Measurement of Time-Dependent $\boldsymbol{CP}$ Violation in $\boldsymbol{B^0 \to K_S^0 \pi^0 \gamma}$ Decays and Contributions to Flavour-Tagging Calibration at Belle II

A Thesis

Submitted to the
Tata Institute of Fundamental Research, Mumbai
Subject Board of Physics
for the degree of Doctor of Philosophy

by

Rishabh Mehta

School of Natural Sciences
Tata Institute of Fundamental Research

Final Version Submitted in September, 2026

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

The work was done under the guidance of Professor Gagan Mohanty at the Tata Institute of Fundamental Research, Mumbai.

(Rishabh Mehta)

In my capacity as the formal supervisor of record of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

(Prof. Gagan Mohanty)

Date: Sept. 1st, 2026

Acknowledgements

First, my supervisor, Prof. Gagan Mohanty. He took me on as a project student in the middle of the COVID-19 pandemic, when opportunities of any kind were hard to come by, and that chance is the reason this thesis exists at all. Since then he has read every draft I have given him, sat through every meeting I needed, and never once lost patience with me even when the analysis did not cooperate. I could not have asked for a better advisor.

Rahul got me started in Belle in the first place. He taught me the software and the conventions when I knew none of it, and stuck with me through this analysis from its messiest early stages right up to the result in this thesis. I owe him a lot.

Soumen and Sagar were the seniors who helped me find my feet when I first joined the group as a project student, at a point when the whole collaboration felt overwhelming. I am grateful for that.

To the DHEP family, Lokesh, Jasmine, Animesh, Ashish, Mukund, Pruthvi, Avik, Santanu, Gursharan, Atul, and Cheshta, thank you for the endless discussions, physics and otherwise, and for just being around on the harder days.

To my friends, Ameya, Manish, Ayush, Erol, Soumya, Sriram, and Gaurav, thank you for the lighthearted moments and for the reminder that there is life outside a PhD.

To my parents, thank you for believing in me and for backing a rather unconventional path after graduation.

Within the Belle II Collaboration, I want to thank Thibaud, Higuchi-san, Michele, and Okubo-san, the TDCPV conveners, for their guidance through the collaboration's review process; Noah, my working-group reader; Xiaodong, the review-committee chair; Giuseppe, my paper-committee reader; Soeren, the physics coordinator; and Bob, the paper-committee chair, for their careful reading and honest feedback at every stage. And to the many others in the collaboration I have not named individually but who answered questions, caught mistakes, and pushed this analysis to be better throughout my PhD: thank you.

Abstract

Violation of the combined charge-conjugation and parity (CP) symmetry is a necessary ingredient for explaining the matter–antimatter asymmetry observed in the Universe, but the amount generated by a single CKM phase in the Standard Model (SM) is far too small to account for it, motivating searches for additional sources of CP violation (CPV) from physics beyond the SM. Radiative $b \to s\gamma$ transitions are loop-mediated and thus highly sensitive to such new physics, particularly right-handed currents that are strongly suppressed in the SM. The decay $B^0 \to K^0_S \pi^0 \gamma$ is a theoretically clean channel for probing new physics via a time-dependent CP asymmetry measurement of the mixing-induced (S) and direct (C) CPV parameters, both predicted to be close to zero in the SM.

This thesis presents several connected pieces of work conducted within the Belle II Collaboration: an early calibration (2022–23) of the mistag parameters of the then-current, BDT-based Flavour Tagger and the decay-time resolution function using hadronic B-meson control decays; a subsequent refactoring of the Flavour Tagger codebase, carried out as TDCPV working-group software liaison; and, as the principal result, a measurement of S and C in the $B^0 \to K^0_S \pi^0 \gamma$ decay using the legacy Belle (711 fb^{-1}) and Belle II (493 fb^{-1}) datasets. This measurement yields $S = 0.09 \pm 0.16 \pm 0.02$ and $C = -0.09 \pm 0.08 \pm 0.04$ in the $K^*(892)$ resonance region, and $S = -0.32 \pm 0.33 \pm 0.09$ and $C = -0.07 \pm 0.17 \pm 0.08$ in the non-$K^*(892)$ region. It improves the precision on the world's best measurement by up to 31% and is consistent with the SM expectation of vanishing CPV. Finally, the thesis describes an ongoing work that extends the CPV programme to a complementary decay $B^0 \to K^0_S \eta \gamma$. My direct contribution is the analysis of the $\eta \to \pi^+\pi^-\pi^0$ channel on Belle and Belle II data, which, combined with another analysis of the $\eta \to \gamma\gamma$ channel, will result in a complementary measurement of CPV in $b \to s\gamma$ transitions.

List of Publications

Papers, notes, and proceedings relevant to this thesis

1. Belle and Belle II Collaborations, "Measurement of time-dependent *CP* violation parameters in $B^0 \to K_S^0 \pi^0 \gamma$ decays at Belle and Belle II," to appear in JHEP, arXiv:2606.04496 [hep-ex].

2. R. Mehta, "$b \to s$ penguin results from Belle and Belle II," to appear in the Proceedings of the 21st International Conference on B-Physics at Frontier Machines (BEAUTY 2026).

3. R. Mehta, R. Tiwary, and G. Mohanty, "Time-Dependent *CP* Violation Measurements in Radiative Penguin Decays of *B* Mesons at Belle and Belle II," Springer Proc. Phys. **432**, 645 (2026) [Proceedings contribution to the 26th DAE-BRNS High Energy Physics Symposium 2024].

4. Cheshta, R. Mehta, S. Dutt, and G. Mohanty, "Measurement of Time-Dependent *CP* Asymmetries in $B^0 \to K_S^0 \eta \gamma$ Decays at Belle and Belle II," Springer Proc. Phys. **432**, 429 (2026) [Proceedings contribution to the 26th DAE-BRNS High Energy Physics Symposium 2024].

5. R. Mehta, "Time-Dependent *CP* Violation Measurements in Radiative Penguin Decays of *B* Mesons at Belle and Belle II," Springer Proc. Phys. **322**, 505 (2026) [Proceedings contribution to the 17th International Conference on Interconnections between Particle Physics and Cosmology, PPC2024].

6. R. Mehta, G. Mohanty, and R. Tiwary, "Measurement of *CP* Asymmetries in $B^0 \to K_S^0 \pi^0 \gamma$ at Belle and Belle II," Belle II internal note, https://docs.belle2.org/pub_data/documents/4576/.

7. T. Humair, R. Mehta, and P. Stavroulakis, "Calibration of the Decay-Time Resolution and Flavor Tagger Using Hadronic *B* Decays Reconstructed in Moriond 2023 Data," Belle II internal note, https://docs.belle2.org/pub_data/documents/495/.

8. Cheshta, S. Dutt, R. Mehta, and G. Mohanty, "Measurement of Time-Dependent CP Asymmetries in $B^0 \to K_S^0\eta\gamma$ at Belle and Belle II," Belle II internal note, https://docs.belle2.org/pub_data/documents/5254/.

Other papers

1. Belle II Collaboration (F. Abudinén *et al.*), "Measurements of the branching fractions for $B \to K^*\gamma$ decays at Belle II," arXiv:2110.08219 [hep-ex].

List of Abbreviations

ACC	Aerogel Cherenkov Counter
ARICH	Aerogel Ring-Imaging Cherenkov detector
AUC	Area Under the (ROC) Curve
BDT	Boosted Decision Tree
BSM	Beyond the Standard Model
CDC	Central Drift Chamber
CKM	Cabibbo–Kobayashi–Maskawa (quark-mixing matrix)
CPV	*CP* Violation
CS	Continuum-Suppression (variable)
ECL	Electromagnetic Calorimeter
FT	Flavour Tagger
GFlaT	GNN-based Flavour Tagger
GNN	Graph Neural Network
IP	Interaction Point
KDE	Kernel Density Estimate
KLM	K^0_L and Muon detector
MC	Monte Carlo (simulation)
MR1	Mass Region 1 ($K^*(892)$ resonant contribution)
MR2	Mass Region 2 (nonresonant + higher kaonic resonances)
MVA	Multivariate Analysis (classifier)
PDF	Probability Density Function
PXD	Pixel Vertex Detector
ROC	Receiver Operating Characteristic
SCF	Self-Cross-Feed
SM	Standard Model
SVD	Silicon Vertex Detector
TD / TI	Time-Dependent / Time-Integrated (fit sample)
TDCPV	Time-Dependent *CP* Violation
TOP	Time-Of-Propagation counter
TOF	Time-Of-Flight (scintillation counter)

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Motivation

1.1 *CP* Violation and Matter–Antimatter Asymmetry

One of the most striking features of the observable Universe is that it is made almost entirely of matter, with negligible traces of primordial antimatter. If the Universe began in a matter–antimatter symmetric state, some dynamical mechanism must have generated this asymmetry. In 1967, Sakharov proposed three necessary conditions for such a baryogenesis to occur dynamically from an initially symmetric state [1]: (i) baryon-number violation, (ii) violation of both charge conjugation (C) and the combined charge-conjugation–parity symmetry (CP), and (iii) departure from thermal equilibrium. The second condition is essential: without CP violation (CPV), any baryon-number-violating process would generate equal numbers of baryons and antibaryons, leaving the net asymmetry unchanged.

The Standard Model (SM) does contain a source for CPV, arising from a single, irreducible phase in the Cabibbo–Kobayashi–Maskawa (CKM) quark-mixing matrix (Section 1.2). However, the amount of CPV this phase can generate through the electroweak baryogenesis mechanism is many orders of magnitude too small to account for the observed baryon-to-photon ratio $\eta_B \approx 6 \times 10^{-10}$ [2], even setting aside the fact that this phase transition is not sufficiently first-order to satisfy the third Sakharov condition [3]. This tension between the observed matter excess and the CPV encoded within the SM is one of the principal motivations for searching for additional, as-yet-unobserved sources of CPV associated with physics beyond the SM (BSM).

CP violation was first observed experimentally in 1964, in the decays of neutral kaons [4]. Since then, the B-meson system has proven to be an especially rich laboratory for CPV studies. The large mass of the B meson permits many competing decay amplitudes, while its relatively long lifetime allows time-dependent effects arising from B^0–$\overline{B}^0$ oscillation to be resolved experimentally at asymmetric-energy e^+e^- colliders operating near the $\Upsilon(4S)$ resonance ("B factories"), as described in Section 1.3. Precision measurements of CPV observables in B-meson decays, including the time-dependent CPV measurement in the radiative decay $B^0 \to K^0_S \pi^0 \gamma$ that forms the central result of this thesis, therefore serve a dual purpose: as stringent, high-precision tests of the SM's CKM mechanism, and as sensitive probes for new sources of CPV that might help explain the cosmological baryon asymmetry.

1.2 The CKM Matrix and Unitarity Triangle

In the SM, the charged-current weak interaction couples the mass eigenstates of the up-type quarks (u, c, t) to the down-type quarks (d, s, b) through the unitary 3×3 CKM matrix [5, 6],

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{1.1}$$

For n quark generations, an arbitrary $n \times n$ complex matrix is described by $2n^2$ real parameters (the real and imaginary parts of each of its n^2 complex entries). Demanding that the matrix be unitary, $V^\dagger_{\text{CKM}} V_{\text{CKM}} = \mathbb{1}$, imposes n^2 real constraints which leaves $2n^2 - n^2 = n^2$ physical parameters, of which $n(n-1)/2$ are mixing angles and the remaining $n^2 - n(n-1)/2 = n(n+1)/2$ are phases. Not all of these phases are physical, however: the $2n$ up and down type quark fields can each be rephased independently, $u_i \to e^{i\alpha_i} u_i$ and $d_i \to e^{i\beta_i} d_i$, without changing any physical prediction. Each such rephasing removes one phase from V_{CKM}. Only the $2n-1$ *relative* phases among these $2n$ redefinitions act on the matrix, since an overall common phase rotation of all $2n$ fields leaves V_{CKM} invariant. Exactly $2n - 1$ of the $n(n+1)/2$ phases can therefore be absorbed this way, leaving

$$n(n+1)/2 - (2n-1) = (n-1)(n-2)/2 \tag{1.2}$$

physical, irreducible phases. For two generations ($n = 2$) this vanishes identically, which is why CPV is entirely absent from a two-generation CKM matrix. The observation of CPV in the kaon system in 1964 [4] was in fact part of the original motivation for postulating a third generation of quarks [6]. For the physical case of three generations ($n = 3$), the matrix is described by 4 parameters: $n(n-1)/2 = 3$ mixing angles, similar to Euler's angles for rotation, and 1 physical phase. It is this single phase that provides the unique source of CPV in the quark sector of the SM. In the Wolfenstein parametrization [7], which expands V_{CKM} in powers of a small parameter $\lambda \equiv |V_{us}| \simeq 0.225$,

$$V_{\text{CKM}} \simeq \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4), \tag{1.3}$$

the imaginary parameter η carries the single physical phase. In this parametrization, η enters V_{ub} and V_{td}. These are the smallest CKM elements, of order λ^3. The phase carried by V_{td}, for example, governs B^0–$\overline{B}^0$ mixing and gives rise to mixing-induced *CP* violation.

Unitarity of V_{CKM} implies several relations among its elements. The relation most relevant to B-meson physics,

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0, \tag{1.4}$$

can be represented as a triangle in the complex plane (the "unitarity triangle") once divided by $V_{cd}V_{cb}^*$. Its interior angles,

$$\phi_1 = \arg\left(-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right), \quad \phi_2 = \arg\left(-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \quad \phi_3 = \arg\left(-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right), \tag{1.5}$$

are all directly accessible through CPV measurements in B-meson decays. The angle ϕ_1 (equivalently β in an alternative labelling convention) is measured with the highest precision, primarily through the "golden channel" $B^0 \to J/\psi\, K_{\mathrm{S}}^0$ (Section 6.4.1). Global fits combining measurements of all three angles and three sides [8, 2] provide a highly overconstrained, consistent picture of the CKM mechanism to date.

Radiative penguin decays such as $B^0 \to K_{\mathrm{S}}^0\pi^0\gamma$ do not directly measure a unitarity-triangle angle, since the leading SM amplitude yields a CPV parameter close to zero (Section 1.4). Instead, their importance lies in testing whether additional decay amplitudes, absent in the SM but generic in many BSM scenarios, contribute at a level detectable over and above this suppressed SM baseline.

1.3 Time-Dependent *CP* Violation at *B* Factories

The study of time-dependent *CP* violation in the radiative decay $B^0 \to K_{\mathrm{S}}^0\pi^0\gamma$ provides a unique probe for physics beyond the SM. The decay occurs predominantly through $b \to s\gamma$ loop transitions, which makes it susceptible to contributions from heavy, virtual particles [9, 10]. As a result, this decay channel can explore physics at energy scales significantly larger than those directly accessible in current collider experiments.

The time-dependent *CP* asymmetry arises from interference between decay amplitudes with and without B^0–$\overline{B}^0$ mixing, where a nonzero phase in the quark-mixing matrix causes differing decay rates for B^0 and $\overline{B}^0$ over time. For coherent $B\overline{B}$ pair production at the $\Upsilon(4S)$ resonance, one B meson (B_{sig}) decays to a final state of $f_{CP}\gamma$, with f_{CP} a *CP* eigenstate, while the other B meson (B_{tag}) decays to a flavour-specific final state; the *CP* asymmetry is then

$$\mathcal{A}_{CP}(\Delta t) = \frac{\Gamma(B_{\mathrm{tag}=B^0}(\Delta t) \to f_{CP}\gamma) - \Gamma(B_{\mathrm{tag}=\overline{B}^0}(\Delta t) \to f_{CP}\gamma)}{\Gamma(B_{\mathrm{tag}=B^0}(\Delta t) \to f_{CP}\gamma) + \Gamma(B_{\mathrm{tag}=\overline{B}^0}(\Delta t) \to f_{CP}\gamma)}. \tag{1.6}$$

Here, Δt is the proper-time difference between the B_{sig} and B_{tag} decays, and $\Gamma(B_{\text{tag}=B^0/\overline{B}^0})$ is the decay rate of B_{sig} to $f_{CP}\gamma$ given that B_{tag} has been identified as a B^0 or $\overline{B}^0$ meson. This asymmetry can be parameterised as

$$\mathcal{A}_{\text{CP}}(\Delta t) = S\sin(\Delta m_d \Delta t) - C\cos(\Delta m_d \Delta t), \tag{1.7}$$

where Δm_d is the mass difference between the two neutral B-meson mass eigenstates, and S and C are the mixing-induced and direct CP-violating parameters, respectively.

Since the B mesons produced at an e^+e^- collider operating around a centre-of-mass energy of 10 GeV travel only a few hundred micrometres before decaying, Δt cannot be measured directly. Instead, it is inferred from the spatial separation Δz between the two decay vertices (Figure 1.1), utilizing the Lorentz boost $\beta\gamma$ associated with the collision. The asymmetric beam energies of the B factories described in Chapter 2 exist specifically to provide this boost. To accurately extract the values of S and C from data, it is essential to have (i) knowledge of the flavour of B_{tag} at the time of its decay, a process known as flavour tagging, and (ii) a precise model of the experimental Δt resolution. Both aspects are elaborated in Chapter 4.

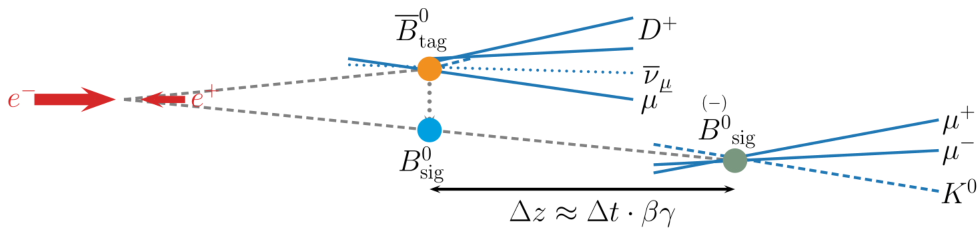


Figure 1.1: Schematic illustration of the boosted $\Upsilon(4S) \to B\overline{B}$ topology exploited in time-dependent CP-violation measurements.

1.4 Radiative $b \to s\gamma$ Transitions

In the SM, the photon emitted in $b \to s\gamma$ transitions is predominantly left-handed due to the chiral structure of the weak interaction. The presence of a polarised photon in the final state suppresses the interference needed for mixing-induced CP violation, leading to small CP-violating effects: S is predicted to be suppressed by a factor of m_s/m_b (the strange-to-bottom quark mass ratio) due to helicity suppression, with deviations expected at the level of a few percent [11]. The SM value of S is calculated to be $(-2.3\pm1.6)\%$ for the resonant $B^0 \to K^{*0}(892)\gamma$ channel, but can be enhanced to $\mathcal{O}(10\%)$ in nonresonant $B^0 \to K^0_S\pi^0\gamma$ decays through long-distance effects, such as charm-loop contributions [12, 13]. The direct CP-violation parameter C is expected to be smaller than 1% [14], although this estimate may not be robust given potentially large uncertainties in predicting strong phases [11].

1.5 New-Physics Sensitivity of $B^0 \to K_S^0\pi^0\gamma$

Any significant contribution from a right-handed photon amplitude would increase S beyond the SM predictions of Section 1.4, providing evidence for new physics such as supersymmetry, extended Higgs sectors, or models with vector-like fermions [15, 16, 17]. A precise measurement of the time-dependent CP asymmetry in $B^0 \to K_S^0\pi^0\gamma$ therefore provides both a stringent test of the SM and a powerful probe for new sources of CP violation, complementary to the unitarity-triangle measurements of Section 1.2: rather than testing the consistency of the known CKM phase, it searches directly for an additional amplitude with a distinct chiral structure that the CKM mechanism alone cannot generate. Since the underlying theoretical mechanisms differ between the $K^{*0}(892)$-dominated region and the remaining phase space [11, 12, 13], this work analyses them separately, denoted MR1 and MR2 respectively (Chapter 5).

1.6 Motivation for This Work

Previous measurements of S and C in this channel from Belle, BaBar, and an earlier Belle II dataset, summarised in Table 1.1, are all statistics-limited, though consistent with SM predictions. The main objective of the work described in this thesis is to improve on this precision using a larger combined Belle and Belle II dataset together with algorithmic improvements: an improved, Graph-Neural-Network-based Flavour Tagger across both experiments (Chapter 4); an improved K_S^0 selection for the Belle dataset; and, for the Belle dataset specifically, the extension of the time-integrated fit component to retain events with poor time resolution, reducing the uncertainty on C (Chapter 6).

Table 1.1: Summary of earlier measurements of S and C for the $B^0 \to K^{*0}(892)\gamma$ and $B^0 \to K_S^0\pi^0\gamma$ (excluding the resonant $K^{*0}(892)$ contribution) channels. The first quoted uncertainty is statistical and the second is systematic.

Channel	Experiment	$N_{\Upsilon(4S)}$	S	C
$K^{*0}(892)\gamma$	Belle [18]	535×10^6	$-0.32^{+0.36}_{-0.33} \pm 0.05$	$0.20 \pm 0.24 \pm 0.05$
	BaBar [19]	467×10^6	$-0.03 \pm 0.29 \pm 0.03$	$-0.14 \pm 0.16 \pm 0.03$
	Belle II [20]	388×10^6	$0.00^{+0.27}_{-0.26} \pm 0.03$	$0.10 \pm 0.13 \pm 0.04$
$K_S^0\pi^0\gamma$, excl. $K^{*0}(892)\gamma$	Belle [18]	535×10^6	$0.50 \pm 0.61 \pm 0.29$	$0.20 \pm 0.37 \pm 0.13$
	BaBar [19]	467×10^6	$-0.78 \pm 0.59 \pm 0.09$	$-0.36 \pm 0.33 \pm 0.04$
	Belle II [20]	388×10^6	$0.04^{+0.45}_{-0.44} \pm 0.23$	$-0.06 \pm 0.25 \pm 0.09$

1.7 Outline of the Thesis

The remainder of this thesis is organized as follows. Chapter 2 describes the Belle and Belle II experiments and the datasets used in this work. Chapter 3 describes the detector-level reconstruction of tracks and calorimeter clusters common to both experiments, and how these are combined with particle-identification information into the standard charged-particle candidates used throughout the rest of this thesis. Chapter 4 presents earlier doctoral contributions to the Belle II flavour-tagging programme: an initial calibration using hadronic control channels, and a subsequent refactoring of the Flavour Tagger software as TDCPV working-group software liaison. Chapters 5–7 present the principal result of this thesis, the measurement of S and C in $B^0 \to K_S^0\pi^0\gamma$ decays: event reconstruction and selection (Chapter 5), the fit strategy and its validation (Chapter 6), and the systematic uncertainties, results, and physics interpretation (Chapter 7). Chapter 8 describes ongoing efforts extending this measurement to the complementary decay $B^0 \to K_S^0\eta\gamma$, while Chapter 9 provides a summary of the thesis and an outlook for the future.

Chapter 2

The Belle and Belle II Experiments

The measurement presented in this thesis combines datasets from two generations of the KEK flavour factory located in Tsukuba, Japan. This chapter introduces each generation of the experiment in turn: first the KEKB collider and the Belle detector it hosted (Section 2.1), then the upgraded SuperKEKB collider and the Belle II detector (Section 2.2). It then discusses material common to both generations: the analysis software used to process and combine the two datasets (Section 2.3), the subdetectors of particular relevance to this analysis (Section 2.4), and the data and simulation samples used (Section 2.5).

2.1 KEKB and Belle

2.1.1 The KEKB Collider

Belle operated at the KEKB asymmetric-energy e^+e^- collider (Figure 2.1) from 1999 to 2010, colliding 8.0 GeV electrons with 3.5 GeV positrons at a centre-of-mass energy of $\sqrt{s} = 10.58$ GeV. The latter corresponds to the mass of the $\Upsilon(4S)$ resonance, which decays almost exclusively to $B\overline{B}$ pairs. The asymmetry between the beam energies imparts a Lorentz boost of $\beta\gamma \approx 0.425$ to the $\Upsilon(4S)$ in the laboratory frame, giving the two B mesons a measurable flight separation before they decay. This is the key experimental handle that makes the time-dependent CP-violation measurements of Section 1.3 possible. Over its operational lifetime of one decade, KEKB delivered, and Belle recorded, an integrated luminosity of 711 fb^{-1} at the $\Upsilon(4S)$ resonance, corresponding to approximately 772×10^6 $B\overline{B}$ events. The full Belle $\Upsilon(4S)$ dataset is used in this work.

2.1.1.1 Beam Profile at the Interaction Point

At KEKB, the electron and positron beams were brought to collision at the interaction point (IP) with a crossing angle of 22 mrad and squeezed by the final-focus magnet system to a small transverse profile, characterised by the horizontal and vertical beta functions at the IP, β_x^* and β_y^*, which set the beam's transverse size for a given emittance.

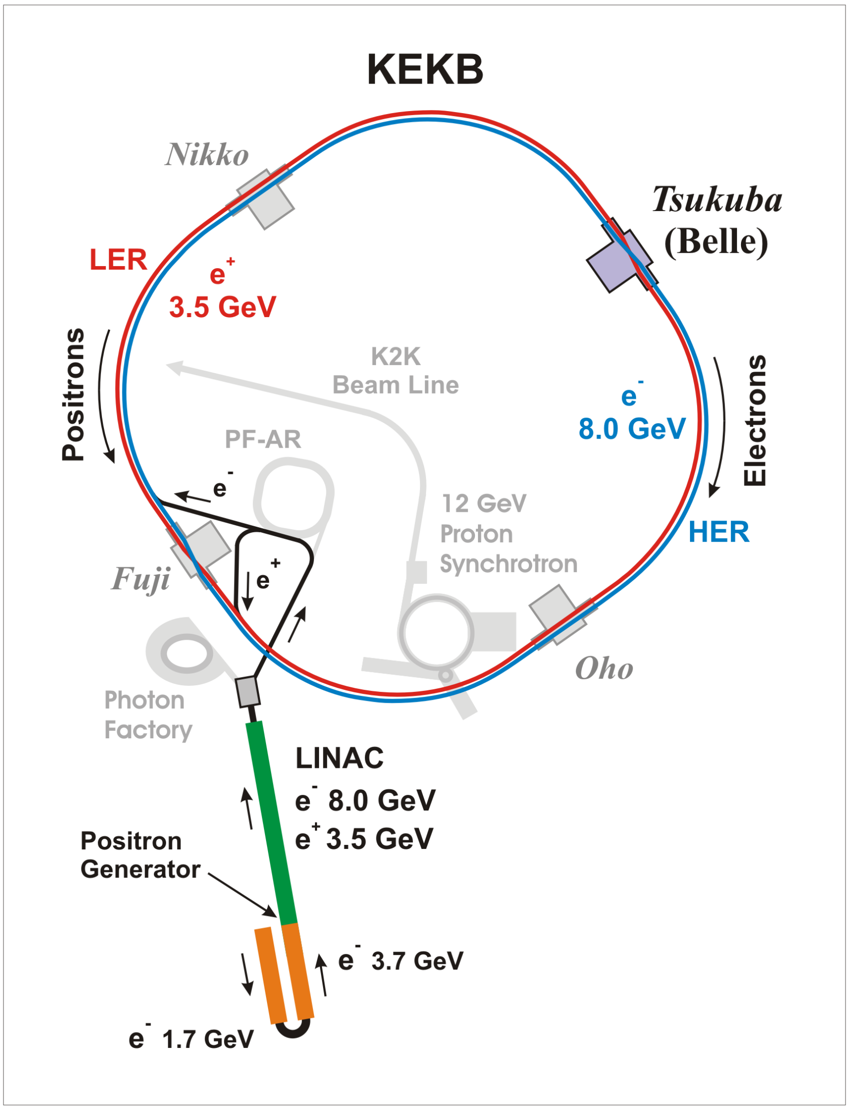


Figure 2.1: Layout of the KEKB accelerator complex, showing the low-energy (positron) and high-energy (electron) storage rings, respectively called LER and HER, the positron generator and LINAC injector, and the location of the Belle detector at the interaction region [21].

Colliding at a finite crossing angle reduces the effective luminosity relative to a true head-on collision, since the two bunches overlap over a shorter effective length. KEKB compensated for this deleterious effect with a novel *crab-crossing* scheme (Fig. 2.2): a superconducting RF deflector cavity in each ring rotates each bunch about its centre by half the crossing angle (11 mrad) immediately before the IP, so that, in the frame of the other beam, the two bunches meet head-on despite their trajectories still crossing at the full 22 mrad angle; a second crab cavity downstream restores each bunch to its original orientation [22]. This recovers the luminosity of the head-on collision while retaining the background-suppression and bucket-filling advantages of a finite crossing angle discussed above. The peak luminosity achievable for given beam currents then scales as

$$\mathcal{L} \propto \frac{\gamma_\pm I_\pm \xi_{y\pm}}{\beta_y^*}, \tag{2.1}$$

where $\gamma_\pm$ is the Lorentz factor, $I_\pm$ the beam current, and $\xi_{y\pm}$ the vertical beam-beam parameter, for each beam species (+ for positrons and − for electrons) [21]. The original design set a target peak luminosity of 1.0×10^{34} cm^{-2}s^{-1}, to be reached with $\beta_y^* = 1.0$ cm and $\xi_y \approx 0.05$ at beam currents of 2.6 A (LER) and 1.1 A (HER) [23]. Over its lifetime, KEKB instead reached this luminosity mainly by squeezing β_y^* substantially below its original design value while increasing the beam currents, ultimately exceeding the design luminosity by a factor of two. Table 2.1 summarises the resulting, final IP beam profile alongside the corresponding SuperKEKB design, discussed further in Section 2.2.1.2.

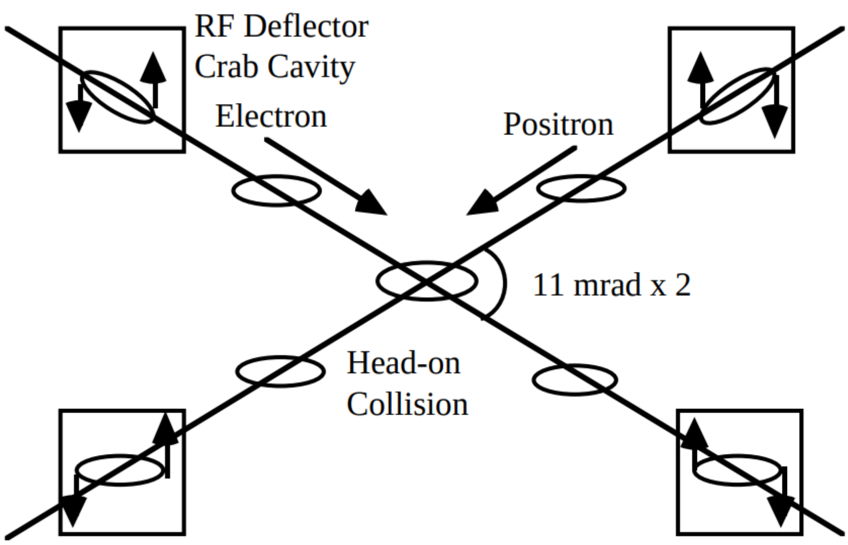


Figure 2.2: Schematic of the KEKB crab-crossing scheme.

2.1.2 The Belle Detector

The Belle detector [24] (Figure 2.3) was a general-purpose spectrometer arranged around the KEKB interaction point and immersed in a 1.5 T magnetic field provided by a superconducting solenoid. Moving outward from the beam pipe, various subsystems of the detector are briefly described below.

2.1.2.1 Beam Pipe

The section of beam pipe traversing the interaction region was a cylindrical, double-walled beryllium tube, originally of 20.0 mm radius, with 0.5 mm-thick beryllium walls separated by a 2.5 mm gap through which helium gas was flowed for cooling [24]; its inner radius was later reduced to 15.0 mm when the silicon vertex detector was upgraded (Section 2.1.2.2) [25]. Beryllium was chosen for its low atomic number, which helps minimize multiple Coulomb scattering and photon conversion for particles originating from B-meson decay vertices close to the IP, directly benefiting the vertex resolution needed for the Δz measurement of Section 1.3.

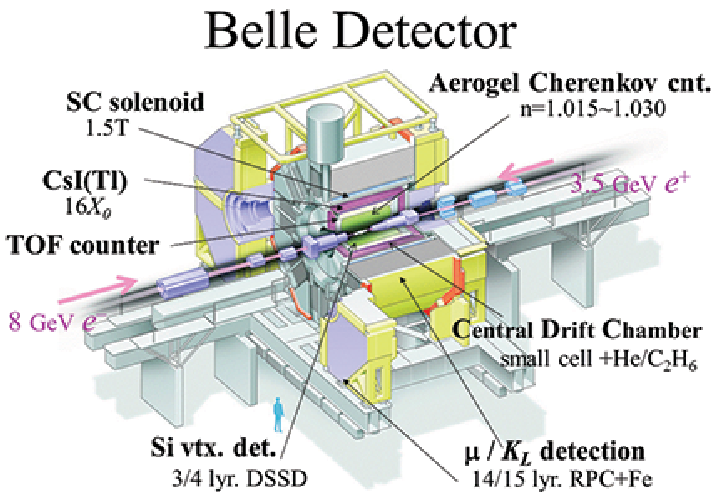


Figure 2.3: A cutaway view of the Belle detector with its subdetector systems [24].

2.1.2.2 Silicon Vertex Detector (SVD)

The SVD used double-sided silicon strip sensors arranged in concentric cylindrical layers immediately outside the beam pipe to precisely measure the position of charged-particle tracks close to their point of origin. The original subdetector, SVD1, comprised three layers (with 8, 10, and 14 ladders respectively) of sensors with a strip pitch of 25.0 µm on the p-side and 50.0 µm on the n-side [24]. Over the lifetime of the experiment the SVD was upgraded once, to SVD2, adding a fourth layer and extending the polar-angle acceptance from 23°–140° to 17°–150°, in order to maintain adequate vertexing performance as KEKB's luminosity and background conditions increased [26].

Since the SVD determines the position of the B-meson decay vertices that enter the Δz measurement of Section 1.3 directly, its momentum-dependent impact parameter resolution, and its improvement from SVD1 to SVD2, is one of the key inputs to this analysis. An early commissioning measurement using cosmic-ray muons compared the transverse ($r\phi$) and longitudinal (z) impact-parameter resolutions of the two subdetectors in terms of a pseudo-momentum-dependent parametrization, finding

$$\begin{aligned} \text{SVD1:}\quad & \sigma_{r\phi} \approx 19.2 \oplus \frac{54.0}{p}\ \mu\text{m}, & \sigma_z \approx 42.2 \oplus \frac{44.3}{p}\ \mu\text{m}, \\ \text{SVD2:}\quad & \sigma_{r\phi} \approx 21.9 \oplus \frac{35.5}{p}\ \mu\text{m}, & \sigma_z \approx 27.8 \oplus \frac{31.9}{p}\ \mu\text{m}, \end{aligned} \tag{2.2}$$

where p is the pseudo-momentum in GeV/c and $\oplus$ denotes the addition in quadrature of a constant term, dominated by the intrinsic sensor and alignment precision, and a momentum-dependent term, dominated by multiple Coulomb scattering in the beam pipe and SVD material itself [26]. Despite SVD2's coarser strip pitch (50.0 µm versus 25.0 µm for SVD1 on the $r\phi$-measuring side), its reduced innermost-layer radius, and the correspondingly reduced multiple scattering, gave it a smaller momentum-dependent term in both coordinates, with the resulting improvement most pronounced at low momentum. The $r\phi$ constant term was slightly larger for SVD2, consistent with its coarser pitch.

2.1.2.3 Central Drift Chamber (CDC)

The CDC was a cylindrical, small-cell multiwire drift chamber filled with a helium+ethane gas mixture, spanning a radius of 8.3–86.3 cm and instrumented with 50 anode (sense-wire) layers and three cathode-strip layers [24]. It provided the primary charged-particle tracking and momentum measurement over the polar-angle range 17°–150°, achieving a transverse position resolution of $\sigma_{r\phi} \approx 130$ µm, a resolution along the wire of $\sigma_z \approx$ 200–1400 µm, and a transverse-momentum resolution of $\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.3\%\sqrt{p_\mathrm{T}^2 + 1}$ (p_T in GeV/c). A supplementary $\mathrm{d}E/\mathrm{d}x$ measurement, with a resolution of 6%, contributed to particle identification together with the ACC and TOF (discussed below), particularly at low momentum.

2.1.2.4 Aerogel Cherenkov Counter (ACC)

The ACC consisted of 960 silica-aerogel counters, segmented into 60 azimuthal cells in the barrel, together with a further 228 counters arranged in five concentric layers in the forward endcap, all pointing towards the IP and read out by photomultiplier tubes. To maintain good pion/kaon separation across the full kinematic range, the aerogel refractive index was varied between 1.015 and 1.030 depending on the polar angle [24]. Charged particles above their Cherenkov threshold for a given radiator produced a typical yield of at least six photoelectrons, giving the ACC good discrimination between charged kaons and pions in the range $1.2 \lesssim p \lesssim 3.5$ GeV/c, above which the CDC's $\mathrm{d}E/\mathrm{d}x$ measurement alone was insufficient.

2.1.2.5 Time-of-Flight Counter (TOF)

The TOF used 128 plastic scintillation counters, each 3.0 m long, arranged at a fixed radius of 1.2 m from the IP and segmented into 128 azimuthal divisions, together with an auxiliary set of 64 thin trigger-scintillation counters mounted just in front of them [24]. With a measured time resolution of $\sigma_t \approx 100\,\text{ps}$, the TOF provided good pion/kaon separation for momenta up to 1.2 GeV/c: combined with the momentum measured by the CDC, this TOF information gave a mass hypothesis for each track, complementing the ACC at these lower momenta.

2.1.2.6 Electromagnetic Calorimeter (ECL)

The ECL was an array of 8736 thallium-doped cesium iodide, CsI(Tl), crystals in a towered geometry, with 6624 in the barrel ($r = 125$–$162\,\text{cm}$) and 1152 (960) in the forward (backward) endcap. Each crystal was deep enough to fully contain electromagnetic showers from photons and electrons in the relevant energy range [24]. It achieved an energy resolution of $\sigma_E/E \approx 1.3\%/\sqrt{E}$ and a position resolution of $\sigma_{\text{pos}} \approx 0.5\ \text{cm}/\sqrt{E}$ (E in GeV). This energy and position resolution made possible the reconstruction of both the prompt photon and the $\pi^0 \to \gamma\gamma$ candidate central to the decay studied in this thesis (Section 5.1).

2.1.2.7 K^0_{L} and Muon Detector (KLM)

The KLM instrumented the iron flux-return yoke of the solenoid with 14 layers of resistive plate chambers (RPCs), two per gap, alternating with 5.0 cm-thick iron plates separated by 4.0 cm gaps [24]. Muons, which penetrate many layers of iron with minimal scattering, were identified by their characteristic hit pattern, with an angular resolution of $\Delta\phi = \Delta\theta \approx 30\,\text{mrad}$ and a hadron fake rate of about 1.0%, while K^0_{L} mesons were identified by the hadronic showers they produced on interacting in the iron.

2.1.2.8 Trigger and Data Acquisition

A hardware-level trigger combined fast signals from the CDC, ECL, TOF, and KLM to select candidate physics events, such as $B\overline{B}$, continuum, and various control-channel topologies, out of the much higher total interaction rate, dominated by quantum electrodynamics (QED) processes, before events were digitized and written to storage by the data-acquisition system.

2.2 SuperKEKB and Belle II

2.2.1 The SuperKEKB Collider

Belle II operates at the upgraded SuperKEKB collider, which collides 7.0 GeV electrons with 4.0 GeV positrons at the same centre-of-mass energy as KEKB, $\sqrt{s} = 10.58\,\text{GeV}$. The accelerator complex (Fig. 2.4) consists of an electron/positron linear injector, a positron damping ring used to reduce the emittance of the LER beam before injection, and the main electron and positron storage rings, which cross at a single interaction region instrumented by the Belle II detector. The analysis reported in this thesis uses up to $493\,\text{fb}^{-1}$ of Belle II data recorded at the $\Upsilon(4S)$ resonance, corresponding to approximately 525×10^6 $B\overline{B}$ events, collected by 2024.

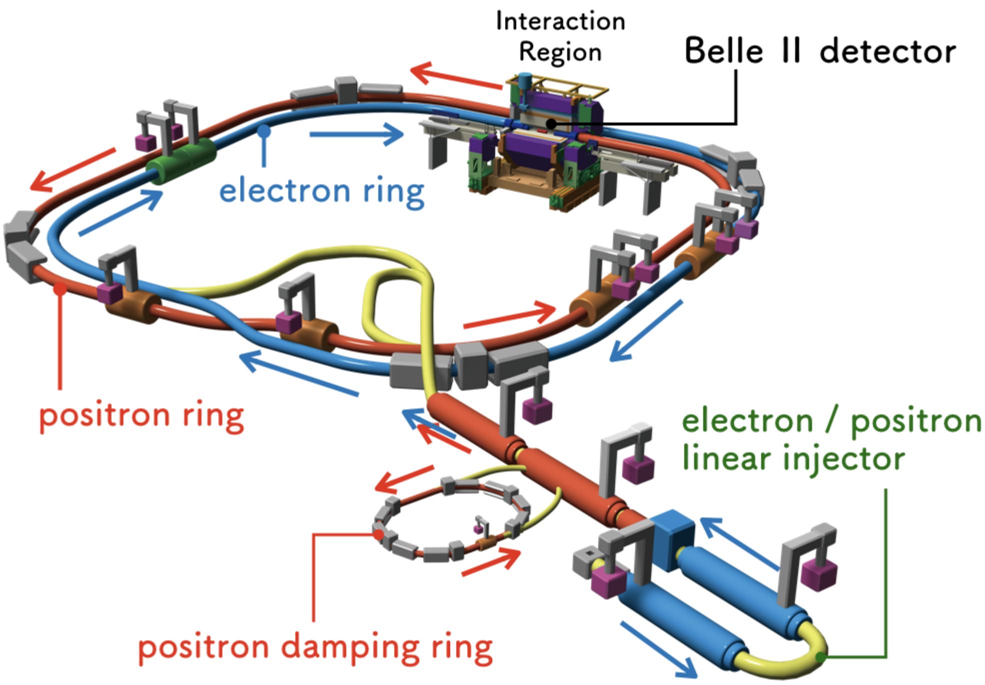


Figure 2.4: Layout of the SuperKEKB accelerator complex [27].

2.2.1.1 The Nano-Beam Scheme

SuperKEKB is the first collider to employ a "nano-beam" final-focus scheme [28]. The peak luminosity of a storage-ring collider scales inversely with the transverse beam size at the IP; at KEKB, this size was limited by the length over which the two bunches overlapped as they crossed at a shallow angle. SuperKEKB instead crosses the beams at a much larger angle and uses strong final-focusing quadrupole magnets, placed closer to the IP than at KEKB, to squeeze both beams to submicron transverse sizes precisely at the crossing region. This confines the luminous region to a far smaller volume than at KEKB for a comparable beam current, allowing a substantially higher luminosity without a proportionate increase in beam-induced background.

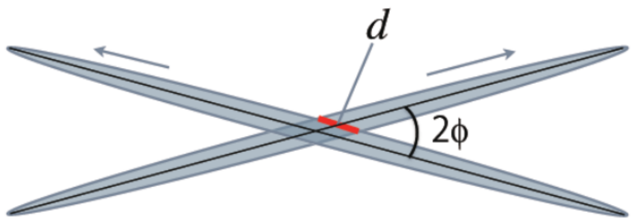


Figure 2.5: Geometry of the nano-beam collision scheme. Two long, thin bunches cross at a large full angle 2ϕ; because each bunch is far longer than its width, the region over which the two bunches physically overlap (shaded red) has a longitudinal extent d that is only a small fraction of the bunch length itself, rather than the full bunch length as in a small-angle, near-head-on crossing [29].

As illustrated in Fig. 2.5, increasing the full crossing angle to 2ϕ shrinks the longitudinal overlap length d between the two long, thin bunches; since it is this overlap region, rather than the full bunch length, that determines the effective longitudinal size relevant to the hourglass effect discussed below (Section 2.2.1.2), a large crossing angle allows β_y^* to be squeezed far beyond what a small-angle, near-head-on geometry like KEKB's could tolerate.

2.2.1.2 IP Beam Profile

Following the luminosity scaling of Eq. 2.1, the nano-beam scheme raises $\mathcal{L}$ primarily by squeezing β_y^* far below what was achievable at KEKB, while simultaneously increasing the beam currents $I_\pm$. Squeezing β_y^* this aggressively at a small crossing angle would cause the beam's longitudinal profile to dominate over its transverse size within the bunch overlap region ("hourglass effect"), degrading rather than improving the luminosity; increasing the full crossing angle to 83 mrad confines the effective overlap length to a small fraction of the bunch length, avoiding this effect and additionally suppressing beam-beam backgrounds outside the immediate IP region [28]. Table 2.1 compares the resulting SuperKEKB design IP beam profile with the values achieved at KEKB (Section 2.1.1.1).

2.2.1.3 Luminosity and Beam-Induced Backgrounds

The nano-beam scheme has allowed SuperKEKB's design luminosity to exceed KEKB's peak luminosity by more than an order of magnitude, and by 2026 it held the world record for peak luminosity at a particle collider, having reached 5.3×10^{34} cm^{-2}s^{-1} [30, 31], en route to a design target of 6.0×10^{35} cm^{-2}s^{-1} and an eventual integrated-luminosity goal of 50 ab^{-1}. This level of luminosity is accompanied by correspondingly higher rates of beam-induced background, principally Touschek scattering, beam-gas interactions, synchrotron radiation, and two-photon QED processes [32]. Together with the intrinsic physics background rate from Bhabha scattering and other QED processes at the $\Upsilon(4S)$, this drove several of the detector upgrades described below relative to Belle, and set the design requirements for the trigger and data-acquisition system of Section 2.2.2.6.

Table 2.1: Final achieved KEKB and design SuperKEKB beam parameters [21, 28].

Parameter	KEKB (achieved)	SuperKEKB (design)
Beam energies e^-/e^+ (GeV)	8.0/3.5	7.0/4.0
Horizontal beta func. β_x^* (mm)	$\sim$1200	32/25 (LER/HER)
Vertical beta func. β_y^* (mm)	$\sim$5.9	0.27/0.30 (LER/HER)
Beam current I, LER/HER (A)	1.64/1.19	2.80/2.00
Number of bunches	$\sim$1600	$\sim$1800
Full crossing angle θ (mrad)	22	83
Peak luminosity (10^{34} cm^{-2}s^{-1})	2.1	60 (design), 5.3 (achieved, 2026)

2.2.2 The Belle II Detector

Belle II [29] (Figure 2.6) follows the same overall concentric layout as Belle, arranged cylindrically around the interaction region and immersed in the same 1.5 T solenoidal field. From the IP outward, the subdetectors are: the pixel detector (PXD) and the SVD, together called the vertex detector (VXD); the CDC; the time-of-propagation (TOP) and aerogel ring-imaging Cherenkov (ARICH) counters; the ECL; and the KLM. Each is described in the following subsections.

2.2.2.1 Vertex Detector: PXD and SVD

The Belle II beam pipe at the IP is a double-walled beryllium cylinder of just 10.0 mm inner radius, half that of Belle's original beam pipe (Section 2.1.2.1). It comprises a 10.0 µm gold layer, a 0.6 mm beryllium wall, a 1.0 mm paraffin coolant layer, and a further 0.4 mm beryllium wall [29]. Immediately outside this beam pipe, a two-layer PXD based on DEPFET (depleted field-effect transistor) active sensor technology provides high-granularity hit information with minimal material budget, essential for resolving decay vertices displaced from the IP by only tens to hundreds of micrometres. It is surrounded by a four-layer SVD of double-sided silicon strip sensors, extending the tracking lever arm out to a radius of a few centimetres and linking VXD hits smoothly onto CDC tracks.

Along with the smaller-radius beam pipe, the PXD and the SVD together replace Belle's single-technology SVD (Section 2.1.2.2), substantially improving the vertex resolution. Design studies project a PXD+SVD combined impact-parameter resolution of

$\sigma_{z_0} \approx 20.0\,\mu\text{m}$ in the high-momentum limit [29], comparable to, or better than, the constant term of Belle's SVD2 resolution (Eq. 2.2), with the larger improvement expected at low momentum, where the closer, thinner PXD layers reduce the multiple-scattering contribution that dominates Belle's resolution. This directly benefits the Δz reconstruction of Section 2.4 and Chapter 5. The second PXD layer was partially installed in 2018, and its full installation was completed in 2024. Due to beam instabilities, the PXD was turned off during the second half of 2024, and the data used in this analysis were partially collected with the PXD off.

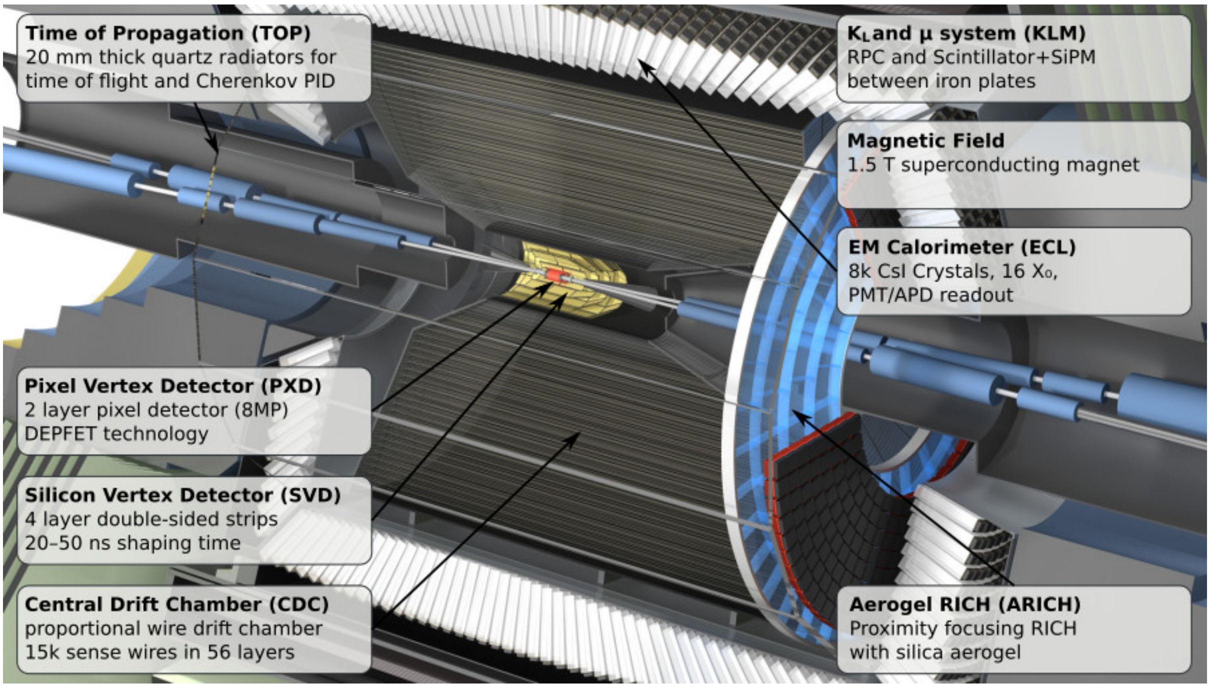


Figure 2.6: A cutaway view of the Belle II detector with its subdetector systems.

2.2.2.2 Central Drift Chamber

The CDC retains the same basic operating concept as Belle's (Section 2.1.2.3): a cylindrical, small-cell helium-ethane drift chamber providing the primary charged-particle tracking, together with a supplementary $\mathrm{d}E/\mathrm{d}x$ measurement for particle identification at low momentum. It is redesigned, however, to cope with SuperKEKB's substantially higher hit occupancy and background rate: it is enlarged to a radius of 16–112 cm (from 8.3–86.3 cm) and a length of -83 cm to $+159$ cm along z, its wire count increased to 56 cylindrical sense-wire layers (32 axial and 24 stereo, replacing Belle's 50 sense-wire and 3 separate cathode-strip layers), and its cell structure and readout electronics retuned accordingly [29]. These changes bring an improved transverse position resolution of $\sigma_{r\phi} \approx 100.0\,\mu\text{m}$ (against Belle's $\approx 130\,\mu\text{m}$) and a $\mathrm{d}E/\mathrm{d}x$ resolution of about 5% (against Belle's 6%), together with a transverse-momentum resolution of $\sigma_{p_\mathrm{T}}/p_\mathrm{T} = \sqrt{(0.2\%\,p_\mathrm{T})^2 + (0.3\%/\beta)^2}$ using CDC information alone, improving to $\sqrt{(0.1\%\,p_\mathrm{T})^2 + (0.3\%/\beta)^2}$ once VXD hits are included in the track fit.

2.2.2.3 Particle Identification: TOP and ARICH

Belle II replaces the ACC and TOF system used in Belle with two dedicated Cherenkov-imaging detectors. In the barrel region, at $r \approx 120\,\mathrm{cm}$, 16 azimuthal segments of the TOP counter each use a 2.75 m-long, 2.0 cm-thick quartz radiator bar: charged particles crossing the quartz emit Cherenkov photons that propagate by total internal reflection to one end of the bar, where their arrival time and position are measured by an array of 4×4-channel micro-channel-plate photomultiplier tubes, yielding around 20 photoelectrons detected per track and an individual-photon timing resolution of $\sigma_t \approx 40\,\mathrm{ps}$, from which the Cherenkov angle, and hence the particle's mass hypothesis, is reconstructed [29]. This gives a K/π separation efficiency above 99% for a pion fake rate below 0.5% in $B \to \rho\gamma$ decays.

In the forward endcap, the ARICH counter uses a 4.0 cm-thick, two-layer focusing aerogel radiator with two different refractive indices, arranged to focus the Cherenkov light cone from each layer onto a common ring on the photon-detector plane, read out by an array of hybrid avalanche photon detectors yielding around 13 photoelectrons detected per track, giving a K/π separation efficiency of 96.0% for a 1.0% pion fake rate at $4\,\mathrm{GeV}/c$. Together, TOP and ARICH maintain kaon/pion discrimination across the full momentum range up to at least $4\,\mathrm{GeV}/c$, comparable to, or exceeding, the combined range of Belle's ACC and TOF (Sections 2.1.2.4–2.1.2.5), despite the substantially higher beam-induced background rates against which they must operate (Section 2.2.1.3). They provide the charged-hadron identification used, together with CDC $\mathrm{d}E/\mathrm{d}x$, in the flavour-tagging algorithm of Chapter 4.

2.2.2.4 Electromagnetic Calorimeter

The ECL reuses the same 8736 CsI(Tl) crystals as Belle (Section 2.1.2.6), each 16.2 radiation lengths deep, covering the barrel and forward/backward endcap regions to give near-hermetic acceptance for photons and electrons [29]. Relative to Belle, the readout electronics have been upgraded to a waveform-sampling design, allowing pulse-shape discrimination that suppresses the impact of beam-induced background pileup on the reconstructed energy and timing of each crystal. The ECL energy and position resolutions enable reconstruction of the $\pi^0 \to \gamma\gamma$ candidate and the prompt signal photon central to the $B^0 \to K^0_S\pi^0\gamma$ decay (Section 5.1).

2.2.2.5 K^0_L and Muon Detector

The KLM instruments the same 14-layer iron flux-return yoke used by Belle (Sect. 2.1.2.7), with the 5.0 cm-iron/4.0 cm-gap structure retained, to sample the showers produced by

K_L^0 mesons interacting hadronically in the iron, and to identify muons by their characteristic penetration through many layers with minimal scattering [29]. It retains the Belle's RPC technology for most of its barrel layers, but replaces it in the endcaps and the innermost two barrel layers, the regions of highest background occupancy, with plastic scintillator strips, read out by wavelength-shifting fibres coupled to silicon photomultipliers, which maintain good efficiency and timing resolution under the higher background rates of SuperKEKB.

2.2.2.6 Trigger and Data Acquisition

At SuperKEKB's design luminosity, the total interaction rate, dominated by QED background processes rather than $\Upsilon(4S)$ decays, exceeds what can be read out and stored in full. This is a substantially tighter constraint than Belle faced (Section 2.1.2.8), given the order-of-magnitude-higher background rates of Section 2.2.1.3. As in Belle, a hardware-based Level 1 trigger uses fast, coarse-grained information from the CDC, ECL, TOP, and KLM to make an initial accept/reject decision within a few microseconds, selecting candidate physics events at a manageable rate. Belle II adds a further, software-based High-Level Trigger stage, running fast online reconstruction similar to that described in Chapter 3, to reduce the rate to a manageable level before events are written to permanent storage.

2.2.3 The Belle II Physics Programme

Belle II pursues a broad physics programme [33], spanning precision *CP*-violation studies and flavour physics in the *B*-meson, charm, and τ-lepton sectors, searches for rare and forbidden decays, and direct searches for light, weakly coupled new particles, exploiting both the clean e^+e^- environment and the very large datasets that SuperKEKB is designed to deliver. The time-dependent *CP*-violation measurements described in this thesis form a major component of this programme.

2.3 Analysis Software

2.3.1 The `basf2` Framework

All Belle II simulation, reconstruction, and much of the offline analysis described in this thesis are performed within the collaboration's common software framework, `basf2` [34]. This provides a unified, version-controlled environment for detector simulation, calibration, event reconstruction, and high-level analysis tools shared across the collaboration, including the Flavour Tagger software described in Chapter 4.

2.3.2 Combining Belle and Belle II Data: `b2bii`

Since this thesis combines data recorded by two generations of detector with different raw formats and reconstruction software, the legacy Belle data and Monte Carlo samples are converted into a `basf2`-compatible format using the dedicated `b2bii` software package [35], which translates Belle `mDST`-format tracks, clusters, and particle-identification information into their Belle II analogues. This allows the same reconstruction, selection, and fitting procedure described in Chapters 5–6 to be applied uniformly to both datasets.

2.4 Subdetectors Relevant to This Analysis

Not every subdetector plays an equal role in the measurements described in Chapters 5–7. The following three classes of subsystem are of particular importance.

Vertex detectors. The SVD in Belle and the combined PXD+SVD system in Belle II are central to reconstructing the spatial separation Δz between the two B-meson decay vertices (Section 1.3), which is the experimentally accessible proxy for the proper-time difference Δt that enters the CP-asymmetry measurement directly. The precision of this vertex reconstruction, and its associated per-event uncertainty, directly determines the decay-time resolution function calibrated in Chapter 4.

Tracking and particle-identification systems. The CDC (both experiments) measures the charged-particle trajectories and momenta, both important for $K^0_S \to \pi^+\pi^-$ reconstruction and $B_{\rm tag}$ vertexing. Together with the ACC+TOF system in Belle and the TOP+ARICH system in Belle II, it also provides particle-identification likelihoods that feed into the multivariate flavour-tagging algorithm described in Chapter 4.

Electromagnetic calorimeter. The ECL, mostly unchanged in concept between the two experiments, is essential for reconstructing both the $\pi^0 \to \gamma\gamma$ candidate and the prompt high-energy photon in the $B^0 \to K^0_S\pi^0\gamma$ signal decay (Section 5.1), and provides the shower-shape and timing information used by several of the background-suppression classifiers described in Section 5.1.

2.5 Data Samples and Monte Carlo Simulation

2.5.1 Data Samples

Both the Belle and Belle II analyses use data recorded at the $\Upsilon(4S)$ resonance, corresponding to the full integrated luminosities listed in Sections 2.1.1 and 2.2.1, together with a smaller off-resonance sample recorded approximately 60 MeV below the $\Upsilon(4S)$

peak, used to characterize the $e^+e^- \to q\bar{q}$, $q \in \{u, d, s, c\}$ continuum background that underlies the continuum-suppression classifiers of Section 5.1. The Belle dataset is processed into the analysis via the `b2bii` conversion of Section 2.3.2, allowing it to be handled uniformly alongside the native Belle II data within `basf2`. For Belle II, approximately 81% of the data sample was collected with both the PXD and SVD operational, with the remaining 19% collected with the PXD switched off [36].

2.5.2 Monte Carlo Event Generation and Detector Simulation

Since this is a blinded analysis, the reconstruction, selection, and fit strategy of Chapters 5–6 are developed and validated entirely on simulated Monte Carlo (MC) samples and data control regions, with the signal region (defined in Section 5.1) in data examined only after these procedures are finalized. Signal and generic $B\bar{B}$ events, from $e^+e^- \to \Upsilon(4S) \to B\bar{B}$, are generated with the EvtGen package [37], while continuum $e^+e^- \to q\bar{q}$, $q \in \{u, d, s, c\}$ events are generated with the KKMC package [38]; both are interfaced to Pythia6 [39] for Belle and Pythia8 [40] for Belle II for parton hadronization. The subsequent detector response is modelled with Geant3 [41] for Belle and Geant4 [42] for Belle II. Both simulated and real data are processed within the `basf2` framework of Section 2.3.1, with Belle samples additionally passed through the `b2bii` conversion of Section 2.3.2 [34, 35, 36].

Table 2.2 summarises the cross sections of the physics processes generated at the $\Upsilon(4S)$ energy for the Belle II generic Monte Carlo production [43]. The continuum and QED processes ($\mu^+\mu^-$, $\tau^+\tau^-$, Bhabha, $\gamma\gamma$, and the two-photon final states) together dominate the total hadronic and leptonic cross section by more than two orders of magnitude over $\Upsilon(4S) \to B\bar{B}$ itself, underscoring why dedicated continuum- and background-suppression classifiers (Section 5.1) are essential to isolate the signal decay.

Table 2.2: Representative cross sections for the physics processes generated at the $\Upsilon(4S)$ energy for the Belle II generic Monte Carlo production [43]. `llXX` and `hhISR` denote, respectively, multilepton and hadronic initial-state-radiation (ISR) final states.

Physics process	Cross section (nb)
$e^+e^- \to \Upsilon(4S) \to B\bar{B}$	1.05
Hadrons from $e^+e^- \to q\bar{q}$ continuum	3.68
$e^+e^- \to \mu^+\mu^-$	1.148
$e^+e^- \to \tau^+\tau^-$	0.919
Bhabha ($e^+e^- \to e^+e^-$)	295.8
$e^+e^- \to \gamma\gamma$	5.10
Two-photon, $e^+e^- \to e^+e^-e^+e^-$	39.55
Two-photon, $e^+e^- \to e^+e^-\mu^+\mu^-$	18.83
`llXX` (multilepton ISR)	2.01
`hhISR` (hadronic ISR)	0.216

2.5.3 Simulated Samples

Large samples of simulated MC events are used throughout the analysis: for signal-efficiency studies, for training and validating the multivariate background-suppression and flavour-tagging classifiers of Chapter 4 and Section 5.1, for deriving fixed probability-density-function shapes in the fit strategy of Chapter 6, and for the closure tests and pseudo-experiment ensembles of Section 6.4. For Belle, these comprise: 1×10^6 signal MC events for both the $K^{*0}(892)$ and nonresonant $B^0 \to K_S^0 \pi^0 \gamma$ decays; six streams of run-dependent continuum $e^+e^- \to q\bar{q}$ MC events, each approximately equivalent to the full data luminosity; ten streams of generic (mixed and charged) $B\overline{B}$ MC events; and dedicated "rare" MC samples enhancing B decays to J/ψ, $\psi(2S)$, $u\ell\nu$, and other rare final states by factors of 20–100, to obtain adequate statistics for otherwise rare peaking backgrounds. The corresponding Belle II samples comprise 1×10^6 signal MC events per mode, four streams of continuum MC events, and four streams of generic $B\overline{B}$ MC events, each again approximately equivalent to the full data luminosity [36]. In addition to these samples, both datasets include dedicated MC samples for the various control channels used for calibration and validation (Section 6.4).

Chapter 3

Tracking, Clustering, Particle ID, and Vertexing

This chapter outlines the detector-level reconstruction common to both experiments. It details how raw hits or energy deposits in the tracking and calorimeter systems, as discussed in Chapter 2, are processed into tracks and clusters (Sections 3.1 and 3.2). It also explains how tracks are combined with particle identification (PID) information to create various charged-particle candidates (Section 3.3), and how these candidates subsequently help reconstruct decay vertices (Section 3.4). Alongside the neutral-particle candidates formed out of clusters, this establishes the common starting point for both the flavour-tagging algorithm of Chapter 4 and the signal-decay reconstruction of Chapter 5.

For Belle II data, the above processing is performed within `basf2` (Section 2.3.1); for Belle data, the equivalent Belle-software objects, reconstructed with an analogous pattern-recognition-plus-Kalman-filter strategy native to the original Belle software, are converted to `basf2`-format tracks, clusters, and PID information by `b2bii` (Section 2.3.2), so that the same higher-level reconstruction and selection procedure can subsequently be applied to both datasets uniformly. Since Belle's own tracking, clustering, and PID algorithms run on different hardware and predate their Belle II counterparts, each of the following sections first describes the Belle II procedure in detail and then summarizes the corresponding Belle-side subdetectors and achieved performance, which `b2bii` carries over unchanged into the common downstream analysis.

3.1 Track Reconstruction

Charged-particle track finding in Belle II proceeds in two largely independent stages, first within the CDC alone and then extending outward into the VXD, before a final track fit is performed [44].

3.1.1 CDC Standalone Track Finding

Two complementary pattern-recognition algorithms are run over the CDC hits for a given event, and their results merged using a multivariate FastBDT classifier. The *global* finder maps each drift-circle hit onto a pair of sinusoidal curves in a Legendre-transformed, (ρ, θ) track-parameter space, where ρ is the trajectory's curvature and θ its azimuthal direction, so that all hits belonging to a common circular trajectory in the transverse plane map to curves intersecting at a single point. A two-dimensional binary search over this space, implemented with a quadtree data structure, locates such intersections efficiently, after which an analogous procedure in the longitudinal (s, z) plane attaches the stereo-layer hits needed to reconstruct the track's polar angle and z position. The *local* finder instead targets short or displaced track segments for which the global method's implicit assumption of a common origin does not hold. This is a category of particular relevance to this analysis, since the K_{S}^0 candidates of Section 5.1 typically decay well outside the IP. It first groups neighbouring hits into short segments using χ^2-weighted hit triplets, and then combines these axial and stereo segments into full tracks via a weighted cellular automaton, a local, iterative algorithm that builds up longer chains of compatible segments by propagating and maximizing a connectivity weight between neighbouring cells, without requiring any assumption about the trajectory's origin.

3.1.2 Extension into the VXD

CDC-only tracks are next extended into the VXD using a Combinatorial Kalman Filter (CKF): each CDC track is iteratively extrapolated outward, using a Runge–Kutta–Nyström integration through the subdetector's inhomogeneous magnetic field and material, and a FastBDT classifier scores candidate SVD or PXD hits at each step based on their spatial and timing compatibility with the extrapolated trajectory. This procedure attaches SVD (PXD) hits with an efficiency of approximately 85% (89%), at a hit purity above 98% (96%) [44]. Low-momentum tracks that curl within, or fail to reach, the CDC's inner radius are instead found independently within the VXD alone by a dedicated algorithm, VXDTF2: each silicon sensor is subdivided into a fine grid of sectors, directed relations between sectors traversed by common MC-trained trajectories are precomputed into a sector map, and background hit combinations are rejected using geometric and timing compatibility filters before a cellular automaton, similar in spirit to the CDC local finder, identifies the longest chains of compatible sectors, with a final triplet-fit quality criterion used to resolve any remaining overlapping candidates. A second CKF pass then pairs any CDC tracks left unmatched by the first pass with these VXDTF2 candidates, so that the two approaches together cover the full momentum range relevant to this analysis.

3.1.3 Final Track Fit

Once a track's full complement of CDC and VXD hits has been assembled, a final fit is performed with the deterministic annealing filter (DAF) implementation of the GENFIT2 toolkit [45], an experiment-independent Kalman-filter-based fitting framework shared with several other experiments; unlike a standard Kalman filter, the DAF iteratively down-weights hits inconsistent with the evolving trajectory, making the fit robust against the occasional mis-assigned hit surviving the pattern-recognition stage. The fit is performed independently under the π, K, and p mass hypotheses, since the small differences in energy loss and multiple scattering between these hypotheses at low momentum can measurably affect the fitted track parameters. This yields the momentum, direction, and point-of-closest-approach, together with its full covariance matrix, used as the starting point for the vertex fits of Section 5.4.

For nominal Belle II beam-background conditions, the overall track-finding and fitting chain achieves an efficiency above 93% for most primary charged particles down to a transverse momentum of $100\,\mathrm{MeV}/c$, and of approximately 90% for the displaced pion tracks of $K^0_S \to \pi^+\pi^-$ decays specifically [44]. Its impact parameters relative to the interaction point (IP) additionally provide a first, loose discriminant against tracks originating from beam-induced background or secondary interactions with detector material, applied at the standard-list-building stage ahead of any decay-specific selection.

3.1.4 Belle Tracking

Belle's own CDC serves the same role as Belle II's, with a broadly similar layout. Pattern recognition is performed independently within the CDC and the SVD, and the resulting hit collections are combined into a single track via a Kalman-filter fit, broadly analogous in spirit to the CDC-plus-VXD combination and DAF-based final fit used for Belle II, though implemented in software native to the original Belle experiment and predating it by close to two decades.

3.2 Cluster Reconstruction

Energy deposits in the ECL are grouped into clusters through a three-stage algorithm designed to maximize the cluster-finding efficiency while containing the full shower energy and avoiding contamination from neighbouring particles or beam background [46]. First, *local maxima* are identified as crystals with an energy above 10.0 MeV that exceeds that of all eight of their immediate neighbours; starting from each local maximum, a *connected region* is grown by iteratively adding neighbouring crystals with energy above 0.5 MeV, continuing the chain outward as long as the newly added crystal itself exceeds 10.0 MeV, and any connected regions that overlap are merged into one. A connected region containing only a single local maximum, the typical case for an isolated, well-separated photon or electron, is reconstructed as a cluster using a 5×5 array of (or 21 excluding the four corner ones) crystals centred on the maximum.

A connected region containing multiple local maxima, corresponding to overlapping showers from nearby particles, is split among the corresponding clusters by an iterative energy-sharing procedure: the energy in each crystal is divided between the candidate clusters according to a distance-dependent weight, where the distance is measured from each crystal to a running estimate of each cluster's centroid, itself computed as a logarithmic-energy-weighted average recalculated after every iteration, until the centroids converge to within 1.0 mm. For both cases, an optimal subset of the highest-energy crystals within the cluster, rather than the full candidate set, is retained for the final energy sum. The number of crystals included is tuned to the estimated per-event noise level (itself estimated from the number of out-of-time, > 5.0 MeV crystals in the event) and the local-maximum energy, and the resulting sum is corrected for the residual shower-leakage and background bias estimated from simulation.

Electron and photon candidates are subsequently formed from the calorimetric clusters after subtracting the expected contribution of any nearby, extrapolated charged track. Their properties, energy, position, and shower-shape variables, together with cluster timing for Belle II (Section 2.2.2.4), are used in the π^0 and prompt-photon reconstruction of Section 5.1, and in several of the background-suppression classifiers described there.

3.3 Particle Identification

Beyond its trajectory, each track is assigned a PID hypothesis among the five stable charged-particle species, namely e, μ, π, K, and p, by combining the response of every subdetector capable of discriminating between them [47, 48]. The CDC's $\mathrm{d}E/\mathrm{d}x$ measurement provides a trimmed, approximately Gaussian-distributed estimator of specific ionization for each hypothesis, obtained by discarding the lowest 5% and highest 25% of individual wire measurements to suppress the effect of long Landau tails, and comparing the resulting mean to its expected, momentum- and polar-angle-dependent value. The TOP and ARICH subdetectors instead construct their likelihoods from individual detected Cherenkov photons. The TOP compares the observed arrival-time distribution of photons in each channel of its photodetector array, relative to the collision time, against the distribution expected for each mass hypothesis, modelled as a sum of weighted Gaussians, while the ARICH, whose individual photodetector pixels are read out only in a binary (fired/not-fired) form, instead evaluates the Poisson probability that each unfired pixel received no Cherenkov photon and each fired pixel received one or more, for the expected Cherenkov ring corresponding to each hypothesis. The ECL contributes through the ratio of cluster energy to track momentum, E/p, which is expected to be close to unity for electrons and characteristically well below it for charged hadrons, modelled using kernel-density-estimated templates. The KLM contributes through the penetration depth and lateral shower shape of associated hits, combined into separate depth and shape likelihoods that are treated as uncorrelated.

For each hypothesis α, the resulting per-subdetector likelihoods $\mathcal{L}_\alpha^d$ are combined into a single global likelihood,

$$\mathcal{L}_\alpha \equiv \prod_d \mathcal{L}_\alpha^d, \tag{3.1}$$

from which either an overall probability for each of the five hypotheses, or a pairwise *binary* likelihood ratio between two specific hypotheses α and β,

$$P_{\alpha/\beta} \equiv \frac{\mathcal{L}_\alpha}{\mathcal{L}_\alpha + \mathcal{L}_\beta}, \tag{3.2}$$

can be formed. The latter, commonly used to discriminate specifically between kaons and pions, is the more common choice for analysis selections, since it depends only on the two hypotheses of interest and is unaffected by a subdetector's inability to distinguish, say, protons from deuterons. When a subdetector's response for a given track is unavailable or judged unreliable, it is simply excluded from the product of Eq. 3.1 by setting $\mathcal{L}_\alpha^d = 1$ for every hypothesis α. A neural network, which combines the per-subdetector log-likelihoods with the track's momentum, direction, and charge, achieves a kaon-identification efficiency of 80–95% and a pion misidentification rate below 5% in the momentum range 0.3–2.0 GeV/c, for a working point of $P_{K/\pi} > 0.8$ [48].

Belle lacks a dedicated imaging detector of the TOP or ARICH type, relying instead on a combination of the CDC's dE/dx measurement with the Cherenkov light yield from the ACC and timing information from the TOF. Kaon/pion separation combines these three measurements into the same product-likelihood and binary-ratio framework of Eqs. 3.1 and 3.2 below; with a likelihood-ratio criterion of $P_{K/\pi} \geq 0.6$, the kaon identification efficiency exceeds 80.0% for a pion misidentification rate below 10.0% in the barrel region [24]. Electron identification instead combines the ECL's E/p and shower-shape response with the CDC dE/dx and ACC light yield, achieving an efficiency of (92.4 ± 0.4)% for a pion misidentification rate of (0.25 ± 0.02)% in the momentum range 1.0–3.0 GeV/c [49].

3.4 Vertex Reconstruction

Decay vertices are reconstructed from their constituent charged-particle candidates, and, where relevant, neutral decay products, using TreeFitter, a global, Kalman-filter-based algorithm that fits an entire decay chain simultaneously, rather than building up individual vertices sequentially from the outside in [50]. Because information is propagated consistently in both directions along the decay tree, a constraint applied at any single point, such as a known intermediate-particle mass, or a requirement that the parent particle originate from a specific production point such as the IP, improves the fitted vertex position and momentum everywhere else in the chain. This includes neutral decay products such as photons or K_L^0 mesons that leave no track, and hence no direct vertex information, of their own. Figure 3.1 illustrates the vertexing for the decay chain central to this thesis, $B^0 \to K_S^0(\to \pi^+\pi^-)\pi^0(\to \gamma\gamma)\gamma$.

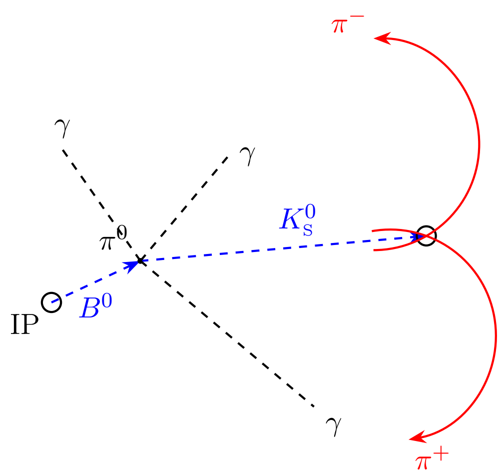


Figure 3.1: Depiction of the $B^0 \to K_S^0(\to \pi^+\pi^-)\pi^0(\to \gamma\gamma)\gamma$ decay chain. The dashed blue line from the IP to B^0 shows the B^0's momentum vector, terminating at its decay vertex; since the π^0 flight length is negligible, its decay vertex coincides with that of the B^0. A further dashed blue line shows the K_S^0 flight trajectory to a displaced vertex. The solid red curves show the trajectory of the K_S^0 decay pions, intersecting at the K_S^0 vertex. The dashed black lines show the extrapolated momenta of the π^0 decay photons and the prompt photon.

For a particle whose decay chain is exclusively reconstructed, an IP constraint can be applied as a point constraint, since the IP profile is known far more precisely than the particle's own decay-vertex resolution. For a particle reconstructed only from the remaining, otherwise-unassigned tracks in an event, however, the transverse vertex position obtained from those tracks alone is often poorly constrained, and a point constraint at the IP provides no additional directional information; instead, a *beam-constrained*, or *IP-tube*, technique exploits overall momentum conservation, inferring the unreconstructed particle's flight direction from the difference between the known total beam momentum and the momentum of the other, exclusively reconstructed particle in the event, and constraining its vertex fit to lie along a narrow tube passing through the IP and oriented along this inferred direction, rather than at a single point [51]. Belle II's substantially smaller and more precisely measured IP profile relative to earlier B factories makes this technique particularly powerful, improving the resulting per-event vertex resolution by roughly a factor of two [51].

Unlike the track-, cluster-, and PID-level algorithms of the preceding sections, vertex reconstruction is not duplicated between the two experiments: since `b2bii` (Section 2.3.2) converts Belle tracks into `basf2`-format objects before any vertex fit is performed, TreeFitter is applied uniformly to both datasets, superseding the Kalman-filter-based helix vertex fit native to the original Belle analysis software for the purposes of this thesis. These generic vertex-fitting techniques underlie every vertex fit performed in this thesis, from the $K^0_S \to \pi^+\pi^-$ vertex and $\pi^0 \to \gamma\gamma$ mass-constrained fit of Section 5.1, to the $B_{\rm sig}$ and $B_{\rm tag}$ vertex reconstruction used to obtain the proper-time difference Δt central to the time-dependent CP-violation measurement (Section 5.4).

Chapter 4

Flavour Tagging and Proper Decay Time

As noted in Section 1.3, extracting the CP-violation parameters S and C from the observed distribution of the proper decay-time difference, Δt, requires (i) a determination, event by event, of the probability that the accompanying "tag-side" B meson decayed as a B^0 or a $\overline{B}^0$ (flavour tagging), and (ii) a probability density function (PDF) describing the resolution with which Δt is reconstructed. This chapter describes two connected pieces of earlier work in this area: an initial calibration of the mistag and resolution-function parameters using hadronic B-decay control channels, conducted for the then-current, boosted-decision-tree (BDT) based Flavour Tagger; and a subsequent refactoring of the Flavour Tagger software itself, carried out as the software liaison for the Belle II TDCPV working group. The Graph-Neural-Network-based Flavour Tagger actually used in the main analysis of Chapters 5–7, and its own control-channel calibration, are introduced at the end of this chapter and described further where applicable.

4.1 Flavour Tagging at B Factories

Flavour tagging is the technique of inferring the heavy-quark flavour content of a neutral meson from the particles produced in its decay. At an e^+e^- flavour factory, the two neutral B mesons produced in $\Upsilon(4S) \to B^0\overline{B}^0$ decay remain in a quantum-entangled state until one of them decays; if the flavour of this first-decaying meson (B_{tag}) can be inferred at that instant, the flavour of the second, not-yet-decayed meson (B_{sig}) is thereby also fixed. Since B_{tag} is not fully reconstructed in a specific decay channel, its flavour must instead be inferred statistically from the charges and identities of particles found in the *rest of the event* (ROE), all reconstructed tracks and clusters not assigned to B_{sig}, using signatures characteristic of flavour-specific b-quark decays, such as the charge of a primary lepton or kaon. A multivariate algorithm combines many such signatures, grouped into physically motivated categories, into a single flavour hypothesis $q_{\mathrm{tag}} = \pm 1$ (for B^0 or $\overline{B}^0$, respectively) together with a dilution factor $r \in [0, 1]$ quantifying the confidence of that hypothesis: $r = 0$ corresponds to no usable flavour information at all, while $r = 1$ corresponds to an unambiguous tag. Table 4.1 lists these categories.

Category	Target for $\overline{B}^0$
Electron	e^-
Intermediate Electron	e^+
Muon	μ^-
Intermediate Muon	μ^+
Kinetic Lepton	ℓ^-
Intermediate Kinetic Lepton	ℓ^+
Kaon	K^-
Kaon-Pion	K^-, π^+
Slow Pion	π^+
Maximum p^*	ℓ^-, π^-
Fast-Slow-Correlated	ℓ^-, π^+
Fast Hadron	π^-, K^-
Lambda	Λ

Table 4.1: The 13 flavour-tagging categories used by the Belle II Flavour Tagger, and their target particle(s) for a true $\overline{B}^0$ tag side ($\ell^\pm$ denotes either $e^\pm$ or $\mu^\pm$, and p^* momentum in the $\Upsilon(4S)$ centre-of-mass frame). Reproduced from Table 1 of Ref. [52].

The following representative $\overline{B}^0$ decay chains illustrate where several of these category targets originate [52]:

- $\overline{B}^0 \to D^{*+}(\to D^0(\to XK^-)\,\pi^+)\,\ell^-\bar{\nu}_\ell$, illustrating the Kinetic Lepton (ℓ^-), Slow Pion (π^+), and Kaon (K^-) targets;
- $\overline{B}^0 \to D^+(\to \overline{K}^0\,\ell^+\nu_\ell)\,\pi^-\,(K^-)$, illustrating the Fast Hadron (π^-, K^-) and Intermediate Kinetic Lepton (ℓ^+) targets;
- $\overline{B}^0 \to \Lambda_c^+(\to \Lambda(\to p\pi^-)\,\pi^+)\,X^-$, illustrating the Lambda (Λ) target.

Since no tagging algorithm is perfect, its response must be calibrated against data with self-tagged control decays: flavour-specific channels whose true tag-side flavour is known independently, for example from the charge of a kaon or pion in the decay. This calibration is expressed through three parameters, evaluated in bins of r: the average wrong-tag probability w, the wrong-tag asymmetry Δw, and the tagging-efficiency asymmetry μ,

$$w = \frac{w_{B^0_{\mathrm{tag}}} + w_{\overline{B}^0_{\mathrm{tag}}}}{2}, \qquad \Delta w = \frac{w_{B^0_{\mathrm{tag}}} - w_{\overline{B}^0_{\mathrm{tag}}}}{2}, \qquad \mu = \frac{\epsilon_{B^0_{\mathrm{tag}}} - \epsilon_{\overline{B}^0_{\mathrm{tag}}}}{\epsilon_{B^0_{\mathrm{tag}}} + \epsilon_{\overline{B}^0_{\mathrm{tag}}}}, \tag{4.1}$$

where $w_{B^0_{\mathrm{tag}}}$ ($w_{\overline{B}^0_{\mathrm{tag}}}$) is the probability of mistagging a true B^0_{tag} ($\overline{B}^0_{\mathrm{tag}}$) decay, and $\epsilon_{B^0_{\mathrm{tag}}}$, $\epsilon_{\overline{B}^0_{\mathrm{tag}}}$ are the corresponding tagging efficiencies. These parameters dilute the physical

CP asymmetry actually observed, as already seen in the signal Δt PDF of Eq. 6.5; a convenient single figure-of-merit encompassing the overall statistical power of a tagging algorithm is its effective tagging efficiency, $\varepsilon_{\text{eff}} = \sum_i f_i(1-2w_i)^2$, summed over the r bin i with fractional population f_i, which is proportional to the statistical precision attainable on S and C for a given event yield.

4.2 Proper Decay-Time Resolution Function

Even with perfect flavour tagging, the finite precision of the vertex reconstruction described in Section 5.4 smears the reconstructed Δt away from its true value, $\Delta t_{\text{res}} \equiv \Delta t - \Delta t_{\text{true}}$. This smearing must be modelled by a resolution function $\mathcal{R}(\Delta t_{\text{res}} \mid \sigma_{\Delta t})$ and convolved with the physical Δt PDF to accurately describe the observed distribution. Physically, three effects are expected to contribute: a dominant Gaussian component reflecting ordinary vertex resolution, whose width scales with the per-event uncertainty $\sigma_{\Delta t}$; an exponential tail component, reflecting B_{tag} vertices reconstructed using tracks from an intermediate, measurably displaced charmed-meson decay (so that the fitted B_{tag} vertex is offset from the true one); and a small, broad outlier component, of width essentially independent of $\sigma_{\Delta t}$, describing rare, poorly reconstructed vertices. The general form

$$\mathcal{R}(\Delta t_{\text{res}} \mid \sigma_{\Delta t}) = (1 - f_{\text{OL}})\,\mathcal{R}^{\text{core}}(\Delta t_{\text{res}}, \sigma_{\Delta t}) + f_{\text{OL}}\,\mathcal{R}^{\text{OL}}(\Delta t_{\text{res}}, \sigma_{\Delta t}), \tag{4.2}$$

where $\mathcal{R}^{\text{core}}$ is a (possibly composite) core resolution term and $\mathcal{R}^{\text{OL}}$ is the outlier term, with relative weight f_{OL}. This functional form of the resolution function is used for the resolution-function calibration described in this chapter, and one without the outlier component is used for the same purpose in the main analysis (Eq. 6.6).

4.3 Calibration with Hadronic Control Channels

As an early doctoral contribution (2022–23), the mistag parameters of the then-current, BDT-based Flavour Tagger and the decay-time resolution function were calibrated using 365 fb^{-1} of Belle II data [53]. The self-tagged hadronic control channels used for this purpose were four neutral B decays, $B^0 \to D^-(\to K^+\pi^-\pi^-)\pi^+$ and $B^0 \to D^{*-}\pi^+$ (with $D^{*-} \to \overline{D}^0\pi^-$ reconstructed in the $\overline{D}^0 \to K^+\pi^-$, $K^+\pi^-\pi^0$, and $K^+\pi^-\pi^+\pi^-$ subchannels), and two charged B decays, $B^+ \to \overline{D}^0(\to K^+\pi^- \,/\, K^+\pi^-\pi^0)\pi^+$. These decays provide high-statistics, effectively flavour-specific samples, well suited to a calibration of this kind. They were combined separately for the neutral and charged subsamples, weighted by their respective branching fractions, into two datasets used for the remainder of the analysis. Alongside 365 fb^{-1} of collision data, dedicated signal MC samples for each individual channel and a 1 ab^{-1}-equivalent generic $B\overline{B}/q\bar{q}$ MC sample (for background studies) were used.

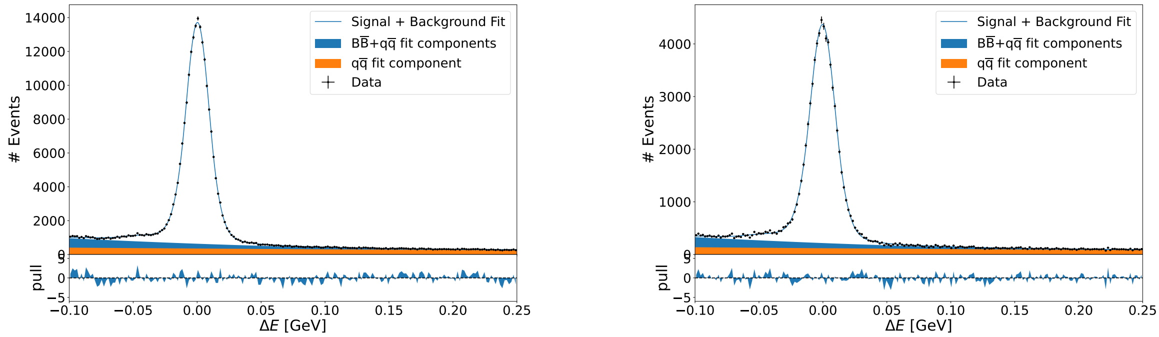
Events were first skimmed with loose requirements on track impact parameters, PID likelihoods, the D/D^* invariant mass, the D^* daughter-pion momentum, and the D^*–D mass difference (Q-value), followed by tighter selections on the two kinematic variables $\Delta E = E^*_B - E^*_{\text{beam}}$ and $M_{\text{bc}} = (1/c^2)\sqrt{(E^*_{\text{beam}})^2 - (\vec{p}^{\,*}_B c)^2}$, requiring $M_{\text{bc}} > 5.20\,\text{GeV}/c^2$ and $\Delta E \in [-0.5, 0.5]\,\text{GeV}$. Continuum background was further suppressed using the normalized second Fox–Wolfram moment [54], $R_2 = H_2/H_0$, where $H_l = \sum_{i,j} |\vec{p}_i||\vec{p}_j| P_l(\cos\theta_{ij})$ is the l-th order Fox–Wolfram moment, summed over pairs of particles i, j in the event, with P_l the l-th Legendre polynomial: this moment quantifies the overall shape of the event's momentum distribution, and is small for the more isotropic $B\overline{B}$ topology and large for jet-like continuum, similarly to the event-shape variables used for continuum suppression in Section 5.6. A *signal region*, enriched in $B\overline{B}$ events ($M_{\text{bc}} > 5.27$ GeV/c^2, $R_2 < 0.4$), and a *sideband*, enriched in continuum ($M_{\text{bc}} < 5.27$ GeV/c^2, $R_2 \in [0.2, 0.4]$), were defined for the subsequent fits.

4.3.1 Signal/Background Separation with the *s𝒫lot* Technique

Since the ultimate quantity of interest, Δt, cannot itself be cleanly separated into signal and background components on an event-by-event basis, the *s𝒫lot* technique [55] was used to statistically unfold the Δt distribution of each component using an uncorrelated *discriminating* variable instead. Here this variable is ΔE, whose distribution for each source of events (signal, $B\overline{B}$ and continuum background) is well understood, and which is required to be uncorrelated with the "control" variable of ultimate interest, Δt, for the technique to be valid. This produces per-event weights (*sWeights*) that, when used to weight the observed Δt distribution, reproduce the distribution each component would have had in isolation, without requiring any prior assumption about the Δt shape itself.

The ΔE shape of correctly reconstructed signal was modelled as the sum of a Gaussian and a double-sided Crystal Ball (DSCB) function, analogous to the single-sided Crystal Ball function of Section 6.2.1 but with independent power-law tails on both sides of the Gaussian core, sharing a common mean, for a total of eight floating shape parameters. A further background specific to hadronic channels, in which a true kaon from the B decay is misidentified as a pion, exhibits a Δt distribution similar to genuine signal and was therefore modelled separately (again with a six-parameter DSCB function) but instructed to be treated as part of the signal component by the *s𝒫lot* algorithm. The remaining, correctly reconstructed $B\overline{B}$ background was modelled with a second-order Chebyshev polynomial, and the continuum background with a single exponential, fitted simultaneously in the signal region and sideband (sharing the same exponential slope) to constrain its shape using the much larger sideband statistics. With the shape of each component fixed from these individual fits, a combined fit to the full dataset, floating only the overall component yields and the exponential slope of the continuum PDF, was used to extract the final sWeights used in the resolution-function and mistag-parameter fits below.

Figure 4.1 shows the combined fit to the full dataset, with all component shapes fixed and only the yields and continuum slope floated, for both generic MC (used to validate the fitting procedure before unblinding the data) and the 365 fb^{-1} Belle II dataset itself.



(a) Generic MC

(b) 365 fb^{-1} Belle II data

Figure 4.1: Combined ΔE fit to (a) generic MC and (b) collision data, with the shape of each component fixed and only the component yields and continuum slope left floating.

4.3.2 Extraction of Mistag and Resolution Parameters

The calibration was performed in seven bins of the dilution factor r, with bin edges

$$r_{\text{edges}} = [0.000,\ 0.100,\ 0.250,\ 0.450,\ 0.600,\ 0.725,\ 0.875,\ 1.000], \tag{4.3}$$

chosen to reflect the differing dominant tagging categories, and hence differing effective resolution and mistag behaviour, populating each bin.

First, a *residual fit* determined the resolution-function parameters alone, via an unbinned, two-dimensional fit to Δt_{res} and $\sigma_{\Delta t}$ in signal MC (using the model of Eq. 4.2), performed simultaneously across all seven r bins. The first six bins shared a common set of parameters. The seventh bin, populated predominantly by semileptonic tagging categories in which at least one lepton track is guaranteed to originate directly from B_{tag}, was instead allowed independent tail parameters, reflecting its measurably different resolution behaviour. Second, a *mixing fit* extracted the mistag parameters themselves: the data were divided into 28 slices, corresponding to the seven r bins and the four combinations of tag-side and signal-side flavour ($q_{\text{tag}}, q_{\text{sig}} = \pm 1$), and a simultaneous two-dimensional unbinned maximum-likelihood fit was performed across all 28 slices, with the resolution-function parameters fixed from the residual fit (except for its tail component, left floating) and w, Δw, and μ floated independently in each r bin; uncertainties were obtained using the HESSE algorithm [56]. This mixing fit was performed on both the sWeighted data and on generic MC, with the fitted MC values consistent with their known generated values, and the sWeighted-data fit providing the calibrated mistag parameters ultimately used, at the time, in TDCPV analyses employing the BDT-based Flavour Tagger.

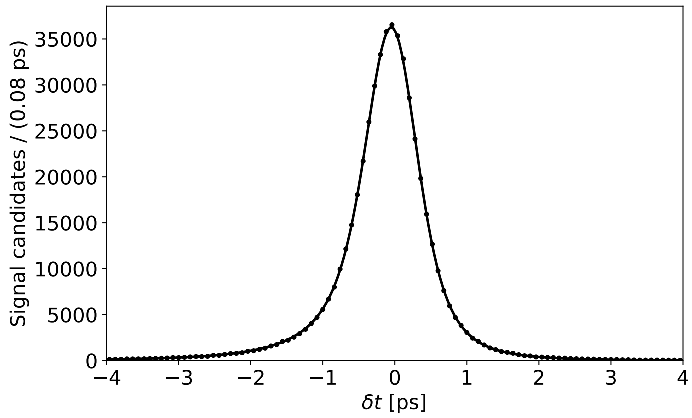
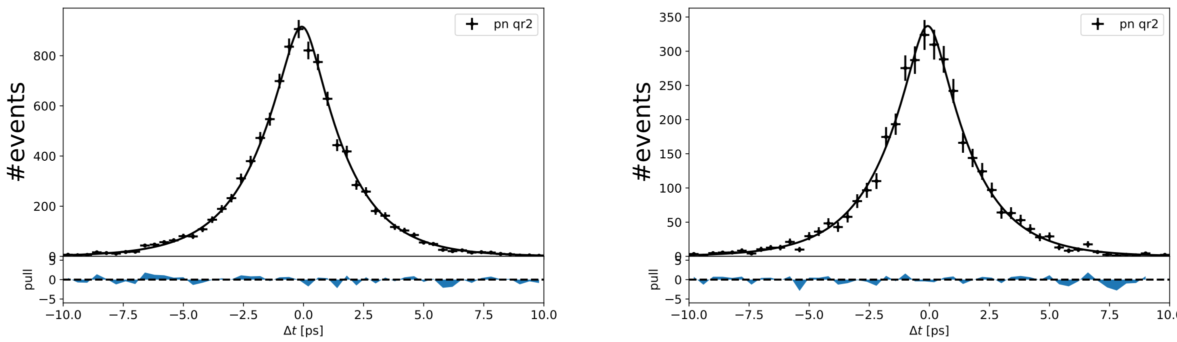


Figure 4.2: Fit to the decay-time residual $\Delta t_{\text{res}}(\delta t) = \Delta t - \Delta t_{\text{true}}$ in signal MC events, combined over all seven r bins of Eq. 4.3.

Figure 4.2 shows the combined residual fit, summed over all seven r bins, together with an example from an individual bin; the sharp, narrow core and long tails extending to several picoseconds illustrate the composite structure of Eq. 4.2. Figure 4.3 shows an example mixing fit, for one combination of r bin and tag/signal flavour, on both generic MC events and sWeighted data.



(a) Generic MC events

(b) sWeighted data

Figure 4.3: Mixing fit, for one representative r bin (3^{rd}) and tag/signal flavour combination ($+1/-1$), on (a) generic MC events and (b) sWeighted collision data.

4.4 Flavour Tagger Software Refactoring

As part of the TDCPV working-group software-liaison duties, the Belle II Flavour Tagger codebase itself was subsequently refactored. The original implementation combined both training/retraining and per-event application logic in a single, undocumented, roughly 1400-line `basf2` script, with no configurable training parameters, making it difficult to maintain, extend, or retrain as the tagger evolved. This work redesigned the package to separate the user-facing application code, which alone remains inside `basf2` and is what is invoked during ordinary event reconstruction (Section 2.3.1), from the expert-side retraining and validation logic. The latter was left for the collaboration to re-implement as a standalone, modular, and documented framework capable of retraining, validating, and updating the corresponding global tags (the collaboration-wide database entries recording which trained tagger version is used for a given data-taking period). The refactored application code was validated to reproduce the output of the original implementation bit-for-bit, and was merged into the Belle II software repository; the companion retraining framework, including full automation of the retraining procedure for the newer, Graph-Neural-Network-based tagger, was left as follow-up work for the working group.

4.5 Graph-Neural-Network-Based Flavour Tagger

The Flavour Tagger used in the principal measurement of this thesis (Chapters 5–7) is a newer, Graph-Neural-Network (GNN)-based algorithm [57], referred to hereafter as the GNN-based Flavour Tagger (GFlaT), which supersedes the BDT-based tagger calibrated in Section 4.3. Rather than processing each tagging category largely independently before combining their outputs, as in the category-based BDT approach, the GFlaT instead represents the particles in the ROE as nodes of a graph and learns correlations between them directly. This allows it to exploit information, such as correlations between multiple particles in the same tagging category, or between different categories, that a category-based approach cannot easily capture. This improves the effective tagging efficiency relative to the earlier BDT-based tagger, as shown in Figure 4.4.

As for the earlier BDT-based tagger, the GFlaT's own mistag parameters were calibrated, in the same seven r bins of Eq. 4.3, against dedicated self-tagged control channels: $B^0 \to J/\psi K^{*0}$ for Belle and $B^0 \to D^{(*)-}\pi^+$ for Belle II. These yield effective tagging efficiencies of $(33.6 \pm 1.2)\%$ and $(37.4 \pm 0.4)\%$, respectively (Section 5.5).

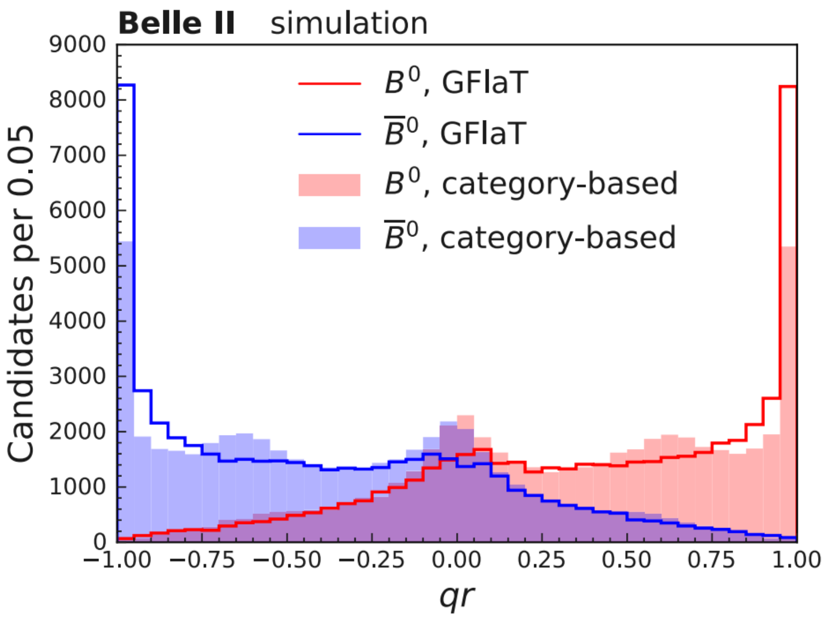


Figure 4.4: Distribution of the flavour-tagger output qr for simulated B^0 and $\overline{B}^0$ candidates in Belle II, comparing the GFlaT (solid lines) with the earlier category-based BDT tagger (filled histograms): the GFlaT concentrates substantially more candidates near $|qr| = 1$, reflecting its higher effective tagging efficiency [57].

Chapter 5

Event Reconstruction and Selection in $B^0 \to K_S^0 \pi^0 \gamma$

This chapter describes how $B^0 \to K_S^0 \pi^0 \gamma$ candidates are reconstructed and selected from the Belle and Belle II datasets. Reconstruction proceeds in a bottom-up manner: a K_S^0 candidate is formed from a pair of oppositely charged tracks assumed to be pions, a π^0 candidate from a pair of ECL photon clusters, and a prompt (high-energy) photon candidate from a further ECL cluster. These three objects are then combined into a B_{sig} candidate, while the remaining tracks and clusters in the event are assigned to B_{tag}.

5.1 Signal Candidate Reconstruction

5.1.1 K_S^0 Selection

K_S^0 candidates are reconstructed from two oppositely charged tracks, assumed to be pions. Both pion candidates are constrained to originate from a common vertex and the resulting invariant mass is required to lie in a window of approximately 3σ around the known K_S^0 mass of 497.6 MeV/c^2 [2], specifically $[468, 528]$ MeV/c^2.

Correctly reconstructed $K_S^0 \to \pi^+\pi^-$ decays need to be distinguished from two dominant sources of misreconstruction: random track combinations, and $\Lambda^0 \to p\pi^-$ decays in which the proton is misidentified as a pion. Both are suppressed using a pair of centrally provided multivariate (MVA) classifiers, each being a single discriminating score computed from several input variables by a trained statistical or machine-learning model, as opposed to a criterion on any one variable alone. The first classifier, a K_S^0-selection MVA, primarily uses the kinematic properties of the K_S^0 candidate and its decay pions together with the K_S^0 flight length (the distance between its decay vertex and the interaction point (IP)), which is a powerful discriminant since genuine K_S^0 mesons typically travel a measurable distance before decaying unlike random track combinations. The second one, a Λ^0-veto MVA (Figure 5.1), uses the proton-hypothesis PID likelihood of each decay pion together with the reconstructed Λ^0 mass obtained by reassigning the proton mass hypothesis to one of the two tracks.

In Belle, these classifiers are based on a NeuroBayes neural network [58], with standard, collaboration-recommended working points of -0.4 on the Λ^0-veto score and 0.5 on the K_S^0-selection (V^0-selection) score. These working points were found to already be close to optimal for this analysis and left unchanged. In Belle II, the analogous classifiers are based on LightGBM [59], a gradient-boosted decision-tree implementation; here the working points of 0.15 on the Λ^0-veto score and 0.9 on the K_S^0-selection score are instead obtained by dedicated optimization of the figure-of-merit

$$\mathrm{FOM} = \frac{N_\mathrm{S}}{\sqrt{N_\mathrm{S} + N_\mathrm{B}}}, \tag{5.1}$$

where N_S and N_B are the expected signal and background yields in the *signal region* used throughout this analysis, defined by $M_\mathrm{bc} > 5.27$ GeV/c^2 and $\Delta E \in [-0.2, 0.1]$ GeV, which retains 80–90% of signal events depending on the mass region; this figure of merit, used throughout this analysis for MVA working-point optimization, is applied here as a distinct background-suppression step (Section 5.6) rather than during preselection itself, and corresponds to background reductions of approximately 7% (for a 2% signal-efficiency loss) and 80% (for a 3% loss), respectively.

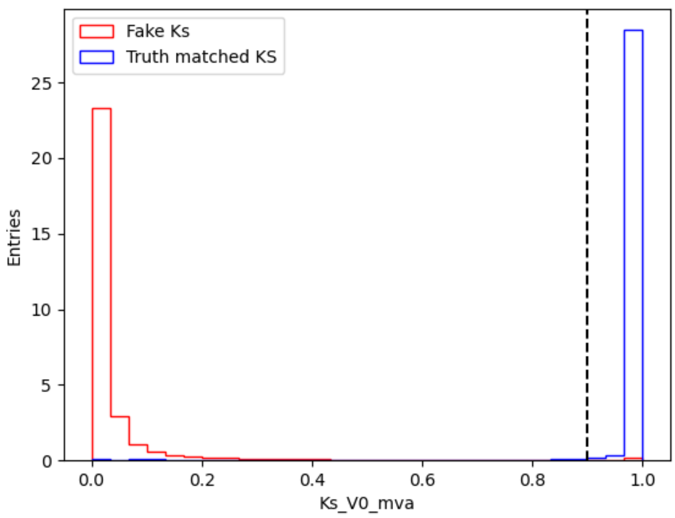


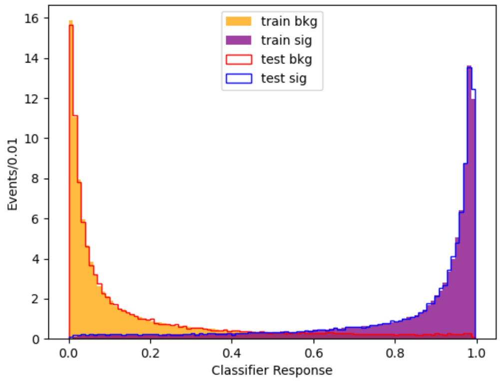
Figure 5.1: Output of the Belle II K_S^0-selection MVA (`Ks_V0_mva`) for genuine, truth-matched K_S^0 candidates and combinatorial ("fake") K_S^0 candidates, showing the strong separation exploited to reject such background; the dashed line indicates the working point used in this analysis.

5.1.2 π^0 Selection

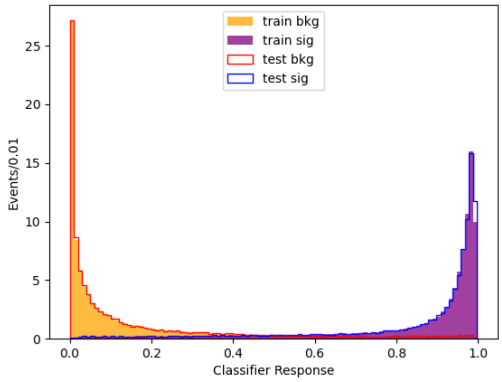
π^0 candidates are reconstructed from pairs of ECL photon clusters, with invariant mass $m_{\gamma\gamma} \in [0.09, 0.16]$ GeV/c^2: the upper edge corresponds to approximately 3σ above the known π^0 mass, while the lower edge is set further away to accommodate a longer tail on the low-mass side of the π^0 peak arising from shower leakage (shower energy that extends beyond the clustered crystals, thereby reducing the reconstructed photon energy). Additional requirements are placed on the number of crystals contributing to each photon cluster (≥ 2) and on the opening angle and azimuthal-angle difference between the two photons (each < 1.3 rad), and, for Belle II, on the ECL cluster timing relative to the beam-crossing time (within ± 200 ns and, additionally, within twice its estimated uncertainty), which suppresses clusters from out-of-time beam-induced background. A mass-constrained fit is then applied to each π^0 candidate, and the χ^2 probability of the fit is used as a discriminant for selecting genuine π^0 candidates.

Since combinatorial photon pairs can also satisfy the above mass window, a dedicated π^0-selection MVA, based on XGBoost [60], is trained to discriminate correctly reconstructed π^0 candidates from combinatorial background. Its features are built from the π^0 candidate itself: its invariant mass, the χ^2 probability of the mass-constrained fit, the cosine of the helicity angle (the angle of one photon relative to the π^0 flight direction, evaluated in the π^0 rest frame), and the opening angle between the two decay photons, with the π^0 transverse momentum in the $\Upsilon(4S)$ frame included as an additional input for Belle. These are combined with, for each photon, a centrally provided fake-photon-suppression score (discriminating genuine ECL clusters from beam-background or hadronic split-off clusters) and either its raw energy (Belle) or a centrally provided beam-background-suppression score (Belle II, reflecting the higher beam-background environment at SuperKEKB). By feature-importance score, the two most powerful discriminants in both experiments are the χ^2 probability of the mass-constrained fit and the photon opening angle, dominant in Belle and Belle II respectively, followed in both cases by the fake-photon-suppression scores of the two decay photons.

The π^0-selection MVA achieves a receiver-operating-characteristic area under the curve (ROC-AUC) of 0.94 (Belle) and 0.95 (Belle II), with its response shown in Figure 5.2. The ROC-AUC is a standard, threshold-independent measure of classifier performance equal to the probability that the classifier ranks a randomly chosen signal candidate above a randomly chosen background one, with 1.0 indicating perfect separation and 0.5 indicating no discriminating power at all. A working-point cut on this score of 0.3 (Belle) and 0.5 (Belle II) removes approximately 75% and 55% of combinatorial background, respectively, for a loss of approximately 10% signal efficiency in the signal region in both cases. Application of this classifier is discussed further in Section 5.6, where the same score is also used for best-candidate selection.



(a) Belle



(b) Belle II

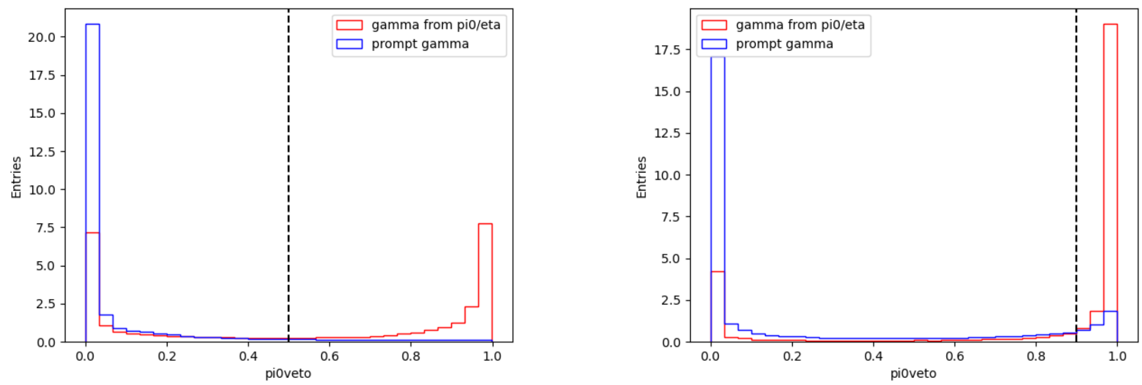
Figure 5.2: Response of the π^0-selection MVA for signal and combinatorial background, in (a) Belle and (b) Belle II, shown separately for the training and test samples.

5.1.3 Prompt Photon Selection

The prompt photon, the high-energy photon emitted directly in the $B^0 \to K_S^0\pi^0\gamma$ transition as opposed to one of the two photons from the π^0 decay, is reconstructed from ECL clusters with an energy in the e^+e^- centre-of-mass frame satisfying $E^* \in [1.4, 3.4]$ GeV, consistent with the kinematics of a two-body-like radiative B decay. Its cluster shape is required to be consistent with a genuine electromagnetic shower via the ratio $E_9/E_{21} > 0.9$, where E_9 and E_{21} are the energies deposited in 3×3 and 5×5 (excluding the four corners) arrays of crystals, respectively, centred on the crystal of highest energy. A photon shower, being compact, deposits nearly all its energy within the smaller region, whereas hadronic interactions or overlapping clusters produce more diffuse energy depositions. As for the π^0 daughters, an ECL cluster-timing requirement is applied for Belle II.

A significant background arises when the prompt-photon candidate is actually one of the two photons from an unrelated, boosted π^0 or η meson elsewhere in the event. This is suppressed with a π^0/η-veto MVA (again XGBoost-based) that attempts to pair the prompt-photon candidate with every other suitably energetic photon in the rest of the event (ROE) and evaluates the invariant mass and fit quality of the resulting π^0 or η candidate; the highest such veto score in the event is retained, and the prompt-photon candidate is rejected if this score is too large. In Belle, this classifier uses the resulting π^0 candidate's invariant mass, the χ^2 probability of its mass-constrained fit, and the cosine of its helicity angle, together with the paired ROE photon's energy and the highest single-crystal energy within its cluster; by feature-importance score, the χ^2 probability is by far the dominant discriminant, roughly an order of magnitude more powerful than the next-most-important feature, the π^0 invariant mass. This classifier achieves an ROC-AUC of 0.97; a working-point criterion at a score of 0.50 removes approximately 50% of background at a cost of 6% signal efficiency. In Belle II, separate, centrally provided

π^0-veto and η-veto classifiers are used instead; working-point criteria of 0.90 and 0.75 on their respective scores remove approximately 70% of background (for 10% signal loss) and 16% of background (for 3% signal loss), respectively; the response of these classifiers is shown in Figure 5.3.



(a) Belle, π^0-veto (b) Belle II, π^0-veto

Figure 5.3: Output of the π^0-veto MVA(s) for genuine prompt photons and for photons that are actually one member of an unrelated π^0 decay: (a) for Belle; (b) for Belle II. Dashed lines indicate the working point used in this analysis, above which the prompt-photon candidate is rejected.

5.2 Kinematic Variables $M_{\rm bc}$ and ΔE

The K^0_S, π^0, and prompt-photon candidates are combined into a $B_{\rm sig}$ candidate, identified using the beam-energy-constrained mass $M_{\rm bc}$ and energy difference ΔE already defined in Section 4.3. Both variables are computed from the $B_{\rm sig}$ momentum obtained after the full $B_{\rm sig}$ vertex fit of Section 5.4, the global decay-chain fit of Section 3.4, applied here with a π^0 mass constraint in addition to the IP constraint, rather than from the pre-fit candidate kinematics; this improves the ΔE resolution. For correctly reconstructed signal, $M_{\rm bc}$ peaks at the known B^0 mass and ΔE peaks at zero. Because the prompt photon is prone to shower leakage in the ECL, the ΔE distribution develops a long tail towards negative values, and since an underestimated photon energy affects both variables, a corresponding correlation develops between $M_{\rm bc}$ and ΔE.

To reduce such effects, a modified definition of $M_{\rm bc}$ is used, in which the π^0 and prompt-photon momenta are rescaled so that their energies sum to $\sqrt{s}/2$ minus the precisely measured K^0_S energy, while preserving the original ratio between the π^0 and photon energies:

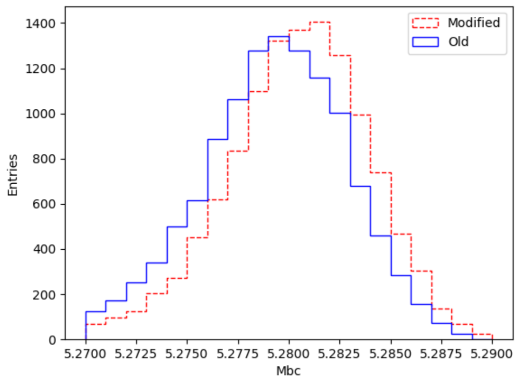
$$\begin{aligned}
E_{\rm res} &= \sqrt{s}/2 - E^*_{K^0_S},\\
E'_{\pi^0} &= \frac{E_{\rm res}}{E^*_{\pi^0}+E^*_\gamma}\,E^*_{\pi^0}, \qquad E'_\gamma = \frac{E_{\rm res}}{E^*_{\pi^0}+E^*_\gamma}\,E^*_\gamma,\\
\vec{p}\,'_{\pi^0} &= \sqrt{(E'_{\pi^0}/c)^2-(M_{\pi^0}c)^2}\,\frac{\vec{p}^{\,*}_{\pi^0}}{|\vec{p}^{\,*}_{\pi^0}|}, \qquad \vec{p}\,'_\gamma = (E'_\gamma/c)\,\frac{\vec{p}^{\,*}_\gamma}{|\vec{p}^{\,*}_\gamma|},\\
M_{\rm bc} &= (1/c^2)\sqrt{s/4-|\vec{p}_{K^0_S}+\vec{p}\,'_{\pi^0}+\vec{p}\,'_\gamma|^2c^2}.
\end{aligned} \tag{5.2}$$

Since the K^0_S momentum is measured more precisely than the π^0 or photon energies, this rescaling propagates the well-measured K^0_S and beam kinematics into $M_{\rm bc}$ while suppressing its sensitivity to shower-leakage-induced biases in the π^0 and photon energies. It improves the signal $M_{\rm bc}$ resolution by 3–7%, depending on the dataset, and significantly reduces the residual $M_{\rm bc}$–ΔE correlation in background events without affecting the signal [61]. Signal candidates are required to satisfy $M_{\rm bc} \in [5.23, 5.29]$ GeV/c^2 and $\Delta E \in [-0.3, 0.2]$ GeV, with the asymmetric ΔE range again reflecting the shower-leakage tail and the upper $M_{\rm bc}$ edge corresponding closely to the kinematic endpoint $\sqrt{s}/2c^2$.

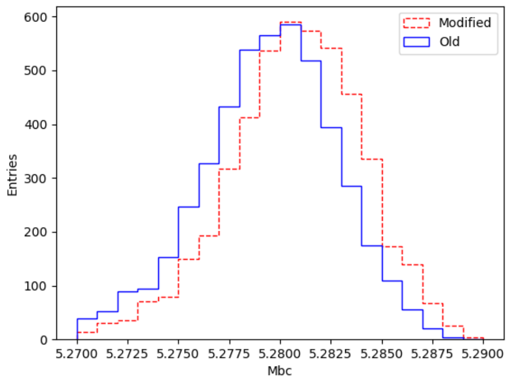
Figure 5.4 compares the resulting signal $M_{\rm bc}$ and ΔE shapes, before and after this modification, in both Belle and Belle II MC events: the modified variables are visibly narrower in both experiments, consistent with the improved resolution. Table 5.1 quantifies this improvement for all four dataset/mass-region combinations, via the central interval containing 68% of signal candidates (equivalent to $\pm 1\sigma$ for a Gaussian distribution); the improvement is consistent across mass regions and experiments, at the few-percent level quoted above.

Table 5.1: Central 68% intervals of the signal $M_{\rm bc}$ and ΔE distributions, for the original and modified variable definitions, in all four dataset/mass-region combinations.

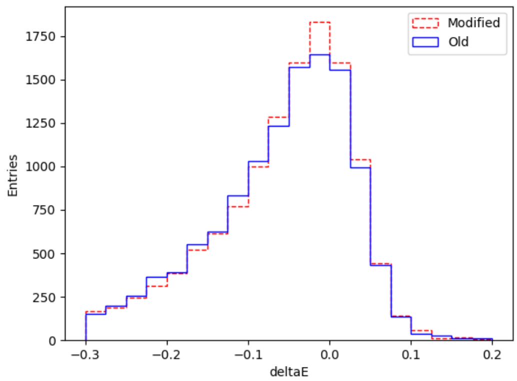
Dataset	Variable	Original interval	Modified interval	Improvement
Belle MR1	$M_{\rm bc}$ (GeV/c^2)	7.52×10^{-3}	7.02×10^{-3}	6.6%
	ΔE (GeV)	1.74×10^{-1}	1.63×10^{-1}	6.3%
Belle MR2	$M_{\rm bc}$ (GeV/c^2)	7.70×10^{-3}	7.26×10^{-3}	5.7%
	ΔE (GeV)	1.80×10^{-1}	1.70×10^{-1}	5.5%
Belle II MR1	$M_{\rm bc}$ (GeV/c^2)	6.90×10^{-3}	6.68×10^{-3}	3.2%
	ΔE (GeV)	1.49×10^{-1}	1.41×10^{-1}	5.4%
Belle II MR2	$M_{\rm bc}$ (GeV/c^2)	6.72×10^{-3}	6.40×10^{-3}	4.8%
	ΔE (GeV)	1.41×10^{-1}	1.32×10^{-1}	6.4%



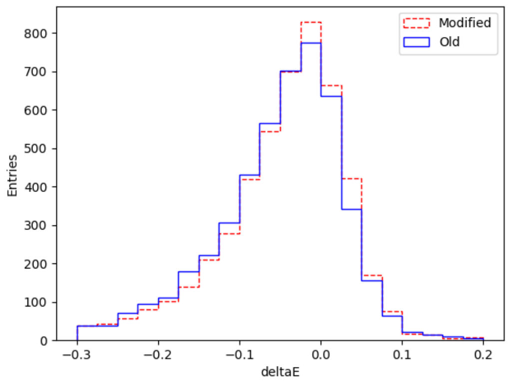
(a) M_{bc}, Belle



(b) M_{bc}, Belle II



(c) ΔE, Belle



(d) ΔE, Belle II

Figure 5.4: Distributions of the original (blue) and modified (red) M_{bc} and ΔE variables for correctly reconstructed signal MC (MR1), in Belle (a, c) and Belle II (b, d); the modified variables are visibly narrower in both experiments.

5.3 Mass Regions MR1 and MR2

The invariant mass of the reconstructed $K^0_S\pi^0$ system, $M_{K^0_S\pi^0}$, is required to lie in the range $[0.8, 1.8]\,\mathrm{GeV}/c^2$, selecting the kinematic region relevant to this analysis while rejecting, at the upper edge, reflections from charmed-meson decays. This range is further divided into two nonoverlapping mass regions, in which the analysis proceeds and the CP-violation parameters are extracted separately:

- **MR1**, $M_{K^0_S\pi^0} \in [0.8, 1.0]\,\mathrm{GeV}/c^2$, dominated by the $K^{*0}(892)$ resonance.
- **MR2**, $M_{K^0_S\pi^0} \in (1.0, 1.8]\,\mathrm{GeV}/c^2$, with contributions from nonresonant $B^0 \to K^0_S\pi^0\gamma$ decay and higher K^* resonances.

As discussed in Section 1.4, the theoretical relationship between S and the underlying $b \to s\gamma$ amplitude differs between these two regions, motivating their separate treatment throughout the reconstruction, background-suppression, and fit strategy.

5.4 B-meson Vertexing and Δt Reconstruction

The B_{sig} decay vertex is obtained from the global decay-chain fit of Section 3.4, applied to its full decay chain with a point constraint to the IP, using the displaced $K^0_S \to \pi^+\pi^-$ vertex as the primary source of vertex information. The B_{tag} vertex is instead reconstructed from the remaining tracks in the event not assigned to B_{sig}, using the beam-constrained, IP-tube technique of Section 3.4, with its flight direction inferred from the (well-measured) B_{sig} momentum. Standard vertex-quality requirements, a positive fit χ^2 probability for B_{sig} and a positive tag-side vertex-fit p-value with at least one effective degree of freedom, are applied to both vertices.

The proper-time difference between the two B decays is calculated as

$$\Delta t = \frac{\Delta z}{\beta\gamma c}, \tag{5.3}$$

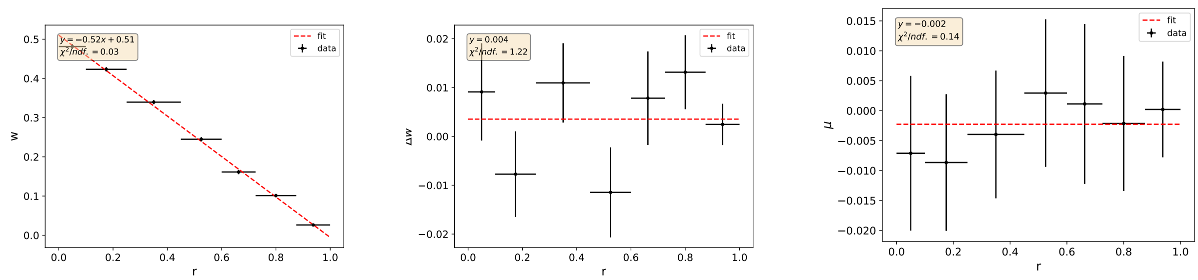
where Δz is the separation between the two reconstructed vertices projected along the boost direction, and $\beta\gamma$ is the Lorentz boost of the $\Upsilon(4S)$ introduced in Section 1.3, equal to 0.425 for Belle and 0.284 for Belle II.

Not every event yields a usable Δt measurement: only K^0_S decays with at least one pion track registering a hit in the vertex detector (the SVD for Belle, either the PXD or SVD for Belle II) provide a sufficiently precise B_{sig} vertex, and a further requirement of $\sigma_{\Delta t} < 3.0\,\mathrm{ps}$ is imposed on the per-event Δt uncertainty. Events satisfying both criteria form the *time-dependent* (TD) sample, used to extract both S and C. The remainder form the *time-integrated* (TI) sample, which lacks a usable Δt but still contributes information on C through its tag-flavour-dependent yields (Chapter 6). The TD sample comprises approximately 50% of selected events in Belle and 75% in Belle II; the larger Belle II fraction follows from its larger-volume vertex detector (PXD and SVD together, versus SVD alone in Belle) and its smaller Lorentz boost, both of which increase the probability that the K_S^0 decays within the instrumented tracking volume.

5.5 Flavour Tagging Application

The flavour of B_{tag} at its decay time, needed to evaluate the CP asymmetry of Eq. 1.7, is determined using the Graph-Neural-Network-based Flavour Tagger (GFlaT) introduced in Section 4.5. Applied to this analysis, the GFlaT returns a flavour hypothesis $q_{\text{tag}} = \pm 1$ and a dilution factor $r \in [0, 1]$ quantifying the tagger's confidence: $r = 1$ corresponds to complete certainty in the flavour assignment, while $r = 0$ corresponds to no tagging information at all. The mistag parameters w, Δw, and μ that translate r into a physical dilution of the measured asymmetry are calibrated, in seven bins of r, using the self-tagged control channels $B^0 \to J/\psi\, K^{*0}$ (Belle) and $B^0 \to D^{(*)-}\pi^+$ (Belle II), as described in Section 4.5. The calibrated mistag parameters are interpolated linearly in r for use in the fit of Chapter 6.

Figure 5.5 shows this r-dependent modelling explicitly: the average mistag probability w decreases linearly with r, as expected since higher r corresponds to a more confident, less-often-wrong tag, while the wrong-tag asymmetry Δw and tagging-efficiency asymmetry μ show no significant r-dependence and are each modelled instead as a single constant across all seven bins.



(a) w, Belle II (b) Δw, Belle II (c) μ, Belle II

Figure 5.5: The seven centrally calibrated r-bin values (black points) of w, Δw, and μ, together with the continuous-in-r model fitted to them (red line) and used in place of the legacy seven-category binning, shown here for Belle II as an example; the same procedure is applied to Belle. Values shown use the "PXD-on" calibration, which applies to the majority (~ 397 fb^{-1} of 493 fb^{-1}) of the Belle II dataset analysed; the "PXD-off" calibration is statistically consistent with it.

5.6 Background Suppression

After the preliminary selection described in Sections 5.1–5.3, the sample remains dominated by combinatorial and physics backgrounds from $B\overline{B}$ and continuum $e^+e^- \to q\bar{q}$ events. Beyond the K_S^0-selection, Λ^0-veto, π^0-selection, and π^0/η-veto classifiers already introduced, a dedicated continuum-suppression MVA is trained to reject the continuum background directly. The training configuration and leading input variables of the π^0/η-veto, π^0-selection, and continuum-suppression classifiers are given in Appendix A.

5.6.1 Continuum-Suppression MVA

Continuum $e^+e^- \to q\bar{q}$ ($q \in \{u, d, s, c\}$) events differ topologically from genuine $\Upsilon(4S) \to B\overline{B}$ events, as illustrated in Figure 5.6: because the two quarks are produced with much higher momentum than the nearly-at-rest B mesons from $\Upsilon(4S)$ decay, continuum events develop a collimated, back-to-back "jet-like" topology, whereas $B\overline{B}$ events are comparatively spherical.

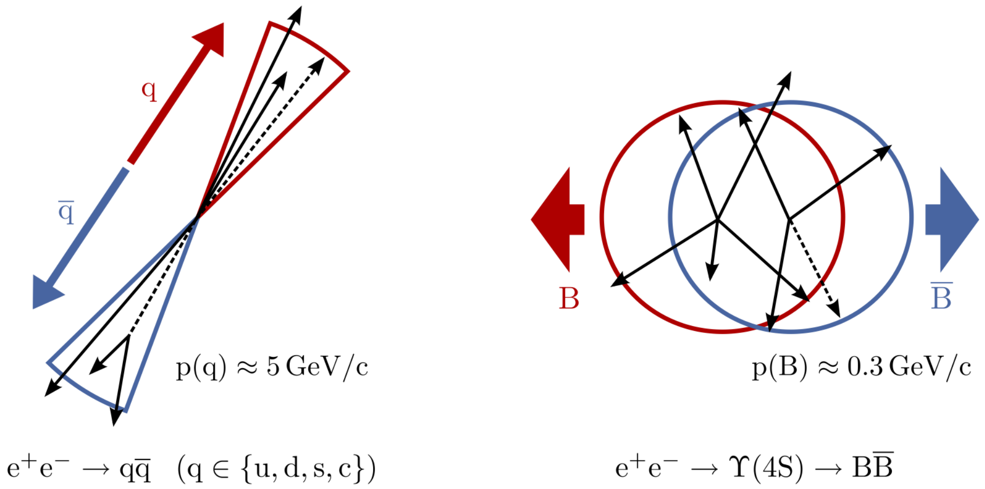
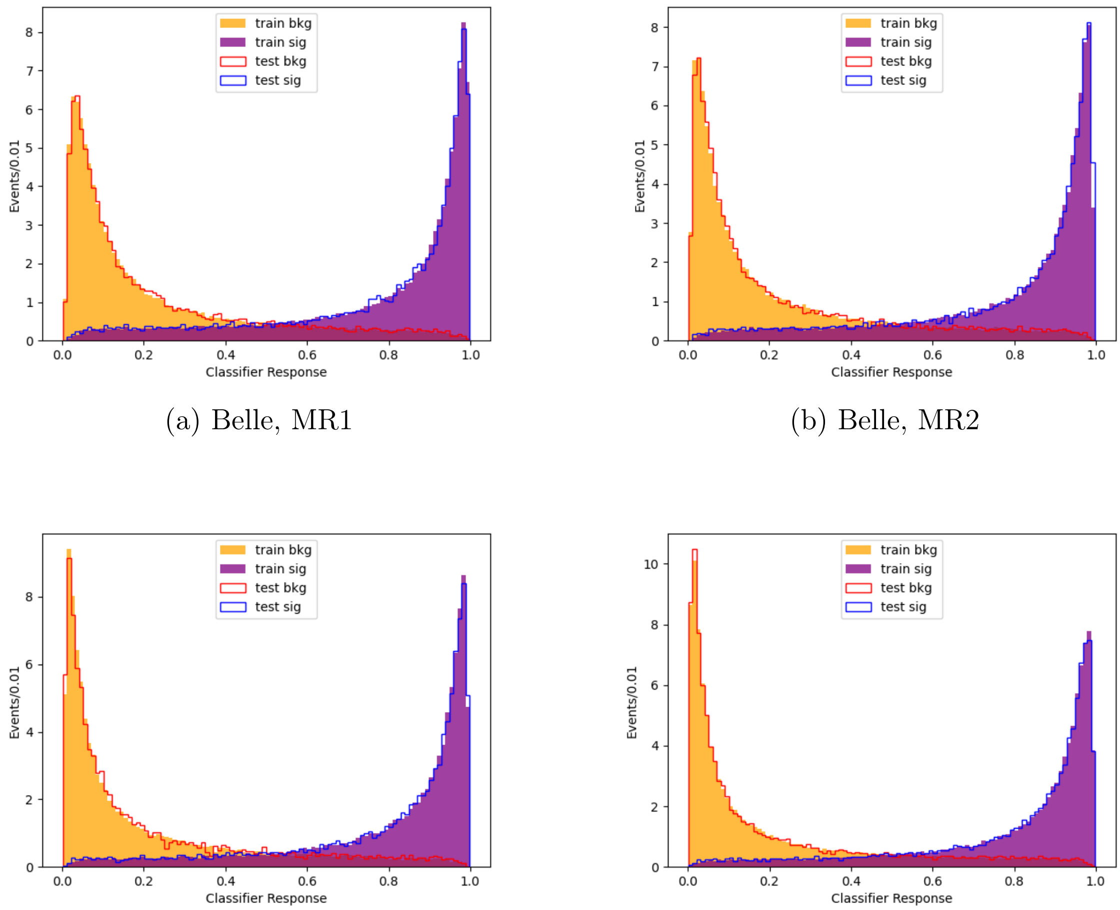


Figure 5.6: Schematic contrast between the collimated, back-to-back topology of continuum $e^+e^- \to q\bar{q}$ events (left) and the comparatively spherical, isotropic topology of $\Upsilon(4S) \to B\overline{B}$ events (right), reflecting the much larger momentum of the continuum quarks relative to the nearly-at-rest B mesons.

This is exploited by an XGBoost classifier trained on event-shape variables, including the thrust [62] (the maximum fraction of total momentum aligned along any single axis, large for jet-like events), Fox–Wolfram moments [54] (a set of rotationally invariant moments of the angular distribution of particle momenta, commonly used to quantify event shape), CLEO cone variables [63] (energies deposited in a series of conical regions around the signal thrust axis), and KSFW moments (a related set of modified Fox–Wolfram moments developed for the B-factory environment). Of these, the most discriminating single feature is the cosine of the angle between the thrust axis of B_{sig} and that of the rest of the event. Separate classifiers are trained for each mass region and each experiment, since the optimal combination of input variables and their relative importance can differ between them. These classifiers achieve an ROC-AUC of 0.92 (Belle) and 0.93 (Belle II) in both mass regions; a working-point cut at a score of 0.6 removes approximately 80% (Belle) to 85% (Belle II) of continuum background for a signal-efficiency loss of approximately 20%, with the classifier response shown in Figure 5.7.



(a) Belle, MR1

(b) Belle, MR2

(c) Belle II, MR1

(d) Belle II, MR2

Figure 5.7: Response of the continuum-suppression MVA for signal and continuum background, in each of the four experiment/mass-region combinations, shown separately for the training and test samples.

The residual distribution of the transformed variable after application of the criterion can be further used for discriminating between signal and background, and this variable is included in the fit. Because the raw classifier output is sharply peaked near unity for signal, making it difficult to model with analytic functions in the fit of Chapter 6, it is transformed into a new variable

$$\mathcal{C}' = \ln\left(\frac{C_{\text{out}} - C_{\text{out}}^{\text{min}}}{C_{\text{out}}^{\text{max}} - C_{\text{out}}}\right), \tag{5.4}$$

where $C_{\rm out}$ is the raw MVA output after the working-point cut, and $C_{\rm out}^{\rm min}$ and $C_{\rm out}^{\rm max}$ are its minimum and maximum values; this log transform spreads out the region near $C_{\rm out} = 1$ into a variable, $\mathcal{C}'$, with a shape that can be described by typical analytic probability density functions.

5.6.2 Data-Simulation Validation and Bias Checks

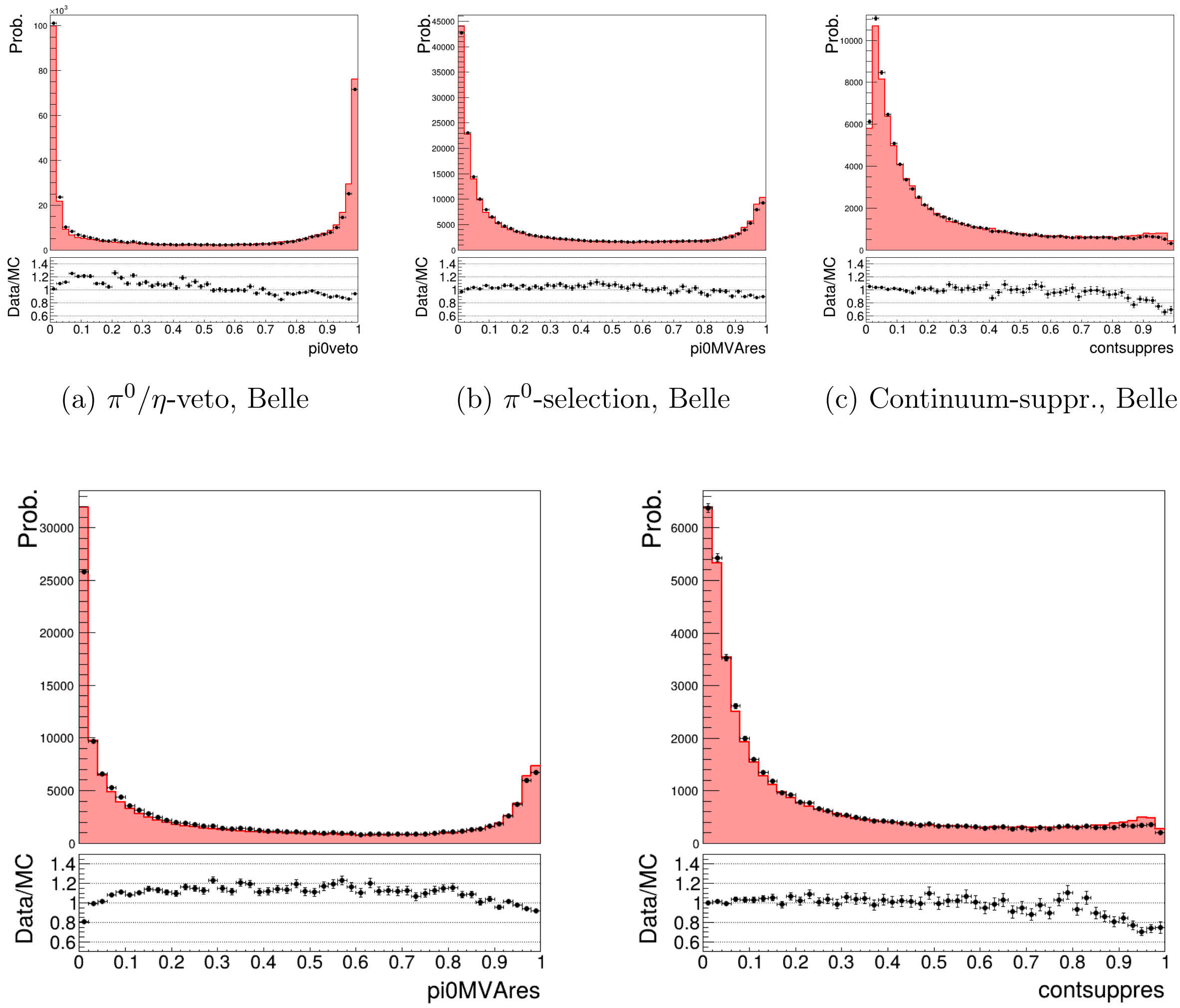
Since every classifier introduced in this chapter is trained on simulation, its output must also be validated directly against data before it can be trusted to select an unbiased data sample. This is done in a background-dominated dataset, by comparing the shape of each classifier's output distribution, and of its input variables individually, between data and MC simulation, with the MC histogram normalized to the data yield to allow a shape-only comparison. Figure 5.8 shows this comparison for the Belle π^0/η-veto, π^0-selection, and continuum-suppression MVA outputs, and for the Belle II π^0-selection and continuum-suppression MVA outputs. In every case, the data/MC ratio is consistent with unity to within approximately 20%, with no evidence of a shape distortion that would bias the selected sample.

A further, more targeted check addresses whether any classifier could introduce a bias correlated with the flavour tag or with Δt itself, which would be far more dangerous than a generic shape mismodelling since it could generate a spurious *CP* asymmetry. Two such checks were performed for the classifier MVAs, using signal MC with known generated Δt: first, the fraction of events retained by the working-point criterion on this classifier was compared before and after tagging the accompanying $B_{\rm tag}$ meson as a B^0 or $\overline{B}^0$, and is flat as a function of generated Δt for both tag flavours to within statistical uncertainties (Figure 5.9); second, no significant change in the shape of the reconstructed Δt distribution itself was observed after applying the cut. Together these confirm that these MVAs do not introduce any tag- or Δt-dependent selection bias capable of mimicking *CP* violation.

5.7 Best-Candidate Selection and Selection Efficiency

After all selection criteria above, 10–12% of events retain more than one $B_{\rm sig}$ candidate, a multiplicity of approximately 1.20 (Belle) and 1.15 (Belle II) among candidates that pass the full selection. Since these ambiguities arise predominantly from misreconstructed π^0 candidates (i.e. alternative photon-pairing combinations satisfying the π^0 mass window), the candidate with the highest π^0-selection MVA score (Section 5.1) is retained in each event; in the rare event of a tie, a candidate is chosen arbitrarily. This best-candidate selection (BCS) procedure is found, from simulation, to select a correctly reconstructed candidate approximately 73% of the time in events with at least one such candidate present, outperforming alternative choices such as selecting the candidate with π^0 mass closest to its nominal value (61%) or choosing a candidate at random (49%).

Table 5.2 summarises the cut-flow through preselection, each background-suppression classifier, and best-candidate selection, for data-equivalent simulated samples in both experiments, together with the resulting cumulative efficiency in MR1 and MR2 (evaluated using the $K^{*0}(892)$-resonant and higher-mass-region signal MC samples, respectively). The overall reconstruction and selection efficiency for signal events, combining correctly reconstructed and self-cross-feed signal, is approximately 19% (MR1) and 13% (MR2) for Belle, and 22% (MR1) and 14% (MR2) for Belle II. The improvement in Belle II reflects its superior detector and algorithm performance.



(a) π^0/η-veto, Belle (b) π^0-selection, Belle (c) Continuum-suppr., Belle

(d) π^0-selection, Belle II (e) Continuum-suppr., Belle II

Figure 5.8: Data-simulation comparison of classifier output shapes in a background-dominated region, with the MC histogram normalized to the data yield; the lower panel of each plot shows the data-MC ratio.

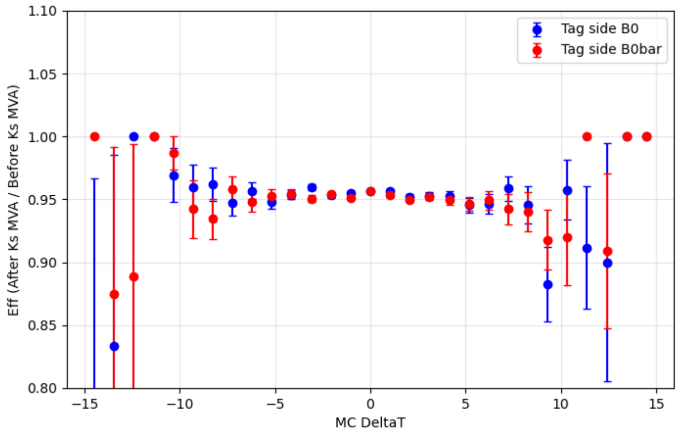


Figure 5.9: Example plot showing the efficiency of the K_S^0-selection MVA working-point criterion as a function of the true (generated) Δt in signal MC, shown separately for events with a $B^0_{\rm tag}$ and a $\overline{B}^0_{\rm tag}$; the efficiency is flat within uncertainties across the full Δt range and shows no significant difference between the two tag flavours.

Table 5.2: Cut-flow of candidate counts (continuum, $B\overline{B}$, correctly reconstructed signal, and self-cross-feed signal, defined in Section 6.2.1) through each selection stage, in data-equivalent simulated samples, together with the cumulative efficiency in MR1 and MR2. The quoted efficiency is for the true-signal-plus-SCF sample, i.e. all signal MC events surviving selection regardless of whether they are correctly reconstructed or SCF.

Selection	$q\bar{q}$	$B\overline{B}$	True signal	SCF signal	MR1 / MR2 eff. (%)
Belle					
Prelim. selection	327329	13361	1874	3276	36.6 / 22.5
π^0/η-veto MVA	155181	10512	1764	3049	34.5 / 21.0
π^0-selection MVA	61233	3156	1647	770	28.8 / 16.6
Continuum MVA	6764	1825	1290	485	22.1 / 13.1
Best-candidate sel.	5970	1503	1226	226	22.1 / 13.1
Belle II					
Prelim. selection	2648828	30683	1973	3379	43.9 / 25.3
K_S^0-selection MVA	447626	8756	1854	3175	41.7 / 23.9
π^0/η-veto MVA	124250	6296	1731	2926	39.2 / 22.1
π^0-selection MVA	38770	1601	1544	447	31.9 / 16.9
Continuum MVA	4248	1112	1249	312	25.3 / 13.7
Best-candidate sel.	3954	990	1206	157	25.3 / 13.7

Chapter 6

Time-Dependent *CP*-Violation Fit

Having reconstructed and selected $B^0 \to K_S^0 \pi^0 \gamma$ candidates (Chapter 5), this chapter describes how the *CP*-violation parameters S and C are extracted from them. Following standard practice in blinded analyses, the fit strategy, probability density functions (PDFs), and validation studies described here were all developed using simulation and data control samples, with the signal region (Section 5.1) in data examined only once all procedures were finalized (Section 6.6).

6.1 Overview of the Fit Strategy

The S and C parameters are extracted with a simultaneous extended unbinned maximum-likelihood fit, performed independently for each dataset (Belle and Belle II) and mass region (MR1 and MR2). Guided by MC studies showing that $M_{\rm bc}$, ΔE, and $\mathcal{C}'$ (Sections 5.2 and 5.6) have negligible correlation with Δt, the full PDF for each fit component i factorizes into a *signal-extraction* (SE) PDF, $P_i^{\rm SE}(M_{\rm bc}, \Delta E, \mathcal{C}')$, discriminating between components, and a *CP*-asymmetry PDF, $P_i^{\rm CPV}(\Delta t \mid \sigma_{\Delta t}, q_{\rm tag}, r)$, encoding the time-dependent asymmetry itself. There are three fit components ($i = 1, 2, 3$): signal, $B\overline{B}$ and continuum background. These two halves of the fit are conceptually distinct: the SE PDFs discriminate signal from background using kinematic and background-suppression variables uncorrelated with decay time, while the CPV PDFs encode the time-dependent physics of interest. Section 6.2 describes both, in turn, for each of the three components, before Section 6.3 combines them into the total fit actually performed.

The fit is further split according to the time-dependent (TD) and time-integrated (TI) classification of Section 5.4. For the TD sample, the full PDF is

$$P_{\rm full}^{\rm TD}(M_{\rm bc}, \Delta E, \mathcal{C}', \Delta t \mid \sigma_{\Delta t}, q_{\rm tag}, r) = \sum_{i=1}^{3} N_i^q(q_{\rm tag}, r)\, f_i^{\rm TD}\, P_i^{\rm SE}(M_{\rm bc}, \Delta E, \mathcal{C}') \times P_i^{\rm CPV}(\Delta t \mid \sigma_{\Delta t}, q_{\rm tag}, r), \tag{6.1}$$

while for the TI sample, lacking a usable Δt, it is

$$P_{\rm full}^{\rm TI}(M_{\rm bc}, \Delta E, \mathcal{C}' \mid q_{\rm tag}, r) = \sum_{i=1}^{3} N_i^q(q_{\rm tag}, r)\,(1 - f_i^{\rm TD})\,P_i^{\rm SE}(M_{\rm bc}, \Delta E, \mathcal{C}'), \tag{6.2}$$

where $N_i^q(q_{\rm tag}, r)$ is the yield of component i for a given tag flavour $q_{\rm tag} = \pm 1$ and dilution r (expressed in terms of C for the signal and $B\overline{B}$ components, Section 6.2.1), and $f_i^{\rm TD}$ is the fraction of component i found in the TD sample, fixed from simulation and corrected using data control samples (Section 6.6).

6.2 Modelling of the Fit Components

The SE and CPV models introduced in Section 6.1 are developed below for each of the three fit components in turn, starting with signal.

6.2.1 Signal

The signal component itself is the sum of two sub-components, weighted by their MC-predicted fractions: correctly reconstructed signal, in which all three of K_S^0, π^0, and the prompt photon are properly matched to the true decay; and *self-cross-feed* (SCF) signal, in which the K_S^0 is correctly reconstructed but the π^0 or the prompt photon is not. The SCF fraction is itself varied as part of the systematic-uncertainty evaluation (Section 7.2).

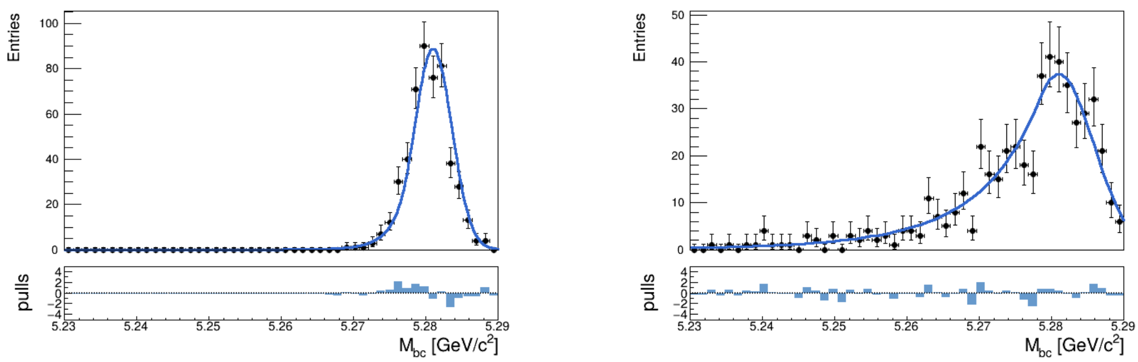
Signal-Extraction Model. For correctly reconstructed signal, the $M_{\rm bc}$ shape is modelled with a Crystal Ball function [64, 65], a Gaussian core smoothly joined to a power-law tail on one side, originally developed to describe energy-loss tails in calorimetric measurements and well suited here to the residual shower-leakage tail in $M_{\rm bc}$. The ΔE and $\mathcal{C}'$ shapes are each instead modelled with a Johnson S_U function [66], a four-parameter distribution capable of describing a wide range of skewed, heavy-tailed shapes beyond what a simple Gaussian or Crystal Ball function can capture. A residual correlation of 5–20% between $M_{\rm bc}$ and ΔE, depending on the dataset, remains despite the modified $M_{\rm bc}$ definition of Section 5.2. Figure 6.1 illustrates the effect directly, comparing independent Crystal Ball fits to the signal $M_{\rm bc}$ shape in a bin centred on $\Delta E \approx 0$ GeV with one far in the shower-leakage tail ($\Delta E \approx -0.27$ GeV): the $M_{\rm bc}$ peak in the latter is visibly broader and less sharply peaked, motivating an explicit treatment of this residual correlation rather than a single, ΔE-independent $M_{\rm bc}$ shape.

This is absorbed by replacing the Crystal Ball width σ and tail-transition parameter α with explicit functions of ΔE,

$$\sigma(\Delta E) = \sigma_0 + \sigma_1\,(\Delta E)^2, \tag{6.3}$$

$$\alpha(\Delta E) = [\alpha_0 + \alpha_1 \exp(-|\Delta E|/\alpha_2)]\, \sigma(\Delta E), \tag{6.4}$$

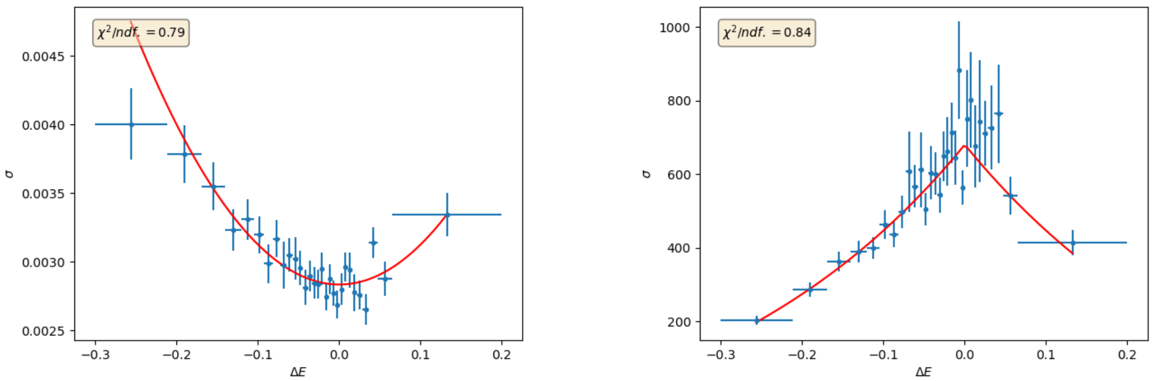
where σ_0 is the width at $\Delta E = 0$ GeV and σ_1 quantifies its quadratic growth away from it, reflecting the progressively degraded M_{bc} resolution as shower leakage worsens towards more negative ΔE; α, rather than being parameterised directly, is instead expressed as a ΔE-dependent multiple of $\sigma(\Delta E)$ itself, with α_0, α_1, and α_2 controlling its constant, exponentially decaying, and decay-length components, respectively. Overall, this conditional M_{bc} model has seven free parameters. This parameterisation is derived and validated by independently fitting the M_{bc} shape in many narrow bins of ΔE and fitting Eqs. 6.3 and 6.4 to the resulting per-bin σ and α/σ values, as shown for one such bin pair in Fig. 6.1; Fig. 6.2 shows the resulting $\sigma(\Delta E)$ and $\alpha(\Delta E)/\sigma(\Delta E)$ trends themselves, for Belle II MR1 as an example, together with the fits of Eqs. 6.3 and 6.4 overlaid; the same procedure is applied in all four dataset/mass-region combinations. The SCF component, being a smaller and less sharply peaked contribution, is instead modelled with a Crystal Ball function in M_{bc} and asymmetric Gaussian functions in ΔE and $\mathcal{C}'$. The means of the correctly reconstructed signal M_{bc} and ΔE PDFs are floated in the fit to data; all other signal shape parameters are fixed from simulation, subject to the control-channel corrections of Section 6.4.2.



(a) $\Delta E \approx 0$ GeV (b) $\Delta E \approx -0.27$ GeV

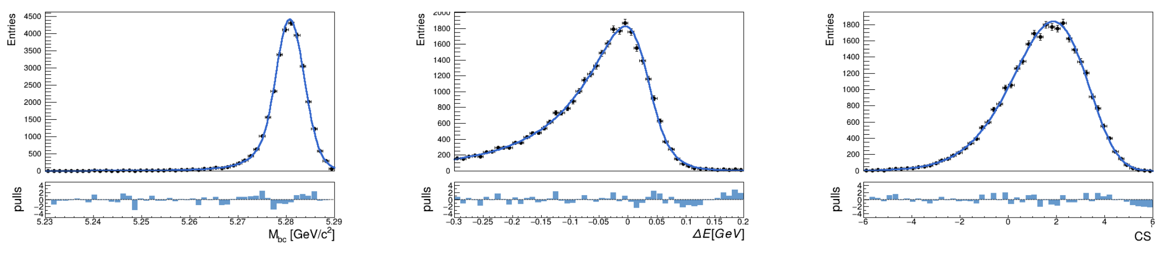
Figure 6.1: Independent Crystal Ball fits to the signal M_{bc} shape in two bins of ΔE, in Belle MR1 MC: (a) a bin centred on $\Delta E \approx 0\,\mathrm{GeV}$, where the π^0 and photon energies are well reconstructed; (b) a bin far in the shower-leakage tail, where the M_{bc} peak is visibly broader and shifted, motivating the parameterisation of the M_{bc} width and tail as functions of ΔE described in the text.

Figure 6.3 shows this signal model fitted individually to the correctly reconstructed signal MC events, in each of the three SE dimensions, for Belle MR1 as an example, illustrating the component's shape in isolation before it is combined with the $B\overline{B}$ and continuum components below; the same procedure is applied in all four dataset/mass-region combinations.



(a) $\sigma(\Delta E)$ (b) $\alpha(\Delta E)/\sigma(\Delta E)$

Figure 6.2: The per-bin σ and α/σ values extracted from independent Crystal Ball fits to $M_{\rm bc}$ in several bins of ΔE, shown for Belle II MR1 as an example, together with the fits of Eqs. 6.3 and 6.4 overlaid (red curve); the quadratic rise in σ towards negative ΔE reflects the worsening $M_{\rm bc}$ resolution in the shower-leakage tail, consistent with Fig. 6.1. The same parametrisation procedure is applied to Belle and to MR2.



(a) $M_{\rm bc}$ (b) ΔE (c) $\mathcal{C}'$

Figure 6.3: The signal-extraction model of this section, fitted individually to correctly reconstructed signal MC events, shown for Belle MR1 as an example, with pull distributions shown below each panel; this isolates the shape of the signal component ahead of the combined fit of Section 6.2.4. The same fit is performed for Belle II and for MR2.

***CP*-Violation Model.** This split into correctly reconstructed and SCF sub-components applies only to the SE PDF: the two signal sub-components necessarily share a single *CP*-violation PDF. This is physically justified because the SCF misreconstruction affects only the π^0 or prompt-photon assignment, not the K_S^0-anchored $B_{\rm sig}$ vertex or the $B_{\rm tag}$ side, and so does not bias the reconstructed Δt or the flavour tag on which that PDF depends. This shared signal *CP*-asymmetry PDF incorporates the physics-motivated time-dependence of Eq. 1.7, modified to account for imperfect flavour tagging via the mistag parameters w, Δw, and μ (Section 5.5), and convolved with a resolution function $\mathcal{R}$ describing the finite precision of the Δt measurement:

$$P_{\rm sig}(\Delta t, q_{\rm tag} \mid \sigma_{\Delta t}, r) = \frac{\exp(-|\Delta t|/\tau_{B^0})}{4\tau_{B^0}} \Big\{1 - q_{\rm tag}\Delta w + q_{\rm tag}\mu(1-2w) + \big[q_{\rm tag}(1-2w) + \mu(1 - q_{\rm tag}\Delta w)\big]\left[S\sin(\Delta m_d \Delta t) - C\cos(\Delta m_d\Delta t)\right]\Big\} \otimes \mathcal{R}(\Delta t \mid \sigma_{\Delta t}). \tag{6.5}$$

The resolution function itself is a sum of a main and a tail Gaussian, the latter partially convolved with an exponential to describe components arising, for example, from secondary vertices of charmed-particle decays within the B decay chain, with the widths of both Gaussians scaled by the per-event uncertainty $\sigma_{\Delta t}$:

$$\begin{aligned}
\mathcal{R}(\Delta t_{\rm res} \mid \sigma_{\Delta t}) &= (1 - f_{\rm tail})\,\mathcal{R}^{\rm main} + f_{\rm tail}\Big[(1-f_{\rm exp})\,\mathcal{R}^{\text{tail-G}} + f_{\rm exp}\,\mathcal{R}^{\text{tail-GE}}\Big],\\
\mathcal{R}^{\rm main} &= G(\Delta t_{\rm res},\, \mu_{\rm main}\sigma_{\Delta t},\, \sigma_{\rm main}\sigma_{\Delta t}),\\
\mathcal{R}^{\text{tail-G}} &= G(\Delta t_{\rm res},\, \mu_{\rm tail}\sigma_{\Delta t},\, \sigma_{\rm tail}\sigma_{\Delta t}),\\
\mathcal{R}^{\text{tail-GE}} &= \mathcal{R}^{\text{tail-G}} \otimes \left[f_{\rm R}\exp\left(\frac{-\Delta t_{\rm res}}{k\sigma_{\Delta t}}\right) + (1-f_{\rm R})\exp\left(\frac{\Delta t_{\rm res}}{k\sigma_{\Delta t}}\right)\right],
\end{aligned} \tag{6.6}$$

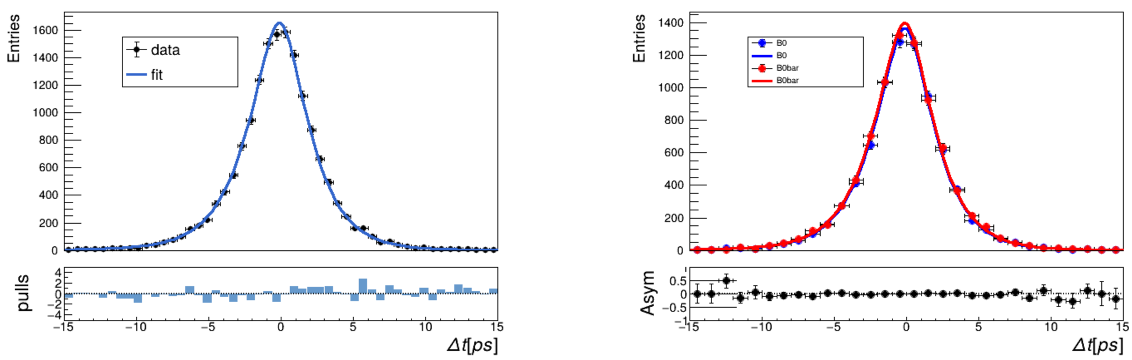
where $\Delta t_{\rm res} \equiv \Delta t - \Delta t_{\rm true}$, G denotes a Gaussian, and the eight parameters $\mu_{\rm main}$, $\sigma_{\rm main}$, $\mu_{\rm tail}$, $\sigma_{\rm tail}$, $f_{\rm tail}$, $f_{\rm exp}$, $f_{\rm R}$, and k are calibrated from the $B^0 \to J/\psi K_S^0$ control channel (Section 6.4.1). The tag-flavour-dependent signal yields directly encode C:

$$N_{\rm sig}^{q}(q_{\rm tag}, r) = \frac{N_{\rm sig}}{2}\left(1 - q_{\rm tag}\Delta w + q_{\rm tag}\mu(1-2w) - C\,\frac{q_{\rm tag}(1-2w) + \mu(1-q_{\rm tag}\Delta w)}{1 + \Delta m_d^2 \tau_{B^0}^2}\right). \tag{6.7}$$

Since the B-meson lifetime measurement is especially sensitive to any resolution-function mis-modelling, a flavour-integrated version of this signal Δt PDF, floating only τ_{B^0} itself with $S = C = 0$, is fitted to signal MC events as a preliminary cross-check, ahead of the full S/C closure test of Section 6.3. Figure 6.4(a) shows the resulting fit, for Belle MR2 as an example, with the fitted lifetime found consistent with the generated value in all four dataset/mass-region combinations. With the lifetime fixed to its generated value, S and C are then floated and fit directly to signal MC events, at 50× the expected signal yield to give the closure test adequate statistical precision; the resulting values, listed in Table 6.1, are consistent with the generated value of zero in every dataset/mass-region combination. Figure 6.4(b) shows the corresponding Δt fit split by B_{tag} flavour, again for Belle MR2, with the raw asymmetry plotted below each panel as the direct visual counterpart of this result.

Table 6.1: S and C from a dedicated fit to signal MC events, at 50× the expected signal yield, against a generated value of zero in both cases.

Dataset	Mass region	S	C
Belle	MR1	0.04 ± 0.03	-0.01 ± 0.02
	MR2	0.01 ± 0.03	0.01 ± 0.02
Belle II	MR1	-0.01 ± 0.02	0.01 ± 0.02
	MR2	-0.02 ± 0.02	0.01 ± 0.01



(a) Lifetime cross-check

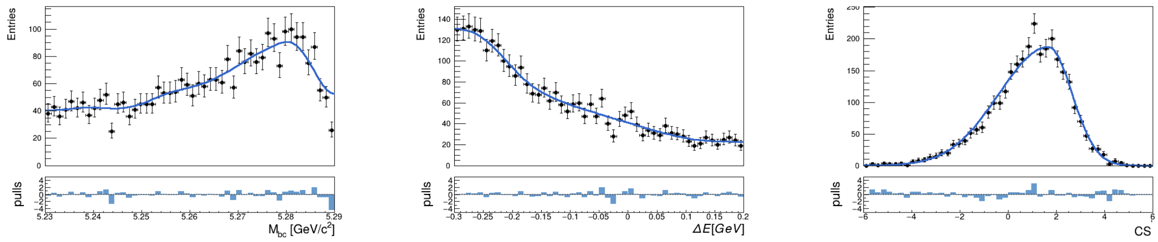
(b) S/C closure test

Figure 6.4: Signal-MC-only Δt fits, shown for Belle MR2 as an example: (a) the flavour-integrated fit used to cross-check the signal lifetime τ_{B^0} ahead of the full S/C closure test; (b) the high-statistics fit, split by B_{tag} flavour, underlying Table 6.1, with the raw asymmetry shown below the panel; its flatness confirms the fitted S and C values are consistent with the generated, *CP*-conserving value of zero. The same fits are performed for Belle II and for MR1.

6.2.2 $B\overline{B}$ Background

Signal-Extraction Model. The $B\overline{B}$ background exhibits a peaking structure in M_{bc}, from partially reconstructed or misreconstructed B decays that nonetheless satisfy the beam-energy constraint, but no corresponding peak in ΔE. To capture the resulting residual correlation without assuming a specific analytic form, its joint M_{bc}–ΔE shape is modelled with a two-dimensional kernel density estimate (KDE) [67], a nonparametric technique that builds a smooth PDF directly from the simulated sample by summing a kernel function centred on each simulated event, rather than fitting the data to a predetermined functional form. Its $\mathcal{C}'$ shape is modelled with an asymmetric Gaussian function.

Figure 6.5 shows this model fitted individually to $B\overline{B}$ background MC events, for Belle II MR2 as an example.

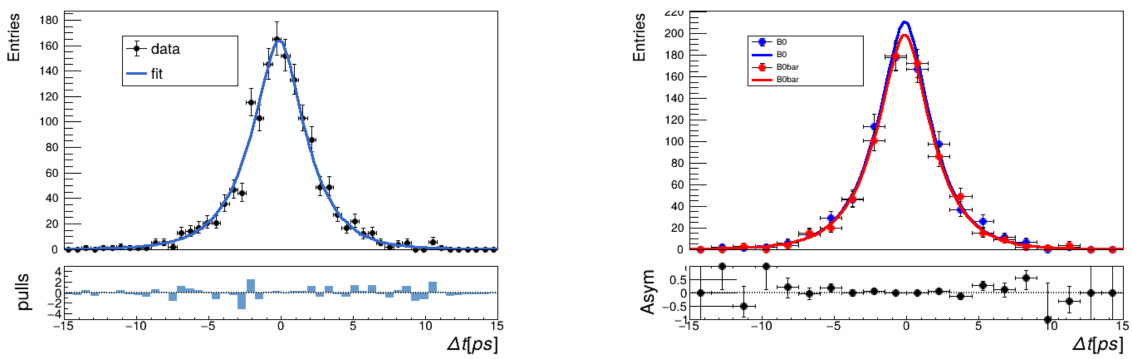


(a) $M_{\rm bc}$ (b) ΔE (c) $\mathcal{C}'$

Figure 6.5: The signal-extraction model of this section, fitted individually to $B\overline{B}$ background MC events, shown for Belle II MR2 as an example, with pull distributions shown below each panel. The same fit is performed for Belle and for MR1.

***CP*-Violation Model.** The $B\overline{B}$ background is modelled with the same functional form as the signal (Section 6.2.1), but with an effective lifetime $\tau_{B\overline{B}}$ left floating in the fit, and with its own *CP*-violation parameters $S_{B\overline{B}}$ and $C_{B\overline{B}}$ fixed to values obtained as weighted averages over its various sub-components in simulation (Appendix B); its tag-flavour-dependent yields are parameterised analogously to Eq. 6.7, with $C_{B\overline{B}}$ in place of C.

Figure 6.6(a) shows the resulting Δt fit, used to extract $\tau_{B\overline{B}}$, applied individually to $B\overline{B}$ background MC events, for Belle MR1 as an example; Fig. 6.6(b) shows the same fit split by $B_{\rm tag}$ flavour.



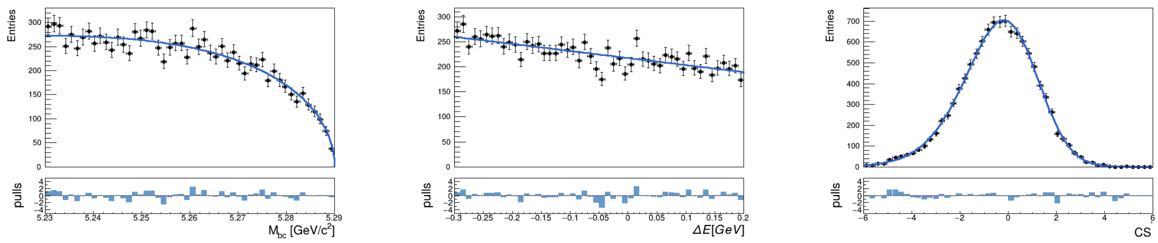
(a) Lifetime fit (b) Tag-flavour split

Figure 6.6: $B\overline{B}$-MC-only Δt fits, shown for Belle MR1 as an example: (a) the fit used to extract the effective $B\overline{B}$ lifetime $\tau_{B\overline{B}}$; (b) the same fit split by $B_{\rm tag}$ flavour, with the raw asymmetry shown below the panel. The same fits are performed for Belle II and for MR2.

6.2.3 Continuum Background

Signal-Extraction Model. The continuum background is modelled as the product of independent one-dimensional PDFs, since its correlations among the three variables are negligible: an ARGUS function [68] for M_{bc}, a first-order polynomial for ΔE, and a Johnson S_U function for $\mathcal{C}'$. The ARGUS function is an empirical shape, developed by the ARGUS collaboration to describe combinatorial backgrounds bounded by the kinematic endpoint of a phase-space-limited process, that falls smoothly to zero at $M_{\text{bc}} = \sqrt{s}/2c^2$.

Figure 6.7 shows this model fitted individually to continuum MC events, for Belle MR1 as an example; the M_{bc} projection visibly falls away towards the kinematic endpoint, the signature the ARGUS function is designed to describe.



(a) M_{bc} (b) ΔE (c) $\mathcal{C}'$

Figure 6.7: The signal-extraction model of this section, fitted individually to continuum MC events, shown for Belle MR1 as an example, with pull distributions shown below each panel. The same fit is performed for Belle II and for MR2.

***CP*-Violation Model.** The continuum background is expected to decay promptly and can be modelled using the same resolution function as the signal (Section 6.2.1), but with the mean and width of its core Gaussian floated, and is assumed to exhibit no *CP* violation, so its yield is split evenly between the two tag flavours. The $c\bar{c}$ component, however, can have displaced vertices from the D-meson decay; nevertheless, it can still be modelled with the same resolution function as the signal.

Figure 6.8 shows this Δt fit applied individually to continuum MC events, for Belle II MR1 as an example; since no *CP* violation is assumed for this component, its yield is split evenly between tag flavours by construction.

6.2.4 Overall Signal-Extraction Model

Combining the three components' SE models above, the overall signal-extraction PDF is

$$P_{\text{tot}}^{\text{SE}}(M_{\text{bc}}, \Delta E, \mathcal{C}') = \sum_{i=1}^{3} N_i \, P_i^{\text{SE}}(M_{\text{bc}}, \Delta E, \mathcal{C}'), \tag{6.8}$$

with the three yields N_i floated in the fit. Figure 6.9 shows the resulting one-dimensional projections of this fit, for Belle II MR1 as an example.

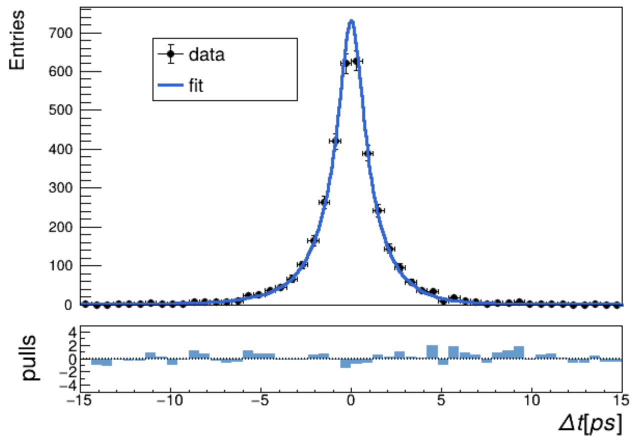


Figure 6.8: The continuum Δt fit, fitted individually to continuum MC events, shown for Belle II MR1 as an example. The same fit is performed for Belle and for MR2.

6.3 Total Fit

The full likelihood is the product of four terms, one for each combination of the TD/TI classification and the two tag flavours $q_{\text{tag}} = \pm 1$:

$$\mathcal{L}_{\text{tot}} = \mathcal{L}_{\text{TD}}^{q_{\text{tag}}=+1} \cdot \mathcal{L}_{\text{TD}}^{q_{\text{tag}}=-1} \cdot \mathcal{L}_{\text{TI}}^{q_{\text{tag}}=+1} \cdot \mathcal{L}_{\text{TI}}^{q_{\text{tag}}=-1}. \tag{6.9}$$

This total fit has five floating observables, the signal, $B\overline{B}$, and continuum yields, and S and C, together with nine floating nuisance parameters absorbing residual data–MC differences: the means of the signal M_{bc} and ΔE PDFs; the M_{bc} endpoint-shape, ΔE slope, and $\mathcal{C}'$ mean and width parameters of the continuum PDF; and the mean, width, and lifetime parameters of the continuum and $B\overline{B}$ Δt shapes.

As a closure test, this fit strategy is applied to a data-equivalent, single-stream MC sample for each dataset and mass region. Representative results for Belle are $N_{\text{sig}} = 707 \pm 36$ (MR1) and 777 ± 44 (MR2), against generated values of 709 and 760 respectively, with S and C both consistent with their generated value of zero within uncertainties in all cases (Belle and Belle II, MR1 and MR2), as expected since no CP violation is generated in this closure MC.

A more statistically stringent version of the same closure test is performed using the full available generic MC statistics, equivalent to six times the Belle dataset and four times the Belle II dataset, so that any small residual fit bias, which might be masked by statistical fluctuations in a single-stream sample, would be more clearly resolved.

Table 6.2 lists the resulting S and C values: all eight results (two datasets, two mass regions, two parameters) remain statistically consistent with the generated value of zero.

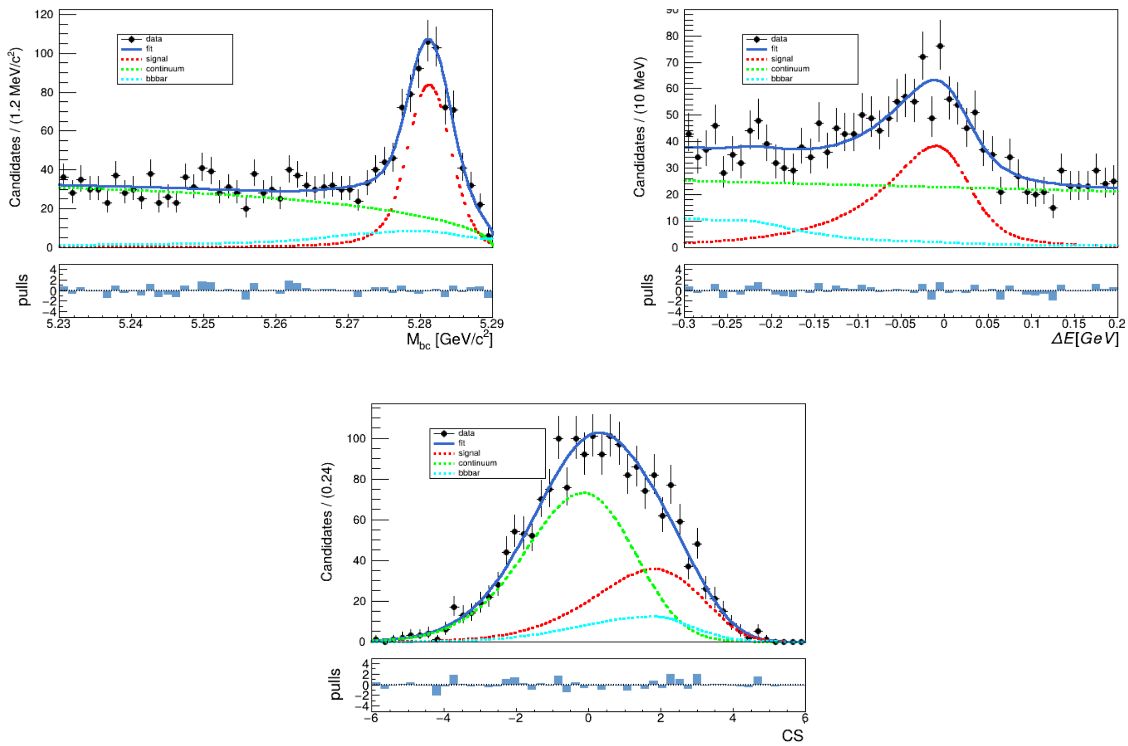


Figure 6.9: Projections of the signal-extraction fit onto $M_{\rm bc}$, ΔE, and $\mathcal{C}'$, shown for Belle II MR1 as an example, with pull distributions shown below each panel; each panel overlays the data, the total fit, and the individual signal, $B\overline{B}$, and continuum component curves. The same fit is performed for Belle and for MR2.

Figure 6.10 shows the Δt distributions and raw asymmetry underlying this closure test, separately by $B_{\rm tag}$ flavour, for signal MC, for Belle II MR1 as an example: the near-perfect overlap of the $B^0_{\rm tag}$- and $\overline{B}^0_{\rm tag}$-tagged distributions, and the flat, null asymmetry, are the visual counterpart of the fitted S and C values being consistent with zero.

Table 6.2: S and C from the total fit applied to high-statistics generic MC (six times the Belle dataset, four times the Belle II dataset), against a generated value of zero in both cases.

Dataset	Mass region	S	C
Belle (6×)	MR1	0.01 ± 0.09	-0.01 ± 0.04
	MR2	0.23 ± 0.10	-0.02 ± 0.05
Belle II (4×)	MR1	-0.08 ± 0.10	-0.07 ± 0.05
	MR2	-0.21 ± 0.09	-0.00 ± 0.05

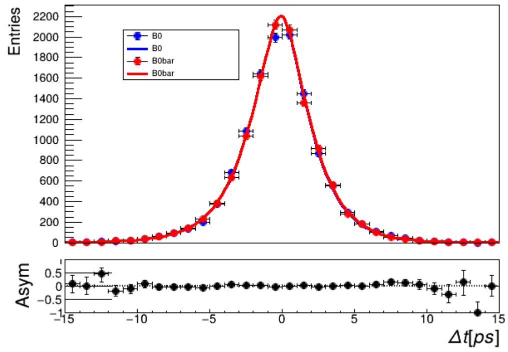


Figure 6.10: Closure-test Δt fit projections for signal MC, separated by B_{tag} flavour, with the corresponding raw asymmetry shown below the panel, shown for Belle II MR1 as an example. The same fit is performed for Belle and for MR2.

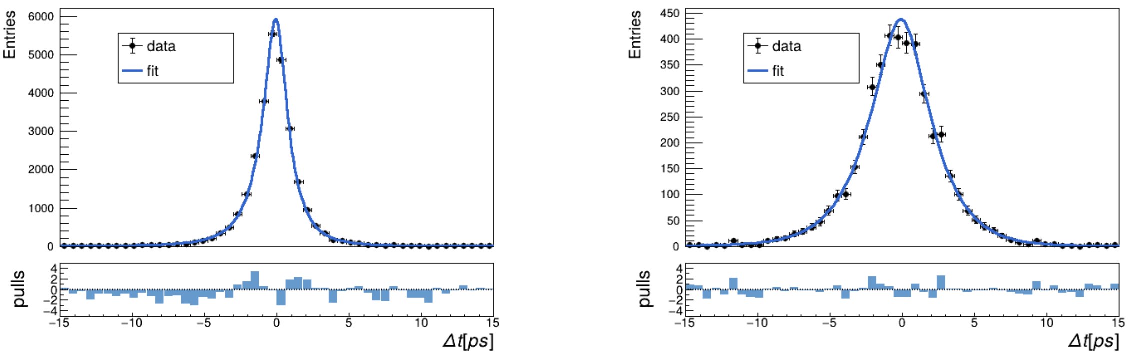
6.4 Control Channels

Two control channels validate different aspects of the fit strategy: a very clean, high-statistics sample of $B^0 \to J/\psi\, K_S^0$ decays, used to calibrate the Δt resolution function and validate the CP-violation fit itself; and a higher-statistics, topologically similar sample of $B^0 \to K^+\pi^-\gamma$ decays, used to derive data–MC correction factors for the signal-extraction PDF shapes.

6.4.1 $B^0 \to J/\psi\, K_S^0$ Control Channel

The channel $B^0 \to J/\psi\, K_S^0$, with $J/\psi \to e^+e^-$ or $\mu^+\mu^-$, is reconstructed using the same K_S^0 selection as the signal channel, with lepton-identification requirements (including bremsstrahlung-photon recovery for electrons, which recovers energy radiated by the electron in the detector material and reassigns it to the electron candidate) applied to the two leptons, and a narrow J/ψ mass window. Tight, slightly asymmetric M_{bc} and ΔE requirements and a continuum-suppression requirement on the event-shape variable R_2 (the normalised Fox–Wolfram moment [54], small for isotropic $B\overline{B}$-like events and large for jet-like continuum) are applied to the B candidate. The B_{sig} vertex fit excludes the J/ψ candidate, so that the vertexing procedure mirrors the signal channel as closely as possible, relying primarily on the displaced K_S^0 vertex; the B_{tag} vertexing and flavour tagging are identical to the signal channel. This selection yields an extremely clean sample.

Decay-Time Resolution Calibration. The Δt_{res} distribution of correctly reconstructed signal MC events is fitted with the resolution-function model of Eq. 6.6, using several streams of simulated events, yielding the calibrated resolution-function parameters. An analogous fit performed on data, with the signal lifetime τ_{sig} and several resolution-function parameters floating, provides the resolution-function parameters ultimately propagated into the signal Δt PDF of Eq. 6.5 for the main $B^0 \to K_S^0 \pi^0 \gamma$ fit. The fitted value of τ_{sig} is consistent with the PDG and serves as a validation of the calibration. Figure 6.11 shows the resulting resolution-function fit and the subsequent lifetime fit to data, for Belle as an example.



(a) Resolution-function fit (Δt_{res})

(b) Lifetime fit to data (Δt)

Figure 6.11: Left: fit of the decay-time-residual model of Eq. 6.6 to signal MC events, shown for Belle as an example. Right: the corresponding fit of the Δt distribution in data, with τ_{sig} and resolution-function parameters floating. The same fits are performed for Belle II.

Validation of the *CP*-Violation Fit. The same control channel validates the *CP*-violation fit itself, since its dominant decay amplitude gives an SM expectation of $S_{J/\psi K_S^0} \simeq \sin 2\phi_1$ (Section 1.2) and negligible direct *CP* violation. Fitting the same fit function used for the main analysis to data yields $S_{J/\psi K_S^0} = 0.65 \pm 0.06$ and $C_{J/\psi K_S^0} = 0.06 \pm 0.03$ for Belle, and $S_{J/\psi K_S^0} = 0.64 \pm 0.05$ and $C_{J/\psi K_S^0} = -0.04 \pm 0.04$ for Belle II, consistent with world averages [2] and thereby validating the end-to-end fit procedure, as illustrated in Fig. 6.12 for Belle II as an example. A further cross-check, fitting data with the tag-side flavour information randomised, yields results consistent with a null asymmetry, as expected in the absence of any genuine tag-flavour correlation.

6.4.2 $B^0 \to K^+ \pi^- \gamma$ Control Channel

The channel $B^0 \to K^{*0}(892)[\to K^+\pi^-]\gamma$ shares the prompt high-energy photon of the signal channel, giving it very similar M_{bc}, ΔE, and $\mathcal{C}'$ distributions, while being both cleaner (lacking a π^0 in its final state) and approximately four times more abundant than the signal channel; it is reconstructed identically to the signal channel except that the $K_S^0\pi^0$ pair is replaced by an oppositely charged, particle-identified $K^+\pi^-$ pair restricted to the $K^{*0}(892)$ mass window.

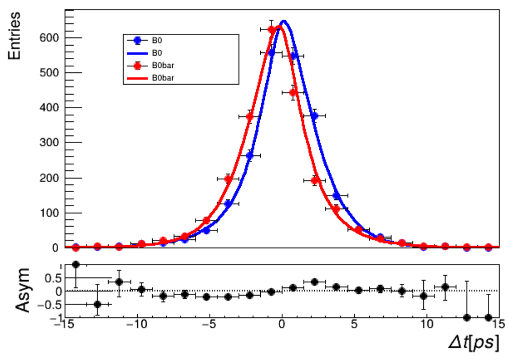
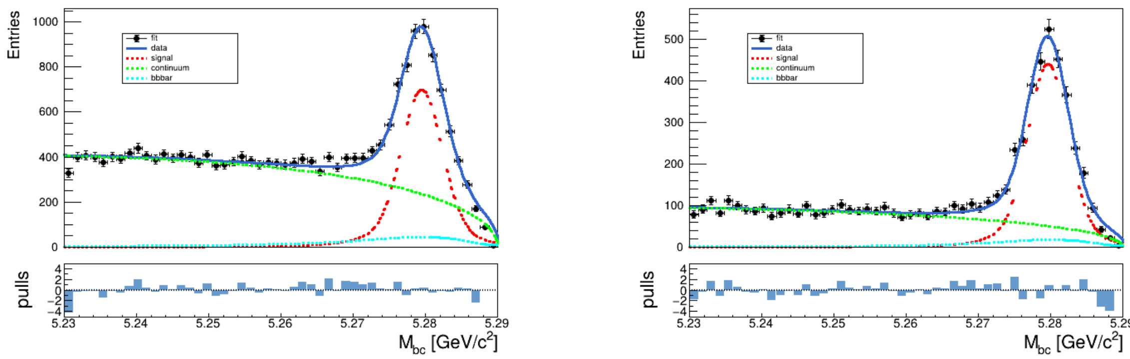


Figure 6.12: Tag-flavour-separated Δt distributions and their fit, for the $B^0 \to J/\psi K^0_S$ sample, shown for Belle II as an example, with the raw flavour asymmetry shown below the panel. The same fit is performed for Belle.

Signal-Extraction Validation. Applying the signal, $B\overline{B}$, and continuum SE models of Section 6.2, with their shape parameters now floated rather than fixed, gives an adequate description of both simulated and real data in this control channel (Fig. 6.13): fits to data-equivalent MC recover the input yields within approximately 2σ, and fits to data itself allow the extraction of scale factors for the widths of the signal $M_{\rm bc}$, ΔE, and $\mathcal{C}'$ PDFs (within a few percent of unity) and fudge factors for the means of the signal and $B\overline{B}$ $\mathcal{C}'$ PDFs (of order 10% for the $B\overline{B}$ component). These correction factors are propagated into the SE fit of the main $B^0 \to K^0_S\pi^0\gamma$ channel.



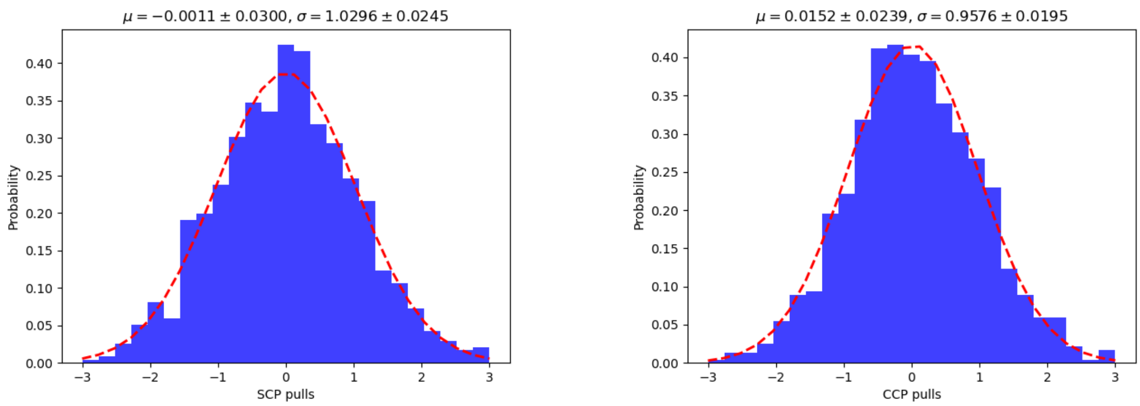
(a) Belle (b) Belle II

Figure 6.13: Signal-extraction fit projection onto $M_{\rm bc}$ for the $B^0 \to K^{*0}(892)[\to K^+\pi^-]\gamma$ control-channel data, in (a) Belle and (b) Belle II, using the same functional forms as the main $B^0 \to K^0_S\pi^0\gamma$ fit (Section 6.2) but with shape parameters floated.

6.5 Fitter Validation with Pseudoexperiments

The fit strategy is further validated using large ensembles of simulated pseudoexperiments ("toys"), in each case generating and fitting 1000 data samples per experiment and mass region, with a size equivalent to one stream of simulated data subject to Poisson fluctuations. For each fitted quantity, a *pull*, the difference between the fitted and generated values divided by the fitted uncertainty, is computed. If the fit is unbiased and correctly estimates its own uncertainties, the resulting pull distribution should be consistent with a standard normal distribution (zero mean, unit width).

Pure toys are sampled directly from the total fit PDF. The resulting pulls for all yields, S, and C are consistent with a standard normal distribution in both mass regions and both datasets, as illustrated in Fig. 6.14 for Belle II MR1 as an example.

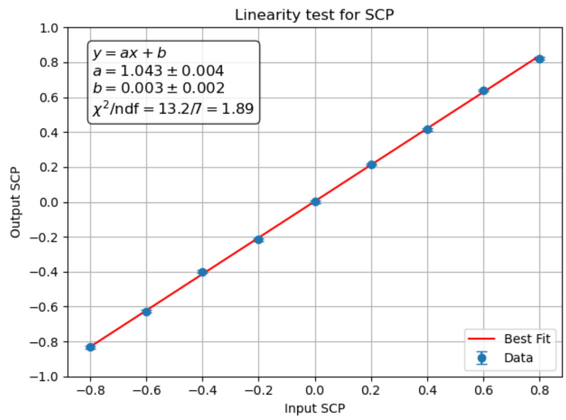


(a) S ($\mu = -0.001 \pm 0.030$, $\sigma = 1.030 \pm 0.025$) (b) C ($\mu = 0.015 \pm 0.024$, $\sigma = 0.958 \pm 0.020$)

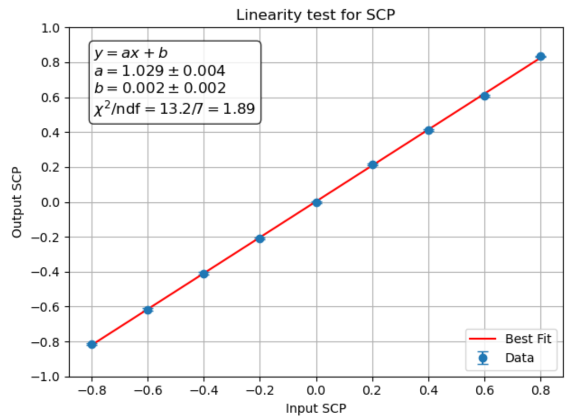
Figure 6.14: Pull distributions of S and C from 1000 pure toys, consistent with the standard normal distribution overlaid in red, shown for Belle II MR1 as an example; this confirms that the fit is unbiased and correctly estimates its own uncertainties for both *CP*-violation parameters. The same toy study is performed for Belle and for MR2.

GSim toys instead resample the signal component, with replacement, directly from signal MC events, while the $B\overline{B}$ and continuum components are still sampled from their analytic PDFs; this tests the sensitivity of the fit to any mismodelling of the signal shape by the analytic PDFs of Section 6.2. These toys reveal a small bias in the C pulls for Belle MR1 and Belle II MR2, but not in the other two combinations. Repeating the same toys with the tag-side flavour information randomised removes this bias entirely, indicating that it originates from a genuine, intrinsic tag-flavour asymmetry present in the underlying signal MC sample itself, rather than from any flaw in the fitter, and is accounted for as a fit-bias systematic uncertainty.

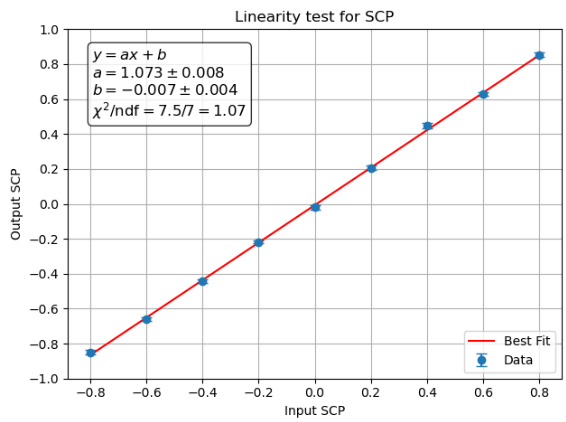
Linearity tests scan the generated values of S and C across a range of inputs and compare the mean fitted value at each point to the generated value. The fitted values track the generated ones within uncertainties across this scan, for S (Fig. 6.15) and C (Fig. 6.16), in all four dataset/mass-region combinations, confirming that the fit responds linearly and without significant bias to the size of a genuine *CP* asymmetry, were one present.



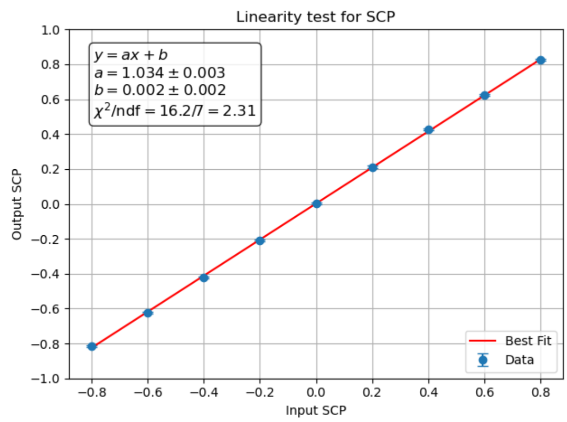
(a) Belle, MR1 (slope 1.043 ± 0.004, intercept 0.003 ± 0.002)



(b) Belle II, MR1 (slope 1.029 ± 0.004, intercept 0.002 ± 0.002)



(c) Belle, MR2 (slope 1.073 ± 0.008, intercept -0.007 ± 0.004)



(d) Belle II, MR2 (slope 1.034 ± 0.003, intercept 0.002 ± 0.002)

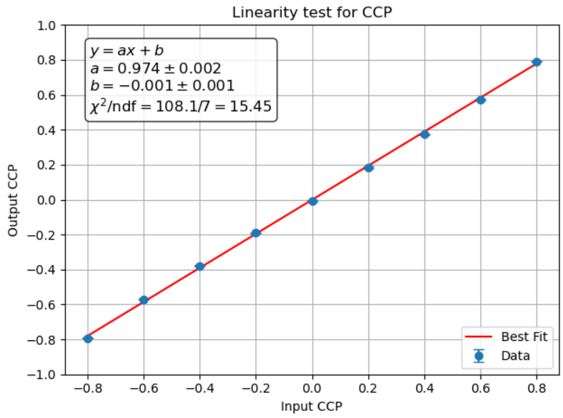
Figure 6.15: Linearity test for S: the mean fitted value at each of several generated input values, together with a linear fit, in all four dataset/mass-region combinations, confirming the fit responds linearly across the full physical range of S.

6.6 Closure Tests and Unblinding Procedure

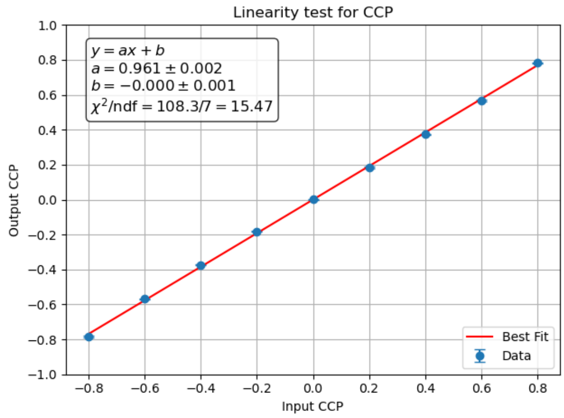
Before the fit is applied to unblinded data, several data–MC corrections and a sequence of blind validation checks are carried out.

6.6.1 Preparing the Fitter for Data

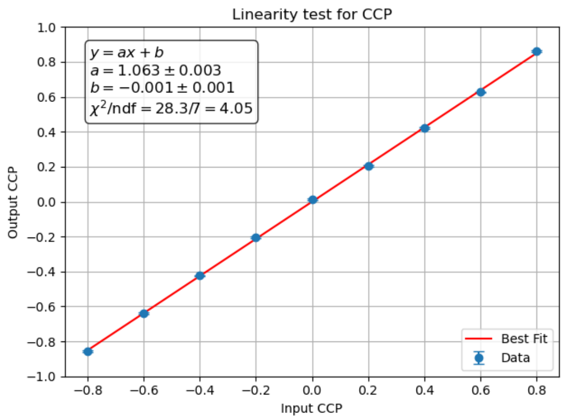
The signal and $B\overline{B}$ shape corrections of Section 6.4.2 are applied to both mass regions, and the mistag parameters w, Δw, and μ are updated using the latest central flavour-tagging calibration, remodelled as continuous functions of r. The resolution-function parameters are replaced by their data-calibrated values from Section 6.4.1.



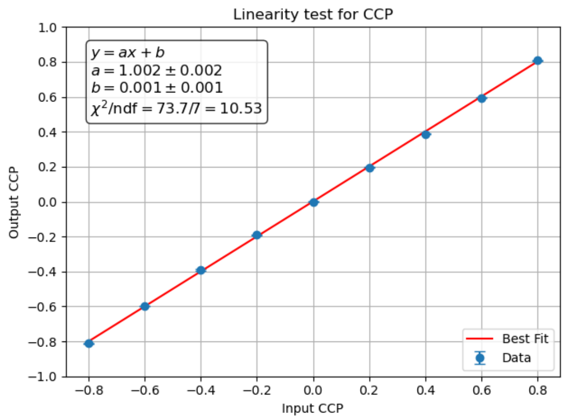
(a) Belle, MR1 (slope 0.974 ± 0.002, intercept -0.001 ± 0.001)



(b) Belle II, MR1 (slope 0.961 ± 0.002, intercept -0.000 ± 0.001)



(c) Belle, MR2 (slope 1.063 ± 0.003, intercept -0.001 ± 0.001)



(d) Belle II, MR2 (slope 1.002 ± 0.002, intercept 0.001 ± 0.001)

Figure 6.16: Linearity test for C: the mean fitted value at each of several generated input values, together with a linear fit, in all four dataset/mass-region combinations, confirming the fit responds linearly across the full physical range of C.

The TD-sample fractions f_i^{TD} of Eqs. 6.1 and 6.2, originally fixed from simulation, are also corrected for data: $f_{\mathrm{sig}}^{\mathrm{TD}}$ is corrected using a data–MC scale factor measured in the $B^0 \to J/\psi K_S^0$ control channel (0.938 for Belle, 1.001 for Belle II); $f_{q\bar{q}}^{\mathrm{TD}}$ is instead measured directly in a continuum-enriched data sideband. Rather than fixing $f_{B\bar{B}}^{\mathrm{TD}}$ from simulation, it is left to float in the fit, constrained by the observed number of TD events, N_{TD}, relative to the total, N_{tot}, via

$$\frac{N_{\mathrm{TD}}}{N_{\mathrm{tot}}} = \frac{N_{\mathrm{sig}} f_{\mathrm{sig}}^{\mathrm{TD}} + N_{B\bar{B}} f_{B\bar{B}}^{\mathrm{TD}} + N_{q\bar{q}} f_{q\bar{q}}^{\mathrm{TD}}}{N_{\mathrm{sig}} + N_{B\bar{B}} + N_{q\bar{q}}}, \tag{6.10}$$

which relates the unknown $f_{B\bar{B}}^{\mathrm{TD}}$ to quantities that are either already floating in the fit or directly measurable by counting events, without needing to trust its simulated value.

6.6.2 Sequential Unblinding Checks

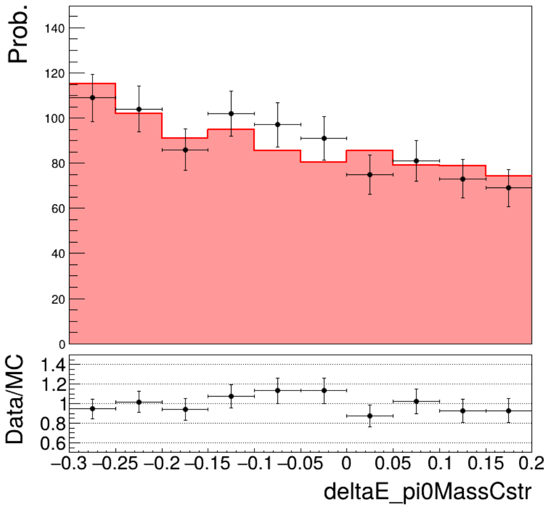
Following these corrections, a series of checks is performed, in order, before the central values of S and C are examined in data:

1. **Background-shape validation**: the M_{bc}, ΔE, and $\mathcal{C}'$ shapes of the $B\overline{B}$ and continuum PDFs are compared between data and simulation in dedicated sideband regions, for example a continuum-enriched sideband ($M_{\mathrm{bc}} \in [5.23, 5.26]\,\mathrm{GeV}/c^2$ with $\mathcal{C}' < 0$, or $\Delta E > 0.1\,\mathrm{GeV}$ with $\mathcal{C}' > 0$) and a $B\overline{B}$-enriched sideband ($M_{\mathrm{bc}} > 5.26\,\mathrm{GeV}/c^2$ with $\Delta E < -0.15\,\mathrm{GeV}$, after subtracting the residual continuum contribution). These are found to agree well, as illustrated for the ΔE shape in the continuum-enriched sideband in Fig. 6.17.

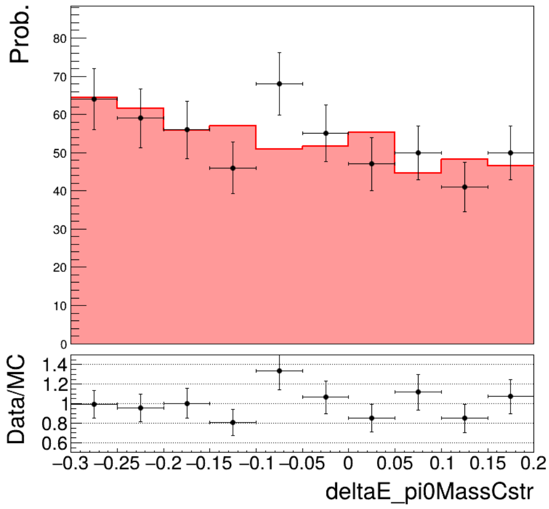
2. **Signal-extraction fits to data**, with the *CP*-violation parameters (but not the yields or shape parameters) still blinded.

3. ***B*-lifetime and *CP*-violation fits to data with the tag-side flavour information randomised**, which allows residual biases in the fitter to be probed without revealing the true *CP* asymmetry: the fitted lifetimes (Fig. 6.18) are consistent with the PDG value of $1.519 \pm 0.004\,\mathrm{ps}$ [2] within uncertainties in all four dataset/mass-region combinations, while repeating the *CP*-violation fit over 5000 independent random tag-flip seeds yields S and C distributions consistent with zero, as expected in the absence of any genuine tag-flavour correlation.

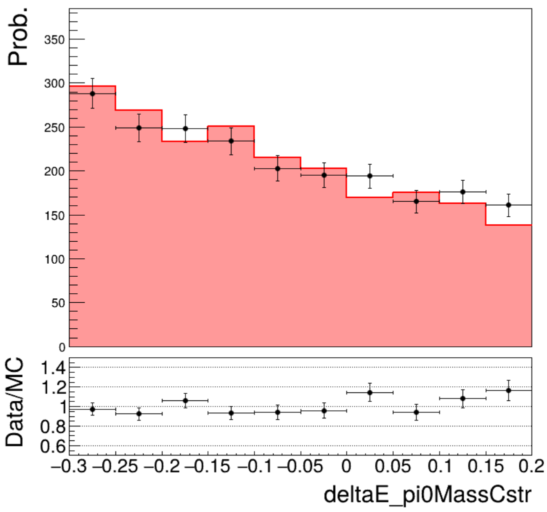
Only once all of these checks are complete are the blinded central values of S and C revealed, yielding the results presented in Chapter 7.



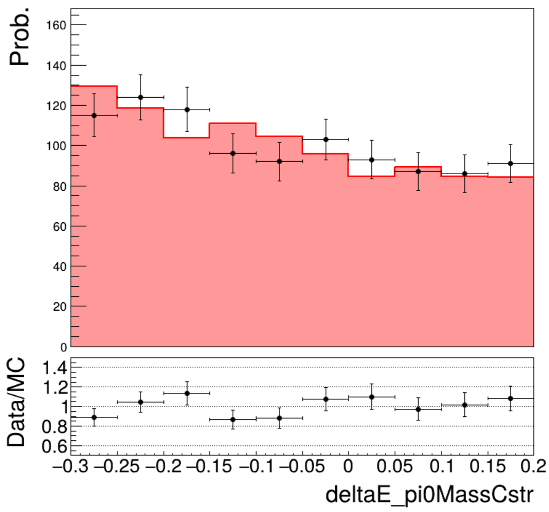
(a) Belle, MR1



(b) Belle II, MR1



(c) Belle, MR2



(d) Belle II, MR2

Figure 6.17: Data-MC comparison of the ΔE shape in the continuum-enriched sideband ($M_{\rm bc} \in [5.23, 5.26]\,{\rm GeV}/c^2$, $\mathcal{C}' < 0$), for all four dataset/mass-region combinations; the good agreement validates the use of simulation-derived continuum shapes in the signal-extraction fit.

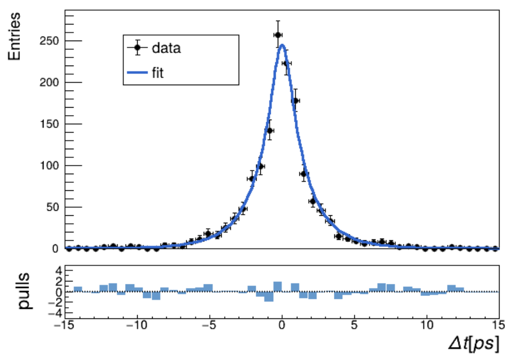


Figure 6.18: Δt fit to data with the tag-side flavour information randomised, floating only the B lifetime, shown for Belle II MR1 as an example.

Chapter 7

Systematics, Results, and Interpretation

Having presented the fit strategy and its validation (Chapter 6), this chapter presents the systematic uncertainties considered in the measurement, the resulting values of S and C for each dataset and mass region, their combination across Belle and Belle II, a comparison with previous measurements, and the physics interpretation of the results.

7.1 Systematic Uncertainties

Nine categories of systematic uncertainty are considered, evaluated separately for each dataset (Belle, Belle II) and mass region (MR1, MR2). For most categories, the same procedure is used: the relevant fixed parameter or parameters are resampled from an appropriate distribution and the full fit to data is repeated using each sampled set of values, with the resulting spread in the fitted S and C values taken as the systematic uncertainty from that source; the resampling distribution itself is either a Gaussian centred on the parameter's nominal value with a width equal to its uncertainty (accounting for correlations between parameters where relevant), or, where a dedicated likelihood is available, direct likelihood sampling via a Markov Chain Monte Carlo (MCMC), which respects the full correlation structure of that likelihood surface without requiring a Gaussian approximation to it. A small number of categories, vertex measurement, tag-side interference, and fit bias, instead rely on dedicated, external studies or direct data-driven checks, described in their respective subsections below.

7.1.1 Flavour Tagging

The centrally provided flavour-tagging calibration parameters (Section 5.5) carry their own uncertainties, which propagate into the mistag parameters w, Δw, and μ modelled as functions of r in the fit. This systematic uncertainty is evaluated by sampling these parameters from a multivariate normal distribution, accounting for correlations, and taking the resulting spread in S and C as the systematic uncertainty (Table 7.1).

Table 7.1: Systematic uncertainty on S and C due to flavour tagging, and their correlation.

	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
S	± 0.010	± 0.019	± 0.002	± 0.004
C	± 0.021	± 0.039	± 0.007	± 0.011
Correlation	-0.176	-0.406	0.715	-0.472

7.1.2 Resolution Function

Similarly, the resolution-function parameters calibrated from the $B^0 \to J/\psi K^0_S$ control channel (Section 6.4.1) carry uncertainties from the finite size of that calibration sample. These are propagated by sampling the resolution-function parameters from the likelihood of the calibration fit itself, using the MCMC method, and refitting the data with each sampled set (Table 7.2). MCMC is an algorithm that draws correlated random samples from a specified probability distribution, here the fit's own likelihood surface, allowing the full correlation structure between resolution parameters to be respected.

Table 7.2: Systematic uncertainty on S and C due to the resolution function, and their correlation.

	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
S	± 0.018	± 0.030	± 0.007	± 0.016
C	± 0.002	± 0.006	± 0.003	± 0.005
Correlation	-0.405	0.841	-0.341	-0.082

7.1.3 $B\overline{B}$ *CP* Asymmetries

The $B\overline{B}$ background comprises many B-decay channels, each with its own (generally unmeasured) *CP*-violation content, whose average behaviour is what is actually fixed in the fit as the effective parameters $S_{B\overline{B}}$ and $C_{B\overline{B}}$ of the $B\overline{B}$ *CP*-violation model (Section 6.2.2). The background composition of $B\overline{B}$ events in simulation is classified by decay type, and central values and uncertainties for S and C are assigned to each: B decays to charmed final states, and all charged-B decays (for S, since a well-defined mixing-induced asymmetry requires a neutral B meson) are assigned zero; channels with a measured *CP* asymmetry are assigned their world-average central value [2]; and all remaining channels are conservatively assigned zero with an uncertainty of ± 0.3 (for C) or ± 0.5 (for S, and only where the final state is a *CP* eigenstate). The full background composition, and the asymmetry assigned to each channel, is summarised in Appendix B (Tables B.1–B.4, for each dataset and mass region).

Weighting each channel's assigned asymmetry by its fractional contribution to the $B\overline{B}$ background gives the $S_{B\overline{B}}$ and $C_{B\overline{B}}$ values ultimately fixed in the fit of Section 6.2.2 (for example, $C_{B\overline{B}} = 0.01 \pm 0.12$ and $S_{B\overline{B}} = 0.01 \pm 0.12$ for Belle MR1, as shown in Table B.1). The corresponding uncertainty is evaluated by resampling each channel's assigned asymmetry within its assigned uncertainty and repeating this weighted average and subsequent fit. This is the dominant systematic uncertainty on S in MR2 for both datasets, reflecting the larger relative $B\overline{B}$ background contribution there (Section 6.2.2) and the larger uncertainties assigned to many of its constituent channels; since the same underlying branching fractions and world-average inputs are used for both experiments, it is treated as 100% correlated between Belle and Belle II, as listed in Table 7.3.

Table 7.3: Systematic uncertainty on S and C due to $B\overline{B}$ CP asymmetries, and their correlation.

	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
S	± 0.013	± 0.076	± 0.010	± 0.081
C	± 0.007	± 0.046	± 0.014	± 0.084
Correlation	0.083	-0.083	0.294	-0.174

7.1.4 Signal/Self-Cross-Feed Fraction

The fraction of self-cross-feed within the signal component (Section 6.2.1), fixed from simulation, is varied by $\pm 20\%$ of its nominal value, and the resulting variation in S and C is taken as the systematic uncertainty (Table 7.4).

Table 7.4: Systematic uncertainty on S and C due to the SCF fraction, and their correlation.

	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
S	± 0.002	± 0.006	± 0.003	± 0.005
C	± 0.001	± 0.001	± 0.002	± 0.002
Correlation	1.0	-1.0	1.0	-1.0

7.1.5 Vertex Measurement

Three sources of vertex-related systematic uncertainty are considered: detector misalignment, adopted from dedicated studies performed for related $B \to K_S^0 \pi^0 \gamma$ and $B \to K_S^0 \eta \gamma$ analyses; uncertainty in the interaction-point profile, adopted from the $B^0 \to J/\psi K_S^0$ analyses in both experiments [57]; and uncertainty in the scale-error correction applied to charged-track trajectory uncertainties, for which no dedicated uncertainty estimate exists. Following the same assumption used in the $B^0 \to J/\psi\, K^0_S$ analysis, this last contribution is instead estimated by supposing that any bias in this correction would manifest as a bias in the measured B^0 lifetime and mixing frequency relative to their world-average values [2]; refitting the data with τ_{B^0} and Δm_d fixed to their (Belle- or Belle II-measured, rather than world-average) values, and taking the resulting shift in S and C as the associated systematic uncertainty. The three contributions are added in quadrature, as summarised in Table 7.5.

Table 7.5: Systematic uncertainty on C and S due to vertex measurement, by source.

Source	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
C				
Detector misalignment	± 0.004	± 0.004	± 0.001	± 0.001
Scale error	± 0.001	± 0.007	± 0.001	± 0.002
IP profile	± 0.004	± 0.004	± 0.002	± 0.002
Total	± 0.006	± 0.009	± 0.002	± 0.003
S				
Detector misalignment	± 0.002	± 0.002	± 0.005	± 0.005
Scale error	± 0.005	± 0.009	± 0.001	± 0.014
IP profile	± 0.022	± 0.022	± 0.002	± 0.002
Total	± 0.023	± 0.024	± 0.005	± 0.015

7.1.6 Tag-Side Interference

Tag-side interference arises from interference between the CKM-favoured ($b \to c\bar{u}d$) and CKM-suppressed ($b \to u\bar{c}d$) amplitudes contributing to certain flavour-specific $B_{\rm tag}$ decay channels [69]; because the tagging algorithm assumes $B_{\rm tag}$ decays are purely flavour-specific, this interference can bias the extracted S and C if left unaccounted for. The resulting systematic uncertainty is estimated using a dedicated tool that scans over the strong phases of the relevant tag-side decay channels to predict the size of this bias. Since it depends on the same underlying hadronic decay parameters for both experiments, it is treated as 100% correlated between Belle and Belle II. This is the dominant systematic uncertainty on C in both mass regions and both datasets (Table 7.6).

Table 7.6: Systematic uncertainty on S and C due to tag-side interference.

	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
S	± 0.003	± 0.004	± 0.006	± 0.001
C	± 0.036	± 0.036	± 0.034	± 0.036

7.1.7 PDF Shape

Two related but distinct procedures are used depending on the fit component. For the analytic signal and continuum shapes with parameters fixed from control-channel or simulation fits (Section 6.2), the same MCMC resampling procedure used for the resolution function is applied to the corresponding fit likelihood. For the nonparametric $B\overline{B}$ KDE shape (Section 6.2.2), which has no natural set of "parameters" to resample in this way, a bootstrap procedure is instead used: the underlying simulated sample is resampled with replacement 1000 times, a new KDE is constructed from each resampled set, and the data is refitted using each of these alternative KDE shapes in turn, with the spread in S and C taken as the systematic uncertainty (Table 7.7).

Table 7.7: Systematic uncertainty on C and S due to PDF shape, by source, and their correlation.

Source	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
C				
Fixed parameters (MCMC)	± 0.002	± 0.006	± 0.004	± 0.002
$B\overline{B}$ KDE (bootstrap)	± 0.002	± 0.011	± 0.006	± 0.011
Total	± 0.003	± 0.013	± 0.007	± 0.011
S				
Fixed parameters (MCMC)	± 0.004	± 0.006	± 0.002	± 0.014
$B\overline{B}$ KDE (bootstrap)	± 0.007	± 0.023	± 0.004	± 0.027
Total	± 0.008	± 0.024	± 0.004	± 0.030
S–C correlation				
Fixed parameters (MCMC)	0.022	-0.559	-0.032	0.385
$B\overline{B}$ KDE (bootstrap)	0.232	-0.555	0.340	0.054

7.1.8 Fixed Physics Parameters

The B^0 lifetime τ_{B^0} and mixing frequency Δm_d, held fixed in the nominal fit, are varied according to their respective world-average values and uncertainties [2], and the data refitted with each sampled pair of values; the resulting spread in S and C is taken as the systematic uncertainty (Table 7.8). This contribution is negligible compared to the other sources considered.

7.1.9 Fit Bias

Any residual bias in the fitter itself, beyond what is captured by the pseudoexperiment studies of Section 6.5, is estimated directly from data: the full fit is repeated on an ensemble of copies of the actual data sample with the tag-side flavour information randomised (Section 6.6). The resulting deviation of the fitted S and C distributions from zero, the expectation in the absence of any genuine tag-flavour correlation, is assigned as the fit-bias systematic uncertainty (Table 7.9).

Table 7.8: Systematic uncertainty on S and C due to the fixed physics parameters τ_{B^0} and Δm_d, and their correlation.

	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
S	±0.001	±0.001	±0.001	±0.003
C	±0.001	±0.001	±0.001	±0.001
Correlation	-0.675	-0.885	-0.912	-0.930

Table 7.9: Systematic uncertainty on S and C due to fit bias, and their correlation.

	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
S	±0.005	±0.011	±0.001	±0.004
C	±0.001	±0.005	±0.003	±0.002
Correlation	0	0	0	0

7.1.10 Total Systematic Uncertainty

Tables 7.10 and 7.11 summarise the contribution of each source described above to the total systematic uncertainty on S and C, respectively, for each dataset and mass region, with the total obtained by adding all contributions in quadrature.

Table 7.10: Systematic uncertainties on S, by source, dataset, and mass region.

Source	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
Flavour tagging	±0.010	±0.019	±0.002	±0.004
Physics parameters	±0.001	±0.001	±0.001	±0.003
$B\overline{B}$ CP asymmetries	±0.013	±0.076	±0.010	±0.081
SCF fraction	±0.002	±0.006	±0.003	±0.005
Resolution function	±0.018	±0.030	±0.007	±0.016
PDF shape	±0.008	±0.024	±0.004	±0.030
Vertex measurement	±0.023	±0.024	±0.005	±0.015
Tag-side interference	±0.003	±0.004	±0.006	±0.001
Fit bias	±0.005	±0.011	±0.001	±0.004
Total	±0.035	±0.091	±0.016	±0.090

The dominant contributions differ between mass regions: in MR1, the resolution function and vertex-measurement categories dominate the uncertainty on S, while tag-side interference dominates C; in MR2, $B\overline{B}$ CP asymmetries dominate both parameters, reflecting the larger and less well-constrained $B\overline{B}$ background contribution in that region (Table 7.12).

Table 7.11: Systematic uncertainties on C, by source, dataset, and mass region.

Source	Belle MR1	Belle MR2	Belle II MR1	Belle II MR2
Flavour tagging	± 0.021	± 0.039	± 0.007	± 0.011
Physics parameters	± 0.001	± 0.001	± 0.001	± 0.001
$B\bar{B}$ CP asymmetries	± 0.007	± 0.046	± 0.014	± 0.084
SCF fraction	± 0.001	± 0.001	± 0.002	± 0.002
Resolution function	± 0.002	± 0.006	± 0.003	± 0.005
PDF shape	± 0.003	± 0.013	± 0.007	± 0.011
Vertex measurement	± 0.006	± 0.009	± 0.002	± 0.003
Tag-side interference	± 0.036	± 0.036	± 0.034	± 0.036
Fit bias	± 0.001	± 0.005	± 0.003	± 0.002
Total	± 0.043	± 0.072	± 0.038	± 0.093

7.2 Results for S_{CP} and C_{CP}

After unblinding (Section 6.6), the full fit of Chapter 6 is applied to the Belle and Belle II datasets in both mass regions. Table 7.12 gives the resulting signal and background yields in the signal region (Section 5.1), and Table 7.13 the corresponding S and C values for each dataset.

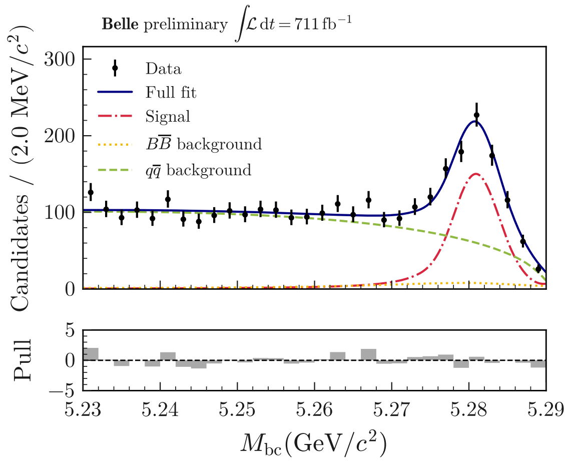
Table 7.12: Signal and background yields obtained from the fits to data, and their ratio.

Mass region	Dataset	$N_{\rm sig}$	$N_{B\bar{B}}$	$N_{q\bar{q}}$	$N_{\rm sig}/(N_{B\bar{B}} + N_{q\bar{q}})$
MR1	Belle	558 ± 29	30 ± 9	350 ± 8	1.47
	Belle II	452 ± 24	34 ± 7	178 ± 6	2.13
MR2	Belle	212 ± 26	216 ± 43	753 ± 21	0.22
	Belle II	213 ± 22	147 ± 14	350 ± 10	0.43

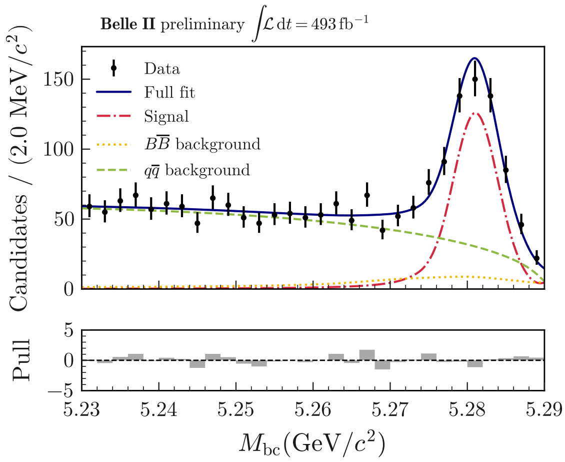
Table 7.13: S and C obtained from the fits to data. The first uncertainty is statistical and the second systematic.

Mass region	Dataset	S	C
MR1	Belle	$0.18 \pm 0.31 \pm 0.04$	$-0.01 \pm 0.12 \pm 0.04$
	Belle II	$0.04 \pm 0.19 \pm 0.02$	$-0.16 \pm 0.10 \pm 0.04$
MR2	Belle	$-0.37 \pm 0.58 \pm 0.09$	$-0.01 \pm 0.29 \pm 0.07$
	Belle II	$-0.29 \pm 0.41 \pm 0.09$	$-0.11 \pm 0.22 \pm 0.09$

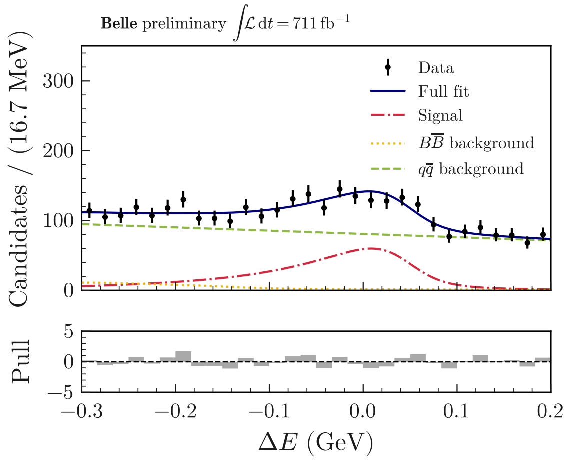
Figures 7.1 and 7.2 show the projections of this same total fit onto $M_{\rm bc}$, ΔE, and $\mathcal{C}'$, for all four dataset/mass-region combinations; unlike the signal-extraction-only projections of Fig. 6.9, these reflect the final, simultaneous fit of Eqs. 6.1 and 6.2, performed jointly with Δt and using the unblinded central values of S and C from Table 7.13.



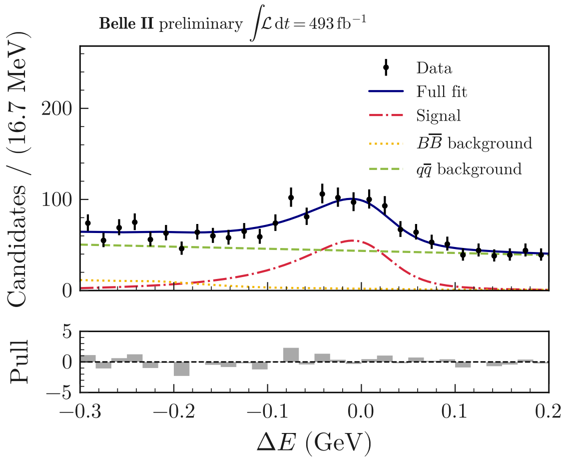
(a) M_{bc}, Belle



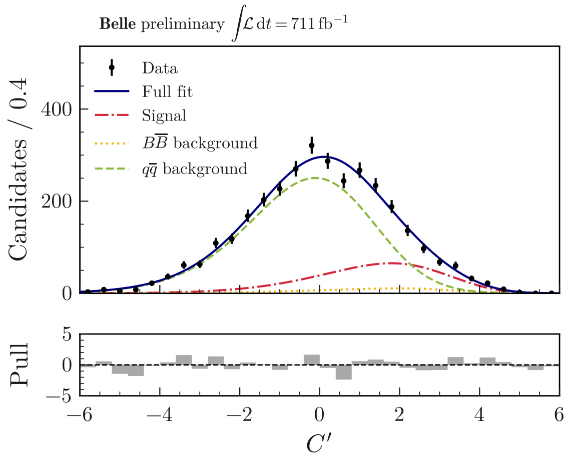
(b) M_{bc}, Belle II



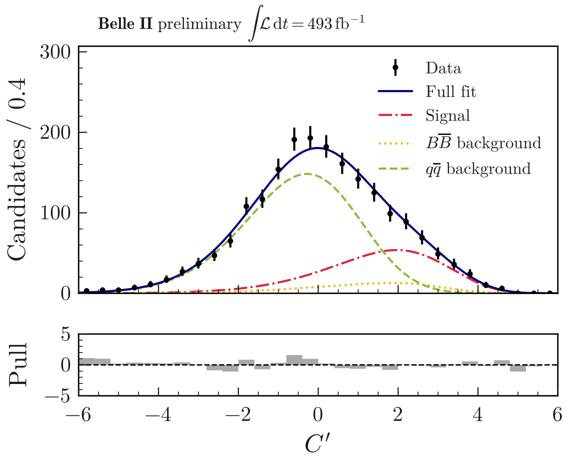
(c) ΔE, Belle



(d) ΔE, Belle II

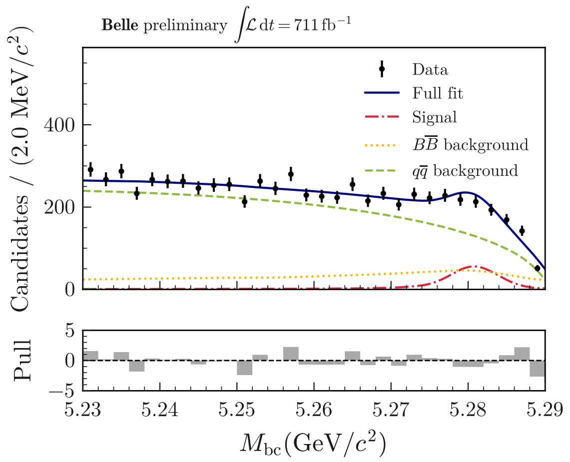


(e) $\mathcal{C}'$, Belle

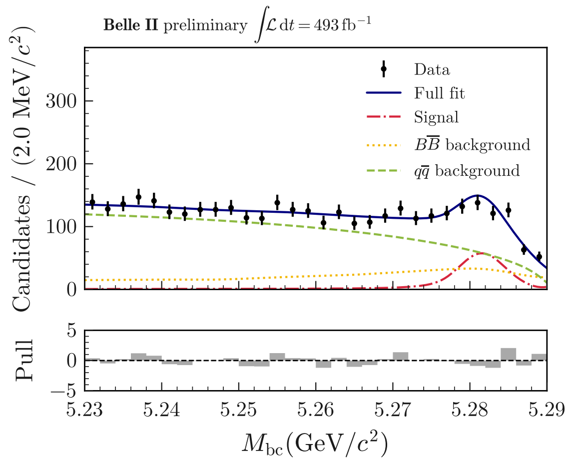


(f) $\mathcal{C}'$, Belle II

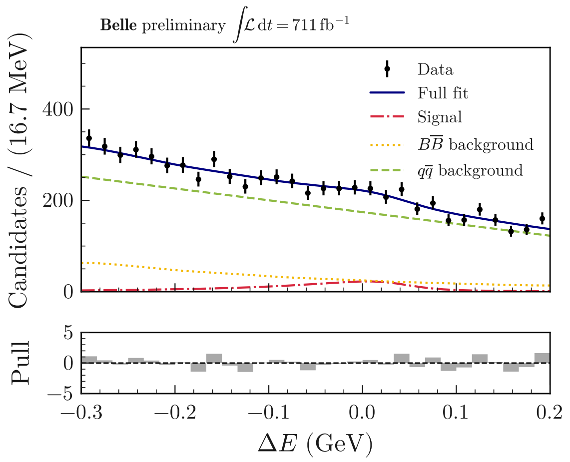
Figure 7.1: Projections of the final, unblinded total fit onto M_{bc}, ΔE, and $\mathcal{C}'$, in MR1, for Belle (left) and Belle II (right), with pull distributions shown below each panel; as in Fig. 6.9, each panel overlays the data, the total fit, and the individual signal, $B\overline{B}$, and continuum component curves.



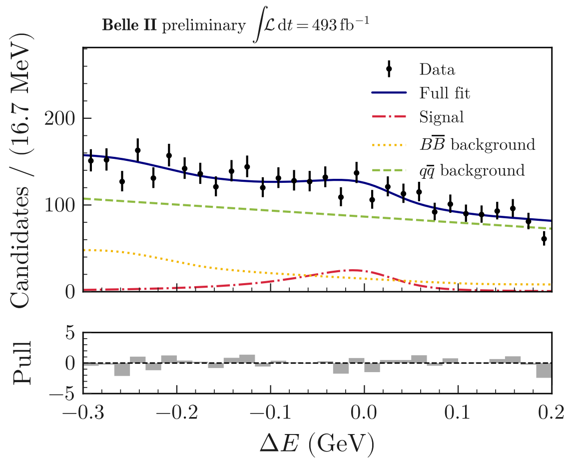
(a) M_{bc}, Belle



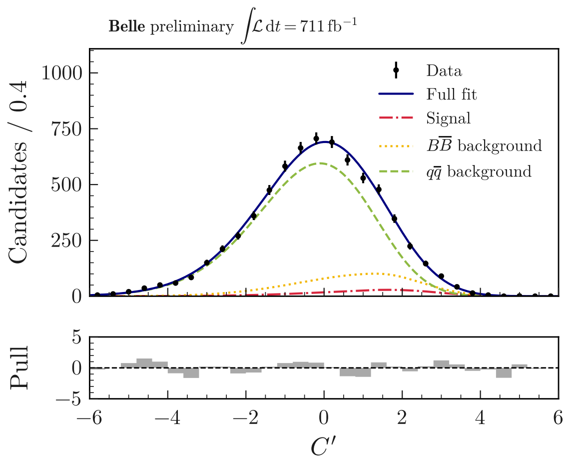
(b) M_{bc}, Belle II



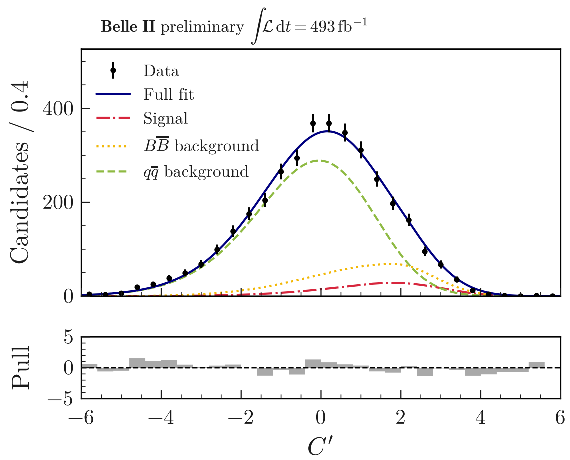
(c) ΔE, Belle



(d) ΔE, Belle II



(e) $\mathcal{C}'$, Belle



(f) $\mathcal{C}'$, Belle II

Figure 7.2: As Figure 7.1, for MR2.

Figure 7.3 shows the background-subtracted Δt distributions and the resulting time-dependent asymmetry, separately for each B_{tag} flavour, in the time-dependent sample: a visible excess of B^0_{tag}-tagged over $\overline{B}^0_{\text{tag}}$ events at positive Δt (and vice versa at negative Δt) would be the signature of a nonzero S value. The C information encoded in these plots corresponds to the TD dataset alone, which is consistent with that obtained from the TI dataset, within 1.4 standard deviations.

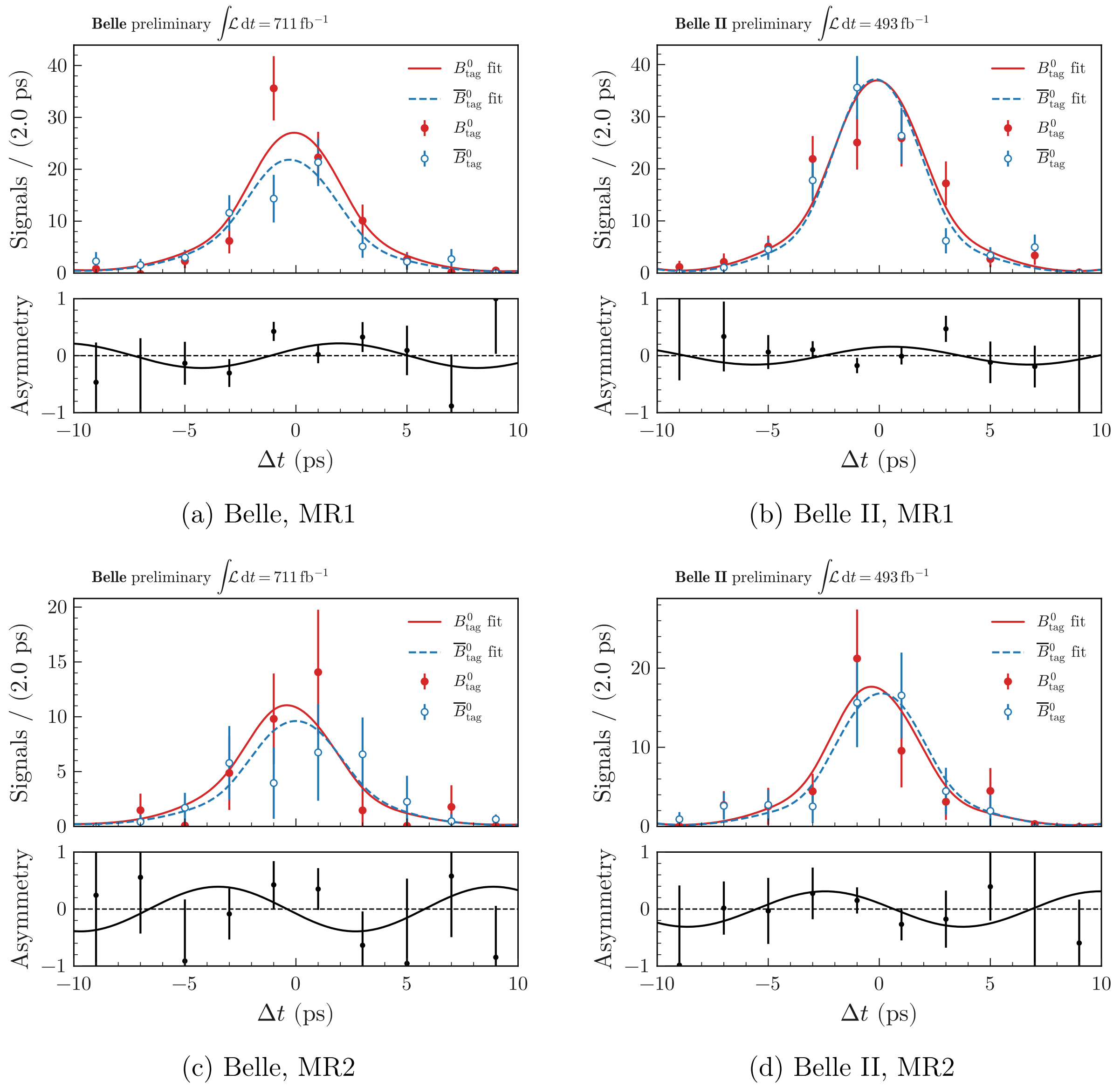


Figure 7.3: Background-subtracted Δt distributions, separated by B_{tag} flavour, for the signal component of the time-dependent sample, with the corresponding fitted asymmetry shown in the lower panel of each plot, for Belle and Belle II in MR1 and MR2.

7.3 Combination of Belle and Belle II Results

The Belle and Belle II results are combined, separately for each mass region, using their weighted average, accounting for correlations in both the statistical and systematic uncertainties using the BLUE (Best Linear Unbiased Estimate) method [70]. Concretely, a full 4×4 covariance matrix is constructed for the two measurements of S and C (Belle and Belle II), combining the statistical covariance with the systematic covariance. Since the two experiments' data samples are entirely independent, the statistical covariance is diagonal between experiments but retains the S–C correlation quoted in Table 7.13 for each experiment separately. In the systematic covariance, the sources treated as 100% correlated between experiments in Section 7.1 (the $B\overline{B}$ CP asymmetries and tag-side interference) contribute off-diagonal terms linking the two experiments. The BLUE weights, which determine each measurement's contribution to the combined result, are then calculated from this covariance matrix so as to minimise the variance of the combined estimate, and the combined central value and its uncertainty follow directly from these weights. Table 7.14 gives the resulting combined values, which constitute the principal result of this thesis, together with the correlation between the combined S and C values in each mass region. Both mass-region results are consistent with the Standard Model (SM) expectation of vanishing CP violation.

Table 7.14: Combined S and C values from the weighted average of the Belle and Belle II measurements, together with their correlation.

Mass region	S	C	Correlation
MR1	$\mathbf{0.09 \pm 0.16 \pm 0.02}$	$\mathbf{-0.09 \pm 0.08 \pm 0.04}$	$+0.05$
MR2	$\mathbf{-0.32 \pm 0.33 \pm 0.09}$	$\mathbf{-0.07 \pm 0.17 \pm 0.08}$	-0.07

7.3.1 Belle II precision improvement over Belle

Table 7.13 shows that the Belle statistical uncertainty on S is substantially larger than the Belle II one, in both mass regions, despite the two experiments' broadly comparable signal yields (Table 7.12). To understand and quantify the origin of this difference, a dedicated pseudoexperiment study was performed, starting from the nominal Belle configuration and successively replacing individual inputs with their Belle II equivalents, one at a time, while holding all other conditions fixed: first the signal-to-background yield ratio $N_{\mathrm{sig}}/N_{q\bar{q}}$, then the flavour-tagging performance (parameterised by the mistag parameters and effective tagging efficiency), and finally the fraction of events entering the time-dependent fit sample, $f_{\mathrm{sig}}^{\mathrm{TD}}$. Table 7.15 shows the resulting S uncertainty in MR1 after each successive replacement.

The flavour-tagging performance and the fraction of events passing the TD selection are the two dominant contributors to the precision difference, together accounting for essentially all of it: once both Belle inputs are replaced by their Belle II equivalents, the resulting uncertainty is statistically indistinguishable from the nominal Belle II value. Both effects trace back to hardware and software improvements in Belle II relative to Belle, respectively the higher effective tagging efficiency of the GFlaT (Section 4.5) and the improved vertex-detector acceptance for the displaced-vertex topology required by the time-dependent sample (Section 5.4). This is rather than any difference in signal purity or fit methodology between the two datasets, validating the overall understanding of the measurement's precision.

Table 7.15: Breakdown, from a dedicated pseudoexperiment study, of the factors contributing to the larger Belle statistical uncertainty on S (MR1) relative to Belle II. Each row successively replaces one more Belle input with its Belle II equivalent, starting from the nominal Belle configuration.

Configuration	$\sigma(S)$	Difference from Belle II
Nominal Belle	0.29	0.09
+ Belle II $N_{\text{sig}}/N_{q\bar{q}}$	0.28	0.08
+ Belle II flavour tagging	0.25	0.05
+ Belle II TD fraction	0.20	0.00
Belle II (nominal)	0.20	—

7.4 Comparison with Previous Measurements

Relative to the previous world-best combination of Belle, BaBar, and an earlier Belle II measurement (Table 1.1), the precision achieved here improves by approximately 24% for S and 31% for C in MR1, and by approximately 8% for S and 19% for C in MR2. This improvement follows from several factors acting together: the larger combined Belle and Belle II dataset analysed here; the improved GFlaT used across both experiments (Chapter 4), with its higher effective tagging efficiency; an improved K_S^0 selection for the Belle dataset relative to earlier analyses; and, for the Belle dataset, the extension of a time-integrated fit component (Section 5.4) to retain events with poor time resolution rather than discarding them, providing additional statistical constraint on C. The results obtained here supersede the earlier Belle and Belle II measurements of this channel.

7.5 Physics Interpretation

Both the MR1 and MR2 results are statistically consistent with the SM expectation of vanishing CP violation in this channel ($S \approx C \approx 0$, Section 1.4), with neither any individual measurement nor the combined result in either mass region deviating from zero by more than approximately 1.5 standard deviations once statistical and systematic uncertainties are combined. In particular, the measured S in MR1, sensitive to the resonant $K^{*0}(892)$ amplitude, shows no evidence for the enhanced right-handed photon amplitude that would be expected from various new-physics scenarios (Section 1.5). Within the precision of this measurement, the results place a constraint on the size of any such right-handed-current contribution, complementary to those placed by branching fraction measurements from the resonant $K^{*0}(892)\gamma$ mode [71, 72, 73]. The larger uncertainties in MR2, driven primarily by the lack of statistics, currently limit its sensitivity to a comparable constraint, motivating the continued pursuit of larger datasets and improved background suppression.

7.5.1 Constraint on the Wilson Coefficient C_7'

Contributions from right-handed photon discussed above can be expressed in a model-independent way through the Wilson coefficient [74] C_7' of the electromagnetic dipole operator O_7' in the effective $b \to s\gamma$ Hamiltonian. This coefficient vanishes in the SM up to corrections of order m_s/m_b. A nonzero value would signal new-physics contributions of the kind mentioned above.

To place the present measurement in this broader context, a global fit to C_7' was performed following the method of Paul and Straub [75], using the `flavio` package [76] together with `wilson` [77] and the WCxf exchange format [78]. The fit combines several observables: the branching fractions of $B^0 \to K^{*0}\gamma$, $B^+ \to K^{*+}\gamma$, $B \to X_s\gamma$, and $B_s^0 \to \phi\gamma$; the mixing-induced asymmetry $A^{\Delta}(B_s^0 \to \phi\gamma)$ [79]; the angular observables $\langle P_1 \rangle$ and $\langle A_T^{\mathrm{Im}} \rangle$ measured by LHCb in $B^0 \to K^{*0}e^+e^-$ decays [80]; and S in the resonant $K^{*0}(892)\gamma$ mode. For the latter, a new world average of 0.06 ± 0.14 is formed by combining the BaBar measurement [19] with the Belle and Belle II result of this thesis [36]. Full details of the fit inputs and method are given in Appendix C.

Figure 7.4 shows the resulting global constraint in the $(\mathrm{Re}\, C_7', \mathrm{Im}\, C_7')$ plane, together with the individual-observable bands. The global fit aligns with the SM point $C_7' = 0$ at a confidence level of 1.0 standard deviation, yielding a χ^2_{min} value of about 6.4 for seven degrees of freedom. The dominant pull away from the SM originates from $A^{\Delta}(B_s^0 \to \phi\gamma)$, whose measured value lies above the prediction. The large uncertainty on this observable keeps the deviation well within a statistical fluctuation.

Removing $S(K^{*0}(892)\gamma)$ from the global fit increases the 1.0 standard-deviation contour in the $(\mathrm{Re}\, C_7', \mathrm{Im}\, C_7')$ plane by approximately 23%, demonstrating this observable's constraining power within the fit. Conversely, a future measurement with Belle II's full projected dataset of $50\,\mathrm{ab}^{-1}$ would shrink the same contour by about 70%, illustrating the considerable further potential of this observable.

Together with the planned extension to $B^0 \to K_{\mathrm{S}}^0\eta\gamma$ described in Chapter 8, this programme continues to tighten constraints on right-handed-current contributions to $b \to s\gamma$ transitions, complementing the broader Belle II physics effort to search for new sources of CP violation beyond the SM.

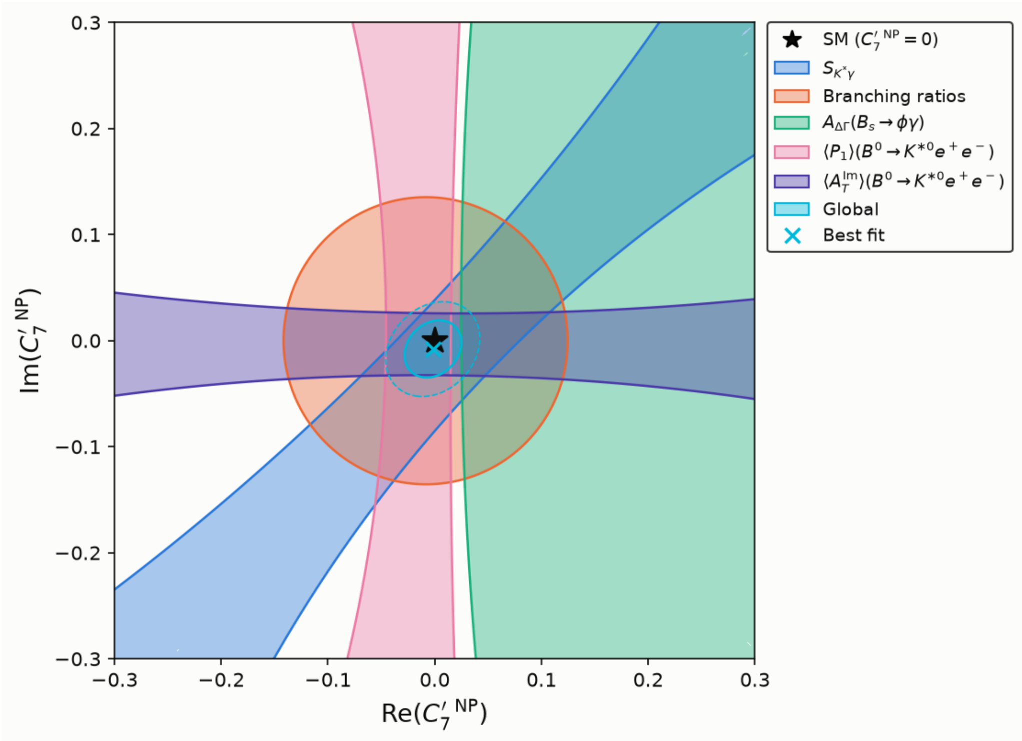


Figure 7.4: Global fit to the Wilson coefficient C_7' in the $(\mathrm{Re}\, C_7', \mathrm{Im}\, C_7')$ plane. The coloured bands show the 1σ constraint from each individual observable, and the cyan-coloured region shows the combined 1σ (solid) and 2σ (dashed) global fit. The S input includes the Belle and Belle II result of this thesis along with the BaBar measurement.

Chapter 8

Towards Time-Dependent *CP* Violation Study in $B^0 \to K_S^0 \eta\gamma$

8.1 Motivation and Complementarity with $K_S^0\pi^0\gamma$

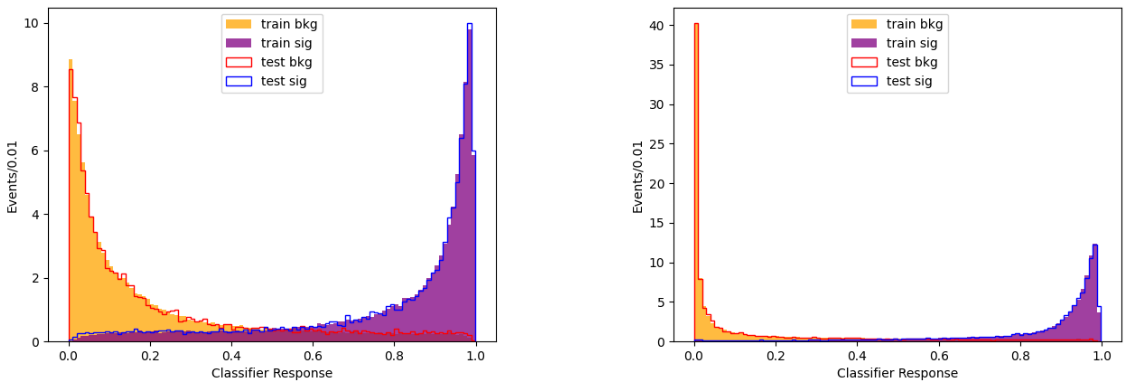
The $B^0 \to K_S^0\eta\gamma$ decay proceeds through the same underlying $b \to s\gamma$ transition as $B^0 \to K_S^0\pi^0\gamma$ (Section 1.4), and is subject to the same theoretical argument that near-maximal photon polarization in the SM suppresses mixing-induced *CP* violation. This makes S in this channel equally sensitive to right-handed-current contributions from physics beyond the SM (Section 1.5). A time-dependent *CP*-violation measurement in $K_S^0\eta\gamma$ is therefore a natural and complementary extension of the $K_S^0\pi^0\gamma$ analysis presented in Chapters 5–7. It probes the same underlying new-physics sensitivity in an independent final state, and shares the same reconstruction technique, background topology, and analysis machinery (Flavour Tagger calibration, resolution function, and fit strategy).

8.2 Reconstruction of the η Meson

The η meson is reconstructed in its $\eta \to \gamma\gamma$ and $\eta \to \pi^+\pi^-\pi^0$ decay channels, with branching fractions of $(39.36 \pm 0.18)\%$ and $(23.02 \pm 0.25)\%$ respectively [2]. The direct contribution of this thesis is to the $\eta \to \pi^+\pi^-\pi^0$ channel described in this chapter, with a complementary analysis of the $\eta \to \gamma\gamma$ channel being carried out separately by another collaborator. While the $\eta \to \gamma\gamma$ channel has a higher branching fraction, the $\eta \to \pi^+\pi^-\pi^0$ channel has a better Δt resolution. Overall, the latter channel has similar sensitivity to S as the former, and the combination of both channels is expected to provide a more precise measurement than either channel alone. In the $\eta \to \pi^+\pi^-\pi^0$ channel, the final state comprises two pairs of oppositely charged pions, one pair coming from $K_S^0 \to \pi^+\pi^-$ and the other from the η, together with $\pi^0 \to \gamma\gamma$ from the η decay and the prompt signal photon. This allows much of the existing $\pi^0 \to \gamma\gamma$ reconstruction and background-suppression machinery from $K_S^0\pi^0\gamma$ (Chapter 5) to be reused directly.

8.3 Background Suppression

Three dedicated multivariate classifiers, based on gradient-boosted decision trees using XGBoost [60], are trained to reject the dominant backgrounds on Belle data: a π^0-selection MVA to suppress combinatorial and peaking photon backgrounds under the π^0 mass peak from the $\eta \to \pi^+\pi^-\pi^0$ decay (ROC-AUC of 0.93); a dedicated η-selection MVA to suppress combinatorial background under the $\eta \to \pi^+\pi^-\pi^0$ mass peak (ROC-AUC of 0.95); and a continuum-suppression MVA exploiting event-shape variables to reject $e^+e^- \to q\bar{q}$ background (ROC-AUC of 0.92), following the same methodology as Section 5.6. Figure 8.1 shows the resulting classifier output for the π^0-selection and η-selection MVAs. The η-selection MVA output is also used for best-candidate selection, retaining the η candidate with the highest classifier score in events with multiple candidates. Applying the full selection, including these three classifiers, yields a signal efficiency roughly double that achieved in the previous Belle-only analysis of this channel [81], substantially improving the expected sensitivity of the eventual S and C measurement.



(a) π^0-selection MVA (b) η-selection MVA

Figure 8.1: Classifier output for the π^0-selection and η-selection MVAs, on Belle training and test samples, for signal and background.

8.4 Current Status and Outlook

Currently, the Belle-side reconstruction, background suppression and best-candidate selection are finalized, and work is progressing on the signal-extraction fit for the Belle sample, following the same three variables (M_{bc}, ΔE, and $\mathcal{C}'$) strategy of Section 6.1. Still to be carried out are: the corresponding Belle II selection and MVA retraining, following the same strategy established for $K_S^0\pi^0\gamma$ in Chapter 5; the Belle II signal-extraction fit; and the time-dependent S and C fit combining both experiments, together with the validation studies of Sections 6.5–6.6 and the systematic-uncertainty evaluation of Section 7.1, analogous to those performed for $K_S^0\pi^0\gamma$.

The ongoing introduction of a new, transformer-based Flavour Tagger, superseding the GFlaT used in Chapters 5–7, is expected to further improve the sensitivity of this measurement to S and C, and its adoption, once validated and calibrated, is expected to take place alongside this work. From simulation studies, this work is projected to improve the expected uncertainty on S for this channel by about 50%.

Chapter 9

Summary and Outlook

This thesis has presented a set of connected contributions to the Belle (II) time-dependent *CP*-violation programme, spanning flavour-tagging calibration, software development, and a measurement of *CP* violation in a radiative *B*-meson decay, together with its interpretation and a complementary ongoing analysis.

Two earlier pieces of work supported this programme (Chapter 4). The mistag and resolution-function parameters of the then-current, BDT-based Flavour Tagger were calibrated using hadronic *B*-decay control channels, providing the tagging and time-resolution inputs needed by any subsequent time-dependent measurement. The Flavour Tagger software itself was then refactored, carried out as the software liaison for the Belle II TDCPV working group, in preparation for the transition to the GFlaT used in the main analysis.

The principal result of this thesis is the measurement of the *CP*-violation parameters S and C in $B^0 \to K^0_S \pi^0 \gamma$ decays (Chapters 5–7), using the legacy Belle (711 fb^{-1}) and Belle II (493 fb^{-1}) datasets. In the $K^{*0}(892)$-dominated region (MR1), the combined result is $S = 0.09 \pm 0.16 \pm 0.02$ and $C = -0.09 \pm 0.08 \pm 0.04$. In the remaining phase space (MR2), which includes higher kaonic resonances, the combined result is $S = -0.32 \pm 0.33 \pm 0.09$ and $C = -0.07 \pm 0.17 \pm 0.08$. Both mass regions are consistent with the SM expectation of vanishing *CP* violation, and this measurement improves the precision on the previous world-best combination by approximately 24% for S and 31% for C in MR1, and by approximately 8% for S and 19% for C in MR2. Interpreted in terms of the Wilson coefficient C'_7 (Section 7.5.1), this result contributes to a global fit that remains consistent with the SM at the 1σ level, and including it in the world average for $S(K^{*0}(892)\gamma)$ tightens the resulting constraint on right-handed currents in $b \to s\gamma$ transitions.

An ongoing extension of this programme to the complementary decay $B^0 \to K^0_S \eta\gamma$ (Chapter 8) proceeds through the same underlying $b \to s\gamma$ transition. The reconstruction, background suppression, and best-candidate selection for the $\eta \to \pi^+\pi^-\pi^0$ channel on Belle data are complete, with the corresponding Belle II selection, the signal-extraction and time-dependent fits, and the associated validation and systematic studies being work in progress. Once completed, this measurement, together with a complementary analysis of the $\eta \to \gamma\gamma$ channel, will provide an independent constraint on the same right-handed-current sensitivity probed by $B^0 \to K_S^0 \pi^0 \gamma$.

Taken together, these results form a key component of the broader Belle II physics programme to search for new sources of *CP* violation beyond the SM. As the Belle II dataset continues to grow and flavour-tagging performance continues to improve, most immediately through the transformer-based tagger mentioned above, the precision achievable on S and C in radiative $b \to s\gamma$ decays will improve correspondingly, sharpening the sensitivity of this programme to right-handed currents and other new-physics contributions to the underlying transition.

Appendix A

Background-Suppression Classifier Details

This appendix supplements Sections 5.1 and 5.6 with the training configuration and leading input variables of the XGBoost-based classifiers trained specifically for this analysis: the π^0/η-veto MVA, the π^0-selection MVA, and the continuum-suppression MVA. The K_S^0-selection and Λ^0-veto classifiers, by contrast, are centrally provided, standard Belle/Belle II selections applied with collaboration-recommended working points rather than trained specifically for this analysis, and are not detailed further here. For each classifier, the input-variable ranking is obtained from its *feature-importance* score. This is the number of times a variable is used to split a node, summed over all trees in the XGBoost ensemble and weighted by the resulting improvement in the training objective, with a higher score indicating a more frequently and more usefully split-upon variable.

A.1 π^0/η-Veto MVA

Table A.1 lists the XGBoost hyperparameters used to train the Belle π^0/η-veto MVA of Section 5.1, and Table A.2 its leading input variables by feature-importance score; the most important variable, χ^2 probability of the pairing candidate's mass-constrained fit, dominates by more than an order of magnitude over the next-ranked variable.

Table A.1: XGBoost hyperparameters for the Belle π^0/η-veto MVA.

Hyperparameter	Value
Objective function	`binary:logistic`
Maximum tree depth	4
Number of estimators	90
Subsample fraction	0.5

Table A.2: Leading input variables of the Belle π^0/η-veto MVA, ranked by feature-importance score.

Feature	Importance score
χ^2 probability of mass-constrained fit	1832.9
Candidate invariant mass	167.6
Cosine of helicity angle	131.1
Energy of second photon	63.6
Highest cluster energy, second photon	54.0

A.2 π^0-Selection MVA

Table A.3 gives the hyperparameters, and Tables A.4 and A.5 the leading input variables, for the π^0-selection MVA of Section 5.1, trained separately for Belle and Belle II.

Table A.3: XGBoost hyperparameters for the π^0-selection MVA.

Hyperparameter	Belle	Belle II
Objective function	`binary:logistic`	`binary:logistic`
Maximum tree depth	3	3
Number of estimators	optimized	optimized

Table A.4: Leading input variables of the Belle π^0-selection MVA, ranked by feature-importance score.

Feature	Importance score
χ^2 probability of mass-constrained fit	2643.3
Opening angle between daughter photons	858.1
Fake-photon suppression, second photon	158.4
Fake-photon suppression, first photon	138.9
Cosine of helicity angle	133.9
Energy of second photon	74.8
Candidate invariant mass	56.3
Energy of first photon	51.4
Centre-of-mass-frame transverse momentum	23.8

A.3 Continuum-Suppression MVA

Table A.6 gives the XGBoost hyperparameters used to train the continuum-suppression MVA of Section 5.6, common to both mass regions and both experiments except for the number of estimators, which is optimized independently for each of the four combinations. Tables A.7, A.8, A.9, and A.10 list the leading input variables in each case; in every combination the single most powerful variable is $\cos\theta_{\mathrm{TBTO}}$, the cosine of the angle between the B_{sig} thrust axis and that of the rest of the event introduced in Section 5.6, consistent with it being quoted there as the most discriminating single feature.

Table A.5: Leading input variables of the Belle II π^0-selection MVA, ranked by feature-importance score.

Feature	Importance score
Opening angle between daughter photons	1321.2
χ^2 probability of mass-constrained fit	503.3
Fake-photon suppression, second photon	257.1
Fake-photon suppression, first photon	207.5
Candidate invariant mass	150.3
Beam-background suppression, first photon	70.6
Cosine of helicity angle	50.2
Beam-background suppression, second photon	34.4

Table A.6: XGBoost hyperparameters for the continuum-suppression MVA (common to all mass regions and experiments).

Hyperparameter	Value
Objective function	`binary:logistic`
Maximum tree depth	3
Number of estimators	optimized
Learning rate	0.2

In every case, the trained classifier was checked for overtraining using the two-sample Kolmogorov–Smirnov test described in Section 5.6, and all classifiers listed in this appendix pass this check.

Table A.7: Leading input variables of the Belle MR1 continuum-suppression MVA, ranked by feature-importance score (top ten of twenty trained input variables shown).

Feature	Importance score
$\cos\theta_{\mathrm{TBTO}}$	1899.4
CLEO cone energy (cone 3, ROE-restricted)	287.2
CLEO cone energy (cone 2, ROE-restricted)	281.9
CLEO cone energy (cone 1, ROE-restricted)	209.3
CLEO cone thrust (cone 5)	132.2
CLEO cone energy (cone 4, ROE-restricted)	115.8
CLEO cone energy (cone 4)	114.9
CLEO cone thrust (cone 6)	111.8
Fox–Wolfram moment R_1	105.3
KSFW moment E_T	68.6

Table A.8: Leading input variables of the Belle MR2 continuum-suppression MVA, ranked by feature-importance score (top ten of twenty trained input variables shown).

Feature	Importance score
$\cos\theta_{\mathrm{TBTO}}$	1588.2
CLEO cone energy (cone 3, ROE-restricted)	229.8
CLEO cone energy (cone 1, ROE-restricted)	226.3
CLEO cone energy (cone 2, ROE-restricted)	213.2
$\cos\theta_{\mathrm{Bz}}$	142.7
Fox–Wolfram moment R_3	122.5
CLEO cone thrust (cone 5)	100.5
CLEO cone energy (cone 2)	98.4
CLEO cone energy (cone 4, ROE-restricted)	91.5
CLEO cone thrust (cone 6)	86.1

Table A.9: Leading input variables of the Belle II MR1 continuum-suppression MVA, ranked by feature-importance score (top ten of twenty-seven trained input variables shown).

Feature	Importance score
$\cos\theta_{\mathrm{TBTO}}$	1394.5
KSFW moment h^2_{oo}	381.1
Harmonic moment of thrust (order 2)	311.6
KSFW moment h^{12}_{so}	275.5
KSFW moment h^{02}_{so}	232.8
Thrust-axis polar-angle moment	146.0
$\cos\theta_{\mathrm{Bz}}$	141.4
CLEO cone energy (cone 1, ROE-restricted)	98.5
CLEO cone energy (cone 2, ROE-restricted)	81.8
Cosine of thrust-axis polar angle	57.9

Table A.10: Leading input variables of the Belle II MR2 continuum-suppression MVA, ranked by feature-importance score (top ten of twenty-six trained input variables shown).

Feature	Importance score
$\cos\theta_{\mathrm{TBTO}}$	1954.0
Harmonic moment of thrust (order 2)	795.2
KSFW moment h_{so}^{12}	436.7
Fox–Wolfram moment R_2	403.0
$\cos\theta_{\mathrm{Bz}}$	261.0
KSFW moment h_{so}^{02}	257.5
KSFW moment h_{oo}^{2}	235.5
Harmonic moment of thrust (order 4)	221.3
CLEO cone energy (cone 2, ROE-restricted)	178.8
CLEO cone energy (cone 1)	152.7

Appendix B

$B\overline{B}$ Background Composition Tables

This appendix supplements Section 7.1.3 with the full background composition of $B\overline{B}$ events in simulation, and the C/S asymmetry assigned to each constituent decay channel, for Belle MR1 (Table B.1), Belle MR2 (Table B.2), Belle II MR1 (Table B.3), and Belle II MR2 (Table B.4).

Table B.1: Background composition of $B\overline{B}$ events in Belle MC, MR1, and the C/S asymmetry assigned to each decay channel (charge conjugation implied).

Decay channel	Frac. (%)	C	S
$B^0 \to X_{\rm sd}\gamma$	14.8	-0.01 ± 0.02	0.0 ± 0.42
$B^0 \to K_1^0\gamma$	5.4	0.0 ± 0.3	0.0 ± 0.37
$B^0 \to K^{*0}\eta$	3.2	0.19 ± 0.05	0.0 ± 0.39
$B^0 \to K^0\eta'$	0.3	-0.08 ± 0.04	0.64 ± 0.05
$B^0 \to K_2^{*0}\gamma$	1.0	-0.08 ± 0.15	0.0 ± 0.37
$B^0 \to X_{\rm sd}\eta$	0.9	0.0 ± 0.3	0.0 ± 0.42
$B^0 \to K^0\pi^0$	0.7	0.00 ± 0.08	0.64 ± 0.13
$B^0 \to K_{\rm S}^0\eta_{\rm c}$	0.1	0.08 ± 0.13	0.93 ± 0.17
$B^0 \to D$, etc.	2.6	0	0
$B^0 \to$ others	2.5	0.0 ± 0.3	0.0 ± 0.5
$B^+ \to X_{\rm su}^+\gamma$	28.6	0.03 ± 0.02	0
$B^+ \to K_1^+\gamma$	21.0	0.0 ± 0.3	0
$B^+ \to K^{*+}\gamma$	9.2	0.01 ± 0.02	0
$B^+ \to K_2^{*+}\gamma$	3.8	0.0 ± 0.3	0
$B^+ \to K^{*+}\eta$	0.7	0.02 ± 0.06	0
$B^+ \to D$, etc.	2.7	0	0
$B^+ \to$ others	1.9	0.0 ± 0.3	0
Total	99.4	0.01 ± 0.12	0.01 ± 0.12

Table B.2: Background composition of $B\overline{B}$ events in Belle MC, MR2, and the C/S asymmetry assigned to each decay channel (charge conjugation implied).

Decay channel	Frac. (%)	C	S
$B^0 \to X_{\rm sd}\gamma$	12.8	-0.01 ± 0.02	0.0 ± 0.37
$B^0 \to K_1^0\gamma$	2.8	0.0 ± 0.3	0.0 ± 0.23
$B^0 \to K^{*0}\eta$	0.7	0.19 ± 0.05	0.0 ± 0.29
$B^0 \to K^0\eta'$	1.2	-0.08 ± 0.04	0.64 ± 0.05
$B^0 \to K_2^{*0}\gamma$	0.8	-0.08 ± 0.15	0.0 ± 0.41
$B^0 \to X_{\rm sd}\eta$	1.3	0.0 ± 0.3	0.0 ± 0.37
$B^0 \to K^0\pi^0$	1.2	0.00 ± 0.08	0.64 ± 0.13
$B^0 \to K_0^0\eta$	1.1	0.0 ± 0.3	0.0 ± 0.40
$B^0 \to K_S^0\eta_c$	0.8	0.08 ± 0.13	0.93 ± 0.17
$B^0 \to D$, etc.	9.3	0	0
$B^0 \to$ others	6.9	0.0 ± 0.3	0.0 ± 0.5
$B^+ \to X_{\rm su}^+\gamma$	23.8	0.03 ± 0.02	0
$B^+ \to K_1^+\gamma$	4.1	0.0 ± 0.3	0
$B^+ \to K^{*+}\gamma$	10.9	0.01 ± 0.02	0
$B^+ \to K_2^{*+}\gamma$	3.6	0.0 ± 0.3	0
$B^+ \to K^{*+}\eta$	1.2	0.02 ± 0.06	0
$B^+ \to D$, etc.	9.3	0	0
$B^+ \to$ others	6.4	0.0 ± 0.3	0
Total	98.2	0.01 ± 0.09	0.02 ± 0.11

Table B.3: Background composition of $B\overline{B}$ events in Belle II MC, MR1, and the C/S asymmetry assigned to each decay channel (charge conjugation implied).

Decay channel	Frac. (%)	C	S
$B^0 \to X_{\rm sd}\gamma$	11.7	-0.01 ± 0.02	0.0 ± 0.41
$B^0 \to K_1^0\gamma$	7.5	0.0 ± 0.3	0.0 ± 0.46
$B^0 \to K^{*0}\eta$	3.6	0.19 ± 0.05	0.0 ± 0.40
$B^0 \to K^{*0}\gamma$	4.0	-0.01 ± 0.01	-0.08 ± 0.40
$B^0 \to K_2^{*0}\gamma$	1.1	-0.08 ± 0.15	0.0 ± 0.36
$B^0 \to K_1'^0\gamma$	1.0	0.0 ± 0.3	0.0 ± 0.46
$B^0 \to D$, etc.	1.8	0	0
$B^0 \to$ others	2.5	0.0 ± 0.3	0.0 ± 0.5
$B^+ \to X_{\rm su}^+\gamma$	19.6	0.03 ± 0.02	0
$B^+ \to K_1^+\gamma$	13.2	0.0 ± 0.3	0
$B^+ \to K^{*+}\gamma$	8.4	0.01 ± 0.02	0
$B^+ \to K^{*+}(1680)\gamma$	7.4	0.0 ± 0.3	0
$B^+ \to K^{*+}(1410)\gamma$	7.1	0.0 ± 0.3	0
$B^+ \to K_2^{*+}\gamma$	2.9	0.0 ± 0.3	0
$B^+ \to K_1'^+\gamma$	2.5	0.0 ± 0.3	0
$B^+ \to D$, etc.	2.3	0	0
$B^+ \to$ others	2.2	0.0 ± 0.3	0
Total	98.8	0.01 ± 0.17	0.0 ± 0.13

Table B.4: Background composition of $B\overline{B}$ events in Belle II MC, MR2, and the C/S asymmetry assigned to each decay channel (charge conjugation implied).

Decay channel	Frac. (%)	C	S
$B^0 \to X_{\mathrm{sd}}\gamma$	13.8	-0.01 ± 0.02	0.0 ± 0.42
$B^0 \to K_1^0\gamma$	2.4	0.0 ± 0.3	0.0 ± 0.44
$B^0 \to K_0^0\eta$	1.3	0.0 ± 0.3	0.0 ± 0.5
$B^0 \to D$, etc.	6.0	0	0
$B^0 \to$ others	8.9	0.0 ± 0.3	0.0 ± 0.5
$B^+ \to X_{\mathrm{su}}^+\gamma$	19.8	0.03 ± 0.02	0
$B^+ \to K_1^+\gamma$	4.8	0.0 ± 0.3	0
$B^+ \to K^{*+}\gamma$	10.1	0.01 ± 0.02	0
$B^+ \to K^{*+}(1680)\gamma$	13.0	0.0 ± 0.3	0
$B^+ \to K^{*+}(1410)\gamma$	3.6	0.0 ± 0.3	0
$B^+ \to K_2^{*+}\gamma$	3.3	0.0 ± 0.3	0
$B^+ \to K_1^{\prime +}\gamma$	0.8	0.0 ± 0.3	0
$B^+ \to D$, etc.	7.5	0	0
$B^+ \to$ others	4.8	0.0 ± 0.3	0
Total	100.0	0.01 ± 0.14	0.07 ± 0.12

Appendix C

Global Fit to the Wilson Coefficient C_7'

This appendix supplements Section 7.5.1 with the full input list, fit method, and best-fit results of the global C_7' fit.

C.1 Fit Method

The fit follows the approach of Paul and Straub [75], extending earlier work on the C_7–C_7' plane [82]. The Wilson coefficient C_7 of the operator O_7 is fixed to its SM value, since it is already tightly constrained by the inclusive branching ratio $\mathrm{BR}(B \to X_s\gamma)$, and only C_7' is left free. The fit is performed at the scale $\mu = 4.8$ GeV using the `FastLikelihood` class of `flavio` [76], with the Wilson coefficient specified in the WCxf exchange format [78] and evolved using `wilson` [77]. For the $B \to K^*$ form factors, the fit uses the default `flavio` choice, the combined light-cone-sum-rule and lattice-QCD fit of Bharucha, Straub, and Zwicky [83].

The best-fit point is obtained by minimizing $\chi^2 = -2\ln\mathcal{L}$ from the SM starting point $C_7' = 0$. Confidence regions are obtained from a 60×60 grid scan over $\mathrm{Re}\,C_7'$, $\mathrm{Im}\,C_7' \in [-0.33, 0.33]$, with the resulting χ^2 surface smoothed by a cubic spline. The 1σ and 2σ contours correspond to $\Delta\chi^2 = 2.30$ and 6.18 respectively, appropriate for a two-parameter fit.

C.2 Fit Inputs

Table C.1 lists the observables used in the fit. All branching ratios and the PDG-quoted asymmetry are taken from the 2025 edition of the Review of Particle Physics [2]. The two angular observables are taken from the LHCb measurement of $B^0 \to K^{*0}e^+e^-$ decays in the low-q^2 region [80], and the mixing-induced asymmetry $A^\Delta(B_s^0 \to \phi\gamma)$ from the LHCb measurement of this observable [79], together with its SM prediction.

The world average for $S(B^0 \to K^{*0}(892)\gamma)$ combines the BaBar measurement, $S =$

$-0.03 \pm 0.29 \pm 0.03$ [19], with the Belle and Belle II result of this thesis, $S = 0.09 \pm 0.16 \pm 0.02$ [36], by inverse-variance weighting of the total uncertainties. This gives 0.06 ± 0.14. This average is quoted for the purpose of this fit only.

Table C.1: Observables used as inputs to the global C_7' fit. The angular observables $\langle P_1 \rangle$ and $\langle A_T^{\mathrm{Im}} \rangle$ are averaged over $q^2 \in [0.0008, 0.257]$ GeV^2.

Observable	Value	Source
$S(B^0 \to K^{*0}(892)\gamma)$	0.06 ± 0.14	BaBar [19] + this thesis [36]
$\mathrm{BR}(B^0 \to K^{*0}\gamma)$	$(4.18 \pm 0.25) \times 10^{-5}$	PDG 2025 [2]
$\mathrm{BR}(B^+ \to K^{*+}\gamma)$	$(3.92 \pm 0.22) \times 10^{-5}$	PDG 2025 [2]
$\mathrm{BR}(B \to X_s\gamma)$, $E_\gamma > 1.6$ GeV	$(3.49 \pm 0.19) \times 10^{-4}$	PDG 2025 [2]
$\mathrm{BR}(B_s^0 \to \phi\gamma)$	$(3.4 \pm 0.4) \times 10^{-5}$	PDG 2025 [2]
$A^\Delta(B_s^0 \to \phi\gamma)$	-0.7 ± 0.4	LHCb [79]
$\langle P_1 \rangle(B^0 \to K^{*0}e^+e^-)$	0.11 ± 0.10	LHCb [80]
$\langle A_T^{\mathrm{Im}} \rangle(B^0 \to K^{*0}e^+e^-)$	0.02 ± 0.10	LHCb [80]

C.3 Best-Fit Results

Table C.2 gives the best-fit point and minimum χ^2 for the global fit, together with single-observable fits used to produce the individual bands of Figure 7.4. A single two-parameter fit to one observable is always underconstrained along one direction, so the reported $\chi^2_{\mathrm{min}} = 0$ for these rows reflects this and only the shape and orientation of the resulting confidence band is meaningful, not the best-fit point itself.

Table C.2: Best-fit values of C_7' and minimum χ^2 for the global fit and for single-observable fits used to produce the individual bands in Figure 7.4. The global fit has seven degrees of freedom.

Fit	$\mathrm{Re}\, C_7'$	$\mathrm{Im}\, C_7'$	χ^2_{min}
Global fit	-0.0019	-0.0075	6.39
$S(K^{*0}\gamma)$ only	0.0143	-0.0080	0
Branching ratios only	-0.0459	-0.0004	2.20
$A^\Delta(B_s^0 \to \phi\gamma)$ only	0.9411	0.0066	0
$\langle P_1 \rangle(K^{*0}e^+e^-)$ only	-0.0148	0.0002	0
$\langle A_T^{\mathrm{Im}} \rangle(K^{*0}e^+e^-)$ only	0.0002	-0.0040	0

The global fit is consistent with the SM at the 1σ level. The largest individual pull comes from $A^\Delta(B_s^0 \to \phi\gamma)$, whose PDG average of -0.7 ± 0.4 lies visibly above the SM prediction of $0.047^{+0.029}_{-0.025}$. Given the size of the experimental uncertainty, this tension does not by itself constitute significant evidence for a non-zero C_7'.

Bibliography

[1] A. D. Sakharov, “Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe,” JETP Lett. **5**, 24 (1967).

[2] Particle Data Group, S. Navas *et al.*, “Review of Particle Physics,” Phys. Rev. D **110**, 030001 (2024), and 2025 update.

[3] L. Canetti, M. Drewes, and M. Shaposhnikov, “Matter and Antimatter in the Universe,” New J. Phys. **14**, 095012 (2012).

[4] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, “Evidence for the 2π Decay of the K_2^0 Meson,” Phys. Rev. Lett. **13**, 138 (1964).

[5] N. Cabibbo, “Unitary Symmetry and Leptonic Decays,” Phys. Rev. Lett. **10**, 531 (1963).

[6] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction,” Prog. Theor. Phys. **49**, 652 (1973).

[7] L. Wolfenstein, “Parametrization of the Kobayashi-Maskawa Matrix,” Phys. Rev. Lett. **51**, 1945 (1983).

[8] J. Charles *et al.* (CKMfitter Group), “CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories,” Eur. Phys. J. C **41**, 1 (2005).

[9] D. Atwood, M. Gronau, and A. Soni, “Mixing-induced CP asymmetries in radiative B decays in and beyond the standard model,” Phys. Rev. Lett. **79**, 185 (1997).

[10] J. L. Hewett, “Probing new physics in the B system,” AIP Conf. Proc. **424**, 328 (1998).

[11] P. Ball, G. W. Jones, and R. Zwicky, “$B \to V\gamma$ beyond QCD factorization,” Phys. Rev. D **75**, 054004 (2007).

[12] B. Grinstein, Y. Grossman, Z. Ligeti, and D. Pirjol, “Photon polarization in $B \to X\gamma$ in the Standard Model,” Phys. Rev. D **71**, 011504 (2005).

[13] B. Grinstein and D. Pirjol, "CP asymmetry in $B^0(t) \to K_S^0\pi^0\gamma$ in the Standard Model," Phys. Rev. D **73**, 014013 (2006).

[14] A. L. Kagan and M. Neubert, "Direct CP violation in $B \to X_s\gamma$ decays as a signature of new physics," Phys. Rev. D **58**, 094012 (1998).

[15] F.-S. Yu, E. Kou, and C.-D. Lü, "Photon polarization in the $b \to s\gamma$ processes in the left-right symmetric model," JHEP **12**, 102 (2013).

[16] H. Eberl, K. Hidaka, E. Ginina, and A. Ishikawa, "Imprint of SUSY in radiative B-meson decays," Phys. Rev. D **104**, 075025 (2021).

[17] M. Jung, X.-Q. Li, and A. Pich, "Exclusive radiative B-meson decays within the aligned two-Higgs-doublet model," JHEP **10**, 063 (2012).

[18] Belle Collaboration (Y. Ushiroda *et al.*), "Time-dependent CP asymmetries in $B^0 \to K_S^0\pi^0\gamma$ transitions," Phys. Rev. D **74**, 111104 (2006).

[19] BaBar Collaboration (B. Aubert *et al.*), "Measurement of Time-Dependent CP Asymmetry in $B^0 \to K_S^0\pi^0\gamma$ Decays," Phys. Rev. D **78**, 071102 (2008).

[20] Belle II Collaboration (I. Adachi *et al.*), "Measurement of CP Asymmetries in $B^0 \to K_S^0\pi^0\gamma$ Decays at Belle II," Phys. Rev. Lett. **134**, 011802 (2025).

[21] S. Kurokawa and E. Kikutani, "Overview of the KEKB accelerators," Nucl. Instrum. Meth. A **499**, 1 (2003).

[22] K. Hosoyama *et al.*, "Superconducting Crab Cavity for KEKB," in Proceedings of the 1st Asian Particle Accelerator Conference (APAC'98), Tsukuba, Japan, paper 6D064 (1998).

[23] "KEKB B-Factory Design Report," KEK Report 95-7, National Laboratory for High Energy Physics (KEK), Tsukuba (1995).

[24] Belle Collaboration (A. Abashian *et al.*), "The Belle detector," Nucl. Instrum. Meth. A**479**, 117 (2002).

[25] J. Brodzicka *et al.* (Belle Collaboration), "Physics Achievements from the Belle Experiment," Prog. Theor. Exp. Phys. **2012**, 04D001 (2012).

[26] R. Abe *et al.*, "BELLE/SVD2 status and performance," Nucl. Instrum. Meth. A **535**, 379 (2004).

[27] Belle II Collaboration, "The SuperKEKB Accelerator," Belle II public web pages, https://www.belle2.org/about_us/superkekb_accelerator/.

[28] K. Akai, K. Furukawa, and H. Koiso (SuperKEKB Design Team), "SuperKEKB Collider," Nucl. Instrum. Meth. A **907**, 188 (2018), arXiv:1809.01958 [physics.acc-ph].

[29] T. Abe *et al.* (Belle II Collaboration), "Belle II Technical Design Report," KEK Report 2010-1, arXiv:1011.0352 [physics.ins-det].

[30] KEK Press Release, "Belle II accumulates the world's largest $\Upsilon(4S)$ dataset," 5 June 2026, https://www.kek.jp/en/press/202606051400.

[31] "SuperKEKB raises the bar," CERN Courier (2026), https://cerncourier.com/a/superkekb-raises-the-bar/.

[32] A. Natochii *et al.*, "Beam background expectations for Belle II at SuperKEKB," arXiv:2203.05731 [physics.ins-det] (2022).

[33] E. Kou and P. Urquijo (eds.) (Belle II Collaboration), "The Belle II Physics Book," Prog. Theor. Exp. Phys. **2019**, 123C01 (2019), arXiv:1808.10567 [hep-ex].

[34] T. Kuhr, C. Pulvermacher, M. Ritter *et al.*, "The Belle II Core Software," Comput. Softw. Big Sci. **3**, 1 (2019).

[35] M. Gelb *et al.*, "B2BII: Data Conversion from Belle to Belle II," Comput. Softw. Big Sci. **2**, 9 (2018), arXiv:1810.00019 [hep-ex].

[36] Belle and Belle II Collaborations, "Measurement of time-dependent CP violation parameters in $B^0 \to K_S^0 \pi^0 \gamma$ decays at Belle and Belle II," arXiv:2606.04496 [hep-ex], https://arxiv.org/abs/2606.04496.

[37] D. J. Lange, "The EvtGen particle decay simulation package," Nucl. Instrum. Meth. A **462**, 152 (2001).

[38] S. Jadach, B. F. L. Ward, and Z. Wąs, "The precision Monte Carlo event generator $\mathcal{KK}$ for two-fermion final states in e^+e^- collisions," Comput. Phys. Commun. **130**, 260 (2000).

[39] T. Sjöstrand, S. Mrenna, and P. Skands, "Pythia 6.4 physics and manual," JHEP **05**, 026 (2006).

[40] T. Sjöstrand, S. Mrenna, and P. Skands, "An introduction to Pythia 8.2," Comput. Phys. Commun. **191**, 159 (2015).

[41] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, and P. Zanarini, "GEANT3," CERN-DD/EE/84-1 (1987).

[42] S. Agostinelli *et al.* (GEANT4 Collaboration), "GEANT4—a simulation toolkit," Nucl. Instrum. Meth. A **506**, 250 (2003).

[43] G. Gaudino (for the Belle II Data Production group), "Run Dependent Monte Carlo at Belle II," PoS **EPS-HEP2025**, 559 (2026), arXiv:2601.19728 [hep-ex].

[44] V. Bertacchi *et al.* (Belle II Tracking Group), "Track finding at Belle II," Comput. Phys. Commun. **259**, 107610 (2021), arXiv:2003.12466 [hep-ex].

[45] T. Bilka *et al.*, "Implementation of GENFIT2 as an experiment-independent track-fitting framework," arXiv:1902.04405 [physics.ins-det] (2019).

[46] F. Wemmer *et al.*, "Photon Reconstruction in the Belle II Calorimeter Using Graph Neural Networks," Comput. Softw. Big Sci. **7**, 13 (2023), arXiv:2306.04179 [hep-ex].

[47] S. Sandilya (Belle II Collaboration), "Particle Identification at Belle II," J. Phys. Conf. Ser. **770**, 012045 (2016), arXiv:1610.00264 [hep-ex].

[48] Belle II Collaboration (I. Adachi *et al.*), "Charged-hadron identification at Belle II," arXiv:2506.04355 [hep-ex] (2025).

[49] K. Hanagaki, H. Kakuno, H. Ikeda, T. Iijima, and T. Tsukamoto, "Electron identification in Belle," Nucl. Instrum. Meth. A **485**, 490 (2002), arXiv:hep-ex/0108044 [hep-ex].

[50] Belle II Analysis Software Group, "Global decay chain vertex fitting at Belle II," Nucl. Instrum. Meth. A**976**, 164269 (2020).

[51] S. Dey and A. Soffer, "Beam-Constrained Vertexing for B Physics at the Belle II Experiment," Springer Proc. Phys. **248**, 411 (2020).

[52] F. Abudinén *et al.* (Belle II Collaboration), "B-flavor tagging at Belle II," Eur. Phys. J. C **82**, 283 (2022), arXiv:2110.00790 [hep-ex].

[53] T. Humair, R. Mehta, and P. Stavroulakis, "Calibration of the Decay-Time Resolution and Flavor Tagger Using Hadronic B Decays Reconstructed in Moriond 2023 Data," Belle II internal note, https://docs.belle2.org/pub_data/documents/495/.

[54] G. C. Fox and S. Wolfram, "Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes," Phys. Rev. Lett. **41**, 1581 (1978).

[55] M. Pivk and F. R. Le Diberder, "sPlot: a statistical tool to unfold data distributions," Nucl. Instrum. Meth. A **555**, 356 (2005).

[56] F. James, "MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1," CERN Program Library Long Writeup D506 (1994).

[57] Belle II Collaboration (I. Adachi *et al.*), "New graph-neural-network flavor tagger for Belle II and measurement of $\sin 2\phi_1$ in $B^0 \to J/\psi K^0_S$ decays," Phys. Rev. D **110**, 012001 (2024).

[58] M. Feindt and U. Kerzel, "The NeuroBayes neural network package," Nucl. Instrum. Meth. A**559**, 190 (2006).

[59] G. Ke, Q. Meng, T. Finley *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Proc. 31st Int. Conf. on Neural Information Processing Systems (NIPS'17)*, 3149 (2017).

[60] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '16)*, 785 (2016).

[61] A. J. Bevan, B. Golob, Th. Mannel, S. Prell, and B. D. Yabsley (eds.), "The Physics of the B Factories," Eur. Phys. J. C **74**, 3026 (2014).

[62] S. Brandt, C. Peyrou, R. Sosnowski, and A. Wroblewski, "The Principal Axis of Jets: An Attempt to Analyse High-Energy Collisions as Two-Body Processes," Phys. Lett. **12**, 57 (1964).

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

[64] J. Gaiser, "Charmonium spectroscopy from radiative decays of the J/ψ and ψ'," Ph.D. thesis, Stanford University, SLAC-R-255 (1982).

[65] T. Skwarnicki, "A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances," Ph.D. thesis, Institute of Nuclear Physics, Krakow, DESY-F31-86-02 (1986).

[66] N. L. Johnson, "Systems of frequency curves generated by methods of translation," Biometrika **36**, 149 (1949).

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

[68] H. Albrecht *et al.* (ARGUS Collaboration), "Search for hadronic $b \to u$ decays," Phys. Lett. B **241**, 278 (1990).

[69] O. Long, M. Baak, R. N. Cahn, and D. Kirkby, "Impact of tag-side interference on time-dependent *CP* asymmetry measurements using coherent $B^0\overline{B}^0$ pairs," Phys. Rev. D **68**, 034010 (2003).

[70] A. Valassi, "Combining correlated measurements of several different physical quantities," Nucl. Instrum. Meth. A **500**, 391 (2003).

[71] Belle Collaboration (M. Nakao *et al.*), "Measurement of the $B \to K^*\gamma$ branching fractions and asymmetries," Phys. Rev. D **69**, 112001 (2004), arXiv:hep-ex/0402042.

[72] BaBar Collaboration (B. Aubert *et al.*), "Measurement of Branching Fractions and CP and Isospin Asymmetries in $B \to K^*(892)\gamma$ Decays," Phys. Rev. Lett. **103**, 211802 (2009), arXiv:0906.2177 [hep-ex].

[73] Belle II Collaboration, "Measurement of $B \to K^*(892)\gamma$ decays at Belle II," JHEP **09**, 024 (2025), arXiv:2411.10127 [hep-ex].

[74] K. G. Wilson, "Non-Lagrangian Models of Current Algebra," Phys. Rev. **179**, 1499 (1969).

[75] A. Paul and D. M. Straub, "Constraints on new physics from radiative B decays," JHEP **04**, 027 (2017), arXiv:1608.02556 [hep-ph].

[76] D. M. Straub, "flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond," arXiv:1810.08132 [hep-ph] (2018).

[77] J. Aebischer, J. Kumar, and D. M. Straub, "Wilson: a Python package for the running and matching of Wilson coefficients above and below the electroweak scale," Eur. Phys. J. C **78**, 1026 (2018), arXiv:1804.05033 [hep-ph].

[78] J. Aebischer *et al.*, "WCxf: an exchange format for Wilson coefficients beyond the Standard Model," Comput. Phys. Commun. **232**, 71 (2018), arXiv:1712.05298 [hep-ph].

[79] LHCb Collaboration (R. Aaij *et al.*), "Measurement of CP-Violating and Mixing-Induced Observables in $B_s^0 \to \phi\gamma$ Decays," Phys. Rev. Lett. **123**, 081802 (2019), arXiv:1905.06284 [hep-ex].

[80] LHCb Collaboration (R. Aaij *et al.*), "Strong constraints on the $b \to s\gamma$ photon polarisation from $B^0 \to K^{*0}e^+e^-$ decays," JHEP **12**, 081 (2020), arXiv:2010.06011 [hep-ex].

[81] Belle Collaboration (H. Nakano *et al.*), "Measurement of time-dependent CP asymmetries in $B^0 \to K_S^0\eta\gamma$ decays," Phys. Rev. D **97**, 092003 (2018).

[82] M. Gronau, Y. Grossman, D. Pirjol, and A. Ryd, "Exploring New Physics in the C_7–C_7' plane," Phys. Rev. Lett. **88**, 051802 (2002), arXiv:hep-ph/0107254.

[83] A. Bharucha, D. M. Straub, and R. Zwicky, "$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules," JHEP **08**, 098 (2016), arXiv:1503.05534 [hep-ph].