the VTX geometries were implemented in the simulation framework without their respective cooling and mechanical support structures, as their design had not yet been fully determined and optimised. Therefore, some deterioration to the reconstruction efficiency is expected in the current VXD due to the additional material budget, though not too large.

3.3 VTX performance study with simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays

The goal of this study is to evaluate the physics performance of the upgraded vertex detector using simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays. To this end, a MC sample containing half a million signal events² was produced for each vertex detector

²Signal events are events where one of the two neutral B mesons produced from the $\Upsilon(4S)$ in each event is forced to decay to the final state of interest, $K_S^0 \pi^+ \pi^- \gamma$, while the other B meson is allowed to decay generically, according to the measured branching fractions of the known decays of B mesons, as they are listed in the Belle II .dec file.

geometry, namely the VXD, the nominal VTX and the alternative VTX. For each sample, signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ events are generated using EVTGEN with the resulting hits in the Belle II detector simulated using GEANT4 inside BASF2.

The beam-induced backgrounds described in Section 3.1 are modeled using GEANT4 and SAD, and so-called “BG overlay” files are created, which are then combined with simulated hits in the detector that result from events containing the previously generated signal decay to make the signal MC samples as realistic as possible. Three different levels of beam-induced background, all at the nominal instantaneous luminosity of SuperKEKB, $6 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$, are also considered. Namely an optimistic, an intermediate, and a pessimistic background (BG) scenario are evaluated, based on projected occupancy levels in the CDC. Figure 3.3 shows the degradation of the tracking efficiency with increasing levels of beam-induced background for each of the tracking sub-detectors, as well as the VTX. The CDC is the least robust against the increasing beam-induced background, while the VTX shows almost no degradation in terms of tracking efficiency.

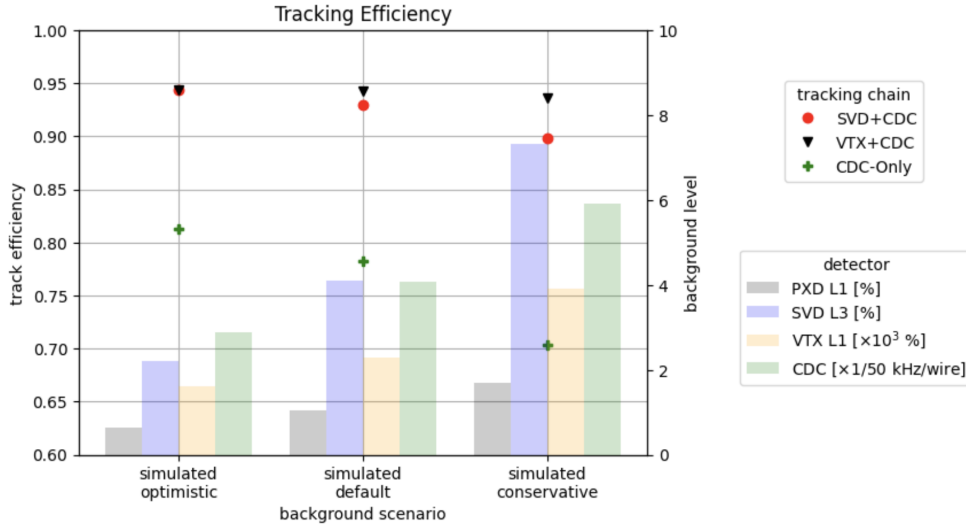


Figure 3.3: Tracking efficiencies (colored markers) for the SVD+CDC, VTX+CDC and CDC-only detector configuration in each of the considered beam-induced background scenarios and the corresponding occupancy levels (colored bars) in the first layer of the PXD, in the third layer of the SVD, in the first layer of the VTX and in the CDC. Taken from [98].

The full B decay chain in each event is then reconstructed and the signal and tag-side B decay vertices are fitted to obtain the B_{sig} and B_{tag} decay positions, respectively. A set of selection criteria is then applied to these samples which is close to the one used in the time-dependent CP asymmetry measurement per-

formed using Belle II data in this thesis (see Section 5.2 for more details). These criteria are not exactly the same ones used in the main analysis, since the selection was not yet fully optimised, but this does not impact this performance study in any way, as the same criteria are used for each geometry configuration.

Separate signal MC samples are produced for each detector geometry and BG scenario, in order to evaluate the performance of the VTX compared to the VXD, in the context of measuring the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays. For this study, a full-scale analysis is not performed with each sample, but rather two crucial metrics are evaluated; the *reconstruction efficiency* and the *decay vertex resolution* for each vertex detector geometry and BG scenario. The findings of the study are presented in the following Sections.

3.3.1 Reconstruction efficiency

Firstly, I estimated the efficiency of reconstructing K_S^0 mesons from a pair of charged pions ($K_S^0 \rightarrow \pi^+ \pi^-$), which is a necessary step on the way to reconstruct the full $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decay in each event. Here, the K_S^0 reconstruction efficiency is defined as the ratio of the number of events in which a K_S^0 meson from the signal-side B decay is correctly reconstructed, after applying all of the selection criteria, to the total number of generated events containing a K_S^0 in the signal-side B decay, which is 500,000 for each MC sample. The results are listed in Table 3.3 for each vertex detector geometry and BG scenario.

Geometry	Reconstruction Efficiency (%)		
	Optimistic BG	Intermediate BG	Pessimistic BG
VXD	64.044 ± 0.068	60.234 ± 0.069	52.380 ± 0.071
Nominal VTX	56.947 ± 0.070	53.281 ± 0.071	46.915 ± 0.071
Alternate VTX	60.955 ± 0.069	57.605 ± 0.070	51.781 ± 0.071

Table 3.3: Reconstruction efficiency of $K_S^0 \rightarrow \pi^+ \pi^-$ for each of the three vertex detector geometries in three different BG scenarios.

From these numbers, it is clear that a significant fraction of K_S^0 's are lost if the third layer of the VTX is not displaced. This is demonstrated more clearly by comparing the K_S^0 reconstruction efficiency of each vertex detector geometry as a function of the transverse distance the K_S^0 travels before decaying (see Figure 3.4).

The point where the K_S^0 reconstruction efficiency drops corresponds to the radius of the third-to-last layer for each of the vertex detector geometries. The reason for this is easily understood by considering the fact that the charged pion tracks coming from K_S^0 's which decay past that point will each lead to only two hits in the subsequent layers, making it much harder to reconstruct them correctly.

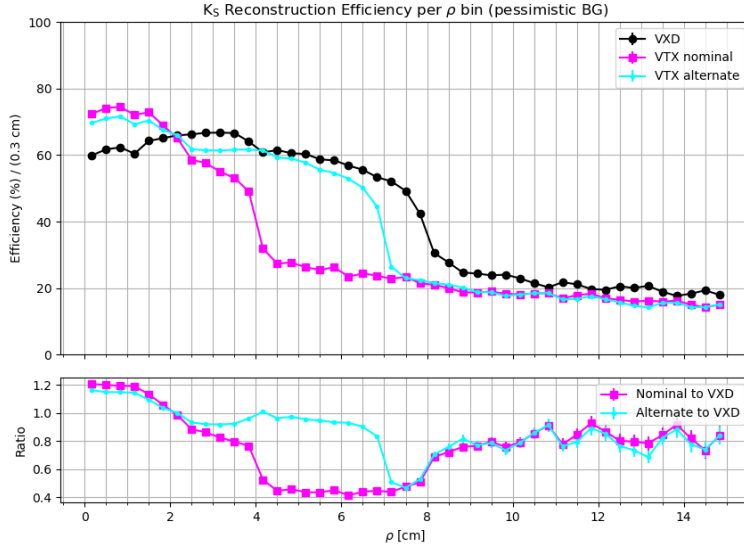


Figure 3.4: K_S^0 reconstruction efficiency as a function of the transverse distance traveled by the K_S^0 . Here, the pessimistic BG scenario is taken as an example.

It is also worth noting that the rate with which the reconstruction efficiency degrades as the hit rate due to beam-induced background increases is slower in the upgraded geometries than in the current vertex detector. This once again demonstrates the higher robustness of both VTX geometries, compared to the VXD, against the harsh beam-induced background conditions caused by the high instantaneous luminosity of SuperKEKB.

Thanks to both the increased robustness of the VTX geometries and the outward shift of the third layer, the K_S^0 reconstruction efficiency in the alternative VTX geometry is nearly equal to that of the VXD in the pessimistic BG scenario. In all other cases, however, the VXD outperforms the VTX in this aspect. This is attributed to the additional detection layer in the VXD, which improves the reliability of reconstructing V^0 particles such as the K_S^0 meson.

Next, I evaluated the efficiency of fully reconstructing the B^0 (and $\bar{B}^0$) in each event using a K_S^0 , two charged pion tracks and a hard photon. Here, the B^0 reconstruction efficiency is defined as the ratio of the number of events in which the signal-side B decay is correctly reconstructed using its daughter particles, after applying all of the selection criteria, to the total number of generated events containing the signal-side B decay, which again is 500,000 for each MC sample. The results are listed in Table 3.4 for each vertex detector geometry and BG scenario.

Geometry	Reconstruction Efficiency (%)		
	Optimistic BG	Intermediate BG	Pessimistic BG
VXD	36.636 ± 0.068	32.127 ± 0.066	21.571 ± 0.058
Nominal VTX	33.970 ± 0.067	29.972 ± 0.065	21.039 ± 0.058
Alternate VTX	35.904 ± 0.068	31.940 ± 0.066	23.020 ± 0.060

Table 3.4: Reconstruction efficiency of $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ for each of the three vertex detector geometries in three different BG scenarios.

The robustness of the VTX geometries is once again demonstrated. Notably, under the pessimistic BG scenario, the alternative VTX geometry achieves a reconstruction efficiency which is higher than that of the current VXD, despite having one fewer detection layer.

3.3.2 Vertex resolution

The second evaluation metric for this study is the B decay vertex resolution, meaning how well the decay position of the B meson is determined by fitting the tracks of its daughters, for each vertex detector geometry in each of the BG scenarios. In each case, the B_{sig} vertex is fitted using TREEFITTER, with the constraint that the momentum of the B originates from the IP. The resolution of the B decay vertex measurement is quantified using the so-called Δz residual, which is defined as $\Delta z_{\text{res}} = \Delta z_{\text{reco}} - \Delta z_{\text{true}}$, where $\Delta z = z_{\text{sig}} - z_{\text{tag}}$ is the signed difference between the reconstructed decay positions of B_{sig} and B_{tag} along the z -direction, in each event. This residual can only be evaluated in simulation, where the true decay vertex positions are known. Similarly, the Δt resolution is quantified using the Δt residual, $\Delta t_{\text{res}} = \Delta t_{\text{reco}} - \Delta t_{\text{true}}$, since $\Delta t = \Delta z / \beta \gamma c$. Figure 3.5 shows the Δt residual distributions of the signal MC samples under study, overlaid for each vertex detector geometry, as well as the ratio between them. This is shown for each BG scenario considered. From these comparisons, it is clear that there is a significant improvement in the resolution in the VTX geometries compared to the VXD. This is expected from the higher granularity of the VTX geometries with respect to the VXD, especially when compared to the PXD layers.

The same Δt residual distributions are shown in Figure 3.6, but the comparison here is made between the different BG scenarios instead and it is shown for each of the three vertex detector geometries. From this set of comparisons, the Δt resolution achieved with the VTX is shown to be robust against the increasing levels of beam-induced background, while the resolution obtained with the VXD shows significant degradation, as expected.

A more quantitative picture is obtained by fitting the Δt residual distributions

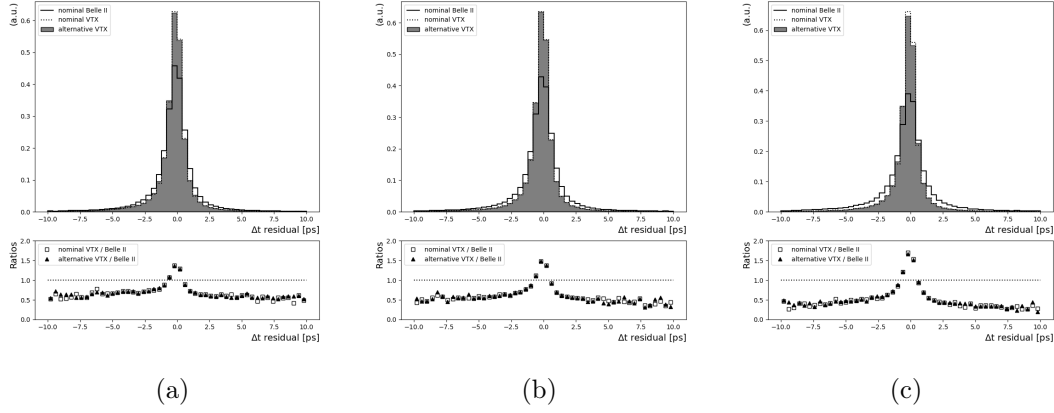


Figure 3.5: Overlaid Δt residual distributions obtained for each vertex detector geometry, and the ratios between them, shown for the (a) optimistic, (b) intermediate and (c) pessimistic BG scenarios.

for each vertex detector geometry and BG scenario, using a sum of three Gaussian distributions. The weighted average of the three Gaussian widths is then used as a quantitative indicator of the Δt resolution for each of the nine configurations (3 geometries $\times$ 3 BG scenarios). Table 3.5 lists the weighted averages obtained from these fits, for each configuration. Figure 3.7 shows as an example the fits to the Δt residual distribution for each of the three vertex detector geometries, in the pessimistic BG scenario.

Geometry	Δt resolution indicator (ps)		
	Optimistic BG	Intermediate BG	Pessimistic BG
VXD	1.098 ± 0.054	1.169 ± 0.063	1.269 ± 0.070
Nominal VTX	0.863 ± 0.034	0.845 ± 0.033	0.819 ± 0.039
Alternate VTX	0.869 ± 0.029	0.843 ± 0.032	0.826 ± 0.037

Table 3.5: Δt resolution for each vertex detector geometry and BG scenario. Here, the weighted average of the three Gaussian widths obtained from fitting the Δt residual distribution using a sum of three Gaussians is quoted as a quantitative indicator of the Δt resolution.

The resulting Δt resolution values are also depicted in Figure 3.8, clearly illustrating the degradation in performance of the current VXD under increasing levels of beam-induced background, as well as the robustness of the proposed VTX geometries. The figure also highlights the differences in Δt resolution achieved by each geometry in a given BG scenario.

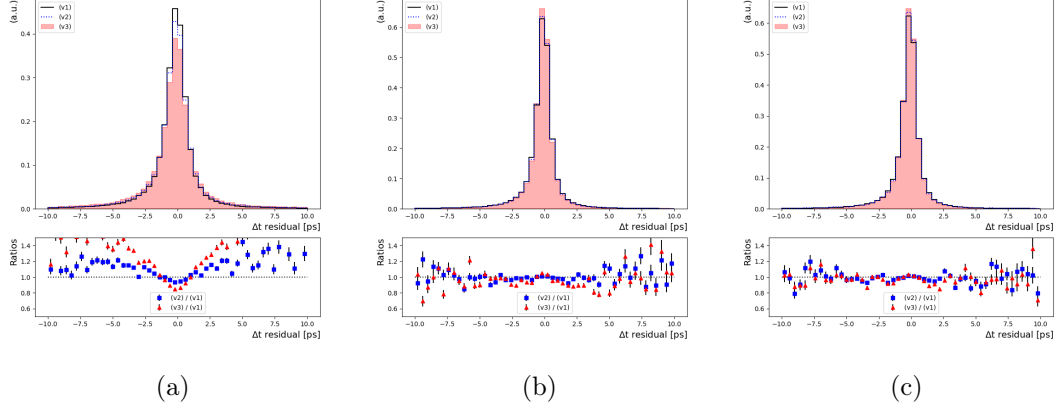


Figure 3.6: Overlaid Δt residual distributions obtained for each BG scenario, and the ratios between them, shown for the (a) VXD, (b) nominal VTX and (c) alternative VTX geometries. Here, v1, v2 and v3 correspond to the optimistic, intermediate and pessimistic BG scenarios, respectively.

3.3.3 Sensitivity study

Lastly, I performed a sensitivity study using toy MC samples to estimate the potential improvement expected in the sensitivity of the time-dependent CP asymmetry measurement resulting from the vertex detector upgrade. In particular, I estimated the expected improvement on the sensitivity of the S parameter and I found a relative improvement of $\sim 15\%$ between the VTX and VXD geometries. This is effectively equivalent to increasing the size of the dataset used to perform this kind of measurement by $\sim 32\%$, just by upgrading the vertex detector.

3.4 Conclusion and outlooks regarding the VTX performance study

At the time of writing this thesis, this is one of the main studies performed in Belle II that demonstrates the capabilities of the proposed vertex detector upgrade, in the context of a real physics analysis, which led to its inclusion in the CDR document. Most of the expected improvements are confirmed in this study, mainly the improvement in decay vertex resolution and the robustness against higher levels of beam-induced background, which makes the VTX proposal highly promising.

It is important to keep in mind, however, that such improvements in the resolution do not affect all time-dependent measurements the same way. For example,

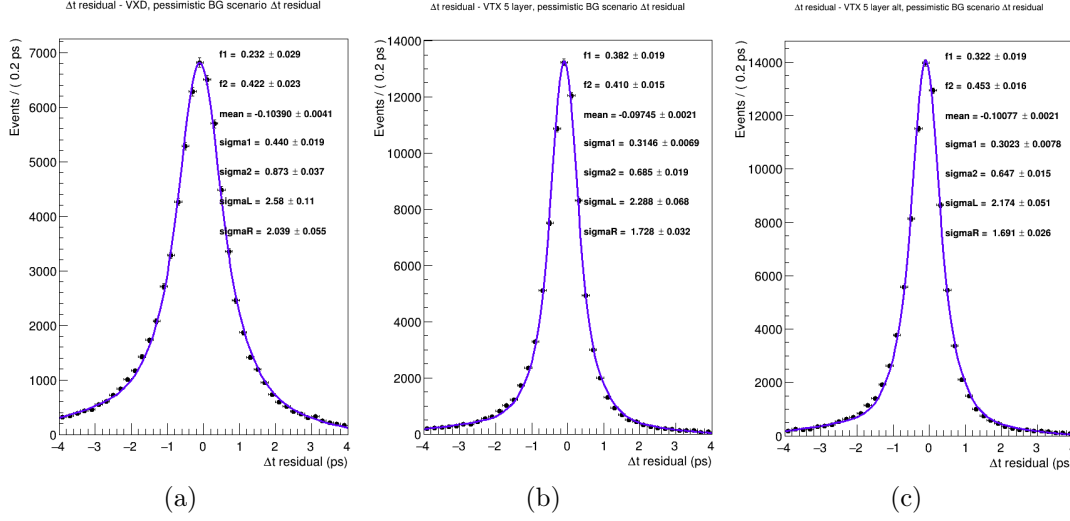


Figure 3.7: Fit of the Δt residual distribution using the sum of three Gaussians for the (a) VXD, (b) nominal VTX and (c) alternative VTX geometries in the pessimistic BG scenario.

when measuring the $\sin 2\beta$ parameter with $B^0 \rightarrow J/\psi K_S^0$ decays, the main driving factor is the observed number of signal events, while, for more rare decays, like $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$, the sensitivity depends just as much on the observed number of signal events, as on other factors like flavor tagging and decay vertex resolution. So, while a factor ~ 2 improvement in the resolution of the vertex detector might not do much for the sensitivity of the former measurement, it can lead to a significant increase in precision in the latter measurement.

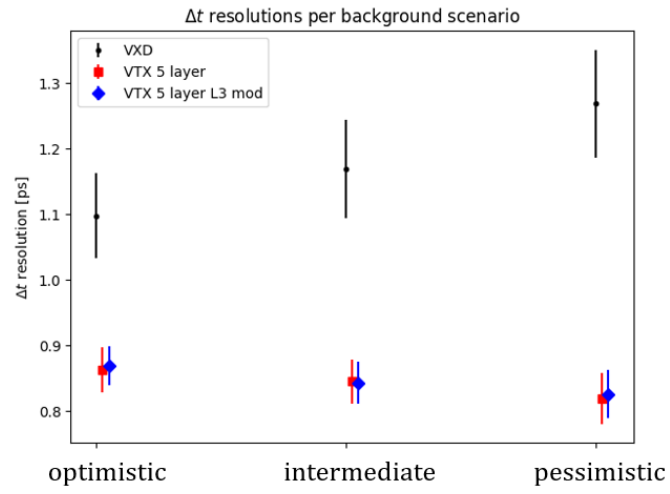


Figure 3.8: Δt resolution indicators for each vertex detector geometry across the three BG scenarios.

Chapter 4

Calibration of the Flavor Tagger and decay time resolution parameters

Contents

4.1 Flavor Tagging at Belle II	69
4.2 Flavor Tagger and Resolution Function calibration	73
4.2.1 Principle behind the calibration of the GNN-based Flavor Tagger	73
4.2.2 Reconstruction and selection of $B^0 \rightarrow D^{(*)-}\pi^+$ decays and selection criteria	77
4.2.3 Estimation of signal and background yields	79
4.2.4 Interlude: Fit of the Δt residual distribution of simulated signal events	89
4.2.5 Production of the background-subtracted Δt distribution with <code>sWeights</code>	92
4.2.6 Fit of the signal Δt distribution	94
4.3 Conclusion and outlooks	108

This chapter elaborates on the concept of “Flavor Tagging” and describes the procedure of evaluating the performance of the latest Flavor Tagging algorithm used in Belle II.

4.1 Flavor Tagging at Belle II

The flavor of a B meson (either B^0 or $\bar{B}^0$) can be determined from the properties of the charged particles produced in its decay. The algorithm developed to estimate

the flavor of a reconstructed B candidate is known as the Flavor Tagger. In the case of Belle II and other B -factories, one of the two B mesons produced in each event, B_{tag} , can be used to *tag* the flavor of the other B meson, B_{sig} , under study. As was already explained in Section 1.4.2, by exploiting the quantum entanglement of the B meson pair produced in the decay of the $\Upsilon(4S)$, the flavor of the B_{sig} is known to be opposite to the flavor estimated for B_{tag} , at the time of the latter's decay. This flavor information is necessary for CP asymmetry measurements and B^0 – $\bar{B}^0$ oscillation and mixing studies.

The principle of Flavor Tagging at Belle II relies on the fact that many decays of neutral B mesons proceed through *flavor-specific* final states. A final state is called flavor-specific when the electric charge of one of its particles directly correlates with the flavor of the parent B meson. For example, in the semileptonic, $\bar{B}^0 \rightarrow D^{*+} \ell^- \bar{\nu}_\ell$ decay, the negative electric charge of the lepton unambiguously tags a $\bar{B}^0$. The same holds true for the charge-conjugate decay, $B^0 \rightarrow D^{*-} \ell^+ \nu_\ell$, where the positive electric charge of the lepton tags a B^0 . Particle identification information helps to identify the lepton, while momentum information can be used to disentangle between leptons from the $b \rightarrow c$ process and oppositely charged leptons from the $c \rightarrow s$ process.

In LHC experiments, such as LHCb, the situation is completely different because the hadrons are produced from the quark hadronisation process and not from an $\Upsilon(4S)$ decay. When two b -quarks are produced, they hadronise into two hadrons, generally of different type, e.g. a B_s and a Λ_b . In this case, flavor tagging uses information from charged particles found in either B decays or from b -quark jets to identify the flavor of the B meson. The tagging performance is much poorer compared to Belle II and other B factory experiments, and depends on the decay channel being studied.

Until 2023, Belle II used a machine-learning algorithm based on a series of Boosted Decision Tree (BDT) classifiers [99] to predict the flavor of the tag-side B meson in each event (see Section 5.2.3 for more information about BDT classifiers). It relies on 13 categories of varying quality to estimate the flavor, corresponding to different types of information, such as the electric charge of the particle identified¹ as a lepton (or kaon) originating from a B decay involving a $b \rightarrow c$ (or $b \rightarrow c \rightarrow s$) transition or the charge of the particle in the B decay with the highest momentum, among others. Hence, this algorithm is hereafter referred to as the *category-based* Flavor Tagger. After determining which of the aforementioned categories is more likely to provide the most accurate flavor signature, the Flavor Tagger provides a prediction q for the B_{tag} flavor ($q = +1$ for B^0 and $q = -1$ for $\bar{B}^0$, by convention) alongside the confidence value r associated to this prediction, ranging from 0 (no

¹using particle identification information from all sub-detectors that provide it, namely ARICH, TOP, ECL, KLM and dE/dx from CDC.

confidence) to 1 (maximum confidence). The value of r itself is directly correlated with the probability assigned by the algorithm to the category that best represents the flavor signature of the B_{tag} decay. The product of these two numbers, $q \cdot r$, constitutes the combined output of the Flavor Tagger, the distribution of which is shown in Figure 4.1 for the category-based Flavor Tagger.

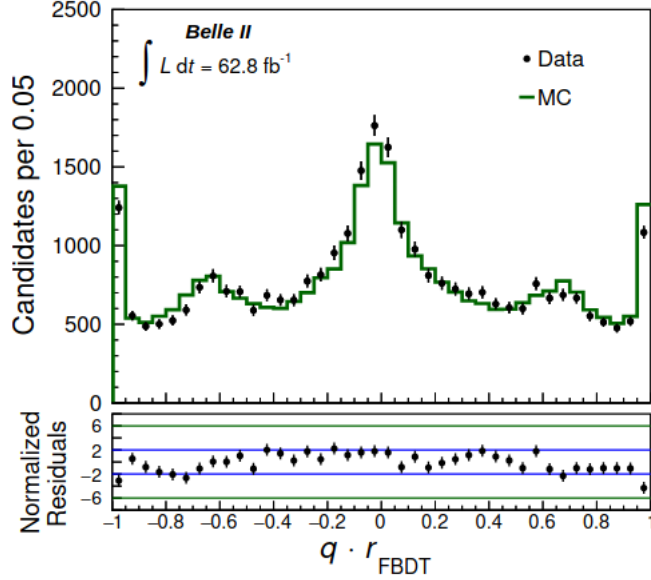


Figure 4.1: Comparison between the $q \cdot r$ distribution of the signal contribution (explained in Section 4.2.5) in 62.8 fb^{-1} of Belle II data and the one of simulated signal events for the category-based Flavor Tagger. Taken from [99].

The distribution of r is usually split into 7 uneven bins: $[0.0, 0.1, 0.25, 0.45, 0.6, 0.725, 0.875, 1.0]$. This binning of the r distribution is based on the different categories of the B_{tag} decays that populate it. For example, leptons from $b \rightarrow cl$ semi-leptonic B_{tag} decays are easier to identify correctly and usually lead to values of r closer to 1, while decays to neutral hadrons, that are generally more difficult to tag correctly, lead to values of r closer to 0.

As one might expect, predicting the B_{tag} flavor in this way is not perfect, meaning it has an efficiency and a purity associated to it. The metric commonly used to evaluate the performance of a Flavor Tagging algorithm is the so-called *effective tagging efficiency* defined as:

$$\varepsilon_{\text{tag}} = \sum_i \varepsilon_i (1 - w_i)^2, \quad (4.1)$$

where ε_i is the fraction of the number of tagged events that lie in a given bin of r , the aforementioned confidence of the B_{tag} flavor prediction, over the total number

of tagged events across all r -bins, w_i is the so-called *wrong tag fraction*, defined as the fraction of events in a given r -bin where the flavor of the B_{tag} was wrongly assigned over all tagged events in that r -bin, and i denotes the r -bin, ranging from 1 to 7². The effective tagging efficiency of the category-based Flavor Tagger was evaluated using the method described in this chapter on 365 fb^{-1} of Belle II data to be $(31.68 \pm 0.45)\%$, where the uncertainty is statistical only [100].

Besides the wrong tag fraction, there are two other parameters that are used to characterize the performance of the Flavor Tagger. The first is the *wrong tag asymmetry*, Δw , defined as the asymmetry between the fraction of wrongly tagged B^0 's and $\bar{B}^0$'s in a given r -bin, $\Delta w_i = w_i^+ - w_i^-$, where w_i^+ (w_i^-) denotes the fraction of events where the B_{tag} is wrongly tagged as $\bar{B}^0$ (B^0) in the i -th r -bin. The second is the *B_{tag} detection efficiency asymmetry*, a_{tag} , defined as the asymmetry between the fraction of correctly tagged B^0 's and $\bar{B}^0$'s in a given r -bin, $a_{\text{tag},i} = \varepsilon_i^+ - \varepsilon_i^-$, where ε_i^+ (ε_i^-) denotes the fraction of events where the B_{tag} is correctly tagged as B^0 ($\bar{B}^0$) in the i -th r -bin. Both of these asymmetries arise from the differing interaction rates, or cross-sections, of positively and negatively charged particles with the detector material. Since electric charge is used to infer the B_{tag} flavor, these asymmetries can lead to differences in the tagging efficiency between B^0 and $\bar{B}^0$ mesons. It is therefore important to take them into account when evaluating the performance of the Flavor Tagger.

Recently, a new Flavor Tagging algorithm was developed in Belle II, called **GF1aT**, that is based on graph-neural-networks (GNNs). It improves upon the category-based algorithm by incorporating relationships between final-state particles, such as their spatial distances, which the previous algorithm did not use, leading to a significant performance gain. Figure 4.2 shows a comparison between the $q \cdot r$ distributions of the category-based Flavor Tagger and **GF1aT** for simulated signal events where the better performance of the latter is already visible. The origins of this improvement are made clearer in Figure 4.3, where the same comparison is broken down into four categories of events depending on whether charged leptons or charged kaons can be found in the decay of the B_{tag} . This allows one to see that the new algorithm can tag the B_{tag} flavor more reliably than the old algorithm, in the cases where the flavor signature does not give enough information.

²Note that, in some figures, the numbering of the r -bins instead ranges from 0 to 6.

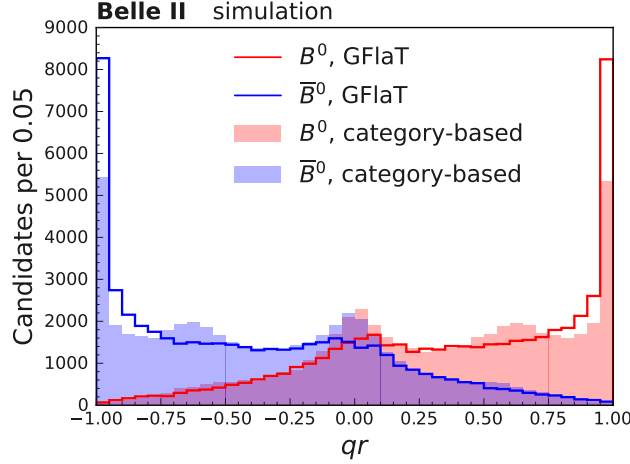


Figure 4.2: Comparison between the $q \cdot r$ distribution of the category-based Flavor Tagger and of GFlaT for simulated signal events. The distribution is split based on whether the B_{sig} in the event is a true B^0 (red) or a true $\bar{B}^0$ (blue).

4.2 Calibration of the GNN-based Flavor Tagger and Δt Resolution Function parameters

This section describes the data-driven evaluation of the performance of the GNN-based Flavor Tagger. This work was part of my service task to qualify as an author of Belle II publications. It was conducted by a team of three physicists leading the $\sin 2\beta$ analysis, with my contribution focused on the calibration of the Flavor Tagger. The calibrated parameter values are considered universal and are provided to the rest of the collaboration for use in measurements related to CP violation and the mixing of B mesons. This work led to an internal analysis note, a published paper [100], and a conference presentation given by myself [101].

4.2.1 Principle behind the calibration of the GNN-based Flavor Tagger

The performance of this new Flavor Tagging algorithm is evaluated by determining the values of w (and therefore ε_{tag}), Δw and a_{tag} in each bin of r . These parameters are all evaluated directly in 365 fb^{-1} of Belle II data³, using self-tagging B decays, i.e. decays of neutral B mesons to flavor-specific final states. Specifically, the following decays are used:

³The Run 1 dataset of Belle II contains 387 million $B\bar{B}$ pairs.

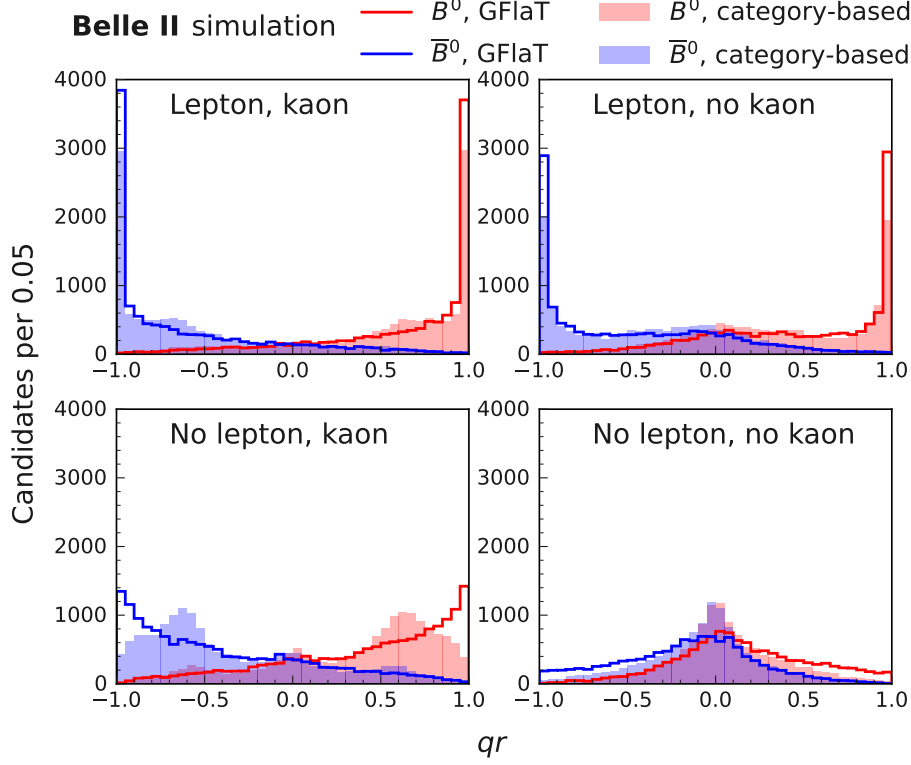


Figure 4.3: The same comparison that is shown in Figure 4.2 with an additional distinction made based on the presence of charged leptons or charged kaons among the decay products of the B_{tag} in each event.

- $B^0 \rightarrow D^- \pi^+$, with $D^- \rightarrow K^+ \pi^- \pi^-$
- $B^0 \rightarrow D^{*-} \pi^+$, with $D^{*-} \rightarrow \bar{D}^0 \pi^-$ and $\bar{D}^0$ decaying to either:
 - $\bar{D}^0 \rightarrow K^+ \pi^-$,
 - $\bar{D}^0 \rightarrow K^+ \pi^- \pi^0$,
 - $\bar{D}^0 \rightarrow K^+ \pi^- \pi^+ \pi^-$,

where the charge-conjugate decays are implied for the rest of this chapter. The reason behind using flavor-specific neutral B decays is to allow the determination of the true flavor in data, so that it can be compared to the output of GFlaT, thus giving a data-driven estimation of the true performance of the algorithm.

The Flavor Tagger parameters, w , Δw and a_{tag} , are determined in each r -bin in a time-dependent manner. Neglecting any inefficiencies due to Flavor Tagging and detector resolution effects, the decay time distribution in the case of flavor-specific

decays is written as:

$$\mathcal{P}_{\text{flav}}(\Delta t; q_{\text{tag}}, q_{\text{sig}}) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \{1 + q_{\text{tag}}q_{\text{sig}} \cos(\Delta m \Delta t)\}, \quad (4.2)$$

where Δt is the difference in time between the decay of the B_{sig} and B_{tag} , q_{tag} is the flavor of the B_{tag} at the time of its decay, q_{sig} is the flavor of the B_{sig} at the time of its decay, τ_{B^0} is the proper B^0 lifetime and Δm is the B^0 – $\bar{B}^0$ oscillation frequency.

After including the performance of the Flavor Tagger, the decay time distribution of flavor-specific decays takes the form:

$$\begin{aligned} \mathcal{P}_{\text{flav}}^{\text{FT}}(\Delta t; q_{\text{tag}}, q_{\text{sig}}, w, \Delta w, a_{\text{tag}}, a_{\text{sig}}) = & (1 + q_{\text{sig}}a_{\text{sig}}) \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} \{1 - q_{\text{tag}}\Delta w + q_{\text{tag}}a_{\text{tag}}(1 - 2w) \\ & + [q_{\text{tag}}(1 - 2w) + a_{\text{tag}}(1 - q_{\text{tag}}\Delta w)]q_{\text{sig}} \cos(\Delta m \Delta t)\}, \end{aligned} \quad (4.3)$$

where w , Δw and a_{tag} are the aforementioned Flavor Tagger parameters and a_{sig} is the B_{sig} detection efficiency asymmetry, defined as the asymmetry between the number of events in which the B_{sig} decays as a B^0 and ones where it decays as a $\bar{B}^0$. Unlike a_{tag} , a_{sig} is not considered to be r -bin dependent.

The reason for following a time-dependent approach instead of a time-integrated one, is that it makes it possible to also calibrate the parameters of the Δt *Resolution Function* in the same procedure, thereby simplifying the treatment of certain systematic uncertainties. The Δt Resolution Function models how the decay time resolution of the detector⁴ affects the measured values of Δt , which makes it a critical tool for all time-dependent analyses, thus highlighting the importance of accurately calibrating its parameters. The Δt Resolution Function takes the form:

$$\begin{aligned} \mathcal{R}(\delta t; \sigma_{\Delta t}) = & (1 - f_t - f_{\text{OL}})G(\delta t; m_{\text{main}}\sigma_{\Delta t}, s_{\text{main}}\sigma_{\Delta t}) \\ & + f_t(\sigma_{\Delta t})R_t(\delta t; m_{\text{tail}}\sigma_{\Delta t}, s_{\text{tail}}\sigma_{\Delta t}, \kappa\sigma_{\Delta t}, f_<, f_>) \\ & + f_{\text{OL}}G(\delta t; 0, \sigma_{\text{OL}}), \end{aligned} \quad (4.4)$$

where $\delta t \equiv \Delta t^{\text{true}} - \Delta t$ quantifies the detector time resolution (also known as Δt residual) and $\sigma_{\Delta t}$ is the uncertainty on the measured value of Δt , that is determined on a per-event basis. This exact form of the Δt Resolution Function was also used in Ref. [102]. The model is comprised of three parts:

- The first component, corresponding to the first term in Equation 4.4, is a Gaussian distribution with a mean, m_{main} , and standard deviation, s_{main} ,

⁴The time resolution is derived almost entirely from the spatial resolution of the vertex detector since, in each event, Δt is determined from the reconstructed decay vertex positions of the two B mesons.

which are scaled by the per-event $\sigma_{\Delta t}$. It accounts for $\sim 70\%$ of the candidates.

- The second component, corresponding to the second term in Equation 4.4, is a weighted sum of a Gaussian distribution and two Gaussian distributions that have each been convolved with an exponential function. It takes the form:

$$\begin{aligned} R_t(x; \mu, \sigma, \kappa, f_<, f_>) = & (1 - f_< - f_>)G(x; \mu, \sigma) \\ & + f_<G(x; \mu, \sigma) \otimes \frac{1}{\kappa}\text{exp}_<(x/\kappa) \\ & + f_>G(x; \mu, \sigma) \otimes \frac{1}{\kappa}\text{exp}_>(-x/\kappa) , \end{aligned} \quad (4.5)$$

where $\text{exp}_>(-x/\kappa) = \exp(-x/\kappa)$ if $x > 0$ and $\text{exp}_>(-x/\kappa) = 0$ otherwise, and similarly for $\text{exp}_<(x/\kappa)$. The exponential tails account for B_{tag} vertices that were poorly determined due to secondary displaced vertices from intermediate charm mesons. The fraction f_t takes on different values depending on the value of $\sigma_{\Delta t}$. This is modeled using a sigmoid distribution with a parameter (f_t^μ) describing the $\sigma_{\Delta t}$ threshold after which f_t takes non-zero values, a parameter (f_t^σ) describing how fast f_t rises to its plateau value and a parameter (f_t^{max}) for the plateau value itself. This part of the model accounts for $\sim 30\%$ of the candidates.

- The third component, corresponding to the third term in Equation 4.4, is a Gaussian distribution accounting for outlier events (OL) where the vertices were poorly reconstructed. It has a mean $m = 0$ ps and a standard deviation $\sigma = 200$ ps and it accounts for $\sim 0.1\%$ of the candidates.

More information about how the Δt Resolution Function models the detector resolution of signal events is given in Section 4.2.4. The Δt Resolution Function is convolved to Equation 4.3, so that the full description of the decay time distribution of flavor-specific decays reads:

$$\mathcal{P}_{\text{flav}}^{\text{FT, res}}(\Delta t; \sigma_{\Delta t}, q_{\text{tag}}, q_{\text{sig}}, w, \Delta w, a_{\text{tag}}, a_{\text{sig}}) = \mathcal{P}_{\text{flav}}^{\text{FT}}(\Delta t; q_{\text{tag}}, q_{\text{sig}}, w, \Delta w, a_{\text{tag}}, a_{\text{sig}}) \otimes \mathcal{R}(\delta t; \sigma_{\Delta t}) , \quad (4.6)$$

In the following, it is assumed that the parameters of the Flavor Tagger, w , Δw and a_{tag} , are uncorrelated to the parameters of the Δt Resolution Function. This assumption will eventually be tested by inspecting the relevant elements of the correlation matrix obtained from the final fit.

The calibration procedure consists of 3 main steps:

1. Reconstruction of the $B^0 \rightarrow D^{(*)-}\pi^+$ decays and application of selection criteria on the reconstructed data sample. This will be explained in Section 4.2.2.
2. Estimation of the signal and background yields in order to statistically subtract the background events in the Δt distribution using the $_s\mathcal{P}lot$ technique [103]. This will be detailed in Sections 4.2.3 and 4.2.5.
3. Fit of the remaining signal Δt distribution to determine the Flavor Tagger and Δt Resolution Function parameters. This will be described in Section 4.2.6.

The reason for choosing to subtract the background events from the Δt distribution, instead of modeling them, is that Δt is not a variable that allows to easily separate between the contributions of signal and the various kinds of backgrounds, while other more discriminating variables, such as ΔE and M_{bc} , previously defined in Equations 2.7 and 2.8, are typically used for this exact purpose. Statistically subtracting the background from the Δt distribution allows to fit the remaining signal contribution using Equation 4.6 and determine the parameters of the Flavor Tagger and Δt Resolution Function, without needing to model all of the background contributions in Δt . Steps 2 and 3 of the calibration procedure were first performed on fully simulated Monte Carlo (MC) events, in order to better control and validate the fit strategy, before performing it on the data recorded by Belle II. In the following subsections, all three steps are described in detail.

4.2.2 Reconstruction and selection of $B^0 \rightarrow D^{(*)-}\pi^+$ decays and selection criteria

The reconstruction of B_{sig} candidates decaying to self-tagging decay modes, in the case of both simulated and real data, is performed in the following way. D^- candidates are reconstructed from $D^- \rightarrow K^+\pi^-\pi^-$ and D^{*-} candidates from $D^{*-} \rightarrow \bar{D}^0\pi^-$ with $\bar{D}^0 \rightarrow K^+\pi^-$, $K^+\pi^-\pi^0$, or $K^+\pi^-\pi^+\pi^-$, requiring all charged particle tracks to originate from the IR and to have polar angles within the CDC acceptance.

π^0 candidates are reconstructed from $\pi^0 \rightarrow \gamma\gamma$ with photon candidates formed from ECL clusters that are not associated to any tracks from charged particles. To reduce the contribution of photons coming from beam-background, each cluster is required to have an energy greater than 120 MeV, 80 MeV, or 30 MeV depending on whether it is in the forward, backward, or barrel region of the calorimeter, corresponding to the polar angle ranges $[12.4, 31.4]^\circ$, $[130.7, 155.1]^\circ$ and $[32.2, 128.7]^\circ$, respectively, in the lab frame. The angle between the momenta of the two

photons must be less than 52° in the lab frame and the diphoton mass must be in the range $[121, 142]$ MeV/ c^2 .

One of the D meson's decay products has to be consistent with being a charged kaon and the invariant mass of each D^- candidate must be within $[1.860, 1.880]$ GeV/ c^2 , which is centered around the known D^- mass. Each $\bar{D}^0$ candidate reconstructed from $K^+\pi^-(\pi^+\pi^-)$ must have an invariant mass within $[1.845, 1.885]$ GeV/ c^2 , which is centered around the known $\bar{D}^0$ mass, while $\bar{D}^0$ candidates reconstructed from $K^+\pi^-\pi^0$ must have an invariant mass within $[1.810, 1.895]$ GeV/ c^2 , which is an asymmetric range around the known $\bar{D}^0$ mass in order to account for energy losses in photon reconstruction.

The π^- from the D^{*-} decay is required to have a momentum below 300 MeV/ c in the C.o.M. frame, since this pion is very soft, i.e. the phase space that is available for it is really small. Moreover, each D^{*-} candidate has to have an energy release, defined as: $m(D^{*-}) - m(\bar{D}^0) - m(\pi^-)$, within $[4.6, 7.0]$ MeV/ c^2 , which is centered around the known energy release for this decay. The clearly defined peak in the energy release allows to remove almost all of the background in the sample.

Neutral B meson candidates on the signal side of the event are reconstructed from D^* candidates and a charged particle track that is consistent with being a $\pi^\pm$. The charge of this pion track is what is used to determine the flavor of the reconstructed signal B meson; π^+ (π^-) corresponds to B^0 ($\bar{B}^0$), neglecting the contribution from the Cabibbo-suppressed decay process ($B^0 \rightarrow D^{(*)+}\pi^-$). In order to determine the decay position of the B_{sig} , the trajectories and momenta of its decay products are fitted according to its decay chain using the TREEFIT algorithm [97], while constraining the B to originate from the IR and the $D^{(*)}$ to its known mass [70].

The tag-side neutral B meson in each event is reconstructed after reconstructing the B_{sig} , using all remaining charged particle tracks and calorimeter clusters in the “Rest-Of-Event” (ROE). The decay position of the B_{tag} is determined by fitting the trajectories of ROE tracks using the RAVE algorithm [96], which accounts for the unknown decay chain of the B_{tag} by constraining its decay position to be consistent with the origin and direction of the B_{sig} . This significantly reduces the impact of tertiary decay vertices due to potential intermediate D mesons in the B_{tag} decay chain.

Events that proceed through the process $e^+e^- \rightarrow q\bar{q}$, where $q = \{u, d, c, s\}$, are suppressed by exploiting their topological differences with events that proceed through $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$, by requiring the ratio of the second to the zeroth Fox-Wolfram moment, R_2 , be less than 0.4 [104].

4.2.3 Estimation of signal and background yields

The number of signal and background events is determined from a fit to the ΔE distribution of the reconstructed $B^0 \rightarrow D^{(*)-}\pi^+$ decays. Using simulated decays, the following types of events are identified and modeled:

- Signal events are modeled in ΔE using the sum of a double-sided Crystal Ball function with a Gaussian function.
- $B^0 \rightarrow D^{(*)-}K^+$ (referred to as BDK) events that were misidentified as signal events are modeled in ΔE using a double-sided Crystal Ball function. They amount to $\sim 2.5\%$ of the expected signal events.
- Background events where neither of the two B mesons decays to the final state of interest (referred to as $B\bar{B}$ background) are modeled in ΔE using a 2^{nd} degree Chebyshev polynomial function.
- Continuum background events (also referred to as $q\bar{q}$ background) are modeled in ΔE using an exponential function.

Details on the mathematical formulae of the functions used to model all the different components can be found in Appendix B. The ΔE shape parameters for the signal and each of the individual background components are determined from fits to fully simulated signal and background samples and are subsequently fixed when fitting the ΔE distribution of the reconstructed $B^0 \rightarrow D^{(*)-}\pi^+$ decays.

Figure 4.4 shows fits to the ΔE distributions of the simulated signal, BDK and $B\bar{B}$ background components. Signal events are expected to peak at $\Delta E \sim 0$ GeV, since in their case the B meson energy is correctly reconstructed and exactly matches half of the energy in the C.o.M. frame. BDK events peak at slightly lower energies, because in their case a K meson is mistakenly reconstructed as a π meson, leading to an offset peak for the energy B meson. It is also clear that, in the case of $B\bar{B}$ background, there is a small number of events that peak around $\Delta E \sim 0$ GeV, which are not modeled, for the sake of simplicity, since the modeling of this background does not impact the final parameter values significantly. With larger datasets in the future, a more detailed treatment will be necessary, but for the present analysis the simplified description is sufficient, with the associated effects covered by the systematic uncertainties (see Section 4.2.6).

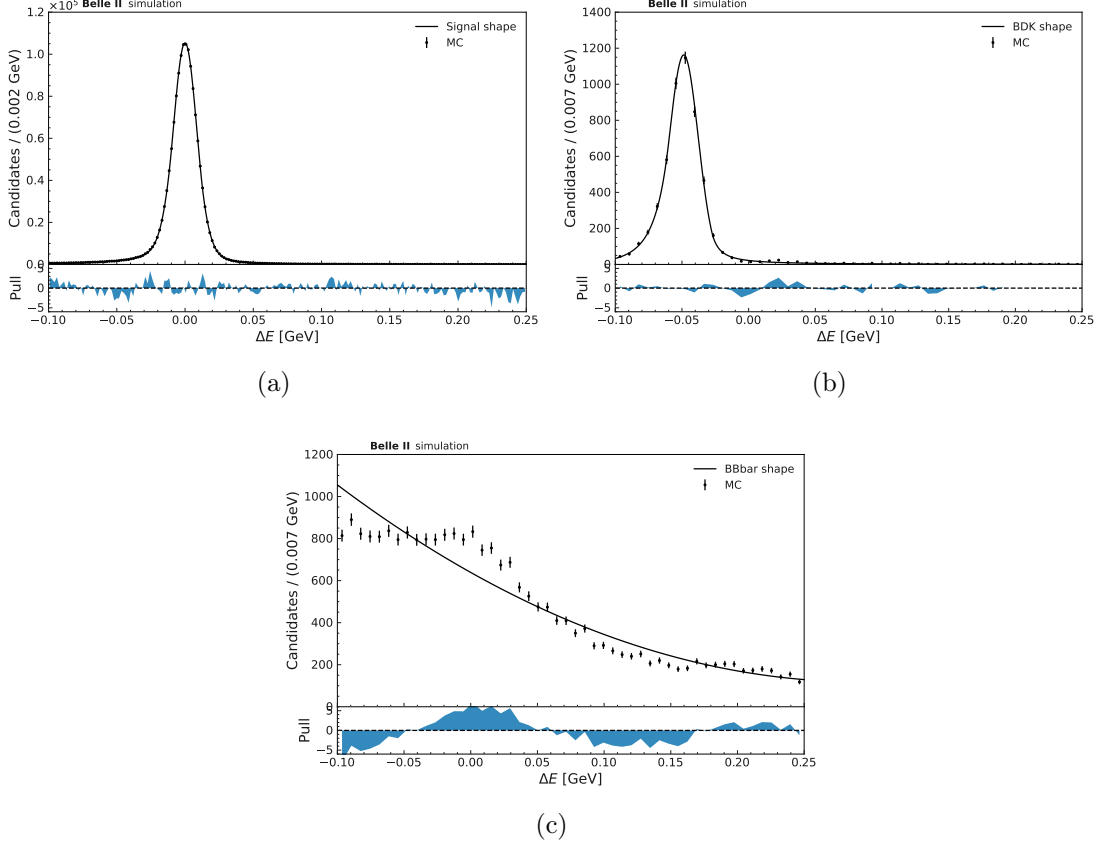


Figure 4.4: ΔE shape fits of the simulated (a) signal, (b) BDK and (c) $B\bar{B}$ background components.

The full Probability Density Function (PDF) model that describes the ΔE distribution of all signal and background contributions to the selected events is then written as:

$$\mathcal{P}^{\text{full}}(\Delta E) = N_{\text{sig-like}} \times (f_{\text{sig}} \mathcal{P}^{\text{sig}} + f_{\text{BDK}} \mathcal{P}^{\text{BDK}} + f_{B\bar{B}} \mathcal{P}^{B\bar{B}}) + N_{q\bar{q}} \times \mathcal{P}^{q\bar{q}}, \quad (4.7)$$

where the fraction between the signal and BDK components, which are considered together as signal-like, as well as the one between the signal+BDK and $B\bar{B}$ background components are both determined and fixed from simulation. This only leaves $N_{\text{sig-like}}$ and $N_{q\bar{q}}$, i.e. the numbers of signal-like and $q\bar{q}$ background events, respectively, free to float in the fit.

To measure the number of observed signal and background events, or yields, an unbinned maximum likelihood fit is performed to the ΔE distribution of all events reconstructed as $B^0 \rightarrow D^{(*)-} \pi^+$ decays, using the PDF model of Equation 4.7. More details about the concept behind the maximum likelihood fit are given in

Appendix B.8. This fit is performed in the signal region, $-0.1 < \Delta E < 0.25$ GeV and $5.27 < M_{bc} < 5.29$ GeV/ c^2 , where most signal events are expected to lie. The ΔE distribution is also fitted simultaneously in the M_{bc} sideband region, $M_{bc} < 5.24$ GeV/ c^2 , where almost no signal events are expected, using a shared exponential slope parameter for the $q\bar{q}$ background between the two regions. This significantly improves our knowledge of the shape of the $q\bar{q}$ background, knowing that the latter region is continuum-dominated.

Besides the signal and background yields, some shape parameters are also allowed to float in the ΔE fit. These include factors that scale the mean and standard deviation of the double-sided Crystal Ball shapes that model the signal and BDK components, the parameters of the 2^{nd} degree polynomial shape that models the $B\bar{B}$ background, as well as the slope parameter of the exponential shape that models the continuum background in the signal and sideband regions. This is done in order to better capture potential data/MC differences in the modeling of the signal and the various background components. Figure 4.5 shows the ΔE fit on 1 ab^{-1} of fully simulated data and Table 4.1 lists the values of the fitted parameters.

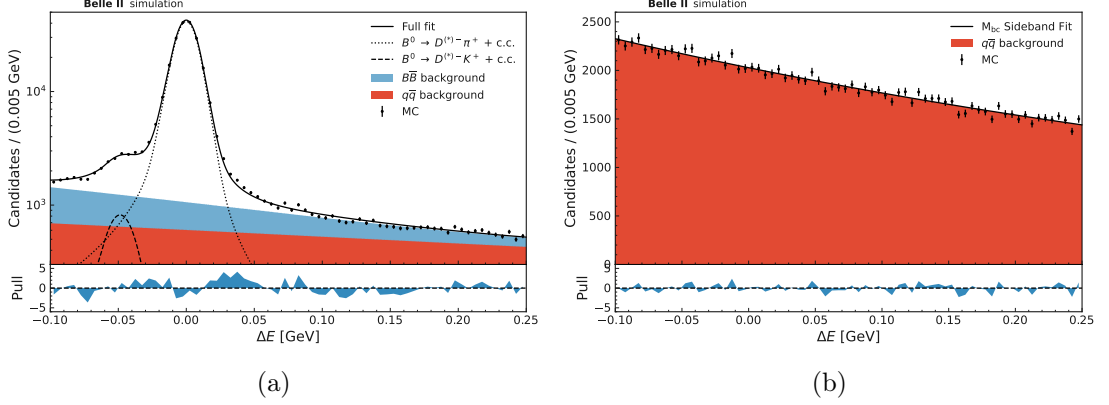


Figure 4.5: ΔE fit of 1 ab^{-1} of fully simulated data (MC) in the (a) signal and (b) sideband regions.

Table 4.1: Yields and shape parameters obtained by fitting the ΔE distribution of 1 ab^{-1} of fully simulated data. c_1 and c_2 are parameters describing the 2^{nd} degree polynomial that models the $B\bar{B}$ background and λ is the parameter describing the exponential shape that models the $q\bar{q}$ background. The BDK and $B\bar{B}$ yields are calculated using the signal yield by scaling it by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.

Parameter	Value	Uncertainty
Signal + BDK yield	209340	510
BDK yield	5220	10
$B\bar{B}$ yield	23400	60
$q\bar{q}$ yield	38450	360
Sideband yield	129280	360
μ shift [$\times 10^{-5}$ GeV]	-1.3391	2.4711
σ scale	1.0137	0.0026
c_1 [GeV $^{-1}$]	-0.9264	0.0222
c_2 [GeV $^{-2}$]	0.1745	0.0170
λ [GeV $^{-1}$]	-1.3737	0.0277

A separate ΔE fit is performed simultaneously for $B^0 \rightarrow D^{(*)-}\pi^+$ events on the one hand, and $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ events on the other hand, in order to determine the value of the a_{sig} parameter of Equation 4.3 that will be subsequently used in the Δt fit, described in Section 4.2.6. This is done without applying any selection criteria related to the tag-side, so as not to bias the value of a_{sig} , a parameter that is exclusively related to the signal-side. A biased value of a_{sig} would in turn lead to biased Flavor Tagger and Δt Resolution Function parameter values. Figure 4.6 shows this fit on 1 ab^{-1} of fully simulated data and Table 4.2 lists the values of the fitted parameters. The results of the ΔE fit performed separately for each B_{sig} flavor are found to be consistent with those from the flavor-integrated ΔE fit.

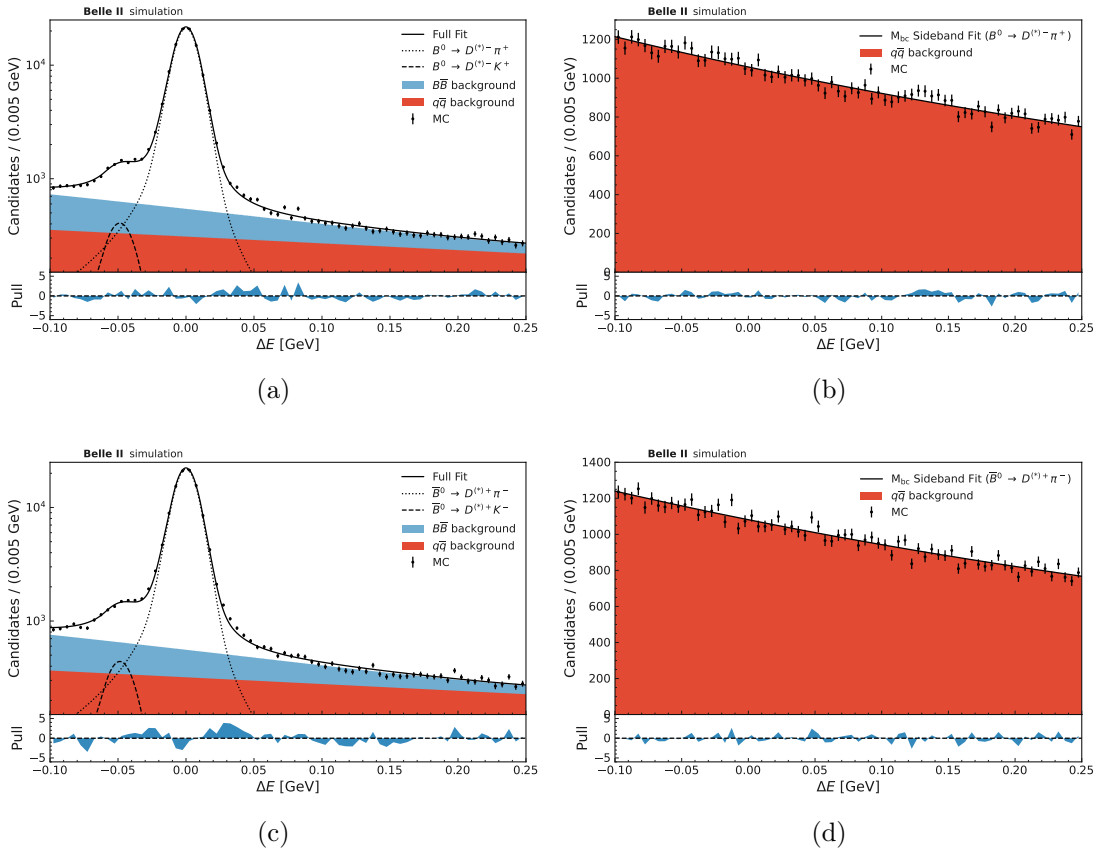


Figure 4.6: ΔE fit of 1 ab^{-1} of fully simulated data (MC) in the (a) signal and (b) sideband regions in the case of $B^0 \rightarrow D^{(*)-}\pi^+$ decays and in the (c) signal and (d) sideband regions in the case of $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ decays.

Table 4.2: Yields and shape parameters obtained by fitting the ΔE distribution of 1 ab^{-1} of fully simulated data simultaneously for $B^0 \rightarrow D^{(*)-}\pi^+$ and $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ events. a_{sig} , the shift in μ , the scale factor of σ as well as the total signal yield are shared between the two fits. The BDK and $B\bar{B}$ yields are separately calculated for B^0 and $\bar{B}^0$ using the signal yields by scaling them by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.

Parameter	Value	Uncertainty
a_{sig}	-0.01359	0.00237
Signal + BDK yield	216640	520
BDK yield (B^0)	2590	10
BDK yield ($\bar{B}^0$)	2810	10
$B\bar{B}$ yield (B^0)	12070	40
$B\bar{B}$ yield ($\bar{B}^0$)	12140	40
$q\bar{q}$ yield (B^0)	19700	250
$q\bar{q}$ yield ($\bar{B}^0$)	20370	260
Sideband yield (B^0)	67430	260
Sideband yield ($\bar{B}^0$)	68960	260
μ shift [$\times 10^{-5}$ GeV]	-1.4980	2.4306
σ scale	1.0141	0.0026
c_{1,B^0} [GeV^{-1}]	-0.8947	0.0305
$c_{1,\bar{B}^0}$ [GeV^{-1}]	-0.9450	0.0306
c_{2,B^0} [GeV^{-2}]	0.1617	0.0237
$c_{2,\bar{B}^0}$ [GeV^{-2}]	0.1785	0.0236
λ_{B^0} [GeV^{-1}]	-1.3767	0.0383
$\lambda_{\bar{B}^0}$ [GeV^{-1}]	-1.3706	0.0379

After having demonstrated that both of the above ΔE fits produce reasonable results with MC, they were performed on 365 fb^{-1} of data recorded by Belle II. Figure 4.7 shows the ΔE fit that does not separate between the two B_{sig} flavors and Table 4.3 lists the measured parameters. Figure 4.8 shows the ΔE fit that separates $B^0 \rightarrow D^{(*)-}\pi^+$ and $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ events and Table 4.4 lists the measured parameters.

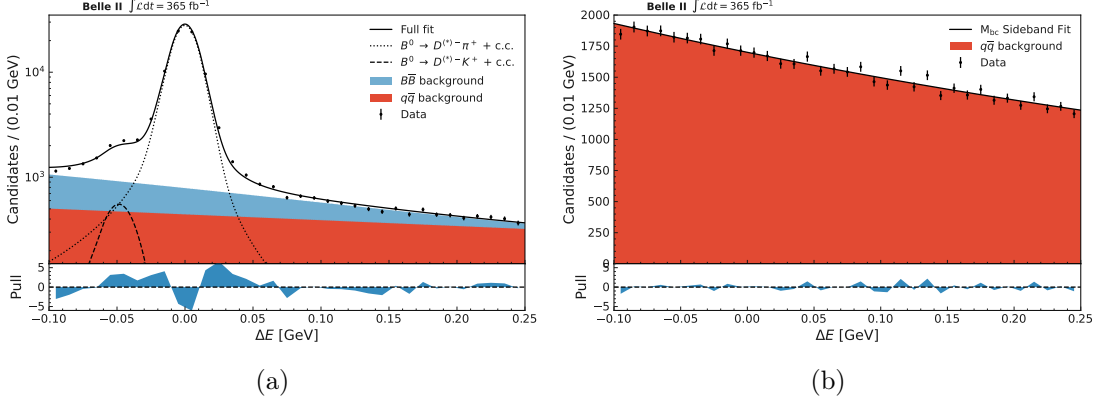


Figure 4.7: ΔE fit of the 365 fb^{-1} Belle II data sample in the (a) signal and (b) sideband regions.

Table 4.3: Yields and shape parameters obtained by fitting the ΔE distribution of the 365 fb^{-1} Belle II data sample. The BDK and $B\bar{B}$ yields are calculated using the signal yield by scaling it by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.

Parameter	Value	Uncertainty
Signal + BDK yield	77130	320
BDK yield	1920	10
$B\bar{B}$ yield	8620	40
$q\bar{q}$ yield	14200	230
Sideband yield	54530	230
μ shift [$\times 10^{-5} \text{ GeV}$]	-3.5875	4.4953
σ scale	1.1120	0.0049
c_1 [GeV^{-1}]	-0.9980	0.0372
c_2 [GeV^{-2}]	0.1562	0.0285
λ [GeV^{-1}]	-1.2777	0.0426

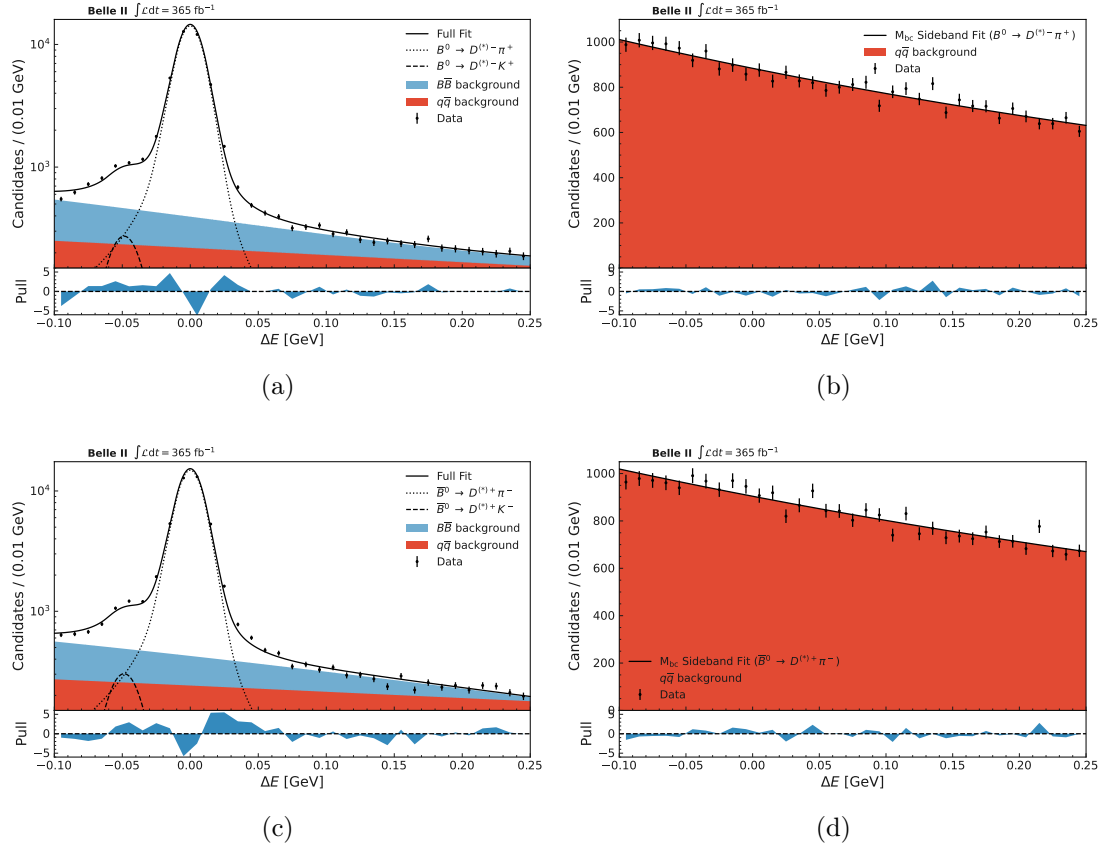


Figure 4.8: ΔE fit of the 365 fb^{-1} Belle II data sample in the (a) signal and (b) sideband regions in the case of $B^0 \rightarrow D^{(*)-} \pi^+$ decays and in the (c) signal and (d) sideband regions in the case of $\bar{B}^0 \rightarrow D^{(*)+} \pi^-$ decays.

Table 4.4: Yields and shape parameters obtained by fitting the ΔE distribution of the 365 fb^{-1} Belle II data sample simultaneously for $B^0 \rightarrow D^{(*)-}\pi^+$ and $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ events. The BDK and $B\bar{B}$ yields are separately calculated for B^0 and $\bar{B}^0$ using the signal yields by scaling them by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.

Parameter	Value	Uncertainty
a_{sig}	-0.02528	0.00391
Signal + BDK yield	80310	320
BDK yield (B^0)	950	10
BDK yield ($\bar{B}^0$)	1050	10
$B\bar{B}$ yield (B^0)	4420	30
$B\bar{B}$ yield ($\bar{B}^0$)	4550	30
$q\bar{q}$ yield (B^0)	7000	150
$q\bar{q}$ yield ($\bar{B}^0$)	7780	160
Sideband $q\bar{q}$ yield (B^0)	28220	160
Sideband $q\bar{q}$ yield ($\bar{B}^0$)	29140	170
μ shift [$\times 10^{-5}$ GeV]	-2.9154	4.4049
σ scale	1.1121	0.0048
c_{1,B^0} [GeV^{-1}]	-0.9758	0.0492
$c_{1,\bar{B}^0}$ [GeV^{-1}]	-1.0150	0.0512
c_{2,B^0} [GeV^{-2}]	0.1947	0.0384
$c_{2,\bar{B}^0}$ [GeV^{-2}]	0.1192	0.0405
λ_{B^0} [GeV^{-1}]	-1.3456	0.0592
$\lambda_{\bar{B}^0}$ [GeV^{-1}]	-1.1969	0.0581

A signal + BDK yield of 77130 ± 320 events is obtained from fitting 365 fb^{-1} of Belle II data, which is more than 4σ larger compared to the 75410 ± 180 events expected from the MC ($\simeq 36\%$ of the signal + BDK yield from the ΔE fit to 1 ab^{-1} of simulated data). This is attributed to the branching fractions of $B^0 \rightarrow D^{(*)-}\pi^+$ decays being slightly mismodeled in the Belle II MC. The signal detection efficiency asymmetry is measured to be: $a_{\text{sig}} = (-2.53 \pm 0.39)\%$, where the negative sign indicates that the reconstruction efficiency of $\bar{B}_{\text{sig}}^0$ is greater than that of B_{sig}^0 . The fact that a_{sig} is negative is attributed to charge asymmetries in kaon identification and in low-momentum tracking. The value of a_{sig} obtained from the fit to data is also larger than compared to the one obtained from the MC, $a_{\text{sig}} = (-1.36 \pm 0.24)\%$, with a significance of 2.8σ , which can be explained by a mismodeling of the detector material, leading to an asymmetry in the interaction cross section between positively and negatively charged particles with the detector material that is smaller in absolute value in the simulation compared to reality.

4.2.4 Interlude: Fit of the Δt residual distribution of simulated signal events

Before describing the treatment of the Δt distribution using the signal and background yields that are obtained from the ΔE fit in Section 4.2.3, a brief demonstration of how the Δt Resolution Function models the detector resolution will be given. Figure 4.9 shows a fit of the Δt Resolution Function of Equation 4.4 to the Δt residual distribution of fully simulated signal events and Table 4.5 lists the fitted parameters. Some of the parameter values obtained from this fit are still allowed to float in the final fit performed on the background-subtracted Δt distribution, described in Section 4.2.6, namely: m_{main} , m_{tail} , s_{main} , s_{tail} , f_t^{max} , m_{main}^7 and m_{tail}^7 . These are the parameters of the Δt Resolution Function that are actually calibrated on data. The rest of the parameters are determined entirely from fitting simulated data, knowing that their values are fully compatible with what was obtained from fitting a smaller subset of Belle II data in previous studies. This significantly improves the stability of the final Δt fit. Moreover, some parameters, such as: c , f_R , f_{exp} , m_{main} and m_{tail} are allowed to take different values for events in the seventh r -bin, i.e. the one with the best quality of flavor assignment, compared to those in the other six, as observed in both simulation and data.

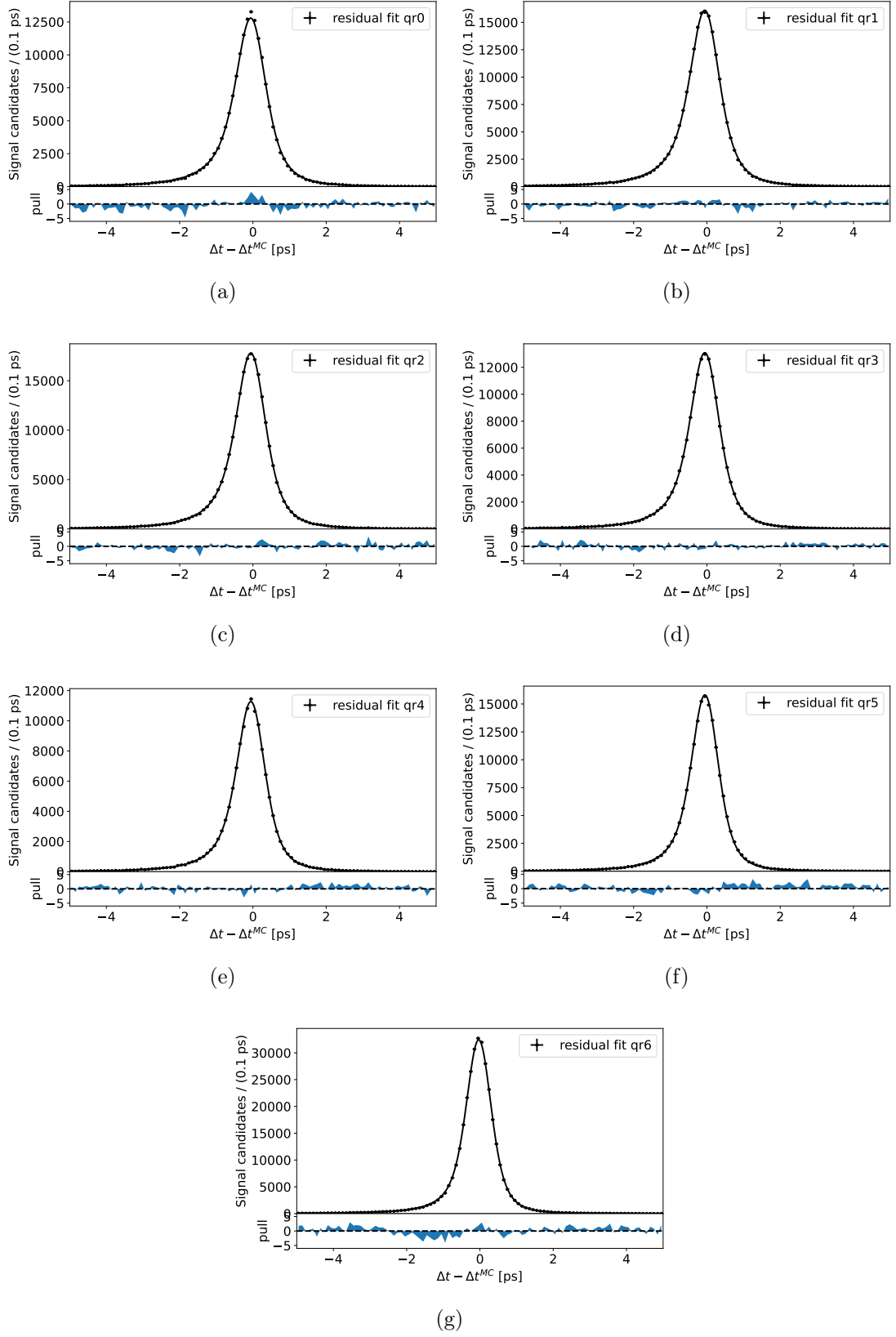


Figure 4.9: Fits of the Δt residual distribution of fully simulated signal events in each of the seven Flavor Tagging r -bins.

Table 4.5: Parameters of the resolution function obtained by fitting the Δt residual distribution of fully simulated signal events. When comparing to Equations 4.4 and 4.5, $f_{>} = f_R \times f_{exp}$ and $f_{<} = (1 - f_R) \times f_{exp}$. Parameters indexed with the number 7 denote the value of that parameter in the last r -bin. The last column shows whether these parameters are left free to float when performing the final Δt fit on the real data.

Parameter	Value	Uncertainty	Floating in Δt fit
m_{main}	-0.18578	0.00207	yes
m_{tail}	-1.03696	0.01796	yes
s_{main}	1.2324	0.0022	yes
s_{tail}	2.4753	0.0180	yes
κ	4.2343	0.0567	no
f_R	0.2637	0.0055	no
f_{exp}	0.2048	0.0045	no
f_t^{max}	0.2553	0.0033	yes
f_t^{μ} [ps]	0.2372	0.0012	no
f_t^{σ} [ps]	0.0914	0.0022	no
κ^7	4.2718	0.0762	no
f_R^7	0.2014	0.0084	no
f_{exp}^7	0.2415	0.0071	no
m_{main}^7	-0.11724	0.00311	yes
m_{tail}^7	-0.77795	0.02759	yes
f_{OL}	1.900×10^{-4}	0.175×10^{-4}	no

4.2.5 Production of the background-subtracted Δt distribution with **sWeights**

As explained in Section 4.2.1, the Δt distribution of signal events in data must first be isolated in order to fit it using Equation 4.6 and finally measure the Flavor Tagger and Δt Resolution Function parameters. To this end, the signal and background yields obtained from the ΔE fit (see Tables 4.1 and 4.3) are used to calculate the so-called **sWeights** which are then used to statistically subtract the background from the Δt distribution. The **sWeights** are per-event numbers that weight the Δt distribution, so that all background events are suppressed and only the distribution of signal events remains.

The formula used to compute the **sWeights** ${}_s\mathcal{W}$ is:

$${}_s\mathcal{W}_n(y_i) = \frac{\sum_{j=1}^{N_S} \mathbf{V}_{nj} \mathcal{P}^j(y_i)}{\sum_{k=1}^{N_S} N_k \mathcal{P}^k(y_i)}, \quad (4.8)$$

where:

- n denotes the component we are interested to retain in the distribution of the control variable, here Δt ,
- y is the discriminating variable that is fitted to determine the **sWeights**, in this case ΔE ,
- i is index denoting the event for which the **sWeight** is being computed,
- N_S is the total number of different event species considered, here 3 for signal+BDK, $B\bar{B}$ background and $q\bar{q}$ background,
- $\mathbf{V}_{nj}$ is the nj -th element of the **covariance matrix** that is obtained directly from the ΔE fit⁵,
- N_k is the yield of component k , as determined from the ΔE fit, and
- $\mathcal{P}^j(y_i)$ is the value of the fitted PDF model of component j , as a function of the discriminating variable y , for event i .

As explained in Ref. [103], the above weights are accurate only in cases where the distribution of the discriminating variable that is fitted to determine the **sWeights** is *uncorrelated* with the distribution of the control variable on which

⁵The elements of the covariance matrix are given by the inverse of the second derivative of the negative log likelihood minimized by the fit, in terms of the yields of the different species considered, $\mathbf{V}_{nj}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_n \partial N_j}$.

the `sWeights` are applied, for each of the different components considered, i.e. signal and the various backgrounds. This was checked to be the case with the maximum correlation observed being $\sim 2\%$.

It is also important to note that BDK events are considered as signal events, in that they are not subtracted from the Δt distribution by the `sWeights` procedure. The reason for that is that the Δt distribution of BDK events is the same as the signal one, because for these events the track of the kaon that is misidentified as a pion still points to the B_{sig} vertex, just like the pion track does in the case of signal events. This results in a slightly larger signal sample, without biasing the Δt distribution in any way.

With the `sWeights` calculated using Equation 4.8, it is possible to recover the $q \cdot r$ distribution of signal events in data, since the output of the Flavor Tagger is not correlated to ΔE for either of the signal and background components, and then compare it to the $q \cdot r$ distribution of simulated signal events. Figure 4.10 shows this comparison and clearly demonstrates a good agreement between the two distributions.

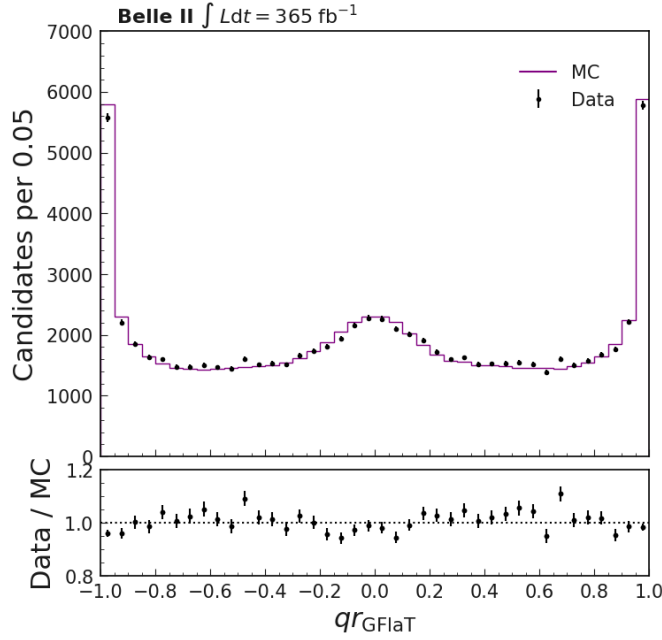


Figure 4.10: Comparison between the background-subtracted $q \cdot r$ distribution observed in data and the simulated signal+BDK contribution for the GNN-based Flavor Tagger. Here the MC distribution is normalised to match the luminosity of the data distribution.

4.2.6 Fit of the signal Δt distribution

After obtaining the background-subtracted Δt distribution, a *weighted* unbinned maximum likelihood fit is performed to it, using the model of Equation 4.6, simultaneously across the seven r -bins. It is also performed simultaneously for events where the B_{sig} and B_{tag} have opposite flavors (called OF) and for events where they have the same flavor (called SF) at the time of their decay. In total, 28 parameters are allowed to float freely in the fit; 21 related to the Flavor Tagger and 7 related to the Δt Resolution Function. In order to make the fit more computationally efficient, the wrong tag fractions (w) and wrong tag asymmetries (Δw) are expressed in terms of values they take depending on whether the B_{tag} was tagged as a B^0 or as a $\bar{B}^0$. They are defined as follows:

$$w \equiv \frac{1}{2}(w^+ + w^-) \quad \text{and} \quad \Delta w \equiv w^+ - w^-, \quad (4.9)$$

where the $+$ ($-$) superscript indicates that the B_{tag} was tagged as a B^0 ($\bar{B}^0$).

As with the ΔE fit, the Δt fit is first checked on the 1 ab^{-1} of fully simulated data before being performed on the 365 fb^{-1} Belle II data sample. Table 4.6 lists the Flavor Tagger and Δt Resolution Function parameters obtained from the fit on the MC. The fit projections in each r -bin individually are shown in Appendix C.1 for both OF and SF events, together with the asymmetry between them. Comparisons of the fitted Flavor Tagger parameters in each of the seven r -bins, with respect to their true values known from the MC are shown in Figures 4.11–4.13 and they show good overall agreement between them. This means that the fit strategy does not introduce any major biases in the evaluation of the Flavor Tagger parameters.

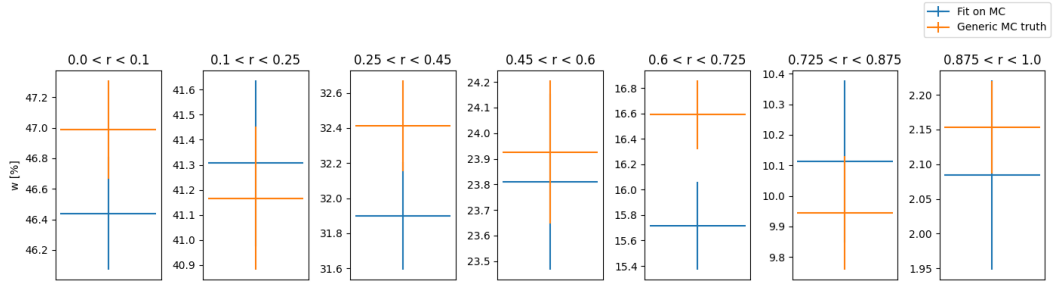


Figure 4.11: Comparison between the wrong tag fractions (w) obtained from the fit to the background-subtracted Δt distribution of the 1 ab^{-1} simulated data sample (in blue) and their true values known from the MC (in orange).

Table 4.6: Flavor Tagger and Δt Resolution Function parameters obtained by fitting the background-subtracted Δt distribution of the 1 ab⁻¹ fully simulated sample.

Parameter	Value	Uncertainty
m_{main}	-0.16683	0.02824
m_{tail}	-0.81315	0.10347
s_{main}	1.18742	0.03291
s_{tail}	2.36873	0.12706
f_{max}	0.31511	0.02494
w_1^+	0.46022	0.00522
w_1^-	0.46856	0.00529
$a_{\text{tag},1}$	0.00101	0.00840
w_2^+	0.41199	0.00467
w_2^-	0.41415	0.00467
$a_{\text{tag},2}$	-0.00774	0.00749
w_3^+	0.31859	0.00434
w_3^-	0.31937	0.00435
$a_{\text{tag},3}$	-0.00807	0.00712
w_4^+	0.23483	0.00479
w_4^-	0.24137	0.00495
$a_{\text{tag},4}$	0.00852	0.00828
w_5^+	0.15742	0.00482
w_5^-	0.15689	0.00502
$a_{\text{tag},5}$	0.02350	0.00894
w_6^+	0.09823	0.00370
w_6^-	0.10403	0.00382
$a_{\text{tag},6}$	0.00199	0.00745
m_{main}^7	-0.04613	0.04687
m_{tail}^7	-0.61763	0.16506
w_7^+	0.01696	0.00188
w_7^-	0.02472	0.00199
$a_{\text{tag},7}$	-0.00897	0.00511

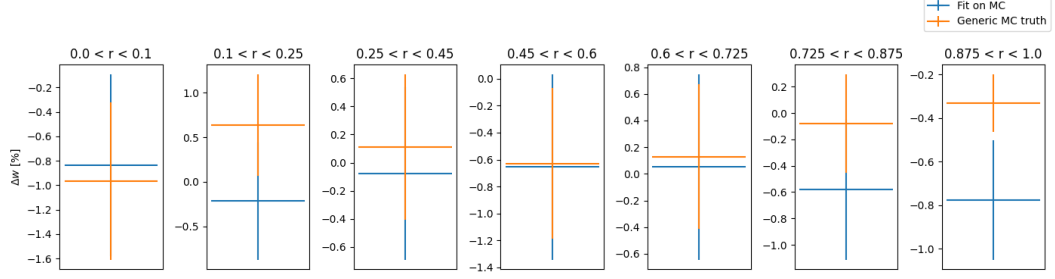


Figure 4.12: Comparison between the wrong tagging asymmetries (Δw) obtained from the fit to the background-subtracted Δt distribution of the 1 ab^{-1} simulated data sample (in blue) and their true values known from the MC (in orange).

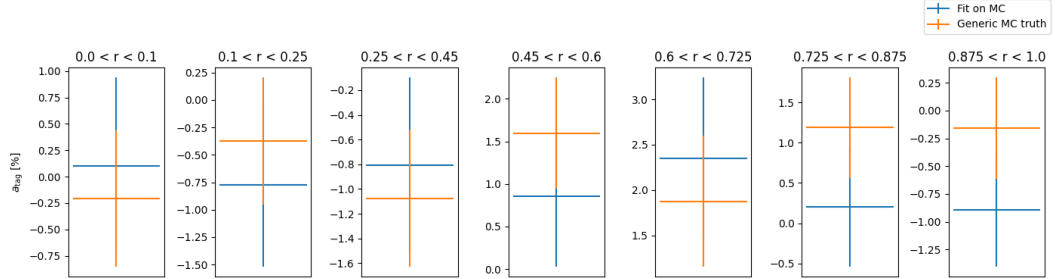


Figure 4.13: Comparison between the tag-side detection efficiency asymmetries (a_{tag}) obtained from the fit to the background-subtracted Δt distribution of the 1 ab^{-1} simulated data sample (in blue) and their true values known from the MC (in orange).

The effective tagging efficiency obtained by fitting the fully simulated data is calculated using Equation 4.1, using the fitted values of w^+ and w^- , and is found to be:

$$\varepsilon_{\text{tag}}^{\text{MC}} = (40.08 \pm 0.24)\%, \quad (4.10)$$

where the uncertainty is purely statistical.

Equation 4.6 was then fitted to the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample. Table 4.7 lists the Flavor Tagger and Δt Resolution Function parameters obtained from the fit to the data. The fit projections in each r -bin individually are shown in Figures 4.14–4.20 for both OF and SF events, together with the asymmetry between them. Comparisons of the fitted Flavor Tagger parameters in each of the seven r -bins, with respect to their true values obtained from the MC are shown in Figures 4.21–4.23 and they clearly show that the obtained wrong tag fractions are significantly worse in the data compared to the MC. This increase of the wrong tag fractions is attributed to the fact that the simulation does not perfectly model the data, thereby justifying the use of data and *self-tagged* decay modes to evaluate the performance of the Flavor Tagger and to calibrate its parameters.

Table 4.7: Flavor Tagger and Δt Resolution Function parameters obtained by fitting the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample.

Parameter	Value	Uncertainty
m_{main}	-0.10358	0.04021
m_{tail}	-0.53601	0.16563
s_{main}	1.03324	0.05148
s_{tail}	2.20154	0.23547
f_{max}	0.31591	0.04404
w_1^+	0.48681	0.00873
w_1^-	0.47902	0.00871
$a_{\text{tag},1}$	-0.01720	0.01387
w_2^+	0.41368	0.00764
w_2^-	0.42779	0.00773
$a_{\text{tag},2}$	-0.00936	0.01229
w_3^+	0.34611	0.00705
w_3^-	0.34646	0.00717
$a_{\text{tag},3}$	-0.00275	0.01153
w_4^+	0.24988	0.00768
w_4^-	0.23346	0.00799
$a_{\text{tag},4}$	0.03205	0.01330
w_5^+	0.17660	0.00793
w_5^-	0.16302	0.00812
$a_{\text{tag},5}$	0.01174	0.01442
w_6^+	0.11369	0.00634
w_6^-	0.11630	0.00638
$a_{\text{tag},6}$	-0.01128	0.01230
m_{main}^7	0.10940	0.07828
m_{tail}^7	-0.92925	0.26822
w_7^+	0.02997	0.00338
w_7^-	0.02242	0.00336
$a_{\text{tag},7}$	-0.00178	0.00864

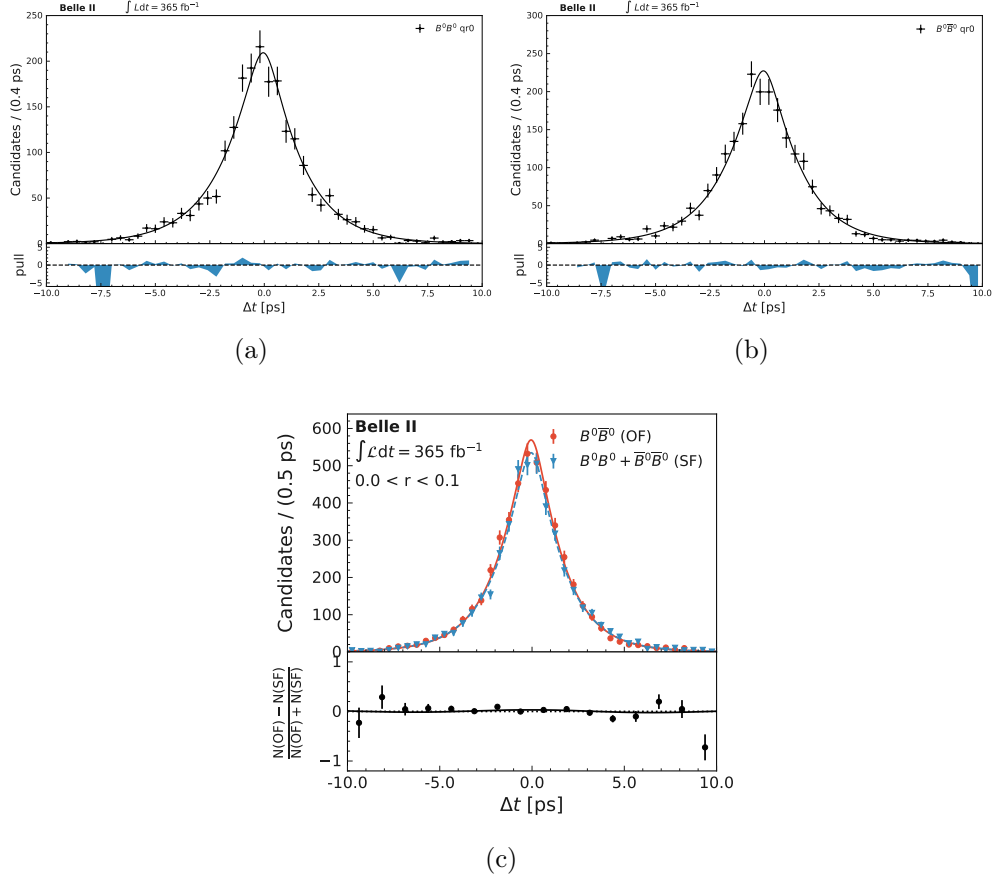


Figure 4.14: Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the first r -bin (i.e. the bin with the worst Flavor Tagging quality) for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).

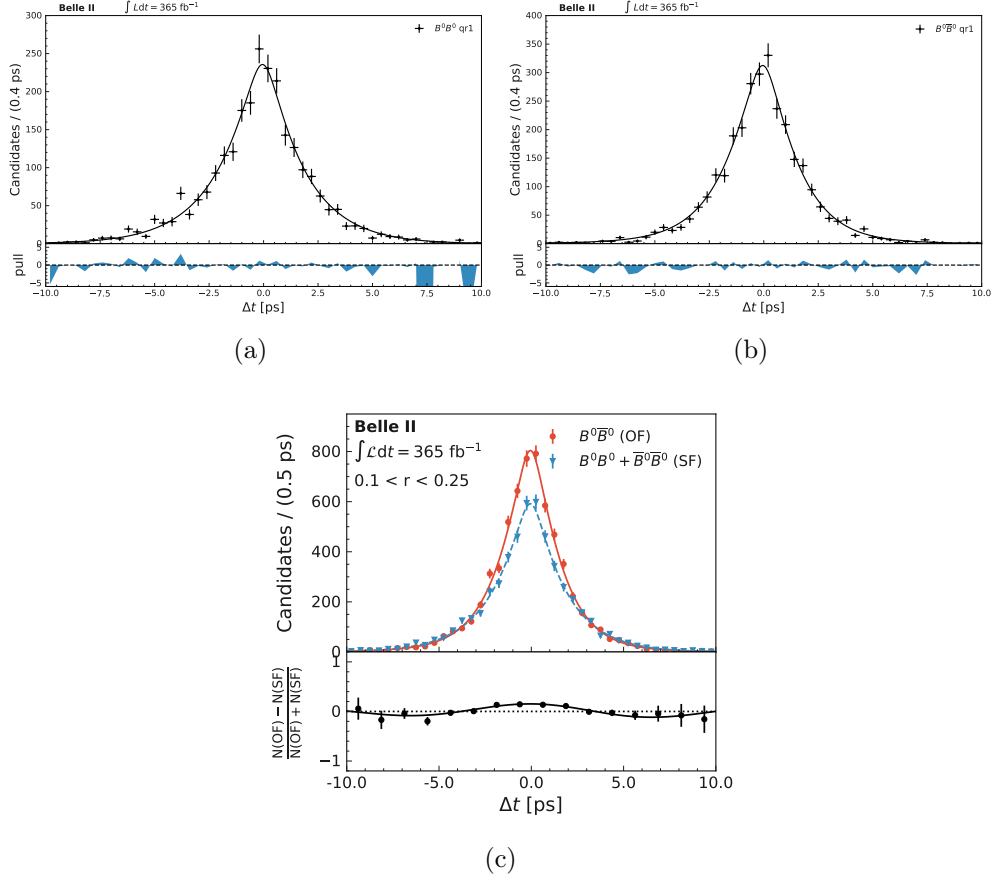


Figure 4.15: Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the second r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).

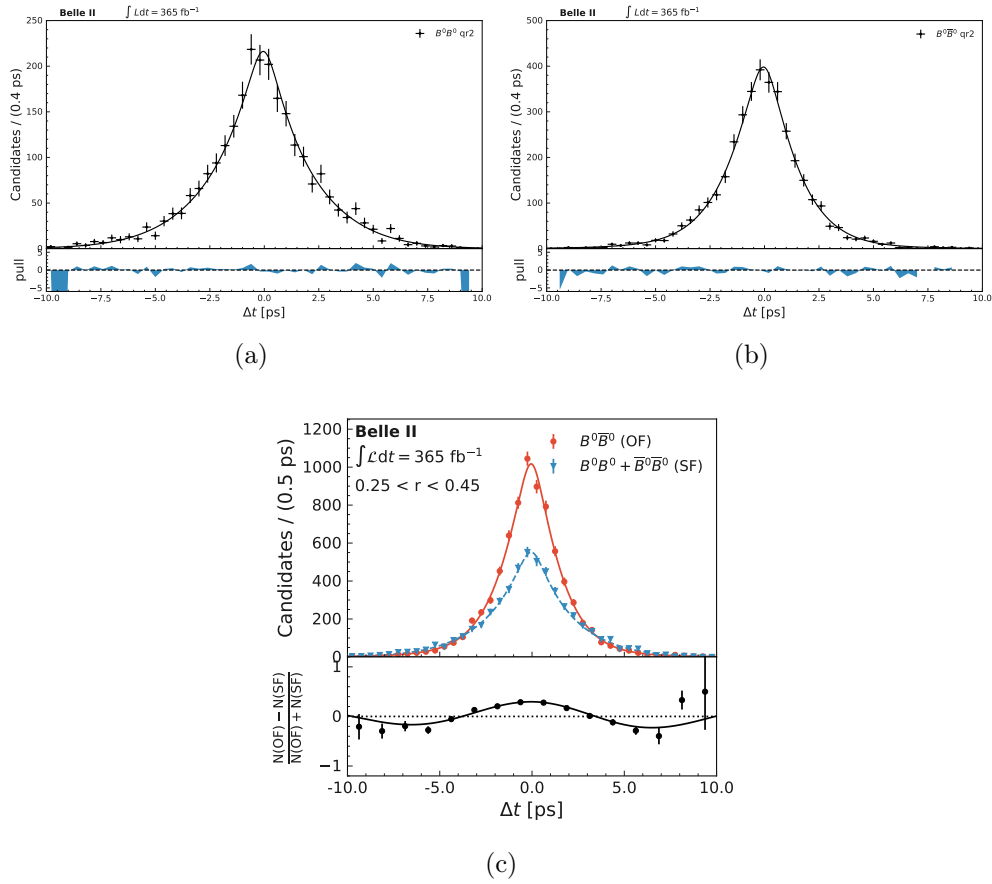


Figure 4.16: Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the third r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).

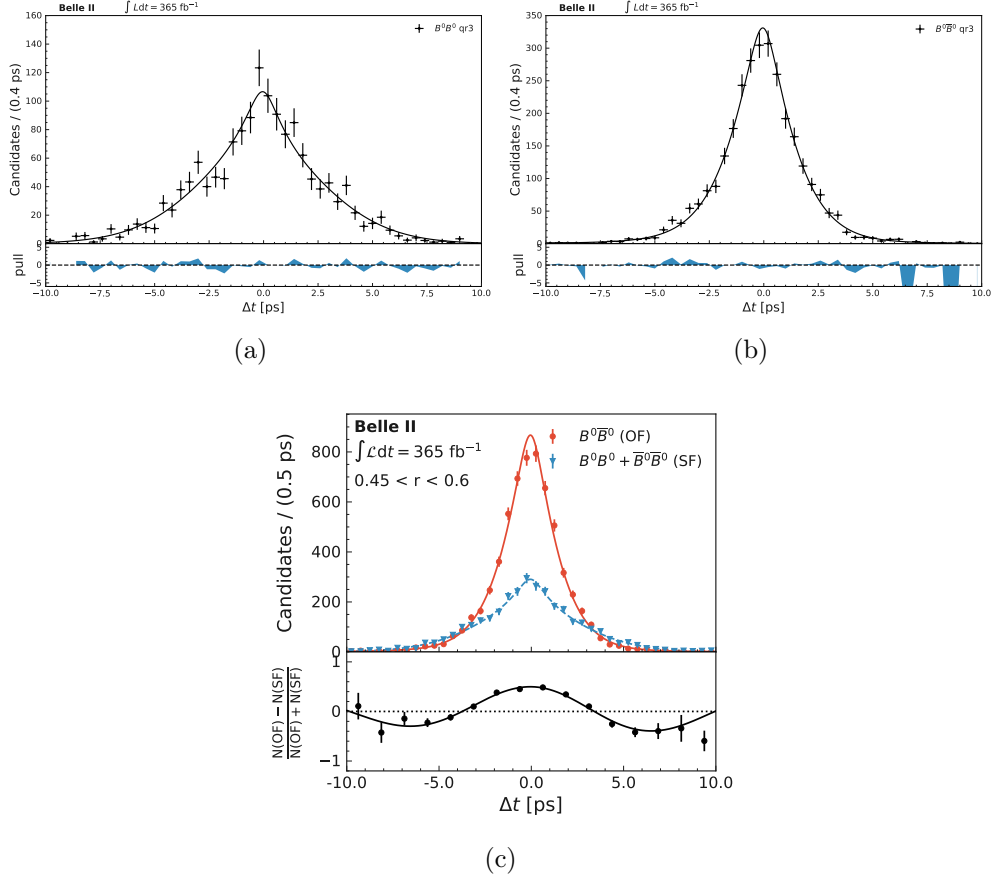


Figure 4.17: Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the fourth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).

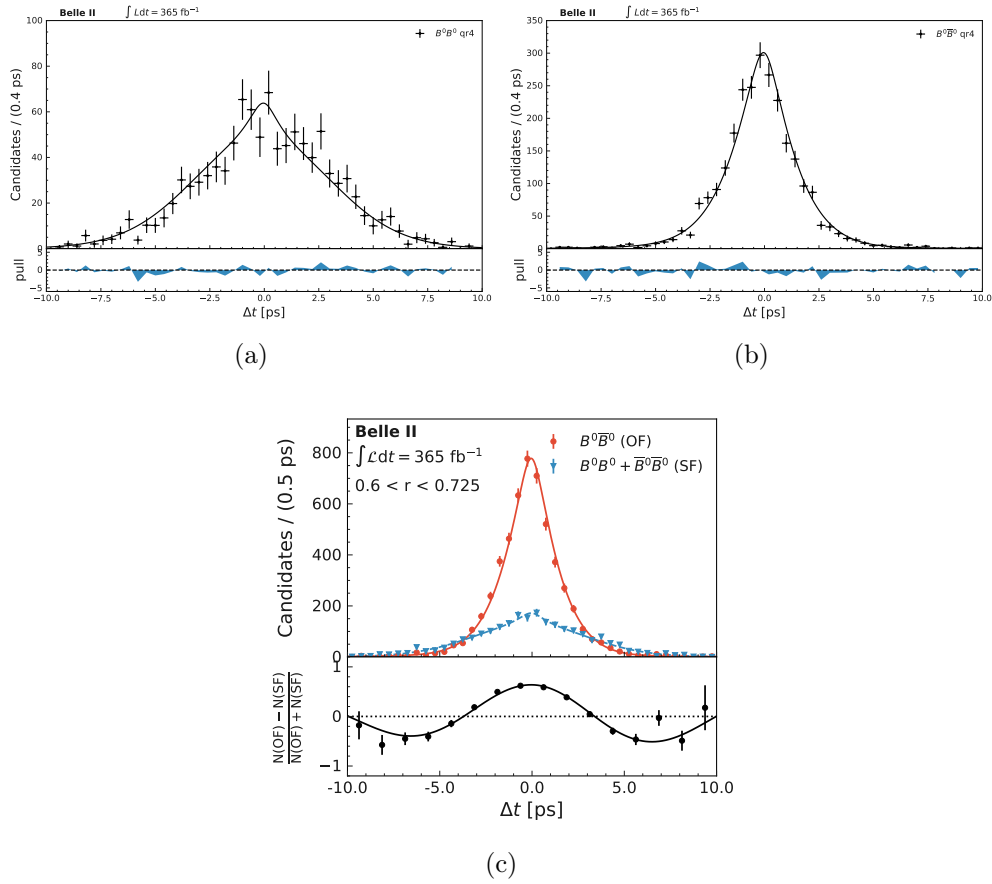


Figure 4.18: Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the fifth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).

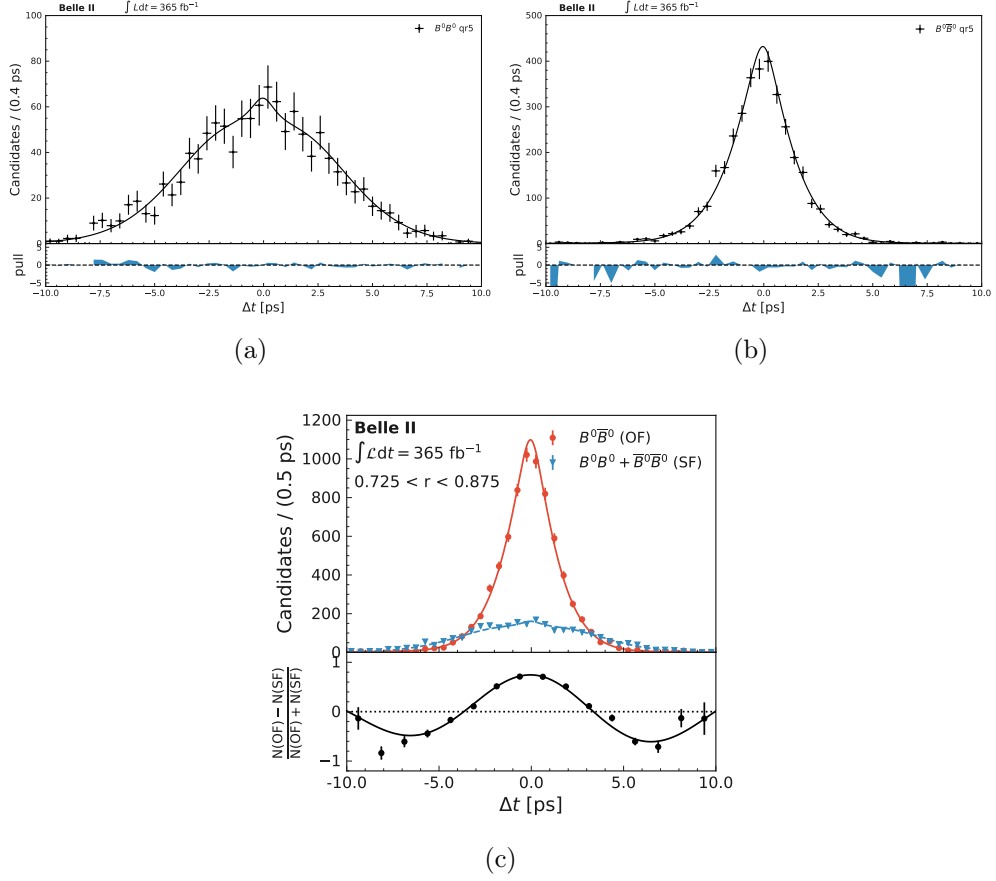


Figure 4.19: Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the sixth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).

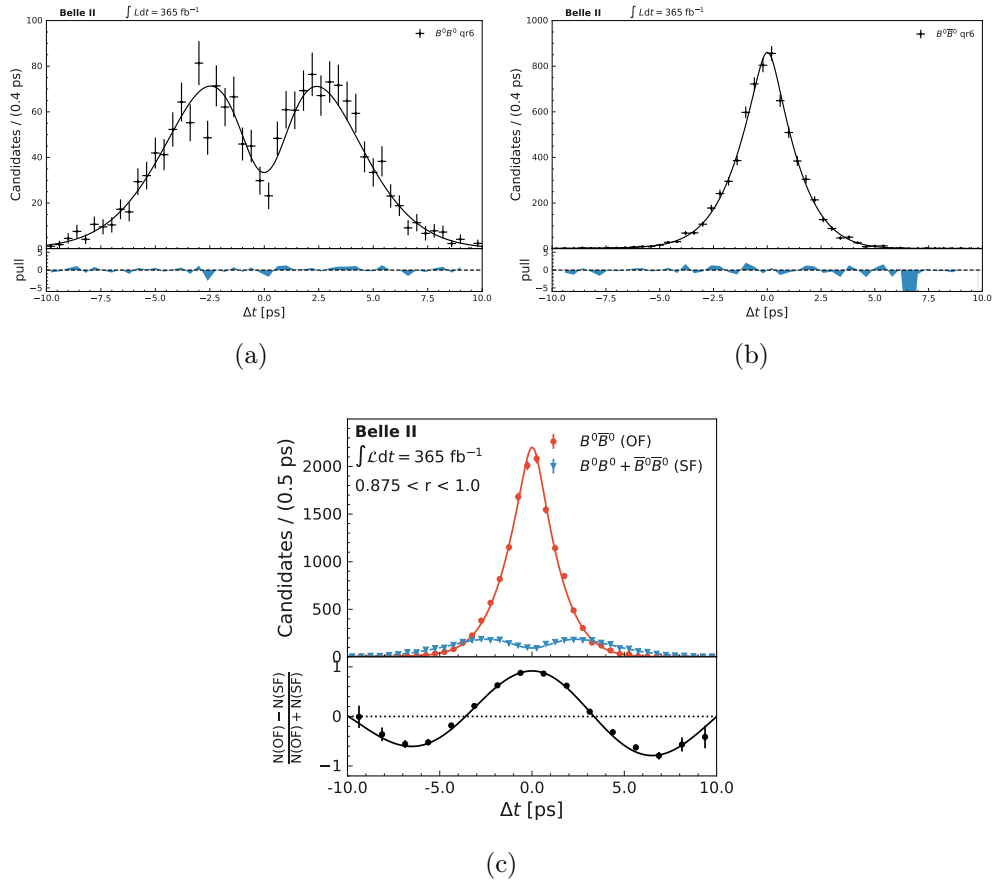


Figure 4.20: Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the seventh r -bin (i.e. the bin with the best tagging quality) for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).

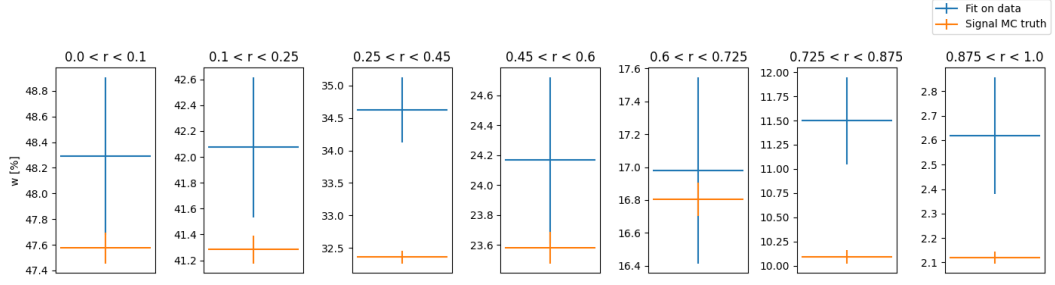


Figure 4.21: Comparison between the wrong tag fractions (w) obtained from the fit to the background-subtracted Δt distribution of the 365 fb^{-1} data sample and their true values obtained from the MC.

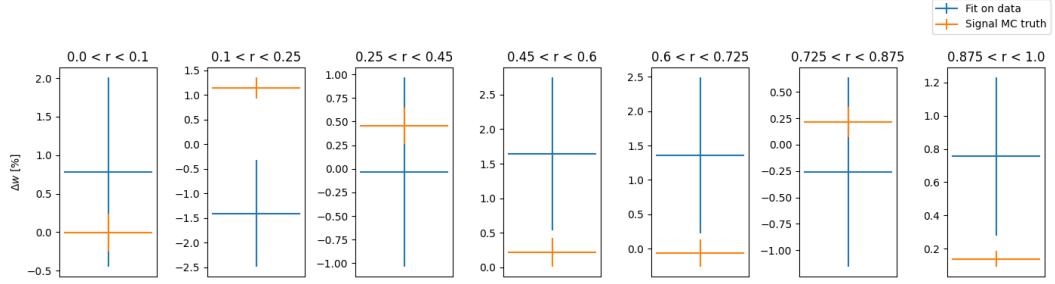


Figure 4.22: Comparison between the wrong tagging asymmetries (Δw) obtained from the fit to the background-subtracted Δt distribution of the 365 fb^{-1} data sample and their true values obtained from the MC.

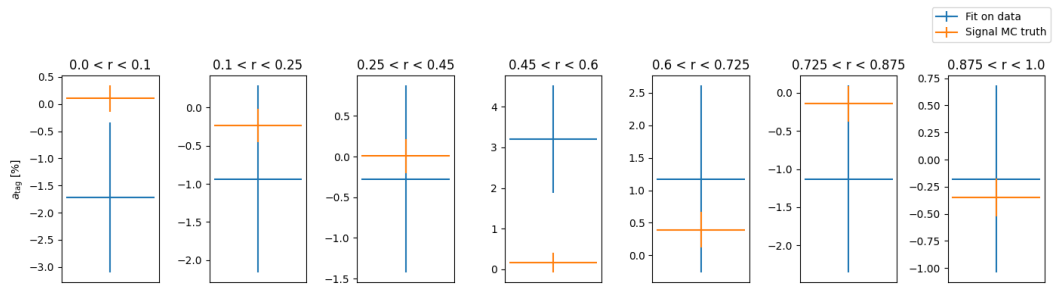


Figure 4.23: Comparison between the tag-side detection efficiency asymmetries (a_{tag}) obtained from the fit to the background-subtracted Δt distribution of the 365 fb^{-1} data sample and their true values obtained from the MC.

The effective tagging efficiency obtained by fitting the real data is once again calculated following Equation 4.1 and is found to be:

$$\varepsilon_{\text{tag}} = (37.40 \pm 0.39)\% , \quad (4.11)$$

where the quoted uncertainty is purely statistical. Due to the fact that `sWeights` were used to evaluate ε_{tag} , the uncertainty provided by the likelihood profile is slightly underestimated with respect to the actual statistical uncertainty expected knowing the size of the data that were analyzed. This is because the background components are not accounted for in the computation of the uncertainty by the likelihood method, and neither is the uncertainty on the value of a_{sig} . To obtain the correct value for the statistical uncertainty on ε_{tag} , the dataset was bootstrapped [105], meaning that events were randomly sampled from it with replacement, effectively generating toy datasets of equal size that exhibit slight variations between them. The full fit procedure explained earlier in this Section is then performed on each of the bootstrap replicas and, from the spread of the resulting distribution of the fitted parameters, the actual value of the statistical uncertainty on ε_{tag} is estimated and is found to be 0.433%.

Systematic uncertainties on the fitted parameters were estimated by taking into account various effects, such as:

- the alignment of the detector,
- the calibration of the parameters of the beam spot,
- the modeling of signal and background events in ΔE ,
- the uncertainty on the world average values of the B^0 lifetime and the $B^0-\bar{B}^0$ oscillation frequency, which are fixed in the Δt fit,
- the uncertainties on the Δt Resolution Function parameters that are fixed in the Δt fit, as explained in Section 4.2.4 and
- the background subtraction method (*sPlot*).

Table 4.8 lists all of the sources of systematic uncertainty that were estimated for ε_{tag} together with the statistical uncertainty for comparison.

The total value of the systematic uncertainty for ε_{tag} , accounting for all of the above sources, amounts to 0.36%. Correcting the value of the statistical uncertainty and adding the value for the systematic uncertainty, we now have:

$$\varepsilon_{\text{tag}} = (37.40 \pm 0.43 \pm 0.36)\% . \quad (4.12)$$

Table 4.8: Systematic uncertainties related to the process of evaluating ε_{tag} using $B^0 \rightarrow D^{(*)-}\pi^+$ decays recorded by Belle II. The uncertainty due to the sample statistics is also listed for comparison.

Source	Uncertainty [%]
Detector alignment	0.080
Interaction region	0.157
Beam energy	0.029
Background ΔE fit shape	0.114
Signal ΔE fit shape	0.083
Background subtraction w/ sWeights	0.243
Fixed Resolution Function parameters	0.067
τ_{B^0} and Δm	0.061
$\sigma_{\Delta t}$ binning	0.039
Δt -fit bias	0.093
Total systematic uncertainty	0.359
Statistical uncertainty	0.433

4.3 Conclusion and outlooks regarding the calibration of the Flavor Tagger

Comparing Equation 4.12 to Equation 4.10, it is clear that the effective tagging efficiency is worse in the data, which is a direct result of the wrong tag fractions being larger in the data compared to the MC. This underscores the importance of evaluating the Flavor Tagger’s performance in a data-driven manner. Notably, the graph-neural-network-based Flavor Tagger achieves an **18% relative improvement** in effective tagging efficiency over the category-based approach. This represents a substantial gain that directly enhances the sensitivity of all Belle II measurements that rely on Flavor Tagging.

Moreover, it was shown (see Appendix C.2), that there is indeed almost no correlation between the parameters of the Flavor Tagger and any of the Δt Resolution Function parameters when measured in this way. This means that the calibrated set of Flavor Tagger parameter values can be used in both time-dependent and time-integrated measurements, without introducing any significant bias or loss of precision.

The first Belle II measurement demonstrating the gain in sensitivity resulting from using **GFlaT** was that of the time-dependent CP asymmetry in $B^0 \rightarrow J/\psi K_S^0$ decays⁶, which is the so-called *golden channel* for measurements of $\sin 2\beta$. This is

⁶with $J/\psi \rightarrow \mu^+\mu^-$ and $K_S^0 \rightarrow \pi^+\pi^-$.

because of the large branching ratio for this decay, due to it being mostly a tree-level process, as well as its reasonable reconstruction efficiency, leading to quite a sizable data sample. This coupled to the fact there is almost no contribution expected from BSM processes, due to the negligible contribution from penguin diagrams, allows for a very precise estimation of one of the fundamental parameters of the SM, that being the unitarity triangle angle β . The measured CP violation parameters, S (or $\sin 2\beta$) and C , were found to be 8% and 7% more precise, respectively, than they would be if measured using the category-based Flavor Tagger. This gain will become more important in the future, with more data having been collected by Belle II, but its impact can already be seen in early measurements like this one that only utilise a small fraction of the expected Belle II dataset.

Chapter 5

Measurement of the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays

Contents

5.1	Overview of the measurement	112
5.2	Signal reconstruction and event selection	114
5.2.1	Signal-side B decay reconstruction	114
5.2.2	Vertex and decay time reconstruction	116
5.2.3	Continuum background suppression	118
5.2.4	Single candidate selection	120
5.2.5	Selection efficiencies	122
5.2.6	Control mode datasets and selection strategy	124
5.3	Measurement strategy	125
5.3.1	3D fit model	125
5.3.2	Signal modeling	127
5.3.3	Background modeling	135
	SCF background	135
	$B\bar{B}$ background	137
	Continuum background	139
5.3.4	3D fit to a simulated sample	142
5.4	Validation of the fit strategy	154
5.4.1	Validation using the signal mode	154

Fit validation with bootstrapped datasets	154
Fit validation with toy datasets	164
Agreement between data and simulation	190
Fit for the B^0 lifetime	192
5.4.2 Validation using the control mode	193
3D fit to a simulated sample	195
Fit validation with bootstrapped datasets	205
3D fit to the data	210
5.5 Systematic uncertainties	219
5.5.1 Component fit shape parameters	219
5.5.2 Flavor Tagger parameters	219
5.5.3 Δt resolution function parameters	220
5.5.4 B^0 lifetime and B^0 – $\bar{B}^0$ oscillation frequency	221
5.5.5 Fixed SCF background fractions	222
5.5.6 CP violation in the $B\bar{B}$ background	223
5.5.7 Effect of biased yields on CP parameters	224
5.5.8 Bootstrapping bias	224
5.5.9 Misalignment	225
5.5.10 Tag-side interference	226
5.5.11 Summary of systematic uncertainties	228
5.6 Results	229

In this last chapter, the measurement of the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays is presented, based on the Run 1 dataset of Belle II corresponding to 387 million $B\bar{B}$ pairs (365 fb^{-1}) collected at the $\Upsilon(4S)$. Although not detailed here, the same measurement was also performed on the full Belle dataset (711 fb^{-1}), and the two results were subsequently combined, yielding significantly improved precision on the measured CP parameters.

5.1 Overview of the measurement

As explained in Section 1.6.2, the photon polarisation in $b \rightarrow s\gamma$ transitions can be probed by measuring the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays. This observable is expressed in terms of two parameters, S and C , describing

the mixing-induced and direct CP violation, respectively, such as:

$$\begin{aligned} \mathcal{A}_{CP}(\Delta t) &\equiv \frac{\Gamma(\bar{B}^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma) - \Gamma(B^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma)}{\Gamma(\bar{B}^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma) + \Gamma(B^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma)} \\ &= S \sin(\Delta m_d \Delta t) - C \cos(\Delta m_d \Delta t) . \end{aligned} \quad (5.1)$$

An additional pair of parameters, S^+ and S^- , is defined by splitting the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane into two halves and measuring S simultaneously in both. The first measurement of these parameters is presented in this thesis as a proof of concept. These two measurements can be translated into additional new constraints on the C_7'/C_7 complex plane when combined with the measurement of two dilution-like factors a and b [27].

The dataset used for this measurement corresponds to 365 fb^{-1} of collision data recorded by Belle II at the $\Upsilon(4S)$ resonance (“on-resonance data”) during the experiment’s first period of data taking, or Run 1. An additional 42 fb^{-1} of data, collected slightly below the $\Upsilon(4S)$ resonance (“off-resonance data”) is used to validate the agreement between data and simulation.

To develop the signal selection strategy and validate the fit procedure, several simulated MC samples are used. Specifically, a run-dependent¹ “generic MC” sample corresponding in size to 4 times the experimental data, i.e. 1.46 ab^{-1} , is used, which includes both $B\bar{B}$ and $q\bar{q}$ events. For comparison with the off-resonance data, an off-resonance MC sample corresponding to 168 fb^{-1} is also used. In addition, a run-dependent “signal MC” sample is produced with ~ 100 times the signal yield expected in the Run 1 dataset. This sample only includes the signal events where the intermediate state corresponds to the CP eigenstate, namely $B^0(\bar{B}^0) \rightarrow K_1(1270)\gamma \rightarrow K_S^0 \rho^0 \gamma \rightarrow K_S^0 \pi^+ \pi^- \gamma$, whereas the signal events in the generic MC sample are generated with all the known intermediate states, including both CP and non- CP eigenstate contributions.

The dedicated signal MC sample is only used in certain aspects of the selection procedure, such as training the continuum-suppression algorithm (see Section 5.2.3), and for determining the shape parameters of the PDFs describing the signal and background components (see Sections 5.3.2 and 5.3.3). Consequently, the fact that all kaonic intermediate states leading to the signal decay are not included at generator level in this sample does not significantly impact the time-dependent CP asymmetry measurement. Lastly, signal MC samples generated with non-zero S values are also produced for the purpose of testing the fit linearity (see Section 5.4). The input S values are specified using the HELAMP model

¹As a reminder, run-dependent MC refers to simulated samples in which beam and detector conditions are varied across runs to better reflect the changing experimental environment, rather than assuming static conditions.

within EVTGEN, by modifying the value of $\sin 2\beta$, which is equivalent to S in the case of B_d mesons [106].

5.2 Signal reconstruction and event selection

The process of reconstructing the decay of the signal-side B meson into the final state of interest, $K_S^0 \pi^+ \pi^- \gamma$, is described in this section, alongside the event selection strategy followed to suppress contributions from background processes.

5.2.1 Signal-side B decay reconstruction

The first step in the reconstruction process is to build the signal-side B decay from its final-state particles. As outlined in Section 5.2, this task is performed within the BASF2 framework, where dedicated modules handle different aspects of the reconstruction.

Prompt pions

The prompt pions are reconstructed from charged particle tracks that satisfy $dr < 0.5$ cm and $dz < 3.0$ cm, corresponding to the transverse and longitudinal displacements from the Interaction Point (IP), respectively. As explained in Section 2.6.2, particle identification information is used to form likelihood ratios for each particle hypothesis, and for pions the likelihood ratio is defined as $\text{PID}_\pi = \mathcal{L}_\pi / \sum_i \mathcal{L}_i$, where $i = \{e, \mu, \pi, K, p, d\}$. A relatively loose selection threshold is set on this likelihood ratio to remove misidentified prompt pion candidates.

As explained in Section 1.6.2, the signal decay of interest must go through an intermediate state, $K_{\text{res}} \rightarrow K_S^0 \rho^0$, with $\rho^0 \rightarrow \pi^+ \pi^-$, in order for the final state to be a CP eigenstate and for interference between the B^0 and $\bar{B}^0$ decay amplitudes to take place. Thus, a requirement on the invariant mass of the prompt pions is imposed, $0.6 < M_{\pi^+ \pi^-} < 0.9$ GeV/ c^2 , symmetric around the ρ^0 meson mass. This ρ^0 mass window enriches the data sample in B_{sig} decays that pass through the ρ^0 resonance, but does not eliminate all the other possible decay paths, which will end up diluting the final measurement of S . Notably, this is the only selection criterion that affects the relative contribution of intermediate states which are not CP eigenstates.

K_S^0 candidates

As described in Section 2.6.3, a K_S^0 candidate is formed using two charged particle tracks, in this case pions, and a vertex fit is performed on them. Only candidates with an invariant mass in the range $0.45 < M_{K_S^0} < 0.55$ GeV after the vertex fit,

and for which the vertex fit did not fail, are kept. A multivariate analysis (MVA) classifier is then used to further purify the remaining set of K_S^0 candidates. This MVA classifier, dubbed K_S^0 MVA, is based on a Boosted Decision Tree (BDT) that can be used to discriminate between correctly and incorrectly reconstructed K_S^0 's. More details about how a BDT classifier works are given in Section 5.2.3.

The K_S^0 MVA has been pre-trained using the kinematical variables related to the K_S^0 and its daughter pions, the K_S^0 flight length, as well as the number of hits in the PXD and the SVD. An optimal cut on the output of the K_S^0 MVA classifier is found by maximising the significance, $N_S/\sqrt{N_S + N_B}$, where N_S is the number of truth-matched K_S^0 candidates and N_B is the number of fake K_S^0 candidates found in the signal MC sample.

Photon candidates

As explained in Section 2.6.4, photon candidates are formed from clusters in the ECL, which have no matched tracks in the CDC. A requirement is placed on the energy of these clusters in the lab frame, $1.5 < E_\gamma < 4.0$ GeV, as well as on the photon momentum polar angle, $-0.87 < \cos \theta_\gamma < 0.95$. To suppress background from high-momentum π^0 decays, the candidate photon is paired with other photons in the event, and for each pair an MVA variable based on a pre-trained FastBDT classifier is evaluated [107]. Candidates consistent with originating from $\pi^0 \rightarrow \gamma\gamma$ decays are vetoed by applying a requirement on the output of the BDT classifier.

The inputs to the BDT classifier are the diphoton invariant mass and various properties for the low-energy photon, such as its energy, polar angle, distance between the ECL cluster and tracks extrapolated from the CDC, and certain Zernike moments which characterise the ECL shower shape [108]. The BDT classifier output, dubbed “ π^0 likeness”, is a continuous variable ranging from 0 (least consistent with a π^0 photon) to 1 (most consistent with a π^0 photon). The selection threshold on the π^0 likeness is optimised by maximising the signal significance, $N_S/\sqrt{N_S + N_B}$, where N_S here is the number of truth-matched signal B candidates and N_B is the number of fake B candidates found in the signal MC sample.

B meson reconstruction

The K_{res} candidates are built from the four-momenta of the K_S^0 , π^+ , and π^- candidates, with a requirement on their invariant mass, $0.9 < M_{K_{\text{res}}} < 1.8$ GeV/ c^2 . The upper bound on the K_{res} invariant mass is imposed to suppress contributions from charm resonances, which become significant above 1.8 GeV/ c^2 .

Finally, the B_{sig} candidates are formed from the K_{res} and photon candidates, with the following requirements: $-0.2 < \Delta E < 0.2$ GeV and $5.2 < M_{\text{bc}} < 5.3$ GeV/ c^2 , where ΔE is the energy difference and M_{bc} the beam-constrained

mass, as they were defined in Equations 2.7 and 2.8 back in Section 2.6.5.

The selection criteria applied to the data sample up to this point are summarised in Table 5.1. An additional set of selection criteria is also applied at the event-level, in order to reduce the contribution from beam-induced background. These are:

- Having at least 4 charged particle tracks in the event,
- At least 3 of these tracks have more than 20 hits in the CDC and also have $|z_0| < 2.0$ cm and $d_0 < 0.5$ cm²,
- There are at least 2 ECL clusters with $17^\circ < \theta_{\text{cluster}} < 150^\circ$ and $E_{\text{cluster}} > 0.2$ GeV,
- The visible energy in the C.o.M. frame is higher than 4 GeV,
- The combined energy deposited in the ECL by pions and photons is lower than 9 GeV.

Table 5.1: List of selection criteria applied to the prompt $\pi^\pm$, K_S^0 , photon and B candidates in the data sample.

Variable	Criterion	Units
Prompt $\pi^\pm$ dr	< 0.5	cm
Prompt $\pi^\pm$ dz	< 3.0	cm
Prompt $\pi^\pm$ PID	> 0.1	–
$M_{K_S^0}$	$0.45 < M_{K_S^0} < 0.55$	MeV/ c^2
K_S^0 MVA	> 0.95	–
E_γ	$1.4 < E_\gamma < 4.0$	GeV
$\cos \theta_\gamma$	$-0.87 < \cos \theta_\gamma < 0.96$	–
Veto $_{\pi^0 \rightarrow \gamma\gamma}$	π^0 likeness < 0.7	–
$M_{K_{\text{res}}}$	$0.9 < M_{K_{\text{res}}} < 1.8$	GeV/ c^2
$M_{\pi^+\pi^-}$	$0.6 < M_{\pi^+\pi^-} < 0.9$	GeV/ c^2
ΔE	$-0.2 < \Delta E < 0.2$	GeV
M_{bc}	$5.2 < M_{\text{bc}} < 5.3$	GeV/ c^2

5.2.2 Vertex and decay time reconstruction

The B_{sig} decay vertex position is determined by fitting the common origin of the two prompt π^+ and π^- tracks, using the TREEFIT algorithm [97], with the

²As a reminder, z_0 is the signed distance of the track perigee from the origin along the z -axis and d_0 is the distance of the perigee from the origin in the transverse plane.

additional constraint that its flight path needs to be consistent with the profile of the IP. Including the pion tracks arising from the K_S^0 decay in the vertex fit has shown no significant improvement on the vertex resolution (see Appendix D.1), therefore these tracks are not used for this purpose. Omitting the K_S^0 from the vertex fit here also allows to make use of the isospin partner of the signal decay mode, $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$, as a control channel, where the charged kaon track is not used in the determination of the B_{sig} vertex. More details on the reconstruction of this mode are discussed in Section 5.2.6.

The position of the B_{tag} decay vertex is determined by fitting to a common origin all remaining tracks in the Rest-Of-Event (ROE) that were not used to reconstruct the B_{sig} candidate. This fit is performed using the RAVE algorithm [96], which also accounts for tertiary tracks coming from D meson decays, as explained in Section 4.2.2. The *tube* constraint is also used in this fit, which ensures the fitted B_{tag} vertex is found within a given cylindrical volume around its flight direction, which is determined from the IP profile. After the B_{tag} vertex fit, the flavor of the B_{sig} in each event is determined at the time of the B_{tag} decay, using the GNN-based Flavor Tagger, GFlaT, in the same manner described in Section 4.1.

As explained back in Section 2.1, the pair of B mesons produced from the $\Upsilon(4S)$ decay is boosted along the z -axis³, which allows to determine the position of the B_{sig} and B_{tag} decay vertices, and subsequently measure Δt , through the relation $\Delta z = \beta \gamma c \Delta t$. However, this is an approximate relation, since the B mesons have a small transverse component to their motion in the $\Upsilon(4S)$ frame. This so-called “Kinematic Approximation” is accounted for using the following transformation:

$$\Delta t^{\text{corr}} = \frac{\Delta t^{\text{approx}} - (\alpha/\beta) \cos\theta_B^* \tau_{B^0}}{1 + s (\alpha/\beta) \cos\theta_B^*} \quad (5.2)$$

where Δt^{approx} is the value of Δt obtained using Δz , $\alpha = p_B^* c / m_B c^2 \approx 0.06$ is the boost factor of the B mesons in the rest frame of the $\Upsilon(4S)$, $\beta = p_{\text{beam}} c / E_{\text{beam}} \approx 0.28$ is the boost factor of the $\Upsilon(4S)$ in the lab frame, $\cos\theta_B^*$ is the cosine of the polar angle of the B^0 in the rest frame of the $\Upsilon(4S)$, s is the sign of Δz and τ_{B^0} is the B^0 lifetime. This transformation has also been used by *BABAR* in time-dependent analyses (see Ref. [45], page 80, for more details).

Figure 5.1 shows a fit to the Δt distribution of simulated signal events found in one of the signal MC samples, to extract the B^0 lifetime, with and without correcting for the small motion of the B mesons in the $\Upsilon(4S)$ rest frame. It can be seen that the value for the B^0 lifetime obtained in the latter case is slightly biased with respect to the lifetime value used during event generation, which in the case of the Belle II Monte Carlo is 1.519 ps. Thus, whenever Δt is used in the following fits, it should be understood that it is Δt^{corr} and not Δt^{approx} .

³The Lorentz boost is along an axis that is almost collinear to the z -axis due to the very small

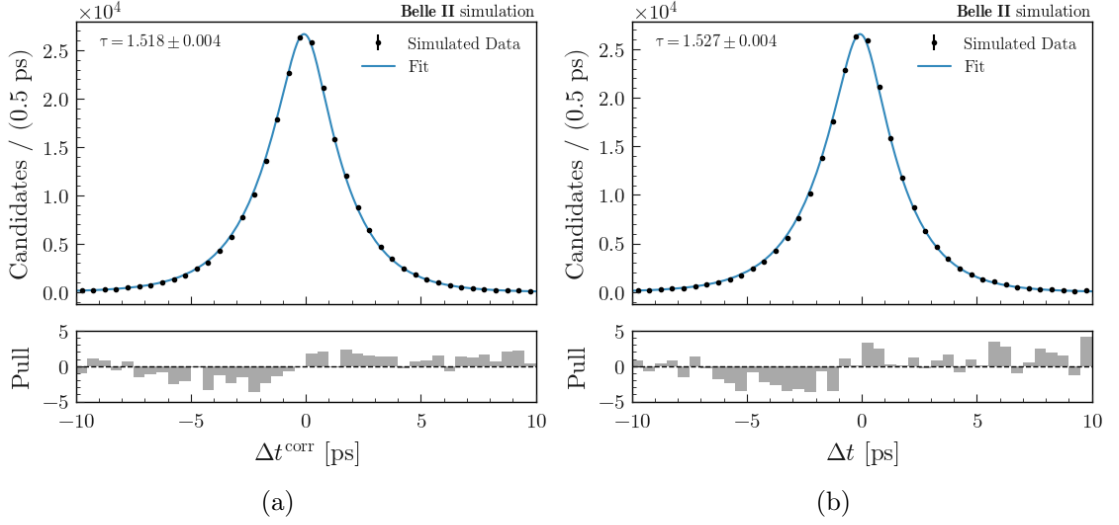


Figure 5.1: Fit for the B^0 lifetime of simulated signal events (a) with and (b) without correcting for the small motion of the B mesons in the $\Upsilon(4S)$ rest frame.

5.2.3 Continuum background suppression

The most dominant source of background for this measurement is due to non-resonant $e^+e^- \rightarrow q\bar{q}$ events, where $q = \{u, d, s, c\}$, also referred to as $q\bar{q}$ background. These events have a jet-like topology, as opposed to $e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$ events which are isotropic due to the B mesons being produced almost at rest in the $\Upsilon(4S)$ frame (see Figure 5.2). This difference in the event topology between $q\bar{q}$ and $B\bar{B}$ events can be exploited in order to suppress the continuum background.

Specifically, a continuum-suppression multivariate analysis (CSMVA) classifier, that is based on a BDT, is specifically built and trained for this measurement. Decision Trees [109] classify events by applying successive cuts on input variables, such as event-shape observables, to distinguish between signal- and background-like event topologies. While a single tree tends to be sensitive to statistical fluctuations and limited in performance, the *boosting* procedure [110] overcomes this by building an ensemble of trees. Each new tree is trained to focus on events that were misclassified by the previous ones, and the final classifier output is obtained as a weighted combination of all trees. The resulting BDT response provides a single continuous variable that acts as a powerful discriminator between signal and continuum background events. The set of input variables used in the BDT are:

- $\cos \theta_B^*$, the polar angle of the B_{sig} in the C.o.M. frame,

crossing angle between the electron and positron beams.

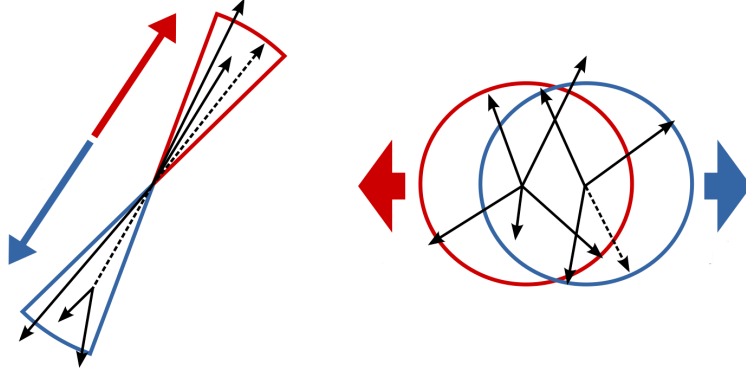


Figure 5.2: Illustration of continuum (left) and $B\bar{B}$ (right) event topologies, with resulting momenta.

- $\cos\theta_{\text{TBTO}}$, the angle between the thrust axis of B_{sig} and that of the ROE. The thrust axis maximises the total projection function $T(\hat{n})$ of a unit vector $\hat{n}$, defined as:

$$T(\hat{n}) = \frac{\sum_i |\vec{p}_i \cdot \hat{n}|}{\sum_i |\vec{p}_i|}, \quad (5.3)$$

where $\vec{p}_i$ is the momentum of the i^{th} daughter particle of the B meson, and

- the following Modified Fox-Wolfram moments: E_T , $h_{(0,0)}^{\text{SO}}$, $h_{(0,2)}^{\text{SO}}$, $h_{(0,4)}^{\text{SO}}$, $h_{(1,0)}^{\text{SO}}$, $h_{(1,2)}^{\text{SO}}$, $h_{(2,0)}^{\text{SO}}$, $h_{(2,2)}^{\text{SO}}$ and $h_{(2,4)}^{\text{SO}}$. Here, E_T is the sum of transverse momenta of all particles in the event. The Fox-Wolfram moments are defined as:

$$H_{(m,l)}^{\text{SO}} = \sum_{i,j} \frac{|p_i||p_j|}{E_{\text{event}}^2} P_l(\cos\theta_{ij}), \quad (5.4)$$

where p_i is the momentum of the i^{th} particle, E_{event} is the collision energy, P_l is the l^{th} Legendre polynomial and θ_{ij} is the angle between the momenta of i^{th} and j^{th} particles [104]. The superscript SO specifies that the moment is calculated between the signal (B_{sig}) and the other particles in the event, while the subscript m denotes whether the other particle is charged ($m = 0$), neutral ($m = 1$), or missing ($m = 2$). The Modified Fox-Wolfram moments are then defined as:

$$h_{(m,l)}^{\text{SO}} = \frac{H_{(m,l)}^{\text{SO}}}{\sum_{i,j} |p_i||p_j|}, \quad (5.5)$$

normalised such that they are of order $\mathcal{O}(1)$.

The continuum-suppression BDT classifier is trained on simulated signal and continuum background candidates which are found in the signal MC and generic

MC samples, respectively, after the selection criteria listed in Table 5.1 have been applied to them. As is standard practice, an independent set of signal and continuum background samples is reserved for testing the BDT. The training is performed using the XGBOOST package [111], while its hyperparameters, such as the number of trees or their maximum depth, are optimised with the OPTUNA library [112]. This specific set of input variables for the BDT was chosen and optimised for the parallel Belle analysis, and BDT classifiers using alternative sets of inputs were also tested, showing little to no impact on the final measurement. The distributions of each input variable, as well as the correlations between them for signal and continuum background candidates, are presented in Appendix D.2. No significant effect is expected from the observed correlations, and any residual impact would only result in a sub-optimal selection rather than a bias.

Figure 5.3 shows the CSMVA classifier output for simulated signal and continuum background candidates, where it can be seen that the classifier achieves good separation power and there is no overtraining present. A selection criterion on the CSMVA classifier output is optimised by maximising the significance, $N_S/\sqrt{N_S + N_B}$, in a signal enhanced window in ΔE and M_{bc} defined by $-0.2 < \Delta E < 0.1$ GeV and $5.27 < M_{bc} < 5.29$ GeV/ c^2 . The distribution of the significance with respect to various cut values is also shown in Figure 5.3 together with the optimal cut value. The optimal cut on the CSMVA classifier output results in $\sim 70\%$ of signal candidates being retained and $\sim 91\%$ of continuum background candidates being rejected.

5.2.4 Single candidate selection

The last stage of the signal selection is to reject multiple B candidates which arise from the same collision event. This is because multiple combinations of the available final state particles candidates in an event ($K_S^0, \pi^\pm, \gamma$) can lead to a B_{sig} candidate being formed, but only one of those combinations represents the true signal B_{sig} candidate. In $\sim 13\%$ of correctly identified signal events, more than one possible B_{sig} candidate is found. Figure 5.4 shows the distribution of the candidate multiplicity. It is worth mentioning that, before any selection is applied, multiple candidates are observed in $\sim 56\%$ of events. These fake candidates arise mainly from mis-reconstructed signal events, rather than from continuum background or alternative B_{sig} decay modes.

In events where only a single B_{sig} candidate is present no action is necessary, whereas, in events with multiple B_{sig} candidates, there are several ways to select the candidate most likely to be the true B_{sig} candidate. For example, B_{sig} candidates in an event can be ranked according to how reconstruction performance, reconstruction uncertainty, lifetime value, etc., and the B_{sig} candidate with the best quality is chosen. This is known as Best Candidate Selection (BCS). For

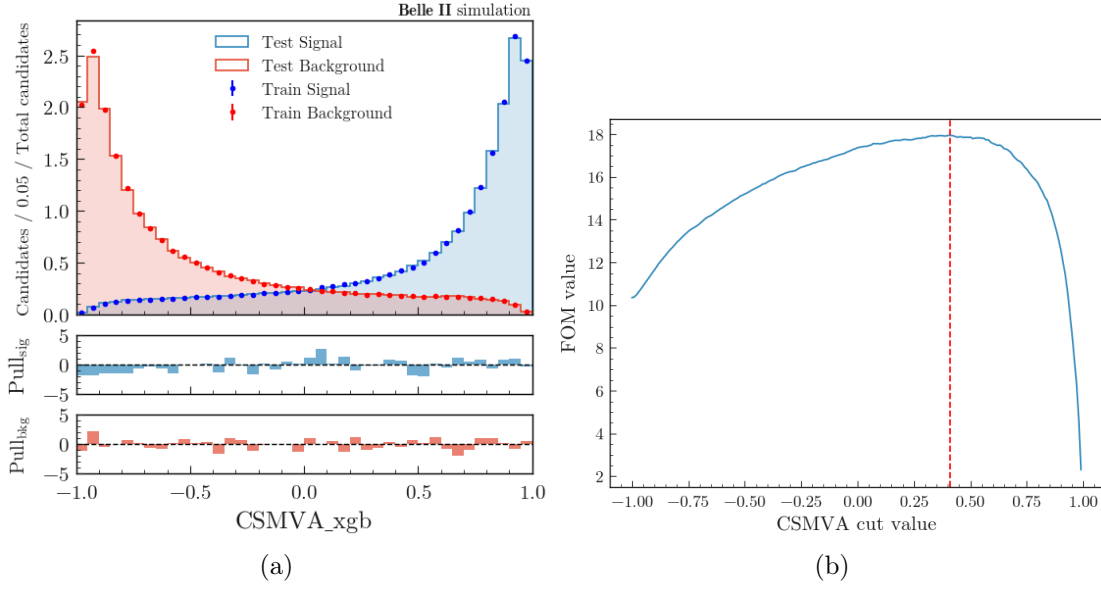


Figure 5.3: Continuum-suppression classifier: (a) normalised train and test distributions for signal and continuum background candidates, (b) Significance as a function of the cut on the CSMVA classifier output. The dashed red line denotes the cut value for which the significance has reached a maximum.

example, good proxies to measure the reconstruction quality are χ^2 distributions such as: track goodness, vertex quality, etc. It is worth noting that the selected proxy used for the BCS may have biasing effects on final distributions that are used in the extraction of the CP parameters, and thus these need to be tested and avoided if possible. Another alternative is to select the B_{sig} candidate randomly, or Random Candidate Selection (RCS), which usually leads to a worse success rate; the metric used to judge the performance of one's single candidate selection strategy. The success rate is defined as the ratio between the number of events with more than one B_{sig} candidate where the true B_{sig} candidate is correctly identified over the total number of events containing more than one B_{sig} candidate.

In this analysis, the initial plan was to select the single candidate based on the χ^2 probability of the B_{sig} vertex fit, since in most events with multiple candidates, the wrong candidates were built using a wrong pion at the prompt pion vertex. However, this led to a bias when determining the B^0 lifetime by fitting the distribution of the generated-level, proper decay time difference, $\Delta\tau$ of true signal events, as shown in Figure 5.5. Compared to Δt , this distribution does not include the resolution effect of the detector in determining the B_{sig} and B_{tag} decay vertices, and it does not include the effect of the small motion of the B mesons in

the $\Upsilon(4S)$ frame. The bias arises from the fact that the quality of the B_{sig} vertex fit is correlated with the B^0 flight length and, therefore, the B^0 lifetime [113]. To avoid this bias, it was decided to select the single candidate randomly, with the cost of a lower success rate ($\sim 45\%$ compared to $\sim 62\%$). The corresponding fit for the B^0 lifetime after random candidate selection is also shown in Figure 5.5 and it is clear that no bias is observed when fitting for the B^0 lifetime after RCS. Overall, switching from BCS to RCS leads to an absolute drop of $\sim 5\%$ in signal efficiency associated to the single candidate selection process ($\sim 89\%$ to $\sim 84\%$).

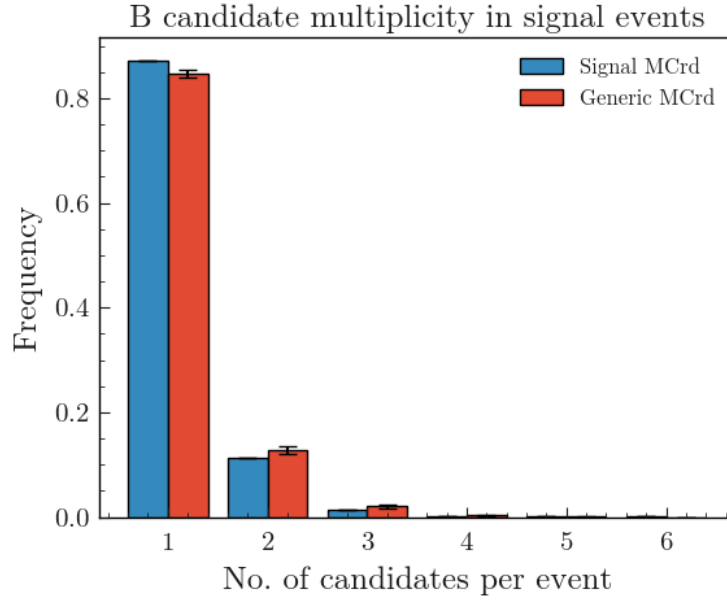


Figure 5.4: Candidate multiplicity after the selection of the signal and cut on the CSMVA classifier output have been applied.

5.2.5 Selection efficiencies

The signal and background efficiencies for the most impactful steps of the signal selection process are summarised in Table 5.2. The explanation of the different backgrounds is given in Section 5.3.

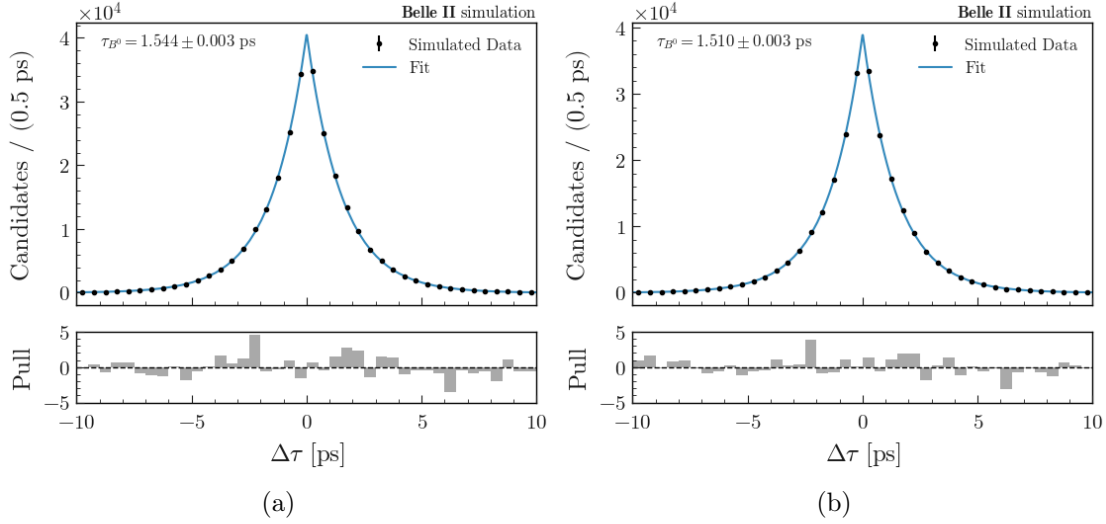


Figure 5.5: Fit of the generated-level $\Delta\tau$ distribution after selecting a single candidate (a) based on the χ^2 probability of the B_{sig} vertex fit and (b) randomly.

Table 5.2: Cumulative signal and background efficiencies for the most impactful steps of the selection procedure. Absolute efficiencies are shown in parentheses. All values are estimated using the generic MC sample.

Selection criterion	Signal (%)	Self-crossfeed (%)	$B\bar{B}$ bkg (%)	$q\bar{q}$ bkg (%)
Prompt $\pi^\pm$ PID	81.93 (81.93)	76.09 (76.09)	67.72 (67.72)	71.99 (71.99)
π^0 veto	69.61 (84.97)	63.19 (83.05)	31.54 (46.57)	17.05 (23.69)
ρ^0 mass window	35.63 (51.19)	24.76 (39.19)	12.77 (40.49)	6.51 (38.21)
K_S^0 MVA > 0.95	32.97 (92.52)	6.95 (28.05)	1.42 (11.15)	0.62 (9.49)
CSMVA > 0.41	23.03 (69.84)	3.45 (49.71)	0.67 (46.86)	0.05 (8.76)
Single candidate	19.38 (85.93)	2.55 (72.50)	0.57 (85.67)	0.05 (91.16)

5.2.6 Control mode datasets and selection strategy

The measurement is validated using the isospin partner mode, $B^+ \rightarrow K^+\pi^+\pi^-\gamma$, as a control mode (see Section 5.4.2). For this purpose, the same dataset is used, corresponding to an integrated luminosity of 365 fb^{-1} . The same generic MC sample is also used here, corresponding to 1.46 ab^{-1} , as well as a dedicated high-statistics signal MC sample that contains ~ 100 times the signal events expected in the Run 1 dataset. Moreover, as with the signal mode, off-resonance samples are used to validate the agreement between data and MC.

The reconstruction and selection of the control mode are designed to closely follow those of the signal mode. The main difference between the two channels lies in the presence of a K^+ in the final state of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ instead of a K_S^0 . Apart from this difference, the reconstruction of the B_{sig} proceeds in the same way. The B_{sig} decay vertex is determined using only the trajectories of the two prompt pions and the same constraint with respect to the IP is applied. The Flavor Tagger is also applied identically, determining the flavor of the B_{sig} (B^+ or B^-) without relying on the kaon charge, even though it is directly correlated with the B_{sig} flavor.

Of course, since there is no K_S^0 present in the final state, the K_S^0 MVA is not utilised. Instead, a requirement is placed on the binary likelihood ratio of the charged kaon, with respect to the pion hypothesis, $\text{PID}(K|\pi) = \mathcal{L}_K/(\mathcal{L}_K + \mathcal{L}_\pi)$, which retains $\sim 90\%$ of the signal and rejects $\sim 40\%$ of the background. This cut was not optimised to reach the highest significance, but to achieve a signal to background ratio that is as similar as possible to that of the signal mode. All of the other selection criteria are identical to those applied to the signal mode samples.

5.3 Measurement strategy

To measure the CP parameters, a *three-dimensional* (3D) *unbinned maximum likelihood fit* (see Appendix B.8) is performed to the distributions of ΔE , M_{bc} and Δt found in the data. Information on the CP observables is present in the signal Δt distribution, while the ΔE , M_{bc} distributions allow to disentangle the signal and background populations.

After the selection procedure described in the previous section, four different populations of candidates are considered.

- Signal: Correctly reconstructed signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ candidates.
- Self-crossfeed (SCF background): Mis-reconstructed true $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ candidates, where, after the full selection has been applied, in $\sim 95\%$ of cases, one of the prompt pions is exchanged with a pion from the tag side, $\sim 4\%$ is due to K_S^0 mis-reconstruction and less than 1% is due to a fake photon. These candidates are considered to be a background contribution, since most of the time the B_{sig} is fitted using the wrong pair of prompt pions, leading to the wrong Δt distribution.
- Continuum ($q\bar{q}$ background): Background candidates arising from non-resonant $e^+e^- \rightarrow q\bar{q}$ events. This is the most dominant source of background after applying the full selection.
- Physical B decays ($B\bar{B}$ background): Candidates arising from generic $B\bar{B}$ decays not leading to the $K_S^0 \pi^+ \pi^- \gamma$ final state. Figure 5.6 shows a breakdown of the different decay modes that make up the $B\bar{B}$ background after the full selection has been applied.

The normalised distributions for each component in ΔE , M_{bc} and Δt as they are found in the generic MC sample are shown in Figure 5.7.

5.3.1 3D fit model

The 3D PDF in ΔE , M_{bc} and Δt is built from the product of each of the three 1D PDFs in the case of signal, SCF and continuum background, while a two-dimensional Kernel Density Estimator (KDE) [114] is used to model $B\bar{B}$ background in M_{bc} and ΔE , which is then combined with the Δt PDF. The reason for this is that the M_{bc} and ΔE distributions show a significant correlation between them for $B\bar{B}$ background events, while for the other three components, correlations between the distributions of the fit variables are found to be negligible, as can be seen in Appendix D.3.

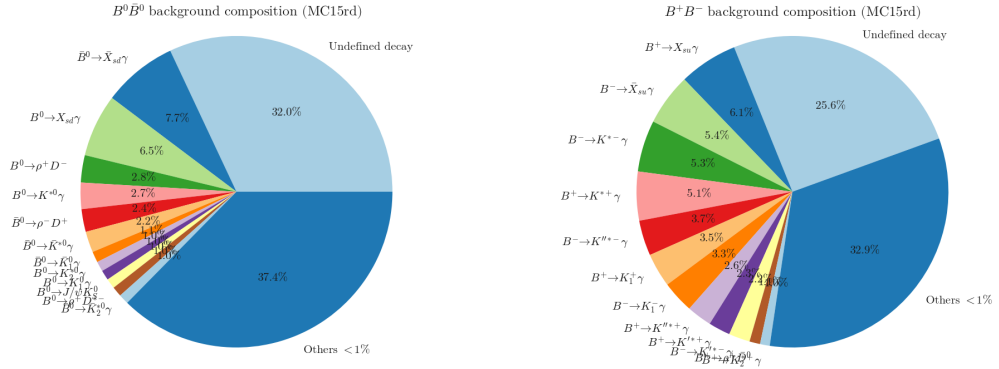


Figure 5.6: Breakdown of $B\bar{B}$ background in the generic MC. The left chart is for $B^0(\bar{B}^0)$ decays, which make up $\sim 27\%$ of the total $B\bar{B}$ background, while the right chart is for $B^+(B^-)$ decays that make up $\sim 73\%$ of the total $B\bar{B}$ background.

Signal and SCF background are considered signal-like, meaning that their total yield is combined and allowed to be free in the 3D fit, while the relative fraction between them is extracted from the simulated samples and is fixed in the 3D fit. The yields of the continuum and $B\bar{B}$ backgrounds are allowed to float separately in the 3D fit. Each of the individual component 3D PDF models (except for $B\bar{B}$ background) has the form:

$$\mathcal{P}_{\text{3D}}^i(\Delta t, \Delta E, M_{\text{bc}}) = \mathcal{P}^i(\Delta t) \times \mathcal{P}^i(\Delta E) \times \mathcal{P}^i(M_{\text{bc}}) \quad (5.6)$$

where i runs over the four fit components. The full 3D PDF model then takes the form:

$$\mathcal{P}_{3\text{D}}^{\text{full}} = N_{\text{sig-like}} \times (f_{\text{sig}} \mathcal{P}_{3\text{D}}^{\text{sig}} + f_{\text{SCF}} \mathcal{P}_{3\text{D}}^{\text{SCF}}) + N_{B\bar{B} \text{ bkg}} \times \mathcal{P}_{3\text{D}}^{B\bar{B} \text{ bkg}} + N_{q\bar{q} \text{ bkg}} \times \mathcal{P}_{3\text{D}}^{q\bar{q} \text{ bkg}} \quad (5.7)$$

The fractions are expressed as: $f_X = N_X / (N_{\text{sig}} + N_{\text{SCF}})$, where N_X is the number of candidates in either the signal or SCF background component, evaluated using a luminosity-equivalent simulated sample corresponding to 25% of the generic MC. Since it is not *a priori* known whether these fractions remain the same across all flavor tagging r -bins, they were initially treated as r -bin dependent. However, after checking the generic MC (see Figure 5.8), it became clear that f_{SCF} , and by extension f_{sig} , is essentially stable across the first six r -bins, with a noticeable deviation in the last r -bin. Therefore, two sets of fractions are adopted: one for the first six r -bins and a separate set for the last r -bin. The corresponding values, together with the average across all r -bins, are listed in Table 5.3.

Table 5.3: Relative fractions between the signal and SCF background components determined from simulation and further fixed in the fit to the data.

Component	f_{avg}	f_{0-5}	f_6
Signal	0.647	0.622	0.731
SCF	0.353	0.378	0.269

The modeling of the ΔE , M_{bc} and Δt distributions of signal events is described in the following section.

5.3.2 Signal modeling

The signal Δt PDF is the one defined in Equation 1.49, albeit with a few modifications. Firstly, inefficiencies due to flavor tagging are accounted for by incorporating the Flavor Tagger parameters into the signal Δt PDF, such as:

$$\begin{aligned} \mathcal{P}_{\text{CP}}^{\text{FT}}(\Delta t; q_{\text{tag}}, w, \Delta w, a_{\text{tag}}) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}} & \left\{ 1 - q_{\text{tag}}\Delta w + q_{\text{tag}}a_{\text{tag}}(1 - 2w) \right. \\ & + [q_{\text{tag}}(1 - 2w) + a_{\text{tag}}(1 - q_{\text{tag}}\Delta w)] \\ & \times [S \sin(\Delta m_d \Delta t) - C \cos(\Delta m_d \Delta t)] \left. \right\}, \end{aligned} \quad (5.8)$$

where q_{tag} is the B_{tag} flavor assignment provided by the Flavor Tagger, and w , Δw and a_{tag} are the wrong tag fraction, wrong tag asymmetry and detection efficiency asymmetry parameters, respectively, which were previously defined in Section 4.1. These parameters are treated as *conditional parameters* in the Δt PDF model. This means that, although their values may vary on an event-by-event basis, they are not considered free fit variables in the maximum likelihood fit, unlike ΔE , M_{bc} and Δt .

Secondly, the effect of the detector's resolution in determining the decay vertex positions of B_{sig} and B_{tag} in each event needs to be included, and Equation 5.8 becomes:

$$\mathcal{P}_{\text{CP}}^{\text{FT}, \mathcal{R}}(\Delta t; q_{\text{tag}}, w, \Delta w, a_{\text{tag}}, \sigma_{\Delta t}) = \mathcal{P}_{\text{CP}}^{\text{FT}}(\Delta t; q_{\text{tag}}, w, \Delta w, a_{\text{tag}}) \otimes \mathcal{R}(\delta t; \sigma_{\Delta t}), \quad (5.9)$$

where $\mathcal{R}(\delta t; \sigma_{\Delta t})$ is the Δt Resolution Function defined in Equations 4.4 and 4.5. Once again, the parameters of the Δt Resolution Function, as well as $\sigma_{\Delta t}$ (the uncertainty in measuring Δt), are treated as conditional parameters in the maximum likelihood fit.

In addition, to avoid the so-called Punzi bias [115], the Δt PDF models for each of the components (signal, SCF, $B\bar{B}$ background and continuum background) are

multiplied by histogram PDFs of their corresponding $\sigma_{\Delta t}$ distributions. The Punzi bias arises when a conditional parameter like $\sigma_{\Delta t}$ is present in a model, but its description is not included in the likelihood. In this case, a histogram PDF might not be the most accurate description, but it is sufficient considering the relatively low statistics available.

The Flavor Tagger and Δt Resolution Function parameter values used here are the ones obtained from the calibration procedure described in Chapter 4. This means that, in the case of simulated data, the parameter values listed in Tables 4.5 and 4.6 are used, while in the case of real data, the values listed in Tables 4.5 and 4.7 are used.

Before using the calibrated Δt Resolution Function, it was first checked whether it can adequately describe the Δt residual distribution ($\delta t \equiv \Delta t^{true} - \Delta t$) of $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays. Figure 5.9 shows the Δt residual distribution of simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ signal events with the calibrated Δt Resolution Function overlaid. A discrepancy is visible, which is expected since the Δt Resolution Function is calibrated using a different decay channel, $B^0 \rightarrow D^{(*)-} \pi^+$, rather than the signal mode of the present measurement. This difference likely arises from the larger amount of tracks available in the B_{sig} vertex fit of the $B^0 \rightarrow D^{(*)-} \pi^+$ channel, as well as from the mass constraint applied to the D meson, both of which lead to a slightly better Δt resolution compared to the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ case. The figure also compares the Δt residual distributions of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ and $B^0 \rightarrow D^{(*)-} \pi^+$ simulated samples, further illustrating this difference. Nevertheless, the effect that such a discrepancy can have on the final measurement is minimal and a dedicated systematic uncertainty is assigned to account for it (see Section 5.5.3).

Using the calibrated Δt Resolution Function, the PDF model of Equation 5.9 is fitted to the Δt distribution of simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ signal events to extract the CP parameters S and C . Figure 5.10 shows this fit across all flavor tagging r -bins, alongside the projection of the same fit in the last r -bin. The signal MC sample used here has been generated with input values $(S, C) = (0.6, 0.0)$ and the fitted parameters are found to be in good agreement with these inputs. Note that the uncertainties on the fitted parameters are ~ 30 times smaller than those expected in the final measurement using 365 fb^{-1} of Belle II data. This is due to the significantly larger statistics present in the signal MC sample compared to the number of signal events expected in the real data.

As an additional check, the same fit is performed to the same signal Δt distribution, with S and C fixed to their input values and the B^0 lifetime left free to float instead. Figure 5.11 shows this fit across all flavor tagging r -bins and in the last r -bin. The value obtained for the B^0 lifetime is in good agreement with the generated value, 1.519 ps. No parameters are fixed in the 3D fit from these fits to the Δt distribution of simulated signal events.

The ΔE and M_{bc} distributions of signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ events are modeled using a double-sided Crystal Ball (DSCB) and a Crystal Ball (CB) function, respectively, whose mathematical definitions are provided in Appendix B. The corresponding fits are shown in Figure 5.12 shows the fits to the ΔE and M_{bc} distributions of signal events found in one of the signal MC samples, and the fitted parameter values are listed in Tables 5.4 and 5.5. The third column in each table indicates whether a parameter is fixed to the given value in the 3D fit. In particular, the Gaussian mean and width of the CB functions describing the signal ΔE and M_{bc} distributions are left free in the 3D fit, in order to account for potential discrepancies between data and simulation.

The modeling of the ΔE , M_{bc} and Δt distributions of SCF, $B\bar{B}$ and continuum background is described in the following section.

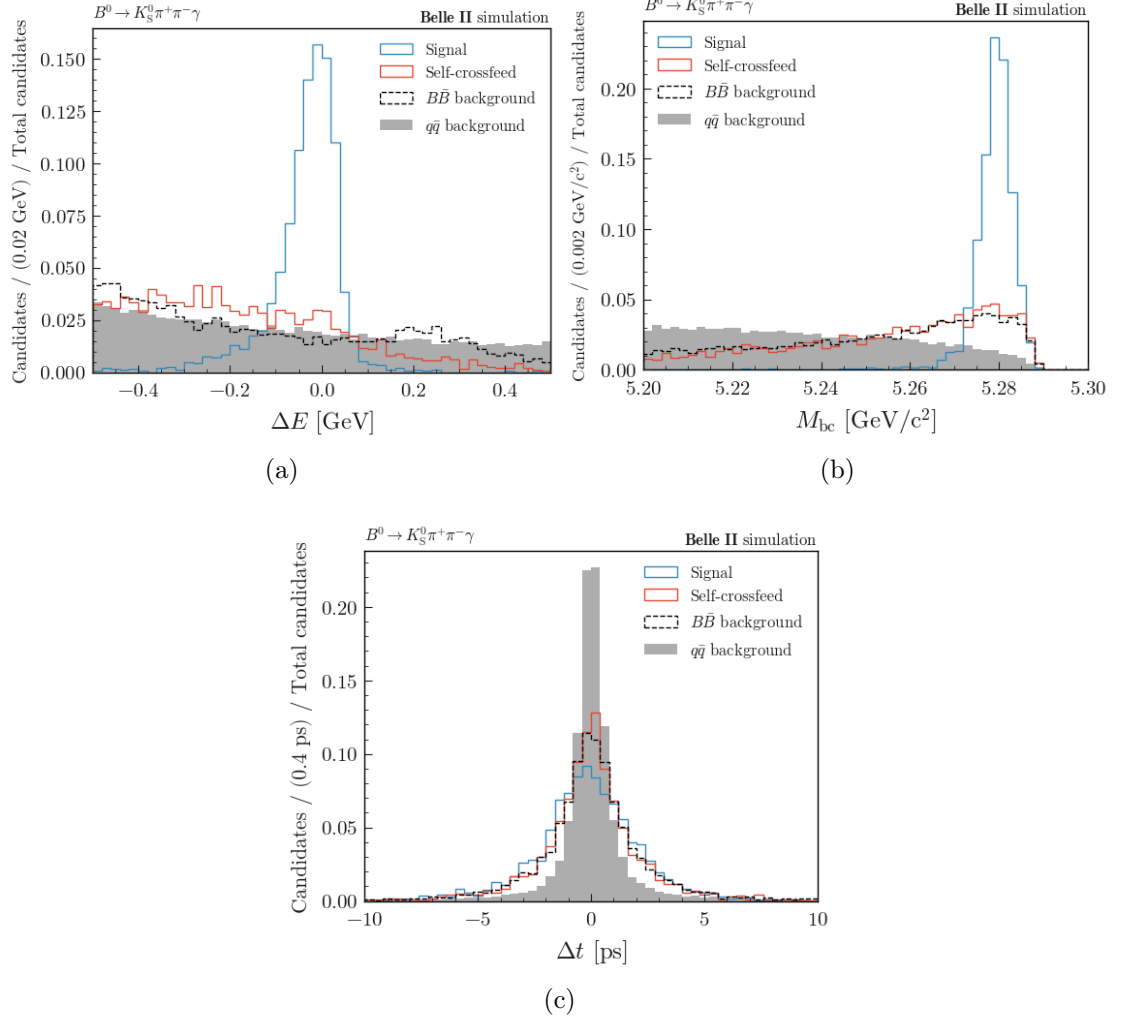


Figure 5.7: Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions between the signal and different background components for simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ events.

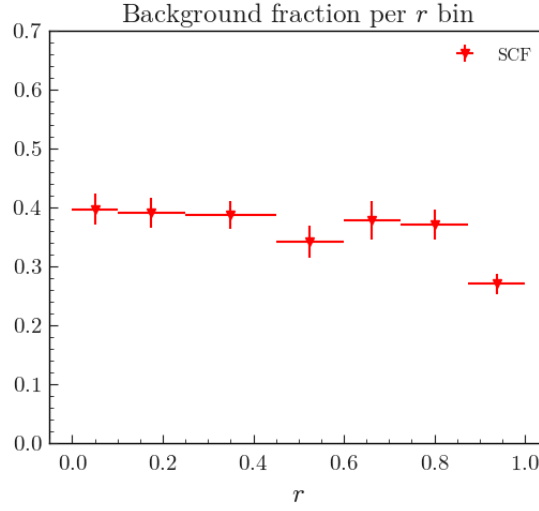


Figure 5.8: Fraction of SCF events over the all signal-like events in each flavor tagging r -bin estimated using the generic MC sample.

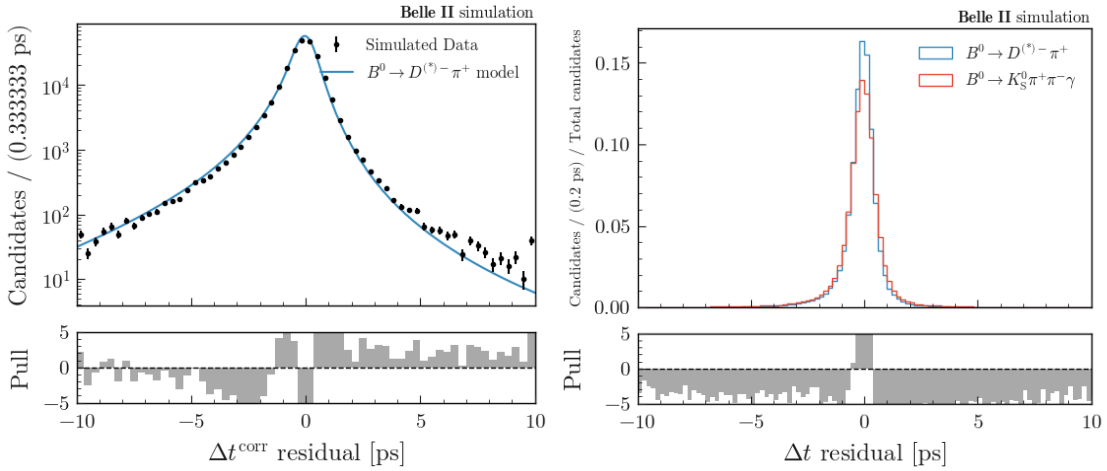


Figure 5.9: Left: Δt residual distribution of simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ signal events overlaid to the Δt Resolution Function that is calibrated using $B^0 \rightarrow D^{(*)-} \pi^+$ decays. Right: Normalized Δt residual distribution of simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ signal events overlaid to the one of simulated $B^0 \rightarrow D^{(*)-} \pi^+$ signal events.

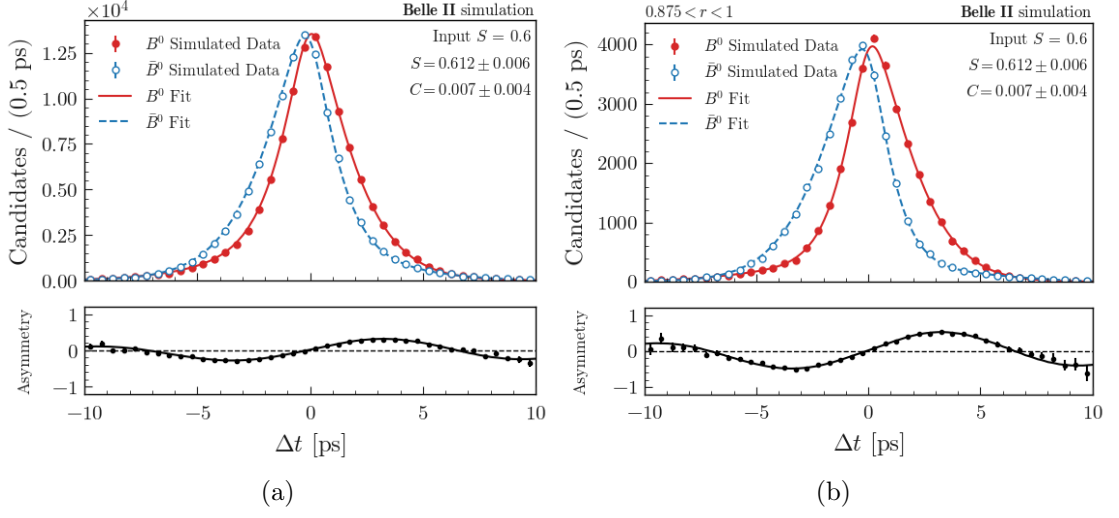


Figure 5.10: Projections of the fit to the Δt distribution of simulated signal events to extract the CP parameters S and C together with the corresponding CP asymmetry (a) in all flavor tagging r -bins and (b) in the last r -bin.

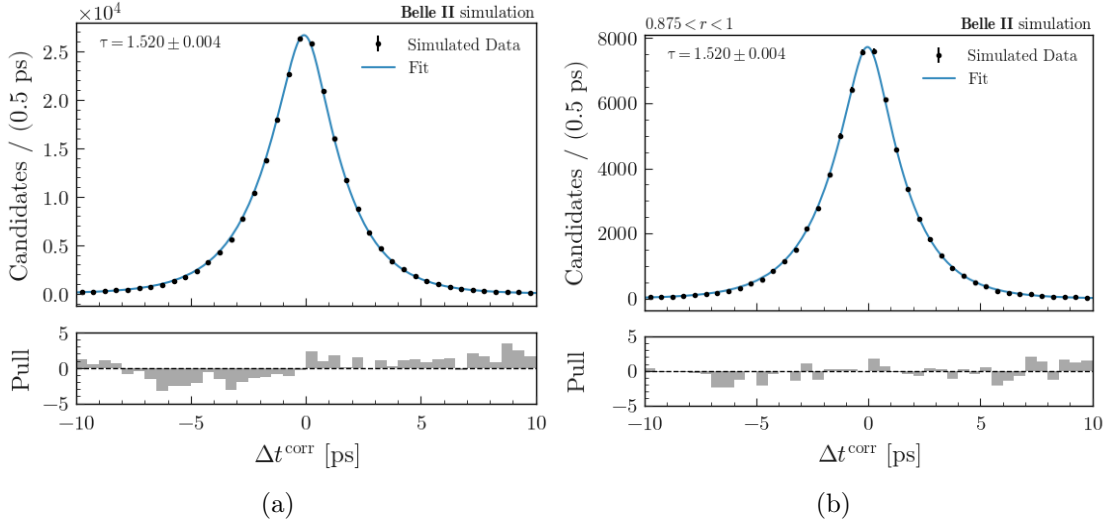


Figure 5.11: Projections of the fit to the Δt distribution of simulated signal events to extract the B^0 lifetime (a) in all flavor tagging r -bins and (b) in the last r -bin.

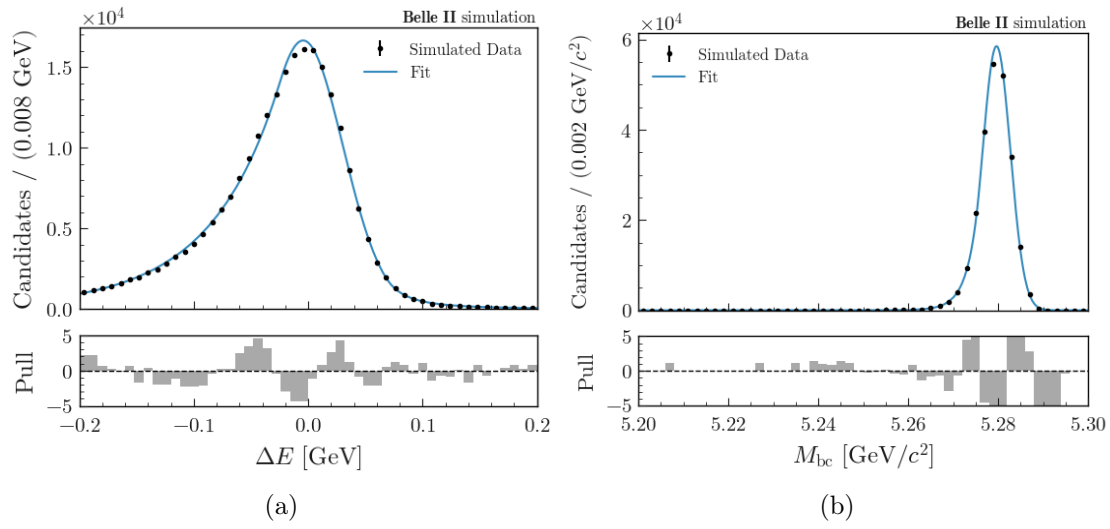


Figure 5.12: Fits of the (a) ΔE and (b) M_{bc} distributions of simulated signal events.

Table 5.4: Parameters of the signal ΔE model determined from a fit to simulated signal events.

Parameter	Fitted value	Fixed
α_L	0.578 ± 0.010	Yes
α_R	2.066 ± 0.030	Yes
n_L	19.7 ± 4.6	Yes
n_R	2.6 ± 0.2	Yes
σ [GeV]	0.0349 ± 0.0002	No
μ [GeV]	-0.0045 ± 0.0002	No

Table 5.5: Parameters of the signal M_{bc} model determined from a fit to simulated signal events.

Parameter	Fitted value	Fixed
α	1.24 ± 0.01	Yes
n	41.7 ± 7.7	Yes
σ [GeV/ c^2]	0.00308 ± 0.00001	No
μ [GeV/ c^2]	5.27970 ± 0.00001	No

5.3.3 Background modeling

SCF background

First, the SCF background modeling in ΔE , M_{bc} and Δt is described. The Δt distribution of SCF background events is modeled using a similar model to the one used for signal events, i.e. the one given in Equation 5.9. The only difference is that some of the Δt Resolution Function parameters and the B^0 lifetime parameter are allowed to take different values to those they have in the signal model. The values of these parameters are determined through a dedicated fit to the Δt distribution of SCF background events found in one of the signal MC samples. Figure 5.13 shows this fit across all flavor tagging r -bins and in the last r -bin, while the values of the fitted parameters are listed in Table 5.6.

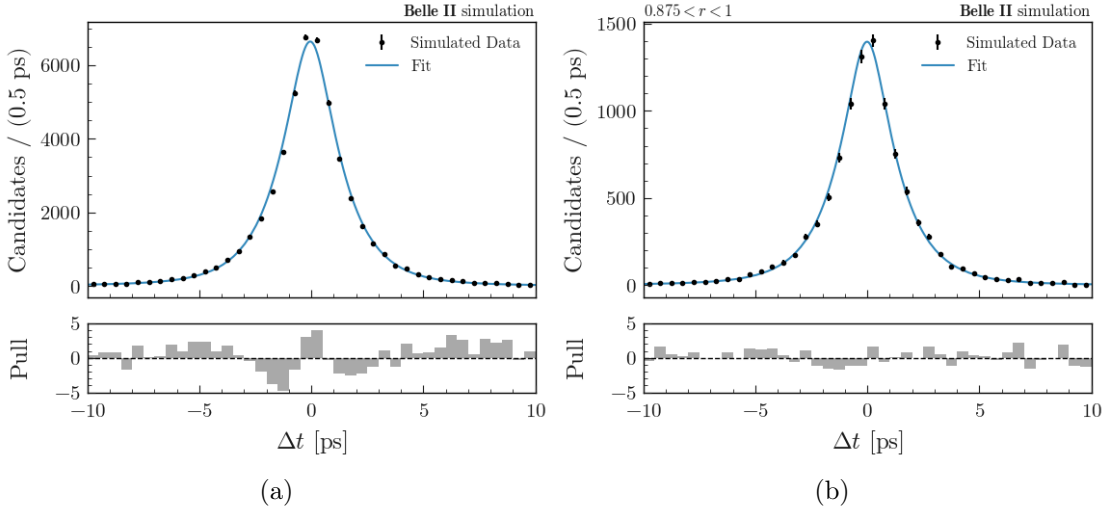


Figure 5.13: Projections of the fit to the Δt distribution of simulated SCF background events (a) in all flavor tagging r -bins and (b) in the last r -bin.

Table 5.6: Parameters of the SCF background Δt model determined from a fit to simulated SCF background events.

Parameter	Fitted value	Fixed
s_{main} [ps]	1.314 ± 0.035	Yes
s_{tail} [ps]	12.736 ± 0.369	Yes
f_{max}	0.1002 ± 0.0047	Yes
$\tau_{B^0}^{\text{eff}}$ [ps]	1.212 ± 0.014	Yes

The fitted lifetime parameter being smaller than expected should not be too surprising, as SCF background consists of mis-reconstructed signal events. In such cases, charged pion tracks from the B_{sig} and B_{tag} decays are often mistakenly exchanged in the vertex fits, causing the two vertices to be reconstructed closer together than they should be and thus leading to an underestimated lifetime.

Furthermore, the fit to extract S and C is performed on all available signal MC samples (see Appendix D.4 for an example of such a fit), and a clear correlation is observed between the fitted S value from SCF background events, S_{SCF} , and the input S used in generating the samples. This is illustrated in Figure 5.14, which shows the fitted values of S_{SCF} versus the input S for the seven available signal MC samples. To account for this effect in the 3D fit, S_{SCF} and C_{SCF} are tied to the corresponding signal S and C values through a proportionality factor κ_{SCF} , defined as $S_{\text{SCF}} \equiv \kappa_{\text{SCF}} S$ and $C_{\text{SCF}} \equiv \kappa_{\text{SCF}} C$. The κ_{SCF} parameter is floated in the 3D fit with a Gaussian constraint⁴, where the mean is set to 0.8, a value estimated from Figure 5.14, and the width is set to 0.2, which has been chosen conservatively. Although this correlation was not explicitly tested for non-zero C values, the assumption that the same scaling applies to both S and C is considered reasonable, since C is expected to remain close to zero.

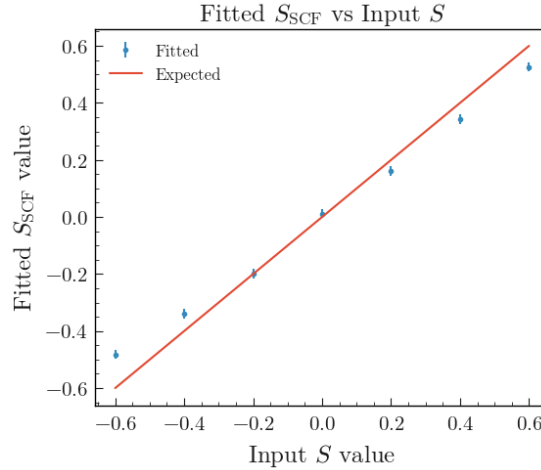


Figure 5.14: Fitted values of S_{SCF} versus the values of S used to generate the signal MC samples. The red solid line corresponds to $S_{\text{SCF}} = S$.

The ΔE distribution of SCF background events is modeled using the sum of 2nd order Chebyshev polynomial and a bifurcated Gaussian function, while the

⁴In the context of a maximum likelihood fit, a parameter is Gaussian constrained by adding a term in the likelihood that penalises deviations from a specified central value with a given uncertainty.

M_{bc} distribution is modeled by the sum of an ARGUS function with a bifurcated Gaussian function. The definitions of these functions can be found in Appendix B. Figure 5.15 presents the fits to the ΔE and M_{bc} distributions of SCF background events found in one of the signal MC samples, and the fitted parameter values are listed in Tables 5.7 and 5.8.

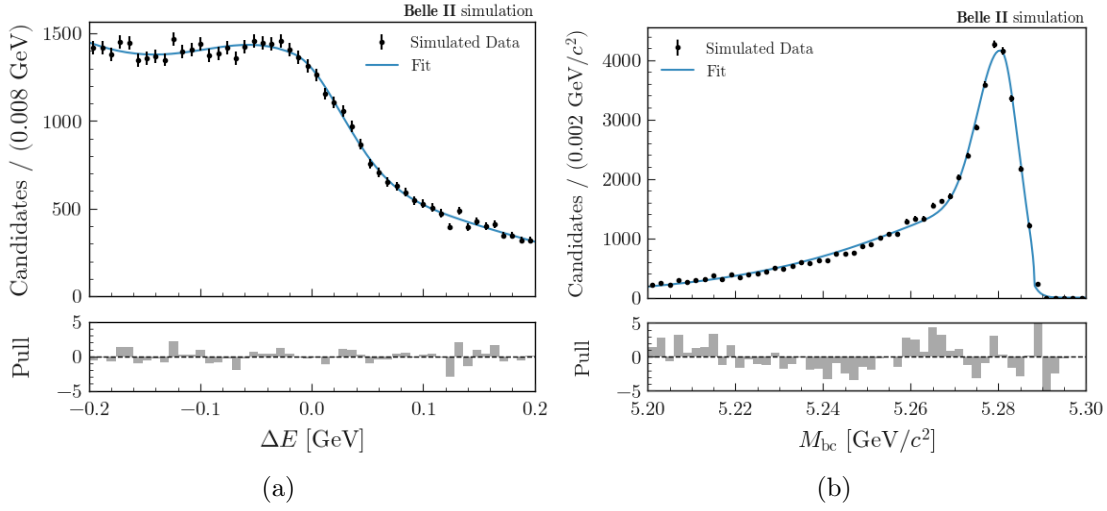


Figure 5.15: Fits of the (a) ΔE and (b) M_{bc} distributions of simulated SCF background events.

Table 5.7: Parameters of the SCF background ΔE model determined from a fit to simulated SCF background events.

Parameter	Fitted value	Fixed
c_1 [GeV] $^{-1}$	-0.6744 ± 0.0161	Yes
c_2 [GeV] $^{-2}$	0.0623 ± 0.0167	Yes
f_{gauss}	0.2252 ± 0.0172	Yes
μ [GeV]	-0.0156 ± 0.0037	Yes
σ_L [GeV]	0.0845 ± 0.0080	Yes
σ_R [GeV]	0.0424 ± 0.0035	Yes

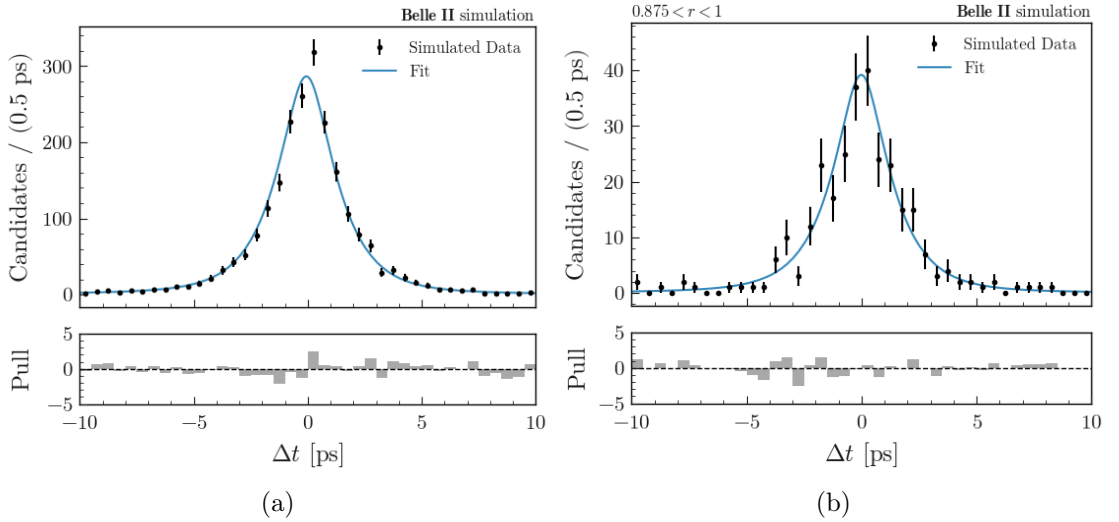
$B\bar{B}$ background

Next, the description of how the $B\bar{B}$ background is modeled in ΔE , M_{bc} and Δt is provided. The Δt distribution of $B\bar{B}$ background events is treated in a similar manner to SCF background events, meaning that the model given in Equation 5.9

Table 5.8: Parameters of the SCF background M_{bc} model determined from a fit to simulated SCF background events.

Parameter	Fitted value	Fixed
c	-108 ± 1.1	Yes
m_0 [GeV/ c^2]	5.28820 ± 0.00001	
f_{gauss}	0.2862 ± 0.0058	Yes
μ [GeV/ c^2]	5.2807 ± 0.0001	Yes
σ_L [GeV/ c^2]	0.0053 ± 0.0002	Yes
σ_R [GeV/ c^2]	0.0033 ± 0.0001	Yes

is used with the lifetime and some of the Δt Resolution Function parameters being allowed to take different values than those they have in the signal model. As with SCF background, the values of these parameters are determined through a dedicated fit to the Δt distribution of $B\bar{B}$ background events found in the generic MC sample. The fit is presented in Figure 5.16 across all flavor tagging r -bins and in the last r -bin, and the values of the fitted parameters are listed in Table 5.9. As an additional check, the fit to extract S and C is also performed on the Δt distribution of $B\bar{B}$ background events. It can be found in Appendix D.4.

Figure 5.16: Projections of the fit to the Δt distribution of simulated $B\bar{B}$ background events (a) in all flavor tagging r -bins and (b) in the last r -bin.

As previously mentioned, the ΔE and M_{bc} distributions of $B\bar{B}$ background events are modeled using a parameterless, two-dimensional KDE, in order to ac-

Table 5.9: Parameters of the $B\bar{B}$ background Δt model determined from a fit to simulated $B\bar{B}$ background events.

Parameter	Fitted value	Fixed
s_{main} [ps]	1.337 ± 0.124	Yes
s_{tail} [ps]	12.042 ± 1.398	Yes
f_{max}	0.1234 ± 0.0201	Yes
$\tau_{B^0}^{\text{eff}}$ [ps]	1.153 ± 0.057	Yes

count for the correlation between the ΔE and M_{bc} distributions of $B\bar{B}$ background. The KDE shape is determined through a fit to the ΔE and M_{bc} distributions of $B\bar{B}$ background events found in the generic MC sample and Figure 5.17 shows the projection of this fit in ΔE and M_{bc} .

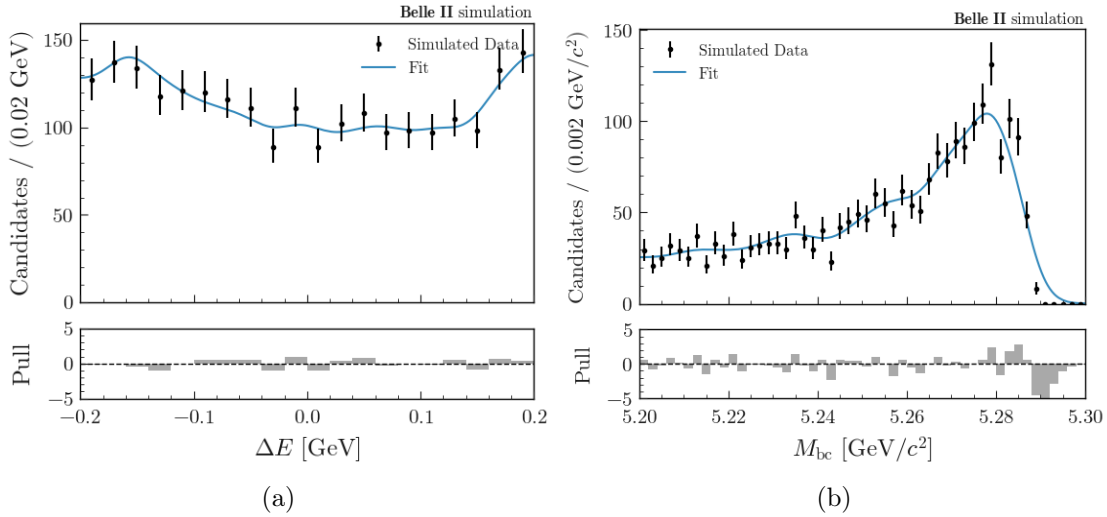


Figure 5.17: Projection of the fit to the (a) ΔE and (b) M_{bc} distributions of simulated $B\bar{B}$ background events.

Continuum background

Lastly, the modeling of the continuum background in ΔE , M_{bc} and Δt is described. The Δt distribution of continuum background events is modeled using a weighted sum of three Gaussian distributions the parameters of which are determined from a fit to the Δt distribution of simulated continuum background events found in the generic MC sample. The first (core) and second (tail) Gaussians have means and widths which are scaled by the per-event $\sigma_{\Delta t}$, similarly to what is done with the Δt Resolution Function, which notably is not used here. The third (outlier)

Gaussian is set to have a fraction of 1%, a mean of 0 ps and a width of 200 ps. Mathematically, the continuum background Δt PDF takes the form:

$$\begin{aligned} \mathcal{P}^{q\bar{q} \text{ bkg}}(\Delta t; \sigma_{\Delta t}) = & (1 - f_{\text{tail}} - f_{\text{OL}})G(\Delta t; m_{\text{core}}\sigma_{\Delta t}, s_{\text{core}}\sigma_{\Delta t}) \\ & + f_{\text{tail}}G(\Delta t; m_{\text{tail}}\sigma_{\Delta t}, s_{\text{tail}}\sigma_{\Delta t}) \\ & + f_{\text{OL}}G(\Delta t; 0, 200) , \end{aligned} \quad (5.10)$$

The fit is presented in Figure 5.18 and the values of the fitted parameters are listed in Table 5.10. Two of the parameters in the continuum background Δt model, the widths of the core and tail Gaussians, are allowed to be free in the 3D fit. This is done to once again account for possible differences between data and simulation.

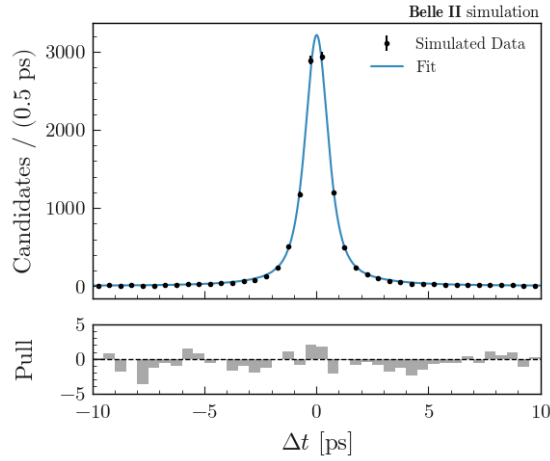


Figure 5.18: Fit to the Δt distribution of simulated continuum background events.

Table 5.10: Parameters of the continuum background Δt model determined from a fit to simulated continuum background events.

Parameter	Fitted value	Fixed
m_{core} [ps]	0.01072 ± 0.00614	Yes
m_{tail} [ps]	0.15103 ± 0.07851	Yes
s_{core} [ps]	1.3658 ± 0.0178	No
s_{tail} [ps]	6.5140 ± 0.2894	No
f_{tail}	0.1466 ± 0.0072	Yes
f_{OL}	0.01	Yes

The ΔE distribution of continuum background events is modeled using an exponential function, while the M_{bc} distribution is modeled by an ARGUS function.

Figure 5.19 presents the fits to the ΔE and M_{bc} distributions of continuum background events found in the generic MC sample, and the fitted parameter values are listed in Tables 5.11 and 5.12. In this case, only the ARGUS threshold is allowed to be free in the 3D fit.

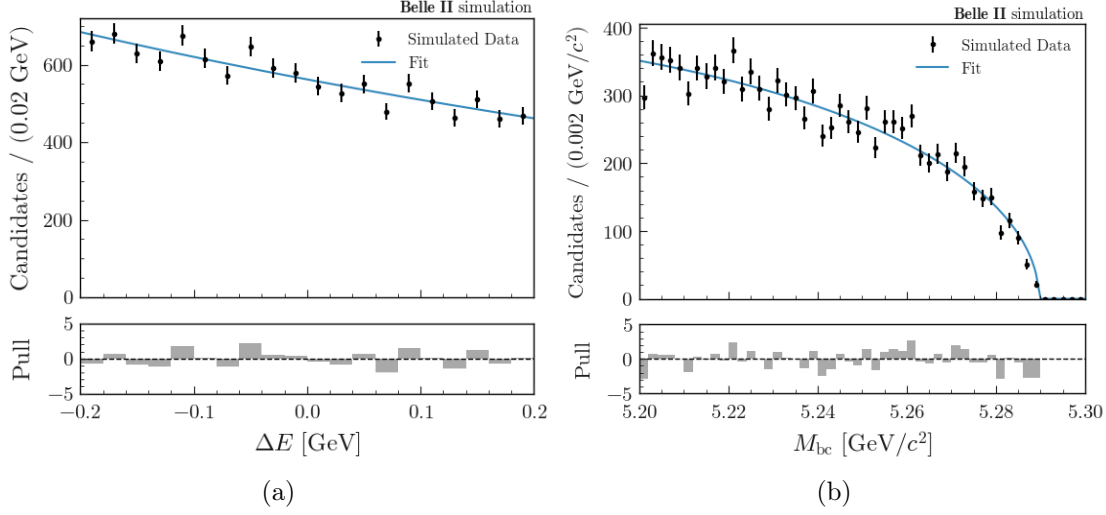


Figure 5.19: Fits of the (a) ΔE and (b) M_{bc} distributions of simulated continuum background events.

Table 5.11: Parameters of the continuum background ΔE model determined from a fit to simulated continuum background events.

Parameter	Fitted value	Fixed
$\lambda \text{ [GeV]}^{-1}$	-0.98 ± 0.08	Yes

Table 5.12: Parameters of the continuum background M_{bc} model determined from a fit to simulated continuum background events.

Parameter	Fitted value	Fixed
c	-4.8 ± 1.1	Yes
$m_0 \text{ [GeV}/c^2]$	5.28970 ± 0.00003	No

5.3.4 3D fit to a simulated sample

A total of 10 parameters and 3 yields are allowed to be free in the fit. The parameters are:

- The CP parameters of interest, S and C .
- The proportionality constant κ_{SCF} , which relates S_{SCF} and C_{SCF} to S and C , though it is Gaussian constrained.
- Two mass-like fudge factors ($f_{\Delta E}^\mu, f_{M_{\text{bc}}}^\mu$) that are directly added to the means of the Gaussian-like cores of the signal ΔE and M_{bc} distributions.
- Two dimensionless fudge factors ($f_{\Delta E}^\sigma, f_{M_{\text{bc}}}^\sigma$) that directly multiply the widths of the Gaussian-like cores of the signal ΔE and M_{bc} components.
- The ARGUS threshold parameter, m_0 , that is considered common for all M_{bc} background components described by an ARGUS function.
- Two dimensionless fudge factors ($f_{\Delta t, \text{core}}^\sigma, f_{\Delta t, \text{tail}}^\sigma$) that scale the widths of the core and tail Gaussians which are used to model the continuum background Δt distribution.

Three yields are allowed to be free in the fit, namely the yield of the continuum background component, the $B\bar{B}$ background yield, and the combined yield of the signal and SCF background components, albeit with the fraction between them fixed from simulation.

The 3D fit is first validated using simulated data. For this purpose, a data-like “MC cocktail” sample is constructed by bootstrapping [105], i.e. sampling with replacement, the signal and generic MC samples. Signal and SCF events are sampled from the signal MC samples, while $B\bar{B}$ background and continuum background events are sampled from the generic MC sample. This MC cocktail sample reproduces the expected signal and background composition found in 25% of the generic MC so as to match the integrated luminosity of the Belle II dataset (365 fb^{-1}). The expected yields of signal, SCF, $B\bar{B}$ background, and continuum background events for each MC cocktail sample are listed in Table 5.13, while Table 5.14 lists the corresponding numbers in a signal enhanced region, defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$, where the ratio of signal over background is approximately equal to 3. Although the fit itself is not restricted to this region, it is useful for illustration purposes.

The ΔE , M_{bc} and Δt distributions found in an MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ are fitted using the PDF model of Equation 5.7. The result of the fit is projected in each of the three fit dimensions and shown in

Figures 5.20 and 5.21. The values of the fitted parameters are shown in Table 5.15. In Figure 5.21, to improve the visibility of the signal component, both the simulated data and fitted models are projected onto a signal enhanced region in ΔE , M_{bc} and r with $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

Table 5.13: Expected number of events for each of the four components as they are found within a data-like MC cocktail sample. These correspond to 25% of the events found in the generic MC.

Component	N_i
Signal	425 ± 20
SCF	230 ± 15
$B\bar{B}$ background	540 ± 25
Continuum	2720 ± 50

Table 5.14: Expected number of events for each of the four components as they are found within a data-like MC cocktail sample, specifically in the signal enhanced region used to display the Δt distributions for tagged B^0 and $\bar{B}^0$ events, shown in Figure 5.21.

Component	N_i^{SER}
Signal	100 ± 10
SCF	10 ± 5
$B\bar{B}$ background	10 ± 5
Continuum	10 ± 5

The statistical uncertainties of S and C obtained from this fit are $\sigma_S = 0.14$ and $\sigma_C = 0.09$, respectively. Moreover, it seems the fitted signal yield is slightly overestimated with respect to the true signal yield in Table 5.13. These findings are later validated through dedicated studies in Section 5.4.

As an additional check, instead of fitting for S and C , the B^0 lifetime is floated in the 3D fit, while S and C are both set to zero. The fit is performed on the same MC cocktail sample as before and the projection of the fit in Δt is shown in Figure 5.22, together with the fitted lifetime value. The fitted lifetime is found to be consistent with the generated value, 1.519 ps, within 1σ .

Table 5.15: Resulting parameters of the Dalitz-integrated fit to a data-like MC cocktail sample, generated with $(S, C) = (0.0, 0.0)$. The fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).

Parameter	Fitted value
C	0.09 ± 0.09
S	-0.08 ± 0.14
κ_{SCF}	0.81 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	0.0033 ± 0.0030
$f_{M_{bc}}^\mu$ [GeV/ c^2]	-0.0001 ± 0.0002
$f_{\Delta E}^\sigma$	0.99 ± 0.08
$f_{M_{bc}}^\sigma$	0.96 ± 0.05
m_0 [GeV/ c^2]	5.2882 ± 0.0002
$f_{\Delta t, \text{core}}^\sigma$	1.01 ± 0.02
$f_{\Delta t, \text{tail}}^\sigma$	0.94 ± 0.08
$N_{\text{sig-like}}$	745 ± 40
$N_{B\bar{B} \text{ bkg}}$	555 ± 65
$N_{q\bar{q} \text{ bkg}}$	2655 ± 70

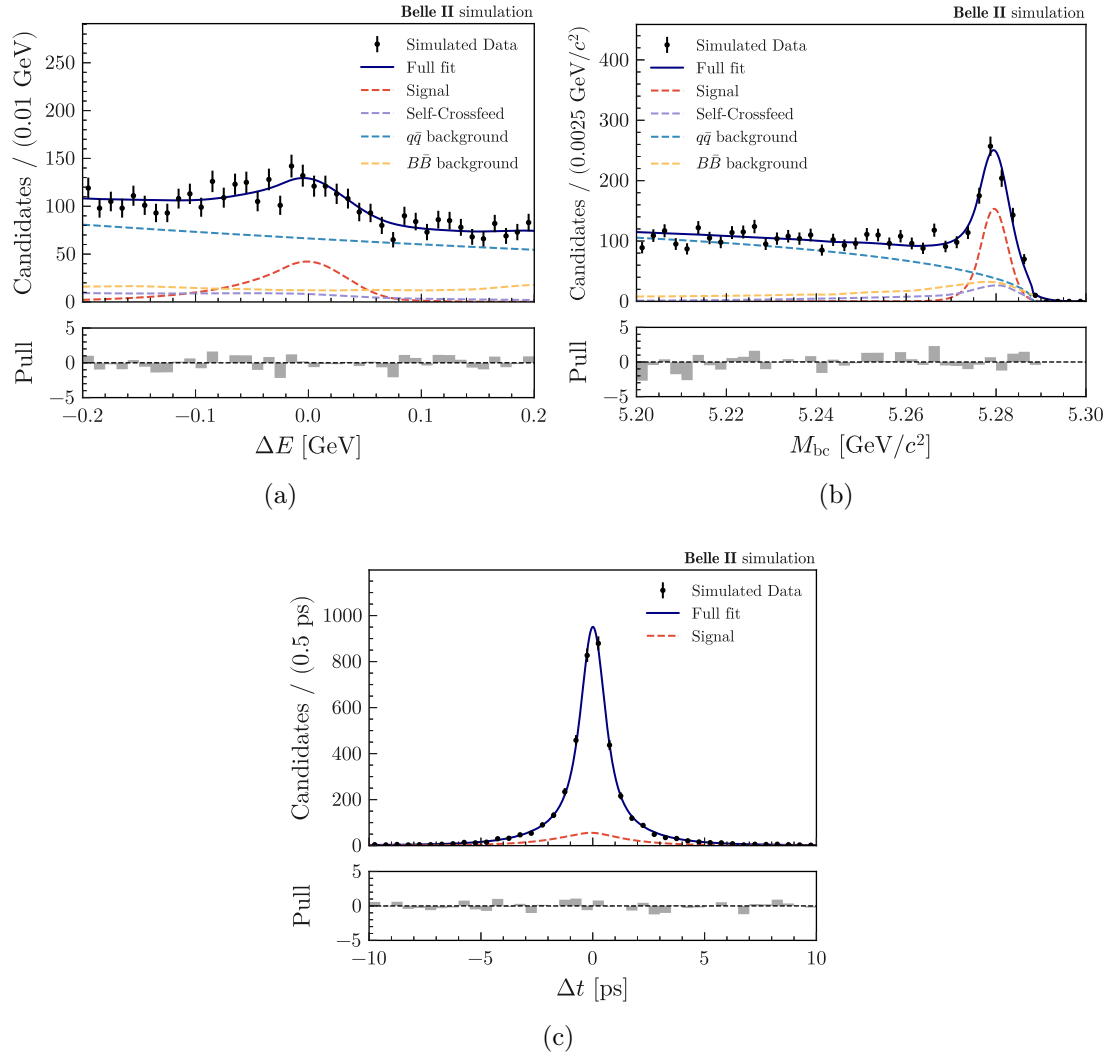


Figure 5.20: Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.

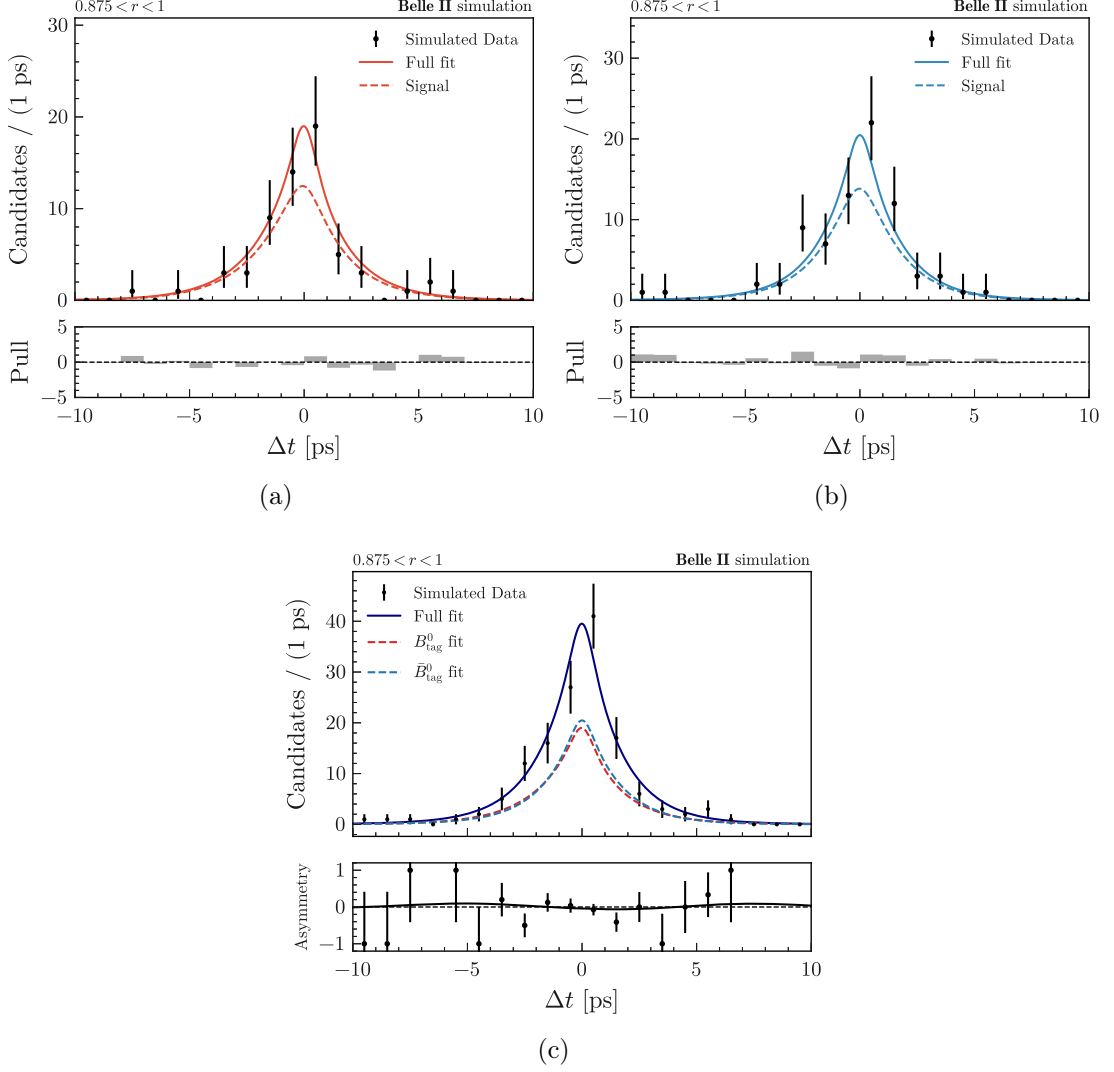


Figure 5.21: Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

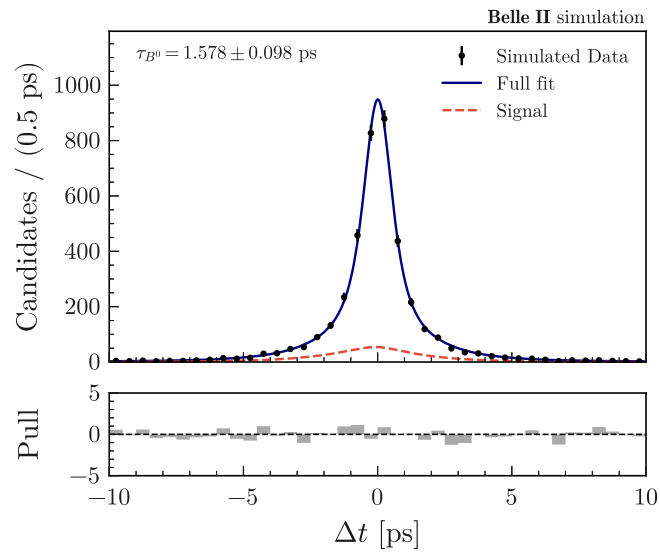


Figure 5.22: Fit of a data-like MC cocktail sample to extract the B^0 lifetime projected onto Δt .

Dalitz-split fit to a simulated sample

As previously stated, one of the goals of this analysis is to also measure the S^+ and S^- parameters. Therefore, the previously mentioned MC cocktail sample is split in the two halves of the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane, separated by the bisector $m_{K_S^0 \pi^+}^2 = m_{K_S^0 \pi^-}^2$. The distribution of simulated events across the Dalitz plane is shown in Appendix D.5. $50.4(49.6) \pm 1.2(1.2)\%$ of signal events are expected to lie in the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 < (>) m_{K_S^0 \pi^-}^2$. As a reminder, the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 < m_{K_S^0 \pi^-}^2$ is referred to as the “Dalitz I” region, while the complementary region is referred to as the “Dalitz $\bar{I}$ ” region.

The fit model of Equation 5.7 is used to fit these two subsets of simulated data simultaneously and the parameters of interest are extracted. In this fit, the yields of each Dalitz region are independent of each other, while the C parameter is also allowed to be free and is considered common between the two Dalitz regions. Moreover, all of the shape parameters as well as the SCF background fraction fixed from simulation are considered common between the two Dalitz regions. A comparison between the global ΔE , M_{bc} and Δt distributions and the distributions for the different fit components in each of the Dalitz regions can be found in Appendix D.6. As expected, the distributions are compatible across the two Dalitz regions.

The proportionality constant, κ_{SCF} , nominally defined as either $\kappa_{\text{SCF}} \equiv S_{\text{SCF}}/S$ or $\kappa_{\text{SCF}} \equiv C_{\text{SCF}}/C$, is also shared between the two Dalitz regions. In the case of the Dalitz-split fit, it is defined as either $\kappa_{\text{SCF}} \equiv S_{\text{SCF}}^I/S^I$ or $\kappa_{\text{SCF}} \equiv S_{\text{SCF}}^{\bar{I}}/S^{\bar{I}}$ or $\kappa_{\text{SCF}} \equiv \frac{C_{\text{SCF}}}{C}$.

The fit projections in each of the two Dalitz regions are shown in Figures 5.23–5.26, while the values of the fitted parameters are listed in Table 5.16.

The statistical uncertainties of S^+ and S^- are found to be on the level of $\sigma_S^+ = 0.28$ and $\sigma_S^- = 0.28$, which is about twice as large as the estimated uncertainty on S . This is expected from how S^+ and S^- are constructed. As a reminder, $S^+ = S^I + S^{\bar{I}}$ and $S^- = S^I - S^{\bar{I}}$, where S^I and $S^{\bar{I}}$ are the measured S parameters in the Dalitz I and $\bar{I}$ regions, respectively.

Table 5.16: Resulting parameters of the Dalitz-split fit to a data-like MC cocktail sample, generated with $(S, C) = (0.0, 0.0)$. Once again, the fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).

Parameter	Fitted value
C	0.10 ± 0.09
S^+	-0.16 ± 0.28
S^-	-0.02 ± 0.28
κ_{SCF}	0.81 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	0.0034 ± 0.0030
$f_{M_{bc}}^\mu$ [GeV/ c^2]	-0.0001 ± 0.0002
$f_{\Delta E}^\sigma$	0.99 ± 0.08
$f_{M_{bc}}^\sigma$	0.95 ± 0.05
m_0 [GeV/ c^2]	5.2887 ± 0.0002
$f_{\Delta t, \text{main}}^\sigma$	1.01 ± 0.02
$f_{\Delta t, \text{tail}}^\sigma$	0.95 ± 0.08
$N_{\text{sig-like}}^I$	365 ± 30
$N_{\text{sig-like}}^I$	375 ± 25
$N_{B\bar{B} \text{ bkg}}^I$	280 ± 45
$N_{B\bar{B} \text{ bkg}}^I$	250 ± 40
$N_{q\bar{q} \text{ bkg}}^I$	1395 ± 50
$N_{q\bar{q} \text{ bkg}}^I$	1285 ± 50

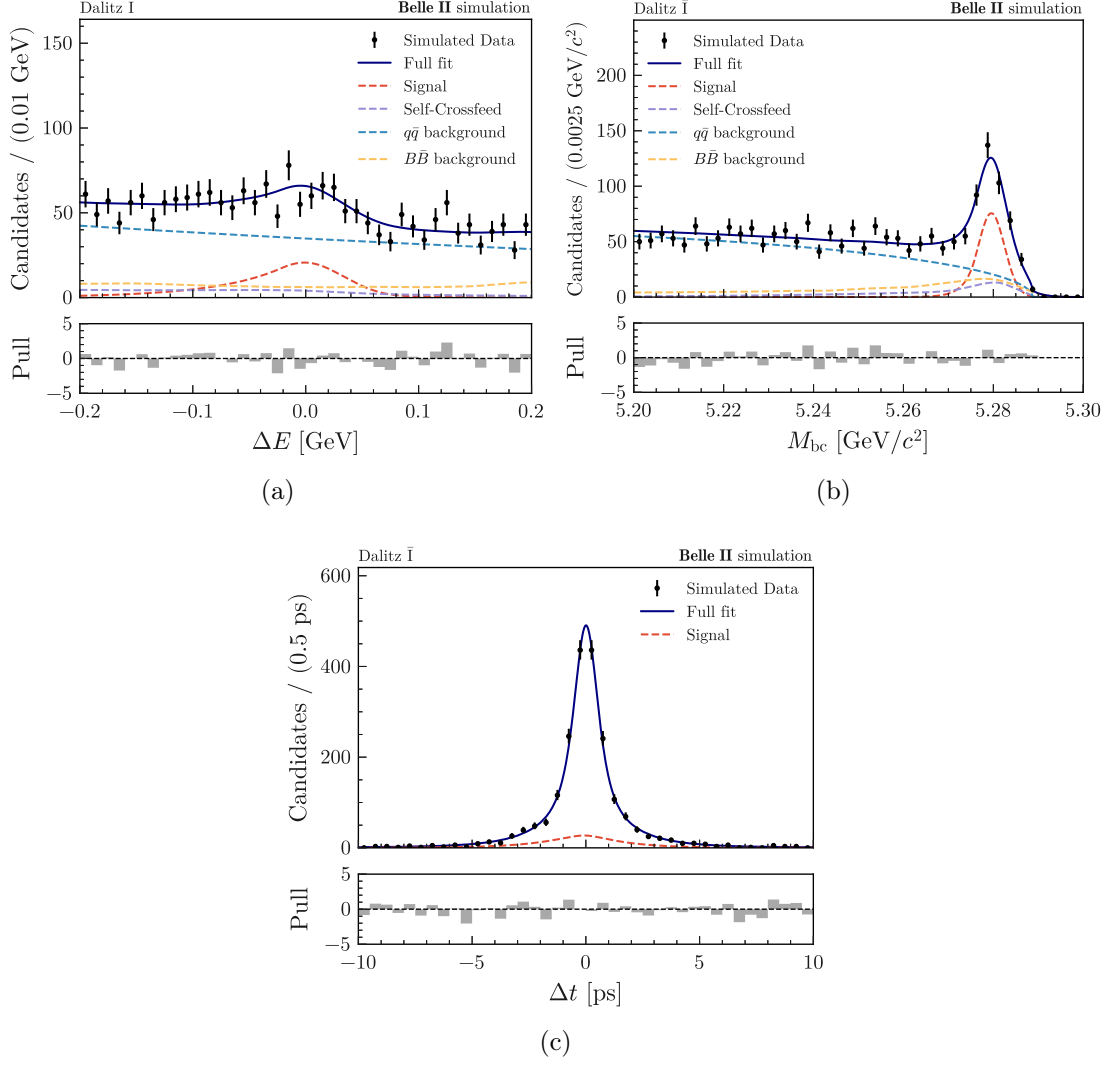


Figure 5.23: Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 > m_{K_S^0 \pi^-}^2$ (Dalitz $\bar{I}$). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.

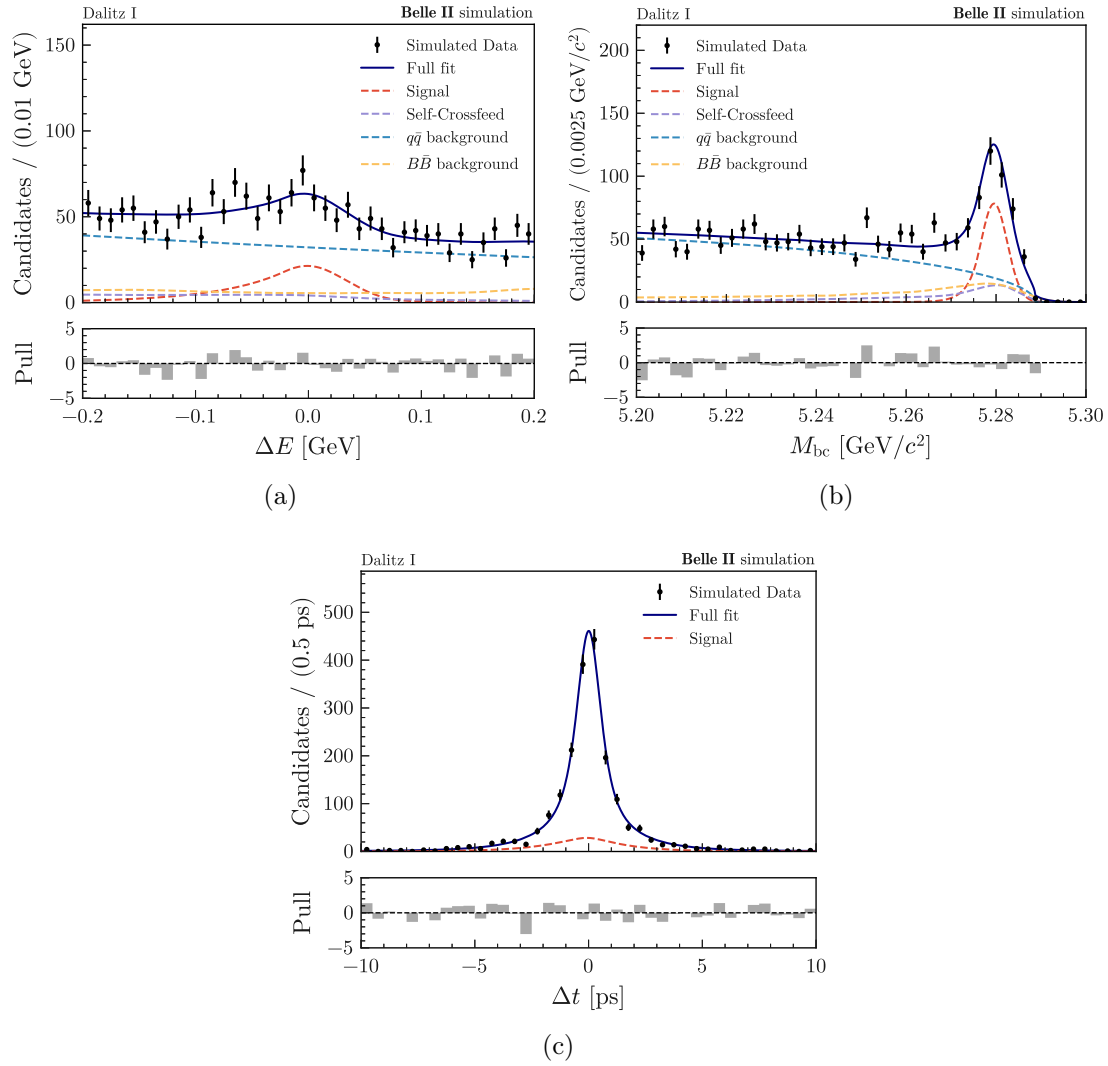


Figure 5.24: Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$ (Dalitz I). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.

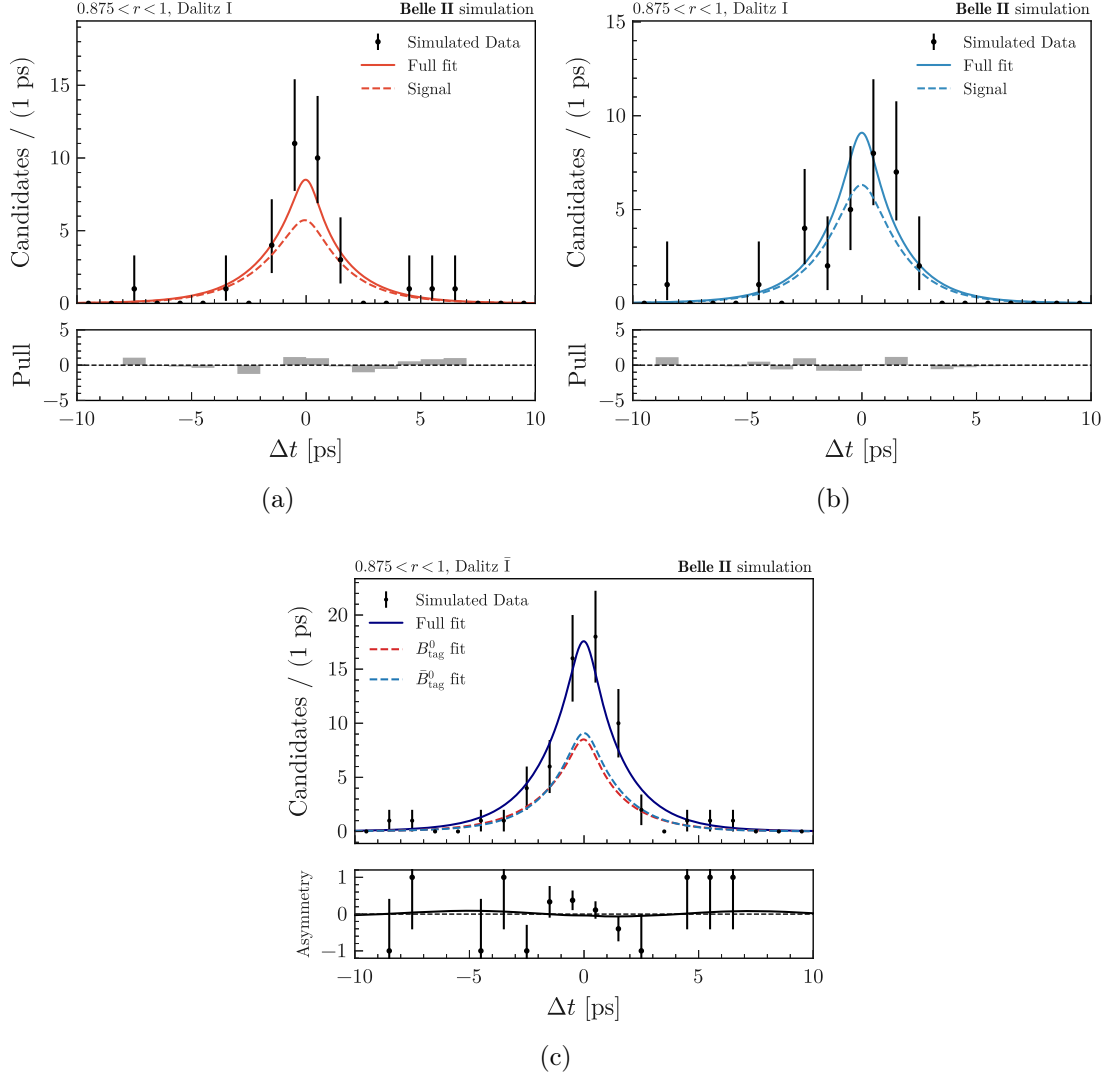


Figure 5.25: Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{I}$ region and in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

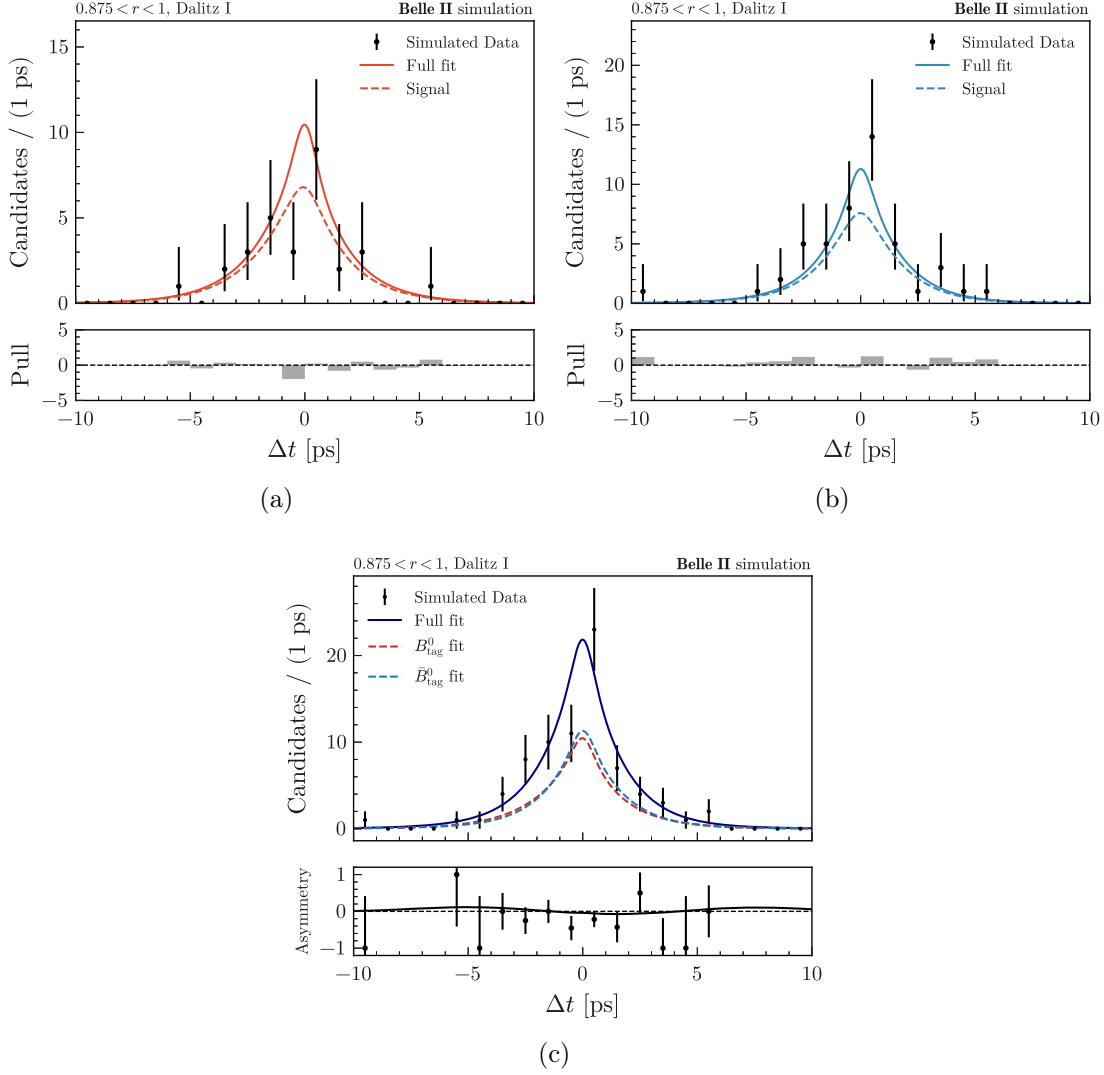


Figure 5.26: Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

5.4 Validation of the fit strategy

In this section, the tests performed to validate the overall fit procedure as well as the agreement between the data and MC are presented. The fit strategy is validated using the aforementioned bootstrapped datasets. In total, 7 sets of 1,000 MC cocktail samples with different input S values are used for the signal mode while 1 set of 1,000 MC cocktail samples is used for the control mode, $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$. The control mode is also used to check the agreement between the data and simulated samples.

5.4.1 Validation using the signal mode

Fit validation with bootstrapped datasets

The fitting procedure developed in the previous section is validated using the 7×1000 MC cocktail samples, *via* a procedure that is referred to as the “ensemble test”, in order to check for possible biases on the fitted CP parameters and to verify that the uncertainties are correctly estimated. The test is performed for seven input (S, C) combinations, namely $(-0.6, 0.0)$, $(-0.4, 0.0)$, $(-0.2, 0.0)$, $(0.0, 0.0)$, $(0.2, 0.0)$, $(0.4, 0.0)$, and $(0.6, 0.0)$, each corresponding to a set of MC cocktail samples.

The MC cocktail samples are constructed by bootstrapping [105] the signal and generic MC samples. For the signal and SCF components, events are drawn from the appropriate signal MC sample corresponding to the chosen input S value, while the $B\bar{B}$ and continuum backgrounds are sampled from the generic MC sample. All parent datasets are taken after the full selection, with only one candidate retained per event. For each bootstrap replica, the yield of each component is drawn from a Poisson distribution centered on the nominal expectation for the Belle II dataset (see Table 5.13).

The 7000 bootstrap replicas are all fitted to extract S and C and then simultaneously in the two Dalitz plane halves to extract S^+ and S^- following the fit strategy presented in Section 5.3. The distributions of the fitted CP parameters are examined for each generated (S, C) pair to check for potential biases. All of the relevant figures are shown in Appendix D.7. As an example, the distributions of the fit pulls, fit estimates and fit errors for S , C , S^+ , S^- and the yields $N_{\text{sig-like}}$, $N_{B\bar{B} \text{ bkg}}$ and $N_{q\bar{q} \text{ bkg}}$ corresponding to the input values $(S, C) = (0.6, 0.0)$ are shown in Figures 5.27, 5.28 and 5.29, where the fit pull distributions are fitted using Gaussian functions. The resulting means and widths of the pulls and of the absolute biases of the CP parameters and yields for all input values of S are listed in Tables 5.17–5.20.

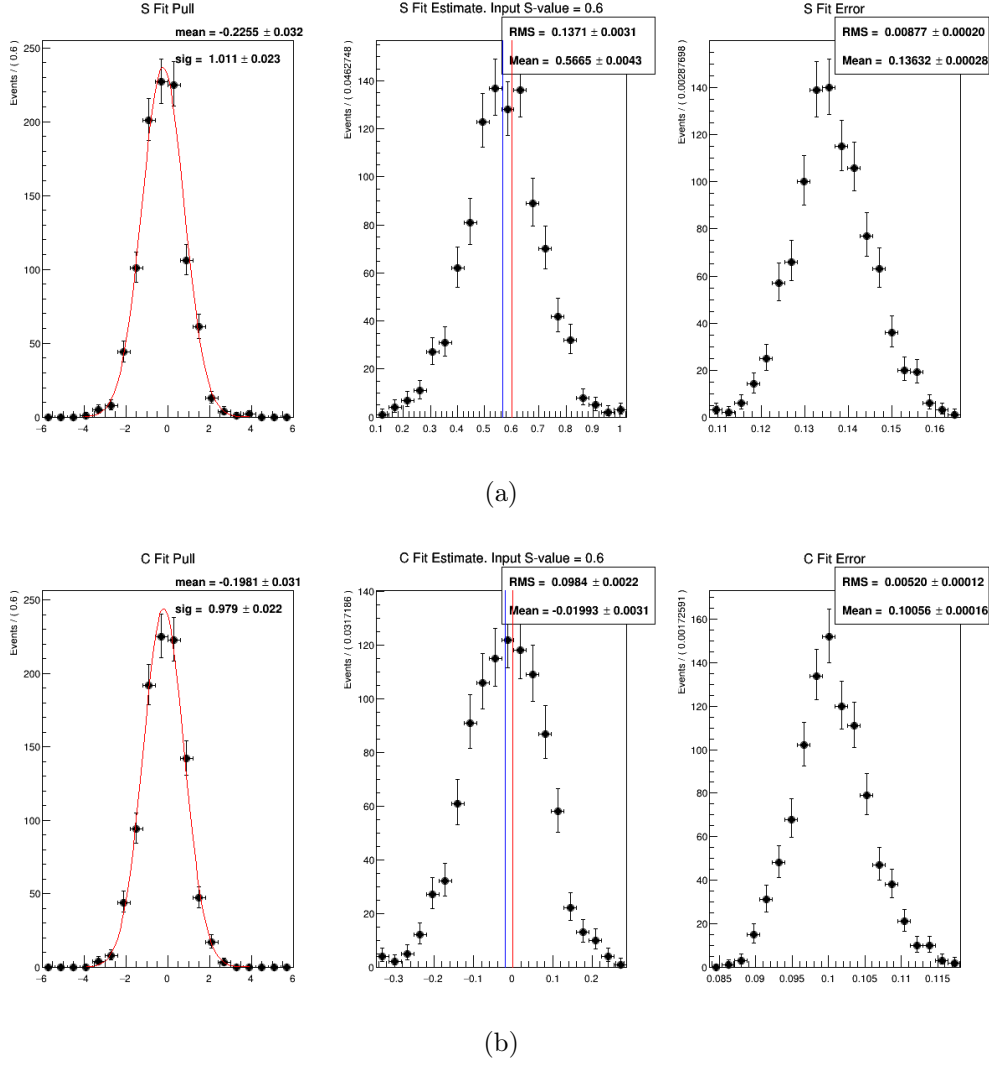


Figure 5.27: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

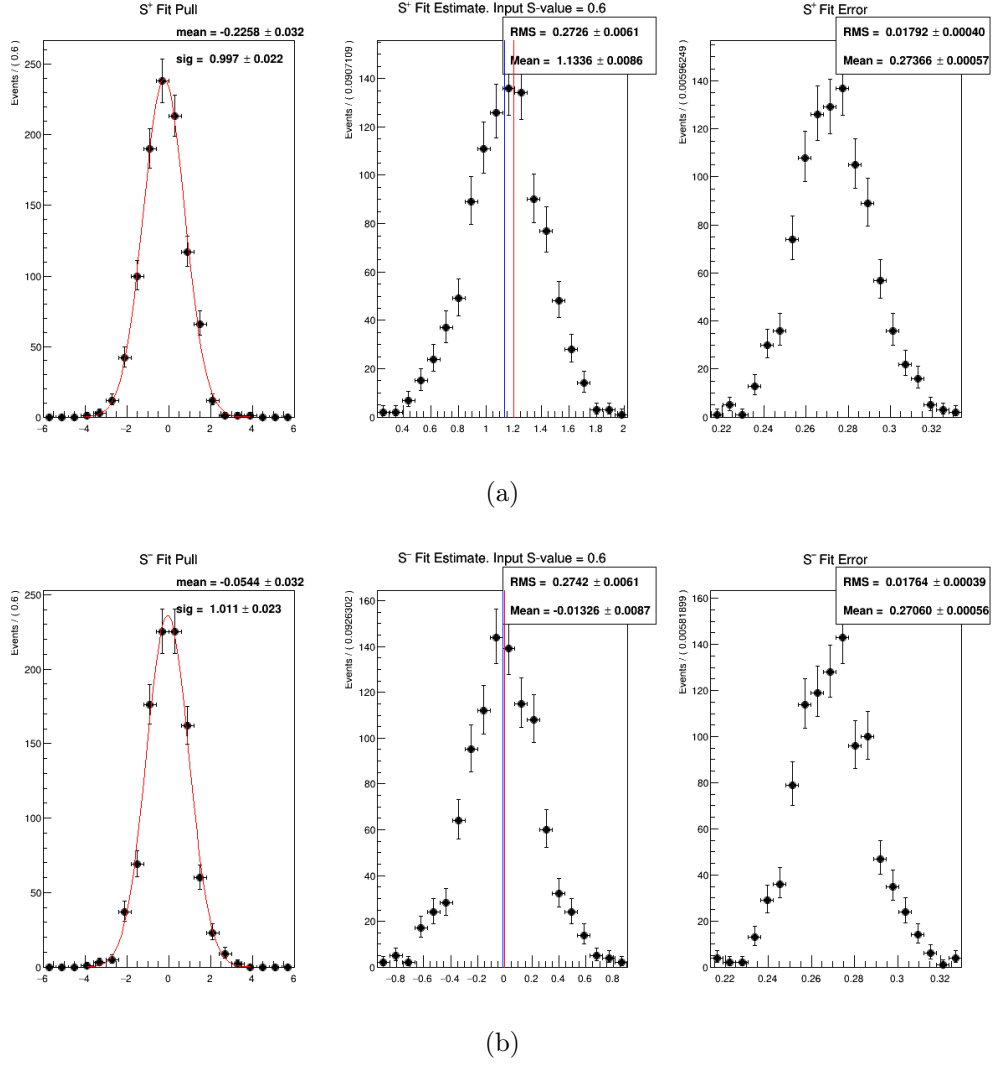


Figure 5.28: Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

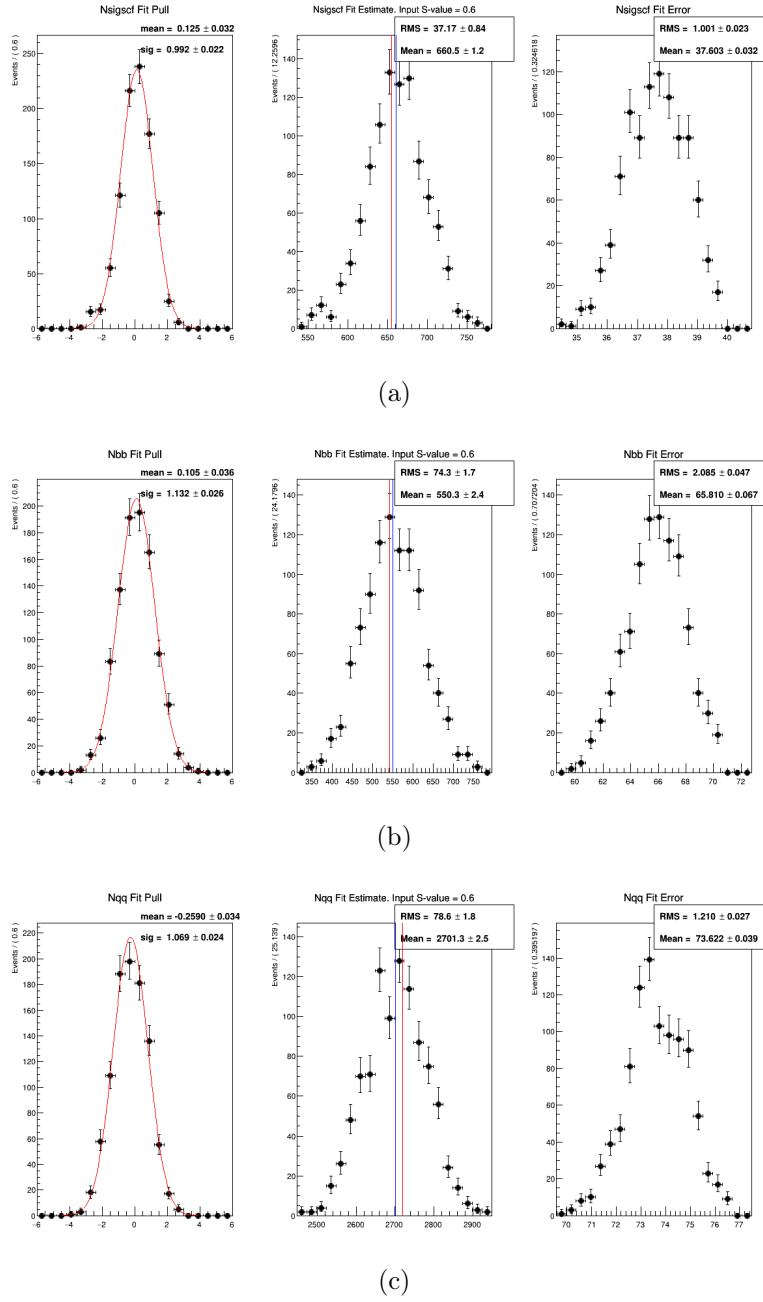


Figure 5.29: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

Table 5.17: Results of the ensemble test using bootstrap samples with varying input S values: Pull mean. All values are expected to be consistent with 0.

Input S	S	C	S^+	S^-
-0.6	-0.27 ± 0.03	-0.24 ± 0.03	-0.27 ± 0.03	0.12 ± 0.03
-0.4	-0.24 ± 0.03	-0.18 ± 0.03	-0.24 ± 0.03	0.07 ± 0.03
-0.2	-0.39 ± 0.03	-0.20 ± 0.03	-0.39 ± 0.03	0.06 ± 0.03
0.0	-0.36 ± 0.03	-0.23 ± 0.03	-0.35 ± 0.03	-0.04 ± 0.03
0.2	-0.30 ± 0.03	-0.10 ± 0.03	-0.31 ± 0.03	0.03 ± 0.03
0.4	-0.25 ± 0.03	-0.16 ± 0.03	-0.25 ± 0.03	-0.04 ± 0.03
0.6	-0.23 ± 0.03	-0.20 ± 0.03	-0.23 ± 0.03	-0.05 ± 0.03

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	0.19 ± 0.03	0.05 ± 0.03	-0.23 ± 0.03
-0.4	0.31 ± 0.03	-0.06 ± 0.04	-0.14 ± 0.04
-0.2	0.31 ± 0.03	-0.07 ± 0.04	-0.17 ± 0.04
0.0	0.54 ± 0.03	0.07 ± 0.04	-0.43 ± 0.03
0.2	0.30 ± 0.03	-0.08 ± 0.04	-0.09 ± 0.03
0.4	0.18 ± 0.03	0.04 ± 0.04	-0.23 ± 0.03
0.6	0.12 ± 0.03	0.10 ± 0.04	-0.26 ± 0.03

Table 5.18: Results of the ensemble test using bootstrap samples with varying input S values: Pull sigma. All values are expected to be consistent with 1.

Input S	S	C	S^+	S^-
-0.6	1.02 ± 0.02	0.97 ± 0.02	1.02 ± 0.02	1.03 ± 0.02
-0.4	1.03 ± 0.02	1.04 ± 0.02	1.03 ± 0.02	1.01 ± 0.02
-0.2	1.03 ± 0.02	1.00 ± 0.02	1.04 ± 0.02	1.00 ± 0.02
0.0	0.99 ± 0.02	1.00 ± 0.02	1.00 ± 0.02	1.00 ± 0.02
0.2	1.02 ± 0.02	1.04 ± 0.02	1.03 ± 0.02	1.00 ± 0.02
0.4	1.02 ± 0.02	0.99 ± 0.02	1.02 ± 0.02	0.98 ± 0.02
0.6	1.01 ± 0.02	0.98 ± 0.02	1.00 ± 0.02	1.01 ± 0.02

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	0.98 ± 0.02	1.08 ± 0.02	1.05 ± 0.02
-0.4	1.01 ± 0.02	1.11 ± 0.03	1.09 ± 0.02
-0.2	1.01 ± 0.02	1.17 ± 0.03	1.12 ± 0.03
0.0	0.98 ± 0.02	1.12 ± 0.03	1.04 ± 0.02
0.2	0.95 ± 0.02	1.10 ± 0.03	1.05 ± 0.02
0.4	1.05 ± 0.02	1.15 ± 0.03	1.09 ± 0.02
0.6	0.99 ± 0.02	1.13 ± 0.03	1.07 ± 0.02

Table 5.19: Results of the ensemble test using bootstrap samples with varying input S values: Absolute bias mean. All values are expected to be consistent with 0.

Input S	S	C	S^+	S^-
-0.6	-0.033 ± 0.004	-0.020 ± 0.003	-0.064 ± 0.008	0.033 ± 0.009
-0.4	-0.033 ± 0.004	-0.022 ± 0.003	-0.063 ± 0.009	0.017 ± 0.008
-0.2	-0.054 ± 0.005	-0.018 ± 0.003	-0.110 ± 0.009	0.024 ± 0.009
0.0	-0.046 ± 0.004	-0.027 ± 0.003	-0.091 ± 0.009	-0.018 ± 0.009
0.2	-0.048 ± 0.005	-0.012 ± 0.003	-0.099 ± 0.009	0.011 ± 0.008
0.4	-0.035 ± 0.004	-0.017 ± 0.003	-0.079 ± 0.009	-0.017 ± 0.009
0.6	-0.031 ± 0.004	-0.017 ± 0.003	-0.061 ± 0.008	-0.014 ± 0.008

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	7 ± 1	4 ± 2	-18 ± 2
-0.4	11 ± 1	-3 ± 2	-8 ± 2
-0.2	10 ± 1	-3 ± 2	-13 ± 3
0.0	21 ± 1	5 ± 2	-32 ± 2
0.2	11 ± 1	-4 ± 2	-4 ± 2
0.4	6 ± 1	1 ± 2	-13 ± 3
0.6	7 ± 1	5 ± 2	-19 ± 3

Table 5.20: Results of the ensemble test using bootstrap samples with varying input S values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.

Input S	S	C	S^+	S^-
−0.6	0.134 ± 0.006	0.098 ± 0.004	0.267 ± 0.012	0.275 ± 0.012
−0.4	0.142 ± 0.006	0.103 ± 0.005	0.287 ± 0.013	0.267 ± 0.012
−0.2	0.144 ± 0.006	0.100 ± 0.004	0.289 ± 0.013	0.288 ± 0.013
0.0	0.140 ± 0.006	0.100 ± 0.004	0.278 ± 0.012	0.286 ± 0.013
0.2	0.148 ± 0.007	0.105 ± 0.005	0.298 ± 0.013	0.262 ± 0.012
0.4	0.141 ± 0.006	0.101 ± 0.005	0.284 ± 0.013	0.271 ± 0.012
0.6	0.130 ± 0.006	0.101 ± 0.005	0.259 ± 0.012	0.265 ± 0.012

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
−0.6	39 ± 2	72 ± 3	77 ± 4
−0.4	39 ± 2	72 ± 3	77 ± 4
−0.2	38 ± 2	75 ± 3	78 ± 4
0.0	39 ± 2	75 ± 3	77 ± 3
0.2	37 ± 2	76 ± 3	77 ± 4
0.4	41 ± 2	72 ± 3	78 ± 4
0.6	36 ± 2	75 ± 3	81 ± 4

The pull distributions for C , S , S^+ and S^- are found to be consistent with Gaussians of $\mu = 0$ and $\sigma = 1$. Slight deviations in the mean ($< 1\sigma$) are observed, however, given the complexity of the fit and the nature of the bootstrapping procedure, where 1,000 samples are drawn with replacement from a parent MC sample containing only four times the expected background yield, such deviations are not unexpected. To assess this further, additional validation is performed using toy datasets (see Section 5.4.1).

The fitted yields are also found to exhibit a small bias, where, on average, the signal-like yield is overestimated by $\sim 0.3\sigma$, while the continuum background yield is underestimated by $\sim 0.2\sigma$. To assess the impact on the CP parameters, the fit is repeated with the yields fixed to the expectations from the generic MC (see Table 5.13), and the effect is found to be negligible. Nevertheless, a dedicated systematic uncertainty is assigned to cover for this yield bias (see Section 5.5.7).

To obtain an overall picture, all the distributions are combined into so-called linearity figures, where the means and widths of each distribution are compared across all (S, C) pairs, to check for trends with respect to the input value of S . Figure 5.30 shows these linearity figures for S , S^+ and S^- , along with plots for the residuals, defined as $(S - S_{\text{true}})$, $(C - C_{\text{true}})$ and $(N_{\text{sig-like}} - N_{\text{sig-like}}^{\text{true}})$ are shown. The uncertainties shown for each (S, C) pair do not account for the intrinsic correlations among the MC cocktail samples, which arise from the limited statistics of the parent MC datasets.

A small negative trend is observed in the S^- residual plot, but otherwise the fitted CP parameters and the signal-like yield do not show any dependence on the input S value. The uncertainties on S , C , S^+ and S^- as they are obtained from this ensemble test also seem to be estimated correctly (the width of the pull distributions is consistent with 1) and are on the level of $\sigma_S = 0.16$, $\sigma_C = 0.11$, $\sigma_{S^+} = 0.31$ and $\sigma_{S^-} = 0.30$, respectively. For S and C , this is already better than the previous measurements by Belle and BABAR.

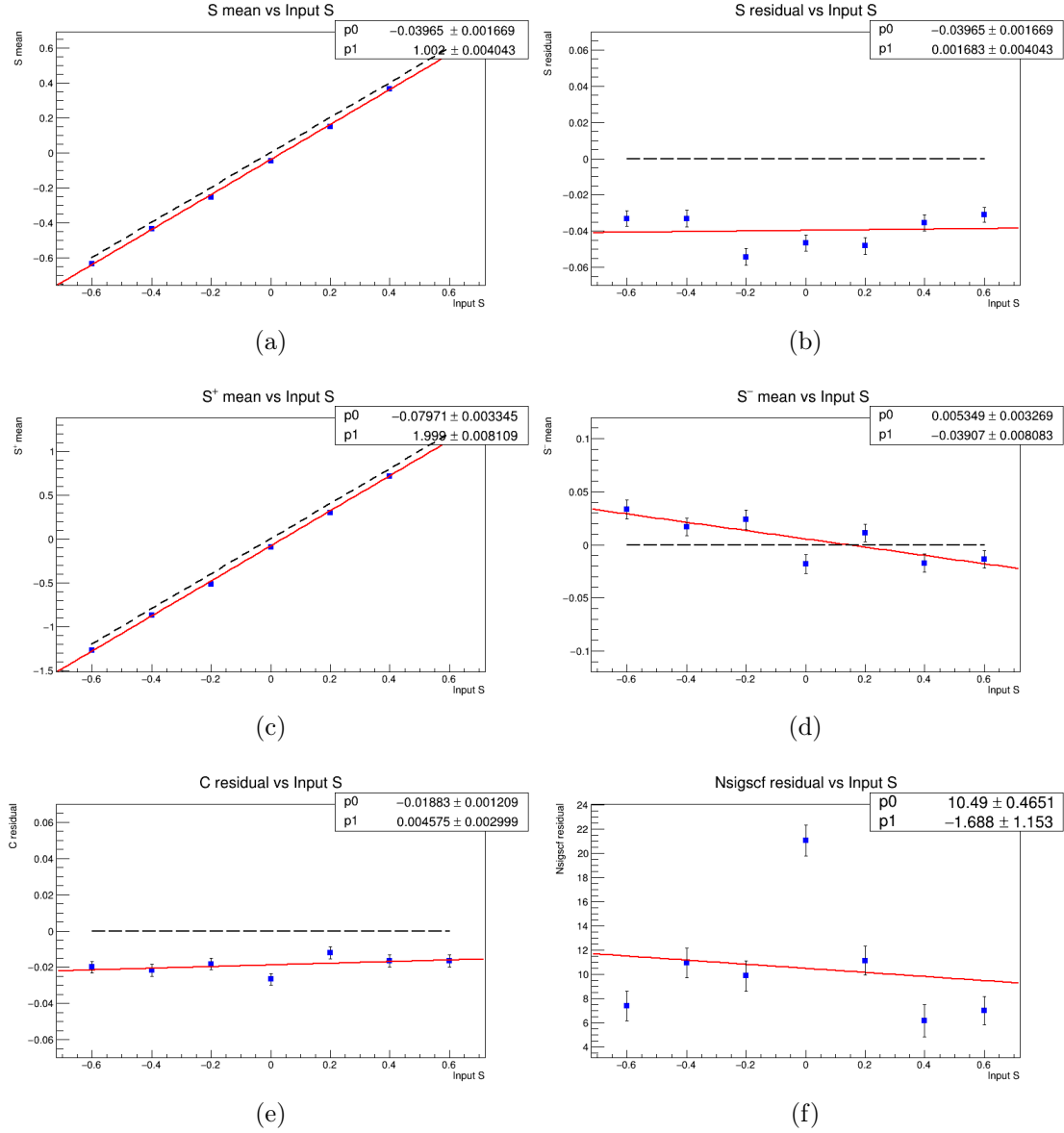


Figure 5.30: Linearity figures for (a) S mean, (b) S residual, (c) S^+ mean, (d) S^- mean, (e) C residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input S values.

Fit validation with toy datasets

The biases observed with bootstrapping might originate either from the choice of fit models or from complications in the fit itself. Thus, the fit procedure is also validated using toy datasets, i.e. datasets with events generated from the PDF model used to fit the simulated data. A small complication arises when generating these toy datasets due to the fact that the conditional parameter $\sigma_{\Delta t}$ enters in the tail fraction parameter of the Δt Resolution Function. Since the `generate` method of ROOFIT is unable to handle this particular case, a workaround was devised where events are generated in several bins of $\sigma_{\Delta t}$ with the fraction of events in each bin taken from simulation. The value of $\sigma_{\Delta t}$ for all events in a given bin is then manually set to the average value of $\sigma_{\Delta t}$ in that bin. In this case, the $\sigma_{\Delta t}$ distribution is split into 10 bins of equal size. Due to the way the fit model is set up, the same treatment is also needed for the Flavor Tagger qr , since it is also considered a conditional parameter. Since the qr distribution is split into 2×7 bins, events are generated in $10 \times 14 = 140$ two-dimensional bins of $\sigma_{\Delta t}$ and qr . A 2D histogram in $\sigma_{\Delta t}$ and qr showing the generated events for a single toy dataset that corresponds to the expected luminosity of the Belle II dataset, is shown in Figure 5.31. A set of 7×1000 toy datasets is generated with different values of S

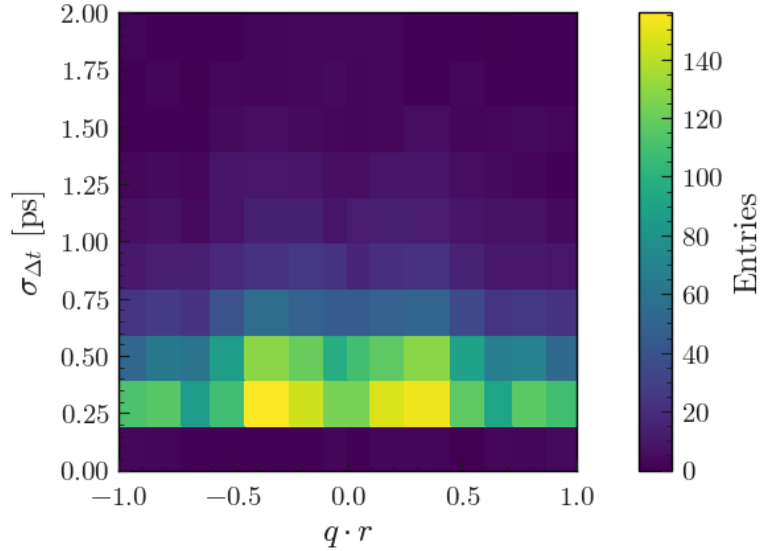


Figure 5.31: 2D $\sigma_{\Delta t}$ - qr distribution of events generated for a single toy dataset with $(S, C) = (0.0, 0.0)$ corresponding to the expected luminosity of the Belle II dataset.

and with $C = 0$, as well as another set of 7×1000 with different values of C and with $S = 0$, with the values being the same ones used in the bootstrapping, i.e.

$\{-0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6\}$ for both S and C . Each of these toy datasets corresponds to the expected luminosity of the Belle II dataset.

The number of events generated corresponding to each component in the fit, i.e. signal, SCF background, $B\bar{B}$ background and $q\bar{q}$ background, is determined from the generic MC sample in each 2D $\sigma_{\Delta t}$ - qr bin and subsequently varied following a Poisson distribution. Appendix D.10 shows comparison plots between the distributions of a toy sample generated with $(S, C) = (0.0, 0.0)$ and those of the generic MC sample.

In the case of the generated toy datasets with different values of C an additional correction is applied to the generated number of signal and SCF events, in order to reproduce the asymmetry expected in the qr distribution of those events as a result of direct CP violation. This correction depends on the value of C that is used to generate the events, as well as the qr bin in which the generated events lie. Specifically, the number of generated signal events in the i -th $\sigma_{\Delta t}$ bin and j -th qr bin with a given value of C is: $N_{i,j}^{\text{sig,corr}} = \{1 - q_{\text{tag}}\Delta w_j - q_{\text{tag}}(1 - 2w_j) \cdot C\} \times N_{i,j}^{\text{sig}}$, where q_{tag} is the B_{tag} flavor and $N_{i,j}^{\text{sig}}$ is the expected number of signal events in the i -th $\sigma_{\Delta t}$ bin and j -th qr bin, as determined from simulation⁵. The correction factor used for SCF background events is similar to the one used for signal, only with an additional factor of 0.8 diluting the value of C , following the treatment discussed in Section 5.3.3.

As was done with the bootstrapped datasets, the two sets of 7×1000 toy datasets are fitted to extract the CP parameters of interest and the fit results are checked for possible biases. As examples, the distributions of the fit pulls, fit estimates and fit errors for S , C , S^+ , S^- and the yields $N_{\text{sig-like}}$, $N_{B\bar{B} \text{ bkg}}$ and $N_{q\bar{q} \text{ bkg}}$ corresponding to the input values $(S, C) = (0.6, 0.0)$ are shown in Figures 5.32, 5.33 and 5.34, while the ones corresponding to the input values $(S, C) = (0.0, 0.6)$ are shown in Figures 5.35, 5.36 and 5.37. The resulting means and sigmas of the pulls and of the absolute biases of the CP parameters and yields are listed in Tables 5.21–5.24 for all input values of S and in Tables 5.25–5.28 for all input values of C . Linearity figures that show the overall picture are once again made and shown in Figures 5.38 and 5.39 for the various input values of S and C , respectively.

⁵Here, it should be understood that Δw_j and w_j are set to the values they have in the corresponding r -bin, so that for example: $w_1^{qr} = w_{14}^{qr} = w_7^r$

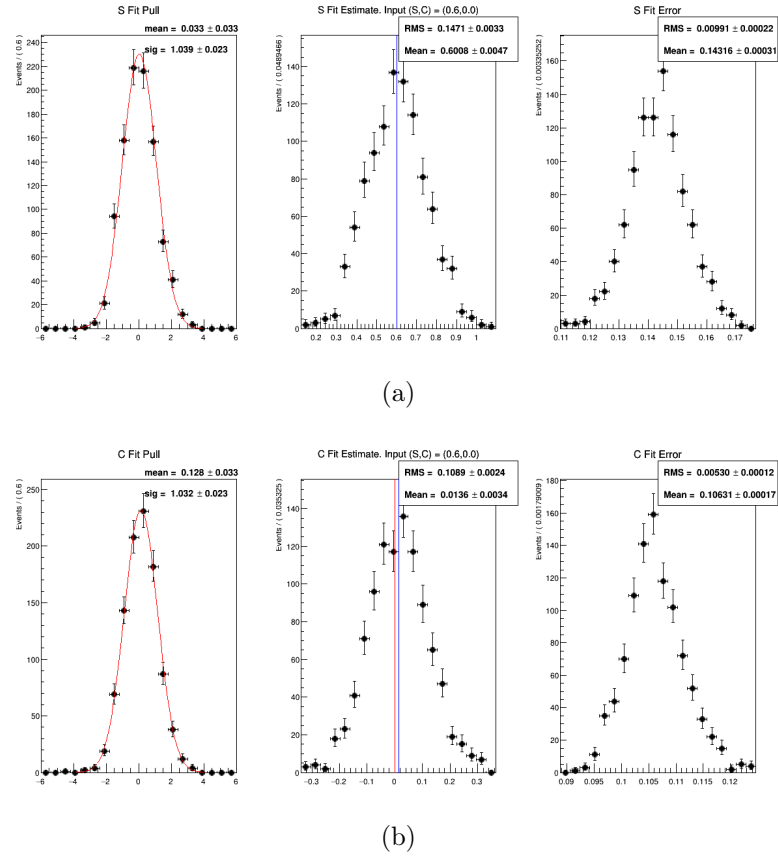


Figure 5.32: Toys: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

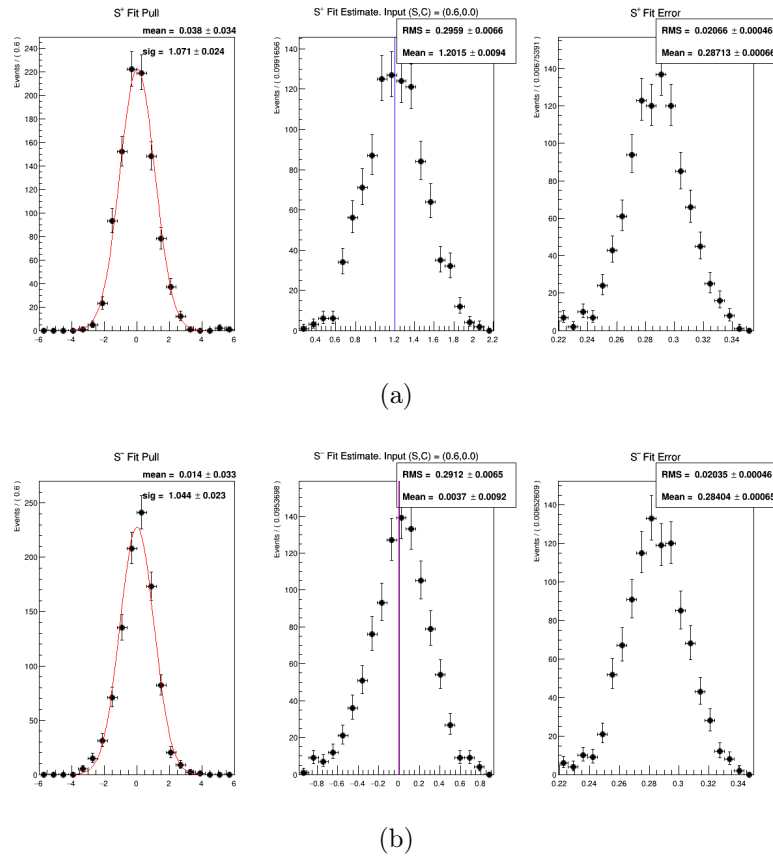


Figure 5.33: Toys: Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

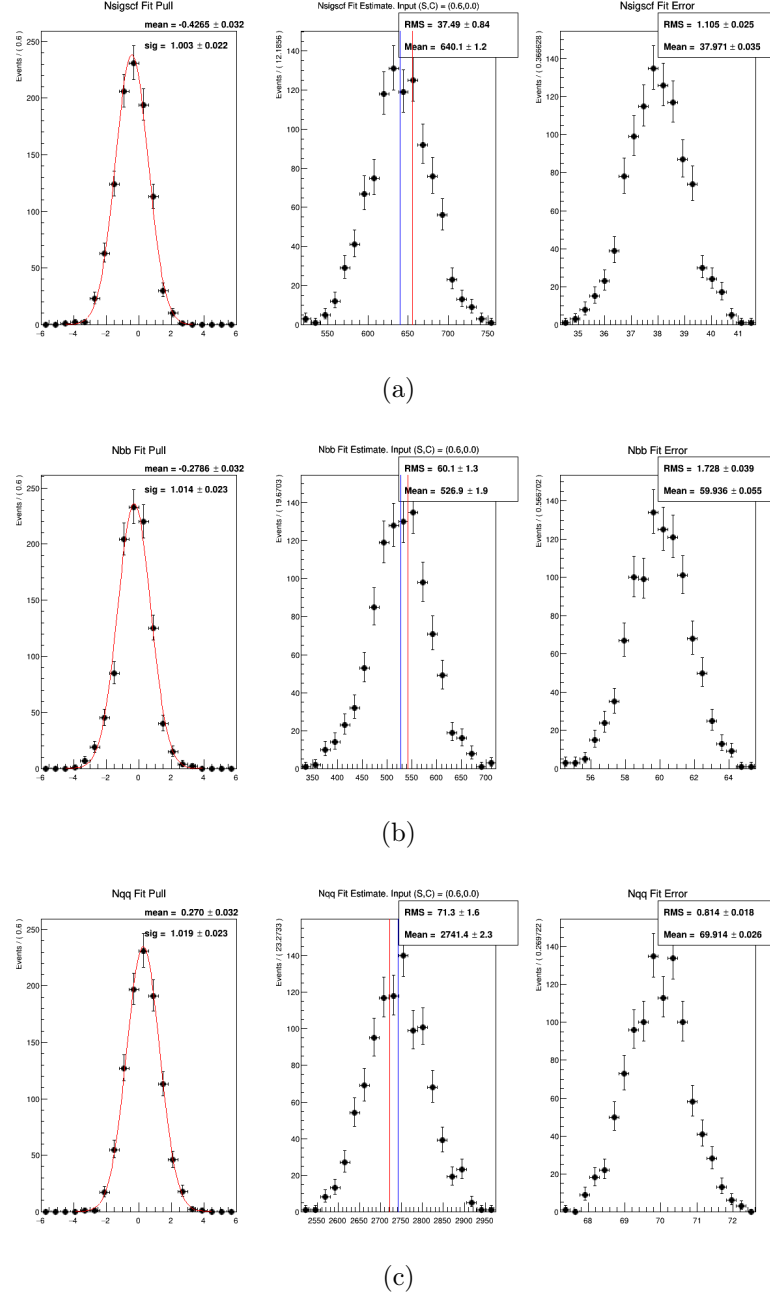


Figure 5.34: Toys: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

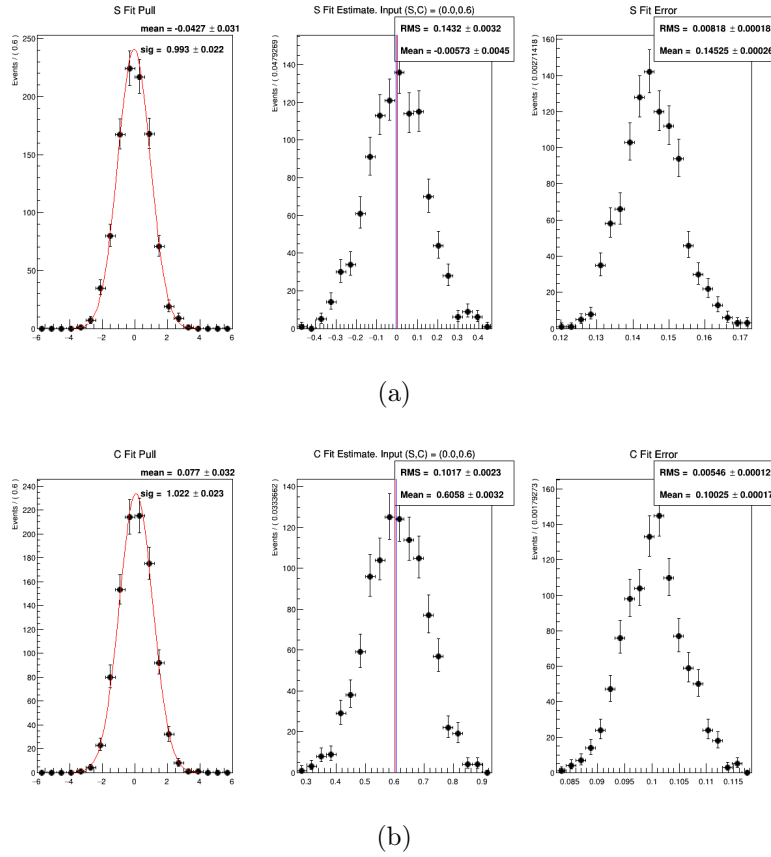


Figure 5.35: Toys: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.0, 0.6)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

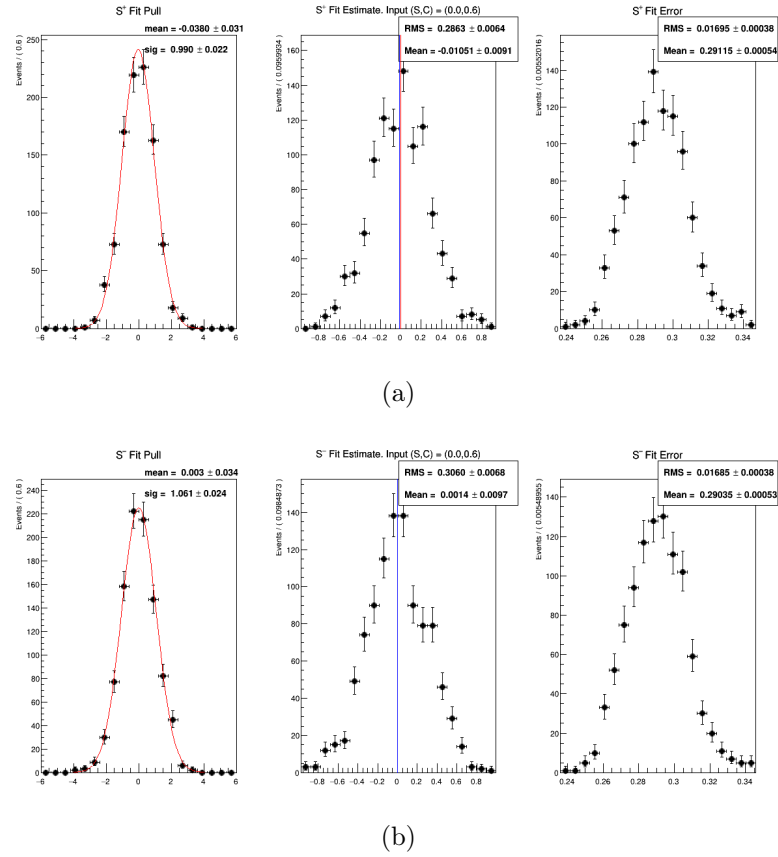


Figure 5.36: Toys: Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.0, 0.6)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

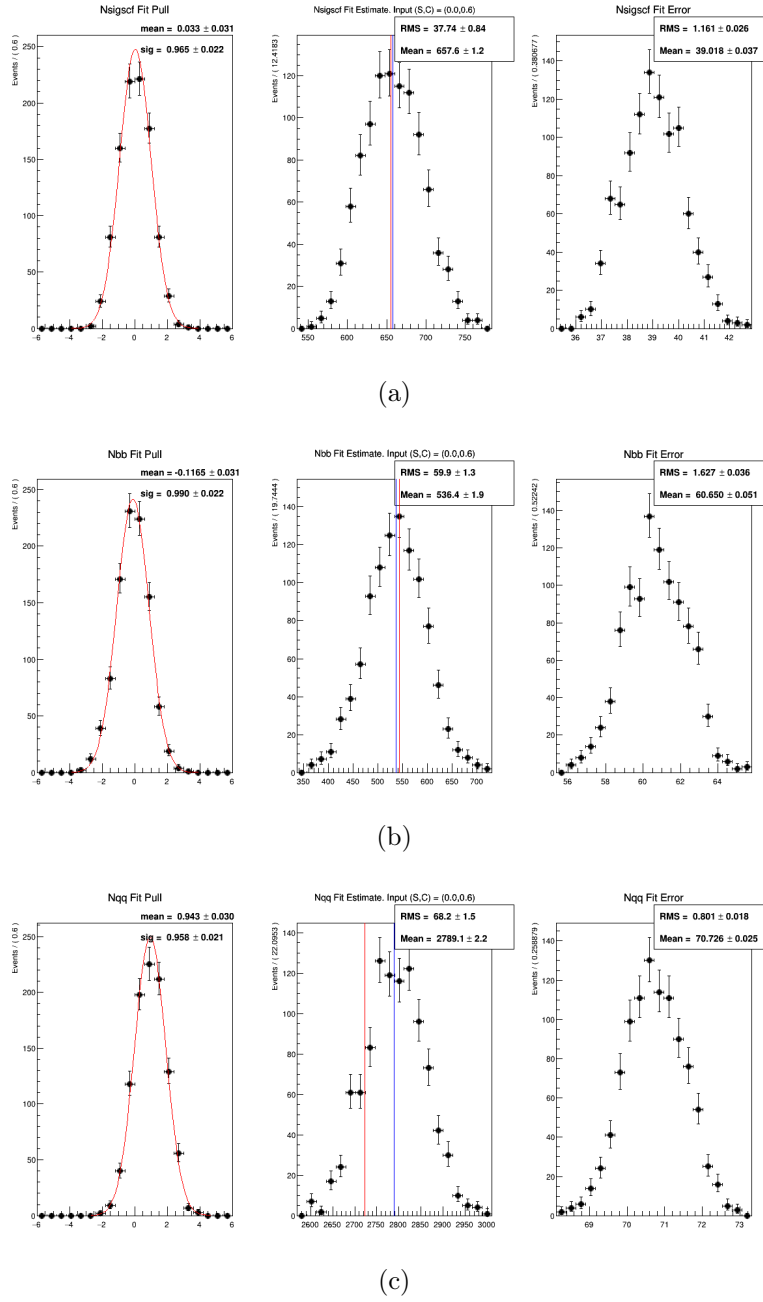


Figure 5.37: Toys: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.0, 0.6)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

Table 5.21: Results of the ensemble test using toy datasets with varying input S values: Pull mean. All values are expected to be consistent with 0.

Input S	S	C	S^+	S^-
-0.6	-0.12 ± 0.03	0.10 ± 0.03	-0.11 ± 0.03	0.03 ± 0.03
-0.4	-0.03 ± 0.03	0.04 ± 0.03	-0.03 ± 0.03	0.01 ± 0.03
-0.2	-0.07 ± 0.03	0.11 ± 0.03	-0.08 ± 0.03	-0.01 ± 0.03
0.0	-0.02 ± 0.03	0.04 ± 0.03	-0.01 ± 0.03	-0.02 ± 0.03
0.2	0.05 ± 0.03	0.08 ± 0.03	0.05 ± 0.03	0.02 ± 0.03
0.4	0.05 ± 0.03	0.06 ± 0.03	0.06 ± 0.03	0.02 ± 0.03
0.6	0.03 ± 0.03	0.13 ± 0.03	0.04 ± 0.03	0.01 ± 0.03

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	-0.49 ± 0.03	-0.29 ± 0.03	0.30 ± 0.03
-0.4	-0.49 ± 0.03	-0.25 ± 0.03	0.27 ± 0.03
-0.2	-0.48 ± 0.03	-0.21 ± 0.03	0.32 ± 0.03
0.0	-0.48 ± 0.03	-0.25 ± 0.03	0.30 ± 0.03
0.2	-0.42 ± 0.03	-0.25 ± 0.03	0.20 ± 0.03
0.4	-0.47 ± 0.03	-0.24 ± 0.03	0.27 ± 0.03
0.6	-0.43 ± 0.03	-0.28 ± 0.03	0.27 ± 0.03

Table 5.22: Results of the ensemble test using toy datasets with varying input S values: Pull sigma. All values are expected to be consistent with 1.

Input S	S	C	S^+	S^-
-0.6	1.04 ± 0.02	1.03 ± 0.02	1.04 ± 0.02	0.99 ± 0.02
-0.4	1.02 ± 0.02	1.00 ± 0.02	1.01 ± 0.02	1.03 ± 0.02
-0.2	1.02 ± 0.02	1.00 ± 0.02	1.02 ± 0.02	1.04 ± 0.02
0.0	1.06 ± 0.02	0.98 ± 0.02	1.07 ± 0.02	1.04 ± 0.02
0.2	1.01 ± 0.02	0.99 ± 0.02	1.01 ± 0.02	1.02 ± 0.02
0.4	1.02 ± 0.02	1.06 ± 0.02	1.02 ± 0.02	0.98 ± 0.02
0.6	1.04 ± 0.02	1.03 ± 0.02	1.07 ± 0.02	1.04 ± 0.02

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	0.99 ± 0.02	1.01 ± 0.02	1.01 ± 0.02
-0.4	1.10 ± 0.02	1.04 ± 0.02	1.01 ± 0.02
-0.2	1.01 ± 0.02	1.02 ± 0.02	1.00 ± 0.02
0.0	1.04 ± 0.02	1.02 ± 0.02	0.99 ± 0.02
0.2	1.02 ± 0.02	1.06 ± 0.02	1.00 ± 0.02
0.4	1.02 ± 0.02	1.03 ± 0.02	0.98 ± 0.02
0.6	1.00 ± 0.02	1.01 ± 0.02	1.02 ± 0.02

Table 5.23: Results of the ensemble test using toy datasets with varying input S values: Absolute bias mean. All values are expected to be consistent with 0.

Input S	S	C	S^+	S^-
-0.6	-0.012 ± 0.005	0.013 ± 0.003	-0.019 ± 0.010	0.005 ± 0.009
-0.4	-0.011 ± 0.005	0.003 ± 0.003	-0.020 ± 0.009	-0.004 ± 0.010
-0.2	-0.014 ± 0.005	0.012 ± 0.004	-0.027 ± 0.010	-0.009 ± 0.010
0.0	-0.006 ± 0.005	0.007 ± 0.003	-0.009 ± 0.010	0.004 ± 0.010
0.2	0.006 ± 0.005	0.007 ± 0.003	0.012 ± 0.009	0.007 ± 0.010
0.4	0.009 ± 0.005	0.005 ± 0.004	0.021 ± 0.010	-0.004 ± 0.009
0.6	0.001 ± 0.005	0.015 ± 0.003	0.001 ± 0.009	0.021 ± 0.009

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	-19 ± 1	-17 ± 2	21 ± 2
-0.4	-20 ± 1	-17 ± 2	18 ± 2
-0.2	-19 ± 1	-12 ± 2	22 ± 2
0.0	-17 ± 1	-15 ± 2	17 ± 2
0.2	-17 ± 1	-10 ± 2	14 ± 2
0.4	-16 ± 1	-13 ± 2	18 ± 2
0.6	-16 ± 1	-15 ± 2	20 ± 2

Table 5.24: Results of the ensemble test using toy datasets with varying input S values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.

Input S	S	C	S^+	S^-
-0.6	0.150 ± 0.007	0.107 ± 0.005	0.302 ± 0.014	0.277 ± 0.012
-0.4	0.150 ± 0.007	0.103 ± 0.005	0.300 ± 0.013	0.302 ± 0.014
-0.2	0.155 ± 0.007	0.115 ± 0.005	0.314 ± 0.014	0.312 ± 0.014
0.0	0.157 ± 0.007	0.109 ± 0.005	0.316 ± 0.014	0.319 ± 0.014
0.2	0.143 ± 0.006	0.105 ± 0.005	0.283 ± 0.013	0.309 ± 0.014
0.4	0.151 ± 0.007	0.116 ± 0.005	0.302 ± 0.014	0.278 ± 0.012
0.6	0.145 ± 0.007	0.108 ± 0.005	0.292 ± 0.013	0.274 ± 0.012

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	35 ± 2	55 ± 3	67 ± 3
-0.4	40 ± 2	63 ± 3	69 ± 3
-0.2	35 ± 2	58 ± 3	70 ± 3
0.0	41 ± 2	59 ± 3	67 ± 3
0.2	37 ± 2	64 ± 3	67 ± 3
0.4	38 ± 2	55 ± 3	64 ± 3
0.6	34 ± 2	55 ± 3	71 ± 3

Table 5.25: Results of the ensemble test using toy datasets with varying input C values: Pull mean. All values are expected to be consistent with 0.

Input C	S	C	S^+	S^-
-0.6	-0.00 ± 0.03	-0.05 ± 0.03	-0.00 ± 0.03	0.04 ± 0.03
-0.4	-0.00 ± 0.03	-0.04 ± 0.03	-0.00 ± 0.03	0.02 ± 0.03
-0.2	0.06 ± 0.03	-0.08 ± 0.03	0.06 ± 0.03	0.00 ± 0.03
0.0	-0.02 ± 0.03	0.04 ± 0.03	-0.01 ± 0.03	-0.02 ± 0.03
0.2	-0.05 ± 0.03	0.08 ± 0.03	-0.05 ± 0.03	-0.01 ± 0.03
0.4	0.01 ± 0.03	-0.06 ± 0.03	0.01 ± 0.03	0.01 ± 0.03
0.6	-0.04 ± 0.03	0.08 ± 0.03	-0.04 ± 0.03	0.00 ± 0.03

Input C	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	-0.64 ± 0.03	-0.43 ± 0.03	-0.43 ± 0.03
-0.4	-0.51 ± 0.03	-0.40 ± 0.03	-0.18 ± 0.03
-0.2	-0.51 ± 0.03	-0.34 ± 0.03	0.08 ± 0.03
0.0	-0.48 ± 0.03	-0.25 ± 0.03	0.30 ± 0.03
0.2	-0.13 ± 0.03	-0.32 ± 0.03	0.55 ± 0.03
0.4	-0.28 ± 0.03	-0.21 ± 0.03	0.74 ± 0.03
0.6	0.03 ± 0.03	-0.12 ± 0.03	0.94 ± 0.03

Table 5.26: Results of the ensemble test using toy datasets with varying input C values: Pull sigma. All values are expected to be consistent with 1.

Input C	S	C	S^+	S^-
-0.6	0.99 ± 0.02	1.00 ± 0.02	1.00 ± 0.02	1.02 ± 0.02
-0.4	0.99 ± 0.02	0.99 ± 0.02	1.00 ± 0.02	1.03 ± 0.02
-0.2	0.99 ± 0.02	0.99 ± 0.02	0.98 ± 0.02	0.94 ± 0.02
0.0	1.06 ± 0.02	0.98 ± 0.02	1.07 ± 0.02	1.04 ± 0.02
0.2	1.04 ± 0.02	1.02 ± 0.02	1.04 ± 0.02	0.99 ± 0.02
0.4	1.02 ± 0.02	1.03 ± 0.02	1.03 ± 0.02	1.04 ± 0.02
0.6	0.99 ± 0.02	1.02 ± 0.02	0.99 ± 0.02	1.06 ± 0.02

Input C	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	1.03 ± 0.02	1.05 ± 0.02	0.97 ± 0.02
-0.4	1.06 ± 0.02	1.02 ± 0.02	0.98 ± 0.02
-0.2	0.99 ± 0.02	0.97 ± 0.02	0.99 ± 0.02
0.0	1.04 ± 0.02	1.02 ± 0.02	0.99 ± 0.02
0.2	1.02 ± 0.02	1.01 ± 0.02	0.97 ± 0.02
0.4	1.02 ± 0.02	1.02 ± 0.02	0.98 ± 0.02
0.6	0.97 ± 0.02	0.99 ± 0.02	0.96 ± 0.02

Table 5.27: Results of the ensemble test using toy datasets with varying input C values: Absolute bias mean. All values are expected to be consistent with 0.

Input C	S	C	S^+	S^-
-0.6	0.005 ± 0.005	-0.006 ± 0.003	0.004 ± 0.010	0.014 ± 0.009
-0.4	-0.003 ± 0.005	-0.001 ± 0.003	-0.009 ± 0.009	0.005 ± 0.010
-0.2	0.004 ± 0.005	-0.012 ± 0.003	0.003 ± 0.010	-0.003 ± 0.010
0.0	-0.006 ± 0.005	0.007 ± 0.003	-0.009 ± 0.010	0.004 ± 0.010
0.2	-0.009 ± 0.005	0.008 ± 0.003	-0.026 ± 0.011	0.000 ± 0.009
0.4	0.012 ± 0.005	-0.004 ± 0.004	0.021 ± 0.010	0.012 ± 0.010
0.6	-0.003 ± 0.005	0.008 ± 0.003	-0.007 ± 0.009	-0.001 ± 0.010

Input C	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	-23 ± 1	-26 ± 2	-28 ± 2
-0.4	-17 ± 1	-23 ± 2	-13 ± 2
-0.2	-18 ± 1	-22 ± 2	2 ± 2
0.0	-17 ± 1	-15 ± 2	17 ± 2
0.2	-5 ± 1	-18 ± 2	37 ± 2
0.4	-12 ± 1	-13 ± 2	50 ± 2
0.6	0 ± 1	-6 ± 2	65 ± 2

Table 5.28: Results of the ensemble test using toys with varying input C values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.

Input C	S	C	S^+	S^-
-0.6	0.153 ± 0.007	0.099 ± 0.004	0.302 ± 0.013	0.298 ± 0.013
-0.4	0.146 ± 0.007	0.104 ± 0.005	0.297 ± 0.013	0.315 ± 0.014
-0.2	0.149 ± 0.007	0.104 ± 0.005	0.303 ± 0.014	0.301 ± 0.013
0.0	0.157 ± 0.007	0.109 ± 0.005	0.316 ± 0.014	0.319 ± 0.014
0.2	0.163 ± 0.007	0.104 ± 0.005	0.336 ± 0.015	0.287 ± 0.013
0.4	0.155 ± 0.007	0.113 ± 0.005	0.308 ± 0.014	0.302 ± 0.013
0.6	0.145 ± 0.006	0.107 ± 0.005	0.287 ± 0.013	0.310 ± 0.014

Input C	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	38 ± 2	58 ± 3	66 ± 3
-0.4	38 ± 2	65 ± 3	69 ± 3
-0.2	36 ± 2	54 ± 2	68 ± 3
0.0	41 ± 2	59 ± 3	67 ± 3
0.2	38 ± 2	64 ± 3	66 ± 3
0.4	36 ± 2	58 ± 3	68 ± 3
0.6	39 ± 2	59 ± 3	67 ± 3

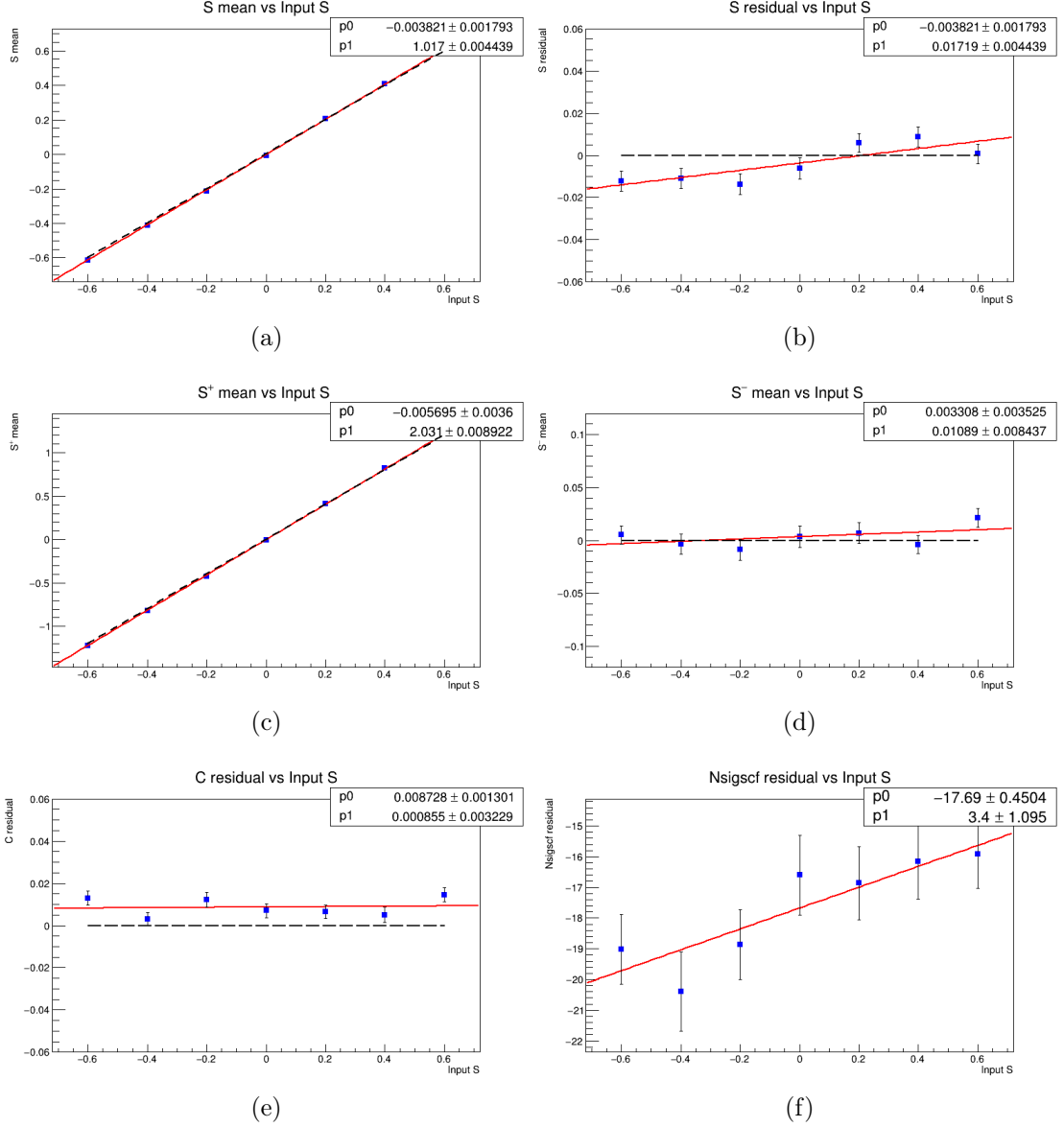


Figure 5.38: Toys: Linearity figures for (a) S mean, (b) S residual, (c) S^+ mean, (d) S^- mean, (e) C residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input S values.

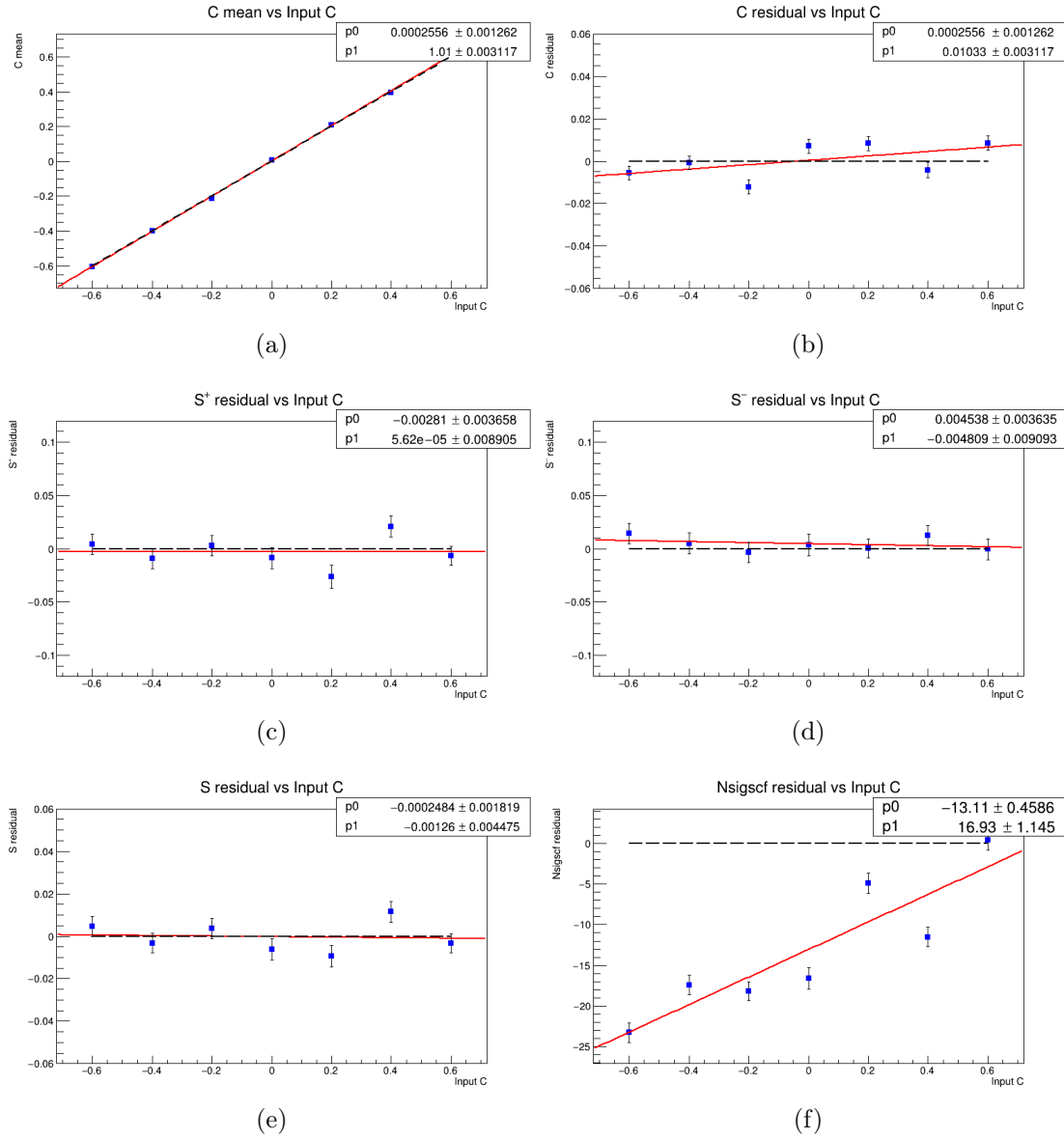


Figure 5.39: Toys: Linearity figures for (a) C mean, (b) C residual, (c) S^+ residual, (d) S^- residual, (e) S residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input C values.

In contrast to the bootstrapped datasets, there is no correlation between the generated signal and background components within each toy dataset, as well as across different toy datasets. Therefore, the error bars on each point in the above linearity figures accurately represent the uncertainty that arises from the fit procedure and the available statistics. That being said, the CP parameters are shown to be unbiased, albeit with a slight trend, while the bias on the signal-like yield is still present, though it is significantly reduced compared to the one seen with the bootstrap samples. Moreover, there seems to be a non-negligible trend in the fitted signal-like yields with respect to the input value of C . This most likely originates from the fact that the correction applied to the signal-like yield when generating the toy datasets itself depends on the C parameter.

The procedure described above makes use of toy datasets with events that were all generated using just the 3D PDF model of Equation 5.7. The results of the ensemble test using toy datasets are not in full agreement with the ones obtained from the test that was carried out using the MC cocktail samples. Thus, in order to obtain a better understanding of these discrepancies the ensemble test was repeated, this time using so called “embedded” toy datasets. These datasets are constructed in a hybrid manner, by bootstrapping signal and SCF events from the signal MC samples, as was done in Section 5.4.1, and by generating continuum and $B\bar{B}$ background events from the corresponding 3D PDF model of Equation 5.6, as was done in the previous section. Finally, these events are combined into a single “embedded” toy dataset. Once again, 7×1000 toy datasets are constructed in this way, with seven different input S values. Performing the ensemble test on these embedded toy datasets will help clarify whether the biases observed on S and the signal-like yield are due to the fit process, a by-product of the bootstrapping procedure, or something that arises due to a mismodeling of the background distributions in either ΔE , M_{bc} or Δt .

As before, the samples are fitted to extract the CP parameters and the results are examined to look for possible biases. As an example, the distributions of the fit pulls, fit estimates and fit errors for S , C , S^+ , S^- and the yields $N_{\text{sig-like}}$, $N_{B\bar{B} \text{ bkg}}$ and $N_{q\bar{q} \text{ bkg}}$ corresponding to the input values $(S, C) = (0.6, 0.0)$ are shown in Figures 5.40, 5.41 and 5.42. The resulting means and sigmas of the pulls and of the absolute biases of the CP parameters and yields are listed in Tables 5.29–5.32 for all input values of S . Linearity figures that show the overall picture are once again made and shown in Figure 5.43 for the various input values of S .

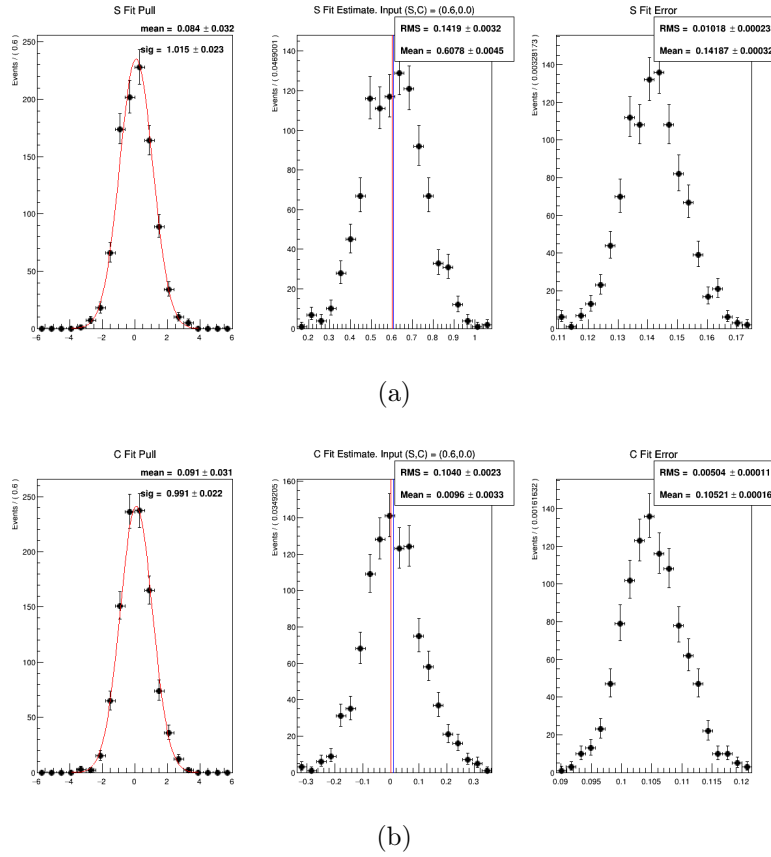


Figure 5.40: Embedded toys: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

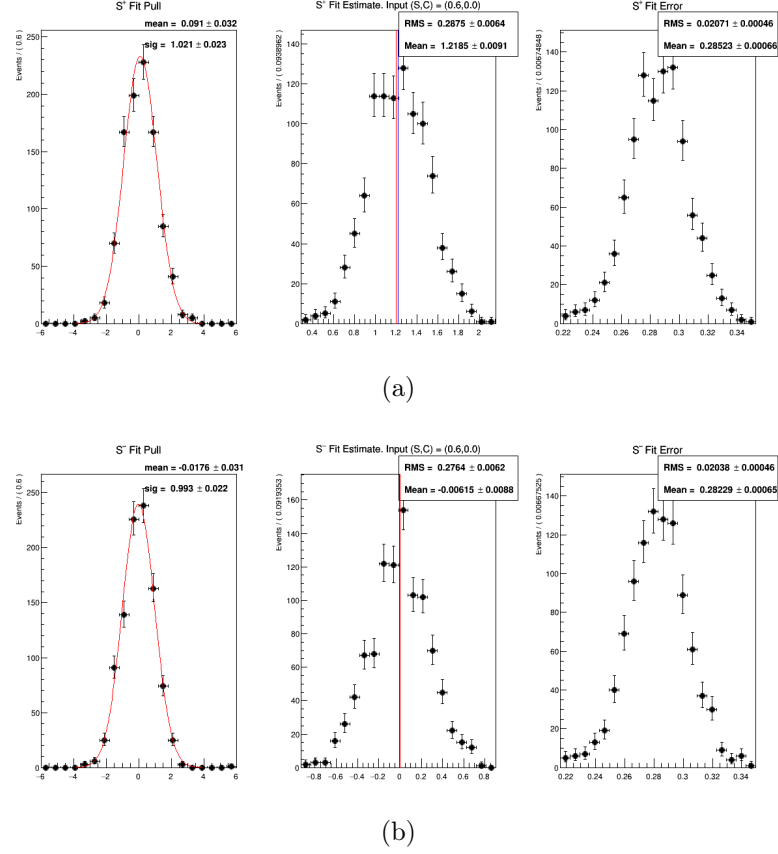


Figure 5.41: Embedded toys: Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

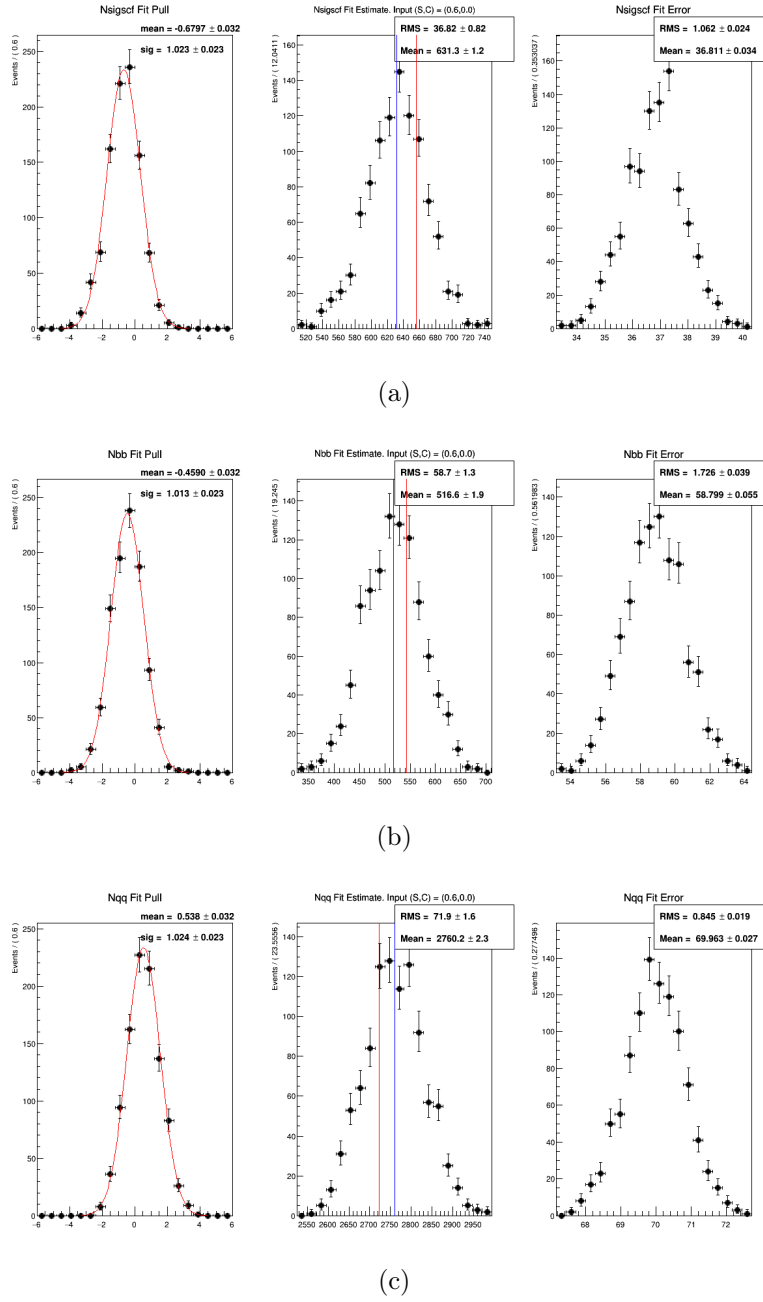


Figure 5.42: Embedded toys: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

Table 5.29: Results of the ensemble test using embedded toy datasets with varying input S values: Pull mean. All values are expected to be consistent with 0.

Input S	S	C	S^+	S^-
-0.6	-0.13 ± 0.03	0.06 ± 0.03	-0.12 ± 0.03	-0.06 ± 0.03
-0.4	-0.07 ± 0.03	0.00 ± 0.03	-0.07 ± 0.03	-0.00 ± 0.03
-0.2	-0.15 ± 0.03	0.00 ± 0.03	-0.15 ± 0.03	-0.01 ± 0.03
0.0	-0.02 ± 0.03	0.08 ± 0.03	-0.01 ± 0.03	-0.03 ± 0.03
0.2	0.02 ± 0.03	0.10 ± 0.03	0.02 ± 0.03	0.09 ± 0.03
0.4	0.11 ± 0.03	0.09 ± 0.03	0.11 ± 0.03	0.02 ± 0.03
0.6	0.08 ± 0.03	0.09 ± 0.03	0.09 ± 0.03	-0.02 ± 0.03

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	-0.77 ± 0.03	-0.42 ± 0.03	0.54 ± 0.03
-0.4	-0.75 ± 0.03	-0.41 ± 0.03	0.52 ± 0.03
-0.2	-0.67 ± 0.03	-0.50 ± 0.03	0.63 ± 0.03
0.0	-0.61 ± 0.03	-0.45 ± 0.03	0.53 ± 0.03
0.2	-0.72 ± 0.03	-0.47 ± 0.03	0.53 ± 0.03
0.4	-0.76 ± 0.03	-0.41 ± 0.03	0.56 ± 0.03
0.6	-0.68 ± 0.03	-0.46 ± 0.03	0.54 ± 0.03

Table 5.30: Results of the ensemble test using embedded toy datasets with varying input S values: Pull sigma. All values are expected to be consistent with 1.

Input S	S	C	S^+	S^-
-0.6	1.05 ± 0.02	1.05 ± 0.02	1.04 ± 0.02	1.01 ± 0.02
-0.4	1.01 ± 0.02	1.03 ± 0.02	1.02 ± 0.02	1.00 ± 0.02
-0.2	1.03 ± 0.02	1.00 ± 0.02	1.03 ± 0.02	1.02 ± 0.02
0.0	1.03 ± 0.02	1.00 ± 0.02	1.03 ± 0.02	1.03 ± 0.02
0.2	1.09 ± 0.02	0.96 ± 0.02	1.09 ± 0.02	1.00 ± 0.02
0.4	0.95 ± 0.02	1.00 ± 0.02	0.96 ± 0.02	1.04 ± 0.02
0.6	1.02 ± 0.02	0.99 ± 0.02	1.02 ± 0.02	0.99 ± 0.02

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	1.06 ± 0.02	1.04 ± 0.02	1.00 ± 0.02
-0.4	1.07 ± 0.02	1.05 ± 0.02	1.00 ± 0.02
-0.2	1.09 ± 0.02	1.04 ± 0.02	1.02 ± 0.02
0.0	1.03 ± 0.02	1.00 ± 0.02	1.00 ± 0.02
0.2	1.02 ± 0.02	1.02 ± 0.02	1.03 ± 0.02
0.4	1.05 ± 0.02	1.04 ± 0.02	1.00 ± 0.02
0.6	1.02 ± 0.02	1.01 ± 0.02	1.02 ± 0.02

Table 5.31: Results of the ensemble test using embedded toy datasets with varying input S values: Absolute bias mean. All values are expected to be consistent with 0.

Input S	S	C	S^+	S^-
-0.6	-0.014 ± 0.005	0.009 ± 0.003	-0.028 ± 0.009	-0.018 ± 0.009
-0.4	-0.009 ± 0.004	-0.000 ± 0.004	-0.018 ± 0.009	-0.011 ± 0.010
-0.2	-0.026 ± 0.005	0.001 ± 0.004	-0.045 ± 0.010	-0.005 ± 0.010
0.0	-0.004 ± 0.005	0.007 ± 0.003	-0.006 ± 0.010	-0.009 ± 0.010
0.2	0.009 ± 0.005	0.010 ± 0.003	0.016 ± 0.011	0.023 ± 0.009
0.4	0.012 ± 0.005	0.010 ± 0.003	0.023 ± 0.009	-0.002 ± 0.010
0.6	0.010 ± 0.005	0.007 ± 0.003	0.020 ± 0.009	0.002 ± 0.009

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
-0.6	-29 ± 1	-27 ± 2	35 ± 2
-0.4	-29 ± 1	-22 ± 2	33 ± 2
-0.2	-24 ± 1	-27 ± 2	44 ± 2
0.0	-23 ± 1	-29 ± 2	37 ± 2
0.2	-27 ± 1	-28 ± 2	38 ± 2
0.4	-28 ± 1	-28 ± 2	35 ± 2
0.6	-24 ± 1	-27 ± 2	35 ± 2

Table 5.32: Results of the ensemble test using embedded toy datasets with varying input S values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.

Input S	S	C	S^+	S^-
−0.6	0.148 ± 0.007	0.104 ± 0.005	0.299 ± 0.013	0.295 ± 0.013
−0.4	0.140 ± 0.006	0.111 ± 0.005	0.282 ± 0.013	0.302 ± 0.014
−0.2	0.155 ± 0.007	0.113 ± 0.005	0.310 ± 0.014	0.300 ± 0.013
0.0	0.158 ± 0.007	0.105 ± 0.005	0.310 ± 0.014	0.301 ± 0.013
0.2	0.165 ± 0.007	0.105 ± 0.005	0.335 ± 0.015	0.295 ± 0.013
0.4	0.146 ± 0.007	0.105 ± 0.005	0.293 ± 0.013	0.308 ± 0.014
0.6	0.148 ± 0.007	0.100 ± 0.004	0.295 ± 0.013	0.270 ± 0.012

Input S	$N_{\text{sig-like}}$	$N_{B\bar{B} \text{ bkg}}$	$N_{q\bar{q} \text{ bkg}}$
−0.6	35 ± 2	61 ± 3	74 ± 3
−0.4	39 ± 2	61 ± 3	66 ± 3
−0.2	39 ± 2	61 ± 3	72 ± 3
0.0	37 ± 2	57 ± 3	67 ± 3
0.2	36 ± 2	56 ± 3	71 ± 3
0.4	37 ± 2	59 ± 3	68 ± 3
0.6	35 ± 2	58 ± 3	69 ± 3

Compared to the ensemble test that used “pure” toy datasets, the results using embedded toys are not significantly different in that, there is no clear bias or trend in the CP parameters and the signal-like yield is still biased on the level of $\sim 0.7\sigma$. These results lead to the conclusion that the biases observed in the ensemble test using the bootstrap samples are only partly due to the fit process. More specifically, it seems that the small bias found on S and S^+ in that test originates either from how the continuum and $B\bar{B}$ backgrounds are modeled in ΔE , M_{bc} and Δt or from the correlations induced between these backgrounds from the bootstrapping process. A systematic uncertainty is assigned to cover for all of these effects and the details are discussed in Section 5.5.8.

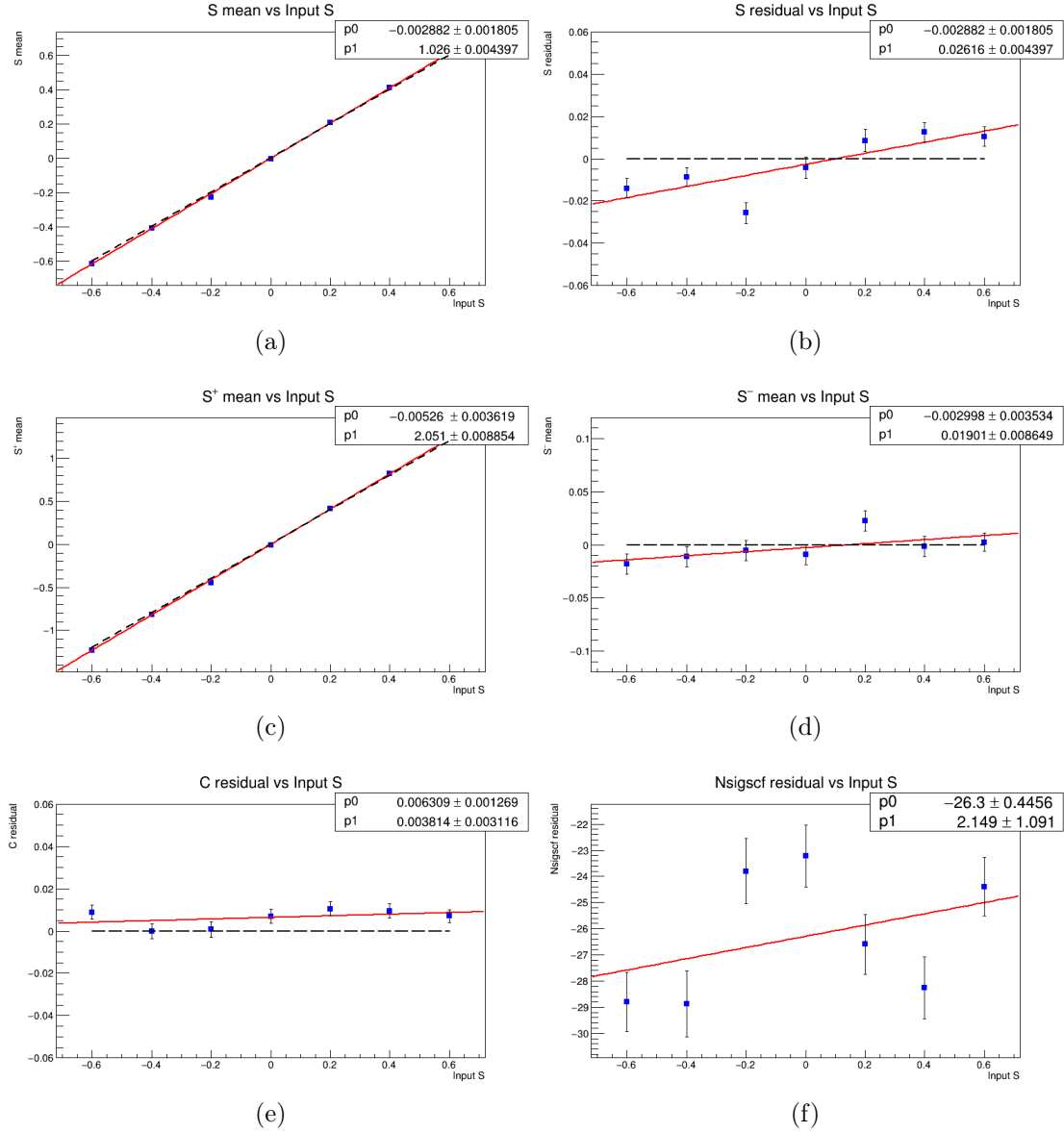


Figure 5.43: Embedded toys: Linearity figures for (a) S mean, (b) S residual, (c) S^+ mean, (d) S^- mean, (e) C residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input S values.

Agreement between data and simulation

The agreement between data and simulation is first checked using the off-resonance samples, which only contain continuum background events. Specifically, the ΔE and Δt distributions⁶ are compared between the off-resonance data and MC samples in Figure 5.44 and it can be seen that there is a significant excess of continuum background in data over MC, on the level of $\sim 20\%$. Such an excess is most likely due to a mismodeling of the $e^+e^- \rightarrow q\bar{q}$ process in the Belle II MC, which is handled by the Pythia generator. This is not expected to bias the measurement in any way, as the continuum background yield is free in the fit, however, it can lead to a false expectation for the precision on S and C , since the signal over background ratio will be lower in the data than what is expected from the simulation.

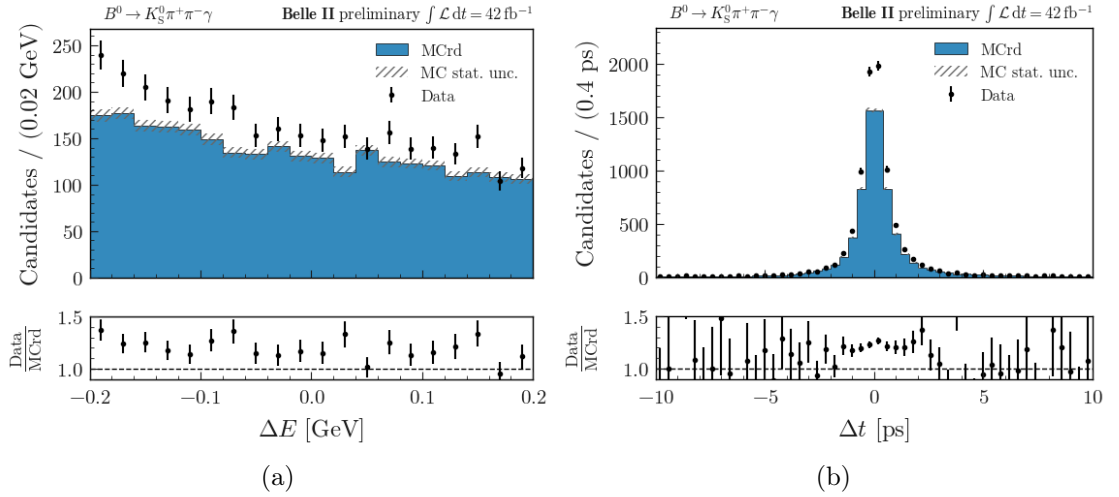


Figure 5.44: Comparison of the (a) ΔE and (b) Δt distributions between the off-resonance data and MC samples. The comparison is made before applying the cut on the CSMVA output.

The data–MC agreement is further tested using the on-resonance samples. The ΔE , M_{bc} and Δt distributions of the on-resonance data and MC samples are compared in Figure 5.45. The excess of continuum background events is also present here, while the signal yield seems to be overestimated in the simulation. Once again, this is not expected to bias the measurement, since the signal-like yield is free in the fit.

⁶The M_{bc} distribution of the off-resonance samples is significantly different to that of the on-resonance samples owing to the lower collision energy, so checking the data–MC agreement in M_{bc} is not really useful.

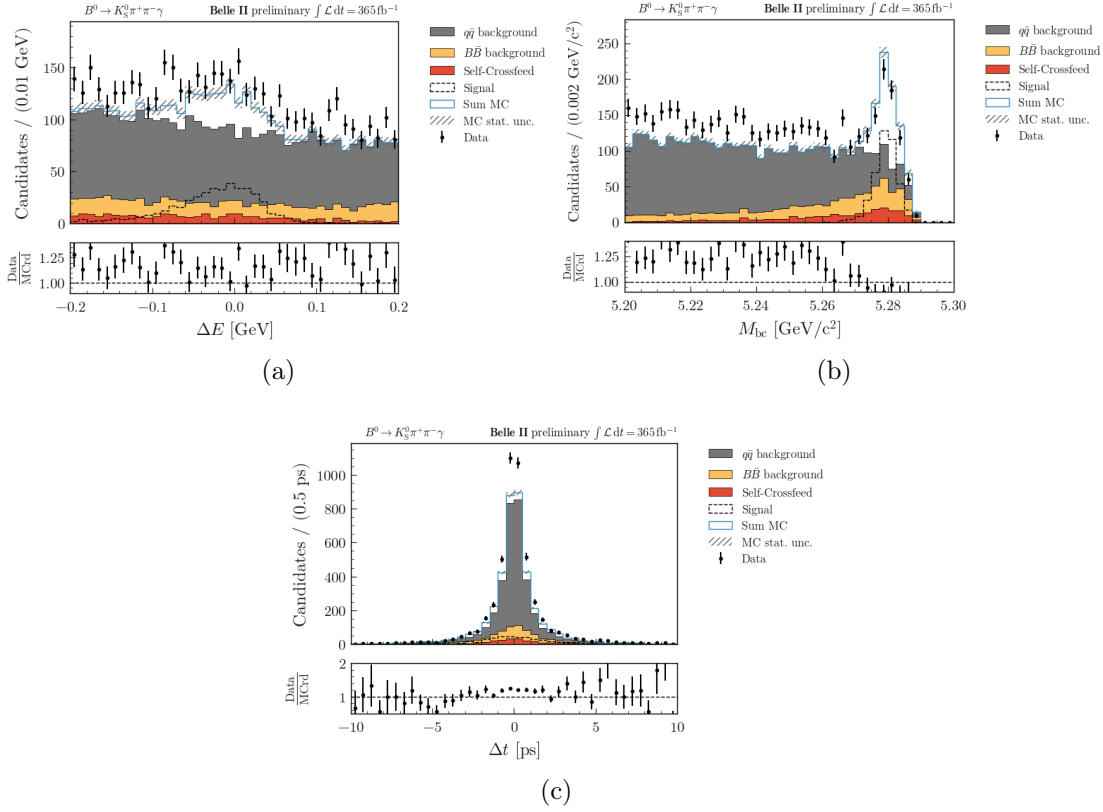


Figure 5.45: Comparison of the (a) ΔE , (b) M_{bc} and (b) Δt distributions of between the on-resonance data and MC samples. The background components are stacked on top of each other while the signal component is not.

To assess the impact of the underestimated continuum background and overestimated signal yield in the simulation on the expected sensitivity of the CP parameters, the ensemble test was repeated on an alternative set of bootstrapped datasets generated with 20% more continuum background and 20% less signal events. The resulting uncertainties on S and C from this ensemble test are found to be $\sim 15\%$ larger than the previously expected values.

Fit for the B^0 lifetime

As an additional check, the 3D fit is then performed on the Belle II data sample to extract the B^0 lifetime instead of the CP parameters, S and C , which are both fixed to zero. Figure 5.46 shows the projection of this fit onto Δt and the value obtained for the B^0 lifetime is compatible with the world average within 1σ .

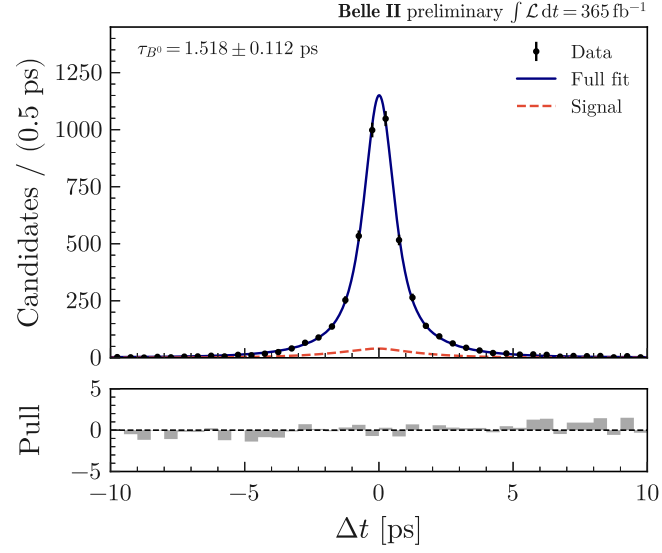


Figure 5.46: Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample to extract the B^0 lifetime projected onto Δt .

5.4.2 Validation using the control mode

The control mode, $B^+ \rightarrow K^+\pi^+\pi^-\gamma$, is used to validate the measurement strategy presented in Section 5.3 and to also check the agreement between data and simulation. The shapes of the ΔE , M_{bc} and Δt distributions for the signal and background components are found to be compatible between the signal and control mode (see Appendix D.8). Therefore, the same PDF models used to fix all the shape parameters in Sections 5.3.2 and 5.3.3 are also used to fix the shape parameters for the control mode (see Appendix D.9).

Due to isospin symmetry, the branching fraction of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays is expected to be similar to that of $B^0 \rightarrow K^0\pi^+\pi^-\gamma$ decays, but because, in the signal decay mode, K^0 are reconstructed only as $K_S^0 \rightarrow \pi^+\pi^-$, a significant portion of signal decays is lost. As a result, the expected amount of signal events in the control mode is ~ 5 times larger than those in the signal mode.

Almost all of the tests that are carried out using bootstrapped datasets in the signal mode are also performed in the control mode. The only major difference is that only a single set of values for S and C is taken into consideration, corresponding to $(S, C) = (0.0, 0.0)$. That is because, in this case, S does not have a physical meaning, since there is no mixing between B^+ and B^- .

In addition, as previously mentioned, the Flavor Tagger is used to tag events as B^0 or $\bar{B}^0$, just like in the signal mode, meaning the information on the charge of the B_{sig} daughters is not used to obtain knowledge on its flavor. The set of Flavor Tagger and Δt Resolution Function parameters used for the control mode validation is the same one used in the signal mode, meaning the set of values listed in Table 4.6 is used when fitting simulated samples and those listed in Table 4.7 is used when fitting the real data sample. It is worth noting, however, that since these parameters were calibrated using neutral B decays, the calibrated parameter values might differ significantly from those determined using MC truth information for the control mode. A comparison for the Flavor Tagger parameters can be seen in Figure 5.47, where a significant discrepancy can be seen between the calibrated values for the wrong tag fraction w and the values determined using MC truth information. The effect of this discrepancy in Flavor Tagger parameters on the final measurement is small and should be covered by the systematic uncertainty which accounts for possible fluctuations in those parameters (see Section 5.5.2).

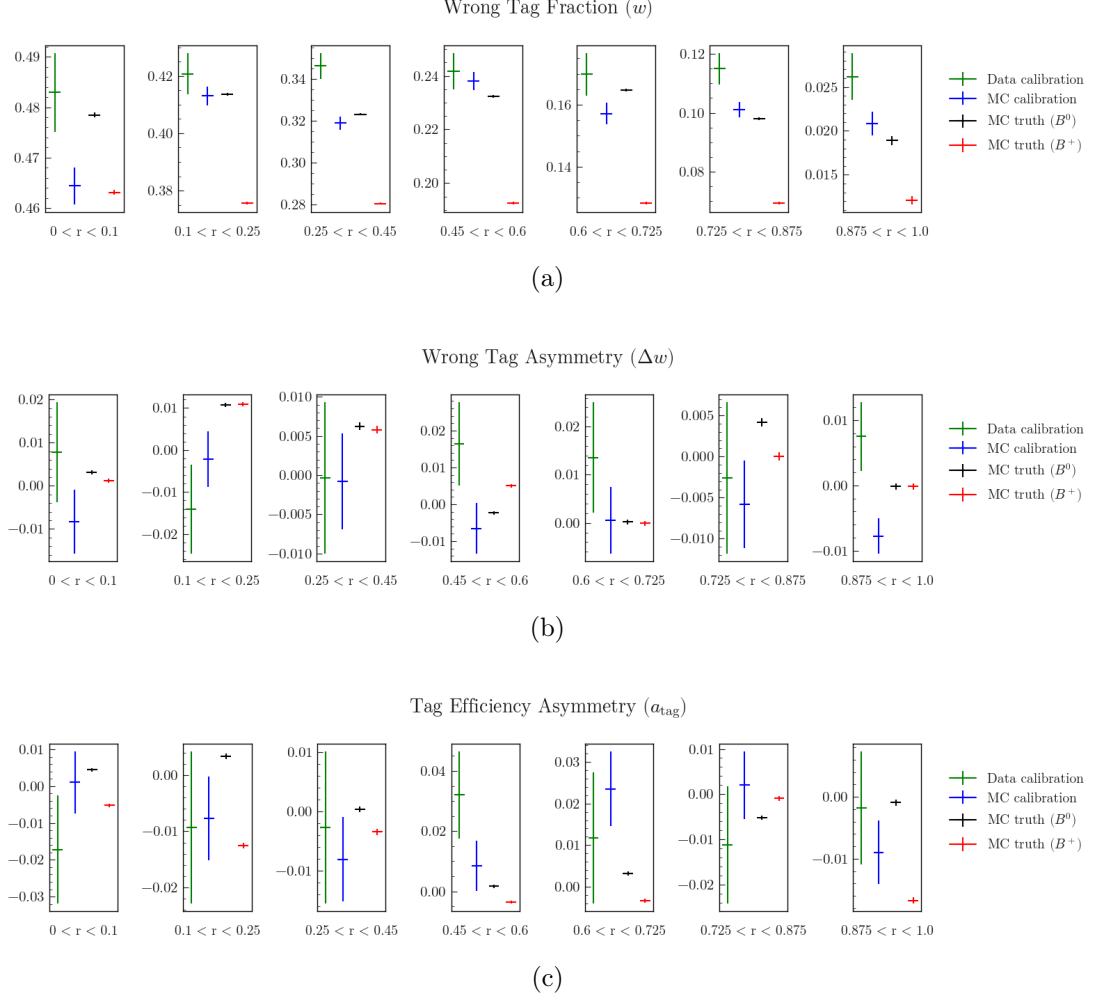


Figure 5.47: Comparison between Flavor Tagger parameters (a) w , (b) Δw and (c) a_{tag} obtained from the calibration using $B^0 \rightarrow D^{(*)-}\pi^+$ events and those obtained using MC truth information in the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ and $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ simulated samples.

3D fit to a simulated sample

As with the signal mode, the 3D fit was first performed on an MC cocktail sample for the control mode. The relative fraction of SCF background is once again fixed to values determined from simulation, listed in Table 5.33. The expected yields are also estimated from the generic MC sample and are shown in Table 5.34.

Table 5.33: Relative fractions between the signal and SCF background components for the control mode determined from simulation and further fixed in the fit to the data.

Component	f_{avg}	f_{0-5}	f_6
Signal	0.687	0.663	0.758
SCF	0.313	0.337	0.242

Table 5.34: Expected number of events for each of the four components as they are found within a data-like MC cocktail sample for the control mode. These correspond to 25% of the events found in the generic MC.

Component	N_i
Signal	2020 ± 45
SCF	920 ± 30
$B\bar{B}$ background	2080 ± 45
$q\bar{q}$ background	13640 ± 115

The result of the 3D fit to the MC cocktail sample is shown in Figures 5.48 and 5.49, while the fitted parameters are listed in Table 5.35. In Figure 5.49, as with the signal mode, to improve the visibility of the signal component, both the simulated data and fitted models are projected onto a signal enhanced region in ΔE , M_{bc} and r with $|\Delta E| < 0.1$ GeV, $M_{\text{bc}} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

The fitted CP parameters and yields are all compatible with their expected values within the statistical uncertainty. As expected the precision on S and C is significantly better compared to the signal mode, which is a direct consequence of the larger amount of signal events present in the control mode sample. The fit is also performed to extract the B^+ lifetime, instead of S and C , and the projection of this fit in Δt is shown in Figure 5.50. The fitted value for the lifetime is compatible with the value used to generate the MC sample, namely 1.638 ps.

As with the signal mode, the 3D fit is performed simultaneously in the two Dalitz plane halves, to extract S^+ and S^- . The fit projections in each of the two Dalitz regions are shown in Figures 5.51–5.54, while the values of the fitted

Table 5.35: Resulting parameters of the Dalitz-integrated fit to a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays. The fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).

Parameter	Fitted value
C	-0.01 ± 0.05
S	0.04 ± 0.06
κ_{SCF}	0.80 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	0.0005 ± 0.0015
$f_{M_{\text{bc}}}^\mu$ [GeV/ c^2]	0.00008 ± 0.00010
$f_{\Delta E}^\sigma$	1.01 ± 0.04
$f_{M_{\text{bc}}}^\sigma$	1.01 ± 0.03
m_0 [GeV/ c^2]	5.28830 ± 0.00005
$f_{\Delta t, \text{main}}^\sigma$	1.06 ± 0.01
$f_{\Delta t, \text{tail}}^\sigma$	0.94 ± 0.04
$N_{\text{sig-like}}$	2990 ± 80
$N_{B\bar{B} \text{ bkg}}$	2090 ± 115
$N_{q\bar{q} \text{ bkg}}$	13430 ± 150

parameters are listed in Table 5.36. The fitted values of S^+ and S^- are both compatible with the expectation and the fitted yields are compatible across the two Dalitz plane halves.

Table 5.36: Resulting parameters of the Dalitz-split fit to a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays. Once again, the fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).

Parameter	Fitted value
C	-0.01 ± 0.05
S^+	0.07 ± 0.13
S^-	0.06 ± 0.13
κ_{SCF}	0.81 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	0.0010 ± 0.0015
$f_{M_{\text{bc}}}^\mu$ [GeV/ c^2]	0.0001 ± 0.0001
$f_{\Delta E}^\sigma$	1.00 ± 0.04
$f_{M_{\text{bc}}}^\sigma$	0.99 ± 0.03
m_0 [GeV/ c^2]	5.2885 ± 0.0001
$f_{\Delta t, \text{main}}^\sigma$	1.06 ± 0.01
$f_{\Delta t, \text{tail}}^\sigma$	0.94 ± 0.04
$N_{\text{sig-like}}^I$	1600 ± 60
$N_{\text{sig-like}}^I$	1485 ± 55
$N_{B\bar{B} \text{ bkg}}^I$	905 ± 80
$N_{B\bar{B} \text{ bkg}}^I$	1090 ± 80
$N_{q\bar{q} \text{ bkg}}^I$	6730 ± 105
$N_{q\bar{q} \text{ bkg}}^I$	6685 ± 105

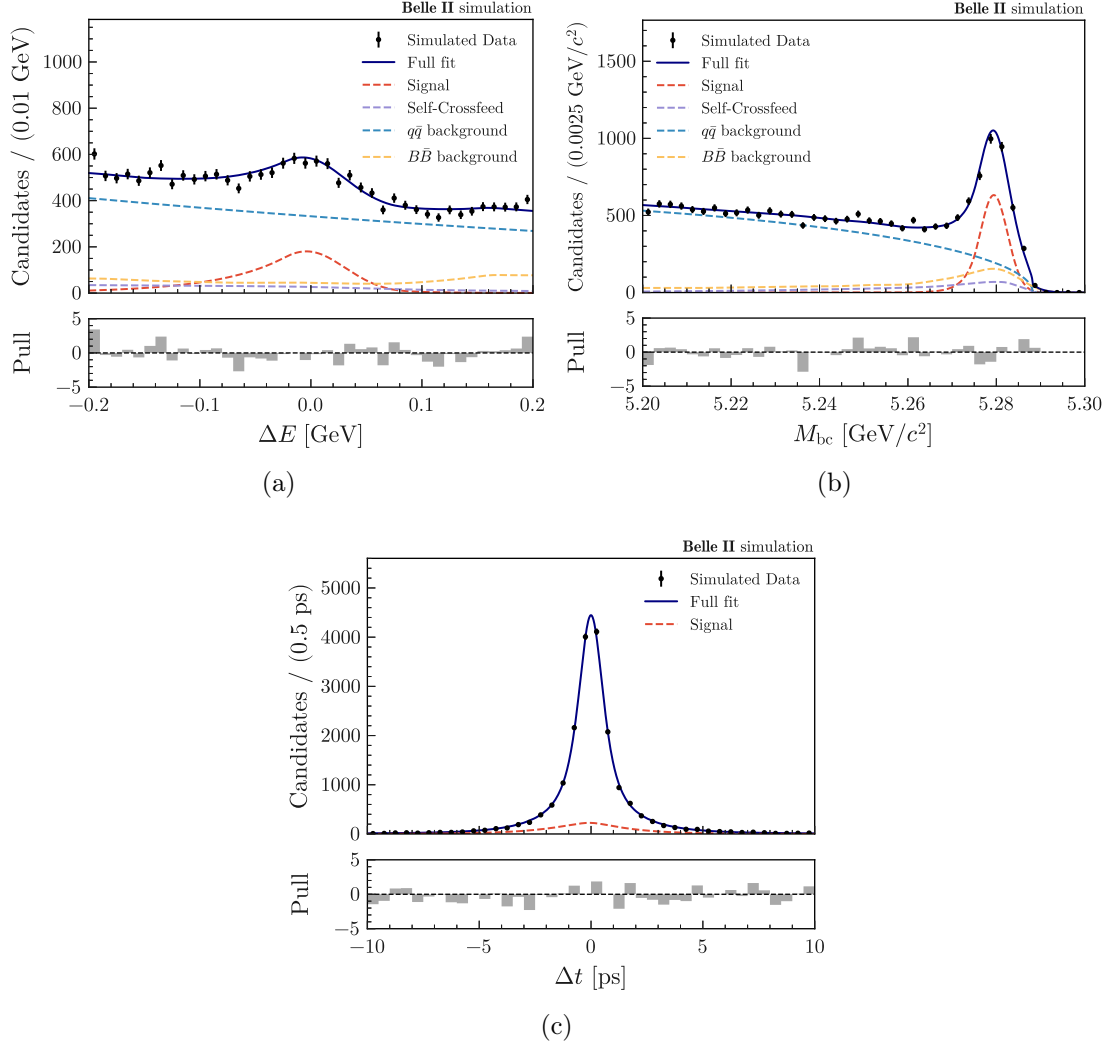


Figure 5.48: Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.

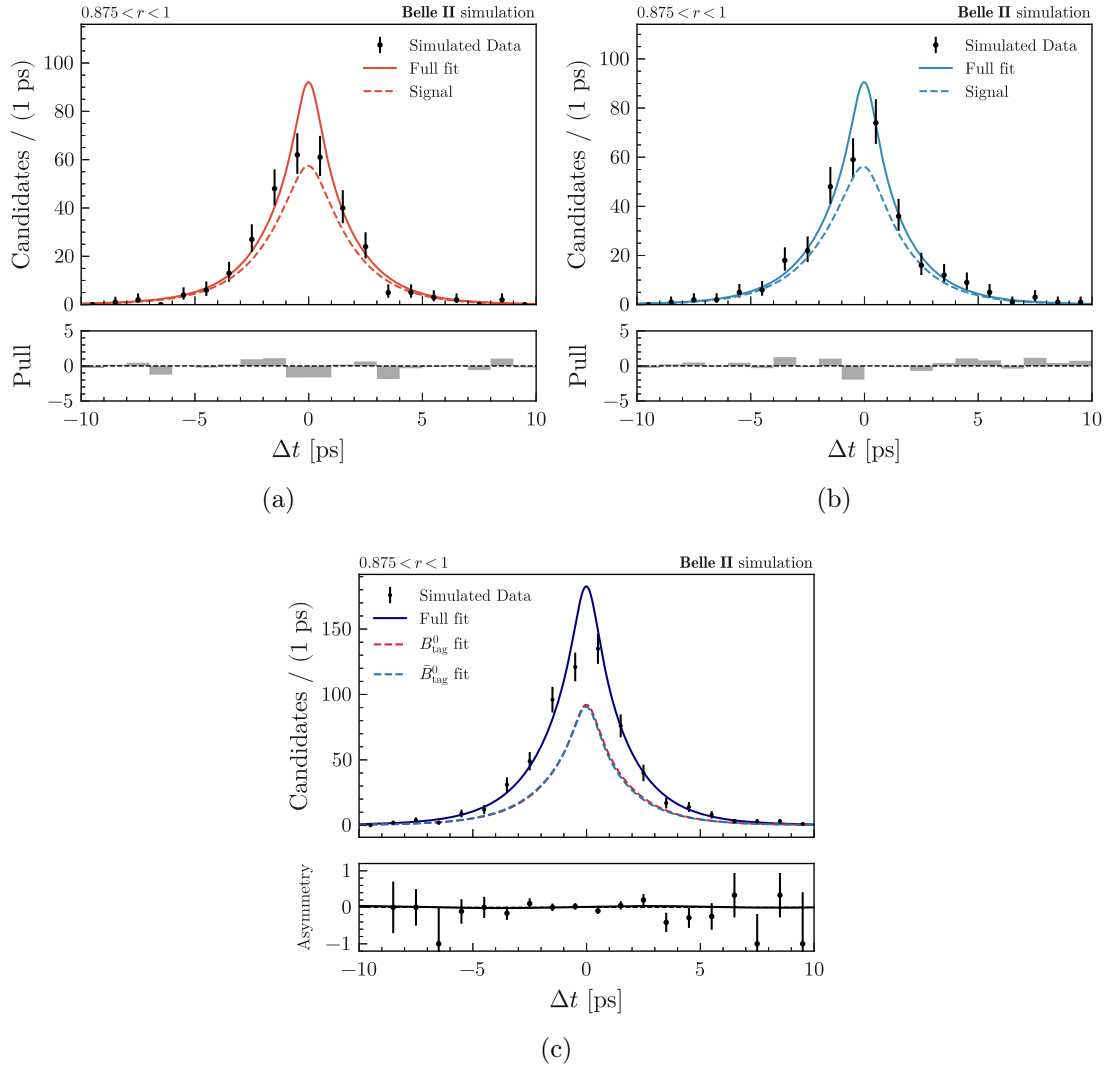


Figure 5.49: Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV, $M_{\text{bc}} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

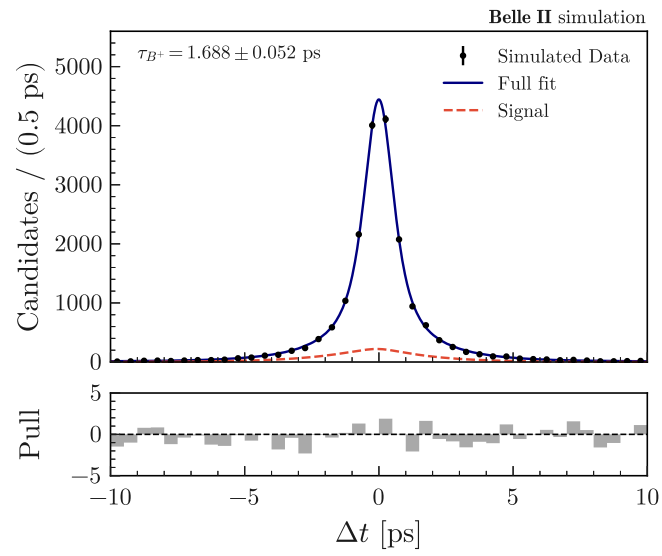


Figure 5.50: Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays to extract the B^+ lifetime projected onto Δt .

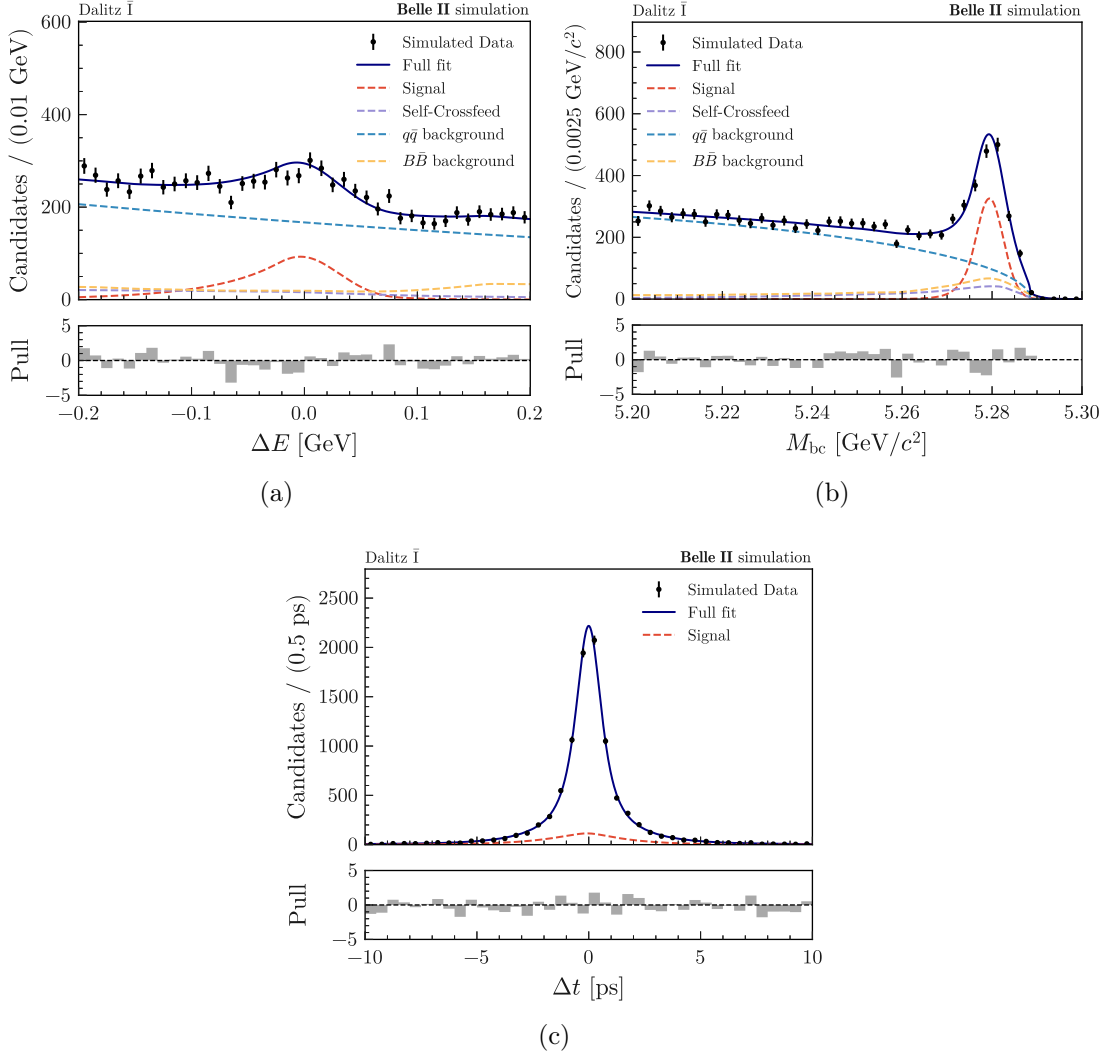


Figure 5.51: Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 > m_{K_S^0 \pi^-}^2$ (Dalitz $\bar{I}$). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.

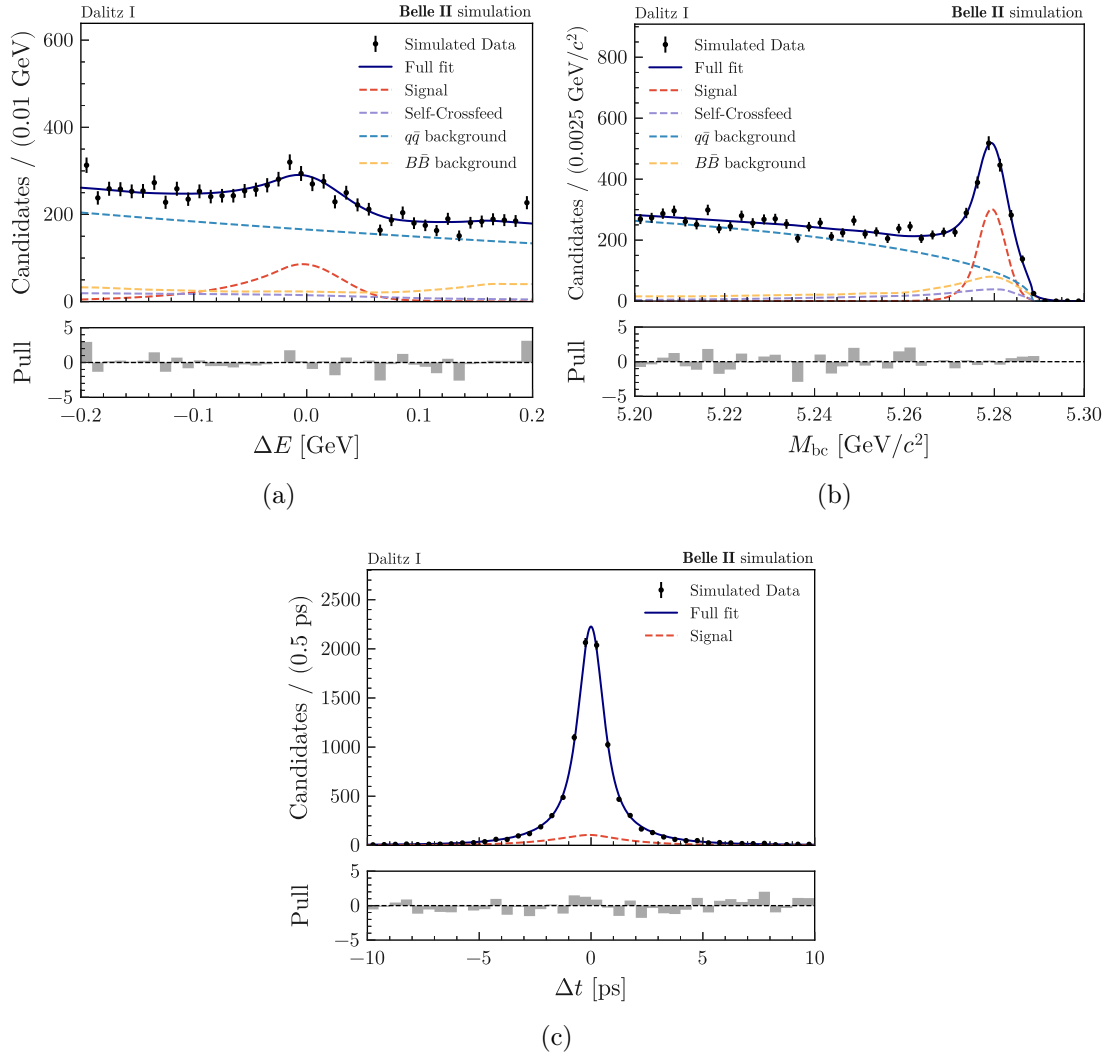


Figure 5.52: Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 < m_{K_S^0 \pi^-}^2$ (Dalitz I). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.

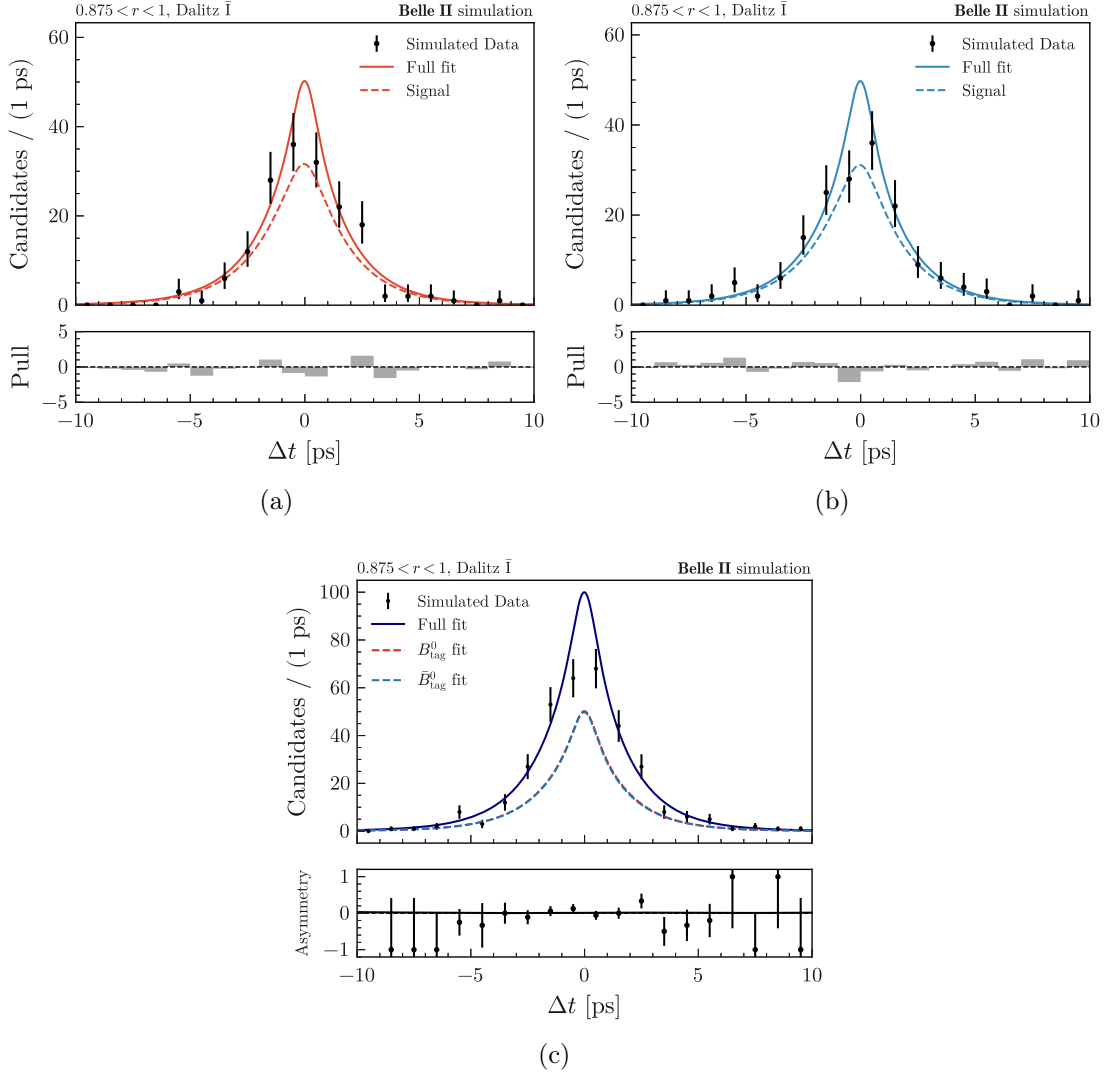


Figure 5.53: Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{I}$ region and in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

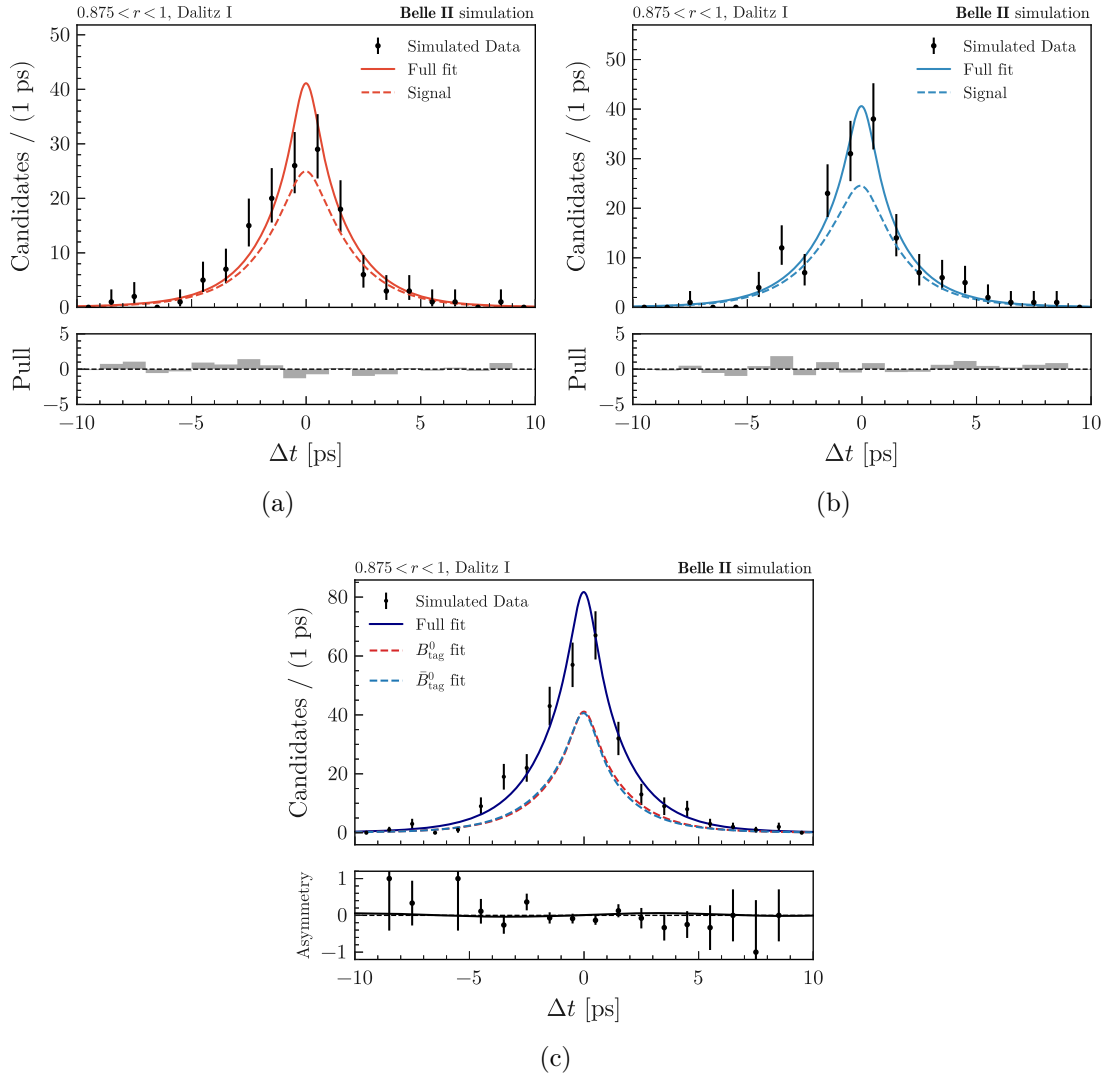


Figure 5.54: Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ decays projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV, $M_{\text{bc}} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

Fit validation with bootstrapped datasets

In a similar manner to the one presented in Section 5.4.1, a set of 1,000 MC cocktail samples are generated via bootstrapping for the purpose of validating the fit to extract the CP parameters. The fit is performed on these MC cocktail samples and the resulting distributions of the fit pull, fit estimate and fit error for S , C , S^+ , S^- and the yields are shown in Figures 5.55 and 5.56.

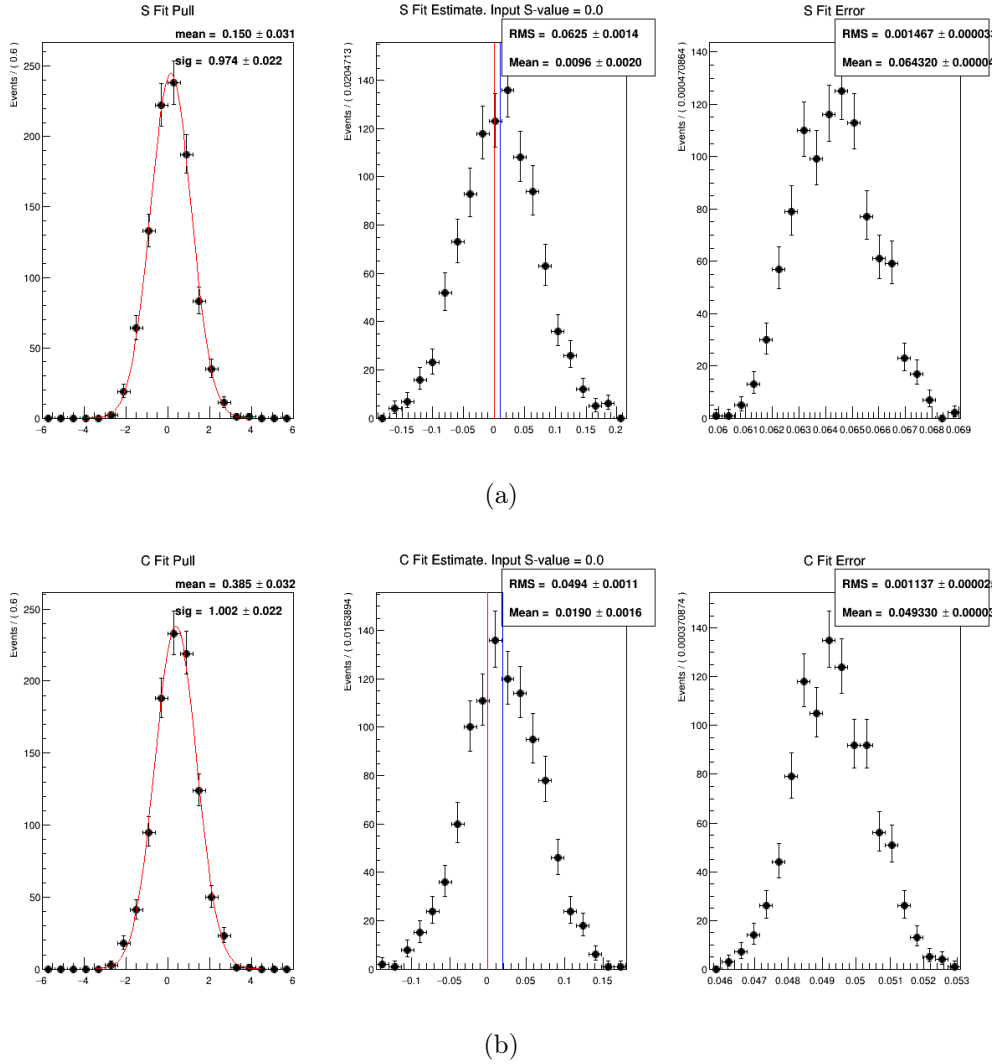


Figure 5.55: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C of the control mode. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

The means of the distributions of the fitted S and C parameters exhibit a small

bias on the level of 20% of the statistical uncertainty, while the signal-like yield is biased on the level of 1.5 times the statistical uncertainty. The results on the CP parameters are coherent with what is observed in the ensemble test on the signal mode, while the bias on the yields is significantly increased when fitting the control mode.

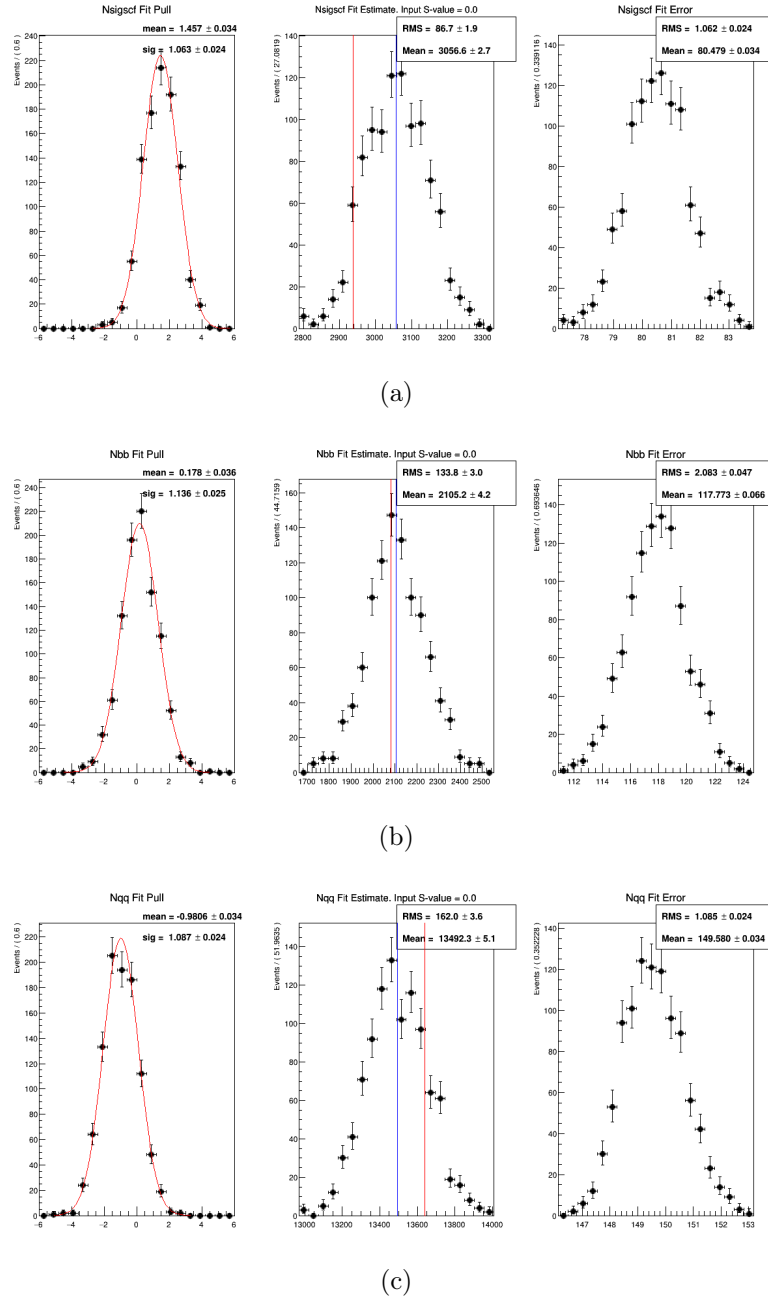


Figure 5.56: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ of the control mode. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

Agreement between data and simulation

The data–MC agreement is also checked using the control mode. Figure 5.57 shows comparisons between data and simulation in the off-resonance samples for the ΔE and Δt distributions. The excess of continuum background observed in the signal mode is also seen here, though it is slightly larger here, on the level of 30%.

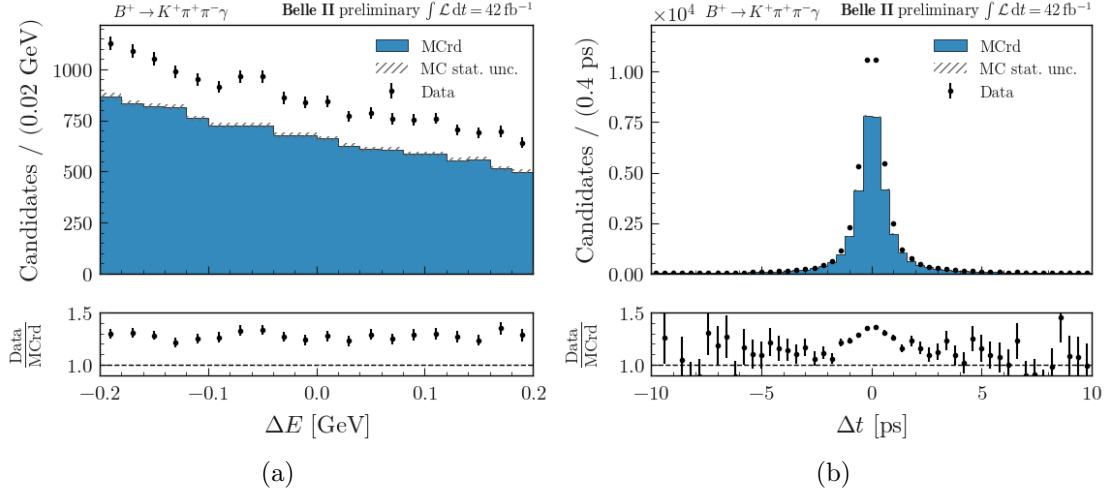


Figure 5.57: Comparison of the (a) ΔE and (b) Δt distributions between the off-resonance data and the off-resonance MC samples for the control mode. The comparison is made before applying the cut on the CSMVA output.

The comparison is also made using the on-resonance samples in Figure 5.58, where the excess of continuum background is confirmed, together with an underestimation of the signal yield, on the level of 30%.

As with the signal mode, the impact on the sensitivity of S and C is checked using an alternative set of bootstrapped datasets that contain 30% more continuum background and 30% less signal events. The uncertainties of S and C are found to be $\sim 20\%$ larger from this ensemble test compared to the previously expected values.

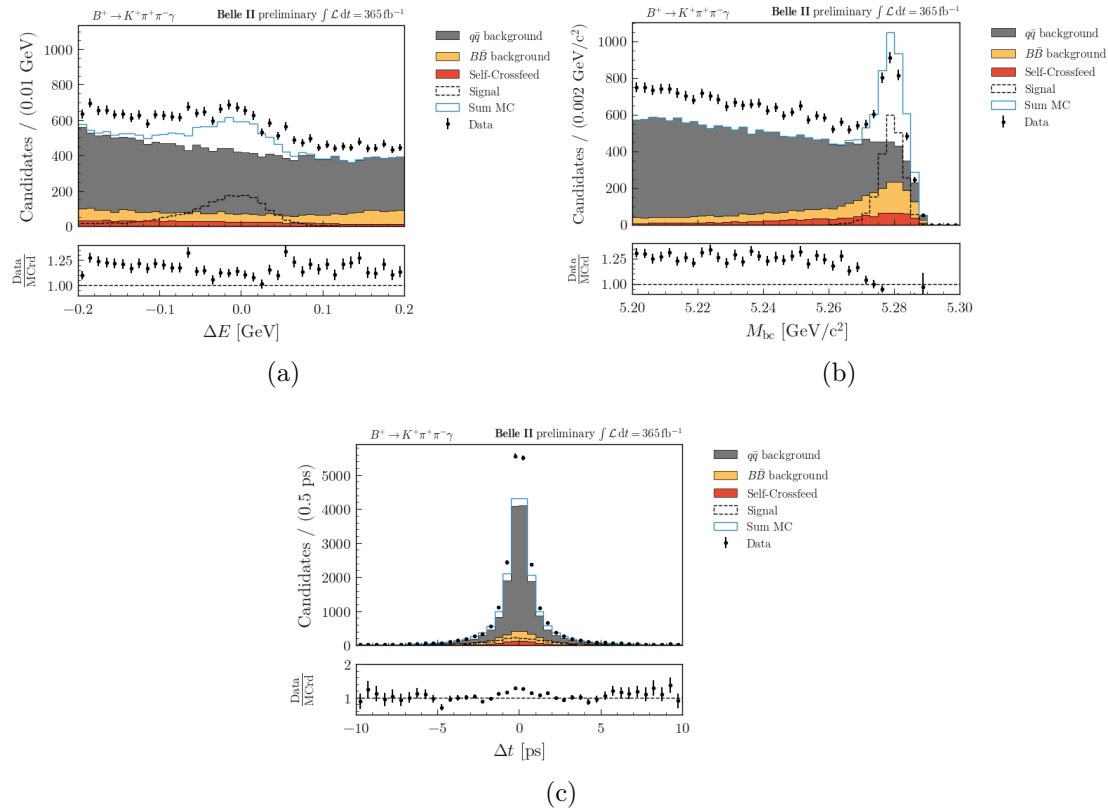


Figure 5.58: Comparison of the (a) ΔE and (b) M_{bc} distributions of between the on-resonance data and MC samples for the control mode. The background components are stacked on top of each other while the signal component is not.

3D fit to the data

Before fitting the data to extract the CP parameters, the 3D fit is performed to extract the B^+ lifetime. Figure 5.59 shows the projection of this fit onto Δt and the measured value for the B^+ lifetime is found to be compatible with the world average value.

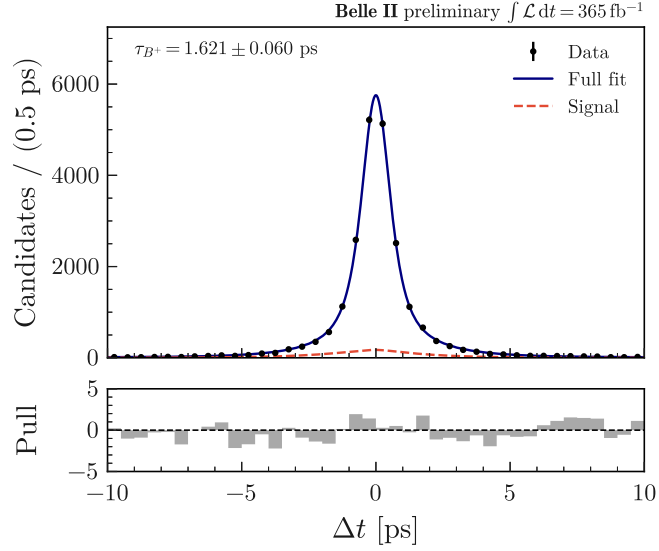


Figure 5.59: Fit of the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample to extract the B^+ lifetime projected onto Δt .

The last validation step involving the control mode is to perform the 3D fit on the Belle II data sample to extract the CP parameters. Figures 5.60 and 5.61 show the projections for the Dalitz-integrated fit and the fitted parameters are listed in Table 5.37.

The value of C is not revealed so as not to introduce potential biases to the ongoing measurement of the time-integrated CP asymmetry in this decay channel. Otherwise, the fitted value for S is found to be compatible with the expectation within the statistical uncertainty.

Finally, the Dalitz-split fit is performed on the Belle II data sample to extract S^+ and S^- in the control mode. Figures 5.62–5.65 show the fit projections and Table 5.38 lists the fitted parameters.

The fitted values of S^+ and S^- are found to be compatible with the expectations within their statistical uncertainties.

Table 5.37: Resulting parameters of the Dalitz-integrated fit to the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample. The central value of the C parameter is blinded to avoid biasing ongoing analyses of the time-integrated CP asymmetry in this mode.

Parameter	Fitted value
C	$— \pm 0.06$
S	0.03 ± 0.08
κ_{SCF}	0.80 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	-0.0046 ± 0.0017
$f_{M_{\text{bc}}}^\mu$ [GeV/ c^2]	-0.00015 ± 0.00011
$f_{\Delta E}^\sigma$	0.98 ± 0.04
$f_{M_{\text{bc}}}^\sigma$	0.89 ± 0.03
m_0 [GeV/ c^2]	5.2885 ± 0.0001
$f_{\Delta t, \text{main}}^\sigma$	0.98 ± 0.01
$f_{\Delta t, \text{tail}}^\sigma$	0.82 ± 0.02
$N_{\text{sig-like}}$	2240 ± 70
$N_{B\bar{B} \text{ bkg}}$	1425 ± 110
$N_{q\bar{q} \text{ bkg}}$	18075 ± 165

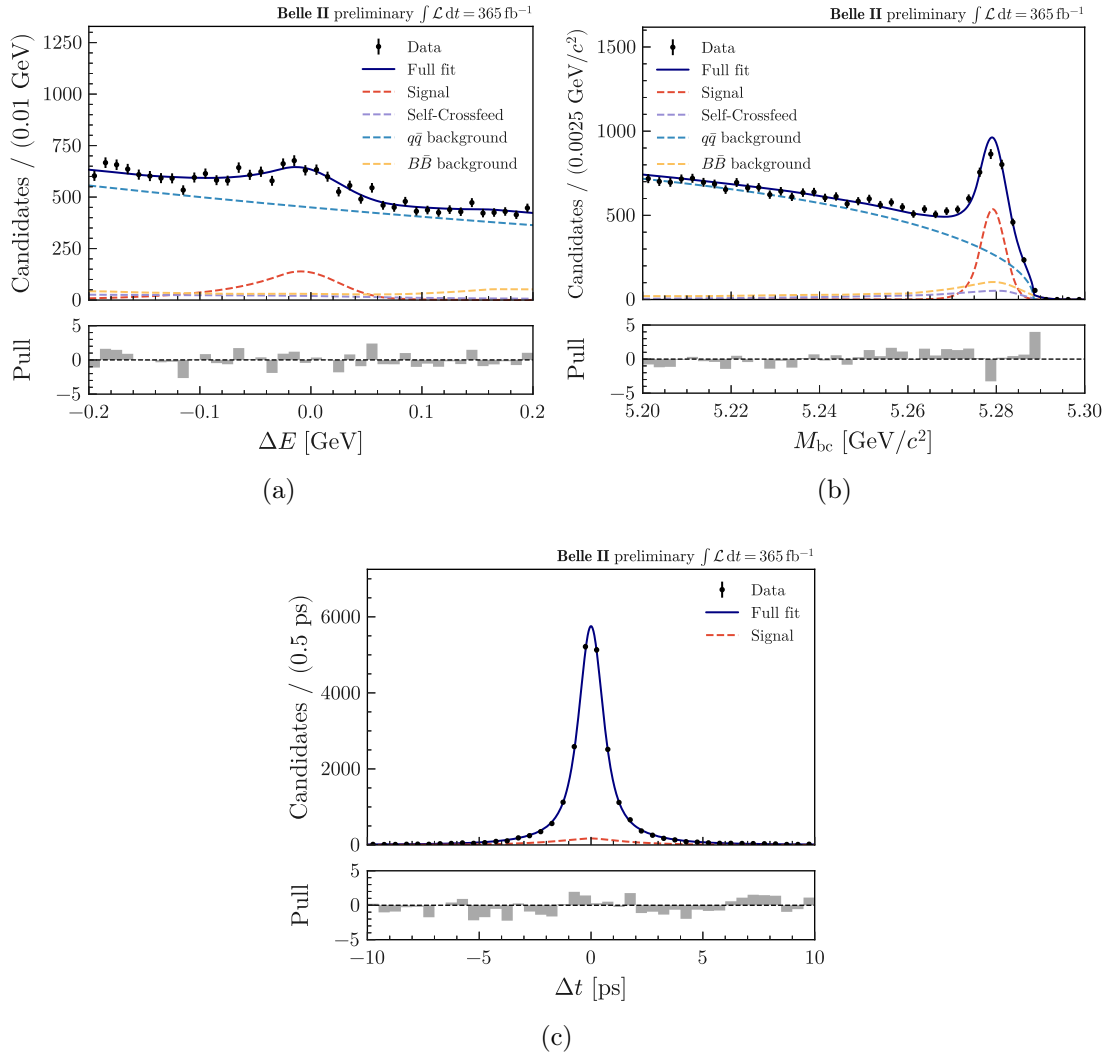


Figure 5.60: Fit of the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.

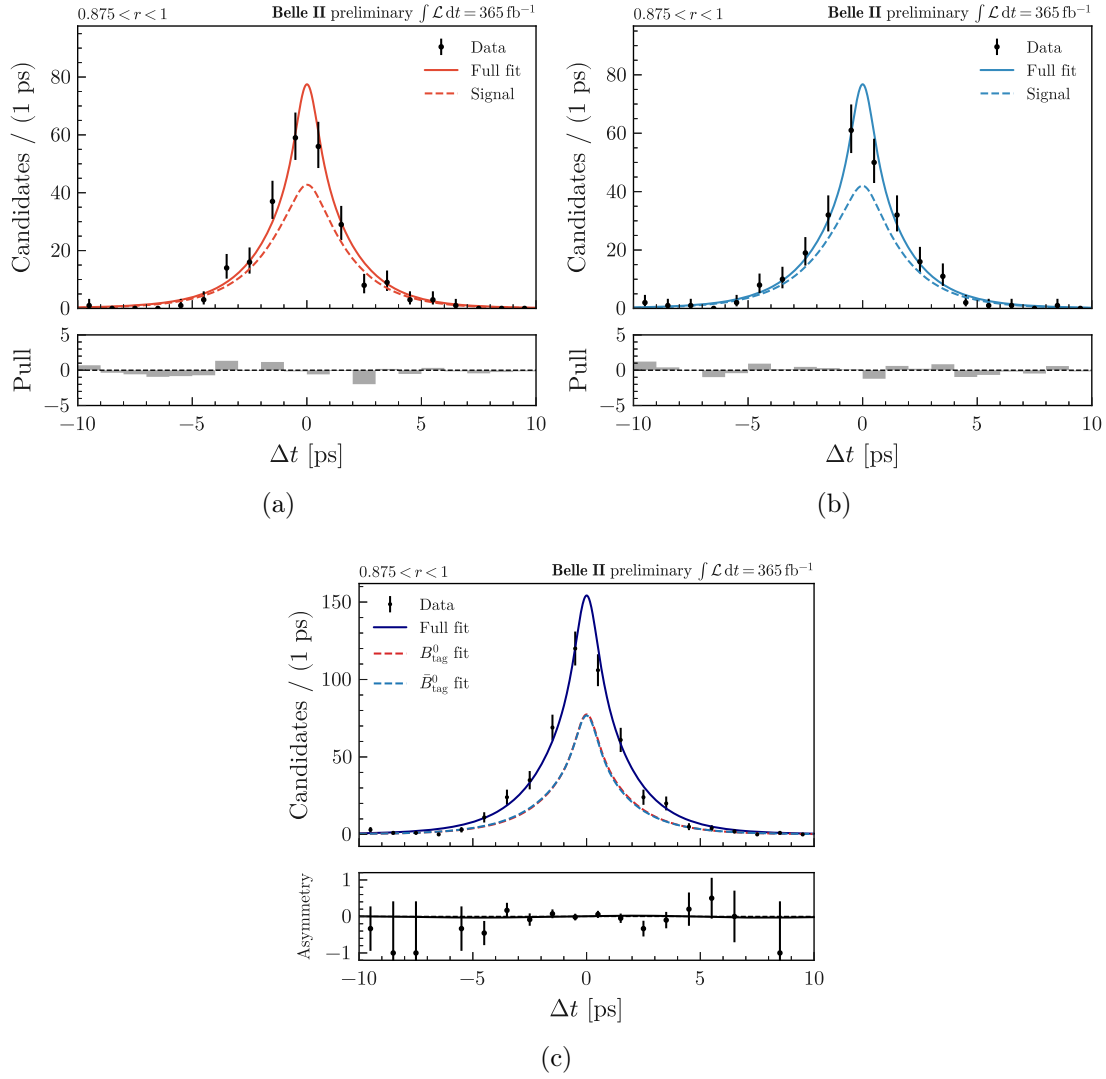


Figure 5.61: Fit of the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$.

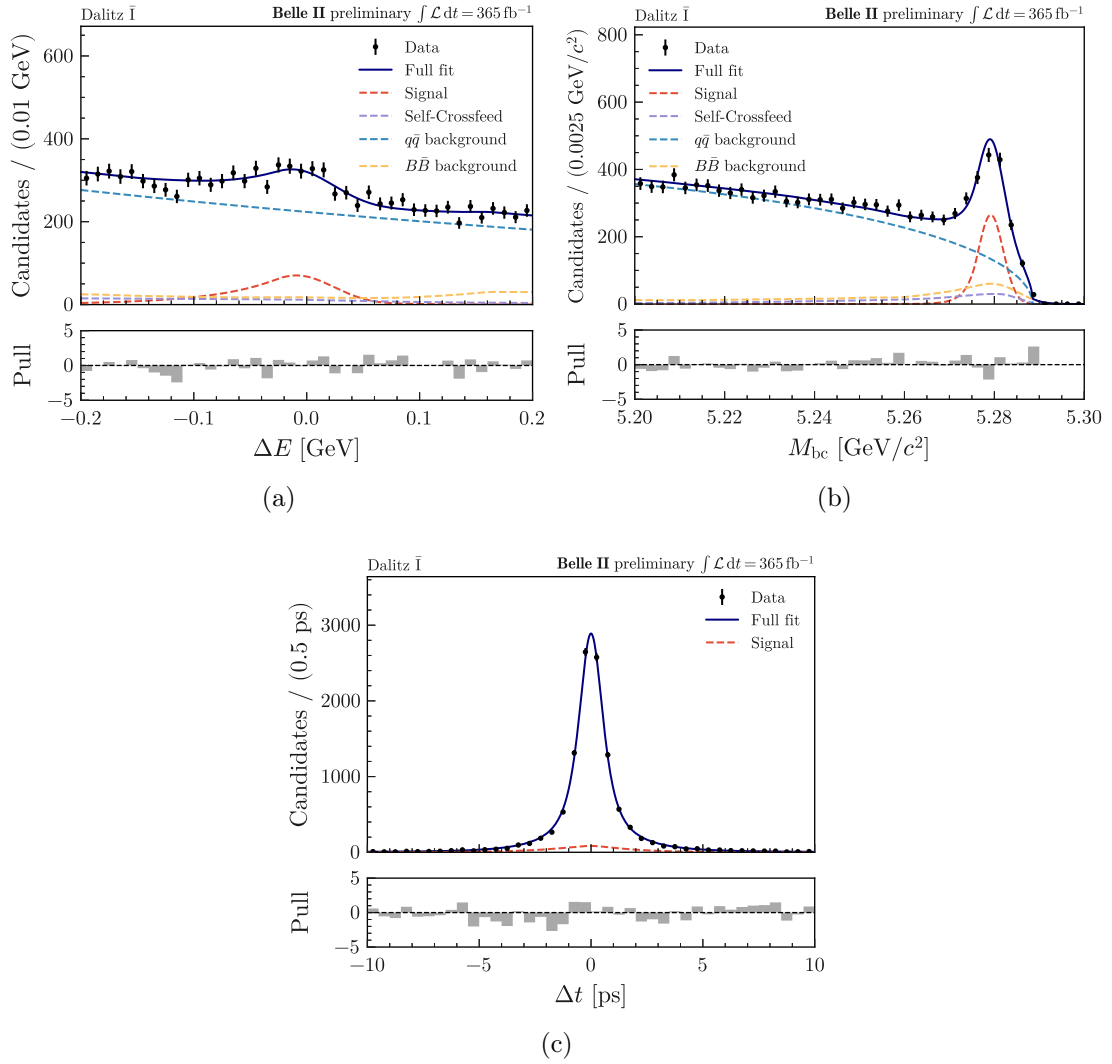


Figure 5.62: Fit of the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+ K_S^0}^2 > m_{\pi^- K_S^0}^2$ (Dalitz $\bar{\text{I}}$). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.

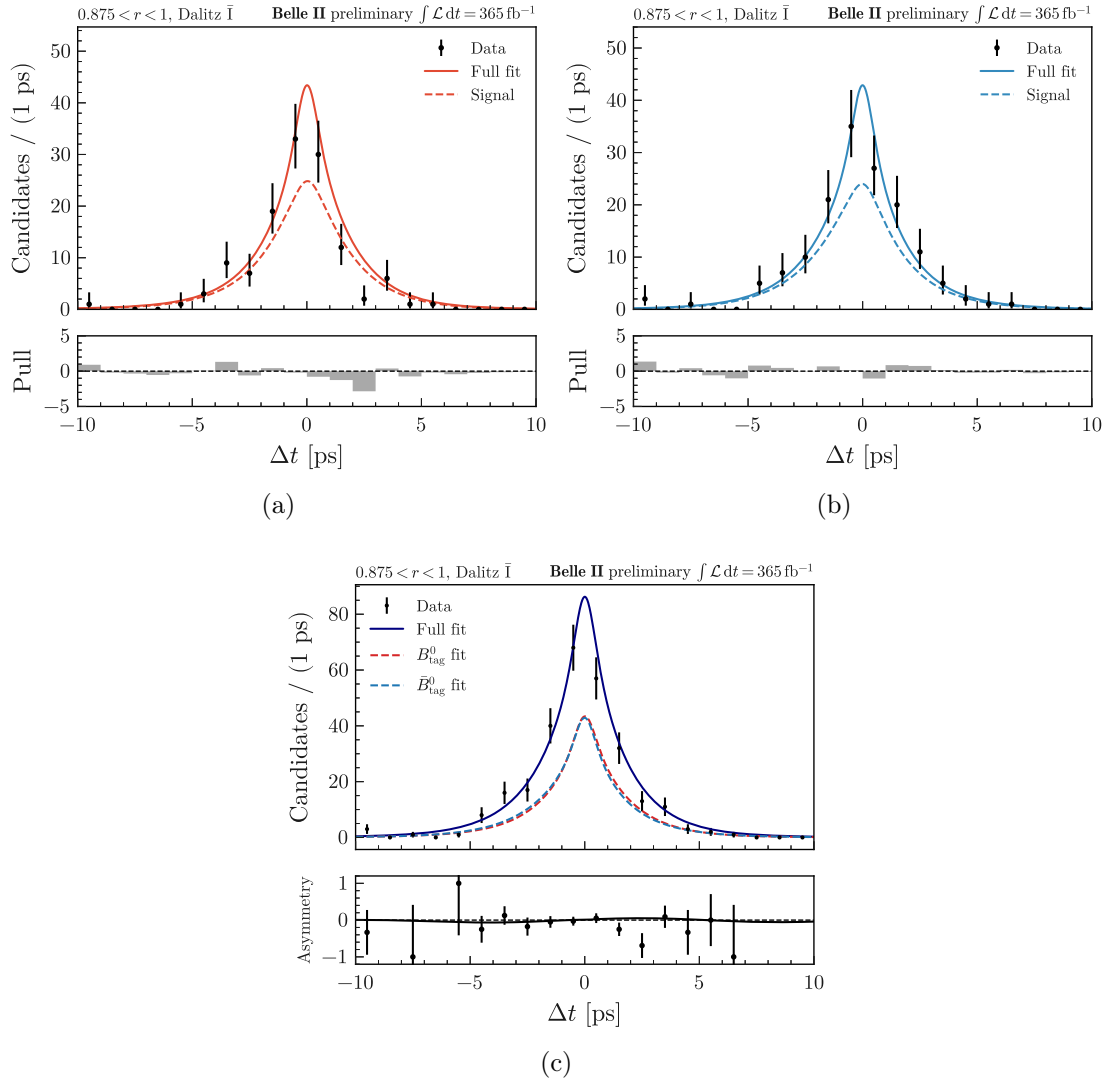


Figure 5.63: Fit of the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{\text{I}}$ region and in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$.

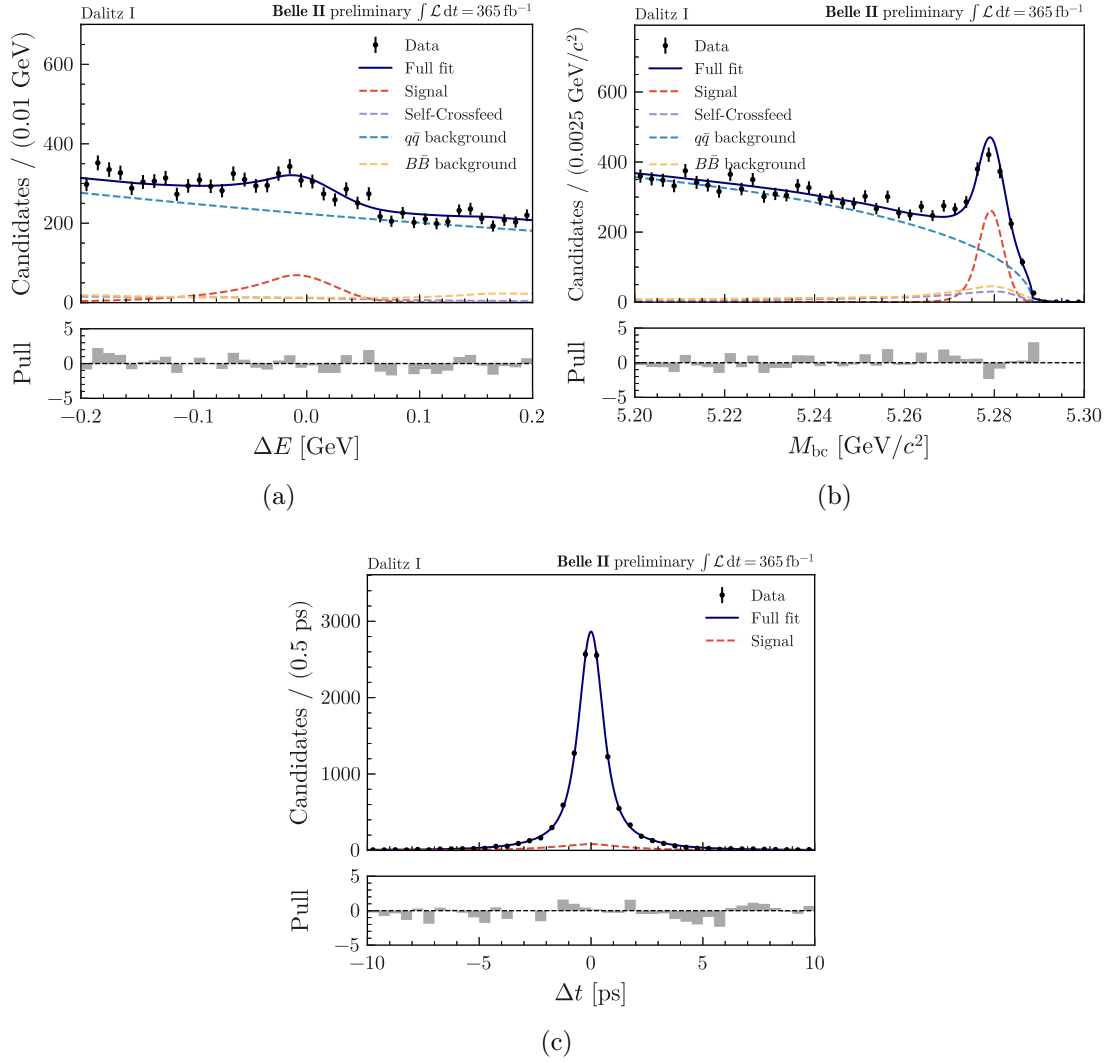


Figure 5.64: Fit of the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+ K_S^0}^2 < m_{\pi^- K_S^0}^2$ (Dalitz I). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.

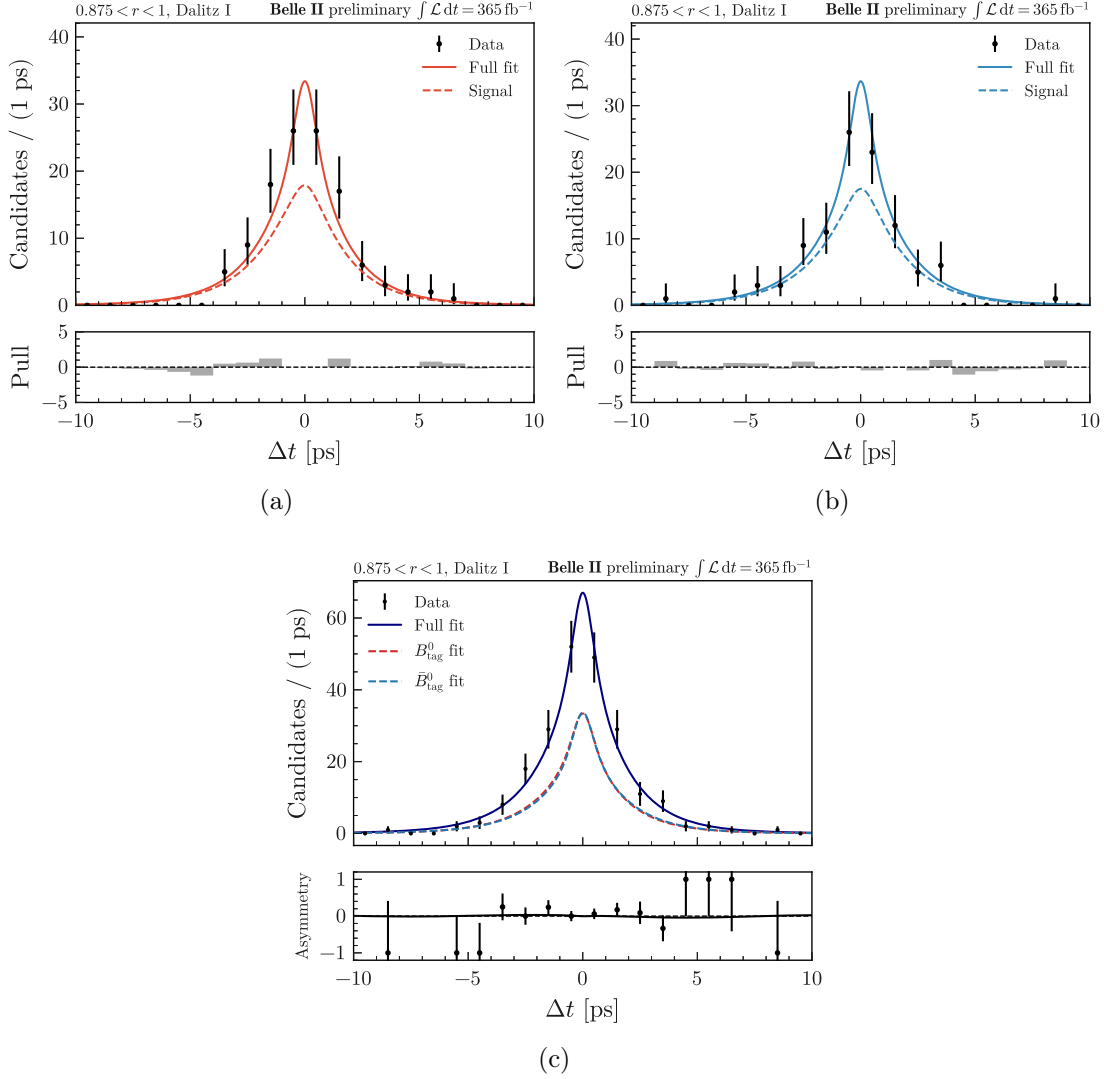


Figure 5.65: Fit of the $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{bc} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$.

Table 5.38: Resulting parameters of the Dalitz-split fit to the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample. The central value of the C parameter is blinded to avoid biasing ongoing analyses of the time-integrated CP asymmetry in this mode.

Parameter	Fitted value
C	$— \pm 0.06$
S^+	0.06 ± 0.15
S^-	-0.03 ± 0.15
κ_{SCF}	0.80 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	-0.0043 ± 0.0017
$f_{M_{\text{bc}}}^\mu$ [GeV/ c^2]	-0.00015 ± 0.00011
$f_{\Delta E}^\sigma$	0.97 ± 0.04
$f_{M_{\text{bc}}}^\sigma$	0.88 ± 0.03
m_0 [GeV/ c^2]	5.2885 ± 0.0001
$f_{\Delta t, \text{main}}^\sigma$	0.98 ± 0.01
$f_{\Delta t, \text{tail}}^\sigma$	0.82 ± 0.02
$N_{\text{sig-like}}^I$	1190 ± 50
$N_{\text{sig-like}}^{\bar{I}}$	1145 ± 50
$N_{B\bar{B} \text{ bkg}}^I$	735 ± 75
$N_{B\bar{B} \text{ bkg}}^{\bar{I}}$	645 ± 75
$N_{q\bar{q} \text{ bkg}}^I$	8745 ± 115
$N_{q\bar{q} \text{ bkg}}^{\bar{I}}$	9290 ± 115

5.5 Evaluation of systematic uncertainties

The evaluation of the various sources of systematic uncertainty considered for C , S , S^+ and S^- is described in detail in this section. Where possible, the systematic uncertainties are estimated in a data-driven way using the Belle II dataset. Otherwise, the simulated MC samples are used.

5.5.1 Component fit shape parameters

Several ΔE , M_{bc} , and Δt shape parameters are fixed in the fit to data. The associated systematic uncertainty is estimated by repeating the fit 1,000 times on the Belle II dataset, each time using a different set of fixed parameter values. These sets are generated by randomly varying the parameters around their central values according to the covariance matrices obtained in the corresponding shape fits (see Sections 5.3.2 and 5.3.3). For the $B\bar{B}$ background in ΔE and M_{bc} , a different KDE shape is built for each trial using events sampled with replacement from the generic MC. The systematic uncertainty for a given CP parameter is then quoted as the width of the Gaussian used to fit its distribution across the 1,000 fits. For asymmetric distributions, a bifurcated Gaussian is used to fit them and the largest of the two widths is quoted. Figure 5.66 shows the resulting distributions of the residuals of C , S , S^+ and S^- . Residual here refers to the signed difference between the parameter values obtained from the 1,000 fits and the parameter value obtained in the nominal case, where the fixed shape parameters have not been varied using the covariance matrix. The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the fixed shape parameters evaluated in this way are $\{0.003, 0.005, 0.005, 0.004\}$.

5.5.2 Flavor Tagger parameters

The parameters of the Flavor Tagger are also fixed in the fit to data. To evaluate the systematic uncertainty arising from this, the fit is repeated 1,000 times on the Belle II dataset, each time with a different set of values for the Flavor Tagger parameters. Each set of parameters is obtained by randomly varying them around their central values using the covariance matrix obtained from the calibration procedure described in Chapter 4. The systematic uncertainty for a given CP parameter is then quoted as the width of the Gaussian used to fit its distribution across the 1,000 fits. Figure 5.67 shows the resulting distributions of the residuals of C , S , S^+ and S^- . The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the Flavor Tagger parameters evaluated in this way are $\{0.018, 0.007, 0.014, 0.012\}$.

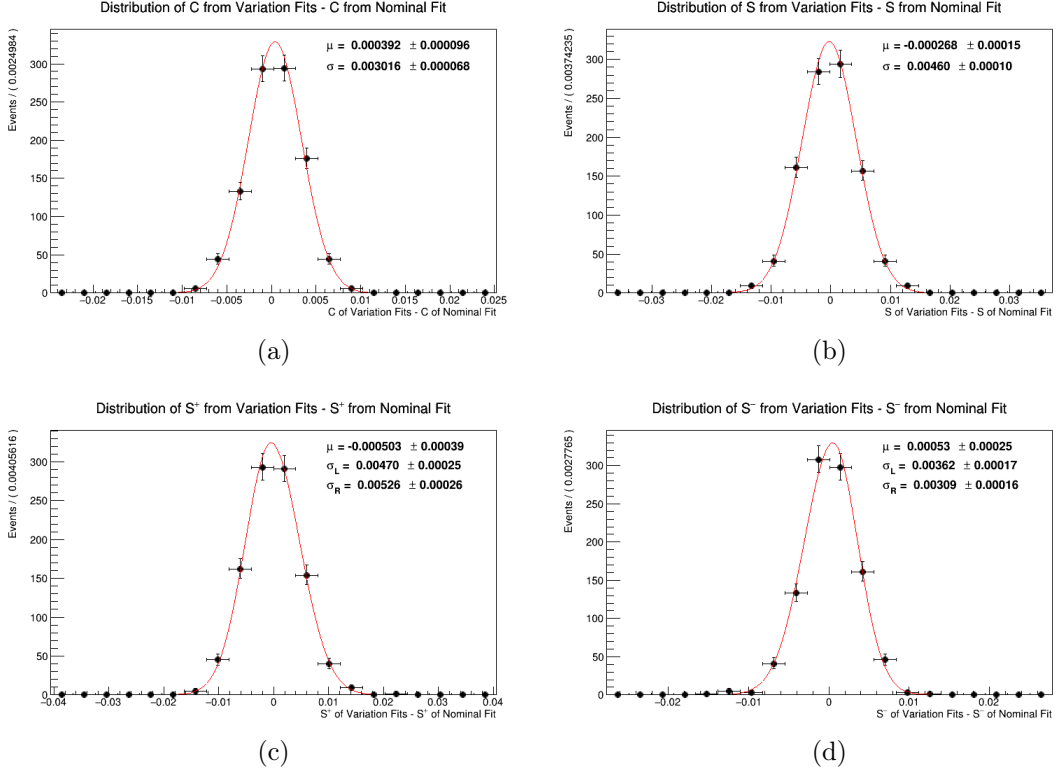


Figure 5.66: Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of fixed shape parameters.

5.5.3 Δt resolution function parameters

The parameters of the Δt Resolution Function are also fixed in the fit to data. The systematic uncertainty arising from this is evaluated by repeating the fit is repeated 1,000 times on the Belle II dataset, each time with a different set of values for the Δt Resolution Function parameters. Each set of parameters is obtained by randomly varying them around their central values using the covariance matrix obtained from the calibration procedure described in Chapter 4. The systematic uncertainty for a given CP parameter is then quoted as the width of the Gaussian used to fit its distribution across the 1,000 fits. For asymmetric distributions, a bifurcated Gaussian is used to fit them and the largest of the two widths is quoted. Figure 5.68 shows the resulting distributions of the residuals of C , S , S^+ and S^- . The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the Δt Resolution Function parameters evaluated in this way are $\{0.005, 0.014, 0.023, 0.018\}$.

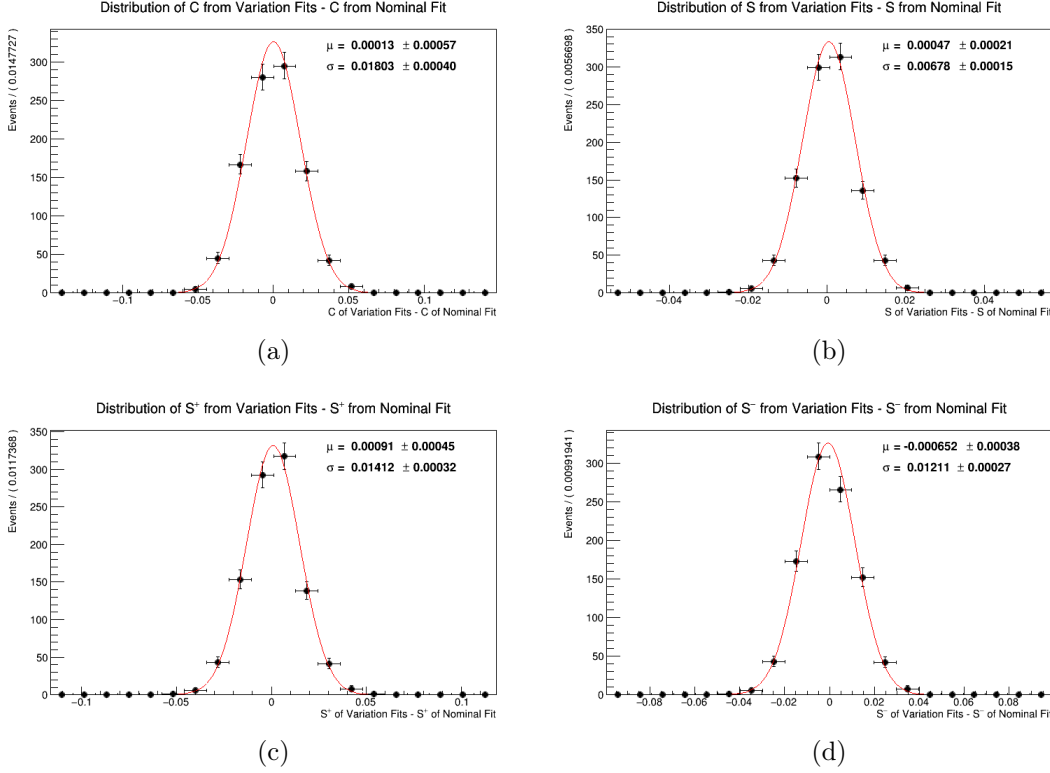


Figure 5.67: Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of Flavor Tagger parameters.

5.5.4 B^0 lifetime and $B^0-\bar{B}^0$ oscillation frequency

The values of the B^0 lifetime, τ_{B^0} , and $B^0-\bar{B}^0$ oscillation frequency, Δm_d , are fixed to their world average values in the fit to data, namely $\tau_{B^0} = 1.517$ ps and $\Delta m_d = 0.5069$ ps⁻¹. In order to evaluate the systematic uncertainty arising from this, the fit is repeated 1,000 times on the Belle II dataset, each time with a different set of values for τ_{B^0} and Δm_d . Each set of parameters is obtained by randomly varying them around their central values and within their PDG uncertainties, namely $\sigma_{\tau_{B^0}} = 0.004$ ps and $\sigma_{\Delta m_d} = 0.0019$ ps⁻¹. The systematic uncertainty for a given CP parameter is then quoted as the width of the Gaussian used to fit its distribution across the 1,000 fits. Figure 5.69 shows the resulting distributions of the residuals of C , S , S^+ and S^- . The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the B^0 lifetime and $B^0-\bar{B}^0$ oscillation frequency parameters evaluated in this way are $\{0.0005, 0.0003, 0.0010, 0.0025\}$.

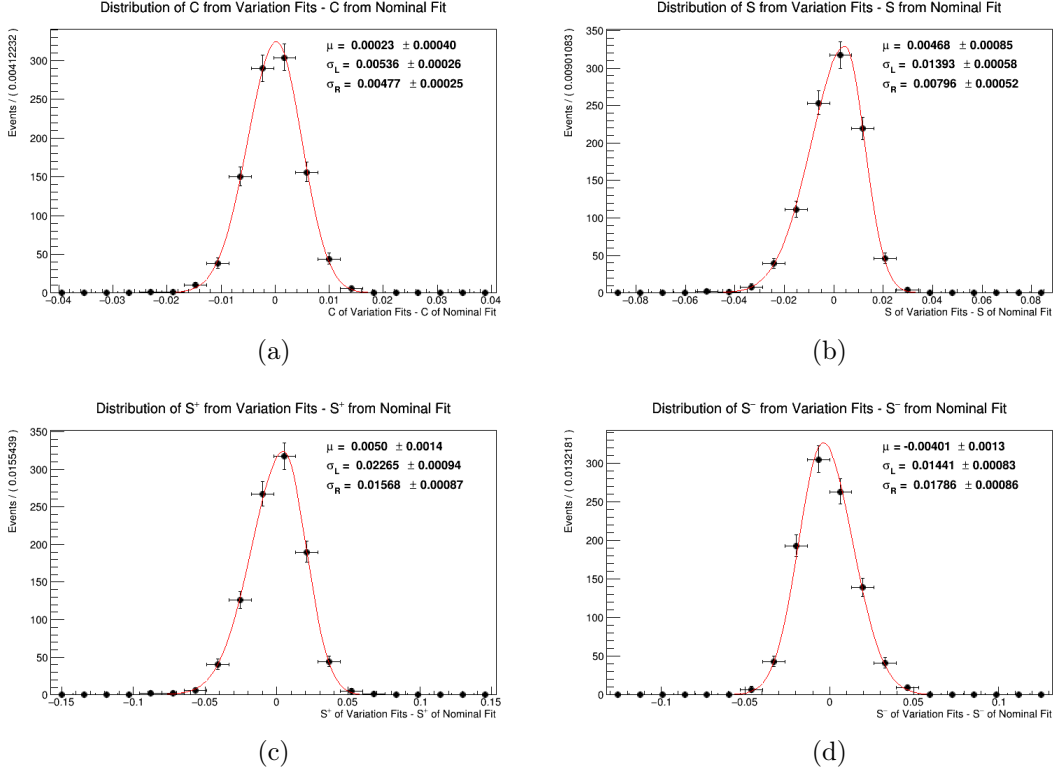


Figure 5.68: Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of Δt Resolution Function parameters.

5.5.5 Fixed SCF background fractions

As mentioned in Section 5.3.1, the fractions of SCF background, namely f_{0-5}^{SCF} and f_6^{SCF} are fixed in the fit to data to values obtained from simulation. To evaluate the systematic uncertainty arising from this, the fit is repeated twice on the Belle II dataset with the SCF fractions varied by either $\pm 20\%$. For each CP parameter, the systematic uncertainty is determined by first computing the absolute difference between the nominal fitted value and the values obtained with the SCF fractions shifted by $+20\%$ and -20% . The larger of these two differences is taken for each fraction, and the final uncertainty is given by the square root of the quadratic sum of these two contributions. Figure 5.70 show the nominal fitted values of C , S , S^+ and S^- together with the fitted values from all the variations. The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the fixed SCF background fractions evaluated in this way are $\{0.006, 0.004, 0.008, 0.011\}$.

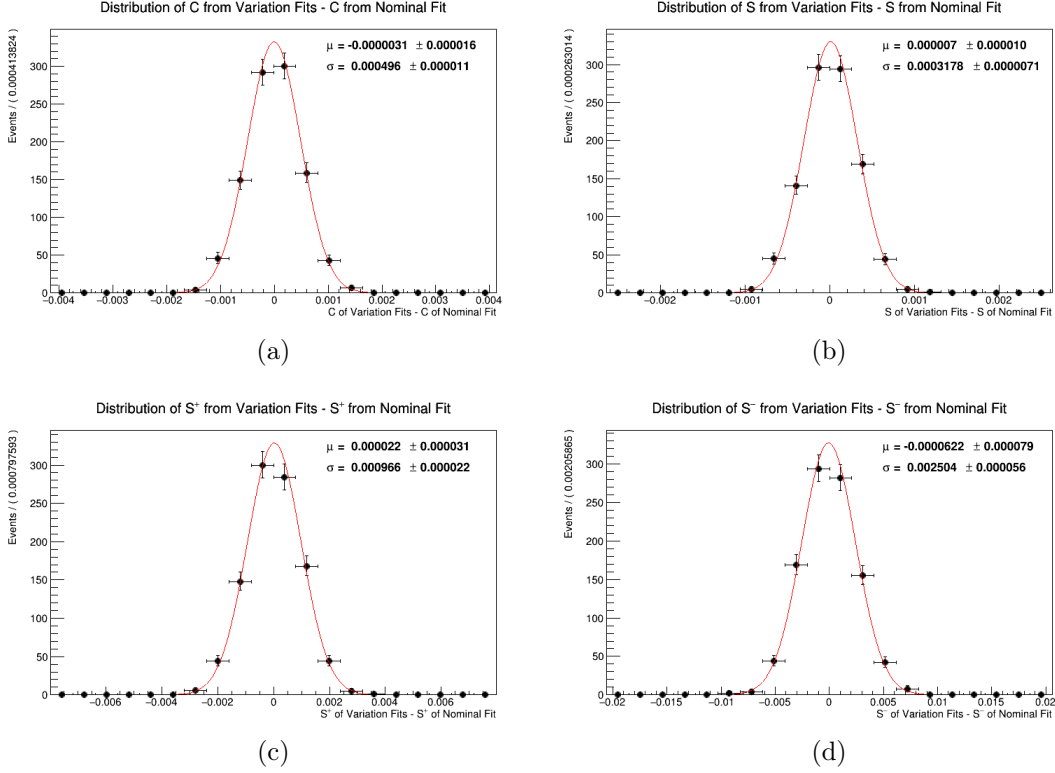


Figure 5.69: Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of values for the B^0 lifetime and B^0 - $\bar{B}^0$ oscillation frequency.

5.5.6 CP violation in the $B\bar{B}$ background

The CP parameters of the $B\bar{B}$ background component are both fixed to zero in the fit to data. To estimate the systematic uncertainty arising from not accounting for potential CP violation in the $B\bar{B}$ background, the fit is repeated on the Belle II dataset with either $S_{B\bar{B}}$ or $C_{B\bar{B}}$ varied by $\pm 10\%$ around zero, i.e. set to ± 0.1 . The largest absolute difference between the alternatively fitted CP parameter and the nominal fitted value is determined when varying each of the CP parameters of the $B\bar{B}$ background separately. The systematic uncertainty is then taken as the square root of the quadratic sum of those differences. Figure 5.71 shows the nominal fitted values of C , S , S^+ and S^- together with the fitted values from all the variations. The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to not accounting for possible CP violation in the $B\bar{B}$ background component evaluated in this way are $\{0.019, 0.017, 0.034, 0.001\}$.

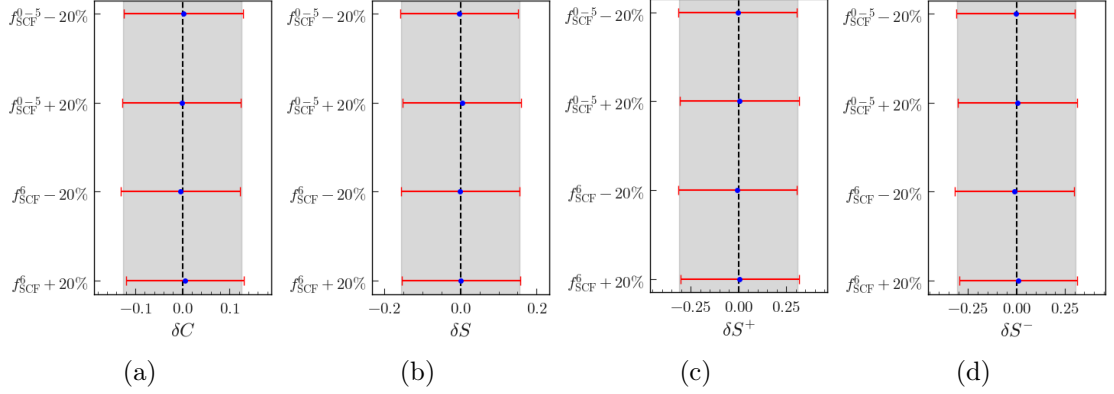


Figure 5.70: Fitted values of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset with the fixed SCF background fractions varied by $\pm 20\%$. The value obtained with the fractions fixed to their nominal value is denoted by the vertical dashed line.

5.5.7 Effect of biased yields on CP parameters

As can be seen in Figure 5.29, there is a significant bias on the fitted yields. In order to quantify the impact of this yield bias on the determination of the CP parameters, the fit is repeated on all of the bootstrapped datasets with the yields fixed to the values that were used to generate the samples and leaving all other free parameters floating. The maximum absolute difference across the seven input S values that is observed between the CP parameters obtained in this way and the ones obtained with the yields left free to float is then quoted as the systematic uncertainty due to the biased yields. Figure 5.72 shows the linearity of the fitted CP parameters in the case of both fixed and free yields. The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the yield bias evaluated in this way are $\{0.007, 0.004, 0.009, 0.018\}$.

5.5.8 Bootstrapping bias

Figure 5.30 shows the behavior of the fitted CP parameters for seven different input values of S , using the bootstrapped datasets. To account for possible biases in the fit procedure and correlations between these datasets, a dedicated systematic uncertainty is assigned to each CP parameter. This uncertainty is defined as the maximum absolute deviation between the fitted value (blue dots) and the expected value (dashed line) across the seven input S values⁷. The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the bootstrapping bias evaluated in this

⁷The deviations are listed in Table 5.19 for all input S values.

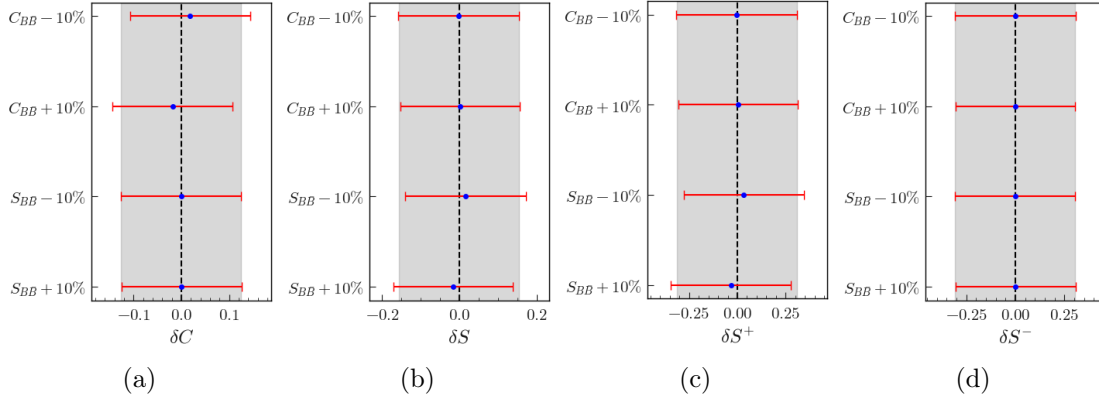


Figure 5.71: Fitted values of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset with $S_{B\bar{B}}$ and $C_{B\bar{B}}$ varied by $\pm 10\%$. The value obtained with $S_{B\bar{B}}$ and $C_{B\bar{B}}$ both fixed to zero is denoted by the vertical dashed line.

way are $\{0.027, 0.054, 0.110, 0.033\}$. This is the dominant contribution for S , S^+ , and S^- , although it was computed conservatively and only amounts to $\sim 30\%$ of the expected statistical uncertainty for these parameters.

5.5.9 Misalignment

To estimate the systematic uncertainty arising from the potential misalignment of the detector, separate signal MC samples are produced for 4 different misalignment scenarios and the 3D fit is subsequently performed on them to extract the CP parameters. The 4 misalignment scenarios are:

- **misA1_1:** Model of systematic misalignment observed in older data with misaligned CDC wires
- **misA1_2:** Residual systematics from simulation studies on misalignment with misaligned CDC wires
- **misA1_3:** Day-to-day alignment difference derived from data without misaligned CDC wires
- **misA1_4:** Residual systematics from simulation studies on misalignment without misaligned CDC wires

The resulting CP parameters are compared to the ones obtained from fitting the nominal (not misaligned) signal MC sample and the systematic uncertainty for each CP parameter is quoted as the maximum deviation observed from the

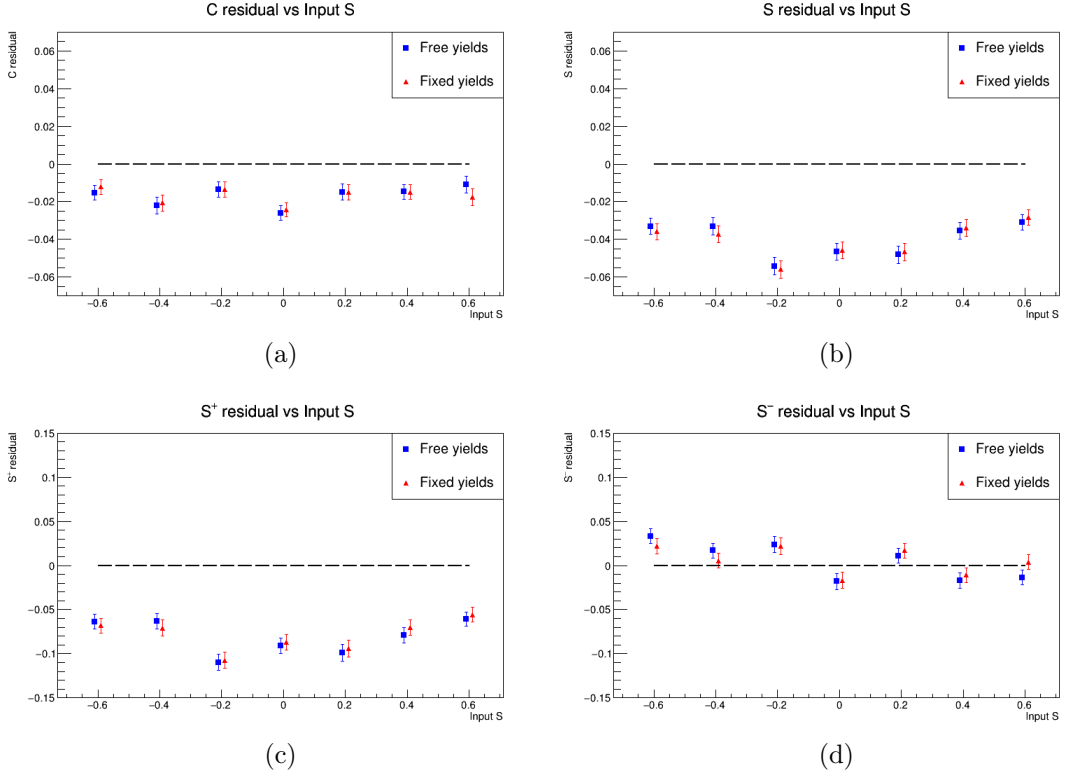


Figure 5.72: Linearity figures for the residuals of (a) C , (b) S , (c) S^+ and (d) S^- . The fitted CP parameters obtained after fixing the yields to their true values are also displayed.

nominal value. Figure 5.73 shows this comparison between the CP parameters for each misalignment scenario and the nominal case. The systematic uncertainties on $\{C, S, S^+, S^-\}$ due to the potential misalignment of the detector are $\{0.005, 0.003, 0.006, 0.012\}$.

5.5.10 Tag-side interference

The CP parameters extracted from the fit are not considering any effects due to tag-side interference. Tag-side interference refers to the interference between CKM-favored $b \rightarrow c\bar{u}d$ and doubly-CKM-suppressed $\bar{b} \rightarrow \bar{u}c\bar{d}$ amplitudes that contribute to the final state of the B_{tag} , which, if not accounted for, can lead to wrong flavor tagging and thus alter the measurement of the time-dependent CP asymmetry. To account for this effect, a dedicated systematic uncertainty is assigned following the suggestions in Ref. [116]. The SM expectation for the direct and mixing-induced CP asymmetry for this mode is close to zero, therefore $|\lambda_{CP}| = 1$ and

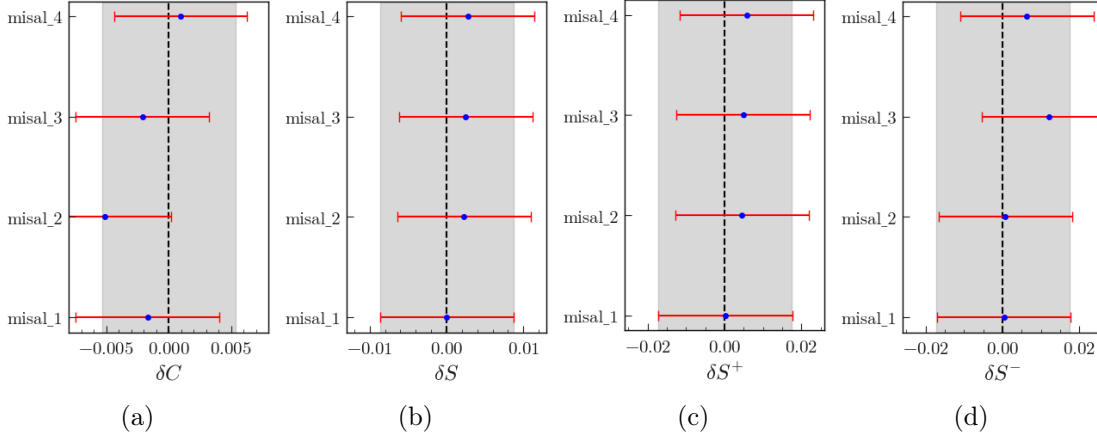


Figure 5.73: Fitted values of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting signal MC samples corresponding to 4 different misalignment scenarios. The value obtained in the case of no misalignment is denoted by the vertical dashed line. Note that the width of the shaded area and the red solid error bars correspond to an uncertainty that is ~ 10 times smaller than the expected statistical uncertainty from fitting the Belle II dataset.

$\text{Im}(\lambda_{CP}) = 0$ is assumed. The values used for the relevant unitarity triangle angles are: $\beta = 22.6^\circ$ and $\gamma = 65.9^\circ$ [70]. The ranges of $r' = [0, 0.04]$ and $\delta' = [0, 360]^\circ$ are scanned and the exact values of $r' = 0.015$ and $\delta' = \pm 90^\circ$ are used to compute the systematic uncertainties on the CP parameters, which are found to be 0.028 for C and 0 for S , S^+ and S^- . This is also the dominant systematic uncertainty for C and only amounts to $\sim 30\%$ of the expected statistical uncertainty.

5.5.11 Summary of systematic uncertainties

The systematic uncertainties for C , S , S^+ and S^- are summarised in Table 5.39. The total systematic uncertainty remains well below the expected statistical uncertainty for all parameters, even though for some sources, such as the bootstrapping bias, the systematic uncertainties are evaluated conservatively.

Table 5.39: Systematic uncertainties for C , S , S^+ and S^- . For each parameter, the total systematic uncertainty is evaluated as the square root of the quadratic sum of the uncertainty due to each source.

Source of uncertainty	C	S	S^+	S^-
Fixed shape parameters	0.003	0.005	0.005	0.004
Flavor Tagger parameters	0.018	0.007	0.014	0.012
Δt Resolution Function parameters	0.005	0.014	0.023	0.018
τ_{B^0} & Δm	< 0.001	0.001	0.001	0.003
Fixed SCF background fractions	0.006	0.004	0.008	0.011
CP violation in $B\bar{B}$ background	0.019	0.017	0.034	0.001
Yield bias	0.005	0.004	0.008	0.014
Bootstrapping bias	0.027	0.054	0.110	0.033
Detector misalignment	0.005	0.003	0.006	0.012
Tag-side interference	0.028	< 0.001	< 0.001	< 0.001
Total systematic uncertainty	0.048	0.059	0.119	0.045
Expected statistical uncertainty	0.110	0.160	0.310	0.300

5.6 Results

Finally, the Belle II data sample is fitted to extract the S and C parameters. Figures 5.74 and 5.75 show the projections of the fit and the values of the fitted parameter are listed in Table 5.40.

Table 5.40: Resulting parameters of the Dalitz-integrated fit to the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample. The quoted uncertainties are purely of statistical nature.

Parameter	Fitted value
C	-0.29 ± 0.13
S	-0.36 ± 0.16
κ_{SCF}	0.81 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	-0.0028 ± 0.0039
$f_{M_{\text{bc}}}^\mu$ [GeV/ c^2]	-0.0001 ± 0.0002
$f_{\Delta E}^\sigma$	1.00 ± 0.09
$f_{M_{\text{bc}}}^\sigma$	0.86 ± 0.06
m_0 [GeV/ c^2]	5.2885 ± 0.0001
$f_{\Delta t, \text{main}}^\sigma$	0.98 ± 0.02
$f_{\Delta t, \text{tail}}^\sigma$	0.81 ± 0.05
$N_{\text{sig-like}}$	540 ± 35
$N_{B\bar{B} \text{ bkg}}$	435 ± 65
$N_{q\bar{q} \text{ bkg}}$	3615 ± 80

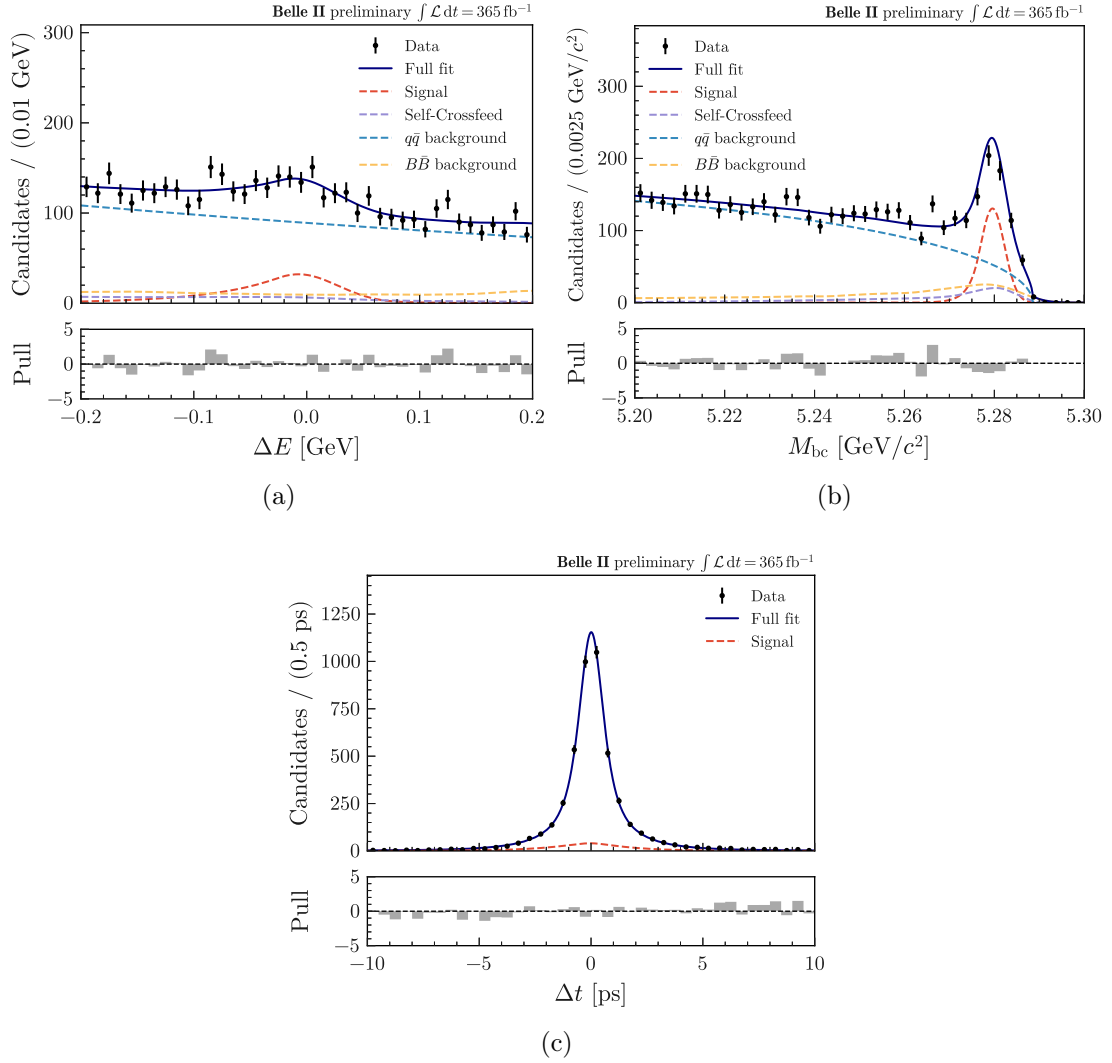


Figure 5.74: Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.

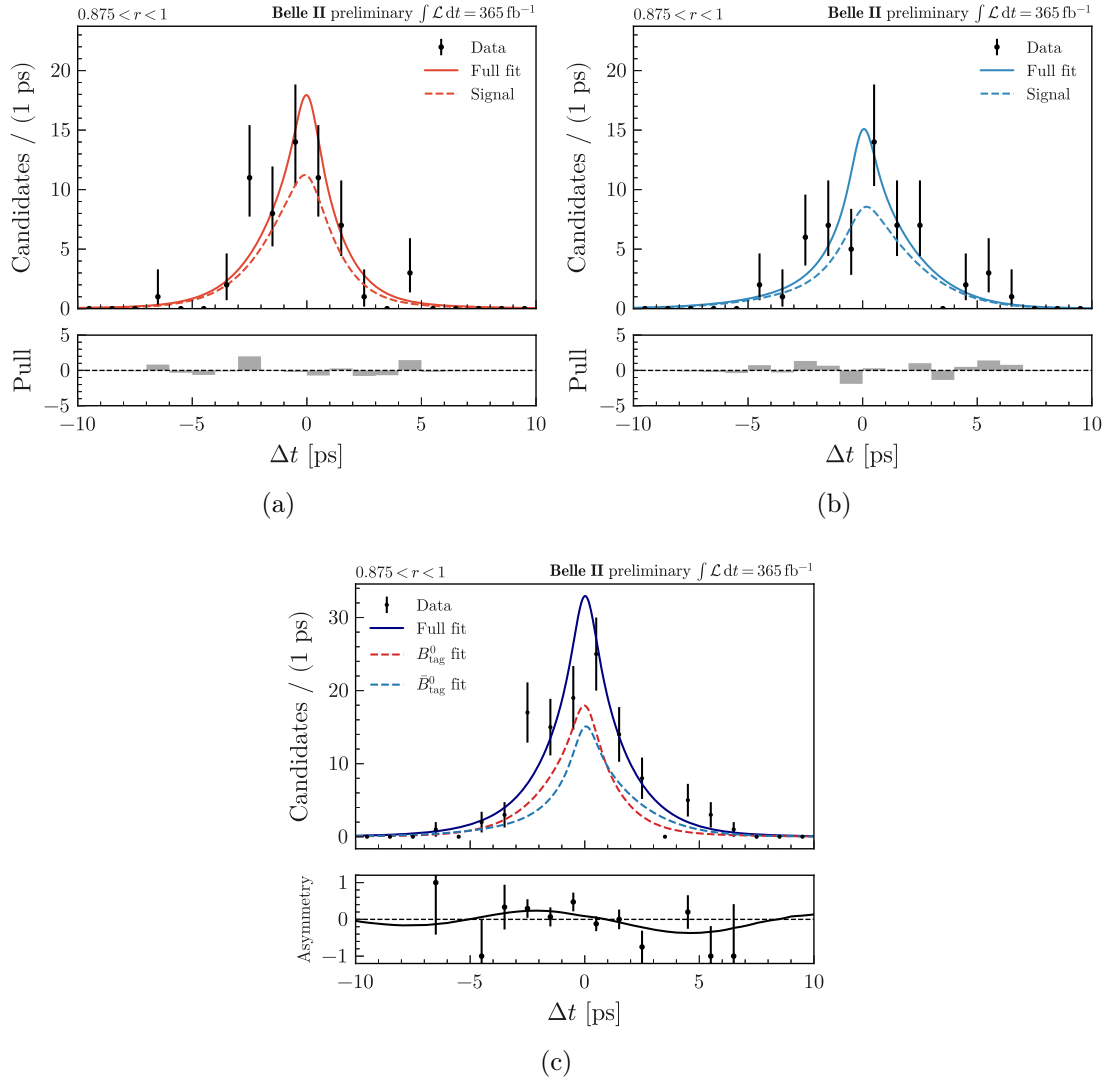


Figure 5.75: Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$.

The measured values of S and C exhibit a mild tension with the SM expectation, corresponding to a significance of 2.3σ , with a precision which surpasses that of previous measurements by Belle and *BABAR* [65, 66]. The significance here is estimated using only the statistical uncertainty. When including the systematic uncertainty, the significance drops to 2.1σ . Given the extensive validation of the fit procedure, the observation of non-zero direct and mixing-induced CP asymmetries is attributed to the data sample itself rather than to any artifact of the fitting method.

Further validation of this result comes from examining the raw B^0 - and $\bar{B}^0$ -tagged Δt distributions in the Belle II data sample, shown in Figure 5.76 in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV and $M_{bc} > 5.27$ GeV/ c^2 , confirming the presence of a non-zero CP asymmetry in the Belle II data sample.

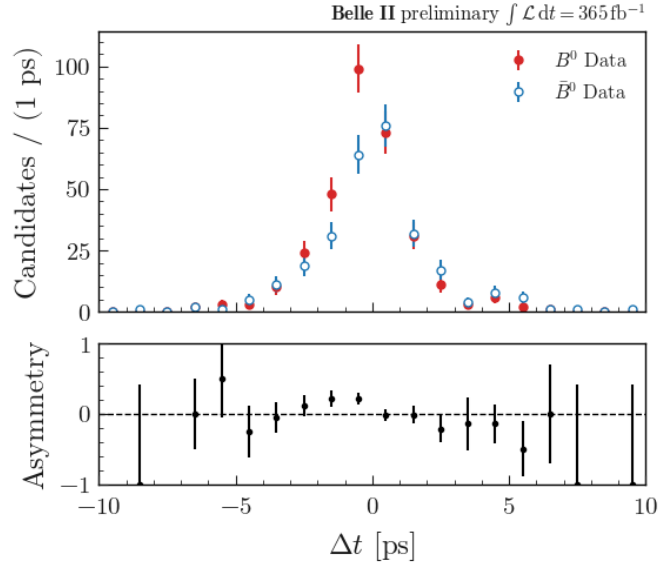


Figure 5.76: Δt distribution of events in the Belle II data sample tagged as B^0 (red) and $\bar{B}^0$ (blue) and asymmetry between them in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV and $M_{bc} > 5.27$ GeV/ c^2 .

The Dalitz-split fit is then performed on the Belle II data sample to extract S^+ and S^- (the distribution of events across the Dalitz plane is shown in Appendix D.5). The fit projections are shown in Figures 5.77–5.80 and the values of the fitted parameters are listed in Table 5.41.

Table 5.41: Resulting parameters of the Dalitz-split fit to the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample. The quoted uncertainties are purely of statistical nature.

Parameter	Fitted value
C	-0.31 ± 0.12
S^+	-0.72 ± 0.31
S^-	0.70 ± 0.30
κ_{SCF}	0.82 ± 0.19
$f_{\Delta E}^\mu$ [GeV]	-0.0026 ± 0.0039
$f_{M_{\text{bc}}}^\mu$ [GeV/ c^2]	-0.0001 ± 0.0002
$f_{\Delta E}^\sigma$	1.00 ± 0.09
$f_{M_{\text{bc}}}^\sigma$	0.86 ± 0.06
m_0 [GeV/ c^2]	5.2885 ± 0.0001
$f_{\Delta t, \text{main}}^\sigma$	0.98 ± 0.02
$f_{\Delta t, \text{tail}}^\sigma$	0.81 ± 0.05
$N_{\text{sig-like}}^I$	270 ± 25
$N_{\text{sig-like}}^I$	265 ± 25
$N_{B\bar{B} \text{ bkg}}^I$	215 ± 40
$N_{B\bar{B} \text{ bkg}}^I$	225 ± 40
$N_{q\bar{q} \text{ bkg}}^I$	1830 ± 55
$N_{q\bar{q} \text{ bkg}}^I$	1775 ± 55

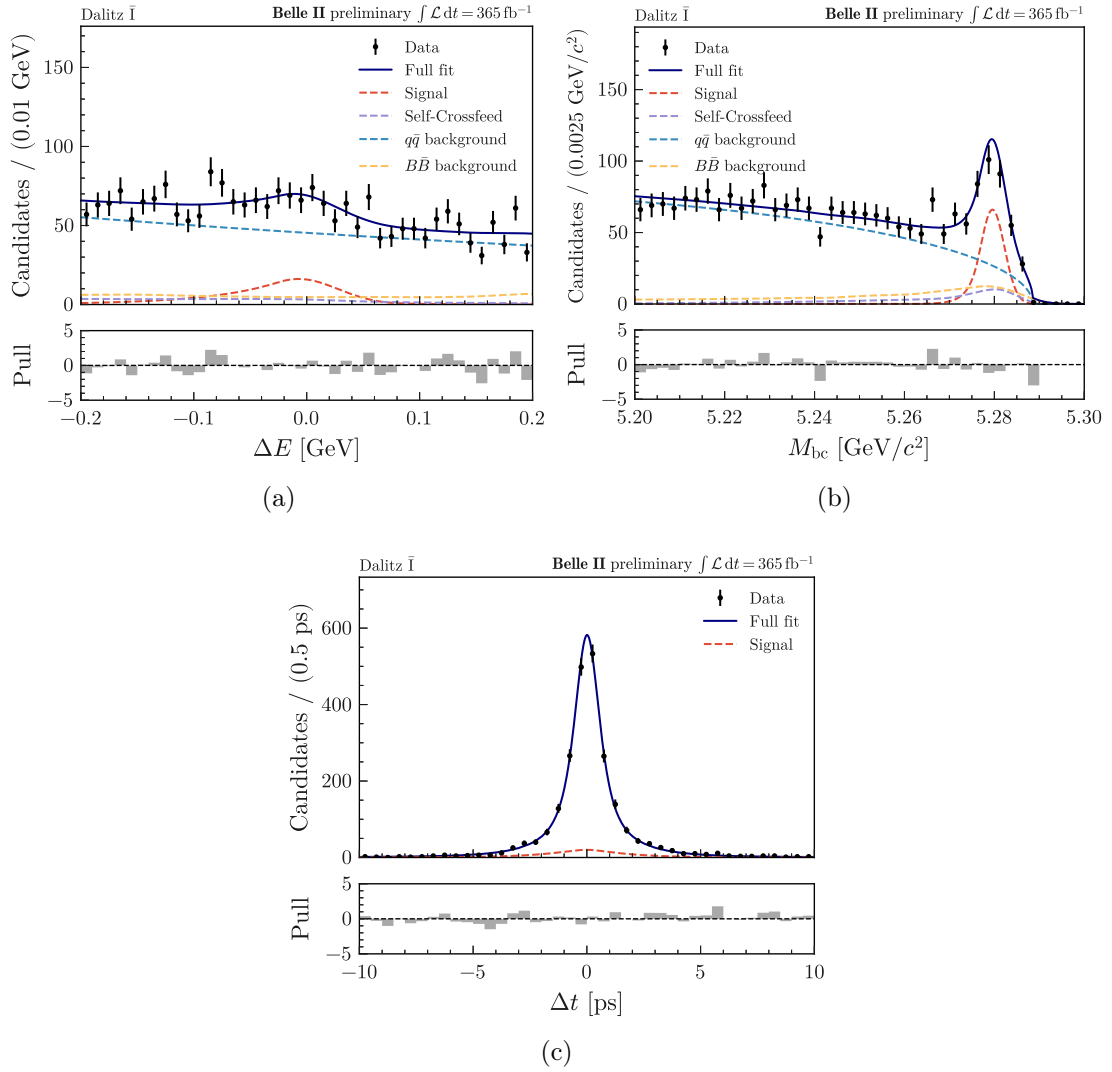


Figure 5.77: Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+ K_S^0}^2 > m_{\pi^- K_S^0}^2$ (Dalitz $\bar{\text{I}}$). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.

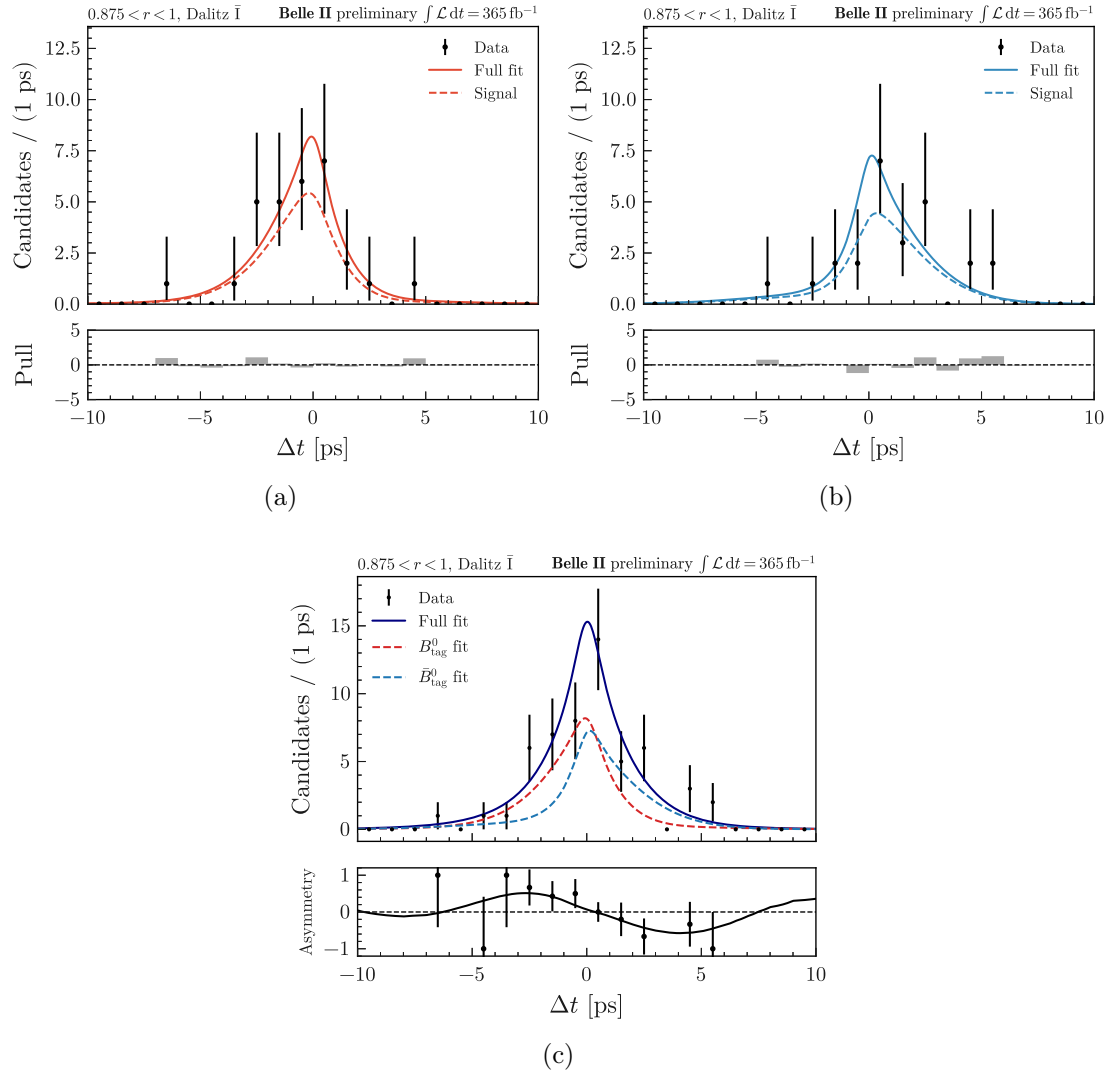


Figure 5.78: Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{\text{I}}$ region and in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$.

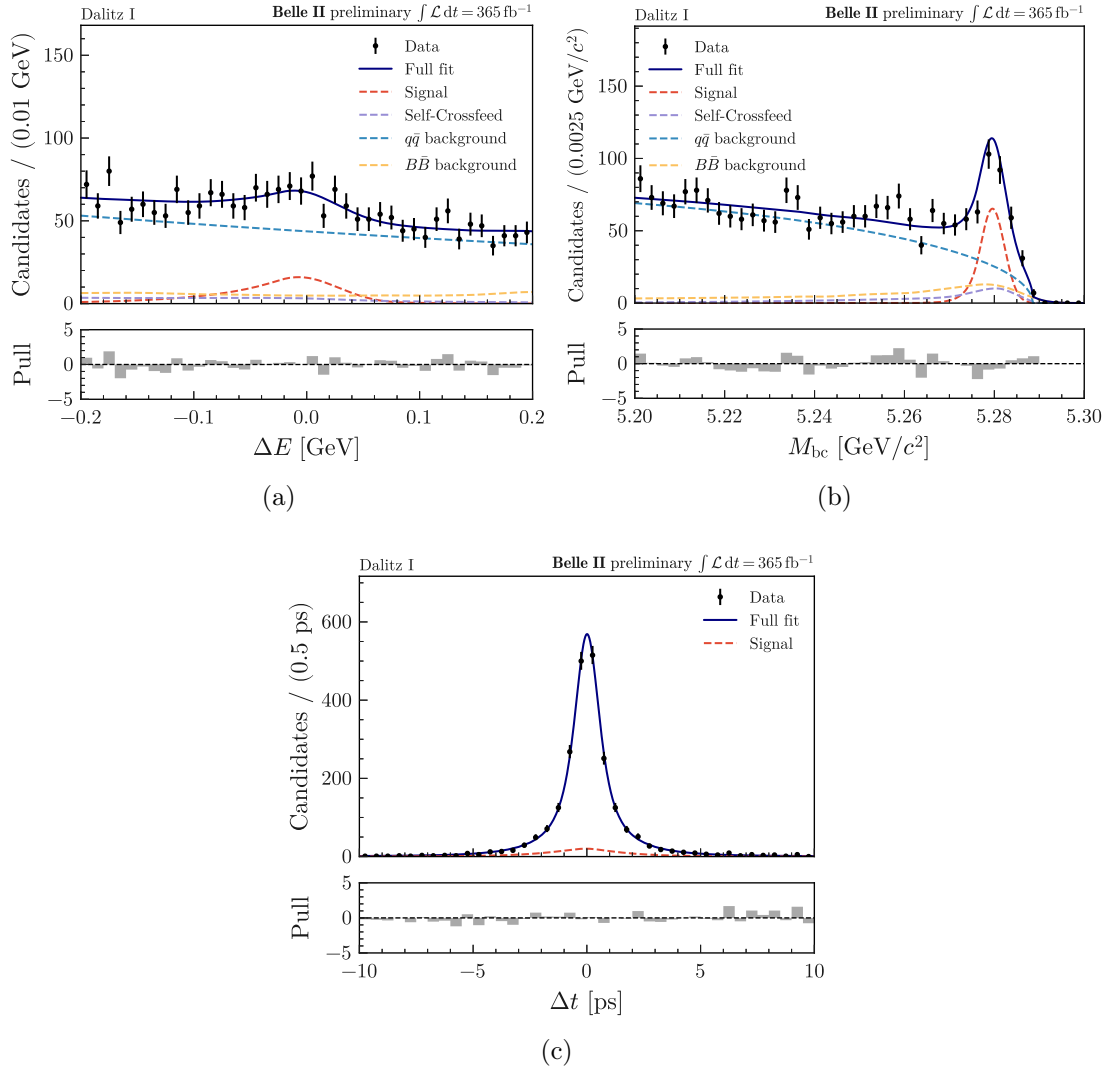


Figure 5.79: Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+ K_S^0}^2 < m_{\pi^- K_S^0}^2$ (Dalitz I). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.

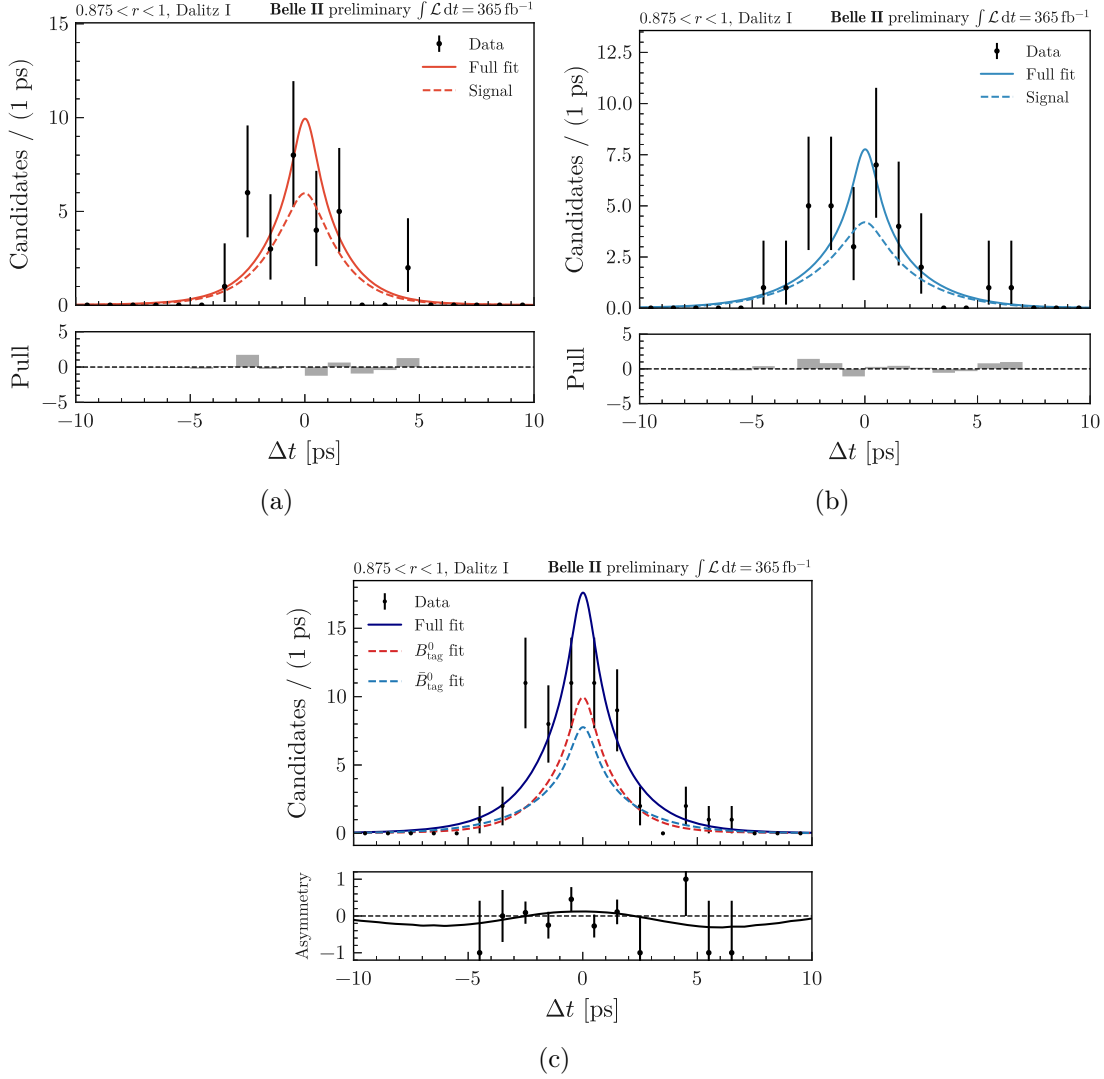


Figure 5.80: Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{bc} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$.

Once again, the measured values of S^+ and S^- show tension with the SM expectation, at a significance of 2.3σ , including both statistical and systematic uncertainties. The value of S^+ is consistent with the measured S , i.e. $S^+ \simeq 2S$, while the measured S^- reveals a clear asymmetry between S^I and $S^{\bar{I}}$. Specifically, the data suggest that $S^{\bar{I}} \simeq S^+$ (or $S^{\bar{I}} \simeq -S^-$) and $S^I \simeq 0$, a feature not observed in any of the MC samples, since no signal MC was generated with $S^I \neq S^{\bar{I}}$.

As in the Dalitz-integrated fit, the raw B^0 - and $\bar{B}^0$ -tagged Δt distributions are examined in Figure 5.81 for the Dalitz I and $\bar{I}$ regions, in a signal enhanced region defined by $|\Delta E| < 0.1$ GeV and $M_{bc} > 5.27$ GeV/ c^2 . These distributions support the interpretation of the fitted values: the observed non-zero mixing-induced CP asymmetry originates mainly from the region of the Dalitz plane with $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$, as indicated by the relative sideways shift between the B^0 - and $\bar{B}^0$ -tagged Δt distributions in Figure 5.78, while in the complementary region the asymmetry is consistent with zero. Interestingly, the non-zero direct CP asymmetry appears to arise predominantly in the region with $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$, reflected in the relative normalisations of the B^0 - and $\bar{B}^0$ -tagged Δt distributions in Figure 5.80.

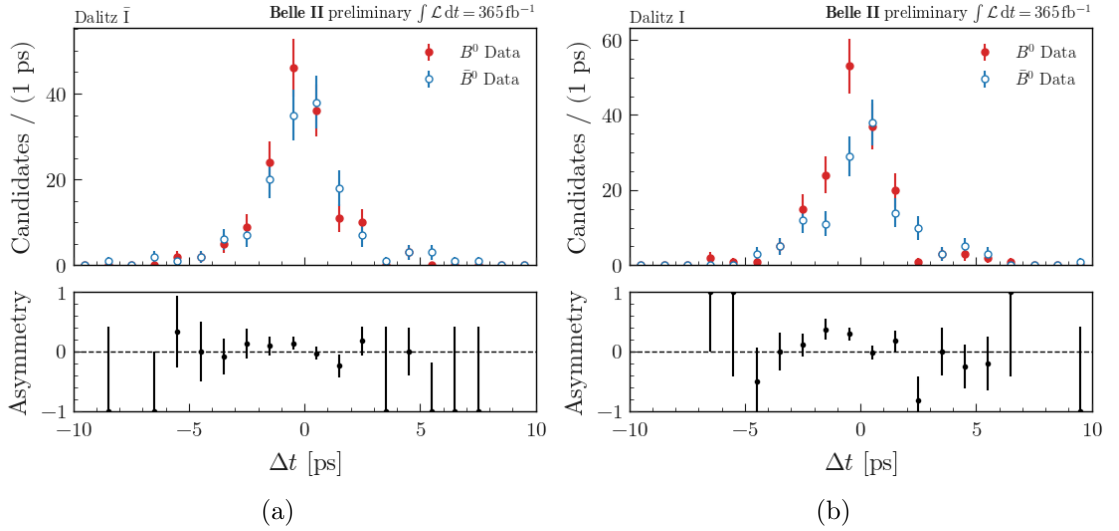


Figure 5.81: Δt distribution of events in the Belle II data sample tagged as B^0 (red) and $\bar{B}^0$ (blue) and asymmetry between them in the (a) Dalitz I region and in the (b) Dalitz $\bar{I}$ within a signal enhanced region defined by $|\Delta E| < 0.1$ GeV and $M_{bc} > 5.27$ GeV/ c^2 .

The final results for the measurement of the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays using the Run 1 dataset of Belle II (365 fb^{-1}) are:

$$C = -0.29 \pm 0.13 \pm 0.06, \quad (5.11)$$

$$S = -0.36 \pm 0.16 \pm 0.05, \quad (5.12)$$

$$S^+ = -0.72 \pm 0.31 \pm 0.12, \quad (5.13)$$

$$S^- = 0.70 \pm 0.30 \pm 0.05, \quad (5.14)$$

where the first uncertainty is statistical and the second systematic. The correlation between S and C is -10% , while the correlation between S^+ and S^- is 7% .

As previously mentioned, this measurement was performed in parallel with an analysis using the entire Belle dataset (711 fb^{-1}), which followed the same selection and fitting strategy⁸. There are however certain differences between the two measurements, mostly due to the different experimental setups. While the Lorentz boost is higher in the case of Belle, which leads to a better kinematic approximation, the Belle II detector has an improved performance for photons, particle identification and vertex reconstruction. The results of this Belle analysis are:

$$C = -0.04 \pm 0.11 \pm 0.07, \quad (5.15)$$

$$S = -0.18 \pm 0.17 \pm 0.08, \quad (5.16)$$

$$S^+ = -0.33 \pm 0.34 \pm 0.15, \quad (5.17)$$

$$S^- = -0.36 \pm 0.38 \pm 0.11, \quad (5.18)$$

where again the first uncertainty is statistical and the second systematic. The precision achieved in this Belle measurement is comparable to that of the Belle II dataset, despite the latter containing only half the luminosity, thanks to the significantly improved detector performance and an improved flavor tagging algorithm. Both measurements are statistically limited and the same sources of systematic uncertainty are considered. The CP parameters measured with the Belle dataset are all consistent with the SM expectation within statistical uncertainties. The two measurements are also mutually compatible, with the exception of S^- , where a tension of about 3σ is observed between the two experiments. The measured value of S^- with the Belle dataset suggests that the mixing-induced CP asymmetry arises predominantly from the opposite Dalitz region compared to the Belle II result. However, this discrepancy could simply be the result of a statistical fluctuation, rather than a genuine physical effect.

⁸When this analysis started, the decision was made to perform the two measurements separately due to incompatibility between the software frameworks used by the two experiments.

The combination of the results is performed following the method presented in Section 3 of Ref. [42], taking into account the correlations between the measured CP parameters. All sources of systematic uncertainty are considered as uncorrelated, except the tag-side interference, which is fully correlated. The combined results using the Belle and Belle II datasets are:

$$C = -0.17 \pm 0.09 \pm 0.04 , \quad (5.19)$$

$$S = -0.29 \pm 0.11 \pm 0.05 , \quad (5.20)$$

$$S^+ = -0.57 \pm 0.23 \pm 0.10 , \quad (5.21)$$

$$S^- = 0.31 \pm 0.24 \pm 0.05 . \quad (5.22)$$

The combined measurement of S and C improves by at least a factor two the precision reached by the previous measurements by Belle and *BABAR* [65, 66]. The measured CP parameters exhibit a mild 1.3σ to 2.4σ tension with the SM prediction, depending on the parameter. In addition, the combined measurement is in agreement with the earlier measurements of S_{eff} in this decay mode by Belle and *BABAR*, which reported $S_{K_S^0\pi^+\pi^-\gamma} = 0.09 \pm 0.27_{-0.07}^{+0.04}$ (1.3σ away) and $S_{K_S^0\pi^+\pi^-\gamma} = 0.14 \pm 0.25 \pm 0.03$ (1.5σ away), respectively. Overall, the combined measurement shows no evidence for physics beyond the Standard Model. Its results are consistent with the earlier Belle and *BABAR* measurements and compatible with the SM expectation once uncertainties are taken into account. Although continued improvements in precision and complementary analyses remain essential for fully mapping the dynamics of this decay, the present study provides a coherent picture in which no significant deviation from established SM behavior is observed.

Conclusion and prospects

In this thesis, three interconnected topics have been presented, each contributing to the overarching goal of measuring the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays.

Firstly, the physics performance of the future upgrade to the Belle II vertex detector, the VTX, has been evaluated using simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays, which serve as an excellent benchmark for testing both the spatial resolution and detection efficiency of alternative detector geometries. The fully pixelated 5-layer vertex detector geometry is found to improve on the decay vertex resolution while retaining a similar efficiency for reconstructing B decays as the current 6-layer design. The VTX is also shown to be a lot more robust against the increasing levels of beam-induced background which are expected at the very high instantaneous luminosity that the SuperKEKB collider intends to reach.

Secondly, the data-driven calibration of the Belle II Flavor Tagger and Δt Resolution Function parameters is described. Through this process the performance of the newly developed GNN-based Flavor Tagger has been evaluated and found to be 18% better relative to the category-based algorithm used previously. The calibrated parameters have been used in the time-dependent analysis presented here, as well as in other Belle II analyses, such as the measurement of $\sin 2\beta$ using $B^0 \rightarrow J/\psi K_S^0$ decays.

Finally, the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays has been measured with the Run 1 Belle II dataset, which corresponds to 387 million $B\bar{B}$ pairs (365 fb^{-1}) collected at the $\Upsilon(4S)$. The direct and mixing-induced CP violation parameters are determined to be $C = -0.29 \pm 0.13 \pm 0.06$ and $S = -0.36 \pm 0.16 \pm 0.05$, already representing an improvement in precision compared to previous measurements by Belle and *BABAR*. The first measurement of the novel observables proposed in Ref. [27] is also reported, yielding $S^+ = -0.72 \pm 0.31 \pm 0.12$ and $S^- = 0.70 \pm 0.30 \pm 0.05$. The measured CP parameters show a slight tension with the SM expectation at the level of 2.3σ .

This measurement is further combined with a newly performed analysis of the full Belle dataset (711 fb^{-1}), resulting in $C = -0.17 \pm 0.09 \pm 0.04$, $S = -0.29 \pm 0.11 \pm 0.05$, $S^+ = -0.57 \pm 0.23 \pm 0.10$, and $S^- = 0.31 \pm 0.24 \pm 0.05$.

The combination improves the precision on S and C by more than a factor of two relative to the previous measurements by Belle and *BABAR*. Nevertheless, the uncertainty remains too large to draw a definitive conclusion on whether the small observed deviation arises from a NP effect or is simply due to a statistical fluctuation in the analyzed data.

The next step to translate this measurement into direct constraints on the C_7 and C'_7 Wilson coefficients that describe NP in the context of an effective theory, will be to calculate the dilution factor relating the measured asymmetry, $S_{K_S^0 \pi^+ \pi^- \gamma}$, to the true CP asymmetry, $S_{K_S^0 \rho^0 \gamma}$, as a result of non- CP eigenstates participating in the final state. To disentangle these contributions, an amplitude analysis must be performed, preferably with the isospin partner mode, $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$, which benefits from larger statistics. As explained in Ref. [27], this amplitude analysis can also be used to determine two dilution-like factors a and b , which, when combined with the measured values of S^+ and S^- , can be used to place additional new constraints on the C'_7/C_7 complex plane.

The current benchmark for constraints on C_7 and C'_7 has been set by LHCb with the angular analysis of $B^0 \rightarrow K^{*0} e^+ e^-$ decays at very low q^2 , using Run 1 and Run 2 data ($3 + 6 \text{ fb}^{-1}$) [67]. These constraints are expected to become even better when the measurement is repeated with Run 3 data, benefiting from the new full-software trigger scheme, which enables a more efficient and comprehensive use of the increased dataset.

Belle II offers a complementary approach where, with its rapidly increasing dataset and the improved detector performance and analysis techniques described in this thesis, measurements of time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays will soon reach unprecedented precision. These results will be crucial in either confirming the observed tension with the SM or in uncovering clear signs of NP in radiative $b \rightarrow s \gamma$ transitions, thereby deepening our understanding of flavor physics and its connection to physics beyond the Standard Model.

Appendix A

Phenomenological context

A.1 Derivation of time-dependent CP asymmetry formula

This derivation closely follows the one in [106], Section 6.6, albeit with a few simplifications, due to focusing on the special case of neutral B mesons. Starting from the time evolution equations that were defined in Section 1.4.1:

$$|B_{t=0}^0(t)\rangle = g_+(t) |B^0\rangle + \frac{q}{p} g_-(t) |\bar{B}^0\rangle \quad (\text{A.1})$$

$$|\bar{B}_{t=0}^0(t)\rangle = g_+(t) |\bar{B}^0\rangle + \frac{p}{q} g_-(t) |B^0\rangle \quad , \quad (\text{A.2})$$

with:

$$g_{\pm}(t) = \frac{1}{2} e^{-\Gamma t/2} e^{-iMt} \times [e^{i\Delta m_d t/2} \pm e^{-i\Delta m_d t/2}] \quad , \quad (\text{A.3})$$

where $\Delta m_d = M_H - M_L > 0$, $M = (M_L + M_H)/2$ and $\Gamma_L = \Gamma_H \approx \Gamma$ in the case of neutral B mesons, which implies that $\Delta\Gamma \approx 0$.

Using the above expressions, we compute the decay rate of a B^0 meson decaying to a final state f that is accessible by both B^0 and $\bar{B}^0$ flavor eigenstates:

$$\Gamma(B^0(t) \rightarrow f) \propto |\langle f | \mathcal{H} | B^0(t) \rangle|^2 = |g_+(t) A(f) + \frac{q}{p} g_-(t) \bar{A}(f)|^2 \quad , \quad (\text{A.4})$$

where we have defined the time-independent decay amplitudes $A(f) = \langle f | \mathcal{H} | B^0 \rangle$

and $\bar{A}(f) = \langle f | \mathcal{H} | \bar{B}^0 \rangle$. We now have:

$$\Gamma(B^0(t) \rightarrow f) \propto |g_+(t)|^2 |A(f)|^2 + \left| \frac{q}{p} \right|^2 |g_-(t)|^2 |\bar{A}(f)|^2 + 2\Re \left[\left(\frac{q}{p} \right) g_+^*(t) g_-(t) A^*(f) \bar{A}(f) \right] \quad (\text{A.5})$$

$$= |A(f)|^2 \left\{ |g_+(t)|^2 + \left| \frac{q}{p} \right|^2 |g_-(t)|^2 |\bar{\rho}(f)|^2 + 2\Re \left[\left(\frac{q}{p} \right) g_+^*(t) g_-(t) \bar{\rho}(f) \right] \right\} , \quad (\text{A.6})$$

using $\rho(f) = \frac{A(f)}{\bar{A}(f)}$ and $\bar{\rho}(f) = \frac{\bar{A}(f)}{A(f)} = \frac{1}{\rho(f)}$. We continue as follows:

$$\Gamma(B^0(t) \rightarrow f) \propto \frac{1}{2} e^{-\Gamma t} |A(f)|^2 \left\{ K_+(t) + K_-(t) \left| \frac{q}{p} \right|^2 |\bar{\rho}(f)|^2 + 2\Re \left[L^*(t) \left(\frac{q}{p} \right) \bar{\rho}(f) \right] \right\} , \quad (\text{A.7})$$

where we have defined the following expressions:

$$K_{\pm}(t) = 2e^{\Gamma t} |g_{\pm}(t)|^2 = 1 \pm \cos(\Delta m_d t) \quad (\text{A.8})$$

$$L^*(t) = 2e^{\Gamma t} g_+^*(t) g_-(t) = i \sin(\Delta m_d t) . \quad (\text{A.9})$$

For the corresponding decay rate of a $\bar{B}^0$ meson decaying to a common final state f , we obtain:

$$\Gamma(\bar{B}^0(t) \rightarrow f) \propto \frac{1}{2} e^{-\Gamma t} |\bar{A}(f)|^2 \left\{ K_+(t) + K_-(t) \left| \frac{p}{q} \right|^2 |\rho(f)|^2 + 2\Re \left[L^*(t) \left(\frac{p}{q} \right) \rho(f) \right] \right\} . \quad (\text{A.10})$$

To make the computation of the time-dependent CP asymmetry formula a bit easier, we express the decay rates as follows:

$$\Gamma(B^0(t) \rightarrow f) \propto e^{-\Gamma t} G_f(t) \quad (\text{A.11})$$

$$\Gamma(\bar{B}^0(t) \rightarrow f) \propto e^{-\Gamma t} \bar{G}_f(t) , \quad (\text{A.12})$$

where:

$$G_f(t) = a + b \cos(\Delta m_d t) + c \sin(\Delta m_d t) \quad (\text{A.13})$$

$$\bar{G}_f(t) = \bar{a} + \bar{b} \cos(\Delta m_d t) + \bar{c} \sin(\Delta m_d t) , \quad (\text{A.14})$$

having defined:

$$a = |A(f)|^2 \left[1 + \left| \frac{q}{p} \right|^2 |\bar{\rho}(f)|^2 \right] \quad (\text{A.15})$$

$$b = |A(f)|^2 \left[1 - \left| \frac{q}{p} \right|^2 |\bar{\rho}(f)|^2 \right] \quad (\text{A.16})$$

$$c = -2|A(f)|^2 \Im \left[\left(\frac{q}{p} \right) \bar{\rho}(f) \right] , \quad (\text{A.17})$$

and:

$$\bar{a} = |\bar{A}(f)|^2 \left[1 + \left| \frac{p}{q} \right|^2 |\rho(f)|^2 \right] \quad (\text{A.18})$$

$$\bar{b} = |\bar{A}(f)|^2 \left[1 - \left| \frac{p}{q} \right|^2 |\rho(f)|^2 \right] \quad (\text{A.19})$$

$$\bar{c} = -2|\bar{A}(f)|^2 \Im \left[\left(\frac{p}{q} \right) \rho(f) \right] . \quad (\text{A.20})$$

Now, for the time-dependent CP asymmetry, we have:

$$\mathcal{A}_{CP}(t) = \frac{\Gamma(\bar{B}^0(t) \rightarrow f) - \Gamma(B^0(t) \rightarrow f)}{\Gamma(\bar{B}^0(t) \rightarrow f) + \Gamma(B^0(t) \rightarrow f)} = \frac{\bar{G}_f(t) - G_f(t)}{\bar{G}_f(t) + G_f(t)} = \quad (\text{A.21})$$

$$= \frac{(\bar{a} - a) + (\bar{b} - b) \cos(\Delta m_d t) + (\bar{c} - c) \sin(\Delta m_d t)}{(\bar{a} + a) + (\bar{b} + b) \cos(\Delta m_d t) + (\bar{c} + c) \sin(\Delta m_d t)} = \quad (\text{A.22})$$

$$= \frac{2\Im((q/p)\rho(f))}{1 + |\bar{\rho}(f)|^2} \sin(\Delta m_d t) - \frac{1 - |\bar{\rho}(f)|^2}{1 + |\bar{\rho}(f)|^2} \cos(\Delta m_d t) , \quad (\text{A.23})$$

where we have made use of the approximation $|q/p| \approx 1$, which is a good approximation for the special case of B_d mesons, as was discussed in Section 1.4.1. The coefficients in front of the sine and cosine terms are the parameters describing *mixing-induced* and *direct* CP violation, defined as:

$$S = \frac{2\Im((q/p)\rho(f))}{1 + |\bar{\rho}(f)|^2}, \quad C = \frac{1 - |\bar{\rho}(f)|^2}{1 + |\bar{\rho}(f)|^2} , \quad (\text{A.24})$$

respectively. Thus, we obtain the famous time-dependent CP asymmetry formula:

$$\mathcal{A}_{CP}(t) = \frac{\Gamma(\bar{B}^0(t) \rightarrow f) - \Gamma(B^0(t) \rightarrow f)}{\Gamma(\bar{B}^0(t) \rightarrow f) + \Gamma(B^0(t) \rightarrow f)} = S \sin(\Delta m_d t) - C \cos(\Delta m_d t) . \quad (\text{A.25})$$

A.2 Derivation of exponential decay part of the Δt distribution

The appearance of an absolute value in the exponential of the time-dependent decay rate (Equation 1.49) can be understood by considering the kinematics of the decays of the two B mesons produced in $\Upsilon(4S)$ decay. Each neutral B meson decays exponentially with lifetime τ_{B^0} . If one meson decays at time t_1 and the other at time t_2 , the joint probability density is given by:

$$\mathcal{P}(t_1, t_2) = \frac{1}{\tau_{B^0}^2} e^{-t_1/\tau_{B^0}} e^{-t_2/\tau_{B^0}} , \quad (\text{A.26})$$

where $t_1, t_2 > 0$. It is convenient to introduce the average decay time T and the decay-time difference Δt :

$$T = \frac{t_1 + t_2}{2} , \quad \Delta t = t_1 - t_2 . \quad (\text{A.27})$$

In terms of these new variables, the individual decay times are:

$$t_1 = T + \frac{\Delta t}{2} , \quad t_2 = T - \frac{\Delta t}{2} . \quad (\text{A.28})$$

Substituting into the joint distribution gives:

$$\mathcal{P}(T, \Delta t) = \frac{1}{\tau_{B^0}^2} e^{-(T+\frac{\Delta t}{2})/\tau_{B^0}} e^{-(T-\frac{\Delta t}{2})/\tau_{B^0}} = \frac{1}{\tau_{B^0}^2} e^{-2T/\tau_{B^0}} . \quad (\text{A.29})$$

The exponential dependence on Δt cancels, but constraints on the integration domain remain. Since $t_1, t_2 \geq 0$, we require:

$$T \pm \frac{\Delta t}{2} \geq 0 \implies T \geq \frac{|\Delta t|}{2} . \quad (\text{A.30})$$

The Δt distribution is obtained by integrating over T :

$$\mathcal{P}(\Delta t) = \int_{|\Delta t|/2}^{\infty} \frac{1}{\tau_{B^0}^2} e^{-2T/\tau_{B^0}} dT \quad (\text{A.31})$$

$$= \frac{1}{\tau_{B^0}^2} \cdot \frac{\tau_{B^0}}{2} e^{-|\Delta t|/\tau_{B^0}} \quad (\text{A.32})$$

$$= \frac{1}{2\tau_{B^0}} e^{-|\Delta t|/\tau_{B^0}} . \quad (\text{A.33})$$

An additional factor of $1/2$ is often included in B factory conventions to account for the symmetrization of the two decays, yielding:

$$\mathcal{P}(\Delta t) = \frac{1}{4\tau_{B^0}} e^{-|\Delta t|/\tau_{B^0}} . \quad (\text{A.34})$$

Appendix B

Auxiliary PDF distributions

B.1 Gaussian function

The standard Gaussian function is defined as:

$$G(x; \mu, \sigma) = \frac{1}{a} \exp \left[\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] , \quad (\text{B.1})$$

where μ and σ are the mean and width of the Gaussian distribution, and $1/a$ is a normalisation constant.

B.2 Bifurcated Gaussian function

The bifurcated Gaussian function is defined as:

$$\text{BG}(x; \mu, \sigma_i) = \frac{1}{a} \exp \left[\frac{1}{2} \left(\frac{x - \mu}{\sigma_i} \right)^2 \right] \begin{cases} i = L & \text{for } (x - \mu) < 0 \\ i = R & \text{for } (x - \mu) > 0 \end{cases} , \quad (\text{B.2})$$

where μ and σ describe a Gaussian, and $1/a$ is a normalisation constant.

B.3 Exponential function

The exponential function is defined as:

$$\text{Exp}(x; \lambda) = \frac{1}{a} e^{\lambda x} , \quad (\text{B.3})$$

where λ is the exponential slope parameter and $1/a$ is a normalisation constant.

B.4 Chebyshev polynomials

The Chebyshev polynomials of the first kind are defined by:

$$T_n(\cos \theta) = \cos(n\theta) . \quad (\text{B.4})$$

They can also be defined using the recurrence relation:

$$T_0(x) = 1 , \quad (\text{B.5})$$

$$T_1(x) = x , \quad (\text{B.6})$$

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) . \quad (\text{B.7})$$

The n^{th} degree Chebyshev polynomial is then given by:

$$\text{CH}(x; c_1, c_2, \dots, c_n) = T_0(x) + c_1 T_1(x) + c_2 T_2(x) + \dots + c_n T_n(x) . \quad (\text{B.8})$$

B.5 Double-sided Crystal Ball function

The double-sided Crystal Ball function is defined as:

$$\text{DSCB}(x; \mu, \sigma, \alpha_L, n_L, \alpha_R, n_R) = N \begin{cases} A_L \left(B_L - \frac{x-\mu}{\sigma} \right)^{-n_L} & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha_L \\ \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) & \text{for } -\alpha_L < \frac{x-\mu}{\sigma} < \alpha_R \\ A_R \left(B_R + \frac{x-\mu}{\sigma} \right)^{-n_R} & \text{for } \frac{x-\mu}{\sigma} \geq \alpha_R \end{cases} , \quad (\text{B.9})$$

where μ and σ are the mean and width of the Gaussian core, and N is a normalisation constant. The constants A_L , B_L and A_R , B_R are chosen such that both the function and its first derivative are continuous at the break points.

B.6 Crystal Ball function

The Crystal Ball function is defined as:

$$\text{DSCB}(x; \mu, \sigma, \alpha, n) = N \begin{cases} A \left(B - \frac{x-\mu}{\sigma} \right)^{-n} & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha \\ \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right) & \text{for } \left| \frac{x-\mu}{\sigma} \right| < \alpha \\ A \left(B + \frac{x-\mu}{\sigma} \right)^{-n} & \text{for } \frac{x-\mu}{\sigma} \geq \alpha \end{cases} , \quad (\text{B.10})$$

where μ and σ are the mean and width of the Gaussian core, and N is a normalisation constant. The constants A and B are chosen such that both the function and its first derivative are continuous at the break points.

B.7 Argus function

The Argus function is defined as:

$$\text{Argus}(x; x_0, c) = N \cdot x \cdot \sqrt{1 - \left(\frac{x}{x_0}\right)^2} \exp \left[c \cdot \sqrt{1 - \left(\frac{x}{x_0}\right)^2} \right], \quad (\text{B.11})$$

where c is the Argus shape parameter and x_0 is the cutoff of the distribution, corresponding to the beam energy, and N is a normalisation constant.

B.8 Maximum Likelihood Fit

The method of maximum likelihood is a standard statistical technique used to estimate parameters of a probability distribution given a set of observed data.

Suppose a dataset consists of N independent observations $x_1, x_2, \dots, x_N$, each described by a Probability Density Function (PDF) $\mathcal{P}(x; \vec{\theta})$ that depends on a set of parameters $\vec{\theta}$. The likelihood function is defined as:

$$\mathcal{L}(\vec{\theta}) = \prod_{i=1}^N \mathcal{P}(x_i; \vec{\theta}), \quad (\text{B.12})$$

which represents the probability of observing the data sample given the model parameters. The principle of maximum likelihood states that the best estimates of $\vec{\theta}$ are those that maximize $\mathcal{L}(\vec{\theta})$, or equivalently minimize the negative log-likelihood (NLL):

$$-2 \ln \mathcal{L}(\vec{\theta}) = -2 \sum_{i=1}^N \ln \mathcal{P}(x_i; \vec{\theta}). \quad (\text{B.13})$$

In cases where the total number of events is also a parameter of interest, the extended likelihood formalism is employed. In this case, the likelihood includes a Poisson term describing the probability of observing N events given the expected number λ :

$$\mathcal{L}_{\text{ext}}(\vec{\theta}, \lambda) = e^{-\lambda} \frac{\lambda^N}{N!} \prod_{i=1}^N \mathcal{P}(x_i; \vec{\theta}). \quad (\text{B.14})$$

Appendix C

Flavor Tagger and decay time resolution calibration

C.1 Fit to the background-subtracted Δt distribution of 1 ab^{-1} of fully simulated data

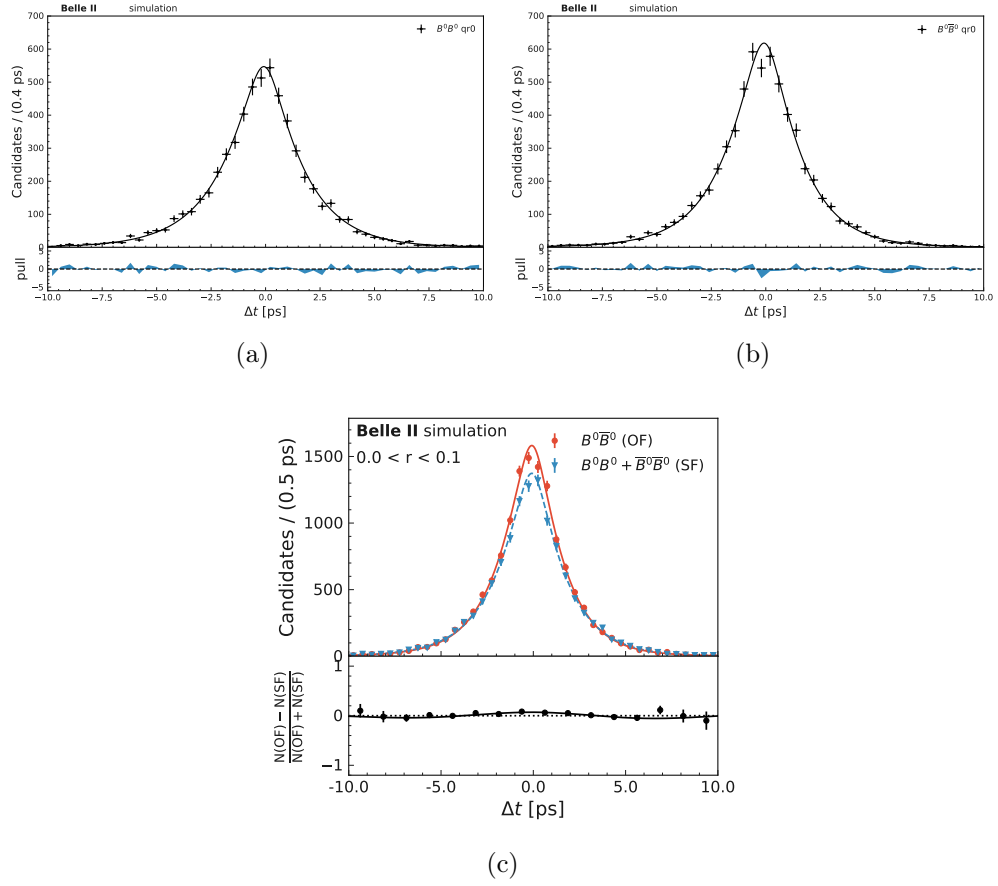


Figure C.1: Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the first r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).

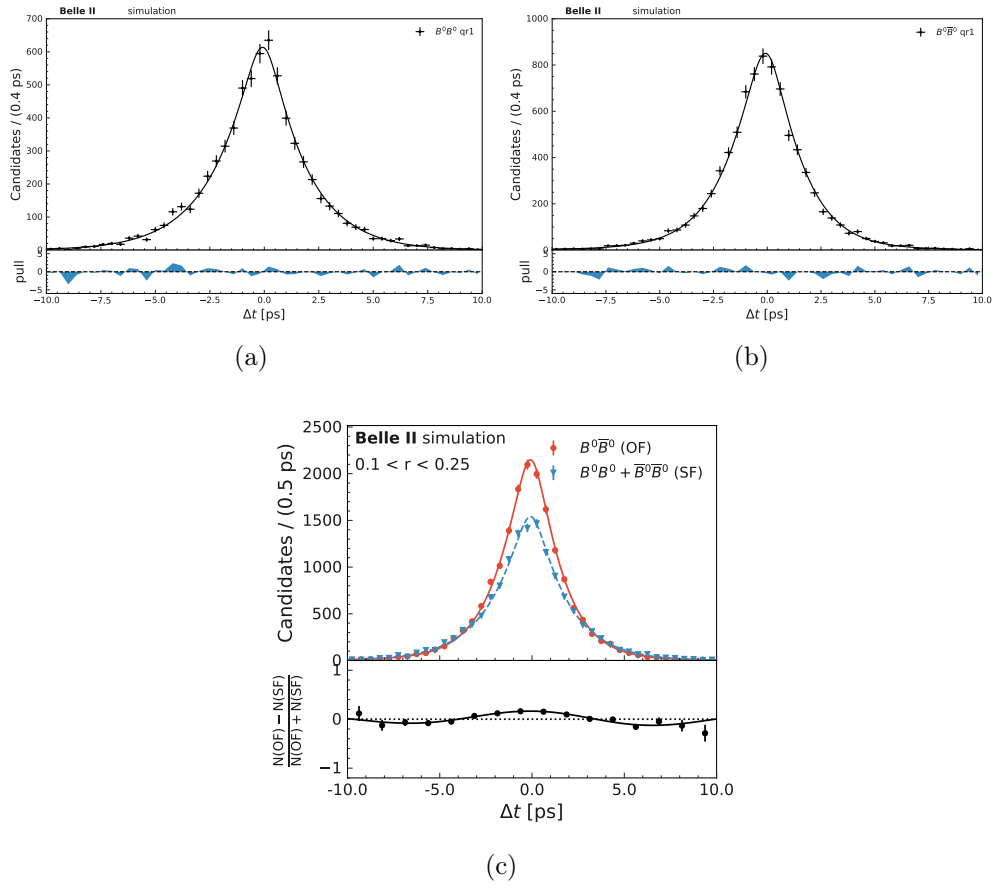


Figure C.2: Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the second r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).

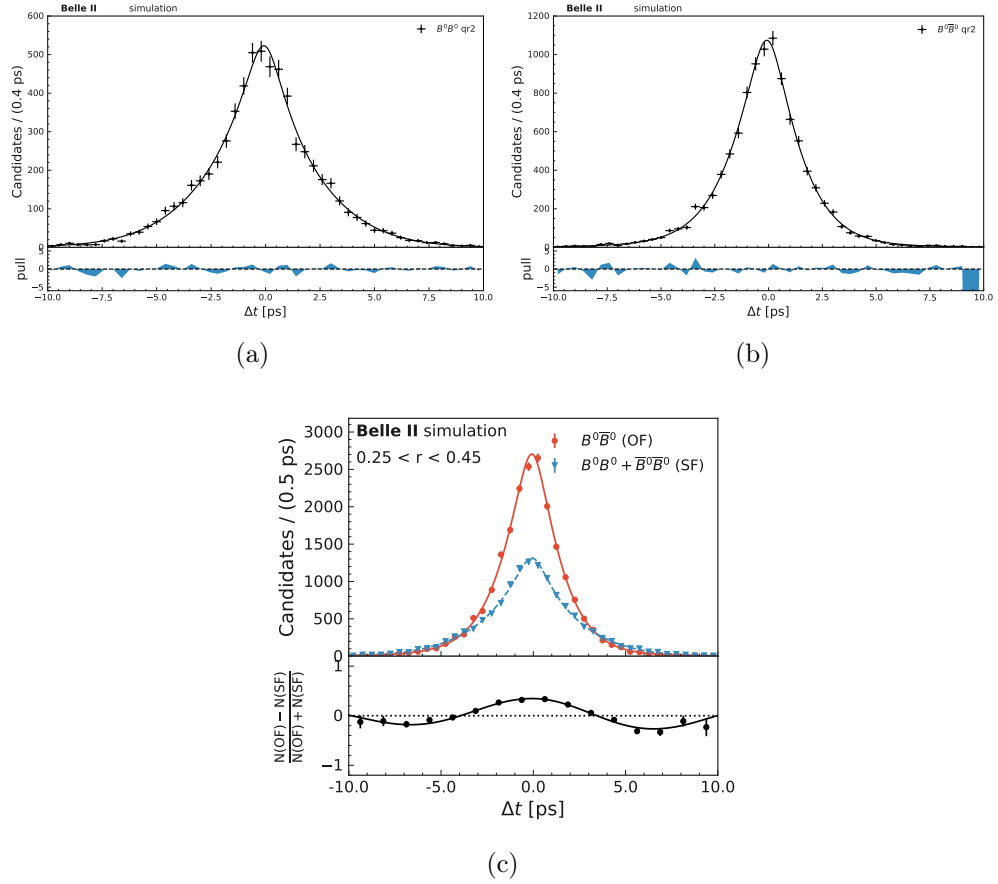


Figure C.3: Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the third r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).

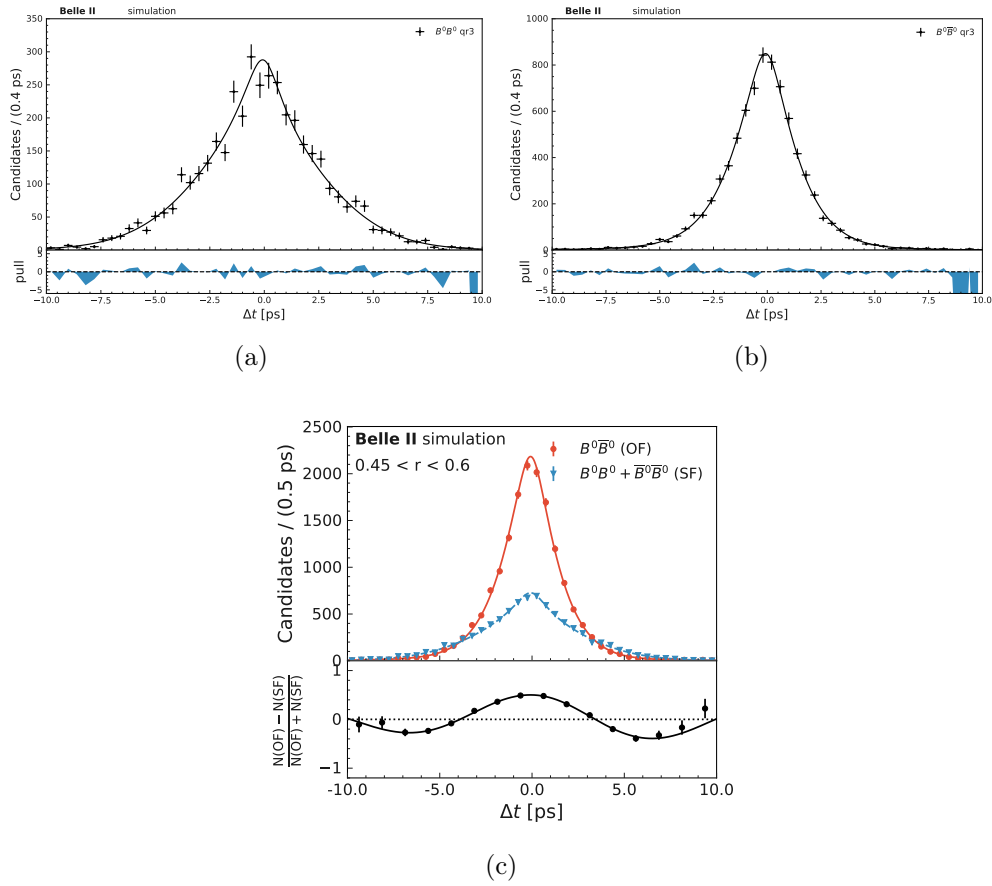


Figure C.4: Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the fourth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).

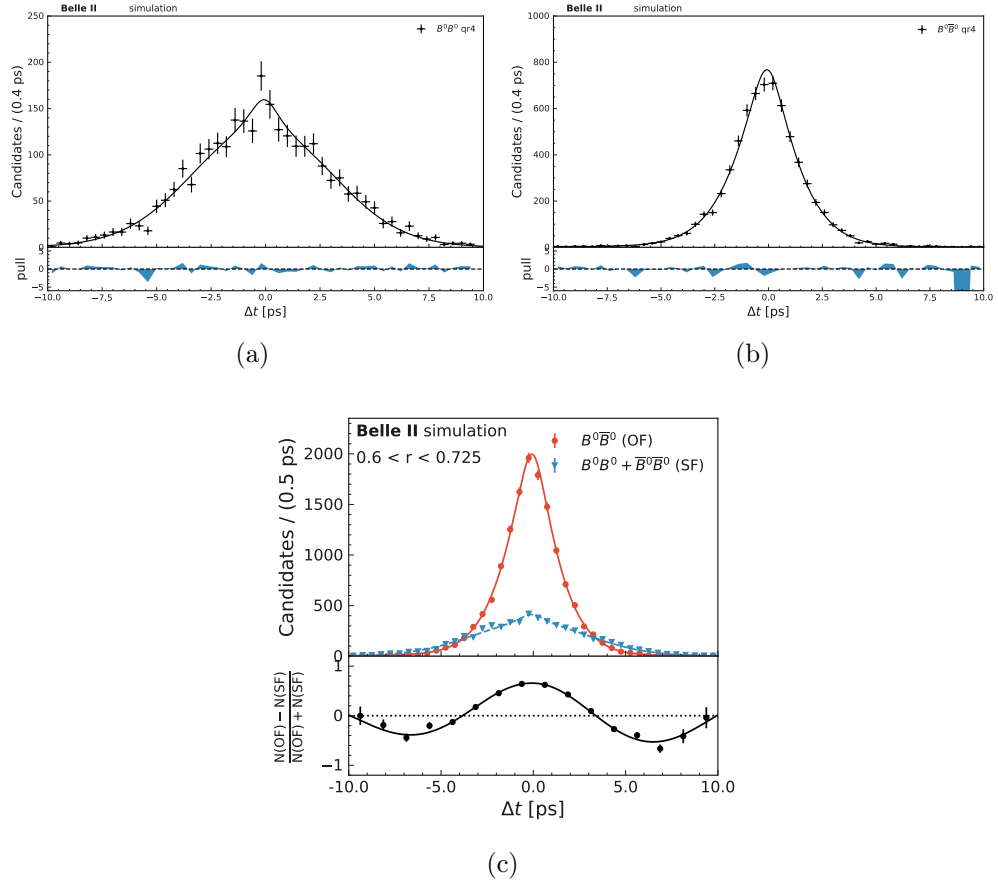


Figure C.5: Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the fifth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).

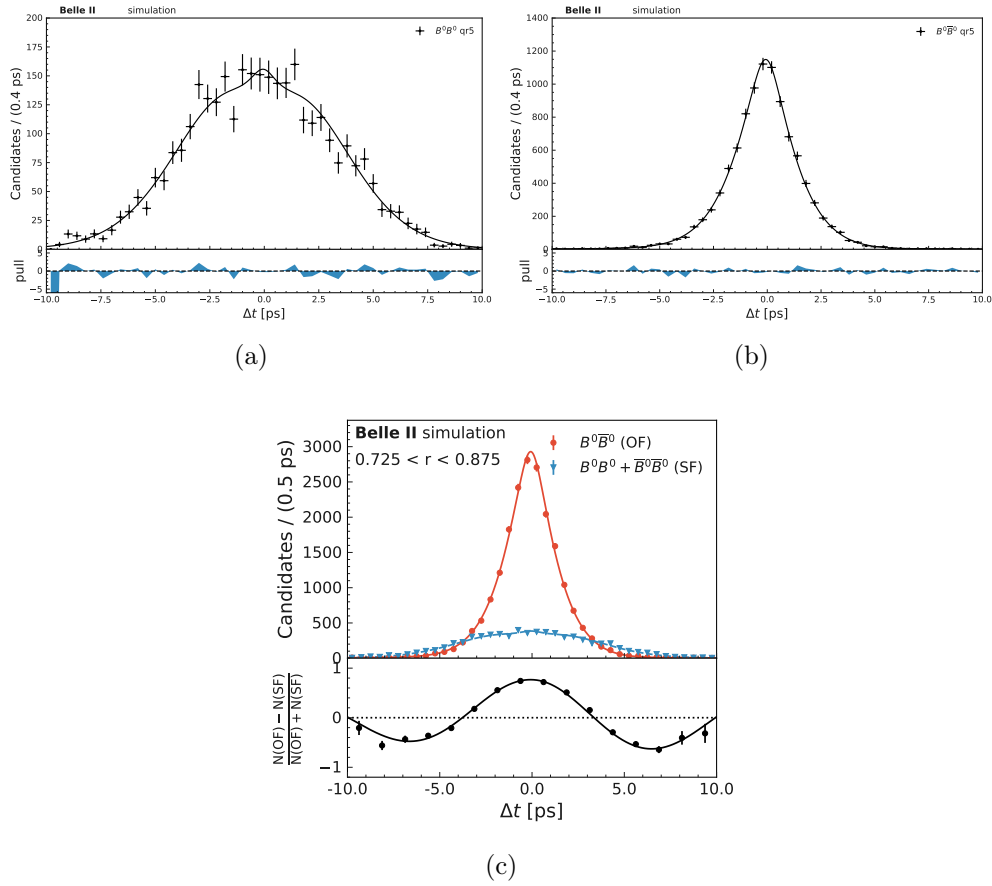


Figure C.6: Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the sixth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).

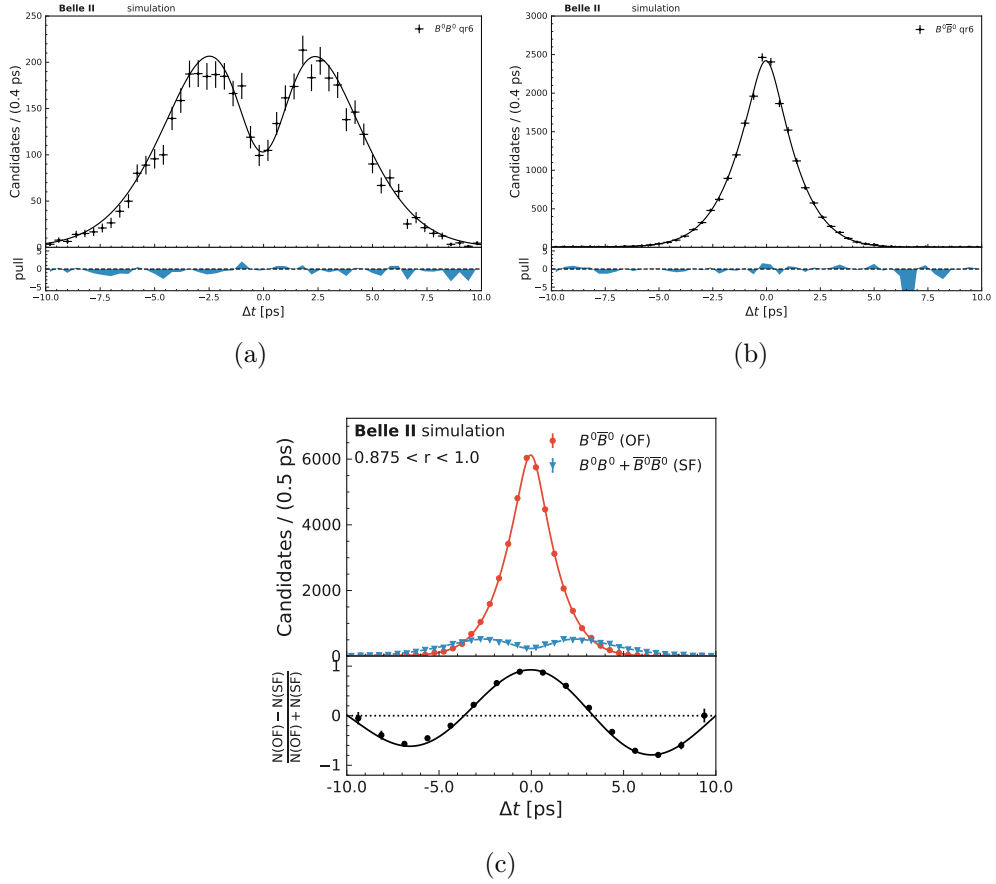


Figure C.7: Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the last r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).

C.2 Correlation matrix from the fit of the background-subtracted Δt distribution of 365 fb^{-1} of Belle II data

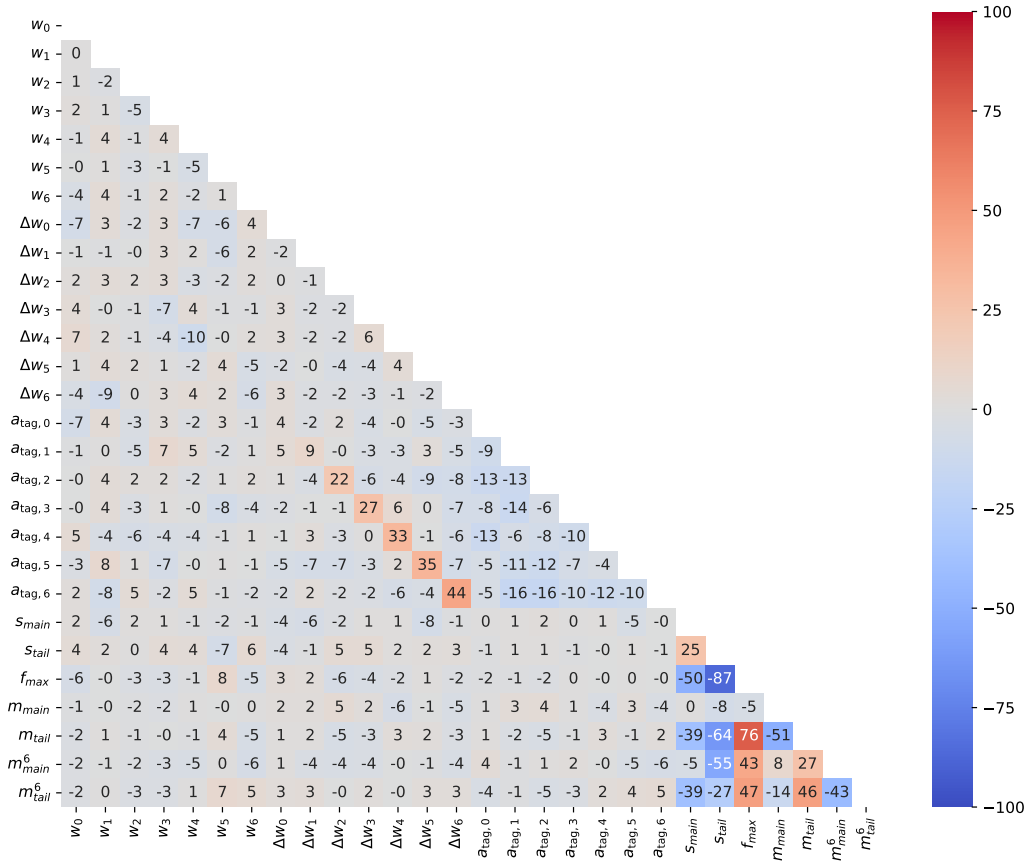


Figure C.8: Correlation between the measured Flavor Tagger and Δt Resolution Function parameters using 365 fb^{-1} of Belle II data. All numbers are in %.

Appendix D

Time-dependent CP asymmetry measurement

D.1 Signal-side B vertex reconstruction

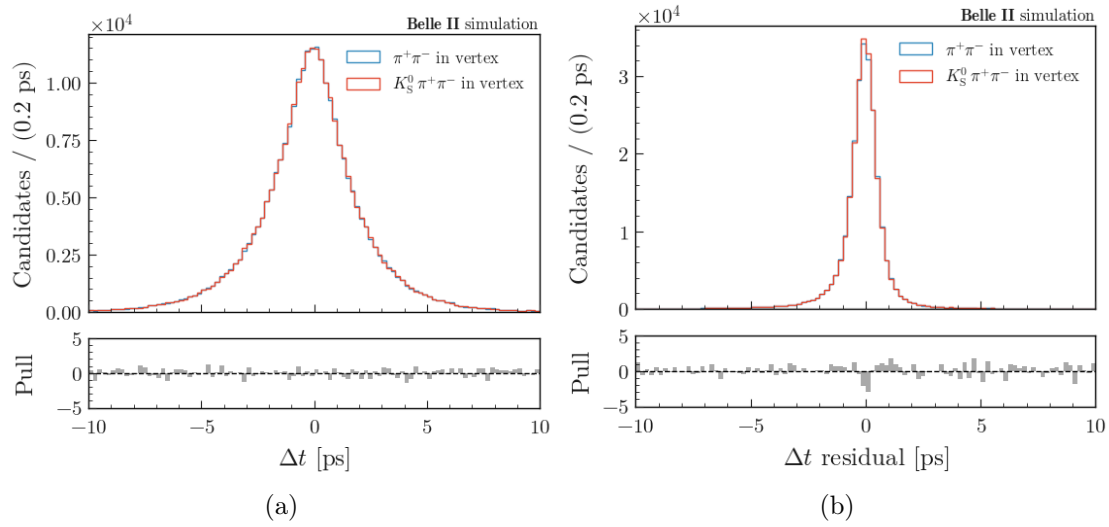


Figure D.1: (a) Δt and (b) Δt residual distribution of signal MC candidates where the position of the B_{sig} vertex was determined with or without considering the pion tracks coming from the K_S^0 .

D.2 BDT input variable distributions for signal and background events

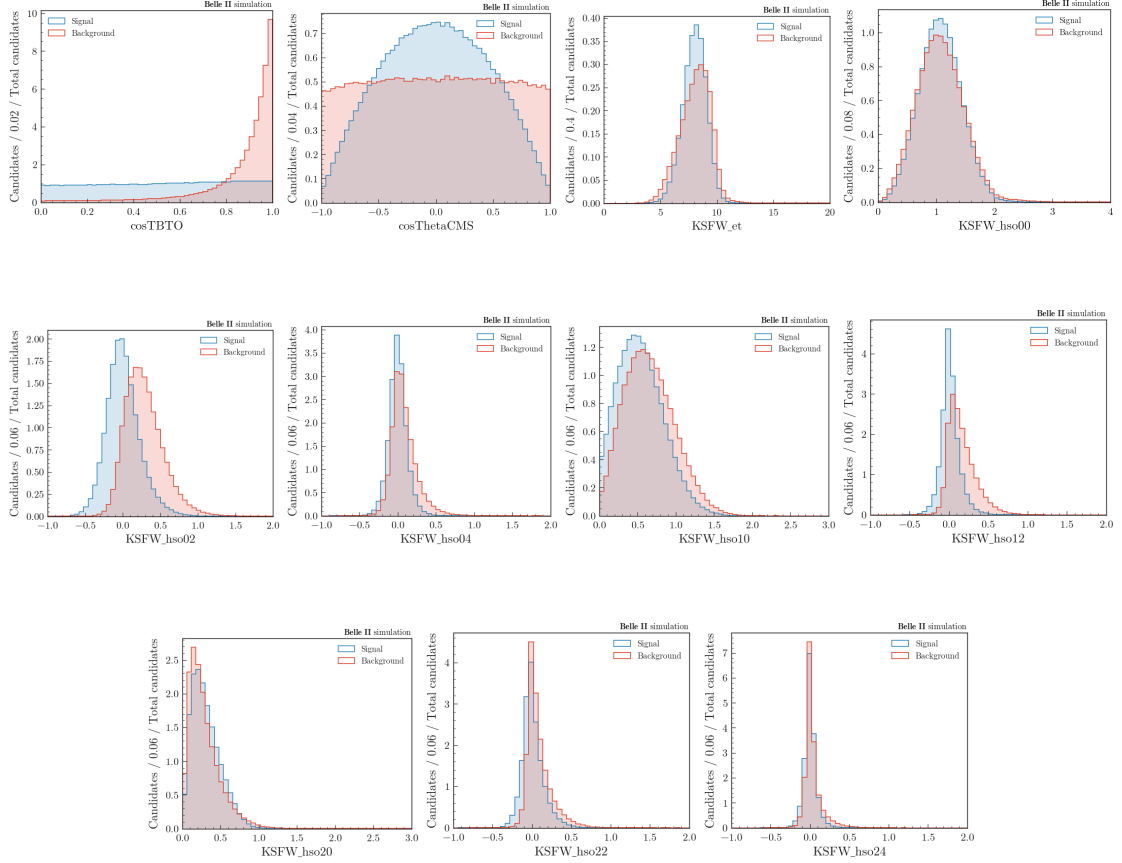


Figure D.2: Distributions of the input variables of the continuum suppression BDT for true signal and continuum background events in simulation.

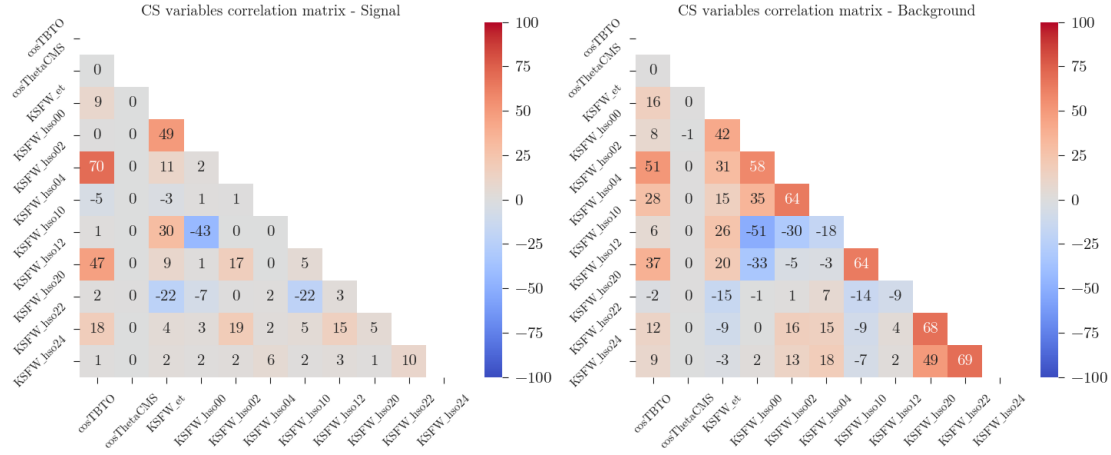


Figure D.3: Correlations between BDT input variables for true signal (left) and continuum background (right) events in simulation.

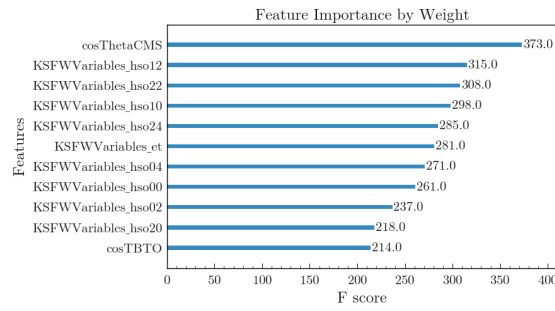


Figure D.4: Variable importance for the BDT classifier.

D.3 Correlations of fit variables

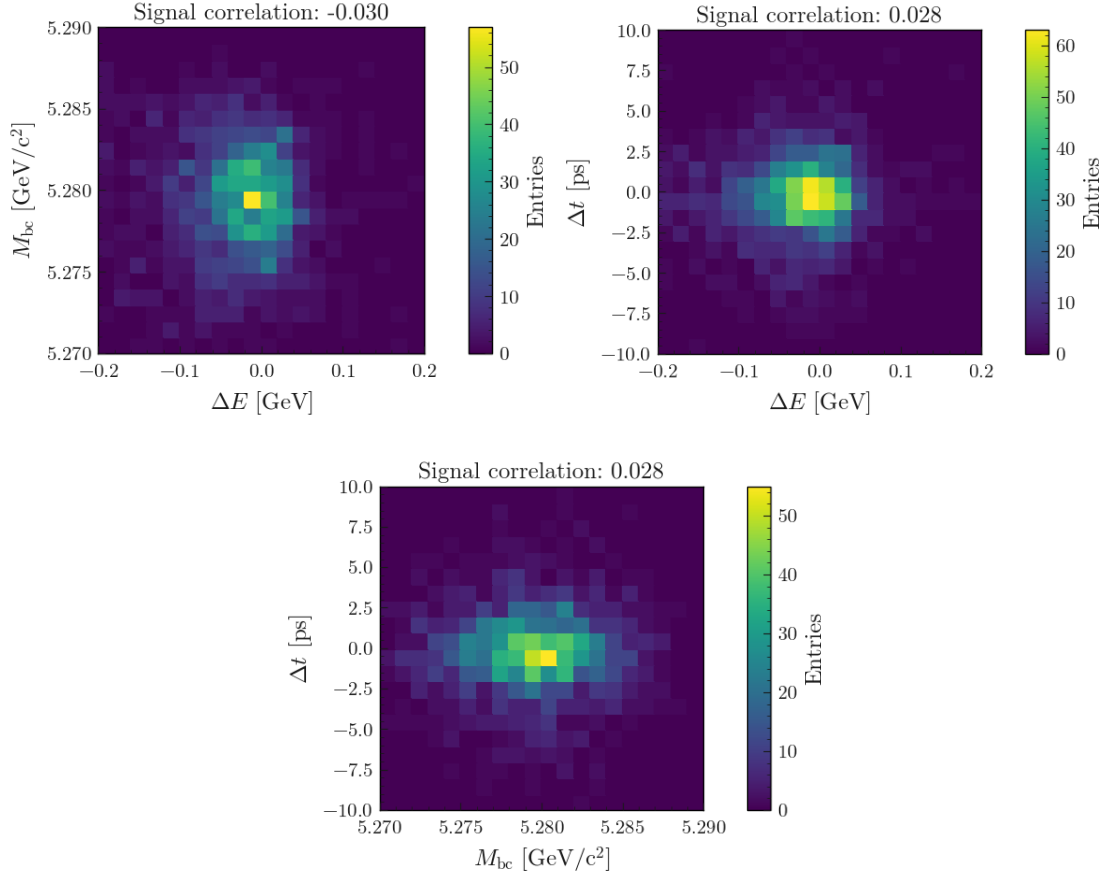


Figure D.5: Pair-wise correlations between fit variables for true signal events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.

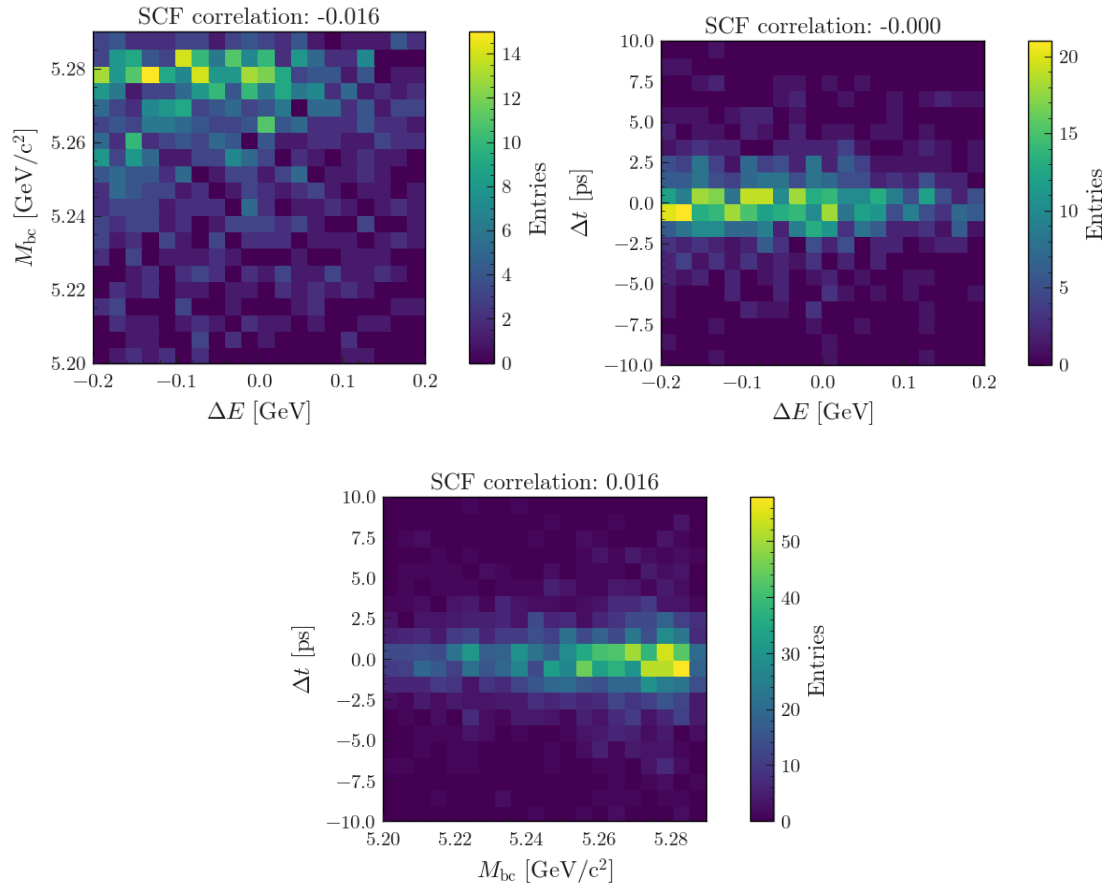


Figure D.6: Pair-wise correlations between fit variables for SCF background events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.

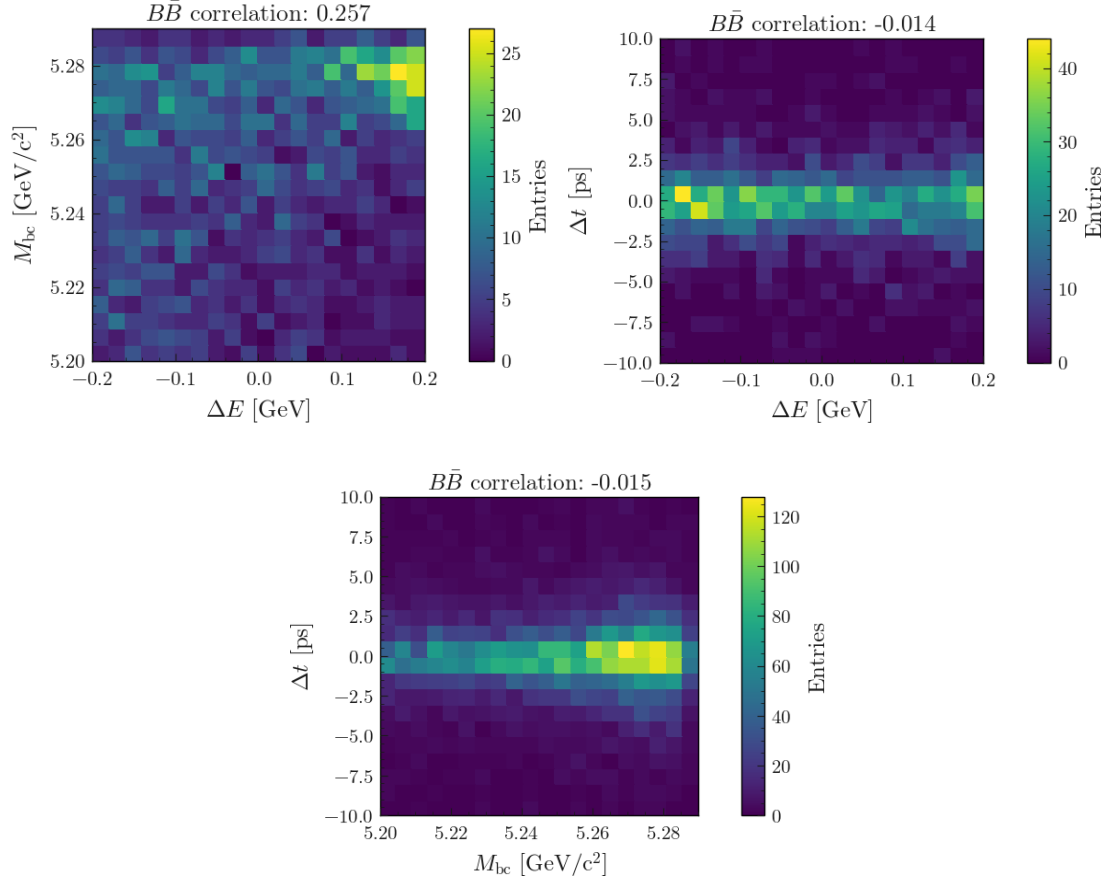


Figure D.7: Pair-wise correlations between fit variables for $B\bar{B}$ background events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.

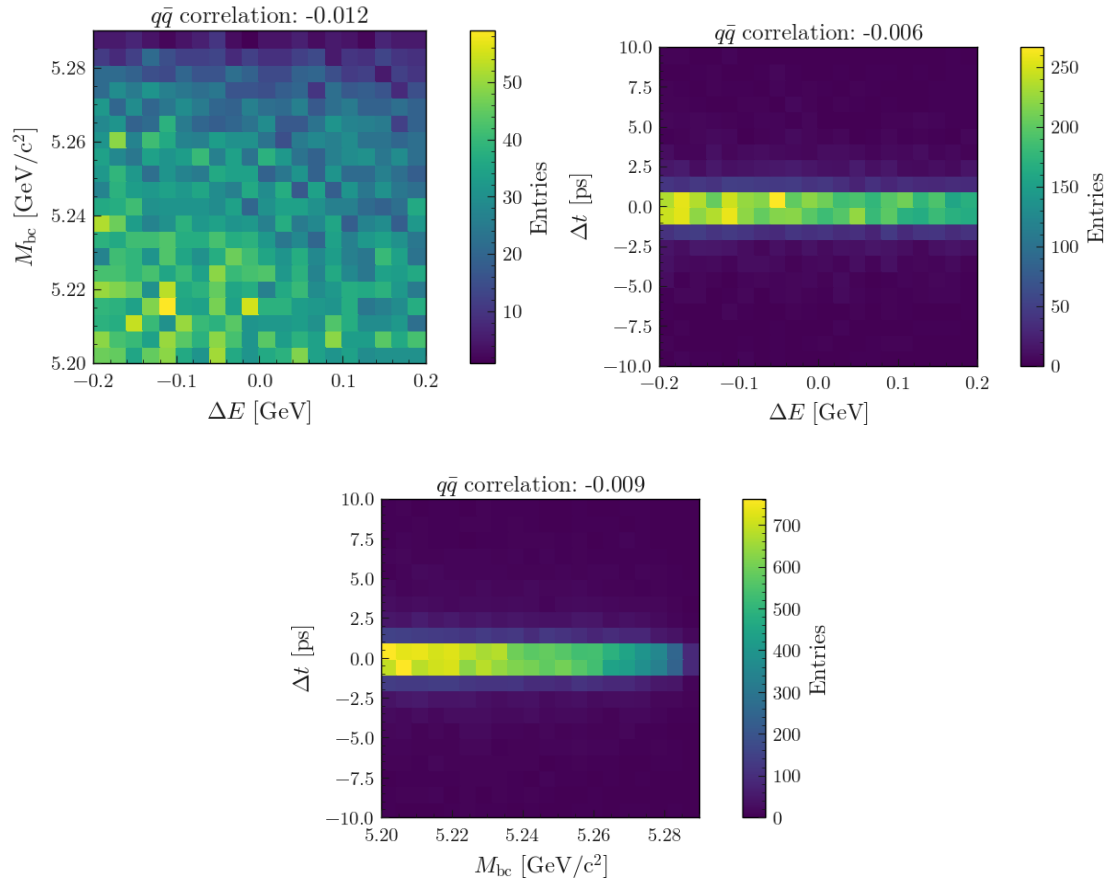


Figure D.8: Pair-wise correlations between fit variables for continuum background events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.

D.4 CP fit of SCF and $B\bar{B}$ backgrounds

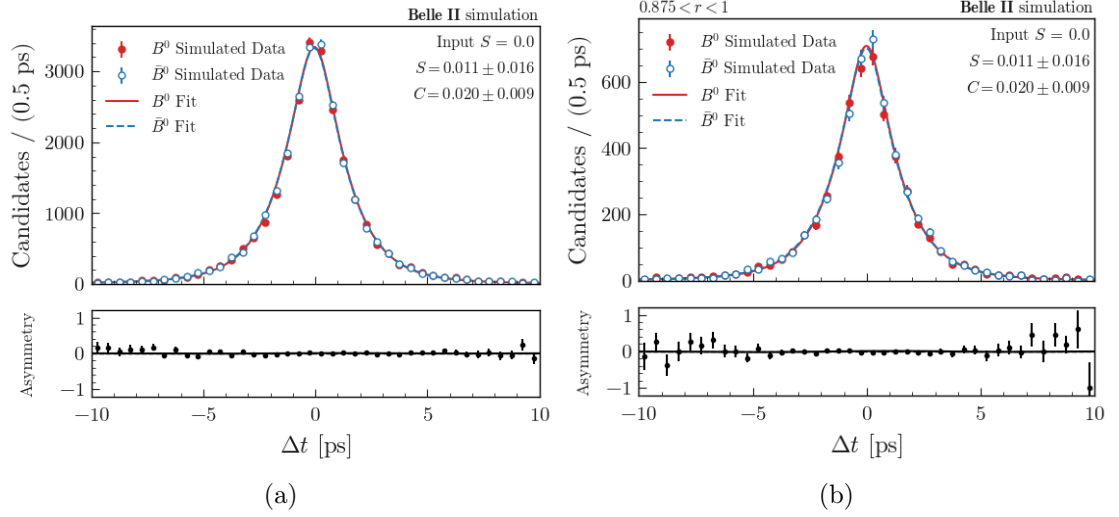


Figure D.9: Projections of the fit to the Δt distribution of simulated SCF background events to extract the CP parameters S and C together with the corresponding CP asymmetry in all Flavor Tagging r -bins (left) and in the last r -bin (right).

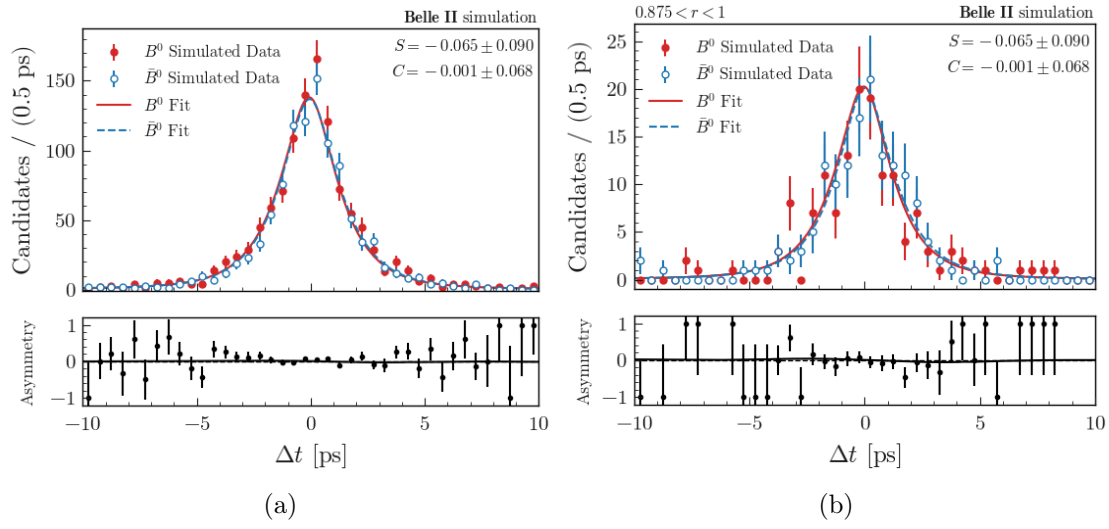


Figure D.10: Projections of the fit to the Δt distribution of simulated $B\bar{B}$ background events to extract the CP parameters S and C together with the corresponding CP asymmetry in all Flavor Tagging r -bins (left) and in the last r -bin (right).

D.5 Dalitz plane plots

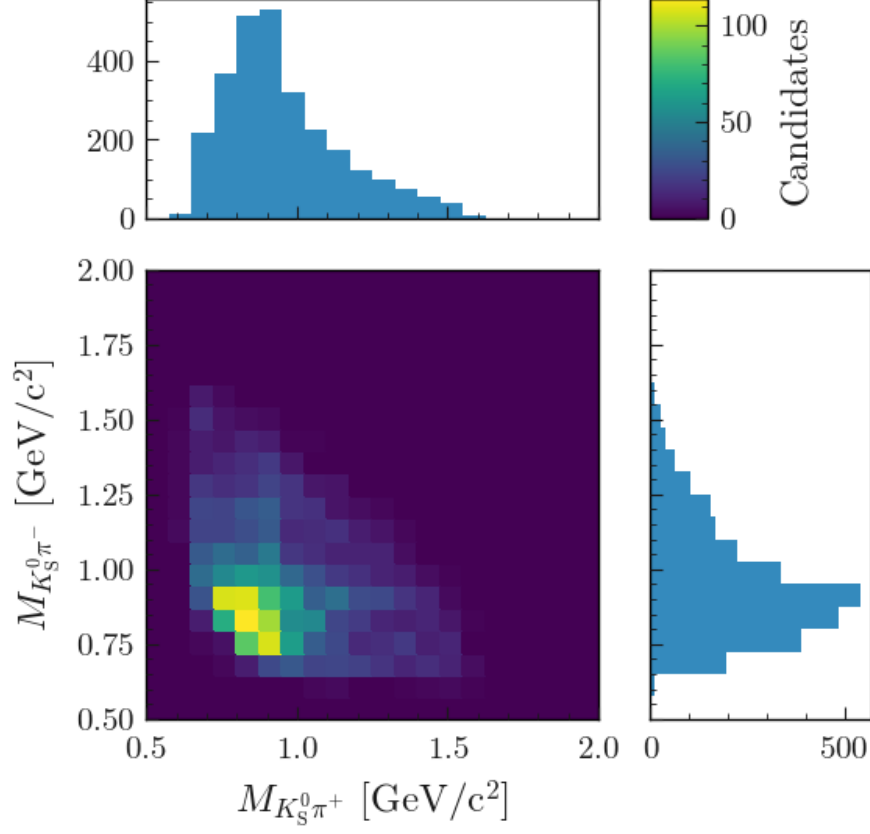


Figure D.11: Distribution of simulated events across the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane in a signal-enhanced region with $|\Delta E| < 0.1 \text{ GeV}$ and $M_{bc} > 5.27 \text{ GeV}/c^2$ found in the generic MC sample.

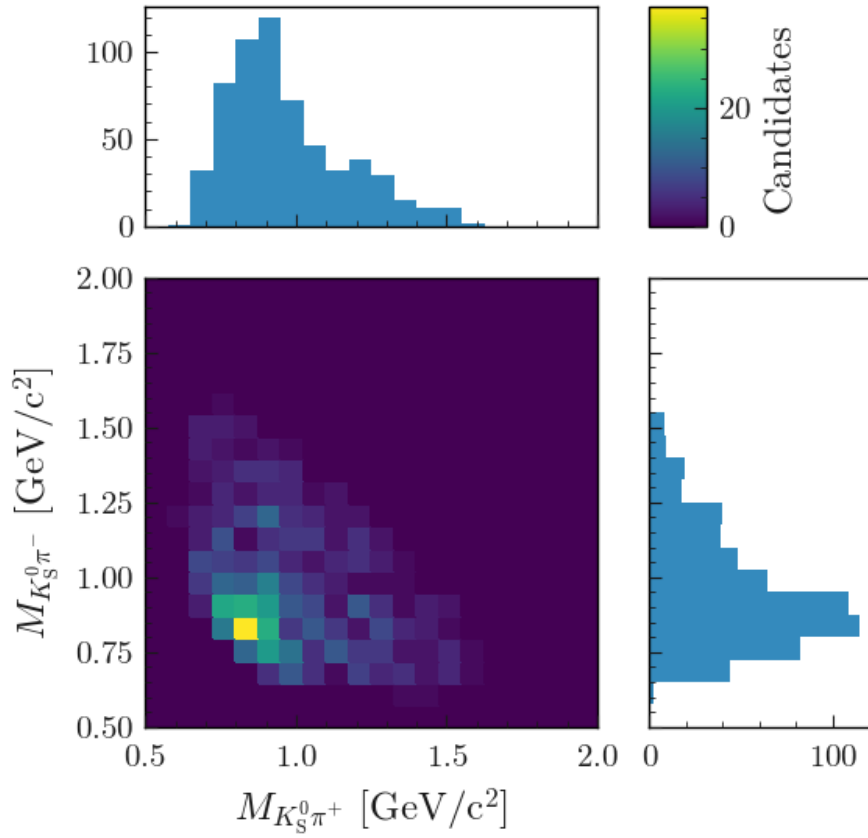


Figure D.12: Distribution of events across the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane in a signal-enhanced region with $|\Delta E| < 0.1$ GeV and $M_{bc} > 5.27$ GeV/ c^2 found in the Belle II dataset.

D.6 Dalitz plane shapes

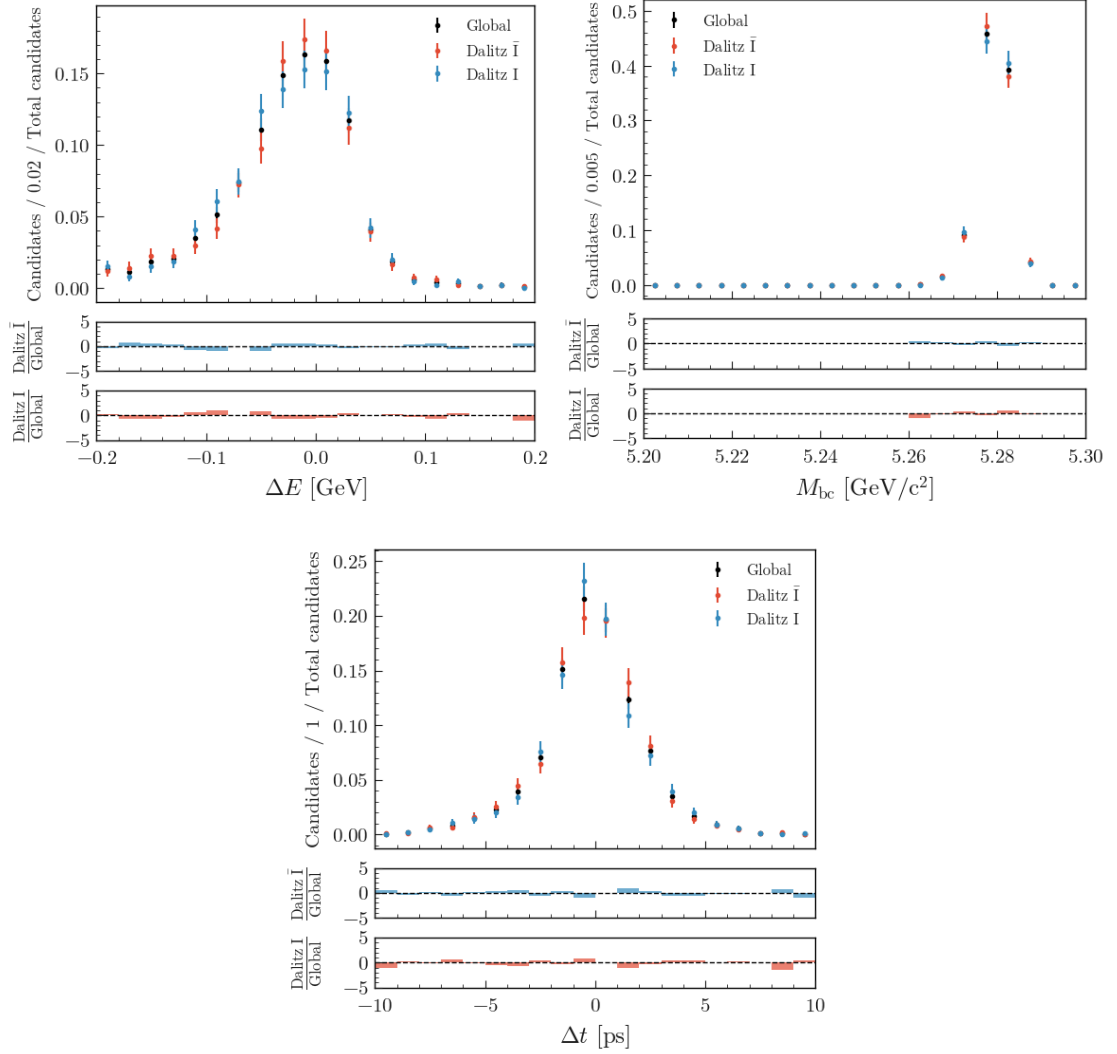


Figure D.13: Distributions of ΔE , M_{bc} and Δt of simulated signal events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$.

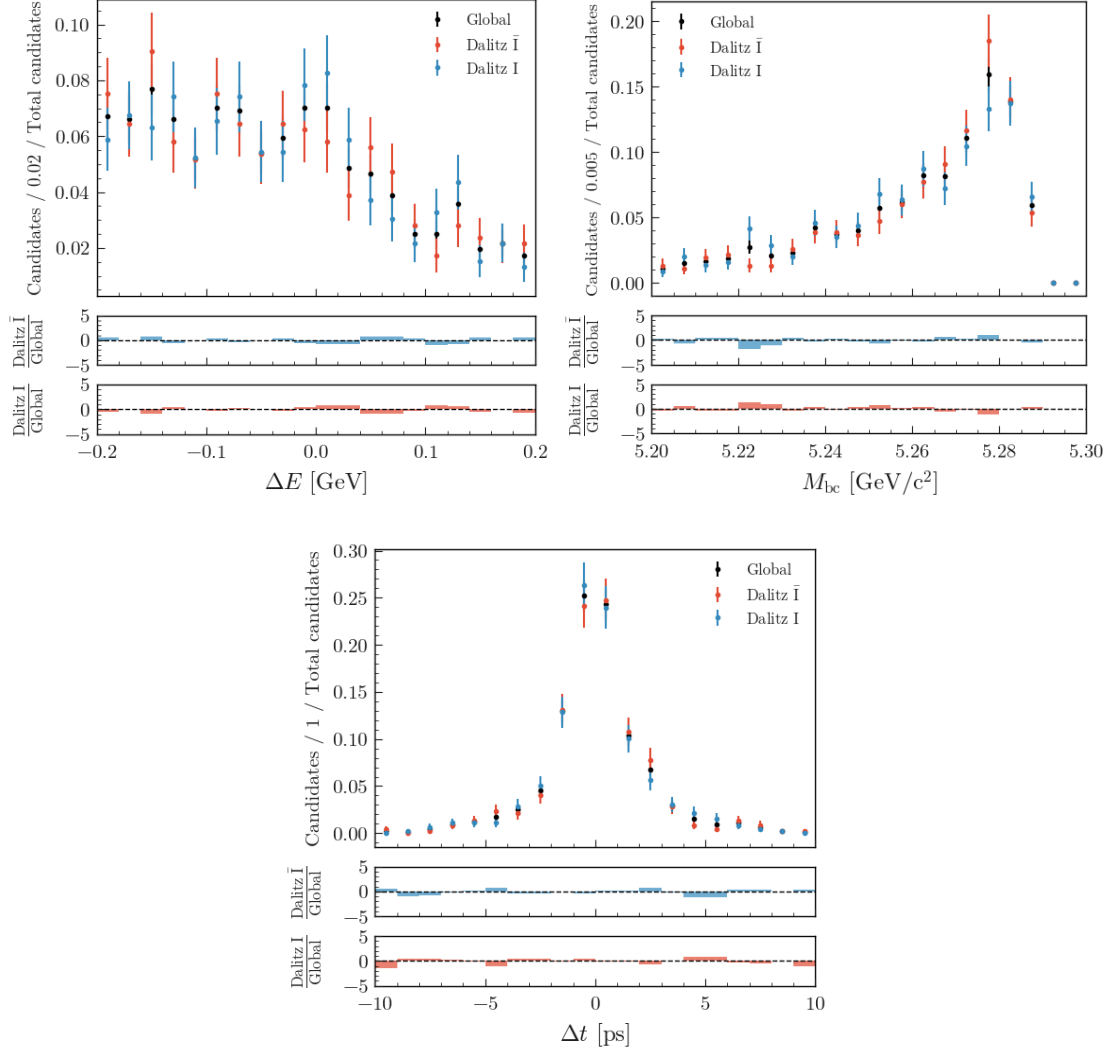


Figure D.14: Distributions of ΔE , M_{bc} and Δt of simulated SCF background events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$.

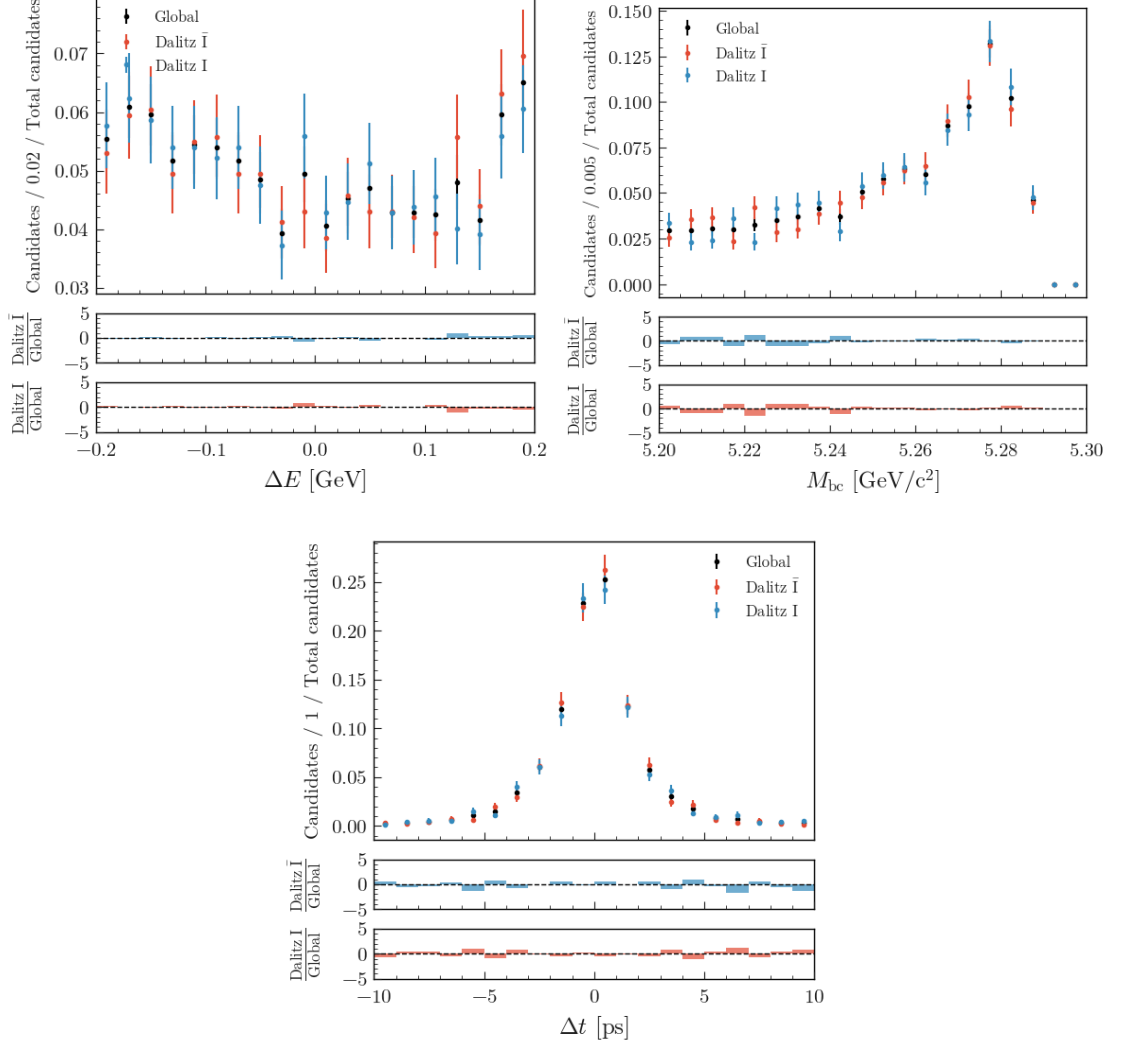


Figure D.15: Distributions of ΔE , M_{bc} and Δt of simulated $B\bar{B}$ background events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$.

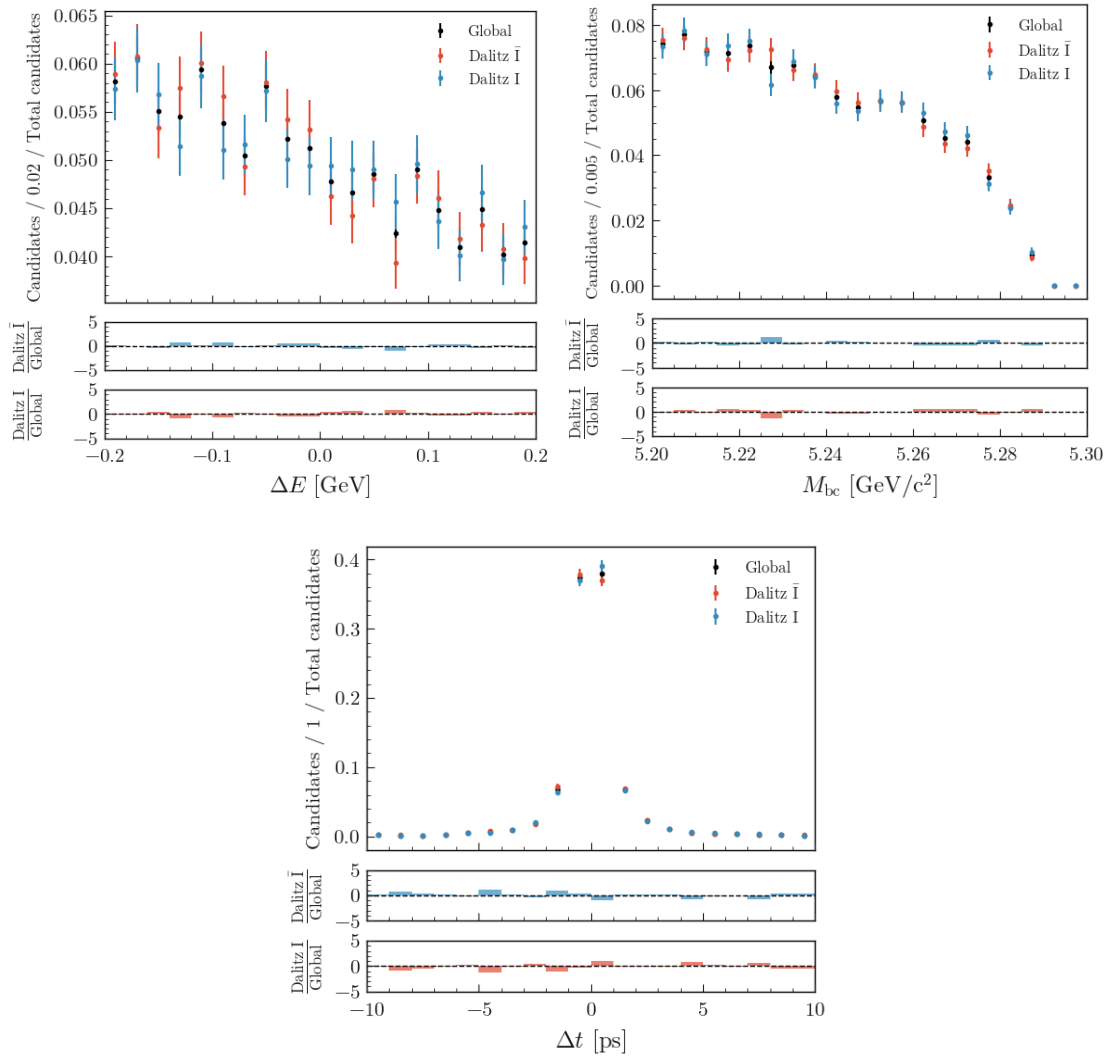


Figure D.16: Distributions of ΔE , M_{bc} and Δt of simulated continuum background events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$.

D.7 Ensemble test figures

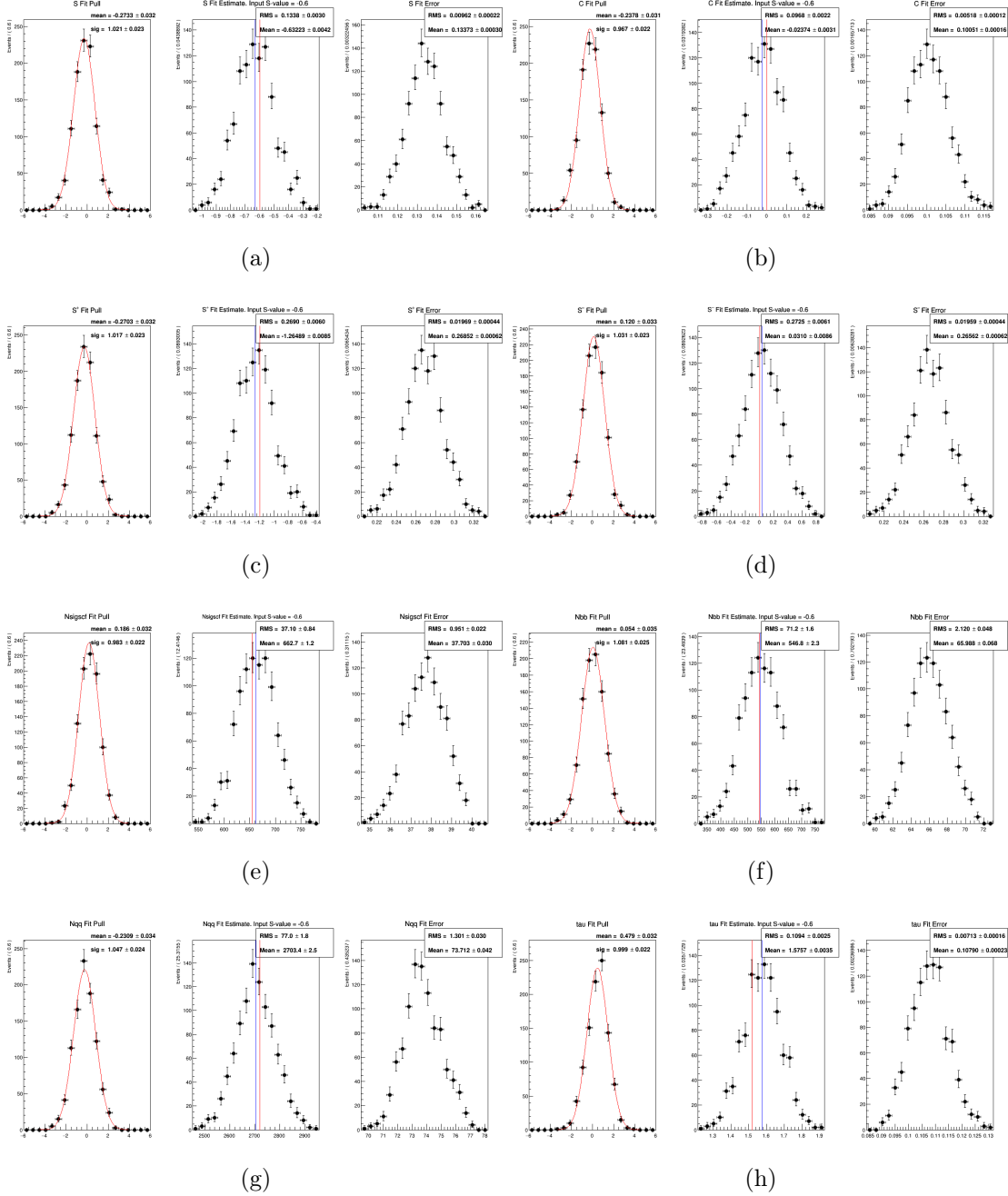


Figure D.17: Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (-0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

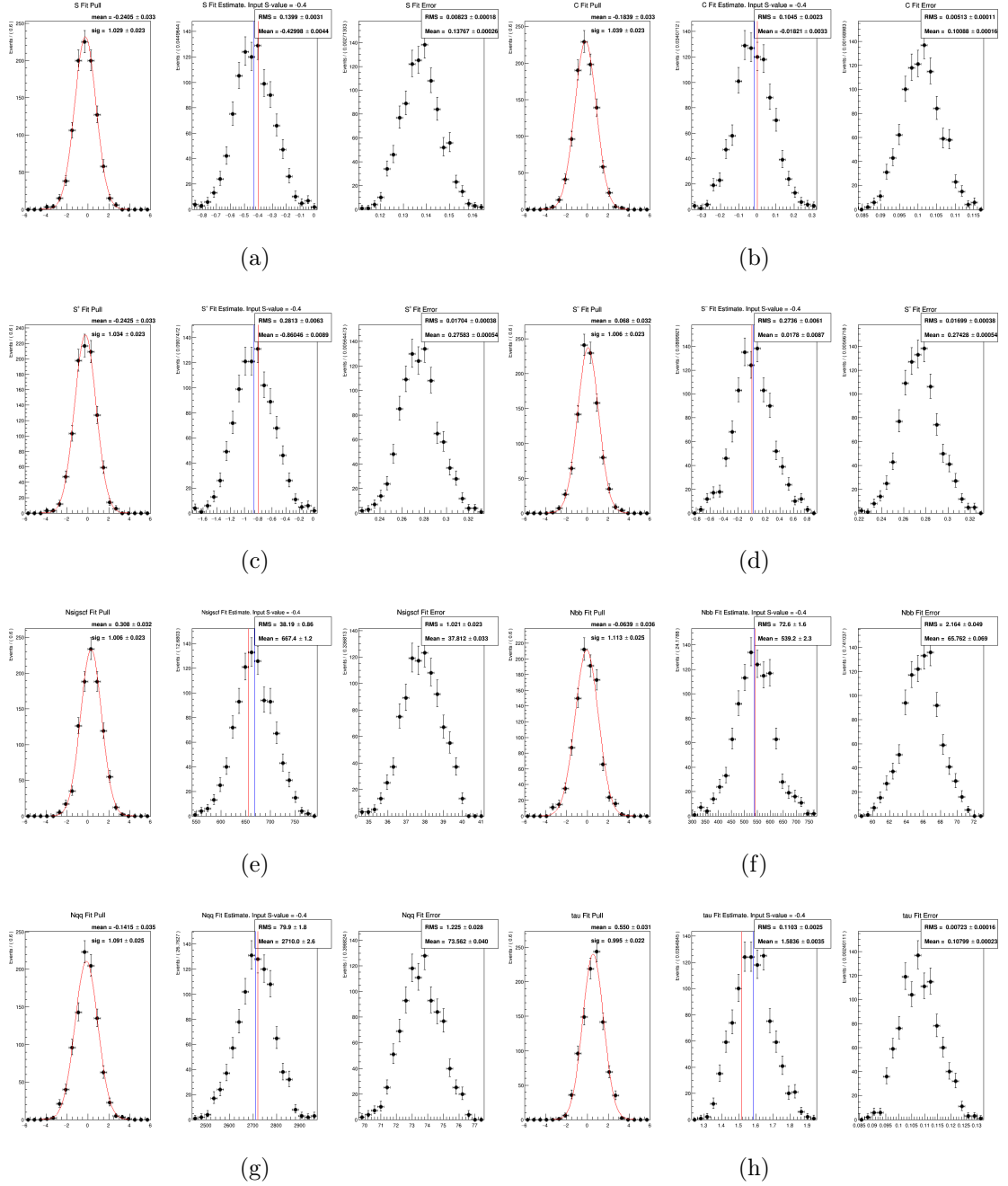


Figure D.18: Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (-0.4, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

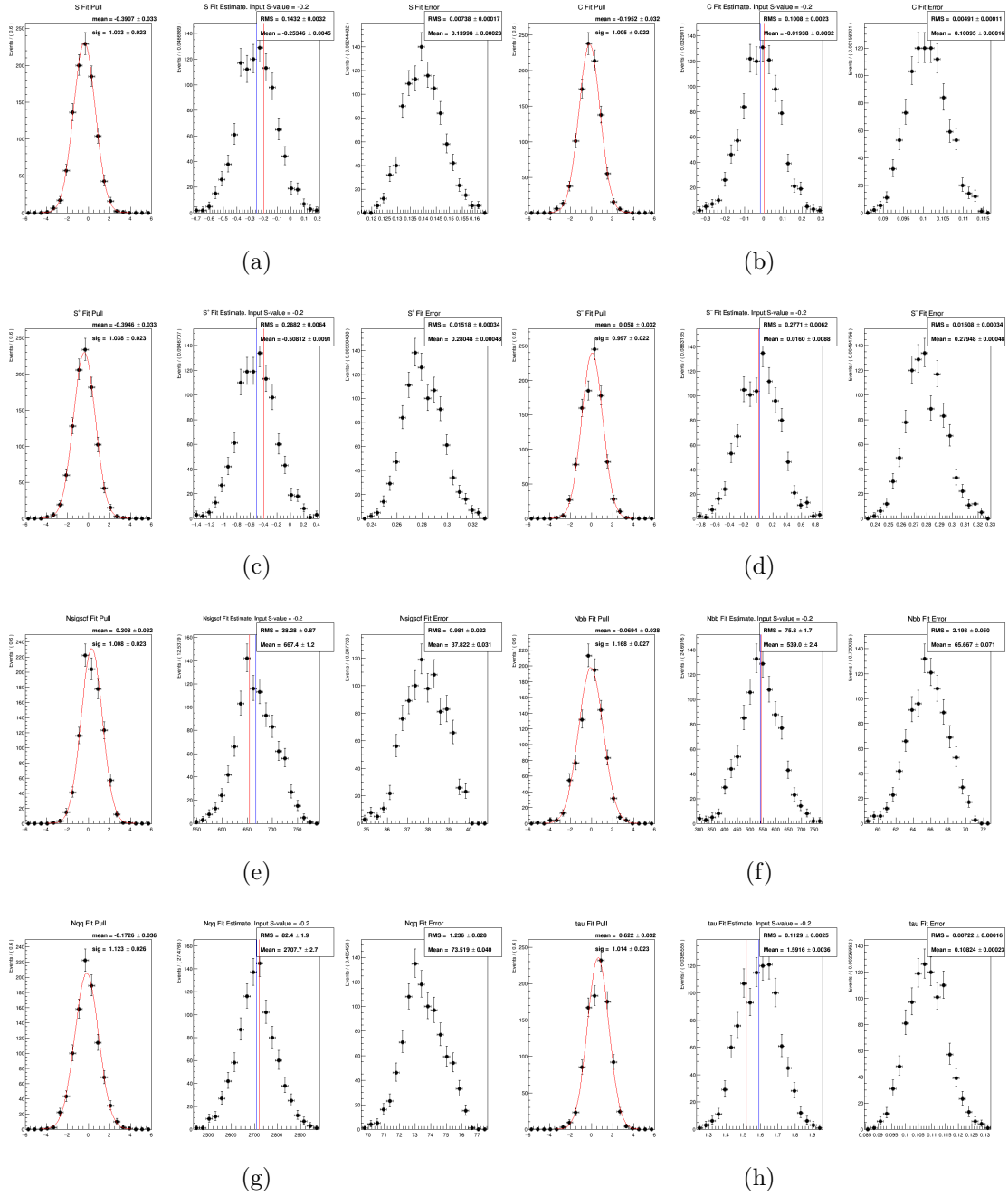


Figure D.19: Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (-0.2, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

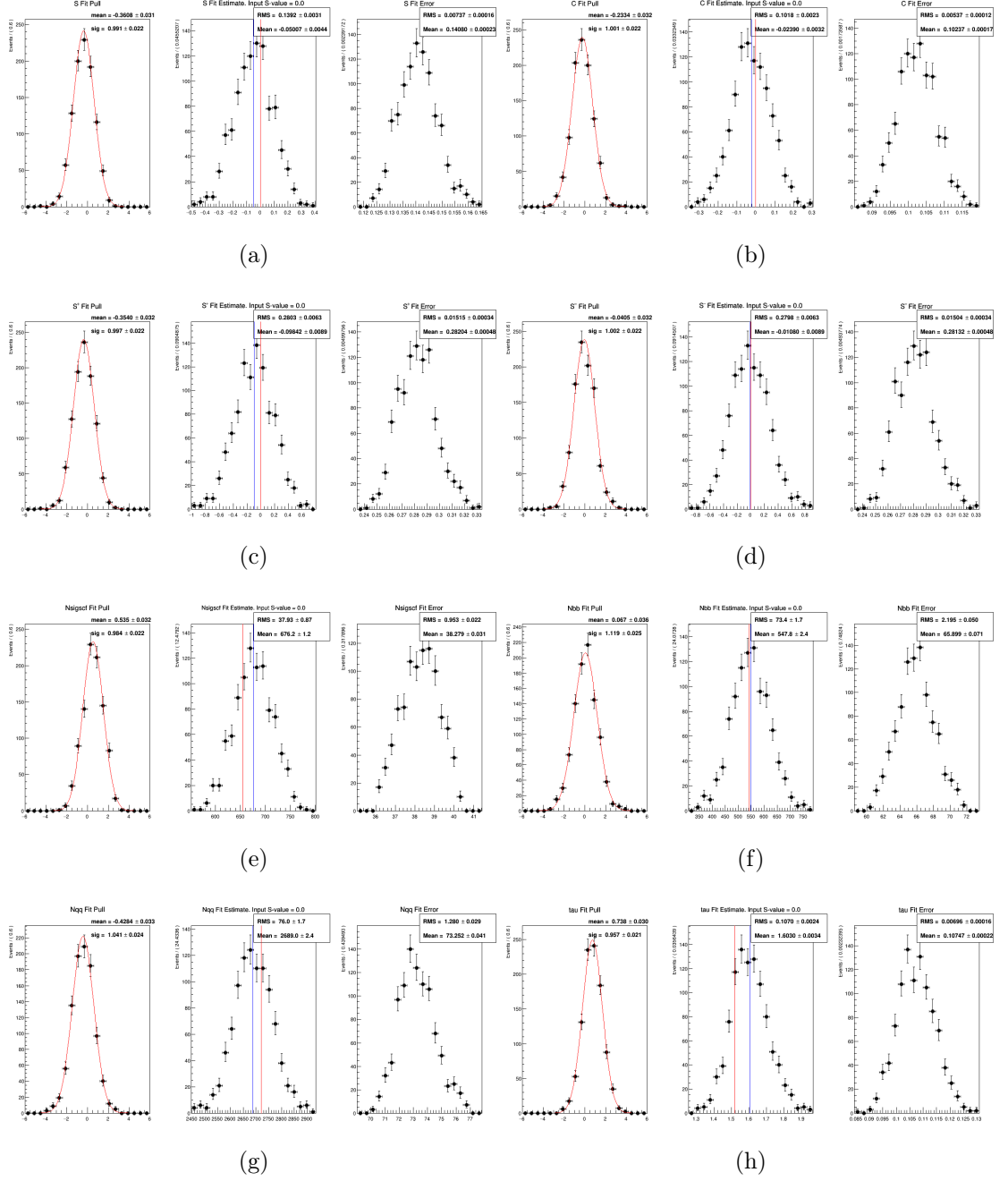


Figure D.20: Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.0, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

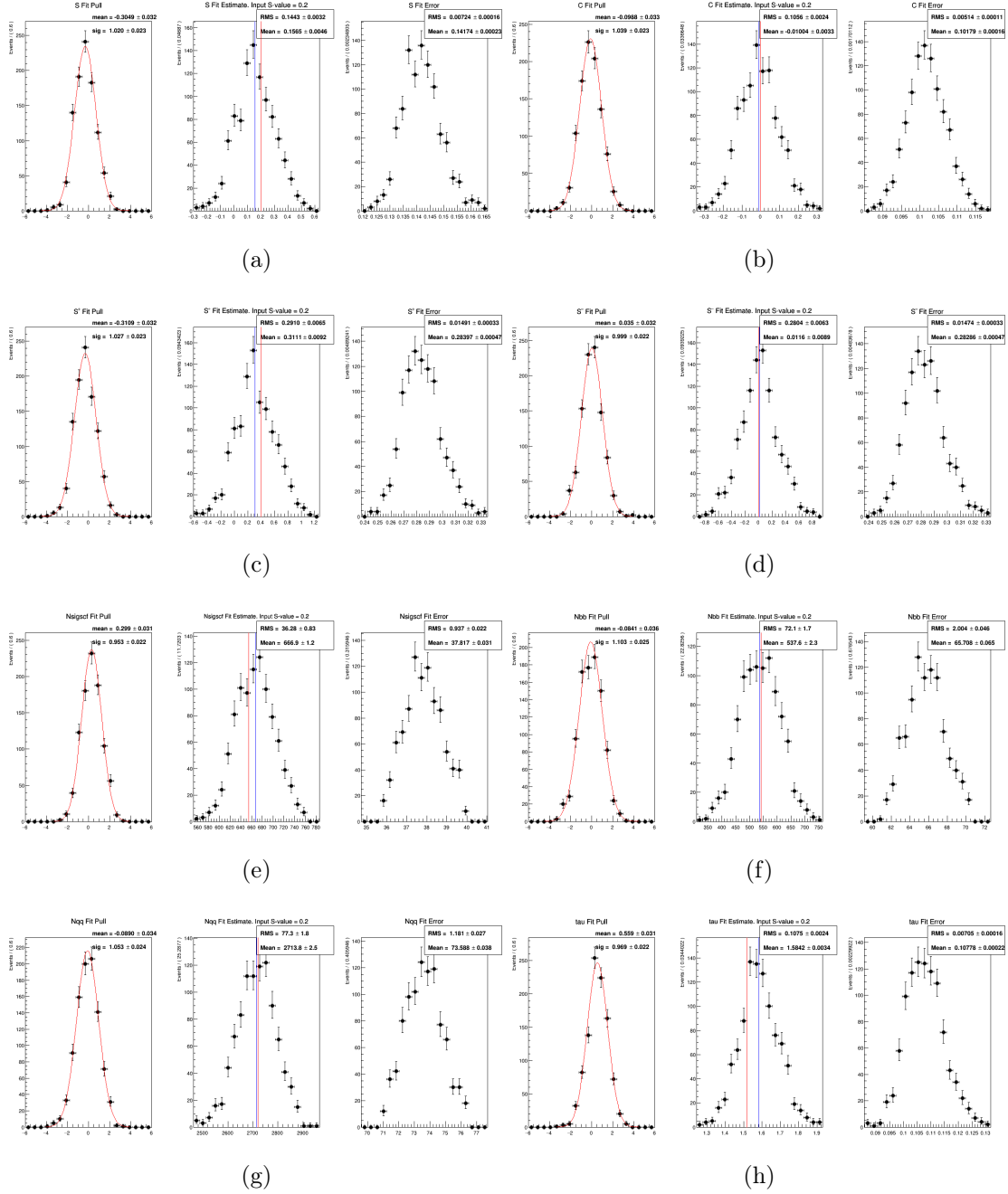


Figure D.21: Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B}\text{bkg}}$, (g) $N_{q\bar{q}\text{bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.2, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

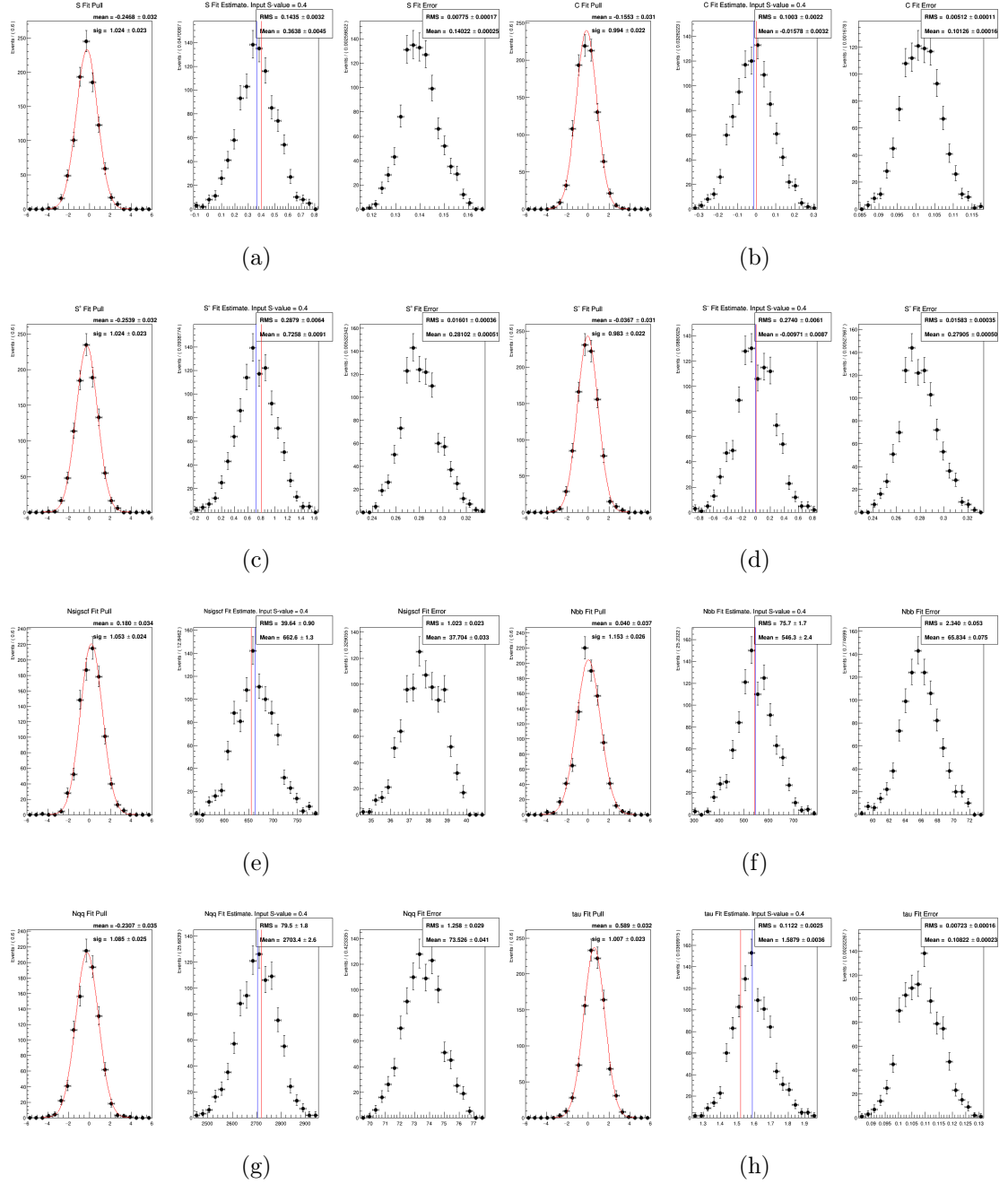


Figure D.22: Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.4, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

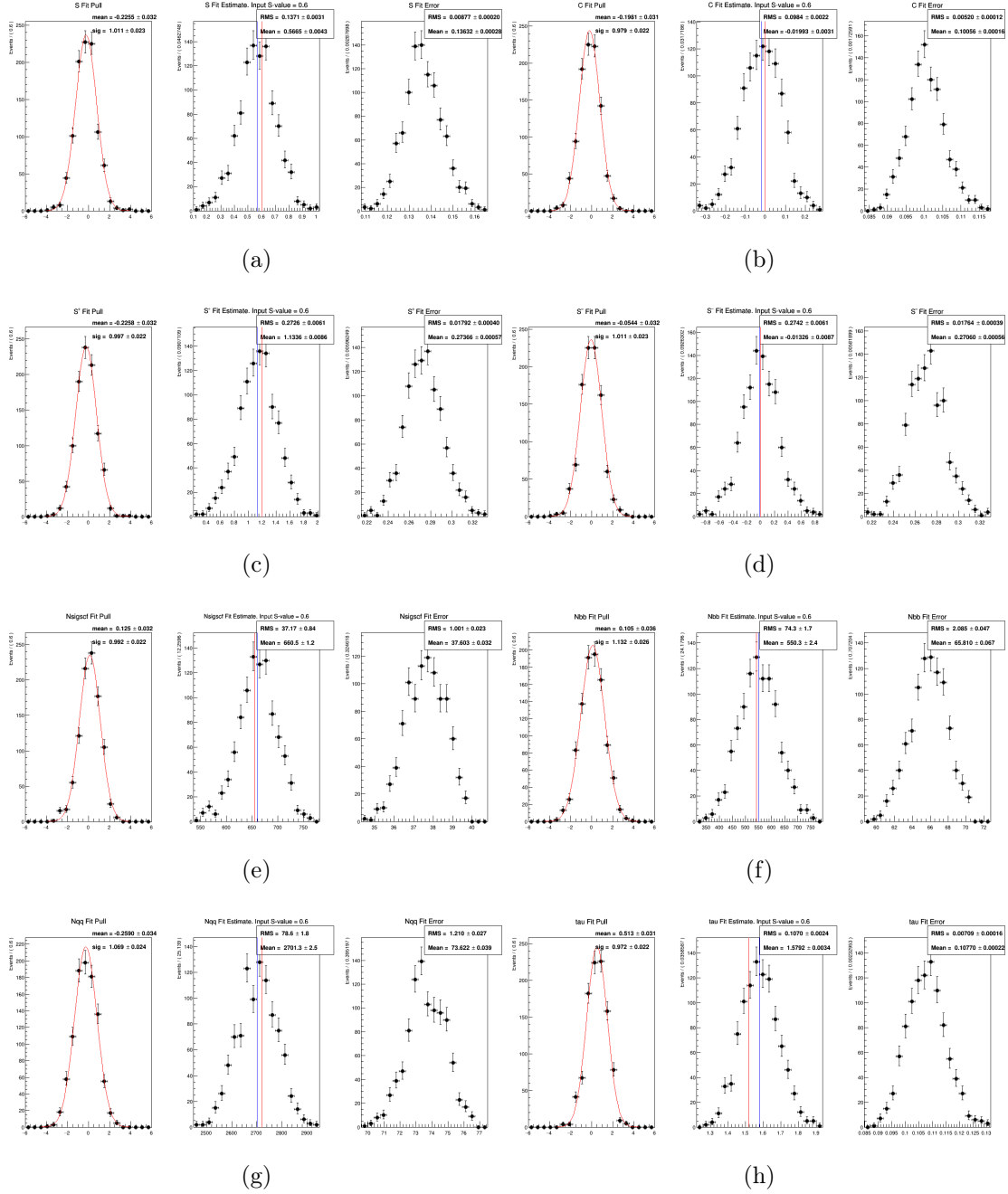


Figure D.23: Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B}\text{bkg}}$, (g) $N_{q\bar{q}\text{bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.

D.8 Component shape comparison between signal and control mode

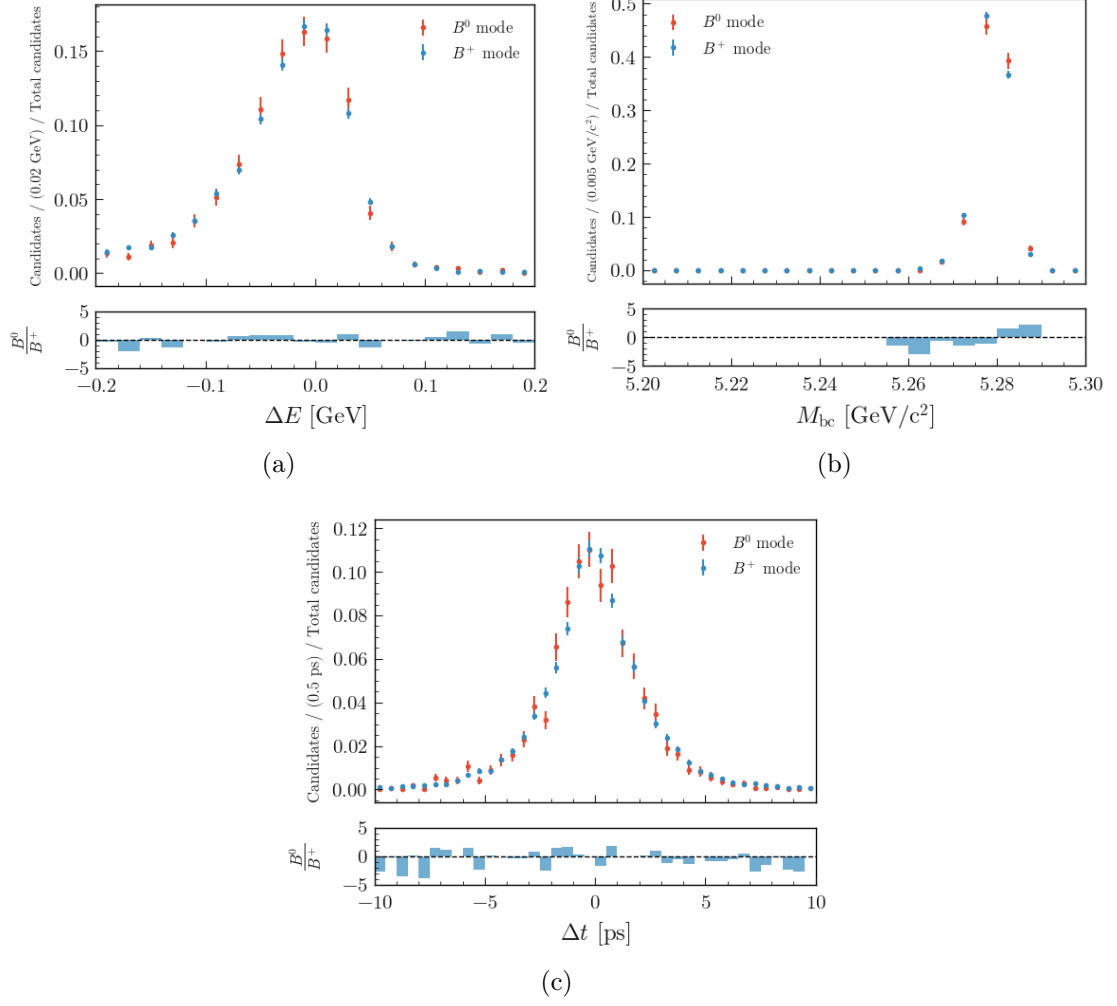


Figure D.24: Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated signal events between the signal and control mode.

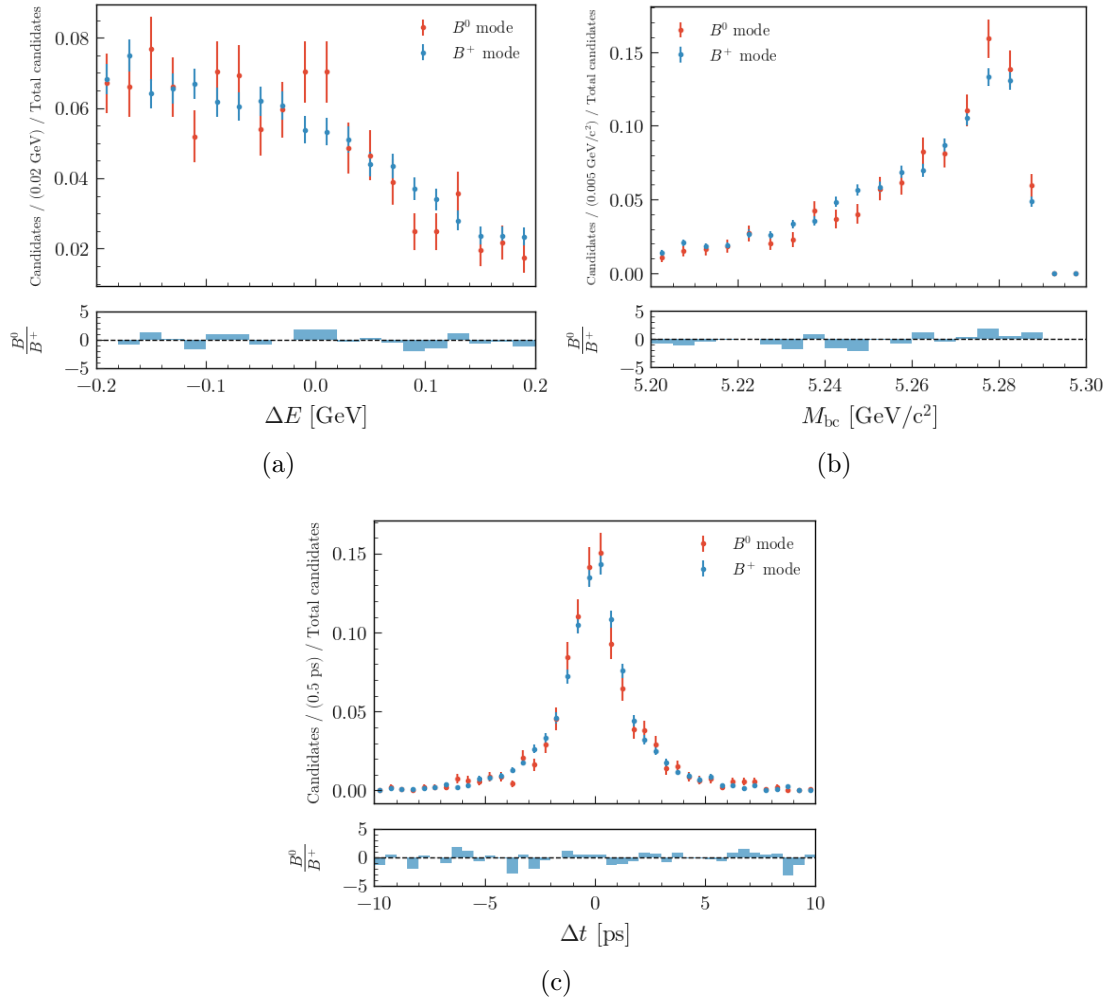


Figure D.25: Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated SCF background events between the signal and control mode.

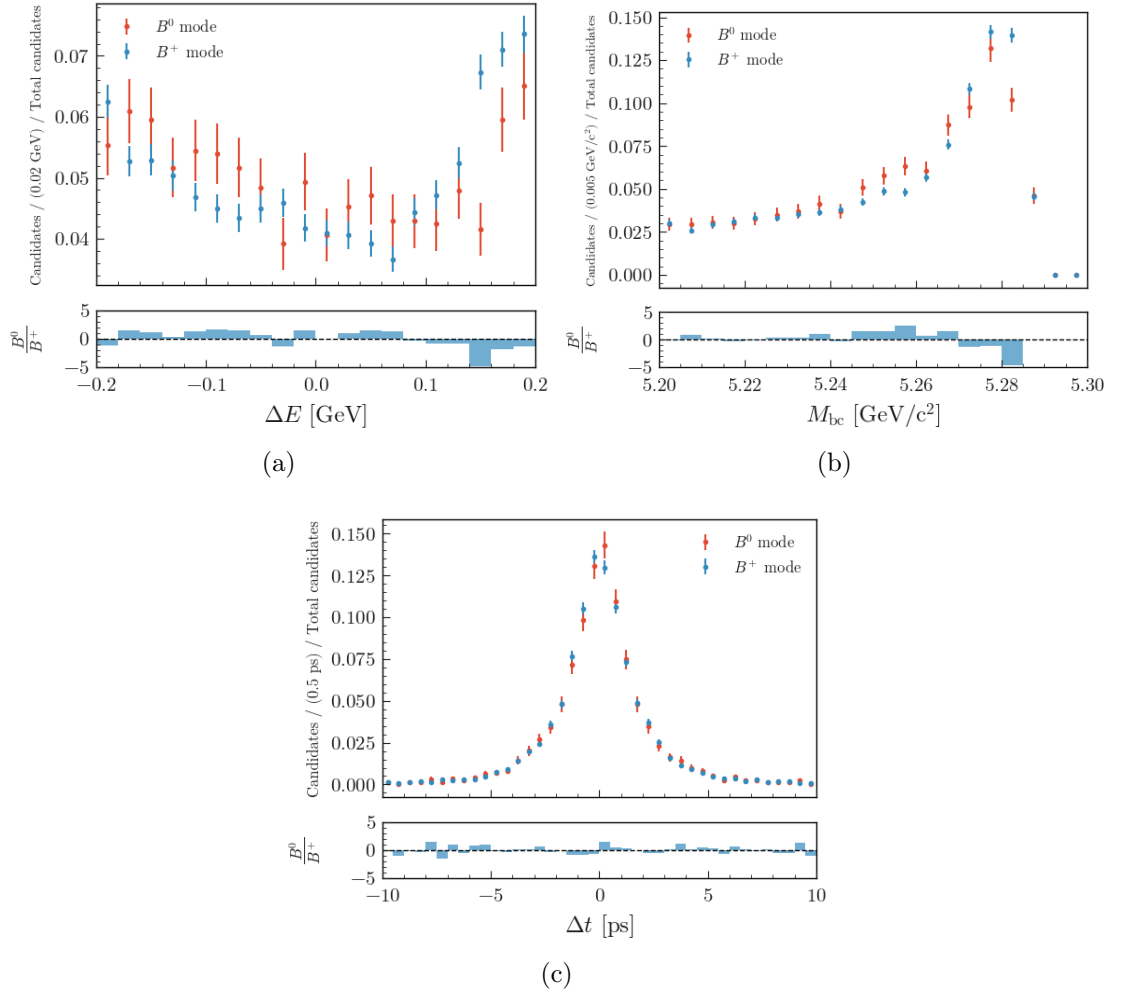


Figure D.26: Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated $B\bar{B}$ background events between the signal and control mode.

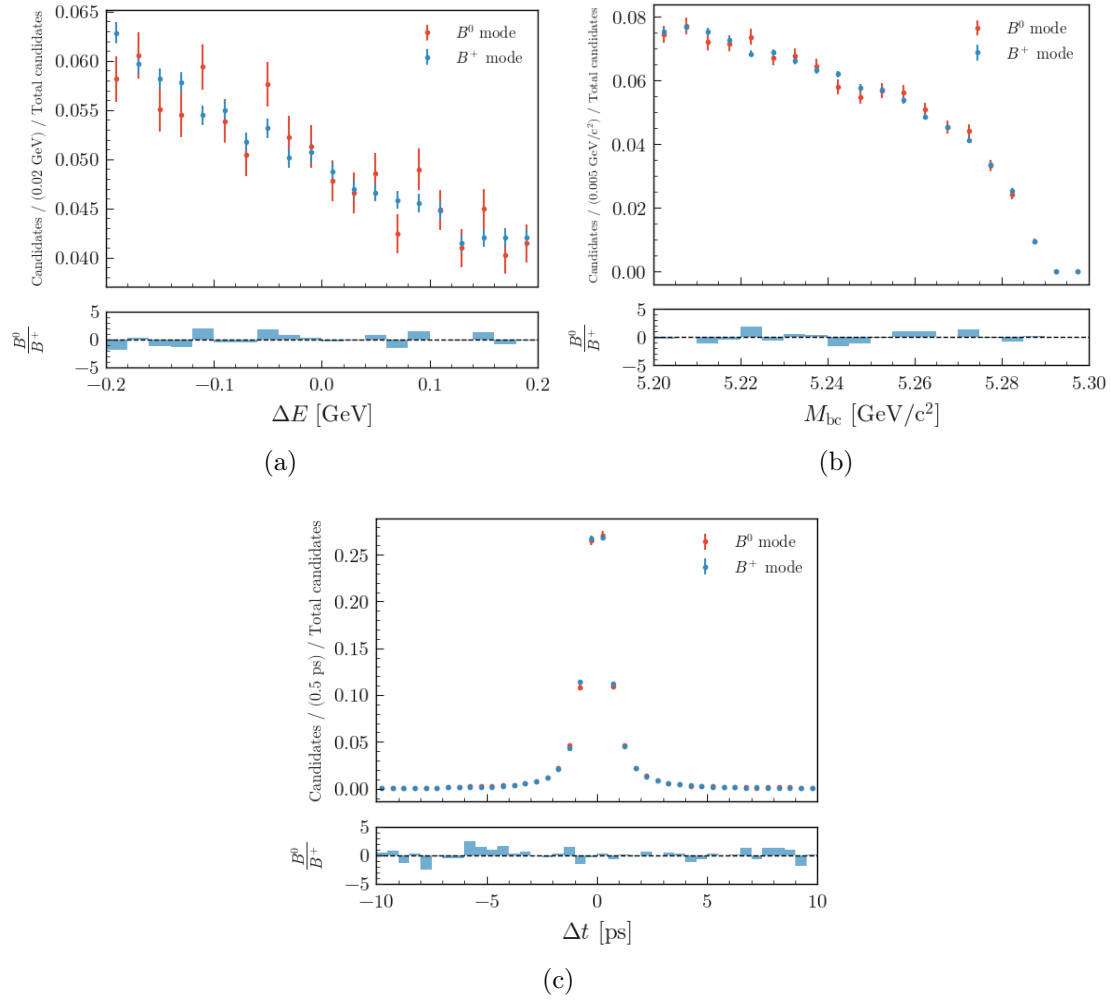


Figure D.27: Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated continuum background events between the signal and control mode.

D.9 Control mode shape fits

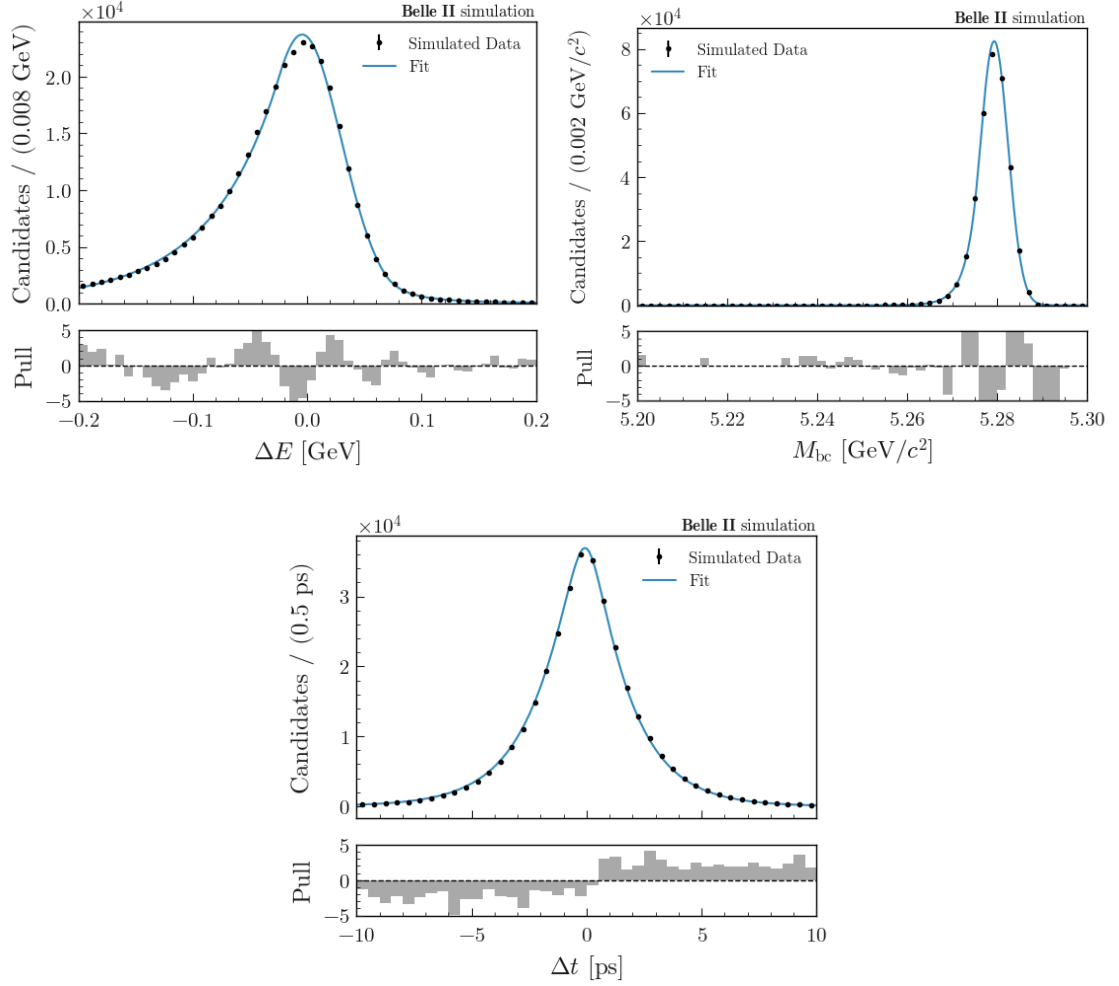


Figure D.28: Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated signal events for the control mode.

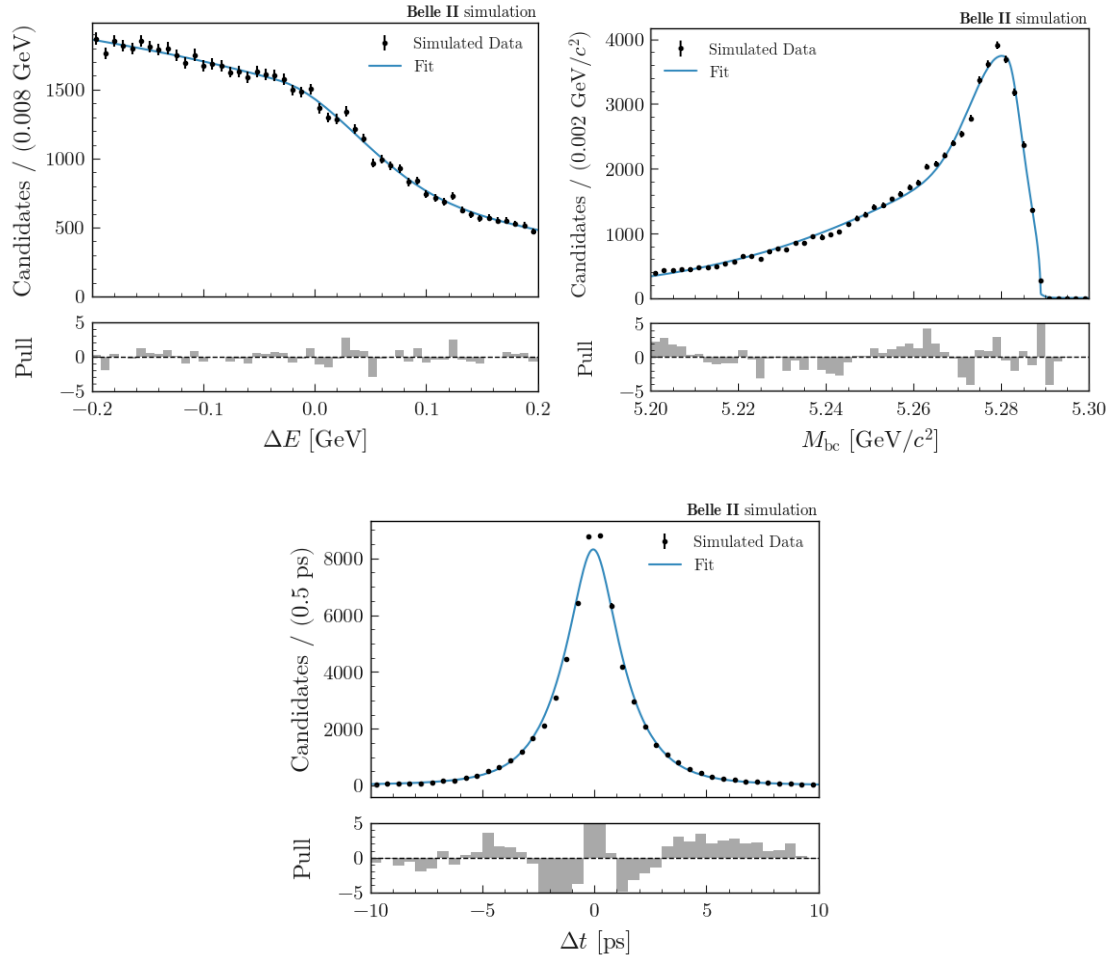


Figure D.29: Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated SCF background events for the control mode.

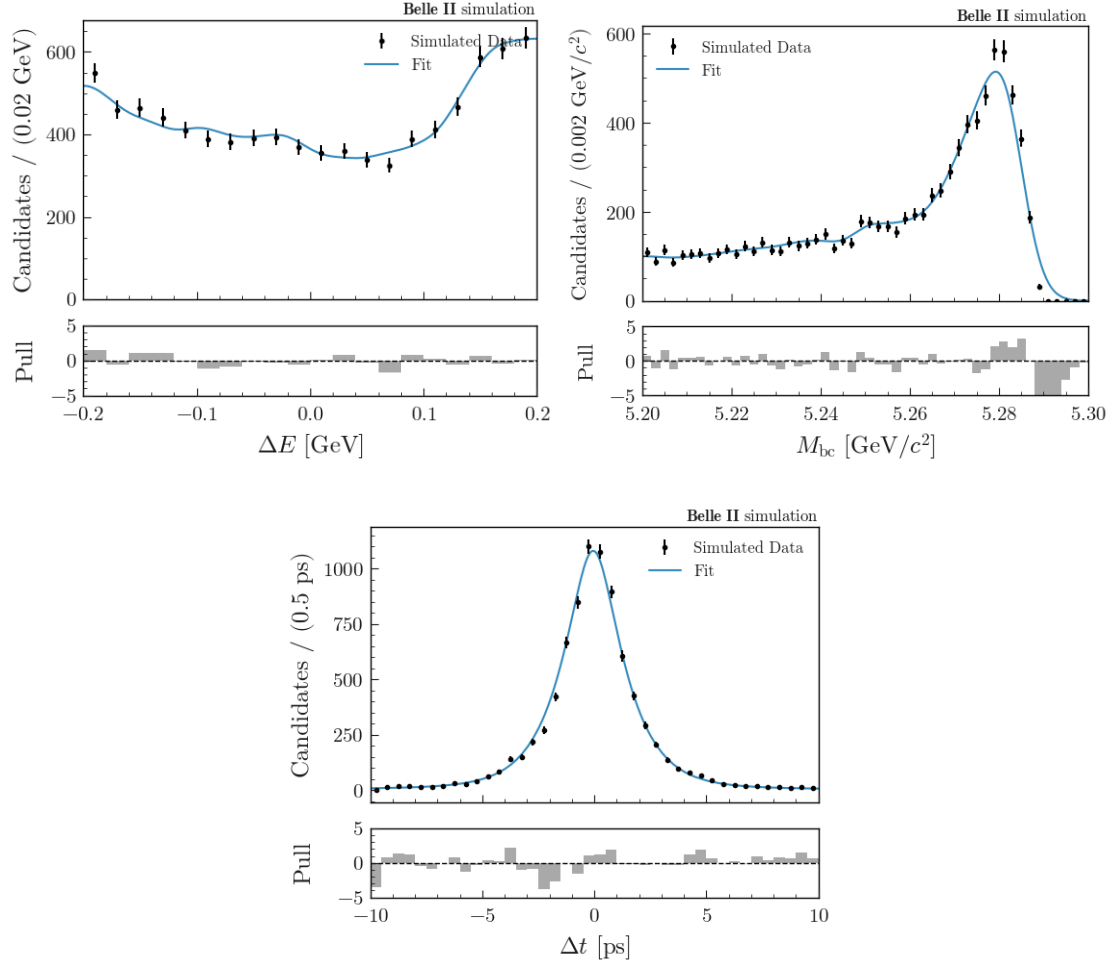


Figure D.30: Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated $B\bar{B}$ background events for the control mode. For ΔE and M_{bc} the projections of the 2D KDE are shown.

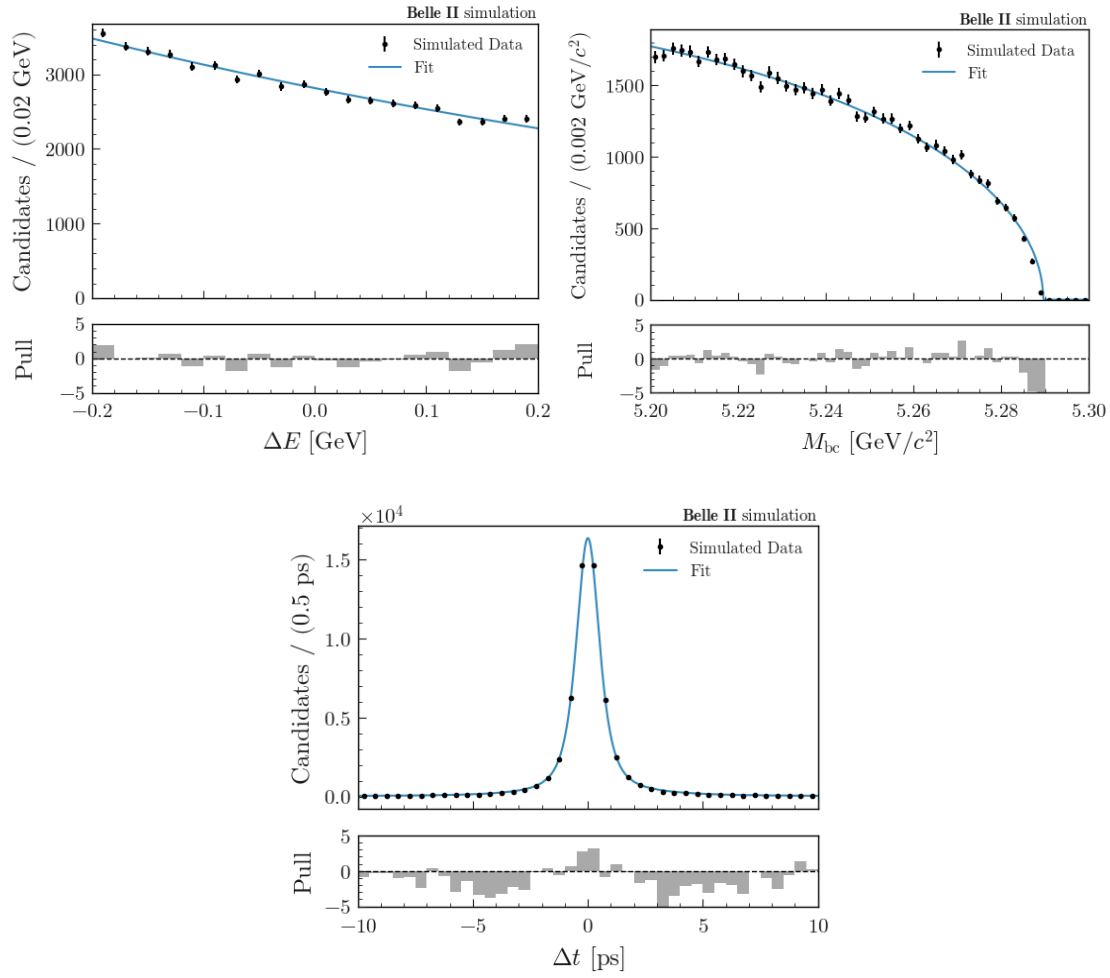


Figure D.31: Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated continuum background events for the control mode.

D.10 Comparison between toy and generic MC distributions

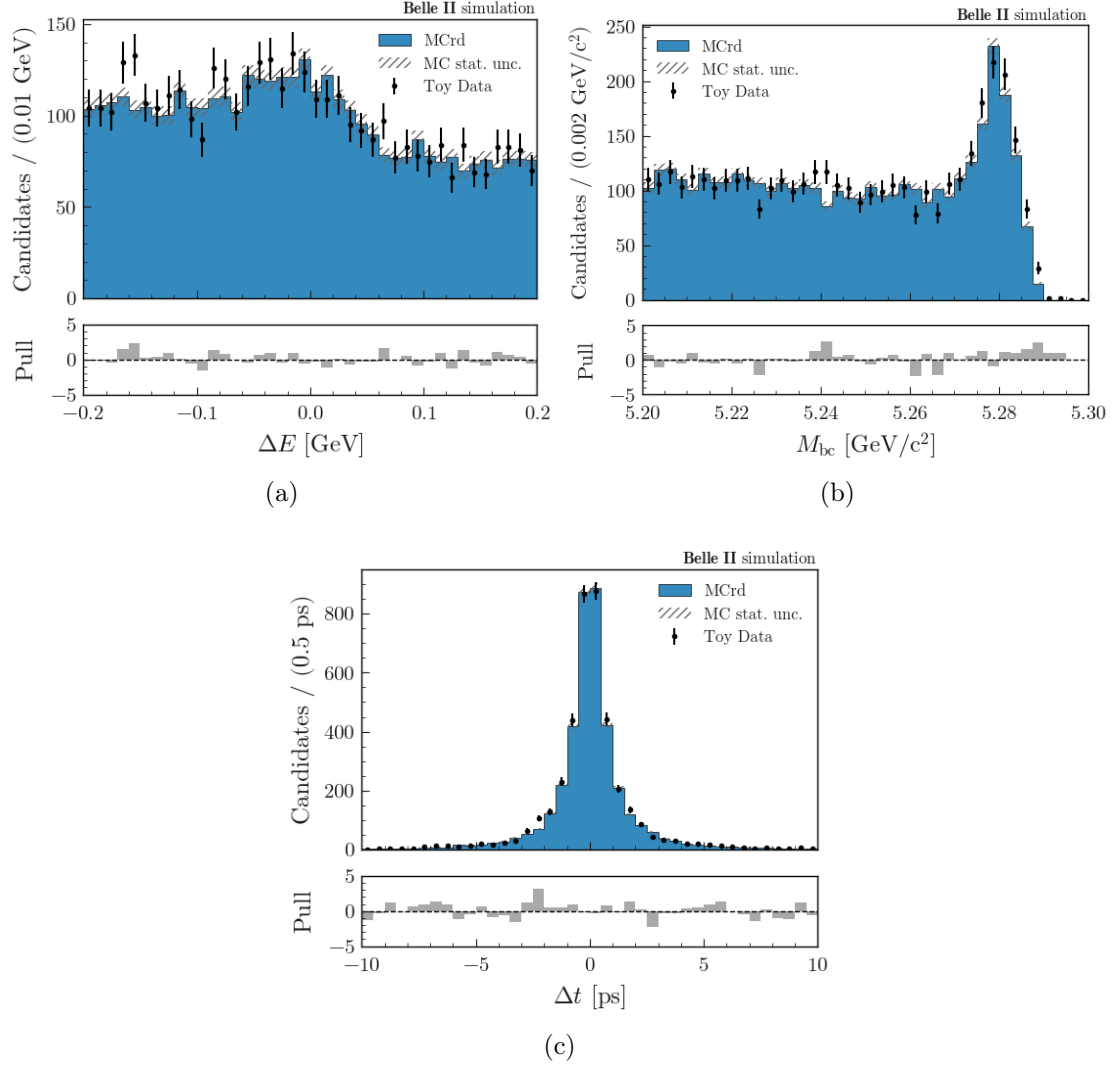


Figure D.32: Comparison between toy MC and generic MC distributions of (a) ΔE , (b) M_{bc} and (c) Δt .

Bibliography

- [1] C.-N. Yang and R. L. Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. In: *Phys. Rev.* 96 (1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191 (Cited on page 1).
- [2] T.-D. Lee and C.-N. Yang. “Question of Parity Conservation in Weak Interactions”. In: *Phys. Rev.* 104 (1956), pp. 254–258. DOI: 10.1103/PhysRev.104.254 (Cited on page 1).
- [3] S. L. Glashow. “Partial Symmetries of Weak Interactions”. In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2 (Cited on page 1).
- [4] M. Gell-Mann. “Symmetries of baryons and mesons”. In: *Phys. Rev.* 125 (1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067 (Cited on page 1).
- [5] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (Cited on pages 1, 4).
- [6] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (Cited on pages 1, 4).
- [7] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264 (Cited on pages 1, 4).
- [8] A. Salam. “Weak and Electromagnetic Interactions”. In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034 (Cited on page 1).
- [9] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* 2 (1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285 (Cited on pages 1, 6).
- [10] S. Schael et al. “Precision electroweak measurements on the Z resonance”. In: *Phys. Rept.* 427 (2006), pp. 257–454. DOI: 10.1016/j.physrep.2005.12.006. arXiv: hep-ex/0509008 (Cited on page 1).

- [11] F. Abe et al. “Observation of top quark production in $\bar{p}p$ collisions”. In: *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002 (Cited on page 1).
- [12] S. Abachi et al. “Observation of the top quark”. In: *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. arXiv: hep-ex/9503003 (Cited on page 1).
- [13] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (Cited on page 1).
- [14] S. Chatrchyan et al. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (Cited on page 1).
- [15] Rubin, V. C. and Ford, W. K. Jr. “Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions”. In: *Astrophys. J.* 159 (Feb. 1970), p. 379. DOI: {10.1086/150317} (Cited on page 1).
- [16] Zwicky, F. “On the Masses of Nebulae and of Clusters of Nebulae”. In: *Astrophys. J.* 86 (Oct. 1937), p. 217. DOI: {10.1086/143864} (Cited on page 1).
- [17] D. Clowe et al. “A direct empirical proof of the existence of dark matter”. In: *Astrophys. J. Lett.* 648 (2006), pp. L109–L113. DOI: 10.1086/508162. arXiv: astro-ph/0608407 (Cited on page 1).
- [18] N. Aghanim et al. “Planck 2018 results. VI. Cosmological parameters”. In: *Astron. Astrophys.* 641 (2020). [Erratum: *Astron. Astrophys.* 652, C4 (2021)], A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO] (Cited on page 1).
- [19] D. P. Aguillard et al. “Measurement of the Positive Muon Anomalous Magnetic Moment to 127 ppb”. In: (June 2025). arXiv: 2506.03069 [hep-ex] (Cited on page 1).
- [20] S. Mertens. “Direct Neutrino Mass Experiments”. In: *J. Phys. Conf. Ser.* 718.2 (2016), p. 022013. DOI: 10.1088/1742-6596/718/2/022013. arXiv: 1605.01579 [nucl-ex] (Cited on page 1).
- [21] M. Fael, M. Prim, and K. K. Vos. “Inclusive $B \rightarrow X_c \ell \bar{\nu}_\ell$ and $B \rightarrow X_u \ell \bar{\nu}_\ell$ decays: current status and future prospects”. In: *Eur. Phys. J. ST* 233.2 (2024), pp. 325–346. DOI: 10.1140/epjs/s11734-024-01090-w (Cited on page 1).

- [22] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Phys. Rev. Lett.* 13 (1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138 (Cited on pages [1](#), [14](#)).
- [23] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497 (Cited on page [1](#)).
- [24] E. Kou, C. D. Lü, and F.-S. Yu. “Photon Polarization in the $b \rightarrow s\gamma$ processes in the Left-Right Symmetric Model”. In: *JHEP* 12 (2013), p. 102. DOI: 10.1007/JHEP12(2013)102. arXiv: 1305.3173 [hep-ph] (Cited on pages [2](#), [18](#)).
- [25] N. Haba et al. “Search for new physics via photon polarization of $b \rightarrow s\gamma$ ”. In: *JHEP* 03 (2015), p. 160. DOI: 10.1007/JHEP03(2015)160. arXiv: 1501.00668 [hep-ph] (Cited on pages [2](#), [18](#)).
- [26] D. Atwood, M. Gronau, and A. Soni. “Mixing induced CP asymmetries in radiative B decays in and beyond the standard model”. In: *Phys. Rev. Lett.* 79 (1997), pp. 185–188. DOI: 10.1103/PhysRevLett.79.185. arXiv: hep-ph/9704272 (Cited on pages [2](#), [24](#)).
- [27] S. Akar et al. “The time-dependent CP asymmetry in $B^0 \rightarrow K_{\text{res}}\gamma \rightarrow \pi^+\pi^-K_S^0\gamma$ decays”. In: *JHEP* 09 (2019), p. 034. DOI: 10.1007/JHEP09(2019)034. arXiv: 1802.09433 [hep-ph] (Cited on pages [2](#), [26](#), [28](#), [113](#), [241](#), [242](#), [315](#)).
- [28] C.-S. Wu et al. “Experimental Test of Parity Conservation in β Decay”. In: *Phys. Rev.* 105 (1957), pp. 1413–1414. DOI: 10.1103/PhysRev.105.1413 (Cited on page [4](#)).
- [29] N. Cabibbo. “Unitary Symmetry and Leptonic Decays”. In: *Phys. Rev. Lett.* 10 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531 (Cited on page [6](#)).
- [30] M. Kobayashi and T. Maskawa. “ CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652 (Cited on page [6](#)).
- [31] L.-L. Chau and W.-Y. Keung. “Comments on the Parametrization of the Kobayashi-Maskawa Matrix”. In: *Phys. Rev. Lett.* 53 (1984), p. 1802. DOI: 10.1103/PhysRevLett.53.1802 (Cited on page [6](#)).
- [32] L. Wolfenstein. “Parametrization of the Kobayashi-Maskawa Matrix”. In: *Phys. Rev. Lett.* 51 (1983), p. 1945. DOI: 10.1103/PhysRevLett.51.1945 (Cited on page [7](#)).

- [33] J. Charles et al. “ CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories”. In: *Eur. Phys. J. C* 41.1 (2005). Updated results and plots available at <http://ckmfitter.in2p3.fr/>, pp. 1–131. DOI: 10.1140/epjc/s2005-02169-1. arXiv: hep-ph/0406184 (Cited on page 9).
- [34] M. Bona et al. “The 2004 UTfit collaboration report on the status of the unitarity triangle in the standard model”. In: *JHEP* 07 (2005). Updated results and plots available at <http://www.utfit.org/>, p. 028. DOI: 10.1088/1126-6708/2005/07/028. arXiv: hep-ph/0501199 (Cited on page 9).
- [35] K. Lande et al. “Observation of Long-Lived Neutral V Particles”. In: *Phys. Rev.* 103 (1956), pp. 1901–1904. DOI: 10.1103/PhysRev.103.1901 (Cited on page 10).
- [36] M. Staric et al. “Evidence for $D^0 - \bar{D}^0$ Mixing”. In: *Phys. Rev. Lett.* 98 (2007). Ed. by Etienne Augé, Bolek Pietrzyk, and Jean Trân Thanh Vân, p. 211803. DOI: 10.1103/PhysRevLett.98.211803. arXiv: hep-ex/0703036 (Cited on page 10).
- [37] B. Aubert et al. “Evidence for $D^0 - \bar{D}^0$ Mixing”. In: *Phys. Rev. Lett.* 98 (2007), p. 211802. DOI: 10.1103/PhysRevLett.98.211802. arXiv: hep-ex/0703020 (Cited on page 10).
- [38] R. Aaij et al. “Measurement of $D^0 - \bar{D}^0$ Mixing Parameters and Search for CP Violation Using $D^0 \rightarrow K^+\pi^-$ Decays”. In: *Phys. Rev. Lett.* 111.25 (2013), p. 251801. DOI: 10.1103/PhysRevLett.111.251801. arXiv: 1309.6534 [hep-ex] (Cited on page 10).
- [39] H. Albrecht et al. “Observation of $B^0 - \bar{B}^0$ Mixing”. In: *Phys. Lett. B* 192 (1987). Ed. by J. Tran Thanh Van, pp. 245–252. DOI: 10.1016/0370-2693(87)91177-4 (Cited on page 10).
- [40] A. Abulencia et al. “Measurement of the $B_s^0 - \bar{B}_s^0$ Oscillation Frequency”. In: *Phys. Rev. Lett.* 97 (2006), p. 062003. DOI: 10.1103/PhysRevLett.97.062003. arXiv: hep-ex/0606027 (Cited on page 10).
- [41] R. Aaij et al. “Observation of $B_s^0 - \bar{B}_s^0$ mixing and measurement of mixing frequencies using semileptonic B decays”. In: *Eur. Phys. J. C* 73.12 (2013), p. 2655. DOI: 10.1140/epjc/s10052-013-2655-8. arXiv: 1308.1302 [hep-ex] (Cited on page 10).
- [42] S. Banerjee et al. “Averages of b -hadron, c -hadron, and τ -lepton properties as of 2023”. In: (Nov. 2024). arXiv: 2411.18639 [hep-ex] (Cited on pages 12, 18, 240).
- [43] A. B. Carter and A. I. Sanda. “ CP Violation in B Meson Decays”. In: *Phys. Rev. D* 23 (1981), p. 1567. DOI: 10.1103/PhysRevD.23.1567 (Cited on page 15).

- [44] I. I. Y. Bigi and A. I. Sanda. “Notes on the Observability of CP Violations in B Decays”. In: *Nucl. Phys. B* 193 (1981), pp. 85–108. DOI: 10.1016/0550-3213(81)90519-8 (Cited on page 15).
- [45] A. J. Bevan et al. “The physics of the B factories”. In: *Eur. Phys. J. C* 74 (2014), p. 3026. DOI: 10.1140/epjc/s10052-014-3026-9. arXiv: 1406.6311 [hep-ex] (Cited on pages 15, 117).
- [46] John R. Ellis et al. “The Phenomenology of the Next Left-Handed Quarks”. In: *Nucl. Phys. B* 131 (1977). [Erratum: Nucl.Phys.B 132, 541 (1978)], p. 285. DOI: 10.1016/0550-3213(77)90374-1 (Cited on page 17).
- [47] M. Misiak, A. Rehman, and M. Steinhauser. “NNLO QCD counterterm contributions to $\bar{B} \rightarrow X_{s\gamma}$ for the physical value of m_c ”. In: *Phys. Lett. B* 770 (2017), pp. 431–439. DOI: 10.1016/j.physletb.2017.05.008. arXiv: 1702.07674 [hep-ph] (Cited on page 18).
- [48] T. Inami and C. S. Lim. “Effects of Superheavy Quarks and Leptons in Low-Energy Weak Processes $K_L^0 \rightarrow \mu^- \mu^+$, $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ and $K^0 \leftrightarrow \bar{K}^0$ ”. In: *Prog. Theor. Phys.* 65 (1981). [Erratum: Prog.Theor.Phys. 65, 1772 (1981)], p. 297. DOI: 10.1143/PTP.65.297 (Cited on page 19).
- [49] S. W. Bosch and G. Buchalla. “Constraining the unitarity triangle with $B \rightarrow V\gamma$ ”. In: *JHEP* 01 (2005), p. 035. DOI: 10.1088/1126-6708/2005/01/035. arXiv: hep-ph/0408231 (Cited on page 21).
- [50] M. Matsumori and A. I. Sanda. “The Mixing-induced CP asymmetry in $B \rightarrow K^* \gamma$ decays with perturbative QCD approach”. In: *Phys. Rev. D* 73 (2006), p. 114022. DOI: 10.1103/PhysRevD.73.114022. arXiv: hep-ph/0512175 (Cited on page 21).
- [51] B. Grinstein et al. “The Photon polarization in $B \rightarrow X\gamma$ in the standard model”. In: *Phys. Rev. D* 71 (2005), p. 011504. DOI: 10.1103/PhysRevD.71.011504. arXiv: hep-ph/0412019 (Cited on page 21).
- [52] B. Grinstein and D. Pirjol. “The CP asymmetry in $B^0(t) \rightarrow K_S^0 \pi^0 \gamma$ in the standard model”. In: *Phys. Rev. D* 73 (2006), p. 014013. DOI: 10.1103/PhysRevD.73.014013. arXiv: hep-ph/0510104 (Cited on page 21).
- [53] R. N. Mohapatra and J. C. Pati. “A Natural Left-Right Symmetry”. In: *Phys. Rev. D* 11 (1975), p. 2558. DOI: 10.1103/PhysRevD.11.2558 (Cited on page 22).
- [54] G. Senjanovic and R. N. Mohapatra. “Exact Left-Right Symmetry and Spontaneous Violation of Parity”. In: *Phys. Rev. D* 12 (1975), p. 1502. DOI: 10.1103/PhysRevD.12.1502 (Cited on page 22).

- [55] H. P. Nilles. “Supersymmetry, Supergravity and Particle Physics”. In: *Phys. Rept.* 110 (1984), pp. 1–162. DOI: 10.1016/0370-1573(84)90008-5 (Cited on page 22).
- [56] H. E. Haber and G. L. Kane. “The Search for Supersymmetry: Probing Physics Beyond the Standard Model”. In: *Phys. Rept.* 117 (1985), pp. 75–263. DOI: 10.1016/0370-1573(85)90051-1 (Cited on page 22).
- [57] F. Gabbiani et al. “A Complete analysis of FCNC and CP constraints in general SUSY extensions of the standard model”. In: *Nucl. Phys. B* 477 (1996), pp. 321–352. DOI: 10.1016/0550-3213(96)00390-2. arXiv: hep-ph/9604387 (Cited on page 22).
- [58] M. Gronau and D. Pirjol. “Photon polarization in radiative B decays”. In: *Phys. Rev. D* 66 (2002), p. 054008. DOI: 10.1103/PhysRevD.66.054008. arXiv: hep-ph/0205065 (Cited on page 23).
- [59] M. Gronau and D. Pirjol. “Reexamining the photon polarization in $B \rightarrow K\pi\pi\gamma$ ”. In: *Phys. Rev. D* 96.1 (2017), p. 013002. DOI: 10.1103/PhysRevD.96.013002. arXiv: 1704.05280 [hep-ph] (Cited on page 23).
- [60] R. Aaij et al. “Strong constraints on the $b \rightarrow s\gamma$ photon polarisation from $B^0 \rightarrow K^{*0}e^+e^-$ decays”. In: *JHEP* 12 (2020), p. 081. DOI: 10.1007/JHEP12(2020)081. arXiv: 2010.06011 [hep-ex] (Cited on page 23).
- [61] R. Aaij et al. “Measurement of the photon polarization in $\Lambda_b^0 \rightarrow \Lambda \gamma$ decays”. In: *Phys. Rev. D* 105.5 (2022), p. L051104. DOI: 10.1103/PhysRevD.105.L051104. arXiv: 2111.10194 [hep-ex] (Cited on page 23).
- [62] Y. Ushiroda et al. “Time-Dependent CP Asymmetries in $B^0 \rightarrow K_S^0\pi^0\gamma$ transitions”. In: *Phys. Rev. D* 74 (2006). Ed. by Alexey Sissakian, Gennady Kozlov, and Elena Kolganova, p. 111104. DOI: 10.1103/PhysRevD.74.111104. arXiv: hep-ex/0608017 (Cited on page 25).
- [63] B. Aubert et al. “Measurement of Time-Dependent CP Asymmetry in $B^0 \rightarrow K_S^0\pi^0\gamma$ Decays”. In: *Phys. Rev. D* 78 (2008), p. 071102. DOI: 10.1103/PhysRevD.78.071102. arXiv: 0807.3103 [hep-ex] (Cited on page 25).
- [64] I. Adachi et al. “Measurement of CP Asymmetries in $B^0 \rightarrow K_S^0\pi^0\gamma$ Decays at Belle II”. In: *Phys. Rev. Lett.* 134.1 (2025), p. 011802. DOI: 10.1103/PhysRevLett.134.011802. arXiv: 2407.09139 [hep-ex] (Cited on page 25).
- [65] J. Li et al. “Time-dependent CP Asymmetries in $B^0 \rightarrow K_S^0\rho^0\gamma$ Decays”. In: *Phys. Rev. Lett.* 101 (2008), p. 251601. DOI: 10.1103/PhysRevLett.101.251601. arXiv: 0806.1980 [hep-ex] (Cited on pages 25, 26, 232, 240).

- [66] P. del Amo Sanchez et al. “Time-dependent analysis of $B^0 \rightarrow K_S^0 \pi^- \pi^+ \gamma$ decays and studies of the $K^+ \pi^- \pi^+$ system in $B^+ \rightarrow K^+ \pi^- \pi^+ \gamma$ decays”. In: *Phys. Rev. D* 93.5 (2016), p. 052013. DOI: 10.1103/PhysRevD.93.052013. arXiv: 1512.03579 [hep-ex] (Cited on pages 25, 26, 232, 240).
- [67] R. Aaij et al. “Angular analysis of the $B^0 \rightarrow K^{*0} e^+ e^-$ decay in the low- q^2 region”. In: *JHEP* 04 (2015), p. 064. DOI: 10.1007/JHEP04(2015)064. arXiv: 1501.03038 [hep-ex] (Cited on pages 28, 242).
- [68] K. Akai, K. Furukawa, and H. Koiso. “SuperKEKB collider”. In: *Nucl. Instrum. Meth. A* 907 (2018), p. 188. DOI: 10.1016/j.nima.2018.08.017. arXiv: 1809.01958 [physics.acc-ph] (Cited on page 30).
- [69] T. Abe et al. “Belle II technical design report”. 2010. arXiv: 1011.0352 [physics.ins-det]. URL: <https://arxiv.org/abs/1011.0352> (Cited on pages 30, 38, 40, 42, 46, 47, 58).
- [70] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001 (Cited on pages 31, 78, 227).
- [71] W. Altmannshofer et al. “The Belle II physics book”. In: *PTEP* 2019.12 (2019). Ed. by E. Kou and P. Urquijo, p. 123C01. DOI: 10.1093/ptep/ptz106. arXiv: 1808.10567 [hep-ex] (Cited on pages 31, 32).
- [72] I. Adachi et al. “Measurement of the integrated luminosity of data samples collected during 2019-2022 by the Belle II experiment*”. In: *Chin. Phys. C* 49.1 (2025), p. 013001. DOI: 10.1088/1674-1137/ad806c. arXiv: 2407.00965 [hep-ex] (Cited on page 32).
- [73] D. Boutigny et al. *The BABAR physics book: Physics at an asymmetric B factory*. Oct. 1998. DOI: 10.2172/979931 (Cited on page 32).
- [74] J. Kemmer et al. “Low Capacity Drift Diode”. In: *Nucl. Instrum. Meth. A* 253 (1987), pp. 378–381. DOI: 10.1016/0168-9002(87)90519-5 (Cited on page 37).
- [75] M. J. French et al. “Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker”. In: *Nucl. Instrum. Meth. A* 466 (2001). Ed. by T. Ohsugi et al., pp. 359–365. DOI: 10.1016/S0168-9002(01)00589-7 (Cited on page 37).
- [76] K. Adamczyk et al. “The design, construction, operation and performance of the Belle II silicon vertex detector”. In: *JINST* 17.11 (2022), P11042. DOI: 10.1088/1748-0221/17/11/P11042. arXiv: 2201.09824 [physics.ins-det] (Cited on pages 38, 58).

- [77] B. Schwartz. “Electromagnetic calorimeter of the Belle II detector”. In: *J. Phys. Conf. Ser.* 928.1 (2017), p. 012021. DOI: 10.1088/1742-6596/928/1/012021 (Cited on page 44).
- [78] F. Wemmer et al. “Photon Reconstruction in the Belle II Calorimeter Using Graph Neural Networks”. In: *Comput. Softw. Big Sci.* 7.1 (2023), p. 13. DOI: 10.1007/s41781-023-00105-w. arXiv: 2306.04179 [hep-ex] (Cited on page 44).
- [79] T. Kuhr et al. “The Belle II Core Software”. In: *Comput. Softw. Big Sci.* 3.1 (2019), p. 1. DOI: 10.1007/s41781-018-0017-9. arXiv: 1809.04299 [physics.comp-ph] (Cited on page 48).
- [80] ISO. *ISO/IEC 14882:2024 Programming languages – C++*. 7th. pub-ISO, Oct. 2024, p. 2104. URL: <https://www.iso.org/standard/83626.html> (Cited on page 48).
- [81] G. Van Rossum and F. L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697 (Cited on page 48).
- [82] R. Brun and F. Rademakers. “ROOT: An object oriented data analysis framework”. In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X (Cited on page 48).
- [83] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4 (Cited on page 48).
- [84] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (Cited on page 48).
- [85] S. Jadach, B. F. L. Ward, and Z. Was. “The precision Monte Carlo event generator KK for two-fermion final states in e^+e^- collisions”. In: *Comput. Phys. Commun.* 130 (2000), p. 260. DOI: 10.1016/S0010-4655(00)00048-5. arXiv: hep-ph/9912214 [hep-ph] (Cited on page 48).
- [86] N. Davidson et al. “Universal Interface of TAUOLA: Technical and Physics Documentation”. In: *Comput. Phys. Commun.* 183 (2012), pp. 821–843. DOI: 10.1016/j.cpc.2011.12.009. arXiv: 1002.0543 [hep-ph] (Cited on page 49).
- [87] G. Balossini et al. “Matching perturbative and parton shower corrections to Bhabha process at flavour factories”. In: *Nucl. Phys. B* 758 (2006), pp. 227–253. DOI: 10.1016/j.nuclphysb.2006.09.022. arXiv: hep-ph/0607181 (Cited on page 49).

- [88] C. M. Carloni Calame et al. “The BABAYAGA event generator”. In: *Nucl. Phys. B Proc. Suppl.* 131 (2004). Ed. by Marco Incagli and G. Venanzoni, pp. 48–55. DOI: 10.1016/j.nuclphysbps.2004.02.008. arXiv: hep-ph/0312014 (Cited on page 49).
- [89] F. A. Berends, P. H. Daverveldt, and R. Kleiss. “Radiative Corrections to the Process $e^+e^- \rightarrow e^+e^-\mu^+\mu^-$ ”. In: *Nucl. Phys. B* 253 (1985), p. 421. DOI: 10.1016/0550-3213(85)90540-1 (Cited on page 49).
- [90] P. Golonka and Z. Was. “PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays”. In: *Eur. Phys. J. C* 45 (2006), pp. 97–107. DOI: 10.1140/epjc/s2005-02396-4. arXiv: hep-ph/0506026 (Cited on page 49).
- [91] Y. Ohnishi et al. “Accelerator design at SuperKEKB”. In: *PTEP* 2013 (2013), 03A011. DOI: 10.1093/ptep/pts083 (Cited on page 49).
- [92] S. Agostinelli et al. “GEANT4: A simulation toolkit”. In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8 (Cited on page 49).
- [93] Y. Sato et al. “Monte Carlo matching in the Belle II software”. In: *EPJ Web Conf.* 251 (2021), p. 03021. DOI: 10.1051/epjconf/202125103021 (Cited on page 49).
- [94] V. Bertacchi et al. “Track finding at Belle II”. In: *Comput. Phys. Commun.* 259 (2021), p. 107610. DOI: 10.1016/j.cpc.2020.107610. arXiv: 2003.12466 [physics.ins-det] (Cited on page 50).
- [95] T. Bilka et al. “Implementation of GENFIT2 as an experiment independent track-fitting framework”. In: (Feb. 2019). arXiv: 1902.04405 [physics.data-an] (Cited on page 50).
- [96] W. Waltenberger et al. “The RAVE/VERTIGO vertex reconstruction toolkit and framework”. In: *J. Phys. Conf. Ser.* 119 (2008). Ed. by Randall Sobie, Reda Tafirout, and Jana Thomson, p. 032037. DOI: 10.1088/1742-6596/119/3/032037 (Cited on pages 51, 54, 78, 117).
- [97] J. F. Krohn et al. “Global decay chain vertex fitting at Belle II”. In: *Nucl. Instrum. Meth. A* 976 (2020), p. 164269. DOI: 10.1016/j.nima.2020.164269. arXiv: 1901.11198 [hep-ex] (Cited on pages 53, 78, 116).
- [98] H. Aihara et al. “The Belle II Detector Upgrades Framework Conceptual Design Report”. In: (June 2024). arXiv: 2406.19421 [hep-ex]. URL: <https://arxiv.org/abs/2406.19421> (Cited on pages 55, 58, 61, 308).

- [99] F. Abudinén et al. “B-flavor tagging at Belle II”. In: *Eur. Phys. J. C* 82.4 (2022), p. 283. DOI: 10.1140/epjc/s10052-022-10180-9. arXiv: 2110.00790 [hep-ex] (Cited on pages 70, 71).
- [100] I. Adachi et al. “New graph-neural-network flavor tagger for Belle II and measurement of $\sin 2\phi_1$ in $B^0 \rightarrow J/\psi K_S^0$ decays”. In: *Phys. Rev. D* 110 (1 July 2024), p. 012001. DOI: 10.1103/PhysRevD.110.012001 (Cited on pages 72, 73, 311).
- [101] P. Stavroulakis. *Graph Neural Network Flavor Tagger and measurement of $\sin 2\beta$ at Belle II*. 2025. arXiv: 2501.17631 [hep-ex]. URL: <https://arxiv.org/abs/2501.17631> (Cited on page 73).
- [102] F. Abudinén et al. “Measurement of the B^0 lifetime and flavor-oscillation frequency using hadronic decays reconstructed in 2019–2021 Belle II data”. In: *Phys. Rev. D* 107 (9 May 2023), p. L091102. DOI: 10.1103/PhysRevD.107.L091102 (Cited on page 75).
- [103] M. Pivk and F. R. Le Diberder. “sPlot: A statistical tool to unfold data distributions”. In: *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083 [physics.data-an] (Cited on pages 77, 92).
- [104] G. C. Fox and S. Wolfram. “Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes”. In: *Phys. Rev. Lett.* 41 (1978), p. 1581. DOI: 10.1103/PhysRevLett.41.1581 (Cited on pages 78, 119).
- [105] B. Efron. “Bootstrap Methods: Another Look at the Jackknife”. In: *Annals Statist.* 7.1 (1979), pp. 1–26. DOI: 10.1214/aos/1176344552 (Cited on pages 107, 142, 154).
- [106] I. I. Bigi and A. I. Sanda. *CP violation*. Vol. 9. Cambridge University Press, Sept. 2009. ISBN: 978-0-511-57728-4, 978-0-521-84794-0, 978-1-107-42430-2. DOI: 10.1017/CB09780511581014 (Cited on pages 114, 243).
- [107] T. Keck. “FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment”. In: *Comput. Softw. Big Sci.* 1.1 (2017), p. 2. DOI: 10.1007/s41781-017-0002-8 (Cited on page 115).
- [108] A. Khotanzad and Y.H. Hong. “Invariant image recognition by Zernike moments”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.5 (1990), pp. 489–497. DOI: 10.1109/34.55109 (Cited on page 115).
- [109] L. Breiman et al. *Classification and Regression Trees*. Wadsworth International Group, 1984. ISBN: 9780412048418. DOI: 10.1201/9781315139470 (Cited on page 118).

- [110] J. H. Friedman. “Greedy Function Approximation: A Gradient Boosting Machine”. In: *Annals of Statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451 (Cited on page 118).
- [111] T. Chen and C. Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: (Mar. 2016). DOI: 10.1145/2939672.2939785. arXiv: 1603.02754 [cs.LG] (Cited on page 120).
- [112] T. Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: (July 2019). arXiv: 1907.10902 [cs.LG] (Cited on page 120).
- [113] H. Tajima et al. “Proper-time resolution function for measurement of time evolution of B mesons at the KEK B-Factor”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 533.3 (2004), pp. 370–386. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2004.07.199>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900204017401> (Cited on page 122).
- [114] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. London: Chapman and Hall, 1986. DOI: 10.1007/978-1-4899-3324-9 (Cited on page 125).
- [115] P. Catastini and G. Punzi. “Bias-Free Estimation in Multicomponent Maximum Likelihood Fits with Component-Dependent Templates”. In: *Statistical Problems in Particle Physics, Astrophysics and Cosmology*. published by Imperial College Press and distributed by World Scientific Publishing Co., May 2006, pp. 60–63. DOI: 10.1142/9781860948985_0014. URL: http://dx.doi.org/10.1142/9781860948985_0014 (Cited on page 127).
- [116] O. Long et al. “Impact of tag side interference on time dependent CP asymmetry measurements using coherent $B^0 \bar{B}^0$ pairs”. In: *Phys. Rev. D* 68 (2003), p. 034010. DOI: 10.1103/PhysRevD.68.034010. arXiv: hep-ex/0303030 (Cited on page 226).

Résumé en Français

Cette thèse présente les recherches que j’ai menées pendant trois ans dans le cadre de l’expérience Belle II, une collaboration internationale réunissant plus de 1000 physiciens. Mes travaux ont porté sur la mesure de l’asymétrie CP en fonction du temps dans les désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ à l’aide de l’ensemble de données 365 fb^{-1} collectées par Belle II. De plus, j’ai calibré les paramètres du Flavor Tagger et de la fonction de résolution sur Δt , deux outils essentiels utilisés dans l’analyse en fonction du temps et largement utilisés dans le cadre de cette collaboration. Enfin, j’ai mené une étude visant à évaluer les performances physiques de la mise à niveau prévue du détecteur de vertex Belle II, sur la base de désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ simulées.

L’expérience Belle II

L’expérience Belle II est menée à l’accélérateur SuperKEKB à Tsukuba, au Japon, où des collisions à haute luminosité se produisent entre des électrons et des positons d’énergies asymétriques proches de la résonance $\Upsilon(4S)$ (voir Figure D.33). À cette énergie, des paires de mésons neutres et chargés B sont créés dans un état quantique cohérent. L’utilisation d’énergies de faisceaux asymétriques, 7 GeV pour les électrons et 4 GeV pour les positons, entraîne une accélération du système $\Upsilon(4S)$ dans la direction du faisceau d’électrons. Cette accélération est cruciale pour les études de violation de CP en fonction du temps, car elle permet de déduire la différence entre les temps de désintégration Δt entre les deux mésons B à partir de la séparation spatiale de leurs points de désintégration le long de la direction du boost.

Le détecteur Belle II est un spectromètre à grand angle solide entourant le point d’interaction, composé de plusieurs sous-détecteurs optimisés pour le vertexing, le suivi et l’identification précise des particules (voir figure D.33). Le plus proche du point d’interaction se trouve le détecteur de vertex (VXD), composé d’un détecteur à pixels (PXD) et d’un détecteur à pistes de silicium (SVD). Ceux-ci fournissent le suivi précis et la résolution de vertex nécessaires pour reconstruire les positions de désintégration des particules à courte durée de vie telles que le méson B^0 et le

K_S^0 , qui sont essentielles pour mesurer avec précision Δt .

À l'extérieur du détecteur de vertex se trouve la chambre à dérive centrale (CDC), qui mesure les impulsions des particules chargées et offre une excellente résolution pour la reconstruction des trajectoires sur un grand volume. La CDC contribue également à l'identification des particules grâce à des mesures de perte d'énergie par ionisation. La CDC est entourée de systèmes dédiés à l'identification des particules : le détecteur TOP (Time-Of-Propagation) dans la région du baril et le détecteur ARICH (Aerogel Ring Imaging Cherenkov) dans la partie avant. Ensemble, ces systèmes permettent une séparation efficace des kaons et des pions, ce qui est une condition essentielle pour l'étiquetage de la saveur dans les analyses dépendantes du temps.

Au-delà des systèmes d'identification des particules, le calorimètre électromagnétique (ECL) constitué de cristaux CsI(Tl) détecte les photons et les électrons avec une grande efficacité et une haute résolution énergétique. Cela est particulièrement important pour la reconstruction des désintégrations radiatives du B telles que $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$, où un photon hautement énergétique est présent dans l'état final. Le sous-détecteur le plus externe est le détecteur K_L^0 et muon (KLM), qui identifie les muons et détecte les kaons neutres à longue durée de vie, fournissant ainsi des informations supplémentaires d'identification et de suppression du bruit de fond.

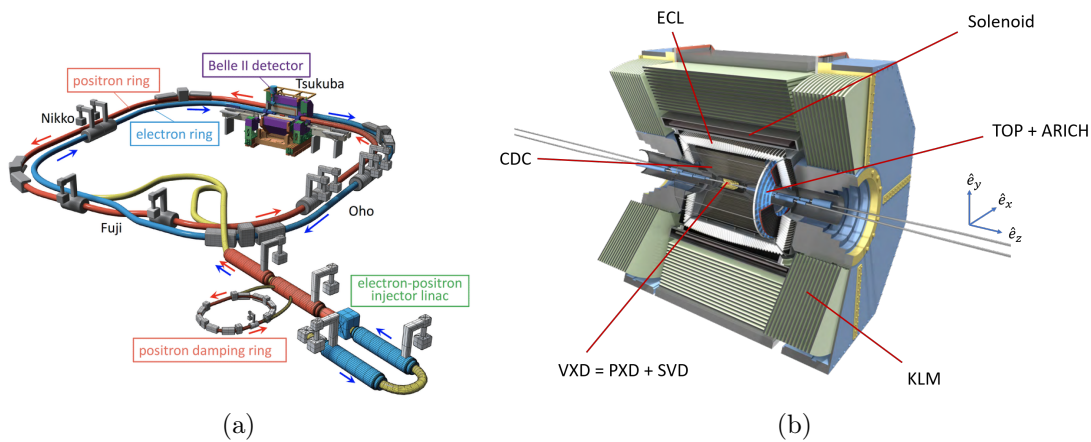


Figure D.33: Vues schématiques (a) de l'accélérateur SuperKEKB et (b) du détecteur Belle II.

Compte tenu des conditions exigeantes prévues pour la luminosité nominale du SuperKEKB, une attention particulière a été accordée au maintien des performances du système de tracking interne. Cela a motivé une étude spécifique de la mise à niveau prévue du détecteur de vertex, dont les performances physiques

attendues sont résumées dans la section suivante.

Étude de performance du VTX

Les performances du futur détecteur de vertex (VTX) ont été évaluées à l'aide de désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ simulées. Ce mode constitue un excellent critère de référence pour évaluer les performances du détecteur de vertex, car il permet de tester simultanément la reconstruction des vertex déplacés à partir des mésons K_S^0 et la précision du vertex de désintégration du B^0 , qui a un impact direct sur la résolution sur Δt essentielle pour les mesures d'asymétrie CP dépendantes du temps. L'étude visait à quantifier les performances physiques attendues du détecteur amélioré par rapport au VXD actuel, en se concentrant sur l'efficacité de la reconstruction et la résolution sur le vertex de désintégration.

La conception VTX remplace les détecteurs PXD et SVD par un seul détecteur entièrement pixellisé basé sur des capteurs CMOS monolithiques. Son objectif est de maintenir ou d'améliorer les performances physiques de Belle II dans les conditions d'occupation et de rayonnement nettement plus élevées attendues suite à l'augmentation de la luminosité du SuperKEKB. La géométrie envisagée au moment de l'étude comprenait cinq couches de détection dont les rayons variaient entre environ 14 mm et 140 mm, assurant une couverture de suivi continue depuis le tube à faisceau vers l'extérieur. Deux configurations ont été évaluées : une géométrie nominale à cinq couches et une configuration alternative dans laquelle la troisième couche était décalée de 30 mm vers l'extérieur. Cette modification était motivée par le désir d'améliorer l'efficacité de la reconstruction du K_S^0 , car le rayon de désintégration typique de $K_S^0 \rightarrow \pi^+ \pi^-$ est de plusieurs centimètres, souvent au-delà des couches les plus internes du détecteur. En augmentant l'espacement entre les troisième et quatrième couches, la probabilité que les deux traces de pions produisent trois impacts ou plus dans le détecteur de vertex est accrue, ce qui conduit à une efficacité de reconstruction plus élevée pour ces vertex déplacés (voir la figure D.34).

Des échantillons Monte Carlo d'événements $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ ont été générés pour chaque géométrie, et la simulation Belle II complète et la chaîne de reconstruction ont été appliquées à l'aide du framework BASF2. Pour chaque événement, la désintégration du méson B côté signal a été reconstruite, et les ajustements de vertex ont été effectués à l'aide de l'algorithme standard Belle II TREEFITTER. Trois niveaux différents de bruit induit par le faisceau ont été superposés, correspondant à des scénarios optimiste, intermédiaire et pessimiste pour un fonctionnement à la luminosité nominale du SuperKEKB de $6 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$.

Les résultats ont montré une nette amélioration de l'efficacité globale de reconstruction du B^0 pour les géométries VTX améliorées, principalement grâce

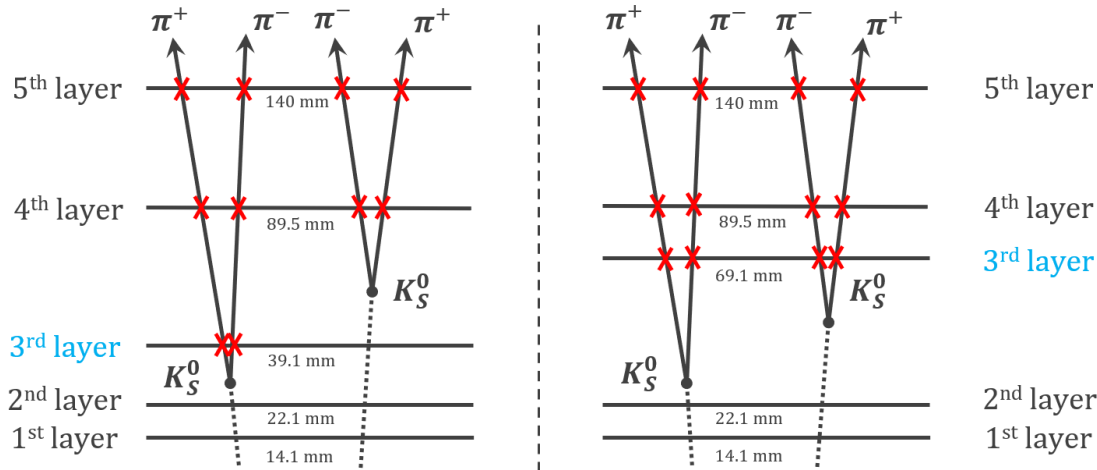


Figure D.34: Illustration de l'amélioration attendue de l'efficacité de reconstruction du K_S^0 grâce au déplacement vers l'extérieur de la troisième couche VTX. À gauche : géométrie VTX nominale. À droite : géométrie VTX alternative.

à de meilleures performances de reconstruction du K_S^0 . Le tableau D.1 résume l'efficacité de reconstruction de la chaîne complète de désintégration du B^0 dans chacune des conditions de fond induites par le faisceau pour chacune des géométries du détecteur de vertex. Dans le scénario pessimiste de fond induit par le faisceau, la configuration VTX alternative apporte une amélioration relative d'environ 10% par rapport au VXD actuel. Cette amélioration est conforme aux attentes liées à la couverture spatiale accrue pour les désintégrations du K_S^0 et à la granularité plus élevée des capteurs de pixels.

Géométrie	Efficacité de reconstruction (%)		
	Optimiste	Intermédiaire	Pessimiste
VXD	36.636 ± 0.068	32.127 ± 0.066	21.571 ± 0.058
VTX nominal	33.970 ± 0.067	29.972 ± 0.065	21.039 ± 0.058
VTX alternatif	35.904 ± 0.068	31.940 ± 0.066	23.020 ± 0.060

Table D.1: Efficacité de reconstruction de $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ pour chacune des trois géométries de détecteur de vertex dans trois scénarios différents de fond induit par le faisceau.

Au-delà de l'efficacité de la reconstruction, l'étude a également examiné l'impact de la mise à niveau sur la résolution sur Δt , qui est un aspect crucial des mesures d'asymétrie CP dépendantes du temps, ayant un impact direct sur la sensibilité aux observables violant CP . La figure D.35 montre la largeur centrale de la résolution sur Δt pour chacune des trois géométries en fonction du niveau de fond induit

par le faisceau. Les résultats indiquent que les géométries VTX offrent une résolution comparable ou légèrement améliorée par rapport au détecteur actuel dans toutes les conditions, malgré un environnement de fond plus difficile. De plus, le déplacement vers l'extérieur de la troisième couche dans la géométrie alternative ne dégrade pas significativement la résolution, ce qui confirme que l'amélioration de l'efficacité de reconstruction du K_S^0 peut être obtenue sans compromettre la résolution du vertex de désintégration.

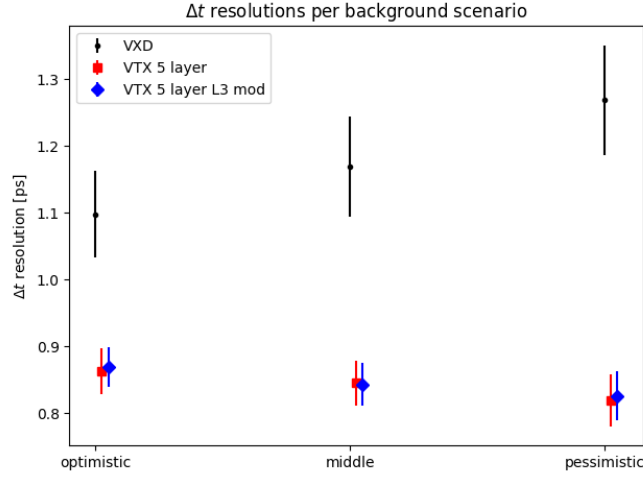


Figure D.35: Indicateurs de résolution sur Δt pour chaque géométrie de détecteur de vertex dans les trois scénarios de bruit induit par le faisceau.

Dans l'ensemble, les résultats démontrent que la mise à niveau VTX permet de maintenir, voire d'améliorer à certains égards, les performances de détection des vertex du détecteur Belle II, même dans les conditions plus exigeantes attendues à la luminosité nominale du SuperKEKB. Les conclusions de cette étude ont joué un rôle clé dans l'élaboration des considérations de conception du détecteur de vertex de nouvelle génération et ont été incluses dans les études de performance physique présentées dans le rapport de conception conceptuelle du cadre de mise à niveau du détecteur Belle II [98].

Calibrage du Flavor Tagger et de la fonction de résolution sur Δt

Les mesures de la violation de CP en fonction du temps à Belle II reposent essentiellement sur la détermination de la saveur du méson B côté signal au moment de sa désintégration. Cela est réalisé grâce l'étiquetage de la saveur, qui consiste

à déduire la saveur du méson B côté tag (B_{tag}) à partir des informations issues de ses produits de désintégration reconstruits. Au niveau de la résonance $\Upsilon(4S)$, les deux mésons neutres B sont produits dans un état quantique intriqué, de sorte que l'identification de la saveur de B_{tag} au moment de sa désintégration permet de déterminer directement la saveur du méson B côté signal (B_{sig}) au même moment. Une étiquetage de la saveur précis et efficace est donc l'un des éléments les plus importants pour mesurer les asymétries CP et les paramètres de mélange.

Jusqu'en 2023, Belle II utilisait un Flavor Tagger basé sur les catégories, un algorithme d'apprentissage automatique fondé sur une série d'arbres de décision boostés qui classait les événements dans l'une des 13 catégories reflétant différentes signatures de saveur, telles que la charge des leptons à haute impulsion ou des kaons dans la désintégration B_{tag} . Chaque événement se voit attribuer une prédiction de saveur $q = \pm 1$ et une valeur de confiance $r \in [0, 1]$, dont le produit, $q \cdot r$, quantifie à la fois la saveur prédite et la fiabilité de l'étiquetage. La distribution de la réponse du Flavor Tagger est généralement divisée en sept classes r de pureté croissante, et la performance globale de l'algorithme est exprimée en termes d'efficacité effective de l'étiquetage :

$$\varepsilon_{\text{tag}} = \sum_i \varepsilon_i (1 - w_i)^2, \quad (\text{D.1})$$

où ε_i désigne la fraction d'événements dans le i -ième intervalle r et w_i la probabilité de l'étiquetage erroné correspondante. Le Flavor Tagger basé sur les catégories a atteint une efficacité effective de l'étiquetage de $(31.68 \pm 0.45)\%$.

Une avancée majeure a été réalisée avec le développement d'un nouvel algorithme d'étiquetage de la saveur, basé sur un réseau de neurones en graphes (GNN), appelé **GFlaT**. En codant les relations géométriques et cinématiques entre les particules de l'état final dans la désintégration du B_{tag} , **GFlaT** dépasse les performances limites de la méthode basée sur les catégories. La figure D.36 montre une comparaison entre la distribution $q \cdot r$ du Flavor Tagger basé sur les catégories et celle de **GFlaT** pour des événements de signal simulés, démontrant que l'approche basée sur le GNN produit une meilleure séparation entre les vrais mésons B^0 et $\bar{B}^0$. Cette amélioration est particulièrement prononcée dans les cas où la désintégration du B côté tag ne contient pas de signature de saveur claire, démontrant ainsi la capacité du Flavor Tagger basé sur GNN à récupérer des informations auparavant inaccessibles.

Les performances de **GFlaT** ont été évaluées à l'aide de 365 fb^{-1} de données Belle II collectées lors de la série de mesures Run 1. La procédure de calibration repose sur la reconstruction des désintégrations auto-étiquetées, $B^0 \rightarrow D^{(*)-} \pi^+$, qui déterminent sans ambiguïté la saveur réelle du B_{sig} . Ces désintégrations permettent une comparaison directe entre les prédictions de l'algorithme et la réalité des données. Après avoir reconstruit les candidats de signal et supprimé le bruit

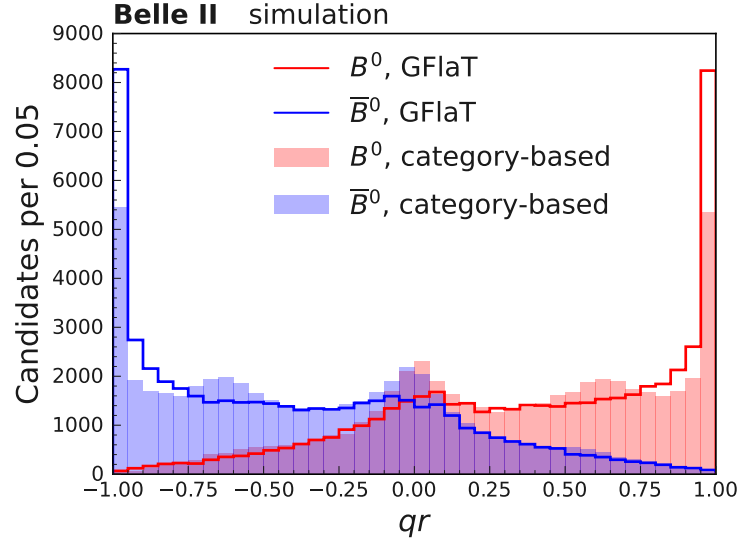


Figure D.36: Comparaison entre la distribution $q \cdot r$ du Flavor Tagger basé sur les catégories et celle de GFlaT pour des événements de signaux simulés. La distribution est divisée selon que le B_{sig} dans l'événement est un B^0 (rouge) ou un $\bar{B}^0$ (bleu).

de fond continu à l'aide de variables de forme d'événement, les efficacités pour le signal et du bruit de fond ont été déterminées à partir d'ajustements de la distribution de ΔE . La figure D.37 montre l'ajustement aux données dans les régions du signal et de la bande latérale.

Le résultat de l'ajustement de ΔE est ensuite utilisé pour soustraire statistiquement le bruit de fond de la distribution de Δt à l'aide de la méthode sPlot. Les fractions de mauvaises étiquetages et les asymétries de mauvaises étiquetages dans chaque bin r sont ensuite extraites en ajustant simultanément la distribution de Δt dont le bruit de fond a été soustrait, pour chaque bin et chaque configuration de saveur. Les ajustements ont été effectués à l'aide d'une méthode de maximum de vraisemblance non binée utilisant la fonction décrivant l'oscillation convoluée avec la fonction de résolution sur Δt , qui est utilisée pour modéliser la résolution du détecteur dans la détermination de Δt . Parallèlement aux paramètres du Flavor Tagger, les paramètres de la fonction de résolution sur Δt ont également été déterminés selon la même procédure. La figure D.38 montre la projection de l'ajustement dans le bin r de la plus haute qualité de prédiction, où le comportement oscillatoire entre les événements de saveur opposée ($B^0 \bar{B}^0$) et de saveur identique ($B^0 B^0$ ou $\bar{B}^0 \bar{B}^0$) est le plus clairement visible. L'étalonnage a démontré que les fractions de les étiquetages erronées dans les données sont systéma-

tiquement plus importantes que dans la simulation, soulignant la nécessité d’une détermination basée sur les données.

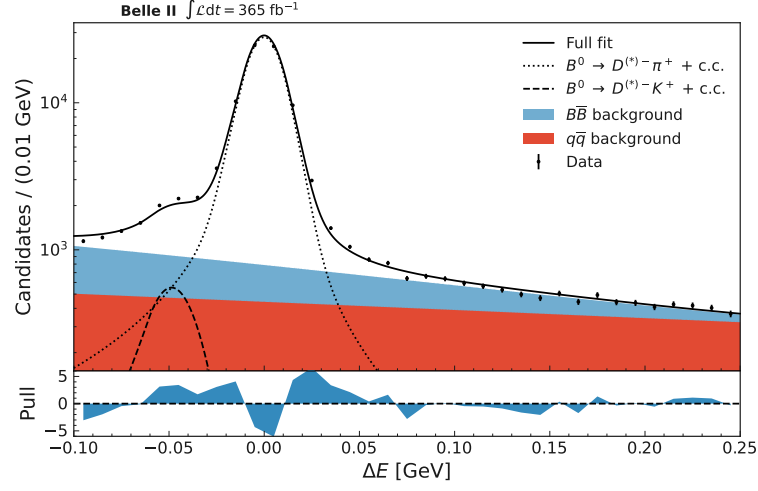
À partir de ces ajustements, l’efficacité effective de l’étiquetage dans les données a été déterminée comme suit :

$$\varepsilon_{\text{tag}} = (37.40 \pm 0.43 \pm 0.36)\% , \quad (\text{D.2})$$

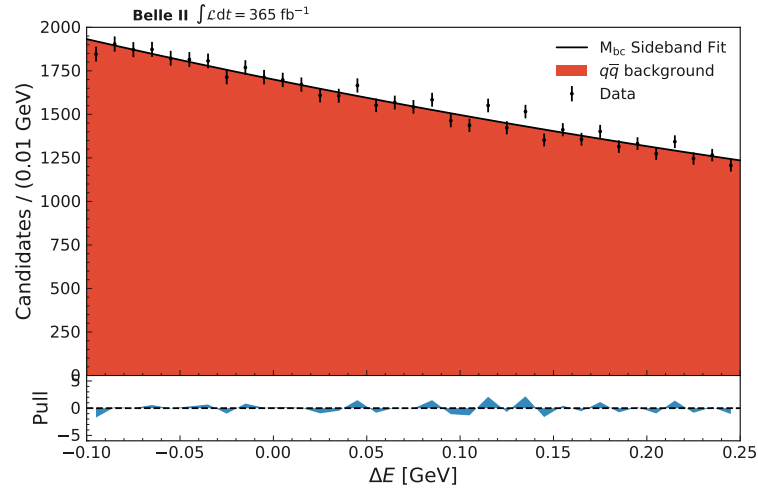
où la première incertitude est statistique et la seconde systématique. La composante systématique tient compte d’effets tels que l’alignement du détecteur, l’étalonnage du spot du faisceau, la modélisation des distributions de ΔE et les incertitudes liées à la fonction de résolution, bien que son ampleur totale reste secondaire par rapport à l’erreur statistique.

L’amélioration par rapport à l’algorithme basé sur les catégories est significative : **GFlaT** permet un gain relatif de **18%** en termes d’efficacité d’étiquetage. Cela se traduit directement par une sensibilité accrue pour toutes les mesures de violation de CP dépendantes du temps, puisque la précision statistique est proportionnelle à $\sqrt{\varepsilon_{\text{tag}}}$. Les premières applications démontrent déjà ce gain : dans le “canal en or” $B^0 \rightarrow J/\psi K_S^0$, la précision des paramètres de CP S et C s’améliore respectivement de 8 % et 7 % par rapport à ce qui aurait été possible avec l’ancien algorithme d’équitage [100].

En conclusion, l’étalonnage du Flavor Tagger basé sur le GNN et de la fonction de résolution sur Δt a démontré la supériorité du premier dans les données réelles et a fourni à la collaboration un ensemble fiable de paramètres à utiliser dans toutes les analyses de violation de CP .



(a)



(b)

Figure D.37: Ajustement de la distribution de ΔE dans 365 fb^{-1} de données dans les régions (a) du signal et (b) de la bande latérale M_{bc} .

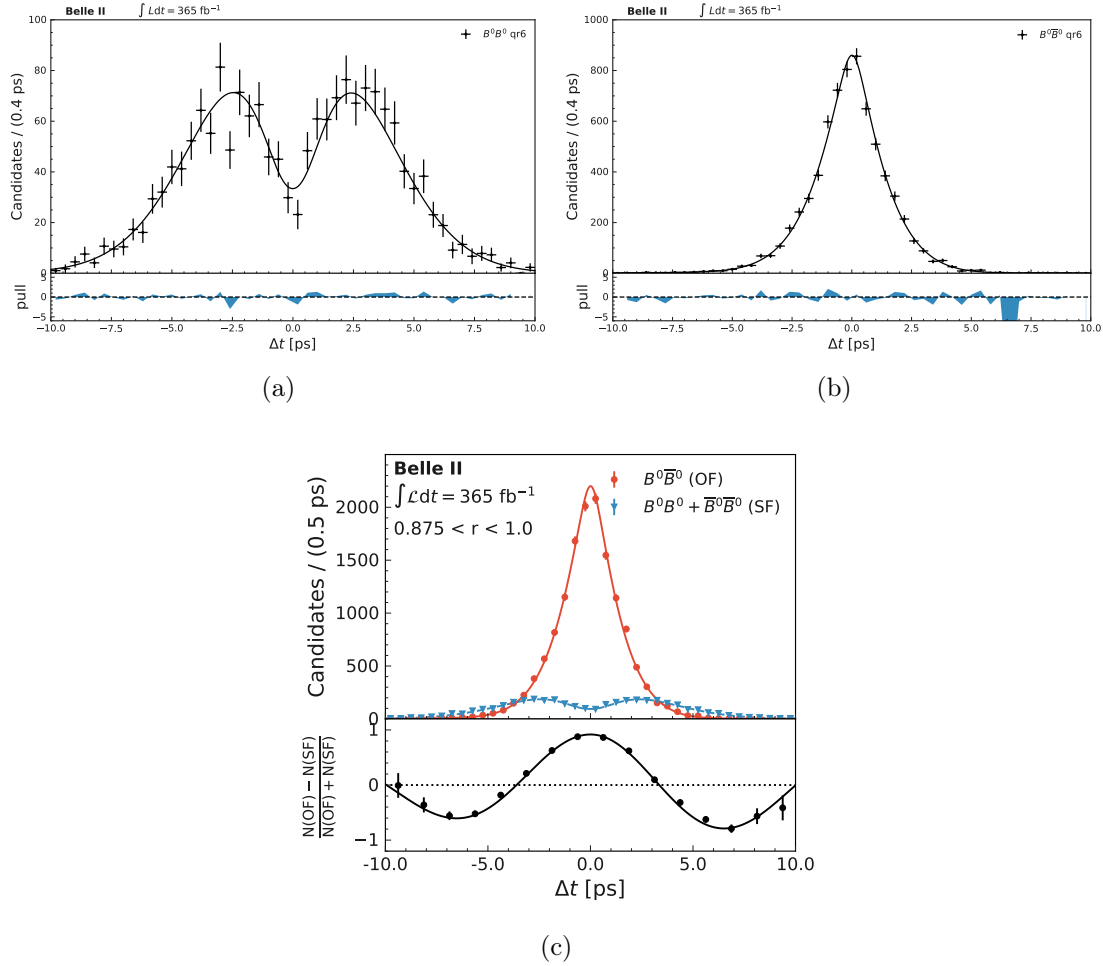


Figure D.38: Ajustement de la distribution de Δt après soustraction du bruit de fond dans le bin r de plus haute qualité de l'étiquetage pour 365 fb^{-1} de données, montrant (a) les événements de même saveur, (b) les événements de saveurs opposées et (c) l'asymétrie entre eux.

Mesure de l'asymétrie CP en fonction du temps dans les désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$

Les processus radiatifs à courant neutre changeant la saveur (FCNC) tels que $b \rightarrow s\gamma$ sont fortement supprimés au niveau de l'arbre dans le Modèle standard (MS) et se déroulent *via* des diagrammes pingouin électrofaibles au niveau de la boucle. Ces transitions sont particulièrement sensibles à la nouvelle physique (NP), car les particules virtuelles au-delà du MS peuvent contribuer à la boucle et modifier l'hélicité du photon. Dans le MS, le photon émis est principalement gaucher dans les désintégrations de b (et droitier dans les désintégrations de $\bar{b}$), avec seulement des petites corrections provenant des masses des quarks légers. Par conséquent, la mesure de la polarisation du photon dans les transitions $b \rightarrow s\gamma$ fournit un test unique du MS et un moyen de contraindre les courants droitiers prédits dans plusieurs scénarios de NP.

Une sonde puissante de la polarisation des photons est l'asymétrie CP en fonction du temps dans les désintégrations telles que $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$. Cette désintégration se produit *via* la transition radiative $b \rightarrow s\gamma$ et se caractérise par un état final qui comprend un K_S^0 , une paire de pions chargés et un photon. Dans le MS, la suppression de l'amplitude droite conduit à une faible asymétrie CP . Cependant, les contributions de NP peuvent renforcer cette asymétrie par des modifications de la polarisation des photons. L'asymétrie CP en fonction du temps est définie comme suit :

$$\begin{aligned} \mathcal{A}_{CP}(\Delta t) &\equiv \frac{\Gamma(\bar{B}^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma) - \Gamma(B^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma)}{\Gamma(\bar{B}^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma) + \Gamma(B^0(\Delta t) \rightarrow K_S^0 \pi^+ \pi^- \gamma)} \\ &= S \sin(\Delta m_d \Delta t) - C \cos(\Delta m_d \Delta t) . \end{aligned} \quad (D.3)$$

Ici, S et C représentent les observables violant la symétrie CP : S résulte de l'interférence entre le mélange et la désintégration, tandis que C quantifie la violation directe de la symétrie CP .

Une innovation de cette analyse est l'utilisation du plan de Dalitz $K_S^0 \pi^+ - K_S^0 \pi^-$ pour définir deux nouvelles observables, S^+ et S^- , qui améliorent la sensibilité à la polarisation des photons. Comme l'illustre la figure D.39, le plan de Dalitz est divisé par la bissectrice $m_{K_S^0 \pi^+}^2 = m_{K_S^0 \pi^-}^2$ en deux régions désignées I et $\bar{I}$. Les paramètres de CP correspondants dans chaque moitié du plan de Dalitz, S^I et $S^{\bar{I}}$, sont combinés en :

$$S^+ = S^I + S^{\bar{I}} , \quad (D.4)$$

$$S^- = S^I - S^{\bar{I}} . \quad (D.5)$$

Ces observables peuvent être combinées avec les résultats d’une analyse en amplitudes du mode partenaire d’isospin, $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$, afin d’imposer de nouvelles contraintes sur les parties réelles et imaginaires du rapport des coefficients de Wilson C_7''/C_7 , qui codifie les contributions des courants droits et gauches [27].

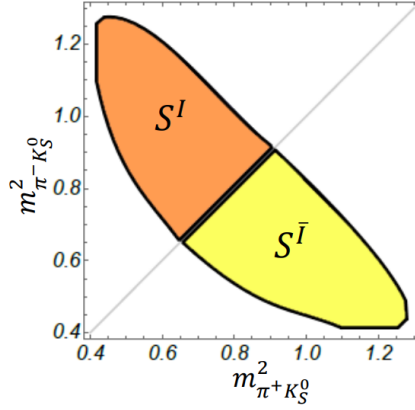


Figure D.39: Illustration représentant le plan de Dalitz $K_S^0 \pi^+ - K_S^0 \pi^-$. La bissectrice $m_{K_S^0 \pi^+}^2 = m_{K_S^0 \pi^-}^2$ divise le plan en deux moitiés et les paramètres mesurés S^+ et S^- sont des combinaisons linéaires des paramètres S mesurés simultanément dans les deux moitiés.

La mesure est effectuée à l’aide de 365 fb^{-1} de données Belle II. Les candidats de signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ sont reconstruits en combinant un $K_S^0 \rightarrow \pi^+ \pi^-$ avec deux pions supplémentaires de charges opposées et un photon. Le bruit de fond continu provenant des processus $e^+ e^- \rightarrow q \bar{q}$ a été supprimé à l’aide d’un arbre de décision boosté multivarié (BDT) entraîné sur des variables de forme d’événement. La saveur du B_{sig} dans chaque événement a été déterminée à l’aide du Flavor Tagger basé sur GNN, GF1aT, ce qui a permis d’améliorer la sensibilité des paramètres de CP mesurés, comme indiqué précédemment.

Le rendement du signal et les paramètres de CP ont été extraits en utilisant une méthode de maximum de vraisemblance non binnée avec trois variables : la différence d’énergie ΔE , la masse contrainte par le faisceau M_{bc} et la différence entre les temps de désintégration Δt . Les formes du signal et du bruit de fond ont été déterminées à partir d’une simulation. La figure D.40 montre les projections de l’ajustement 3D sur chacune des trois observables. Dans le bin r de plus haute pureté ($0.875 < r < 1$) de l’étiquetage de la saveur, des oscillations claires entre les désintégrations B^0 et $\bar{B}^0$ sont visibles dans les distributions de Δt , comme le montre la figure D.41.

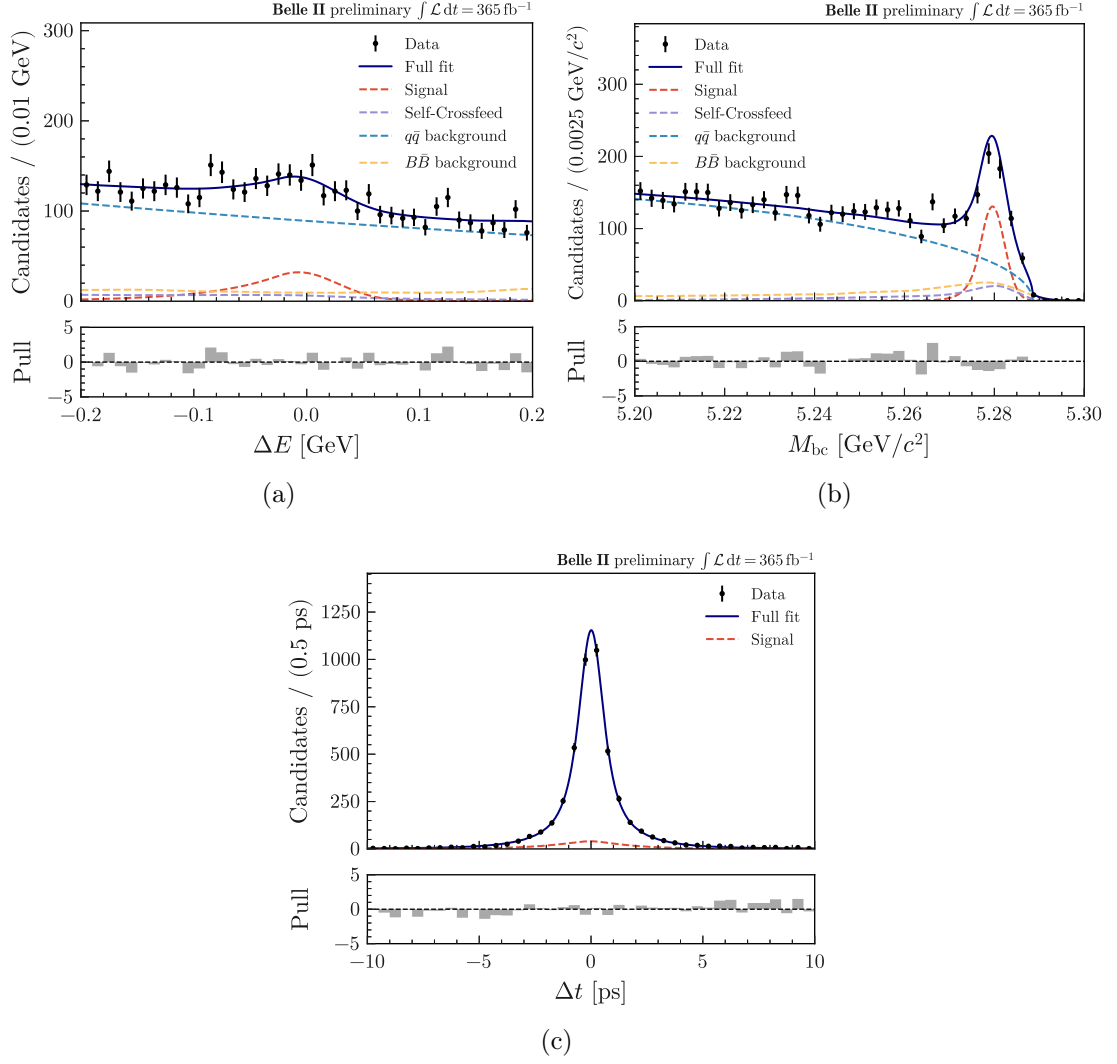


Figure D.40: Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur (a) ΔE (b) M_{bc} et (c) Δt . Les différentes composantes sont superposées dans les projections M_{bc} et ΔE , mais pas dans la projection Δt pour plus de clarté.

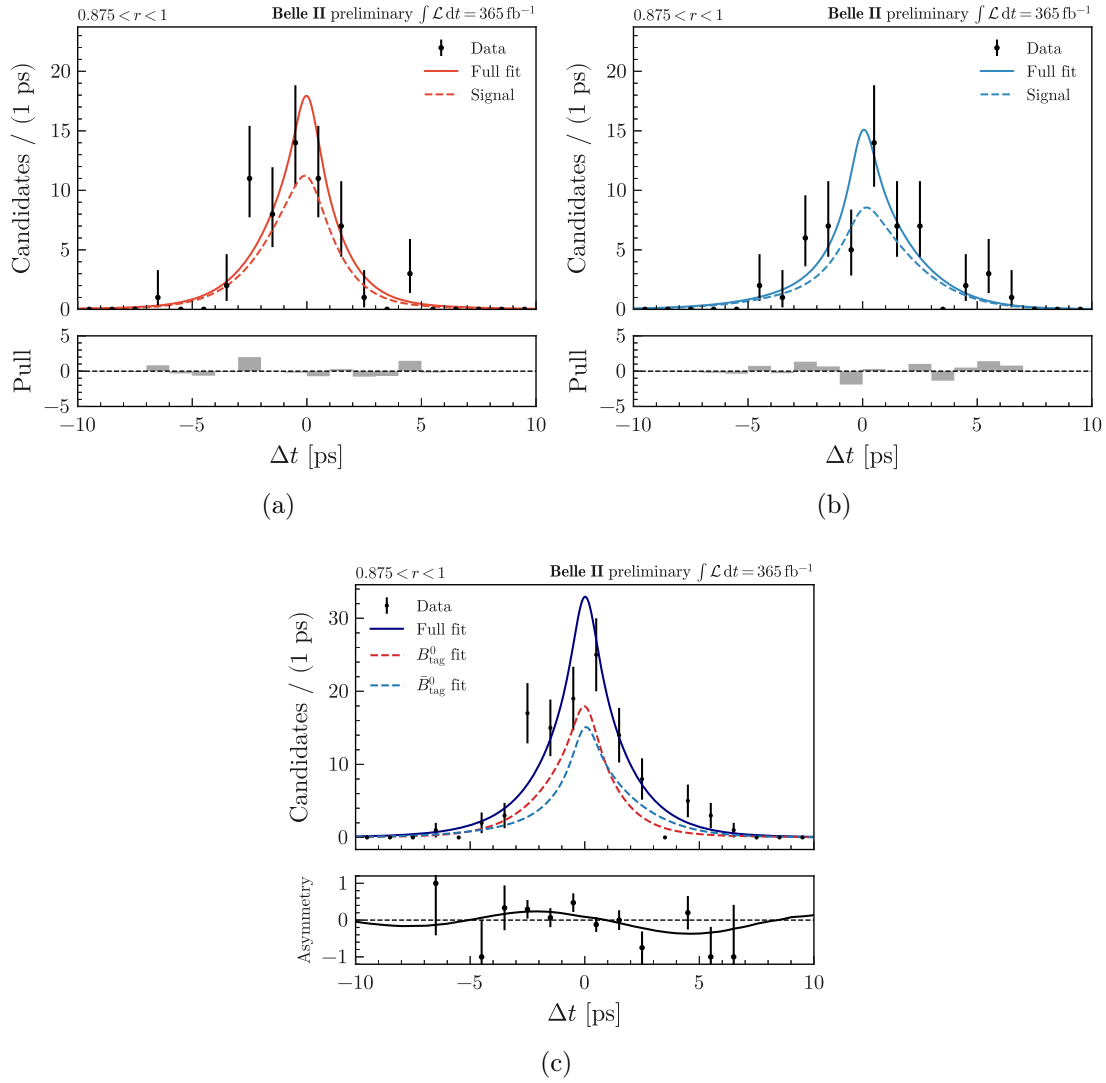


Figure D.41: Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur Δt pour (a) les événements étiquetés comme B^0 , (b) les événements étiquetés comme $\bar{B}^0$ et (c) tous les événements présentant une asymétrie associée entre les événements étiquetés B^0 et $\bar{B}^0$, dans une région enrichie en signal définie par $|\Delta E| < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ et $0.875 < r < 1$.

Pour mesurer les observables $S^\pm$, l'ensemble de données a été divisé selon le plan de Dalitz comme décrit ci-dessus et l'ajustement 3D a été effectué simultanément dans les régions $\bar{\text{I}}$ (où $m_{\pi^+K_S^0}^2 > m_{\pi^-K_S^0}^2$) et I (la région complémentaire). Les projections de l'ajustement sont présentées dans les figures D.42–D.45. Ces figures montrent que la majeure partie de l'asymétrie CP induite par le mélange provient de la région $\bar{\text{I}}$, bien que les statistiques soient très limitées.

Afin de garantir la robustesse de la stratégie d'ajustement et l'interprétation statistique des résultats, plusieurs études de validation ont été menées. Elles ont notamment consisté à générer et à ajuster des ensembles de données bootstrappées avec des paramètres d'entrée connus afin d'évaluer la stabilité du modèle ajusté et de vérifier l'absence de biais. De légers biais ont été détectés dans les distributions des pulls des paramètres de CP ajustés, qui ont également été utilisés pour confirmer que les incertitudes statistiques étaient correctement estimées. Ces biais sont de l'ordre de $\sim 30\%$ de l'incertitude statistique.

Des ensembles de données Toy Monte Carlo générés à partir du modèle PDF ont également été ajustés afin de vérifier si les biais observés avec la méthode de bootstrapping étaient dus au modèle ajusté ou à une mauvaise modélisation de certaines composantes de fond dans le Monte Carlo. Ce test semble indiquer que le modèle ajusté en lui-même n'entraîne aucun biais significatif. Néanmoins, une incertitude systématique a été attribuée afin de couvrir le biais d'ajustement observé.

L'analyse complète a également été réalisée avec un canal de contrôle, $B^+ \rightarrow K^+\pi^+\pi^-\gamma$, qui présente une topologie similaire mais sans asymétrie CP en fonction du temps, ce qui permet de tester directement l'ensemble de la stratégie d'analyse. La reconstruction et la sélection du canal de contrôle ont été conçues pour être aussi similaires que possible à celles du canal de signal afin de garantir une validation équitable de la stratégie d'ajustement. Comme prévu, aucune asymétrie CP significative n'a été observée dans le canal de contrôle.

Les incertitudes systématiques ont été évaluées pour toutes les sources pertinentes, incluant la modélisation des formes du signal et du bruit de fond, l'alignement des détecteurs, les paramètres de la fonction de résolution sur Δt , ainsi que l'impact des interférences du côté du B_{tag} . Les contributions dominantes sont dues au biais d'ajustement et à la possible violation de CP dans le bruit de fond $B\bar{B}$ qui n'est pas prise en compte dans l'ajustement. Cependant, dans tous les cas, les incertitudes systématiques ne représentent qu'une petite fraction des incertitudes statistiques.

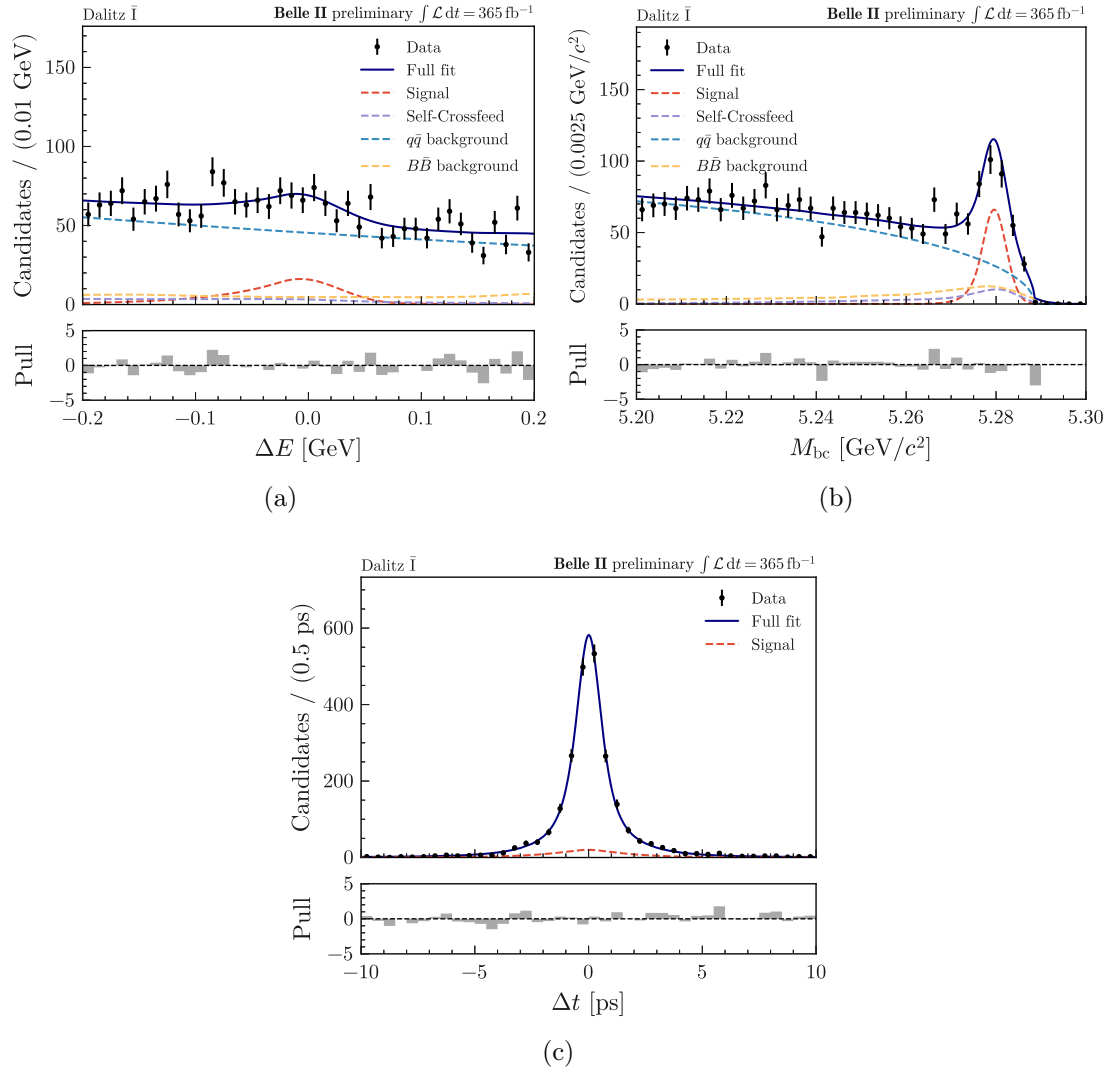


Figure D.42: Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur (a) ΔE (b) M_{bc} et (c) Δt dans la région du plan de Dalitz où $m_{\pi^+ K_S^0}^2 > m_{\pi^- K_S^0}^2$ (Dalitz $\bar{\text{I}}$). Les différentes composantes sont superposées dans les projections M_{bc} et ΔE , mais pas dans la projection Δt pour plus de clarté.

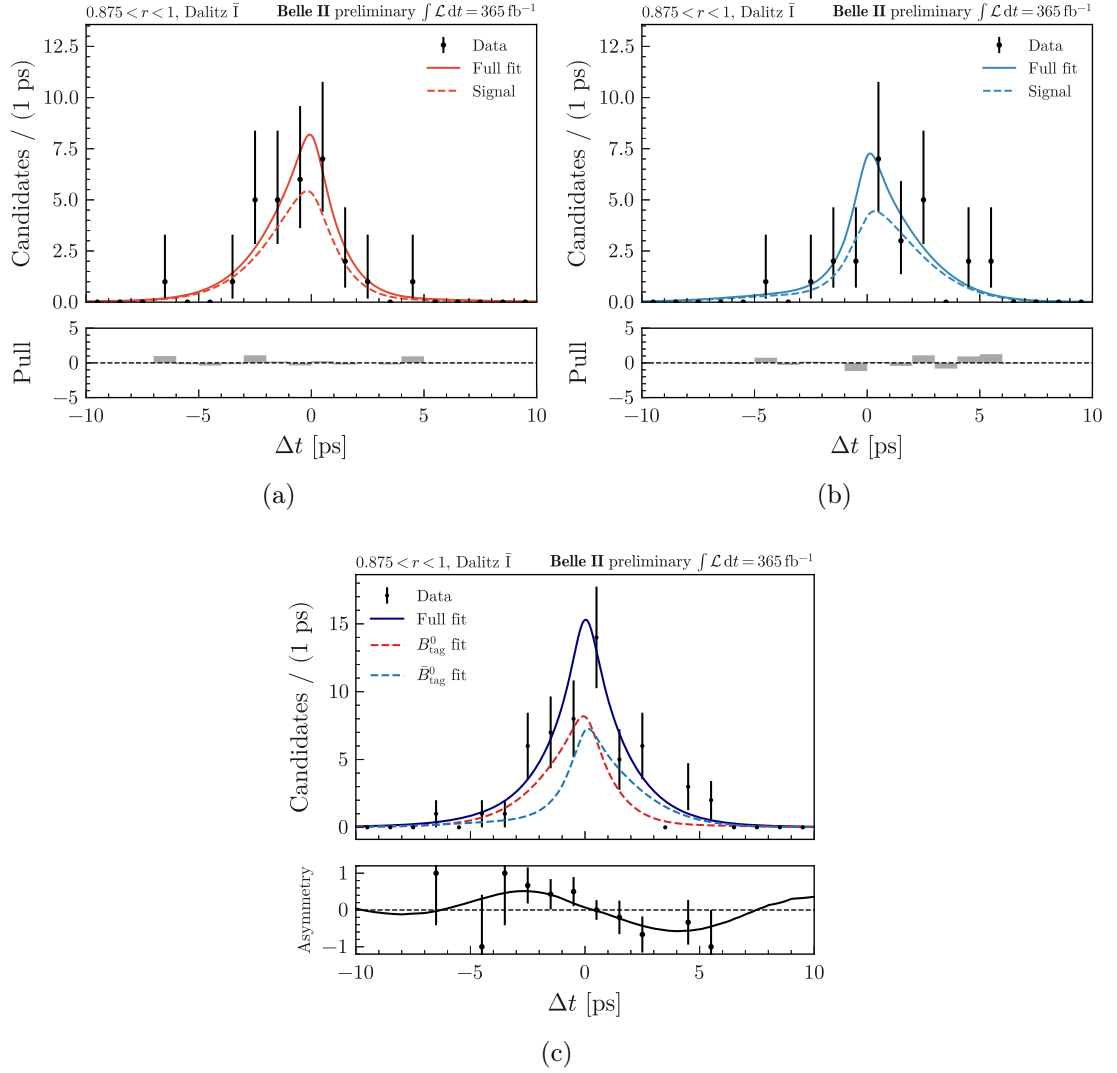


Figure D.43: Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur Δt pour (a) les événements étiquetés comme B^0 , (b) les événements étiquetés comme $\bar{B}^0$ et (c) tous les événements présentant une asymétrie associée entre B^0 et $\bar{B}^0$ étiquetés dans la région de Dalitz $\bar{I}$ et dans une région enrichie en signal définie par $|\Delta E| < 0.1 \text{ GeV}$, $M_{bc} > 5.27 \text{ GeV}/c^2$ et $0.875 < r < 1$.

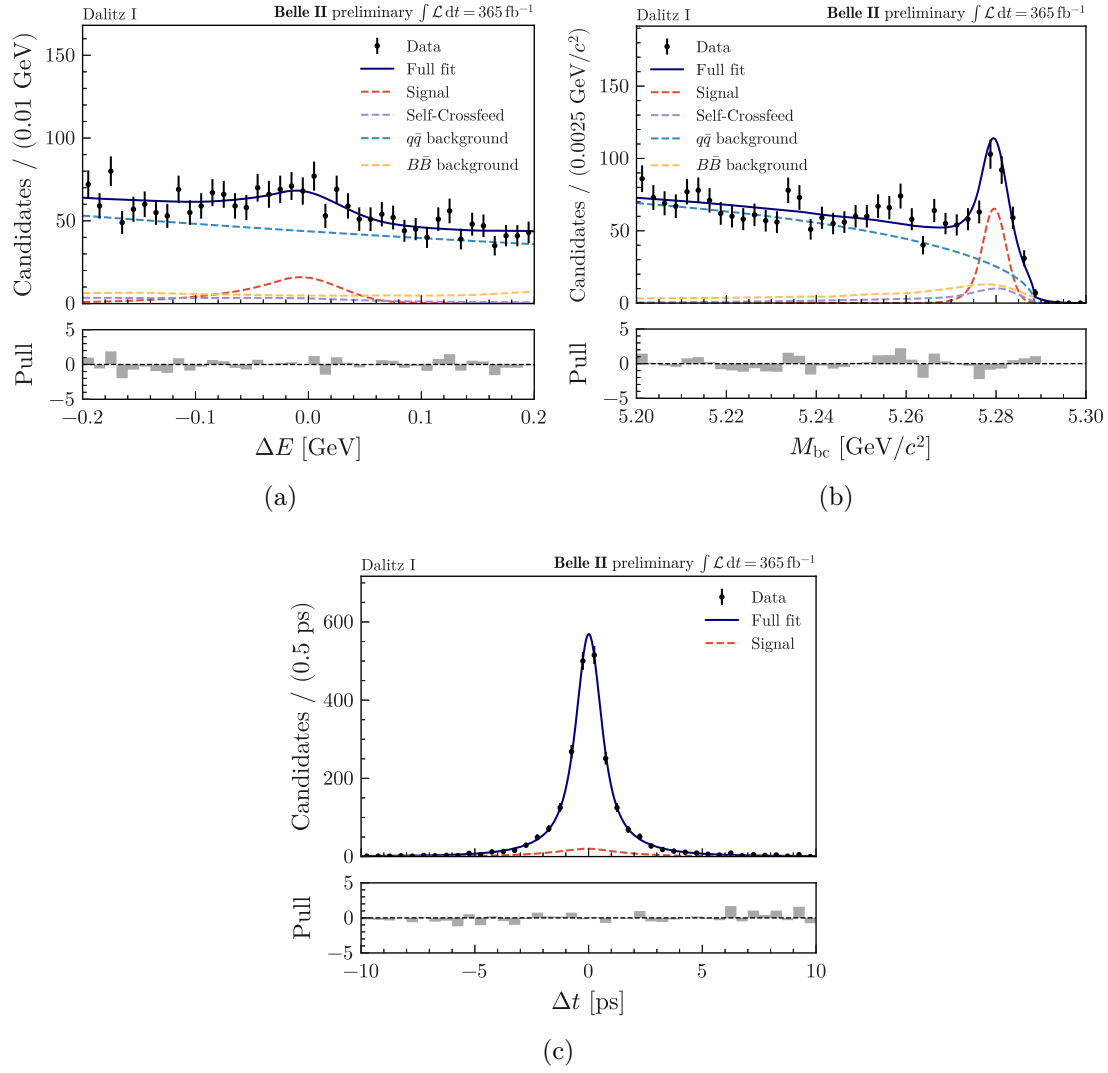


Figure D.44: Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur (a) ΔE (b) M_{bc} et (c) Δt dans la région du plan de Dalitz où $m_{\pi^+ K_S^0}^2 < m_{\pi^- K_S^0}^2$ (Dalitz I). Les différentes composantes sont superposées dans les projections M_{bc} et ΔE , mais pas dans la projection Δt pour plus de clarté.

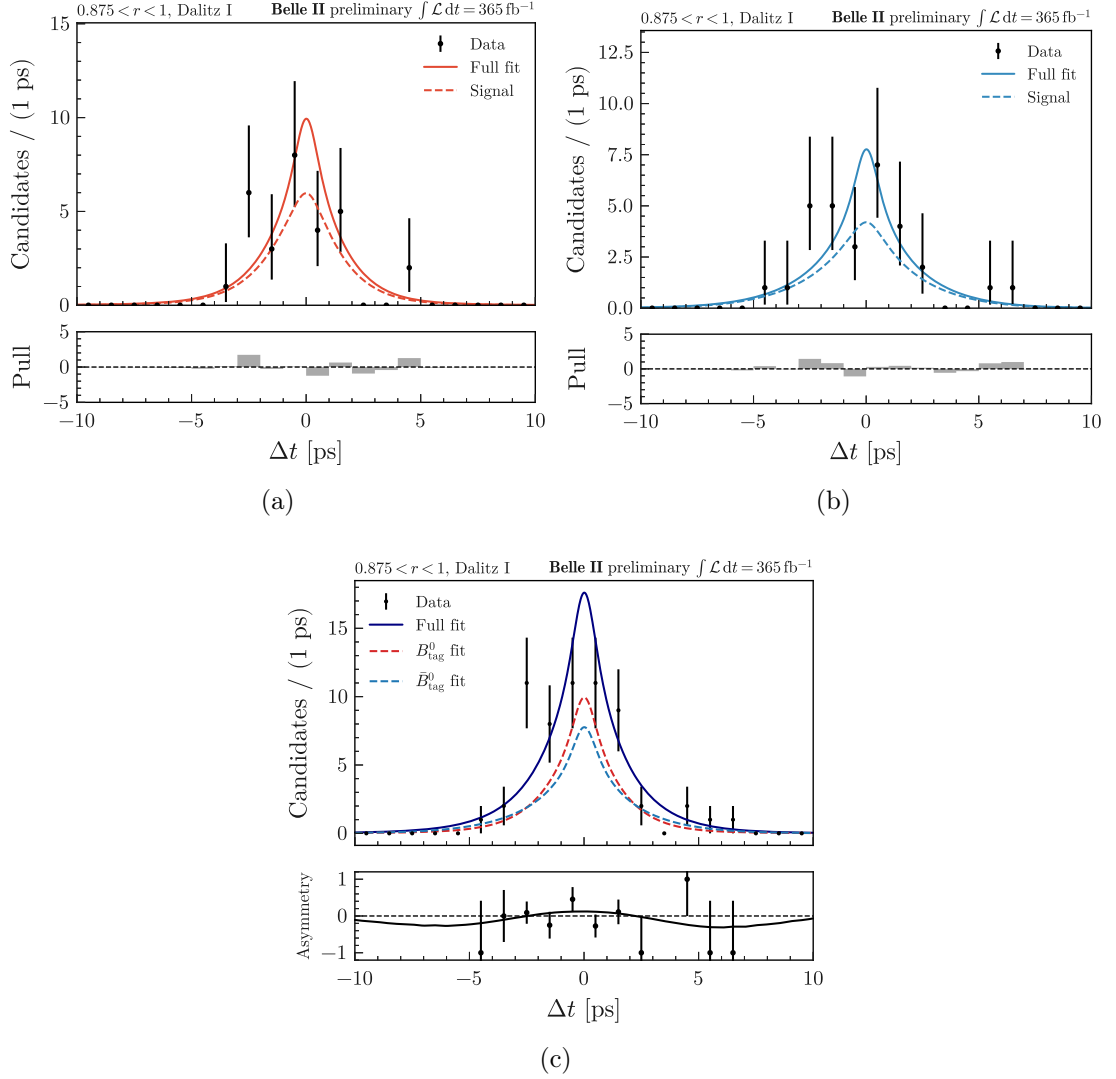


Figure D.45: Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur Δt pour (a) les événements étiquetés comme B^0 , (b) les événements étiquetés comme $\bar{B}^0$ et (c) tous les événements présentant une asymétrie associée entre B^0 et $\bar{B}^0$ étiquetés dans la région de Dalitz I et dans une région enrichie en signal définie par $|\Delta E| < 0.1 \text{ GeV}$, $M_{bc} > 5.27 \text{ GeV}/c^2$ et $0.875 < r < 1$.

Les paramètres de CP obtenus sont les suivants :

$$C = -0.29 \pm 0.13 \pm 0.06 , \quad (D.6)$$

$$S = -0.36 \pm 0.16 \pm 0.05 , \quad (D.7)$$

$$S^+ = -0.72 \pm 0.31 \pm 0.12 , \quad (D.8)$$

$$S^- = 0.70 \pm 0.30 \pm 0.05 , \quad (D.9)$$

où la première incertitude est statistique et la seconde systématique. Un léger écart de l'ordre de 2.3σ par rapport au MS est observé. Les résultats de Belle II ont également été combinés avec une nouvelle mesure utilisant l'ensemble complet des données de Belle représentant à 711 fb^{-1} , selon une stratégie de mesure similaire. Les valeurs combinées sont les suivantes :

$$C = -0.17 \pm 0.09 \pm 0.04 , \quad (D.10)$$

$$S = -0.29 \pm 0.11 \pm 0.05 , \quad (D.11)$$

$$S^+ = -0.57 \pm 0.23 \pm 0.10 , \quad (D.12)$$

$$S^- = 0.31 \pm 0.24 \pm 0.05 . \quad (D.13)$$

Il s'agit des mesures les plus précises à ce jour de l'asymétrie CP en fonction du temps dans les désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$.

Conclusion et perspectives

Dans cette thèse, une étude complète de la violation de CP en fonction du temps dans la désintégration rare $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ est présentée, à partir de 365 fb^{-1} de données collectées par l'expérience Belle II. En tant que transition radiative $b \rightarrow s\gamma$, cette désintégration fournit un outil sensible pour étudier la physique au-delà du Modèle standard grâce à des modifications potentielles de la polarisation des photons. Les paramètres S^+ et S^- , définis à l'aide du plan de Dalitz $K_S^0 \pi^+ - K_S^0 \pi^-$, sont également mesurés pour la première fois. Ces observables pourront être utilisées ultérieurement en combinaison avec une analyse en amplitudes du mode partenaire d'isospin pour imposer des contraintes sur le plan complexe des coefficients de Wilson.

En parallèle, un travail d'étalonnage essentiel a été effectué afin de préparer les algorithmes nécessaires à cette analyse et à d'autres. Les performances d'un nouveau Flavor Tagger basé sur des GNN, **GFlaT**, ont été évaluées à partir des données, démontrant un gain relatif de 18 % en termes d'efficacité d'étiquetage par rapport à l'algorithme précédent basé sur les catégories. Cela se traduit directement par une sensibilité accrue pour toutes les analyses dépendantes du temps à Belle II. Les paramètres de la fonction de résolution sur Δt ont également été calibrés selon

la même procédure, garantissant une modélisation précise de la résolution sur les temps de désintégration mesurés par le détecteur.

Une étude dédiée aux performances de la future mise à niveau du détecteur de vertex Belle II, VTX, a également été réalisée. À l'aide de désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ simulées, l'étude a évalué l'efficacité de la reconstruction et la résolution sur Δt dans diverses conditions de fond induit par le faisceau pour différentes géométries VTX. Les résultats montrent que l'amélioration du détecteur de vertex proposée permet une meilleure efficacité de reconstruction des $KS0$ sans dégrader la résolution sur le vertex, en particulier dans l'option avec une troisième couche déplacée. Ces résultats ont contribué au document décrivant la conception du VTX et montrent que l'amélioration du détecteur de vertex améliorera les précisions des mesures de physique.

Les résultats présentés dans cette thèse constituent une avancée significative vers des analyses plus sensibles à l'avenir. À mesure que Belle II collecte des ensembles de données de plus en plus volumineux et bénéficie de l'amélioration des performances des détecteurs, le potentiel d'observation des écarts par rapport aux prédictions du Modèle standard continuera de croître. De plus, l'analyse future de l'amplitude du mode partenaire d'isospin permettra de traduire la mesure de l'asymétrie CP en fonction du temps dans les désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ en contraintes significatives sur les modèles de nouvelle physique.

List of Figures

1.1	Unitarity triangle corresponding to the unitarity relation $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$ of the CKM matrix.	8
1.2	Experimental constraints on the sides and angles of the unitarity triangle, (a) by the CKMFitter group [33] and (b) by the UTFit group [34], using the results available in summer 2023.	9
1.3	Feynman diagrams describing the mixing of B mesons <i>via</i> a second-order weak transition. These are also referred to as <i>box diagrams</i>	10
1.4	Main Feynman diagrams contributing to the $B^0 \rightarrow J/\psi K_S^0$ decay. The left diagram illustrates the tree-level contribution, while the right diagram shows the loop-level (penguin) contribution.	17
1.5	Feynman diagram of the $b \rightarrow s\gamma$ quark-level transition in the SM. Since the top quark dominates in the loop, the weak couplings can be taken as $ V_{tb} \sim 1$ and $ V_{ts}^* \sim \lambda^2$. The photon helicity is predominantly left-handed in this transition, while in the $\bar{b} \rightarrow \bar{s}\gamma$ transition, it is right-handed.	20
1.6	Constraints at the two standard deviations level on the real and imaginary parts of the ratio of right- and left-handed Wilson coefficients, C_7'/C_7 . The most stringent constraint comes from the angular analysis of $B^0 \rightarrow K^{*0}e^+e^-$ at very low q^2 . From the LHCb collaboration [60].	23
1.7	Illustration showing the $K_S^0\pi^+ - K_S^0\pi^-$ Dalitz plane. The $m_{K_S^0\pi^+}^2 = m_{K_S^0\pi^-}^2$ bisector splits the plane into two halves and the measured S^+ and S^- parameters are linear combinations of the S parameters measured simultaneously in both halves.	27

1.8	Expected constraints on $\Re(C_7'/C_7)$ and $\Im(C_7'/C_7)$ at the three standard deviations level, derived from the measurement of the new proposed observables S^+ and S^- with 10 ab^{-1} and 50 ab^{-1} of Belle II data. The red, green and blue constraints are derived using three different sets of values for $\{S^+, S^-, a \text{ and } b\}$, where a and b are two dilution-like factors. The dashed (solid) green contour is derived from the angular analysis of $B \rightarrow K^* e^+ e^-$ decays at low q^2 using LHCb Run I data [67] (Run II for the left and Run III for the right plot). Taken from [27].	28
2.1	Schematic view of the SuperKEKB accelerator. The LINAC is shown as well as the HER and LER rings. Collisions occur at the Interaction Point, at the center of the Belle II detector. Taken from [68].	30
2.2	Cross sections of $e^+ e^-$ processes observed at $\sqrt{s} = 10.58 \text{ GeV}$. Cross section numbers taken from [71].	31
2.3	Amount of data (in fb^{-1}) collected per week by Belle II during its first physics run period. From the Belle II collaboration.	33
2.4	Schematic representation of a collision event where an $\Upsilon(4S)$ is produced and decays in a pair of neutral B mesons. The Lorentz boost of the produced $B\bar{B}$ pair ensures they will travel far enough for their individual decay positions to be resolvable. The decay of the <i>tag-side</i> B meson tags the flavor state of the <i>CP-side</i> B meson at the time of the former's decay.	34
2.5	Schematic view of the Belle II detector with all sub-detectors denoted. The Belle II coordinate system is centered at the interaction point of the beams. The z -axis is defined to be nearly collinear to the electron beam direction, the x -axis points horizontally outward from the center of the accelerator rings, and the y -axis is vertical, pointing upwards. The polar angle θ is defined over the interval $[-\pi, +\pi]$, with $\theta = 0$ corresponding to the $+z$ direction and $\theta = \pm\pi$ corresponding to the $-z$ direction. The azimuthal angle ϕ spans the $[0, 2\pi]$ range around the xy -plane. Adapted from Belle II collaboration resources.	35
2.6	Schematic representation of the Pixel Detector with both layers. The gray colored areas correspond to the DEPFET pixel sensors. From the Belle II PXD group.	37
2.7	Cross section of the DEPFET sensor. Taken from [69].	38
2.8	Photograph of the fully assembled VXD before installation.	39
2.9	Schematic of the CDC structure. Taken from [69].	40

2.10	Top: Schematic overview of the TOP counter. Bottom: Schematic side-view of TOP counter illustrating how different particle species lead to Cherenkov photons being emitted at different angles. Taken from [69].	42
2.11	Left: Working principle of the ARICH counter. Right: Photo of the ARICH before installation. From the Belle II ARICH group.	43
2.12	Schematic of the ECL showing the arrangement of thallium-doped CsI crystals in the barrel and endcap regions. From the Belle II ECL group.	45
2.13	Schematic side view of the KLM, located outside the ECL and solenoid. The nominal polar angle acceptance of Belle II is marked by the gray lines that meet at the IP. Taken from [69].	46
2.14	Schematic overview of the L1 Trigger. Information from each sub-trigger is fed to the GDL which makes a trigger decision. Taken from [69].	47
2.15	ΔE distribution for simulated signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ and continuum background events.	53
2.16	M_{bc} distribution for simulated signal $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ and continuum background events.	54
3.1	Transverse view of vertex detector geometries in simulation. Left is the current VXD, together with cooling and mechanical support structure, center is the nominal 5-layer VTX upgrade, and right is an alternate version of the 5-layer VTX geometry with the third layer placed 30 mm closer to the fourth layer. The diameter of the vertex detector in each schematic is ~ 28 cm.	59
3.2	Schematic drawing demonstrating the expected improvement in K_S^0 reconstruction efficiency, as a result of displacing the third layer of the VTX. The left drawing represents the nominal proposed VTX geometry and the right drawing represents the alternative one.	60
3.3	Tracking efficiencies (colored markers) for the SVD+CDC, VTX+CDC and CDC-only detector configuration in each of the considered beam-induced background scenarios and the corresponding occupancy levels (colored bars) in the first layer of the PXD, in the third layer of the SVD, in the first layer of the VTX and in the CDC. Taken from [98].	61
3.4	K_S^0 reconstruction efficiency as a function of the transverse distance traveled by the K_S^0 . Here, the pessimistic BG scenario is taken as an example.	63

3.5	Overlaid Δt residual distributions obtained for each vertex detector geometry, and the ratios between them, shown for the (a) optimistic, (b) intermediate and (c) pessimistic BG scenarios.	65
3.6	Overlaid Δt residual distributions obtained for each BG scenario, and the ratios between them, shown for the (a) VXD, (b) nominal VTX and (c) alternative VTX geometries. Here, v1, v2 and v3 correspond to the optimistic, intermediate and pessimistic BG scenarios, respectively.	66
3.7	Fit of the Δt residual distribution using the sum of three Gaussians for the (a) VXD, (b) nominal VTX and (c) alternative VTX geometries in the pessimistic BG scenario.	67
3.8	Δt resolution indicators for each vertex detector geometry across the three BG scenarios.	68
4.1	Comparison between the $q \cdot r$ distribution of the signal contribution (explained in Section 4.2.5) in 62.8 fb^{-1} of Belle II data and the one of simulated signal events for the category-based Flavor Tagger. Taken from [99].	71
4.2	Comparison between the $q \cdot r$ distribution of the category-based Flavor Tagger and of GFlaT for simulated signal events. The distribution is split based on whether the B_{sig} in the event is a true B^0 (red) or a true $\bar{B}^0$ (blue).	73
4.3	The same comparison that is shown in Figure 4.2 with an additional distinction made based on the presence of charged leptons or charged kaons among the decay products of the B_{tag} in each event.	74
4.4	ΔE shape fits of the simulated (a) signal, (b) BDK and (c) $B\bar{B}$ background components.	80
4.5	ΔE fit of 1 ab^{-1} of fully simulated data (MC) in the (a) signal and (b) sideband regions.	82
4.6	ΔE fit of 1 ab^{-1} of fully simulated data (MC) in the (a) signal and (b) sideband regions in the case of $B^0 \rightarrow D^{(*)-}\pi^+$ decays and in the (c) signal and (d) sideband regions in the case of $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ decays.	83
4.7	ΔE fit of the 365 fb^{-1} Belle II data sample in the (a) signal and (b) sideband regions.	85
4.8	ΔE fit of the 365 fb^{-1} Belle II data sample in the (a) signal and (b) sideband regions in the case of $B^0 \rightarrow D^{(*)-}\pi^+$ decays and in the (c) signal and (d) sideband regions in the case of $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ decays.	86
4.9	Fits of the Δt residual distribution of fully simulated signal events in each of the seven Flavor Tagging r -bins.	90

4.10	Comparison between the background-subtracted $q \cdot r$ distribution observed in data and the simulated signal+BDK contribution for the GNN-based Flavor Tagger. Here the MC distribution is normalised to match the luminosity of the data distribution.	93
4.11	Comparison between the wrong tag fractions (w) obtained from the fit to the background-subtracted Δt distribution of the 1 ab^{-1} simulated data sample (in blue) and their true values known from the MC (in orange).	94
4.12	Comparison between the wrong tagging asymmetries (Δw) obtained from the fit to the background-subtracted Δt distribution of the 1 ab^{-1} simulated data sample (in blue) and their true values known from the MC (in orange).	96
4.13	Comparison between the tag-side detection efficiency asymmetries (a_{tag}) obtained from the fit to the background-subtracted Δt distribution of the 1 ab^{-1} simulated data sample (in blue) and their true values known from the MC (in orange).	96
4.14	Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the first r -bin (i.e. the bin with the worst Flavor Tagging quality) for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).	99
4.15	Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the second r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).	100
4.16	Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the third r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).	101
4.17	Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the fourth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).	102

4.18	Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the fifth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).	103
4.19	Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the sixth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).	104
4.20	Fit of the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample in the seventh r -bin (i.e. the bin with the best tagging quality) for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between opposite-flavor (OF) and same-flavor (SF) events is also shown in (c).	105
4.21	Comparison between the wrong tag fractions (w) obtained from the fit to the background-subtracted Δt distribution of the 365 fb^{-1} data sample and their true values obtained from the MC.	106
4.22	Comparison between the wrong tagging asymmetries (Δw) obtained from the fit to the background-subtracted Δt distribution of the 365 fb^{-1} data sample and their true values obtained from the MC.	106
4.23	Comparison between the tag-side detection efficiency asymmetries (a_{tag}) obtained from the fit to the background-subtracted Δt distribution of the 365 fb^{-1} data sample and their true values obtained from the MC.	106
5.1	Fit for the B^0 lifetime of simulated signal events (a) with and (b) without correcting for the small motion of the B mesons in the $\Upsilon(4S)$ rest frame.	118
5.2	Illustration of continuum (left) and $B\bar{B}$ (right) event topologies, with resulting momenta.	119
5.3	Continuum-suppression classifier: (a) normalised train and test distributions for signal and continuum background candidates, (b) Significance as a function of the cut on the CSMVA classifier output. The dashed red line denotes the cut value for which the significance has reached a maximum.	121
5.4	Candidate multiplicity after the selection of the signal and cut on the CSMVA classifier output have been applied.	122

5.5	Fit of the generated-level $\Delta\tau$ distribution after selecting a single candidate (a) based on the χ^2 probability of the B_{sig} vertex fit and (b) randomly.	123
5.6	Breakdown of $B\bar{B}$ background in the generic MC. The left chart is for $B^0(\bar{B}^0)$ decays, which make up $\sim 27\%$ of the total $B\bar{B}$ background, while the right chart is for $B^+(B^-)$ decays that make up $\sim 73\%$ of the total $B\bar{B}$ background.	126
5.7	Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions between the signal and different background components for simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ events.	130
5.8	Fraction of SCF events over the all signal-like events in each flavor tagging r -bin estimated using the generic MC sample.	131
5.9	Left: Δt residual distribution of simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ signal events overlaid to the Δt Resolution Function that is calibrated using $B^0 \rightarrow D^{(*)-} \pi^+$ decays. Right: Normalized Δt residual distribution of simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ signal events overlaid to the one of simulated $B^0 \rightarrow D^{(*)-} \pi^+$ signal events.	131
5.10	Projections of the fit to the Δt distribution of simulated signal events to extract the CP parameters S and C together with the corresponding CP asymmetry (a) in all flavor tagging r -bins and (b) in the last r -bin.	132
5.11	Projections of the fit to the Δt distribution of simulated signal events to extract the B^0 lifetime (a) in all flavor tagging r -bins and (b) in the last r -bin.	132
5.12	Fits of the (a) ΔE and (b) M_{bc} distributions of simulated signal events.	133
5.13	Projections of the fit to the Δt distribution of simulated SCF background events (a) in all flavor tagging r -bins and (b) in the last r -bin.	135
5.14	Fitted values of S_{SCF} versus the values of S used to generate the signal MC samples. The red solid line corresponds to $S_{\text{SCF}} = S$	136
5.15	Fits of the (a) ΔE and (b) M_{bc} distributions of simulated SCF background events.	137
5.16	Projections of the fit to the Δt distribution of simulated $B\bar{B}$ background events (a) in all flavor tagging r -bins and (b) in the last r -bin.	138
5.17	Projection of the fit to the (a) ΔE and (b) M_{bc} distributions of simulated $B\bar{B}$ background events.	139
5.18	Fit to the Δt distribution of simulated continuum background events.	140

5.19	Fits of the (a) ΔE and (b) M_{bc} distributions of simulated continuum background events.	141
5.20	Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.	145
5.21	Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $ \Delta E < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$	146
5.22	Fit of a data-like MC cocktail sample to extract the B^0 lifetime projected onto Δt	147
5.23	Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$ (Dalitz $\bar{\text{I}}$). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.	150
5.24	Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$ (Dalitz I). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.	151
5.25	Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{\text{I}}$ region and in a signal enhanced region defined by $ \Delta E < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$	152
5.26	Fit of a data-like MC cocktail sample generated with $(S, C) = (0.0, 0.0)$ projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $ \Delta E < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$	153
5.27	Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	155

5.28	Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	156
5.29	Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	157
5.30	Linearity figures for (a) S mean, (b) S residual, (c) S^+ mean, (d) S^- mean, (e) C residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input S values.	163
5.31	2D $\sigma_{\Delta t} - qr$ distribution of events generated for a single toy dataset with $(S, C) = (0.0, 0.0)$ corresponding to the expected luminosity of the Belle II dataset.	164
5.32	Toys: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	166
5.33	Toys: Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	167
5.34	Toys: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	168
5.35	Toys: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.0, 0.6)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	169
5.36	Toys: Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.0, 0.6)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	170
5.37	Toys: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.0, 0.6)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	171

5.38	Toys: Linearity figures for (a) S mean, (b) S residual, (c) S^+ mean, (d) S^- mean, (e) C residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input S values.	178
5.39	Toys: Linearity figures for (a) C mean, (b) C residual, (c) S^+ residual, (d) S^- residual, (e) S residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input C values.	179
5.40	Embedded toys: Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	181
5.41	Embedded toys: Distribution of the fit pull, fit estimate and fit error for (a) S^+ and (b) S^- for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	182
5.42	Embedded toys: Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ for input $(S, C) = (0.6, 0.0)$. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	183
5.43	Embedded toys: Linearity figures for (a) S mean, (b) S residual, (c) S^+ mean, (d) S^- mean, (e) C residual and (f) signal-like yield residual. All plots show the behavior of a given fitted parameter with different input S values.	189
5.44	Comparison of the (a) ΔE and (b) Δt distributions between the off-resonance data and MC samples. The comparison is made before applying the cut on the CSMVA output.	190
5.45	Comparison of the (a) ΔE , (b) M_{bc} and (b) Δt distributions of between the on-resonance data and MC samples. The background components are stacked on top of each other while the signal component is not.	191
5.46	Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample to extract the B^0 lifetime projected onto Δt	192
5.47	Comparison between Flavor Tagger parameters (a) w , (b) Δw and (c) a_{tag} obtained from the calibration using $B^0 \rightarrow D^{(*)-} \pi^+$ events and those obtained using MC truth information in the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ and $B^+ \rightarrow K^+ \pi^+ \pi^- \gamma$ simulated samples.	194

5.48	Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.	198
5.49	Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $ \Delta E < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$	199
5.50	Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays to extract the B^+ lifetime projected onto Δt	200
5.51	Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$ (Dalitz $\bar{\text{I}}$). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.	201
5.52	Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$ (Dalitz I). The different components are overlaid in the ΔE and M_{bc} projections, but not in the Δt projection for clarity.	202
5.53	Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{\text{I}}$ region and in a signal enhanced region defined by $ \Delta E < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$	203
5.54	Fit of a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $ \Delta E < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$	204
5.55	Distribution of the fit pull, fit estimate and fit error for (a) S and (b) C of the control mode. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	205

5.56	Distribution of the fit pull, fit estimate and fit error for (a) $N_{\text{sig-like}}$, (b) $N_{B\bar{B} \text{ bkg}}$ and (c) $N_{q\bar{q} \text{ bkg}}$ of the control mode. In the fit estimate plot, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	207
5.57	Comparison of the (a) ΔE and (b) Δt distributions between the off-resonance data and the off-resonance MC samples for the control mode. The comparison is made before applying the cut on the CSMVA output.	208
5.58	Comparison of the (a) ΔE and (b) M_{bc} distributions of between the on-resonance data and MC samples for the control mode. The background components are stacked on top of each other while the signal component is not.	209
5.59	Fit of the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample to extract the B^+ lifetime projected onto Δt	210
5.60	Fit of the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.	212
5.61	Fit of the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $ \Delta E < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$	213
5.62	Fit of the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+K_S^0}^2 > m_{\pi^-K_S^0}^2$ (Dalitz $\bar{\text{I}}$). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.	214
5.63	Fit of the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{\text{I}}$ region and in a signal enhanced region defined by $ \Delta E < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$	215
5.64	Fit of the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+K_S^0}^2 < m_{\pi^-K_S^0}^2$ (Dalitz I). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.	216
5.65	Fit of the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $ \Delta E < 0.1 \text{ GeV}$, $M_{\text{bc}} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$	217

5.66	Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of fixed shape parameters.	220
5.67	Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of Flavor Tagger parameters.	221
5.68	Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of Δt Resolution Function parameters.	222
5.69	Distributions of the residuals of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset 1,000 times, each time with a different set of values for the B^0 lifetime and B^0 - $\bar{B}^0$ oscillation frequency.	223
5.70	Fitted values of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset with the fixed SCF background fractions varied by $\pm 20\%$. The value obtained with the fractions fixed to their nominal value is denoted by the vertical dashed line.	224
5.71	Fitted values of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting the Belle II dataset with $S_{B\bar{B}}$ and $C_{B\bar{B}}$ varied by $\pm 10\%$. The value obtained with $S_{B\bar{B}}$ and $C_{B\bar{B}}$ both fixed to zero is denoted by the vertical dashed line.	225
5.72	Linearity figures for the residuals of (a) C , (b) S , (c) S^+ and (d) S^- . The fitted CP parameters obtained after fixing the yields to their true values are also displayed.	226
5.73	Fitted values of (a) C , (b) S , (c) S^+ and (d) S^- obtained by fitting signal MC samples corresponding to 4 different misalignment scenarios. The value obtained in the case of no misalignment is denoted by the vertical dashed line. Note that the width of the shaded area and the red solid error bars correspond to an uncertainty that is ~ 10 times smaller than the expected statistical uncertainty from fitting the Belle II dataset.	227
5.74	Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt . The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity.	230
5.75	Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events, in a signal enhanced region defined by $ \Delta E < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 and $0.875 < r < 1$	231

- 5.76 Δt distribution of events in the Belle II data sample tagged as B^0 (red) and $\bar{B}^0$ (blue) and asymmetry between them in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$ and $M_{bc} > 5.27 \text{ GeV}/c^2$ 232
- 5.77 Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+ K_S^0}^2 > m_{\pi^- K_S^0}^2$ (Dalitz $\bar{\text{I}}$). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity. 234
- 5.78 Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz $\bar{\text{I}}$ region and in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{bc} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$ 235
- 5.79 Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto (a) ΔE , (b) M_{bc} and (c) Δt in the region of the Dalitz plane where $m_{\pi^+ K_S^0}^2 < m_{\pi^- K_S^0}^2$ (Dalitz I). The different components are overlaid in the M_{bc} and ΔE projections, but not in the Δt projection for clarity. 236
- 5.80 Fit of the $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ data sample projected onto Δt for (a) events tagged as B^0 , (b) events tagged as $\bar{B}^0$ and (c) all events with the associated asymmetry between B^0 and $\bar{B}^0$ tagged events in the Dalitz I region and in a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$, $M_{bc} > 5.27 \text{ GeV}/c^2$ and $0.875 < r < 1$ 237
- 5.81 Δt distribution of events in the Belle II data sample tagged as B^0 (red) and $\bar{B}^0$ (blue) and asymmetry between them in the (a) Dalitz I region and in the (b) Dalitz $\bar{\text{I}}$ within a signal enhanced region defined by $|\Delta E| < 0.1 \text{ GeV}$ and $M_{bc} > 5.27 \text{ GeV}/c^2$ 238
- C.1 Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the first r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c). 252
- C.2 Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the second r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c). 253

C.3	Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the third r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).	254
C.4	Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the fourth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).	255
C.5	Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the fifth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).	256
C.6	Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the sixth r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).	257
C.7	Fit of the background-free Δt distribution of 1 ab^{-1} of fully simulated data projected in the last r -bin for the case where the flavor of the B mesons in the event is the same (a) and opposite (b) at the time of their decay. The asymmetry between same-flavor (SF) and opposite-flavor (OF) events is also shown in (c).	258
C.8	Correlation between the measured Flavor Tagger and Δt Resolution Function parameters using 365 fb^{-1} of Belle II data. All numbers are in %.	259
D.1	(a) Δt and (b) Δt residual distribution of signal MC candidates where the position of the B_{sig} vertex was determined with or without considering the pion tracks coming from the K_S^0 .	261
D.2	Distributions of the input variables of the continuum suppression BDT for true signal and continuum background events in simulation.	262
D.3	Correlations between BDT input variables for true signal (left) and continuum background (right) events in simulation.	263
D.4	Variable importance for the BDT classifier.	263
D.5	Pair-wise correlations between fit variables for true signal events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.	264

D.6	Pair-wise correlations between fit variables for SCF background events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.	265
D.7	Pair-wise correlations between fit variables for $B\bar{B}$ background events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.	266
D.8	Pair-wise correlations between fit variables for continuum background events in simulation. The numbers quoted for the correlations are the Pearson correlation coefficients.	267
D.9	Projections of the fit to the Δt distribution of simulated SCF background events to extract the CP parameters S and C together with the corresponding CP asymmetry in all Flavor Tagging r -bins (left) and in the last r -bin (right).	268
D.10	Projections of the fit to the Δt distribution of simulated $B\bar{B}$ background events to extract the CP parameters S and C together with the corresponding CP asymmetry in all Flavor Tagging r -bins (left) and in the last r -bin (right).	269
D.11	Distribution of simulated events across the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane in a signal-enhanced region with $ \Delta E < 0.1$ GeV and $M_{bc} > 5.27$ GeV/ c^2 found in the generic MC sample.	270
D.12	Distribution of events across the $K_S^0 \pi^+ - K_S^0 \pi^-$ Dalitz plane in a signal-enhanced region with $ \Delta E < 0.1$ GeV and $M_{bc} > 5.27$ GeV/ c^2 found in the Belle II dataset.	271
D.13	Distributions of ΔE , M_{bc} and Δt of simulated signal events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 > m_{K_S^0 \pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 < m_{K_S^0 \pi^-}^2$	272
D.14	Distributions of ΔE , M_{bc} and Δt of simulated SCF background events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 > m_{K_S^0 \pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 < m_{K_S^0 \pi^-}^2$	273
D.15	Distributions of ΔE , M_{bc} and Δt of simulated $B\bar{B}$ background events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 > m_{K_S^0 \pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0 \pi^+}^2 < m_{K_S^0 \pi^-}^2$	274

D.16	Distributions of ΔE , M_{bc} and Δt of simulated continuum background events compared between the two Dalitz plane halves. Black dots are the total MC statistics, red dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 > m_{K_S^0\pi^-}^2$ and blue dots correspond to the region of the Dalitz plane where $m_{K_S^0\pi^+}^2 < m_{K_S^0\pi^-}^2$	275
D.17	Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (-0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	277
D.18	Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (-0.4, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	278
D.19	Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (-0.2, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	279
D.20	Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.0, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	280
D.21	Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.2, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	281
D.22	Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.4, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	282
D.23	Distribution of the fit pull, fit estimate and fit error for (a) S , (b) C , (c) S^+ and (d) S^- , (e) $N_{\text{sig-like}}$, (f) $N_{B\bar{B} \text{ bkg}}$, (g) $N_{q\bar{q} \text{ bkg}}$ and (h) B^0 lifetime for input $(S, C) = (0.6, 0.0)$. In the fit estimate plots, the red vertical line denotes the true value of the parameter and the blue line denotes the mean of the distribution.	283

D.24 Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated signal events between the signal and control mode.	284
D.25 Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated SCF background events between the signal and control mode.	285
D.26 Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated $B\bar{B}$ background events between the signal and control mode.	286
D.27 Comparison of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated continuum background events between the signal and control mode.	287
D.28 Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated signal events for the control mode.	288
D.29 Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated SCF background events for the control mode.	289
D.30 Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated $B\bar{B}$ background events for the control mode. For ΔE and M_{bc} the projections of the 2D KDE are shown.	290
D.31 Fits of the (a) ΔE , (b) M_{bc} and (c) Δt distributions of simulated continuum background events for the control mode.	291
D.32 Comparison between toy MC and generic MC distributions of (a) ΔE , (b) M_{bc} and (c) Δt	292
D.33 Vues schématiques (a) de l'accélérateur SuperKEKB et (b) du détecteur Belle II.	305
D.34 Illustration de l'amélioration attendue de l'efficacité de reconstruction du K_S^0 grâce au déplacement vers l'extérieur de la troisième couche VTX. À gauche : géométrie VTX nominale. À droite : géométrie VTX alternative.	307
D.35 Indicateurs de résolution sur Δt pour chaque géométrie de détecteur de vertex dans les trois scénarios de bruit induit par le faisceau. . .	308
D.36 Comparaison entre la distribution $q \cdot r$ du Flavor Tagger basé sur les catégories et celle de GFlaT pour des événements de signaux simulés. La distribution est divisée selon que le B_{sig} dans l'événement est un B^0 (rouge) ou un $\bar{B}^0$ (bleu).	310
D.37 Ajustement de la distribution de ΔE dans 365 fb^{-1} de données dans les régions (a) du signal et (b) de la bande latérale M_{bc}	312
D.38 Ajustement de la distribution de Δt après soustraction du bruit de fond dans le bin r de plus haute qualité de l'étiquetage pour 365 fb^{-1} de données, montrant (a) les événements de même saveur, (b) les événements de saveurs opposées et (c) l'asymétrie entre eux. . .	313

- D.39 Illustration représentant le plan de Dalitz $K_S^0 \pi^+ - K_S^0 \pi^-$. La bissectrice $m_{K_S^0 \pi^+}^2 = m_{K_S^0 \pi^-}^2$ divise le plan en deux moitiés et les paramètres mesurés S^+ et S^- sont des combinaisons linéaires des paramètres S mesurés simultanément dans les deux moitiés. 315
- D.40 Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur (a) ΔE (b) M_{bc} et (c) Δt . Les différentes composantes sont superposées dans les projections M_{bc} et ΔE , mais pas dans la projection Δt pour plus de clarté. 316
- D.41 Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur Δt pour (a) les événements étiquetés comme B^0 , (b) les événements étiquetés comme $\bar{B}^0$ et (c) tous les événements présentant une asymétrie associée entre les événements étiquetés B^0 et $\bar{B}^0$, dans une région enrichie en signal définie par $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 et $0.875 < r < 1$ 317
- D.42 Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur (a) ΔE (b) M_{bc} et (c) Δt dans la région du plan de Dalitz où $m_{\pi^+ K_S^0}^2 > m_{\pi^- K_S^0}^2$ (Dalitz $\bar{\text{I}}$). Les différentes composantes sont superposées dans les projections M_{bc} et ΔE , mais pas dans la projection Δt pour plus de clarté. 319
- D.43 Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur Δt pour (a) les événements étiquetés comme B^0 , (b) les événements étiquetés comme $\bar{B}^0$ et (c) tous les événements présentant une asymétrie associée entre B^0 et $\bar{B}^0$ étiquetés dans la région de Dalitz $\bar{\text{I}}$ et dans une région enrichie en signal définie par $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 et $0.875 < r < 1$ 320
- D.44 Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur (a) ΔE (b) M_{bc} et (c) Δt dans la région du plan de Dalitz où $m_{\pi^+ K_S^0}^2 < m_{\pi^- K_S^0}^2$ (Dalitz I). Les différentes composantes sont superposées dans les projections M_{bc} et ΔE , mais pas dans la projection Δt pour plus de clarté. 321
- D.45 Ajustement de l'échantillon de données $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ projeté sur Δt pour (a) les événements étiquetés comme B^0 , (b) les événements étiquetés comme $\bar{B}^0$ et (c) tous les événements présentant une asymétrie associée entre B^0 et $\bar{B}^0$ étiquetés dans la région de Dalitz I et dans une région enrichie en signal définie par $|\Delta E| < 0.1$ GeV, $M_{bc} > 5.27$ GeV/ c^2 et $0.875 < r < 1$ 322

List of Tables

2.1	Specifications of the SVD detector layers. Information taken from [76].	38
2.2	Specifications of the Belle and Belle II CDCs. Information taken from [69].	40
3.1	Specifications of the ladders and sensors of the VXD (PXD+SVD). Information taken from [69, 76].	58
3.2	Specifications of the ladders and sensors of the proposed VTX. The 3a (3b) layer corresponds to the nominal (alternative) VTX geometry. Information taken from [98].	58
3.3	Reconstruction efficiency of $K_S^0 \rightarrow \pi^+\pi^-$ for each of the three vertex detector geometries in three different BG scenarios.	62
3.4	Reconstruction efficiency of $B^0 \rightarrow K_S^0 \pi^+\pi^-\gamma$ for each of the three vertex detector geometries in three different BG scenarios.	64
3.5	Δt resolution for each vertex detector geometry and BG scenario. Here, the weighted average of the three Gaussian widths obtained from fitting the Δt residual distribution using a sum of three Gaussians is quoted as a quantitative indicator of the Δt resolution. . . .	65
4.1	Yields and shape parameters obtained by fitting the ΔE distribution of 1 ab^{-1} of fully simulated data. c_1 and c_2 are parameters describing the 2^{nd} degree polynomial that models the $B\bar{B}$ background and λ is the parameter describing the exponential shape that models the $q\bar{q}$ background. The BDK and $B\bar{B}$ yields are calculated using the signal yield by scaling it by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.	82

4.2	Yields and shape parameters obtained by fitting the ΔE distribution of 1 ab^{-1} of fully simulated data simultaneously for $B^0 \rightarrow D^{(*)-}\pi^+$ and $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ events. a_{sig} , the shift in μ , the scale factor of σ as well as the total signal yield are shared between the two fits. The BDK and $B\bar{B}$ yields are separately calculated for B^0 and $\bar{B}^0$ using the signal yields by scaling them by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.	84
4.3	Yields and shape parameters obtained by fitting the ΔE distribution of the 365 fb^{-1} Belle II data sample. The BDK and $B\bar{B}$ yields are calculated using the signal yield by scaling it by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.	85
4.4	Yields and shape parameters obtained by fitting the ΔE distribution of the 365 fb^{-1} Belle II data sample simultaneously for $B^0 \rightarrow D^{(*)-}\pi^+$ and $\bar{B}^0 \rightarrow D^{(*)+}\pi^-$ events. The BDK and $B\bar{B}$ yields are separately calculated for B^0 and $\bar{B}^0$ using the signal yields by scaling them by the fractions of BDK and $B\bar{B}$ to signal events, respectively, which are fixed from the MC.	87
4.5	Parameters of the resolution function obtained by fitting the Δt residual distribution of fully simulated signal events. When comparing to Equations 4.4 and 4.5, $f_> = f_R \times f_{\text{exp}}$ and $f_< = (1 - f_R) \times f_{\text{exp}}$. Parameters indexed with the number 7 denote the value of that parameter in the last r -bin. The last column shows whether these parameters are left free to float when performing the final Δt fit on the real data.	91
4.6	Flavor Tagger and Δt Resolution Function parameters obtained by fitting the background-subtracted Δt distribution of the 1 ab^{-1} fully simulated sample.	95
4.7	Flavor Tagger and Δt Resolution Function parameters obtained by fitting the background-subtracted Δt distribution of the 365 fb^{-1} Belle II data sample.	98
4.8	Systematic uncertainties related to the process of evaluating ε_{tag} using $B^0 \rightarrow D^{(*)-}\pi^+$ decays recorded by Belle II. The uncertainty due to the sample statistics is also listed for comparison.	108
5.1	List of selection criteria applied to the prompt $\pi^\pm$, K_S^0 , photon and B candidates in the data sample.	116
5.2	Cumulative signal and background efficiencies for the most impactful steps of the selection procedure. Absolute efficiencies are shown in parentheses. All values are estimated using the generic MC sample.	123

5.3	Relative fractions between the signal and SCF background components determined from simulation and further fixed in the fit to the data.	127
5.4	Parameters of the signal ΔE model determined from a fit to simulated signal events.	134
5.5	Parameters of the signal M_{bc} model determined from a fit to simulated signal events.	134
5.6	Parameters of the SCF background Δt model determined from a fit to simulated SCF background events.	135
5.7	Parameters of the SCF background ΔE model determined from a fit to simulated SCF background events.	137
5.8	Parameters of the SCF background M_{bc} model determined from a fit to simulated SCF background events.	138
5.9	Parameters of the $B\bar{B}$ background Δt model determined from a fit to simulated $B\bar{B}$ background events.	139
5.10	Parameters of the continuum background Δt model determined from a fit to simulated continuum background events.	140
5.11	Parameters of the continuum background ΔE model determined from a fit to simulated continuum background events.	141
5.12	Parameters of the continuum background M_{bc} model determined from a fit to simulated continuum background events.	141
5.13	Expected number of events for each of the four components as they are found within a data-like MC cocktail sample. These correspond to 25% of the events found in the generic MC.	143
5.14	Expected number of events for each of the four components as they are found within a data-like MC cocktail sample, specifically in the signal enhanced region used to display the Δt distributions for tagged B^0 and $\bar{B}^0$ events, shown in Figure 5.21.	143
5.15	Resulting parameters of the Dalitz-integrated fit to a data-like MC cocktail sample, generated with $(S, C) = (0.0, 0.0)$. The fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).	144
5.16	Resulting parameters of the Dalitz-split fit to a data-like MC cocktail sample, generated with $(S, C) = (0.0, 0.0)$. Once again, the fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).	149
5.17	Results of the ensemble test using bootstrap samples with varying input S values: Pull mean. All values are expected to be consistent with 0.	158

5.18	Results of the ensemble test using bootstrap samples with varying input S values: Pull sigma. All values are expected to be consistent with 1.	159
5.19	Results of the ensemble test using bootstrap samples with varying input S values: Absolute bias mean. All values are expected to be consistent with 0.	160
5.20	Results of the ensemble test using bootstrap samples with varying input S values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.	161
5.21	Results of the ensemble test using toy datasets with varying input S values: Pull mean. All values are expected to be consistent with 0.	172
5.22	Results of the ensemble test using toy datasets with varying input S values: Pull sigma. All values are expected to be consistent with 1.	173
5.23	Results of the ensemble test using toy datasets with varying input S values: Absolute bias mean. All values are expected to be consistent with 0.	173
5.24	Results of the ensemble test using toy datasets with varying input S values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.	174
5.25	Results of the ensemble test using toy datasets with varying input C values: Pull mean. All values are expected to be consistent with 0.	175
5.26	Results of the ensemble test using toy datasets with varying input C values: Pull sigma. All values are expected to be consistent with 1.	175
5.27	Results of the ensemble test using toy datasets with varying input C values: Absolute bias mean. All values are expected to be consistent with 0.	176
5.28	Results of the ensemble test using toys with varying input C values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.	177
5.29	Results of the ensemble test using embedded toy datasets with varying input S values: Pull mean. All values are expected to be consistent with 0.	184
5.30	Results of the ensemble test using embedded toy datasets with varying input S values: Pull sigma. All values are expected to be consistent with 1.	185
5.31	Results of the ensemble test using embedded toy datasets with varying input S values: Absolute bias mean. All values are expected to be consistent with 0.	186

5.32	Results of the ensemble test using embedded toy datasets with varying input S values: Absolute width. The values represent the expected statistical uncertainty as it is obtained directly from the fit.	187
5.33	Relative fractions between the signal and SCF background components for the control mode determined from simulation and further fixed in the fit to the data.	195
5.34	Expected number of events for each of the four components as they are found within a data-like MC cocktail sample for the control mode. These correspond to 25% of the events found in the generic MC.	195
5.35	Resulting parameters of the Dalitz-integrated fit to a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays. The fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).	196
5.36	Resulting parameters of the Dalitz-split fit to a data-like MC cocktail sample of $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ decays. Once again, the fudge factors in the simulated samples are expected to be close to zero for the means (f^μ) and close to one for the widths (f^σ).	197
5.37	Resulting parameters of the Dalitz-integrated fit to the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample. The central value of the C parameter is blinded to avoid biasing ongoing analyses of the time-integrated CP asymmetry in this mode.	211
5.38	Resulting parameters of the Dalitz-split fit to the $B^+ \rightarrow K^+\pi^+\pi^-\gamma$ data sample. The central value of the C parameter is blinded to avoid biasing ongoing analyses of the time-integrated CP asymmetry in this mode.	218
5.39	Systematic uncertainties for C , S , S^+ and S^- . For each parameter, the total systematic uncertainty is evaluated as the square root of the quadratic sum of the uncertainty due to each source.	228
5.40	Resulting parameters of the Dalitz-integrated fit to the $B^0 \rightarrow K_S^0\pi^+\pi^-\gamma$ data sample. The quoted uncertainties are purely of statistical nature.	229
5.41	Resulting parameters of the Dalitz-split fit to the $B^0 \rightarrow K_S^0\pi^+\pi^-\gamma$ data sample. The quoted uncertainties are purely of statistical nature.	233
D.1	Efficacité de reconstruction de $B^0 \rightarrow K_S^0\pi^+\pi^-\gamma$ pour chacune des trois géométries de détecteur de vertex dans trois scénarios différents de fond induit par le faisceau.	307

Acknowledgements

I would like to express my gratitude to all the individuals without whom this thesis would be vastly different, or perhaps not completed at all.

First, I would like to thank my supervisors, Christian and Isabelle, for introducing me to the wonderful world of particle physics research. Isabelle, in particular, sparked my interest in flavor physics and showed remarkable patience throughout the writing of my thesis. I am grateful for her time and for the invaluable feedback she provided, which undoubtedly improved this manuscript. She was always available to discuss both physics and everyday matters, and provided me with numerous opportunities to present my work, which greatly contributed to my visibility and development as a researcher.

I would like to also thank the other members of the IPHC Belle II group, namely Giulio, who, through all our discussions, greatly improved my intuition for the field of particle physics, and Jérôme, who, besides being the supervisor of my master internship, continued to advise me all throughout my PhD.

Of course, I must also thank the two postdoctoral researchers who I worked closely with during my PhD, Varghese and Pere, as their guidance was essential to bringing this analysis to completion. While Varghese was present during the first year and a half, providing me with the know-how and intuition needed to carry out a time-dependent CP asymmetry measurement at Belle II, Pere was there in the latter half, handling the review process and allowing me the time I needed to finish my manuscript. I really appreciate all of these individuals who treated me as their peer and helped me grow as a researcher. My thanks also go out to all my fellow PhD students for their support in times of need and for all the ping pong and board game sessions; you know who you are. Special thanks to Tristan for showing me how to make beautiful plots.

I would like to also thank Marie-Hélène Schune, Anton Poluektov, Iouri Belikov and Thibaud Humair, for accepting to be part of my thesis committee and for their insightful comments. I am especially grateful to Thibaud, who patiently supervised my work on the Flavor Tagger calibration and contributed significantly to the internal review of the main analysis.

Lastly, I am deeply grateful for my family and friends back in Greece, for their immense support during my time abroad, as well as for reminding me that there is more to life than physics. I know that my time with you back home was never enough, but I tried my best.

Measurement of the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays in the Belle II experiment

Résumé

Cette thèse présente trois études interconnectées au sein de l'expérience Belle II. Premièrement, la performance expérimentale du futur détecteur de vertex VTX est évaluée à l'aide de désintégrations simulées $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$; une géométrie entièrement pixellisée à 5 couches améliore la résolution sur les vertex par rapport à la conception actuelle à 6 couches et reste robuste face aux bruits de fond induits par les faisceaux à très haute luminosité instantanée. Deuxièmement, le Flavor Tagger de Belle II et les paramètres de la fonction de résolution en Δt sont calibrés en utilisant directement les données. Le nouveau Flavor Tagger basé sur des GNN montre une amélioration de 18% de la performance d'étiquetage. Enfin, l'asymétrie CP dépendante du temps dans les désintégrations $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ est mesurée à l'aide de 365 fb^{-1} de données du Run 1, résultant en les paramètres de CP $C = -0.29 \pm 0.13 \pm 0.06$ et $S = -0.36 \pm 0.16 \pm 0.05$, ainsi que les premières déterminations des observables $S^+ = -0.72 \pm 0.31 \pm 0.12$ et $S^- = 0.70 \pm 0.30 \pm 0.05$. Ces résultats permettront de contraindre d'éventuelles contributions de physique au-delà du Modèle Standard dans les transitions $b \rightarrow s\gamma$.

Mots-clés: Belle II, SuperKEKB, VTX, Flavor Tagging, GNN, Physique des Particules, Modèle Standard, mésons B , violation de CP , transitions $b \rightarrow s\gamma$, Nouvelle Physique

Résumé en anglais

This thesis presents three interconnected studies within Belle II. First, the physics performance of the future VTX vertex detector is evaluated using simulated $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays; a fully pixelated 5-layer geometry improves the decay vertex resolution compared to the current 6-layer design and remains robust against beam-induced backgrounds at high instantaneous luminosity. Second, the Belle II Flavor Tagger and Δt Resolution Function parameters are calibrated using data-driven methods, with the newly developed GNN-based Flavor Tagger achieving an 18% improvement in tagging performance. Finally, the time-dependent CP asymmetry in $B^0 \rightarrow K_S^0 \pi^+ \pi^- \gamma$ decays is measured using 365 fb^{-1} of Run 1 data, yielding $C = -0.29 \pm 0.13 \pm 0.06$ and $S = -0.36 \pm 0.16 \pm 0.05$, together with the first determinations of the observables $S^+ = -0.72 \pm 0.31 \pm 0.12$ and $S^- = 0.70 \pm 0.30 \pm 0.05$. These results will allow to constrain possible contributions in $b \rightarrow s\gamma$ transitions from physics beyond the Standard Model.

Keywords: Belle II, SuperKEKB, VTX, Flavor Tagging, GNN, Particle Physics, Standard Model, B mesons, CP violation, $b \rightarrow s\gamma$ transitions, New Physics

